Design and analysis of experiments embedded in complex sample surveys

Ontwerp en analyse van experimenten binnen steekproefonderzoeken met een complex ontwerp

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam op gezag van de Rector Magnificus Prof.dr.ir. J.H. van Bemmel en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op 20 december 2001 om 13.30 uur

door

Jan Arie van den Brakel

geboren te Baarn

Promotiecommissie

Promotoren:	Prof.dr. S.E. Fienberg
	Prof.dr.ir. P. Kooiman
Overige leden:	Prof.dr. J.D.F. Habbema
	Prof.dr. T.A.B. Snijders
	Prof.dr. M.J.C.M. Verbeek
Copromotor:	Dr. R.H. Renssen

Contents

The separation of sample surveys from experimental designs is an accident in the recent history of science, and it will diminish.

Kish, 1965, pp. 595.

Chapter 1

Introduction

Statistical methodology of experimental designs provides a powerful instrument for research on the improvement of the sample survey process. Large-scale field experiments embedded in ongoing surveys are particularly appropriate for obtaining quantitative information on the influence of different design parameters of a survey process on response behavior and estimates of population parameters of a sample survey. At Statistics Netherlands, for example, the effects of alternative questionnaire designs, different approach strategies or different advance letters have been investigated this way. Methods from experimental designs can also support quality control of the survey process. For instance, the bias and various variance components in total measurement error models (Forsman, 1989 and Biemer et al., 1991) can be estimated by using techniques from the field of experimental design.

Fienberg and Tanur (1987, 1988, 1989, 1996) reviewed the development of statistical methods in the fields of randomized sampling and randomized experimentation by Fisher and Neyman in the 1920's and 1930's. They discuss the differences and parallels between the theory of experimental designs and the theory of finite population sampling. An important difference concerns the validity of statistical inference. Principles of design and analysis of experiments are mainly intended to guarantee a sufficient internal validity of an experiment, i.e. the cause-effect relationship between treatments and the observed effects within the experiment itself. Design-based methods from sampling theory are mainly intended to guarantee a sufficient external validity of a survey, i.e. the extent to which sample results can be generalized to the finite target population. Despite such differences, however, Fienberg and Tanur emphasize that the statistical methodology employed in both fields is essentially the same. Several statistical methods from the theory of experimental designs and sampling theory can be combined in a useful and natural way in the design and analysis of embedded experiments, thereby increasing the internal as well as the external validity of the inferences drawn from these experiments (Fienberg and Tanur, 1987, 1988 and Van den Brakel and Renssen 1998). This is especially interesting when the effect of alternative survey methodologies on estimates of finite population parameters is investigated. On the one hand, methodology from sampling theory provides more or less design-unbiased estimates for the several parameters of interest and guarantees the external validity of the results. On the other hand, methods from design and analysis of experiments quantify the effect of different survey strategies on these estimates, which guarantees the internal validity of the results. This not only suggests embedding experimental designs in ongoing surveys but also requires a suitable analysis of such designs.

Unfortunately there is no precise statistical theory supporting such an analysis. Generally the purpose of embedded experiments is the estimation of finite population parameters under the different survey implementations or treatments as well as testing hypotheses about the differences between those parameter estimates. Mainstream experimental design theory employs a model-based approach building on the assumption of identically and independently distributed observations (although a few authors such as Kempthorne (1952) and Hinkelman and Kempthorne (1994) give the details of the parallel design-based randomization theory). These assumptions do not reflect the complexities usually exhibited by the actual sampling designs employed in large-scale surveys by statistical agencies or research institutions, cf. Skinner et al (1989). Using standard model-based theory, inferences concern the parameters of some hypothetical superpopulation and not the parameters of the finite target population from which the sample is drawn. In addition results of the analysis may be incommensurable with the sampling results of the ongoing survey, which complicates the interpretation of the analysis. This thesis develops and discusses a design-based approach for the analysis of embedded field experiments to solve these problems.

The typical situation considered in this thesis is a field experiment designed to compare the impact on the estimates of population parameters of $K-1$ alternative survey implementations (or treatments) with respect to a standard implementation of a regular survey. The sample of a regular survey, drawn from a finite target population by means of some complex sampling scheme, is divided randomly into K similarly designed interpenetrating subsamples according to some experimental design. Each subsample is assigned to one of the K treatments. We consider both completely randomized designs (CRD's) and randomized block designs (RBD's). RBD's enable us to apply local control over sampling structures through randomization within sample strata, clusters, primary sampling units, interviewers and the like. The subsamples are not necessarily of equal size. In many practical situations, one of the subsamples is relative large compared to the other subsamples, because it underlies the regular publication purposes of the survey and simultaneously serves as the control group in the experiment. The remaining (smaller) subsamples are assigned to the alternative treatments to be conducted in parallel with the regular survey.

To test hypotheses about differences between finite population parameters, a design-based analysis procedure is explored. Each subsample can be considered as a two-phase sample from the finite target population of the survey where the design of the survey sample and the experimental design determine the design of the first and the second phase, respectively. The

Horvitz-Thompson estimator as well as the generalized regression estimator can be applied to derive (nearly) design-unbiased estimators for the population parameters under each of the K treatments as well as the covariance matrix of the $K-1$ contrasts between these estimates. A design-based Wald statistic, or a design-based $t-$statistic for the embedded two-sample problem as a special case, is obtained to test hypotheses about treatment effects of the finite population parameters that explicitly take into account both the probability structure of the applied sampling design, the weighting scheme applied in the estimation procedure of the regular survey as well as the randomization mechanism of the experimental design. Compared to the Horvitz-Thompson estimator, the generalized regression estimator might decrease the design-variance as well as the bias due to selective non-response and therefore it may increase the efficiency of the experiment as a whole. In the present context the generalized regression estimator therefore represents a design-based analogy of covariance analysis in standard experimental design methodology.

Since the subsamples are drawn without replacement from a finite population according to a general complex sampling design, we expected a rather complicated expression for the covariance matrix with non-zero off-diagonal entries. However, despite the complexity of the sampling design, it appears that this covariance matrix has a relative simple structure as if all observations were drawn independently from each other. For example, in the case of simple random sampling without replacement this result entails that the finite population correction factor should be disregarded in estimating the variance of contrasts. As a result a Wald-statistic or a $t-$statistic, derived from a design-based perspective under general complex sampling designs, is obtained that still has the attractive relative simple structure of standard model-based analysis procedures.

The research presented in this thesis is inspired by the papers of Fienberg and Tanur (1987, 1988, 1989). The lack of design-based analysis procedures for embedded experiments was recognized in several consultancy projects performed by the Statistical Methodology Department of Statistics Netherlands (Van den Brakel and Renssen (1995, 1996a). In a first attempt to fill this gap a design-based procedure for the simplest case, i.e. the analysis of the embedded two-treatment experiment was developed by Van den Brakel and Renssen (1996b). These results were extended to the analysis of embedded CRD's by Van den Brakel and Renssen (1996c, 1997). Based on these first results, a general theory for the analysis of embedded CRD's and RBD's is proposed in this thesis. Results have been applied successfully in several consultancy projects for the Dutch Labour Force Survey, see e.g. Van den Brakel and Van Berkel (2000) and Van den Brakel (2001). Van den Brakel and Binder (2000) explored alternative variance estimation procedures.

The structure of the thesis is as follows. In chapter 2 the design aspects of embedded experiments are discussed as well as the necessity for a design-based approach for the analysis of such experiments. A design-based theory for the analysis of embedded experiments based on the Horvitz-Thompson estimator is developed in chapter 3. This theory is extended to the generalized regression estimator in chapter 4. As a result an analysis procedure for embedded field experiments is obtained to test hypotheses about the effect of alternative survey implementations on the estimates of the population parameters as they are defined in the survey. Chapter 5 and 6 further elaborate on the specific aspects of CRD's as well as the efficiency of blocking on different sampling structures like strata, primary sampling units, clusters or interviewers. In the theory expounded in chapter 3 through 6, it is assumed that the experimental units in the experimental design and the sampling units of the sampling design are of the same level. In chapter 7, the theory is extended to situations where the experimental units do not necessarily coincide with the sampling units. This design-based approach naturally leads to statistical procedures for the design and analysis of embedded experiments, which combine the internal validity guaranteed by methods from randomized experimentation with the external validity obtained from the theory of randomized sampling. In chapter 8 an experiment embedded in the Dutch Labour Force Survey, aimed to test effects of a new questionnaire, is analyzed with the design-based methods proposed in this thesis. The results are compared with the results of a standard model-based analysis. In chapter 9 a second example of an embedded experiment is described. In this experiment we test the effects of the interviewers' workload on the response rates and the outcomes of the Dutch Labour Force Survey.

Chapter 2

Embedding experiments in sample surveys

2.1 Introduction

Traditionally, the design and analysis of experiments and sampling theory form two separate domains of applied statistics. These two fields, however, come together in situations where experiments are conducted to investigate possible improvements of a sample survey process. In this chapter we describe how statistical methods from the theory of experimental designs can support research aimed at improving the survey process. In section 2.2, we briefly review the principles of experimental designs and sampling theory. We distinguish between small-scale experiments conducted in laboratory settings and large-scale field experiments embedded in ongoing surveys. With respect to the large-scale field experiments we are mainly interested in effects on finite population parameters. We discuss how parallels between experimental designs and sampling theory can be used in the design of efficient embedded experiments (section 2.3) and illustrate this with a series of practical examples of embedded experiments conducted at Statistics Netherlands (section 2.4). For the analysis of large-scale field experiments embedded in ongoing surveys, Fienberg and Tanur (1987, 1989) emphasize model-based estimation. As an alternative, we advocate the use of design-based methods in section 2.5. If the interest in embedded experiments focuses on the effects of treatments on estimates of finite population parameters, then we argue that the complexity of the sampling design should be taken into account in the analysis of the experiment. This is the motivation for the design-based methods for the analysis of experiments embedded in complex sampling schemes proposed in the following chapters of this thesis. Substantial parts of this chapter were published in Van den Brakel and Renssen (1998).

2.2 Concepts of experimental designs and sampling theory

2.2.1 Experimental designs

The objective of experimental designs is to obtain relevant quantitative information about the effects of different treatments and their mutual interactions. The principles of experimental designs, such as replication, randomization, local control for sources of variation by skillful grouping of the experimental units (e.g. randomized block designs and row and column designs), the use of factorial designs and covariance analysis were developed mainly on the basis of Fisher's work (1935).

The existence of extraneous variation or experimental errors tends to mask the effects of the treatments in an experiment. A minimum number of replications of the treatments is required to discover statistically significant treatment effects if they exist. Randomization is used to avoid that experimental errors systematically distort the measurement of the treatment effects, by ensuring that each treatment has an equal probability of being favored or handicapped by an extraneous source of variation. Local control by means of randomized block designs, row and column designs, or covariance analysis is applied to eliminate sources of extraneous variation in order to obtain more precise estimates of the treatment effects. Simultaneous testing of different treatments by means of factorial designs is applied to increase the efficiency of experimentation and to test possible interactions between the different treatments.

Statistical methods from the theory of design and analysis of experiments are basically model dependent. Often, the observations in an experiment, taken at different levels of different factors, are assumed to be the outcome of a stochastic variable, which is modeled according to a linear regression model with the different factors (and possibly also covariables) as explanatory variables. The disturbance terms are assumed to be identically and independently distributed (IID) with expectation zero. The factors may concern treatment variables as well as local control variables for sources of extraneous variation (e.g. block variables, row and column variables or covariables). The disturbance terms represent the extraneous variation insofar as it is not explained by local control and/or by the covariables. An important distinction in the model assumptions concerns the question whether the levels at which a factor is measured are fixed or whether they have to be considered as a random sample from a larger population. In the former case the effects are called fixed, whereas in the latter the effects are random. A set of effects is often considered fixed when the statistical inference only concerns the levels included in the experiment, whereas they are considered random when the inference extends to a population from which the levels were supposed to be drawn from.

The main purpose of experimental designs is to make statistical inferences about the treatment effects. Statistical models plays a central role in testing hypotheses about the significance of model parameters, which are assumed to reflect the treatment effects and their interactions, and to explore relationships between variables. Based on the regression model and its assumptions, efficient test statistics are derived in order to test hypotheses about the corresponding regression parameters. When the regression model is correctly specified, the treatment effects are indeed reflected in the corresponding regression parameters and the stated hypotheses about the regression parameters also concern the treatment effects. Note that the principles of experimentation (such as replication, randomization, local control and factorial experimentation) are applied to ensure that observed treatment effects actually can be attributed to the parameters of the statistical model. If the model is misspecified, the validity of the conclusions with respect to the treatment effects depends on the robustness of the test statistics against the type of misspecification involved. In general, conclusions based on efficient test statistics derived under a specific model cannot be generalized to situations outside that model. Kempthorne (1952) and Hinkelmann and Kempthorne (1994) suggested a design-based approach for the analysis of experiments by elaborating on randomization theory in a way similar to the approach in sampling theory.

To summarize, statistical methods of experimental designs are mainly intended to guarantee a sufficient internal validity of the experiment, and to estimate differences in treatment effects as precisely as possible. The internal validity of an experiment is defined as the extent to which the observed effects in an experiment can be attributed to the differences in the treatments. It thus relates to the cause-effect relationship between the treatments and the observed effects within the experiment itself.

2.2.2 Sampling theory

The purpose of sample surveys is to obtain statistical information about a given finite population by estimating finite population parameters such as means, totals and fractions. Until the beginning of the twentieth century such population parameters were obtained by complete enumeration of the finite population. The concept of random sampling has been developed, mainly on the basis of the work of Neyman (1934), as a method to obtain valid estimators for finite population parameters based on representative samples rather than on complete censuses. Neyman (1934) introduced random sampling with unequal selection probabilities by treating optimal allocation in stratified sampling. The concept of random sampling with unequal selection probabilities has been generalized by Hansen and Hurwitz (1943) for random sampling with replacement and by Horvitz and Thompson (1952) for random sampling without replacement as a method to improve the precision of population parameter estimates.

In sampling theory, a minimum sample size is required to estimate population parameters with some precision. Observations obtained from the sampling units are regarded as fixed. The randomness is introduced because a probability sample is observed instead of the entire target population. In random sampling, the concept of random selection is applied in order to draw statistical inferences about finite population parameters and generalize results from the observed sample to the finite target population from which the sample is drawn.

Statistical methods from sampling theory can be considered as design-based or distribution free, that is no assumptions are made on the frequency distribution of the finite population. A proper combination between sampling design and estimation procedure (i.e. the sampling strategy) should give unbiased or nearly unbiased estimates with a minimum variance for the finite population parameters under consideration. An important tool to achieve this is the use of auxiliary information. Such information can be utilized in the sampling design and/or the estimation procedure. In the design stage, techniques like stratification and lattice sampling are applied to increase the precision of the estimators by excluding the variation between homogeneous groups in the finite population from the sampling error. These techniques are similar to randomized block designs and row and column designs from the theory of experimental design. In the estimation procedure, auxiliary information is utilized by means of the regression estimator (with post stratification as a special case) to obtain more precise estimators. This is equivalent to the technique of covariance analysis from the theory of experimental design.

Although auxiliary information was originally used in the design and estimation procedure to decrease the sampling variance of estimators, nowadays it is an important tool to decrease the bias due to selective non-response. Estimators using auxiliary information are generally more robust against selective non-response than estimators that do not use auxiliary information (e.g., see Särndal and Swenson, 1987 and Bethlehem, 1988).

Statistical models have traditionally played a minor role in the analysis of sample surveys. In the model-assisted approach (see Särndal et al., 1992) it is assumed that the value of each element of the finite population with respect to a certain target variable is a realization of a stochastic variable. This stochastic variable is modeled, e.g. according to a linear regression model, with the values of the auxiliary variables as covariates. Based on the assumed relationship between the target variable on the one hand and the auxiliary variables on the other, a general regression estimator can be derived of which most well-known estimators are special cases. After this estimator is derived, it is judged by its design-based properties, such as design expectation and design variance. The derived formulas hold irrespective of the validity of the model. If the regression model used to derive the estimator does not hold for the finite population, this will result only in higher design variances, but not in invalid estimators.

In conclusion, statistical methods from sampling theory are mainly intended to guarantee a sufficient external validity of a survey, i.e., the extent to which the sample statistics are valid estimates of the unknown characteristics of a given finite target population.

2.2.3 Comparison between design and analysis of experiments and sampling theory

Fienberg and Tanur (1987, 1988, 1989) and Van den Brakel and Renssen (1995) discussed the parallels and differences between the statistical methods from experimental design and sampling theory. Some typical parallels between the concepts of experimental design and sampling theory are summarized in table 2.1.

Table 2.1: Parallels between design and analysis of experiments and sampling theory.

EXPERIMENTAL DESIGN	SAMPLING THEORY
random assignment of experimental units to treatments	random sampling of sampling units from a finite population
replication of the treatments	sample size
randomized block designs	stratified sampling designs
row and column designs	lattice sampling or deep stratification
split-plot designs	two-stage sampling designs
covariance analysis	the general regression estimator

Besides these parallels, there are also some typical differences between experimental design and sampling theory, as indicated in the preceding sections. They are summarized in table 2.2. See also Fienberg and Tanur (1987, 1988, 1989) and Van den Brakel and Renssen (1995).

Table 2.2: Differences between design and analysis of experiments and sampling theory.

EXPERIMENTAL DESIGN	SAMPLING THEORY
stochasticity introduced because observations are assumed to be the outcome of a random variable described by a statistical model	stochasticity introduced because a probability sample is drawn from a finite population
predominantly model-based	predominantly design-based
statistical methods mainly intended to guarantee a sufficient internal validity	statistical methods mainly intended to guarantee a sufficient external validity
inference about a hypothetical infinite super population	inference about a given finite population

2.3 Design of embedded field experiments

Conducting small-scale experiments in laboratory settings is an appropriate and regularly used tool to develop questionnaire designs, interview procedures or to investigate non-sampling errors in survey processes more systematically (Fienberg and Tanur 1989). The advantage of laboratory experiments is the relative ease with which the effects of a large number of factors can be tested with a high degree of internal validity. The external validity of the results of such experiments, however, is generally not assured. To test the generalization of significant results obtained in such experiments to finite survey populations, large-scale field experiments embedded in sample surveys are very appropriate, in particularly when interest goes out to the quantification of the effect of alternative survey methodologies on estimates of finite population parameters. The statistical methods from the theory of experimental designs and sampling theory can be combined in a useful and natural way in the design and analysis of such embedded experiments (see Fienberg and Tanur, 1987, 1988). Parallels between the concepts of randomized experiments and random sampling should be exploited in the design of embedded experiments to improve the accuracy of estimated treatment effects and to draw correct conclusions about the observed effects. Thus, the sampling design of the survey forms a prior framework for the design of an embedded field experiment. This will result in a higher degree of the internal validity of the results obtained in an embedded experiment. To ensure external validity, statistical methods from sampling theory should support the analysis of these experiments. Because experimental units are selected by means of a probability sample from a finite population, it becomes possible to draw conclusions concerning population parameters. The remainder of this chapter is mainly devoted to design aspects of embedded experiments. In the next chapters we address the technical aspects of the analysis of such embedded experiments from the design-based perspective.

2.3.1 Design of embedded experiments within simple random sampling designs

Suppose that field experiments embedded in ongoing surveys are designed as split sample experiments in order to test the effects of K treatments. In a split sample experiment the sample is randomly divided into K similarly designed interpenetrating subsamples, not necessarily of equal subsample size. Each subsample can be considered as a probability sample from the population and is assigned to one of the K treatments. Now, if the original sample is drawn by means of simple random sampling, then the experiment is in fact a completely randomized design (CRD) (Cochran and Cox, 1957, Ch.4). Generally, this is not the most efficient design available, because no advantage is taken of the possibility to apply local control for sources of extraneous variation.

An important source of extraneous variation, for example, is the interviewer effect. In conducting an experiment, it should be avoided that treatments are systematically favored or handicapped because only experienced or inexperienced interviewers are assigned to one particular treatment. It is likely that respondents interviewed by the same interviewer produce more homogeneous answers than respondents interviewed by different interviewers and therefore it is efficient to apply local control for interviewers by means of randomized block designs (RBD) (Cochran and Cox, 1957, Ch.4) with interviewers as block variables (Fienberg and Tanur, 1988). When the statistical inference concerning the interviewer effects has to be extended to a larger population from which the interviewers are supposedly drawn, interviewers can be modeled as random components. This leads to an RBD with blocks as random effects, which is equivalent to a split-plot design.

2.3.2 Design of embedded experiments within more complex sampling designs

Sampling designs are usually more complex than simple random sampling. For instance stratified sampling and two-stage or cluster sampling are frequently applied. In a stratified sampling design there are two ways to divide the sample into K subsamples. Firstly, the whole sample is randomly divided into K subsamples, irrespective of the applied stratification. This may cause some differences in the frequency distribution of the experimental units over the strata among the K subsamples. Secondly, the sample is randomly divided into K subsamples in each stratum. Here the frequency distribution of the experimental units over the strata can be held equal for the K subsamples insofar as this is not disturbed by non-response. The first option, i.e. disregarding the stratification in the applied sampling design, leads to a CRD. The second option, i.e. dividing the sample in K subsamples in each stratum, leads naturally to an RBD with strata as block variables (Fienberg and Tanur, 1988). As sampling units from the same stratum are generally more homogeneous than sampling units from different strata, the second option may be more efficient than the first option. Also crossings between two or more control variables can be used as block variables (e.g. interviewers and strata). In the first option, local control for the stratification can be applied by means of covariance analysis.

In two-stage sampling designs, there are three different ways to divide the sample into K subsamples: Firstly, ignoring the primary sampling units, the secondary sampling units of the sample are divided into K subsamples. Secondly, the secondary sampling units within each primary sampling unit are divided into K subsamples. Thirdly, the primary sampling units are divided into K subsamples, and thus all secondary sampling units within a primary sampling unit are assigned to the treatment concerned. Disregarding the structure of the sampling design, the first option leads to a CRD where the secondary sampling units are the experimental units. Only if sufficient secondary sampling units are assigned to each of the K treatments within each primary sampling unit (large primary sampling units), is it possible to apply local control for the primary sampling units by means of covariance analysis. In the second approach the K treatments are randomly assigned to the secondary sampling units within each primary sampling unit or cluster. This type of randomization leads to a split-plot design. Primary sampling units in the sampling design correspond to the whole plots of the split-plot design and the secondary sampling units correspond to the split plots. Fienberg and Tanur (1988) argue that this parallel can be used to design experiments embedded in two-stage samples as split-plot designs in order to eliminate the variance between the whole plots (i.e. the primary sampling units) from the analysis of the treatment effects. This approach is appropriate when sampling units from the same primary sampling units are more homogeneous than sampling units from different primary sampling units. The third approach leads to a CRD with the primary sampling units as the experimental units. This approach is appropriate if the variation within the primary sampling units is large and the variation between the primary sampling units is small or if the size of the primary sampling units is small compared to the number of treatments.

2.3.3 Factorial designs

If the action of different types of treatments or factors is investigated, it is efficient to test these factors simultaneously in one factorial design instead of conducting separate experiments for each of these factors (Cochran and Cox, 1957, Ch.5). There are several advantages of factorial experimentation. In the first place, if the different factors are independent and are tested in one factorial design of size n, the treatment effect of each factor is estimated with the same precision as if each factor were tested in a separate CRD of size n. This results in a considerable savings in time and costs. In the second place, if the factors are not independent it becomes possible to estimate interaction effects of the different factors and to search for the optimal combination of the levels of these factors. Especially if there are interactions between the different factors, the validity of the conclusions of the experiment will be improved by simultaneously testing both factors in one factorial design instead of conducting separate experiments (correct model formulation). For this reason it can be necessary to introduce other factors into the experiment, which are not of interest by themselves, because they influence the effect of the factors of interest by means of interaction.

2.3.4 Design of (double) blind experiments

Field experiments have the advantage that they are conducted in the natural setting of the respondents who don't necessarily know that they participate in an experiment. If the experimental units or those who conduct the experiment (for instance the interviewers) know that they are participating in an experiment, then the behavior of the experimental units may be altered, perhaps even unconsciously. This type of biased results can be avoided by designing blind or double blind experiments. For example, in experiments where the effects of different questions or different sequences of questions in a questionnaire are compared, skilful use can be made of the flexibility offered by computer assisted interviewing. The alternative questions, or different order of questions, can be implemented in the supporting software package. If the routing of the questionnaire already depends on the response of the respondent, so that even without an experiment each interview is more or less unique, interviewer and respondent don't have to know that they are participating in an experiment. In many situations, however, it will be difficult or even impossible to design (double) blind experiments due to the nature of the treatments.

The disadvantages of blocking on interviewers in such situations is that interviewers are aware that they participate in an experiment. Consequently there is the danger that the interviewers introduce a bias due to selective behavior. In such situations, the experimenter is faced with the choice between a double blind experiment or an RBD with interviewers as block variables but consequently not double blind. This choice partially depends on the number and the type of treatments and the experience of the interviewers. If the introduction of a substantial bias due to a systematic interviewer effect can be expected because interviewers know that they participate in an experiment, then a double blind experiment where interviewers are randomly allotted over the treatments is preferable. A less precise comparison is generally less harmful than a systematic influence on one of the treatments.

Conducting an embedded not double blind experiment involves the danger that the regular survey, which besides publication purposes also serves as the control group, will acquire a higher priority than the experimental group and consequently distorts the conclusions of the experiment. This is illustrated by an experiment conducted by the US Bureau of Labor Statistics and the US Bureau of the Census in redesigning the US Current Population Survey (CPS) (O'-Muircheartaigh, 1997). In this experiment the new CPS and the old CPS were run in parallel for some period. First, the new CPS was run in parallel with the old CPS as the regular survey, indicating that the new CPS would lead to an increase of the estimated level of unemployment. After the changeover the old CPS was continued in parallel with the new CPS as the regular survey, as a further check on the observed effect. Unfortunately, in this case the old CPS showed an higher estimate of unemployment. O'Muircheartaigh attributes this outcome to the extra effort going into the regular survey on both occasions and to the fact that the interviewers were not blinded. Nevertheless it is a good methodological practice to run the two versions in parallel both before and after the changeover.

2.4 Embedded experiments conducted at Statistics Netherlands

To illustrate the problems encountered in the design of embedded field experiments, a series of embedded field experiments conducted at Statistics Netherlands are described and discussed in the following sections. In the experiments discussed in subsection 2.4.1 through 2.4.3, the sample of an ongoing survey has been randomly divided into one relatively large subsample and one or more smaller subsamples. The large subsample was in fact the regular survey and, besides publication purposes, served as the standard methodology or control group. The other subsamples were assigned to the alternative treatments so the experiment was conducted in parallel with the regular survey. The experiment discussed in subsection 2.4.4 is an example of a large scale field experiment, which was conducted completely separate from the field work of the regular survey.

2.4.1 Labour Force Survey

Several field experiments have been conducted to improve the data quality of the Dutch Labour Force Survey (LFS). The LFS is based on a stratified two-stage sample design with households as the ultimate sampling units. At the first stage a stratified sample of municipalities is drawn, where strata are formed by geographic areas. At the second stage a sample of addresses is drawn, from each selected municipality. All households on a selected address (with a maximum of three) are included in the sample. The data are collected in personal interviews with hand-held computers (CAPI).

Most of the information of the LFS is gathered by means of retrospective questions, which are often biased by memory effects. In 1990 an embedded experiment was conducted to investigate whether or not the quality of these retrospective data could be improved in terms of consistency and completeness. In the experimental group, a personal calendar indicating important events like festivals, holidays, birthdays, etc. was used during the completion of the retrospective questionnaires. This is supposed to give the respondent mnemonics in answering the survey questions, which should minimize memory effects.

Each interviewer worked in a particular interview district. In the experiment, interviewers were randomly allocated to a control group and an experimental group. The interviewers in the experimental group received special instruction in the use of the calendar to assist the completion of the questionnaires. Consequently, this experiment could not be conducted double blind. Respondents in a particular interview district were assigned to the group of the corresponding interviewer. So they were assigned to the experimental group or the control group, depending on the interviewer district they live in.

A significant decrease of memory errors was observed in the experimental group. However, due to the experimental design that was chosen at that time, it is unclear whether this decrease was caused by the experimental treatment or by an interviewer effect. The extra training of the interviewers in the experimental group could have favored this group systematically, due to extra motivation or attentiveness, and consequently distort the analysis of this experiment. These problems could have been avoided by randomizing households within interview districts over the two different treatments, leading to an RBD with interviewers as block variables or a split-plot design where interviewers correspond with the main plots and households with the split plots. Moreover, if respondents from the same household were more homogeneous with respect to their target variables, then the precision of such an experiment could have been improved by randomizing respondents within each household over the two treatments. This would have led to a split-plot design with interviewers as (fixed) block variables and whereby households correspond to the main plots and respondents to the split plots or a split-plot design with three randomization levels.

2.4.2 National Travel Survey

The sample design of the Dutch National Travel Survey (NTS) is similar to the sample design of the LFS (a stratified two-stage sample design with households as the ultimate sampling units). Only the geographical zoning of the applied stratification between the two surveys is different.

The data are collected in a telephone interview as well as a journey diary sent by mail. A few days after sending an advance letter, one of the members of the household is contacted by telephone and asked to provide some information about the household situation. Next, diaries are mailed to all household members. Each individual is asked to keep a record of all of his/her journeys for one day.

The journey diary has been adjusted several times, for example in order to collect more detailed information concerning travel behavior or to measure carpool behavior. Before new questions are added or other changes in the journey diary are implemented as a standard, it is tested by means of embedded experiments if such changes result in significant differences in the response rates and the estimates of the population parameters of the NTS. This enables us to explain and quantify trend changes in the time series of the population parameter estimates. In an other embedded experiment the possible effects of the implementation of an informed consent paragraph in the advance letter of the NTS on response rates and estimates of population parameters have been tested (Van den Brakel, Luppes and Moritz, 1995).

In each of these experiments households were randomly divided into a (large) control group and a (small) experimental group. These experiments were double blind. This was easily achieved, because the experimental factors concerned adjustments in the journey diary (sent by mail) or in the advance letter. In the analysis of the response rates, strata were incorporated as block variables and were found highly significant (Van den Brakel, Luppes and Moritz, 1995). In the carpool experiment, a few respondents with extremely long distance travels happened to be assigned to the experimental group. Because the experimental group was relatively small, these outliers had quite a large impact on the estimates of the population parameters and consequently resulted in an apparently highly significant treatment effect.

In randomized experiments, the concept of randomization ensures that each treatment has an equal probability of being favored or handicapped by an extraneous source of variation and consequently the observed effects can be assigned to the experimental treatments; however, there is always a small possibility that the results of an experiment are distorted due to an unfavorable outcome of the randomization of the experimental units over the treatments with respect to a covariate. If an experimenter believes that he is in this situation, the best option is to conduct another experiment, but usually this is not feasible due to time and money constraints. As an alternative the two-sample test of Wilcoxon and the two-sample Kolmogorov-Smirnov test were applied to analyze these experiments because these tests are robust against outliers and violations of the normality assumption. These tests could not find a significant difference between the estimates of the population parameters of the experimental group and the control group.

A third experiment was used to test whether working with two alternative calling schedules could improve the response rates and affect estimates of population parameters of the NTS. In these alternative calling schedules, the same number of attempts to contact respondents was more equally spread across the weekdays and the hours of the day. Within each interviewer households were randomly allotted to one control group and two experimental groups. Consequently an RBD was obtained with three treatments and with interviewers as block variables. Interviewer effects turned out to be highly significant with respect to the response rates. The implementation of the alternative calling schedules in the supporting computer system was quite complicated. To preclude distortion of the experiment due to initial problems with the computer system and the behavior of the interviewers in the new situations of the alternative treatments, one week of pretesting preceded the actual experiment. Many unforeseen practical problems, which arose during this pretesting were solved and consequently saved the experiment from systematic distortion of the experimental groups.

2.4.3 Justice and Security Survey

From 1980 until 1992, the Victimization Survey (VS) was conducted in order to measure the frequency of occurrence of particular types of crime. The survey was kept unchanged as long as possible in order to construct crime trends. In 1993 the VS was transformed into the Justice and Security Survey (JSS). In this new survey, several necessary and unavoidable changes were simultaneously implemented (Huys and Rooduijn, 1994).

The VS was originally based on a stratified three-stage sampling design with persons as the ultimate sampling units. In the first two stages, households were drawn in a way identical to the sampling design of the LFS. In the third stage one person was randomly selected from the household. In the VS people were interviewed about events in the preceding calendar year. Interviewing was carried out in January and February by means of CAPI. In the JSS this survey approach was fundamentally changed. In the third stage of the sample design of the JSS two persons (if possible) are randomly selected. The JSS is a continuous survey conducted every month. The figures to be published refer to the twelve months preceding the interview month. The JSS covers more items and new topics than the VS, and the wording of the questions has been modified.

To maintain the possibility of constructing crime trends, the effects of the differences between the JSS and the VS on parameter estimates were quantified by means of an experiment. During one year (1992), both surveys were conducted concurrently and treated as regular surveys. In this experiment two separate samples were drawn for the JSS and the VS, both with a sample size equal to the size the regular VS used to have. Interviewers were randomly allotted to the two treatments in order to keep the experiment double blind. Many estimates of the crime figures based on the JSS turned out to be significantly higher than in the VS. In this experiment only the total effect of all the changes introduced simultaneously could be quantified. Simple indexations were derived in order to keep the figures based on the VS and the JSS comparable. A factorial design should be applied to quantify the effect of the individual changes and their possible interaction.

2.4.4 The Fertility and Family Survey

The Netherlands Fertility and Family Survey (NFFS) was conducted in 1977, 1982, 1988, 1993 and 1998. Data were collected by means of computer assisted personal interviewing (CAPI). Response rates of the NFFS tend to be low, especially among men. The response rates of the NFFS 1993 were less then 50%. Therefore, we conducted a field experiment in an attempt to improve the response rates for the NFFS 1998 by means of an incentive. We tested the effect of a small monetary incentive: one promised, one prepaid. A telephone card with a value of NLG 2.50 with a specially designed NFFS logo and the Statistics Netherlands logo was used as a modern monetary incentive. In addition we tested the effect of presenting a shortened questionnaire to respondents who refused to complete the regular questionnaire of the NFFS 1998. This provides at least some information from non-respondents on the most relevant variables of the NFFS. However, giving the interviewers the opportunity of offering potential non-respondents a shortened questionnaire should not result in a decrease of the response rates for the regular questionnaire. Before implementing the two measures as a standard in the NFFS 1998 a controlled field experiment was conducted during the winter of 1997 to test the effect of both measures on the response rates.

The experiment was designed as a 3×2 factorial randomized block design. The first factor was an incentive by means of a telephone card tested at three levels; no incentive (control group), a prepaid incentive and a promised incentive. The second factor was the possibility for the interviewer to offer a shortened questionnaire tested at two levels; the interviewer did not have the possibility to offer a shortened questionnaire (control group) and the interviewer was asked to offer respondents a shortened questionnaire if they refused to participate with the regular questionnaire. Testing both factors in one factorial design is much more efficient than conducting two separate experiments (section 2.3.3).

From a theoretical point of view, it could be worthwhile to use interviewers as block variables in an RBD. The disadvantage of doing so, however, is that each interviewer must conduct each of the six different treatment combinations. This increases the risk that interviewers will confuse the different treatments in conducting the fieldwork, and consequently distort the experiment. Moreover, if interviewers know that they are participating in the experiment there is always the risk that they will introduce a bias due to selective behavior. Because response rates tend to correlate with urbanization levels, we decided to use municipalities with different urbanization levels as block variables in an RBD as an alternative.

Municipalities in the Netherlands are classified into five different levels of urbanization. In this experiment we selected the municipalities from the five different urbanization levels as follows. First, the two biggest cities (Amsterdam and Rotterdam) with the highest degree of urbanization were selected. Then, one municipality was drawn randomly from each of the five different urbanization levels. This resulted into to seven blocks or municipalities. Within each municipality, six interviewers were drawn randomly from Statistics Netherlands' field staff. Subsequently, a sample of persons from the target population of the NFFS was drawn in the interview area of each interviewer. Within each block the six interviewers together with the sample drawn in their interview area were randomized over the six different treatment combinations of the experiment. The total sample size of this experiment amounted 550 persons. Neither the interviewers nor the respondents knew that they were participating in an experiment. A separate instruction session was held for each of the six different treatment combinations to explain to the interviewers how to conduct the NFFS with their specific treatment combination.

The experiment was analyzed with a logistic regression model. No significant interaction effects between the two treatments (offering an incentive and a shortened questionnaire) could be found. The prepaid incentive resulted in the highest and most significant effect on the response rates. The effect of the promised incentive on the response rates was positive but not significant. The effect of the option to offer a shortened questionnaire was significantly negative. Apparently, there was a tendency for the response rates of the regular questionnaire to decline if the interviewer had the alternative of offering a shortened questionnaire to potential non-respondents.

Finally the block variable had a strong significant influence on the response rates. Having only one interviewer in each municipality for each treatment confounds the effect of interviewer and municipality. Consequently the block effects are a mixture of interviewer and urban effects. In this study, however, our interest goes out to the treatment effects and this block variable was only used to increase the precision of the estimated treatment effects. The confounding of interviewers and municipalities could have been avoided by having two or more interviewers in each municipality for each treatment. We decided not to implement this into the experimental design, due to budget constraints.

Based on the results of this field experiment, the prepaid incentive in the form of a telephone card, was implemented as standard in the NFFS 1998. Furthermore, it was decided not to use the shortened questionnaire to obtain at least some relevant information from potential non-respondents. See Van den Brakel and Renssen (1999, 2000) for a detailed description of the design, analysis and practical implications of this experiment.

2.5 Analysis of embedded field experiments

In section 2.3 we pointed out how the principles of experimental design can be applied in order to design efficient embedded field experiments. Here we will discuss how statistical methods from sampling theory can be used to support the analysis of embedded experiments if we are interested in testing hypotheses about finite population parameters.

In embedded experiments, experimental units are selected by some complex sampling design from a finite population. Statistical methods traditionally used in the analysis of experimental designs are model dependent and typically require IID observations. The stochastic assumptions underlying these techniques, however, do not reflect the complexity that is usually exhibited by the applied sampling design and the finite survey population from which the experimental units were drawn (Skinner et al, 1989, Ch.1). In these cases the assumption that the observations are IID is usually not tenable. The application of multivariate procedures based on the assumption of IID observations in the analysis of data obtained from complex sampling designs can lead to misleading results (Skinner et al, 1989, Ch.3). To deal with this problem, Skinner et al. (1989) proposes to adjust the analysis results of traditional multivariate procedures that are based on the IID-assumption ex-post by means of the so-called misspecification effect, which embodies the complexity of the sampling design.

Fienberg and Tanur (1987, 1988, 1989) advocated a model-based approach for the analysis of embedded field experiments. The internal validity is ensured by the application of such fundamentals as randomization, and local control on sampling structures such as strata, clusters or interviewers in the design as well as in the analysis of the embedded field experiment. The external validity is achieved by incorporating certain local control variables, such as interviewers or clusters, as random components in a mixed model analysis. Fienberg and Tanur (1988) showed that the weights of the applied sampling design can be ignored in the analysis when the selection of the sampling units depends only on prior variables which are conditioned on in the statistical model and are independent of the target variables.

From the examples mentioned in section 2.4, at least two purposes are distinguished for the analysis of embedded experiments:

1. Investigation of causal relationships between variables. For example the effects of different factors on the response rates in sample surveys, or factors that might improve the quality of retrospective questions.

2. Quantifying and testing the effect of alternative survey implementations on the estimates of finite population parameters of a sample survey.

In the first situation, conclusions are not necessarily generalized to the finite target population of one specific survey. In such situations a standard model-based analysis, possibly based on random or mixed linear models, might be appropriate. For example the experiments aimed to test effects on response rates can be analyzed properly with logistic regression models. In such situations it might be even preferable to design and conduct such field experiments completely separate from the fieldwork of an ongoing sample survey, instead of embedding them into a sample survey in order to avoid distortion of the experiment, e.g. by the conduction of the fieldwork.

In the second situation, experiments are used for example to detect and quantify possible trend disruptions in the time series of the estimated population parameters due to modifications in the survey process. This enables a survey manager to choose between the traditional survey implementation and one or more alternative approaches and accept expected trend changes, if there are any. For this purpose the design of experiments embedded in ongoing surveys will be very efficient since these experiments allow generalizing the observed results from the sample to the finite target population of the sample survey. In this case the analysis of an embedded experiment generally serves two purposes:

1. Estimation of finite population parameters under the different survey implementations (or treatments).

2. Testing hypothesis about the differences between the estimated population parameters obtained under the different survey implementations.

A first step in the analysis of embedded experiments is the estimation of the finite population parameters obtained under the K different survey implementations. This is necessary in order to quantify the effect of alternative approaches in relation to the size of these population parameters. In an embedded experiment, the sample of an ongoing survey is, according to an experimental design, divided into K subsamples s_k of size n_k. The elements of each subsample are assigned to one of the K treatments in the experiment. Since each subsample s_k can be considered as a (two-phase) survey sample from the finite target population, a design-unbiased estimator for the population parameter as well as its design variance under each of the K different treatments can be derived. As a result, K interval estimates for the finite population parameter under the K different treatments are obtained. The application of randomized sampling and design-unbiased estimators for the finite population parameters guarantees a correct generalization of the results obtained from the sample to the finite target population. As a result the external validity of this approach will be strong. These interval estimates, however, are not sufficient to test if the differences between the estimated population parameters are significant or not. Since each subsample is drawn from the same finite population without replacement, there is a nonzero design covariance between the subpopulation parameter estimates. This design covariance will generally be negative. The variance of the differences between the subsample estimates will be underestimated by ignoring this covariance (i.e. the subsample estimates are treated as if they are independent).

If it is ignored that the experimental units are drawn from a finite population by some complex sampling design, then the observations can be modeled according to an appropriate linear model for the applied experimental design. Under the assumption that the observations are normally and independently distributed, the hypothesis of no treatment effects can be tested by using e.g. an appropriate $F-$test (see, e.g. Scheffé, 1959 or Searle, 1971). Such an analysis guarantees a high degree of internal validity since the inference focus on the treatment differences within the experiment. The disadvantages of using a model-based approach for such problems is that the inference concerns model parameters from some superpopulation and not the estimates of the parameters of the finite survey population, even if the external validity is guaranteed by the use of random or mixed models. Since the analysis is conducted conditional on the realization of the sample, the external validity of the results will generally be low. First, ignoring the sampling design in the analysis might result in design-biased estimates of the treatment effects as well as misleading variance estimates in a sense that they are incompatible with the design-based parameter and variance estimates of the ongoing sample survey. Second, the dependent variables in the linear model applied to analyze the experiment doesn't necessarily coincide with the target parameters as they are defined in the sample survey. Therefore it is not always obvious how the analysis results obtained in a model-based inference are related to the population parameters of the sample survey, which complicates the interpretation of the results obtained in a model-based procedure.

To cope with these disadvantages we explore a design-based approach. In embedded field experiments, a large number of experimental units are drawn from a finite population by means of a random sampling design and are randomized to the different treatments according to the experimental design. As a result, it becomes possible to draw inferences on finite population parameters that do depend on a probability structure imposed by the design of the survey and not on model parameters from a superpopulation that depends on an assumed probability distribution. To this end, the analysis can be based on the estimates of finite population parameters. A natural approach, based on the objective of the experiment, is to formulate a hypothesis about the differences between the finite population parameters obtained under the different survey implementations and to construct efficient statistics to test this hypothesis. Based on the K interpenetrating subsamples, a design-unbiased estimator for the population parameters as well as the covariance matrix of the $K-1$ contrasts between these K population parameters can be derived under the randomization mechanism of the sampling design as well as the experimental design. This result in a design-based Wald$-$ or $t-$statistic that directly draws inferences about (approximate) design-unbiased estimates of population parameters. Both purposes of the analysis of embedded experiments are met in this procedure, i.e. estimation of population parameters under the different survey implementations as well as testing the hypothesis about differences between the population parameters.

The Horvitz-Thompson estimator as well as the generalized regression estimator, known from sampling theory, can be used to construct such test statistics, which take into account that experimental units are selected from a finite population by some complex sampling design with possibly unequal inclusion probabilities and/or clustering. Besides the sampling design, this estimator must also take into account the randomization of the experimental design applied to assign the experimental units to the different treatments. Furthermore, the use of auxiliary information by means of the generalized regression estimator might not only increase the precision of the analysis of embedded experiments, but it also makes the analysis more robust against the negative effects of selective non-response. Since such test statistics draw inferences on design-unbiased estimates of finite population parameters, a sufficient degree of external validity of the experiment is guaranteed. Together with the application of principles of randomized experimentation in the design stage, this results in statistical methods that combines the internal validity obtained from principals from experimental design with the external validity obtained from the design-based methods of sampling theory.

2.6 Discussion

Field experiments embedded in ongoing surveys are particular appropriate if we are interested in testing hypotheses about the effect of alternative survey methodologies or treatments on estimates of finite population parameters of a sample survey. Statistical methods from experimental designs and sampling theory can be combined in order to develop efficient methods for design and analysis of such experiments. Principles of experimentation should be applied in the design and analysis of embedded experiments to improve the precision of the estimated treatment effects and to avoid distorting the cause-effect relationship between treatments and outcomes. Trying to implement experiments embedded in ongoing surveys, where the regular survey is also used as the control group (see the examples in section 2.4), may come up against the danger that the regular survey takes priority over the alternative treatments by the conduction of the fieldwork. Therefore it may be efficient to conduct embedded experiments completely separate from the field work of regular surveys (for example, the NFFS-experiment discussed in subsection 2.4.4). Nevertheless, in large-scale field experiments it remains very difficult to standardize the application of treatments and to eliminate or exercise sufficient control over external influences. Usually the data collection of a field experiment involves many people (e.g. interviewers), which makes it difficult to standardize protocols of experimentation and to supervise compliance. So there are many sources of extraneous variation that can mask or bias the results of the experiment and distort the cause-effect relationship between treatments and observed effects. The principles of experimental designs can be applied in designing embedded field experiments in order to minimize the negative effects of these disturbances. Parallels between structures of experimental designs and sampling theory can be exploited in a straightforward manner by designing efficient embedded experiments based on these principles. The structure of the survey design forms a framework for the design of the experiment, e.g. local control by means of randomization within strata, clusters or interviewers.

In this chapter we advocate a design-based analysis in order to draw inferences about finite population parameters of the survey. Such statistical procedures are currently not generally available. The next chapters are devoted to develop such methods by deriving a design-based method for the analysis of the embedded CRD's and RBD's. RBD's enables us to apply local control on sampling structures like strata, PSU's, clusters and interviewers. This will lead naturally to statistical procedures for the design and analysis of embedded experiments, combining the internal validity guaranteed by methods from randomized experimentation with the external validity obtained from the theory of randomized sampling.

Chapter 3

A theoretical framework for the analysis of embedded experiments

3.1 Introduction

In this chapter we develop a design-based theory for the analysis of embedded experiments that is based on the Horvitz-Thompson estimator. In section 3.2 we introduce measurement error models for embedded experiments for two crucial reasons. Firstly, investigating treatment effects implicitly presumes the existence of measurement errors and hence we need some theoretical background to allow for such errors. Secondly, measurement error models naturally fit in the model-assisted approach of Särndal et al. (1992), by means of which treatment effects can be properly included in the field of finite population sampling. As we will see in section 3.3, the measurement error model serves as a vehicle to properly define treatment effects and to link these effects to systematic differences between finite population parameters observed under different treatments or survey implementations. This enables us also to properly define differences in finite population parameters in the presence of measurement errors. In section 3.4 we take notice of related problems of comparing domain means and repeated measurements in order to avoid confusion with our problem. The remaining parts of this chapter are devoted to developing and discussing a design-based Wald-statistic based on the Horvitz-Thompson estimator.

3.2 Measurement error models

Testing systematic effects of different survey methodologies on the observations obtained from the elements in the sample implies the existence of measurement errors. Therefore we start with the specification of measurement error models for the observations obtained under the different survey implementations or treatments of the experiment. We will follow the approach to measurement error modeling in surveys of Biemer et al. (1991). Consider a finite population U of N individuals. Let variable y_{ik}^{α} denote the response of the $i-$th individual $(i = 1, 2, \ldots, N)$ observed by means of treatment k $(k = 1, 2, \ldots, K)$ on the $\alpha-$th occasion. In sampling theory it is generally assumed that the observations obtained from the sampling units are fixed. In contrary to this we assume here that the response variable is random, which is expressed by adding the superscript α. The most simple measurement error model specifies that an observation y_{ik}^{α} is the sum of three terms, a true, intrinsic value u_i of individual i, a treatment effect β_k and an error component $\varepsilon_{ik}^{\alpha}$. Consequently, we can write $y_{ik}^{\alpha} = u_i + \beta_k + \varepsilon_{ik}^{\alpha}$. The treatment effects as well as the error terms are conceptual. For each individual in the population it is assumed that there is an infinite population of potential errors for each of the K different treatments (or survey methodologies). The error term $\varepsilon_{ik}^{\alpha}$ is assumed to be a realization of a hypothetical distribution obtained on the $\alpha-$th occasion that individuals i true value u_i is measured by means of treatment k. This model allows for slight variations in the response of individual i by repeated measurements on different occasions. Since we have defined K response variables for each individual obtained under K different treatments, we can express the basic measurement error model in matrix notation as

$$\mathbf{y}_i^{\alpha} = \mathbf{j}u_i + \beta + \varepsilon_i^{\alpha}, \tag{3.1}$$

where $\mathbf{y}_i^{\alpha} = (y_{i1}^{\alpha}, \ldots, y_{iK}^{\alpha})^t$, $\beta = (\beta_1, \ldots, \beta_K)^t$, $\varepsilon_i^{\alpha} = (\varepsilon_{i1}^{\alpha}, \ldots, \varepsilon_{iK}^{\alpha})^t$ and $\mathbf{j}$ a vector of order K with each element equal to one. Let E_{α} and Cov_{α} denote the expectation and the covariance with respect to the measurement error model. The following assumptions are made:

$$\mathrm{E}_{\alpha}(\varepsilon_i^{\alpha}) = \mathbf{0}, \tag{3.2}$$

$$\mathrm{Cov}_{\alpha}(\varepsilon_i^{\alpha}, \varepsilon_{i'}^{\alpha^t}) = \begin{cases} \mathbf{\Sigma}_i & : \quad i = i' \\ \mathbf{O} & : \quad i \neq i' \end{cases}, \tag{3.3}$$

where $\mathbf{0}$ is a vector of order K with each element zero and $\mathbf{O}$ a matrix of order $K \times K$ with each element zero. The model specified by (3.1), (3.2) and (3.3) will be referred to as the basic measurement error model for the K survey strategies. Random errors between individuals are modeled to be independent, while random errors within individuals may be dependent.

If there are interactions between treatments and individuals, then the treatment effects differ between the individuals. A measurement error model with additive treatment effects is nevertheless still applicable as long as we are interested in the average treatment effect. If β_{ik} denotes the effect of treatment k at the $i-$th individual, then β_k in (3.1) denotes the average treatment effect in the finite population, i.e. $\beta_k = (1/N)\sum_{i=1}^{N}\beta_{ik}$. Since we ignore these interactions in the measurement error model, the conclusions of our analysis in such situations concern the treatment effects averaged over the N elements in the finite population (see e.g. Scheffé, 1959, Ch. 4.1). This can also be concluded from formula (3.11) in section 3.3.

In many situations there may be some systematic clustering or correlation between the response of the individuals that violates (3.2) or (3.3). For example, due to interviewer effects measured values collected by the same interviewer might be correlated or systematically biased. This effect is, however, not unique to interviewers. Other common factors, such as coders and supervisors, may also induce correlation or bias among the response of the individuals. We will refer to this factor as an interviewer, keeping in mind that there are other possible sources for correlated response or bias. The basic measurement error model can be extended with a variable that describes this type of correlation or bias. Let J denote the number of interviewers available to carry out the fieldwork of a survey and s_j the subsample with individuals assigned to the $j-$th interviewer, $j = 1, \ldots, J$. The measurement error model for K survey strategies with interviewer effects is defined as

$$\mathbf{y}_i^\alpha = \mathbf{j}u_i + \beta + \mathbf{j}\gamma_i^\alpha + \varepsilon_i^\alpha, \tag{3.4}$$

where γ_i^α denotes the interviewer effect of the $i-$th individual. This interviewer effect is modeled as

$$\gamma_i^\alpha = \psi_{f(i)} + \xi_{f(i)}^\alpha \tag{3.5}$$

where f is a function, which assigns the $i-$th individual to one of the J interviewers. Without loss of generality it is assumed that the $i-$th individual is assigned to the $j-$th interviewer, i.e. $f(i) = j$. Here, ψ_j denotes the systematic error component due to the fixed interviewer effect of the $j-$th interviewer. This error is the same for all individuals assigned to this interviewer. The component ξ_j^α denotes the random error due to the $j-$th interviewer and is assumed to be a realization from a hypothetical error distribution with an expectation equal to zero and a variance equal to τ_j^2 on the $\alpha-$th occasion that individual i is interviewed by interviewer j. The realization of ξ_j^α may differ between the individuals that are assigned to this interviewer. Besides model assumptions (3.2) and (3.3), the following model assumptions are made:

$$\mathrm{E}_\alpha(\xi_{f(i)}^\alpha) = 0, \tag{3.6}$$

$$\mathrm{Cov}_\alpha(\xi_{f(i)}^\alpha, \xi_{f(i')}^\alpha) = \begin{cases} \tau_j^2 & : \quad f(i) = f(i') = j \\ 0 & : \quad f(i) \neq f(i') \end{cases}, \tag{3.7}$$

$$\mathrm{Cov}_\alpha(\varepsilon_{ik}^\alpha, \xi_{f(i)}^\alpha) = 0, \quad k = 1, 2, \ldots, K. \tag{3.8}$$

This measurement error model allows for mixed interviewer effects. If $\psi_j = 0$, then a model with only random interviewer effects is obtained. This measurement error model might be realistic if an interviewer, by repeated measuring on different occasions and different individuals, does not bias the response systematically but only increase the variance of the measurement errors. If $\tau_j^2 = 0$, then a model with only fixed interviewer effects is obtained. This measurement error model might be realistic if an interviewer, by repeated measuring on different occasions and different individuals, systematically biases the response with a fixed value. There is an extensive body of literature about random and fixed effect models for situations where data are assumed to be identically and independently distributed. We refer to Scheffé (1959) and Searle (1971) for a more detailed discussion.

There may be interactions between treatments and interviewers resulting in fixed interviewer effects that depend on the treatment (i.e. ψ_{jk}). As long as our interest goes out to the treatment effects averaged over the interviewers available to conduct the field work, measurement error model (3.4) is still applicable. This follows also formula (3.11) in section 3.3.

It follows from the model assumptions that

$$\mathrm{E}_\alpha(\mathbf{y}_i^\alpha) = \mathbf{j}u_i + \mathbf{j}\psi_{f(i)} + \beta, \tag{3.9}$$

and

$$\mathrm{Cov}_\alpha(\mathbf{y}_i^\alpha, \mathbf{y}_{i'}^\alpha) = \begin{cases} \mathbf{\Sigma}_i + \mathbf{J}\tau_j^2 & : \quad i = i' \text{ and } f(i) = j \\ \mathbf{J}\tau_j^2 & : \quad i \neq i' \text{ and } f(i) = f(i') = j \\ \mathbf{O} & : \quad i \neq i' \text{ and } f(i) \neq f(i') \end{cases} \tag{3.10}$$

where $\mathbf{J}$ is a $K \times K$ matrix with each element equal to one. Obviously, any correlation between the response of different individuals assigned to the same interviewer can be modeled by means of random interviewer effects. Any fixed interviewer effects influences the bias of the response values.

Knowledge about the sampling design or possible interviewer effects plays an important role in the design of embedded field experiments (see chapter 2). An RBD might be efficient if we can divide the individuals in the sample s into groups or blocks that are more or less homogeneous with respect to their measurements $\mathbf{y}_i^\alpha$. Firstly, homogeneity in $\mathbf{y}_i^\alpha$ can arise due to homogeneity in the individuals true values u_i. This is especially the case in stratified sampling, cluster sampling or two-stage sampling, since the target variables u_i of individuals from the same stratum, cluster or PSU tend to be more homogenous compared with individuals from different strata, clusters or PSU's. Therefore, under the basic measurement error model, the application of an RBD with sampling structures like strata, PSU's or clusters as block variables might be efficient. Secondly, due to random or fixed interviewer effects the response of sampling units interviewed by the same interviewer might be more homogeneous. Therefore, under the measurement error model with interviewer effects, the application of an RBD with interviewers as block variables might be efficient. When no such effects are expected, a CRD may be more suitable so as to avoid unnecessarily complications, e.g. in the field work (see e.g. section 2.3).

3.3 Hypotheses testing

A substantial part of the analysis of embedded experiments concerns of testing the hypotheses of no treatment effects with respect to finite population parameters. Especially under a measurement error model with interviewer effects, it is important that such parameters are defined precisely. Therefore we conceptually divide the population U of size N into J groups U_j of size N_j, $j = 1, \ldots, J$, such that all individuals within a group are potentially interviewed by the same interviewer and individuals between groups by different interviewers. Let $\bar{\mathbf{Y}}^\alpha = (\bar{Y}_1^\alpha, \bar{Y}_2^\alpha, \ldots, \bar{Y}_K^\alpha)^t$

denote the K dimensional vector of population means of $\mathbf{y}_i^{\alpha}$, i.e.

$$\bar{\mathbf{Y}}^{\alpha} = \mathbf{j}\frac{1}{N}\sum_{i=1}^{N} u_i + \beta + \mathbf{j}\sum_{j=1}^{J}\frac{N_j}{N}\psi_j + \mathbf{j}\sum_{j=1}^{J}\frac{N_j}{N}\xi_j^{\alpha} + \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^{\alpha}. \tag{3.11}$$

The objective of the experiment is to investigate whether there are systematic differences between the K population means of $\bar{\mathbf{Y}}^{\alpha}$ due to the K different survey strategies or treatments. Therefore, random deviations between the components of $\bar{\mathbf{Y}}^{\alpha}$ due to other types of measurement errors (including the interviewer effects) should not lead to significant differences in the analysis. This can be accomplished by formulating hypotheses about

$$\mathrm{E}_{\alpha}(\bar{\mathbf{Y}}^{\alpha}) = \mathbf{j}\frac{1}{N}\sum_{i=1}^{N} u_i + \mathbf{j}\sum_{j=1}^{J}\frac{N_j}{N}\psi_j + \beta \equiv \bar{\mathbf{Y}} \tag{3.12}$$

instead of (3.11), where the expectation is taken over the measurement error model. Since we are mainly interested in systematic treatment effects the following hypothesis is formulated:

$$\begin{aligned} H_0 &: \ \mathrm{E}_{\alpha}(\bar{Y}_1^{\alpha}) = \mathrm{E}_{\alpha}(\bar{Y}_2^{\alpha}) = \ldots = \mathrm{E}_{\alpha}(\bar{Y}_K^{\alpha}), \text{ (there are no treatment effects)}, \\ H_1 &: \ \mathrm{E}_{\alpha}(\bar{Y}_k^{\alpha}) \neq \mathrm{E}_{\alpha}(\bar{Y}_{k'}^{\alpha}), \ (k, k' = 1, 2, \ldots, K \text{ and } k \neq k' \text{ for at least one pair}). \end{aligned} \tag{3.13}$$

Hypothesis (3.13) can also be written in matrix notation as:

$$\begin{aligned} H_0 &: \ \mathbf{C}\mathrm{E}_{\alpha}\bar{\mathbf{Y}}^{\alpha} = \mathbf{0}, \\ H_1 &: \ \mathbf{C}\mathrm{E}_{\alpha}\bar{\mathbf{Y}}^{\alpha} \neq \mathbf{0}, \end{aligned} \tag{3.14}$$

where $\mathbf{0}$ denotes a vector of order $K-1$ with each element equal to zero and $\mathbf{C}$ a $((K-1)\times K)$ matrix containing the $K-1$ contrasts between the elements of $\mathrm{E}_{\alpha}\bar{\mathbf{Y}}^{\alpha}$, for example:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{pmatrix} = \left(\mathbf{I} \ \vdots \ \mathbf{0}\right) + \left(\mathbf{0} \ \vdots \ -\mathbf{I}\right) \tag{3.15}$$

or,

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 & 0 \\ 1 & 0 & -1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & -1 \end{pmatrix} = \left(\mathbf{j} \ \vdots \ -\mathbf{I}\right), \tag{3.16}$$

where $\mathbf{j}$ is a vector with ones of order $(K-1)$, and $\mathbf{I}$ the identity matrix of order $((K-1)\times(K-1))$. Since $\mathbf{Cj} = \mathbf{0}$, it follows that $\mathbf{C}\bar{\mathbf{Y}} = \mathbf{C}\beta$ and hypothesis (3.14) concerns the treatment effects as represented by β. The differences between the population parameters neatly correspond to the treatment effects. For reasons that are given in section 2.5 we emphasis, however, that this hypothesis will be tested by estimating $\bar{\mathbf{Y}}$ instead of β, taking into account both the sampling design and the experimental design. This why this approach is called design-based, even though the hypothesis can be expressed in terms of a model parameter. The measurement error model is only used to properly define the finite population parameters in case of measurement errors such as treatment effects. In the remainder of this section, we only sketch this approach. It will be further developed in sections 3.5 through 3.7 and in chapter 4.

To test the hypothesis specified by (3.14), a sample s of n individuals is available, where s is drawn from a finite population U of size N by some complex sampling design with first-order inclusion probabilities π_i for sampling unit i and second-order inclusion probabilities $\pi_{ii'}$ for sampling units i and i'. According to the experimental design the n individuals (or experimental units) are randomized over K subsamples s_k of size n_k. The experimental units of subsample s_k are assigned to one of the K treatments. As a consequence, only one of the components of $\mathbf{y}_i^{\alpha}$, $i \in s$, is actually observed.

In section 2.5 it was emphasized that in the analysis of embedded experiments, the assumption of IID observations is generally not tenable, which complicates the application of standard model-based procedures. Generally, the purpose of embedded experiments is to estimate finite population parameters under different survey implementations and to test hypotheses about the differences between these population parameter estimates. Consider a sampling scheme with unequal inclusion probabilities or a self-weighted sampling scheme where different response rates between geographical regions lead to unequal net inclusion probabilities. Ignoring the sampling design in the analysis of an embedded experiment will generally result in design-biased estimates for the finite population parameters as well as misleading variance estimates. Moreover, a model-based procedure does not necessarily test hypotheses about effects of alternative survey approaches on the estimates of finite population parameters. Therefore the analysis results obtained in a standard model-based procedure might be incommensurable with the sampling results of the ongoing survey.

To draw inferences about $\bar{\mathbf{Y}}$, the analysis needs explicitly take into account the probability structure imposed by the chosen sampling design used to draw s, as well as the randomization mechanism of the experimental design used to divide s into K subsamples. The Wald test (Wald, 1943), is frequently applied in design-based testing procedures (see e.g. Skinner et al 1989 or Lehtonen and Pahkinen 1995). To test hypotheses (3.14) we will derive design-based Wald statistics under the joint probability structure of the sampling design and the experimental design. Let $\hat{\bar{\mathbf{Y}}}^{\alpha}$ denote a design-unbiased estimator for $\mathrm{E}_{\alpha}\bar{\mathbf{Y}}^{\alpha}$ and $\mathbf{V}$ the covariance matrix of $\hat{\bar{\mathbf{Y}}}^{\alpha}$ with respect to the sampling design, the experimental design and the measurement error model. If $\mathbf{V}$ is known, then (3.14) can be tested by the Wald statistic

$$W = \hat{\bar{\mathbf{Y}}}^{\alpha^t} \mathbf{C}^t \left(\mathbf{C}\mathbf{V}\mathbf{C}^t\right)^{-1} \mathbf{C}\hat{\bar{\mathbf{Y}}}^{\alpha}. \tag{3.17}$$

In general $\mathbf{V}$ is unknown and has to be estimated. As only one of the components of $\mathbf{y}_i^{\alpha}$, $i \in s$, is actually observed, a design-unbiased estimator for $\mathbf{V}$ will be hard to derive. Van den Brakel and Binder (2000) tried to overcome this problem by imputing the unobserved components.

The usefulness of their results, however, depends on the correctness of the imputation model. Fortunately, as is apparent from (3.17), it suffices, to derive an (approximately) design-unbiased estimator for the covariance matrix of the $K-1$ contrasts of $\hat{\hat{\mathbf{Y}}}^{\alpha}$, denoted $\widehat{\mathbf{CVC}^t}$. As will follows, this covariance matrix of contrasts does not suffer from the fact that only one component of $\mathbf{y}_i^{\alpha}$ is observed. Hypothesis (3.14) can be tested by means of the following design-based Wald statistic:

$$W = \hat{\hat{\mathbf{Y}}}^{\alpha^t} \mathbf{C}^t \left(\widehat{\mathbf{CVC}^t} \right)^{-1} \mathbf{C} \hat{\hat{\mathbf{Y}}}^{\alpha}. \tag{3.18}$$

If $\mathbf{A}$ is nonsingular, any $\mathbf{C}' = \mathbf{AC}$ gives the same statistic (3.17). For reasons of mathematical convenience, we prefer the contrast matrix (3.16). Besides, this matrix has the following nice interpretation. Generally the purpose of an embedded experiment is to compare the population parameter estimates of the ongoing survey with the estimates obtained under alternative approaches. Therefore the regular survey serves, besides the official publication purposes, also as the control group in the experiment. Suppose, without loss of generality, that the ongoing survey corresponds with the first subsample. Then the first column of (3.16) obviously corresponds with the control group.

We propose using the Horvitz-Thompson estimator and the generalized regression estimator as (approximately) design-unbiased estimators for $\mathrm{E}_{\alpha}\bar{\mathbf{Y}}^{\alpha}$ and $\widehat{\mathbf{CVC}^t}$. As a result, a statistical test is obtained that takes into account that experimental units are selected from a finite population by some complex sampling design with possibly unequal inclusion probabilities and/or clustering. Furthermore, such a design-based approach makes it possible to incorporate the weighting scheme of the survey in the analysis of an embedded experiment. Sections 3.5 and 3.6 further elaborate on the estimators $\hat{\hat{\mathbf{Y}}}^{\alpha}$ and $\widehat{\mathbf{CVC}^t}$ in the Wald statistic by means of the Horvitz-Thompson estimator. In chapter 4, we extend this theory to the generalized regression estimator. Before we start with this, some related problems are discussed first in section 3.4.

3.4 Related problems

At first sight, the $K-$sample problem seems a similar problem as comparing K domain means of a finite population in classical sampling theory. The problem of testing hypotheses concerning domain means is, nevertheless, conceptually different from the problem of testing treatment effects in an experiment. This can be seen as follows. In an experiment, for each individual K observations for the target parameter $(y_{i1}^{\alpha}, \ldots, y_{iK}^{\alpha})^t$ under the K different treatments are defined. The experimental design determines which of these K variables are observed at each experimental unit. From this point of view, the analysis of (embedded) experiments can be regarded as a multivariate problem with missing data (see for example Kempthorne (1952), Hinkelmann and Kempthorne (1994) and Gelman et al. (1995, Ch.7)). This concept is illustrated in figure 3.1.

In the case of comparing K domain means, for each unit in the finite population as well as the

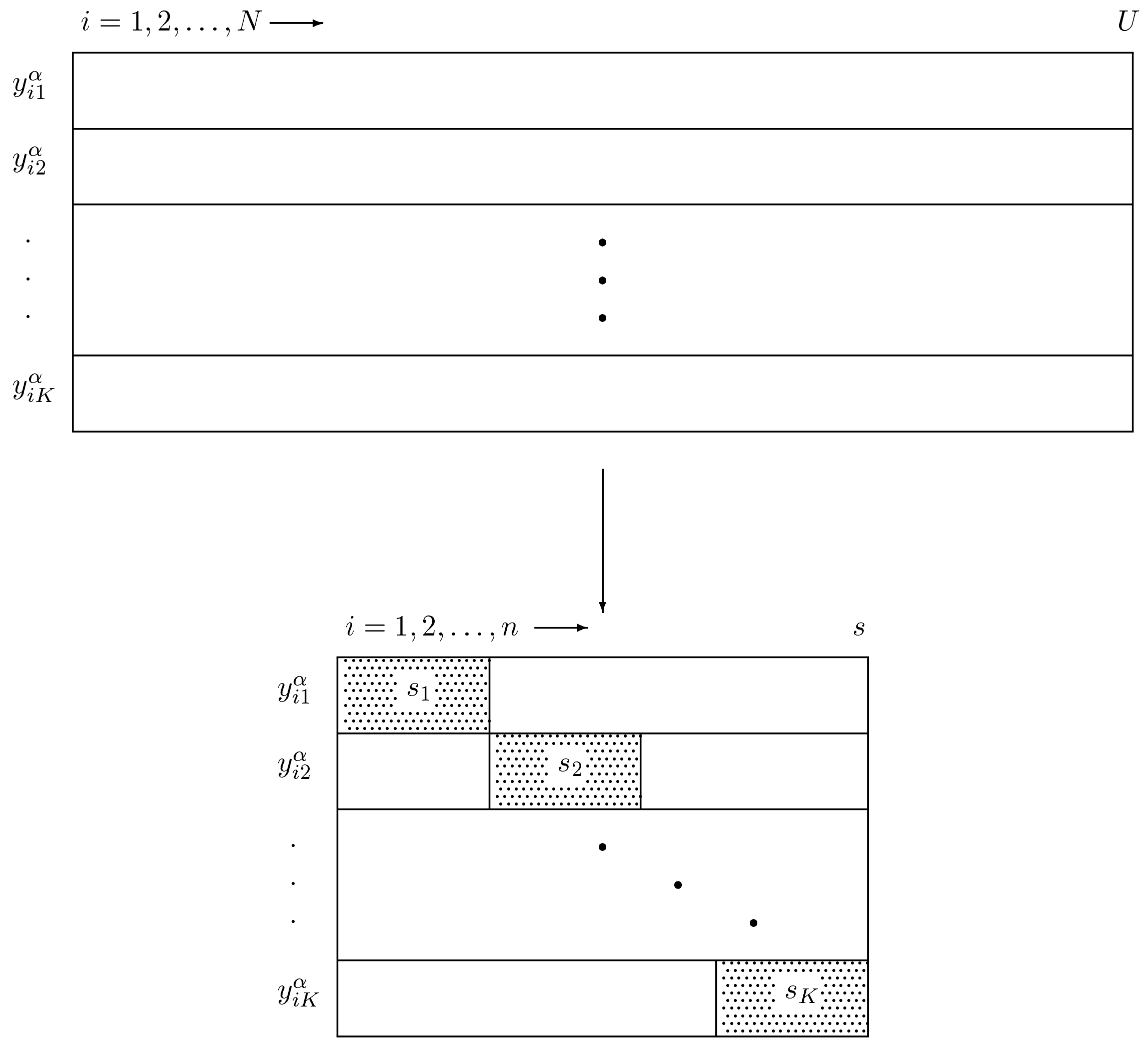

Figure 3.1: *Concept of embedded experiments. From a finite population U of size N (the upper rectangular box) a sample s of size n is drawn by means of a complex sample design (the lower rectangular box). For each element in the population and the sample, K observations under the different treatments are defined. According to the experimental design, sample s is divided into K subsamples s_k. Only one of the K observations are obtained within each subsample. This is expressed by means of the small shaded boxes. The unshaded part of the lower box symbolizes the unobserved observations in s.*

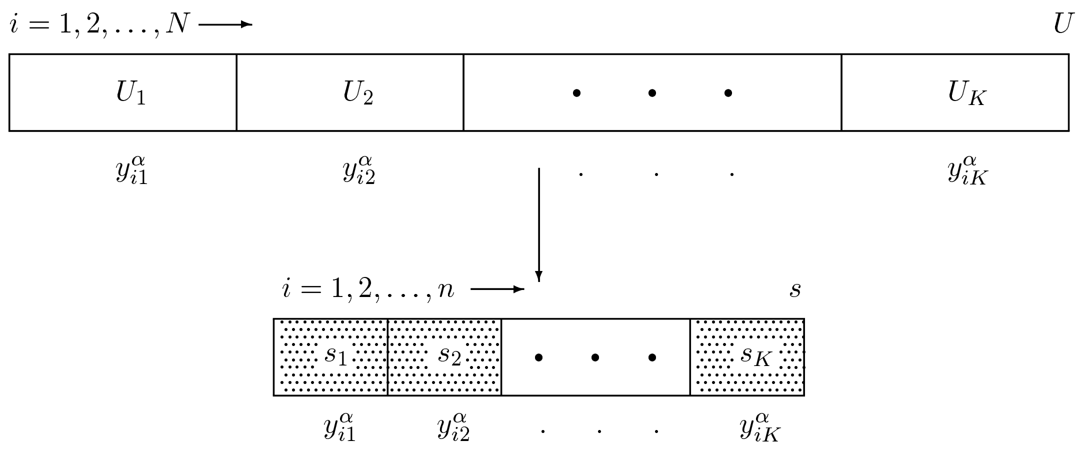

Figure 3.2: *Concept of comparison between domain means. A population U of size N can be divided into K subpopulations or domains (the upper rectangular box). A sample s of size n is drawn from this population (the lower rectangular box). For each element only one observation is defined. The observations obtained from the sampling units can be classified according to the K domains in order to estimate the corresponding domain means (shaded boxes).*

sample only one observation for the target parameter is defined. The population is divided into K different subpopulations or domains. The elements in the sample are used to estimate means, totals and absolute or relative sizes of these domains. This is basically an univariate problem. Furthermore, in the analysis of domain means only the probability structure of the sampling design is involved, while in the case of embedded experiments there is also the randomization mechanism of the experimental design. The concept of comparing domain means is illustrated in figure 3.2.

Hypotheses concerning exact equality of domain means are not sensible, and will generally be rejected because domain means in finite populations are generally different. Therefore, Cochran (1977, section 2.14), suggests testing the hypothesis that the domains were drawn from infinite populations having the same mean, whichs leads to standard model-based theory. Under the null hypothesis of an experiment, however, the finite population means under the different treatments are exactly equal. Since the sampling units are randomized over the K treatments by the experimental design, the estimated means under the different treatments will not be significantly different under the null hypothesis. This conceptual difference emphasizes the important role of randomization in experimental designs. It protects the analysis against distortion due to selective bias.

The conceptual difference between the analysis of embedded experiments and the comparison between domain means, is also illustrated by a complete enumeration of the finite population

U. Then, in the case of comparing domain means, each domain is completely observed. Consequently, the design variance of each domain estimate is equal to zero. In the case of an embedded experiment, however, according to the experimental design the entire population is divided into K subsamples of sizes n_k (with $\sum_{k=1}^{K} n_k = N$). Each subsample can be regarded as a survey sample. So despite the complete enumeration, the sample means under the different treatments are estimated and have a nonzero design variance, which is induced by the randomization mechanism of the experimental design. Thus, a complete enumeration for the sample means under each treatment is impossible under these experimental designs.

