Tutkimuksia Forskningsrapporter Research Reports 231

Kari Djerf

Properties of some estimators under unit nonresponse

Tilastokeskus
Statistikcentralen
Statistics Finland

TUTKIMUKSIA FORSKNINGSRAPPORTER RESEARCH REPORTS 231

Kari Djerf

PROPERTIES OF SOME ESTIMATORS UNDER UNIT NONRESPONSE

Tilastokeskus
Statistikcentralen
Statistics Finland

Cover
Maija Sohlman

Layout
Hilkka Lehikoinen

ISSN 0355–2071
ISBN 951–727–857–8

Hakapaino Oy, Helsinki 2001

FOREWORD

This study reviews some reweighting methodology widely used in statistical agencies to adjust for unit nonresponse in sample surveys. As an academic thesis it covers many theoretical sections, but the ultimate goal was to find out which methods perform best in simulation where a real data set was used. The results confirmed some earlier empirical findings on the usefulness of modern calibration estimators. Thus, the decision to replace traditional types of estimators by calibration estimators in many of Statistics Finland's surveys was justified from the nonresponse adjustment point of view as well.

I would like to thank many colleagues both at Statistics Finland, academic institutions and international scientific networks. My collaborators at the Statistical Research and Development section and the Social Statistics unit gave me a lot of support and could devote their time for fruitful discussions. Professors Carl-Erik Särndal (Université de Montréal) and Erkki Pahkinen (University of Jyväskylä) as supervisors with Dr Risto Lehtonen of Statistics Finland were the real motivators of my study with their helpful comments. My family members suffered from the time needed for the study and therefore, I would like to thank them once more for their understanding.

Ms Merja Ryhänen carefully checked and improved my English in this study. Finally, financial support both from the Academy of Finland and Statistics Finland, and technical support from the Department of Statistics, University of Jyväskylä provided me with material opportunities to complete this licentiate thesis.

Helsinki, September 2000

CONTENTS

SUMMARY

Missing and incomplete observations are present in practically all surveys. A totally missing observation is called unit nonresponse. Sometimes totally missing observations are not reacted to in any way. Therefore, one could assume that nonresponse is harmless, i.e. distribution in the nonresponse is the same as in the response. However, in most cases such assumption is wrong, leading to biased estimates.

In the past few years attention has focused on the consequences of nonresponse. A variety of new methods have been developed in the sampling theory for correcting the consequences of unit nonresponse. Especially the so-called reweighting methods have been used frequently. The properties of the unknown mechanism which generates response or nonresponse have also been investigated. The response process can be interpreted as a second phase sampling, conditioned on the original sampling procedure. Application of the two-phase sampling theory complicates the derivation of the estimators of sampling variances.

The aim of this study is to compare the statistical properties, especially bias and precision, of some commonly used estimators when there is unit nonresponse in the data set. The estimators commonly used in surveys conducted by Statistics Finland include, for example, the Horvitz-Thompson estimator, and the following reweighting methods: weighting-class estimator, post-stratified estimator and calibration estimator. Their properties are evaluated by empirical data and by a Monte Carlo simulation. A real data set from the Finnish Labour Force Survey was used as the basis for the simulation study. In order to mimic the original response structure as exactly as possible, a response/nonresponse indicator was created in the data set. Subsequently, 1,000 independent SRSWOR samples, each containing 1,165 elements, were selected from the data. The properties of the estimators were assessed on the basis of these samples.

The result from the simulation study was clear: the auxiliary information used should always correlate with both the response mechanism and the study parameters. If both conditions are met,

bias will remain negligible and precision will be good. In general, the best results in this respect were obtained using calibration estimators. However, the effect is not necessarily carried to the domains of the study. To minimise bias and maximise precision, the effect of the auxiliary information should also extend to the domains.

TIIVISTELMÄ

Puuttuvat ja puutteelliset havainnot aiheuttavat ongelmia miltei kaikissa survey-tutkimuksissa. Koko havainnon puuttumista kutsutaan yksikkökadoksi. Havaintoyksiköiden puuttumiseen ei joskus reagoida mitenkään. Silloin oletetaan, että kato olisi harmitonta eli se jakautuisi täsmälleen samalla tavalla kuin saadut havainnotkin. Useimmiten oletus harmittomasta kadosta osoittautuu vääräksi, jolloin kuvattu menettely johtaa harhaisiin estimaatteihin.

Yksikkökadon vaikutukseen on viime vuosina kiinnitetty huomiota. Otantateoriassa on kehitetty useita eri menetelmiä, joilla kadon vaikutusta kyettäisiin eliminoimaan. Yleisimmin käytetään niin sanottuja uudelleenpainotusmenetelmiä. Samoin vastaamista ja katoa generoivan (tuntemattoman) mekanismin ominaisuuksia on pyritty jäljittämään. Vastaamisprosenssi voidaan tulkita toisen vaiheen todennäköisyysotannaksi ehdolla alkuperäinen otanta. Kaksivaiheisen teorian soveltaminen tekee otosvarianssien estimaattorien johtamisen monimutkaiseksi.

Tämän työn tavoitteena on vertailla joidenkin yleisesti käytettyjen estimaattorien tilastollisia ominaisuuksia, varsinkin harhaa ja täsmällisyyttä yksikkökadon vallitessa. Valittuja estimaattoreita käytetään muun muassa Tilastokeskuksen otantatutkimuksissa. Ne poikkeavat toisistaan sekä survey-tutkimuksen sisäisen että sen ulkopuolisista lähteistä saatavan lisäinformaation käytettävyyden suhteen. Esimerkkeinä mainittakoon tavallinen Horvitz-Thompson -estimaattori, ja varsinaisista uudelleenpainotusmenetelmistä painotusluokka-estimaattori, jälkiositusestimaattori tai kalibrointiestimaattori. Estimaattorien ominaisuuksia arvioidaan sekä empiirisen aineiston perusteella että empiiriseen aineistoon perustuvalla Monte Carlo -simuloinnilla. Simulointia varten aineistoon generoitiin vastaamista ja katoa kuvaava indikaattori siten, että se kuvaisi mahdollisimman tarkasti alkuperäisen aineiston vastaamisprosessia. Aineistosta poimittiin tuhat toisistaan riippumatonta 1 165 alkion suuruista yksinkertaista satunnaisotosta, joiden perusteella estimaattorien ominaisuuksia on kyetty arvioimaan.

Tutkimuksen tulokset ovat selkeät: havaintoja painotettaessa on syytä käyttää sellaista lisäinformaatiota, jolla on selkeä ja mieluiten voimakas riippuvuus yhtäältä vastaamista jäljittävän prosessin ja toisaalta tutkimuksen kohteena olevien parametrien kanssa. Silloin harha pysyy pienenä ja estimaatit ovat täsmällisiä. Yleisesti ottaen parhaat tulokset saavutettin kalibrointiestimaattoreilla. Vaikutus ei kuitenkaan välttämättä ulotu koko populaatiosta osajoukkoihin saakka. Harhattomuuden ja täsmällisyyden turvaamiseksi lisäinformaatio kyettävä ulottamaan myös osajoukkoon.

I
UNIT NONRESPONSE MODELS

1 INTRODUCTION TO MISSING DATA

Traditional sampling theory assumes complete measurement for all sampling units (and elements) and all study variables. This ideal situation can be achieved in some circumstances, for example when a sample is selected from administrative records which contain the whole target population. In sample surveys, however, this ideal can rarely be achieved. In true surveys we are faced with numerous types of data imperfections. The number of the sampled unit may eventually be reduced from the original sample size in different phases of survey operations. Reductions may result from coverage errors which are actually caused by coverage problems in the sampling frame; from unit nonresponse during the field work, or from the rejections of some responses during the data checking process. Survey contents may also be reduced from the basic survey design due to item nonresponse or other imperfections in data obtained by surveys. This paper fives a brief outline of sources of errors in the Finnish context.

1.1. Coverage errors

The Finnish system of registers is quite up-to-date, especially in register data on individual persons. The Central Population Register (CPR) containing in principle all resident persons, is the most frequently used sampling frame. Certain coverage problems exist, though which will be dealt with briefly.

Overcoverage can be measured. In Statistics Finland's samples for surveys on individual persons overcoverage is normally about 1.5 per cent. The principal reasons for overcoverage are the death of the sampled person or moving either to an institution (hospital, prison) or abroad.

The size of undercoverage is more difficult to evaluate. In surveys on individuals undercoverage most often originates from moves away from institutions or from abroad back home. The mechanism of undercoverage arises due to delays either in updating the frames or starting the survey. Undercoverage has generally not been viewed as particularly harmful in the surveys conducted by Statistics Finland.

Although unit coverage errors remain fairly small, other types of coverage problems may be substantial. The contents of the CPR itself are not very comprehensive. General demographic and housing information is normally considered to fulfil the needs of basic types of surveys. The updating delay in the CPR is normally less than one month. There remains another, far more important delay factor – the lag between the primary sample selection and the actual start of the survey.

In some surveys basic demographic information is not comprehensive enough. In such cases other administrative records or register information is merged into the original sample. Despite the technical feasibility of merging large data sets, the problem of the quality of the auxiliary data becomes far more pervasive. In many cases the additional data sources may have much longer updating delays either due to administrative reason (e.g. taxation data is at least one year old), or lags in updating the register. Such lags cause major problems in business survey data. When the sampling frame is known to contain errors or when using multiple frames from different points in time, etc., polishing the final sampling frame is a very important task. It may require the use of statistical matching, frame estimation and other ways to obtain the information necessary for calculating merely the basic inclusion probabilities.

The means of dealing with coverage errors differ significantly from one survey to another. In the surveys on individual persons or households there is good reason to believe that coverage errors are of minor importance. The coverage errors are therefore normally taken into account in weighting. The original size measures, for example population size N^* (which may contain over- and undercoverage) is replaced by the "corrected" $\tilde{N}$. The latter is always an estimate of the true N. In normal circumstances we can obtain a fairly good approximation of N for deriving the inclusion probabilities.

1.2. Nonresponse error

The problems caused by nonresponse can be far more serious than those caused by coverage errors. In this presentation we deal with unit nonresponse. The size of unit nonresponse varies considerably from one survey to another. Nonresponse rates in the Labour Force Survey remained quite low until 1996; the rate for the first wave was about 7.5 per cent on average. In the household surveys conducted by Statistics Finland, instead, the nonresponse rates are much higher, ranging from 15 to 40 per cent, depending of the response burden in the survey in question. Some features of nonresponse are quite common but, of course, there are features specific to each survey (c.f. Djerf 1996a).

Various ways of dealing with unit nonresponse are available. Reweighting is perhaps the most common method. The idea of reweighting the data set is to adjust the original inclusion probabilities by the response probabilities. The following chapters will deal with different types of reweighting adjustment.

Another technique, unit imputation, is common in some countries and in some research institutes. Unit imputation means that the missing unit is replaced by another unit which is close enough in a metric sense, for example by using the nearest neighbour technique. So far, unit imputation has been used seldom in the surveys conducted by Statistics Finland (c.f. Laaksonen 1992).

Besides the above techniques several other methods exist which can be constructed as a remedy against nonresponse, such as randomized response, substitution, and quota sampling. Some of the methods have a fairly sound statistical background and thus their use can be recommended. But the widely used techniques of substitution and quota sampling fall short in this respect (c.f., however, Deville 1991). As our study is restricted to reweighting, none of these methods will be discussed.

Another type of nonresponse is item nonresponse, where the response is generally acceptable although some items are missing. Some questions may be ones that the respondent either cannot or is not willing to answer. Principally item nonresponse is always specific to the survey in question. Imputation is a technique commonly used to cover the blanks in the data set due to item nonresponse. This is also the case at Statistics Finland. Some forms of imputation are known as logical or deductive imputation: in some cases a missing item can be imputed with probability 1, while in other cases we must use either some kind of prediction (a simple regression or other type of modelling), or to locate out a suitable host whose value can be brought to the missing case (hot decking), or both. Laaksonen (op cit.) has given examples of different methods applied at Statistics Finland.

1.3. Measurement errors

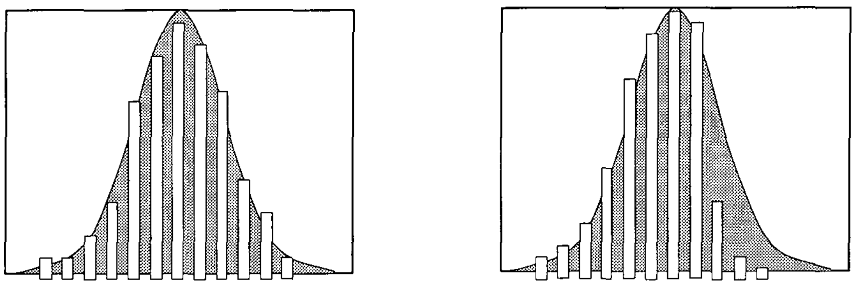
Measurement errors are the third general type of errors which exist in sample surveys. They are also probably the most difficult errors because they are difficult to detect. Measurement errors may derive from various sources: insufficient design of the survey and/or the questionnaire, misleading interviewer (or respondent) instructions, wording and language problems, deliberately or undeliberately given misleading answers, technical problems etc.

Efforts are made to trace measurement errors by applying various logical checks, data edits and occasionally by modelling. Nonetheless measurement errors exist in every survey and are also to be found in the "cleaned" data sets. In the broad sense of the term, certain problems arising from data missingness can also be regarded as measurement errors. However, measurement errors as such are not the topic of this paper, so in this case the reader should refer to other sources. A comprehensive treatment of the topic is presented, for example, in the handbook by Biemer et al. (1991): *Measurement Errors in Surveys*.

1.4. Consequences of missing data

In an ideal situation missing data are not harmful. The term "harmless" means that all statistical properties of the data set remain untouched in analysis except the number of cases. Little and Rubin also use the concept "missing at random" (1987). Consequently, the term "harmful" means that some kind of problem will be present when the data set is analysed by statistical methods (i.e. not missing at random). The distinction can easily be illustrated by two distributions. Let us take two hypothetical data sets from the same phenomenon obtained by comparable procedures. In data set A data missingness is random and therefore we can proceed in analysing our study variable Z. However, in data set B we have lost information in a non-random manner. The distribution of the study variable Z therefore no longer presents the true population distribution. Harmful missingness thus leads to bi-

Figure 1. Examples of missing data: harmless and harmful cases.



A: Harmless case B. Harmful case

ased estimators (See Little and Rubin 1987 for different examples of the problems).

In case A we can assume that there is no actual change caused by nonresponse in the distribution of our study variable, i.e. $E(Z_A|m_A) \approx E(Z_A|n_A) = Z$, where n is the sample size and m the number of respondents ($m_A \leq n_A$). By contrast it is easy to recognise from case B that the expectations differ: $E(Z_B|m_B) < E(Z_B|n_B)$. Also the sampling variance of case B will be smaller than that obtained from case A due to reduced variability of Z.

"Ignorable" and "non-ignorable" are terms closely related to "harmless" and "harmful". Loosely speaking both terms refer to the same phenomenon. There exists, however, theoretical differences between the two. Rubin advocated the use of former terms in Bayesian analysis. Data missingness (and nonresponse in particular) is non-ignorable when the statistic in question contains a parameter where information on the data missingness procedure is given. Normally it is a parameter reflecting the true population distribution. In the Bayesian analysis such a parameter can be solved, and the estimates are improved in reflecting the population distribution.

1.5. Aim of the study

The main goal in this study is to compare different types of adjustment methods used for unit nonresponse. Various methods developed for the purpose are presented and the most significant ones are further empirically evaluated. Firstly, the methods are tested using an empirical data set in the estimation problem of the Finnish Labour Force Survey. Secondly, they are tested in a Monte Carlo simulation study. The latter reveals statistical properties of the estimators in the nonresponse case: bias, mean square error and coverage rates.

The choice of methods which are compared is also connected with the use of auxiliary informal-ion. There are methods which use information at the sample level, and some other at the population level. The best case being, of course, the one where information is available both at the sample and at the population level. The availability of information will also be discussed.

2. THEORY OF UNIT NONRESPONSE

2.1. Introduction

The impact of unit nonresponse on surveys, has been known a long time, with efforts presumably having been targeted to overcome the problem after the first large scale surveys were introduced. Current statistical methods to deal with unit nonresponse date back to the 1940s. Although we deal only with the statistical methods in this study it is worth of remembering that improving the survey setup and operations in general can result in greatly reduced nonresponse.

Perhaps the first method to be adopted in widespread use was callbacks and reminders, introduced to surveys in the 1940s. In the original callback (here remainder) method the nonrespondents are sent one or more reminders after the first data collection deadline is over. They are dealt with separately in order to locate possible differences due to early or late responding. The current callback method is clearly related to reminding the sampled persons or firms already during the fieldwork period to respond, or interviewing the reluctant part of the sample using a reduced set of core questions. In any case the method improves the quality of the survey and may reduce bias as well.

The first well-known adjustment method was based on a similar idea. Households "being not at home" upon the arrival of the interviewer were given adjustment weights, a method was suggested by Hartley and implemented by Politz and Simmons (Oh and Scheuren 1983, Cochran 1983).

Subsampling of nonrespondents was first devised by Hansen and Hurwitz in 1946, where subsample is generally interviewed with a reduced set of questions. Subsampling may reveal the characteristics of respondents and nonrespondents, making it possible to reduce eventual bias by other methods.

The Hansen and Hurwitz method leads to the so-called reweighting approach, which has become a widespread means of reducing bias due to unit nonresponse. Reweighting is based on the idea certain characteristics differ in the subsets of the sample, namely the respondents and the nonrespondents. Thus the main task is to locate these characteristics from the whole sample and then the reweight the respondents so that the original sample (or population) distributions can be maintained. There are numerous technical options available: post-stratification (e.g. Thomsen 1973, Bethlehem 1988), weighting-class estimators (Oh and Scheuren 1983), the raking-ratio method (originally Deming 1940) and the prediction of individual response probabilities using a model estimated from the obtained sample (e.g. Ekholm and Laaksonen 1991). Furthermore, more generalised estimation techniques may be used, especially the regression and the calibration estimators (Bethlehem and Keller 1987, Deville and Särndal 1992, Särndal et al. 1992). Many of these techniques are discussed later in this presentation.

However it took time before the problem of nonresponse was clearly addressed in the sampling theory textbooks. For example Cochran (1977) and Sukhatme et al. (1984) included chapters dealing with the problem in general terms, but restricted the treatment to the basic methods. Kish (1965) provides a more comprehensive treatment of the question using a fairly practical approach, while Särndal et al. (1992) supplied a longer chapter on the problem of the nonresponse, also offering a number of up-to-date solutions to deal with nonresponse. In Lehtonen and Pahkinen (1995) the discussion is more limited but works along the lines of Särndal et al. Even the two recent books by Särndal et al. and by Lehtonen and Pahkinen only mention the more complicated model-based approach presented in Ekholm and Laaksonen (1992), and Rosenbaum and Rubin (1983). Probably the most comprehensive treatment of nonresponse is to be found in a three-volume set of books under the title *Incomplete Data in Sample Surveys* (1983).

In the following chapters we introduce some reweighting methods which can be used to tackle bias resulting from nonresponse.

2.2. Basic concepts in sample selection

Let $U = \{1, \ldots, k, \ldots, N\}$ be a finite population and y be the variable of interest (study variable) which has values in the whole population U (y_k is the value of y for the kth element). We are interested in estimating a parameter of y (e.g. the total value of y denoted by t) from U. Allow for an arbitrary sampling design $p(\cdot)$ such that $p(s)$ is the probability of selecting a sample s with size n_s. We can denote that the samples s are subsets of U ($s \subseteq U$). Then the first and second order inclusion probabilities can be denoted by

$$\pi_k = \sum_{s \ni k} p(s)\ ,\ \ and\ \ \pi_{kl} = \sum_{s \ni k,l} p(s)$$

where $\pi_{kk} = \pi_k$. Let $\Delta_{kl} = \pi_{kl} - \pi_k\pi_l$. Let us assume for simplicity, that both the first and second order inclusion probabilities are strictly positive. Then the π or Horvitz-Thompson (abbreviated as H-T) estimator for the population total is

$$\hat{t}_{HT} = \sum_{k=1}^{n} \frac{y_k}{\pi_k} \tag{2.1}$$

The simple expansion or H-T estimator is easily shown to be design unbiased. Let $I\,(s)$ be an indicator for the sample membership such that $E\,(I_k) = \pi_k$. Then the expectation for the H-T estimator is

$$E(\hat{t}_\pi) = E\left(\sum_{k=1}^{N} I_k \frac{y_k}{\pi_k}\right) = \sum_{k=1}^{N} \pi_k \frac{y_k}{\pi_k} = t$$

The design variance for the π estimator i

$$V_{HT}(t) = \sum\sum_{U} \Delta_{kl}(y_k/\pi_k)(y_l/\pi_l) \tag{2.2}$$

An unbiased estimator for the variance is

$$\hat{V}_{HT}(\hat{t}) = \sum\sum_{s} \left(1 - \frac{\pi_k\pi_l}{\pi_{kl}}\right)\left(\frac{y_k}{\pi_k}\right)\left(\frac{y_l}{\pi_l}\right) \tag{2.3}$$

Various sampling designs can be treated under the general principles of the B estimator (c.f. e.g. Särndal et al. 1992).

The π estimator assumes that we can obtain information from all units which belong to the target population U. For different reasons, however, we will have missing units in our sample s. Let us denote the responding set by r ($\subseteq s$) and $r - s$ the nonresponding set by nr. Instead of the original sample of size n_s we receive complete responses for m_r.

Furthermore, we have to assume that the response probabilities are strictly positive for all elements k. This means that no hard-core refusers exist whose response probability is exactly zero. Loosely speaking, every sampled element is willing to at least to consider responding to a survey.

The most common method of handling nonresponse is probably reweighting, where the original (first phase) inclusion probabilities are deflated by the relative size of response. The sampling weights are inflated by the reciprocal of the measure of response, say θ ($0 < \theta \leq 1$). Hence we obtain $w_k = 1/(\pi_k \theta_k)$.

Uniform response mechanism

A naive way would be to assume that the response generating mechanism is stochastic and uniform over the obtained sample (and more generally over all samples in the sample space). Then the response probability is fixed to θ $(= m/n)$. The use of a globally uniform response model is acceptable provided that

(a) the (unknown) process generating the response probabilities is a random process for all elements in the target population

(b) there is no association between the study variables and the response probabilities.

If both conditions are met the researcher may wish to continue without further problems, the sample size n is replaced by the number of respondents m $(= \theta n)$ in weighting. This "doing nothing" approach leads to the estimator

$$\hat{t}_r = N\tilde{y}_r = N \frac{\sum_r y_k / \pi_k \theta}{\sum_r 1/\pi_k \theta} = N \frac{\sum_r y_k / \pi_k}{\sum_r 1/\pi_k} \tag{2.4}$$

which in this case would be practically unbiased.

In this case the only effect of nonresponse is observed in the form of increased sampling variance. In the case of simple random sampling without replacement, sampling variance can be derived conditionally; first the original sampling variance and followed by the variance due to conditional subsampling of respondents from the obtained sample. Thus the total variance is:

$$V(\hat{t}|n,m) = N^2(\frac{1}{n} - \frac{1}{N})V + N^2(\frac{1}{m} - \frac{1}{n})V = N^2(\frac{1}{m} - \frac{1}{N})V \tag{2.5}$$

where $V = \sum_{k=1}^{N} (y_k - \bar{y}_U)N^2 / (N-1)$

The unbiased estimator of the variance is obtained directly from the ordinary SRSWOR sampling variance estimator which is slightly bigger in the nonresponse case than in the full response case since $m \le n$ ($m > 1$):

$$\hat{V}(\hat{t}_s)=N^2\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{k=1}^{n}\frac{(y_k-\bar{y}_s)^2}{n-1}\le \qquad (2.6)$$
$$N^2\left(\frac{1}{m}-\frac{1}{N}\right)\sum_{k=1}^{m}\frac{(y_k-\bar{y}_r)^2}{m-1}=\hat{V}(\hat{t}_r)$$

Although optimal circumstances occur quite seldom, this approach prevails in the traditional sampling theory. However, some authors advocating the traditional approach have warned about the bias. For example, Cochran, while presenting one version of the model (using separate strata for respondents and nonrespondents), also shows that fixed deterministic nonresponse treatment can lead to substantial bias. We divide the population U into the subsets of respondents and nonrespondents as above, so that they are treated as strata: in stratum U_r the response probability is 1 and in stratum U_{nr} it is 0. It is easy to indicate that bias is a function of the differences of the means in each stratum of the population:

$$B(\hat{t}_r)=E(\hat{t}_r)-t \doteq \hat{t}_r-(\hat{t}_r+\hat{t}_{nr})=N_{nr}\ (\bar{y}_r-\bar{y}_{nr}) \qquad (2.7)$$

(Cochran 1977, 360–361).

If the response probabilities vary (i.e. $\Pr\{k \in r \mid s\}=\theta_k$) and are associated with the study variables, there will again be bias:

$$B(\hat{t}_r)=E(\hat{t}_r)-t \doteq N\frac{\sum_U y_k\theta_k}{\sum_U \theta_k}-t=(N-1)\frac{S_{y\theta U}}{\bar{\theta}_U}$$
$$=\frac{(N-1)}{N}t\,R_{y\theta U}\ cv_{yU}\ cv_{\theta U}\approx t\,R_{y\theta U}\ cv_{yU}\ cv_{\theta U} \qquad (2.8)$$

(Särndal et al. 1992, 577).

The size of bias depends on the correlation between the study variable and the response probability in population ($R_{y\theta U}$), and the coefficient of variation of the study variable (cv_{yU}) and the response probabilities in population ($cv_{\theta U}$). Since both coefficients of variation have some positive values, the magnitude of bias is inherent in the correlation coefficient. Non-zero correlation clearly violates assumption (b) above. This type of a missing data mechanism is harmful.

It can be claimed that one of the two conditions is almost always violated in real sample surveys. Without appropriate modification of the estimator(s) the results will be biased. A more advanced way of handling nonresponse is presented in the following section.