Another related problem is the analysis of repeated measurements. In this case, observations are collected under the K treatments of the experiment at different times on each individual, which is included in the sample. Each sampling unit serves as its own control. The analysis of repeated measurements can be regarded as a multivariate problem and is usually analyzed by means of a Hotelling T^2 statistic (Morrison, 1990, Ch.4). We don't consider this type of experiment, because there are many practical objections to collecting observations under all K treatments at each sampling unit in a large-scaled embedded field experiment. For example the response burden would become unacceptably high. Furthermore, in many applications, the nature of the treatments excludes the possibility to collect observations at the same unit under the different treatments of the experiment. The concept of repeated measurements is illustrated in figure 3.3.

3.5 Estimation of treatment effects

In the derivation of design-unbiased estimators for the Wald statistic, three sources of stochasticity are distinguished. Firstly, the stochasticity of the measurement error model as discussed in section 3.2. Secondly, the stochasticity induced by the sampling design. That is the stochasticity of the survey sample statistics by repeatedly drawing a survey sample s of size n from the finite population U of size N according to the same sampling design. In sampling theory, the randomization mechanism of the sampling design is usually described by the first and second-order inclusion probabilities π_i and $\pi_{ii'}$ of the $i-$th and i, $i'-$th individual(s), respectively. See Särndal et al. (1992, Ch.2) for an overview of the properties of these first and second-order inclusion probabilities. Thirdly, the stochasticity induced by the experimental design. That is the stochasticity of the subsample statistics given the survey sample s of size n by repeatedly dividing the given survey sample into K subsamples s_k of size n_k according to the same experimental design.

The randomization mechanism of the experimental design can also be described with conditional selection probabilities. We distinguish between CRD's and RBD's. In the case of a CRD, sample s of size n is randomly divided into K subsamples of size n_k. As a result, the conditional probability that the $i-$th individual is selected in subsample s_k, given that sample s is selected, equals n_k/n. In the case of an RBD, the n sampling units are deterministically divided

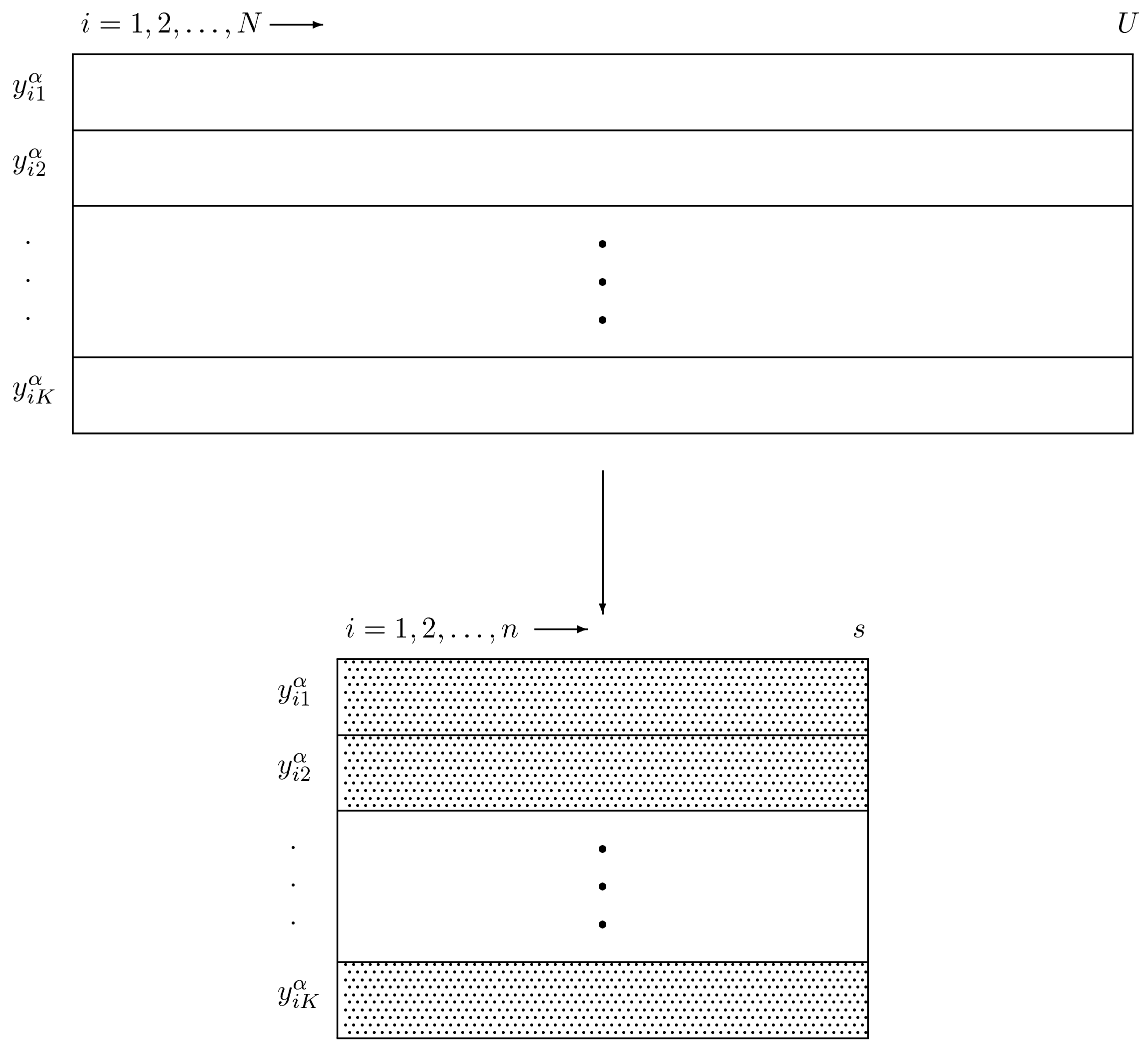

Figure 3.3: *Concept of repeated measurements. From a population U of size N (the upper rectangular box) a sample s of size n is drawn by means of a complex sample (the lower rectangular box). At each element in the population and the sample K observations are defined. Observations are obtained under the K treatments at each sampling unit in s (shaded boxes).*

in J blocks s_j of size n_j. Within each block, n_j experimental units are randomized over the K treatments. Let n_{jk} denote the number of experimental units in block j assigned to treatment k. It follows that the size of subsample s_k is equal to $n_k = \sum_{j=1}^{J} n_{jk}$. Furthermore it follows that $n = \sum_{j=1}^{J} n_j$ and $n_j = \sum_{k=1}^{K} n_{jk}$. The conditional probability that the $i-$th individual is selected in subsample s_k, given that sample s is selected and $i \in s_j$, equals n_{jk}/n_j.

Any estimator of the population parameter $\mathrm{E}_\alpha \bar{Y}_k^\alpha$ is necessarily based on the n_k experimental units in subsample s_k. Therefore we have to know the first-order inclusion probabilities of the experimental units assigned to subsample s_k, which can be derived in a straightforward manner. Let E_s denote the expectation with respect to the sampling design used to draw sample s and E_ε the expectation with respect to the randomization mechanism of the experimental design. Let a_i be the sample membership indicator of element i for sample s and a_{ik} the sample membership indicator of unit i for subsample s_k. In the case of a CRD, let a_{ik}^* denote the sample membership indicator for s_k given $i \in s$. Note that $a_{ik} = a_{ik}^* \times a_i$. Due to the fact that in a CRD sample s of size n is randomly divided into K subsamples s_k of size n_k, it follows that $\mathrm{E}_\varepsilon(a_{ik}^* \mid s) = n_k/n$. Then the first-order inclusion probabilities with respect to s_k are:

$$\mathrm{E}(a_{ik}) = \mathrm{E}_s \mathrm{E}_\varepsilon(a_{ik}^* \times a_i \mid s) = \frac{n_k}{n} \mathrm{E}_s(a_i) = \frac{n_k}{n} \pi_i. \tag{3.19}$$

In the case of an RBD, let a_{ik}^* denote the sample membership indicator for s_k given $i \in s_j$. Within each block, n_{jk} of the n_j experimental units are randomly assigned to treatment k. Consequently, $\mathrm{E}_\varepsilon(a_{ik}^* \mid s_j) = n_{jk}/n_j$. It follows that the first-order inclusion probabilities with respect to s_k are:

$$\mathrm{E}(a_{ik}) = \mathrm{E}_s \mathrm{E}_\varepsilon(a_{ik}^* \times a_i \mid s_j) = \frac{n_{jk}}{n_j} \mathrm{E}_s(a_i) = \frac{n_{jk}}{n_j} \pi_i. \tag{3.20}$$

These results can also be obtained as follows. Each subsample s_k can be considered as a two-phase sample. The first-order inclusion probabilities π_i of the first phase are obtained from the sampling design used to draw s. The conditional first-order inclusion probabilities of the second phase are obtained from the experimental design, used to divide s at random in K subsamples and are given by $\frac{n_k}{n}$ or $\frac{n_{jk}}{n_j}$ for a CRD or an RBD, respectively. From this point of view, the first-order inclusion probabilities with respect to s_k are given by $\frac{n_k}{n}\pi_i$ in the case of a CRD and $\frac{n_{jk}}{n_j}\pi_i$ in the case of an RBD.

Let $\hat{\bar{Y}}_k^\alpha$ denote the Horvitz-Thompson estimator for $\mathrm{E}_\alpha \bar{Y}_k^\alpha$ based on the experimental units of subsample s_k. For a CRD it follows that

$$\hat{\bar{Y}}_k^\alpha = \frac{n}{N n_k} \sum_{i \in s_k} \frac{y_{ik}^\alpha}{\pi_i}, \quad k = 1, 2, \ldots, K. \tag{3.21}$$

For an RBD we have

$$\hat{\bar{Y}}_k^\alpha = \frac{1}{N} \sum_{j=1}^{J} \sum_{i \in s_{jk}} \frac{n_j}{n_{jk}} \frac{y_{ik}^\alpha}{\pi_i}, \quad k = 1, 2, \ldots, K, \tag{3.22}$$

where s_{jk} denotes the subsample of individuals in block j, which are assigned to treatment k. Using (3.24) and (3.25), Horvitz-Thompson estimators (3.21) and (3.22) can be rewritten more generally as:

$$\hat{\bar{Y}}_k^\alpha = \frac{1}{N}\sum_{i\in s}\frac{\mathbf{p}_{ik}^t\mathbf{y}_i^\alpha}{\pi_i}, \tag{3.23}$$

where $\mathbf{p}_{ik}$ are $K-$vectors that describe the randomization mechanism of the experimental design. For a CRD, it follows from the identity between (3.21) and (3.23) that

$$\mathbf{p}_{ik} \equiv \begin{cases} \frac{n}{n_k}\mathbf{e}_k & \text{if} \quad i \in s_k \\ \mathbf{0} & \text{if} \quad i \notin s_k \end{cases}. \tag{3.24}$$

For an RBD, it follows from the identity between (3.22) and (3.23) that

$$\mathbf{p}_{ik} \equiv \begin{cases} \frac{n_j}{n_{jk}}\mathbf{e}_k & \text{if} \quad i \in s_{jk} \\ \mathbf{0} & \text{if} \quad i \notin s_{jk} \end{cases}. \tag{3.25}$$

Here $\mathbf{e}_k$ denotes the unit vector of order K with the $k-$th element equal to one and the other elements equal to zero and $\mathbf{0}$ a vector of order K with each element equal to zero. By means of these vectors the randomization mechanism of the experimental design is described. Together with the first and second-order inclusion probabilities π_i and $\pi_{ii'}$ for the sampling design they enable us to strictly separate the randomization mechanism of the sampling design and the experimental design in deriving the design-covariance matrix $\mathbf{CVC}^t$. In appendix 3.9.1 and 3.9.2, properties of the vectors $\mathbf{p}_{ik}$ are derived for CRD's and RBD's, respectively.

Note that given the sample s, $\hat{\bar{Y}}_k^\alpha$ given by (3.23) is still stochastic due to the randomization mechanism of the experimental design. This is expressed by the vectors $\mathbf{p}_{ik}$. The vector $\hat{\bar{\mathbf{Y}}}_{\mathbf{s}_k}^\alpha = (\hat{\bar{Y}}_1^\alpha, \ldots, \hat{\bar{Y}}_K^\alpha)^t$ is proposed as a design-unbiased estimator for $\bar{\mathbf{Y}}^\alpha = (\bar{Y}_1^\alpha, \ldots, \bar{Y}_K^\alpha,)^t$. The subscript $\mathbf{s}_k$ is added in order to emphasize that the vector $\hat{\bar{\mathbf{Y}}}_{\mathbf{s}_k}^\alpha$ consists of K Horvitz-Thompson estimators $\hat{\bar{Y}}_k^\alpha$, each based on the n_k elements of subsample s_k. By definition $\hat{\bar{\mathbf{Y}}}_{\mathbf{s}_k}^\alpha$ is also a design-unbiased estimator for $\bar{\mathbf{Y}} = \mathrm{E}_\alpha\bar{\mathbf{Y}}^\alpha$ and can therefore be used for the design-based Wald statistic (3.18).

It follows from $\bar{\mathbf{Y}}$, given by (3.12), that under a measurement error model with fixed interviewer effects the estimates of the population parameters are biased with a weighted mean of these fixed interviewer effects. Principally, under a CRD as well as an RBD, the bias in the parameter estimates due to fixed interviewer effects is under each of the K treatments the same. As a result this bias cancels out in the $K-1$ contrasts between the K parameter estimates. Under a CRD this holds true as long as each experimental unit assigned to an interviewer j, has a nonzero probability to be assigned to each of the K treatments. In many practical situations, however, experiments are designed in such a way that each interviewer only participates in one of the K treatments (see e.g. the example in section 2.4.4.). Such experiments are strictly speaking not CRD's since each treatment is handicapped with a different set of fixed interviewer effects. As a result the contrasts between the parameter estimates will be biased with fixed interviewer effects.

3.6 Variance estimation of treatment effects

3.6.1 Derivation of the covariance matrix

In this section a design-unbiased estimator is derived for the covariance matrix of the $K-1$ contrasts of $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$. Let $\mathbf{V}$ denote the covariance matrix of $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ and $\mathbf{C}\mathbf{V}\mathbf{C}^t$ the covariance matrix of the $K-1$ contrasts of $\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$. Since $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ is based on K interpenetrating subsamples drawn from a finite population, the K elements of $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ are dependent. Consequently, $\mathbf{V}$ is not a diagonal matrix. To estimate the covariance terms on the off-diagonal elements of $\mathbf{V}$ vectors $\mathbf{y}_i^{\alpha}$, containing the observations of all K treatments obtained from each experimental unit, are required. Due to the fact that in the experimental designs under consideration each experimental unit is assigned to one of the K treatments, only one of the K variables y_{ik}^{α} are observed at each experimental unit i. Consequently, we cannot derive a design-unbiased estimator for $\mathbf{V}$. In order to estimate the design-based Wald statistic (3.18), however, our interest goes out to the covariance matrix of $\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ instead of the covariance matrix of $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ itself. To cope with the fact that the complete vectors $\mathbf{y}_i^{\alpha}$ for each experimental unit are not observed, we concentrate on the covariance matrix of the contrasts $\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$.

Let Cov_{α}, Cov_s and $\mathrm{Cov}_{\varepsilon}$ denote the covariance with respect to the measurement error model, the sampling design and the experimental design, respectively. Consider the following variance decomposition:

$$\begin{aligned}\mathbf{C}\mathbf{V}\mathbf{C}^t &= \mathrm{Cov}_{\alpha}\mathrm{E}(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha) + \mathrm{E}_{\alpha}\mathrm{Cov}(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha) \\ &= \mathrm{Cov}_{\alpha}\mathrm{E}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s) + \mathrm{E}_{\alpha}\mathrm{Cov}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s) + \mathrm{E}_{\alpha}\mathrm{E}_s\mathrm{Cov}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s). \quad (3.26)\end{aligned}$$

The first two components on the right-hand side of (3.26) can be evaluated as follows:

$$\mathrm{Cov}_{\alpha}\mathrm{E}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s) = \frac{1}{N^2}\sum_{i=1}^{N}\mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t, \quad (3.27)$$

and

$$\mathrm{E}_{\alpha}\mathrm{Cov}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s) = \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\mathrm{Var}(\hat{N}) + \frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{1}{\pi_i} - 1\right)\mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t, \quad (3.28)$$

where

$$\mathrm{Var}(\hat{N}) = \sum_{i=1}^{N}\sum_{i'=1}^{N}\frac{(\pi_{ii'} - \pi_i\pi_{i'})}{\pi_i\pi_{i'}} \quad (3.29)$$

denotes the design variance of the estimated population size

$$\hat{N} = \sum_{i=1}^{n}\frac{1}{\pi_i}, \quad (3.30)$$

based on the n elements of sample s. Proofs are given in appendix 3.9.3 and 3.9.4, respectively. The third component on the right-hand side of (3.26) can be evaluated as

$$\mathrm{E}_{\alpha}\mathrm{E}_s\mathrm{Cov}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s) = \mathrm{E}_{\alpha}\mathrm{E}_s\mathbf{C}\mathbf{D}\mathbf{C}^t + \mathrm{E}_{\alpha}\mathrm{E}_s\mathbf{C}\mathbf{A}\mathbf{C}^t, \quad (3.31)$$

with

$$\mathbf{D} = \mathbf{diag}(d_1, \ldots d_K).$$

For a CRD it follows that:

$$d_k = \frac{1}{(n-1)} \frac{1}{n_k} \sum_{i=1}^{n} \left(\frac{n y_{ik}^{\alpha}}{N\pi_i} - \frac{1}{n} \sum_{i'=1}^{n} \frac{n y_{i'k}^{\alpha}}{N\pi_{i'}} \right)^2 \equiv \frac{S_k^2}{n_k}, \tag{3.32}$$

and

$$\mathbf{A} = -\frac{n}{(n-1)} \left(\sum_{i=1}^{n} \left(\frac{\mathbf{y}_i^{\alpha}}{N\pi_i} - \frac{1}{n} \sum_{i'=1}^{n} \frac{\mathbf{y}_{i'}^{\alpha}}{N\pi_{i'}} \right) \left(\frac{\mathbf{y}_i^{\alpha}}{N\pi_i} - \frac{1}{n} \sum_{i'=1}^{n} \frac{\mathbf{y}_{i'}^{\alpha}}{N\pi_{i'}} \right)^t \right). \tag{3.33}$$

For an RBD it follows that:

$$d_k = \sum_{j=1}^{J} \frac{1}{(n_j - 1)} \frac{1}{n_{jk}} \sum_{i=1}^{n_j} \left(\frac{n_j y_{ik}^{\alpha}}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{n_j y_{i'k}^{\alpha}}{N\pi_{i'}} \right)^2 \equiv \frac{S_{jk}^2}{n_{jk}}, \tag{3.34}$$

and

$$\mathbf{A} = -\sum_{j=1}^{J} \frac{n_j}{(n_j - 1)} \left(\sum_{i=1}^{n_j} \left(\frac{\mathbf{y}_i^{\alpha}}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{y}_{i'}^{\alpha}}{N\pi_{i'}} \right) \left(\frac{\mathbf{y}_i^{\alpha}}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{y}_{i'}^{\alpha}}{N\pi_{i'}} \right)^t \right). \tag{3.35}$$

Proofs are given in appendix 3.9.5. Note that (3.34) and (3.35) are J independent replications of the variance structure under a CRD given by (3.32) and (3.33).

Now we need to evaluate the expectation of $\mathbf{CAC}^t$ with respect to the measurement error model and the sampling design. Let

$$\tilde{\text{Var}}(\hat{N}) = \frac{1}{n} \left(\sum_{i=1}^{N} \frac{n}{\pi_i} - N^2 \right) \tag{3.36}$$

denote the variance of $\hat{N}$ as if the n individuals of s are drawn with replacement with selection probabilities π_i/n (see Cochran, 1977, section 9A.3). For a CRD it is proved in appendix 3.9.6 that

$$\text{E}_{\alpha}\text{E}_s\mathbf{CAC}^t = -\frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t}{\pi_i} - \frac{n}{(n-1)} \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2} \left(\tilde{\text{Var}}(\hat{N}) - \frac{1}{n}\text{Var}(\hat{N}) \right). \tag{3.37}$$

Note that under the null hypothesis $\mathbf{C}\beta = \mathbf{0}$ and under the alternative hypothesis $\mathbf{C}\beta \neq \mathbf{0}$. If the results (3.27), (3.28), (3.31) and (3.37) obtained for a CRD are substituted into (3.26), then it follows that

$$\mathbf{CVC}^t = \text{E}_{\alpha}\text{E}_s\mathbf{CDC}^t - \frac{n}{(n-1)} \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2} \left(\tilde{\text{Var}}(\hat{N}) - \text{Var}(\hat{N}) \right), \tag{3.38}$$

where the diagonal elements of $\mathbf{D}$ are defined by (3.32).

In order to evaluate $\text{E}_{\alpha}\text{E}_s\mathbf{CAC}^t$ for RBD's we distinguish two situations. First, consider RBD's where the block variables are directly linked with structures of the sampling design, e.g. when strata, clusters or PSU's are used as block variables. Second, consider RBD's where the block variables are not directly linked with structures of the sampling design, for example when interviewers are used as block variables. In this situation, the sampling design doesn't necessarily determine to which block an individual belongs, which implies that the blocks should be considered as domains. We will discuss each of these cases separately.

RBD's where PSU's are block variables.

Consider an RBD embedded in a two-stage sampling design, where PSU's are used as block variables. In the first stage, J PSU's or blocks are drawn from a finite population of J_u blocks. As a result the number of blocks becomes random with respect to the sampling design. In the second stage, a sample of n_j individuals is drawn from the PSU's, selected in the first stage. These individuals are randomized over the K treatments according to the experimental design. Let π_j and $\pi_{jj'}$ denote the first and second-order inclusion probabilities of the first stage of the sampling design. Let $\pi_{i|j}$ denote the first-order inclusion probability of individual i in the second stage conditional on the realization of the first stage. It follows that the first-order inclusion probabilities are $\pi_i = \pi_j \pi_{i|j}$. Furthermore $\pi_{ii'|j}$ denotes the second-order inclusion probabilities of individual i and i' in the second stage of the sampling scheme.

Let

$$\hat{N}_j = \sum_{i=1}^{n_j} \frac{1}{\pi_{i|j}} \tag{3.39}$$

denote the estimated population size of block j based on the n_j elements in block j. It follows that

$$\text{Var}(\hat{N}_j) = \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} \frac{(\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j})}{\pi_{i|j}\ \pi_{i'|j}}, \tag{3.40}$$

is the design variance of $\hat{N}_j$ and that

$$\tilde{\text{Var}}(\hat{N}_j) = \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j}{\pi_{i|j}} - N_j^2 \right), \tag{3.41}$$

is the variance of $\hat{N}_j$ as if the individuals of s_j are drawn with replacement with selection probabilities $\pi_{i|j}/n_j$ from a population of size N_j, (see Cochran, 1977, section 9A.3). For an RBD embedded in a two-stage sampling design where PSU's are block variables, it is proved in appendix 3.9.6 that

$$\text{E}_\alpha \text{E}_s \mathbf{CAC}^t = -\frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{C}\mathbf{\Sigma}_i \mathbf{C}^t}{\pi_i} - \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2} \sum_{j=1}^{J_u} \frac{1}{\pi_j (n_j - 1)} \frac{n_j}{} \left(\tilde{\text{Var}}(\hat{N}_j) - \frac{1}{n_j} \text{Var}(\hat{N}_j) \right). \tag{3.42}$$

If the results (3.27), (3.28), (3.31) and (3.42) are substituted into (3.26), then it follows that

$$\mathbf{CVC}^t = \text{E}_\alpha \text{E}_s \mathbf{CDC}^t + \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2} \left(\text{Var}(\hat{N}) - \sum_{j=1}^{J_u} \frac{1}{\pi_j (n_j - 1)} \frac{n_j}{} \left(\tilde{\text{Var}}(\hat{N}_j) - \frac{1}{n_j} \text{Var}(\hat{N}_j) \right) \right), \tag{3.43}$$

where the diagonal elements of $\mathbf{D}$ are defined by (3.34).

RBD's where strata are block variables.

Consider an RBD embedded in a stratified sampling design, where strata are used as block variables. In this situation the number of blocks in the finite population equals the number of blocks in the sample. A stratified sampling design can be considered as a two-stage sampling design where the first stage is completely observed. As a result, an expression for $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t$ for an RBD embedded in a stratified sampling design where strata are used as block variables follows directly as a special case from (3.42) with $\pi_j = 1$, $\pi_{i|j} = \pi_i$, $\pi_{ii'|j} = \pi_{ii'}$ and $J = J_u$. Thus

$$\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t = -\frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t}{\pi_i} - \frac{\mathbf{C}\beta\left(\mathbf{C}\beta\right)^t}{N^2}\sum_{j=1}^{J}\frac{n_j}{(n_j-1)}\left(\tilde{\mathrm{Var}}(\hat{N}_j) - \frac{1}{n_j}\mathrm{Var}(\hat{N}_j)\right), \tag{3.44}$$

where $\hat{N}_j$ is defined by (3.39), $\mathrm{Var}(\hat{N}_j)$ by (3.40) and $\tilde{\mathrm{Var}}(\hat{N}_j)$ by (3.41), where $\pi_{i|j}$ and $\pi_{ii'|j}$ are replaced by π_i and $\pi_{ii'}$, respectively.

If the results (3.27), (3.28), (3.31) and (3.44) are substituted into (3.26), then it follows for an RBD where strata are block variables that

$$\mathbf{CVC}^t = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t + \frac{\mathbf{C}\beta\left(\mathbf{C}\beta\right)^t}{N^2}\left(\mathrm{Var}(\hat{N}) - \sum_{j=1}^{J}\frac{n_j}{(n_j-1)}\left(\tilde{\mathrm{Var}}(\hat{N}_j) - \frac{1}{n_j}\mathrm{Var}(\hat{N}_j)\right)\right), \tag{3.45}$$

where the diagonal elements of $\mathbf{D}$ are defined by (3.34). Since in the case of a stratified sampling design,

$$\mathrm{Var}(\hat{N}) = \sum_{j=1}^{J}\mathrm{Var}(\hat{N}_j),$$

it follows that (3.45) can be simplified to

$$\mathbf{CVC}^t = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t - \frac{\mathbf{C}\beta\left(\mathbf{C}\beta\right)^t}{N^2}\sum_{j=1}^{J}\frac{n_j}{(n_j-1)}\left(\tilde{\mathrm{Var}}(\hat{N}_j) - \mathrm{Var}(\hat{N}_j)\right). \tag{3.46}$$

The expressions $\mathrm{Var}(\hat{N}_j)$ and $\tilde{\mathrm{Var}}(\hat{N}_j)$ are the design variances of the estimated block size. If the sampling scheme within each block is self-weighted and the sample size is fixed, then $\hat{N}_j = N_j$ and consequently $\mathrm{Var}(\hat{N}_j)$ and $\tilde{\mathrm{Var}}(\hat{N}_j)$ are equal to zero. That $\tilde{\mathrm{Var}}(\hat{N}_j)$ is equal to zero in this situation, follows directly from expression (3.41) for $\pi_{i|j} = n_j/N_j$. That $\mathrm{Var}(\hat{N}_j)$ is equal to zero can be proved as follows. If the sample within each block has a fixed size n_j then

$$\sum_{\substack{i'=1 \\ i'\neq i}}^{N_j}\pi_{ii'|j} = (n_j-1)\pi_{i|j}, \tag{3.47}$$

see Särndal et al. (1992), result 2.6.2. If the sampling scheme within each block is also self-weighted, i.e. $\pi_{i|j} = n_j/N_j$, then it follows for (3.40) that

$$\begin{aligned}
\mathrm{Var}(\hat{N}_j) &= \sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\frac{(\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j})}{\pi_{i|j}\pi_{i'|j}} \\
&= \left(\frac{N_j}{n_j}\right)^2\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\pi_{ii'|j} - N_j^2 \\
&= \left(\frac{N_j}{n_j}\right)^2\left((n_j-1)\sum_{i=1}^{N_j}\pi_{i|j} + \sum_{i=1}^{N_j}\pi_{i|j}\right) - N_j^2 = 0.
\end{aligned} \tag{3.48}$$

For the same arguments, the variances $\text{Var}(\hat{N})$ and $\tilde{\text{Var}}(\hat{N})$ in (3.29) and (3.36) are equal to zero in the case of self-weighted sampling designs with a fixed sample size or in the case of stratified sampling design where within each stratum a self-weighted sampling design with a fixed sample size is applied.

RBD's where clusters are block variables.

Consider an RBD embedded in a cluster sample, where clusters are used as block variables. A cluster sample can be considered as a two-stage sample where the second stage is completely observed. Therefore an expression for $\text{E}_\alpha \text{E}_s \mathbf{CAC}^t$ and $\mathbf{CVC}^t$ can be obtained for an RBD where clusters are block variables from (3.42) and (3.43) by taking $\pi_{i|j} = 1$, $\pi_{ii'|j} = 1$ and $n_j = N_j$. This implies that $\text{Var}(\hat{N}_j)$ and $\tilde{\text{Var}}(\hat{N}_j)$ defined by (3.40) and (3.41) respectively, are equal to zero.

RBD's where interviewers are block variables.

Consider an RBD embedded in a generally complex sampling design, where interviewers are used as block variables. Since the block variables are not directly linked with the sampling design in this case, the allocation of the sampling units to the blocks is random. To take into account for this extra variation, blocks should be considered as domains in the derivation of the covariance matrix of the contrasts between the subsample estimates. Therefore the finite population U is conceptually divided in J subpopulations U_j, which are used as block variables in the experiment. Let a_{ij} denote the membership indicator of individual i in block j;

$$a_{ij} = \begin{cases} 1 & : \; i \in U_j \\ 0 & : \; i \notin U_j \end{cases} . \tag{3.49}$$

Furthermore let

$$N_{a_j} = \sum_{i=1}^{N} a_{ij} \tag{3.50}$$

denote the size of the $j-$th block in the finite population. It can be interpreted as the number of sampling units in the finite population, which are assigned to the $j-$th interviewer if they are included in the sample s. The Horvitz-Thompson estimator for N_{a_j} is defined as

$$\hat{N}_{a_j} = \sum_{i=1}^{n} \frac{a_{ij}}{\pi_i}. \tag{3.51}$$

It follows that

$$\text{Var}(\hat{N}_{a_j}) = \sum_{i=1}^{N} \sum_{i'=1}^{N} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{a_{ij} \; a_{i'j}}{\pi_i \; \pi_{i'}}, \tag{3.52}$$

is the design variance of $\hat{N}_{a_j}$ and that

$$\tilde{\text{Var}}(\hat{N}_{a_j}) = \frac{1}{n_j} \left(\sum_{i=1}^{N} \frac{n_j a_{ij}}{\pi_i} - N_j^2 \right), \tag{3.53}$$

is the variance of $\hat{N}_{a_j}$ as if the individuals of s are drawn with replacement with selection probabilities π_i/n_j, (see Cochran, 1977, section 9A.3). It is proved in appendix 3.9.6 that

$$\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t = -\frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t}{\pi_i} - \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\sum_{j=1}^{J}\frac{n_j}{(n_j-1)}\left(\tilde{\mathrm{Var}}(\hat{N}_{a_j}) - \frac{1}{n_j}\mathrm{Var}(\hat{N}_{a_j})\right). \quad (3.54)$$

If the results (3.27), (3.28), (3.31) and (3.54) are substituted into (3.26), then it follows for an RBD where interviewers are block variables that

$$\mathbf{CVC}^t = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t + \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\left(\mathrm{Var}(\hat{N}) - \sum_{j=1}^{J}\frac{n_j}{(n_j-1)}\left(\tilde{\mathrm{Var}}(\hat{N}_{a_j}) - \frac{1}{n_j}\mathrm{Var}(\hat{N}_{a_j})\right)\right), \quad (3.55)$$

where the diagonal elements of $\mathbf{D}$ are defined by (3.34). Generally, the variance components $\mathrm{Var}(\hat{N}_{a_j})$ and $\tilde{\mathrm{Var}}(\hat{N}_{a_j})$ will be nonzero. Only if interviewer regions coincide exactly with the strata or PSU's and the sampling design within each stratum or PSU is self-weighted with a fixed sample size, then both variance components are equal to zero.

3.6.2 Estimation of the covariance matrix

Our next challenge is to find design-unbiased estimators for $\mathbf{CVC}^t$ defined by (3.38) in the case of a CRD and for (3.46), (3.43) and (3.55) in the case of an RBD. Note that $\mathbf{CVC}^t$ consists of one major component $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t$ and several components, which basically concerns the variance of population totals, i.e. $\tilde{\mathrm{Var}}(\hat{N})$ and $\mathrm{Var}(\hat{N})$ in the case of a CRD and $\mathrm{Var}(\hat{N})$, $\tilde{\mathrm{Var}}(\hat{N}_j)$, and $\mathrm{Var}(\hat{N}_j)$, $\tilde{\mathrm{Var}}(\hat{N}_{a_j})$ and $\mathrm{Var}(\hat{N}_{a_j})$ in the case of RBD's.

Without further elaborating on the expectation with respect to the measurement error model and the sampling design we can derive unbiased estimators for $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{D}$. Expressions for $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{D}$ are derived in chapters 5 and 6 for CRD's and RBD's embedded in different sampling designs. An unbiased estimator for $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{D}$ in the case of a CRD is given by

$$\hat{\mathbf{D}} = \mathbf{diag}(\hat{d}_1, \ldots, \hat{d}_K), \quad (3.56)$$

with

$$\hat{d}_k = \frac{1}{n_k}\frac{1}{(n_k-1)}\sum_{i=1}^{n_k}\left(\frac{n y_{ik}^\alpha}{N\pi_i} - \frac{1}{n_k}\sum_{i'=1}^{n_k}\frac{n y_{i'k}^\alpha}{N\pi_{i'}}\right)^2 \equiv \frac{\hat{S}_k^2}{n_k}. \quad (3.57)$$

Because in subsample s_k, n_k elements with variables $(n\, y_{ik}^\alpha)/(\pi_i\, N)$ are drawn by means of simple random sampling without replacement from sample s of size n, it follows that $\hat{S}_k^2$ in (3.57) is an unbiased estimator for S_k^2 in (3.32) (Cochran, 1977, Ch.2, Theorem 2.4). An unbiased estimator for $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{D}$ in the case of an RBD is given by $\hat{\mathbf{D}}$, with

$$\hat{d}_k = \sum_{j=1}^{J}\frac{1}{n_{jk}}\frac{1}{(n_{jk}-1)}\sum_{i=1}^{n_{jk}}\left(\frac{n_j y_{ik}^\alpha}{N\pi_i} - \frac{1}{n_{jk}}\sum_{i'=1}^{n_{jk}}\frac{n_j y_{i'k}^\alpha}{N\pi_{i'}}\right)^2 \equiv \sum_{j=1}^{J}\frac{\hat{S}_{jk}^2}{n_{jk}}. \quad (3.58)$$

Because in block j of subsample s_k, n_{jk} elements with variables $(n_j\, y_{ik}^\alpha)/(\pi_i\, N)$ are drawn by means of simple random sampling without replacement from block j of size n_j, it follows that

$\hat{S}^2_{jk}$ defined in (3.58) is an unbiased estimator for S^2_{jk} defined in (3.34) (Cochran, 1977, Ch.2, Theorem 2.4). Note that this estimator is unbiased under all kind of sources of variation, which are used as a block variable in an RBD.

The remaining variance components in $\mathbf{CVC}^t$ are the variances of population totals $\mathrm{Var}(\hat{N})$, $\tilde{\mathrm{Var}}(\hat{N})$, $\mathrm{Var}(\hat{N}_j)$, $\tilde{\mathrm{Var}}(\hat{N}_j)$, $\mathrm{Var}(\hat{N}_{a_j})$ and $\tilde{\mathrm{Var}}(\hat{N}_{a_j})$, which are multiplied with an unknown constant $\mathbf{C}\beta\,(\mathbf{C}\beta)^t$. It will be shown how these components can be estimated. For computational convenience, however, we will first argue that they are zero or at least negligible with respect to $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t$. Under the null hypotheses these variance components are equal to zero, since $\mathbf{C}\beta = \mathbf{0}$. Furthermore, these components concern the design variance of the population size or the block size in the finite population. For sampling designs for which it holds that $\hat{N} = N$, $\hat{N}_j = N_j$ or $\hat{N}_{a_j} = N_{a_j}$ it follows that these variance components are equal to zero. For sampling designs for which these properties do not hold, it follows that the size of the variance components under the alternative hypothesis depends on the variation between the first-order inclusion probabilities π_i of the sampling scheme. These terms might be negligible compared with the major term $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t$ as long as there is no extreme variation between the first-order inclusion probabilities. Furthermore, we will see in chapter 4 that under general conditions of the generalized regression estimator these covariance terms are equal to zero under the null hypothesis as well as the alternative hypothesis. As a result, it follows that

$$\mathbf{CVC}^t = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t. \tag{3.59}$$

Unbiased estimators for the diagonal elements of $\mathbf{D}$ are given by (3.57) for a CRD and (3.58) for an RBD.

For the sake of completeness, we show how the covariance component which concerns the variance of the population totals multiplied with the $K-1$ contrasts between the treatment effects β in (3.38), (3.46), (3.43) and (3.55) can be estimated. The estimation procedure of these covariance terms is illustrated for a CRD. It will be convenient to express the covariance component in (3.38) as

$$\frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\left(\mathrm{Var}(\hat{N}) - \frac{n}{(n-1)}\left(\tilde{\mathrm{Var}}(\hat{N}) - \frac{1}{n}\mathrm{Var}(\hat{N})\right)\right). \tag{3.60}$$

Recall that the first term of (3.60) arises from $\mathrm{E}_\alpha \mathrm{Cov}_s \mathrm{E}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{Y}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s)$ in (3.28) and the second term from $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t$ in (3.37). Since the variance components $\mathrm{Var}(\hat{N})$ and $\tilde{\mathrm{Var}}(\hat{N})$ are the same for each of the k subsamples, the entire sample s can be used for the estimation of these variance components.

The treatment effects β are identified by taking the $K-1$ contrasts between β, i.e. $\mathbf{C}\beta$. Since the treatment effects are assumed to be constants in the measurement error models, it follows from (3.12) that an estimator for $\mathbf{C}\beta$ is given by $\mathbf{C}\hat{\hat{\mathbf{Y}}}^\alpha_{\mathbf{s}_k}$.

A design-unbiased estimator for $\mathrm{Var}(\hat{N})$ in (3.60) is given by

$$\sum_{i=1}^{n}\sum_{i'=1}^{n}\frac{(\pi_{ii'} - \pi_i\pi_{i'})}{\pi_{ii'}}\frac{1}{\pi_i\pi_{i'}}.$$

A design-unbiased estimator for

$$\tilde{\mathrm{Var}}(\hat{N}) - \frac{1}{n}\mathrm{Var}(\hat{N}) \tag{3.61}$$

is given by

$$\sum_{i=1}^{n}\left(\frac{1}{\pi_i} - \frac{1}{n}\sum_{i'=1}^{n}\frac{1}{\pi_{i'}}\right)^2. \tag{3.62}$$

The proof that (3.62) is an unbiased estimator for (3.61) follows analogous to the derivation of result (3.110) in appendix 3.9.6. If these estimators are combined, then it follows that an estimator for (3.60) is given by

$$\frac{\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}\left(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}\right)^t}{N^2}\left[\sum_{i=1}^{n}\sum_{i'=1}^{n}\frac{(\pi_{ii'}-\pi_i\pi_{i'})}{\pi_{ii'}}\frac{1}{\pi_i\pi_{i'}} - \frac{n}{(n-1)}\sum_{i=1}^{n}\left(\frac{1}{\pi_i} - \frac{1}{n}\sum_{i'=1}^{n}\frac{1}{\pi_{i'}}\right)^2\right]. \tag{3.63}$$

Especially the computation of (3.61) might be time consuming due to the second-order inclusion probabilities.

In an equivalent way, for RBD's where PSU's are used as block variables, an estimator must be derived for

$$\frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\left(\mathrm{Var}(\hat{N}) - \sum_{j=1}^{J_u}\frac{1}{\pi_j}\frac{n_j}{(n_j-1)}\left(\tilde{\mathrm{Var}}(\hat{N}_j) - \frac{1}{n_j}\mathrm{Var}(\hat{N}_j)\right)\right). \tag{3.64}$$

Expressions for the corresponding variance components for RBD's where strata or clusters are block variables follow as a special case from (3.64). For RBD's where interviewers are used as block variables an estimator must be derived for

$$\frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\left(\mathrm{Var}(\hat{N}) - \sum_{j=1}^{J}\frac{n_j}{(n_j-1)}\left(\tilde{\mathrm{Var}}(\hat{N}_{a_j}) - \frac{1}{n_j}\mathrm{Var}(\hat{N}_{a_j})\right)\right). \tag{3.65}$$

An estimator for both components (3.64) and (3.65) is given by

$$\frac{\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}\left(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}\right)^t}{N^2}\left[\sum_{i=1}^{n}\sum_{i'=1}^{n}\frac{(\pi_{ii'}-\pi_i\pi_{i'})}{\pi_{ii'}}\frac{1}{\pi_i\pi_{i'}} - \sum_{j=1}^{J}\frac{n_j}{(n_j-1)}\sum_{i=1}^{n_j}\left(\frac{1}{\pi_i} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{1}{\pi_{i'}}\right)^2\right]. \tag{3.66}$$

3.6.3 Interpretation of the covariance matrix

In section 3.6.2 we argued that the covariance matrix of the $K-1$ contrasts between the estimated treatment effects is given by $\mathbf{CVC}^t = \mathrm{E}_\alpha\mathrm{E}_s\mathbf{CDC}^t$, where the diagonal elements of $\mathbf{D}$ are defined by (3.32) in the case of a CRD and (3.34) in the case of an RBD. The interpretation of these results is introduced by elaborating on the diagonal elements of $\mathbf{D}$ for the special case of a self-weighted sampling design (i.e. $\pi_i = n/N$). For a CRD embedded in a self-weighted sampling design, it follows from (3.32) that

$$d_k = \frac{1}{n_k}\frac{1}{(n-1)}\sum_{i=1}^{n}\left(y_{ik}^{\alpha} - \frac{1}{n}\sum_{i'=1}^{n}y_{i'k}^{\alpha}\right)^2. \tag{3.67}$$

This is the ordinary variance of a sample mean as if n_k elements are drawn by means of simple random sampling with replacement from a population of size n. Formula (3.32) has the same structure where the observations are weighted with a factor $n/(N\pi_i)$. Conditional on the realization of α and s, (3.32) might be interpreted as the variance of a sample mean as if n_k elements are drawn by means of simple random sampling with replacement where the observations y_{ik}^{α} are weighted with a factor $n/(N\pi_i)$. In an equivalent way $\hat{d}_k$ defined by (3.57) can be interpreted as the variance estimator of a sample mean as if n_k elements are drawn by means of simple random sampling with replacement from a population of size n, where the observations y_{ik}^{α} are weighted with a factor $n/(N\pi_i)$. Unconditional on α and s, $\hat{d}_k$ are ordinary variance estimators for the sample means as if the sample elements are selected with unequal probabilities (π_i/n) with replacement from the finite population U of size N (Cochran, 1977, equation (9A.16)).

In the case of an RBD embedded in a self-weighted sample design, it follows from (3.34) that

$$d_k = \sum_{j=1}^{J} \left(\frac{n_j}{n}\right)^2 \frac{1}{n_{jk}} \frac{1}{(n_j-1)} \sum_{i=1}^{n_j} \left(y_{ik}^{\alpha} - \frac{1}{n_j} \sum_{i'=1}^{n_j} y_{i'k}^{\alpha} \right)^2 . \tag{3.68}$$

This can be recognized as the ordinary variance of a sample mean as if n_k elements are drawn by means of stratified simple random sampling with replacement from a population of size n, divided in J strata of size n_j. Consequently, conditional on the realization of α and s, formula (3.34) might be interpreted as the variance of a sample mean as if n_k elements are drawn by means of stratified simple random sampling with replacement where the observations y_{ik}^{α} are weighted with a factor $n_j/(N\pi_i)$ from a population of size n, divided in J strata of size n_j. In an equivalent way $\hat{d}_k$ defined by (3.58) can be interpreted as the variance estimator of a sample mean as if n_k elements are drawn by means of stratified simple random sampling with replacement where the observations y_{ik}^{α} are weighted with a factor $n_j/(N\pi_i)$. Unconditional on α and s, $\hat{d}_k$ can be interpreted as the ordinary variance estimators for the sample means as if the sample elements are selected with unequal probabilities (π_i/n_j) with replacement within each block j (Cochran, 1977, equation (9A.16)).

It follows that the covariance matrix of the $K-1$ contrasts of $\hat{\bar{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ in (3.59) has the structure as if K independent subsamples are drawn by means of simple random sampling with replacement in the case of a CRD, or stratified simple random sampling with replacement in the case of an RBD. The probability structure of the sampling design is incorporated in the estimation procedure by means of a reweighting of the observations with a factor, which contains the first-order inclusion probabilities. Estimators have been derived for the covariance matrix of the contrasts between $\hat{\bar{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ as if the subsamples s_k were drawn independently from each other. These variance estimators only depend on the first-order inclusion expectations. No second-order inclusion expectations are required, which simplifies the variance estimation procedure considerably.

It is remarkable that the second-order inclusion probabilities of the sampling design are vanished. This is a consequence of several factors coming together. Since we concentrate on the variance of the contrasts between the subsample estimates, the two terms $\mathrm{Cov}_\alpha \mathrm{E}_s \mathrm{E}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{Y}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s)$ and $\mathrm{E}_\alpha \mathrm{Cov}_s \mathrm{E}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{Y}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s)$ in (3.26), which are given by (3.27) and (3.28) respectively, only contains measurement errors and additive treatment effects. The intrinsic values as well as the interviewer effects cancel out since they do not depend on the different treatments. Due to the assumption of independence of the measurement errors between the individuals, the cross products of the measurement errors between individuals, which contain the second-order inclusion probabilities, vanish. The components of $\mathbf{CVC}^t$, which contains second-order inclusion probabilities, concern the covariance of the additive treatment effects β (see (3.38), (3.46)), (3.43)) and (3.55)). Under the null hypothesis the covariance matrices of these constants are equal to zero. Under the alternative hypothesis, the size of these covariance terms depends on the variation between the first-order inclusion probabilities. In many practical situations, these terms might be negligible or even zero.

The structure of the third covariance matrix $\mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{Y}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s)$ in (3.26) is mainly determined by the randomization mechanism of the experimental design. Conditional on the realization α and s, the allocation of the experimental units can be regarded as simple random sampling without replacement in the case of a CRD and as stratified simple random sampling without replacement in the case of an RBD. Since we concentrate on the covariance matrix of the contrasts between subsample means and due to the randomization mechanism of the experimental design, the finite population corrections in the variance of the subsample means of the intrinsic values cancels out against the covariances between the subsample means. Therefore a variance estimation procedure could be derived as if the K subsamples are drawn independently from each other by means of simple random sampling with replacement for CRD's and stratified simple random sampling with replacement for RBD's. This phenomenon is illustrated in the remaining part of this subsection.

The randomization mechanism of the experimental design can be recognized in the structure of $\mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{Y}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s) = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t + \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t$, see (3.31). For a CRD, this covariance matrix has the structure equivalent to that K subsamples of size n_k drawn by means of simple random sampling without replacement from a finite population of size n. For an RBD, this covariance matrix has the structure associated with K subsamples of size n_k drawn by means of stratified simple random sampling from a population divided in J strata of size n_j. Consider for example a CRD, embedded in a self-weighted sampling scheme. Recall that the diagonal elements d_k, given by (3.67), have the structure of simple random sampling with replacement. Furthermore, it follows from (3.33) that the diagonal elements of $\mathbf{A}$, denoted a_{kk}, equals

$$a_{kk} = -\frac{1}{n}\frac{1}{(n-1)}\sum_{i=1}^{n}\left(y_{ik}^\alpha - \frac{1}{n}\sum_{i'=1}^{n} y_{i'k}^\alpha\right)^2. \tag{3.69}$$

Note that

$$d_k + a_{kk} = \frac{(1-n_k/n)}{n_k}\frac{1}{(n-1)}\sum_{i=1}^{n}\left(y_{ik}^\alpha - \frac{1}{n}\sum_{i'=1}^{n} y_{i'k}^\alpha\right)^2, \tag{3.70}$$

is the variance of a sample mean as if n_k elements are drawn by means of simple random sampling without replacement from a population of size n. The diagonal elements a_{kk} are in fact the finite population corrections. For the off-diagonal elements of $\mathbf{A}$, denoted $a_{kk'}$, it follows that

$$a_{kk'} = -\frac{1}{n}\frac{1}{(n-1)}\sum_{i=1}^{n}\left(y_{ik}^{\alpha} - \frac{1}{n}\sum_{i'=1}^{n} y_{i'k}^{\alpha}\right)\left(y_{ik'}^{\alpha} - \frac{1}{n}\sum_{i'=1}^{n} y_{i'k'}^{\alpha}\right). \tag{3.71}$$

This can be interpreted as the covariance between two sample means obtained from two interpenetrating samples of size n_k and $n_{k'}$, both drawn by means of simple random sampling without replacement from a population of size n.

For an RBD embedded in a self-weighted sampling design it follows that

$$d_k + a_{kk} = \sum_{j=1}^{J}\left(\frac{n_j}{n}\right)^2 \frac{(1-n_{jk}/n_j)}{n_{jk}}\frac{1}{(n_j-1)}\sum_{i=1}^{n_j}\left(y_{ik}^{\alpha} - \frac{1}{n_j}\sum_{i'=1}^{n_j} y_{i'k}^{\alpha}\right)^2, \tag{3.72}$$

and

$$a_{kk'} = -\sum_{j=1}^{J}\left(\frac{n_j}{n}\right)^2 \frac{1}{n_j}\frac{1}{(n_j-1)}\sum_{i=1}^{n_j}\left(y_{ik}^{\alpha} - \frac{1}{n_j}\sum_{i'=1}^{n_j} y_{i'k}^{\alpha}\right)\left(y_{ik'}^{\alpha} - \frac{1}{n_j}\sum_{i'=1}^{n_j} y_{i'k'}^{\alpha}\right). \tag{3.73}$$

Formula (3.72) is the variance of a sample mean as if n_k elements are drawn by means of stratified simple random sampling without replacement from a population divided in J strata of size n_j. The diagonal elements a_{kk} contain the finite population corrections of the sample means within each block. The off-diagonal elements $a_{kk'}$ contain the covariance between the sample means within each block.

Let $\hat{\hat{U}}_{s_k}$ denote the subsample mean of the intrinsic values of the target parameter of subsample s_k. If we concentrate on the variance of the contrast between two subsample means, conditional on the realization of α and s, then it follows that the finite population corrections in the design variance of $\hat{\hat{U}}_{s_k}$ and $\hat{\hat{U}}_{s_{k'}}$ cancels out against the design covariance of these two susbample means. For a CRD, for example, it follows for the intrinsic values that:

$$\mathrm{Var}_{\varepsilon}(\hat{\hat{U}}_{s_k}) = \frac{1-n_k/n}{n_k}\frac{1}{(n-1)}\sum_{i=1}^{n}\left(\frac{nu_i}{N\pi_i} - \frac{1}{n}\sum_{i'=1}^{n}\frac{nu_{i'}}{N\pi_{i'}}\right)^2 \equiv \frac{1-n_k/n}{n_k}S_u^2, \tag{3.74}$$

$$\mathrm{Var}_{\varepsilon}(\hat{\hat{U}}_{s_{k'}}) = \frac{1-n_{k'}/n}{n_{k'}}\frac{1}{(n-1)}\sum_{i=1}^{n}\left(\frac{nu_i}{N\pi_i} - \frac{1}{n}\sum_{i'=1}^{n}\frac{nu_{i'}}{N\pi_{i'}}\right)^2 \equiv \frac{1-n_{k'}/n}{n_{k'}}S_u^2, \tag{3.75}$$

$$\mathrm{Cov}_{\varepsilon}(\hat{\hat{U}}_{s_k}, \hat{\hat{U}}_{s_k}) = -\frac{1}{n}\frac{1}{(n-1)}\sum_{i=1}^{n}\left(\frac{nu_i}{N\pi_i} - \frac{1}{n}\sum_{i'=1}^{n}\frac{nu_{i'}}{N\pi_{i'}}\right)^2 \equiv -\frac{1}{n}S_u^2. \tag{3.76}$$

The conditional variance of a contrast equals

$$\begin{aligned}\mathrm{Var}_{\varepsilon}(\hat{\hat{U}}_{s_k} - \hat{\hat{U}}_{s_k}) &= \mathrm{Var}_{\varepsilon}(\hat{\hat{U}}_{s_k}) + \mathrm{Var}_{\varepsilon}(\hat{\hat{U}}_{s_{k'}}) - 2\mathrm{Cov}_{\varepsilon}(\hat{\hat{U}}_{s_k}, \hat{\hat{U}}_{s_{k'}}) \\ &= \frac{1-n_k/n}{n_k}S_u^2 + \frac{1-n_{k'}/n}{n_{k'}}S_u^2 + \frac{2}{n}S_u^2 = \frac{S_u^2}{n_k} + \frac{S_u^2}{n_{k'}} \end{aligned} \tag{3.77}$$

For an RBD, it follows in an equivalent way that

$$\begin{aligned}
\mathrm{Var}_\varepsilon(\hat{\hat{U}}_{s_k} - \hat{\hat{U}}_{s_k}) &= \sum_{j=1}^{J}\Big(\mathrm{Var}_\varepsilon(\hat{\hat{U}}_{s_{jk}}) + \mathrm{Var}_\varepsilon(\hat{\hat{U}}_{s_{jk'}}) - 2\mathrm{Cov}_\varepsilon(\hat{\hat{U}}_{s_{jk}}, \hat{\hat{U}}_{s_{jk'}})\Big) \\
&= \sum_{j=1}^{J}\left(\frac{1-n_{jk}/n_j}{n_{jk}}S_{u_j}^2 + \frac{1-n_{jk'}/n_j}{n_{jk'}}S_{u_j}^2 + \frac{2}{n_j}S_{u_j}^2\right) \\
&= \sum_{j=1}^{J}\left(\frac{S_{u_j}^2}{n_{jk}} + \frac{S_{u_j}^2}{n_{jk'}}\right),
\end{aligned}$$

where

$$S_{u_j}^2 = \frac{1}{(n_j-1)}\sum_{i=1}^{n_j}\left(\frac{n_j u_i}{N\pi_i} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{n_j u_{i'}}{N\pi_{i'}}\right)^2.$$

These considerations illustrate why second-order inclusion probabilities vanish in the variance component of the intrinsic values, conditional on the realization of α and s. Second-order inclusion probabilities only appear in the real variance-covariance matrix of the contrasts of $\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ and arise if the expectation with respect to the sampling design in the expressions of $\mathrm{E}_\alpha\mathrm{E}_s\mathbf{D}$ is worked out (see chapter 5 and 6).

3.7 Wald test for the hypothesis of no treatment effects

3.7.1 The design-based Wald statistic

A design-unbiased estimator for $\mathbf{C}\bar{\mathbf{Y}}$ is given by $\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$, where the elements of $\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ are defined by (3.23) in section 3.5. A design-unbiased estimator for the covariance matrix of $\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ is given by $\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t$, where the diagonal elements of $\hat{\mathbf{D}}$ are defined by (3.57) for a CRD and (3.58) for an RBD in section 3.6.2. Inserting these estimators into the design-based Wald statistic (3.18) in section 3.3 leads to

$$W = \hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha^t}\mathbf{C}^t\left(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t\right)^{-1}\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}. \tag{3.78}$$

It is proved in appendix 3.9.7 that

$$\mathbf{C}^t\left(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t\right)^{-1}\mathbf{C} = \hat{\mathbf{D}}^{-1} - \frac{1}{\mathrm{trace}(\hat{\mathbf{D}})^{-1}}\hat{\mathbf{D}}^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}^{-1}. \tag{3.79}$$

Due to the diagonal structure of $\hat{\mathbf{D}}$, the expression for the Wald statistic (3.78) can be further simplified by means of (3.79) as follows

$$\begin{aligned}
W &= \hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha^t}\left(\hat{\mathbf{D}}^{-1} - \frac{1}{\mathrm{trace}(\hat{\mathbf{D}})^{-1}}\hat{\mathbf{D}}^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}^{-1}\right)\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \\
&= \sum_{k=1}^{K}\frac{\hat{\hat{Y}}_k^{\alpha^2}}{\hat{d}_k} - \frac{1}{\sum_{k=1}^{K}\frac{1}{\hat{d}_k}}\left(\sum_{k=1}^{K}\frac{\hat{\hat{Y}}_k^{\alpha}}{\hat{d}_k}\right)^2.
\end{aligned} \tag{3.80}$$

Expression (3.80) can also be written as

$$W = \sum_{k=1}^{K} \frac{(\hat{\hat{Y}}_k^{\alpha} - \hat{\hat{Y}}^{\alpha})^2}{\hat{d}_k}, \tag{3.81}$$

where

$$\hat{\hat{Y}}^{\alpha} = \frac{\sum_{k=1}^{K} \frac{\hat{\hat{Y}}_k^{\alpha}}{\hat{d}_k}}{\sum_{k=1}^{K} \frac{1}{\hat{d}_k}}. \tag{3.82}$$

Under the null hypothesis the K subsample means $\hat{\hat{Y}}_k^{\alpha}$ are unbiased estimators for $\bar{Y}^{\alpha} = \bar{Y}_k^{\alpha}$ for $k = 1, 2, \ldots, K$. Then $\hat{\hat{Y}}^{\alpha}$ is the weighted sum of unbiased estimators $\hat{\hat{Y}}_k^{\alpha}$, weighted with the inverse of the estimated variance of $\hat{\hat{Y}}_k^{\alpha}$. Note that if $\sum_{k=1}^{K} \hat{\mathrm{Var}}(\vartheta_k \hat{\hat{Y}}_k^{\alpha})$ is minimized with respect to the weights ϑ_k under the restriction $\sum_{k=1}^{K} \vartheta_k = 1$, then the same weights $\vartheta_k = \frac{1}{\hat{d}_k} / (\sum_{k=1}^{K} \frac{1}{\hat{d}_k})$ are obtained. Expression (3.81) for the Wald statistic can be regarded as the variance between the subsample means $\hat{\hat{Y}}_k^{\alpha}$ weighted with the inverse of the variance of $\hat{\hat{Y}}_k^{\alpha}$. Under the null hypothesis $\hat{\hat{Y}}^{\alpha} \approx \hat{\hat{Y}}_k^{\alpha}$ for $k = 1, 2, \ldots, K$ and thus the value of W will be close to zero. Under the alternative hypothesis W will tend to larger values.

It is emphasized in section 3.6.3 that in the variance estimation procedure the K subsamples can be treated as if they are drawn independently from each other and that no second-order inclusion probabilities are required. This implies that the design-based Wald statistic is relatively simple to evaluate as expression (3.80) demonstrates. The efficiency of the sample design tends to vanish by the comparison of the subsample means $\hat{\hat{Y}}_k^{\alpha}$. This result seems to be in conformity with the results of Kish and Frankel (1974). They empirically found that the design effect for differences between subclass means tends towards one from below for stratified sampling designs with proportional allocation. Also for cluster samples they empirically found that the design effect of a positive intraclass correlation for differences between subclass means is less than for separate means.

The results obtained for the design-based Wald statistic are related to the literature on testing interviewer differences and other non-sampling or measurement errors. Mahalanobis (1946) used interpenetrating subsamples to study interviewer differences. Under simple random sampling and the assumption of equal workload an estimate of the total variance, which includes nonsampling errors due to e.g. interviewers, is obtained (Cochran, 1977, section 13.15). Hartley and Rao (1978) provided a general theory, using mixed linear models, to estimate the overall variance for stratified multistage sampling designs in which the last stage units are drawn with simple random sampling. If in the CRD's considered in this paper the subsamples are assigned to different interviewers, then the Wald statistic (3.81) can be interpreted as a weighted sum of squares of interviewer means $\hat{\hat{Y}}_k^{\alpha}$.

3.7.2 Construction of critical regions

The hypothesis of no treatment effects, given by (3.14) in section 3.3, can be tested with the design-based Wald statistic W defined in (3.78) and critical region $[\xi_{(1-\gamma)}, \infty)$, where $\xi_{(1-\gamma)}$ denotes the $(1-\gamma)$th quantile of W and γ the size of the test. In order to construct critical regions, the probability distribution of the test statistic (3.78) has to be known. For general sampling designs, the (asymptotic) distribution of this test statistic is unknown. However, if the sampling design is simple random sampling without replacement and the experimental design is a CRD, then, for given α, Lehmann (1975, appendix 8) gives sufficient conditions under which $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ is asymptotically multivariate normal distributed with mean $\mathrm{E}_s\mathrm{E}_\varepsilon(\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid s, \alpha) = \bar{\mathbf{Y}}^{\alpha}$ and covariance matrix $\mathbf{V}^{\alpha} = \mathrm{Cov}_s\mathrm{E}_\varepsilon(\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid s, \alpha) + \mathrm{E}_s\mathrm{Cov}_\varepsilon(\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid s, \alpha)$ if $n_k \to \infty$ and $N - n_k \to \infty$:

$$(\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha) \to \mathcal{N}\left(\bar{\mathbf{Y}}^{\alpha}, \mathbf{V}^{\alpha}\right).$$

See also Hájek (1960). So, we have

$$(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha) \to \mathcal{N}\left(\mathbf{C}\bar{\mathbf{Y}}^{\alpha}, \mathbf{C}\mathbf{V}^{\alpha}\mathbf{C}^t\right).$$

From (3.11) it follows that $\mathbf{C}\bar{\mathbf{Y}}^{\alpha} = \mathbf{C}\beta + \mathbf{C}\bar{\mathbf{E}}^{\alpha}$, where $\mathbf{C}\bar{\mathbf{E}}^{\alpha} = 1/N\sum_{i=1}^{N}\mathbf{C}\varepsilon_i^{\alpha}$. Since the $\mathbf{C}\varepsilon_i^{\alpha}$ are mutually independent random variables with mean zero and covariance matrix $\mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t$ we have by the ordinary central limit theorem

$$(\mathbf{C}\bar{\mathbf{E}}^{\alpha}) \to \mathcal{N}\left(\mathbf{0}, (1/N^2)\sum_{i=1}^{N}\mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t\right).$$

Combining both limit distributions we obtain unconditionally that

$$(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}) \to \mathcal{N}\left(\mathbf{C}\beta, \mathbf{C}\mathbf{V}\mathbf{C}^t\right).$$

Consequently, if sample s is drawn by means of simple random sampling without replacement, then the design-based Wald statistic W (3.78) is, under the null hypothesis, asymptotically distributed as a chi-squared random variable with $K-1$ degrees of freedom:

$$W \simeq \chi^2_{[K-1]}. \tag{3.83}$$

This is shown in appendix 3.9.8. In the case of stratified simple random sampling and an RBD where strata correspond with the block variables, this result remains valid as $n_{jk} \to \infty$ and $N_j - n_{jk} \to \infty$. No limit distribution for test statistic (3.78) is known for more complex sampling designs. Usually in survey literature the normality assumption for estimators based on complex sampling designs is assumed to be valid. If it is assumed that a limit theorem holds so that $(\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}) \to \mathcal{N}\left(\mathbf{C}\beta, \mathbf{C}\mathbf{V}\mathbf{C}^t\right)$, then the design-based Wald statistic is still asymptotically chi-squared distributed with $K-1$ degrees of freedom. In these cases critical regions for W can be constructed using the chi-squared distribution. Then $\xi_{(1-\gamma)}$ can be taken as $\chi^2_{[K-1]}(1-\gamma)$, where $\chi^2_{[K-1]}(1-\gamma)$ is the $(1-\gamma)$th quantile of the chi-squared distribution with $K-1$ degrees of freedom. If the normality assumption for $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ is doubtful, a permutation test might be an alternative to obtain approximate critical regions for W. See for example Efron and Tibshirani (1993) for an introduction to hypotheses testing with permutation tests. See Shao and Tu (1995) for the use of resampling methods in finite population settings.

We have conducted simulation studies to the sampling distribution of the test statistics in the case of the two-treatment experiment and the CRD under stratified sampling, cluster sampling, two-stage sampling and stratified two-stage sampling. In all these sampling schemes we used unequal inclusion probabilities. The assumption that the $t-$statistic and the Wald statistic are asymptotically standard normal respectively chi-squared distributed has been confirmed by these simulation studies (Huisman, 1998).

3.8 Discussion

A procedure for the analysis of experiments embedded in ongoing surveys is proposed in this chapter. The Horvitz-Thompson estimator is applied to derive design-unbiased estimators for the population parameters under the K different treatments of the experiment as well as the covariance matrix of the $K-1$ contrasts between these population parameters. As a result a Wald test is obtained, which draws inferences on the estimates of finite population parameters. Nevertheless, this is not a full design-based approach, since the results are obtained under specific measurement error models. These models assume that each observation is the result of a non-observable intrinsic value, being distorted by a possible additive treatment effect, a random or fixed interviewer effect and a measurement error. It is assumed that measurement errors between individuals are independent and that possible treatment effects are additive. Based on these assumptions an estimator is proposed for the covariance matrix, which has the structure as if the subsamples are drawn with replacement with unequal selection probabilities. This simplifies the computations remarkably. First no second-order inclusion probabilities are required in the variance estimators. Second, due to the diagonal structure of the covariance matrix, matrix operations can be avoided in the computation of the Wald statistic. In the proposed variance estimation procedure it is assumed that the variance components, which concern the design variance of the the additive treatment effects can be ignored. First these components are equal to zero under the null hypotheses. Second, these components can be interpreted as the design variances of the estimated population size or block size. Since these population sizes can be estimated without error in many practical situations, the corresponding variance components will be equal to zero or at least negligible. Estimators for these components are given for the sake of completeness. They, however, complicate the required calculations considerably. Moreover, in the next chapter these results are extended to the generalized regression estimator. It will be shown that under very general conditions these components always equal zero under the null as well as under the alternative hypothesis.

In Van den Brakel and Renssen (1996b, 1996c), a full design-based approach for the analysis of embedded experiments is followed. They derived the same variance estimation procedure as proposed in this chapter without assuming any type of measurement error models. These variance estimators are, however, only design-unbiased under the null hypothesis of no treatment effects. Under the alternative hypothesis the variances are generally overestimated.

3.9 Appendix

3.9.1 Properties of the randomization vectors $\mathbf{p}_{ik}$ for a CRD

For a CRD the randomization vectors $\mathbf{p}_{ik}$ are defined as

$$\mathbf{p}_{ik} = \begin{cases} \frac{n}{n_k}\mathbf{e}_k & \text{if} \quad i \in s_k \\ \mathbf{0} & \text{if} \quad i \notin s_k \end{cases} .$$

As a consequence of the randomization mechanism of a CRD, the vectors $\mathbf{p}_{ik}$ are random with the following conditional probability mass function

$$P\left(\mathbf{p}_{ik} = \frac{n}{n_k}\mathbf{e}_k \mid s\right) = \frac{n_k}{n} \quad \text{and} \quad P\left(\mathbf{p}_{ik} = \mathbf{0} \mid s\right) = 1 - \frac{n_k}{n},$$

with $\mathbf{e}_k$ the unit vector of order K. Also the following conditional probabilities follows from the randomization mechanism of a CRD:

$$\mathbf{p}_{ik}\mathbf{p}_{ik}^t = \begin{cases} \left(\frac{n}{n_k}\right)^2 \mathbf{e}_k\mathbf{e}_k^t & \text{with probability}: \quad \frac{n_k}{n} \\ \mathbf{O} & \text{with probability}: \quad 1 - \frac{n_k}{n} \end{cases} ,$$

$$\mathbf{p}_{ik}\mathbf{p}_{ik'}^t = \begin{cases} \frac{n}{n_k}\frac{n}{n_{k'}}\mathbf{e}_k\mathbf{e}_{k'}^t & \text{with probability}: \quad 0 \\ \mathbf{O} & \text{with probability}: \quad 1 \end{cases} ,$$

$$\mathbf{p}_{ik}\mathbf{p}_{i'k'}^t = \begin{cases} \frac{n}{n_k}\frac{n}{n_{k'}}\mathbf{e}_k\mathbf{e}_{k'}^t & \text{with probability}: \quad \frac{n_k}{n}\frac{n_{k'}}{(n-1)} \\ \mathbf{O} & \text{with probability}: \quad 1 - \frac{n_k}{n}\frac{n_{k'}}{(n-1)} \end{cases} ,$$

$$\mathbf{p}_{ik}\mathbf{p}_{i'k}^t = \begin{cases} \left(\frac{n}{n_k}\right)^2 \mathbf{e}_k\mathbf{e}_k^t & \text{with probability}: \quad \frac{n_k}{n}\frac{(n_k-1)}{(n-1)} \\ \mathbf{O} & \text{with probability}: \quad 1 - \frac{n_k}{n}\frac{(n_k-1)}{(n-1)} \end{cases} .$$

The expectation of $\mathbf{p}_{ik}$ with respect to the experimental design can be derived as:

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}) = P\left(\mathbf{p}_{ik} = \frac{n}{n_k}\mathbf{e}_k\right)\frac{n}{n_k}\mathbf{e}_k + P\left(\mathbf{p}_{ik} = \mathbf{0}\right)\mathbf{0} = \mathbf{e}_k. \tag{3.84}$$

The following expectations can be derived in an equivalent way:

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{ik}^t) = \frac{n}{n_k}\mathbf{e}_k\mathbf{e}_k^t,$$

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{ik'}^t) = \mathbf{O},$$

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{i'k'}^t) = \frac{n}{(n-1)}\mathbf{e}_k\mathbf{e}_{k'}^t,$$

$$\mathrm{E}_{\varepsilon}(\mathbf{p}_{ik}\mathbf{p}_{i'k}^{t}) = \frac{n}{n_k}\frac{(n_k-1)}{(n-1)}\mathbf{e}_k\mathbf{e}_k^t.$$

Since $\mathrm{Cov}(x, y) = \mathrm{E}(xy) - \mathrm{E}(x)\mathrm{E}(y)$, the following covariances with respect to the experimental design can be derived from these expected values;

$$\mathrm{Cov}_{\varepsilon}(\mathbf{p}_{ik}\mathbf{p}_{ik}^{t}) = \frac{(n-n_k)}{n_k}\mathbf{e}_k\mathbf{e}_k^t, \tag{3.85}$$

$$\mathrm{Cov}_{\varepsilon}(\mathbf{p}_{ik}\mathbf{p}_{ik'}^{t}) = -\mathbf{e}_k\mathbf{e}_{k'}^t, \tag{3.86}$$

$$\mathrm{Cov}_{\varepsilon}(\mathbf{p}_{ik}\mathbf{p}_{i'k'}^{t}) = \frac{1}{(n-1)}\mathbf{e}_k\mathbf{e}_{k'}^t, \tag{3.87}$$

$$\mathrm{Cov}_{\varepsilon}(\mathbf{p}_{ik}\mathbf{p}_{i'k}^{t}) = -\frac{(n-n_k)}{n_k}\frac{1}{(n-1)}\mathbf{e}_k\mathbf{e}_k^t. \tag{3.88}$$

3.9.2 Properties of the randomization vectors $\mathbf{p}_{ik}$ for an RBD

Let block j be denoted by s_j. For an RBD the randomization vectors $\mathbf{p}_{ik}$ are defined as

$$\mathbf{p}_{ik} = \begin{cases} \frac{n_j}{n_{jk}}\mathbf{e}_k & \text{if} \quad i \in s_{jk} \\ \mathbf{0} & \text{if} \quad i \notin s_{jk} \end{cases}.$$

As a consequence of the randomization mechanism of an RBD, the vectors $\mathbf{p}_{ik}$ are random with the following conditional probability mass function

$$P\left(\mathbf{p}_{ik} = \frac{n_j}{n_{jk}}\mathbf{e}_k \mid s_j\right) = \frac{n_{jk}}{n_j} \quad \text{and} \quad P\left(\mathbf{p}_{ik} = \mathbf{0} \mid s_j\right) = 1 - \frac{n_{jk}}{n_j},$$

with $\mathbf{e}_k$ the unit vector of order K. Furthermore, the following conditional probabilities follows from the randomization mechanism of an RBD:

$$\mathbf{p}_{ik}\mathbf{p}_{ik}^t = \begin{cases} \left(\frac{n_j}{n_{jk}}\right)^2 \mathbf{e}_k\mathbf{e}_k^t & \text{with probability : } \frac{n_{jk}}{n_j} & \text{if } i \in s_j \\ \mathbf{O} & \text{with probability : } 1 - \frac{n_{jk}}{n_j} & \end{cases},$$

$$\mathbf{p}_{ik}\mathbf{p}_{ik'}^t = \begin{cases} \frac{n_j}{n_{jk}}\frac{n_j}{n_{jk'}}\mathbf{e}_k\mathbf{e}_{k'}^t & \text{with probability : } 0 & \text{if } i \in s_j \\ \mathbf{O} & \text{with probability : } 1 & \end{cases},$$

$$\mathbf{p}_{ik}\mathbf{p}_{i'k'}^t = \begin{cases} \frac{n_j}{n_{jk}}\frac{n_j}{n_{jk'}}\mathbf{e}_k\mathbf{e}_{k'}^t & \text{with probability : } \frac{n_{jk}}{n_j}\frac{n_{jk'}}{(n_j-1)} & \text{if } i \in s_j \text{ and } i' \in s_j \\ \mathbf{O} & \text{with probability : } 1 - \frac{n_{jk}}{n_j}\frac{n_{jk'}}{(n_j-1)} & \text{if } i \in s_j \text{ and } i' \in s_j \\ \frac{n_j}{n_{jk}}\frac{n_{j'}}{n_{j'k'}}\mathbf{e}_k\mathbf{e}_{k'}^t & \text{with probability : } \frac{n_{jk}}{n_j}\frac{n_{j'k'}}{n_{j'}} & \text{if } i \in s_j \text{ and } i' \in s_{j'} \\ \mathbf{O} & \text{with probability : } 1 - \frac{n_{jk}}{n_j}\frac{n_{j'k'}}{n_{j'}} & \text{if } i \in s_j \text{ and } i' \in s_{j'} \end{cases},$$

$$\mathbf{p}_{ik}\mathbf{p}_{i'k}^t = \begin{cases} \left(\frac{n_j}{n_{jk}}\right)^2 \mathbf{e}_k\mathbf{e}_k^t & \text{with probability : } \frac{n_{jk}}{n_j}\frac{(n_{jk}-1)}{(n_j-1)} & \text{if } i \in s_j \text{ and } i' \in s_j \\ \mathbf{O} & \text{with probability : } 1 - \frac{n_{jk}}{n_j}\frac{(n_{jk}-1)}{(n_j-1)} & \text{if } i \in s_j \text{ and } i' \in s_j \\ \frac{n_j}{n_{jk}}\frac{n_{j'}}{n_{j'k}}\mathbf{e}_k\mathbf{e}_k^t & \text{with probability : } \frac{n_{jk}}{n_j}\frac{n_{j'k}}{n_{j'}} & \text{if } i \in s_j \text{ and } i' \in s_{j'} \\ \mathbf{O} & \text{with probability : } 1 - \frac{n_{jk}}{n_j}\frac{n_{j'k}}{n_{j'}} & \text{if } i \in s_j \text{ and } i' \in s_{j'} \end{cases}.$$

The expectation of $\mathbf{p}_{ik}$ with respect to the experimental design can be derived:

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}) = P\left(\mathbf{p}_{ik} = \frac{n_j}{n_{jk}}\mathbf{e}_k\right)\frac{n_j}{n_{jk}}\mathbf{e}_k + P\left(\mathbf{p}_{ik} = \mathbf{0}\right)\mathbf{0} = \mathbf{e}_k. \tag{3.89}$$

In a similar way it follws that

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{ik}^t) = \frac{n_j}{n_{jk}}\mathbf{e}_k\mathbf{e}_k^t \text{ if } i \in s_j,$$

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{ik'}^t) = \mathbf{O},$$

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{i'k'}^t) = \begin{cases} = \frac{n_j}{(n_j-1)}\mathbf{e}_k\mathbf{e}_{k'}^t & \text{if } i \in s_j \text{ and } i' \in s_j \\ = \mathbf{e}_k\mathbf{e}_{k'}^t & \text{if } i \in s_j \text{ and } i' \in s_{j'} \end{cases},$$

$$\mathrm{E}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{i'k}^t) \quad = \quad \begin{cases} = \frac{n_j}{(n_j-1)}\frac{(n_{jk}-1)}{n_{jk}}\mathbf{e}_k\mathbf{e}_k^t & \text{if } i \in s_j \text{ and } i' \in s_j \\ = \mathbf{e}_k\mathbf{e}_k^t & \text{if } i \in s_j \text{ and } i' \in s_{j'} \end{cases}.$$

The following covariances with respect to the experimental design can be derived from these expectations:

$$\mathrm{Cov}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{ik}^t) = \frac{(n_j - n_{jk})}{n_{jk}}\mathbf{e}_k\mathbf{e}_k^t, \quad \text{if } i \in s_j, \tag{3.90}$$

$$\mathrm{Cov}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{ik'}^t) = -\mathbf{e}_k\mathbf{e}_{k'}^t, \tag{3.91}$$

$$\mathrm{Cov}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{i'k'}^t) = \begin{cases} \frac{1}{(n_j-1)}\mathbf{e}_k\mathbf{e}_{k'}^t & \text{if } i \in s_j \text{ and } i' \in s_j \\ \mathbf{O} & \text{if } i \in s_j \text{ and } i' \in s_{j'} \end{cases}, \tag{3.92}$$

$$\mathrm{Cov}_\varepsilon(\mathbf{p}_{ik}\mathbf{p}_{i'k}^t) = \begin{cases} -\frac{(n_j-n_{jk})}{n_{jk}}\frac{1}{(n_j-1)}\mathbf{e}_k\mathbf{e}_k^t & \text{if } i \in s_j \text{ and } i' \in s_j \\ \mathbf{O} & \text{if } i \in s_j \text{ and } i' \in s_{j'} \end{cases}. \tag{3.93}$$

3.9.3 Proof of formula (3.27)

In this appendix it is proved that

$$\mathrm{Cov}_\alpha\mathrm{E}_s\mathrm{E}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^\alpha \mid \alpha, s) \quad = \quad \frac{1}{N^2}\sum_{i=1}^N \mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t.$$

Let $\hat{\hat{\mathbf{Y}}}_{\mathbf{s}}^\alpha = (\hat{\hat{Y}}_{1_\mathbf{s}}^\alpha, \hat{\hat{Y}}_{2_\mathbf{s}}^\alpha, \ldots, \hat{\hat{Y}}_{K_\mathbf{s}}^\alpha)^t$ denote a vector of order K with each element the Horvitz-Thompson estimator for $\mathrm{E}_\alpha\bar{Y}_k^\alpha$ based on the n sampling units of sample s:

$$\hat{\hat{\mathbf{Y}}}_{\mathbf{s}}^\alpha = \frac{1}{N}\sum_{i\in s}\frac{\mathbf{y}_i^\alpha}{\pi_i}. \tag{3.94}$$

Note that this estimator is not observable, since only n_k of the n experimental units are assigned to treatment k and thus only one of the K elements of each vector $\mathbf{y}_i^\alpha$ is observed instead of the whole vector. Using result (3.84) under a CRD (appendix 3.9.1) or (3.89) under an RBD (appendix 3.9.2), the expectation of $(\hat{\hat{Y}}_k^{\alpha} \mid \alpha, s)$ with respect to the experimental design can be evaluated as follows:

$$\begin{aligned}
\mathrm{E}_{\varepsilon}(\hat{\hat{Y}}_k^{\alpha} \mid \alpha, s) &= \mathrm{E}_{\varepsilon}\left(\frac{1}{N}\sum_{i\in s}\frac{\mathbf{p}_{ik}^t\mathbf{y}_i^{\alpha}}{\pi_i} \mid \alpha, s\right) = \frac{1}{N}\sum_{i\in s}\frac{\mathrm{E}_{\varepsilon}(\mathbf{p}_{ik}^t \mid \alpha, s)\mathbf{y}_i^{\alpha}}{\pi_i} \\
&= \frac{1}{N}\sum_{i\in s}\frac{\mathbf{e}_k^t\mathbf{y}_i^{\alpha}}{\pi_i} = \hat{Y}_{k_\mathbf{s}}^{\alpha}, \quad k = 1, 2, \ldots, K.
\end{aligned}$$

Consequently, it follows that

$$\mathrm{E}_{\varepsilon}(\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s) = \hat{\mathbf{Y}}_{\mathbf{s}}^{\alpha}. \tag{3.95}$$

Since $\mathbf{Cj} = \mathbf{0}$, it follows from the measurement error model (3.1) and (3.4) in section 3.1 that

$$\mathbf{Cy}_i^{\alpha} = \mathbf{C}\beta + \mathbf{C}\varepsilon_i^{\alpha}. \tag{3.96}$$

Using (3.94), (3.96) and the measurement error model assumption (3.3), it follows that

$$\begin{aligned}
\mathrm{Cov}_{\alpha}\mathrm{E}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s) &= \mathrm{Cov}_{\alpha}\mathrm{E}_s(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}}^{\alpha} \mid \alpha) = \mathrm{Cov}_{\alpha}(\mathbf{C}\bar{\mathbf{Y}}^{\alpha}) \\
&= \mathrm{Cov}_{\alpha}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Cy}_i^{\alpha}\right) = \mathrm{Cov}_{\alpha}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{C}(\beta + \varepsilon_i^{\alpha})\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\mathbf{C}\mathrm{Cov}_{\alpha}(\varepsilon_i^{\alpha}\varepsilon_i^{\alpha^t})\mathbf{C}^t + \frac{1}{N^2}\sum_{i}^{N}\sum_{i'\neq i}^{N}\mathbf{C}\mathrm{Cov}_{\alpha}(\varepsilon_i^{\alpha}\varepsilon_{i'}^{\alpha^t})\mathbf{C}^t \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\mathbf{C\Sigma}_i\mathbf{C}^t, \quad \mathbf{QED}.
\end{aligned}$$

3.9.4 Proof of formula (3.28)

In this appendix it is proved that

$$\mathrm{E}_{\alpha}\mathrm{Cov}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s) = \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\mathrm{Var}(\hat{N}) + \frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{1}{\pi_i} - 1\right)\mathbf{C\Sigma}_i\mathbf{C}^t.$$

Using (3.94), (3.96) and the measurement error model assumptions (3.2) and (3.3), it follows that

$$\begin{aligned}
&\mathrm{E}_{\alpha}\mathrm{Cov}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s) = \mathrm{E}_{\alpha}\mathrm{Cov}_s(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}}^{\alpha} \mid \alpha) \\
&= \mathrm{E}_{\alpha}\left(\mathrm{Cov}_s\left(\frac{1}{N}\sum_{i=1}^{n}\frac{\mathbf{Cy}_i^{\alpha}}{\pi_i}\right) \mid \alpha\right) \\
&= \mathrm{E}_{\alpha}\left(\mathrm{Cov}_s\left(\frac{1}{N}\sum_{i=1}^{n}\frac{\mathbf{C}(\beta + \varepsilon_i^{\alpha})}{\pi_i}\right) \mid \alpha\right) \\
&= \mathrm{E}_{\alpha}\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{\mathbf{C}(\beta + \varepsilon_i^{\alpha})(\beta + \varepsilon_{i'}^{\alpha})^t\mathbf{C}^t}{\pi_i\pi_{i'}}\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{\mathbf{C}\beta\beta^t\mathbf{C}^t}{\pi_i\pi_{i'}} + \frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{1}{\pi_i} - 1\right)\mathbf{C\Sigma}_i\mathbf{C}^t \\
&= \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\mathrm{Var}(\hat{N}) + \frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{1}{\pi_i} - 1\right)\mathbf{C\Sigma}_i\mathbf{C}^t, \quad \mathbf{QED}.
\end{aligned}$$

Note that the cross products between β and ε_i^α cancels out, due to measurement error model assumption (3.2). Also note that the cross products between ε_i^α and $\varepsilon_{i'}^\alpha$ cancels out since measurement errors between individuals are assumed to independent (model assumption (3.3)).

3.9.5 Proof of formula (3.31)

In this appendix it is proved that

$$\mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\mathbf{Y}}_{\mathbf{s}_k}^\alpha \mid \alpha, s) = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t + \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t.$$

The proof is given for an RBD. Results for a CRD follow directly as a special case of an RBD. Using properties (3.90) and (3.93) derived in appendix 3.9.2, we can elaborate on the diagonal elements of $\mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon \left(\hat{\mathbf{Y}}_{\mathbf{s}_k}^\alpha \mid \alpha, s\right)$ as follows

$$\begin{aligned}
\mathrm{Var}_\varepsilon\left(\hat{Y}_k^\alpha \mid \alpha, s\right) &= \mathrm{Cov}_\varepsilon\left(\sum_{i=1}^n \frac{\mathbf{p}_{ik}^t \mathbf{y}_i^\alpha}{\pi_i N}, \sum_{i=1}^n \frac{\mathbf{p}_{ik}^t \mathbf{y}_i^\alpha}{\pi_i N} \mid \alpha, s\right) \\
&= \sum_{i=1}^n \frac{\mathbf{y}_i^{\alpha t}}{\pi_i N} \mathrm{Cov}_\varepsilon\left(\mathbf{p}_{ik}, \mathbf{p}_{ik}^t \mid \alpha, s\right) \frac{\mathbf{y}_i^\alpha}{\pi_i N} + \sum_i^n \sum_{i' \neq i}^n \frac{\mathbf{y}_i^{\alpha t}}{\pi_i N} \mathrm{Cov}_\varepsilon\left(\mathbf{p}_{ik}, \mathbf{p}_{i'k}^t \mid \alpha, s\right) \frac{\mathbf{y}_{i'}^\alpha}{\pi_{i'} N} \\
&= \sum_{j=1}^J \left(\sum_{i=1}^{n_j} \frac{\mathbf{y}_i^{\alpha t}}{\pi_i N} \mathrm{Cov}_\varepsilon\left(\mathbf{p}_{ik}, \mathbf{p}_{ik}^t \mid \alpha, s\right) \frac{\mathbf{y}_i^\alpha}{\pi_i N} + \sum_i^{n_j} \sum_{i' \neq i}^{n_j} \frac{\mathbf{y}_i^{\alpha t}}{\pi_i N} \mathrm{Cov}_\varepsilon\left(\mathbf{p}_{ik}, \mathbf{p}_{i'k}^t \mid \alpha, s\right) \frac{\mathbf{y}_{i'}^\alpha}{\pi_{i'} N}\right) \\
&= \sum_{j=1}^J \left(\frac{(n_j - n_{jk})}{n_{jk}} \sum_{i=1}^{n_j} \frac{y_{ik}^{\alpha 2}}{\pi_i^2 N^2} - \frac{(n_j - n_{jk})}{n_{jk}} \frac{1}{(n_j - 1)} \sum_i^{n_j} \sum_{i' \neq i}^{n_j} \frac{y_{ik}^\alpha y_{i'k}^\alpha}{\pi_i N^2 \pi_{i'}}\right) \\
&= \sum_{j=1}^J \left(\frac{n_j}{(n_j - 1)} \frac{(n_j - n_{jk})}{n_{jk}} \sum_{i=1}^{n_j} \frac{y_{ik}^{\alpha 2}}{\pi_i^2 N^2} - \frac{(n_j - n_{jk})}{n_{jk}} \frac{1}{(n_j - 1)} \left(\sum_{i=1}^{n_j} \frac{y_{ik}^\alpha}{\pi_i N}\right)^2\right) \\
&= \sum_{j=1}^J \left(\frac{n_j}{(n_j - 1)} \frac{n_j}{n_{jk}} \sum_{i=1}^{n_j} \left(\frac{y_{ik}^\alpha}{\pi_i N} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{y_{i'k}^\alpha}{\pi_{i'} N}\right)^2 - \frac{n_j}{(n_j - 1)} \sum_{i=1}^{n_j} \left(\frac{y_{ik}^\alpha}{\pi_i N} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{y_{i'k}^\alpha}{\pi_{i'} N}\right)^2\right). \quad (3.97)
\end{aligned}$$

A CRD can be considered as a special case of an RBD, where $J = 1$, $n_j = n$ and $n_{jk} = n_k$. Therefore, it follows from (3.97) that the diagonal elements of $\mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon(\hat{\mathbf{Y}}_{\mathbf{s}_k}^\alpha \mid \alpha, s)$ for a CRD are given by

$$\mathrm{Var}_\varepsilon\left(\hat{Y}_k^\alpha \mid \alpha, s\right) = \frac{n}{(n-1)} \frac{n}{n_k} \sum_{i=1}^n \left(\frac{y_{ik}^\alpha}{\pi_i N} - \frac{1}{n} \sum_{i'=1}^n \frac{y_{i'k}^\alpha}{\pi_{i'} N}\right)^2 - \frac{n}{(n-1)} \sum_{i=1}^n \left(\frac{y_{ik}^\alpha}{\pi_i N} - \frac{1}{n} \sum_{i'=1}^n \frac{y_{i'k}^\alpha}{\pi_{i'} N}\right)^2. \quad (3.98)$$

Using (3.91) and (3.92) derived in appendix 3.9.2 we can elaborate on the off-diagonal elements of $\mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon \left(\hat{\mathbf{Y}}_{\mathbf{s}_k}^\alpha \mid \alpha, s\right)$ for an RBD as follows

$$\begin{aligned}
\mathrm{Cov}_\varepsilon\left(\hat{Y}_k^\alpha, \hat{Y}_{k'}^\alpha \mid \alpha, s\right) &= \mathrm{Cov}_\varepsilon\left(\sum_{i=1}^n \frac{\mathbf{p}_{ik}^t \mathbf{y}_i^\alpha}{\pi_i N}, \sum_{i=1}^n \frac{\mathbf{p}_{ik'}^t \mathbf{y}_i^\alpha}{\pi_i N} \mid \alpha, s\right) \\
&= \sum_{i=1}^n \frac{\mathbf{y}_i^{\alpha t}}{\pi_i N} \mathrm{Cov}_\varepsilon\left(\mathbf{p}_{ik}, \mathbf{p}_{ik'}^t \mid \alpha, s\right) \frac{\mathbf{y}_i^\alpha}{\pi_i N} + \sum_i^n \sum_{i' \neq i}^n \frac{\mathbf{y}_i^{\alpha t}}{\pi_i N} \mathrm{Cov}_\varepsilon\left(\mathbf{p}_{ik}, \mathbf{p}_{i'k'}^t \mid \alpha, s\right) \frac{\mathbf{y}_{i'}^\alpha}{\pi_{i'} N} \\
&= \sum_{j=1}^J \left(\sum_{i=1}^{n_j} \frac{\mathbf{y}_i^{\alpha t}}{\pi_i N} \mathrm{Cov}_\varepsilon\left(\mathbf{p}_{ik}, \mathbf{p}_{ik'}^t \mid \alpha, s\right) \frac{\mathbf{y}_i^\alpha}{\pi_i N} + \sum_i^{n_j} \sum_{i' \neq i}^{n_j} \frac{\mathbf{y}_i^{\alpha t}}{\pi_i N} \mathrm{Cov}_\varepsilon\left(\mathbf{p}_{ik}, \mathbf{p}_{i'k'}^t \mid \alpha, s\right) \frac{\mathbf{y}_{i'}^\alpha}{\pi_{i'} N}\right)
\end{aligned}$$

$$= \sum_{j=1}^{J}\left(-\sum_{i=1}^{n_j}\frac{y_{ik}^{\alpha}y_{ik'}^{\alpha}}{\pi_i^2N^2}+\frac{1}{(n_j-1)}\sum_{i=1}^{n_j}\sum_{i'\neq i=1}^{n_j}\frac{y_{ik}^{\alpha}y_{i'k'}^{\alpha}}{\pi_iN^2\pi_{i'}}\right)$$
$$= \sum_{j=1}^{J}\left(-\frac{n_j}{(n_j-1)}\sum_{i=1}^{n_j}\frac{y_{ik}^{\alpha}y_{ik'}^{\alpha}}{\pi_i^2N^2}+\frac{1}{(n_j-1)}\sum_{i=1}^{n_j}\sum_{i'=1}^{n_j}\frac{y_{ik}^{\alpha}y_{i'k'}^{\alpha}}{\pi_iN^2\pi_{i'}}\right)$$
$$= -\sum_{j=1}^{J}\frac{n_j}{(n_j-1)}\sum_{i=1}^{n_j}\left(\frac{y_{ik}^{\alpha}}{\pi_iN}-\frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{y_{i'k}^{\alpha}}{\pi_{i'}N}\right)\left(\frac{y_{ik'}^{\alpha}}{\pi_iN}-\frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{y_{i'k'}^{\alpha}}{\pi_{i'}N}\right). \tag{3.99}$$

An expression for the off-diagonal elements of $\mathrm{E}_\alpha\mathrm{E}_s\mathrm{Cov}_\varepsilon(\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s)$ for a CRD follows directly from (3.99) with $J = 1$, $n_j = n$ and $n_{jk} = n_k$ and is given by:

$$\mathrm{Cov}_\varepsilon\left(\hat{\hat{Y}}_k^{\alpha},\hat{\hat{Y}}_{k'}^{\alpha} \mid \alpha, s\right) = -\frac{n}{(n-1)}\sum_{i=1}^{n}\left(\frac{y_{ik}^{\alpha}}{\pi_iN}-\frac{1}{n}\sum_{i'=1}^{n}\frac{y_{i'k}^{\alpha}}{\pi_{i'}N}\right)\left(\frac{y_{ik'}^{\alpha}}{\pi_iN}-\frac{1}{n}\sum_{i'=1}^{n}\frac{y_{i'k'}^{\alpha}}{\pi_{i'}N}\right). \tag{3.100}$$

For a CRD, the results (3.98) and (3.100) can be written in matrix notation as follows:

$$\mathrm{E}_\alpha\mathrm{E}_s\mathrm{Cov}_\varepsilon\left(\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s\right) = \mathrm{E}_\alpha\mathrm{E}_s\mathbf{D} + \mathrm{E}_\alpha\mathrm{E}_s\mathbf{A},$$

where $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_K)$ is defined in (3.32) and $\mathbf{A}$ defined in (3.33). For an RBD, the results (3.97) and (3.99) can be written in matrix notation as follows:

$$\mathrm{E}_\alpha\mathrm{E}_s\mathrm{Cov}_\varepsilon\left(\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s\right) = \mathrm{E}_\alpha\mathrm{E}_s\mathbf{D} + \mathrm{E}_\alpha\mathrm{E}_s\mathbf{A},$$

where the elements d_k of the diagonal matrix $\mathbf{D}$ are defined in (3.34) and $\mathbf{A}$ is defined in (3.35). Consequently we have for a CRD as well as an RBD

$$\mathrm{E}_\alpha\mathrm{E}_s\mathrm{Cov}_\varepsilon\left(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s\right) = \mathrm{E}_\alpha\mathrm{E}_s\mathbf{CDC}^t + \mathrm{E}_\alpha\mathrm{E}_s\mathbf{CAC}^t, \quad \textbf{QED}.$$

3.9.6 Proof of formula (3.37), (3.42) and (3.54)

In this appendix expressions for $\mathrm{E}_\alpha\mathrm{E}_s\mathbf{CAC}^t$ for RBD's are derived. Results for CRD's follow as a special case.

Proof of formula (3.54) for an RBD

From (3.35) and (3.96) it follows that

$$\mathbf{CAC}^t = -\sum_{j=1}^{J}\frac{n_j}{n_j-1}\sum_{i=1}^{n_j}\left(\frac{\mathbf{C}(\beta+\varepsilon_i^{\alpha})}{N\pi_i}-\frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{\mathbf{C}(\beta+\varepsilon_{i'}^{\alpha})}{N\pi_{i'}}\right)\left(\frac{\mathbf{C}(\beta+\varepsilon_i^{\alpha})}{N\pi_i}-\frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{\mathbf{C}(\beta+\varepsilon_{i'}^{\alpha})}{N\pi_{i'}}\right)^t, \tag{3.101}$$

Note that the sum of squares of the cross products between β and ε_i^{α} is equal to zero, since $\mathrm{E}_\alpha(\beta{\varepsilon_i^{\alpha}}^t) = \mathbf{O}$. Now we concentrate on the sum of squares of ε_i^{α}, i.e.

$$-\mathrm{E}_\alpha\mathrm{E}_s\sum_{j=1}^{J}\frac{n_j}{n_j-1}\sum_{i=1}^{n_j}\left(\frac{\mathbf{C}\varepsilon_i^{\alpha}}{N\pi_i}-\frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{\mathbf{C}\varepsilon_{i'}^{\alpha}}{N\pi_{i'}}\right)\left(\frac{\mathbf{C}\varepsilon_i^{\alpha}}{N\pi_i}-\frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{\mathbf{C}\varepsilon_{i'}^{\alpha}}{N\pi_{i'}}\right)^t. \tag{3.102}$$

First we give the proof for RBD's where interviewers are block variables. In this situation the block variables should be considered as domains since the sampling design doesn't determine to which block an individual belongs. Since the block size n_j is random, we condition on the realization of n_j. Let a_{ij} denote the membership indicator of individual i in block s_j defined by (3.49). Furthermore, let

$$\bar{\Delta}_{a_j} = \sum_{i=1}^{N} \frac{a_{ij}\varepsilon_i^{\alpha}}{N}$$

denote the population mean of the measurement errors of the individuals of block j. Then

$$\hat{\bar{\Delta}}_{a_j} = \sum_{i=1}^{n} \frac{a_{ij}\varepsilon_i^{\alpha}}{N\pi_i}$$

denotes the Horvitz-Thompson estimator for $\bar{\Delta}_{a_j}$. We need the following intermediate results:

$$\begin{aligned}
&\mathrm{E}_s \sum_{i=1}^{n_j} \left(\frac{\varepsilon_i^{\alpha}}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\varepsilon_{i'}^{\alpha}}{N\pi_{i'}} \right) \left(\frac{\varepsilon_i^{\alpha}}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\varepsilon_{i'}^{\alpha}}{N\pi_{i'}} \right)^t \\
&= \frac{1}{n_j^2} \mathrm{E}_s \sum_{i=1}^{n_j} \left(\frac{n_j \varepsilon_i^{\alpha}}{N\pi_i} - \bar{\Delta}_{a_j} + \bar{\Delta}_{a_j} - \hat{\bar{\Delta}}_{a_j} \right) \left(\frac{n_j \varepsilon_i^{\alpha}}{N\pi_i} - \bar{\Delta}_{a_j} + \bar{\Delta}_{a_j} - \hat{\bar{\Delta}}_{a_j} \right)^t \\
&= \frac{1}{n_j^2} \mathrm{E}_s \sum_{i=1}^{n_j} \left(\frac{n_j \varepsilon_i^{\alpha}}{N\pi_i} - \bar{\Delta}_{a_j} \right) \left(\frac{n_j \varepsilon_i^{\alpha}}{N\pi_i} - \bar{\Delta}_{a_j} \right)^t - \frac{1}{n_j} \mathrm{E}_s \left(\hat{\bar{\Delta}}_{a_j} - \bar{\Delta}_{a_j} \right) \left(\hat{\bar{\Delta}}_{a_j} - \bar{\Delta}_{a_j} \right)^t . \qquad (3.103)
\end{aligned}$$

Note that the second term of (3.103) can be recognized as the design variance of $\bar{\Delta}_{a_j}$ divided by n_j, denoted by $(1/n_j)\mathrm{Cov}(\hat{\bar{\Delta}}_{a_j})$. The first term of (3.103) can be further elaborated as

$$\begin{aligned}
&\frac{1}{n_j^2} \mathrm{E}_s \sum_{i=1}^{n_j} \left(\frac{n_j \varepsilon_i^{\alpha}}{N\pi_i} - \bar{\Delta}_{a_j} \right) \left(\frac{n_j \varepsilon_i^{\alpha}}{N\pi_i} - \bar{\Delta}_{a_j} \right)^t \\
&= \frac{1}{n_j^2} \left(\mathrm{E}_s \sum_{i=1}^{n} \frac{n_j^2 a_{ij} \varepsilon_i^{\alpha} \varepsilon_i^{\alpha^t}}{N^2 \pi_i^2} - \bar{\Delta}_{a_j} \mathrm{E}_s \sum_{i=1}^{n} \frac{n_j a_{ij} \varepsilon_i^{\alpha^t}}{N\pi_i} - \mathrm{E}_s \sum_{i=1}^{n} \frac{n_j a_{ij} \varepsilon_i^{\alpha}}{N\pi_i} \bar{\Delta}_{a_j}^t + n_j \bar{\Delta}_{a_j} \bar{\Delta}_{a_j}^t \right) \\
&= \frac{1}{n_j^2} \left(\sum_{i=1}^{N} \frac{n_j^2 a_{ij} \varepsilon_i^{\alpha} \varepsilon_i^{\alpha^t}}{N^2 \pi_i} - \bar{\Delta}_{a_j} \sum_{i=1}^{N} \frac{n_j a_{ij} \varepsilon_i^{\alpha^t}}{N} - \sum_{i=1}^{N} \frac{n_j a_{ij} \varepsilon_i^{\alpha}}{N} \bar{\Delta}_{a_j}^t + n_j \bar{\Delta}_{a_j} \bar{\Delta}_{a_j}^t \right) \\
&= \frac{1}{n_j^2} \left(\sum_{i=1}^{N} \frac{n_j^2 a_{ij} \varepsilon_i^{\alpha} \varepsilon_i^{\alpha^t}}{N^2 \pi_i} - n_j \bar{\Delta}_{a_j} \bar{\Delta}_{a_j}^t \right) \\
&= \frac{1}{n_j} \left(\sum_{i=1}^{N} \frac{n_j a_{ij} \varepsilon_i^{\alpha} \varepsilon_i^{\alpha^t}}{N^2 \pi_i} - \bar{\Delta}_{a_j} \bar{\Delta}_{a_j}^t \right) \equiv \tilde{\mathrm{Cov}}(\hat{\bar{\Delta}}_{a_j}). \qquad (3.104)
\end{aligned}$$

Note that $\tilde{\mathrm{Cov}}(\hat{\bar{\Delta}}_{a_j})$ can be recognized as the design variance of $\hat{\bar{\Delta}}_{a_j}$ as if the individuals are drawn with replacement with selection probabilities π_i/n_j. Taking the expectation of $(1/n_j)\mathrm{Cov}(\hat{\bar{\Delta}}_{a_j})$ with respect to the measurement error model and using model assumption (3.3) gives:

$$\begin{aligned}
\frac{1}{n_j} \mathrm{E}_{\alpha} \mathrm{E}_s \left(\hat{\bar{\Delta}}_{a_j} - \bar{\Delta}_{a_j} \right) \left(\hat{\bar{\Delta}}_{a_j} - \bar{\Delta}_{a_j} \right)^t &= \frac{1}{n_j} \mathrm{E}_{\alpha} \sum_{i=1}^{N} \sum_{i'=1}^{N} \frac{\pi_{ii'} - \pi_i \pi_{i'}}{\pi_i \pi_{i'}} \frac{a_{ij} \varepsilon_i^{\alpha}}{N} \frac{a_{i'j} \varepsilon_{i'}^{\alpha^t}}{N} \\
&= \frac{1}{N^2} \frac{1}{n_j} \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1 \right) a_{ij} \mathbf{\Sigma}_i. \qquad (3.105)
\end{aligned}$$

Taking the expectation of $\tilde{\text{Cov}}(\hat{\bar{\Delta}}_{a_j})$ with respect to the measurement error model and using model assumption (3.3) gives:

$$\begin{aligned}
\frac{1}{n_j}\text{E}_\alpha\left(\sum_{i=1}^{N}\frac{n_j a_{ij}\varepsilon_i^\alpha {\varepsilon_i^\alpha}^t}{N^2\pi_i} - \bar{\Delta}_{a_j}\bar{\Delta}_{a_j}^t\right) &= \frac{1}{n_j}\left(\sum_{i=1}^{N}\frac{n_j a_{ij}\boldsymbol{\Sigma}_i}{N^2\pi_i} - \sum_{i=1}^{N}\frac{a_{ij}\boldsymbol{\Sigma}_i}{N^2}\right) \\
&= \frac{1}{N^2}\frac{1}{n_j}\sum_{i=1}^{N}\left(\frac{n_j}{\pi_i} - 1\right) a_{ij}\boldsymbol{\Sigma}_i. \qquad (3.106)
\end{aligned}$$

Result (3.105) and (3.106) gives

$$\tilde{\text{Cov}}(\hat{\bar{\Delta}}_{a_j}) - \frac{1}{n_j}\text{Cov}(\hat{\bar{\Delta}}_{a_j}) = \frac{1}{N^2}\frac{1}{n_j}\sum_{i=1}^{N}\left(\frac{n_j}{\pi_i} - \frac{1}{\pi_i}\right) a_{ij}\boldsymbol{\Sigma}_i. \qquad (3.107)$$

With result (3.107) we can elaborate on formula (3.102) as

$$\begin{aligned}
&-\text{E}_\alpha\text{E}_s\sum_{j=1}^{J}\frac{n_j}{n_j-1}\sum_{i=1}^{n_j}\left(\frac{\mathbf{C}\varepsilon_i^\alpha}{N\pi_i} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{\mathbf{C}\varepsilon_{i'}^\alpha}{N\pi_{i'}}\right)\left(\frac{\mathbf{C}\varepsilon_i^\alpha}{N\pi_i} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{\mathbf{C}\varepsilon_{i'}^\alpha}{N\pi_{i'}}\right)^t \\
&= -\sum_{j=1}^{J}\frac{n_j}{n_j-1}\frac{1}{N^2}\frac{1}{n_j}\sum_{i=1}^{N}\left(\frac{n_j}{\pi_i} - \frac{1}{\pi_i}\right) a_{ij}\mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t \\
&= -\sum_{j=1}^{J}\frac{1}{N^2}\sum_{i=1}^{N}\frac{a_{ij}\mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t}{\pi_i} \\
&= -\frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t}{\pi_i}. \qquad (3.108)
\end{aligned}$$

Finally we have to elaborate on the sum of squares of β. Let

$$N_{a_j} = \sum_{i=1}^{N} a_{ij}$$

denote the block size in the finite population and

$$\hat{N}_{a_j} = \sum_{i=1}^{n}\frac{a_{ij}}{\pi_i}$$

the Horvitz-Thompson estimator for N_{a_j}. Equivalent to the derivation of (3.103) and (3.104) it can be shown that

$$\text{E}_s\sum_{i=1}^{n_j}\left(\frac{\beta}{N\pi_i} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{\beta}{N\pi_{i'}}\right)\left(\frac{\beta}{N\pi_i} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{\beta}{N\pi_{i'}}\right)^t = \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2}\left(\tilde{\text{Var}}(\hat{N}_{a_j}) - \frac{1}{n_j}\text{Var}(\hat{N}_{a_j})\right),$$

where

$$\tilde{\text{Var}}(\hat{N}_{a_j}) = \frac{1}{n_j}\left(\sum_{i=1}^{N}\frac{n_j a_{ij}}{\pi_i} - N_j^2\right),$$

and

$$\text{Var}(\hat{N}_{a_j}) = \sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{a_{ij}\ a_{i'j}}{\pi_i\ \pi_{i'}}.$$

Since β is fixed with respect to the measurement error model it follows that

$$\begin{aligned}
&\mathrm{E}_\alpha \mathrm{E}_s - \sum_{j=1}^{J} \frac{n_j}{n_j-1} \sum_{i=1}^{n_j} \left(\frac{\mathbf{C}\beta}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\beta}{N\pi_{i'}} \right) \left(\frac{\mathbf{C}\beta}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\beta}{N\pi_{i'}} \right)^t \\
&= - \frac{\mathbf{C}\beta \left(\mathbf{C}\beta\right)^t}{N^2} \sum_{j=1}^{J} \frac{n_j}{(n_j-1)} \left(\tilde{\mathrm{Var}}(\hat{N}_{a_j}) - \frac{1}{n_j} \mathrm{Var}(\hat{N}_{a_j}) \right). \qquad (3.109)
\end{aligned}$$

For an RBD with interviewers as block variables, it follows from (3.108) and (3.109) that

$$\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t = - \frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{C\Sigma}_i \mathbf{C}^t}{\pi_i} - \frac{\mathbf{C}\beta \left(\mathbf{C}\beta\right)^t}{N^2} \sum_{j=1}^{J} \frac{n_j}{(n_j-1)} \left(\tilde{\mathrm{Var}}(\hat{N}_{a_j}) - \frac{1}{n_j} \mathrm{Var}(\hat{N}_{a_j}) \right), \quad \mathbf{QED}.$$

Proof of formula (3.42)

Now we consider RBD's where PSU's are block variables. In a two-stage sampling scheme, J blocks or PSU's are drawn from a finite population of J_u blocks with first-order selection probabilities π_j. Within each PSU, n_j SSU's are drawn in the second stage with first and second-order inclusion probabilities $\pi_{i|j}$ and $\pi_{ii'|j}$. The first-order inclusion probabilities of the individuals in the sample are $\pi_i = \pi_j \pi_{i|j}$. Furthermore, let

$$\bar{\Delta}_j = \sum_{i=1}^{N_j} \frac{\varepsilon_i^\alpha}{N_j}$$

denote the population mean of the measurement errors of the individuals of block j. Then

$$\hat{\bar{\Delta}}_j = \sum_{i=1}^{n_j} \frac{\varepsilon_i^\alpha}{N_j \pi_{i|j}}$$

denotes the Horvitz-Thompson estimator for $\bar{\Delta}_j$. Let E_{s_I} denote the expectation with respect to the fist stage of the sampling design and $\mathrm{E}_{s_{II}}$ the expectation with respect to the second stage of the sampling design. Now we have

$$\begin{aligned}
&\mathrm{E}_s \sum_{j=1}^{J} \sum_{i=1}^{n_j} \left(\frac{\varepsilon_i^\alpha}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\varepsilon_{i'}^\alpha}{N\pi_{i'}} \right) \left(\frac{\varepsilon_i^\alpha}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\varepsilon_{i'}^\alpha}{N\pi_{i'}} \right)^t \\
&= \mathrm{E}_{s_I} \mathrm{E}_{s_{II}} \sum_{j=1}^{J} \frac{1}{n_j^2 \, \pi_j^2} \left(\frac{N_j}{N} \right)^2 \sum_{i=1}^{n_j} \left(\frac{n_j \varepsilon_i^\alpha}{N_j \pi_{i|j}} - \hat{\bar{\Delta}}_j + \bar{\Delta}_j - \bar{\Delta}_j \right) \left(\frac{n_j \varepsilon_i^\alpha}{N_j \pi_{i|j}} - \hat{\bar{\Delta}}_j + \bar{\Delta}_j - \bar{\Delta}_j \right)^t \\
&= \mathrm{E}_{s_I} \sum_{j=1}^{J} \frac{1}{\pi_j^2} \left(\frac{N_j}{N} \right)^2 \left(\frac{1}{n_j^2} \mathrm{E}_{s_{II}} \sum_{i=1}^{n_j} \left(\frac{n_j \varepsilon_i^\alpha}{N_j \pi_{i|j}} - \bar{\Delta}_j \right) \left(\frac{n_j \varepsilon_i^\alpha}{N_j \pi_{i|j}} - \bar{\Delta}_j \right)^t \right. \\
&\qquad \left. - \frac{1}{n_j} \mathrm{E}_{s_{II}} \left(\hat{\bar{\Delta}}_j - \bar{\Delta}_j \right) \left(\hat{\bar{\Delta}}_j - \bar{\Delta}_j \right)^t \right) \\
&= \mathrm{E}_{s_I} \sum_{j=1}^{J} \left(\frac{N_j}{N} \right)^2 \frac{1}{\pi_j^2} \left(\frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j \varepsilon_i^\alpha \varepsilon_i^{\alpha t}}{N_j^2 \pi_{i|j}} - \bar{\Delta}_j \bar{\Delta}_j^t \right) - \frac{1}{n_j} \, \mathrm{Cov}(\hat{\bar{\Delta}}_j) \right) \\
&= \sum_{j=1}^{J_u} \frac{1}{\pi_j} \left(\frac{N_j}{N} \right)^2 \left(\tilde{\mathrm{Cov}}(\hat{\bar{\Delta}}_j) - \frac{1}{n_j} \mathrm{Cov}(\hat{\bar{\Delta}}_j) \right). \qquad (3.110)
\end{aligned}$$

Taking the expectation of $\text{Cov}(\hat{\bar{\Delta}}_j)$ with respect to the measurement error model gives

$$\begin{aligned} \text{E}_\alpha \text{Cov}(\hat{\bar{\Delta}}_j) &= \text{E}_\alpha \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j}) \frac{\varepsilon_i^\alpha \ \varepsilon_{i'}^{\alpha^t}}{\pi_{i|j}\pi_{i'|j}} \\ &= \frac{1}{N_j^2} \sum_{i=1}^{N_j} \left(\frac{1}{\pi_{i|j}} - 1 \right) \mathbf{\Sigma}_i. \end{aligned} \tag{3.111}$$

Taking the expectation of $\tilde{\text{Cov}}(\hat{\bar{\Delta}}_j)$ with respect to the measurement error model gives

$$\begin{aligned} \text{E}_\alpha \tilde{\text{Cov}}(\hat{\bar{\Delta}}_j) &= \text{E}_\alpha \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j \varepsilon_i^\alpha \varepsilon_i^{\alpha^t}}{N_j^2 \pi_{i|j}} - \bar{\Delta}_j \bar{\Delta}_j^t \right) \\ &= \frac{1}{n_j \ N_j^2} \sum_{i=1}^{N_j} \left(\frac{n_j}{\pi_{i|j}} - 1 \right) \mathbf{\Sigma}_i. \end{aligned} \tag{3.112}$$

With results (3.110), (3.111) and (3.112) we can elaborate on formula (3.102) as

$$\begin{aligned} &-\text{E}_\alpha \text{E}_s \sum_{j=1}^{J} \frac{n_j}{n_j - 1} \sum_{i=1}^{n_j} \left(\frac{\mathbf{C}\varepsilon_i^\alpha}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\varepsilon_{i'}^\alpha}{N\pi_{i'}} \right) \left(\frac{\mathbf{C}\varepsilon_i^\alpha}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\varepsilon_{i'}^\alpha}{N\pi_{i'}} \right)^t \\ &= -\text{E}_\alpha \sum_{j=1}^{J_u} \frac{1}{\pi_j} \left(\frac{N_j}{N} \right)^2 \frac{n_j}{(n_j - 1)} \mathbf{C} \left(\tilde{\text{Cov}}(\hat{\bar{\Delta}}_j) - \frac{1}{n_j} \text{Cov}(\hat{\bar{\Delta}}_j) \right) \mathbf{C}^t \\ &= -\sum_{j=1}^{J_u} \frac{1}{\pi_j} \frac{1}{N^2} \sum_{i=1}^{N_j} \frac{\mathbf{C}\mathbf{\Sigma}_i \mathbf{C}^t}{\pi_{i|j}} \\ &= -\frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{C}\mathbf{\Sigma}_i \mathbf{C}^t}{\pi_i}. \end{aligned} \tag{3.113}$$

In order to elaborate on the sum of squares of β, let

$$\hat{N}_j = \sum_{i=1}^{n_j} \frac{1}{\pi_{i|j}} \tag{3.114}$$

denote the Horvitz-Thompson estimator for the size of block or PSU j. According to the derivation of (3.110) it follows that

$$\begin{aligned} &\text{E}_s \sum_{j=1}^{J} \sum_{i=1}^{n_j} \left(\frac{\beta}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\beta}{N\pi_{i'}} \right) \left(\frac{\beta}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\beta}{N\pi_{i'}} \right)^t \\ &= \frac{\mathbf{C}\beta \, (\mathbf{C}\beta)^t}{N^2} \sum_{j=1}^{J_u} \frac{1}{\pi_j} \left(\tilde{\text{Var}}(\hat{N}_j) - \frac{1}{n_j} \text{Var}(\hat{N}_j) \right), \end{aligned} \tag{3.115}$$

where

$$\text{Var}(\hat{N}_j) = \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} \frac{(\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j})}{\pi_{i|j} \ \pi_{i'|j}},$$

and

$$\tilde{\text{Var}}(\hat{N}_j) = \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j}{\pi_{i|j}} - N_j^2 \right). \tag{3.116}$$

Since β is a constant with respect to the measurement error model and from result (3.115) it follows that

$$-\mathrm{E}_\alpha \mathrm{E}_s \sum_{j=1}^{J} \frac{n_j}{n_j-1} \sum_{i=1}^{n_j} \left(\frac{\mathbf{C}\beta}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\beta}{N\pi_{i'}} \right) \left(\frac{\mathbf{C}\beta}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\beta}{N\pi_{i'}} \right)^t$$
$$= - \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2} \sum_{j=1}^{J_u} \frac{1}{\pi_j} \frac{n_j}{(n_j-1)} \left(\tilde{\mathrm{Var}}(\hat{N}_j) - \frac{1}{n_j} \mathrm{Var}(\hat{N}_j) \right). \tag{3.117}$$

Finally it follows from (3.113) and (3.117) for an RBD with PSU's as block variables that

$$\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t = -\frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{C}\mathbf{\Sigma}_i \mathbf{C}^t}{\pi_i} - \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2} \sum_{j=1}^{J_u} \frac{1}{\pi_j} \frac{n_j}{(n_j-1)} \left(\tilde{\mathrm{Var}}(\hat{N}_j) - \frac{1}{n_j} \mathrm{Var}(\hat{N}_j) \right), \quad \mathbf{QED}.$$

Proof of formula (3.37) for a CRD

From (3.33) and (3.96) it follows that

$$\mathbf{CAC}^t = -\frac{n}{n-1} \sum_{i=1}^{n} \left(\frac{\mathbf{C}(\beta+\varepsilon_i^\alpha)}{N\pi_i} - \frac{1}{n} \sum_{i'=1}^{n} \frac{\mathbf{C}(\beta+\varepsilon_{i'}^\alpha)}{N\pi_{i'}} \right) \left(\frac{\mathbf{C}(\beta+\varepsilon_i^\alpha)}{N\pi_i} - \frac{1}{n} \sum_{i'=1}^{n} \frac{\mathbf{C}(\beta+\varepsilon_{i'}^\alpha)}{N\pi_{i'}} \right)^t. \tag{3.118}$$

Let

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{\pi_i}$$

denote the Horvitz-Thompson estimator for the population size and

$$\tilde{\mathrm{Var}}(\hat{N}) = \frac{1}{n} \left(\sum_{i=1}^{N} \frac{n}{\pi_i} - N^2 \right),$$

and

$$\mathrm{Var}(\hat{N}) = \sum_{i=1}^{N} \sum_{i'=1}^{N} \frac{(\pi_{ii'} - \pi_i \pi_{i'})}{\pi_i\ \pi_{i'}},$$

the variances of $\hat{N}$. The expectation of (3.118) with respect to the sampling design and the measurement error model is given by

$$\mathrm{E}_\alpha \mathrm{E}_s \mathbf{CAC}^t = -\frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{C}\mathbf{\Sigma}_i \mathbf{C}^t}{\pi_i} - \frac{n}{(n-1)} \frac{\mathbf{C}\beta\,(\mathbf{C}\beta)^t}{N^2} \left(\tilde{\mathrm{Var}}(\hat{N}) - \frac{1}{n} \mathrm{Var}(\hat{N}) \right), \quad \mathbf{QED}.$$

The proof of this result for a CRD follows analogous to the proof for an RBD where PSU's are block variables, with $\pi_j = 1$, $\pi_{i|j} = \pi_i$, $\pi_{ii'|j} = \pi_{ii'}$, $J = 1$, $N_j = N$, $n_j = n$ and $n_{jk} = n_k$.

3.9.7 Proof of formula (3.79)

Matrix $\hat{\mathbf{D}}$ can be partitioned as follows:

$$\hat{\mathbf{D}} = \begin{pmatrix} \hat{d}_1 & \mathbf{0}^t \\ \mathbf{0} & \hat{\mathbf{D}}_* \end{pmatrix}.$$

First it is proved that

$$\left(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t\right)^{-1} = \hat{\mathbf{D}}_*^{-1} - \frac{1}{\text{trace}(\hat{\mathbf{D}}_*^{-1})}\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}. \tag{3.119}$$

Note that

$$\left(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t\right) = \begin{pmatrix} \mathbf{j} & -\mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{d}_1 & \mathbf{0}^t \\ \mathbf{0} & \hat{\mathbf{D}}_* \end{pmatrix} \begin{pmatrix} \mathbf{j}^t \\ -\mathbf{I} \end{pmatrix} = \hat{d}_1\mathbf{j}\mathbf{j}^t + \hat{\mathbf{D}}_*.$$

According to Bartlett's identity (Morisson, 1990, Ch.2) it follows that:

$$\begin{aligned}
\left(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t\right)^{-1} &= \left(\hat{d}_1\mathbf{j}\mathbf{j}^t + \hat{\mathbf{D}}_*\right)^{-1} = \hat{\mathbf{D}}_*^{-1} - \frac{\hat{d}_1}{1+\hat{d}_1\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j}}\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1} \\
&= \hat{\mathbf{D}}_*^{-1} - \frac{1}{\text{trace}(\hat{\mathbf{D}}^{-1})}\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}.
\end{aligned}$$

Define $p = \text{trace}(\hat{\mathbf{D}}^{-1})$. Now it follows that

$$\begin{aligned}
\mathbf{C}^t\left(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t\right)^{-1}\mathbf{C} &= \mathbf{C}^t\hat{\mathbf{D}}_*^{-1}\mathbf{C} - \frac{1}{p}\mathbf{C}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{C} \\
&= \begin{pmatrix} \mathbf{j}^t \\ -\mathbf{I} \end{pmatrix}\hat{\mathbf{D}}_*^{-1}\begin{pmatrix} \mathbf{j} & -\mathbf{I} \end{pmatrix} - \frac{1}{p}\begin{pmatrix} \mathbf{j}^t \\ -\mathbf{I} \end{pmatrix}\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\begin{pmatrix} \mathbf{j} & -\mathbf{I} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j} & -\mathbf{j}^t\hat{\mathbf{D}}_*^{-1} \\ -\hat{\mathbf{D}}_*^{-1}\mathbf{j} & \hat{\mathbf{D}}_*^{-1} \end{pmatrix} - \frac{1}{p}\begin{pmatrix} \mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j} & -\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1} \\ -\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j} & \hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j} - 1/p\left(\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j}\right)^2 & \left(1-1/p\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j}\right)\left(-\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\right) \\ \left(-\hat{\mathbf{D}}_*^{-1}\mathbf{j}\right)\left(1-1/p\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\mathbf{j}\right) & \hat{\mathbf{D}}_*^{-1} - 1/p\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1} \end{pmatrix} \\
&= \begin{pmatrix} p - 1/d_1 - 1/p\,(p-1/d_1)^2 & (1-1/p\,(p-1/d_1))\left(-\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\right) \\ (1-1/p\,(p-1/d_1))\left(-\hat{\mathbf{D}}_*^{-1}\mathbf{j}\right) & \hat{\mathbf{D}}_*^{-1} - 1/p\hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1} \end{pmatrix} \\
&= \begin{pmatrix} 1/d_1 & \mathbf{0}^t \\ \mathbf{0} & \hat{\mathbf{D}}_*^{-1} \end{pmatrix} - \frac{1}{p}\begin{pmatrix} (1/d_1)^2 & 1/d_1\left(\mathbf{j}^t\hat{\mathbf{D}}_*^{-1}\right) \\ 1/d_1\left(\hat{\mathbf{D}}_*^{-1}\mathbf{j}\right) & \hat{\mathbf{D}}_*^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}_*^{-1} \end{pmatrix} \\
&= \hat{\mathbf{D}}^{-1} - \frac{1}{\text{trace}(\hat{\mathbf{D}}^{-1})}\hat{\mathbf{D}}^{-1}\mathbf{j}\mathbf{j}^t\hat{\mathbf{D}}^{-1}, \quad \textbf{QED}.
\end{aligned}$$

3.9.8 Proof of formula (3.83)

It is assumed that a limit theorem holds true such that $\mathbf{C}\hat{\bar{\mathbf{Y}}}^{\alpha} \rightarrow \mathcal{N}\left(\mathbf{C}\bar{\mathbf{Y}}, \mathbf{CVC}^t\right)$. From Searle (1971), theorem 2, Ch.2, it follows that if $\mathbf{C}\hat{\bar{\mathbf{Y}}}^{\alpha} \simeq \mathcal{N}\left(\mathbf{C}\bar{\mathbf{Y}}, \mathbf{CVC}^t\right)$ and $\mathbf{A} = (\mathbf{CVC}^t)^{-1}$, then $W = (\mathbf{C}\hat{\bar{\mathbf{Y}}}^{\alpha})^t\mathbf{A}(\mathbf{C}\hat{\bar{\mathbf{Y}}}^{\alpha}) \simeq \chi^2[\text{rang}(\mathbf{A}), 1/2(\mathbf{C}\bar{\mathbf{Y}})^t\mathbf{A}(\mathbf{C}\bar{\mathbf{Y}})]$ if and only if $\mathbf{ACVC}^t$ is idempotent, with rang($\mathbf{A}$) the number of degrees of freedom and $1/2(\mathbf{C}\bar{\mathbf{Y}})^t\mathbf{A}(\mathbf{C}\bar{\mathbf{Y}})$ the non-central parameter of the chi-squared distribution.

Since $\mathbf{ACVC}^t = \mathbf{I}$, it follows directly that $\mathbf{ACVC}^t$ is idempotent.

Because matrix $\mathbf{CVC}^t$ is nonsingular, $\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{ACVC}^t)$. Because $\mathbf{ACVC}^t$ is idempotent, $\text{rang}(\mathbf{ACVC}^t) = \text{trace}(\mathbf{ACVC}^t)$. From this it follows that $\text{rang}(\mathbf{A}) = \text{trace}(\mathbf{ACVC}^t) = \text{trace}(\mathbf{I}) = K - 1$.

Under the null hypothesis it holds true that $\mathbf{C}\bar{\mathbf{Y}} = \mathbf{0}$, and it follows that the non-central parameter $1/2(\mathbf{C}\bar{\mathbf{Y}})^t\mathbf{AC}\bar{\mathbf{Y}} = 0$.

Finally, if there is a limit theorem such that $\mathbf{C}\hat{\bar{\mathbf{Y}}}^{\alpha} \to \mathcal{N}\left(\mathbf{C}\bar{\mathbf{Y}}, \mathbf{CVC}^t\right)$, then it follows under the null hypothesis for the Wald statistic that $W \to \chi^2_{[K-1]}$, **QED**.

Chapter 4

The analysis of embedded experiments with the generalized regression estimator

4.1 Introduction

In sampling theory, the generalized regression estimator is frequently used in estimation procedures. This has several reasons. Firstly, to make advantage of auxiliary information in an attempt to reduce the design variance of the estimates of population parameters and to correct, at least partially, for the design bias due to selective non-response and coverage errors. Therefore, the use of suitable auxiliary information by means of the generalized regression estimator might increase the accuracy of the Horvitz-Thompson estimator, see e.g. Bethlehem and Keller (1987) or Särndal et al. (1992). Secondly, to ensure that for auxiliary variables the weighted observations in the sample sum up to the known population totals. This enables us to calibrate the sampling weights such that the margins of different publication tables are exactly equal to the known subpopulation totals. In doing so, consistency between the margin totals of different publication tables can be enforced.

In chapter 3, a design-based theory for the analysis of embedded experiments using the Horvitz-Thompson estimator is developed. In this section we extend the theory to the generalized regression estimator, using the model-assisted approach of Särndal et al. (1992). This enables us to incorporate the weighting scheme of the survey in the analysis of embedded experiments as auxiliary information. This might increase the precision of the analysis of treatment effects and make the analysis, at least partially, more robust against bias due to selective non-response. This approach is comparable with the technique of covariance analysis from the theory of experimental designs.

Interest is still focussed on the test of the hypothesis of no treatment effects compared with the unrestricted alternative:

$$H_0: \quad \mathbf{C}\mathrm{E}_\alpha \bar{\mathbf{Y}}^\alpha = \mathbf{0},$$
$$H_1: \quad \mathbf{C}\mathrm{E}_\alpha \bar{\mathbf{Y}}^\alpha \neq \mathbf{0},$$

with $\mathbf{C}$ defined by (3.16) in section 3.3. This hypothesis can be tested by means of the Wald statistic:

$$W = \hat{\bar{\mathbf{Y}}}^{\alpha^t} \mathbf{C}^t \left(\mathbf{C}\widehat{\mathbf{V}}\mathbf{C}^t \right)^{-1} \mathbf{C}\hat{\bar{\mathbf{Y}}}^\alpha. \tag{4.1}$$

In this section a design-based estimator for the contrasts of $\bar{\mathbf{Y}}^\alpha$ and its covariance matrix in the Wald statistic (4.1) is derived using the generalized regression estimator.

4.2 Measurement error models

In order to use the generalized regression estimator in the analysis of embedded experiments, the measurement error models introduced in section 3.2 have to be extended first. This is accomplished by modeling the true, intrinsic value u_i by means of a linear regression model.

Let $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iH})^t$ denote a vector containing the H auxiliary variables x_{ih} of individual i. If the model assisted approach of Särndal et al. (1992) is followed, the true values u_i for each element in the population are assumed to be a realization of the following linear regression model:

$$u_i = \mathbf{b}^t \mathbf{x}_i + e_i, \quad k = 1, 2, \ldots, K, \tag{4.2}$$

$$\mathrm{E}_\xi(e_i) = 0, \tag{4.3}$$

$$\mathrm{Cov}_\xi(e_i, e_{i'}) = \begin{cases} \omega_i^2 & : \ i = i' \\ 0 & : \ i \neq i' \end{cases}, \tag{4.4}$$

where E_ξ and Cov_ξ are the expectation and the covariance with respect to the regression model, $\mathbf{b} = (b_{1k}, \ldots, b_H)^t$ a vector of order H with the regression coefficients and e_i the residuals. It is required that all ω_i^2 be known up to a common scale factor; that is $\omega_i^2 = v_i \omega^2$, with v_i known. In the model assisted approach of Särndal et al. (1992), the finite population of the intrinsic values u_i is considered to be a realization from an underlying superpopulation model defined by (4.2). In this case, the residuals e_i are random variables. From a strictly design-based point of view, as followed by Bethlehem and Keller (1987), there is no need to adopt a superpopulation model. In that case the residuals e_i are considered fixed, intrinsic values and there is no need to introduce model assumptions (4.3) and (4.4). The regression coefficients in both cases are considered finite population characteristics. It is assumed that the auxiliary variables are not affected by the treatments and are intrinsic values, which can be observed without measurement errors.

The basic measurement error model for K survey strategies can be extended by inserting (4.2) into measurement error model (3.1), giving:

$$\mathbf{y}_i^\alpha = \mathbf{j}(\mathbf{b}^t \mathbf{x}_i + e_i) + \beta + \varepsilon_i^\alpha, \tag{4.5}$$

with model assumptions (4.3), (4.4) and

$$\mathrm{E}_\alpha(\varepsilon_i^\alpha) = \mathbf{0}, \tag{4.6}$$

$$\mathrm{Cov}_\alpha(\varepsilon_i^\alpha, \varepsilon_{i'}^{\alpha^t}) = \begin{cases} \boldsymbol{\Sigma}_i & : \quad i = i' \\ \mathbf{O} & : \quad i \neq i' \end{cases} . \tag{4.7}$$

The measurement error model for K survey strategies with interviewer effects can be extended by inserting (4.2) into (3.4), yielding:

$$\mathbf{y}_i^\alpha = \mathbf{j}(\mathbf{b}^t\mathbf{x}_i + e_i) + \beta + \mathbf{j}\gamma_i^\alpha + \varepsilon_i^\alpha, \tag{4.8}$$

where the interviewer effect γ_i^α of the $i-$th individual assigned to the $j-$th interviewer is modeled as

$$\gamma_i^\alpha = \psi_{f(i)} + \xi_{f(i)}^\alpha.$$

Here ψ_j denotes the fixed interviewer effect and ξ_j^α the random interviewer effect of the $j-$th interviewer. Besides the model assumptions (4.3), (4.4), (4.6) and (4.7), the following model assumptions are made

$$\mathrm{E}_\alpha(\xi_{f(i)}^\alpha) = 0, \tag{4.9}$$

$$\mathrm{Cov}_\alpha(\xi_{f(i)}^\alpha, \xi_{f(i')}^\alpha) = \begin{cases} \tau_j^2 & : \quad f(i) = f(i') = j \\ 0 & : \quad f(i) \neq f(i') \end{cases} , \tag{4.10}$$

$$\mathrm{Cov}_\alpha(\varepsilon_{ik}^\alpha, \xi_{f(i)}^\alpha) = 0, \qquad k = 1, \ldots, K. \tag{4.11}$$

4.3 Estimation of treatment effects

Under the measurement error models (4.5) or (4.8) the generalized regression estimator can be used to estimate the population parameters $\mathrm{E}_\alpha\bar{\mathbf{Y}}^\alpha$. According to a sampling design a sample s of size n is drawn from a finite population U of size N with first and second-order inclusion probabilities π_i and $\pi_{ii'}$. According to the experimental design, s is randomly divided in K subsamples s_k of size n_k. Each subsample is assigned to one of the K treatments. At each individual in subsample s_k an observation y_{ik}^α is obtained under treatment k.

In section 3.5 the first-order inclusion probabilities for the sampling units in the K subsamples are derived. In the case of a CRD, the first-order inclusion probabilities of the elements of subsample s_k are equal to $\pi_i^* = (n_k/n)\pi_i$. In the case of an RBD the sample s is divided in J blocks s_j of size n_j. Within each block the individuals are randomized over the K treatments. Let n_{jk} denote the number of individuals in block j assigned to treatment k. It is proved in section 3.5 that under the randomization mechanism of an RBD the first-order inclusion probabilities of the elements of subsample s_k are equal to $\pi_i^* = (n_{jk}/n_j)\pi_i$.

Besides the observations y_{ik}^α, for each element in the sample a vector with H auxiliary variables $\mathbf{x}_i$ is observed. It is assumed that the auxiliary variables are not effected by the treatments and are intrinsic values that can be observed without measurement errors. The H population means of the auxiliary variables are assumed to be known and are denoted by the vector $\bar{\mathbf{X}} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_H)^t$. Let $\hat{\bar{\mathbf{X}}}_{\mathbf{s}_k} = (\hat{\bar{X}}_1, \hat{\bar{X}}_2, \ldots, \hat{\bar{X}}_H)^t$ denote the Horvitz-Thompson estimator for the population means of the auxiliary variables $\bar{\mathbf{X}}$ based on the sampling units of subsample s_k. In the case of a CRD it follows that

$$\hat{\bar{\mathbf{X}}}_{\mathbf{s}_k} = \frac{n}{Nn_k} \sum_{i=1}^{n_k} \frac{1}{\pi_i} \mathbf{x}_i = \frac{1}{N} \sum_{i=1}^{n_k} \frac{\mathbf{x}_i}{\pi_i^*},$$

and for an RBD it follows that

$$\hat{\bar{\mathbf{X}}}_{\mathbf{s}_k} = \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{n_{jk}} \frac{n_j}{n_{jk}\ \pi_i} \mathbf{x}_i = \frac{1}{N} \sum_{i=1}^{n_k} \frac{\mathbf{x}_i}{\pi_i^*}.$$

For a complete enumeration of the finite population, the best linear unbiased estimator for the regression coefficients $\mathbf{b}$ of model (4.2) is given by:

$$\mathbf{b} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^t}{\omega_i^2} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i u_i}{\omega_i^2} \right). \tag{4.12}$$

Note that (4.12) can be expressed in matrix notation as

$$\mathbf{b} = (\mathbf{X}^t \boldsymbol{\Lambda}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \boldsymbol{\Lambda}^{-1} \mathbf{u},$$

where $\mathbf{X}$ denotes a $N \times H$ matrix with each row the auxiliary variables $\mathbf{x}_i$ of individual i, $\mathbf{u}$ a vector of order N with each element the intrinsic values u_i of individual i, and $\boldsymbol{\Lambda}$ a $N \times N$ diagonal matrix with each diagonal element the variance ω_i^2 of (4.4) of individual i in the finite population.

The true values u_i are not observable due to measurement errors and treatment effects. Consequently (4.12) cannot be computed, even in the case of a complete enumeration. In the case of a complete enumeration, however,

$$\mathbf{b}_k^{\alpha} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^t}{\omega_i^2} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i y_{ik}^{\alpha}}{\omega_i^2} \right), \quad k = 1, 2, \ldots, K, \tag{4.13}$$

is defined as the finite population regression coefficients observed under the $k-$th treatment on the $\alpha-$th occasion of the complete enumeration of the finite population. Note that it follows from measurement error model (3.4) that $\mathrm{E}_{\alpha} y_{ik}^{\alpha} = u_i + \beta_k + \psi_j$. Therefore

$$\mathbf{b}_k = \mathrm{E}_{\alpha} \mathbf{b}_k^{\alpha} = \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^t}{\omega_i^2} \right)^{-1} \left(\sum_{i=1}^{N} \frac{\mathbf{x}_i (u_i + \psi_j + \beta_k)}{\omega_i^2} \right), \quad k = 1, 2, \ldots, K, \tag{4.14}$$

is defined as the finite population regression coefficients under the $k-$th treatment.

The regression coefficients $\mathbf{b}_k^{\alpha}$ and $\mathbf{b}_k$ are finite population characteristics, which can be estimated using the sample data from subsample s_k. The Horvitz-Thompson estimator of $\mathbf{b}_k^{\alpha}$ based on the n_k observations of subsample s_k is defined as

$$\hat{\mathbf{b}}_k^{\alpha} = \left(\sum_{i=1}^{n_k} \frac{\mathbf{x}_i \mathbf{x}_i^t}{\omega_i^2 \pi_i^*} \right)^{-1} \left(\sum_{i=1}^{n_k} \frac{\mathbf{x}_i y_{ik}^{\alpha}}{\omega_i^2 \pi_i^*} \right), \quad k = 1, 2, \ldots, K. \tag{4.15}$$

The vector $\hat{\mathbf{b}}_k^{\alpha}$ (4.15) is the estimator proposed by Särndal et al. (1992, section 6.4) for the population parameter $\mathbf{b}_k^{\alpha}$. Since $\hat{\mathbf{b}}_k^{\alpha}$ is an estimator for $\mathbf{b}_k^{\alpha}$, it is also an estimator for $\mathbf{b}_k$, by definition.

The generalized regression estimator for $\mathrm{E}_{\alpha}\bar{Y}_k^{\alpha}$ based on the n_k observations of subsample s_k is given by:

$$\hat{\bar{Y}}_{k_R}^{\alpha} = \hat{\bar{Y}}_k^{\alpha} + \hat{\mathbf{b}}_k^{\alpha^t}(\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}), \quad k = 1, 2, \ldots, K, \tag{4.16}$$

with $\hat{\bar{Y}}_k^{\alpha}$ the Horvitz-Thompson estimator for $\mathrm{E}_{\alpha}\bar{Y}_k^{\alpha}$ defined by (3.23) in section 3.5. To describe the randomization mechanism of the experimental design we introduced in section 3.5 the vectors $\mathbf{p}_{ik}$ by (3.24) for a CRD and (3.25) for an RBD. In order to express the randomization mechanism of the applied experimental design explicitly in the generalized regression estimator, we rewrite (4.16) for a CRD as well as an RBD as follows:

$$\hat{\bar{Y}}_{k_R}^{\alpha} = \sum_{i=1}^{n} \left(\frac{\mathbf{p}_{ik}^t(\mathbf{y}_i^{\alpha} - \hat{\mathbf{B}}^{\alpha^t}\mathbf{x}_i)}{\pi_i \; N} \right) + \hat{\mathbf{b}}_k^{\alpha^t}\bar{\mathbf{X}}, \tag{4.17}$$

where $\hat{\mathbf{B}}^{\alpha}$ denote a matrix of order $H \times K$ whose columns are the $H-$vectors $\hat{\mathbf{b}}_k^{\alpha}$ defined in (4.15). Now $\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha} = (\hat{\bar{Y}}_{1_R}^{\alpha}, \ldots, \hat{\bar{Y}}_{K_R}^{\alpha})^t$ is an approximately design-unbiased estimator for $\bar{\mathbf{Y}}^{\alpha}$ as well as $\bar{\mathbf{Y}}$ defined by (3.11) and (3.12) in section 3.3 (Särndal et al., 1992, ch.6). The subscript $\mathbf{s}_k$ is added in order to emphasize that $\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}$ consists of K generalized regression estimators $\hat{\bar{Y}}_{k_R}^{\alpha}$, each based on the n_k elements of subsample s_k.

In (4.2) each intrinsic value u_i is decomposed into a term that can be explained by the auxiliary information and a remainder term e_i. If the design variance of this remainder term is smaller than the design variance of the intrinsic value, then the precision of the Horvitz-Thompson estimator can be increased by means of the auxiliary information used in the generalized regression estimator.

Under the null hypothesis it follows that $\mathbf{b}_k^{\alpha} = \mathbf{b}_{k'}^{\alpha}$ for all k and k'. Since the finite population regression coefficients are not biased with treatment effects under the null hypothesis, it might be efficient to estimate $\mathbf{b}$ in (4.2) by means of the pooled Horvitz-Thompson estimator

$$\hat{\mathbf{b}}^{\alpha} = \left(\sum_{i=1}^{n} \frac{\mathbf{x}_i\mathbf{x}_i^t}{\omega_i^2 \pi_i} \right)^{-1} \left(\sum_{k=1}^{K} \sum_{i=1}^{n_k} \frac{\mathbf{x}_i y_{ik}^{\alpha}}{\omega_i^2 \pi_i} \right). \tag{4.18}$$

In the generalized regression estimator (4.16) and (4.17), $\hat{\mathbf{b}}_k^{\alpha}$ can be substituted for $\hat{\mathbf{b}}^{\alpha}$. This approach might look efficient at first sight since only H instead of $K \times H$ regression coefficients have to be estimated on the basis of n observations of sample s. As a result, the pooled estimates of the regression coefficients $\hat{\mathbf{b}}^{\alpha}$ will be more precise. This advantage, however, is only minor since $\hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}$ is still based on the n_k observations of subsample s_k, which limits the possibility to specify larger weighting schemes that are comparable with the weighting scheme of the regular survey. Another disadvantage of this approach is that $\hat{\mathbf{b}}^{\alpha}$ is biased with a mixture of treatment effects under the alternative hypothesis. Finally, for the auxiliary variables, the weighted observations in the subsamples do not necessarily sum up to the known population totals.

As an alternative we can write the generalized regression estimator (4.16) in terms of design and correction weights, see Särndal et al. (1992, section 6.5)

$$\begin{aligned}
\hat{\hat{Y}}^{\alpha}_{k_R} &= \hat{\hat{Y}}^{\alpha}_{k} + \hat{\mathbf{b}}^{\alpha^t}_{k}(\bar{\mathbf{X}} - \hat{\hat{\mathbf{X}}}_{\mathbf{s}_k}) \\
&= \frac{1}{N}\sum_{i=1}^{n_k}\frac{y^{\alpha}_{ik}}{\pi^*_i} + \left[\left(\sum_{i=1}^{n_k}\frac{\mathbf{x}_i\mathbf{x}^t_i}{\omega^2_i\pi^*_i}\right)^{-1}\left(\sum_{i=1}^{n_k}\frac{\mathbf{x}_i y^{\alpha}_{ik}}{\omega^2_i\pi^*_i}\right)\right]^t\left(\bar{\mathbf{X}} - \frac{1}{N}\sum_{i=1}^{n_k}\frac{\mathbf{x}_i}{\pi^*_i}\right) \\
&= \frac{1}{N}\sum_{i=1}^{n_k}\frac{y^{\alpha}_{ik}}{\pi^*_i}\left[1 + \frac{\mathbf{x}^t_i}{\omega^2_i}\left(\sum_{i=1}^{n_k}\frac{\mathbf{x}_i\mathbf{x}^t_i}{\omega^2_i\pi^*_i}\right)^{-1}(N\bar{\mathbf{X}} - \sum_{i=1}^{n_k}\frac{\mathbf{x}_i}{\pi^*_i})\right] \\
&\equiv \frac{1}{N}\sum_{i=1}^{n_k}\frac{y^{\alpha}_{ik}}{\pi^*_i}g^*_i. \qquad (4.19)
\end{aligned}$$

If the pooled estimator for the regression coefficients (4.18), based on the n sampling units of sample s, is substituted in the generalized regression estimator (4.16) and (4.17), then it is not possible to rewrite the generalized regression estimator in terms of design and correction weights. This emphasizes the artificiality of the application of a pooled estimator for the regression coefficient.

4.4 Variance estimation of treatment effects

Let $\mathbf{V}_R$ denote the variance-covariance matrix of the generalized regression estimator $\hat{\hat{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$. As we emphasized in section 3.6 there is a nonzero design covariance between the elements of $\hat{\hat{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$ since they are based on K interpenetrating subsamples drawn from a finite population without replacement. To estimate the covariance terms of $\mathbf{V}_R$, vectors $\mathbf{y}^{\alpha}_i$ containing observations under all K treatments obtained from each experimental unit are required. Since we haven't collect observations of the target variable under all of the K treatments at each experimental unit, it will be cumbersome to estimate the covariance terms on the off-diagonal elements of $\mathbf{V}_R$. Therefore, we follow the approach applied in section 3.6 and concentrate on the covariance matrix of the $K-1$ contrasts of $\hat{\hat{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$, denoted $\mathbf{C}\mathbf{V}_R\mathbf{C}^t$.

Since the generalized regression estimator of $\mathrm{E}_{\alpha}\bar{\mathbf{Y}}^{\alpha}$ is nonlinear, it is generally not possible to obtain exact results on the covariance matrix of $\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$. A Taylor series linearization can be applied, though, to obtain approximate expressions for the variance and approximate estimates for the variance of the generalized regression estimator (Särndal et al. (1992)). Expressing (4.16) as a function of $(\hat{\hat{Y}}^{\alpha}_{k}, \hat{\mathbf{b}}^{\alpha}_{k}, \hat{\hat{\mathbf{X}}}_{\mathbf{s}_k})$, the generalized regression estimator can be approximated by means of a first-order Taylor series linearization about $(\bar{Y}_k, \mathbf{b}_k, \bar{\mathbf{X}})$, where $\bar{Y}_k = \mathrm{E}_{\alpha}(\bar{Y}^{\alpha}_k)$ and $\mathbf{b}_k = \mathrm{E}_{\alpha}(\mathbf{b}^{\alpha}_k)$. This gives:

$$\hat{\hat{Y}}^{\alpha}_{k_R} \doteq \hat{\hat{Y}}^{\alpha}_{k} + \mathbf{b}^t_k(\bar{\mathbf{X}} - \hat{\hat{\mathbf{X}}}_{\mathbf{s}_k}) = \hat{\hat{E}}^{\alpha}_{k} + \mathbf{b}^t_k\bar{\mathbf{X}}, \quad k = 1, 2, \ldots, K, \qquad (4.20)$$

with

$$\hat{\hat{E}}^{\alpha}_{k} = \hat{\hat{Y}}^{\alpha}_{k_R} - \mathbf{b}^t_k\hat{\hat{\mathbf{X}}}_{\mathbf{s}_k}$$

$$= \sum_{i \in s} \left(\frac{\mathbf{p}_{ik}^t (\mathbf{y}_i^\alpha - \mathbf{B}^t \mathbf{x}_i)}{\pi_i \; N} \right), \quad k = 1, 2, \ldots, K, \tag{4.21}$$

and $\mathbf{B}$ a $H \times K$ matrix whose columns are the $H-$vectors $\mathbf{b}_k$ (see appendix 4.7.1). The vector $\mathbf{p}_{ik}$, defined by (3.24) and (3.25) in section 3.5, enables us to take into account for the randomization mechanism of the experimental design in deriving design variances and covariances of the subsample estimates. The covariance matrix of the contrasts of $\hat{\hat{\mathbf{Y}}}^\alpha_{R\mathbf{s}_k}$ is approximated with the covariance matrix of the first-order Taylor series approximation of the of the contrasts of $\hat{\hat{\mathbf{Y}}}^\alpha_{R\mathbf{s}_k}$, which is obtained by that of $\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} = (\hat{\hat{E}}^\alpha_1, \ldots, \hat{\hat{E}}^\alpha_K)^t$.

If we condition on the measurement error model, the sampling design and the experimental design, then it follows that a first-order Taylor series approximation of $\mathbf{C}\mathbf{V}_R\mathbf{C}^t$ equals

$$\mathbf{C}\mathbf{V}_R\mathbf{C}^t \doteq \mathrm{Cov}_\alpha \mathrm{E}_s \mathrm{E}_\varepsilon (\mathbf{C}\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \,|\, \alpha, s) + \mathrm{E}_\alpha \mathrm{Cov}_s \mathrm{E}_\varepsilon (\mathbf{C}\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \,|\, \alpha, s) + \mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon (\mathbf{C}\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \,|\, \alpha, s). \tag{4.22}$$

If there exists a constant $H-$vector $\mathbf{a}$ such that $\mathbf{a}^t\mathbf{x}_i = 1$ for all $i \in U$, then it follows that

$$\mathbf{C}(\mathbf{y}_i^\alpha - \mathbf{B}^t\mathbf{x}_i) = \mathbf{C}\varepsilon_i^\alpha. \tag{4.23}$$

A proof is given in appendix 4.7.2. The stated condition, which is quite common in finite population sampling, implicitly assumes that the size of the finite population is known and is used as auxiliary information in the estimation procedure.