Two-phase sampling theory applied to nonresponse

A random process which determines that some sample elements will be nonrespondents and that the rest are respondents is actually second phase sampling conditional to the realization of the original sampling design: $\Pr\{\cdot \mid s\} = \theta_k$. This is sometimes called quasi-randomisation (Oh and Scheuren 1983) or response homogeneity group models (Särndal et al. 1992). It should be observed that the true response distribution remains undetectable. However, we may attempt to approximate the true distribution by modelling the empirical responses and then use the estimated response probabilities. One useful way is to assume they are results from individual Bernoulli trials over the sample set. Thus $\Pr\{k \in r \mid s\} = \hat{\theta}_k$ and $\Pr\{k \,\&\, l \in r \mid s\} = \hat{\theta}_k \hat{\theta}_l$. The latter inclusion probability is not true if we assume that the sampled elements do not respond independently of each other.

The variance of the two-phase design with the general Horvitz-Thompson type of presentation is presented by Särndal and Swensson as follows:

$$V(\hat{t}_{2P}) = \sum\sum_{U} (\pi_{kl} - \pi_k \pi_l) \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} + E_a\left[\sum\sum_{s} \left(\pi_{kl|s} - \pi_{k|s}\pi_{l|s}\right) \frac{y_k}{\pi_{k|s}} \frac{y_l}{\pi_{l|s}}\right] \tag{2.9}$$

and its unbiased variance estimator as:

$$\hat{V}(\hat{t}_{2P}) = \sum\sum_{r} \frac{\pi_{kl} - \pi_k \pi_l}{\pi_{kl}\hat{\theta}_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} + \sum\sum_{r} \frac{\hat{\theta}_{kl} - \hat{\theta}_k \hat{\theta}_l}{\hat{\theta}_{kl}} \frac{y_k}{\pi_k \hat{\theta}_k} \frac{y_l}{\pi_l \hat{\theta}_l}$$

$$= \sum\sum_{r} \frac{\pi_{kl}\hat{\theta}_{kl} - \pi_k \hat{\theta}_k \pi_l \hat{\theta}_l}{\pi_{kl}\hat{\theta}_{kl}} \frac{y_k}{\pi_k \hat{\theta}_k} \frac{y_l}{\pi_l \hat{\theta}_l} \qquad (2.10)$$

(Särndal and Swensson 1987, 280–281).

This general formula consists of two parts; the first being the variance estimator for the realised sample with the condition that only r responses out of n are obtained. The other component is the variance of selecting r elements for responding from the sample of n.

Application of the quasi-randomization theory leads to adjustment of the original sampling weights (design weights) by the empirical or predicted response probabilities conditioned to the obtained sample. When the sample elements are divided into mutually exclusive categories each having similar response probabilities, the empirical nonresponse treatment is based on adjustment cells. This approach is also known as the uniform response mechanism in subpopulations (Oh and Scheuren 1983).

Different methods exist, depending on the use of the sample or population information or both. In any case one uses the mean of the empirical response probabilities in each adjustment cell.

It is possible to go even further and use strictly model-based adjustment. In model-based adjustment the predicted (sometimes actual) response probabilities are grouped according to a parametric model. The number of response homogeneity groups g can easily grow quite large and the predicted probabilities $\hat{\theta}_g$ also vary greatly from one group to another, and in an extreme case each element may end up with its own response probability.

In the broad sense of statistical thinking all methods developed for reweighting presume some type of model. Thus they can be regarded as model-assisted methods. The next subsection will present briefly the most common practices which are based on this theory.

2.3. Alternative reweighting strategies

2.3.1. Weighting-class estimators

Probably the most well-known method of adjusting sampling weights is what is known as the weighting-class estimator (WTC estimator, c.f. e.g. Oh and Scheuren 1983), where the sample is divided into homogenous and mutually exclusive groups ($\hat{\theta}_g > 0, \forall g$) with respect to the response generating process. We have to assume that the response process in each group is random, and that no other information is available besides the empirical information of the sample. Thus only sample count information is used in the reweighting $\{n_g$ and $m_g\}$.

$$\hat{t}_{WTC} = \frac{N}{n} \sum_{g=1}^{G} \frac{n_g}{m_g} \sum_{k=1}^{m_g} y_k \tag{2.11}$$

The variance estimator in the WTC estimator is slightly cumbersome. In the case of the SRSWOR design it becomes:

$$\hat{V}(\hat{t}_{WTC}) = N^2 \frac{1-n/N}{n} \left[\sum_{g=1}^{G} \frac{n_g}{n} \left(1 - \frac{1-n_g/n}{m_g} \frac{n}{n-1} \right) S_{yg}^2 \right.$$

$$\left. + \frac{n}{n-1} \sum_{g=1}^{G} \frac{n_g}{n} (\bar{y}_g - \tilde{y})^2 \right]$$

$$+ N^2 \sum_{g=1}^{G} \left(\frac{n_g}{n} \right)^2 \frac{1-m_g/n_g}{m_g} S_{yg}^2 = \hat{V}_1 + \hat{V}_2 \tag{2.12}$$

where $S_{yg}^2 = \frac{1}{m_g - 1} \sum_{k=1}^{m_g} (y_{gk} - \bar{y}_g)^2$, $\bar{y}_g = \frac{1}{m_g} \sum_{k=1}^{m_g} y_{gk}$, $\tilde{y} = \sum_{g=1}^{G} \frac{n_g}{n} \bar{y}$,

and $\bar{y} = \frac{1}{m} \sum_{k=1}^{m} y_k$

The first sampling variance component of the WTC estimator ($\hat{V}_1$) is practically the same as the unconditional variance estimator for post-stratified sampling, while the second component inflates the sampling variance by the "missing units", and disappears if $m = n$.

The weighting-class estimator is unconditionally unbiased, but conditional bias can arise when a difference between the distribution of a (incorrectly) weighted population and the sample level based weighted population exists. Oh and Scheuren (1983, 152) show that conditional bias cannot be directly derived. However, an average can be obtained from mean square error derivation. Kalton and Maligalig (1991, 414) derive slightly simpler approximation

$$B(\hat{t}_{WTC}|n,m) \doteq \sum_{g=1}^{G} (\tilde{N}_g - N_g)\, \bar{y}_g \tag{2.13}$$

where $\tilde{N}$ is the estimate for the population in each subgroup based on the respondents.

2.3.2. Post-stratification

The post-stratification approach and weighting-class approach share common features but use information in different ways. In post-stratification we have information on subpopulations (RHG's) both at the sample and at the population level. Subpopulations can be treated as strata because we know both N_g, n_g and m_g. Besides the nonresponse adjustment, post-stratification can correct some coverage errors. The post-stratified estimator for a total is:

$$\hat{t}_{POST} = \sum_{g=1}^{G} N_g\, \bar{y}_g = \sum_{g=1}^{G} \frac{N_g}{m_g} \sum_{k=1}^{m_g} y_k \tag{2.14}$$

Let us assume that the sample has been selected by simple random sampling without replacement and the second phase sampling process can be approximated without bias. Two variance estimators can be given. The first one is the "ordinary" one, as follows:

$$\hat{V}(\hat{t}_{POST}) \doteq N^2 \frac{1-m/N}{m} \left[\sum_{g=1}^{G} \frac{N_g}{N} S_{yg}^2 + \frac{1}{m} \sum_{g=1}^{G} \left(1-\frac{N_g}{N}\right) S_{yg}^2 \right] \tag{2.15}$$

The "second-phase" conditional formula applied to the stratified design more complicated again:

$$\hat{V}(\hat{t}_{POST-2P}) = N(N-1)\sum_{g=1}^{G}\left[\frac{n_g-1}{n-1}-\frac{m_g-1}{N-1}\right]\left(\frac{n_g}{n}\right)S^2_{y1g}/m_h$$
$$+\frac{N(N-n)}{n-1}\sum_{g=1}^{G}\frac{n_g}{n}(\bar{y}_g-\tilde{y})^2 \qquad (2.16)$$

where $S^2_{yg} = \frac{1}{m_g-1}\sum_{k=1}^{m_g}(y_{gk}-\bar{y}_g)^2, \bar{y}_g = \frac{1}{m_g}\sum_{k=1}^{m_g} y_{gk}, \tilde{y} = \sum_{g=1}^{G}\frac{n_g}{n}\bar{y},$

and $\bar{y} = \frac{1}{m}\sum_{k=1}^{m} y_k$

n is the sample size, m is the number of respondents, and g is the response homogeneity group.

(Särndal et al. 1992, 289, Särndal and Swensson 1987, 282–283).

The post-stratified estimator is most often be design unbiased if the population totals are known and if there is no covariance between the study variable and the response probabilities:

$$B(\hat{t}_{POST}|n,m) = \sum_{g=1}^{G} N_g(\hat{t}_g - t_g) = \sum_{g=1}^{G} N_g C_{\theta Yg}/\bar{\theta}_g \qquad (2.17)$$

where $C_{\theta Yg} = \sum_{k=1}^{N_g}(\theta_{gk}-\bar{\theta}_g)(y_{gk}-\bar{y}_g)/N_g$

(Bethlehem 1988, 257).

If the population totals are unknown for some strata they must be estimated using the sample level information. In this case bias can occur for reasons similar to those of the weighting class estimator above. (Oh and Scheuren 1983, 150).

Where the traditional approach is chosen and the response mechanism is regarded as fixed, two sources of bias exist: (A) the weighted difference of the means between the respondent and the nonrespondent stratum and (B) the variability of the response rates within the strata (Thomsen 1973 and 1978).

Owing to the existence of different interpretations, it is worth citing Bethlehem (1987, 258):

1. Construct strata which are homogenous with respect to the target variable .... the covariance $C_{\theta Yg}$ will be close to zero.

2. Construct strata which are homogenous with respect to the response probabilities. Then again the covariance will be close to zero.

As Bethlehem points out, the construction of such post-strata simultaneously is difficult.

2.3.3. Raking ratio adjustment

Raking ratio adjustment was developed by Deming and his colleagues in the early 1940s. The idea was to adjust survey estimates to be consistent with the known population counts. The method was commonly used in the U.S. Bureau of the Census and at Statistics Canada from the 1950s onwards. The use of the method also covers nonresponse adjustment, resembling post-stratification because the information is used both at the population level and at the sample level. However, in raking ratio adjustment either the population information or sample information is restricted to the marginal distributions, whereas post-stratification uses all information on the adjustment cells.

Here we treat the raking procedure in the so-called population-based adjustment mode prevalent in the Finnish context (c.f. Kalton and Maligalig 1991, 417; and comments by Little op cit., 441–443). In broad terms raking can be described as follows: consider a two-way table consisting of cell counts. The sample count is n_{hj}, and the number of respondents m_{hj}, respectively, $h = \{1, \dots H\}$ and $j = \{1, \dots J\}$. We know the marginal distributions of N_{h+} and N_{+j} from the population. Since the joint distribution (i.e. population cell counts) N_{hj} are unknown they must be estimated from the marginal distributions using the sample information. Let us denote the estimate by $\tilde{N}_{hj.}$ The raking ratio estimator results from an iterative process where the row and column counts are calculated from each cell in turn. The totals are adjusted in each iteration to the known marginal and gradually the adjusted cell count estimates $\tilde{n}_{hj}$ provide correct estimates for both the row and the column distributions. Ireland and Kullback (1968) demonstrated that in the case of the SRS design the population cell count estimates for large samples are asymptotically unbiased, i.e.

$$\left(\frac{N}{n}\right)E(\tilde{n}_{hj}) \doteq N_{hj}$$

They also demonstrated that the raking estimators are best asymptotically normal (BAN) estimators for the unknown cell frequencies.

In the case of nonresponse Oh and Scheuren postulated that

1. Within each subgroup the responses are generated by a Bernoulli sampling process with common probability $\theta_{hj} > 0$.

2. The response mechanism is independent from one subpopulation to another, and

3. The response probabilities θ_{hj} have following structure over population subgroups: $ln\ (\theta_{hj}/(1-\theta_{hj})) = \alpha_h + \beta_j$. Thus they depend linearly on row and column margins, not on cells hj.

(Oh and Scheuren 1983, 164).

In addition, zero cells must not exist. The raking estimator for the total under nonresponse will be

$$\hat{t}_{RAK} = \sum_h \sum_j \tilde{N}_{hj} \bar{y}_{hj} \tag{2.18}$$

with the condition of $\tilde{N}_{hj} = \tilde{a}_h \tilde{b}_j m_{hj}$ where $\tilde{a}_h \tilde{b}_j$ are the factors needed for the convergence of marginal distributions.

The conditional bias of the raking method is comparable to the WTC estimator above:

$$B(\hat{t}_{RAK} \mid n, m) = \sum_{h=1}^{H} \sum_{j=1}^{J} (\tilde{N}_{hj} - N_{hj})(\bar{Y}_{hj} - \bar{Y}_{h+} - \bar{Y}_{+j} + \bar{Y}) \tag{2.19}$$

(Kalton & Maligalig 1991, 419).

The unconditional bias will be zero over repeated samples because $E\ \tilde{N}_{hj} = N_{hj}$.