If the stated condition is satisfied, then the three components on the right-hand side of (4.22) can be evaluated as follows. For the first and the second term it follows from (3.84) for a CRD or (3.89) for an RBD and measurement error model assumption (4.7) that

$$\begin{aligned}
\mathrm{Cov}_\alpha \mathrm{E}_s \mathrm{E}_\varepsilon (\mathbf{C}\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s) &= \mathrm{Cov}_\alpha \mathrm{E}_s \mathrm{E}_\varepsilon \left(\sum_{i=1}^n \frac{\mathbf{p}_{ik}^t \mathbf{C}(\mathbf{y}_i^\alpha - \mathbf{B}^t\mathbf{x}_i)}{\pi_i N} \mid \alpha, s \right) \\
&= \mathrm{Cov}_\alpha \mathrm{E}_s \mathrm{E}_\varepsilon \left(\sum_{i=1}^n \frac{\mathbf{p}_{ik}^t \mathbf{C}\varepsilon_i^\alpha}{\pi_i N} \mid \alpha, s \right) \\
&= \mathrm{Cov}_\alpha \mathrm{E}_s \left(\sum_{i=1}^n \frac{\mathbf{C}\varepsilon_i^\alpha}{\pi_i N} \mid \alpha \right) = \mathrm{Cov}_\alpha \left(\sum_{i=1}^N \frac{\mathbf{C}\varepsilon_i^\alpha}{N} \right) \\
&= \frac{1}{N^2} \sum_{i=1}^N \mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t,
\end{aligned} \tag{4.24}$$

$$\begin{aligned}
\mathrm{E}_\alpha \mathrm{Cov}_s \mathrm{E}_\varepsilon (\mathbf{C}\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s) &= \mathrm{E}_\alpha \mathrm{Cov}_s \mathrm{E}_\varepsilon \left(\sum_{i=1}^n \frac{\mathbf{p}_{ik}^t \mathbf{C}(\mathbf{y}_i^\alpha - \mathbf{B}^t\mathbf{x}_i)}{\pi_i N} \mid \alpha, s \right) \\
&= \mathrm{E}_\alpha \mathrm{Cov}_s \left(\sum_{i=1}^n \frac{\mathbf{C}\varepsilon_i^\alpha}{\pi_i N} \mid \alpha \right) \\
&= \mathrm{E}_\alpha \frac{1}{N^2} \sum_{i=1}^N \sum_{i'=1}^N (\pi_{ii'} - \pi_i\pi_{i'}) \frac{\mathbf{C}\varepsilon_i^\alpha \varepsilon_{i'}^{\alpha^t} \mathbf{C}^t}{\pi_i\pi_{i'}} \\
&= \frac{1}{N^2} \sum_{i=1}^N \left(\frac{1}{\pi_i} - 1 \right) \mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t.
\end{aligned} \tag{4.25}$$

In appendix 4.7.3 it is proved that

$$\mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s) = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{C}\mathbf{D}_R\mathbf{C}^t - \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t}{\pi_i}, \tag{4.26}$$

where $\mathbf{D}_R$ is a $K \times K$ diagonal matrix with elements

$$d_{k_R} = \frac{1}{(n-1)}\frac{1}{n_k}\sum_{i=1}^{n}\left(\frac{n(y^\alpha_{ik} - \mathbf{b}^t_k\mathbf{x}_i)}{N\pi_i} - \frac{1}{n}\sum_{i'=1}^{n}\frac{n(y^\alpha_{i'k} - \mathbf{b}^t_k\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2 \equiv \frac{S^2_{E_k}}{n_k}, \tag{4.27}$$

in the case of a CRD and

$$d_{k_R} = \sum_{j=1}^{J}\frac{1}{(n_j-1)}\frac{1}{n_{jk}}\sum_{i=1}^{n_j}\left(\frac{n_j(y^\alpha_{ik} - \mathbf{b}^t_k\mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{n_j(y^\alpha_{i'k} - \mathbf{b}^t_k\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2 \equiv \sum_{j=1}^{J}\frac{S^2_{E_{jk}}}{n_{jk}}, \tag{4.28}$$

in the case of an RBD. If the results obtained in (4.24), (4.25) and (4.26) are substituted into (4.22), then it follows that

$$\mathbf{C}\mathbf{V}_R\mathbf{C}^t = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{C}\mathbf{D}_R\mathbf{C}^t. \tag{4.29}$$

Before we evaluate the expectations E_α and E_s of $\mathrm{E}_\alpha\mathrm{E}_s\mathbf{D}_R$ in chapter 5 for CRD's and chapter 6 for RBD's, an unbiased estimators for $\mathrm{E}_\alpha\mathrm{E}_s\mathbf{D}_R$ is derived. In the case of a CRD, an approximately design-unbiased estimator for $\mathrm{E}_\alpha\mathrm{E}_s\mathbf{D}_R$ is given by the diagonal matrix $\hat{\mathbf{D}}_R$ with diagonal elements

$$\hat{d}_{k_R} = \frac{1}{n_k}\frac{1}{(n_k-1)}\sum_{i=1}^{n_k}\left(\frac{n(y^\alpha_{ik} - \hat{\mathbf{b}}^{\alpha^t}_k\mathbf{x}_i)}{N\pi_i} - \frac{1}{n_k}\sum_{i'=1}^{n_k}\frac{n(y^\alpha_{i'k} - \hat{\mathbf{b}}^{\alpha^t}_k\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2 \equiv \frac{\hat{S}^2_{E_k}}{n_k}. \tag{4.30}$$

Conditional on the realization of α and s, the allocation of the experimental units to the subsamples s_k by means of a CRD can be considered as simple random sampling without replacement from s with variables $(n(y^\alpha_{ik} - \hat{\mathbf{b}}^{\alpha^t}_k\mathbf{x}_i))/(\pi_i N)$. Therefore it follows that $\hat{S}^2_{E_k}$ in (4.30) is an approximately design-unbiased estimator for $S^2_{E_k}$ in (4.27). An approximately design-unbiased estimator for $\mathrm{E}_\alpha\mathrm{E}_s\mathbf{D}_R$ for an RBD is given by the diagonal matrix $\hat{\mathbf{D}}_R$ with diagonal elements

$$\begin{aligned}\hat{d}_{k_R} &= \sum_{j=1}^{J}\frac{1}{n_{jk}}\frac{1}{(n_{jk}-1)}\sum_{i=1}^{n_{jk}}\left(\frac{n_j(y^\alpha_{ik} - \hat{\mathbf{b}}^{\alpha^t}_k\mathbf{x}_i)}{N\pi_i} - \frac{1}{n_{jk}}\sum_{i'=1}^{n_{jk}}\frac{n_j(y^\alpha_{i'k} - \hat{\mathbf{b}}^{\alpha^t}_k\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2 \\ &\equiv \sum_{j=1}^{J}\frac{\hat{S}^2_{E_{jk}}}{n_{jk}}.\end{aligned} \tag{4.31}$$

In the case of an RBD, conditional on the realization of α and s, the allocation of the experimental units within each block to the subsamples s_{jk} can be considered as simple random sampling without replacement with variables $(n_j(y^\alpha_{ik} - \hat{\mathbf{b}}^{\alpha^t}_k\mathbf{x}_i))/(\pi_i N)$. As a result it follows that $\hat{S}^2_{E_{jk}}$ in (4.31) is an approximately design-unbiased estimator for $S^2_{E_{jk}}$ in (4.28).

The residuals $(y^\alpha_{ik} - \hat{\mathbf{b}}^{\alpha^t}_k\mathbf{x}_i)$ in the variance estimators (4.30) and (4.31) can be multiplied with the $g-$weights g^*_i defined in (4.19) as an alternative. Then the variance estimator for the generalized regression estimator, proposed by Särndal et al. (1992, result 6.6.1.) is obtained.

No second-order inclusion probabilities are required in the estimator for $\mathbf{C}\mathbf{V}_R\mathbf{C}^t$. This is a consequence of the following factors:

1. The assumption of additive treatment effects in the measurement error models (i.e. β_k, for all $i \in U$ observed under treatment k).

2. The assumption that measurement errors between individuals are independent.

3. The first-order Taylor series approximation.

4. A properly chosen set of auxiliary information in the weighting scheme such that the condition $\mathbf{a}^t\mathbf{x}_i = 1$ for all $i \in U$ is satisfied.

5. The fact that variances are calculated for the contrasts between subsample means.

Since the generalized regression estimator is nonlinear, a first-order Taylor series approximation is applied to obtain a linearized first-order approximation. As a result an approximation of the design variance of the generalized regression estimator is obtained by the design variance of the residuals $(y_{ik}^{\alpha} - \mathbf{b}_k^t\mathbf{x}_i)$. Under the condition that there is a vector $\mathbf{a}$, such that $\mathbf{a}^t\mathbf{x}_i = 1$, for all $i \in U$, it follows that the additive treatment effects β_k vanish in the residuals $(y_{ik}^{\alpha} - \mathbf{b}_k^t\mathbf{x}_i)$. The stated condition implies that the weighting scheme of the generalized regression estimator at least utilizes the size of the finite population as auxiliary information. Therefore we have $\sum_{i=1}^{n_k} w_{ik} = \hat{N} = N$, where w_{ik} are the regression weights of the elements in subsample s_k. This implies that a constant, for example the additive treatment effect β_k, will be estimated without error. Consequently the additive treatment effects vanish from the residuals. In the residuals three terms remains:

1. The residual of the linear regression model of the intrinsic value: $e_i = u_i - \mathbf{b}^t\mathbf{x}_i$.

2. A term, concerning the bias due to interviewer effects. This term equals $(\gamma_i^{\alpha} - \mathbf{d}^t\mathbf{x}_i)$, where $\mathbf{d}$ denotes the regression of the interviewer effects on the auxiliary variables $\mathbf{x}_i$ (see (4.39) in appendix 4.7.2).

3. The measurement error $\varepsilon_{ik}^{\alpha}$.

Since we concentrate on the variance of the contrasts between the subsample means, the two terms $\mathrm{Cov}_{\alpha}\mathrm{E}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{E}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s)$ and $\mathrm{E}_{\alpha}\mathrm{Cov}_s\mathrm{E}_{\varepsilon}(\mathbf{C}\hat{\hat{\mathbf{E}}}_{\mathbf{s}_k}^{\alpha} \mid \alpha, s)$, which are given by (4.24) and (4.25) respectively, only contains the measurement errors $\varepsilon_{ik}^{\alpha}$. The residuals of the intrinsic values e_i and the bias due to interviewer effects cancels out in these two terms since they do not depend on the different treatments. Finally, due to the assumption of independence of the measurement errors between individuals, the cross products between individuals, which contain the second-order inclusion probabilities in (4.24) and (4.25) vanish.

Conditional on the realization of α and s, the allocation of the experimental units to the treatments comes down to the selection of K subsamples of size n_k from s of size n by means of simple random sampling without replacement in the case of a CRD or stratified simple random sampling without replacement where strata correspond to the blocks of the experiment in the case of an RBD. The covariance structure of the third term in (4.22), which is given by (4.26), is mainly determined by these randomization mechanisms of the experimental design. Under these two randomization mechanisms, it is shown in section 3.6.3 that in the variance of the contrasts between two subsample means of the intrinsic values, conditional on α and s, the finite population corrections in the variance of the subsample means cancel out against the covariances between these two subsample means. As a result, the leading term of (4.26), i.e. $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{C}\mathbf{D}_R\mathbf{C}^t$, has the structure as if the K subsamples are drawn independently from each other by means of simple random sampling with replacement in the case of a CRD, or stratified simple random sampling with replacement in the case of an RBD. The second variance component in (4.26) arises from the measurement errors and cancels out against the variance components in (4.25). As a result a variance estimation procedure could be derived as if the subsamples were drawn independently from each other.

The main difference between the results for the generalized regression estimator and the Horvitz-Thompson estimator is that in the variance estimation procedure under the generalized regression estimator, the second-order inclusion probabilities completely vanish. For the Horvitz-Thompson estimator a covariance component, which concerns the variance of the fixed, additive treatment effects, remains under the alternative hypotheses. If this component is not neglected, then second-order inclusion probabilities are required in the variance estimation procedure. Under the generalized regression estimator with a properly chosen set of auxiliary information these components vanish, even under the alternative hypothesis. The condition under which these results hold true, implies that at least the size of the finite population is used as auxiliary information in the weighting scheme. The minimum use of auxiliary information is a weighting scheme where $\mathbf{x}_i = (1)$ and $\omega_i^2 = \omega^2$ for all $i \in U$. This implies that regression model (4.2) equals $u_i = b + e_i$, which is the common mean model discussed by Särndal et al. (1992, section 7.4). Under this weighting scheme, it follows from (4.15) that the regression coefficient

$$\hat{b}_k^\alpha = \left(\sum_{i=1}^{n_k} \frac{1}{\pi_i^*}\right)^{-1} \left(\sum_{i=1}^{n_k} \frac{y_{ik}^\alpha}{\pi_i^*}\right) = \tilde{y}_{s_k}. \tag{4.32}$$

If this regression coefficient is substituted into the generalized regression estimator defined by (4.16), then it follows that

$$\begin{aligned}\hat{\hat{Y}}{}^\alpha_{k_R} &= \hat{\hat{Y}}{}^\alpha_k + \left(\sum_{i=1}^{n_k} \frac{1}{\pi_i^*}\right)^{-1} \left(\sum_{i=1}^{n_k} \frac{y_{ik}^\alpha}{\pi_i^*}\right) \left(1 - \frac{1}{N}\sum_{i=1}^{n_k} \frac{1}{\pi_i^*}\right) \\ &= \left(\sum_{i=1}^{n_k} \frac{y_{ik}^\alpha}{\pi_i^*}\right) \left(\sum_{i=1}^{n_k} \frac{1}{\pi_i^*}\right)^{-1} = \frac{1}{\hat{N}} \sum_{i=1}^{n_k} \frac{y_{ik}^\alpha}{\pi_i^*},\end{aligned} \tag{4.33}$$

which can be recognized as the extended Horvitz-Thompson estimator or the ratio estimator for a population mean, originally proposed by Hájek (1971). Approximately design-unbiased estimates for the covariance matrix of the treatment effects are given by (4.30) and (4.31), where $\hat{\mathbf{b}}_k^{\alpha^t} \mathbf{x}_i$ is replaced by $\tilde{y}_{s_k}$ defined by (4.32).

For situations where the sum over the design weights equals the population total, i.e. $\sum_{i=1}^{n_k} \frac{1}{\pi_i^*} = \hat{N} = N$, the ratio estimator (4.33) and its variance estimator, correspond with the results obtained for the Horvitz-Thompson estimator. This condition is met for example in the case of CRD's and RBD's embedded in a simple random sampling design, an RBD embedded in a stratified simple random sampling design where strata are used as block variables or a CRD embedded in a stratified simple random sampling design with proportional allocation. (Under a CRD the fraction of sampling units in stratum j assigned to treatment k, i.e. n_{jk}/n_k is random and will therefore be different for the different strata. As a result $\hat{N} = N$ only in the case of a self-weighted sampling scheme, which is obtained by proportional allocation in the case of stratified sampling. For other allocations it follows that $\hat{N} \neq N$. Consequently, for embedded experiments where this condition is met, the variance estimators based on the Horvitz-Thompson estimator in section 3.6 are design unbiased under the alternative hypothesis, since the covariance components of the additive treatment effects are equal to zero under such sampling designs. For other type of sampling schemes, it might be preferable to apply the ratio estimator defined by (4.33) instead of the Horvitz-Thompson estimator. Firstly for this estimator the variance estimator is always approximately design-unbiased since it is a special case of the generalized regression estimator. Secondly this estimator avoids the extreme estimates sometimes obtained with the Horvitz-Thompson estimator under variable size sampling schemes (see e.g. Särndal et al. 1992, example 7.4.1).

A second difference with the results obtained under the Horvitz-Thompson estimator is the use of auxiliary information in the estimation procedure. In the variance estimators, the observations y_{ik}^{α} are replaced by their corresponding residuals $y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t} \mathbf{x}_i$. The structure of these variance estimators, however, is equal to the structure of the variance estimators derived under the Horvitz-Thompson estimator.

4.5 Wald statistic for the hypothesis of no treatment effects

An approximately design-unbiased estimator for $\mathbf{C}\bar{\mathbf{Y}}$ is given by $\mathbf{C}\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}$, where the elements of $\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}$ are defined by (4.16) or (4.17). An approximately design-unbiased estimator for the covariance of $\mathbf{C}\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}$ is given by $\mathbf{C}\hat{\mathbf{D}}_R\mathbf{C}^t$, where the diagonal elements of $\hat{\mathbf{D}}_R$ are defined by (4.30) and (4.31) for CRD's and RBD's respectively. Inserting these estimators into the design-based Wald statistic (4.1) leads to

$$W = \hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha^t} \mathbf{C}^t \left(\mathbf{C}\hat{\mathbf{D}}_R\mathbf{C}^t\right)^{-1} \mathbf{C}\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}. \tag{4.34}$$

Equivalent to the derivation of (3.80) under the Horvitz-Thompson estimator in section 3.7.1, expression (4.34) can be simplified to

$$\begin{aligned} W &= \sum_{k=1}^{K} \frac{\hat{\bar{Y}}_{k_R}^{\alpha^2}}{\hat{d}_{k_R}} - \frac{1}{\sum_{k=1}^{K} \frac{1}{\hat{d}_{k_R}}} \left(\sum_{k=1}^{K} \frac{\hat{\bar{Y}}_{k_R}^{\alpha}}{\hat{d}_{k_R}} \right)^2 \qquad (4.35) \\ &= \sum_{k=1}^{K} \frac{(\hat{\bar{Y}}_{k_R}^{\alpha} - \hat{\bar{Y}}_{R}^{\alpha})^2}{\hat{d}_{k_R}}, \end{aligned}$$

with

$$\hat{\hat{Y}}^{\alpha}_{R} = \frac{\sum_{k=1}^{K} \frac{\hat{\hat{Y}}^{\alpha}_{k_R}}{\hat{d}_{k_R}}}{\sum_{k=1}^{K} \frac{1}{\hat{d}_{k_R}}}.$$

In order to test the hypothesis of no treatment effects, critical regions for the design-based Wald statistic are required. According to the same arguments as given for the Horvitz-Thompson estimator in subsection 3.7.2, it follows that if the sampling design is simple random sampling and the experimental design is a CRD and $n_k \to \infty$ and $N - n_k \to \infty$, then

$$(\mathbf{C}\hat{\hat{\bar{\mathbf{E}}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha) \to \mathcal{N}(\mathbf{C}\bar{\mathbf{E}}^{\alpha}, \mathbf{C}\mathbf{V}_R{}^{\alpha}\mathbf{C}^t),$$

where $\bar{\mathbf{E}}^{\alpha} = \mathrm{E}_s\mathrm{E}_{\varepsilon}(\hat{\hat{\bar{\mathbf{E}}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s)$ and $\mathbf{V}^{\alpha}_{R} = \mathrm{Cov}_s\mathrm{E}_{\varepsilon}(\hat{\hat{\bar{\mathbf{E}}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s) + \mathrm{E}_s\mathrm{Cov}_{\varepsilon}(\hat{\hat{\bar{\mathbf{E}}}}^{\alpha}_{\mathbf{s}_k} \mid \alpha, s)$. Since $\mathbf{C}\bar{\mathbf{E}}^{\alpha} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{C}\varepsilon_i^{\alpha}$ and the $\mathbf{C}\varepsilon_i^{\alpha}$ are mutually independent random variables with mean zero and covariance matrix $\mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t$ it follows from the ordinary central limit theorem that

$$(\mathbf{C}\bar{\mathbf{E}}^{\alpha}) \to \mathcal{N}(\mathbf{0}, (\frac{1}{N^2})\sum_{i=1}^{N} \mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t). \tag{4.36}$$

If both limit theorems are combined, then it follows unconditionally that

$$(\mathbf{C}\hat{\hat{\bar{\mathbf{E}}}}^{\alpha}_{\mathbf{s}_k}) \to \mathcal{N}(\mathbf{0}, \mathbf{C}\mathbf{V}_R\mathbf{C}^t),$$

and thus

$$(\mathbf{C}\hat{\hat{\bar{\mathbf{Y}}}}^{\alpha}_{R\mathbf{s}_k}) \to \mathcal{N}(\mathbf{C}\beta, \mathbf{C}\mathbf{V}_R\mathbf{C}^t). \tag{4.37}$$

For more complex sampling designs it is assumed that there is a limit theorem such that (4.37) holds true. In that case the Wald statistic, under the null hypothesis, is asymptotically distributed as a chi-squared random variable with $K - 1$ degrees of freedom (see appendix 3.9.8).

4.6 Discussion

In this chapter the design-based approach for the analysis of experiments developed in chapter 3, is extended to the generalized regression estimator. In doing so, the weighting scheme of the survey, can be incorporated in the analysis of an embedded experiment. This not only increases the precision, but it also makes the analysis more robust against the bias due to selective nonresponse. The use of auxiliary information by means of the generalized regression estimator in the analysis of embedded experiments represents a design-based analogy of the technique of covariance analysis in the standard theory of design and analysis of experiments.

The estimator for the variance-covariance matrix $\mathbf{C}\mathbf{V}_R\mathbf{C}^t$ has the structure as if the K subsamples has been drawn independently from each other. The probability structure of the sampling design is incorporated in the variance estimation procedure by means of a reweighting of the observations with a factor, which contains the first-order inclusion probabilities $(n/(N\pi_i)$

in the case of a CRD and $n_j/(N\pi_i)$ in the case of an RBD). The derived variance estimators for the subsample means have the structure as if the sample elements are selected with unequal probabilities (π_i/n) with replacement in the case of a CRD or (π_i/n_j) with replacement within each block in the case of an RBD. Since no second-order inclusion probabilities are required in the variance estimation procedure, the analysis of embedded experiments is simplified considerably. Second-order inclusion probabilities only appear in the covariance matrix if the expectation with respect to the sampling design in the expressions of $\mathrm{E}_\alpha \mathrm{E}_s \mathbf{DR}$ is worked out (see chapters 5 and 6).

Under the generalized regression estimator with a properly chosen set of auxiliary information, this variance estimation procedure is approximately design-unbiased. Since this condition is satisfied by Hájeks ratio estimator for a population mean (also known as the common mean model), this estimator is preferable to the Horvitz-Thompson estimator in many situations.

4.7 Appendix

4.7.1 Taylor series approximation for the generalized regression estimator

A first-order Taylor series approximation of the generalized regression estimator $\hat{\bar{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$ about $(\mathrm{E}_\alpha \bar{\mathbf{Y}}^{\alpha}, \mathbf{b}_k, \bar{\mathbf{X}})$ can be derived as follows (see e.g. Särndal et al., 1992, section 5.5 and 6.6). The generalized regression estimator for each element of $\hat{\bar{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$ is expressed as a function of $(\hat{\bar{Y}}^{\alpha}_k, \hat{\mathbf{b}}^{\alpha}_k, \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k})$

$$\hat{\bar{Y}}^{\alpha}_{k_R} = \hat{\bar{Y}}^{\alpha}_k + \hat{\mathbf{b}}^{\alpha^t}_k (\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}) = f(\hat{\bar{Y}}^{\alpha}_k, \hat{\mathbf{b}}^{\alpha}_k, \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}).$$

The partial derivatives of $f(\hat{\bar{Y}}^{\alpha}_k, \hat{\mathbf{b}}^{\alpha}_k, \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k})$ are given by

$$\begin{aligned}
\frac{\partial f(\hat{\bar{Y}}^{\alpha}_k, \hat{\mathbf{b}}^{\alpha}_k, \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k})}{\partial \hat{\bar{Y}}^{\alpha}_k} &= 1, \\
\frac{\partial f(\hat{\bar{Y}}^{\alpha}_k, \hat{\mathbf{b}}^{\alpha}_k, \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k})}{\partial \hat{\mathbf{b}}^{\alpha}_k} &= (\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}), \\
\frac{\partial f(\hat{\bar{Y}}^{\alpha}_k, \hat{\mathbf{b}}^{\alpha}_k, \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k})}{\partial \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}} &= -\hat{\mathbf{b}}^{\alpha}_k.
\end{aligned}$$

These partial derivatives are evaluated in their true value points, where $\hat{\bar{Y}}^{\alpha}_k = \mathrm{E}_\alpha \bar{Y}^{\alpha}_k$, $\hat{\mathbf{b}}^{\alpha}_k = \mathbf{b}_k$ and $\hat{\bar{\mathbf{X}}}_{\mathbf{s}_k} = \bar{\mathbf{X}}$. (Note that $\hat{\bar{Y}}^{\alpha}_k$, $\hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}$ and $\hat{\mathbf{b}}^{\alpha}_k$ are consistent estimators of $\mathrm{E}_\alpha \bar{Y}^{\alpha}_k$, $\bar{\mathbf{X}}$ and $\mathbf{b}_k$.) Now, a first-order Taylor series approximation of the generalized regression estimator is obtained by:

$$\begin{aligned}
f(\hat{\bar{Y}}^{\alpha}_k, \hat{\mathbf{b}}^{\alpha}_k, \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}) &= \mathrm{E}_\alpha \bar{Y}^{\alpha}_k + \mathbf{b}^t_k(\bar{\mathbf{X}} - \bar{\mathbf{X}}) + (\hat{\bar{Y}}^{\alpha}_k - \mathrm{E}_\alpha \bar{Y}^{\alpha}_k) \\
&\quad + (\bar{\mathbf{X}} - \bar{\mathbf{X}})^t(\hat{\mathbf{b}}^{\alpha}_k - \mathbf{b}_k) - \mathbf{b}^t_k(\hat{\bar{\mathbf{X}}}_{\mathbf{s}_k} - \bar{\mathbf{X}}) + R \\
&\doteq \hat{\bar{Y}}^{\alpha}_k + \mathbf{b}^t_k(\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{\mathbf{s}_k}) = \sum_{i \in s} \left(\frac{\mathbf{p}^t_{ik}(\mathbf{y}^{\alpha}_i - \mathbf{B}^t_k \mathbf{x}_i)}{\pi_i \; N} \right) + \mathbf{b}^t_k \bar{\mathbf{X}}. \qquad (4.38)
\end{aligned}$$

4.7.2 Proof of formula (4.23)

It is proved that

$$\mathbf{C}(\mathbf{y}_i^\alpha - \mathbf{B}^t\mathbf{x}_i) = \mathbf{C}\varepsilon_i^\alpha,$$

under the condition that there exists a constant $H-$vector $\mathbf{a}$ such that $\mathbf{a}^t\mathbf{x}_i = 1$ for all $i \in U$. Under the stated condition it follows that $\mathbf{b}_k^\alpha$ in (4.14) can be evaluated as

$$\begin{aligned}
\mathrm{E}_\alpha(\mathbf{b}_k^\alpha) &= \mathrm{E}_\alpha \left(\sum_{i=1}^N \frac{\mathbf{x}_i\mathbf{x}_i^t}{\omega_i^2}\right)^{-1} \sum_{i=1}^N \frac{\mathbf{x}_i y_{ik}^\alpha}{\omega_i^2} \\
&= \left(\sum_{i=1}^N \frac{\mathbf{x}_i\mathbf{x}_i^t}{\omega_i^2}\right)^{-1} \sum_{i=1}^N \frac{\mathbf{x}_i u_i}{\omega_i^2} + \left(\sum_{i=1}^N \frac{\mathbf{x}_i\mathbf{x}_i^t}{\omega_i^2}\right)^{-1} \sum_{j=1}^J \sum_{i=1}^{N_j} \frac{\mathbf{x}_i \psi_j}{\omega_i^2} + \mathbf{a}\beta_k \\
&= \mathbf{b} + \mathbf{d} + \mathbf{a}\beta_k, \qquad (4.39)
\end{aligned}$$

where $\mathbf{b}$ are the regression coefficients defined in (4.2) and $\mathbf{d}$ denotes the regression coefficients from the regression function of the interviewer effects on the auxiliary variables $\mathbf{x}_i$.
From result (4.39) it follows that

$$\begin{aligned}
\mathbf{B}^t\mathbf{x}_i &= \begin{bmatrix} \mathbf{b}_1^t \\ \vdots \\ \mathbf{b}_K^t \end{bmatrix} \mathbf{x}_i \\
&= \mathbf{j}(\mathbf{b}^t\mathbf{x}_i) + \mathbf{j}(\mathbf{d}^t\mathbf{x}_i) + \beta(\mathbf{a}^t\mathbf{x}_i) \\
&= \mathbf{j}(\mathbf{b}^t\mathbf{x}_i + \mathbf{d}^t\mathbf{x}_i) + \beta. \qquad (4.40)
\end{aligned}$$

Since $\mathbf{Cj} = \mathbf{0}$ and from measurement error model (4.8) and result (4.40) it follows that

$$\begin{aligned}
\mathbf{C}\left(\mathbf{y}_i^\alpha - \mathbf{B}^t\mathbf{x}_i\right) &= \mathbf{C}\left(\mathbf{j}u_i + \mathbf{j}\gamma_i^\alpha + \beta + \varepsilon_i^\alpha - \mathbf{j}(\mathbf{b}^t + \mathbf{d}^t)\mathbf{x}_i - \beta\right) \\
&= \mathbf{C}\varepsilon_i^\alpha, \quad \mathbf{QED}. \qquad (4.41)
\end{aligned}$$

4.7.3 Proof of formula (4.26)

In this appendix it is proved that

$$\mathrm{E}_\alpha\mathrm{E}_s\mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{E}}}_{\mathbf{s}_k}^\alpha \mid \alpha, s) = \mathrm{E}_\alpha\mathrm{E}_s\mathbf{C}\mathbf{D}_R\mathbf{C}^t - \frac{1}{N^2}\sum_{i=1}^N \frac{\mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t}{\pi_i}.$$

The proof is given for an RBD. Results for a CRD follows analogous to the proof for an RBD with $J = 1$, $n_j = n$ and $n_{jk} = n_k$.
First we derive an expression for $\mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{E}}}_{\mathbf{s}_k}^\alpha \mid \alpha, s)$. Note that

$$\mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{E}}}_{\mathbf{s}_k}^\alpha \mid \alpha, s) = \mathbf{C}\mathrm{Cov}_\varepsilon(\hat{\hat{\mathbf{E}}}_{\mathbf{s}_k}^\alpha \mid \alpha, s)\mathbf{C}^t, \qquad (4.42)$$

with $\hat{\hat{\mathbf{E}}}_{\mathbf{s}_k}^\alpha = (\hat{\hat{E}}_1^\alpha, \ldots, \hat{\hat{E}}_K^\alpha)$ and

$$\hat{\hat{E}}_k^\alpha = \sum_{i=1}^n \left(\frac{\mathbf{p}_{ik}^t(\mathbf{y}_i^\alpha - \mathbf{B}^t\mathbf{x}_i)}{\pi_i\ N}\right), \quad k = 1, 2, \ldots, K.$$

According to the same derivations applied in appendix 3.9.5, it follows that the diagonal elements of $\mathrm{Cov}_\varepsilon(\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s)$ equals

$$\mathrm{Var}_\varepsilon(\hat{\hat{E}}^\alpha_k \mid \alpha, s) = \sum_{j=1}^{J} \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \sum_{i=1}^{n_j} \left(\frac{(y^\alpha_{ik} - \mathbf{b}^t_k \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{(y^\alpha_{i'k} - \mathbf{b}^t_k \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2$$
$$- \sum_{j=1}^{J} \frac{n_j}{(n_j-1)} \sum_{i=1}^{n_j} \left(\frac{(y^\alpha_{ik} - \mathbf{b}^t_k \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{(y^\alpha_{i'k} - \mathbf{b}^t_k \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2, \qquad (4.43)$$

and that the off-diagonal elements of $\mathrm{Cov}_\varepsilon(\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s)$ equals

$$\mathrm{Cov}_\varepsilon(\hat{\hat{E}}^\alpha_k \hat{\hat{E}}^\alpha_{k'} \mid \alpha, s) = \qquad (4.44)$$
$$-\sum_{j=1}^{J} \frac{n_j}{(n_j-1)} \sum_{i=1}^{n_j} \left(\frac{(y^\alpha_{ik} - \mathbf{b}^t_k \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{(y^\alpha_{i'k} - \mathbf{b}^t_k \mathbf{x}_{i'})}{N\pi_{i'}} \right) \left(\frac{(y^\alpha_{ik'} - \mathbf{b}^t_{k'} \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{(y^\alpha_{i'k'} - \mathbf{b}^t_{k'} \mathbf{x}_{i'})}{N\pi_{i'}} \right).$$

Results (4.43) and (4.44) can be written in matrix notation as follows

$$\mathrm{Cov}_\varepsilon(\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s) = \mathbf{D}_R$$
$$-\sum_{j=1}^{J} \frac{n_j}{(n_j-1)} \sum_{i=1}^{n_j} \left(\frac{(\mathbf{y}^\alpha_i - \mathbf{B}^t \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{(\mathbf{y}^\alpha_{i'} - \mathbf{B}^t \mathbf{x}_{i'})}{N\pi_{i'}} \right) \left(\frac{(\mathbf{y}^\alpha_i - \mathbf{B}^t \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{(\mathbf{y}^\alpha_{i'} - \mathbf{B}^t \mathbf{x}_{i'})}{N\pi_{i'}} \right)^t,$$

where $\mathbf{D}_R$ denotes a $K \times K$ diagonal matrix with elements d_{k_R} defined by (4.28). According to (4.23) it follows that

$$\mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s) = \mathbf{C}\mathbf{D}_R\mathbf{C}^t$$
$$-\sum_{j=1}^{J} \frac{n_j}{(n_j-1)} \sum_{i=1}^{n_j} \left(\frac{\mathbf{C}\varepsilon^\alpha_i}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\varepsilon^\alpha_{i'}}{N\pi_{i'}} \right) \left(\frac{\mathbf{C}\varepsilon^\alpha_i}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\varepsilon^\alpha_{i'}}{N\pi_{i'}} \right)^t. \qquad (4.45)$$

In order to take the expectation with respect to the sampling design of the second term of (4.45) we distinguished in section 3.6 between RBD's where the block variables are directly linked with the sampling scheme (e.g. if strata, clusters or PSU's as block variables) and RBD's where the block variables are not directly linked with the sampling scheme (e.g. if interviewers are block variables). In appendix 3.9.6 it is proved for these situations that

$$\mathrm{E}_\alpha \mathrm{E}_s \sum_{j=1}^{J} \frac{n_j}{(n_j-1)} \sum_{i=1}^{n_j} \left(\frac{\mathbf{C}\varepsilon^\alpha_i}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\varepsilon^\alpha_{i'}}{N\pi_{i'}} \right) \left(\frac{\mathbf{C}\varepsilon^\alpha_i}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{\mathbf{C}\varepsilon^\alpha_{i'}}{N\pi_{i'}} \right)^t$$
$$= \frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t}{\pi_i}. \qquad (4.46)$$

Now it follows from (4.46) that

$$\mathrm{E}_\alpha \mathrm{E}_s \mathrm{Cov}_\varepsilon(\mathbf{C}\hat{\hat{\mathbf{E}}}^\alpha_{\mathbf{s}_k} \mid \alpha, s) = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{C}\mathbf{D}_R\mathbf{C}^t - \frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{C}\mathbf{\Sigma}_i\mathbf{C}^t}{\pi_i}, \quad \textbf{QED}.$$

Chapter 5

Completely randomized designs

5.1 Introduction

In chapter 3 and 4 a general framework for the analysis of embedded experiments has been derived. This chapter further elaborates on the specific details of CRD's. Expressions for the variances $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ are derived for CRD's under general sampling designs in section 5.2. These expressions are used to evaluate the efficiency of RBD's under different sampling designs in chapter 6. In section 5.3, it is shown under which conditions the design-based Wald statistic for a CRD is equal to the $F-$statistic of the one-way ANOVA in a model-based analysis. A design-based $t-$statistic is derived for the analysis of the embedded two-treatment experiment as a special case of CRD's where $K = 2$ in section 5.4.

5.2 Variance of treatment effects

5.2.1 The Horvitz-Thompson estimator

In section 3.5 the Horvitz-Thompson estimator $\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ is proposed as a design-unbiased estimator for $\bar{\mathbf{Y}}$, which contains the population means under the K different treatments of the experiment. Expressions for the elements of vector $\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ are given by (3.21) or (3.23). The covariance matrix of the $K-1$ contrasts of $\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ is given by $\mathbf{CVC}^t$, defined by (3.38) in section 3.6. With the exception of a small covariance term of the additive treatment effects, which will generally be negligible or even zero, it followed that $\mathbf{CVC}^t = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{CDC}^t$. The elements d_k of diagonal matrix $\mathbf{D}$ are defined by (3.32). Design-unbiased estimators for d_k are defined by (3.57). We can further evaluate the variance of the estimated treatment effects by working out the expectations of $\mathbf{D}$ with respect to the sampling design and the measurement error model.

The Horvitz-Thompson estimator for the population mean of the intrinsic values $\bar{U}$ based on the n elements of sample s is given by

$$\hat{\hat{U}}_{\mathbf{s}} = \frac{1}{N}\sum_{i=1}^{n}\frac{u_i}{\pi_i}.$$

Let $\bar{U}_k$ denote the population mean of the intrinsic values biased with the treatment effect β_k of treatment k, i.e. $\bar{U}_k = \bar{U} + \beta_k$. The Horvitz-Thompson estimator for $\bar{U}_k$, based on the n elements of sample s is defined as

$$\hat{\bar{U}}_{k_{\mathbf{s}}} = \frac{1}{N}\sum_{i=1}^{n}\frac{(u_i + \beta_k)}{\pi_i}.$$

Then

$$\text{Var}(\hat{\bar{U}}_{k_{\mathbf{s}}}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(u_i + \beta_k)(u_{i'k} + \beta_k)}{\pi_i\pi_{i'}} \tag{5.1}$$

denotes the real variance of $\hat{\bar{U}}_{k_{\mathbf{s}}}$, i.e. the variance with respect to the sampling design used to draw sample s and

$$\tilde{\text{Var}}(\hat{\bar{U}}_{k_{\mathbf{s}}}) = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{n(u_i + \beta_k)^2}{\pi_i N^2} - \bar{U}_k^2\right) \tag{5.2}$$

denotes the variance of $\hat{\bar{U}}_{k_{\mathbf{s}}}$ as if the sample s has been drawn with replacement with selection probabilities π_i/n. Under the basic measurement error model (3.1) in section 3.2, it is proved in appendix 5.5.1 that:

$$\text{E}_\alpha\text{E}_s d_k = \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\text{Var}}(\hat{\bar{U}}_{k_{\mathbf{s}}}) - \frac{1}{n}\text{Var}(\hat{\bar{U}}_{k_{\mathbf{s}}})\right) + \frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{\pi_i N^2}, \tag{5.3}$$

where σ_{ik}^2 denotes the $k-$th diagonal element of $\mathbf{\Sigma}_i$ defined in (3.3) in section 3.2. An expression of the covariance matrix of $\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ is obtained if the elements $\text{E}_\alpha\text{E}_s d_k$ defined in (5.3) are substituted into the expression of $\mathbf{CVC}^t$ in (3.38) in section 3.6.1.

An expression for the design-based Wald statistic for a CRD based on the Horvitz-Thompson estimator follows directly from (3.80) in section 3.7 and reads as

$$W = \sum_{k=1}^{K}\frac{n_k}{\hat{S}_k^2}\hat{\bar{Y}}_k^{\alpha 2} - \frac{1}{\sum_{k=1}^{K}\frac{n_k}{\hat{S}_k^2}}\left(\sum_{k=1}^{K}\frac{n_k}{\hat{S}_k^2}\hat{\bar{Y}}_k^{\alpha}\right)^2, \tag{5.4}$$

where $\hat{\bar{Y}}_k^{\alpha}$ is defined in (3.21) or (3.23) in section 3.5 and $\hat{S}_k^2$ is defined in (3.57) in section 3.6.

5.2.2 The generalized regression estimator

In section 4.3 the generalized regression estimator $\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}$ is proposed as an approximately design-unbiased estimator for $\bar{\mathbf{Y}}$. Expressions for the elements of vector $\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}$ are given by (4.17). The covariance matrix of the $K-1$ contrasts of the first-order Taylor series approximation of $\hat{\bar{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}$ is given by $\text{E}_\alpha\text{E}_s\mathbf{CD}_R\mathbf{C}^t$. The elements d_{k_R} of the diagonal matrix $\mathbf{D}_R$ are defined by (4.27) in section 4.4. Design-unbiased estimators for d_{k_R} are defined by (4.30) in section 4.4. In this section, the variance of the estimated treatment effects for the generalized regression estimator is derived by taking the expectations with respect to the sampling design and the measurement error model of d_{k_R}.

Let $\hat{\hat{U}}_{R_{\mathbf{s}}}$ denote the generalized regression estimator for $\bar{U}$ based on the n sampling units of s. The first-order Taylor series approximation of $\hat{\hat{U}}_{R_{\mathbf{s}}}$ is given by

$$\hat{\hat{U}}_{R_{\mathbf{s}}} \doteq \hat{\hat{U}}_{\mathbf{s}} + \mathbf{b}^t(\bar{\mathbf{X}} - \hat{\hat{\mathbf{X}}}_{\mathbf{s}}) = \hat{\hat{E}}_{\mathbf{s}} + \mathbf{b}^t\bar{\mathbf{X}}, \tag{5.5}$$

with

$$\hat{\hat{E}}_{\mathbf{s}} = \hat{\hat{U}}_{\mathbf{s}} - \mathbf{b}^t\hat{\hat{\mathbf{X}}}_{\mathbf{s}} = \frac{1}{N}\sum_{i=1}^{n}\frac{(u_i - \mathbf{b}^t\mathbf{x}_i)}{\pi_i}. \tag{5.6}$$

Here $\hat{\hat{\mathbf{X}}}_{\mathbf{s}}$ is the Horvitz-Thompson estimator of $\bar{\mathbf{X}}$ based on the n sampling units of s and $\mathbf{b}$ the regression coefficients of the linear regression model (4.2) in section 4.2. The design variance of the first-order Taylor series approximation of $\hat{\hat{U}}_{R_{\mathbf{s}}}$ is given by

$$\mathrm{Var}(\hat{\hat{E}}_{\mathbf{s}}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(u_i - \mathbf{b}^t\mathbf{x}_i)}{\pi_i}\frac{(u_{i'} - \mathbf{b}^t\mathbf{x}_{i'})}{\pi_{i'}}. \tag{5.7}$$

Furthermore, let

$$\tilde{\mathrm{Var}}(\hat{\hat{E}}_{\mathbf{s}}) = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{n(u_i - \mathbf{b}^t\mathbf{x}_i)^2}{\pi_i N^2} - (\bar{U} - \mathbf{b}^t\bar{\mathbf{X}})^2\right) \tag{5.8}$$

denote the variance of the first-order Taylor series approximation of $\hat{\hat{U}}_{R_{\mathbf{s}}}$ as if the sample s has been drawn with replacement with selection probabilities π_i/n.

Under the basic measurement error model (4.5) in section 4.2 it is proved in appendix 5.5.2 that:

$$\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} = \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\mathrm{Var}}(\hat{\hat{E}}_{\mathbf{s}}) - \frac{1}{n}\mathrm{Var}(\hat{\hat{E}}_{\mathbf{s}})\right) + \frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{\pi_i N^2}. \tag{5.9}$$

Note that $\mathrm{Var}(\hat{\hat{E}}_{\mathbf{s}})$ and $\tilde{\mathrm{Var}}(\hat{\hat{E}}_{\mathbf{s}})$ do not depend on treatment effect β_k. Therefore $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ only depends on treatment k through the variance of the measurement errors σ_{ik}^2.

For the generalized regression estimator based on the common mean model (or the extended Horvitz-Thompson estimator) it follows from (4.2) in section 4.2 that $\mathbf{b} = \bar{U}$. Consequently

$$\mathrm{Var}(\hat{\hat{E}}_{\mathbf{s}}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(u_i - \bar{U})}{\pi_i}\frac{(u_{i'} - \bar{U})}{\pi_{i'}},$$

and

$$\tilde{\mathrm{Var}}(\hat{\hat{E}}_{\mathbf{s}}) = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{n(u_i - \bar{U})^2}{\pi_i N^2}\right).$$

For situations where $\hat{N} = N$ it follows that the Horvitz-Thompson estimator coincides with the extended Horvitz-Thompson estimator, e.g. CRD's embedded in a simple random sampling design or a stratified simple random sampling design with proportional allocation (see also section 4.4). Under this situation, $\mathrm{Var}(\hat{\hat{U}}_{k_{\mathbf{s}}})$ and $\tilde{\mathrm{Var}}(\hat{\hat{U}}_{k_{\mathbf{s}}})$ of the Horvitz-Thompson estimator equals $\mathrm{Var}(\hat{\hat{E}}_{\mathbf{s}})$ and $\tilde{\mathrm{Var}}(\hat{\hat{E}}_{\mathbf{s}})$ of the extended Horvitz-Thompson estimator. In this situation the variance (5.3) for the Horvitz-Thompson estimator doesn't depend on the treatment effects β_k. In the case of simple random sampling, for example, it follows for the Horvitz-Thompson estimator as well as the extended Horvitz-Thompson estimator that

$$\text{Var}(\hat{\hat{U}}_{\mathbf{s}}) = \frac{(1-f)}{n}\frac{1}{(N-1)}\sum_{i=1}^{N}(u_i - \bar{U})^2,$$

and

$$\tilde{\text{Var}}(\hat{\hat{U}}_{\mathbf{s}}) = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{u_i^2}{N} - \bar{U}^2\right).$$

Recall that no second-order inclusion probabilities are required in the estimator for the variance-covariance matrix of the contrasts of the subsample estimates, derived in sections 3.6 and 4.4. As follows from result (5.3) and (5.9), second-order inclusion probabilities do appear in the real variance-covariance matrix of the contrasts of the estimated population means, though. Expressions for $\text{E}_\alpha\text{E}_s d_k$ and $\text{E}_\alpha\text{E}_s d_{k_R}$ under measurement error models with interviewer effects in the case of CRD's are given in chapter 6.

An approximately design-unbiased estimator for the Wald statistic based on the generalized regression estimator reads as

$$W_R = \sum_{k=1}^{K}\frac{n_k}{\hat{S}_{E_k}^2}\hat{\hat{Y}}_{k_R}^{\alpha 2} - \frac{1}{\sum_{k=1}^{K}\frac{n_k}{\hat{S}_{E_k}^2}}\left(\sum_{k=1}^{K}\frac{n_k}{\hat{S}_{E_k}^2}\hat{\hat{Y}}_{k_R}^{\alpha}\right)^2, \tag{5.10}$$

where $\hat{\hat{Y}}_{k_R}^{\alpha}$ is defined in (4.16) or (4.17) in section 4.3 and $\hat{S}_{E_k}^2$ is defined in (4.30) in section 4.4.

5.3 Equal population variances and self-weighted sampling designs

The variance estimation procedure can be further improved by utilizing the model assumptions of additive treatment effects and identical and independently distributed measurement errors. In the case of the Horvitz-Thompson estimator,

$$\hat{S}_k^2 = \frac{1}{(n_k-1)}\sum_{i=1}^{n_k}\left(\frac{ny_{ik}^{\alpha}}{\pi_i N} - \frac{1}{n_k}\sum_{i'=1}^{n_k}\frac{ny_{i'k}^{\alpha}}{\pi_{i'} N}\right)^2 \tag{5.11}$$

defined in (3.57) in section 3.6.2 is an estimator of

$$S_k^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}\left(\frac{ny_{ik}^{\alpha}}{\pi_i N} - \frac{1}{n}\sum_{i'=1}^{n}\frac{ny_{i'k}^{\alpha}}{\pi_{i'} N}\right)^2 \tag{5.12}$$

defined in (3.32) in section 3.6.1. Parameter (5.12) can be regarded as the population variance of target variable y_{ik}^{α} (weighted with a factor $n/(\pi_i N)$) of the n individuals of sample s observed by means of treatment k. It is assumed that the covariance matrix in the measurement error model (3.1), section 3.2, is equal to $\mathbf{\Sigma}_i = \sigma_i^2\mathbf{I}$. Under this assumption it follows that the expectation of S_k^2 with respect to the measurement error model and the sampling design only depends on treatment k through the additive treatment effect β_k. Consequently, it follows that the expected values of these population variances are equal, i.e. $\mathrm{E}_\alpha\mathrm{E}_s S_1^2 = \mathrm{E}_\alpha\mathrm{E}_s S_2^2 = \ldots = \mathrm{E}_\alpha\mathrm{E}_s S_K^2 = \mathrm{E}_\alpha\mathrm{E}_s S^2$. Under this assumption a more efficient estimator for the population variance $\mathrm{E}_\alpha\mathrm{E}_s S^2$ is obtained from the pooled variance estimator

$$\hat{S}^2 = \frac{1}{n-K}\left(\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(\frac{ny_{ik}^\alpha}{\pi_i N} - \frac{1}{n_k}\sum_{i'=1}^{n_k}\frac{ny_{i'k}^\alpha}{\pi_{i'}N}\right)^2\right). \tag{5.13}$$

The pooled variance estimator can also be written as

$$\hat{S}^2 = \frac{1}{n-K}\left(\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(\frac{ny_{ik}^\alpha}{\pi_i N}\right)^2 - \sum_{k=1}^{K}\frac{1}{n_k}\left(\sum_{i=1}^{n_k}\frac{ny_{ik}^\alpha}{\pi_i N}\right)^2\right).$$

In this situation we can evaluate the Wald statistic (5.4) as follows:

$$\begin{aligned} W &= \sum_{k=1}^{K}\frac{n_k}{\hat{S}_k^2}\hat{\bar{Y}}_k^{\alpha 2} - \frac{1}{\sum_{k=1}^{K}\frac{n_k}{\hat{S}_k^2}}\left(\sum_{k=1}^{K}\frac{n_k}{\hat{S}_k^2}\hat{\bar{Y}}_k^{\alpha}\right)^2 \\ &= \frac{1}{\hat{S}^2}\left(\sum_{k=1}^{K}n_k\hat{\bar{Y}}_k^{\alpha 2} - \frac{1}{n}\left(\sum_{k=1}^{K}n_k\hat{\bar{Y}}_k^{\alpha}\right)^2\right) \\ &= \frac{1}{\hat{S}^2}\left(\sum_{k=1}^{K}\frac{1}{n_k}\left(\sum_{i=1}^{n_k}\frac{ny_{ik}^\alpha}{\pi_i N}\right)^2 - \frac{1}{n}\left(\sum_{k=1}^{K}\sum_{i=1}^{n_k}\frac{ny_{ik}^\alpha}{\pi_i N}\right)^2\right). \end{aligned} \tag{5.14}$$

Let $\tilde{y}_{ik}^\alpha$ denote the observations y_{ik}^α weighted with a factor $n/(\pi_i N)$, i.e. $\tilde{y}_{ik}^\alpha = (ny_{ik}^\alpha)/(\pi_i N)$. Note that the expression within the exterior set of parenthesis in (5.13) corresponds with the sum of squares within treatments of the one way ANOVA of $\tilde{y}_{ik}^\alpha$. In the analysis of variance, this term is generally called the error sum of squares, and is usually abbreviated as SS_E. Consequently, $\hat{S}^2$ in (5.13) is equal to the mean squares within treatments, also called the error mean squares, and is usually abbreviated as MS_E. Also note that the expression within the exterior set of parenthesis in (5.14) corresponds with the sum of squares between treatments of the one way ANOVA of $\tilde{y}_{ik}^\alpha$, also called the sums of squares for testing the hypothesis of no treatment effects and abbreviated as SS_T. In a model-based analysis, the hypothesis of no treatment effects can be tested by means of the $F-$statistic

$$F = \frac{SS_T/(K-1)}{SS_E/(n-K)} = \frac{MS_T}{MS_E}, \tag{5.15}$$

where MS_T denotes the mean squares between treatments. Under the assumption that the observations are normally and independently distributed, the $F-$statistic follows under the null hypothesis of no treatment effects a $F-$distribution with $K-1$ and $n-K$ degrees of freedom: $F \simeq \mathcal{F}_{(n-K)}^{(K-1)}$. The hypothesis of no treatment effects can be tested with the $F-$test of size γ, with statistic (5.15) and critical region $[\mathcal{F}_{(n-K)}^{(K-1)}(1-\gamma), \infty)$.

From (5.14) it follows that if the pooled variance estimator is used, then $W/(K-1)$ corresponds with the $F-$statistic (5.15), where the observations y_{ik}^α are weighted with a factor

$n/(\pi_i N)$. If sample s is drawn by means of a self-weighted sampling design, i.e. $\pi_i = n/N$, and the pooled variance estimator (5.13) is used, then the Wald statistic can be further simplified to

$$W = \frac{1}{\hat{S}^2}\left(\sum_{k=1}^{K}\frac{1}{n_k}\left(\sum_{i=1}^{n_k} y_{ik}^{\alpha}\right)^2 - \frac{1}{n}\left(\sum_{k=1}^{K}\sum_{i=1}^{n_k} y_{ik}^{\alpha}\right)^2\right),$$

with

$$\hat{S}^2 = \frac{1}{n-K}\left(\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(y_{ik}^{\alpha} - \frac{1}{n_k}\sum_{i'=1}^{n_k} y_{i'k}^{\alpha}\right)^2\right). \tag{5.16}$$

It follows that test statistic $W/(K-1)$ is equal to the $F-$statistic (5.15).

Under the assumption of normally and independently distributed observations, it follows that under the null hypothesis $F \simeq \mathcal{F}_{[n-K]}^{[K-1]}$. In the case of simple random sampling test statistic W is under the null hypothesis asymptotically chi-squared distributed with $K-1$ degrees of freedom. If n tends to infinity, then $\mathcal{F}_{[n-K]}^{[K-1]}$ tends to $\chi^2_{[K-1]}/(K-1)$ and consequently the test statistic W and the $F-$statistic have the same asymptotical distribution.

In section 4.4 it is derived for the generalized regression estimator that

$$\hat{S}_{E_k}^2 = \frac{1}{n_k-1}\sum_{i=1}^{n_k}\left(\frac{n(y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_i)}{N\pi_i} - \frac{1}{n_k}\sum_{i'=1}^{n_k}\frac{n(y_{i'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2 \tag{5.17}$$

is an approximately design-unbiased estimator for

$$S_{E_k}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{n(y_{ik}^{\alpha} - \mathbf{b}_k^{t}\mathbf{x}_i)}{N\pi_i} - \frac{1}{n}\sum_{i'=1}^{n}\frac{n(y_{i'k}^{\alpha} - \mathbf{b}_k^{t}\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2. \tag{5.18}$$

The residuals $(y_{ik}^{\alpha} - \mathbf{b}_k^t\mathbf{x}_i)$, however, only depend on the $k-$th treatment through the measurement error $\varepsilon_{ik}^{\alpha}$ (see (5.36) in appendix 5.5.2). If it is assumed that the covariance matrix of measurement error model (4.5) in section 4.2 is equal to $\mathbf{\Sigma}_i = \sigma_i^2\mathbf{I}$, then it follows that $\mathrm{E}_{\alpha}\mathrm{E}_s S_{E_k}^2$ are identical parameters, i.e. $\mathrm{E}_{\alpha}\mathrm{E}_s S_{E_1}^2 = \ldots = \mathrm{E}_{\alpha}\mathrm{E}_s S_{E_K}^2 = \mathrm{E}_{\alpha}\mathrm{E}_s S_{E}^2$. Since $\hat{S}_{E_k}^2$, $k = 1, \ldots, K$ are K estimators for the same parameter $\mathrm{E}_{\alpha}\mathrm{E}_s S_E^2$, a more efficient estimator for S_E^2 is the pooled variance estimator

$$\hat{S}_E^2 = \frac{1}{n}\left(\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(\frac{n(y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_i)}{N\pi_i} - \frac{1}{n}\sum_{k'=1}^{K}\sum_{i'=1}^{n_k}\frac{n(y_{i'k'}^{\alpha} - \hat{\mathbf{b}}_{k'}^{\alpha^t}\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2\right). \tag{5.19}$$

An alternative pooled variance estimator resembling the mean square error of a one way ANOVA is

$$\hat{S}_E^2 = \frac{1}{(n-K)}\left(\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(\frac{n(y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_i)}{N\pi_i} - \frac{1}{n_k}\sum_{i'=1}^{n_k}\frac{n(y_{i'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2\right). \tag{5.20}$$

Under this pooled variance estimator, the Wald statistic (5.10) can be expressed as

$$W = \frac{1}{\hat{S}_E^2}\left(\sum_{k=1}^{K} n_k \hat{\hat{Y}}_{k_R}^{\alpha^2} - \frac{1}{n}\left(\sum_{k=1}^{K} n_k \hat{\hat{Y}}_{k_R}^{\alpha}\right)^2\right).$$

5.4 Two-sample problem

The two-sample problem or the two-treatment embedded experiment can be considered as a special case of a CRD with $K = 2$. The objective of the two-sample problem is to investigate whether there is a significant difference between two population parameters $\mathrm{E}_\alpha(\bar{Y}_1^\alpha)$ and $\mathrm{E}_\alpha(\bar{Y}_2^\alpha)$ by testing the following hypotheses

$$\begin{aligned} H_0 : &\quad \mathrm{E}_\alpha(\bar{Y}_1^\alpha) = \mathrm{E}_\alpha(\bar{Y}_2^\alpha), \\ H_1 : &\quad \mathrm{E}_\alpha(\bar{Y}_1^\alpha) \neq \mathrm{E}_\alpha(\bar{Y}_2^\alpha) \;\text{ or }\; \mathrm{E}_\alpha(\bar{Y}_1^\alpha) > \mathrm{E}_\alpha(\bar{Y}_2^\alpha) \;\text{ or }\; \mathrm{E}_\alpha(\bar{Y}_1^\alpha) < \mathrm{E}_\alpha(\bar{Y}_2^\alpha). \end{aligned} \tag{5.21}$$

To draw inferences about the finite population parameters $\mathrm{E}_\alpha(\bar{Y}_1^\alpha)$ and $\mathrm{E}_\alpha(\bar{Y}_2^\alpha)$ a design-based $t-$statistic derived under the randomization mechanism of the sampling design as well as the experimental design should be applied to test hypotheses (5.21). Such a design-based $t-$statistic follows directly from the results obtained for the design-based Wald statistic in chapters 3 and 4. The enumerator of this $t-$statistic contains the difference between a design-unbiased estimator for $\mathrm{E}_\alpha(\bar{Y}_1^\alpha)$ and $\mathrm{E}_\alpha(\bar{Y}_2^\alpha)$. The denominator contains a design-unbiased estimator for the standard deviation of the difference between these two estimators.

Under the basic measurement error model (3.1) in section 3.2 where $K = 2$, the Horvitz-Thompson estimator can be used for the estimation of the unknown parameters in the design-based $t-$statistic. As a result we obtain

$$t = \frac{\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha}{\sqrt{\hat{\mathrm{Var}}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha)}}, \tag{5.22}$$

to test hypotheses (5.21) The estimators $\hat{\bar{Y}}_1^\alpha$ and $\hat{\bar{Y}}_2^\alpha$ are defined in (3.21) or (3.23) in section 3.5. An expression for $\mathrm{Var}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha)$ follows directly from $\mathrm{E}_\alpha\mathrm{E}_s\mathbf{CDC}^t$ for $K = 2$ and the expression for the diagonal elements $\mathrm{E}_\alpha\mathrm{E}_s d_k$ from formula (5.3):

$$\mathrm{Var}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha) \;=\; \sum_{k=1}^{2} \frac{n}{(n-1)} \frac{n}{n_k} \left(\tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_s}) - \frac{1}{n}\mathrm{Var}(\hat{\bar{U}}_{k_s}) \right) + \frac{n}{n_k} \sum_{i=1}^{N} \frac{\sigma_{ik}^2}{\pi_i N^2}, \tag{5.23}$$

where $\mathrm{Var}(\hat{\bar{U}}_{k_s})$ and $\tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_s})$ are defined by (5.1) and (5.2), respectively. An unbiased estimator for $\mathrm{Var}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha)$ follows from (3.57) in section 3.6.2 and is given by

$$\hat{\mathrm{Var}}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha) \;=\; \sum_{k=1}^{2} \frac{1}{n_k} \frac{1}{(n_k-1)} \sum_{i=1}^{n_k} \left(\frac{n y_{ik}^\alpha}{N\pi_i} - \frac{1}{n_k} \sum_{i'=1}^{n_k} \frac{n y_{i'k}^\alpha}{N\pi_{i'}} \right)^2.$$

In the case of a self-weighted sampling design, i.e. $\pi_i = n/N$, $\hat{\bar{Y}}_1^\alpha$ and $\hat{\bar{Y}}_2^\alpha$ reduces to the unweighted sample means of the subsamples s_1 and s_2. An unbiased estimator for $\mathrm{Var}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha)$ is given by

$$\hat{\mathrm{Var}}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha) = \frac{1}{n_1} \frac{1}{(n_1-1)} \sum_{i=1}^{n_1} \left(y_{i1}^\alpha - \frac{1}{n_1} \sum_{i'=1}^{n_1} y_{i'1}^\alpha \right)^2 + \frac{1}{n_2} \frac{1}{(n_2-1)} \sum_{i=1}^{n_2} \left(y_{i2}^\alpha - \frac{1}{n_2} \sum_{i'=1}^{n_2} y_{i'2}^\alpha \right)^2.$$

In this case $t-$statistic (5.22) reduces to the Welch's $t'-$test statistic (Miller, 1986).

Under the assumption of identical and independently distributed measurement errors a more efficient estimator for the population variance is obtained from the pooled variance estimator

$$\hat{S}^2 = \frac{1}{n-2}\left(\sum_{i=1}^{n_1}\left(\frac{ny_{ik}^\alpha}{\pi_i N} - \frac{1}{n_1}\sum_{i'=1}^{n_1}\frac{ny_{i'k}^\alpha}{\pi_{i'} N}\right)^2 + \sum_{i=1}^{n_2}\left(\frac{ny_{ik}^\alpha}{\pi_i N} - \frac{1}{n_2}\sum_{i'=1}^{n_2}\frac{ny_{i'k}^\alpha}{\pi_{i'} N}\right)^2\right). \tag{5.24}$$

An unbiased estimator for $\text{Var}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha)$ is given by

$$\hat{\text{Var}}(\hat{\bar{Y}}_1^\alpha - \hat{\bar{Y}}_2^\alpha) = \frac{1}{\hat{S}^2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

It follows from (3.21) that $\hat{\bar{Y}}_1^\alpha$ and $\hat{\bar{Y}}_2^\alpha$ are the weighted sample means of s_1 and s_2 where the observations y_{i1}^α and y_{i2}^α are weighted with a factor $n/(\pi_i N)$. As a result, in the case of the pooled variance estimator, $t-$statistic (5.22) is equal to the ordinary model-based $t-$statistic where the observations are weighted with a factor $n/(\pi_i N)$ regardless the second-order inclusion probabilities of the applied sampling design. In the case of a self-weighted sampling design $\hat{\bar{Y}}_1^\alpha$ and $\hat{\bar{Y}}_2^\alpha$ reduce to the sample means of the subsamples s_1 and s_2, respectively. The pooled variance estimator is equal to

$$\hat{S}^2 = \frac{1}{n-2}\left(\sum_{i=1}^{n_1}\left(y_{ik}^\alpha - \frac{1}{n_1}\sum_{i'=1}^{n_1} y_{i'k}^\alpha\right)^2 + \sum_{i=1}^{n_2}\left(y_{ik}^\alpha - \frac{1}{n_2}\sum_{i'=1}^{n_2} y_{i'k}^\alpha\right)^2\right). \tag{5.25}$$

As a result, in the case of a self-weighted sampling design and the pooled variance estimator, the design-based $t-$statistic (5.22) reduces to the ordinary model-based $t-$statistic, regardless the second-order inclusion probabilities of the applied sampling design.

Under the basic measurement error model (4.5) in section 4.2 with $K = 2$, the generalized regression estimator can be used for the estimation of the unknown parameters in the design-based $t-$statistic. It follows that:

$$t = \frac{\hat{\bar{Y}}_{1_R}^\alpha - \hat{\bar{Y}}_{2_R}^\alpha}{\sqrt{\hat{\text{Var}}(\hat{\bar{Y}}_{1_R}^\alpha - \hat{\bar{Y}}_{2_R}^\alpha)}}. \tag{5.26}$$

Expressions for $\hat{\bar{Y}}_{1_R}^\alpha$ and $\hat{\bar{Y}}_{2_R}^\alpha$ are given by (4.16) or (4.17) in section 4.3. It follows from $\text{E}_\alpha\text{E}_s\mathbf{C}\mathbf{D}_R\mathbf{C}^t$ for $K = 2$ and the expression for the diagonal elements $\text{E}_\alpha\text{E}_s d_{k_R}$ in (5.9) section 4.4 that

$$\text{Var}(\hat{\bar{Y}}_{1_R}^\alpha - \hat{\bar{Y}}_{2_R}^\alpha) = \sum_{k=1}^{2}\frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\text{Var}}(\hat{\bar{U}}_{k_{R_s}}) - \frac{1}{n}\text{Var}(\hat{\bar{U}}_{k_{R_s}})\right) + \frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{\pi_i N^2}, \tag{5.27}$$

where $\text{Var}(\hat{\bar{U}}_{k_{R_s}})$ and $\tilde{\text{Var}}(\hat{\bar{U}}_{k_{R_s}})$ are defined defined in (5.7) and (5.8), respectively. An approximately unbiased estimator for $\text{Var}(\hat{\bar{Y}}_{1_R}^\alpha - \hat{\bar{Y}}_{2_R}^\alpha)$ follows from (4.30) in section 4.4 and is given by

$$\hat{\text{Var}}(\hat{\bar{Y}}_{1_R}^\alpha - \hat{\bar{Y}}_{2_R}^\alpha) = \sum_{k=1}^{2}\frac{1}{n_k}\frac{1}{(n_k-1)}\sum_{i=1}^{n_k}\left(\frac{n(y_{ik}^\alpha - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_i)}{N\pi_i} - \frac{1}{n_k}\sum_{i'=1}^{n_k}\frac{n(y_{i'k}^\alpha - \hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2.$$

Under the assumption that the measurement errors are identically and independently distributed the pooled variance estimator (5.19) or (5.20) can be applied.

Hypotheses (5.21) can be tested with the design-based $t-$statistic (5.22) or (5.26). If s is drawn by means of simple random sampling it follows from sections 3.7.2 and 4.5 that the $t-$statistic has an asymptotically standard normal distribution. In the case of more complex sampling designs, it is assumed that a limit theorem holds true such that the $t-$statistic as an approximately standard normal distribution. This assumption has been confirmed by simulation studies for different sampling designs. As a result, the standard normal distribution can be used to construct critical regions, which yield very nearly $(1-\gamma)\%$ coverage, where γ denotes the size of the test.

5.5 Appendix

5.5.1 Proof of formula (5.3)

For the Horvitz-Thompson estimator it is proved that:

$$\mathrm{E}_\alpha \mathrm{E}_s d_k = \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_\mathbf{s}}) - \frac{1}{n}\mathrm{Var}(\hat{\bar{U}}_{k_\mathbf{s}})\right) + \frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{\pi_i N^2}.$$

Let $\hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}}$ denote the Horvitz-Thompson estimator of $\bar{Y}^{\alpha}_k$ based on the n sampling units of s. Then

$$\mathrm{Var}(\hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{y_{ik}^{\alpha}y_{i'k}^{\alpha}}{\pi_i\pi_{i'}}, \tag{5.28}$$

denotes the real design variance of $\hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}}$ and

$$\tilde{\mathrm{Var}}(\hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}}) = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{n{y_{ik}^{\alpha}}^2}{\pi_i N^2} - {\bar{Y}_k^{\alpha}}^2\right), \tag{5.29}$$

denotes the variance of $\hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}}$ as if sample s has been drawn with replacement with selection probabilities π_i/n.

The diagonal elements d_k are defined by (3.32) in section 3.6. Firstly we evaluate the expectation of d_k with respect to the sampling design as follows

$$\begin{aligned}
&\mathrm{E}_s\frac{1}{(n-1)}\frac{1}{n_k}\sum_{i=1}^{n}\left(\frac{ny_{ik}^{\alpha}}{\pi_i N} - \frac{1}{n}\sum_{i'=1}^{n}\frac{ny_{i'k}^{\alpha}}{\pi_{i'}N}\right)^2 \\
&= \mathrm{E}_s\frac{1}{(n-1)}\frac{1}{n_k}\sum_{i=1}^{n}\left(\frac{ny_{ik}^{\alpha}}{\pi_i N} - \bar{Y}_k^{\alpha} + \bar{Y}_k^{\alpha} - \hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}}\right)^2 \\
&= \frac{1}{(n-1)}\frac{1}{n_k}\mathrm{E}_s\left(\sum_{i=1}^{n}\left(\frac{ny_{ik}^{\alpha}}{\pi_i N} - \bar{Y}_k^{\alpha}\right)^2 - n\left(\hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}} - \bar{Y}_k^{\alpha}\right)^2\right) \\
&= \frac{1}{(n-1)}\frac{1}{n_k}\left(\mathrm{E}_s\left(\sum_{i=1}^{n}\frac{n^2{y_{ik}^{\alpha}}^2}{\pi_i^2N^2} - 2n\bar{Y}_k^{\alpha}\hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}} + n{\bar{Y}_k^{\alpha}}^2\right) - n\mathrm{Var}(\hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}})\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(n-1)}\frac{1}{n_k}\left(\sum_{i=1}^{N}\frac{n^2 y_{ik}^{\alpha 2}}{\pi_i N^2} - n\bar{Y}_k^{\alpha 2} - n\mathrm{Var}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}})\right)\\
&= \frac{n}{(n-1)}\frac{n}{n_k}\left(\frac{1}{n}\left(\sum_{i=1}^{N}\frac{n y_{ik}^{\alpha 2}}{\pi_i N^2} - \bar{Y}_k^{\alpha 2}\right) - \frac{1}{n}\mathrm{Var}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}})\right)\\
&= \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\mathrm{Var}}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}}) - \frac{1}{n}\mathrm{Var}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}})\right). \qquad (5.30)
\end{aligned}
$$

Now we will further evaluate the expectation with respect to the measurement error model under the basic measurement error model (3.1) in section 3.2. If the basic measurement error model (3.1) is substituted into the expression of $\mathrm{Var}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}})$ given by (5.28), then it follows from the model assumption (3.2) and (3.3) that

$$
\begin{aligned}
\mathrm{E}_\alpha \mathrm{Var}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}}) &= \mathrm{E}_\alpha \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(u_i+\beta_k+\varepsilon_{ik}^{\alpha})}{\pi_i}\frac{(u_{i'}+\beta_k+\varepsilon_{i'k}^{\alpha})}{\pi_{i'}}\\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(u_i+\beta_k)}{\pi_i}\frac{(u_{i'}+\beta_k)}{\pi_{i'}} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{(1-\pi_i)}{\pi_i}\sigma_{ik}^2\\
&= \mathrm{Var}(\hat{\bar{U}}_{k_{\mathbf{s}}}) + \frac{1}{N^2}\sum_{i=1}^{N}\frac{(1-\pi_i)}{\pi_i}\sigma_{ik}^2. \qquad (5.31)
\end{aligned}
$$

The cross-products $(u_i+\beta_k)\varepsilon_{ik}^{\alpha}$ cancels out due to model assumption (3.2). The cross-products $\varepsilon_{ik}^{\alpha}\varepsilon_{i'k}^{\alpha}$ cancels out since the measurement errors between individuals are independent according to model assumption (3.3). If the basic measurement error model (3.1) is substituted into the expression of $\tilde{\mathrm{Var}}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}})$ given by (5.29), it follows that

$$
\begin{aligned}
\mathrm{E}_\alpha \tilde{\mathrm{Var}}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}}) &= \mathrm{E}_\alpha \frac{1}{n}\left(\sum_{i=1}^{N}\frac{n(u_i+\beta_k+\varepsilon_{ik}^{\alpha})^2}{\pi_i N^2} - (\bar{U}+\beta_k+\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{ik}^{\alpha})^2\right)\\
&= \frac{1}{n}\left(\sum_{i=1}^{N}\frac{n(u_i+\beta_k)^2}{\pi_i N^2} - (\bar{U}+\beta_k)^2\right) + \frac{1}{n}\frac{1}{N^2}\sum_{i=1}^{N}\frac{(n-\pi_i)}{\pi_i}\sigma_{ik}^2\\
&= \tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_{\mathbf{s}}}) + \frac{1}{n}\frac{1}{N^2}\sum_{i=1}^{N}\frac{(n-\pi_i)}{\pi_i}\sigma_{ik}^2. \qquad (5.32)
\end{aligned}
$$

If results (5.31) and (5.32) are substituted into (5.30), then it follows that

$$
\begin{aligned}
\mathrm{E}_\alpha \mathrm{E}_s(d_k) &= \mathrm{E}_\alpha \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\mathrm{Var}}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}}) - \frac{1}{n}\mathrm{Var}(\hat{\hat{Y}}^{\alpha}_{k_{\mathbf{s}}})\right)\\
&= \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_{\mathbf{s}}}) - \frac{1}{n}\mathrm{Var}(\hat{\bar{U}}_{k_{\mathbf{s}}})\right) + \frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{\pi_i N^2}, \quad \textbf{QED}.
\end{aligned}
$$

5.5.2 Proof of formula (5.9)

For the generalized regression estimator it is proved that:

$$
\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} = \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\mathrm{Var}}(\hat{\bar{E}}_{\mathbf{s}}) - \frac{1}{n}\mathrm{Var}(\hat{\bar{E}}_{\mathbf{s}})\right) + \frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{\pi_i N^2}.
$$

Let $\hat{\bar{Y}}^{\alpha}_{k_{R_\mathbf{s}}}$ denote the generalized regression estimator of $\bar{Y}^{\alpha}_k$ based on the n sampling units of s. The first-order Taylor series approximation of $\hat{\bar{Y}}^{\alpha}_{k_{R_\mathbf{s}}}$ is given by:

$$\hat{\bar{Y}}^{\alpha}_{k_{R_\mathbf{s}}} \doteq \hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}} + \mathbf{b}^t_k(\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_\mathbf{s}) = \hat{\bar{E}}^{\alpha}_{k_\mathbf{s}} + \mathbf{b}^t_k\bar{\mathbf{X}},$$

where

$$\hat{\bar{E}}^{\alpha}_{k_\mathbf{s}} = \hat{\bar{Y}}^{\alpha}_{k_\mathbf{s}} - \mathbf{b}^t_k\hat{\bar{\mathbf{X}}}_\mathbf{s} = \frac{1}{N}\sum_{i=1}^{n}\frac{(y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i)}{\pi_i}.$$

The design variance of the first-order Taylor series approximation of $\hat{\bar{Y}}^{\alpha}_{k_{R_\mathbf{s}}}$ is given by

$$\mathrm{Var}(\hat{\bar{E}}^{\alpha}_{k_\mathbf{s}}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i)}{\pi_i}\frac{(y^{\alpha}_{i'k} - \mathbf{b}^t_k\mathbf{x}_{i'})}{\pi_{i'}}, \tag{5.33}$$

and

$$\tilde{\mathrm{Var}}(\hat{\bar{E}}^{\alpha}_{k_\mathbf{s}}) = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{n(y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i)^2}{\pi_i N^2} - (\bar{Y}^{\alpha}_k - \mathbf{b}^t_k\bar{\mathbf{X}})^2\right), \tag{5.34}$$

denotes the approximate variance of $\hat{\bar{Y}}^{\alpha}_{k_{R_\mathbf{s}}}$ as if sample s has been drawn with replacement with selection probabilities π_i/n.