The derivation of the variance estimator is always rather complicated and only approximate solutions exist. The conditional variance of the raked estimator under nonresponse is

$$\begin{aligned} V(\hat{t}_{RAK-2P} | n, m) &= \sum_{h=1}^{H} \sum_{j=1}^{J} \tilde{N}_{hj}^2 \left(1 - \frac{m_{hj}^*}{N_{hj}}\right) \frac{1}{m_{hj}^*} V_{hj}(\hat{t}_{hj}) \\ &= \sum_{h=1}^{H} \sum_{j=1}^{J} (\tilde{a}_h^2 \tilde{b}_j^2 m_{hj}) \left(1 - \frac{m_{hj}}{N_{hj}}\right) V_{hj}(\hat{t}_{hj}) \end{aligned} \tag{2.20}$$

$$\text{where } m^*_{hj} = \begin{cases} m_{hj}, \text{ for } m_{hj} > 0 \\ (n/N)\hat{\theta}_{hj} N_{hj}, \text{ for } m_{hj} = 0 \end{cases}$$

The approximate variance estimator is very lengthy and will not be presented here, c.f. Kalton & Maligalig for details (1991, 427).

2.3.4. Adjustment based on explicit response modelling

The previous strategies to adjust for nonresponse were constructed on the idea of using adjustment cells, where implicit models to compensate for nonresponse are employed. Next we will treat explicit response modelling. The model-based adjustment strategy has much in common with the weighting cell approach but also a number of differing elements exist. The model-based sampling theory is closely related to the superpopulation theory and Bayesian inference. The theory presented in the Bayes context can, for instance, be found in Rubin (1983), Rosenbaum and Rubin (1983), Little (1983a and 1983b) or Rosenbaum (1987). The use of the explicit model also contains the assumption of a deterministic response mechanism, and the modeler's task is to determine the correct empirical model for each data set.

The approach of Ekholm and Laaksonen (1991) will be presented here, where a logistic regression model for explaining the response mechanism is applied, as follows :

$$\text{logit}\left(\Pr(\theta_k = 1|\mathbf{x}_k)\right) = \log\left(\frac{\theta_k}{1-\theta_k}\right) = \mathrm{B}\ \mathbf{x}_k \tag{2.21}$$

Response probabilities for all sample elements were predicted after estimation and nonresponse adjustment cells g were created according to the characteristics of the covariates in $\mathbf{x}$. Thus the response probabilities are model-based estimates ($0 < \hat{\theta}_g \leq 1$). This approach has been advocated, for example, by Rosenbaum who claims that besides nonresponse adjustment, the model-based estimates can also compensate for differences between the population and sample proportions (1987, 391). The estimator for a total is thus

$$\hat{t}_{MOD} = \sum_{g=1}^{G} \frac{\hat{t}_{g,HT}}{\hat{\theta}_{g,MOD}} \tag{2.22}$$

The conditional variance estimator of the model-based response probability estimator according to Ekholm and Laaksonen is a modification of the ordinary H-T variance estimator although not a true two-phase sampling estimator:

$$\hat{V}(\hat{t}_{MOD}|n,m) = \frac{m}{m-1}\sum_{k=1}^{m}\left(\frac{y_k}{\pi_k\hat{\theta}_k} - \bar{y}\right)^2 + m(1-m/n)\left(\overline{\frac{y_k}{\pi_k\hat{\theta}_k}}\right)^2 \quad (2.23)$$

(Ekholm and Laaksonen 1991, 333).

Ekholm and Laaksonen give proof that the estimator is asymptotically unbiased (1990). Ekholm and Laaksonen assume that (a) nonresponse is believed to be ignorable (i.e. harmless) in each adjustment cell, and (b) the applied model is correct in explaining the true response distribution. Assumption (a) essentially relates the model of Ekholm and Laaksonen to the quasi-randomization approach (c.f. Little 1983a, 337–340). Assumption (b) is a key to evaluating unbiasedness. Little (1983a and 1983b) considers numerous different models which could be applied to the nonresponse problem. The result in general terms is that model-based predicted response probabilities tend to control bias (Little 1986).

Model-based adjustment and the generalized regression estimator

The model-based adjustment estimator can actually be considered as a special case of the generalized regression estimator. Bethlehem (1988) uses the following normal formulation of the estimator for complete response:

$$\hat{t}_{GREG} = \hat{t}_{HT} + (\mathbf{x} - \hat{\mathbf{x}}_{HT})'\hat{\beta} \quad (2.24)$$

where the parameter vector $\hat{\beta}$ is the outcome of the model predicting the auxiliary information. Bethlehem considers the estimator in its ordinary form and discusses post-stratification only as a special case.

Kott (1994) goes further by using the obtained model for explaining response probabilities. A short description of the possibility of considering the GREG estimator as a tool in fighting nonresponse bias is given below.

Let us assume that nonresponse is harmful. Then the H-T estimator $\hat{t}_{HT,r} \neq \hat{t}_{HT,s}$. We can, however, improve our estimate using a model, provided that the information available to us is associated either with the study variable(s) or the response probabilities or both. If our model is good it ensures that $E(\hat{\beta}\mathbf{x}_r) = \mathbf{x}_s$.

According to basic assumptions our original sampling design is unbiased, so E (x_s) = x. Now we can use the GREG estimator which "shifts" the biased estimate ($\hat{t}_{HT,r}$) by the amount of estimated bias:

$$\hat{t}_{GREG} = \hat{t}_{HT,r} + (\mathbf{x} - \hat{\mathbf{x}}_{HT,r})'\hat{\beta} \tag{2.25}$$

The GREG estimator is asymptotically unbiased provided that the applied model is correct. However, it is not very easy to show that it will hold in practise in the case of nonresponse. Isaki and Fuller (1982) and Kott (1994) demonstrate that the obtained estimators are design consistent (or quasi-design consistent). Thus the bias and the sampling variance will have a limit at zero for (very) large samples. If the model fails it will yield biased estimates.

2.3.5. Calibration of sampling weights

The calibration estimator was introduced in two articles by Deville and Särndal (1992) and by Deville et al. (1993), although the ideas were developed in a long sequence of studies originating to Deming and Stephan (1940). Deming stressed the importance of obtaining a good population structure in samples, and recommended the raking ratio estimator. This estimator was widely used despite the lack of a good approximate variance estimator. Other techniques were also developed with stratification and post-stratification appearing particularly useful. These two techniques have the advantage over raking of a straightforward way of calculating sampling variances.

The general idea of calibration is to modify the sampling weights so that marginal distributions become correct. This can be achieved using a suitable distance measure so that the measure minimizes the distance between the design weights (assumed to be design unbiased) and the calibrated weights. The distance function to be minimized is generally of the form: $E_p\{\Sigma_s (w_k - d_k)^2 / d_k q_k\}$. Deville and Särndal (1992) show that there are numerous distance functions which yield fairly close results. Four of them have been included in the computer program CALMAR: the linear distance function ("ordinary linear regression"), the exponential distance function (raking), the bounded logistic distance function (logit regression) and the bounded linear distance function (truncated regression) (Sautory, 1993).

The calibration estimator is flexible comprising elements of both post-stratification and the ratio estimator. Where auxiliary information is available in the form of total counts of marginal distributions, the calibration estimator can be regarded as incomplete post-stratification. Where auxiliary information is continuous, the calibration estimator resembles the ratio estimator. The two basic approaches can also be combined in order to obtain a form of the generalized regression estimator.

Särndal et al. show various approximate variance estimators which differ from each other in the use of auxiliary information (1992, 583–589). Särndal and Hidiroglou (1995) and Dupont (1995) have also derived new results for the calibration estimator in a two-phase sampling situation. This idea can be extended to multi-phase designs in the manner presented by Breidt and Fuller (1993).

Calibration in the complete response case

The calibration estimator can be presented in a numerous ways. Here we present it as an estimator for study variables in the form of a generalized regression estimator. The derivation itself refers to the calibration of sampling design weights which are denoted here by d_k (c.f. Deville and Särndal 1992).

$$\hat{t}_{yCAL} = \sum_s w_k y_k = \hat{t}_{y\pi} + (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})' \hat{\mathbf{B}}_s$$

$$\text{where } \hat{\mathbf{t}}_{x\pi} = \sum_s d_k \mathbf{x}_k,$$

$$\hat{\mathbf{B}}_s = \left(\sum_s d_k q_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_s d_k q_k \mathbf{x}_k y_k, \tag{2.26}$$

$$q_k = 1 / x_{p,k}, \text{ and}$$

$$w_k = d_k (1 + q_k \mathbf{x}'_k \lambda)$$

$$= d_k \left[1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})' \left(\sum_s d_k q_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \mathbf{x} q_k \right]$$

q_k is a reciprocal of a size measure contained in a vector x_p.

The calibration estimator is proven by Deville and Särndal to be design consistent. They also point out that due to minimizing the distance between $\hat{t}_{y\pi}$ and $\hat{t}_{yw}$ the latter is also at least asymptotically design unbiased (1992, 379).

Since the estimator itself is one form of the generalized regression estimator, the appropriate variance approximation in the variance formula is easy to derive using the residuals

$$\hat{V}_{CAL}(\hat{t}) = \sum_{s}\sum \left(\frac{\pi_{kl} - \pi_k \pi_l}{\pi_{kl}}\right)(w_k e_k)(w_l e_l) \tag{2.27}$$

where $e_k = y_k - \mathbf{x}'_k \hat{\mathbf{B}}_s$

(Deville and Särndal 1992, 380).

Calibration estimator and nonresponse

Owing to close kinship with the generalized regression estimator, the calibration estimator can also be regarded as one tool in dealing with nonresponse bias. The role of the unbiased sampling design is necessary; it ensures that the calibrated weights contain the original design structure and hence changes from the design weights are minimized. However, certain differences between the nonresponse modelling and the calibration do also exist; the calibration estimator provides the user with individually modified weights for all responding elements and does not presume that the response probabilities remain strictly positive. Moreover, the calibration estimator can be formulated as an unbalanced main-effects ANOVA model. Thus it works reasonably well in cases where the response distribution can be approximated by the marginal distribution of the "explanatory variables". Furthermore, it does not easily run into problems with degrees of freedom even with very small sample sizes.

If information is available only at the sample level for the calibration of sampling weights, then the final weighting resembles the response homogeneity group approach described above. A second case, seldom encountered in the Finnish context is where information at the population level exists but does not exist at the sample level. This case is not dealt as it has little practical relevance. The third case, the most common one is where auxiliary information exists both at the population and the sample level.

By applying the variance formula for the complete response case and allow for the ordinary formula (2.27) to be inflated by the estimated response probability, we obtain

$$\hat{V}_{CAL}(\hat{t}) = \sum_{r}\sum \left(\frac{\pi_{kl}\hat{\theta}_{kl} - \pi_k\hat{\theta}_k\pi_l\hat{\theta}_l}{\pi_{kl}\hat{\theta}_{kl}}\right)(w_k e_k)(w_l e_l) \tag{2.28}$$

The two-phase estimator is of the form:

$$\hat{t}_{yCAL-RHG} = \hat{t}_{y\pi} + (\hat{\mathbf{t}}_{\mathbf{x}} - \hat{\mathbf{t}}_{1\mathbf{x}})'\hat{\mathbf{B}}_1 + (\hat{\mathbf{t}}_{\mathbf{x}} - \hat{\mathbf{t}}_{\mathbf{x}})'\hat{\mathbf{B}}_r$$

$$= \sum_U \hat{y}_{1k} + \sum_{g=1}^{G}\left(\sum_{s_g} \frac{\hat{y}_k - \hat{y}_{1k}}{\pi_k} + \frac{1}{m_g / n_g}\sum_{r_g} \frac{y_k - \hat{y}_k}{\pi_k}\right)$$

$$\text{where} \quad \hat{y}_k = \mathbf{x}'_k \hat{\mathbf{B}}_r = \mathbf{x}'_k \left(\sum_{g=1}^{G} \frac{1}{m_g / n_g} \sum_{r_g} \frac{\mathbf{x}_k \mathbf{x}'_k}{\sigma_k^2 \pi_k}\right)^{-1} * \sum_{g=1}^{G} \frac{1}{m_g / n_g} \sum_{r_g} \frac{\mathbf{x}_k y_k}{\sigma_k^2 \pi_k} \tag{2.29}$$

$$\text{and} \quad \hat{y}_{1k} = \mathbf{x}'_{1k} \hat{\mathbf{B}}_1 = \mathbf{x}'_{1k} \left(\sum_{g=1}^{G} \frac{1}{m_g / n_g} \sum_{r_g} \frac{\mathbf{x}_{1k} \mathbf{x}'_{1k}}{\sigma_{1k}^2 \pi_k}\right)^{-1} * \sum_{g=1}^{G} \frac{1}{m_g / n_g} \sum_{r_g} \frac{\mathbf{x}_{1k} y_k}{\sigma_{1k}^2 \pi_k}$$

Here $\mathbf{x}_k$ refers to information at the sample level and $\mathbf{x}_{1k}$ at the population level. As Särndal and Swensson note, we must assume that variances $V(y_k)$ are constant ($V(y_k) = \sigma^2 > 0$). Although unknown, they disappear when estimating the residuals (1987, 284).