The diagonal elements d_{k_R} are defined by (4.27) in section 4.4. Equivalent to the derivation of result (5.30) in appendix 5.5.1 it follows that the expectation of d_{k_R} with respect to the sampling design equals

$$\begin{aligned} &\mathrm{E}_s\frac{1}{(n-1)}\frac{1}{n_k}\sum_{i=1}^{n}\left(\frac{n(y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i)}{\pi_i N} - \frac{1}{n}\sum_{i'=1}^{n}\frac{n(y^{\alpha}_{i'k} - \mathbf{b}^t_k\mathbf{x}_{i'})}{\pi_{i'} N}\right)^2 \\ &= \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\mathrm{Var}}(\hat{\bar{E}}^{\alpha}_{k_\mathbf{s}}) - \frac{1}{n}\mathrm{Var}(\hat{\bar{E}}^{\alpha}_{k_\mathbf{s}})\right). \end{aligned} \tag{5.35}$$

Under the basic measurement error model (4.5) in section 4.2 and the condition that there exists a constant $H-$vector $\mathbf{a}$ such that $\mathbf{a}^t\mathbf{x}_i = 1$ for all $i \in U$, it follows from (4.39) in appendix 4.7.2 that $\mathbf{b}_k = \mathbf{b} + \mathbf{a}\beta_k$. Under the stated condition it follows for a basic measurement error model that

$$y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i = u_i - \mathbf{b}^t\mathbf{x}_i + \varepsilon^{\alpha}_{ik}. \tag{5.36}$$

If (5.36) is substituted into the expression of $\mathrm{Var}(\hat{\bar{E}}^{\alpha}_{k_\mathbf{s}})$ given by (5.33), then it follows from the measurement error model assumptions (4.6) and (4.7) that

$$\begin{aligned} \mathrm{E}_\alpha\mathrm{Var}(\hat{\bar{E}}^{\alpha}_{k_\mathbf{s}}) &= \mathrm{E}_\alpha\frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(u_i - \mathbf{b}^t\mathbf{x}_i + \varepsilon^{\alpha}_{ik})}{\pi_i}\frac{(u_{i'} - \mathbf{b}^t\mathbf{x}_{i'} + \varepsilon^{\alpha}_{ik})}{\pi_{i'}} \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(u_i - \mathbf{b}^t\mathbf{x}_i)}{\pi_i}\frac{(u_{i'} - \mathbf{b}^t\mathbf{x}_{i'})}{\pi_{i'}} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{(1-\pi_i)}{\pi_i}\sigma^2_{ik} \\ &= \mathrm{Var}(\hat{\bar{E}}_\mathbf{s}) + \frac{1}{N^2}\sum_{i=1}^{N}\frac{(1-\pi_i)}{\pi_i}\sigma^2_{ik}. \end{aligned} \tag{5.37}$$

If (5.36) is substituted into the expression of $\tilde{\text{Var}}(\hat{\bar{E}}^{\alpha}_{k_{\mathbf{s}}})$ given by (5.34), then it follows that

$$\begin{aligned} \text{E}_\alpha \tilde{\text{Var}}(\hat{\bar{E}}^{\alpha}_{k_{\mathbf{s}}}) &= \text{E}_\alpha \frac{1}{n}\left(\sum_{i=1}^{N} \frac{n(u_i - \mathbf{b}^t\mathbf{x}_i + \varepsilon^{\alpha}_{ik})^2}{\pi_i N^2} - \left(\bar{U} - \mathbf{b}^t\bar{\mathbf{X}} + \frac{1}{N}\sum_{i=1}^{N}\varepsilon^{\alpha}_{ik}\right)^2\right) \\ &= \frac{1}{n}\left(\sum_{i=1}^{N} \frac{n(u_i - \mathbf{b}^t\mathbf{x}_i)^2}{\pi_i N^2} - \left(\bar{U} - \mathbf{b}^t\bar{\mathbf{X}}\right)^2\right) + \frac{1}{n}\frac{1}{N^2}\sum_{i=1}^{N}\frac{(n-\pi_i)}{\pi_i}\sigma^2_{ik} \\ &= \tilde{\text{Var}}(\hat{\bar{E}}_{\mathbf{s}}) + \frac{1}{n}\frac{1}{N^2}\sum_{i=1}^{N}\frac{(n-\pi_i)}{\pi_i}\sigma^2_{ik}. \end{aligned} \tag{5.38}$$

If results (5.37) and (5.38) are substituted into (5.35), then it follows that

$$\begin{aligned} \text{E}_\alpha \text{E}_s(d_{k_R}) &= \text{E}_\alpha \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\text{Var}}(\hat{\bar{E}}^{\alpha}_{k_{\mathbf{s}}}) - \frac{1}{n}\text{Var}(\hat{\bar{E}}^{\alpha}_{k_{\mathbf{s}}})\right) \\ &= \frac{n}{(n-1)}\frac{n}{n_k}\left(\tilde{\text{Var}}(\hat{\bar{E}}_{\mathbf{s}}) - \frac{1}{n}\text{Var}(\hat{\bar{E}}_{\mathbf{s}})\right) + \frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma^2_{ik}}{\pi_i N^2}, \quad \textbf{QED}. \end{aligned}$$

Chapter 6

Randomized block designs

6.1 Introduction

Design-based methods for the analysis of randomized block designs embedded in complex sample surveys where each possible source of extraneous variation can be used as a block variable, are developed in chapters 3 and 4. To test the hypothesis of no treatment effects with respect to the finite population parameters, the analysis is based on a design-based Wald statistic derived under the randomization mechanism of the sampling design and the experimental design by means of the Horvitz-Thompson estimator (chapter 3) or the generalized regression estimator (chapter 4). Let $\hat{\bar{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ denote the Horvitz-Thompson estimator and $\hat{\bar{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$ the generalized regression estimator for the population means observed under the K different treatments. The covariance matrices of the $K-1$ contrasts of $\hat{\bar{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}$ and $\hat{\bar{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$ are given by $\mathrm{E}_{\alpha}\mathrm{E}_{s}\mathbf{C}\mathbf{D}\mathbf{C}^t$ and $\mathrm{E}_{\alpha}\mathrm{E}_{s}\mathbf{C}\mathbf{D}_R\mathbf{C}^t$, respectively. Here $\mathbf{D}$ and $\mathbf{D}_R$ are diagonal matrices with elements d_k defined in (3.34), section 3.6.1 and d_{k_R} defined in (4.28), section 4.4. Approximately design-unbiased estimators for $\mathrm{E}_{\alpha}\mathrm{E}_{s}d_k$ and $\mathrm{E}_{\alpha}\mathrm{E}_{s}d_{k_R}$ are given by $\hat{d}_k$ defined in (3.58), section 3.6.2 and $\hat{d}_{k_R}$ defined in (4.31), section 4.4 respectively.

This chapter elaborates on the specific details of RBD's. Expressions for $\mathrm{E}_{\alpha}\mathrm{E}_{s}d_k$ and $\mathrm{E}_{\alpha}\mathrm{E}_{s}d_{k_R}$ are derived by elaborating on the expectations with respect to the measurement error model and the sampling design. The comparison of these design variances obtained under RBD's with the variances obtained under CRD's in chapter 5 enables us to quantify the efficiency of blocking under different measurement error models and sampling designs.

In order to evaluate the expectation with respect to the measurement error model and the sampling design of $\mathrm{E}_{\alpha}\mathrm{E}_{s}d_k$ or $\mathrm{E}_{\alpha}\mathrm{E}_{s}d_{k_R}$ we distinguish between two situations. First, consider RBD's where the block variables are directly linked with structures of the sampling design. For example RBD's where strata, PSU's or clusters are used as block variables. Since the block variables coincide exactly with structures of the sampling design, the randomization mechanism of the experimental design coincides with the randomization mechanism of the sampling design. This simplifies the derivation of the design variances considerably. In section 6.2, 6.3 en 6.4

expressions for $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ are derived if PSU's, strata and clusters are used as block variables, respectively. Second, consider RBD's where block variables are not directly linked with structures of the sampling design. For example RBD's where interviewers are block variables. In these situations the randomization mechanism of the experimental design cuts arbitrarily through the randomization mechanism of the sampling design. This implies that blocks should be considered as domains if we evaluate the variance with respect to the sampling design. Sections 6.5 and 6.6 deal with RBD's where interviewers are block variables under simple random sampling and more complex sampling schemes, respectively. In section 6.7 we show how the variance estimation procedure can be improved by means of a pooled variance estimator. In section 6.8 a design-based $t-$statistic is derived for the analysis of the embedded two-treatment experiment as a special case of RBD's where $K = 2$.

6.2 Primary sampling units as block variables

Consider an experiment embedded in a two-stage sampling design conducted under the basic measurement error model (3.1) in section 3.2 or (4.5) in section 4.2. In the first stage of the sampling design, J_s PSU's of size N_j are drawn randomly from a finite population of J_u PSU's. In the second stage a sample s_j of size n_j SSU's is drawn from each of the J_s PSU's drawn in the first stage. According to the experimental design, the n_j SSU's within each PSU are randomized over the K different treatments. This randomization mechanism naturally leads to an RBD where PSU's correspond to the block variables. In this experimental design, a sample of J_s blocks is drawn from a finite population of J_u blocks according to the first stage of the sampling scheme.

Let π_j denote the first-order inclusion probability of PSU j in the first stage of the sample design, $\pi_{jj'}$ the second-order inclusion probability of PSU j and j' in the first stage, $\pi_{i|j}$ the first-order inclusion probability of SSU i in the second stage, conditional on the realization of the first stage and $\pi_{ii'|j}$ the second-order inclusion probability of SSU i and i' in the second stage, conditional on the realization of the first stage of the sample design. It follows that

$$\pi_i = \pi_j \pi_{i|j}, \tag{6.1}$$

$$\pi_{ii'} = \begin{cases} \pi_j \pi_{ii'|j} & \text{if } i \in j \text{ and } i' \in j \\ \pi_{jj'} \pi_{i|j} \pi_{i'|j'} & \text{if } i \in j \text{ and } i' \in j' \end{cases} . \tag{6.2}$$

Expressions for the design variances $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ are derived by taking the expectation with respect to the measurement error model and the sampling design of d_k and d_{k_R} for an RBD embedded in a two-stage sampling design where PSU's are block variables. In the case of the Horvitz-Thompson estimator, the diagonal elements d_k are defined by (3.34) in section 3.6.1. The number of blocks in (3.34) equals J_s. Let $\bar{U}_j$ denote the population mean of the intrinsic values in block or PSU j. Then $\bar{U}_{jk} = \bar{U}_j + \beta_k$ denotes the population mean biased with the treatment effect β_k in the $j-$th block. The Horvitz-Thompson estimator for the mean

$\bar{U}_{jk}$ of block or PSU j based on the n_j SSU's drawn from PSU j is defined as

$$\hat{\hat{U}}_{jk_{\mathbf{s}}} = \frac{1}{N_j} \sum_{i=1}^{n_j} \frac{(u_i + \beta_k)}{\pi_{i|j}}. \tag{6.3}$$

Then

$$\mathrm{Var}(\hat{\hat{U}}_{jk_{\mathbf{s}}}) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j}) \frac{(u_i + \beta_k)}{\pi_{i|j}} \frac{(u_{i'} + \beta_k)}{\pi_{i'|j}} \tag{6.4}$$

denotes the design variance of the Horvitz-Thompson estimator $\hat{\hat{U}}_{jk_{\mathbf{s}}}$, i.e. the variance with respect to the second stage of the sampling design used to draw n_j SSU's from PSU j and

$$\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_{\mathbf{s}}}) = \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j (u_i + \beta_k)^2}{\pi_{i|j} N_j^2} - \bar{U}_{jk}^2 \right) \tag{6.5}$$

denotes the variance of $\hat{\hat{U}}_{jk_{\mathbf{s}}}$ as if n_j SSU's from PSU j were drawn with replacement with selection probabilities $\pi_{i|j}/n_j$ in the second stage of the sample design. Under the basic measurement error model (3.1) in section 3.2 it is proved in appendix 6.9.1 for an RBD embedded in a two-stage sampling design where PSU's are block variables that

$$\begin{aligned} \mathrm{E}_\alpha \mathrm{E}_s d_k &= \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{n_j}{(n_j - 1)} \frac{n_j}{n_{jk}} \left(\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_{\mathbf{s}}}) - \frac{1}{n_j} \mathrm{Var}(\hat{\hat{U}}_{jk_{\mathbf{s}}}) \right) \\ &\quad + \sum_{j=1}^{J_u} \frac{n_j}{n_{jk}} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_i N^2}. \end{aligned} \tag{6.6}$$

The covariance matrix of $\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ is obtained if the expression of $\mathrm{E}_\alpha \mathrm{E}_s d_k$ in (6.6) is substituted into $\mathbf{CVC}^t$ in (3.43) in section 3.6.1

In the case of the generalized regression estimator, the diagonal elements d_{k_R} are defined by (4.28) in section 4.4. Let $\hat{\hat{U}}_{j_{R_{\mathbf{s}}}}$ denote the generalized regression estimator for $\bar{U}_j$ based on the n_j elements of sample s_j in block j. The first-order Taylor series approximation of $\hat{\hat{U}}_{j_{R_{\mathbf{s}}}}$ is given by

$$\hat{\hat{U}}_{j_{R_{\mathbf{s}}}} \doteq \hat{\hat{U}}_{j_{\mathbf{s}}} + \mathbf{b}^t (\bar{\mathbf{X}}_j - \hat{\hat{\mathbf{X}}}_{j_{\mathbf{s}}}) = \hat{\hat{E}}_{j_{\mathbf{s}}} + \mathbf{b}^t \bar{\mathbf{X}}_j, \tag{6.7}$$

where

$$\hat{\hat{E}}_{j_{\mathbf{s}}} = \hat{\hat{U}}_{j_{\mathbf{s}}} - \mathbf{b}^t \hat{\hat{\mathbf{X}}}_{j_{\mathbf{s}}} = \frac{1}{N_j} \sum_{j=1}^{n_j} \frac{(u_i - \mathbf{b}^t \mathbf{x}_i)}{\pi_{i|j}}. \tag{6.8}$$

Here $\hat{\hat{\mathbf{X}}}_{j_{\mathbf{s}}}$ is the Horvitz-Thompson estimator of $\bar{\mathbf{X}}_j$ based on the n_j sampling units of s_j and $\mathbf{b}$ the regression coefficients of linear regression model (4.2) in section 4.2. The design variance of the first-order Taylor series approximation of $\hat{\hat{U}}_{j_{R_{\mathbf{s}}}}$ is given by

$$\mathrm{Var}(\hat{\hat{E}}_{j_{\mathbf{s}}}) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j}) \frac{(u_i - \mathbf{b}^t \mathbf{x}_i)}{\pi_{i|j}} \frac{(u_{i'} - \mathbf{b}^t \mathbf{x}_{i'})}{\pi_{i'|j}}. \tag{6.9}$$

Furthermore,

$$\tilde{\mathrm{Var}}(\hat{\hat{E}}_{j_\mathbf{s}}) = \frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n_j(u_i - \mathbf{b}^t\mathbf{x}_i)^2}{\pi_{i|j}N_j^2} - (\bar{U}_j - \mathbf{b}^t\bar{\mathbf{X}}_j)^2\right) \tag{6.10}$$

denotes the variance of the first-order Taylor series approximation of $\hat{\hat{U}}_{jk_{R_\mathbf{s}}}$ as if n_j elements in block j were drawn with replacement with selection probabilities $\pi_{i|j}/n_j$. Under the basic measurement error model (4.5) in section 4.2, it is proved in appendix 6.9.2 that

$$\mathrm{E}_\alpha\mathrm{E}_s d_{k_R} = \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\left(\tilde{\mathrm{Var}}(\hat{\hat{E}}_{j_\mathbf{s}}) - \frac{1}{n_j}\mathrm{Var}(\hat{\hat{E}}_{j_\mathbf{s}})\right) + \sum_{j=1}^{J_u}\frac{n_j}{n_{jk}}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{\pi_i N^2}. \tag{6.11}$$

The variance expressions $\mathrm{E}_\alpha\mathrm{E}_s d_k$ for the Horvitz-Thompson estimator (given by (6.6)) and $\mathrm{E}_\alpha\mathrm{E}_s d_{k_R}$ for the generalized regression estimator (given by (6.11)) consist of two terms. The first term is the design variance of the target parameter and can be attributed to the randomization mechanisms of the applied sampling design and experimental design. It follows that the variance between the PSU's of the estimates of the target parameters is eliminated from this variance term, since PSU's are used as block variables. In this experiment a random sample of J_s blocks is drawn from a finite population of J_u blocks. In a model-based approach this might be taken into account by modeling the blocks as a random component in a random or mixed model. In the design-based approach followed here, it shows up as an inflation of the variance component of each block with a factor $1/\pi_j$. The second term concerns the variance of the individual disturbance terms and can be attributed to stochasticity of the measurement error model.

The efficiency of blocking on PSU's can be quantified by calculating the increase in $\mathrm{E}_\alpha\mathrm{E}_s d_k$ or $\mathrm{E}_\alpha\mathrm{E}_s d_{k_R}$ due to the application of a CRD instead of an RBD under two-stage sampling. In the case of a CRD analyzed with the Horvitz-Thompson estimator, a general expression for $\mathrm{E}_\alpha\mathrm{E}_s d_k$ is given by (5.3) in section 5.2. In the case of a CRD under two-stage sampling, the terms $\tilde{\mathrm{Var}}(\hat{\hat{U}}_{k_\mathbf{s}})$ and $\mathrm{Var}(\hat{\hat{U}}_{k_\mathbf{s}})$ in $\mathrm{E}_\alpha\mathrm{E}_s d_k$ in (5.3) can be evaluated as follows:

$$\begin{aligned}
\tilde{\mathrm{Var}}(\hat{\hat{U}}_{k_\mathbf{s}}) &= \frac{1}{n}\left(\sum_{i=1}^{N}\frac{nu_{ik}^2}{\pi_i N^2} - \bar{U}_k^2\right) \\
&= \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n_j u_{ik}^2}{\pi_{i|j}N_j^2} - \bar{U}_{jk}^2\right) + \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\frac{1}{n_j}\bar{U}_{jk}^2 - \frac{1}{n}\bar{U}_k^2 \\
&= \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\tilde{\mathrm{Var}}(\hat{\bar{U}}_{jk_\mathbf{s}}) + \frac{1}{n}\left(\sum_{j=1}^{J_u}\frac{n}{n_j}\frac{U_{jk}^2}{\pi_j\; N^2} - \bar{U}_k^2\right),
\end{aligned} \tag{6.12}$$

where $U_{jk} = N_j\bar{U}_{jk}$. The second term can be interpreted as the variance between the PSU totals as if the PSU's has been drawn with replacement with selection probabilities $(n_j\pi_j)/n$. Furthermore,

$$\mathrm{Var}(\hat{\hat{U}}_{k_\mathbf{s}}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(\pi_{ii'} - \pi_i\pi_{i'})\frac{u_{ik}u_{i'k}}{\pi_i\pi_{i'}}$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{j=1}^{J_u}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\left(\pi_j\pi_{ii'|j} - \pi_j^2\pi_{i|j}\pi_{i'|j}\right)\frac{u_{ik}u_{i'k}}{\pi_j^2\pi_{i|j}\pi_{i'|j}} \\
&\quad + \frac{1}{N^2}\sum_{j=1}^{J_u}\sum_{i=1}^{N_j}\sum_{j'\neq j=1}^{J_u}\sum_{i'=1}^{N_{j'}}\left(\pi_{jj'}\pi_{i|j}\pi_{i'|j'} - \pi_j\pi_{j'}\pi_{i|j}\pi_{i'|j'}\right)\frac{u_{ik}u_{i'k}}{\pi_j\pi_{i|j}\pi_{j'}\pi_{i'|j'}} \\
&= \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\bar{U}}_{jk_s}) + \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\bar{U}_{jk}^2 - \sum_{j=1}^{J_u}(N_j/N)^2\,\bar{U}_{jk}^2 \\
&\quad + \sum_{j=1}^{J_u}\sum_{j'\neq j=1}^{J_u}\left(\pi_{jj'} - \pi_j\pi_{j'}\right)\frac{(N_j/N)\,\bar{U}_{jk}}{\pi_j}\frac{(N_{j'}/N)\,\bar{U}_{j'k}}{\pi_{j'}} \\
&= \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\bar{U}}_{jk_s}) + \frac{1}{N^2}\sum_{j=1}^{J_u}\sum_{j'=1}^{J_u}\left(\pi_{jj'} - \pi_j\pi_{j'}\right)\frac{U_{jk}}{\pi_j}\frac{U_{j'k}}{\pi_{j'}}. \qquad (6.13)
\end{aligned}
$$

The second term is the design variance between the PSU totals with respect to the first stage of the sampling design. If the results obtained in (6.12) and (6.13) are substituted into expression (5.3), then it follows that

$$
\begin{aligned}
\mathrm{E}_\alpha\mathrm{E}_s d_k &= \frac{n}{n-1}\frac{n}{n_k}\sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\left(\tilde{\mathrm{Var}}(\hat{\bar{U}}_{jk_s}) - \frac{1}{n}\mathrm{Var}(\hat{\bar{U}}_{jk_s})\right) \\
&\quad + \frac{n}{n-1}\frac{n}{n_k}\left[\frac{1}{n}\left(\sum_{j=1}^{J_u}\frac{n\;U_{jk}^2}{n_j\pi_j N^2} - \bar{U}_k^2\right) - \frac{1}{n}\frac{1}{N^2}\sum_{j=1}^{J_u}\sum_{j'=1}^{J_u}(\pi_{jj'} - \pi_j\pi_{j'})\frac{U_{jk}\;U_{j'k}}{\pi_j\;\pi_{j'}}\right] \\
&\quad + \frac{n}{n_k}\sum_{j=1}^{J_u}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{\pi_i N^2}. \qquad (6.14)
\end{aligned}
$$

The efficiency of blocking on PSU's follows by comparing the expressions (6.6) and (6.14). To that purpose it is aasumed for the sake of convenience that $n/(n-1) \approx 1$ and $n_j/(n_j-1) \approx 1$. Furthermore it is assumed that the fraction of SSU's assigned to treatment k is the same within each PSU; i.e. $n_k/n = n_{jk}/n_j$ for all $j = 1, \ldots, J_u$. It follows that the increase in the variance expressions $\mathrm{E}_\alpha\mathrm{E}_s d_k$ due to the application of a CRD under two-stage sampling instead of an RBD with PSU's as block variables, is approximately equal to

$$
\begin{aligned}
&\frac{n}{n-1}\frac{n}{n_k}\left[\frac{1}{n}\left(\sum_{j=1}^{J_u}\frac{n\;U_{jk}^2}{n_j\pi_j N^2} - \bar{U}_k^2\right) - \frac{1}{n}\frac{1}{N^2}\sum_{j=1}^{J_u}\sum_{j'=1}^{J_u}(\pi_{jj'} - \pi_j\pi_{j'})\frac{U_{jk}\;U_{j'k}}{\pi_j\;\pi_{j'}}\right] \\
&+ \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\left(\frac{1}{n_{jk}} - \frac{1}{n}\right)\mathrm{Var}(\hat{\bar{U}}_{jk_s}). \qquad (6.15)
\end{aligned}
$$

The first term concerns a between-PSU variance and the second term concerns a within PSU variance. If the sampling units from the same PSU's are homogeneous and the variation between the PSU's is large, then the size of the first term will be substantial and the application of an RBD with PSU's as block variables will be efficient.

The efficiency of blocking on PSU's under the generalized regression estimator can be evaluated in an equivalent way. An expression for the variance reduction due to the application of an RBD with PSU's as block variables under the generalized regression estimator equals (6.15), where the variance terms of the Horvitz-Thompson estimator are replaced by the variance terms of the first-order Taylor series approximation of the generalized regression estimator. Equivalent to the results obtained for the Horvitz-Thompson estimator it follows that the between-PSU variance is eliminated from the variance expressions of $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ if PSU's are used as block variables in an RBD.

In summary the conclusion is that it might be efficient to use PSU's as block variables since the between-PSU variance is eliminated from the variance expressions $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$. This effect will be substantial if the intrinsic values u_i of the sampling units from the same PSU's tend to be more homogeneous than sampling units from different PSU's. Furthermore the randomization mechanism of an RBD where PSU's are block variables guarantees that all the PSU's selected in the sample s are also included in each of the subsamples s_k.

6.3 Strata as block variables

In this section we consider experiments embedded in a stratified sampling design conducted under the basic measurement error model (3.1) in section 3.2 or (4.5) in section 4.2. From each of the J strata in the population, a sample s_j of size n_j is drawn by means of a complex sampling design. According to the experimental design, the n_j sampling units within each stratum are randomized over the K different treatments. This randomization mechanism naturally leads to an RBD where strata are block variables. In this situation all the blocks in the finite population are selected in the sample. Therefore the number of bocks in the finite population is equal to the number of blocks selected in the sample.

A stratified sampling design can be considered as a special case of a two-stage sampling design where the first stage is completely observed. Therefore, expressions for the variances $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ for RBD's embedded in stratified sampling designs where strata are used as block variables follow as a special case from the results obtained under two-stage sampling in section 6.2.

In the case of the Horvitz-Thompson estimator, the diagonal elements d_k are defined in (3.34), section 3.6.1. Let $\bar{U}_j$ denote the population mean of the intrinsic values in stratum or block j. Then $\bar{U}_{jk} = \bar{U}_j + \beta_k$ denotes the population mean biased with the treatment effect β_k in block j. The Horvitz-Thompson estimator for $\bar{U}_{jk}$ based on the n_j elements in the sample s_j of block j is defined as:

$$\hat{\bar{U}}_{jk_s} = \frac{1}{N_j} \sum_{i=1}^{n_j} \frac{(u_i + \beta_k)}{\pi_i}, \tag{6.16}$$

where N_j denotes the population total of block j. Then

$$\mathrm{Var}(\hat{\bar{U}}_{jk_s}) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{(u_i + \beta_k)(u_{i'} + \beta_k)}{\pi_i \pi_{i'}} \tag{6.17}$$

denotes the design variance of $\hat{\hat{U}}_{jk_{\mathbf{s}}}$, i.e. the variance, which takes into account the real sampling design used to draw n_j elements from block j and

$$\tilde{\text{Var}}(\hat{\hat{U}}_{jk_{\mathbf{s}}}) = \frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n_j(u_i+\beta_k)^2}{\pi_i N_j^2} - \bar{U}_{jk}^2\right) \tag{6.18}$$

denotes the variance of $\hat{\hat{U}}_{jk_{\mathbf{s}}}$ as if n_j elements were drawn from block j with replacement with selection probabilities π_i/n_j. Under the basic measurement error model (3.1) in section 3.2 and an RBD where strata are used as block variables, it follows that

$$\begin{aligned} \text{E}_\alpha\text{E}_s d_k &= \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\left(\tilde{\text{Var}}(\hat{\hat{U}}_{jk_{\mathbf{s}}}) - \frac{1}{n_j}\text{Var}(\hat{\hat{U}}_{jk_{\mathbf{s}}})\right) \\ &\quad + \sum_{j=1}^{J}\frac{n_j}{n_{jk}}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{\pi_i N^2}. \end{aligned} \tag{6.19}$$

The proof of this result follows analogous to the proof of (6.6), appendix 6.9.1, for an RBD where PSU's are block variables, with $\pi_j = 1$, $\pi_{i|j} = \pi_i$ and $\pi_{ii'|j} = \pi_{ii'}$. An expression for the covariance matrix of $\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ is obtained if the elements $\text{E}_\alpha\text{E}_s d_k$ defined in (6.20) are substituted into the expression of $\mathbf{CVC}^t$ in (3.46) in section 3.6.1.

In the case of the generalized regression estimator, d_{k_R} is defined by (4.28) in section 4.4. Let $\hat{\hat{U}}_{jR_{\mathbf{s}}}$ denote the generalized regression estimator for $\bar{U}_j$ based on the n_j elements of sample s_j in block j. The first-order Taylor series approximation of $\hat{\hat{U}}_{jR_{\mathbf{s}}}$ is given by

$$\hat{\hat{U}}_{jR_{\mathbf{s}}} \doteq \hat{\hat{U}}_{j_{\mathbf{s}}} + \mathbf{b}^t(\bar{\mathbf{X}}_j - \hat{\hat{\mathbf{X}}}_{j_{\mathbf{s}}}) = \hat{\hat{E}}_{j_{\mathbf{s}}} + \mathbf{b}^t\bar{\mathbf{X}}_j, \tag{6.20}$$

where

$$\hat{\hat{E}}_{j_{\mathbf{s}}} = \hat{\hat{U}}_{j_{\mathbf{s}}} - \mathbf{b}^t\hat{\hat{\mathbf{X}}}_{j_{\mathbf{s}}} = \frac{1}{N_j}\sum_{j=1}^{n_j}\frac{(u_i - \mathbf{b}^t\mathbf{x}_i)}{\pi_i}. \tag{6.21}$$

Here $\hat{\hat{\mathbf{X}}}_{j_{\mathbf{s}}}$ is the Horvitz-Thompson estimator of $\bar{\mathbf{X}}_j$ based on the n_j sampling units of s_j and $\mathbf{b}$ the regression coefficients of linear regression model (4.2) in section 4.2. The design variance of the first-order Taylor series approximation of $\hat{\hat{U}}_{jR_{\mathbf{s}}}$ is given by

$$\text{Var}(\hat{\hat{E}}_{j_{\mathbf{s}}}) = \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'} - \pi_i\pi_{i'})\frac{(u_i - \mathbf{b}^t\mathbf{x}_i)}{\pi_i}\frac{(u_{i'} - \mathbf{b}^t\mathbf{x}_{i'})}{\pi_{i'}}. \tag{6.22}$$

Furthermore,

$$\tilde{\text{Var}}(\hat{\hat{E}}_{j_{\mathbf{s}}}) = \frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n(u_i - \mathbf{b}^t\mathbf{x}_i)^2}{\pi_i N^2} - (\bar{U}_j - \mathbf{b}^t\bar{\mathbf{X}}_j)^2\right) \tag{6.23}$$

denotes the approximate variance of $\hat{\hat{U}}_{jk_{R_{\mathbf{s}}}}$ as if n_j elements in block j were drawn with replacement with selection probabilities π_i/n_j. Under the basic measurement error model (4.5) in section 4.2, and an RBD with strata as block variables, it follows that

$$\text{E}_\alpha\text{E}_s d_{k_R} = \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2\frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\left(\tilde{\text{Var}}(\hat{\hat{E}}_{j_{\mathbf{s}}}) - \frac{1}{n_j}\text{Var}(\hat{\hat{E}}_{j_{\mathbf{s}}})\right) + \sum_{j=1}^{J}\frac{n_j}{n_{jk}}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{\pi_i N^2}. \tag{6.24}$$

The proof of this result follows as a special case of the proof of (6.11), appendix 6.9.2, for an RBD where PSU's are block variables, with $\pi_j = 1$, $\pi_{i|j} = \pi_i$ and $\pi_{ii'|j} = \pi_{ii'}$.

The expressions of $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ consist of two terms. The first term of (6.20) and (6.24) is the variance of the Horvitz-Thompson estimator or the generalized regression estimator of the population mean of the intrinsic values of the target parameter and can be attributed to the randomization mechanisms of the sampling design and the experimental design. It follows from this component that the variance between the strata is eliminated from $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$, since strata are used as block variables in an RBD. The second term of (6.20) and (6.24) concerns the variance of the individual disturbance terms from the measurement error model.

Note that the variance of the generalized regression estimator doesn't depend on the treatment effect β_k. For situations where the extended Horvitz-Thompson estimator coincides with the Horvitz-Thompson estimator, the variance of the Horvitz-Thompson estimator also doesn't depend on the treatment effect β_k. This holds true if $\hat{N}_j = N_j$; e.g. in the case of stratified simple random sampling where strata are block variables.

The efficiency of blocking on strata can be quantified by calculating the increase in $\mathrm{E}_\alpha \mathrm{E}_s d_k$ due to the application of a CRD instead of an RBD embedded in a stratified sampling scheme. A general expression for $\mathrm{E}_\alpha \mathrm{E}_s d_k$ in the case of a CRD is given by (5.3) in section 5.2. In the case of a CRD embedded in a stratified sampling, we can evaluate the term $\tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_s})$ in $\mathrm{E}_\alpha \mathrm{E}_s d_k$ in (5.3) as follows:

$$\begin{aligned}
\tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_s}) &= \frac{1}{n}\left(\sum_{i=1}^{N} \frac{n u_{ik}^2}{\pi_i N^2} - \bar{U}_k^2\right) \\
&= \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \frac{1}{n_j}\left(\sum_{i=1}^{N_j} \frac{n_j u_{ik}^2}{\pi_i N_j^2} - \bar{U}_{jk}^2\right) + \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \frac{1}{n_j}\bar{U}_{jk}^2 - \frac{1}{n}\bar{U}_k^2 \\
&= \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \tilde{\mathrm{Var}}(\hat{\bar{U}}_{jk_s}) + \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \frac{1}{n_j}\bar{U}_{jk}^2 - \frac{1}{n}\bar{U}_k^2. \qquad (6.25)
\end{aligned}$$

Under stratified sampling it also follows that

$$\mathrm{Var}(\hat{\bar{U}}_{k_s}) = \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \mathrm{Var}(\hat{\bar{U}}_{jk_s}). \qquad (6.26)$$

If the results (6.25) and (6.26) are substituted into (5.3) in section 5.2 it follows that $\mathrm{E}_\alpha \mathrm{E}_s d_k$ for a CRD embedded in a stratified sampling design equals

$$\begin{aligned}
\mathrm{E}_\alpha \mathrm{E}_s d_k &= \frac{n}{n-1}\frac{n}{n_k}\sum_{j=1}^{J} W_j^2 \left(\tilde{\mathrm{Var}}(\hat{\bar{U}}_{jk_s}) - \frac{1}{n}\mathrm{Var}(\hat{\bar{U}}_{jk_s})\right) \\
&\quad + \frac{n}{n-1}\frac{n}{n_k}\left(\sum_{j=1}^{J} \frac{(W_j \bar{U}_{jk})^2}{n_j} - \frac{1}{n}\bar{U}_k^2\right) + \frac{n}{n_k}\sum_{j=1}^{J}\sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_i N^2}, \qquad (6.27)
\end{aligned}$$

where $W_j = N_j/N$ and $\mathrm{Var}(\hat{\bar{U}}_{jk_s})$ and $\tilde{\mathrm{Var}}(\hat{\bar{U}}_{jk_s})$ are defined by (6.17) and (6.18), respectively.

To quantify the efficiency of an RBD in this situation, expression (6.27) should be compared with (6.20). Note that $\bar{U}_k = \sum_{j=1}^{J}(N_j/N)\bar{U}_{jk}$. It is assumed for the sake of convenience that $n/(n-1) \approx 1$ and $n_j/(n_{jk}-1) \approx 1$ and that the fraction of sampling units assigned to treatment k is the same within each stratum, i.e. $n_k/n = n_{jk}/n_j$, for all $j = 1, \ldots, J$. The increase in $\mathrm{E}_\alpha \mathrm{E}_s d_k$ under a CRD with respect to an RBD under stratified sampling is given by

$$\frac{n}{n-1}\frac{n}{n_k}\left(\sum_{j=1}^{J}\frac{(W_j\bar{U}_{jk})^2}{n_j} - \frac{1}{n}\bar{U}_k^2\right) + \sum_{j=1}^{J} W_j^2 \left(\frac{1}{n_{jk}} - \frac{1}{n_k}\right)\mathrm{Var}(\hat{\bar{U}}_{jk_s}). \tag{6.28}$$

The first term of (6.28) can be broken into

$$\frac{n}{n-1}\frac{1}{n_k}\sum_{j=1}^{J} W_j\left(\bar{U}_{jk} - \bar{U}_k\right)^2 + \frac{n}{n-1}\frac{1}{n_k}\sum_{j=1}^{J} W_j\left(\frac{W_j}{w_j} - 1\right)\bar{U}_{jk}^2, \tag{6.29}$$

with $w_j = n_j/n$. This is the major part of the increase in $\mathrm{E}_\alpha \mathrm{E}_s d_k$. The first term of (6.29) is the variance between the strata. In the case of stratification with proportional allocation it follows that $W_j = w_j$ and consequently the second term of (6.29) is equal to zero. So, for proportional allocation, (6.29) equals the variance between strata. For other allocations, (6.29) approximates the variance between strata. From sampling theory it is well known that stratified sampling is particularly interesting when the population stratum means differ substantially. It can be conclude from (6.28) that this efficiency property of stratified designs disappears in CRD's, but remains in RBD's with strata as block variables. CRD's nullify the effect of stratifying as randomizing the sampling units over the subsamples ignores these strata. Using strata as block variables not only preserves the efficiency of stratification, but also ensures that each stratum is sufficiently represented within each subsample s_k. This last property is especially important in applications where the subsamples assigned to the alternative treatments are small compared with the subsample assigned to the regular survey, which simultaneously serves as the control group in the experiment.

The efficiency of blocking on strata under the generalized regression estimator, can be evaluated in an equivalent way. As in the case of the Horvitz-Thompson estimator it follows for a CRD embedded in a stratified sampling design, that the efficiency gain of a stratified sampling design is essentially nullified by the randomization mechanism of the experimental design. In the special case of proportional allocation, the increase in the variance due to the application of a CRD instead of an RBD exactly equals the variance between strata.

6.4 Clusters as block variables

Consider an experiment embedded in a cluster sample, conducted under the basic measurement error model (3.1) in section 3.2 or (4.5) in section 4.2. According to the sampling design, J_s clusters of size N_j are drawn from the J_u clusters in the finite population by means of first-order inclusion probabilities π_j and second-order inclusion probabilities $\pi_{jj'}$. All individuals in a selected cluster are included in the sample s. According to the experimental design, the N_j

individuals of cluster j are randomized over the K different treatments. This randomization mechanism naturally leads to an RBD where clusters are block variables.

A cluster sample can be considered as a special case of a two-stage sampling design where the second stage is completely observed. Therefore, expressions for the variances $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ for RBD's embedded in cluster samples where clusters are used as block variables follow as a special case from the results obtained under two-stage sampling in section 6.2.

In the case of the Horvitz-Thompson estimator, an expression for $\mathrm{E}_\alpha \mathrm{E}_s d_k$ follows directly from (6.6) in section 6.2 by taking $\pi_{i|j} = 1$, $\pi_{ii'|j} = 1$ and $n_j = N_j$ and is given by

$$\mathrm{E}_\alpha \mathrm{E}_s d_k = \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{N_j}{(N_j-1)} \frac{N_j}{n_{jk}} \tilde{\mathrm{Var}}(\hat{\bar{U}}_{jk_{\mathbf{s}}}) + \sum_{j=1}^{J_u} \frac{N_j}{\pi_j n_{jk} N^2} \left(\sum_{i=1}^{N_j} \sigma_{ik}^2 \right). \tag{6.30}$$

Since all the N_j individuals of a cluster are selected in the sample, it follows that $\mathrm{Var}(\hat{\bar{U}}_{jk_{\mathbf{s}}})$ defined by (6.4) in section 6.2 is equal to zero. Furthermore, it follows from (6.5) (section 6.2) that

$$\tilde{\mathrm{Var}}(\hat{\bar{U}}_{jk_{\mathbf{s}}}) = \frac{1}{N_j} \left(\sum_{i=1}^{N_j} \frac{u_{ik}^2}{N_j} - \bar{U}_{jk}^2 \right), \tag{6.31}$$

which can be interpreted as the variance of $\hat{\bar{U}}_{jk_{\mathbf{s}}}$ as if N_j elements from cluster j were drawn with replacement with selection probabilities $1/N_j$.

In an equivalent way an expression for $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ for the generalized regression estimator follows as a special case of (6.11) in section 6.2 and is given by

$$\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} = \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{N_j}{(N_j-1)} \frac{N_j}{n_{jk}} \tilde{\mathrm{Var}}(\hat{\bar{E}}_{j_{\mathbf{s}}}) + \sum_{j=1}^{J_u} \frac{N_j}{\pi_j n_{jk} N^2} \left(\sum_{i=1}^{N_j} \sigma_{ik}^2 \right), \tag{6.32}$$

where

$$\tilde{\mathrm{Var}}(\hat{\bar{E}}_{j_{\mathbf{s}}}) = \frac{1}{N_j} \left(\sum_{i=1}^{N_j} \frac{(u_i - \mathbf{b}^t \mathbf{x}_i)^2}{N_j} - (\bar{U}_j - \mathbf{b}^t \bar{\mathbf{X}}_j)^2 \right). \tag{6.33}$$

The efficiency of blocking on clusters can be quantified by calculating the increase in the variance expressions $\mathrm{E}_\alpha \mathrm{E}_s d_k$ or $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ due to the application of a CRD instead of an RBD. For a CRD analyzed by means of the Horvitz-Thompson estimator, an expression for $\mathrm{E}_\alpha \mathrm{E}_s d_k$ follows as a special case from (6.14) in section 6.2 and is equal to

$$\begin{aligned}
\mathrm{E}_\alpha \mathrm{E}_s d_k \quad = \quad & \frac{n}{n-1} \frac{n}{n_k} \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \tilde{\mathrm{Var}}(\hat{\bar{U}}_{jk_{\mathbf{s}}}) \\
& + \frac{n}{n-1} \frac{n}{n_k} \left[\frac{1}{n} \left(\sum_{j=1}^{J_u} \frac{n}{N_j \pi_j N^2} \frac{U_{jk}^2}{} - \bar{U}_k^2 \right) - \frac{1}{n} \frac{1}{N^2} \sum_{j=1}^{J_u} \sum_{j'=1}^{J_u} (\pi_{jj'} - \pi_j \pi_{j'}) \frac{U_{jk}}{\pi_j} \frac{U_{j'k}}{\pi_{j'}} \right] \\
& + \frac{n}{n_k} \sum_{j=1}^{J_u} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_j N^2}.
\end{aligned} \tag{6.34}$$

The efficiency of blocking on clusters follows by comparing (6.34) and (6.30). It is assumed that $n/(n-1) \approx 1$, $N_j/(N_j-1) \approx 1$ and that $n_k/n = n_{jk}/N_j$ for all the blocks $j = 1, \ldots, J_u$. The increase of $\mathrm{E}_\alpha \mathrm{E}_s d_k$ due to the application of a CRD under cluster sampling instead of an RBD with clusters as block variables, is approximately equal to

$$\frac{n}{n-1} \frac{n}{n_k} \left[\frac{1}{n} \left(\sum_{j=1}^{J_u} \frac{n \ U_{jk}^2}{N_j \pi_j N^2} - \bar{U}_k^2 \right) - \frac{1}{n} \frac{1}{N^2} \sum_{j=1}^{J_u} \sum_{j'=1}^{J_u} (\pi_{jj'} - \pi_j \pi_{j'}) \frac{U_{jk} \ U_{j'k}}{\pi_j \ \pi_{j'}} \right]. \tag{6.35}$$

This is the between-cluster variance. The first term between brackets can be interpreted as the variance between the cluster totals as if clusters are drawn with replacement with selection probabilities $(N_j \pi_j)/n$. The second term between brackets is the design variance of the sampling scheme. If the sampling units from the same clusters are homogeneous and the variation between the clusters is large, then the size of expression (6.35) will be substantial and the application of an RBD with clusters as block variables will be efficient.

An expression for the variance reduction due to the application of an RBD with clusters as block variables instead of a CRD for the generalized regression estimator equals (6.35) where the variance of the Horvitz-Thompson estimator is replaced by the variance of the first-order Taylor series approximation of the generalized regression estimator.

It follows that the variance between the cluster totals is eliminated from the total variance of $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ due to the application of an RBD where clusters are block variables. Consequently, if the clusters are homogeneous with respect to the intrinsic values u_i, then it might be efficient to use clusters as block variables in an RBD. This randomization mechanism also ensures that each cluster included in the sample s, is also included in each of the K subsample s_k.

Instead of randomizing the SSU's within clusters over the K treatments in an RBD, it is also possible to randomize the clusters over the K treatments in a CRD. This might be efficient if the variance between cluster totals is small and the variance within clusters large. If clusters are used as experimental units in CRD, the analysis concerns cluster totals. Therefore the variance of the estimated treatment effects only concerns the variance between cluster totals. This in contrast with an RBD with clusters as block variables, where the variance of the estimated treatment effects only concerns the variance within clusters.

6.5 Interviewers as block variables under simple random sampling without replacement

In many social surveys, data are collected by means of computer-assisted telephone interviewing (CATI) or computer-assisted personal interviewing (CAPI). In such situations a measurement error model with random or fixed interviewer effects might be realistic. Due to the interviewer effects specified in measurement error models (3.4), section 3.2 and (4.8), section 4.2, observations $\mathbf{y}_i^\alpha$ obtained from individuals assigned to the same interviewer tend to be more homogenous than individuals assigned to different interviewers. Therefore, it might be efficient to design an experiment embedded in a survey where data are collected by means of CAPI or CATI as an RBD with interviewers as block variables.

If interviewers are used as block variables, then the blocks of the experimental design are not directly linked with the sampling design. In this section we consider experiments conducted under simple random sampling without replacement.

The population U of size N can be divided in J groups U_j of size N_j. The individuals of each subgroup U_j are assigned to interviewer j if they are included in the sample. From this population a simple random sample of size n is drawn. After realizing this sample, the distribution $(n_1, \dots, n_J)$ of the sampling units over the J interviewers is obtained. Consequently, the number of sampling units assigned to each interviewer (or block) must be treated as random. Because n elements are drawn by means of simple random sampling without replacement from a population of size N, which can be divided in J groups of size $(N_1, \dots, N_J)$, it follows that the vector $(n_1, \dots, n_J)$ has a multivariate hypergeometric distribution. According to the experimental design, the n_j individuals assigned to the $j-$th interviewer, are randomized over the K treatments. Note that the sample space of simple random sampling without replacement is equal to the sample space of stratified simple random sampling from J strata multiplied with the probability of drawing a vector $(n_1, \dots, n_J)$ from the hypergeometric distribution where each element corresponds to the sample size of the strata. Now we can evaluate the expectation with respect to the sampling design as follows. Conditional on the realized distribution of the sampling units over the blocks (or interviewers), we can take the expectation as if a stratified sample was drawn where the blocks are the strata. Then we can take the expectation of the number of sampling units within each block as if this is a realization from the hypergeometric distribution, see appendix 6.9.3 for technical details.

In the case of the Horvitz-Thompson estimator, under simple random sampling generally called the direct estimator, the diagonal elements d_k are defined in (3.34), section 3.6 where $\pi_i = n/N$. Under a measurement error model with random as well as fixed interviewer effects, it is proved in appendix 6.9.3 that:

$$\mathrm{E}_\alpha \mathrm{E}_s d_k = \sum_{j=1}^{J} \frac{N_j}{N} \frac{n_j}{n_{jk}\, n} \frac{1}{(N_j-1)} \sum_{i=1}^{N_j} \left(u_i - \frac{1}{N_j} \sum_{i'=1}^{N_j} u_{i'} \right)^2 + \sum_{j=1}^{J} \frac{n_j}{n_{jk}\, n} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{N}. \tag{6.36}$$

In the case of the generalized regression estimator, the diagonal elements d_{k_R} are defined in (4.28), section 4.4 with $\pi_i = n/N$. Under a measurement error model with random as well as fixed interviewer effects, it is proved in appendix 6.9.3 that:

$$\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} =$$
$$\sum_{j=1}^{J} \frac{N_j}{N} \frac{n_j}{n_{jk} n} \frac{1}{(N_j-1)} \sum_{i=1}^{N_j} \left((u_i - \mathbf{b}^t \mathbf{x}_i) + (\psi_j - \mathbf{d}^t \mathbf{x}_i) - \frac{1}{N_j} \sum_{i'=1}^{N_j} \Big((u_{i'} - \mathbf{b}^t \mathbf{x}_{i'}) + (\psi_j - \mathbf{d}^t \mathbf{x}_{i'}) \Big) \right)^2$$

$$+\sum_{j=1}^{J}\frac{n_j}{n_{jk}\,n}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{N}, \tag{6.37}$$

where $\mathbf{d}$ denotes the multiple regression coefficients from the regression function of the interviewer effects on the auxiliary variables $\mathbf{x}_i$ (see (4.39) in appendix 4.7.2). We see that under the direct estimator, interviewer effects are eliminated from the variance since interviewers are used as block variables. Under the generalized regression estimator, however, this efficiency gain is partially nullified since the residuals of the fixed interviewer effects $\psi_j - \mathbf{d}\mathbf{x}_i$ generally do not cancel out in (6.37). They do cancel out under two specific situations: First if the extended Horvitz-Thompson estimator is applied. In this situation all the residuals of the fixed interviewer effects $\psi_j - \mathbf{d}^t\mathbf{x}_i$ within each block are equal to $\psi_j - \bar{\psi}$, where $\bar{\psi} = \sum_{j=1}^{J}(N_j/N)\psi_j$. Since these residuals are constant within each block, they consequently cancel out in (6.37). Since $\mathbf{b}^t = (\bar{U})$, (6.37) equals (6.36) in this case. Second, if within each block a separate regression estimator is applied minimally using the block size as auxiliary information. In this situation the fixed interviewer effects cancel out from the residuals $(y_{ik}^{\alpha} - \mathbf{b}_k^t\mathbf{x}_i)$, since $\mathbf{d}\mathbf{x}_i = \psi_j$, $\forall i$ (see (6.84) in appendix 6.9.3). In this situation (6.37) simplifies to

$$\mathrm{E}_{\alpha}\mathrm{E}_s d_{k_R} = \sum_{j=1}^{J}\frac{N_j}{N}\frac{n_j}{n_{jk}\,n}\frac{1}{(N_j-1)}\sum_{i=1}^{N_j}\left((u_i - \mathbf{b}^t\mathbf{x}_i) - \frac{1}{N_j}\sum_{i'=1}^{N_j}(u_{i'} - \mathbf{b}^t\mathbf{x}_{i'})\right)^2 + \sum_{j=1}^{J}\frac{n_j}{n_{jk}\,n}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{N}.$$

6.5.1 Efficiency of blocking on interviewers with random interviewer effects

The efficiency of blocking on interviewers follows if the increase in the variance expressions of $\mathrm{E}_{\alpha}\mathrm{E}_s d_k$ or $\mathrm{E}_{\alpha}\mathrm{E}_s d_{k_R}$ due to the application of a CRD instead of an RBD with interviewers as block variables is quantified under a measurement error model with fixed or random interviewer effects. In order to keep formula's manageable, measurement error models with only random or only fixed interviewer effects are treated separately. We start with random interviewer effects in this subsection. The efficiency of blocking on interviewers under fixed interviewer effects is dealt with in the next subsection.

In the case of a CRD analyzed with the direct estimator under a measurement error model with random interviewer effects ((3.4), section 3.2 where $\psi_j = 0$) it is proved in appendix 6.9.4 that

$$\begin{aligned}\mathrm{E}_{\alpha}\mathrm{E}_s d_k &= \frac{1}{n_k}\frac{1}{N-1}\sum_{i=1}^{N}\left(u_i - \frac{1}{N}\sum_{i'=1}^{N}u_{i'}\right)^2 + \frac{1}{n_k}\frac{1}{N-1}\sum_{j=1}^{J}N_j\tau_j^2\left(1-\frac{N_j}{N}\right)\\ &\quad + \frac{1}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{N}.\end{aligned} \tag{6.38}$$

Expression (6.36) must be compared with (6.38) in order to quantify the efficiency of blocking on interviewers. The population variance of the target variable in (6.38) (i.e. the first term on the right-hand side of the equality sign) can be rewritten as follows

$$\frac{1}{N-1}\sum_{i=1}^{N}(u_i - \bar{U})^2 = \frac{1}{N-1}\sum_{j=1}^{J}\sum_{i=1}^{N_j}(u_i - \bar{U})^2$$

$$= \frac{1}{N-1}\left(\sum_{j=1}^{J}\sum_{i=1}^{N_j}\left(u_i - \bar{U}_j\right)^2 + \sum_{j=1}^{J} N_j\left(\bar{U}_j - \bar{U}\right)^2\right)$$

$$\approx \sum_{j=1}^{J}\frac{N_j}{N}\frac{1}{N_j-1}\sum_{i=1}^{N_j}\left(u_i - \bar{U}_j\right)^2 + \sum_{j=1}^{J}\frac{N_j}{N}\left(\bar{U}_j - \bar{U}\right)^2. \tag{6.39}$$

In an RBD with interviewers as block variables the fraction of experimental units in block or interviewer j assigned to treatment k equals n_{jk}/n_j. If this fraction is (approximately) equal for each interviewer, then it follows that

$$\frac{n_j}{n_{jk}n} = \frac{1}{n_k}. \tag{6.40}$$

If (6.40) is substituted into (6.36) and (6.39) is substituted into (6.38), then it follows that the increase in $\mathrm{E}_\alpha \mathrm{E}_s d_k$ due to the application of a CRD instead of an RBD with interviewers as block variables under a measurement error model with random interviewer effects is equal to

$$\frac{1}{n_k}\frac{1}{N-1}\sum_{j=1}^{J} N_j \tau_j^2\left(1 - \frac{N_j}{N}\right) + \frac{1}{n_k}\sum_{j=1}^{J}\frac{N_j}{N}\left(\bar{U}_j - \bar{U}\right)^2.$$

Under a CRD the variance expressions of $\mathrm{E}_\alpha \mathrm{E}_s d_k$ are increased with two components. The first component is the variance due to the random interviewer effects. The second variance component concerns the variance of the intrinsic values between the blocks (i.e. the interviewers). In surveys where data are collected by means of CAPI, interviewers work in separate relatively small regions. In such situations, the variance between the means of the target variables might be substantial. In surveys where data are collected by means of CATI, telephone numbers are assigned randomly to an interviewer. Consequently interviewers work through the entire survey population. In this situation, the variance between the means of the target variables might be relatively small.

For the generalized regression estimator the efficiency of blocking on interviewers under a measurement error model with random interviewer effects can be quantified in an equivalent way. An expression for $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ in the case of a CRD is given by (6.38) where u_i is replaced by the residuals $u_i - \mathbf{b}^t\mathbf{x}_i$. The increase in $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ due to the application of CRD instead of an RBD where interviewers are block variables equals

$$\frac{1}{n_k}\frac{1}{N-1}\sum_{j=1}^{J} N_j \tau_j^2\left(1 - \frac{N_j}{N}\right) + \frac{1}{n_k}\sum_{j=1}^{J}\frac{N_j}{N}\left(\bar{E}_j - \bar{E}\right)^2,$$

where

$$\bar{\mathrm{E}}_j = \frac{1}{N_j}\sum_{i=1}^{N_j} u_i - \mathbf{b}^t\mathbf{x}_i,$$

and

$$\bar{\mathrm{E}} = \frac{1}{N}\sum_{i=1}^{N} u_i - \mathbf{b}^t\mathbf{x}_i.$$

The first term is a variance component of random interviewer effects. The second variance component concerns the variance between the block means of the first-order Taylor series approximation of the generalized regression estimator.

6.5.2 Efficiency of blocking on interviewers with fixed interviewer effects

In the case of a CRD analyzed with the direct estimator conducted under a measurement error model with fixed interviewer effects ((3.4), section 3.2 where $\xi_j^\alpha = 0$) it is proved in appendix 6.9.4 that:

$$\begin{aligned}\mathrm{E}_\alpha \mathrm{E}_s d_k &= \frac{1}{n_k}\frac{1}{N-1}\sum_{i\in U}\left(u_i - \frac{1}{N}\sum_{i'\in U} u_{i'}\right)^2 + \frac{1}{n_k}\frac{1}{N-1}\left(\sum_{j=1}^J N_j\psi_j^2 - \frac{1}{N}\left(\sum_{j=1}^J N_j\psi_j\right)^2\right) \\ &+2\frac{1}{n_k}\frac{1}{N-1}\sum_{i\in U}\left(u_i - \frac{1}{N}\sum_{i'\in U} u_{i'}\right)\left(\psi_j - \frac{1}{N}\sum_{j=1}^J N_j\psi_j\right) + \frac{1}{n_k}\sum_{i\in U}\frac{\sigma_{ik}^2}{N}. \qquad (6.41)\end{aligned}$$

The efficiency of an RBD is quantified if (6.36) is compared with (6.41). If (6.40) is substituted into (6.36) and (6.39) is substituted into (6.41), then it follows that the increase in $\mathrm{E}_\alpha \mathrm{E}_s d_k$ due to the application of a CRD instead of an RBD with interviewers as block variables equals

$$\begin{aligned}&\frac{1}{n_k}\frac{1}{N-1}\left(\sum_{j=1}^J N_j\psi_j^2 - \frac{1}{N}\left(\sum_{j=1}^J N_j\psi_j\right)^2\right) \\ &+2\frac{1}{n_k}\frac{1}{N-1}\sum_{i\in U}\left(u_i - \frac{1}{N}\sum_{i'\in U} u_{i'}\right)\left(\psi_j - \frac{1}{N}\sum_{j=1}^J N_j\psi_j\right) + \frac{1}{n_k}\sum_{j=1}^J\frac{N_j}{N}\left(\bar{U}_j - \bar{U}\right)^2.\end{aligned}$$

The variance reduction, due to the application of an RBD where interviewers are block variables under a measurement error model with fixed interviewer effects, consists of three terms. The first term is the variance of the fixed interviewer effects. The second term is the covariance between the fixed interviewer effects and the target variables. The third term is the between-block variance of the intrinsic values of the target variables.

Results for the generalized regression estimator read as follows. In the case of a CRD conducted under a measurement error model with fixed interviewer effects ((4.8), section 4.2 where $\xi_j^\alpha = 0$) it follows that

$$\begin{aligned}\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} &= \frac{1}{n_k}\frac{1}{N-1}\sum_{i=1}^N\left((u_i - \mathbf{b}^t\mathbf{x}_i) + (\psi_j - \mathbf{d}^t\mathbf{x}_i) - \frac{1}{N}\sum_{i'=1}^N\left((u_{i'} - \mathbf{b}^t\mathbf{x}_{i'}) + (\psi_j - \mathbf{d}^t\mathbf{x}_{i'})\right)\right)^2 \\ &+\frac{1}{n_k}\sum_{i=1}^N\frac{\sigma_{ik}^2}{N}. \qquad (6.42)\end{aligned}$$

The proof of (6.42) is equivalent to the proofs for the direct estimator (6.41). The increase in variance in $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ due to the application of a CRD instead of blocking on interviewers follows if (6.42) is compared with (6.37) and is approximately equal to

$$\frac{1}{n_k}\sum_{j=1}^J\frac{N_j}{N}\left((\bar{U}_j - \mathbf{b}^t\bar{\mathbf{X}}_j) - (\bar{U} - \mathbf{b}^t\bar{\mathbf{X}}) + (\psi_j - \mathbf{d}^t\bar{\mathbf{X}}_j) - (\bar{\psi} - \mathbf{d}^t\bar{\mathbf{X}})\right)^2,$$

where $\bar{\psi} = \sum_{j=1}^J (N_j/N)\psi_j$. From (6.37) it follows that the generalized regression estimator can partially nullify the efficiency of local control on interviewers by means of an RBD. Therefore the variance reduction due to the application of an RBD instead of a CRD only concerns the between-block variance of the target parameter, the fixed interviewer effects and the covariance between these two parameters.

Under a CRD, each individual assigned to an interviewer has a nonzero probability to be assigned to one of the K treatments. In this case the subsample estimates $\hat{\hat{Y}}_k^{\alpha}$ are biased with the same fixed interviewer effect $\bar{\psi} = \sum_{j=1}^{J}(N_j/N)\psi_j$. In the $K-1$ contrasts of these estimates, this bias conveniently cancels out. In many situations, however, interviewers are for practical reasons assigned to only one of the K treatments. Strictly speaking this are not CRD's. Since each treatment is conducted with a different set of interviewers, each estimate is biased with the mean of a different set of fixed interviewer effects. This bias, consequently, does not cancel out in the $K-1$ contrasts of the estimated population parameters.

6.6 Interviewers as block variables under complex sampling schemes

In the case of an RBD where interviewers are block variables, the blocks of the experimental design are generally not directly linked with the sampling scheme. Consequently, the number of individuals allocated to each block is random. Under simple random sampling, considered in the preceding section, the expectation with respect to the sampling design was evaluated as follows. Conditional on the realized distribution of the sampling units over the blocks, we can take the expectation as if a stratified simple random sample was drawn where the strata correspond with the blocks of the experimental design. Then we can take the expectation of the number of individuals allocated to each block as if this is a realization from the hypergeometric distribution. This approach can be applied under simple random sampling, since the sample space of simple random sampling is equal to the sample space of stratified simple random sampling multiplied with the probability of drawing a vector from the hypergeometric distribution where each element corresponds to the sample size of the strata. This approach is not applicable under more complex sampling schemes. If it is ignored that the number of individuals allocated to block j is random, then we can still evaluate the expectation with respect to the sampling design by treating the blocks as domains. This implies that the variances are derived conditional on the realized sample s.

6.6.1 Random interviewer effects

For an RBD with interviewers as block variables analyzed with the Horvitz-Thompson estimator under a measurement error model with only random interviewer effects (i.e. model (3.4), section 3.2 with $\psi_j = 0$) it is proved in appendix 6.9.5 that

$$\mathrm{E}_{\alpha}\mathrm{E}_s d_k = \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\left[\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_{\mathbf{s}}}) + \frac{\tau_j^2}{N_j^2}\tilde{\mathrm{Var}}(\hat{N}_j) - \frac{1}{n_j}\left(\mathrm{Var}(\hat{\hat{U}}_{jk_{\mathbf{s}}}) + \frac{\tau_j^2}{N_j^2}\mathrm{Var}(\hat{N}_j)\right)\right]$$

$$+\frac{1}{N^2}\sum_{j=1}^{J}\frac{n_j}{n_{jk}}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{\pi_i}. \tag{6.43}$$

Here $\hat{\hat{U}}_{jk_s}$, $\mathrm{Var}(\hat{\hat{U}}_{jk_s})$ and $\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_s})$ are defined by (6.16), (6.17) and (6.18) in section 6.3. Recall that

$$\hat{N}_j=\sum_{i=1}^{n_j}\frac{1}{\pi_i}, \tag{6.44}$$

is the Horvitz-Thompson estimator for the size of the $j-$th block. It follows that

$$\mathrm{Var}(\hat{N}_j)=\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\frac{(\pi_{ii'}-\pi_i\pi_{i'})}{\pi_i\pi_{i'}}, \tag{6.45}$$

is the variance of the estimated block size with respect to the real sampling design and

$$\tilde{\mathrm{Var}}(\hat{N}_j)=\frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n_j}{\pi_i}-N_j^2\right), \tag{6.46}$$

the variance of the estimated block size as if the sampling units were drawn with replacement with inclusion probabilities π_i/n_j.

For an RBD analyzed with the generalized regression estimator, it is proved in appendix 6.9.6 that

$$\begin{aligned}\mathrm{E}_\alpha\mathrm{E}_s d_{k_R} \;=\; & \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2\frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\left[\tilde{\mathrm{Var}}(\hat{\hat{E}}_{j_s})+\frac{\tau_j^2}{N_j^2}\tilde{\mathrm{Var}}(\hat{N}_j)-\frac{1}{n_j}\left(\mathrm{Var}(\hat{\hat{E}}_{j_s})+\frac{\tau_j^2}{N_j^2}\mathrm{Var}(\hat{N}_j)\right)\right]\\ & +\frac{1}{N^2}\sum_{j=1}^{J}\frac{n_j}{n_{jk}}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{\pi_i},\end{aligned} \tag{6.47}$$

where $\hat{\hat{E}}_{j_s}$, $\mathrm{Var}(\hat{\hat{E}}_{j_s})$ and $\tilde{\mathrm{Var}}(\hat{\hat{E}}_{j_s})$ are defined by (6.21), (6.22) and (6.23), section 6.3.

The expressions $\mathrm{Var}(\hat{N}_j)$ and $\tilde{\mathrm{Var}}(\hat{N}_j)$ are the design variances of the estimated block size. These variance components are generally nonzero since we cannot estimate these domain sizes without error under general complex sampling designs. If the sampling scheme within each block is self-weighted and the sample size is fixed, then $\hat{N}_j = N_j$ and consequently $\mathrm{Var}(\hat{N}_j)$ and $\tilde{\mathrm{Var}}(\hat{N}_j)$ are equal to zero. This condition is met if e.g. the interviewer regions coincide exactly with strata or PSU's and the sampling design within each stratum or PSU is self-weighted with a fixed sample size n_j. That $\tilde{\mathrm{Var}}(\hat{N}_j)$ is equal to zero in this situation, follows directly from expression (6.46) for $\pi_i = n_j/N_j$. That $\mathrm{Var}(\hat{N}_j)$ is equal to zero follows from proof (3.48) in subsection 3.6.1.

The expressions of $\mathrm{Var}(\hat{\hat{U}}_{jk_s})$, $\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_s})$, $\mathrm{Var}(\hat{\hat{E}}_{j_s})$ and $\tilde{\mathrm{Var}}(\hat{\hat{E}}_{j_s})$ can be interpreted as the variances of the block means of the intrinsic values, conditional on the realized allocation of individuals over the J blocks in the sample. Under the Horvitz-Thompson estimator, $\mathrm{Var}(\hat{\hat{U}}_{jk_s})$ and $\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_s})$ contain the treatment effects β_k. If interviewer regions coincide exactly with the strata or PSU's of the sampling design and the sampling design within each stratum or PSU is self-weighted with a fixed sample size n_j, then these treatment effects vanish from these variance components.

6.6.2 Fixed interviewer effects

For an RBD with interviewers as block variables analyzed with the Horvitz-Thompson estimator under a measurement error model with only fixed interviewer effects (i.e. model (3.4), section 3.2 with $\xi_j^\alpha = 0$) it is proved in appendix 6.9.5 that

$$\begin{aligned} \mathrm{E}_\alpha \mathrm{E}_s d_k \;=\; & \sum_{j=1}^{J} \left(\frac{N_j}{N}\right)^2 \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \Big[\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_s}) + \tilde{\mathrm{Var}}(\hat{\psi}_j) + 2\tilde{\mathrm{Cov}}(\hat{\hat{U}}_{jk_s}, \hat{\psi}_j) \\ & -\frac{1}{n_j}\Big(\mathrm{Var}(\hat{\hat{U}}_{jk_s}) + \mathrm{Var}(\hat{\psi}_j) + 2\mathrm{Cov}(\hat{\hat{U}}_{jk_s}, \hat{\psi}_j)\Big) \Big] \\ & +\frac{1}{N^2} \sum_{j=1}^{J} \frac{n_j}{n_{jk}} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_i}. \end{aligned} \tag{6.48}$$

Here

$$\mathrm{Var}(\hat{\psi}_j) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{\psi_j^2}{\pi_i \pi_{i'}} = \frac{\psi_j^2}{N_j^2} \mathrm{Var}(\hat{N}_j), \tag{6.49}$$

and

$$\tilde{\mathrm{Var}}(\hat{\psi}_j) = \frac{1}{n_j} \left(\frac{n_j \psi_j^2}{\pi_i N_j^2} - \psi_j^2 \right) = \frac{\psi_j^2}{N_j^2} \tilde{\mathrm{Var}}(\hat{N}_j), \tag{6.50}$$

where $\mathrm{Var}(\hat{N}_j)$ and $\tilde{\mathrm{Var}}(\hat{N}_j)$ are defined in (6.45) and (6.46), respectively. Furthermore,

$$\mathrm{Cov}(\hat{\hat{U}}_{jk_s}, \hat{\psi}_j) = \frac{\psi_j}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{u_{ik}}{\pi_i \pi_{i'}} \equiv \frac{\psi_j}{N_j^2} \mathrm{Cov}(\hat{U}_{jk_s}, \hat{N}_j), \tag{6.51}$$

and

$$\tilde{\mathrm{Cov}}(\hat{\hat{U}}_{jk_s}, \hat{\psi}_j) = \frac{\psi_j}{N_j^2} \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j u_{ik}}{\pi_i} - \bar{U}_{jk} N_j^2 \right) \equiv \frac{\psi_j}{N_j^2} \tilde{\mathrm{Cov}}(\hat{U}_{jk_s}, \hat{N}_j). \tag{6.52}$$

Note that $\mathrm{Var}(\hat{\psi}_j)$ and $\tilde{\mathrm{Var}}(\hat{\psi}_j)$ can be interpreted as the within block variances of

$$\hat{\psi}_j = \frac{\psi_j}{N_j} \sum_{i=1}^{n_j} \frac{1}{\pi_i},$$

with the differences that $\mathrm{Var}(\hat{\psi}_j)$ can be recognized as the design variance of the Horvitz-Thompson estimator and $\tilde{\mathrm{Var}}(\hat{\psi}_j)$ the design variance as if n_j elements in this block were drawn with replacement with selection probabilities π_i/n_j. Equivalently, formulas (6.51) and (6.52) are the within block covariances between the Horvitz-Thompson estimator of the fixed interviewer effect and the Horvitz-Thompson estimator of the target parameter, i.e. $\hat{\hat{U}}_{jk_s}$, with respect to the real sampling design and as if n_j elements in this block were drawn with replacement with selection probabilities π_i/n_j. The variance and covariance expressions (6.49) through (6.52) arise since the block sizes cannot be estimated without error under general complex sampling designs.

If interviewer regions coincide exactly with strata or PSU's of the sampling scheme and the sampling design within each stratum or PSU is self-weighted with a fixed sample size n_j, then the block size can be estimated without error, i.e. $\hat{N}_j = N_j$. In this situation we have the following simplifications in the variance expressions. First the treatment effects β_k vanish in the expressions of $\mathrm{Var}(\hat{\hat{U}}_{jk_\mathbf{s}})$, $\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_\mathbf{s}})$, $\mathrm{Cov}(\hat{\hat{U}}_{jk_\mathbf{s}}, \hat{\psi})$ and $\tilde{\mathrm{Cov}}(\hat{\hat{U}}_{jk_\mathbf{s}}, \hat{\psi})$. Second, the variances $\mathrm{Var}(\hat{\psi}_j)$ and $\tilde{\mathrm{Var}}(\hat{\psi}_j)$ as well as the covariance $\mathrm{Cov}(\hat{\hat{U}}_{jk_\mathbf{s}}, \hat{\psi}_j)$ and $\tilde{\mathrm{Cov}}(\hat{\hat{U}}_{jk_\mathbf{s}}, \hat{\psi}_j)$ are equal to zero. That $\tilde{\mathrm{Cov}}(\hat{\hat{U}}_{jk_\mathbf{s}}, \hat{\psi}_j)$ is equal to zero in this situation follows directly from expression (6.52) for $\pi_i = n_j/N_j$. For $\mathrm{Cov}(\hat{\hat{U}}_{jk_\mathbf{s}}, \hat{\psi}_j)$ it follows from (3.47) in subsection 3.6.1 and $\pi_i = n_j/N_j$ that

$$\begin{aligned}
\mathrm{Cov}(\hat{\hat{U}}_{jk_\mathbf{s}}, \hat{\psi}_j) &= \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{u_{ik}\psi_j}{\pi_i \pi_{i'}} \\
&= \frac{1}{N_j^2} \left(\left(\frac{N_j}{n_j} \right)^2 \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} \pi_{ii'} u_{ik} \psi_j - N_j^2 \bar{U}_{jk} \psi_j \right) \\
&= \frac{1}{N_j^2} \left(\left(\frac{N_j}{n_j} \right)^2 \left[\sum_{i=1}^{N_j} u_{ik}\psi_j \sum_{\substack{i'=1 \\ i' \neq i}}^{N_j} \pi_{ii'} + \sum_{i=1}^{N_j} \pi_i u_{ik} \psi_j \right] - N_j^2 \bar{U}_{jk} \psi_j \right) \\
&= \frac{1}{n_j^2} \left((n_j - 1) \sum_{i=1}^{N_j} \pi_i u_{ik} \psi_j + \sum_{i=1}^{N_j} \pi_i u_{ik} \psi_j \right) - \bar{U}_{jk} \psi_j = 0.
\end{aligned}$$

For an RBD analyzed with the generalized regression estimator it is proved in appendix 6.9.6 that

$$\begin{aligned}
\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} = &\sum_{j=1}^{J} \left(\frac{N_j}{N} \right)^2 \frac{n_j}{(n_j - 1)} \frac{n_j}{n_{jk}} \Big[\tilde{\mathrm{Var}}(\hat{\hat{E}}_{j_\mathbf{s}}) + \tilde{\mathrm{Var}}(\hat{\psi}_{jR}) + 2\tilde{\mathrm{Cov}}(\hat{\hat{E}}_{j_\mathbf{s}}, \hat{\psi}_{jR}) \\
&\qquad - \frac{1}{n_j} \left(\mathrm{Var}(\hat{\hat{E}}_{j_\mathbf{s}}) + \mathrm{Var}(\hat{\psi}_{jR}) + 2\mathrm{Cov}(\hat{\hat{E}}_{j_\mathbf{s}}, \hat{\psi}_{jR}) \right) \Big] \\
&+ \frac{1}{N^2} \sum_{j=1}^{J} \frac{n_j}{n_{jk}} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_i},
\end{aligned} \tag{6.53}$$

where

$$\mathrm{Var}(\hat{\psi}_{jR}) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{(\psi_j - \mathbf{d}^t \mathbf{x}_i)}{\pi_i} \frac{(\psi_j - \mathbf{d}^t \mathbf{x}_{i'})}{\pi_{i'}}, \tag{6.54}$$

$$\tilde{\mathrm{Var}}(\hat{\psi}_{jR}) = \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j (\psi_j - \mathbf{d}^t \mathbf{x}_i)^2}{\pi_i N_j^2} - \left(\psi_j - \mathbf{d}^t \bar{\mathbf{X}}_j \right)^2 \right), \tag{6.55}$$

$$\mathrm{Cov}(\hat{\hat{E}}_{j_\mathbf{s}}, \hat{\psi}_{jR}) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{(u_i - \mathbf{b}^t \mathbf{x}_i)}{\pi_i} \frac{(\psi_j - \mathbf{d}^t \mathbf{x}_{i'})}{\pi_{i'}}, \tag{6.56}$$

and

$$\tilde{\mathrm{Cov}}(\hat{\hat{E}}_{j_\mathbf{s}}, \hat{\psi}_{jR}) = \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j (u_i - \mathbf{b}^t \mathbf{x}_i)(\psi_j - \mathbf{d}^t \mathbf{x}_i)}{\pi_i N_j^2} - (\bar{U}_j - \mathbf{b}^t \bar{\mathbf{X}}_j)(\psi_j - \mathbf{d}^t \bar{\mathbf{X}}_j) \right). \tag{6.57}$$

Here $\mathbf{d}$ denotes the multiple regression coefficients from the regression function of the interviewer effects on the auxiliary variables $\mathbf{x}_i$ (see (4.39) in appendix 4.7.2). Note that $\text{Var}(\hat{\psi}_{j_R})$ and $\tilde{\text{Var}}(\hat{\psi}_{j_R})$ can be interpreted as the within block variances of

$$\frac{1}{N_j}\sum_{i=1}^{n_j}\frac{\psi_j - \mathbf{d}^t\mathbf{x}_i}{\pi_i}. \tag{6.58}$$

Equivalently, $\tilde{\text{Cov}}(\hat{\hat{E}}_{j_\mathbf{s}}, \hat{\psi}_{j_R})$ and $\tilde{\text{Cov}}(\hat{\hat{E}}_{j_\mathbf{s}}, \hat{\psi}_{j_R})$ are the within block covariances between the Horvitz-Thompson estimators (6.58) and $\hat{\hat{E}}_{j_\mathbf{s}}$.