Dupont (1995) produced several estimators which can be regarded as refined forms of the above general estimator with distinctions between the different forms arising from the availability and nature of auxiliary information. Our estimator best coincides with Dupont's strategy 2.

The approximate variance estimator for a two-phase generalized regression estimator is:

$$\hat{V}(\hat{t}_{yCAL-RHG}) = \sum\sum_r \frac{\pi_{kl}\hat{\theta}_{kl} - \pi_k\hat{\theta}_k\pi_l\hat{\theta}_l}{\pi_{kl}\hat{\theta}_{kl}} \tilde{e}_{1kr}\, \tilde{e}_{1lr} + \sum_{g=1}^{G} n_g^2 \frac{1 - m_g / n_g}{m_g} S^2_{\tilde{e}r_g}$$

$$\text{where} \quad \tilde{e}_{1kr} = \frac{y_k - \hat{y}_{1k}}{\pi_k\hat{\theta}_k} \quad \text{and} \quad S^2_{\tilde{e}r_g} = \frac{1}{m}\sum_{r_g}\left(\frac{y_k - \hat{y}_k}{\pi_k\hat{\theta}_k} - \left(\overline{\frac{y_k - \hat{y}_k}{\pi_k\hat{\theta}_k}}\right)\right)^2 \tag{2.30}$$

(Särndal et al. 1992, 584).

It can be seen that when applying the exponential distance function (raking approach), the point estimates of the two-phase model will coincide exactly with the ordinary calibration estimates because we measure the **x** variables both at the population and at the sample level. Using the same variable additionally at the sample level only would not provide us with any extra information and thus $\hat{y}_k = \hat{y}_{1k}$.

This does not hold for variance form, although it has empirically been found that the original variance approximation and the two-phase variance approximation yield very close results.

Dupont (1994) showed that only the raking approach provides a consistent estimator in the case of nonresponse. The result is theoretically sound: the exponential distance function leads to a formula where the nonresponse is modelled by $ln\ (\theta_{hj} /(1- \theta_{hj})) = \alpha_h + \beta_j$. The calibration estimator is design unbiased in the case of nonresponse with one condition: the main effects model exploits the unknown response probability information and no interactions exist between the levels of α and β.

2.4. Evaluation of different reweighting approaches

Some discussion exist in sampling literature and comparisons of available methods of reweighting can be found. For example, Oh and Scheuren (1983), Kalton and Maligalig, and Laaksonen (1992) have reported their results. Laaksonen's results refer to real survey data sets while others have used mainly artificially created data sets in their simulations. In the next sections some comparisons based both on direct comparison and simulation of various estimators are presented. The idea behind using real data sets is to investigate methods which could later be applied on the survey practice. The methods compared consist of the weighting-class estimator, post-stratification and the calibration estimator. The methods are also combined in order to determine whether a combination of two or more methods yields better results.

II
EMPIRICAL COMPARISON OF DIFFERENT ESTIMATORS

3. EMPIRICAL FINDINGS FROM VARIOUS REWEIGHTING STRATEGIES

3.1. Introduction

The many alternative reweighting methods presented in the preceding chapter are tested here using a real survey data set. Two exceptions, however, must be mentioned. Firstly, the ordinary raking ratio estimator has not been tested separately because it is included in the calibration estimator. And secondly, the explicit modelling of response probabilities is a good method for improving the weighting of large-scale surveys conducted fairly seldom. Our data set is large but the survey is carried out very frequently. Searching the best model available is not possible in the production of real statistics due to time constraints, however.

Data set, sampling design and estimation design

We have selected the Labour Force Survey (LFS) data set of Statistics Finland dating to March 1993. The monthly LFS has a rather complicated design being a rotating panel design with five waves in the course of 15 months. However, the data set of one month can be regarded as a simple random sample of individuals aged from 15 to 74 years, although the true selection procedure is actually systematic sampling from the Central Population Register, where the Register is sorted according to the domicile codes before the sample selection. It means that the sample has implicit geographical stratification. So far we have not encountered any indications of selection bias due to systematic sampling, so the selection procedure can be approximated by simple random sampling without replacement (SRSWOR).

The ordinary LFS data set is reweighted using post-stratification. The purpose of post-stratification is to guarantee correct population distribution according to gender, age and geographic distribution. The total number of post-strata is 312 (2 * 13 * 12).

The total sample size in March 1993 was 12,804 and the number of respondents 12,092. Some of the post-strata contain only a few individuals. The strata of Autonomous Territory of the Åland Islands proved to be particularly problematic.

The properties of the estimation method are described in Djerf (1997 and 1996b) and in Djerf and Väisänen (1993). Djerf and Väisänen began to examine the statistical properties of the current estimation design as the gap between the unemployment figures generated by the LFS and those of the Unemployed Job-Seekers' Register of the Ministry of Labour started to increase since 1988. It was found that the post-stratified estimator of the LFS performed reasonably well for providing the total estimates of study variables such as being in the labour force, being retired, etc. The estimates were also consistent and efficient. For the unemployment figures, however, the post-stratified estimator did not yield efficient estimates. The design effect statistics remained about 1.

The precision of unemployment estimates was improved though by introducing an indicator whether or not the sampled person was an unemployed job-seeker in the Ministry's register. The information is linked directly to each record using a unique personal identification number (PIN). The correlation between the two concepts of being unemployed proved to be fairly high, about 0.80.

Djerf obtained further results by showing that the post-stratification estimator provided biased estimates for the register indicator because the share of unemployed persons was much higher among the nonrespondents than among the respondents. Hence the register indicator could serve as an explanatory variable for nonresponse as well. He advocated the use of the calibration estimator and the inclusion of register information in order to improve both the accuracy and precision of the estimates (1997). Some of the results are discussed in more detail in this paper.

3.2. Study variables and estimators

We are interested here in two major study variables in the LFS; namely the size of the labour force and the number of unemployed people. The concepts are measured using the interview information. The decision rule for the labour force status is rather complicated. First each person is classified according to his/her main activity indicating whether the respondent belongs to the labour force or not. Those in labour force are further classified according to their activities and working hours to those employed and those unemployed.

The following comparison serves mainly for determining the effect of using auxiliary information either at the sample level or both at the sample and at the population level. When using information at the sample level only it is assumed we actually apply a weighting-class estimator connected to the original sampling or estimation design estimator. In allowing our auxiliary information to contain the population counts, we can treat the information as an additional post-stratification variable or as a calibration variable. It was also applied as a weighting-class estimator. The estimators to be compared consist of:

Estimator	**Note**
Horvitz-Thompson	modified: sample size replaced by the number of respondents
Horvitz-Thompson – 2 PHASE	H-T estimator in the first phase, the Ministry of Labour's indicator of unemployed job-seeker in the second phase (as a weighting-class estimator)
Post-stratification (sex, age, province)	ordinary post-stratification used for official LFS statistics
Post-stratification (sex, age, province, ue-indicator)	as above but the original strata are further stratified by the ue-indicator
Post-stratification – 2 PHASE (sex, age, province)	ordinary post-stratification in the first phase, the Ministry of Labour's indicator of unemployed job-seeker in the uniform nonresponse model; second phase applied to the total sample level (as a weighting-class estimator)
Post-stratification – 2 PHASE (sex, age, province)	ordinary post-stratification in the first phase, the Ministry of Labour's indicator of unemployed job-seeker in the separate nonresponse model; second phase applied in each stratum (as a weighting-class estimator in each stratum)
Calibration (sex, age, province, ue-indicator)	The calibration of weights: both the raking and linear models were applied
Calibration (sex, age, province, ue-indicator) – 2 PHASE	The calibration of weights: both the raking and linear models were applied, the Ministry of Labour's indicator of unemployed job-seeker in the second phase (as a weighting-class estimator)

We present the results according to one domain only, namely the gender of the respondent (the formulae are in Appendix 1).

3.3. Estimates for totals

Table 1 confirms the findings by Djerf (1997) both in reducing the sampling variance and increasing the estimate for the unemployed persons. In principle all methods that use some information on the Register of Unemployed Job-Seekers should yield about the same point estimates because the indicator has only two outcomes. However, some surprising differences between the two occur. Especially the weighting-class method using only sample level information (H-T – 2 PHASE) performs much worse than the other methods in terms of sampling variance. The point estimates are also substantially lower than those obtained by using a more efficient estimator, e.g. the post-stratified estimator with a weighting-class estimator.

Table 1. Comparison of estimates for totals of LFS, March 1993. (Estimated standard errors in parentheses).

(a) Labour force

Estimator	Male	Female	Total
Horvitz-Thompson	1,223,208	1,173,878	2,397,086
(SRSWOR)	(16,118)	(15,940)	(16,647)
Horvitz-Thompson	1,227,048	1,175,648	2,402,696
– 2 PHASE	(16,127)	(15,951)	(16,580)
POST-STRATIFICATION	**1,276,273**	**1,155,012**	**2,431,285**
(sex, age, province)	(8,551)	(9,268)	(12,610)
Post-stratification	1,280,468	1,156,630	2,437,098
(sex, age, province, ue-indicator)	(8,587)	(9,221)	(12,445)
Post-stratification – 2 PHASE	1,252,242	1,159,406	2,411,649
(sex, age, province)	(9,556)	(10,126)	(14,201)
Uniform nonresponse model			
Post-stratification – 2 PHASE	1,277,224	1,154,758	2,431,982
(sex, age, province)	(8,532)	(9,250)	(12,584)
Separate nonresponse model			
in each stratum			
Calibration (raking)	1,273,467	1,161,977	2,435,444
(sex, age, province, ue-indicator)	(9,994)	(10,453)	(12,243)
Calibration (raking) – 2 PHASE	1,273,467	1,161,977	2,435,444
(sex, age, province, ue-indicator)	(9,993)	(10,452)	(12,243)
Calibration (linear)	1,273,037	1,162,438	2,436,475
(sex, age, province, ue-indicator)	(9,995)	(10,454)	(12,244)
Calibration (linear) – 2 PHASE	1,273,092	1,162,336	2,435,428
(sex, age, province, ue-indicator)	(9,994)	(10,453)	(12,243)

Table 1. Comparison of estimates for totals of LFS, March 1993. (Estimated standard errors in parentheses). (Contd.)

(b) Unemployment

Estimator	Male	Female	Total
Horvitz-Thompson (SRSWOR)	237,541 (8,352)	176,899 (7,268)	414,439 (10,754)
Horvitz-Thompson – 2 PHASE	244,555 (8,460)	182,355 (7,397)	426,910 (10,734)
POST-STRATIFICATION (sex, age, province)	**250,170** (8,331)	**171,922** (6,894)	**422,091** (10,914)
Post-stratification (sex, age, province, ue-indicator)	257,893 (6,863)	177,442 (6,093)	435,335 (6,802)
Post-stratification – 2 PHASE (sex, age, province) Uniform nonresponse model	252,343 (8,455)	178,972 (7,184)	431,315 (11,095)
Post-stratification – 2 PHASE (sex, age, province) Separate nonresponse model in each stratum	258,076 (8,255)	173,363 (6,826)	431,440 (10,712)
Calibration (raking) (sex, age, province, ue-indicator)	254,897 (6,852)	180,189 (6,125)	435,086 (6,722)
Calibration (raking) – 2 PHASE (sex, age, province, ue-indicator)	254,897 (6,851)	180,189 (6,124)	435,086 (6,722)
Calibration (linear) (sex, age, province, ue-indicator)	254,620 (6,849)	180,437 (6,128)	435,057 (6,722)
Calibration (linear) – 2 PHASE (sex, age, province, ue-indicator)	254,794 (6,850)	180,269 (6,125)	435,063 (6,722)

The estimated standard errors produced results that were predictable; the smallest standard errors are found when auxiliary information is available both at the sample and at the population level. Owing to a strong correlation between the register concept of being unemployed and our study variable, the unemployment estimates gain in precision. In using the same variable in the sample information only (Post-stratification – 2 PHASE), the sampling variance becomes substantially greater. By using slightly more information and by solving the response probabilities separately in each stratum the sampling variance can be marginally reduced. Nonetheless the model remains a poor one because real response behaviour should be checked much more carefully (op cit. Djerf and Väisänen 1993).

Table 2 also illustrates that the register variable used in the RHG model violates the assumption that no association between

Table 2. Response behaviour and unemployed job-seeker status in the Ministry of Labour's register.

Sample	Not in the Register for Unemployed Job Seekers	In the Register for Unemployed Job Seekers	Total
Respondents	10,714	1,378	12,092
	(88.6%)	(11.4%)	
Nonrespondents	577	135	712
	(81.0%)	(19.0%)	
Total	11,291	1,513	12,804
	(88.2%)	(11.8%)	

the study variables and the response generating mechanism exists. This is easily detected either from correlation between the two variables (0.8) or by comparing the estimated number of unemployed by the register concept, which stood at 440,458 for respondents, while the correct the figure was 457,453 (c.f. Djerf 1997).

A Pearson X^2 test shows that the assumption that the rows and columns are independent can be rejected (X^2 =36.9, df=1, p<0.001). The concept of being an unemployed job seeker is thus associated with the response mechanism in the Labour Force Survey sample.

The use of register information as a response homogeneity group for nonresponse adjustment did not change the result from that of the calibration in the expected manner. Hence the variance component resulted in $\hat{V}_2 = 0$.