The variance and covariance components (6.54), (6.55), (6.56) and (6.57) vanish when within each block a separate regression estimator is applied minimally using the block size as auxiliary information, since then $\mathbf{d}^t\mathbf{x}_i = \psi_j$, $\forall i$, see (6.84) in appendix 6.9.3. This condition might be met in situations where interviewer regions coincide exactly with the strata or the PSU's of the sampling design. We emphasis that if the block variables are not incorporated in the weighting scheme of the regression estimator, the components (6.54), (6.55), (6.56) and (6.57) are nonzero so that the generalized regression estimator might increase the variance of the estimated treatment effects.

6.6.3 Efficiency of blocking on interviewers

In the following section we elaborate on the expecations E_α and E_s once more, but now for CRD's. This enables us to study the effect of blocking on interviewers under general complex sampling designs. The derivation of $\text{E}_\alpha\text{E}_s d_k$ for a CRD conducted under a measurement error model with random interviewer effects, analyzed with the Horvitz-Thompson estimator proceeds equivalently to the derivation of (6.43) in appendix 6.9.5. As a result we have

$$\begin{aligned}\text{E}_\alpha\text{E}_s d_k \;=\; & \frac{n}{(n-1)}\frac{n}{n_k}\Big[\tilde{\text{Var}}(\hat{\hat{U}}_{k_\mathbf{s}}) + \sum_{j=1}^{J}\left(\frac{\tau_j^2}{N^2}\tilde{\text{Var}}(\hat{N}_j) + \frac{N_j^2}{N^2}\frac{(n-n_j)}{n}\tau_j^2\right) \\ & -\frac{1}{n}\left(\text{Var}(\hat{\hat{U}}_{k_\mathbf{s}}) + \sum_{j=1}^{J}\frac{\tau_j^2}{N^2}\text{Var}(\hat{N}_j)\right)\Big] + \frac{1}{N^2}\frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{\pi_i}.\end{aligned} \tag{6.59}$$

Here $\text{Var}(\hat{\hat{U}}_{k_\mathbf{s}})$ and $\tilde{\text{Var}}(\hat{\hat{U}}_{k_\mathbf{s}})$ are defined in (5.1) and (5.2) in section 5.2. An expression of $\text{E}_\alpha\text{E}_s d_{k_R}$ for a CRD conducted under a measurement error model with random interviewer effects, analyzed with the generalized regression estimator is given by (6.59), where $\tilde{\text{Var}}(\hat{\hat{U}}_{k_\mathbf{s}})$ and $\text{Var}(\hat{\hat{U}}_{k_\mathbf{s}})$ are replaced by $\tilde{\text{Var}}(\hat{\hat{E}}_\mathbf{s})$ and $\text{Var}(\hat{\hat{E}}_\mathbf{s})$, respectively. Expressions for $\text{Var}(\hat{\hat{E}}_\mathbf{s})$ and $\tilde{\text{Var}}(\hat{\hat{E}}_\mathbf{s})$ are defined in (5.7) and (5.8) in section 5.2.

The derivation of $\text{E}_\alpha\text{E}_s d_k$ for a CRD conducted under a measurement error model with fixed interviewer effects, analyzed with the Horvitz-Thompson estimator, proceeds equivalently to the derivation of (6.48) in appendix 6.9.5. As a result we have

$$\begin{aligned}\text{E}_\alpha\text{E}_s d_k \;=\; & \frac{n}{(n-1)}\frac{n}{n_k}\Big[\tilde{\text{Var}}(\hat{\hat{U}}_{k_\mathbf{s}}) + \tilde{\text{Var}}(\hat{\hat{\psi}}) + 2\tilde{\text{Cov}}(\hat{\hat{U}}_{k_\mathbf{s}}, \hat{\hat{\psi}}) \\ & -\frac{1}{n}\left(\text{Var}(\hat{\hat{U}}_{k_\mathbf{s}}) + \text{Var}(\hat{\hat{\psi}}) + 2\text{Cov}(\hat{\hat{U}}_{k_\mathbf{s}}, \hat{\hat{\psi}})\right)\Big] + \frac{1}{N^2}\frac{n}{n_k}\sum_{i=1}^{N}\frac{\sigma_{ik}^2}{\pi_i}.\end{aligned} \tag{6.60}$$

Here

$$\mathrm{Var}(\hat{\bar{\psi}}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'=1}^{N} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{\psi_j \psi_{j'}}{\pi_i \pi_{i'}},$$

$$\tilde{\mathrm{Var}}(\hat{\bar{\psi}}) = \frac{1}{n} \left(\sum_{i=1}^{N} \frac{n \psi_j^2}{\pi_i N^2} - \bar{\psi}^2 \right),$$

$$\mathrm{Cov}(\hat{\bar{U}}_{k_{\mathbf{s}}}, \hat{\bar{\psi}}) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'=1}^{N} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{u_{ik}}{\pi_i} \frac{\psi_j}{\pi_{i'}},$$

$$\tilde{\mathrm{Cov}}(\hat{\bar{U}}_{k_{\mathbf{s}}}, \hat{\bar{\psi}}) = \frac{1}{n} \left(\sum_{i=1}^{N} \frac{n u_{ik} \psi_j}{\pi_i N^2} - \bar{U} \bar{\psi} \right),$$

where $\bar{\psi} = \sum_{j=1}^{J} N_j / N \psi_j$ denote the weighted mean of the fixed interviewer effects in the finite population and

$$\hat{\bar{\psi}} = \frac{1}{N} \sum_{i=1}^{n} \frac{\psi_j}{\pi_i}$$

the Horvitz-Thompson estimator of $\bar{\psi}$. An expression of $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ for a CRD conducted under a measurement error model with fixed interviewer effects, analyzed with the generalized regression estimator is given by (6.60), where $\mathrm{Var}(\hat{\bar{U}}_{k_{\mathbf{s}}})$ is replaced by $\mathrm{Var}(\hat{\bar{E}}_{\mathbf{s}})$, $\tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_{\mathbf{s}}})$ by $\tilde{\mathrm{Var}}(\hat{\bar{E}}_{\mathbf{s}})$, $\mathrm{Var}(\hat{\bar{\psi}})$ by

$$\mathrm{Var}(\hat{\bar{\psi}}_R) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'=1}^{N} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{(\psi_j - \mathbf{d}^t \mathbf{x}_i)}{\pi_i} \frac{(\psi_{j'} - \mathbf{d}^t \mathbf{x}_{i'})}{\pi_{i'}},$$

$\tilde{\mathrm{Var}}(\hat{\bar{\psi}})$ by

$$\tilde{\mathrm{Var}}(\hat{\bar{\psi}}_R) = \frac{1}{n} \left(\sum_{i=1}^{N} \frac{n (\psi_j - \mathbf{d}^t \mathbf{x}_i)^2}{\pi_i N^2} - (\bar{\psi} - \mathbf{d}^t \bar{\mathbf{X}})^2 \right),$$

$\mathrm{Cov}(\hat{\bar{U}}_{k_{\mathbf{s}}}, \hat{\bar{\psi}})$ by

$$\mathrm{Cov}(\hat{\bar{E}}_{\mathbf{s}}, \hat{\bar{\psi}}_R) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'=1}^{N} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{(u_i - \mathbf{b}^t \mathbf{x}_i)}{\pi_i} \frac{(\psi_j - \mathbf{d}^t \mathbf{x}_{i'})}{\pi_{i'}},$$

and $\tilde{\mathrm{Cov}}(\hat{\bar{U}}_{k_{\mathbf{s}}}, \hat{\bar{\psi}})$ by

$$\tilde{\mathrm{Cov}}(\hat{\bar{E}}_{\mathbf{s}}, \hat{\bar{\psi}}_R) = \frac{1}{n} \left(\sum_{i=1}^{N} \frac{n (u_i - \mathbf{b}^t \mathbf{x}_i)(\psi_j - \mathbf{d}^t \mathbf{x}_i)}{\pi_i N^2} - (\bar{U} - \mathbf{b}^t \bar{\mathbf{X}})(\bar{\psi} - \mathbf{d}^t \bar{\mathbf{X}}) \right).$$

It is assumed that the covariance between the intrinsic values and the fixed interviewer effects is equal to zero. Now, when applying a CRD instead of an RBD with interviewers as block variables, the variance increase in $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ is mainly caused by three factors:

- A between-block variance of the fixed interviewer effects as far as it is not explained by the auxiliary information in the case of the generalized regression estimator.

- Random interviewer effects as expressed by τ_j^2.
- A between-block variance of the intrinsic values as it is not explained by the auxiliary information in the case of the generalized regression estimator.

We conclude that blocking on interviewers not only eliminates any interviewer effects, but may also be fruitful if the population means of the intrinsic values between the different interviewer regions differ substantially. The latter may be especially expected in surveys where data are collected by CAPI, for then interviewers work in separate relative small regions. In surveys where data are collected by means of CATI, telephone numbers are assigned randomly to an interviewer. Consequently, interviewers work through the entire survey population, so the variance between the means of the target variables are expected to be small.

6.7 Equal population variances and self-weighted sampling designs

In this section we will further improve the variance estimation procedure for RBD's by utilizing the model assumptions of additive treatment effects and IID measurement errors. In the case of the Horvitz-Thompson estimator

$$\hat{S}_{jk}^2 = \frac{1}{(n_{jk}-1)} \sum_{i=1}^{n_{jk}} \left(\frac{n_j y_{ik}^{\alpha}}{N\pi_i} - \frac{1}{n_{jk}} \sum_{i'=1}^{n_{jk}} \frac{n_j y_{i'k}^{\alpha}}{N\pi_{i'}} \right)^2, \tag{6.61}$$

defined by (3.58) in section 3.6.2, is an estimator of

$$S_{jk}^2 = \frac{1}{(n_j-1)} \sum_{i=1}^{n_j} \left(\frac{n_j y_{ik}^{\alpha}}{\pi_i N} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{n_j y_{i'k}^{\alpha}}{\pi_{i'} N} \right)^2,$$

defined by (3.34) in section 3.6.1. The parameter S_{jk}^2 can be regarded as the population variance of the target variable y_{ik}^{α} (weighted with a factor $n_j/(\pi_i N)$) of the n_j individuals of sample s in block j observed by means of treatment k. In the case of the generalized regression estimator

$$\hat{S}_{E_{jk}}^2 = \frac{1}{(n_{jk}-1)} \sum_{i=1}^{n_{jk}} \left(\frac{n_j (y_{ik}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t} \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_{jk}} \sum_{i'=1}^{n_{jk}} \frac{n_j (y_{i'k}^{\alpha} - \hat{\mathbf{b}}_k^{\alpha^t} \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2, \tag{6.62}$$

defined by (4.31) in section 4.4, is an estimator of:

$$S_{E_{jk}}^2 = \frac{1}{(n_j-1)} \sum_{i=1}^{n_j} \left(\frac{n_j (y_{ik}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{n_j (y_{i'k}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2, \tag{6.63}$$

defined by (4.28) in section 4.4. The parameter $S_{E_{jk}}^2$ can be regarded as the population variance of the residuals $(y_{ik}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_i)$ (weighted with a factor $n_j/(\pi_i N)$) of the n_j individuals of sample s in block j observed by means of treatment k.

In the case of an RBD the sample s of size n is divided in J blocks and each block is divided in K subsamples. Consequently, the n individuals of s are divided into JK groups of size n_{jk}. For each of these JK subsamples separate population variances $\hat{S}^2_{jk}$ or $\hat{S}^2_{E_{jk}}$ have to be estimated. If the number of experimental units n_{jk} available for the estimation of these population variances becomes too small, then these estimates might become unstable. In such situations, more stable estimates can be obtained by pooling estimates of the population variances within the blocks.

In the case of the Horvitz-Thompson estimator, it is assumed that the covariance matrices in the measurement error models discussed in section 3.2 are equal to $\mathbf{\Sigma}_i = \sigma_i^2\mathbf{I}$. Under this assumption it follows that the expectation of S^2_{jk} with respect to the measurement error model and the sampling design within each block only depends on treatment k through the additive treatment effect β_k. As a result it follows that the expected values of these population variances within each block are equal, i.e. $\mathrm{E}_\alpha \mathrm{E}_s S^2_{j1} = \ldots = \mathrm{E}_\alpha \mathrm{E}_s S^2_{jK} = \mathrm{E}_\alpha \mathrm{E}_s S^2_j$, for $j = 1, 2, \ldots, J$. Under this assumption, a more efficient estimate of the population variance $\mathrm{E}_\alpha \mathrm{E}_s S^2_j$ is obtained if the estimates of the population variances of the K experimental groups within each block are pooled:

$$\hat{S}^2_j = \frac{1}{n_j - K} \sum_{k=1}^{K} \sum_{i=1}^{n_{jk}} \left(\frac{n_j y^\alpha_{ik}}{\pi_i N} - \frac{1}{n_{jk}} \sum_{i'=1}^{n_{jk}} \frac{n_j y^\alpha_{i'k}}{\pi_{i'} N} \right)^2 . \tag{6.64}$$

In the case of the generalized regression estimator, the residuals $(y^\alpha_{ik} - \mathbf{b}^t_k \mathbf{x}_i)$ only depend on the $k-$th treatment effect through the measurement errors ε^α_{ik}. Under the assumption that the covariance matrices in the measurement error models discussed in section 4.2 are equal to $\mathbf{\Sigma}_i = \sigma_i^2\mathbf{I}$, it follows that $\mathrm{E}_\alpha \mathrm{E}_s S^2_{E_{jk}}$ within each block are identical parameters, i.e. $\mathrm{E}_\alpha \mathrm{E}_s S^2_{E_{j1}} = \ldots = \mathrm{E}_\alpha \mathrm{E}_s S^2_{E_{jK}} = \mathrm{E}_\alpha \mathrm{E}_s S^2_{E_j}$, for $j = 1, 2, \ldots, J$. Under this assumption the $\hat{S}^2_{E_{jk}}$'s, within each block, are K estimators for the same parameter $\mathrm{E}_\alpha \mathrm{E}_s S^2_{E_j}$. A more efficient estimator for $\mathrm{E}_\alpha \mathrm{E}_s S^2_{E_j}$ is the pooled variance estimator

$$\hat{S}^2_{E_j} = \frac{1}{(n_j - 1)} \sum_{k=1}^{K} \sum_{i=1}^{n_{jk}} \left(\frac{n_j (y^\alpha_{ik} - \hat{\mathbf{b}}^{\alpha^t}_k \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{k'=1}^{K} \sum_{i'=1}^{n_{jk'}} \frac{n_j (y^\alpha_{i'k'} - \hat{\mathbf{b}}^{\alpha^t}_{k'} \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2 \tag{6.65}$$

or as an alternative

$$\hat{S}^2_{E_j} = \frac{1}{(n_j - K)} \sum_{k=1}^{K} \sum_{i=1}^{n_{jk}} \left(\frac{n_j (y^\alpha_{ik} - \hat{\mathbf{b}}^{\alpha^t}_k \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_{jk}} \sum_{i'=1}^{n_{jk}} \frac{n_j (y^\alpha_{i'k} - \hat{\mathbf{b}}^{\alpha^t}_k \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2 . \tag{6.66}$$

There are several special cases where the design-based Wald statistic coincides with the $F-$ statistics known from more standard model-based analysis procedures. Consider an RBD embedded in a self-weighted sampling design where the treatments are allocated proportionally over the blocks, i.e. $\pi_i = n/N$ and $n_{jk}/n_j = n_k/n$ for all $j = 1, \ldots, J$. Then, it follows from the results obtained for the ratio estimator (4.33), section 4.4 that $\hat{\bar{Y}}^\alpha_{k_R} = \frac{1}{n_k} \sum_{i=1}^{n_k} y^\alpha_{ik} \equiv \bar{y}^\alpha_k$ and $\hat{\mathbf{b}}^{\alpha^t}_k \mathbf{x}_i = \bar{y}^\alpha_k$. Furthermore, denote $\bar{y}^\alpha_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y^\alpha_{ik}$ and $\bar{y}^\alpha = \frac{1}{n} \sum_{k=1}^{K} \sum_{i=1}^{n_k} y^\alpha_{ik}$, then it follows that

$$\frac{1}{n_j} \sum_{k=1}^{K} \sum_{i=1}^{n_{jk}} y^\alpha_{ik} = \bar{y}^\alpha_j ,$$

and

$$\frac{1}{n_j}\sum_{k=1}^{K}\sum_{i=1}^{n_{jk}}\hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_i = \frac{1}{n_j}\sum_{k=1}^{K}\sum_{i=1}^{n_{jk}}\bar{y}_k^{\alpha} = \sum_{k=1}^{K}\frac{n_{jk}}{n_j}\bar{y}_k^{\alpha} = \sum_{k=1}^{K}\frac{n_k}{n}\bar{y}_k^{\alpha} = \bar{y}^{\alpha}.$$

If $n_j \approx n_j - 1$, then it follows under the pooled variance estimator (6.65) that

$$\begin{aligned}\hat{d}_k &= \sum_{j=1}^{J}\frac{n_j}{n_{jk}}\frac{n_j}{n_j-1}\sum_{k=1}^{K}\sum_{i=1}^{n_{jk}}\left(\frac{y_{ik}^{\alpha}-\hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_i}{N\pi_i} - \frac{1}{n_j}\sum_{k'=1}^{K}\sum_{i'=1}^{n_{jk'}}\frac{y_{i'k'}^{\alpha}-\hat{\mathbf{b}}_{k'}^{\alpha^t}\mathbf{x}_{i'}}{N\pi_{i'}}\right)^2 \\ &\approx \frac{1}{n_k}\frac{1}{n}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{i=1}^{n_{jk}}\left(y_{ik}^{\alpha}-\bar{y}_k^{\alpha}-\bar{y}_j^{\alpha}+\bar{y}^{\alpha}\right)^2 \equiv \frac{\hat{d}}{n_k}.\end{aligned} \tag{6.67}$$

Denote $\bar{y}_{jk}^{\alpha} = \frac{1}{n_{jk}}\sum_{i=1}^{n_{jk}} y_{ik}^{\alpha}$. Under the pooled variance estimator (6.66) it follows that

$$\begin{aligned}\hat{d}_k &= \sum_{j=1}^{J}\frac{n_j}{n_{jk}}\frac{n_j}{n_j-K}\sum_{k=1}^{K}\sum_{i=1}^{n_{jk}}\left(\frac{y_{ik}^{\alpha}-\hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_i}{N\pi_i} - \frac{1}{n_{jk}}\sum_{i'=1}^{n_{jk}}\frac{y_{i'k}^{\alpha}-\hat{\mathbf{b}}_k^{\alpha^t}\mathbf{x}_{i'}}{N\pi_{i'}}\right)^2 \\ &\approx \frac{1}{n_k}\frac{1}{n}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{i=1}^{n_{jk}}\left(y_{ik}^{\alpha}-\bar{y}_{jk}^{\alpha}\right)^2 \equiv \frac{\hat{d}}{n_k}.\end{aligned} \tag{6.68}$$

If these pooled variance estimators are substituted into the Wald statistic (4.35) in section 4.5, then it follows that

$$W = \frac{1}{\hat{d}}\left(\sum_{k=1}^{K} n_k\left(\bar{y}_k^{\alpha}\right)^2 - n\left(\bar{y}^{\alpha}\right)^2\right), \tag{6.69}$$

where $\hat{d}$ is given by (6.67) or (6.68). It can be recognized that $W/(K-1)$ in (6.69) with the pooled variance estimator (6.67) corresponds with the $F-$statistic of an ANOVA for a two-way layout without interactions. If (6.68) is applied, then $W/(K-1)$ corresponds with the $F-$statistic of an ANOVA for a two-way layout with interactions (Scheffé (1959, ch.4). Under the assumption of normally and independently distributed observations, it follows that under the null hypothesis $F \simeq \mathcal{F}_{[n-K]}^{[K-1]}$. If n tends to infinity, then $\mathcal{F}_{[n-K]}^{[K-1]}$ tends to $\chi^2_{[K-1]}/(K-1)$ and consequently the Wald statistic and the $F-$statistic asymptotically have the same distribution.

6.8 Two sample problem

In section 5.4, the two sample problem is treated as a special case of a CRD. It is, however, also possible to design a two treatment embedded experiment as an RBD. Hypotheses of interest are defined by (5.21) and can be tested by means of the design-based $t-$statistic, defined by (5.22) for the Horvitz-Thompson estimator and (5.26) for the generalized regression estimator. For the Horvitz-Thompson estimator expressions of the difference between the two subsample means $\hat{\hat{Y}}_1^{\alpha}$ and $\hat{\hat{Y}}_2^{\alpha}$, in the enumerator of the $t-$statistic, are given by (3.22) or (3.23) in section 3.5. Expressions for the generalized regression estimators $\hat{\hat{Y}}_{1_R}^{\alpha}$ and $\hat{\hat{Y}}_{2_R}^{\alpha}$ are defined by and (4.16) or (4.17) in section 4.3. The variance of the difference between the two subsample means in the denominator of the $t-$statistic is given by $\text{Var}(\hat{\hat{Y}}_1^{\alpha} - \hat{\hat{Y}}_2^{\alpha}) = d_1 + d_2$, for the Horvitz-Thompson estimator and $\text{Var}(\hat{\bar{Y}}^{\alpha}_{1_R} - \hat{\bar{Y}}^{\alpha}_{2_R}) = d_{1_R} + d_{2_R}$ for the generalized regression estimator. Expressions for d_k and d_{k_R} are derived in sections 6.3 through 6.6. Design-unbiased estimators for these variances are obtained by $\hat{d}_k$ given in (3.58) section 3.5 for the Horvitz-Thompson estimator and $\hat{d}_{k_R}$ defined in (4.31) section 4.3 for the generalized regression estimator. The standard normal distribution can be used to construct critical regions for the derived $t-$statistic, which yield very nearly $(1-\gamma)\%$ coverage, where γ denotes the size of the test (see sections 3.7.2, 4.5 and 5.4).

6.9 Appendix

6.9.1 Proof of formula (6.6)

Consider an RBD embedded in a two-stage sampling scheme where PSU's are block variables, conducted under the basic measurement error model (3.1) in section 3.2 and analyzed with the Horvitz-Thompson estimator. It is proved that

$$\begin{aligned} \text{E}_{\alpha}\text{E}_{s} d_k &= \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\text{Var}}(\hat{\bar{U}}_{jk_{\mathbf{s}}}) - \frac{1}{n_j} \text{Var}(\hat{\bar{U}}_{jk_{\mathbf{s}}}) \right) \\ &+ \sum_{j=1}^{J_u} \frac{n_j}{n_{jk}} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_i N^2}. \end{aligned}$$

Let $\hat{\bar{Y}}^{\alpha}_{jk_{\mathbf{s}}}$ denote the Horvitz-Thompson estimator for $\bar{Y}^{\alpha}_{jk}$, i.e. the population mean in block j, based on the n_j individuals of sample s_j. Then

$$\text{Var}(\hat{\bar{Y}}^{\alpha}_{jk_{\mathbf{s}}}) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j}) \frac{y^{\alpha}_{ik} y^{\alpha}_{i'k}}{\pi_{i|j}\pi_{i'|j}}$$

denotes the design variance of $\hat{\bar{Y}}^{\alpha}_{jk_{\mathbf{s}}}$ and

$$\tilde{\text{Var}}(\hat{\bar{Y}}^{\alpha}_{jk_{\mathbf{s}}}) = \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j {y^{\alpha}_{ik}}^2}{\pi_{i|j} N_j^2} - \bar{Y}^2_{jk} \right)$$

the variance of $\hat{\bar{Y}}^{\alpha}_{jk_{\mathbf{s}}}$ as if n_j elements are drawn with replacement from block j of size N_j with selection probabilities $\pi_{i|j}/n_j$.

The diagonal elements d_k are defined by (3.34) in section 3.6.1. Let E_{s_I} denote the expectation with respect to the first stage of the sampling design and $\text{E}_{s_{II}}$ the expectation with respect to the second stage of the sampling design conditional on the realization of the first stage. In order to take the expectation with respect to the sampling design we condition on the first stage of the sampling design as follows; $\text{E}_s(.) = \text{E}_{s_I}[\text{E}_{s_{II}}(.) \mid s_I]$ where E_{s_I} denotes the expectation with respect to the sampling design of the first stage and $\text{E}_{s_{II}}(.) \mid s_I$ the the expectation with respect to the second stage of the sampling design, conditional on the first stage. Using the derivation applied in (5.30) in appendix 5.5.1, we can elaborate on the expectation with respect to the sampling design as follows:

$$\begin{aligned}
&\mathrm{E}_s \sum_{j=1}^{J_s} \frac{1}{(n_j-1)} \frac{1}{n_{jk}} \sum_{i=1}^{n_j} \left(\frac{n_j y_{ik}^{\alpha}}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{n_j y_{i'k}^{\alpha}}{N\pi_{i'}} \right)^2 \\
&= \mathrm{E}_{s_I} \mathrm{E}_{s_{II}} \left[\sum_{j=1}^{J_s} \frac{(N_j/N)^2}{\pi_j^2} \frac{1}{(n_j-1)} \frac{1}{n_{jk}} \sum_{i=1}^{n_j} \left(\frac{n_j y_{ik}^{\alpha}}{\pi_{i|j} N_j} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{n_j y_{i'k}^{\alpha}}{\pi_{i'|j} N_j} \right)^2 \mid s_I \right] \\
&= \mathrm{E}_{s_I} \sum_{j=1}^{J_s} \frac{(N_j/N)^2}{\pi_j^2} \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\mathrm{Var}}(\hat{\hat{Y}}_{jk_s}^{\alpha}) - \frac{1}{n_j} \mathrm{Var}(\hat{\hat{Y}}_{jk_s}^{\alpha}) \right) \\
&= \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\mathrm{Var}}(\hat{\hat{Y}}_{jk_s}^{\alpha}) - \frac{1}{n_j} \mathrm{Var}(\hat{\hat{Y}}_{jk_s}^{\alpha}) \right). \qquad (6.70)
\end{aligned}$$

Now we will further evaluate the expectation with respect to the measurement error model under the basic measurement error model (3.1) in section 3.2. Equivalent to the derivation of (5.31) in appendix 5.5.1, it follows that

$$\mathrm{E}_{\alpha} \mathrm{Var}(\hat{\hat{Y}}_{jk_s}^{\alpha}) = \mathrm{Var}(\hat{\hat{U}}_{jk_s}) + \frac{1}{N_j^2} \sum_{i=1}^{N_j} \frac{(1-\pi_{i|j})}{\pi_{i|j}} \sigma_{ik}^2, \qquad (6.71)$$

where $\mathrm{Var}(\hat{\hat{U}}_{jk_s})$ is defined in (6.4) in section 6.2. Equivalent to the derivation of (5.32) in appendix 5.5.1, it follows that

$$\mathrm{E}_{\alpha} \tilde{\mathrm{Var}}(\hat{\hat{Y}}_{jk_s}^{\alpha}) = \tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_s}) + \frac{1}{n_j} \frac{1}{N_j^2} \sum_{i=1}^{N_j} \frac{(n_j-\pi_{i|j})}{\pi_{i|j}} \sigma_{ik}^2, \qquad (6.72)$$

where $\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_s})$ is defined in (6.5) in section 6.2. If results (6.71) and (6.72) are substituted into (6.70), then it follows that

$$\begin{aligned}
\mathrm{E}_{\alpha} \mathrm{E}_s(d_k) &= \mathrm{E}_{\alpha} \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\mathrm{Var}}(\hat{\hat{Y}}_{jk_s}^{\alpha}) - \frac{1}{n_j} \mathrm{Var}(\hat{\hat{Y}}_{jk_s}^{\alpha}) \right) \\
&= \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_s}) - \frac{1}{n_j} \mathrm{Var}(\hat{\hat{U}}_{jk_s}) \right) \\
&\quad + \sum_{j=1}^{J_u} \frac{n_j}{n_{jk}} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_i N^2}, \quad \textbf{QED.}
\end{aligned}$$

6.9.2 Proof of formula (6.11)

For an RBD embedded in a two-stage sampling scheme where PSU's are block variables, conducted under the basic measurement error model (4.5) in section 4.2 and analyzed with the generalized regression estimator, it is proved that

$$\mathrm{E}_{\alpha} \mathrm{E}_s d_{k_R} = \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\mathrm{Var}}(\hat{\hat{E}}_{j_s}) - \frac{1}{n_j} \mathrm{Var}(\hat{\hat{E}}_{j_s}) \right) + \sum_{j=1}^{J_u} \frac{n_j}{n_{jk}} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_i N^2}.$$

Let $\hat{\bar{Y}}^{\alpha}_{jk_{R_{\mathbf{s}}}}$ denote the generalized regression estimator of $\bar{Y}^{\alpha}_{jk}$ based on the n_j sampling units of s_j in block j. The first-order Taylor series approximation of $\hat{\bar{Y}}^{\alpha}_{jk_{R_{\mathbf{s}}}}$ is given by:

$$\hat{\bar{Y}}^{\alpha}_{jk_{R_{\mathbf{s}}}} \doteq \hat{\bar{Y}}^{\alpha}_{jk_{\mathbf{s}}} + \mathbf{b}^t_k(\bar{\mathbf{X}}_j - \hat{\bar{\mathbf{X}}}_{j_{\mathbf{s}}}) = \hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}} + \mathbf{b}^t_k\bar{\mathbf{X}}_j,$$

where

$$\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}} = \hat{\bar{Y}}^{\alpha}_{jk_{\mathbf{s}}} - \mathbf{b}^t_k\hat{\bar{\mathbf{X}}}_{j_{\mathbf{s}}} = \frac{1}{N}\sum_{i=1}^{n}\frac{(y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i)}{\pi_{i|j}}.$$

The design variance of the first-order Taylor series approximation of $\hat{\bar{Y}}^{\alpha}_{jk_{R_{\mathbf{s}}}}$ is given by

$$\mathrm{Var}(\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}}) = \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j})\frac{(y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i)}{\pi_{i|j}}\frac{(y^{\alpha}_{i'k} - \mathbf{b}^t_k\mathbf{x}_{i'})}{\pi_{i'|j}}. \tag{6.73}$$

Furthermore,

$$\tilde{\mathrm{Var}}(\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}}) = \frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n_j(y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i)^2}{\pi_{i|j}N_j^2} - (\bar{Y}^{\alpha}_{jk} - \mathbf{b}^t_k\bar{\mathbf{X}}_j)^2\right) \tag{6.74}$$

denotes the variance of the first-order Taylor series approximation of $\hat{\bar{Y}}^{\alpha}_{jk_{R_{\mathbf{s}}}}$ as if sample s_j in block j has been drawn with replacement with selection probabilities $\pi_{i|j}/n_j$.

The diagonal elements d_{k_R} are defined by (4.28) in section 4.4. Equivalent to the derivation of (6.70) in appendix 6.9.1 it follows that the expectation of d_{k_R} with respect to the sampling design equals

$$\begin{aligned}
&\mathrm{E}_s\sum_{j=1}^{J_s}\frac{1}{(n_j-1)}\frac{1}{n_{jk}}\sum_{i=1}^{n_j}\left(\frac{n_j(y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{n_j(y^{\alpha}_{i'k} - \mathbf{b}^t_k\mathbf{x}_{i'})}{N\pi_{i'}}\right)^2\\
&= \sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\left(\tilde{\mathrm{Var}}(\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}}) - \frac{1}{n_j}\mathrm{Var}(\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}})\right).
\end{aligned} \tag{6.75}$$

Under the basic measurement error model (4.5) in section 4.2 and the condition that there exists a constant $H-$vector $\mathbf{a}$ such that $\mathbf{a}^t\mathbf{x}_i = 1$ for all $i \in U$, it follows from (4.39) in appendix 4.7.2 that $\mathbf{b}_k = \mathbf{b} + \mathbf{a}\beta_k$. Consequently, $y^{\alpha}_{ik} - \mathbf{b}^t_k\mathbf{x}_i = u_i - \mathbf{b}^t\mathbf{x}_i + \varepsilon^{\alpha}_{ik}$. Equivalent to the derivation of (5.37) and (5.38) in appendix 5.5.2 it follows that

$$\mathrm{E}_{\alpha}\mathrm{Var}(\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}}) = \mathrm{Var}(\hat{\bar{E}}_{j_{\mathbf{s}}}) + \frac{1}{N_j^2}\sum_{i=1}^{N_j}\frac{(1-\pi_{i|j})}{\pi_{i|j}}\sigma^2_{ik}, \tag{6.76}$$

and

$$\mathrm{E}_{\alpha}\tilde{\mathrm{Var}}(\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}}) = \tilde{\mathrm{Var}}(\hat{\bar{E}}_{j_{\mathbf{s}}}) + \frac{1}{n_j}\frac{1}{N_j^2}\sum_{i=1}^{N_j}\frac{(n_j - \pi_{i|j})}{\pi_{i|j}}\sigma^2_{ik}. \tag{6.77}$$

If (6.76) and (6.77) are substituted into (6.75), then it follows that

$$\mathrm{E}_{\alpha}\mathrm{E}_s d_{k_R} = \mathrm{E}_{\alpha}\sum_{j=1}^{J_u}\frac{(N_j/N)^2}{\pi_j}\frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\left(\tilde{\mathrm{Var}}(\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}}) - \frac{1}{n_j}\mathrm{Var}(\hat{\bar{E}}^{\alpha}_{jk_{\mathbf{s}}})\right)$$

$$= \sum_{j=1}^{J_u} \frac{(N_j/N)^2}{\pi_j} \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\text{Var}}(\hat{\hat{E}}_{j_\mathbf{s}}) - \frac{1}{n_j} \text{Var}(\hat{\hat{E}}_{j_\mathbf{s}}) \right)$$

$$+ \sum_{j=1}^{J_u} \frac{n_j}{n_{jk}} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{\pi_i N^2}, \quad \textbf{QED.}$$

6.9.3 Proof of formula (6.36) and (6.37)

Consider an RBD with arbitrarily chosen block variables (e.g. interviewers) embedded in a simple random sampling design. The finite population might be divided in J groups of size $(N_1, \ldots, N_J)$, which correspond with the blocks of the experimental design. A sample of size n is drawn by means of simple random sampling without replacement from this population. Consequently, the number of sampling units within each block $(n_1, \ldots, n_J)$ is random. In order to evaluate the expectation with respect to the sampling design we condition on the realized number of sampling units within each block as follows. The probability to draw a simple random sample of size n without replacement is equal to

$$P(s) = \binom{N}{n}^{-1} = \binom{N}{n}^{-1} \left(\prod_{j=1}^{J} \binom{N_j}{n_j} \right) \left(\prod_{j=1}^{J} \binom{N_j}{n_j}^{-1} \right). \tag{6.78}$$

Note that

$$\binom{N}{n}^{-1} \prod_{j=1}^{J} \binom{N_j}{n_j} \tag{6.79}$$

is the probability that a vector $(n_1, \ldots, n_J)$ is drawn from the multivariate hypergeometric distribution and that

$$\prod_{j=1}^{J} \binom{N_j}{n_j}^{-1} \tag{6.80}$$

is the probability of drawing a particular stratified simple random sample of given size $(n_1, \ldots, n_J)$ without replacement. It follows that the sample space of simple random sampling without replacement is equal to product of the probability of drawing a vector $(n_1, \ldots, n_J)$ from the multivariate hypergeometric distribution and the sample space of stratified sampling from J strata where in each stratum n_j elements are drawn by means of simple random sampling without replacement from a population of size N_j. Since n sampling units are drawn by means of simple random sampling without replacement from a population of size N, which can be classified into J groups of size $(N_1, \ldots, N_J)$, it follows that the distribution of the n sampling units over the J groups $(n_1, \ldots, n_J)$ is the multivariate hypergeometric distribution (Johnson and Kotz, 1969, Ch. 11.5). Consequently, the marginal distribution of n_j is the hypergeometric distribution. Now we can evaluate the expectation with respect to the sampling design by conditioning on the realized number of sampling units within each block as follows:

$$\text{E}_s(.) = \text{E}_h \left[\text{E}_{str}(.) \mid n_j \right],$$

with E_{str} the expectation of stratified simple random sample without replacement and E_h the expectation of n_j with respect to the hypergeometric distribution. First we condition on the realized distribution of the sampling units over the J blocks $(n_1, \ldots, n_J)$ and take the expectation as if a stratified simple random sample without replacement is drawn from J strata with $\pi_i = n/N$. Strata are formed by the blocks of the experimental design and within each stratum n_j elements are drawn by means of simple random sampling from a population of size N_j where $N_j = N(n_j/n)$. Second we can evaluate the expectation of the number of sampling units within each stratum (block) as if the sample size n_j in stratum or block j is a realization from the hypergeometric distribution. Here we use $\mathrm{E}_h(n_j/n) = (N_j/N)$. Within each block the n_j experimental units are according to the experimental design randomized over the K treatments with probability $f_{jk} = n_{jk}/n_j$. Therefore, f_{jk} is fixed with respect to the sampling design.

Proof of formula (6.36)

It is proved for the direct estimator under a measurement error model with interviewer effects that

$$\mathrm{E}_\alpha \mathrm{E}_s d_k = \sum_{j=1}^{J} \frac{N_j}{N} \frac{n_j}{n_{jk}} \frac{1}{n\,(N_j-1)} \sum_{i=1}^{N_j} \left(u_i - \frac{1}{N_j} \sum_{i'=1}^{N_j} u_{i'} \right)^2 + \sum_{j=1}^{J} \frac{n_j}{n_{jk}\, n} \sum_{i=1}^{N_j} \frac{\sigma_{ik}^2}{N}.$$

The diagonal elements d_k are defined in (3.34) in section 3.6.1. First we take the expectation with respect to the sampling design.

$$\begin{aligned}
&\mathrm{E}_s \sum_{j=1}^{J} \frac{1}{(n_j-1)} \frac{1}{n_j} \sum_{i=1}^{n_j} \left(\frac{n_j y_{ik}^\alpha}{\pi_i N} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{n_j y_{i'k}^\alpha}{\pi_{i'} N} \right)^2 \\
&= \mathrm{E}_h \left[\mathrm{E}_{str} \sum_{j=1}^{J} \frac{n_j}{n} \frac{n_j}{n_{jk}} \frac{1}{n} \frac{1}{(n_j-1)} \sum_{i=1}^{n_j} \left(y_{ik}^\alpha - \frac{1}{n_j} \sum_{i'=1}^{n_j} y_{i'k}^\alpha \right)^2 \mid n_j \right] \\
&= \mathrm{E}_h \sum_{j=1}^{J} \frac{n_j}{n} \frac{1}{f_{jk}} \frac{1}{n} \frac{1}{(N_j-1)} \sum_{i=1}^{N_j} \left(y_{ik}^\alpha - \frac{1}{N_j} \sum_{i'=1}^{N_j} y_{i'k}^\alpha \right)^2 \\
&= \sum_{j=1}^{J} \frac{N_j}{N} \frac{1}{f_{jk}} \frac{1}{n} \frac{1}{(N_j-1)} \sum_{i=1}^{N_j} \left(y_{ik}^\alpha - \frac{1}{N_j} \sum_{i'=1}^{N_j} y_{i'k}^\alpha \right)^2 . \qquad (6.81)
\end{aligned}$$

If the measurement error model with random interviewer effects (3.4) in section 3.2 is substituted into (3.34), then it follows from model assumptions (3.2), (3.3), (3.6), (3.7) and (3.8) that

$$\begin{aligned}
&\mathrm{E}_\alpha \sum_{i=1}^{N_j} \left(y_{ik}^\alpha - \frac{1}{N_j} \sum_{i'=1}^{N_j} y_{i'k}^\alpha \right)^2 \\
&= \mathrm{E}_\alpha \sum_{i=1}^{N_j} \left((u_i + \beta_k + \psi_j + \xi_j^\alpha + \varepsilon_{ik}^\alpha) - \frac{1}{N_j} \sum_{i'=1}^{N_j} (u_{i'} + \beta_k + \psi_j + \xi_j^\alpha + \varepsilon_{i'k}^\alpha) \right)^2 \\
&= \mathrm{E}_\alpha \sum_{i=1}^{N_j} \left[\left(u_i - \frac{1}{N_j} \sum_{i'=1}^{N_j} u_{i'} \right) + \left(\beta_k - \frac{1}{N_j} \sum_{i'=1}^{N_j} \beta_k \right) + \left(\psi_j - \frac{1}{N_j} \sum_{i'=1}^{N_j} \psi_j \right) \right.
\end{aligned}$$

$$+\left(\xi_j^\alpha - \frac{1}{N_j}\sum_{i'=1}^{N_j}\xi_j^\alpha\right) + \left(\varepsilon_{ik}^\alpha - \frac{1}{N_j}\sum_{i'=1}^{N_j}\varepsilon_{i'k}^\alpha\right)\Bigg]^2$$

$$= \sum_{i=1}^{N_j}\left(u_i - \frac{1}{N_j}\sum_{i'=1}^{N_j}u_{i'}\right)^2 + \frac{(N_j-1)}{N_j}\sum_{i=1}^{N_j}\sigma_{ik}^2. \qquad (6.82)$$

We illustrate for some of the cross-products how they are worked out

$$\mathrm{E}_\alpha\sum_{i=1}^{N_j}\left(\varepsilon_{ik}^\alpha - \frac{1}{N_j}\sum_{i'=1}^{N_j}\varepsilon_{i'k}^\alpha\right)^2 = \mathrm{E}_\alpha\left(\sum_{i=1}^{N_j}{\varepsilon_{ik}^\alpha}^2 - \frac{1}{N_j}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\varepsilon_{ik}^\alpha\varepsilon_{i'k}^\alpha\right)$$

$$= \frac{(N_j-1)}{N_j}\sum_{i=1}^{N_j}\sigma_{ik}^2.$$

$$\mathrm{E}_\alpha\sum_{i=1}^{N_j}\left(\xi_j^\alpha - \frac{1}{N_j}\sum_{i'=1}^{N_j}\xi_j^\alpha\right)^2 = \mathrm{E}_\alpha\left(\sum_{i=1}^{N_j}{\xi_j^\alpha}^2 - \frac{1}{N_j}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\xi_j^\alpha\xi_j^\alpha\right)$$

$$= \sum_{i=1}^{N_j}\tau_j^2 - \frac{1}{N_j}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\tau_j^2 = 0.$$

$$\mathrm{E}_\alpha\sum_{i=1}^{N_j}\left(u_i - \frac{1}{N_j}\sum_{i'=1}^{N_j}u_{i'}\right)\left(\varepsilon_{ik}^\alpha - \frac{1}{N_j}\sum_{i'=1}^{N_j}\varepsilon_{i'k}^\alpha\right) = 0.$$

If result (6.82) is substituted into result (6.81), then it follows that

$$\mathrm{E}_\alpha\mathrm{E}_s d_k = \sum_{j=1}^{J}\frac{N_j}{N}\frac{n_j}{n_{jk}\,n}\frac{1}{(N_j-1)}\sum_{i=1}^{N_j}\left(u_{ik} - \frac{1}{N_j}\sum_{i'=1}^{N_j}u_{i'k}\right)^2 + \sum_{j=1}^{J}\frac{n_j}{n_{jk}\,n}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{N}, \quad \textbf{QED}.$$

Proof of formula (6.37)

For the generalized regression estimator with a measurement error model with interviewer effects it is proved that

$$\mathrm{E}_\alpha\mathrm{E}_s d_{k_R} =$$
$$\sum_{j=1}^{J}\frac{N_j}{N}\frac{n_j}{n_{jk}n(N_j-1)}\sum_{i=1}^{N_j}\left((u_i-\mathbf{b}^t\mathbf{x}_i)+(\psi_j-\mathbf{d}^t\mathbf{x}_i)-\frac{1}{N_j}\sum_{i'=1}^{N_j}\Big((u_{i'}-\mathbf{b}^t\mathbf{x}_{i'})+(\psi_j-\mathbf{d}^t\mathbf{x}_{i'})\Big)\right)^2$$
$$+\sum_{j=1}^{J}\frac{n_j}{n_{jk}\,n}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{N}.$$

The diagonal elements d_{k_R} are defined in (4.28) in section 4.4. According to the derivation of (6.81) it follows that the expectation of d_{k_R} with respect to the sampling design equals

$$\mathrm{E}_s\sum_{j=1}^{J}\frac{1}{(n_j-1)}\frac{1}{n_j}\sum_{i=1}^{n_j}\left(\frac{n_j(y_{ik}^\alpha-\mathbf{b}_k^t\mathbf{x}_i)}{\pi_i N} - \frac{1}{n_j}\sum_{i'=1}^{n_j}\frac{n_j(y_{i'k}^\alpha-\mathbf{b}_k^t\mathbf{x}_{i'})}{\pi_{i'}N}\right)^2$$

$$= \sum_{j=1}^{J}\frac{N_j}{N}\frac{1}{f_{jk}}\frac{1}{n}\frac{1}{(N_j-1)}\sum_{i=1}^{N_j}\left((y_{ik}^\alpha-\mathbf{b}_k^t\mathbf{x}_i) - \frac{1}{N_j}\sum_{i'=1}^{N_j}(y_{i'k}^\alpha-\mathbf{b}_k^t\mathbf{x}_{i'})\right)^2. \qquad (6.83)$$

Under a measurement error model with interviewer effects (4.8), section 4.2, and the condition that there exists a constant $H-$vector such that $\mathbf{a}^t\mathbf{x}_i = 1$ for all $i \in U$ it follows from (4.39) in appendix 4.7.2 that $\mathbf{b}_k = \mathbf{b} + \mathbf{d} + \mathbf{a}\beta_k$, where $\mathbf{d}$ denote the regression coefficients from the regression function of the interviewer effects on the auxiliary variables $\mathbf{x}_i$. Consequently,

$$y_{ik}^{\alpha} - \mathbf{b}_k^t\mathbf{x}_i = u_i - \mathbf{b}^t\mathbf{x}_i + \psi_j - \mathbf{d}^t\mathbf{x}_i + \xi_j^{\alpha} + \varepsilon_{ik}^{\alpha}. \tag{6.84}$$

Using (6.84) we can evaluate the expectation of the sum of squares of the residuals in each block in (6.83) with respect to the measurement error model as follows:

$$\begin{aligned}
&\mathrm{E}_\alpha \sum_{i=1}^{N_j}\left((y_{ik}^{\alpha} - \mathbf{b}_k^t\mathbf{x}_i) - \frac{1}{N_j}\sum_{i'=1}^{N_j}(y_{i'k}^{\alpha} - \mathbf{b}_k^t\mathbf{x}_{i'})\right)^2 \\
&= \mathrm{E}_\alpha \sum_{i=1}^{N_j}\left(u_i - \mathbf{b}^t\mathbf{x}_i + \psi_j - \mathbf{d}^t\mathbf{x}_i + \xi_j^{\alpha} + \varepsilon_{ik}^{\alpha}) - \frac{1}{N_j}\sum_{i'=1}^{N_j}(u_{i'} - \mathbf{b}^t\mathbf{x}_{i'} + \psi_j - \mathbf{d}^t\mathbf{x}_{i'} + \xi_j^{\alpha} + \varepsilon_{i'k}^{\alpha})\right)^2 \\
&= \mathrm{E}_\alpha \sum_{i=1}^{N_j}\left[\left((u_i - \mathbf{b}^t\mathbf{x}_i) + (\psi_j - \mathbf{d}^t\mathbf{x}_i) - \frac{1}{N_j}\sum_{i'=1}^{N_j}(u_{i'} - \mathbf{b}^t\mathbf{x}_{i'}) + (\psi_j - \mathbf{d}^t\mathbf{x}_{i'})\right) + \right. \\
&\qquad \left. + \left(\xi_j^{\alpha} - \frac{1}{N_j}\sum_{i'=1}^{N_j}\xi_j^{\alpha}\right) + \left(\varepsilon_{ik}^{\alpha} - \frac{1}{N_j}\sum_{i'=1}^{N_j}\varepsilon_{i'k}^{\alpha}\right)\right]^2 \\
&= \sum_{i=1}^{N_j}\left((u_i - \mathbf{b}^t\mathbf{x}_i) + (\psi_j - \mathbf{d}^t\mathbf{x}_i) - \frac{1}{N_j}\sum_{i'=1}^{N_j}(u_{i'} - \mathbf{b}^t\mathbf{x}_{i'}) + (\psi_j - \mathbf{d}^t\mathbf{x}_{i'})\right)^2 \\
&\qquad + \frac{(N_j - 1)}{N_j}\sum_{i=1}^{N_j}\sigma_{ik}^2.
\end{aligned} \tag{6.85}$$

If result (6.85) is substituted into (6.83), then it follows that

$$\begin{aligned}
&\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} = \\
&\sum_{j=1}^{J}\frac{N_j}{N}\frac{n_j}{n_{jk}}\frac{1}{n(N_j-1)}\sum_{i=1}^{N_j}\left((u_i - \mathbf{b}^t\mathbf{x}_i) + (\psi_j - \mathbf{d}^t\mathbf{x}_i) - \frac{1}{N_j}\sum_{i'=1}^{N_j}\left((u_{i'} - \mathbf{b}^t\mathbf{x}_{i'}) + (\psi_j - \mathbf{d}^t\mathbf{x}_{i'})\right)\right)^2 \\
&\quad + \sum_{j=1}^{J}\frac{n_j}{n_{jk}\,n}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{N}, \quad \textbf{QED}.
\end{aligned}$$

6.9.4 Proof of formula (6.38)

In this appendix we evaluate the expectation with respect to the sampling design and the measurement error model of $\mathrm{E}_\alpha \mathrm{E}_s d_k$ for CRD under simple random sampling and a measurement error model with random interviewer effects.

$$\mathrm{E}_\alpha \mathrm{E}_s d_k \;=\; \mathrm{E}_\alpha \mathrm{E}_s \frac{1}{n_k}\frac{1}{(n-1)}\sum_{i=1}^{n}\left(y_{ik}^{\alpha} - \frac{1}{n}\sum_{i'=1}^{n}y_{i'k}^{\alpha}\right)^2$$

$$
\begin{aligned}
&= \mathrm{E}_\alpha \frac{1}{n_k}\frac{1}{(N-1)}\sum_{i=1}^{N}\left(y_{ik}^\alpha - \frac{1}{N}\sum_{i'=1}^{N} y_{i'k}^\alpha\right)^2 \\
&= \mathrm{E}_\alpha \frac{1}{n_k}\frac{1}{(N-1)}\sum_{i=1}^{N}\left((u_i+\beta_k+\xi_j^\alpha+\varepsilon_{ik}^\alpha) - \frac{1}{N}\sum_{i'=1}^{N}(u_{i'}+\beta_k+\xi_j^\alpha+\varepsilon_{i'k}^\alpha)\right)^2 \\
&= \mathrm{E}_\alpha \frac{1}{n_k}\frac{1}{(N-1)}\sum_{i=1}^{N}\left[\left(u_i - \frac{1}{N}\sum_{i'=1}^{N} u_{i'}\right) + \left(\beta_k - \frac{1}{N}\sum_{i'=1}^{N}\beta_k\right)\right. \\
&\qquad \left. + \left(\xi_j^\alpha - \frac{1}{N}\sum_{i'=1}^{N}\xi_j^\alpha\right) + \left(\varepsilon_{ik}^\alpha - \frac{1}{N}\sum_{i'=1}^{N}\varepsilon_{i'k}^\alpha\right)\right]^2 \\
&= \frac{1}{n_k}\frac{1}{(N-1)}\sum_{i=1}^{N}\left(u_i - \frac{1}{N}\sum_{i'=1}^{N}u_{i'}\right)^2 + \frac{1}{n_k}\frac{1}{(N-1)}\sum_{j=1}^{J}N_j\tau_j^2\left(1-\frac{N_j}{N}\right) \\
&\qquad + \frac{1}{n_k}\sum_{i=1}\frac{\sigma_{ik}^2}{N}.
\end{aligned} \tag{6.86}
$$

We illustrate for some cross-products how they are worked out:

$$
\begin{aligned}
\mathrm{E}_\alpha \sum_{i\in U}\left(\xi_j^\alpha - \frac{1}{N}\sum_{i'\in U}\xi_j^\alpha\right)^2 &= \mathrm{E}_\alpha\left(\sum_{i\in U}\xi_j^{\alpha 2} - \frac{1}{N}\sum_{i\in U}\sum_{i'\in U}\xi_j^\alpha\xi_{j'}^\alpha\right) \\
&= \mathrm{E}_\alpha\left(\sum_{j=1}^{J}N_j\xi_j^{\alpha 2} - \frac{1}{N}\sum_{j=1}^{J}N_j^2\xi_j^{\alpha 2} - \frac{1}{N}\sum_{j=1}^{J}\sum_{j'=1}^{J}N_j\xi_j^\alpha N_{j'}\xi_{j'}^\alpha\right) \\
&= \sum_{j=1}^{J}\left(N_j\tau_j^2 - \frac{N_j^2}{N}\tau_j^2\right).
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}_\alpha \sum_{i=1}^{N}\left(\varepsilon_{ik}^\alpha - \frac{1}{N}\sum_{i'=1}^{N}\varepsilon_{i'k}^\alpha\right)^2 &= \mathrm{E}_\alpha\left(\sum_{i=1}^{N}\varepsilon_{ik}^{\alpha 2} - \frac{1}{N}\sum_{i=1}^{N}\sum_{i'=1}^{N}\varepsilon_{ik}^\alpha\varepsilon_{i'k}^\alpha\right) \\
&= \frac{N-1}{N}\sum_{i=1}^{N}\sigma_{ik}^2, \quad \textbf{QED}.
\end{aligned}
$$

6.9.5 Proof of formula (6.43) and (6.48)

In this appendix we evaluate the expectation with respect to the measurement error model and the sampling design of d_k for RBD's where interviewers are block variables under a general complex sampling schemes analyzed with the Horvitz-Thompson estimator. The block variables are not directly linked with the sampling design. Therefore, the sampling scheme doesn't determine to which block an individual belongs. In order to evaluate the expectation with respect to the sampling design, blocks are treated as domains. Here it is ignored that the number of individuals n_j assigned to block j is random. Let a_{ij} denote the membership indicator of individual i in block j, defined by (3.49) in section 3.6.1. Let

$$
\bar{Y}_{a_{jk}}^\alpha = \frac{1}{N}\sum_{i=1}^{N} a_{ij}y_{ik}^\alpha,
$$

denote the population mean of the observations y_{ik}^{α} in block j. Then

$$\hat{\bar{Y}}_{a_{jk}}^{\alpha} = \frac{1}{N} \sum_{i=1}^{n} \frac{a_{ij} y_{ik}^{\alpha}}{\pi_i} \tag{6.87}$$

denotes the Horvitz-Thompson estimator of $\bar{Y}_{a_{jk}}^{\alpha}$.

The diagonal elements d_k are defined in (3.34) in section 3.6.1. First we take the expectation of d_k with respect to the sampling design.

$$\begin{aligned}
& \mathrm{E}_s \sum_{j=1}^{J} \frac{1}{(n_j-1)} \frac{1}{n_{jk}} \sum_{i=1}^{n_j} \left(\frac{n_j y_{ik}^{\alpha}}{\pi_i N} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{n_j y_{i'k}^{\alpha}}{\pi_{i'} N} \right)^2 \\
&= \mathrm{E}_s \sum_{j=1}^{J} \frac{1}{(n_j-1)} \frac{1}{n_{jk}} \sum_{i=1}^{n_j} \left(\frac{n_j y_{ik}^{\alpha}}{\pi_i N} - \bar{Y}_{a_{jk}}^{\alpha} + \bar{Y}_{a_{jk}}^{\alpha} - \hat{\bar{Y}}_{a_{jk}}^{\alpha} \right)^2 \\
&= \sum_{j=1}^{J} \frac{1}{(n_j-1)} \frac{1}{n_{jk}} \left(\mathrm{E}_s \sum_{i=1}^{n_j} \left(\frac{n_j y_{ik}^{\alpha}}{\pi_i N} - \bar{Y}_{a_{jk}}^{\alpha} \right)^2 - n_j \mathrm{E}_s \left(\hat{\bar{Y}}_{a_{jk}}^{\alpha} - \bar{Y}_{a_{jk}}^{\alpha} \right)^2 \right).
\end{aligned} \tag{6.88}$$

The second term of (6.88) can be recognized as the design variance of $\hat{\bar{Y}}_{a_{jk}}^{\alpha}$. Consequently we have

$$\begin{aligned}
\mathrm{E}_s \left(\hat{\bar{Y}}_{a_{jk}}^{\alpha} - \bar{Y}_{a_{jk}}^{\alpha} \right)^2 &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'=1}^{N} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{a_{ij} y_{ik}^{\alpha}}{\pi_i} \frac{a_{i'j} y_{i'k}^{\alpha}}{\pi_{i'}} \\
&= \left(\frac{N_j}{N} \right)^2 \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{y_{ik}^{\alpha} y_{i'k}^{\alpha}}{\pi_i \, \pi_{i'}} \\
&\equiv \left(\frac{N_j}{N} \right)^2 \mathrm{Var}(\hat{\bar{Y}}_{jk_s}^{\alpha}).
\end{aligned} \tag{6.89}$$

For the first term of (6.88) it follows that the expectation with respect to the sampling design is given by

$$\begin{aligned}
& \mathrm{E}_s \sum_{i=1}^{n_j} \left(\frac{n_j a_{ij} y_{ik}^{\alpha}}{\pi_i N} - \bar{Y}_{a_{jk}}^{\alpha} \right)^2 \\
&= \mathrm{E}_s \left(\sum_{i=1}^{n_j} \frac{n_j^2 a_{ij} y_{ik}^{\alpha 2}}{N^2 \pi_i^2} - 2 n_j \hat{\bar{Y}}_{a_{jk}}^{\alpha} \bar{Y}_{a_{jk}}^{\alpha} + n_j \bar{Y}_{a_{jk}}^{\alpha 2} \right) \\
&= n_j \left(\sum_{i=1}^{N} \frac{n_j a_{ij} y_{ik}^{\alpha 2}}{N^2 \pi_i} - \bar{Y}_{a_{jk}}^{\alpha 2} \right) \\
&= \left(\frac{N_j}{N} \right)^2 n_j^2 \left[\frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j y_{ik}^{\alpha 2}}{N_j^2 \pi_i} - \bar{Y}_{jk}^{\alpha 2} \right) \right] \equiv \left(\frac{N_j}{N} \right)^2 n_j^2 \tilde{\mathrm{Var}}(\hat{\bar{Y}}_{jk_s}^{\alpha}).
\end{aligned} \tag{6.90}$$

If results (6.89) and (6.90) are substituted into (6.88), then it follows that

$$\mathrm{E}_s d_k = \sum_{j=1}^{J} \left(\frac{N_j}{N} \right)^2 \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\mathrm{Var}}(\hat{\bar{Y}}_{jk_s}^{\alpha}) - \frac{1}{n_j} \mathrm{Var}(\hat{\bar{Y}}_{jk_s}^{\alpha}) \right). \tag{6.91}$$

Now we evaluate the expectation with respect to the measurement error model of $\mathrm{Var}(\hat{\bar{Y}}_{jk_s}^{\alpha})$ and $\tilde{\mathrm{Var}}(\hat{\bar{Y}}_{jk_s}^{\alpha})$ under a measurement error model with random interviewer effects. For $\mathrm{Var}(\hat{\bar{Y}}_{jk_s}^{\alpha})$ we obtain

$$\begin{aligned}
\mathrm{E}_\alpha \mathrm{Var}(\hat{\hat{Y}}^{\alpha}_{jk_{\mathbf{s}}}) &= \mathrm{E}_\alpha \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'}-\pi_i\pi_{i'})\frac{(u_i+\beta_k+\xi_j^\alpha+\varepsilon_{ik}^\alpha)}{\pi_i}\frac{(u_{i'}+\beta_k+\xi_j^\alpha+\varepsilon_{i'k}^\alpha)}{\pi_{i'}} \\
&= \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'}-\pi_i\pi_{i'})\frac{(u_i+\beta_k)}{\pi_i}\frac{(u_{i'}+\beta_k)}{\pi_{i'}} \\
&\quad +\frac{\tau_j^2}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\frac{(\pi_{ii'}-\pi_i\pi_{i'})}{\pi_i\pi_{i'}}+\frac{1}{N_j^2}\sum_{i=1}^{N_j}\frac{(\pi_i-\pi_i^2)}{\pi_i}\sigma_{ik}^2 \\
&= \mathrm{Var}(\hat{\hat{U}}_{jk_{\mathbf{s}}})+\frac{\tau_j^2}{N_j^2}\mathrm{Var}(\hat{N}_j)+\frac{1}{N_j^2}\sum_{i=1}^{N_j}\frac{(1-\pi_i)}{\pi_i}\sigma_{ik}^2, \qquad (6.92)
\end{aligned}$$

where $\mathrm{Var}(\hat{\hat{U}}_{jk_{\mathbf{s}}})$ is defined by (6.17), section 6.3 and $\mathrm{Var}(\hat{N}_j)$ by (6.45), section 6.6. Taking the expectation of $\tilde{\mathrm{Var}}(\hat{\hat{Y}}^{\alpha}_{jk_{\mathbf{s}}})$ with respect to the measurement error model gives

$$\begin{aligned}
\mathrm{E}_\alpha \tilde{\mathrm{Var}}(\hat{\hat{Y}}^{\alpha}_{jk_{\mathbf{s}}}) &= \mathrm{E}_\alpha \frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n_j(u_i+\beta_k+\xi_j^\alpha+\varepsilon_{ik}^\alpha)^2}{N_j^2\pi_i}-\left(\bar{U}_{jk}+\xi_j^\alpha+\frac{1}{N_j}\sum_{i=1}^{N_j}\varepsilon_{ik}^\alpha\right)^2\right) \\
&= \frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n_j(u_i+\beta_k)^2}{N_j^2\pi_i}-(\bar{U}_{jk})^2\right)+\frac{\tau_j^2}{N_j^2}\frac{1}{n_j}\left(\sum_{i=1}^{N_j}\frac{n_j}{\pi_i}-N_j^2\right) \\
&\quad +\frac{1}{n_j}\frac{1}{N_j^2}\sum_{i=1}^{N_j}\left(\frac{(n_j-\pi_i)}{\pi_i}\right)\sigma_{ik}^2 \\
&= \tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_{\mathbf{s}}})+\frac{\tau_j^2}{N_j^2}\tilde{\mathrm{Var}}(\hat{N}_j)+\frac{1}{n_j}\frac{1}{N_j^2}\sum_{i=1}^{N_j}\left(\frac{(n_j-\pi_i)}{\pi_i}\right)\sigma_{ik}^2, \qquad (6.93)
\end{aligned}$$

where $\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_{\mathbf{s}}})$ is defined by (6.18), section 6.3 and $\tilde{\mathrm{Var}}(\hat{N}_j)$ by (6.46), section 6.6.
If results (6.92) and (6.93) are substituted into (6.91), then it follows that

$$\begin{aligned}
\mathrm{E}_\alpha \mathrm{E}_s d_k &= \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2\frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\Big[\tilde{\mathrm{Var}}(\hat{\hat{U}}_{jk_{\mathbf{s}}})+\frac{\tau_j^2}{N_j^2}\tilde{\mathrm{Var}}(\hat{N}_j) \\
&\quad -\frac{1}{n_j}\left(\mathrm{Var}(\hat{\hat{U}}_{jk_{\mathbf{s}}})+\frac{\tau_j^2}{N_j^2}\mathrm{Var}(\hat{N}_j)\right)\Big]+\frac{1}{N^2}\sum_{j=1}^{J}\frac{n_j}{n_{jk}}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{\pi_i}, \qquad \textbf{QED.}
\end{aligned}$$

The proof of (6.48) follows in an equivalent way, if the expectation of $\mathrm{Var}(\hat{\hat{Y}}^{\alpha}_{jk_{\mathbf{s}}})$ and $\tilde{\mathrm{Var}}(\hat{\hat{Y}}^{\alpha}_{jk_{\mathbf{s}}})$ under a measurement error model with fixed interviewer effects is taken and the results are substituted into (6.91).

6.9.6 Proof of formula (6.47) and (6.53)

In this appendix we evaluate the expectation with respect to the measurement error model and the sampling design of d_{k_R} for RBD's where interviewers are block variables under a general complex sampling schemes analyzed with the generalized regression estimator. The diagonal elements of d_{k_R} are defined in (4.28), section 4.4. Equivalent to the derivation of (6.91) in appendix 6.9.5 it follows that

$$\begin{aligned} \mathrm{E}_s d_{k_R} &= \mathrm{E}_s \sum_{j=1}^{J} \frac{1}{(n_j-1)} \frac{1}{n_{jk}} \sum_{i=1}^{n_j} \left(\frac{n_j(y_{ik}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_i)}{N\pi_i} - \frac{1}{n_j} \sum_{i'=1}^{n_j} \frac{n_j(y_{i'k}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_{i'})}{N\pi_{i'}} \right)^2 \\ &= \sum_{j=1}^{J} \left(\frac{N_j}{N}\right)^2 \frac{n_j}{(n_j-1)} \frac{n_j}{n_{jk}} \left(\tilde{\mathrm{Var}}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha}) - \frac{1}{n_j} \mathrm{Var}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha}) \right), \end{aligned} \tag{6.94}$$

where

$$\mathrm{Var}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha}) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{(y_{ik}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_i)}{\pi_i} \frac{(y_{i'k}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_{i'})}{\pi_{i'}},$$

and

$$\tilde{\mathrm{Var}}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha}) = \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j (y_{ik}^{\alpha} - \mathbf{b}_k^t \mathbf{x}_i)^2}{N_j^2 \pi_i} - \left(\bar{Y}_{jk}^{\alpha} - \mathbf{b}_k^t \bar{\mathbf{X}}_j \right)^2 \right).$$

Now we evaluate the expectation with respect to the measurement error model of $\mathrm{Var}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha})$ and $\tilde{\mathrm{Var}}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha})$ for a measurement error model with fixed interviewer effects. If (6.84) in appendix 6.9.3 is substituted into the expression of $\mathrm{Var}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha})$, then it follows that

$$\begin{aligned} &\mathrm{E}_{\alpha} \mathrm{Var}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha}) \\ &= \mathrm{E}_{\alpha} \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'} - \pi_i \pi_{i'}) \frac{\left[(u_i - \mathbf{b}^t \mathbf{x}_i) + (\psi_j - \mathbf{d}^t \mathbf{x}_i) + \varepsilon_{ik}^{\alpha}\right]}{\pi_i} \frac{\left[(u_{i'} - \mathbf{b}^t \mathbf{x}_{i'}) + (\psi_j - \mathbf{d}^t \mathbf{x}_{i'}) + \varepsilon_{i'k}^{\alpha}\right]}{\pi_{i'}} \\ &= \mathrm{Var}(\hat{\hat{E}}_{j\mathbf{s}}) + \mathrm{Var}(\hat{\psi}_{jR}) + 2\mathrm{Cov}(\hat{\hat{E}}_{j\mathbf{s}}, \hat{\psi}_{jR}) + \frac{1}{N_j^2} \sum_{i=1}^{N_j} \frac{(\pi_i - \pi_i^2)}{\pi_i^2}, \end{aligned} \tag{6.95}$$

where $\mathrm{Var}(\hat{\hat{E}}_{j\mathbf{s}})$ is defined by (6.22) in section 6.3, $\mathrm{Var}(\hat{\psi}_{jR})$ by (6.54) in section 6.6 and $\mathrm{Cov}(\hat{\hat{E}}_{j\mathbf{s}}, \hat{\psi}_{jR})$ by (6.56) in section 6.6. If (6.84) in appendix 6.9.3 is substituted into the expression of $\tilde{\mathrm{Var}}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha})$, then it follows that

$$\begin{aligned} &\mathrm{E}_{\alpha} \tilde{\mathrm{Var}}(\hat{\hat{E}}_{jk_{\mathbf{s}}}^{\alpha}) \\ &= \mathrm{E}_{\alpha} \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j \left((u_i - \mathbf{b}^t \mathbf{x}_i) + (\psi_j - \mathbf{d}^t \mathbf{x}_i) + \varepsilon_{ik}^{\alpha}\right)^2}{N_j^2 \pi_i} - \left((\bar{U}_j - \mathbf{b}^t \bar{\mathbf{X}}_j) + (\psi_j - \mathbf{d}^t \bar{\mathbf{X}}_j) + \frac{1}{N_j} \sum_{i=1}^{N_j} \varepsilon_{ik}^{\alpha} \right)^2 \right) \\ &= \frac{1}{n_j} \left(\sum_{i=1}^{N_j} \frac{n_j \left((u_i - \mathbf{b}^t \mathbf{x}_i) + (\psi_j - \mathbf{d}^t \mathbf{x}_i)\right)^2 + \sigma_{ik}^2}{N_j^2 \pi_i} - \left((\bar{U}_j - \mathbf{b}^t \bar{\mathbf{X}}_j) + (\psi_j - \mathbf{d}^t \bar{\mathbf{X}}_j) \right)^2 - \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sigma_{ik}^2 \right) \\ &= \tilde{\mathrm{Var}}(\hat{\hat{E}}_{j\mathbf{s}}) + \tilde{\mathrm{Var}}(\hat{\psi}_{jR}) + 2\tilde{\mathrm{Cov}}(\hat{\hat{E}}_{j\mathbf{s}}, \hat{\psi}_{jR}) + \frac{1}{n_j} \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sigma_{ik}^2 \left(\frac{(n_j - \pi_i)}{\pi_i} \right), \end{aligned} \tag{6.96}$$

where $\tilde{\mathrm{Var}}(\hat{\hat{E}}_{j\mathbf{s}})$ is defined by (6.23) in section 6.3, $\tilde{\mathrm{Var}}(\hat{\psi}_{jR})$ by (6.55) in section 6.6 and $\mathrm{Cov}(\hat{\hat{E}}_{j\mathbf{s}}, \hat{\psi}_{jR})$ by (6.57) in section 6.6.

If results (6.95) and (6.96) are substituted into results (6.94), then it follows that

$$\begin{aligned}\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} &= \sum_{j=1}^{J}\left(\frac{N_j}{N}\right)^2 \frac{n_j}{(n_j-1)}\frac{n_j}{n_{jk}}\Big[\tilde{\mathrm{Var}}(\hat{\tilde{E}}_{j_\mathbf{s}}) + \tilde{\mathrm{Var}}(\hat{\psi}_{jR}) + 2\tilde{\mathrm{Cov}}(\hat{\tilde{E}}_{j_\mathbf{s}}, \hat{\psi}_{jR}) \\ &-\frac{1}{n_j}\Big(\mathrm{Var}(\hat{\tilde{E}}_{j_\mathbf{s}}) + \mathrm{Var}(\hat{\psi}_{jR}) + 2\mathrm{Cov}(\hat{\tilde{E}}_{j_\mathbf{s}}, \hat{\psi}_{jR})\Big)\Big] \\ &+\frac{1}{N^2}\sum_{j=1}^{J}\frac{n_j}{n_{jk}}\sum_{i=1}^{N_j}\frac{\sigma_{ik}^2}{\pi_i}, \quad \textbf{QED}.\end{aligned}$$

The proof of (6.47) follows in an equivalent way, if the expectation of $\mathrm{Var}(\hat{\tilde{E}}^{\alpha}_{jk_\mathbf{s}})$ and $\tilde{\mathrm{Var}}(\hat{\tilde{E}}^{\alpha}_{jk_\mathbf{s}})$ under a measurement error model with random interviewer effects is taken and the results are substituted into (6.94).

Chapter 7

Embedded experiments with different randomization levels for experimental units and sampling units

7.1 Introduction

In the experiments considered in the preceding chapters the sampling units of the sampling design are randomized over K different treatments, according to the experimental design. This implies that the sampling units of the sampling design coincide with the experimental units of the experimental design. In many practical situations, however, this is impossible. Consider an experiment, embedded in a two-stage sampling design. Due to the nature of the treatments in the experiment, it might be necessary to use the PSU's as experimental units in the experiment, while the sampling units are the SSU's. For example if PSU's are households, SSU's are persons and the treatments in the experiment concern different approach strategies to optimize response rates. In such situations it might, from a practical point of view, be infeasible to apply different treatments within the same household. Consequently, households are randomized over the different treatments. In this chapter the design-based approach for the analysis of embedded experiments developed in the preceding chapters is extended to situations where the experimental units correspond with the PSU's of a two-stage sampling design.

Consider an experiment embedded in a two-stage sampling design conducted to test the effect of K different survey methodologies or treatments on the estimates of the finite population parameters obtained from a sample survey. The aim of the experiment is to test the hypothesis of no treatment effects:

$$H_0 : \quad \mathbf{C}\mathrm{E}_\alpha \bar{\mathbf{Y}}^\alpha = \mathbf{0},$$

$$H_1: \quad \mathbf{C}\mathrm{E}_\alpha \bar{\mathbf{Y}}^\alpha \neq \mathbf{0}, \tag{7.1}$$

with $\mathbf{C}$ defined by (3.16) in section 3.3. This hypothesis can be tested by means of the Wald statistic:

$$W = \hat{\hat{\mathbf{Y}}}^{\alpha^t} \mathbf{C}^t \left(\widehat{\mathbf{C}\mathbf{V}\mathbf{C}^t}\right)^{-1} \mathbf{C}\hat{\hat{\mathbf{Y}}}^\alpha.$$

A design-based estimator for this Wald-statistic will be derived using the Horvitz-Thompson estimator and the generalized regression estimator for CRD's as well as RBD's with different randomization levels for the experimental units and the sampling units.

7.2 Measurement error models

The measurement error models introduced in section 3.2 and 4.2 are adjusted for the experiments considered in this chapter. Let y^α_{ijk} denote the observation of individual i from PSU j observed by means of treatment k, measured on the $\alpha-$th occasion, u_{ij} the intrinsic value of individual i from PSU j and β_k the $k-$th treatment effect. Furthermore, let ε^α_{ijk} the error term obtained on the $\alpha-$th occasion that individuals i true value u_{ij} is measured by means of the $k-$th treatment. If the Horvitz-Thompson estimator is applied in the analysis, then the observations y^α_{ijk} are modeled with the basic measurement error model for K survey strategies

$$\mathbf{y}^\alpha_{ij} = \mathbf{j}u_{ij} + \beta + \varepsilon^\alpha_{ij}, \tag{7.2}$$

with model assumptions

$$\mathrm{E}_\alpha(\varepsilon^\alpha_{ij}) = \mathbf{0}, \tag{7.3}$$

$$\mathrm{Cov}_\alpha(\varepsilon^\alpha_{ij}, \varepsilon^{\alpha^t}_{i'j'}) = \begin{cases} \mathbf{\Sigma}_{ij} & : \quad i = i' \\ \mathbf{O} & : \quad i \neq i' \end{cases}. \tag{7.4}$$

Here $\mathbf{y}^\alpha_{ij} = (y^\alpha_{ij1}, \ldots, y^\alpha_{ijK})^t$, $\beta = (\beta_1, \ldots, \beta_K)^t$ and $\varepsilon^\alpha_{ij} = (\varepsilon^\alpha_{ij1}, \ldots, \varepsilon^\alpha_{ijK})^t$. Extensions to measurement error models with random or fixed interviewer effects follows directly from (3.4) given in section 3.2.

If the experiment is analyzed with the generalized regression estimator, then the intrinsic values u_{ij} are modeled with a linear regression model. Let $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijH})^t$ denote a vector of order H with each element x_{ijh} an auxiliary variable of individual i from PSU j. According to the approach followed in section 4.2, it is assumed that each intrinsic value u_{ij} of element i from PSU j is a realization of the linear regression model

$$u_{ij} = \mathbf{b}^t \mathbf{x}_{ij} + e_{ij}, \tag{7.5}$$

where $\mathbf{b}$ is a $H-$vector with regression coefficients and e_{ij} the residuals of the linear regression model. The following model assumptions are adopted:

$$\mathrm{E}_\xi(e_{ij}) = 0, \tag{7.6}$$

$$\mathrm{Cov}_\xi(e_{ij}, e_{i'j'}) = \begin{cases} \omega^2_{ij} & : \quad i = i' \\ 0 & : \quad i \neq i' \end{cases}, \tag{7.7}$$

where ω_{ij} denotes the variance of the residuals e_{ij}. The basic measurement error model (7.2) can be extended by inserting (7.5) into (7.2):

$$\mathbf{y}_{ij}^{\alpha} = \mathbf{j}\left(\mathbf{b}^t\mathbf{x}_{ij} + e_{ij}\right) + \beta + \varepsilon_{ij}^{\alpha}, \tag{7.8}$$

with model assumptions (7.3), (7.4), (7.6) and(7.7). Extensions to models with random or fixed interviewer effects follows directly from (4.8) in section 4.2.

7.3 Estimation of treatment effects

Consider a finite population U that consists of M PSU's. Each PSU j consists of N_j SSU's. Let $N = \sum_{j=1}^{M} N_j$ denote the size of the population. To test the hypothesis of no treatment effects (7.1) a two-stage sample s of size n, drawn from the finite population, is available. In the first stage m PSU's are drawn with first and second-order inclusion probabilities π_j and $\pi_{jj'}$. In the second stage a sample of n_j SSU's is drawn from each PSU selected in the first stage with first and second-order inclusion probabilities $\pi_{i|j}$ and $\pi_{ii'|j}$. The first and second-order inclusion probabilities of the SSU's in a two-stage sample design are defined by

$$\begin{aligned}\pi_i &= \pi_j\pi_{i|j}, \\ \pi_{ii'} &= \begin{cases} \pi_j\pi_{ii'|j} & \text{if } i \in j \text{ and } i' \in j \\ \pi_{jj'}\pi_{i|j}\pi_{i'|j'} & \text{if } i \in j \text{ and } i' \in j' \end{cases} .\end{aligned}$$

According to the experimental design, the m PSU's are randomly divided into K subsamples of size m_k. The number of SSU's in each subsample equals $n_k = \sum_{j=1}^{m_k} n_j$. In the experimental designs considered in the preceding chapters, j is used as index for the block variables. Since sampling structures like PSU's were potential block variables under that type of experimental designs, it was convenient to use the same index for block variables and PSU's (or other sampling structures). In the experimental designs considered in this chapter, PSU's cannot correspond with block variables since we randomize over the PSU's. Since j is used as index for the PSU's, we change to b as the index for the block variables. Potential block variables are strata (in the case of stratified two-stage sampling) or interviewers (in the case of a measurement error model with interviewer effects).

The randomization mechanism of the experimental design can be described with conditional selection probabilities. In the case of a CRD, the m PSU's are randomly divided into K subsamples of size m_k. Therefore, the conditional probability that the $j-$th PSU is selected in subsample s_k equals m_k/m. In the case of an RBD, the experimental units are deterministically divided in B blocks s_b. Let m_b denote the number of experimental units in block b and m_{bk} the number of experimental units in block b assigned to treatment k. In an RBD, m_{bk} of the m_b PSU's or experimental units within each block are randomly assigned to treatment k. The conditional probability that the $j-$th PSU is selected in subsample s_k, given that sample s is selected and that the $j-$th PSU is in block b, equals m_k/m_{bk}.

For the estimation of the population parameters $\mathrm{E}_\alpha \bar{Y}_k^\alpha$ the first-order inclusion probabilities of the n_k sampling units in subsample s_k must be known. Equivalent to the derivation of (3.19) in section 3.5 it follows for a CRD that the first-order inclusion probabilities for the elements of subsample s_k are equal to $\pi_i^* = \frac{m_k}{m}\pi_i$. Equivalent to the derivation of (3.20) it follows for an RBD that the first-order inclusion probabilities of the sampling units in subsample s_k are equal to $\pi_i^* = \frac{m_{bk}}{m_b}\pi_i$.

The Horvitz-Thompson estimator for $\mathrm{E}_\alpha \bar{Y}_k^\alpha$, based on subsample s_k for a CRD is given by

$$\hat{\bar{Y}}_k^\alpha = \frac{m}{Nm_k}\sum_{j=1}^{m_k}\sum_{i=1}^{n_j}\frac{y_{ijk}^\alpha}{\pi_i} \equiv \frac{m}{Nm_k}\sum_{j=1}^{m_k}\frac{\hat{y}_{jk}^\alpha}{\pi_j}, \tag{7.9}$$

where

$$\hat{y}_{jk}^\alpha = \sum_{i=1}^{n_j}\frac{y_{ijk}^\alpha}{\pi_{i|j}} \tag{7.10}$$

denotes the Horvitz-Thompson estimator for the total of the $j-$th PSU. For an RBD, the Horvitz-Thompson estimator for $\mathrm{E}_\alpha \bar{Y}_k^\alpha$ is given by

$$\hat{\bar{Y}}_k^\alpha = \frac{1}{N}\sum_{b=1}^{B}\sum_{j=1}^{m_{bk}}\sum_{i=1}^{n_j}\frac{m_b}{m_{bk}}\frac{y_{ijk}^\alpha}{\pi_i} \equiv \frac{1}{N}\sum_{b=1}^{B}\sum_{j=1}^{m_{bk}}\frac{m_b}{m_{bk}}\frac{\hat{y}_{jk}^\alpha}{\pi_j}. \tag{7.11}$$

Using (7.14) and (7.15), the Horvitz-Thompson estimators for $\mathrm{E}_\alpha \bar{Y}_k^\alpha$ for a CRD and an RBD can be defined more generally as

$$\hat{\bar{Y}}_k^\alpha = \frac{1}{N}\sum_{j=1}^{m}\frac{\mathbf{p}_{jk}^t\hat{\mathbf{y}}_j^\alpha}{\pi_j}, \tag{7.12}$$

where

$$\hat{\mathbf{y}}_j^\alpha = \sum_{i=1}^{n_j}\frac{\mathbf{y}_{ij}^\alpha}{\pi_{i|j}}, \tag{7.13}$$

and $\mathbf{p}_{jk}$ are $K-$vectors that describe the randomization mechanism of the experimental design. For a CRD it follows from the identity between (7.9) and (7.12) that

$$\mathbf{p}_{jk} \equiv \begin{cases} \frac{m}{m_k}\mathbf{e}_k & \text{if} \quad j \in s_k \\ \mathbf{0} & \text{if} \quad j \notin s_k \end{cases}, \tag{7.14}$$

where $\mathbf{e}_k$ denotes the unit vector of order K with the $k-$th element equal to one and the other elements equal to zero. As a consequence of the randomization mechanism of a CRD, the vectors $\mathbf{p}_{jk}$ are random with the following conditional probability mass function

$$P\left(\mathbf{p}_{jk} = \frac{m}{m_k}\mathbf{e}_k \mid s\right) = \frac{m_k}{m} \quad \text{and} \quad P\left(\mathbf{p}_{jk} = \mathbf{0} \mid s\right) = 1 - \frac{m_k}{m}.$$

For an RBD it follows from the identity between (7.11) and (7.12) that

$$\mathbf{p}_{jk} \equiv \begin{cases} \frac{m_b}{m_{bk}}\mathbf{e}_k & \text{if} \quad j \in s_{bk} \\ \mathbf{0} & \text{if} \quad j \notin s_{bk} \end{cases}, \tag{7.15}$$

where s_{bk} denotes the subsample of m_{bk} PSU's in block b that are assigned to treatment k. As a consequence of the randomization mechanism of an RBD, the vectors $\mathbf{p}_{jk}$ are random with the following conditional probability mass function

$$P\left(\mathbf{p}_{jk}=\frac{m_b}{m_{bk}}\mathbf{e}_k \mid s_j\right)=\frac{m_{bk}}{m_b} \quad \text{and} \quad P\left(\mathbf{p}_{jk}=\mathbf{0} \mid s_j\right)=1-\frac{m_{bk}}{m_b}.$$

Properties of the randomization vectors $\mathbf{p}_{jk}$ follow directly from appendix 3.9.1 and 3.9.2.

The vector $\hat{\hat{\mathbf{Y}}}^{\alpha}_{\mathbf{s}_k}=(\hat{\hat{Y}}^{\alpha}_1,\hat{\hat{Y}}^{\alpha}_2,\dots,\hat{\hat{Y}}^{\alpha}_K)^t$ is proposed as a design-unbiased estimator for $\mathrm{E}_\alpha\bar{\mathbf{Y}}^\alpha=(\mathrm{E}_\alpha\bar{Y}^\alpha_1,\mathrm{E}_\alpha\bar{Y}^\alpha_2,\dots,\mathrm{E}_\alpha\bar{Y}^\alpha_K)^t$.

The generalized regression estimator for $\mathrm{E}_\alpha\bar{\mathbf{Y}}^\alpha$ is derived as follows. The best linear unbiased estimator for the regression coefficients $\mathbf{b}$ in (7.5) based on a complete enumeration of the finite population is given by

$$\mathbf{b}=\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\mathbf{x}_{ij}\mathbf{x}_{ij}^t}{\omega_{ij}^2}\right)^{-1}\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\mathbf{x}_{ij}u_{ij}}{\omega_{ij}^2}\right). \tag{7.16}$$

The finite population regression coefficients $\mathbf{b}$ observed under the $k-$th treatment on the $\alpha-$th occasion is defined as

$$\mathbf{b}_k^\alpha=\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\mathbf{x}_{ij}\mathbf{x}_{ij}^t}{\omega_{ij}^2}\right)^{-1}\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\mathbf{x}_{ij}y_{ijk}^\alpha}{\omega_{ij}^2}\right), \quad k=1,\dots,K. \tag{7.17}$$

The expectation of $\mathbf{b}_k^\alpha$ with respect to the measurement error model equals

$$\mathbf{b}_k=\mathrm{E}_\alpha\mathbf{b}_k^\alpha=\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\mathbf{x}_{ij}\mathbf{x}_{ij}^t}{\omega_{ij}^2}\right)^{-1}\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\mathbf{x}_{ij}(u_{ij})+\beta_k)}{\omega_{ij}^2}\right), \quad k=1,\dots,K. \tag{7.18}$$

Let $\bar{\mathbf{X}}$ denote a vector of order H with each element the known population totals of the auxiliary variables. The Horvitz-Thompson estimator for $\bar{\mathbf{X}}$ in the case of a CRD is given by

$$\hat{\hat{\mathbf{X}}}_{\mathbf{s}_k}=\frac{m}{Nm_k}\sum_{j=1}^{m_k}\sum_{i=1}^{n_j}\frac{1}{\pi_i}\mathbf{x}_{ij}\equiv\frac{m}{Nm_k}\sum_{j=1}^{m_k}\frac{\hat{\mathbf{x}}_j}{\pi_j},$$

where

$$\hat{\mathbf{x}}_j=\sum_{i=1}^{n_j}\frac{\mathbf{x}_{ij}}{\pi_{i|j}} \tag{7.19}$$

denotes the Horvitz-Thompson estimator of the totals of the auxiliary variables of the $j-$th PSU. The Horvitz-Thompson estimator for $\bar{\mathbf{X}}$ in the case of an RBD is given by

$$\hat{\hat{\mathbf{X}}}_{\mathbf{s}_k}=\frac{1}{N}\sum_{b=1}^{B}\sum_{j=1}^{m_{bk}}\sum_{i=1}^{n_j}\frac{m_b}{m_{bk}}\frac{1}{\pi_i}\mathbf{x}_{ij}\equiv\frac{1}{N}\sum_{b=1}^{B}\sum_{j=1}^{m_{bk}}\frac{m_b}{m_{bk}}\frac{\hat{\mathbf{x}}_j}{\pi_j}.$$

The generalized regression estimator for $\mathrm{E}_\alpha\bar{Y}_k^\alpha$ based on the observations obtained in subsample s_k for a CRD as well as an RBD is given by

$$\hat{\hat{Y}}^{\alpha}_{k_R}=\hat{\hat{Y}}^{\alpha}_k+\hat{\mathbf{b}}_k^{\alpha^t}(\bar{\mathbf{X}}-\hat{\hat{\mathbf{X}}}_{\mathbf{s}_k}), \quad k=1,2,\dots,K, \tag{7.20}$$

where $\hat{\hat{Y}}_k^{\alpha}$ is the Horvitz-Thompson estimator defined in (7.12). The Horvitz-Thompson estimator for $\mathbf{b}_k^{\alpha}$, based on the observations obtained from subsample s_k, is given by

$$\hat{\mathbf{b}}_k^{\alpha} = \left(\sum_{j=1}^{m_k}\sum_{i=1}^{n_j}\frac{\mathbf{x}_{ij}\mathbf{x}_{ij}^t}{\omega_{ij}^2\pi_i^*}\right)^{-1}\left(\sum_{j=1}^{m_k}\sum_{i=1}^{n_j}\frac{\mathbf{x}_{ij}y_{ijk}^{\alpha}}{\omega_{ij}^2\pi_i^*}\right), \quad k = 1, \ldots, K. \tag{7.21}$$

For the derivation of the covariance matrix of the contrasts between the subsample estimates, it is convenient to express the generalized regression estimator (7.20) in terms of randomization vectors $\mathbf{p}_{jk}$ as follows:

$$\hat{\hat{Y}}_{k_R}^{\alpha} = \sum_{j\in s}\left(\frac{\mathbf{p}_{jk}^t(\hat{\mathbf{y}}_j^{\alpha} - \hat{\mathbf{B}}^{\alpha^t}\hat{\mathbf{x}}_j)}{\pi_j\ N}\right) + \hat{\mathbf{b}}_k^{\alpha^t}\bar{\mathbf{X}}, \tag{7.22}$$

where $\hat{\mathbf{y}}_j^{\alpha}$ and $\hat{\mathbf{x}}_j$ are defined by respectively (7.13) and (7.19) and $\hat{\mathbf{B}}^{\alpha}$ denotes a matrix of order $H \times K$ whose columns are the H dimensional vectors of $\hat{\mathbf{b}}_k^{\alpha}$. The vector $\hat{\hat{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha} = (\hat{\hat{Y}}_{1_R}^{\alpha}, \hat{\hat{Y}}_{2_R}^{\alpha}, \ldots, \hat{\hat{Y}}_{K_R}^{\alpha})^t$ denotes the generalized regression estimator for $\mathrm{E}_{\alpha}\bar{\mathbf{Y}}^{\alpha}$.

7.4 Variance estimation of treatment effects

In this section (approximately) design-unbiased estimators for the covariance matrix of the $K-1$ contrasts of the Horvitz-Thompson estimator $\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ and the generalized regression estimator $\hat{\hat{\mathbf{Y}}}_{R\mathbf{s}_k}^{\alpha}$ are given.

For the Horvitz-Thompson estimator the covariance matrix of the $K-1$ contrasts $\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ is given by $\mathbf{CVC}^t = \mathrm{E}_{\alpha}\mathrm{E}_s\mathbf{CDC}^t$, where $\mathbf{D}$ denotes the $K \times K$ diagonal matrix with the diagonal elements d_k. Under the alternative hypothesis a generally small component, which concerns the design variance of estimated population totals is ignored. See section 3.6 for more details. For a CRD it follows that

$$d_k = \frac{m}{(m-1)}\frac{m}{m_k}\sum_{j=1}^{m}\left(\frac{\hat{y}_{jk}^{\alpha}}{N\pi_j} - \frac{1}{m}\sum_{j'=1}^{m}\frac{\hat{y}_{j'k}^{\alpha}}{N\pi_{j'}}\right)^2. \tag{7.23}$$

For an RBD it follows that

$$d_k = \sum_{b=1}^{B}\frac{m_b}{(m_b-1)}\frac{m_b}{m_{bk}}\sum_{j=1}^{m_b}\left(\frac{\hat{y}_{jk}^{\alpha}}{N\pi_j} - \frac{1}{m_b}\sum_{j'=1}^{m_b}\frac{\hat{y}_{j'k}^{\alpha}}{N\pi_{j'}}\right)^2. \tag{7.24}$$

where $\hat{y}_{jk}^{\alpha}$ is defined by (7.10).

An unbiased estimator for the covariance matrix of $\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{s}_k}^{\alpha}$ is given by $\widehat{\mathbf{CVC}}^t = \mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t$ with $\hat{\mathbf{D}} = \mathbf{diag}(\hat{d}_1 \ldots \hat{d}_K)$. For a CRD it follows that

$$\hat{d}_k = \frac{1}{m_k}\frac{1}{(m_k-1)}\sum_{j=1}^{m_k}\left(\frac{m\hat{y}_{jk}^{\alpha}}{N\pi_j} - \frac{1}{m_k}\sum_{j'=1}^{m_k}\frac{m\hat{y}_{j'k}^{\alpha}}{N\pi_{j'}}\right)^2. \tag{7.25}$$

For an RBD it follows that

$$\hat{d}_k = \sum_{b=1}^{B}\frac{1}{m_{bk}}\frac{1}{(m_{bk}-1)}\sum_{j=1}^{m_{bk}}\left(\frac{m_b\hat{y}_{jk}^{\alpha}}{N\pi_j} - \frac{1}{m_{bk}}\sum_{j'=1}^{m_{bk}}\frac{m_b\hat{y}_{j'k}^{\alpha}}{N\pi_{j'}}\right)^2. \tag{7.26}$$

For the generalized regression estimator, the covariance matrix of the $K-1$ contrasts the first-order Taylor series approximation of $\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$ is given by $\mathbf{C}\mathbf{V}_R\mathbf{C}^t = \mathrm{E}_\alpha \mathrm{E}_s \mathbf{C}\mathbf{D}_R\mathbf{C}^t$, where $\mathbf{D}_R$ denotes the $K \times K$ diagonal matrix with the diagonal elements d_{k_R}. Expression for d_{k_R} for a CRD and an RBD are given by (7.23) and (7.24), respectively, where $\hat{y}_j^\alpha$ is replaced by

$$\hat{y}_{jk}^\alpha - \mathbf{b}_k^t \hat{\mathbf{x}}_j = \sum_{i=1}^{n_j} \frac{(y_{ijk}^\alpha - \mathbf{b}_k^t \mathbf{x}_{ij})}{\pi_{i|j}}. \tag{7.27}$$

An approximately unbiased estimator for the covariance matrix of $\mathbf{C}\hat{\hat{\mathbf{Y}}}^{\alpha}_{R\mathbf{s}_k}$ is given by $\mathbf{C}\widehat{\mathbf{V}_R\mathbf{C}}^t = \mathbf{C}\hat{\mathbf{D}}_R\mathbf{C}^t$ with $\hat{\mathbf{D}}_R = \mathbf{diag}(\hat{d}_{1_R} \ldots \hat{d}_{K_R})$. An expression of $\hat{d}_{k_R}$ for a CRD and an RBD is given by (7.25) and (7.26), respectively, where $\hat{y}_{jk}^\alpha$ is replaced by

$$\hat{y}_{jk}^\alpha - \hat{\mathbf{b}}_k^{\alpha t} \hat{\mathbf{x}}_j = \sum_{i=1}^{n_j} \frac{(y_{ijk}^\alpha - \hat{\mathbf{b}}_k^{\alpha t} \mathbf{x}_{ij})}{\pi_{i|j}}. \tag{7.28}$$

An outline of the proof of these results is given in appendix 7.7.1.

A special case of the generalized regression estimator is the extended Horvitz-Thompson estimator. The extended Horvitz-Thompson estimator $\tilde{y}_{s_k}^\alpha$ is defined by (4.33) in section 4.4. An expression for the variance estimator is obtained by (7.25) for a CRD and (7.26) for an RBD where $\hat{y}_{jk}^\alpha$ is replaced by

$$\hat{y}_{jk}^\alpha - \tilde{y}_{s_k}^\alpha \hat{n}_j = \sum_{i=1}^{n_j} \frac{(y_{ijk}^\alpha - \tilde{y}_{s_k}^\alpha)}{\pi_{i|j}}.$$

Generally it will be preferable to apply the extended Horvitz-Thompson estimator instead of the Horvitz-Thompson estimator since the variance estimator of the extended Horvitz-Thompson estimator is approximately design-unbiased under the null as well as the alternative hypothesis. The Horvitz-Thompson estimator overestimates the variance under the alternative hypotheses if $N \neq \sum_{i \in s_k}(1/\pi_i^*)$.

7.5 Hypotheses testing

Hypotheses of no treatment effects can be tested with the design-based Wald statistic defined by (3.80) in section 3.7 for the Horvitz-Thompson estimator or (4.35) in section 4.5 for the generalized regression estimator. For the two-sample embedded experiment, the design-based $t-$statistic proposed in sections 5.4 or 6.8 can be used. Estimators for the population parameters and the variance of the contrasts of these population parameters are derived in the preceding sections. Critical regions or $p-$values are calculated by using the chi-squared distribution with $K-1$ degrees of freedom for the Wald-statistic and the standard normal distribution for the $t-$statistic.

7.6 Variance of the estimated treatment effects

In this section expressions for $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ are derived for a CRD under the basic measurement error model. This enables us to compare the efficiency between an experiment designed as an RBD with PSU's as block variables (see section 6.2) and an experiment designed as a CRD where the PSU's are the experimental units. We don't further evaluate the variance of $\mathrm{E}_\alpha \mathrm{E}_s d_k$ and $\mathrm{E}_\alpha \mathrm{E}_s d_{k_R}$ for RBD's where the PSU's are the experimental units and strata or interviewers are block variables. This results in principally the same conclusions concerning the efficiency of blocking on sampling structures like strata or interviewers as obtained in sections 6.3, 6.5 and 6.6.

Let u_j denote the population total of the intrinsic values u_{ij} of the $j-$th PSU. Then $u_{jk} = u_j + N_j \beta_k$ denotes the population total of the intrinsic values of the $j-$th PSU biased with treatment effect β_k. Let $\bar{U}_k = \bar{U} + \beta_k$ denote the population mean of the intrinsic values in the target population, biased with treatment effect β_k. The Horvitz-Thompson estimator of the population mean $\bar{U}_k$ based on the m PSU totals observed in sample s is given by

$$\hat{\bar{U}}_{k_{\mathbf{s_I}}} = \frac{1}{N} \sum_{j=1}^{m} \frac{u_{jk}}{\pi_j}. \tag{7.29}$$

Note that (7.29) is not observable in the experiment since only m_k out of the m PSU's in sample s are assigned to treatment k. Let

$$\mathrm{Var}(\hat{\bar{U}}_{k_{\mathbf{s_I}}}) = \frac{1}{N^2} \sum_{j=1}^{M} \sum_{j'=1}^{M} \left(\pi_{jj'} - \pi_j \pi_{j'}\right) \frac{u_{jk} u_{j'k}}{\pi_j \pi_{j'}} \tag{7.30}$$

denote the design variance of $\hat{\bar{U}}_{k_{\mathbf{s_I}}}$, i.e. the variance with respect to the sampling design used to draw the PSU's in the first stage and let

$$\tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_{\mathbf{s_I}}}) = \frac{1}{m} \left(\sum_{j=1}^{M} \frac{m u_{jk}^2}{\pi_j N^2} - \bar{U}_k^2 \right) \tag{7.31}$$

denotes the variance of $\hat{\bar{U}}_{k_{\mathbf{s_I}}}$ as if the m PSU's in the first stage were drawn with replacement with selection probabilities π_j/m. Let $\bar{u}_{jk} = \bar{u}_j + \beta_k$ denotes the mean of the intrinsic values $u_{ijk} = u_{ij} + \beta_k$ of the $j-$th PSU. Then

$$\hat{\bar{u}}_{jk} = \frac{1}{N_j} \sum_{i=1}^{n_j} \frac{u_{ijk}}{\pi_{i|j}} \tag{7.32}$$

denotes the Horvitz-Thompson estimator of $\bar{u}_{jk}$ and

$$\mathrm{Var}(\hat{\bar{u}}_{jk}) = \frac{1}{N_j^2} \sum_{i=1}^{N_j} \sum_{i'=1}^{N_j} (\pi_{ii'|j} - \pi_{i|j} \pi_{i'|j}) \frac{u_{ijk} u_{i'jk}}{\pi_{i|j} \pi_{i'|j}} \tag{7.33}$$

denotes the design variance of $\hat{\bar{u}}_{jk}$, i.e. the variance with respect to the sampling design used to draw the SSU's from the $j-$th PSU in the second stage, conditional on the realization of the first stage. Note that $\hat{\bar{u}}_{jk}$ defined by (7.32) is equally defined as $\hat{\bar{U}}_{jk_{\mathbf{s}}}$ in (6.3) section 6.2. Also note that (7.33) is equivalent to $\text{Var}(\hat{\hat{U}}_{jk_{\mathbf{s}}})$ defined in (6.4), section 6.2. In a CRD where the PSU's are the experimental units, all the SSU's of PSU j are assigned to one of the K treatments. Consequently, the Horvitz-Thompson estimator (7.32) is observable. This in contrast with $\hat{\hat{U}}_{jk_{\mathbf{s}}}$ for an experimental design where the SSU's are randomized over the K treatments within each PSU.

For a CRD conducted under the assumption of a basic measurement error model (7.2) it is proved in the appendix 7.7.2 that:

$$\begin{aligned}
\text{E}_{\alpha}\text{E}_{s}d_{k} &= \frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\text{Var}}(\hat{\hat{U}}_{k_{\mathbf{s_I}}}) - \frac{1}{m}\text{Var}(\hat{\hat{U}}_{k_{\mathbf{s_I}}})\right) \\
&\quad + \frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\text{Var}(\hat{\bar{u}}_{jk}) + \frac{m}{m_k}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\sigma^2_{ijk}}{\pi_i N^2}.
\end{aligned} \tag{7.34}$$

Similar expressions can be derived for the generalized regression estimator. Let $\hat{\hat{U}}_{R_{\mathbf{s_I}}}$ denote the generalized regression estimator of the population mean $\bar{U}$ based on the m PSU totals in sample s. The first-order Taylor series approximation of $\hat{\hat{U}}_{R_{\mathbf{s_I}}}$ is given by

$$\hat{\hat{U}}_{R_{\mathbf{s_I}}} \doteq \hat{\hat{U}}_{\mathbf{s_I}} + \mathbf{b}^t(\bar{\mathbf{X}} - \hat{\hat{\bar{\mathbf{X}}}}_{\mathbf{s_I}}) = \hat{\hat{E}}_{\mathbf{s_I}} + \mathbf{b}^t\bar{\mathbf{X}},$$

where

$$\hat{\hat{E}}_{\mathbf{s_I}} = \frac{1}{N}\sum_{j=1}^{M}\frac{(u_j - \mathbf{b}^t\mathbf{x}_j)}{\pi_j}, \tag{7.35}$$

and $\hat{\hat{U}}_{\mathbf{s_I}}$ and $\hat{\hat{\bar{\mathbf{X}}}}_{\mathbf{s_I}}$ denotes the Horvitz-Thompson estimators of $\bar{U}$ and $\bar{\mathbf{X}}$ based on the m PSU totals in s. Then

$$\text{Var}(\hat{\hat{E}}_{\mathbf{s_I}}) = \frac{1}{N^2}\sum_{j=1}^{M}\sum_{j'=1}^{M}(\pi_{jj'} - \pi_j\pi_{j'})\frac{(u_j - \mathbf{b}^t\mathbf{x}_j)(u_{j'} - \mathbf{b}^t\mathbf{x}_{j'})}{\pi_j\pi_{j'}} \tag{7.36}$$

denotes the approximate design variance of $\hat{\hat{U}}_{R_{\mathbf{s_I}}}$. Furthermore

$$\tilde{\text{Var}}(\hat{\hat{E}}_{\mathbf{s_I}}) = \frac{1}{m}\left(\sum_{j=1}^{M}\frac{m(u_j - \mathbf{b}^t\mathbf{x}_j)^2}{\pi_j N^2} - \left(\bar{U} - \mathbf{b}^t\bar{\mathbf{X}}\right)^2\right) \tag{7.37}$$

denotes the approximate variance of $\hat{\hat{U}}_{R_{\mathbf{s_I}}}$ as if the m PSU's in the first stage are drawn with replacement with selection probabilities π_j/m. Finally

$$\text{Var}(\hat{\bar{e}}_j) = \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j})\frac{(u_{ij} - \mathbf{b}^t\mathbf{x}_{ij})}{\pi_{i|j}}\frac{(u_{i'j} - \mathbf{b}^t\mathbf{x}_{i'j})}{\pi_{i'|j}} \tag{7.38}$$

denotes the approximate design variance of the generalized regression estimator of $\bar{u}_j$. Note that this variance is equivalent to $\text{Var}(\hat{\hat{E}}_{jk_{\mathbf{s}}})$ defined by (6.9) in section 6.2. For a CRD conducted under measurement error model (7.8) and analyzed with the generalized regression estimator, it is proved in appendix 7.7.3 that:

$$
\begin{aligned}
\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} &= \frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\mathrm{Var}}(\hat{\hat{E}}_{\mathbf{s_I}}) - \frac{1}{m}\mathrm{Var}(\hat{\hat{E}}_{\mathbf{s_I}})\right) \\
&\quad + \frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\hat{e}}_j) + \frac{m}{m_k}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\sigma_{ijk}^2}{\pi_i N^2}. \qquad (7.39)
\end{aligned}
$$

The first term on the right-hand side of the equality sign in (7.34) or (7.39) concerns the variance between the PSU-totals. They have the typical variance structure of an embedded CRD but now on the level of the PSU-totals since PSU's are randomized over the K treatments, see section 5.2. The second term concerns the variance between the SSU's within the PSU's. This term arises since the PSU's are not completely observed in two-stage sampling designs. The third term concerns the variance of the error component of the measurement error model.

The efficiency of designing an experiment embedded in a two-stage sampling design as an RBD with PSU's as block variables instead of a CRD with PSU's as experimental units follows if we compare the variance expressions under both designs, i.e. formula (6.6) from section 6.2 with (7.34) for the Horvitz-Thompson estimator or formula (6.11) from section 6.2 with (7.39) for the generalized regression estimator. If the variance between the PSU's is large and the variance between the SSU's within the PSU's is small, then the first term on the right-hand side of the equality sign in (7.34) or (7.39) will be large. Since this variance component will be eliminated in an RBD with PSU's as block variables, blocking on PSU's will be very efficient in such situations.

If the variance between the PSU-totals is small and the variance within the PSU's is large, then a large variance component, which concerns the variance between the SSU's within the PSU's, appears in the variances d_k or d_{k_R} under both experimental designs. In this situation an RBD with PSU's as block variables still has several advantages. First the SSU's are the experimental units, which has the advantage of more degrees of freedom in the estimation procedure. Second the variance between the PSU's, though small, is eliminated from the variances d_k or d_{k_R}. In section 6.4 we saw for cluster samples that it might be efficient to use clusters as experimental units in a CRD if the variance between clusters is small and the variance within clusters large. Since clusters are completely observed, the variance within clusters is eliminated if clusters are used as experimental units. This doesn't hold true for two-stage sampling designs, since PSU's are not completely observed in these sampling designs.

If an experiment is embedded in a two-stage sampling design it will generally be efficient to design it as an RBD with PSU's as block variables. For practical reasons, however, this might be impossible. Especially in sampling designs where households are PSU's there can be a strong duality between the efficiency obtained with blocking on PSU's and the practical advantages of using PSU's as experimental units. If observations of household members are highly correlated, then blocking on households will generally be very efficient. On the other hand there might be strong practical arguments against the application of different treatments within the same household.

7.7 Appendix

7.7.1 Outline of the proof of results (7.23), (7.24), (7.25), (7.26), (7.27) and (7.28)

The results for the covariance matrix of the contrasts between the subsample estimates and the variance estimation procedure are obtained analogous to the derivations given in section 3.6 and 4.4. Now the derivations are principally applied on the level of the PSU's (which are the experimental units) with $\hat{\mathbf{y}}_j^\alpha$ and $\hat{\mathbf{x}}_j$ defined in (7.13) and (7.19). The properties of the randomization vectors $\mathbf{p}_{jk}$ are defined on the level of the PSU's. Properties of $\mathbf{p}_{jk}$ for a CRD follow directly from appendix 3.9.1 where n and n_k are replaced by m and m_k, respectively. Properties of $\mathbf{p}_{jk}$ for an RBD follow directly from appendix 3.9.2 where n_j and n_{jk} are replaced by m_b and m_{bk}, respectively. For the generalized regression estimator it is assumed that there exists a constant $H-$vector $\mathbf{a}$ such that $\mathbf{a}^t\mathbf{x}_{ij} = 1$, for all $i \in U$.

7.7.2 Proof of formula (7.34)

Consider a CRD where the experimental units are the PSU's of a two-stage sampling design, conducted under the basic measurement error model (7.2) in section 7.2. It is proved for the Horvitz-Thompson estimator that

$$\begin{aligned} \mathrm{E}_\alpha \mathrm{E}_s d_k &= \frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\mathrm{Var}}(\hat{\hat{U}}_{k_{\mathbf{s_I}}}) - \frac{1}{m}\mathrm{Var}(\hat{\hat{U}}_{k_{\mathbf{s_I}}})\right) \\ &+\frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\bar{u}}_{jk}) + \frac{m}{m_k}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\sigma_{ijk}^2}{\pi_i N^2}. \end{aligned}$$

Let

$$\hat{\hat{Y}}^\alpha_{k_{\mathbf{s_I}}} = \frac{1}{N}\sum_{j=1}^{m}\frac{y_{jk}^\alpha}{\pi_j}$$

denote the Horvtiz-Thompson estimator of $\bar{Y}_k^\alpha$. Then we can define

$$\mathrm{Var}(\hat{\hat{Y}}^\alpha_{k_{\mathbf{s_I}}}) = \frac{1}{N^2}\sum_{j=1}^{M}\sum_{j'=1}^{M}(\pi_{jj'} - \pi_j\pi_{j'})\frac{y_{jk}^\alpha y_{j'k}^\alpha}{\pi_j\pi_{j'}}, \tag{7.40}$$

and

$$\tilde{\mathrm{Var}}(\hat{\hat{Y}}^\alpha_{k_{\mathbf{s_I}}}) = \frac{1}{m}\left(\sum_{j=1}^{M}\frac{m y_{jk}^{\alpha 2}}{\pi_j N^2} - \bar{Y}_k^{\alpha 2}\right). \tag{7.41}$$

Furthermore,

$$\mathrm{Var}(\hat{\bar{y}}_{jk}^\alpha) = \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j})\frac{y_{ijk}^\alpha y_{i'jk}^\alpha}{\pi_{i|j}\pi_{i'|j}} \tag{7.42}$$

denotes the variance of

$$\hat{\bar{y}}_{jk}^{\alpha} = \frac{1}{N_j}\sum_{i=1}^{N_j}\frac{y_{ijk}^{\alpha}}{\pi_{i|j}}.$$

The diagonal elements d_k are defined by (7.23) in section 7.4. We can evaluate the expectation with respect to the sampling design as follows.

$$\begin{aligned}
\mathrm{E}_s d_k &= \mathrm{E}_s \frac{m}{(m-1)}\frac{m}{m_k}\sum_{j=1}^{m}\left(\frac{\hat{y}_{jk}^{\alpha}}{N\pi_j} - \frac{1}{m}\sum_{j'=1}^{m}\frac{\hat{y}_{j'k}^{\alpha}}{N\pi_{j'}}\right)^2 \\
&= \mathrm{E}_s \frac{m}{(m-1)}\frac{m}{m_k}\frac{1}{N^2}\left(\sum_{j=1}^{m}\frac{\hat{y}_{jk}^{\alpha 2}}{\pi_j^2} - \frac{1}{m}\sum_{j=1}^{m}\sum_{j'=1}^{m}\frac{\hat{y}_{jk}^{\alpha}\hat{y}_{j'k}^{\alpha}}{\pi_j\pi_{j'}}\right) \\
&= \mathrm{E}_s \frac{m}{m_k}\frac{1}{N^2}\left(\sum_{j=1}^{m}\frac{\hat{y}_{jk}^{\alpha 2}}{\pi_j^2} - \frac{1}{(m-1)}\sum_{j=1}^{m}\sum_{j'\neq j}\frac{\hat{y}_{jk}^{\alpha}\hat{y}_{j'k}^{\alpha}}{\pi_j\pi_{j'}}\right). \qquad (7.43)
\end{aligned}$$

The expectation of the single summation term of (7.43) can be evaluated as follows:

$$\begin{aligned}
&\mathrm{E}_s\frac{1}{N^2}\sum_{j=1}^{m}\frac{\hat{y}_{jk}^{\alpha 2}}{\pi_j^2} = \mathrm{E}_{s_I}\left[\mathrm{E}_{s_{II}}\frac{1}{N^2}\sum_{j=1}^{m}\frac{1}{\pi_j^2}\sum_{i=1}^{n_j}\sum_{i'=1}^{n_j}\frac{y_{ijk}^{\alpha}y_{i'jk}^{\alpha}}{\pi_{i|j}\pi_{i'|j}} \mid s_I\right] \\
&= \mathrm{E}_{s_I}\frac{1}{N^2}\sum_{j=1}^{m}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\frac{y_{ijk}^{\alpha}y_{i'jk}^{\alpha}}{\pi_{i|j}\pi_{i'|j}}\frac{\pi_{ii'|j}}{\pi_j^2} = \frac{1}{N^2}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\frac{y_{ijk}^{\alpha}y_{i'jk}^{\alpha}}{\pi_{i|j}\pi_{i'|j}}\frac{\pi_{ii'|j}}{\pi_j} \\
&= \sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\left(\frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j})\frac{y_{ijk}^{\alpha}y_{i'jk}^{\alpha}}{\pi_{i|j}\pi_{i'|j}}\right) + \sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\frac{y_{ijk}^{\alpha}y_{i'jk}^{\alpha}}{N_j^2} \\
&= \sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\bar{y}}_{jk}^{\alpha}) + \frac{1}{N^2}\sum_{j=1}^{M}\frac{y_{jk}^{\alpha 2}}{\pi_j}. \qquad (7.44)
\end{aligned}$$

The expectation of the double summation term of (7.43) can be evaluated as follows:

$$\begin{aligned}
&\mathrm{E}_s\frac{1}{N^2}\sum_{j=1}^{m}\sum_{j'\neq j}^{m}\frac{\hat{y}_{jk}^{\alpha}\hat{y}_{j'k}^{\alpha}}{\pi_j\pi_{j'}} = \mathrm{E}_{s_I}\left[\mathrm{E}_{s_{II}}\frac{1}{N^2}\sum_{j=1}^{m}\sum_{j'\neq j}^{m}\sum_{i=1}^{n_j}\sum_{i'=1}^{n_j'}\frac{y_{ijk}^{\alpha}y_{i'j'k}^{\alpha}}{\pi_{i|j}\pi_{i'|j'}}\frac{1}{\pi_j\pi_{j'}} \mid s_I\right] \\
&= \mathrm{E}_{s_I}\frac{1}{N^2}\sum_{j=1}^{m}\sum_{j'\neq j}^{m}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_{j'}}\frac{y_{ijk}^{\alpha}y_{i'j'k}^{\alpha}}{\pi_j\pi_{j'}} = \mathrm{E}_{s_I}\frac{1}{N^2}\sum_{j=1}^{m}\sum_{j'\neq j}^{m}\frac{y_{jk}^{\alpha}y_{j'k}^{\alpha}}{\pi_j\pi_{j'}} \\
&= \frac{1}{N^2}\sum_{j=1}^{M}\sum_{j'\neq j}^{M}\frac{y_{jk}^{\alpha}y_{j'k}^{\alpha}}{\pi_j\pi_{j'}}\pi_{jj'} \\
&= \frac{1}{N^2}\sum_{j=1}^{M}\sum_{j'=1}^{M}(\pi_{jj'} - \pi_j\pi_{j'})\frac{y_{jk}^{\alpha}y_{j'k}^{\alpha}}{\pi_j\pi_{j'}} + \frac{1}{N^2}\sum_{j=1}^{M}\sum_{j'=1}^{M}y_{jk}^{\alpha}y_{j'k}^{\alpha} - \frac{1}{N^2}\sum_{j=1}^{M}\frac{y_{jk}^{\alpha 2}}{\pi_j} \\
&= \mathrm{Var}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha}) + \bar{Y}_k^{\alpha 2} - \frac{1}{N^2}\sum_{j=1}^{M}\frac{y_{jk}^{\alpha 2}}{\pi_j}. \qquad (7.45)
\end{aligned}$$

If the results obtained in (7.44) and (7.45) are substituted into (7.43), then it follows that

$$\mathrm{E}_s(d_k) = \frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\bar{y}}_{jk}^{\alpha}) - \frac{1}{(m-1)}\frac{m}{m_k}\mathrm{Var}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha})$$

$$-\frac{1}{(m-1)}\frac{m}{m_k}\frac{1}{N^2}\bar{Y}_k^{\alpha 2}+\frac{m}{(m-1)}\frac{m}{m_k}\frac{1}{N^2}\sum_{j=1}^{M}\frac{y_{jk}^{\alpha 2}}{\pi_j}$$

$$= \frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\bar{y}}_{jk}^{\alpha})-\frac{1}{(m-1)}\frac{m}{m_k}\mathrm{Var}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha})$$

$$+\frac{1}{(m-1)}\frac{m}{m_k}\left(\sum_{j=1}^{M}\frac{my_{jk}^{\alpha 2}}{\pi_j N^2}-\bar{Y}_k^{\alpha 2}\right)$$

$$= \frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\mathrm{Var}}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha})-\frac{1}{m}\mathrm{Var}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha})\right)+\frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\bar{y}}_{jk}^{\alpha}). \quad (7.46)$$

Now we will evaluate the expectation with respect to the measurement error model. If the basic measurement error model (7.2) in section 7.2 is substituted into the expressions of $\mathrm{Var}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha})$, $\tilde{\mathrm{Var}}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha})$ and $\mathrm{Var}(\hat{\bar{y}}_{jk}^{\alpha})$, then the following results are obtained.

$$\begin{aligned}
\mathrm{E}_\alpha\mathrm{Var}(\hat{\bar{y}}_{jk}^{\alpha}) &= \mathrm{E}_\alpha\frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'|j}-\pi_{i|j}\pi_{i'|j})\frac{(u_{ij}+\beta_k+\varepsilon_{ijk}^{\alpha})}{\pi_{i|j}}\frac{(u_{i'j}+\beta_k+\varepsilon_{i'jk}^{\alpha})}{\pi_{i'|j}} \\
&= \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'|j}-\pi_{i|j}\pi_{i'|j})\frac{(u_{ij}+\beta_k)}{\pi_{i|j}}\frac{(u_{i'j}+\beta_k)}{\pi_{i'|j}}+\frac{1}{N_j^2}\sum_{i=1}^{N_j}\frac{(1-\pi_{i|j})}{\pi_{i|j}}\sigma_{ijk}^2 \\
&= \mathrm{Var}(\hat{\bar{u}}_{jk})+\frac{1}{N_j^2}\sum_{i=1}^{N_j}\frac{(1-\pi_{i|j})}{\pi_{i|j}}\sigma_{ijk}^2, \qquad (7.47)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}_\alpha\mathrm{Var}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha}) &= \mathrm{E}_\alpha\frac{1}{N^2}\sum_{j=1}^{M}\sum_{j'=1}^{M}(\pi_{jj'}-\pi_j\pi_{j'})\frac{(u_j+N_j\beta_k+\varepsilon_{jk}^{\alpha})}{\pi_j}\frac{(u_{j'}+N_{j'}\beta_k+\varepsilon_{j'k}^{\alpha})}{\pi_{j'}} \\
&= \mathrm{Var}(\hat{\bar{U}}_{k_{\mathbf{s_I}}})+\mathrm{E}_\alpha\frac{1}{N^2}\sum_{j=1}^{M}\sum_{j'=1}^{M}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_{j'}}\frac{(\pi_{jj'}-\pi_j\pi_{j'})}{\pi_j\pi_{j'}}\varepsilon_{ijk}^{\alpha}\varepsilon_{i'j'k}^{\alpha} \\
&= \mathrm{Var}(\hat{\bar{U}}_{k_{\mathbf{s_I}}})+\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{(1-\pi_j)}{\pi_j}\sigma_{ijk}^2, \qquad (7.48)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}_\alpha\tilde{\mathrm{Var}}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha}) &= \mathrm{E}_\alpha\frac{1}{m}\left(\sum_{j=1}^{M}\frac{m(u_j+N_j\beta_k+\varepsilon_{jk}^{\alpha})^2}{\pi_j N^2}-\left(\bar{U}+\beta_k+\frac{1}{N}\sum_{j=1}^{M}\varepsilon_{jk}^{\alpha}\right)^2\right) \\
&= \tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_{\mathbf{s_I}}})+\mathrm{E}_\alpha\frac{1}{m}\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}\frac{m\varepsilon_{ijk}^{\alpha}\varepsilon_{i'jk}^{\alpha}}{\pi_j N}-\left(\frac{1}{N}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\varepsilon_{ijk}^{\alpha}\right)^2\right) \\
&= \tilde{\mathrm{Var}}(\hat{\bar{U}}_{k_{\mathbf{s_I}}})+\frac{1}{m}\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{m\sigma_{ijk}^2}{\pi_j N^2}-\frac{1}{N^2}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\sigma_{ijk}^2\right). \qquad (7.49)
\end{aligned}$$

If the results obtained in (7.47), (7.48) and (7.49) are substituted into (7.46), then it follows that

$$\mathrm{E}_\alpha\mathrm{E}_s d_k = \mathrm{E}_\alpha\frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\mathrm{Var}}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha})-\frac{1}{m}\mathrm{Var}(\hat{\bar{Y}}_{k_{\mathbf{s_I}}}^{\alpha})\right)+\frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\bar{y}}_{jk}^{\alpha})$$

$$
\begin{aligned}
&= \frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\text{Var}}(\hat{\hat{U}}_{k_{\mathbf{s_I}}}) - \frac{1}{m}\text{Var}(\hat{\hat{U}}_{k_{\mathbf{s_I}}})\right) \\
&\quad + \frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\text{Var}(\hat{u}_{jk}) + \frac{m}{m_k}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\sigma_{ijk}^2}{\pi_i N^2}, \quad \textbf{QED}.
\end{aligned}
$$

7.7.3 Proof of formula (7.39)

Consider a CRD where the experimental units are the PSU's of a two-stage sampling design, conducted under the basic measurement error model (7.8) in section 7.2. It is proved for the generalized regression estimator that

$$
\begin{aligned}
\text{E}_\alpha \text{E}_s d_{k_R} &= \frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\text{Var}}(\hat{\hat{E}}_{\mathbf{s_I}}) - \frac{1}{m}\text{Var}(\hat{\hat{E}}_{\mathbf{s_I}})\right) \\
&\quad + \frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\text{Var}(\hat{e}_j) + \frac{m}{m_k}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\sigma_{ijk}^2}{\pi_i N^2}.
\end{aligned}
$$

Equivalent to the derivation of (7.46) in appendix 7.7.2 it follows that

$$
\begin{aligned}
\text{E}_s d_{k_R} &= \text{E}_s \frac{m}{(m-1)}\frac{m}{m_k}\sum_{j=1}^{m}\left(\frac{(\hat{y}_{jk}^\alpha - \mathbf{b}_k^t \hat{\mathbf{x}}_j)}{N\pi_j} - \frac{1}{m}\sum_{j'=1}^{m}\frac{(\hat{y}_{j'k}^\alpha - \mathbf{b}_k^t \hat{\mathbf{x}}_{j'})}{N\pi_{j'}}\right)^2 \\
&= \frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\text{Var}}(\hat{\hat{E}}_{k_{\mathbf{s_I}}}^\alpha) - \frac{1}{m}\text{Var}(\hat{\hat{E}}_{k_{\mathbf{s_I}}}^\alpha)\right) + \frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\text{Var}(\hat{e}_{jk}^\alpha), \qquad (7.50)
\end{aligned}
$$

where

$$
\text{Var}(\hat{\hat{E}}_{k_{\mathbf{s_I}}}^\alpha) = \frac{1}{N^2}\sum_{j=1}^{M}\sum_{j'=1}^{M}(\pi_{jj'} - \pi_j\pi_{j'})\frac{(y_{jk}^\alpha - \mathbf{b}_k^t \mathbf{x}_j)}{\pi_j}\frac{(y_{j'k}^\alpha - \mathbf{b}_k^t \mathbf{x}_{j'})}{\pi_{j'}},
$$

$$
\tilde{\text{Var}}(\hat{\hat{E}}_{k_{\mathbf{s_I}}}^\alpha) = \frac{1}{m}\left(\sum_{j=1}^{M}\frac{m(y_{jk}^\alpha - \mathbf{b}_k^t \mathbf{x}_j)^2}{\pi_j N^2} - (\bar{Y}_k^\alpha - \mathbf{b}_k^t \bar{\mathbf{X}})^2\right),
$$

and

$$
\text{Var}(\hat{e}_{jk}^\alpha) = \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{i'=1}^{N_j}(\pi_{ii'|j} - \pi_{i|j}\pi_{i'|j})\frac{(y_{ijk}^\alpha - \mathbf{b}_k^t \mathbf{x}_{ij})}{\pi_{i|j}}\frac{(y_{i'jk}^\alpha - \mathbf{b}_k^t \mathbf{x}_{i'j})}{\pi_{i'|j}}.
$$

Under the assumption that there exists a constant $H-$vector $\mathbf{a}$ such that $\mathbf{a}^t\mathbf{x}_{ij} = 1$, for all $i \in U$, it follows that $\mathbf{b}_k = \mathbf{b} + \mathbf{a}\beta_k$. Consequently,

$$
y_{ijk}^\alpha - \mathbf{b}_k^t \mathbf{x}_{ij} = u_{ij} - \mathbf{b}^t \mathbf{x}_{ij} + \varepsilon_{ijk}^\alpha,
$$

and

$$
y_{jk}^\alpha - \mathbf{b}_k^t \mathbf{x}_j = u_j - \mathbf{b}^t \mathbf{x}_j + \varepsilon_{jk}^\alpha.
$$

Now we can evaluate the expectation with respect to the measurement error model as follows. Equivalent to the derivation of (7.47), (7.48) and (7.49) in appendix 7.7.2 it follows that

$$
\text{E}_\alpha \text{Var}(\hat{\hat{e}}_{jk}^\alpha) = \text{Var}(\hat{\hat{e}}_{jk}) + \frac{1}{N_j^2}\sum_{i=1}^{N_j}\frac{(1-\pi_{i|j})}{\pi_{i|j}}\sigma_{ijk}^2, \qquad (7.51)
$$

$$\mathrm{E}_\alpha \mathrm{Var}(\hat{\hat{E}}^{\alpha}_{k_{\mathbf{s_I}}}) = \mathrm{Var}(\hat{\hat{E}}_{k_{\mathbf{s_I}}}) + \sum_{j=1}^{M}\sum_{i=1}^{N_j} \frac{(1-\pi_j)}{\pi_j}\sigma^2_{ijk}, \tag{7.52}$$

and

$$\mathrm{E}_\alpha \tilde{\mathrm{Var}}(\hat{\hat{E}}^{\alpha}_{k_{\mathbf{s_I}}}) = \tilde{\mathrm{Var}}(\hat{\hat{E}}_{k_{\mathbf{s_I}}}) + \frac{1}{m}\left(\sum_{j=1}^{M}\sum_{i=1}^{N_j} \frac{m\sigma^2_{ijk}}{\pi_j N^2} - \frac{1}{N^2}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\sigma^2_{ijk}\right). \tag{7.53}$$

If (7.51), (7.52) and (7.53) are substituted into (7.50), then it follows that

$$\begin{aligned}\mathrm{E}_\alpha \mathrm{E}_s d_{k_R} \quad = \quad & \frac{m}{(m-1)}\frac{m}{m_k}\left(\tilde{\mathrm{Var}}(\hat{\hat{E}}_{\mathbf{s_I}}) - \frac{1}{m}\mathrm{Var}(\hat{\hat{E}}_{\mathbf{s_I}})\right) \\ & + \frac{m}{m_k}\sum_{j=1}^{M}\frac{(N_j/N)^2}{\pi_j}\mathrm{Var}(\hat{\tilde{e}}_j) + \frac{m}{m_k}\sum_{j=1}^{M}\sum_{i=1}^{N_j}\frac{\sigma^2_{ijk}}{\pi_i N^2}, \quad \textbf{QED}.\end{aligned}$$

Chapter 8

Testing effects of a new questionnaire in the Dutch Labour Force Survey

8.1 Introduction

The objective of the Dutch Labour Force Survey (LFS) is to provide reliable information about the situation on the labour market, especially about employment and unemployment rates. The LFS has been carried out as a continuous survey since 1987. Each month a sample of addresses is selected from which, during the data collection, households are identified that can be regarded as the ultimate sampling units. Data are collected by means of Computer Assisted Personal Interviewing (CAPI) using hand-held computers. The questionnaire was developed by means of a DOS-version of Blaise 2. Since the LFS changed in October 1999 from a cross-sectional survey to a rotating panel survey, a new questionnaire was developed in a Windows version of Blaise 4. Before the new questionnaire was implemented as a standard in the LFS, the impact of this conversion on the main outcomes of the LFS was tested by means of a field experiment. This experiment was designed as a two-sample RBD embedded in the sample of the current LFS. To test hypotheses about possible effects due to the differences between the two questionnaires on estimates of the target parameters of the LFS, the design-based analysis procedure proposed in the preceding chapters was applied and compared with the results obtained with a standard model-based analysis procedure. As a result of the observed effects in the experiment, the new questionnaire was adjusted before it was implemented as a standard in the LFS.

This chapter describes the design, analysis and the results of this experiment. The survey design of the LFS is described in section 2. A detailed description of the methodology of the LFS is given by Hilbink, Van Berkel and Van den Brakel (2000). In section 3 the experimental design used to test the effects of the new questionnaire is described. In section 4 the design-based methods applied in the analysis of this particular experiment are described. In section 5 we analyze the LFS experiment with these design-based methods and compare the results obtained with a standard model-based procedure. Conclusions are drawn in section 6.

8.2 Design of the Dutch Labour Force Survey

8.2.1 Sample design

The target population of the LFS consists of the non-institutionalised population aged over 15 residing in the Netherlands. The sampling frame is derived from a register of all known addresses in the Netherlands (Geographical Base Register). The addresses in the sampling frame are strictly speaking postal delivery points. The LFS design is based on a stratified two-stage cluster sample of addresses. Municipalities are considered as Primary Sampling Units (PSU) and addresses as Secondary Sampling Units (SSU). Strata are formed by geographical areas based on a cross-classification of Corop regions (Corop is the Dutch abbreviation for Committee for the Coordination of Regional Research Program) and RBA regions (RBA is the Dutch abbreviation for Regional Employment Board). In the first stage a sample of municipalities is drawn with inclusion probabilities proportional to the number of addresses. Municipalities with more than 75.000 addresses are self-selecting and are included in the sample for each month. Non-self selecting PSU's are drawn in such a way that all (or nearly all) municipalities are included in the sample for at least one month of the calendar year. At the second stage a sample of minimal 12 addresses is drawn without replacement from each selected PSU. Principally, all households residing on an address, with a maximum of three, are included in the sample. The monthly sample size amounts to approximately 10.000 addresses.

Since the LFS has to provide accurate outcomes for the monthly publication on registered unemployment, addresses that occur in the register of the Employment Exchange are oversampled. Since most target parameters of the LFS concern people aged between 15 and 65, addresses with only persons aged over 65 are undersampled. Technically this is carried out by drawing a sample that is three times as large as necessary for the LFS (30.000 addresses per month). The number of addresses drawn from each municipality in the second stage is determined such that a self-weighted sample of addresses is obtained. This sample is randomly divided into three subsamples of equal size. From the first subsample 60% of the addresses with only people aged over 65 are removed. The remaining addresses of the first subsample are included in the final sample. From the second subsample only the addresses which occur in the register of the Employment Exchange (employed or unemployed) are included. From the third subsample only the addresses of persons who are registered as unemployed are included. This results in a sample of approximately 10.000 addresses per month. Due to reduced capacity of the field staff during the holiday season, the sample size is halved in the months July and August.

8.2.2 Data collection

Data are collected in personal interviews with hand-held computers. These computers contain the questionnaire supported by the Blaise system. About 500 interviewers are working on the data collection of the LFS in areas around their place of residence. For all members of the selected households, demographic variables are observed. For the target variables only persons aged over 15 are interviewed.

At most four persons in a household are interviewed. The so-called household kernel is always interviewed. The members of the household kernel are the wage earner and his or her partner. When there are more than four household members the other respondents are selected first from the members aged between 15 and 65. If the number of four is not reached after this selection, persons aged over 65 are selected too.

When a household member cannot be contacted the required information can be obtained by means of a proxy-interview. This means that another member of the household, preferably a member of the household kernel, answers the questions concerning the absent person. By allowing proxy-interviews the number of partially responding households is reduced. Households in which one or more of the selected persons do not respond for themselves or in a proxy interview, are treated as nonresponding households.

8.2.3 Weighting procedure

The weighting procedure of the LFS starts with deriving inclusion weights for the responding households. In doing this we take into account:

1. The oversampling of addresses that occur in the register of the Employment Exchange.
2. The undersampling of addresses with only persons aged over 65.
3. The month of interview (i.e. held in July or August or in the other months).
4. The different response rates between geographical regions. These regions are formed by the self-selecting municipalities. The non-self selecting municipalities are grouped into regions by province.

In the weighting procedure, inclusion weights are adjusted using the generalized regression estimator. Regression-based methods use auxiliary variables that are observed in the sample for which the population totals are known. For the LFS only categorical auxiliary variables are available and the corresponding population totals are present as population counts. The weighting scheme is based on a combination of different social-demographic categorical variables. The weighting procedure is based on the integrated method for weighting persons and families of Lemaître and Dufour (1987) to ensure a single weight for each member of the same household. Finally a bounding algorithm is applied to avoid negative weights.

8.3 Experimental design

To investigate the effects of a new questionnaire supported by a Windows version of Blaise 4 on the main outcomes of the LFS, a large scale field experiment was carried out during the period

April through September 1999. The experiment was performed as a two-treatment embedded RBD. For the experiment 90 experimental areas or blocks were selected. Each block consists of the union of two neighboring interview-areas of two interviewers. The addresses in the monthly sample of the LFS in each block were randomly divided into two subsamples of equal size. One subsample was assigned to the new questionnaire, the other subsample to the regular one. The two interviewers in each block were randomly assigned to the regular or the new questionnaire. Both interviewers collected data from the households on the addresses of the subsample with the questionnaire to which they were assigned. It follows that both interviewers worked through the entire area of a block. Using regions as block variables enables us to control for regional effects in the analysis of the experiment. It is emphasized that the observations obtained with the regular questionnaire from the households outside these blocks were not used in the analysis of this experiment. If those observations were also incorporated in the analysis, then the estimated treatment effects would have been biased with regional effects.

It was decided not to use interviewers as block variables, since this implies that each interviewer must execute the LFS with the new as well as the regular questionnaire. This has the important drawback that interviewers easily confuse the two questionnaires during the data collection, which might disturb the experiment. Furthermore, since it was not possible to run both questionnaires on one hand-held computer, the interviewer would be forced to visit addresses with two different questionnaires.

Blocks were selected as follows. From the available interviewers with at least one year of experience with the LFS, a list of potential couples was formed. The union of the interview areas of each couple formed a potential block. From this list of potential blocks a stratified sample of 90 blocks was drawn using Corop as a stratification variable. Since the selected blocks were equally divided among the Netherlands by region, and since a random sample of addresses was drawn in each block for both treatments, the results observed in the experiment can be generalized to the entire target population of the LFS. According to this experimental design, each month approximately 15% of the sample size of the LFS was assigned to the new questionnaire.

One month of pretesting, during March, preceded the experiment so as to preclude distortion of the experiment due to initial problems with the Windows version of Blaise 4 on the new hand-held computers, and in the way the interviewers dealth with the new questionnaire. Many unexpected problems with the new software were solved during this month of pretesting.

8.4 Methods

The aim of this experiment is to test hypotheses about the effects of the new questionnaire on the main outcomes of the LFS. The design-based analysis procedure for embedded experiments proposed in the preceeding chapters is applied for the analysis of this experiment. Let $\bar{Y}_k^{\alpha}$ denote the finite poplution mean observed under treatment k. From the purpose of the experiment it follows that we are interested in the hypothesis

$$\begin{aligned} H_0: &\quad \mathrm{E}_\alpha(\bar{Y}_1^\alpha) = \mathrm{E}_\alpha(\bar{Y}_2^\alpha), \\ H_1: &\quad \mathrm{E}_\alpha(\bar{Y}_1^\alpha) \neq \mathrm{E}_\alpha(\bar{Y}_2^\alpha). \end{aligned} \tag{8.1}$$

Based on the observations obtained in the experiment, two estimates were obtained for each population parameter of the LFS. One estimate was based on the subsample where data were collected with the regular questionnaire. The other was based on the subsample where data were collected with the new questionnaire. Hypothesis (8.1), about possible effects on the estimates of the population parameters of the LFS induced by the different questionnaires, can be tested using a design-based $t-$statistic

$$t = \frac{\hat{\bar{Y}}^\alpha_{1_R} - \hat{\bar{Y}}^\alpha_{2_R}}{\sqrt{\hat{\mathrm{Var}}(\hat{\bar{Y}}^\alpha_{1_R} - \hat{\bar{Y}}^\alpha_{2_R})}}. \tag{8.2}$$

The extended Horvitz-Thompson estimator as well as the generalized regression estimator is applied for the estimation of the population parameters as well as the variance of the differences between these two estimators.

Since households are both the sampling and the experimental units, the analysis is carried out at the household level. Let y^α_{ijk} denote the observation obtained from person i from household j assigned to treatment k. Then $y^\alpha_{jk} = \sum_{i=1}^{n_j} y^\alpha_{ijk}$ is the target parameter of household j assigned to treatment k with n_j the number of persons in household j aged over 15. It is assumed that the observations obtained from the households are a realization of the basic measurement error model (4.5) with model assumptions (4.6) and (4.7) in section 4.2.

The first-order inclusion probabilities of households are derived based on the realized net sample within each block. At this we take into account that addresses which occur in the register of the Employment Exchange are oversampled, addresses where only persons aged over 65 live are undersampled, and the sample size in the months July and August is reduced. See Hilbink, Van Berkel and Van den Brakel (2000) for technical details about the derivation of the first-order inclusion probabilities.

To allow for the weighting procedure of the LFS, the analysis must be based on the generalized regression estimator. Let $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijH})^t$ denote a vector of order H with each element x_{ijh} an auxiliary variable of individual i from household j. Then $\mathbf{x}_j = \sum_{i=1}^{n_j} \mathbf{x}_{ij}$ denotes a vector of order H with the household totals of the auxiliary variables of the persons aged over 15. A linear regression model for the intrinsic values of the households in the population is defined by (4.2) with model assumptions (4.3) and (4.4) in section 4.2. The weighting procedure is carried out at the level of household totals. The variance structure in the linear regression model (4.2) is chosen proportional to the size of the household: $\omega_j^2 = n_j\omega^2$ with n_j the number of persons aged over 15 in household j. Nieuwenbroek (1993) shows that in this situation the weighting procedure corresponds to the integrated method for weighting persons and households proposed by Lemaître and Dufour (1987).

The net sample sizes of both subsamples are relatively small compared with the sample of the LFS (see section 8.5). Therefore we could not apply the regular weighting scheme of the LFS in the analysis of this experiment. The following weighting scheme, which contains the most important auxiliary information of the regular weighting scheme of the LFS, was applied in the analysis of this experiment:

$$Age \times Region + Sex \times Region + Marital\ Status \times Region, \quad (8.3)$$

where the four variables are categorical.

If the experiment is analyzed as an RBD using the generalized regression estimator, then an expression for $\hat{\hat{Y}}^{\alpha}_{k_R}$ in the enumerator of (8.2) is given by (4.16) or (4.17) in section 4.3. An expression for the variance in the denominator of (8.2) is given by

$$\hat{\mathrm{Var}}(\hat{\hat{Y}}^{\alpha}_{1_R} - \hat{\hat{Y}}^{\alpha}_{2_R}) = \hat{d}_{1_R} + \hat{d}_{2_R}, \quad (8.4)$$

where $\hat{d}_{k_R}$ is defined by (4.31), section 4.4.

An expression for the $t-$statistic analyzed with the extended Horvitz-Thompson estimator follows as a special case from the results obtained for the generalized regression estimator. The minimum use of auxiliary information is a weighting scheme where $\mathbf{x}_{ij} = (1)$, hence $\mathbf{x}_j = (n_j)$, for all elements in the population. Furthermore it is assumed that $\omega_j^2 = n_j\omega^2$. This weighting scheme corresponds with the common mean model. An expression for $\hat{\hat{Y}}^{\alpha}_{k_R}$ in the enumerator of (8.2) is given by $\tilde{y}_{s_k}$ defined in (4.33), section 4.4. An expression for the variance is obtained by (8.4), where where $\hat{d}_{k_R}$ is given by (4.31), with $(y^{\alpha}_{jk} - \hat{\mathbf{b}}^{\alpha^t}_k \mathbf{x}_j)$ replaced by $(y^{\alpha}_{jk} - n_j\tilde{y}_{s_k})$.

To quantify the efficiency of the RBD, this experiment is also analyzed as if it were designed as a CRD. Then expressions for the variance in (8.2) are given by (8.4) with $\hat{d}_{k_R}$ defined by (4.30), section 4.4.

It is assumed that a finite population central limit theorem holds so that the two estimated population parameters have a multivariate normal distribution. Consequently, the $t-$statistic under the null-hypothesis of no treatment effects is standard-normally distributed.

8.5 Results

A total of 16647 addresses were selected in the sample of the LFS for this experiment. From this sample, 8262 addresses were assigned to the subsample where data were collected with the regular questionnaire and 8385 addresses to the subsample where data were collected with the new questionnaire. Response rates in both subsamples amount to approximately 50%. The response obtained in a block during one month was used in the analysis only if each of the two interviewers in that block obtained a response of at least 3 households. In several blocks one of the two interviewers didn't have any response at all during a month due to e.g. illness or vacation. In such situations the response obtained in a block for the other interviewer during that month was excluded from the analysis in order to avoid any systematic influence in the analysis caused by seasonal effects. Finally 3037 households interviewed with the regular questionnaire and 3159 households interviewed with the new questionnaire were used in the analysis of the experiment.

The following five parameters of the LFS were analyzed: the Employed Labour Force, the Unemployed Labour Force, Registered Unemployment, the Employed Labour Force according to the International Labour Organization (abbreviated as ILO Employed) and the Registered at the Employment Exchange. All parameters are expressed as percentages of the population aged between 15 and 65. The first analysis is based on the $t-$statistic where the population parameters and the variance of the corresponding contrasts were estimated with the generalized regression estimator for an RBD with the weighting scheme defined by (8.3). This is the most appropriate analysis since it takes both the experimental design into account and the design of the LFS including the estimation and weighting procedure. Results are given in table 8.1.

Table 8.1: Results with the generalized regression estimator for an RBD.

	LFS questionnaire					
	regular	new	diff.	std. error	$t-$statistic	$p-$value
Employed Labour Force	64.710	63.046	-1.663	1.006	-1.653	0.098
Unemployed Labour Force	2.403	3.051	0.649	0.337	1.923	0.054
Registered Unemployment	2.136	2.612	0.476	0.324	1.471	0.141
ILO Employed	71.490	69.741	-1.749	0.985	-1.776	0.076
Registered at Empl. Exc.	7.640	9.072	1.433	0.568	2.524	0.012

Parameters are expressed as percentages of the population aged between 15 and 65.

The second analysis is based on the $t-$statistic where the population parameters and the variance of the corresponding contrasts were estimated with the Horvitz-Thompson estimator for an RBD. Results are given in table 8.2.

Table 8.2: Results with the Horvitz-Thompson estimator for an RBD.

	LFS questionnaire					
	regular	new	diff.	std. error	$t-$statistic	$p-$value
Employed Labour Force	65.306	62.667	-2.639	1.763	-1.497	0.134
Unemployed Labour Force	2.246	2.954	0.709	0.337	2.103	0.035
Registered Unemployment	1.935	2.515	0.580	0.327	1.773	0.076
ILO Employed	72.197	69.263	-2.934	1.877	-1.563	0.118
Registered at Empl. Exc.	7.265	8.542	1.277	0.543	2.351	0.019

Parameters are expressed as percentages of the population aged between 15 and 65.

Compared with the Horvitz-Thompson estimator the generalized regression estimator has two effects on the analysis. Firstly, the size of the treatment effects is reduced. The estimates of the parameters under both treatments are more similar due to the application of the auxiliary information in the weighting scheme. Secondly, the variance of the estimated treatment effects is reduced. For the Unemployed Labour Force and Registered Unemployment this finally resulted in less significant treatment effects while for the other parameters the treatment effects are more significant.

In order to compare the efficiency of designing this experiment as an RBD, the experiment was analyzed as if it were designed as a completely randomized design (CRD). Results for the generalized regression estimator are given in table 8.3 and for the Horvitz-Thompson estimator in table 8.4.

Table 8.3: Results with the generalyzed regression estimator for a CRD.

	LFS questionnaire					
	regular	new	diff.	std. error	$t-$statistic	$p-$value
Employed Labour Force	64.711	63.067	-1.643	0.986	-1.666	0.096
Unemployed Labour Force	2.396	3.025	0.629	0.338	1.860	0.063
Registered Unemployment	2.152	2.511	0.359	0.313	1.147	0.251
ILO Employed	71.499	69.694	-1.805	0.968	-1.865	0.062
Registered at Empl. Exc.	7.495	9.132	1.636	0.546	2.997	0.003

Parameters are expressed as percentages of the population aged between 15 and 65.

Table 8.4: Results with the Horvitz-Thompson estimator for a CRD.

	LFS questionnaire					
	regular	new	diff.	std. error	$t-$statistic	$p-$value
Employed Labour Force	64.779	63.014	-1.764	1.871	-0.943	0.346
Unemployed Labour Force	2.265	2.975	0.710	0.343	2.069	0.039
Registered Unemployment	1.913	2.471	0.558	0.343	1.627	0.104
ILO Employed	71.721	69.495	-2.226	1.993	-1.117	0.264
Registered at Empl. Exc.	7.062	8.732	1.670	0.552	3.023	0.003

Parameters are expressed as percentages of the population aged between 15 and 65.

If the results of an analysis for an RBD are compared with a CRD, both analyzed with the Horvitz-Thompson estimator (tables 8.2 and 8.4), then it follows that the variance of the estimated treatment effects under a CRD increases only slightly. If the results of an analysis for an RBD are compared with a CRD using the generalized regression estimator (tables 8.1 and 8.3), then it follows that the variance of the estimated treatment effects is approximately equal or even slightly smaller under a CRD. The efficiency of blocking is small in this application, which can be explained as follows. Firstly, the variation of the parameters of the LFS between the blocks appears to be small. Secondly, only if the fraction of households assigned to a treatment within each block is equal for each block (i.e. $n_{bk}/n_b = n_{b'k}/n_{b'}$, where n_b denotes the number of households in block b and n_{bk} the number of households in block b assigned to treatment k), then it can be proved that the variance of the estimated treatment effects for a CRD is larger than for an RBD (see section 6.3). For the gross sample these fractions were equal by the design of this experiment. However, due to unequal response rates of the two interviewers within each block, the equality of these fractions for the net sample were severely disturbed. This resulted in an extra variation of the sample weights $n_b/(n_{bk}\pi_i)$ in the analysis of an RBD. This explains the small variance reduction due to the application of an RBD for the Horvitz-Thompson estimator. The weighting scheme of the generalized regression estimator eliminates some regional variation from the estimated treatment effects. Therefore, the negative influence of the unequal allocation of the households over the treatments within the blocks of an RBD becomes more obvious if the analysis is carried out with the generalized regression estimator. This explains why the variance of the treatment effects are even slightly smaller under a CRD for the generalized regression estimator. In summary, due to different response rates obtained by the interviewers, the optimality of the RBD was at least partially disturbed. This might have been avoided if interviewers had been used as block variables.

The effect on the analysis of the application of the generalized regression estimator compared with the Horvitz-Thompson estimator in the case of CRD follows from tables 8.3 and 8.4. As in the case of an RBD (tables 8.1 and 8.2) it follows that the generalized regression estimator reduces the differences between the treatments as well as the variance of the estimated treatment effects.

To compare the effect on the analysis of a design-based with a model-based analysis, this experiment was also analyzed with the standard model-based $t-$statistic and a standard ANOVA for an RBD, using SPSS. To approximate the target variables of the LFS as well as possible in these analyses, the dependent variables are the household means of the target parameters. The observations obtained from the experimental units are modeled in a linear regression model containing one factor with two levels that models the treatments in the experiment and, in the case of an RBD, one factor that models the block structure. It is assumed that the residuals are normally and independently distributed with expectation zero and equal variance. Results are summarized in tables 8.5 and 8.6. The complete ANOVA tables are omitted here; only the analysis results of the treatment effects are included in table 8.6.

Table 8.5: Results with the standard $t-$statistic (unweighted).

	LFS questionnaire					
	regular	new	diff.	std. error	$t-$statistic	$p-$value
Employed Labour Force	62.13	60.49	-1.646	0.991	-1.661	0.097
Unemployed Labour Force	3.716	4.708	0.992	0.419	2.370	0.018
Registered Unemployment	3.836	4.839	1.003	0.441	2.275	0.023
ILO Employed	68.68	66.44	-2.240	0.983	-2.279	0.023
Registered at Empl. Exc.	14.68	16.04	1.361	0.787	1.728	0.084

Parameters are expressed as percentages of the population aged between 15 and 65.