It seems that the calibration estimator can utilize information on the respondents and on the nonrespondents when fairly high correlation with the study variable exists. It would be with examining whether this finding holds with poorer correlation. The possibility of bias also requires a further Monte Carlo study.

3.4. Discussion

Despite fairly clear statistical evidence some fears of introducing register information on unemployed in the LFS estimation still exist. The main argument against the use of the Job-Seekers Register is a major one: it is claimed that there are problems in updating the register data which may result in bias in estimates. The time lag in updating does in fact exist because the unemployed job-seekers are given fairly long intervals between the obligatory visits to the employment office. However, if the indicator for each individual and the respective population total of the indicator refer to the same period in time no discrepancy should occur. Hence measurement error is reduced if not removed completely. Secondly, it has been argued that part of those registered as unemployed are not viewed as unemployed according to the LFS concept, being instead claimed to have left the labour force because of long-term unemployment or given social problems. No clear evidence supporting such a claim has been obtained.

With the introduction of the revised LFS in January 1997, the estimation procedures were changed. The old post-stratified estimator was replaced by the calibration estimator. In order to maintain the population distribution as correct as possible, weights are first post-stratified, and then calibrated both on the marginal distributions of the population and the register indicator of the status of unemployed job-seeker. The indicator is further divided into three categories according to the duration of unemployment in the register. Thus the current LFS estimator exhausts a lot of available auxiliary information.

4. MONTE CARLO STUDY

4.1. The setup of the simulation study

4.1.1. General setup

The performance of various estimators were examined in Monte Carlo simulations. We created a population which is actually a pooled data set of LFS respondents from three consecutive months in 1993. A total of 1,000 independent samples were drawn. Each sample consisted of generated response and nonresponse subsets which mimic the true population structure. The average number of respondents was 1,000.

The main purpose of this experiment was to examine the statistical properties of various estimators when nonresponse was present and auxiliary information was available. The next chapters present the data creation procedure, the selected estimators, the Monte Carlo estimators, and finally the results of the study.

4.1.2. Study population

Original data set

In the Monte Carlo study we used a real data set for which both real and artificial variables were generated. The original data set was pooled from three months (March, April and May, 1993) of the Finnish Labour Force Survey. The elements, i.e. individual persons in those samples, were independent of each other and the pooled samples were mutually exclusive (non-overlapping). The total size of the pooled sample was 38,246.

However, our data matrix is not complete due to nonresponse. The number of nonrespondents was 2,351 (6.15 per cent) being fairly stabile over the selected months:

	Respondents	Nonres-pondents	Total
March	12,092	712 5.56%	12,804
April	11,930	780 6.14%	12,710
May	11,873	859 6.75%	12,732
Total	35,895	2,351 6.15%	38,246

Although the relative size of nonresponse is small, the distribution of nonrespondents differs from that of the respondents. Most study variables were lacking for nonrespondents but fortunately some background information from various register sources is available. We could therefore produce a model using a logistic regression model which predicted values for the 2 parameter for each person. The model is presented in Appendix 2.

The register dummy indicating whether the person is an unemployed job-seeker or not will be serve as a basic variable to create other indicators with different correlation with the unemployment measure. As indicated above, this indicator correlates strongly with the survey question involving unemployment ($r = 0.8$). Additional auxiliary variables were generated from the register indicator by assigning randomly selected individuals with "incorrect" values so that the correlation structure of the generated auxiliary variables varied between [0,0.8] with the increment of $|0.4|$.

Creating the population frame

Next we discarded the original nonrespondents and used the responding set (35,895 persons) as our sampling frame. Nonrespondents were generated by using model prediction and random selection. The model prediction was used as a base for the selection. However, the proportion of men among the nonrespondents would have grown too large had the model prediction been used as such. We therefore decided to perform random selection among the lowest response probabilities (predicted response probability less than 12 per cent) separately for men and women. The size of nonresponse was 6.15 per cent (2,206 persons). Thus the number of "respondents" according to the responding indicator was 33,689 its distribution being close to the original nonresponse distribution for most variables. Some of the distributions are compared in Appendix 3.

Figure 2. Data matrix of the population used in the Monte Carlo simulation study.

White area: nonrespondents in the LFS data
Light grey: respondents predicted to be nonrespondents in the simulation study
Dark grey: respondents predicted to be respondents in the simulation study

Case	X1	X2	...	Resp.	Est. Resp.	Y1	Y2	...
1	$x_{1,1}$	$x_{1,2}$		1	1	$y_{1,1}$	$y_{1,2}$	
2	$x_{2,1}$	$x_{2,2}$		1	1	$y_{2,1}$	$y_{2,2}$	
33689	$x_{33689,1}$	$x_{33689,2}$		1	1	$y_{33689,1}$	$y_{33689,2}$	
33690	$x_{33690,1}$	$x_{33690,2}$		1	0	$y_{33690,1}$	$y_{33690,2}$	
33691	$x_{33691,1}$	$x_{33691,2}$		1	0	$y_{33691,1}$	$y_{33691,2}$	
35895	$x_{35895,1}$	$x_{35859,2}$		1	0	$y_{35895,1}$	$y_{35895,2}$	
35896	$x_{356896,1}$	$x_{35896,2}$		0	–	–	–	
38246	$x_{38246,1}$	$x_{38246,2}$		0	–	–	–	

As a starting point we used the original register variable indicating whether a person is an unemployed job-seeker . The variable was called RUEP8 and has a stable Pearson correlation of 0.8 with the study variable of being unemployed. Next we created an indicator RUEP4 with Pearson correlation of 0.4 by selecting about 5,500 persons and assigning them with a randomly chosen indicator with the value of either 0 or 1. And finally a random assignment indicator RUE0 was created from the pseudo-random generator (uniform distribution), which supposedly should have zero correlation with the unemployment variable. The table below shows the correlations in the frame and among the model-based "respondents".

	Frame (n=35,895)		**Respondents (n=33,689)**	
	Labour Force	**Unemployed**	**Labour Force**	**Unemployed**
RUEP8	0.232	0.801	0.230	0.799
RUEP4	0.121	0.410	0.116	0.399
RUE0	0.010	–0.004	0.009	–0.003

4.1.3. Study variables

Our principal interest lay in the estimates for totals. The variables chosen for the study were the number of people in the labour force and the number of unemployed. Our aim was to evaluate the effect of both nonresponse and different correlations of the auxiliary variables to the study variables. We also evaluated them according to the gender and the age of the respondent.

The true values of those variables according to respondent's gender were

	Labour Force	Number of Unemployed	Unemployment Rate %
Male	11,880	2,300	19.36
Female	10,984	1,647	14.99
Total	22,864	3,947	17.26

In addition, we were interested in including the ratio of the two totals, the unemployment rate:

$$\hat{UR} = \frac{\hat{t}(UE)}{\hat{t}(LBF)} * 100$$

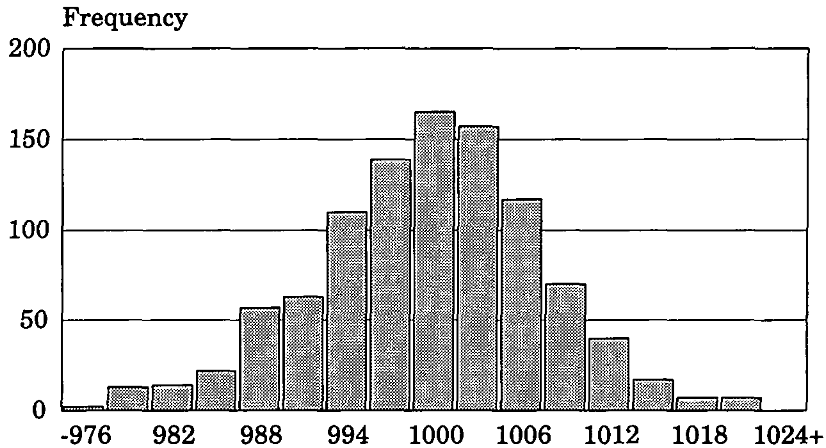
The Monte Carlo estimator for linear parameters are unbiased and consistent. However, for nonlinear parameters such as proportions (ratios in general), correlation coefficients, etc., the Monte Carlo estimator is known to be slightly biased, but remains consistent. Generally estimators which prove to be good for linear parameters are fairly good for functions of linear parameters, too. Here we are dealing with a situation where we compare the mean of the estimated proportions to the proportion of the estimated totals. Since in this Monte Carlo setting we only add up and divide with a fixed J (i.e. the weight is equal to 1), we can expect the bias problem to remain fairly small.

4.1.4. Monte Carlo samples

A total of 1,000 independent samples were selected from the frame by simple random sampling. We used a pseudo-random number generator of uniform distribution for selecting fixed size samples. The sample size was 1,065. The samples were selected without replacement.

We could expect that the mean of the nonresponse indicator would be close to 6.15 per cent over all 1,000 samples. Thus we would have on average 1,000 respondents. The mean proved to be 999.55 and the median 1,000. The distribution of respondents was normal but relatively flat, giving the impression of randomly varying nonresponse patterns. The smallest number of respondents was 973 and the largest 1,022. The estimated nonresponse rates varied from 4.0 to 8.6, corresponding rather well with the real nonresponse behaviour of the LFS (c.f. Djerf 1996a, 207–211).

Figure 3. Distribution of the number of respondents in 1,000 independent SRSWOR samples.



A few estimators were calculated for each sample, with the order is principally chosen according to the information contents:

Estimator	Variance formula
Horvitz-Thompson estimator	2.2 (2.6 with m in denom.)
Horvitz-Thompson estimator + weighting class	2.12
Post-stratification (sex, age*region)	2.15
Post-stratification (sex, age*region) + weighting class	2.16
Post-stratification (sex, age*region) + calibration (sex, age*region, ue-indicator)	2.29
Calibration (sex, age*region)	2.29
Calibration (sex, age*region, ue-indicator)	2.29

We constructed post-stratification in a manner comparable to the original post-stratification applied in the LFS, i.e. according to the respondent's sex, age and region. Since the size of the sample handled was relatively small, the number of post-strata was reduced to 24.

The post-strata were formed in the following way:

Variable	Categories	Description
Sex	2	M, F
Age	6	10 year groups starting from 15 to 24
Region	2	1 if the person was living in Southern Finland (Province of Uusimaa, Turku and Pori, Häme or in the Autonomous Territory of the Åland Islands) 2 elsewhere
Total	24	

The domains in estimation were closely related to stratification categories but were not identical: The first variable was the respondent's sex and the crossing variable was a combination between age and region: Four age categories (15–24, 25–49, 50–64, 65–74) were crossed to three regions (the Greater Helsinki Area, other parts of Southern Finland and the rest of Finland).

4.1.5. Program

We used a SAS macro program for calculating the point-estimates and their standard errors, CLAN v. 9510 (Andersson and Nordberg, 1992, 1995). Although a rather slow and not very practical program for handling a large number of calculations, the SAS macro program possesses a number of good qualities; the calibrated weights and the respective generalized regression estimator for sampling variance can be calculated inside the program itself. We were also able to program the whole sequence of estimators in one and the same run.

4.2. Measures of accuracy and precision

We were interested in evaluating both the bias and the precision properties of various estimators chosen for the study. The asymptotic properties of estimators can easily calculated in Monte Carlo studies by merely reestimating the parameter value as many times as needed. Let us denote the population parameter we are interested in (study variable) by Θ and its estimated value in the sample j by $\hat{\Theta}_j$. The Monte Carlo estimator of the parameter is then the arithmetic mean of the parameter estimates over all the samples.

$$\hat{\Theta}_{MC} = \sum_{j}^{J} \frac{\hat{\Theta}_j}{J} \tag{4.1}$$

Bias can be calculated directly from the Monte Carlo estimate: $\boldsymbol{B} = \hat{\Theta}_{MC} - \Theta$. Relative bias is achieved when the true value is plugged in the estimator above:

$$Rel-Bias = \frac{\hat{\Theta}_{MC} - \Theta}{\Theta} = \frac{1}{J} \frac{\sum_{j=1}^{J} (\hat{\Theta}_j - \Theta)}{\Theta} \tag{4.2}$$

The relative bias in the table below is given percentage points.

Since our estimators are only approximate unbiased estimators it is necessary to calculate the mean square error (MSE):

$$MSE(\hat{\Theta}) = \hat{V}(\hat{\Theta}) + (B(\hat{\Theta}))^2 = \sum_{j=1}^{J} \frac{(\hat{\Theta}_j - \Theta)^2}{J} \tag{4.3}$$

Here we apply the root mean square error, which is merely the square root of the MSE statistic:

$$RMSE(\hat{\Theta}) = \sqrt{MSE(\hat{\Theta})}$$

It is, however, also interesting to compare the estimated variance with the mean of the estimated variance of the study variable in question. For comparison we have used the RMSE from the Monte Carlo samples, the standard error estimates (i.e. square root of the sampling variance estimates from the Monte Carlo samples) and the mean of the estimated standard error of each estimator.

Domain estimates

Due to a restricted number of classifying variables chosen for the simulations the only combinations available to us are constructed according to gender, region and age. In each run we calculated point estimates and their standard errors for totals and in 24 domains; a combination of four age categories and three regions times gender. Here we concentrate separately on the domains of gender and the combination of age and region. The small sample size makes it very difficult to calculate good measures of precision over all domains of interest. For example, the mean of the absolute relative domain error might become too large for all estimators. For this reason only a given number of relative bias figures and coverage rates were calculated for the domains of interest.

Coverage rates

The coverage of the Monte Carlo experiment can be calculated in various ways. Here we have calculated the normal 95 per cent confidence interval from the estimated parameter and its standard deviation, e.g. $\left[\hat{y}-1.96\sqrt{\hat{V}(\hat{y})}, \hat{y}+1.96\sqrt{\hat{V}(\hat{y})}\right]$.

The standard deviation of a binary variable for the 95 per cent coverage rate is 0.006892. Thus the acceptance region for the coverage rate equals [93.65, 96.35].

4.3. Simulation results

4.3.1. Accuracy of the estimators

The point estimates and the respective relative bias for each estimator are presented in table 3 below. Here we analyse the results separately for each study variable. Auxiliary information containing either modest (+0.4) or zero correlation with the study variable were not included in the subsequent analysis. The improvement in results using the modest correlation +0.4 was so negligible that our decision may be justified.

Labour force participation

All but the Horvitz-Thompson estimator yield practically unbiased estimates for the study variable labour force participation. It is evident that information on the population distribution is necessary and sufficient for any estimator. For example, the relative bias of the original post-stratification and the respective calibration estimator is less than 0.1 per cent. When the register dummy of being an unemployed job-seeker is included, the point estimates of the labour force tend to increase slightly. In the case of the H-T estimator the negative bias is reduced but in other cases the use of auxiliary information causes some extra bias. The bias of the estimates still remained below a half per cent, though.

Unemployment

The lack of the auxiliary information causes serious negative bias in unemployment estimates. The largest relative bias stays from the H-T estimator (over 5 per cent), and is about 4.5 per cent for estimators that use information on the population distribution (post-stratification and calibration 1).

The use of register information on the status of an unemployed job-seeker clearly improves the accuracy of the point estimates. The use of sample level information only (weighting-class estimator for nonresponse) decreases the relative bias to about one per cent. Where the register information is used at the population level (calibrated estimators), the relative bias is further decreased to about 0.6 per cent.

Unemployment rate

Estimators which perform well both for labour force participation and the status of unemployment have fairly modest bias also for unemployment rates. In this case one of the problems connected with the ratio estimator becomes evident; namely the H-T estimator with the weighting class adjustment for nonresponse provides the least biased estimate for the unemployment rate although it is not least biased either for the nominator or the denominator.

Table 3. Accuracy of various estimators in Monte Carlo simulation under nonresponse.

Estimator		**Labour force**	**Unemployment**	**Unemployment rate**
		(t = 22,864)	(t = 3,947)	(R = 17.26)
Horvitz-Thompson	$\hat{t}_{MC}$	22,517.67	3,739.98	16.61
	Rel-Bias	–1.51	–5.24	–3.78
Horvitz-Thompson + weighting class	$\hat{t}_{MC}$	22,594.41	3,908.66	17.30
	Rel-Bias	–1.18	–0.97	0.22
Post-stratification (sex, age, region)	$\hat{t}_{MC}$	22,881.03	3,774.21	16.49
	Rel-Bias	0.07	–4.38	–4.45
Post-stratification (sex, age, region) + weighting class	$\hat{t}_{MC}$	22,918.22	3,910.99	17.06
	Rel-Bias	0.24	–0.91	–1.15
Post-stratification (sex, age, region) + Calibration (sex, age, region, ue-indicator)	$\hat{t}_{MC}$	22,926.60	3,921.40	17.11
	Rel-Bias	0.27	–0.65	–0.90
Calibration (sex, age, region)	$\hat{t}_{MC}$	22,878.16	3,776.76	16.51
	Rel-Bias	0.06	–4.31	–4.38
Calibration (sex, age, region, ue-indicator)	$\hat{t}_{MC}$	22,924.40	3,922.77	17.11
	Rel-Bias	0.26	–0.61	–0.86

Conclusions

It can be concluded that in the Labour Force Survey it is not possible to obtain unbiased estimates on unemployment without using proper auxiliary information. It is obvious that labour force participation can be accurately estimated when sufficient information on population structure is available. A distribution of population according to the gender, age and region fulfills these requirements. The result gives support to earlier findings by Djerf and Väisänen who claimed that the current post-stratified estimator of the LFS is ample for large domains of study (1993).

When estimating unemployment and related study variables, however, more information would be necessary. It seems that the use of register information on unemployed job-seekers improves such estimates and reduce bias to an acceptable level. The disadvantage of using strongly correlating information is also evident, calling for slight bias in labour force participation which in this case amounts to positive bias in employment estimates. The magnitude of the bias remains fairly small.

4.3.2. Mean square error and variance

Since the estimators were found to contain bias we will present both the MSE and the variance estimates of the linear parameters based on the Monte Carlo simulation.

Here the role of the sampling and estimation design as well as use of information becomes crucial. Horvitz-Thompson estimator provides the largest mean square errors and variances in general. Post-stratification and calibration yield estimates which are similar to each other.

If we mow use auxiliary information as a response homogeneity model we do not gain in precision, but if we use the information at maximum we can improve the precision significantly. This is the case when we look at post-stratification and use the auxiliary information as a weighting class. Variance for the number of unemployed is about 110,600 while it is reduced to 46,500 when we use the same variable as a simple indirect post-stratification variable (Deville et al. 1993). Thus we can improve precision by 60 per cent by merely using information to the fullest.

Table 4. Mean square error and variance for various estimators in the Monte Carlo simulation

Estimator		Labour force	Unemployment	Unemployment rate
Horvitz-Thompson	*MSE*	395,023	154,461	0.00025311
	VAR	285,082	111,604	0.00021056
Horvitz-Thompson + weighting class	*MSE*	344,664	112,368	0.00020826
	VAR	271,982	110,898	0.00020811
Post-stratification (sex, age, region)	*MSE*	164,364	142,303	0.00026329
	VAR	164,074	112,448	0.00020420
Post-stratification (sex, age, region) + weighting class	*MSE*	166,492	111,921	0.00020427
	VAR	163,560	110,625	0.00020033
Post-stratification (sex, age, region) + Calibration (sex, age, region, ue-indicator)	*MSE*	161,112	47,193	0.00009239
	VAR	157,193	46,537	0.00008996
Calibration (sex, age, region)	*MSE*	163,013	140,789	0.00025941
	VAR	162,812	111,809	0.00020230
Calibration (sex, age, region, ue-indicator)	*MSE*	159,502	47,034	0.00009146
	VAR	155,854	46,447	0.00008926

4.3.3. Standard error estimates and coverage rates

The mean of the estimated standard errors can be compared to the standard deviation and RMSE estimate of the Monte Carlo estimates. In general, the mean of the estimated standard errors was close to the Monte Carlo standard deviation. Where the estimators were biased, the mean of the standard error estimate was quite distant from the RMSE, which means the confidence intervals become too close.

In the case of nearly unbiased estimates two different phenomena could be discerned. Generally the mean of the standard error estimates was greater than the standard deviation of the estimates. (See the labour force estimates), which means the confidence intervals can become slightly too distant. When using information on registered unemployment, instead, the calibrated estimates appear to result in confidence intervals that are slightly too close because the mean of the standard error estimates are smaller than the respective standard deviation. Table 5 below confirms the assumption empirically. Estimators which result in modest bias but contain large standard errors actually perform better than unbiased estimators with high efficiency.

Table 5. Coverage rates of various estimators in Monte Carlo simulation. (Asterisk indicates estimates which did not fall into the 95 per cent acceptance range)

	Labour force	Unemployment	Unemployment rate
Horvitz-Thompson	91.0 *	90.2 *	92.7 *
Horvitz-Thompson + weighting class	93.4 *	95.7	95.6
Post-stratification (sex, age, region)	96.1	91.8 *	92.0 *
Post-stratification (sex, age, region) + weighting class	95.8	95.4	94.8
Post-stratification (sex, age, region) + calibration (incl. ue-indicator)	96.0	93.9	94.0
Calibration (sex, age, region)	96.2	91.4 *	92.0 *
Calibration (sex, age, region, ue-indicator)	95.5	93.7	94.2

Table 6. Mean square error and variance for various estimators in the Monte Carlo simulation

Estimator		Labour force	Unemployment	Unemployment rate
Horvitz-Thompson	*RMSE*	628.51	393.01	0.0159
	STD_{MC}	524.48	334.07	0.0145
	SE	541.29	341.75	0.0146
Horvitz-Thompson + weighting class	*RMSE*	587.08	335.21	0.0144
	STD_{MC}	521.52	333.01	0.0144
	SE	538.36	342.91	0.0146
Post-stratification (sex, age, region)	*RMSE*	405.42	377.23	0.0162
	STD_{MC}	405.06	335.33	0.0143
	SE	422.62	344.51	0.0146
Post-stratification (sex, age, region) + weighting class	*RMSE*	408.03	334.55	0.0143
	STD_{MC}	404.43	332.60	0.0142
	SE	421.18	342.85	0.0145
Post-stratification (sex, age, region) + Calibration (sex, age, region, ue-indicator)	*RMSE*	401.39	217.24	0.0096
	STD_{MC}	396.48	215.72	0.0095
	SE	413.42	212.50	0.0094
Calibration (sex, age, region)	*RMSE*	403.75	375.22	0.0161
	STD_{MC}	403.50	334.38	0.0142
	SE	418.86	340.99	0.0145
Calibration (sex, age, region, ue-indicator)	*RMSE*	399.38	216.87	0.0096
	STD_{MC}	394.78	215.52	0.0094
	SE	409.78	210.14	0.0093

4.4. Domain estimates

4.4.1. Bias consideration

Labour force

Estimators which use auxiliary information on the population distribution perform quite well. Even the relative bias of the Horvitz-Thompson estimator remains small when females are concerned. Auxiliary information tends to induce some bias in estimates but the magnitude remains quite small.

Unemployed

The use of auxiliary register information on the unemployed jobseekers seems to also reduce bias in the gender domain. However, there is some evidence of "pushing"; the relative bias of women tends to grow from negative to positive whereas the bias of men is constantly reduced but remains negative.

Table 7. Relative bias of various estimators for the study variables of labour force participation and unemployment.

	Labour force		Unemployment	
Estimator	**Male**	**Female**	**Male**	**Female**
Horvitz-Thompson	–3.24	0.36	–7.60	–1.96
Horvitz-Thompson + weighting class	–2.80	0.57	–3.45	2.49
Post-stratification (sex, age, region)	0.03	0.12	–5.29	–3.11
Post-stratification (sex, age, region)+ weighting class	0.27	0.21	–0.81	–1.06
Post-stratification (sex, age, region)+ calibration (sex, age, region, ue-ind.)	0.25	0.30	–1.67	0.78
Calibration (sex, age, region)	–0.16	0.30	–5.42	–2.77
Calibration (sex, age, region, ue-ind.)	0.06	0.48	–1.90	1.18

The results in table 7 prompted us to investigate the role of gender in more detail in domain estimation. The register indicator was calculated separately for men and women and applied to the calibration estimators. Table 8 shows that the change was worth doing: the relative bias is clearly reduced when information is applied on this domain level.

Table 8. Relative bias of various estimators for the study variables of labour force participation and unemployment.

Estimator	Labour force Male	Labour force Female	Unemployment Male	Unemployment Female
Horvitz–Thompson	–3.24	0.36	–7.60	–1.96
Horvitz–Thompson + weighting class	–2.80	0.57	–3.45	2.49
Post-stratification (sex, age, region)	0.03	0.12	–5.29	–3.11
Post-stratification (sex, age, region)+ weighting class	0.27	0.21	–0.81	–1.06
Post-stratification (sex, age, region)+ calibration (sex, age, region, ue-indicator)	0.25	0.30	–1.67	0.78
Post-stratification (sex, age, region)+ calibration (sex, age, region, ue-indicator by sex)	0.30	0.24	–1.02	–0.12
Calibration (sex, age, region)	–0.16	0.30	–5.42	–2.77
Calibration (sex, age, region, ue-indicator)	0.06	0.48	–1.90	1.18
Calibration (sex, age, region, ue-indicator by sex)	0.14	0.40	–1.05	–0.00

We can, however, go further and check whether the use of more detailed information also improves the estimates of other domains. In table 9 the domains cut across the structure of the use of auxiliary information in estimation.