Table 8.6: Results with the ANOVA for an RBD (unweighted).

	contrast	std. error	$t-$statistic	$p-$value
Employed Labour Force	-1.372	1.00	-1.370	0.171
Unemployed Labour Force	1.025	0.40	2.413	0.016
Registered Unemployment	1.028	0.40	2.306	0.021
ILO Employed	-1.795	1.00	-1.818	0.069
Registered at Empl. Exc.	0.678	0.80	0.860	0.390

Contrasts of the parameters are expressed as percentages of the population aged between 15 and 65.

The influence of a model-based versus a design-based analysis differs for the five parameters of the LFS. It follows from table 8.5 that the Unemployed Labour Force, Registered Unemployment and the Registered at Employment Exchange are severely overestimated while the Employed Labour Force and the ILO Employed are underestimated. This is mainly caused by the oversampling of addresses that occur in the register of the Employment Exchange, which is ignored in a standard model-based analysis. This consequently results in biased estimates for the population parameters. Especially the treatment effects for the Unemployed Labour Force, Registered Unemployment and the ILO Employed are larger and more significant in a model-based analysis. In a design-based analysis the variance of the estimated treatment effects might be increased due to the fluctuation of the sampling weights. Nevertheless, the variance of the estimated treatment effects for the Unemployed Labour Force, Registered Unemployment and the ILO Employed in the model-based analysis are slightly higher than the corresponding variances obtained with the design-based analysis. It is unclear why, contrary to the results observed for the Unemployed Labour Force and Registered Unemployment, the estimated treatment effect of those Registered at the Employment Exchange is smaller and less significant in a model-based approach.

The efficiency of an RBD in a model-based analysis follows if the results of tables 8.5 and 8.6 are compared. Although in the ANOVA for an RBD, the $F-$statistic for the blocks is significant at a level of 0.05, the reduction in the error sum of squares is very small. Consequently, the variance reduction of the estimated treatment effects due to the application of an RBD, is minor. For the Employed Labour Force, the ILO Employed and the persons registered at the employment exchange, the contrasts are smaller due to the correction for the block variables, resulting in less significant treatment effects.

In conclusion, it follows that the results obtained with the model-based analysis with the t-test as well as the ANOVA for an RBD are biased since the specific features of the weighting procedure of the LFS, especially the oversampling of addresses with persons registered at Employment Exchange, is ignored in these analyses.

The final analysis of the experiment was based on the generalized regression estimator for an RBD (table 8.1) since this estimator allows for both the sampling design and the weighting procedure of the LFS. Tests were carried out at a significance level of 0.05.

The Employed Labour Force as well as the ILO Employed observed with the new questionnaire are about 1.7 percentage points smaller than with the regular one. Although the tests of no treatment effects for these parameters are not rejected, the differences are substantial. Further investigation showed that these differences are caused by the Self-Employed. With the new questionnaire the percentage of the Self-Employed were found to be 1.8 percentage points smaller than with the regular one. Since the corresponding $p-$value amounts 0.01, this difference is significant. Changes in the routing of the new questionnaire caused many self-employed to answer the question "Do you currently have a paid job?" in the negative. Consequently these self-employed are erroneously not classified as active Labour Force. This effect was confirmed by the debriefings of the interviewers. This resulted in the observed differences for the Employed Labour Force and the ILO Employed. The new questionnaire was adjusted in order to avoid this under-reporting before it was actually implemented as the standard questionnaire in the LFS.

The Unemployed Labour Force and Registered Unemployment measured with the new questionnaire are about 0.5 percentage points higher than with the regular questionnaire. These differences are substantial and almost significant for the Unemployed Labour Force.

For the parameter Registered at the Employment Exchange the observed difference of 1.4 percentage points is significant. In the regular questionnaire, the question about registration at the Employment Exchange was preceded by several questions concerning benefits. In the new questionnaire the questions about benefits were skipped. Possible interactions between the questions concerning benefits and registration at the Employment Exchange might explain the observed difference. Despite this observed difference the questions about benefits are not included in the new questionnaire since this information will be available from registrations.

8.6 Conclusions

This experiment illustrates the necessity to investigate if and how changes in the questionnaire, or other aspects of the survey process, effect the outcomes of a survey. This should be done by means of embedded experiments, which can be used to detect and quantify possible trend changes in the time series of estimated population parameters due to an alternative survey process. This enables the survey manager to adjust the new process or to accept the expected trend changes. The effects found in this experiment might not have been detected by means of small-scale laboratory experiments since they were provoked by very small groups in the target population, wich could not have been foreseen.

Standard model-based techniques for the analysis of experiments don't allow for the sampling design and the weighting procedure of the survey. The model assumptions that the data are normally and independently distributed are likely to be violated due to the complexity of the sampling design. This results in biased parameter and variance estimates. Moreover, a model-based approach draws inferences about the parameters of a linear regression model. Consequently, tests of treatment effects do not directly apply to the target population parameters of the survey. To cope with these disadvantages, the design-based analysis procedure for embedded experiments proposed in the preceeding chapters is applied to the analysis of the LFS experiment.

Turning to the practical consequences of this field experiment the new questionnaire was adjusted in order to avoid an expected under-reporting of the Self Employed. The new questionnaire might result in a significant decrease of the persons registered at Employment Exchange. This could be the result of the omission of the questions about benefits in the new questionnaire. Since information concerning benefits can be obtained from registrations more efficiently, this didn't result in an adjustment of the new questionnaire.

Chapter 9

Effects of interviewers' workload on the Dutch Labour Force Survey

9.1 Introduction

Response rates to Statistics Netherlands' social surveys declined alarmingly during the nineties. During the same period, Statistics Netherlands' field staff, who collect data by means of computer assisted personal interviewing (CAPI), faced increasing capacity problems, so that more sample addresses were not being visited at all. The question was also raised whether these capacity problems were a significant factor in the increasing non-response of the households that were contacted. The increase in non-response rates as well as the number of households that were not visited has a negative effect on the outcomes of the social surveys. First it results in less precise estimates since the net sample sizes are reduced (i.e. an increase of the design variance). Second it might result in less accurate estimates due to selective response (i.e. the estimates are biased). Not only might the non-response be selective, also the households that are not visited might be selective, since the interviewers are allowed to decide by themselves which addresses are visited and which not. It is not unlikely that interviewers concentrate on neighborhoods where the highest response rates are expected in case they are not able to visit all the addresses assigned to them.

Interviewer workload, i.e. the number of sample cases assigned to an interviewer to complete during the data collection period, affect the level of effort of the interviewer applied to contacting and obtaining cooperation from each sampling unit. Therefore, interviewer workload is one of the design factors of household surveys that influence response rates (Groves and Couper, 1998, Ch. 10). To shed some light on these issues a field experiment has been carried out where we systematically varied the workload of approximately 70 interviewers for the Dutch Labour Force Survey (LFS). Possible effects on response rates, the rates of household that were not visited, as well as the main outcomes of the LFS have been investigated. The design-based methods for the analysis of embedded experiments developed in the preceding chapters have been applied to test hypotheses about parameter estimates of the LFS.

In this chapter the design, analysis and results of this experiment are described. For a description of the design of the LFS, we refer to section 8.2. The objective of this study as well as the experimental design is given in sections 9.2 and 9.3 respectively. Results are presented in section 9.4. The major conclusions are summarized in section 9.5.

9.2 Objective

The objective of the experiment is to investigate the influence of the interviewers' workload on response rates. We confined ourselves to the field staff collecting data by means of CAPI since the capacity problems mainly occur in this group. All interviewers work in an interview area around their place of residence. Interviewers visit the households living on addresses in his or her interview area that are drawn in the samples of the different social surveys. Data are collected in face-to-face interviews, supported by electronic questionnaires on a hand-held computer (i.e. CAPI).

In this experiment workload is defined as the number of addresses per month assigned to an interviewer. Among other factors the total workload of a package of addresses is determined by

1. The time required to complete a questionnaire of a survey.
2. The travel distance between the addresses. This depends on the degree of urbanization of the interview area.
3. The average response rates.

In this experiment we systematically varied with the number of addresses for the LFS, since this is one of Statistics Netherlands' largest continuously conducted social surveys. This entails that all the interviewers visit addresses to collect data for the LFS each month throughout the year. Therefore the LFS is an ideal survey to execute a large field experiment with a group of interviewers that is representative of Statistics Netherlands' field staff.

In the response analysis we controlled for the workload that an interviewer received from other surveys. At this, the interview time required for the different surveys is taken into account. Simple ratios between the time required completing a questionnaire of the different surveys and the LFS were derived. The number off addresses for other surveys weighted with these ratios are used as a covariable in the response analysis.

The objective of this experiment is to test the following hypotheses.

Hypothesis 1: The rate of households that were not visited is not influenced by interviewers' workload versus the rate of households that were not visited is influenced by interviewers' workload.

For this purpose, we divided the gross sample into the following groups:

1. the number of households visited,

2. the number of households not visited due to high workload,

3. the number of households not visited due to other reasons (illness, vacation).

Hypothesis 2: The response rate of the households that were visited is not influenced by interviewers' workload, versus the response rate of the households that were visited is influenced by interviewers' workload.

For this purpose the group of households that were visited is divided into

1. the number of completely responding households,

2. the number of partially responding households,

3. the number of non-responding households (refusal, not at home at least three times, language problems, illness, vacation, special circumstances).

Hypothesis 3: The main outcomes of the LFS are not influenced by interviewers' workload, versus the main outcomes of the LFS are influenced by interviewers' workload. The following three parameters of the LFS were analyzed: the Employed Labour Force, the Unemployed Labour Force and Registered Unemployment.

9.3 Experimental design

The experiment is designed as an RBD where interviewers are the block variables. For each interviewer who participated in the experiment we systematically varied with three different levels of workload, i.e. the number of addresses for the LFS assigned to an interviewer to complete during a data collection period of one month. The degree of urbanization of the interviewer region is a contributory factor of the final workload for a specific number of addresses, which are assigned to an interviewer each month. Therefore the three workload levels in the experiment depended on the degree of urbanization of the interviewer area. To this end the interviewers are classified into three different groups based on the degree of urbanization of their interviewer region. Within each group the average number of addresses per month for the LFS (during 1997) was calculated. In the first group, more then 50% of the interview area is highly urbanised. In the second group, more then 50% of the interview area has a moderate or low degree of urbanization. The interview areas with a moderate or low degree of urbanization are taken together, since the average number of LFS addresses per month in these groups are equal. The third group is a rest group. The interview areas in this group don't consist of one dominating degree of urbanization. These three different groups are indicated as workload area. Each interviewer received three different workloads;

1. low workload, i.e. 75% of the average workload,

2. average workload, i.e. the average number of LFS addresses per month of the interviewers' workload area during 1997,

3. high workload, i.e. 125% of the average workload.

The three different workloads for the three different workload areas are summarized in table 9.1.

Table 9.1: Number of LFS addresses per month for different workload levels in three different workload areas.

workload area	treatment		
	low workload	average workload	high workload
1	23	31	39
2	17	22	29
3	19	26	33

A random sample of 72 interviewers from Statistics Netherlands' field staff was selected to participate in the experiment. First, the interviewers who met the following requirements were selected:

1. at least 2 years experience with the LFS,

2. no extremely low or high response rates,

3. no extremely low or high number of addresses for the LFS per month.

This resulted in a frame of 382 interviewers. A sample of 72 out of these 382 interviewers was drawn by means of proportional stratified simple random sampling where the three workload areas are the strata. Each interviewer received each workload level for a period of three consecutive months. As a result the experiment was carried out in 9 months from September 1998 through June 1999. (December 1999 is excluded from the experiment.) It could be the case that the effect of a change in the monthly workload on response rates becomes manifest only after a certain period of time, since the interviewers must get accustomed to this new workload level. Therefore each first month that the interviewers received a new workload level in the experiment, is excluded from the analysis.

In order to preclude seasonal effects, interviewers were equally divided over the three different treatment levels at each point in time. This is accomplished by distinguishing three different sequences of workload levels. Within each workload area, interviewers are randomly assigned to one of the three sequences. This design is based on a Latin square and is represented in table 9.2. Interviewers as well as respondents didn't know that they participated in an experiment to avoid that they altered their behavior, perhaps even unconsciously.

Table 9.2: Experimental design.

workload area	number of interv.	month								
		1	2	3	4	5	6	7	8	9
1	5	low workl. (23)			average workl. (31)			high workl. (39)		
	5	average workl. (31)			high workl. (39)			low workl. (23)		
	5	high workl. (39)			low workl. (23)			average workl. (31)		
2	15	low workl. (17)			average workl. (23)			high workl. (29)		
	15	average workl. (23)			high workl. (29)			low workl. (17)		
	14	high workl. (29)			low workl. (17)			average workl. (23)		
3	4	low workl. (19)			average workl. (26)			high workl. (33)		
	5	average workl. (26)			high workl. (33)			low workl. (19)		
	4	high workl. (33)			low workl. (19)			average workl. (26)		

The numbers between parentheses are the number of addresses assigned to each interviewer per month for the LFS.

9.4 Results

9.4.1 Effects on response rates

During the experiment 7 interviewers dropped out because they stopped working for Statistics Netherlands, or where not available for several months due to illness, vacation or family circumstances. There were no indications that they dropped out due to their participation with the experiment. Finally 65 interviewers participated 9 months in the experiment. From these 65 interviewers the addresses of the LFS of the last two months of each workload level are used in the analysis of the experiment. This resulted in a gross sample size of 9952 households. In table 9.3 the gross sample is itemized into the number of households visited, the number of households not visited due to higher workload and the number of households not visited due to other reasons (e.g. illness or vacation). It clearly follows that the number of households not visited increases with the workload.

Given the group of households that were visited, we can study the effect of the interviewers' workload on their response rates. Therefore the number of visited households is itemized into the number of fully responding households, partially responding households and non-responding households in table 9.4. It seems that, conditional on the group of households that were visited, there is no effect of the interviewers' workload on their response rates. The number of partially responding households is extremely low since proxy interviews are allowed in the LFS (see section 8.2).

Table 9.3: Gross sample itemized to different levels of treatments and visitation accounts.

treatment (workload)	number of households			
	visited	not visited due to workload	not visited due to other reasons	total
low	2215 (91%)	42 (2%)	178 (7%)	2435
average	2905 (87%)	106 (3%)	328 (10%)	3339
high	3465 (83%)	319 (8%)	394 (9%)	4178
total	8585 (86%)	467 (5%)	900 (9%)	9952

The numbers between parentheses are percentages based on the row totals.

Table 9.4: Visited households itemized to different levels of treatments and response accounts.

treatment (workload)	number of households			
	fully responding households	partially responding households	non-responding households	total
low	1093 (49%)	42 (2%)	1080 (49%)	2215
average	1455 (50%)	61 (2%)	1389 (48%)	2905
high	1699 (49%)	59 (2%)	1707 (49%)	3465
total	4247 (49%)	162 (2%)	4176 (49%)	8585

The numbers between parentheses are percentages based on the row totals.

For each interviewer b we have observed under workload level k, a vector $\mathbf{p}_{bk} = (p_{bk1}, p_{bk2}, p_{bk3})^t$ that represents the proportion of the visitation account over the three different categories households visited (p_{bk1}), households not visited due to workload (p_{bk2}) and households not visited due to other reasons (p_{bk3}) during the last two months that the $b-$th interviewer was exposed to the $k-$th workload level. These proportions are defined as

$$p_{bkl} = \frac{n_{bkl}}{n_{bk+}},$$

where n_{bk1} denotes the number of households visited, n_{bk2} the number of households not visited due to workload and n_{bk3} the number of households not visited due to other reasons. Furthermore $n_{bk+} = \sum_{l=1}^{3} n_{bkl}$ denotes the total number of households assigned to the $b-$th interviewer under workload level k.

In an equivalent way we have observed for each interviewer b under workload level k a vector $\mathbf{q}_{bk} = (q_{bk1}, q_{bk2}, q_{bk3})^t$ that represents the response account of the visited households over the three different categories of fully responding households (q_{bk1}), partially responding households (q_{bk2}) and non-responding households (q_{bk3}). These proportions are defined as

$$q_{bkl} = \frac{m_{bkl}}{n_{bk1}},$$

where m_{bk1} denotes number of fully responding households, m_{bk2} the number of partially responding households and m_{bk3} the number of non-responding households. Note that $n_{bk1} = \sum_{l=1}^{3} m_{bkl}$.

Since these vectors represent proportions of three categories of some whole, they are subject to the constraints

$$\sum_{l=1}^{3} p_{bkl} = 1, \quad p_{bkl} \geq 0, \quad \sum_{l=1}^{3} q_{bkl} = 1, \quad q_{bkl} \geq 0. \tag{9.1}$$

Vectors representing proportions of some whole are called compositions. A convenient way of displaying the variability of 3-part compositions is what is called a ternary diagram, also called a reference triangle or a barycentric coordinate space. A ternary diagram is an equilateral triangle with unit altitude. Each vertex corresponds with one of the three parts of the composition and represents the point where the composition completely consists of that part. The side opposite of a vertex corresponds with the area where this part doesn't occur in the composition. The proportion of each part of the composition equals the perpendicular value from a point in the triangle to the side opposite the vertex of that part. The larger the proportion of a component is, the nearer the representative point in the triangle lies to its corresponding vertex.

For each workload level, the compositions for visitation and response account are plotted in a ternary diagram in figure 9.1 and 9.2, respectively. The figures show that the proportion of households not visited increases with the workload level (figure 9.1) and that, conditional on the group of households that were visited, the proportions fully responding, partially responding and non-responding households are not influenced by the workload level (figure 9.2).

The compositions for the visitation account and the response account can be used to test hypotheses 1 and 2 from section 9.2. The sample space of these 3-part compositions is the 2 dimensional simplex. Aitchison (1986), discusses the statistical analysis of such compositional data. Due to the constraints given by (9.1) compositional vectors have a degenerate and singular covariance structure. Also the assumption of a multivariate normal distribution for compositions observed on a constrained space, as the simplex, is doubtful. Therefore standard statistical methods can lead to distorted inferences. To analyze compositional data with standard multivariate procedures a class of additive logistic normal distributions on the simplex is defined (Aitchison, 1986, Ch.6). A composition is said to have an additive logistic normal distribution, if the logratio transformed compositions have a multivariate normal distribution. This logratio transformation transforms the 3-part compositions from the two-dimensional simplex to the two-dimensional real space. For the time being it is assumed that the categories of each composition are strictly positive. For the compositions of the visitation account the logratio transformation

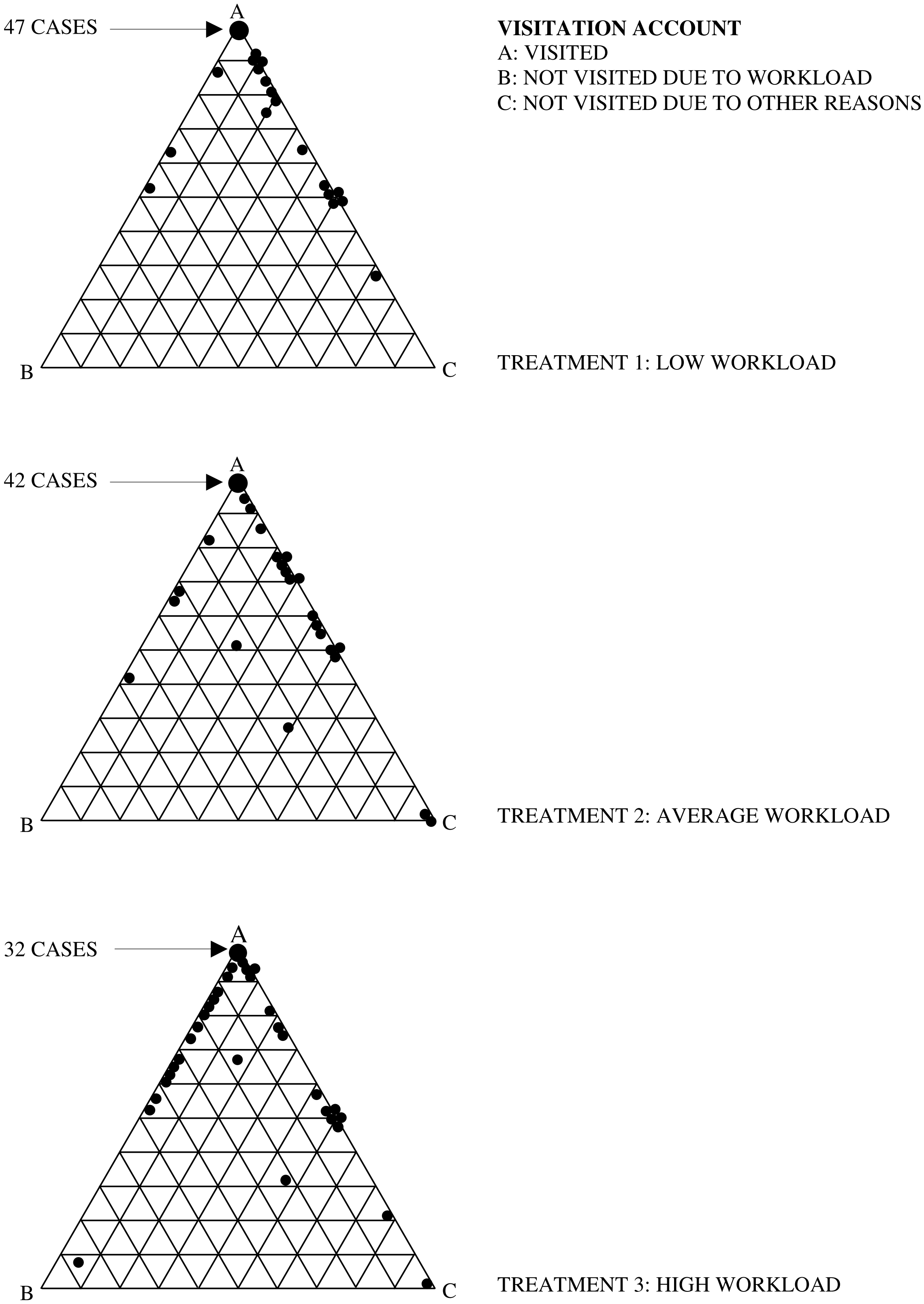

Figure 9.1: *Ternary diagrams of the compositions for visitation account for each workload level.*

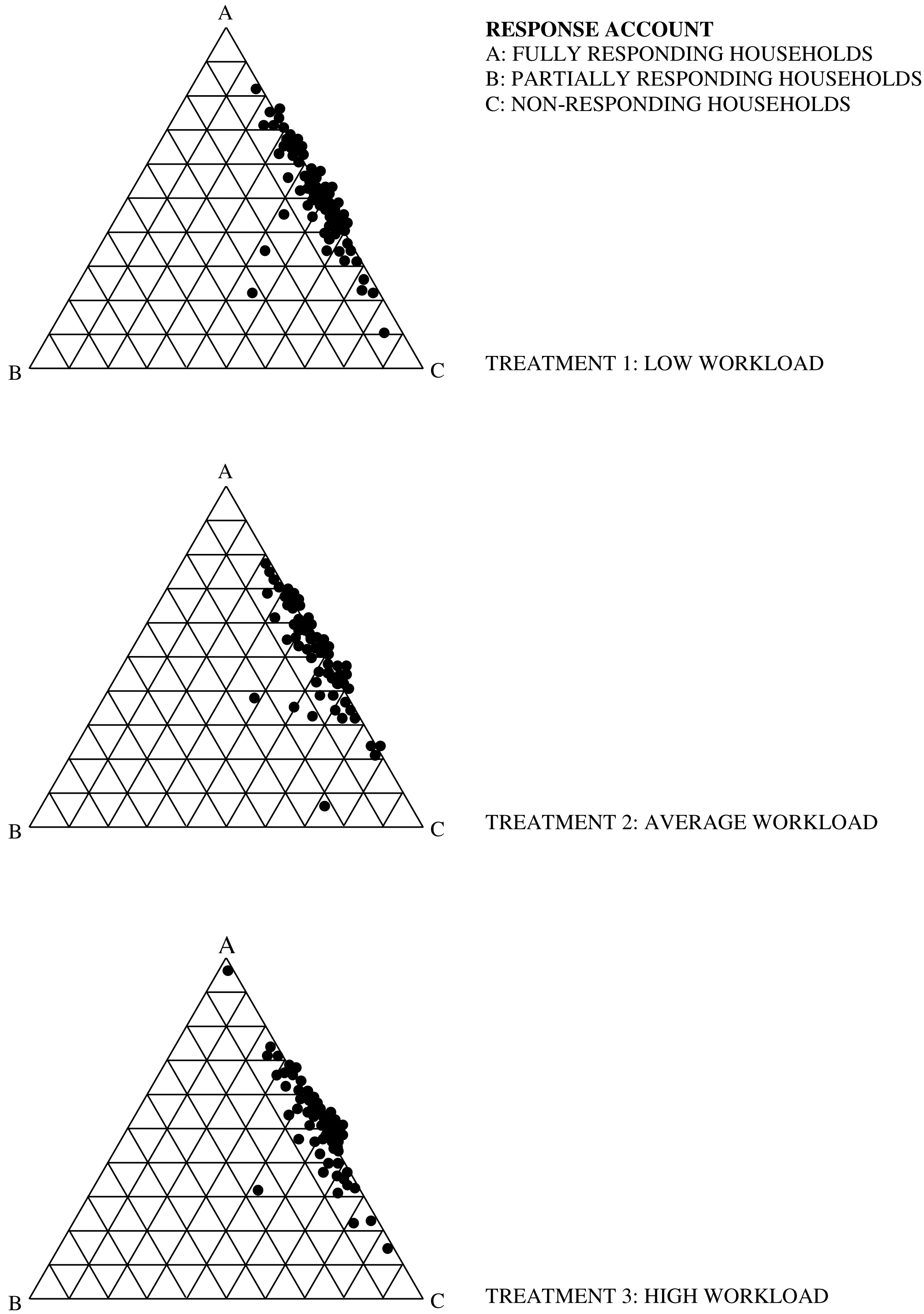

Figure 9.2: *Ternary diagrams of the compositions for response account for each workload level.*

is defined as

$$\mathbf{y}_{bk} = (y_{bk1}, y_{bk2})^t = \left(\log\left(\frac{p_{bk1}}{p_{bk3}}\right),\ \log\left(\frac{p_{bk2}}{p_{bk3}}\right)\right)^t,$$

where "log" denotes the natural logarithm. The logratio transformed compositions of the response account are equivalently defined as

$$\mathbf{z}_{bk} = (z_{bk1}, z_{bk2})^t = \left(\log\left(\frac{q_{bk1}}{q_{bk3}}\right),\ \log\left(\frac{q_{bk2}}{q_{bk3}}\right)\right)^t.$$

If the compositions $\mathbf{p}_{bk}$ and $\mathbf{q}_{bk}$ are independent realizations from a bi-variate logistic normal distribution, then the logratio transformed compositions $\mathbf{y}_{bk}$ and $\mathbf{z}_{bk}$ are independent realizations from a bi-variate normal distribution. After this transformation, standard multivariate techniques can be applied to analyze the compositions for visitation and response accounts.

The logratio transformation, however, is not defined for compositions that are lying on the edge of the simplex, i.e. at least one component of the composition is equal to zero. The additive logistic normal distribution is a very flexible class of distributions to model compositional data, if most of data do not occur on the edges of the simplex. It follows from figures 9.1 and 9.2 for the visitation account as well as the response account, however, that a main part of the data mass lies on the edges of the simplex. Models based on the bi-variate logistic normal distribution will generally fit such data patterns poorly. It might be possible to separate out the zeros by some form of conditional modeling. The data pattern of the visitation account for example can be modeled with a finite mixture for the probability mass on the vertex A and two univariate logistic normal distributions for the data that occur on the A-B and the A-C edges. Then we are in the position to construct a likelihood function corresponding to a compositional data set with such zero components and test parametric hypotheses.

An analysis based on an explicit modeling of the zero components is very laborious. Moreover, it follows directly from the figures 9.1 and 9.2 that workload has a strong influence on the visitation account and nearly no influence on the response account. Therefore we confined ourselves to a quicker but less sophisticated approach. Aitchison (1986, Ch.11) proposed a general procedure, which replaces zeros by a positive value, which is smaller than the smallest recordable value. Let δ denote the maximum rounding-off error of the observed compositions. According to this procedure any $D-$part composition with C zero and $D-C$ non-zero components are replaced by a composition in which the zeros become $\delta(C+1)(D-C)/D^2$ and the positive components are each reduced by $\delta C(C+1)/D^2$. If a compositional data set has been adjusted for zeros, then the analysis under consideration should be subjected to some form of a sensitivity analysis. Therefore we study the effect on the statistical inference for variation in the choice of δ in the zero replacement procedure. Aitchison (1986, Ch.11) proposed to repeat the analysis within a range of $\delta_r/5 \le \delta \le 2\delta_r$, with the δ_r the maximum round-off error. In this application $\delta_r = 1\%$ and we repeated the analyses with δ equal to 0.2%, 0.5%, 1.5% and 2%.

The compositional data for the visitation and response accounts are modeled in two multivariate linear regression models. Then, hypotheses 1 and 2 (section 9.2) are tested with a

MANOVA. The dependent variables are the logratio transformed compositions $\mathbf{y}_{bk}$ and $\mathbf{z}_{bk}$. The following explanatory variables are used to model the compositions:

1. intercept,
2. interviewer (block variable); a factor with 65 levels,
3. workload (treatment); a factor with 3 levels,
4. workload area; a factor with 3 levels,
5. time period; a factor with 3 levels indicating if the composition is observed in the first, second or third three-month period of the experiment,
6. workload obtained from other surveys; a continuous variable, which contains the number of addresses for other surveys weighted with a ratio that indicates the time required to complete these questionnaires as compared to a questionnaire of the LFS,
7. experience of the interviewer; a continuous variable, which contains the number of months that the interviewer has carried out the LFS.

Finally the following model was selected to test hypotheses 1 and 2

$$Intercept + Interviewer + Workload + Workload\ other\ surveys + Time\ period.$$

The $p-$value of workload on the interviewers' visitation and response accounts in both MANOVA's was 0.001 and 0.873 respectively. Therefore it can be concluded that workload has a significant effect on the interviewers' visitation account but no significant effect on the interviewers' response account. It also follows that the inference based on both MANOVA's was not sensitive for the choice of δ in the zero replacement procedure.

9.4.2 Effects on outcomes of the Labour Force Survey

In the preceding section we saw that a higher workload results in an increase of the proportion of households not visited. If the households that are not visited are a selective group with respect to their target parameters of the LFS, then the interviewers' workload might influence the main outcomes of the LFS. Therefore we will test hypothesis 3 from section 9.2. Let $\bar{\mathbf{Y}}^{\alpha} = (\bar{Y}_1^{\alpha}, \bar{Y}_2^{\alpha}, \bar{Y}_3^{\alpha})^t$ where $\bar{Y}_1^{\alpha}$, $\bar{Y}_2^{\alpha}$ and $\bar{Y}_3^{\alpha}$ denotes a parameter of the LFS based on data collected with interviewers assigned to the low, average and high workload, respectively. Now the hypothesis of no treatment effects in the parameters of the LFS can be formulated more formally by (3.14) in section 3.3. According to the experimental design described in section 9.3, three subsample are obtained. The households in subsample s_1 are visited by interviewers assigned to the low workload, in subsample s_2 by interviewers assigned to the average workload and the households in subsample s_3 by interviewers assigned to the high workload. Based on these three subsamples estimates of the parameters of the LFS observed under the different workloads can be obtained and the hypothesis of no treatment effect can be tested using the design-based Wald statistic proposed in chapters 3 and 4.

The analysis procedure is largely equivalent to the procedure described in section 8.4. The analysis is carried out at the household level, since households are the sampling units. Let y^{α}_{ijk} denote the observation obtained from the $i-$th individual from household j observed under treatment k. Then $y^{\alpha}_{jk} = \sum_{i=1}^{n_j} y^{\alpha}_{ijk}$ denotes the household total of the n_j household members aged over 15. These household totals y^{α}_{jk} can be modeled with a measurement error model, which allows for interviewer effects (4.8) with model assumptions (4.6), (4.7), (4.9), (4.10) and (4.11), section 4.2. To incorporate the estimation procedure of the LFS in the analysis of this experiment, the analysis is based on the generalized regression estimator. First-order inclusion probabilities are derived, which take into account that different types of addresses are drawn with different inclusion probabilities (section 8.2). To apply the generalized regression estimator, it is assumed that the household totals of the intrinsic values u_{jk} can be modeled according to linear regression model (4.2) with model assumptions (4.3) and (4.4), section 4.2. For this analysis the following weighting scheme was applied:

$$Age + Sex + Marital\ Status + Region,$$

where the four variables are categorical. To ensure a single weight for each member of the same household, the integrated method for weighting persons and families of Lemaître and Dufour (1987) is applied as described in section 8.4.

Based on the observations obtained from the households in the three susbamples, estimates of $\bar{\mathbf{Y}}^{\alpha}$ are obtained by the generalized regression estimator (4.16) in section 4.3. The variances of the treatment effects are estimated with (4.31), section 4.4, where households are the experimental units and interviewers the block variables. To test the hypothesis of no treatment effects, the design-based Wald statistic (4.35), section 4.5 is applied.

The net sample sizes of the three subsamples are given in table 9.5. These numbers are lower than the numbers of fully responding households given in table 9.4 for two reasons. First, the numbers of fully responding households in table 9.4 are based on the definition of Statistics Netherlands' Data Collection Department. Part of the completed questionnaires of these households are, nevertheless, not of sufficient quality to be used for publication purposes of the LFS. These households are also excluded from this analysis. Second, in one block there were no responding households obtained in one of the three treatments and in an other block there were only three fully responding households obtained in one of the treatments. In this situation we cannot estimate the population variance for this block-treatment combination (see formula (4.31), section 4.4). Moreover, this treatment is extremely underrepresented in this block. Therefore, all the households assigned to these two blocks are also excluded from the analysis.

Table 9.5: Number of households of the subsamples.

workload level	sample size
low	977
average	1340
high	1530

It is assumed that a finite population central theorem holds so that the estimates of a population parameter under the three different treatments have multivariate normal distribution. Then the Wald statistic, under the null hypothesis of no treatment effects, is asymptotically distributed as a chi-squared random variable with two degrees of freedom.

The Employed Labour Force, the Unemployed Labour Force and Registered Unemployment are the main parameters of the LFS, which are analyzed in this experiment. These parameters are expressed as percentages of the population aged between 15 and 65. Results are given in table 9.6.

Table 9.6: Analysis results for three parameters of the Dutch Labour Force Survey.

parameter	workload			Wald	$p-$value
	low	average	high	statistic	
Employed Labour Force	66.1	60.9	63.4	9.15	0.01
Unemployed Labour Force	3.28	3.67	2.79	3.06	0.22
Registered Unemployment	2.13	2.23	1.68	3.28	0.19

Only the three estimates of the Employed Labour Force are significantly different, at a significance level of 0.05. It is not clear why the Employed Labour Force for the average and high workload is lower with respect to the low workload. The observed differences might be a result of selectivity in the households, which are not visited by the interviewer. In the preceding section we saw that the proportion of households that are not visited increases with the workload. Some selectivity in the outcomes of the LFS might be introduced, since interviewers decide by themselves which households are visited and which not. It might also be an artifact, however, since the observed effect of workload on the Employed Labour Force is not monotone.

9.5 Conclusions

In this experiment we first investigated the effect of the interviewers' workload on their visitation account, i.e. the proportion of households visited, not visited due to workload, and not visited for other reasons. We also investigated the influence of interviewers' workload on the response account of the households that were visited, i.e. the proportion of fully, partially and non-responding households. For the visitation account it follows that the proportion of households that were visited declines if the interviewers' workload increases. It appears that the current organization of Statistics Netherlands' fieldwork allows interviewers to decide how many addresses they visit per month. If the number of addresses assigned to an interviewer exceeds the limit of the interviewers' personal capacities, then this results in an increase of households that were not visited. The response account of the households that were visited, however, seems not to be influenced by the interviewers' workload. There are indications that the target parameters the LFS of the households that are not visited are systematically different with respect to the households that are visited. As a result this interviewers' selection mechanism has an undesirable influence on the outcomes of the LFS and probably also on outcomes of other surveys.

From these results, it follows that an improvement of the response rates can't be expected from a reduction of the interviewers' workload. The capacity problems of Statistics Netherlands' field staff, however, resulted in an undesirably large proportion of households that were not visited. Therefore the planning system of the fieldwork has been improved in such a way that the available capacity of the field staff is known at a detailed regional level. Based on this information system the sample sizes of the surveys can be adjusted if necessary. This should minimize the proportion of households that are not visited and the accompanying negative effects on the outcomes of our surveys. Reducing the original sample sizes until this match with the available interviewer capacity, however, is an unsatisfactory and temporary solution. In order to settle these capacity problems more permanently, the free lance status of Statistics Netherlands' field staff will be changed to permanent contracts. This should make the fieldwork more attractive, which enables Statistics Netherlands to recrute more field staff that is obligated to visit a minimum number of households during the data collection period of a survey.

Summary

Large-scale field experiments embedded in ongoing surveys are particularly appropriate to test effects of alternative survey methodologies on finite population parameters of a survey in the daily practice of ongoing sample surveys. In national statistical institutes, such experiments are executed for different purposes. Sample surveys are generally kept unchanged as long as possible in order to construct time series of the estimated population parameters. It is inevitable, however, to adjust these survey processes from time to time. In such situations, embedded experiments can be applied to detect and quantify trend breaks in the time series of estimated population parameters due to necessary adjustments of a survey process before they are implemented as a standard. Another broad area of applications for embedded experiments is the research aimed at the improvement of sample survey processes and the quantification of the bias due to all kinds of non-sampling errors.

In an embedded experiment the sample of an ongoing survey is randomly divided into K subsamples, according to an experimental design. Generally there is one relatively large subsample, which is assigned to the regular survey for regular publication purposes and which is also considered as the control group in the experiment. The remaining, generally smaller, subsamples are assigned to alternative survey approaches (treatments) to be conducted in parallel with the regular survey. Based on the observations obtained in these K subsamples we can estimate the finite population parameters under the K different treatments and test hypotheses about the differences between these parameter estimates.

Usually many people are involved in conducting embedded field experiments. This entails the risk that many sources of extraneous variation can mask or bias the results of such experiments and distort the cause-effect relationship between treatments and observed effects. In order to minimize these risks, principles of experimental designs can be applied in designing embedded field experiments. Parallels between the theory of experimental designs and sampling theory can be exploited in a straightforward manner. The structure of the survey design forms a framework for the design of the experiment, e.g. local control by means of randomization within strata, primary sampling units (PSU's), clusters or interviewers. This results in embedded experiments with a sufficient degree of internal validity, mainly borrowed from statistical methods known from experimental design.

The purpose of embedded experiments is to estimate finite population parameters under

different treatments and to test hypotheses about the differences between population parameter estimates. Therefore the formal analysis of such embedded experiments needs to explicitly take into account the probability structure imposed by the chosen sampling design, used to draw the sample as well as the randomization mechanism of the experimental design used to divide this sample into K subsamples. An analysis based on the assumption of identical and independently distributed data might result in design-biased parameter estimates as well as misleading variance estimates, in a sense that they are incompatible with the parameter and variance estimates of the ongoing survey. Moreover it is not always obvious how the analysis results obtained in such a model-based inference are related to the population parameters of the sample survey. This thesis shows that statistical methods from sampling theory can be used to construct design-based test statistics, which take into account that experimental units are selected from a finite population by some complex sampling design with possibly unequal inclusion probabilities and/or clustering. We start with the formulation of measurement error models. Since embedded experiments are aimed to investigate different sources of non-sampling errors, the concept of measurement error models fits naturally in a general framework for the analysis of such experiments. In these models the observations obtained from the sampling units are assumed to be a realization of the individuals true and intrinsic value confounded with an additive treatment effect. These measurement error models also allow for correlated or biased response due to e.g. random or fixed interviewer effects.

Each subsample can be considered as a two-phase probability sample from a finite population, where the design of the survey sample and the experimental design determine the design of the first and the second phase, respectively. Therefore statistical methods from sampling theory can be used to derive design-unbiased estimators for the population parameters observed under the K different treatments as well as the covariance matrix of the $K-1$ contrasts between these parameter estimates. This results in a design-based Wald-type test statistic. This Wald statistic tests hypotheses about estimates of finite population parameters since it is derived under the different measurement error models, the probability structure imposed by the chosen sampling design, and the randomization mechanism imposed by the experimental design. Expressions for the Wald statistic are worked out for the analysis of completely randomized designs (CRD's) and randomized block designs (RBD's) embedded in general complex sampling designs both using the Horvitz-Thompson estimator and the generalized regression estimator. The use of the generalized regression estimator has the advantage that the weighting scheme of the regular survey is incorporated in the analysis of the experiment. This might increase the precision of the analysis and make the analysis, at least partially, more robust against bias due to selective non-response. The application of design-based procedures in the analysis of embedded experiments enables us to generalize the results of the experiment from the specific sample to the entire survey population from which the sample is drawn. In other words, it guarantees a high degree of external validity that is mainly borrowed from statistical methods known from sampling theory. This leads to statistical procedures for the design and analysis of embedded experiments that combine the internal validity guaranteed by methods from randomized experimentation with the external validity obtained from the theory of randomized sampling.

Since the K subsamples are drawn without replacement from a finite population, there is a nonzero design covariance between the subsample estimates. For the estimation of this covariance, we need an observation under each of the K treatments for each individual. Since an individual is assigned to only one of the K treatments, these paired observations are not available. As a result we cannot estimate the covariance matrix of the K subsample estimates. The problem of the missing values is circumvented by deriving an estimation procedure for the covariance matrix of the $K-1$ contrasts between the K subsample estimates. For the generalized regression estimator an approximately design-unbiased estimator for this covariance matrix is derived, which has the structure as if the subsamples are drawn independently from each other. The probability structure of the sampling design is incorporated in the variance estimation procedure, by weighting the residuals of the generalized regression estimator with a factor containing the first-order inclusion probabilities of the sampling design. No second-order inclusion probabilities are required, which simplifies the analysis considerably. This remarkable result is mainly a consequence of the fact that we estimate variances of contrasts between subsample means, the assumption that the measurement errors between individuals are independent, the assumption of additive treatment effects in the measurement error model, and the condition that the weighting scheme of the generalized regression estimator at least utilizes the size of the finite population as auxiliary information.

The variance estimation procedure for the Horvitz-Thompson estimator is almost equivalent to the variance estimation procedure for the generalized regression estimator. The minimum use of auxiliary information in the weighting scheme of the generalized regression estimator for which these results hold true is the common mean model (Särnal et al., 1992, ch. 7.4). Under this weighting scheme the generalized regression estimator equals the extended Horvitz-Thompson estimator or the ratio estimator for a population mean, originally proposed by Hájek (1971). For the regular Horvitz-Thompson estimator a design-unbiased estimator for the covariance matrix of the contrasts between the subsample means consists of two components. The leading component is a diagonal covariance matrix, which has the same structure as the covariance matrix of the generalized regression estimator. Instead of residuals, observations are weighted with a factor containing the first-order inclusion probabilities of the sampling design. The second component is a non-diagonal covariance matrix that involves second-order inclusion probabilities. It can be interpreted as the product of the variance of the estimated population size and a matrix containing the cross products of the estimated treatment effects. In many practical situations this second component will be zero or negligible. There are several special cases where the Horvitz-Thompson estimator coincides with the extended Horvitz-Thompson estimator. For example, in the case of a CRD embedded in a simple random sampling design, a CRD embedded in a stratified simple random sampling design with proportional allocation, and an RBD embedded in a stratified simple random sampling design under all kinds of allocations with strata as block variables. In these situations the non-diagonal covariance component with second-order inclusion probabilities for the Horvitz-Thompson estimator is equal to zero. In other situations we can ignore this component, since it will generally be negligible, or use the extended Horvitz-Thompson estimator as an alternative.

There are several special cases where the Wald statistic coincides with the test statistics known from more standard model-based analysis procedures. A CRD analyzed with the Horvitz-Thompson estimator and the pooled variance estimator of the K subsample means gives rise to a Wald statistic, which, up to a factor $K-1$, equals the $F-$statistic of the ANOVA of the one-way layout, where the observations are weighted with a factor containing the first-order inclusion probabilities of the sampling design. In case of a self-weighted sampling design it follows that the Wald statistic divided by a factor $K-1$ equals the $F-$statistic of the ANOVA of the one-way layout. In an equivalent way, for an RBD embedded in a self-weighted sampling design where treatments are allocated proportionally over the blocks, we can distinguish two pooled variance estimators which gives rise to a Wald statistic, which, up to a factor $K-1$, equals the $F-$statistic of the ANOVA for a two-way layout with and without interactions. For the analysis of the embedded two-treatment experiment a design-based version of the $t-$statistic can be derived as a special case from the Wald statistic. In the case of a self-weighted sampling design, this test statistic reduces to the Welch's $t-$statistic, regardless the second-order inclusion probabilities of the sampling design. If the pooled variance estimator is applied under a self-weighted sampling design, then the test statistic reduces to the usual model-based $t-$statistic.

The application of an RBD might be efficient if the individuals in the sample can be divided into groups or blocks with relatively homogeneous observations, e.g. individuals from the same stratum, primary sampling unit (PSU) or cluster, or individuals assigned to the same interviewer. The efficiency of blocking can be quantified by comparing the variance of the estimated treatment effects under a CRD and an RBD. It follows that if an experiment is embedded in stratified, two-stage or cluster sampling designs, the variance between strata, PSU's or clusters can be eliminated from the variance of the estimated treatment effects if these sampling structures are used as block variables in an RBD. For a stratified sampling design this implies that the efficiency gain of stratification is nullified in the analysis of an embedded experiment if this experiment is designed as a CRD instead of an RBD with strata as block variables.

If an experiment is embedded in a two-stage sample, then we can choose between the PSU's or the SSU's as experimental units to randomize over the treatments. Sampling units within PSU's are generally more homogeneous than sampling units from different PSU's. Consequently, it will generally be efficient to design such experiments as an RBD with PSU's as block variables and to use the secondary sampling units (SSU's) as experimental units. This also holds for cluster samples, if clusters are homogeneous. Using clusters as block variables, however, implies that clusters are not completely observed in the different subsamples. As a result, a nonzero within-cluster variance component appears in the variance of the estimated treatment effects. Therefore it might be efficient to use clusters as experimental units if the variance between clusters is small and the variance within clusters large.

There can be a conflict between the efficiency gain obtained by blocking on PSU's or clusters and the practical advantages of using PSU's or clusters as experimental units. Consider for example sampling designs where households are PSU's or clusters. Using households as block variables will be efficient if the observations of household members are highly correlated (i.e. a high degree of homogeneity within clusters). Due to the nature of the treatments, however, it might be infeasible to apply different treatments within the same household.

Under a measurement error model with fixed or random interviewer effects, it might be efficient to use interviewers as block variables in an RBD. The efficiency gain in the estimated treatment effects due to the application of an RBD with interviewers as block variables instead of a CRD consists of two components. The first component concerns the variance of random and/or fixed interviewer effects, as well as the covariance between fixed interviewer effects and the block means of the target variables. The second component concerns the variance between the means of the target variables of the individuals assigned to the same interviewer. Especially in the case of CAPI where interviewers work in separate interviewer areas around their place of residence, this second term might be substantial.

The explicit modeling of non-sampling errors like interviewer effects in a measurement error model emphazes the important role of concepts like randomization or local control by means of blocking on interviewers. Under a measurement error model with fixed interviewer effects, the subsample means are biased with a weighted average of these interviewer effects. In the case of an RBD, each subsample mean is biased with the same interviewer effect. Also in a CRD there is a nonzero probability that the individuals of each interviewer are assigned to each of the K treatments of the experiment. This implies that the design expectation of each subsample mean is biased with the same weighted average of interviewer effects. This bias, consequently, cancels out in the contrasts between these subsample means under RBD's as well as CRD's.

Despite these theoretical arguments there may be practical reasons to consider experiments where interviewers participate in only one of the K treatments. For example, it might be efficient to avoid that interviewers know that they are participating in an experiment or to avoid that other measurement errors are introduced since interviewers confuse the interviewer instructions for the different treatments. Under a measurement error model with fixed interviewer effects, each susbsample mean is biased with a weighted average of a different set of interviewers, which doesn't necessarily cancel out in the contrasts between the subsample means. In such situations the assignment of interviewers to the treatments should be random to preclude systematical bias of the treatments.

The methods proposed in this thesis are applied in the analysis of two different experiments embedded in the Dutch Labour Force Survey. The first experiment was designed to test the effect of a new questionnaire on the main outcomes of this survey. In the second experiment, effects of interviewers' workload on the repsonse rates and the outcomes of the LFS were investigated. These examples illustrate several aspects. First that the application of standard model-based methods may distort the analysis. Secondly that non-sampling errors can have a significant impact on the accuracy of the outcomes of a sample survey. This emphasizes the relevance of paying much attention to the improvement of sample survey processes in order to minimize non-sampling errors. It also emphasizes the relevance of quantifying trend breaks due to necessary adjustments or a total redesign of a sample survey process.

Samenvatting

Het doel van statistische bureaus zoals het Centraal Bureau voor de Statistiek (CBS) is het beschrijven van uiteenlopende verschijnselen en ontwikkelingen in de samenleving. Dit gebeurt door aan de hand van steekproefonderzoeken van diverse kenmerken populatietotalen, gemiddelden of fracties te schatten. Op het CBS worden regelmatig grootschalige veldexperimenten, die zijn ingebouwd of ingebed in lopende steekproefonderzoeken, uitgevoerd om te onderzoeken wat het effect is van een of meerdere aanpassingen in het surveyproces op de uitkomsten van dergelijke steekproefonderzoeken. Steekproefonderzoeken worden doorgaans zo lang mogelijk ongewijzigd gelaten. Hierdoor ontstaat de mogelijkheid om van de gepubliceerde kenmerken tijdreeksen op te bouwen die ontwikkelingen door de tijd heen zo goed mogelijk beschrijven. Het blijft echter onvermijdelijk om van tijd tot tijd het surveyproces op één of meerdere onderdelen aan te passen. In dergelijke situaties kan door middel van een grootschalig veldexperiment worden gekwantificeerd wat het effect is van één of meerdere aanpassingen in het surveyproces op de uitkomsten van het steekproefonderzoek. Op deze wijze wordt voorkomen dat ontwikkelingen die worden beschreven door middel van tijdreeksen worden verstoord. Daarnaast worden experimenten uitgevoerd in het kader van het onderzoek naar de verbetering van kwaliteit en efficiëntie van het surveyproces.

Bij een veldexperiment wordt de steekproef van het betreffende onderzoek door middel van een proefopzet in K deelsteekproeven gedeeld. Doorgaans is er één relatief grote deelsteekproef die wordt gebruikt voor het reguliere onderzoek en tegelijk dient als de controlegroep in het experiment. De overige, meestal relatief kleinere, deelsteekproeven worden gebruikt om de effecten van één of meerdere aanpassingen op het reguliere surveyproces te onderzoeken. Op grond van de deelsteekproeven worden K schattingen van de belangrijkste doelparameters van het steekproefonderzoek gemaakt, die zijn gebaseerd op waarnemingen verkregen onder de K verschillende behandelingen van het experiment. Op basis hiervan kunnen hypothesen worden getoetst omtrent mogelijke verschillen tussen deze parameterschattingen.

Bij de uitvoering van veldexperimenten zijn er doorgaans een groot aantal externe factoren die het experiment kunnen verstoren en waarvoor moeilijk gecontroleerd kan worden. Dit wordt met name veroorzaakt doordat bij het dataverzamelingsproces meestal veel mensen betrokken zijn. Hierdoor bestaat het gevaar dat de interne validiteit van het experiment, dat wil zeggen de mate waarin het waargenomen effect kan worden toegeschreven aan de behandeling, laag is.

Vanuit de statistische proeftechniek zijn ten behoeve van het ontwerpen van experimenten een aantal methoden bekend die de nauwkeurigheid waarmee behandelingseffecten worden geschat kunnen verhogen en kunnen voorkomen dat behandelingseffecten systematisch worden vertekend door die externe factoren. Bijvoorbeeld randomiseren, het toepassen van lokale controle door middel van blokken, het blind of dubbel blind uitvoeren van proeven en het simultaan testen van meerdere behandelingen in één factoriële proef. De interne validiteit van in steekproefonderzoeken ingebouwde veldexperimenten wordt zo goed mogelijk gewaarborgd door dergelijke methoden uit de statistische proeftechniek toe te passen in de ontwerpfase van het experiment. Het steekproefontwerp biedt daartoe diverse aanknopingspunten, bijvoorbeeld door strata, clusters, primaire steekproefeenheden of interviewers te gebruiken als blokvariabelen in de proefopzet.

Het doel van in steekproefonderzoeken ingebouwde veldexperimenten is het schatten van eindige populatieparameters die zijn waargenomen onder de verschillende surveyimplementaties of behandelingen en het toetsen van hypothesen met betrekking tot verschillen tussen deze parameterschattingen. Bij een dergelijke analysedoelstelling moeten de toetsingsgrootheden betrekking hebben op de parameterschattingen van het steekproefonderzoek. Dit betekent, dat bij de analyse expliciet rekening moet worden gehouden met het kansmechanisme van het steekproefontwerp waarmee de steekproef uit de eindige populatie is getrokken en het kansmechanisme van het experiment waarmee de steekproef in K deelsteekproeven is verdeeld. Een standaard model-based analyse, gebaseerd op de veronderstelling dat de data identiek en onafhankelijk verdeeld zijn, kan resulteren in verkeerde parameter- en variantieschattingen. Analyseresultaten gebaseerd op dergelijke model-based methoden zijn doorgaans niet of moeilijk vergelijkbaar met de parameter- en variantieschattingen zoals die uit het lopende steekproefonderzoek komen. In dit proefschrift zijn design-based toetsingsgrootheden voor het analyseren van veldexperimenten ontwikkeld waarbij rekening wordt gehouden met het steekproefontwerp waarmee experimentele eenheden zijn getrokken uit de eindige doelpopulatie van het steekproefonderzoek. De observaties uit de steekproef worden gemodelleerd met behulp van een meetfoutmodel. Deze modellen veronderstellen dat iedere observatie het resultaat is van een intrinsieke waarde van de experimentele eenheid, een additief behandelingseffect, een random of vast interviewereffect en een meetfout. Het doel van in steekproefonderzoeken ingebouwde veldexperimenten is doorgaans het onderzoeken van het effect van niet-steekproeffouten op de uitkomsten van het steekproefonderzoek. Om die reden past het concept van meetfoutmodellen goed in een design-based theorie voor het analyseren van veldexperimenten.

Iedere deelsteekproef kan worden opgevat als een tweefasensteekproef uit de eindige doelpopulatie van het steekproefonderzoek, waarbij de eerste fase is gedefinieerd door het steekproefontwerp en de tweede fase door de proefopzet. Vervolgens kunnen statistische methoden uit de steekproeftheorie worden gebruikt voor het afleiden van design-zuivere schatters voor de populatieparameters waargenomen onder de K verschillende behandelingen en voor de covariantiematrix van de $K-1$ contrasten tussen deze parameterschatters. Dit resulteert in een design-based Wald-toets waarmee hypothesen met betrekking tot eindige populatieparameters kunnen worden getoetst. Uitdrukkingen voor deze Wald-toetsingsgrootheden zijn afgeleid voor volledig gewarde proeven (CRD's) en blokkenproeven (RBD's) ingebouwd in complexe steekproefontwerpen, geanalyseerd met de Horvitz-Thompson-schatter of de gegeneraliseerde regressieschatter. Design-based $t-$toetsingsgrootheden voor de analyse van het tweedeelsteekproevenprobleem volgen als speciale gevallen van de resultaten van de Wald-toetsingsgrootheden. De gegeneraliseerde regressieschatter heeft als voordeel dat bij de analyse van het experiment rekening kan worden gehouden met het weegschema van het steekproefonderzoek. Het gebruik van hulpinformatie door middel van de gegeneraliseerde regressieschatter kan de precisie van de analyse verhogen en het kan de analyse meer robuust maken tegen vertekening ten gevolge van selectieve nonrespons. Door het toepassen van design-based analysemethoden wordt generalisatie van resultaten waargenomen in het experiment naar de eindige doelpopulatie van het steekproefonderzoek beter mogelijk. Met andere woorden, dankzij het toepassen van design-based technieken uit de steekproeftheorie kan een sterke externe validiteit van veldexperimenten worden verkregen. Dit leidt tot statistische procedures voor het ontwerpen en analyseren van in lopende steekproefonderzoeken ingebouwde veldexperimenten die een zo hoog mogelijke interne validiteit, gewaarborgd via methoden uit de statistische proeftechniek, combineren met een zo hoog mogelijke externe validiteit, gewaarborgd via methoden uit de steekproeftheorie.

Omdat de K deelsteekproeven zijn getrokken zonder terugleggen uit een eindige populatie, zijn de covarianties tussen de deelsteekproefschatters ongelijk aan nul. Voor het schatten van deze covarianties is het noodzakelijk dat voor ieder individu een observatie onder iedere behandeling is verkregen. Deze zogenaamde gepaarde waarnemingen zijn echter niet beschikbaar omdat ieder individu in het experiment is toegewezen aan slechts een van de K behandelingen. Hierdoor is het niet mogelijk om een schatter voor de covariantiematrix van de K deelsteekproefschatters af te leiden. Het probleem van deze ontbrekende waarnemingen is omzeild door een schatter voor de covariantiematrix van de $K-1$ contrasten tussen de populatieparameterschatters af te leiden. Voor de gegeneraliseerde regressieschatter is een bij benadering design-zuivere schatter voor deze covariantiematrix afgeleid die een structuur heeft alsof de waarnemingen onderling onafhankelijk verdeeld zijn. In de variantieschatters wordt rekening gehouden met het steekproefontwerp door middel van een herweging van de waarnemingen met een factor die is gebaseerd op de eerste orde insluitkans van het steekproefontwerp. Omdat voor de variantieschatters geen tweede orde insluitkansen nodig zijn, kunnen de variantieberekeningen aanmerkelijk worden vereenvoudigd. Dit resultaat wordt onder andere verkregen doordat varianties van contrasten tussen parameterschattingen worden berekend, de aanname wordt gemaakt dat meetfouten tussen individuen onderling onafhankelijk zijn, de behandelingseffecten additief zijn en ten minste het populatietotaal als hulpinformatie wordt gebruikt in het weegschema van de gegeneraliseerde regressieschatter.

De varianties van de Horvitz-Thompson-schatter komen in grote lijnen overeen met die van de gegeneraliseerde regressieschatter. Het populatietotaal is het minimum aan hulpinformatie dat in het weegschema van de gegeneraliseerde regressieschatter moet zijn opgenomen om een variantieschatter af te kunnen leiden, die bij benadering design-zuiver is. Onder dit weegschema correspondeert de gegeneraliseerde regressieschatter met de extended Horvitz-Thompson-schatter, ook wel aangeduid als Hàjeks ratioschatter voor een populatiegemiddelde. Bij de Horvitz-Thompson-schatter bestaat een schatter voor de covariantiematrix uit twee componenten. De hoofdcomponent is een diagonaalmatrix die dezelfde vorm heeft als die van de gegeneraliseerde regressieschatter. De tweede component is een niet-diagonaalmatrix waarbij ook tweede orde insluitkansen noodzakelijk zijn. Deze component kan worden geïnterpreteerd als het product van de variantie van het geschatte populatietotaal en de matrix die de kruisproducten van de behandelingseffecten bevat. In veel situaties is deze tweede component gelijk aan nul of verwaarloosbaar ten opzichte van de hoofdcomponent. In een aantal situaties komt de Horvitz-Thompson-schatter overeen met de extended Horvitz-Thompson-schatter. Bijvoorbeeld in het geval van een CRD ingebed in een enkelvoudig aselect steekproefontwerp, of een gestratificeerd steekproefontwerp met proportionele allocatie, of een RBD ingebed in een gestratificeerd enkelvoudig aselect steekproefontwerp onder willekeurige allocaties. In deze situaties is de niet-diagonaalmatrix in geval van de Horvitz-Thompson-schatter gelijk aan nul.

Er zijn een aantal situaties waarin de Wald-toetsingsgrootheid overeenkomt met toetsingsgrootheden uit meer standaard model-based analyseprocedures. In een aantal specifieke situaties kan worden aangetoond dat de Wald-toetsingsgrootheid, op een factor $K-1$ na, overeenkomt met de $F-$toetsingsgrootheid van een ANOVA voor proefopzetten met één of twee factoren. Op overeenkomstige wijze zijn er voor het tweedeelsteekproevenprobleem een aantal specifieke situaties waar de design-based $t-$toetsingsgrootheid overeenkomt met Welch's $t-$toetsingsgrootheid en de standaard model-based $t-$toetsingsgrootheid.

Het toepassen van een RBD is efficiënt indien de steekproefeenheden kunnen worden ingedeeld in min of meer homogene groepen, ook wel blokken genaamd. De variantie tussen strata, primaire steekproefeenheden of clusters van het steekproefontwerp kan uit de variantie van de behandelingseffecten worden gehaald door deze variabelen van het steekproefontwerp te gebruiken als blokvariabelen in een RBD. Hieruit volgt bijvoorbeeld dat de variantiereductie die wordt bereikt met behulp van stratificatie teniet wordt gedaan indien een experiment ingebouwd in een gestratificeerd steekproefontwerp wordt ontworpen als een CRD.

Indien een experiment wordt ingebed in een tweetrapssteekproef kunnen zowel de primaire als de secundaire steekproefeenheden worden gebruikt als de experimentele eenheden waarover gerandomiseerd wordt. In veel steekproefonderzoeken zijn de secundaire steekproefeenheden die tot dezelfde primaire steekproefeenheid behoren homogener dan de secundaire steekproefeenheden afkomstig van verschillende primaire steekproefeenheden. Om die reden is het doorgaans efficiënt om een in een tweetrapssteekproef ingebouwd veldexperiment te ontwerpen als een RBD, waarbij de primaire steekproefeenheden worden gebruikt als blokvariabelen en de secundaire steekproefeenheden als experimentele eenheden. Hetzelfde geldt voor clustersteekproeven indien de elementen behorende tot hetzelfde cluster homogeen zijn ten opzichte van elementen behorende tot verschillende clusters. Het gebruiken van clusters als blokvariabelen impliceert echter dat clusters niet meer integraal worden waargenomen, waardoor de variantie bij de analyse van het experiment wordt verhoogd met een binnen-clustervariantiecomponent. Het kan daarom efficiënt zijn om clusters als experimentele eenheden te gebruiken indien de binnen-clustervariantie groot is ten opzichte van de tussen-clustervariantie.

Tegenover de theoretische voordelen van de variantiereductie, die wordt verkregen door een experiment te ontwerpen als een RBD waarbij primaire steekproefeenheden of clusters worden gebruikt als blokvariabelen, staan een aantal praktische bezwaren. Bijvoorbeeld indien de primaire steekproefeenheden of clusters corresponderen met huishoudens kan het, afhankelijk van de aard van de behandelingen, erg moeilijk worden om binnen één huishouden de verschillende behandelingen van het experiment toe te passen.

Onder een meetfoutmodel met vaste of random interviewereffecten kan het efficiënt zijn om interviewers te gebruiken als blokvariabelen in een RBD. De variantiereductie die met deze proefopzet gepaard gaat, bestaat uit twee componenten. Ten eerste wordt de variantie van de vaste en random interviewereffecten uit de variantie van de behandelingseffecten geëlimineerd. Ten tweede wordt de tussen-blokvariantie van de doelvariabelen uit de variantie van de behandelingseffecten verwijderd. Met name indien de dataverzameling wordt uitgevoerd door interviewers die steekproefpersonen of -huishoudens in hun eigen regionale werkgebied thuis bezoeken, kan dit een substantiële component zijn. Een nadeel van het gebruik van interviewers als blokvariabelen is dat iedere interviewer alle behandelingen van het experiment moet uitvoeren. Afhankelijk van de aard van de behandelingen kan het experiment worden verstoord doordat de interviewer behandelingen verwart dan wel een voorkeur voor bepaalde behandelingen ontwikkelt. Hierdoor kan systematische vertekening in de parameterschattingen worden geïntroduceerd. Dit kan worden voorkomen door interviewers aselect aan een behandeling toe te wijzen en er voor te zorgen dat de interviewers niet weten dat zij in een experiment participeren.

De methoden die in dit proefschrift zijn ontwikkeld, zijn toegepast bij de analyse van twee experimenten, beiden ingebouwd in de steekproef van de Enquête beroepsbevolking (EBB). Het eerste experiment is uitgevoerd om te onderzoeken wat het effect is van een nieuwe vragenlijst op de belangrijkste doelvariabelen van de EBB. Het tweede experiment had als doel om te onderzoeken wat het effect is van de werklast van interviewers, dat wil zeggen het aantal steekproefadressen dat een interviewer per maand krijgt toegewezen, op de respons en de belangrijkste doelvariabelen van de EBB. Deze twee toepassingen illustreren dat het gebruik van standaard model-based analysemethoden kan leiden tot misleidende analyseresultaten. Verder illustreren beide experimenten de gevoeligheid van uitkomsten van steekproefonderzoeken voor niet-steekproeffouten, zoals vragenlijsteffecten. Dit onderstreept de noodzaak om enerzijds continu onderzoek te doen naar kwaliteitsverbetering van surveyprocessen teneinde de negatieve effecten van niet-steekproeffouten te minimaliseren en anderzijds de noodzaak om effecten van aanpassingen in het surveyproces aan de hand van veldexperimenten te kwantificeren om de continuïteit van opgebouwde tijdreeksen te waarborgen.

Bibliography

[1] Aitchison, J. (1986). The Statisitcal Analysis of Compositional Data. Chapman & Hall, London.

[2] Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. Journal of Official Statistics, Vol. 4, No. 3, pp. 251-260.

[3] Bethlehem, J.G. and Keller, W.G. (1987). Linear Weighting of Sample Survey Data. Journal of Official Statistics, vol. 3, no. 2, pp. 141-153.

[4] Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. and Sudman, S. (1991). Measurement Errors in Surveys. John Wiley & Sons, New York.

[5] Cochran, W.G. (1977). Sampling Techniques. John Wiley & Sons, New York.

[6] Cochran, W.G. and Cox, G.M. (1957). Experimental Designs. John Wiley & Sons, New York.

[7] Efron, B. and R.J. Tibshirani, (1993). An Introduction to the Bootstrap. Chapman & Hall, New York.

[8] Fienberg, S.E. and Tanur, J.M. (1987). Experimental and Sampling Structures: Parallels Diverging and Meeting. International Statistical Review, 55, 1, pp. 75-96.

[9] Fienberg, S.E. and Tanur, J.M. (1988). From the inside out and the outside in: Combining experimental and sampling structures. The Canadian Journal of Statistics, Vol. 16, No. 2, pp. 135-151.

[10] Fienberg, S.E. and Tanur, J.M. (1989). Combining Cognitive and Statistical Approaches to Survey Design. Science, Vol. 243, pp 1017-1022.

[11] Fienberg, S.E. and Tanur, J.M. (1996). Reconsidering the Fundemental Contributions of Fisher and Neyman on Experimentation and Sampling. International Statistical Review, Vol. 64, No. 3, pp. 237-253.

[12] Fisher, R.A. (1935). The Design of Experiments. Oliver & Boyd. Edinburgh.

[13] Forsman, G. (1989). Early Survey Models and their use in Survey Quality Work. Journal of Official Statistics, Vol. 5, No. 1, pp. 41-55.

[14] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). Bayesian Data Analysis. Chapman & Hall, London.

[15] Groves, R.M. and M.P. Couper (1998). Non-response in household interview surveys. John Wiley & Sons, New York.

[16] Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. Mathematical Institute of the Hungarian Academy of Sciences, No. 5, pp. 361-374.

[17] Hájek, J. (1971). Comment on a paper by D. Basu. In: Godambe, V.P., and Sprott, D.A. (eds.) Foundations of Statistical Inference. Toronto: Holt, Rinehart and Winston, p. 236.

[18] Hansen, M.H. and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. Annals of Mathematical Statistics. Vol. 14, pp. 333-362.

[19] Hartley, H.O. and Rao, J.N.K. (1978). Estimation of nonsampling variance components in sample surveys. In N.K. Namboodiri (ed.), Survey Sampling and Measurement, Academic Press, New York, pp. 35-43.

[20] Hilbink, K., K. van Berkel and J. van den Brakel, (2000). Methodology of the Dutch Labour Force Survey, 1987-1999. Research paper, BPA nr.: 2297-00-RSM, Department of Statistical Methods, Statistics Netherlands, Heerlen.

[21] Hinkelmann, K. and Kempthorne, O. (1994). Design and Analysis of Experiments. John Wiley & Sons, New York.

[22] Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association. Vol. 47, pp. 663-685.

[23] Huisman, A.E. (1998). Monte Carlo methods for the analysis of experiments embedded in complex sampling designs (unpublished report in Dutch). Final project for Master thesis at the Wageningen Agricultural University.

[24] Huys, H. and Rooduijn, J. (1994). A new survey on justice and security. Netherlands Official Statistics, Vol. 9, spring 1994, pp. 47-51.

[25] Johnson, N.L. and Kotz, S. (1969). Discrete distributions. John Wiley & Sons, New York.

[26] Kempthorne, O. (1952). The design and analysis of experiments. John Wiley & Sons, New York.

[27] Kish, L. (1965). Survey Sampling. John Wiley & Sons, New York.

[28] Kish, L. and Frankel, M.R. (1974). Inference from Complex Samples. Journal of the Royal Statistical Society, B, 36, pp 1-37.

[29] Lehmann, E.L. (1975). Nonparametrics: Statistical Methods Based on Ranks. McGraw-Hill, New York.

[30] Lehtonen, R. and Pahkinen, E.J. (1995). Practical Methods for Design and Analysis of Complex Surveys. John Wiley & Sons, New York.

[31] Lemaître, G. and Dufour, J. (1987). An Integrated Method for Weighting Persons and Families. Survey Methodology, Vol. 13, pp. 199-207.

[32] Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, Vol. 109, pp. 325-370.

[33] Miller, R.G. (1986). Beyond ANOVA, Basics of Applied Statistics. John Wiley & Sons, New York.

[34] Morisson, D.F., (1990). Multivariate Statistical Methods. McGraw-Hill, Singapore.

[35] Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society, Vol. 97, pp. 558-625.

[36] O'Muircheartaigh, C. (1997). Casm: Successes, [Failures,] and Potential. Methodology Institute London School of Economics and Political Science.

[37] Nieuwenbroek, N.J. (1993). An Integrated Method for Weighting Characteristics of Persons and Households using the Generalized Regression Estimator. Research paper, BPA nr.: 8445-93-M1, Department of Statistical Methods, Statistics Netherlands, Heerlen. John Wiley & Sons, New York.

[38] Särndal, C.E. and Swensson, B. (1987). A general view of estimation for two phases of selection with application to two-phase sampling and nonresponse. International Statistical Review, Vol 55, No. 3, pp. 279-294.

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

[40] Searle, S.R. (1971). Linear Models. John Wiley & Sons, New York.

[41] Scheffé, H. (1959). The Analysis of Variance. John Wiley & Sons, New York.

[42] Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap. Springer Verlag, New York.

[43] Skinner, C.J., Holt, D. and Smith, T.M.F. (1989). Analysis of Complex Surveys. John Wiley & Sons, New York.

[44] Van den Brakel, J.A. (2001). Effects of interviewers' workload on the Dutch Labour Force Survey. Research paper (BPA no. 1817-01-TMO), Department of Statistical Methods, Statistics Netherlands.

[45] Van den Brakel, J.A. and Binder, D. (2000). Variance estimation for experiments embedded in complex sampling schemes. Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, Indianapolis 2000, American Statistical Association, pp. 805-810.

[46] Van den Brakel, J.A., Luppes, M. and Moritz, G. (1995). Testing effects of informed consent in the National Travel Survey. Netherlands Official Statistics, Vol. 10, winter 1995, pp. 37-43.

[47] Van den Brakel, J.A. and Renssen, R.H. (1995). Application of experimental designs at statistical bureau's. Research paper (BPA no.: 5443-95-RSM), Department of Statistical Methods, Statistics Netherlands.

[48] Van den Brakel, J.A. and Renssen, R.H. (1996a). Application of experimental designs in survey methodology. Proceedings of the International Conference on Survey Measurement and Process Quality, American Statistical Association, pp. 151-156.

[49] Van den Brakel, J.A. and Renssen, R.H. (1996b). The analysis of the two-sample problem embedded in complex sampling designs. Research paper (BPA no. 1027-96-RSM), Department of Statistical Methods, Statistics Netherlands.

[50] Van den Brakel, J.A. and Renssen, R.H. (1996c). The analysis of completely randomized designs embedded in complex sampling designs. Research paper (BPA no. 8090-96-RSM), Department of Statistical Methods, Statistics Netherlands.

[51] Van den Brakel, J.A. and Renssen, R.H. (1997). Design and analysis of experiments embedded in complex sample surveys. Proceedings of Statistics Canada Symposium 97, "New Directions in Surveys and Censuses", pp. 385-388.

[52] Van den Brakel, J.A. and Renssen, R.H. (1998). Design and analysis of experiments embedded in sample surveys. Journal of Official Statistics, Vol. 14, No. 3, pp. 277-295.

[53] Van den Brakel, J.A. and Renssen, R.H. (1999). Testing Effects of Incentives and a Condensed Questionnaire on Response Rates. Netherlands Official Statistics, Vol. 14, summer 1999, pp.9-12.

[54] Van den Brakel, J.A. and Renssen, R.H. (2000). A field experiment to test effects of an incentive and a condensed questionnaire on response rates in the Netherlands Fertility and Family Survey. Research in Official Statistics, Vol. 3, No. 1, pp. 55-63.

[55] Van den Brakel, J.A. and Van Berkel, C. (2000). Testing effects of a new questionnaire in the Dutch Labour Force Survey. Research paper (BPA no. 3702-00-RSM), Department of Statistical Methods, Statistics Netherlands.

[56] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. American Mathematical Society, 54, pp. 426-482.

Curriculum vitae

Jan van den Brakel was born on May 15, 1966 in Baarn, the Netherlands. From 1978 to 1982 he attended Lower Agricultural Vocational Education in Nijkerk and from 1982 to 1985 Intermediate Agricultural Vocational Education in Utrecht. In 1985 he started studying Forestry at the College for Forestry and Land and Water Management in Velp. He wrote his bachelor's thesis in rural estate management and received his Bachelor of Science degree in Forestry in 1989. The same year he started studying Biometrics at Wageningen Agricultural University. He wrote his master's thesis in statistics "Dendrochronology, State-Space Models and the Kalman Filter" in 1992. This research was conducted in cooperation with KEMA Environmental Services. In the same year he received his Master of Science degree in Biometrics (cum laude). After graduation he started as research fellow at the Interuniversity Centre for Social Science Theory and Methodology (ICS) at Utrecht University where he did research on statistical methods for the analysis of compositional data. Since 1994 he is employed at the Department of Statistical Methods of Statistics Netherlands where he conducts research and consultancy work in the areas of sampling theory, design and analysis of experiments and multivariate data analysis. He is also involved in statistical auditing projects. In 2001 he finished his Ph.D. thesis on "Design and Analysis of Experiments Embedded in Complex Sample Surveys", which is based on his research and consultancy work on experimental design at Statistics Netherlands.