Table 9. Relative bias of various estimators for the study variables labour force participation and unemployment for domains which do not coincide with strata or marginal distributions used in calibration

(a) Region: Greater Helsinki Area; Age category: 15–24

Estimator	Labour force Male	Labour force Female	Unemployment Male	Unemployment Female
Horvitz–Thompson	–11.35	–11.42	–11.74	–12.74
Horvitz–Thompson + weighting class	–11.12	–11.52	–7.45	–9.46
Post-stratification (sex, age, region)	–4.87	–8.01	–5.52	–9.72
Post-stratification (sex, age, region)+ weighting class	–4.82	–8.04	–1.33	–8.95
Post-stratification (sex, age, region)+ calibration (sex, age, region, ue-indicator)	–4.85	–8.13	–1.52	–6.65
Post-stratification (sex, age, region)+ calibration (sex, age, region, ue-indicator by sex)	–4.81	–8.14	–0.76	–7.62
Calibration (sex, age, region)	–5.61	–8.68	–6.21	–10.08
Calibration (sex, age, region, ue-indicator)	–5.59	–8.80	–2.52	–7.08
Calibration (sex, age, region, ue-indicator by sex)	–5.55	–8.81	–1.61	–8.15

Table 9. Relative bias of various estimators for the study variables labour force participation and unemployment for domains which do not coincide with strata or marginal distributions used in calibration (Contd.)

(b) Region: Greater Helsinki Area; Age category: 50–64

	Labour force		Unemployment	
Estimator	**Male**	**Female**	**Male**	**Female**
Horvitz–Thompson	–8.13	–7.64	–10.59	–10.83
Horvitz–Thompson + weighting class	–7.99	–7.78	–6.91	–6.91
Post-stratification (sex, age, region)	–3.31	–5.23	–5.68	–9.09
Post-stratification (sex, age, region)+ weighting class	–3.19	–5.37	–2.41	–8.63
Post-stratification (sex, age, region)+ calibration (sex, age, region, ue-indicator)	–3.20	–5.15	–2.43	–5.26
Post-stratification (sex, age, region)+ calibration (sex, age, region, ue-indicator by sex)	–3.20	–5.14	–2.07	–6.17
Calibration (sex, age, region)	–3.21	–5.88	–5.86	–10.00
Calibration (sex, age, region, ue-indicator)	–3.11	–5.80	–2.77	–6.17
Calibration (sex, age, region, ue-indicator by sex)	–3.11	–5.78	–2.30	–7.14

The overall impression is quite similar to earlier findings in the use of auxiliary information in general. However, now the estimates for one domain (respondent's gender) are not improved although the information is given in that domain level (i.e. ue-indicator by sex). It can therefore be said that auxiliary information should be used as comprehensively as possible when the most important domains are expected to gain in precision.

4.4.2. Coverage rates of the domain estimates

We evaluated the accuracy and precision of the domain estimates by coverage rates. Here the estimators refer only to the domain of the respondent's gender.

Table 10 demonstrates that for labour force participation all estimators other than the H-T estimator perform in a consistent manner. Instead, the unemployment estimates generate slightly controversial results. Estimators that use the most detailed information, i.e. register-based unemployment separately for both sexes (post-stratification combined with calibration and calibration), are controversial. In the case of the calibration estimator all estimates stayed within the acceptance region, whereas a similar estimator that uses even more information (post-stratification combined with calibration) failed altogether three times.

Table 10. Coverage rates of various estimators by respondent's gender in Monte Carlo simulation
(Asterisk indicates estimates which did not fall into the 95 per cent acceptance region)

(a) Labour force	**Male**	**Female**	**Total**
Horvitz-Thompson	88.6 *	95.2	91.0 *
Horvitz-Thompson + weighting class	90.4 *	94.3	93.4 *
Post-stratification (sex, age, region)	95.2	95.4	96.1
Post-stratification (sex, age, region) + weighting class	94.8	95.5	95.8
Post-stratification (sex, age, region) + calibration (sex, age, region, ue-indicator)	94.7	95.2	96.0
Post-stratification (sex, age, region) + calibration (sex, age, region, ue-indicator by sex)	94.8	96.0	95.9
Calibration (sex, age, region)	94.7	94.9	96.2
Calibration (sex, age, region, ue-indicator)	93.8	94.4	95.5
Calibration (sex, age, region, ue-indicator by sex)	94.1	94.6	95.6
(b) Unemployment			
Horvitz-Thompson	87.5 *	94.4	90.2 *
Horvitz-Thompson + weighting class	92.9 *	96.2	95.7
Post-stratification (sex, age, region)	91.0 *	93.5 *	91.8 *
Post-stratification (sex, age, region) + weighting class	94.7	95.3	95.4
Post-stratification (sex, age, region) + calibration (sex, age, region, ue-indicator)	93.5 *	95.2	93.9
Post-stratification (sex, age, region) + calibration (sex, age, region, ue-indicator by sex)	93.5 *	96.0	93.6 *
Calibration (sex, age, region)	90.4 *	94.1	91.4 *
Calibration (sex, age, region, ue-indicator)	92.4 *	95.2	93.7
Calibration (sex, age, region, ue-indicator by sex)	94.6	95.3	93.8
(c) Unemployment rate			
Horvitz-Thompson	91.8 *	94.4	92.7 *
Horvitz-Thompson + weighting class	95.6	96.2	95.6
Post-stratification (sex, age, region)	91.4 *	93.5 *	92.0 *
Post-stratification (sex, age, region) + weighting class	94.9	95.3	94.8
Post-stratification (sex, age, region) + calibration (sex, age, region, ue-indicator)	94.1	95.2	94.0
Post-stratification (sex, age, region) + calibration (sex, age, region, ue-indicator by sex)	93.6 *	96.0	94.4
Calibration (sex, age, region)	90.1 *	94.1	92.0 *
Calibration (sex, age, region, ue-indicator)	94.1	95.2	94.2
Calibration (sex, age, region, ue-indicator by sex)	94.3	95.3	93.9

4.4.3. Conclusions

It seems evident that the information-intensive methods perform best when measuring bias and precision also for the domain level. In our case the nonresponse structure was not the same according the gender of the respondent. Hence the use of auxiliary information for overall totals did not improve the domain estimates as much as one could have expected. On the contrary, these rather strong methods induced some extra bias in estimates. We can thus conclude that information should always be targeted as precisely as possible to the correct domains.

5. CONCLUSIONS AND FURTHER STUDY

The aim of the study was to compare a number of fairly straightforward estimation methods when unit nonresponse is present in the data set. Unit nonresponse is a prevailing and probably increasing problem in all kinds of surveys. Traditionally only "cosmetic" adjustments were used to decrease the possibility of biased estimates in some domains of surveys. This problem was given new insight in the early 1980s when empirical nonresponse modelling was introduced, although post-stratification and raking ratio estimators had been used since the 1960s. The calibration estimator was introduced in the 1990s and was thought to provide new opportunities for dealing nonresponse bias.

The estimators in this study are generally used in sample surveys. They differ in their ability to use information; while the Horvitz-Thompson estimator needs no information other than inclusion probability, calibration estimators can utilize a lot of information both at the sample level and at the population level. The key issue is, of course, whether the information available is suitable for adjusting for nonresponse. It should also be borne in mind fairly strong correlation either between the auxiliary information and the response mechanism or between the auxiliary information and the study variable(s) or both must exist.

In this study we compared estimators using empirical data sets where nonresponse was not a major problem in general. However, nonresponse was not distributed in a way that the traditional way of dividing the data set into two strata according to the response could be justified. The nonresponse was in fact found to be associated with one of the major parameters of the survey, namely unemployment. In such circumstances it may be expected that the methods that are incapable of capturing the nonresponse mechanism will fail to provide unbiased estimates. This was confirmed in our findings.

The Horvitz-Thompson estimator performed clearly worst, while the traditional post-stratification estimator and the calibration estimator using only information not directly linked to the response mechanism did not perform much better for the parameter of unemployment. Nonetheless, they were much more accurate and precise than the H-T estimator in all other respects.

The introduction of an register indicator that correlates strongly both with the study variable of unemployment and with the response mechanism improved the problem of bias substantially. The type of information usage had clear effects on the precision of the estimates; precision was not improved very much when information was used at the sample level (weighting-class estimator), whereas when information was available also at the population level large gains in precision could be achieved.

The difference between the post-stratified and the calibration estimator was minor, which was to be expected because in categorical auxiliary information it is only at the level of availability that information differs. Although both estimators belong to generalized regression estimators, the calibration estimator has advantage over post-stratification in that the ratio type estimator can be used at the same time as the total counts. Both perform well when information is good.

The estimates for domains revealed that the use of auxiliary information should be targeted as far as possible when there was clear evidence of unequal nonresponse mechanisms at the levels of the pertinent domain. In this study we found that the calibration estimator can introduce some bias in such cases.

It can be generally said that the efforts of introducing unit nonresponse adjustment is worthwhile provided that the efforts are correct. The calibration estimator is a very powerful tool in this respect since it can also cover the explicit nonresponse modelling approach.

There is definitely room for further study in the field of unit nonresponse adjustment. In this study we concentrated on very simple cases and simple sampling designs. The results need to be confirmed with more complex (multistage) designs and especially in panel surveys. Another important task would be to determine how these types of estimators could be applied in nonresponse problems prevailing in business surveys.

REFERENCES

Andersson, C., and L. Nordberg (1992). *CLAN – ett SAS-program för skattningar av medelfel.* SCB, R&D Report 1992:11.

Andersson, C., and L. Nordberg (1994). A Method for variance estimation of non-linear functions of totals in surveys – theory and a software implementation. *Journal of Official Statistics,* Vol. 10, No. 4, 395–406.

Andersson, C., and L. Nordberg (1996). New features of CLAN9510. *Statistics Sweden* (mimeo in Swedish).

Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics,* Vol. 4, No. 3, 251–260.

Bethlehem, J.G., and W.J. Keller (1987). Linear Weighting of Sample Survey Data. *Journal of Official Statistics,* Vol. 3, No. 2, 141–153.

Breidt, F.J., and W.A. Fuller (1993). Regression Weighting for Multiphase Samples. *Sankhya,* Vol. 55, Ser. B, No. 3, 297–309.

Cassel, C.-M., C.-E. Särndal and J.H. Wretman (1983). Some Uses of Statistical Models in Connection with the Nonresponse Problem (with discussion), in *Incomplete Data in Sample Surveys,* eds. W.G. Madow, I. Olkin and D.B. Rubin, Vol 3., 143–170. New York: Academic Press.

Cochran, W.G. (1977). *Sampling Techniques,* 3rd ed. New York: John Wiley & Sons.

Deming, W.E. (1943). *Statistical Adjustment of Data.* New York: Wiley (reprint 1964 by Dover Publications).

Deming, W.E., and F.F. Stephan (1940). On least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics,* Vol. 11, 427–444.

Deville, J-C. (1991). A Theory of Quota Samples. *Survey Methodology,* Vol. 17, 163–181.

Deville, J-C., and F. Dupont (1993). Calage et redressement de la non-réponse totale. *Journées de Méthodologie.*

Deville, J.-C., and C.-E. Särndal (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association,* Vol. 87, No. 418, 376–382.

Deville, J.-C., C.-E. Särndal and O. Sautory (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association,* Vol. 88, No. 423, 1013–1020.

Djerf, K. (1996a). Trends of nonresponse in the surveys conducted by Statistics Finland: Labour Force Survey, in S. Laaksonen (ed.) *International Perspectives on Nonresponse,* Statistics Finland, Research Reports 219, 207–211,

Djerf, K. (1996b). Properties of the GREG estimators. Empirical findings from the Finnish LFS. (Unpublished mimeo, Statistics Finland.)

Djerf, K. (1997). Effects of Post-stratification on the Estimates of the Finnish Labour Force Survey. *Journal of Official Statistics,* Vol. 13, No. 1, 29–39.

Djerf, K. and P. Väisänen (1993). Effects of Post-stratification on the Estimates of the Finnish Labour Force Survey. *Bulletin of the International Statistical Institute, 49th session, Contributed Papers, Book* 1. Firenze, 375–376.

Dupont, F. (1994). Calibration Used as a Nonresponse Adjustment. *I.N.S.E.E.* (unpublished mimeo, in French).

Dupont, F. (1995). Alternative Adjustments Where There Are Several Levels of Auxiliary Information. *Survey Methodology,* Vol. 21, No. 2, 125–135.

Ekholm, A., and S. Laaksonen (1990). Reweighting via Nonresponse Modeling in the Finnish Household Budget Survey. *University of Helsinki, Department of Statistics, Research Reports* 68 2nd ed.

Ekholm, A., and S. Laaksonen (1991). Weighting via Response Modeling in the Finnish Household Budget Survey. *Journal of Official Statistics,* Vol. 7, No. 3, 325–337.

Estevao, V., M. Hidiroglou and C.-E. Särndal (1995). Methodological Principles for a Generalized Estimation System at Statistics Canada, *Journal of Official Statistics,* Vol. 11, No. 2, 181–204.

Hidiroglou, M. A., and C.-E. Särndal (1995). Use of Auxiliary Information for Two-phase Sampling. *American Statistical Association, 1995 Proceedings of the Section on Survey Research Methods, Vol II, 873–878.*

Incomplete Data in Sample Surveys, Vol. 1 Report and Case Studies (Eds. Madow, Nisselsohn and Olkin); Vol. 2 Theory and Bibliographies (Eds. Madow, Olkin and Rubin); Vol. 3 Proceedings of the Symposium (Eds. Madow and Olkin). New York: Academic Press.

Isaki, G., and W.A. Fuller (1982). Survey Design under the Regression Superpopulation Model. *Journal of the American Statistical Association,* Vol. 77, 89–96.

Kalton, G. (1983). *Compensating for Missing Survey Data.* The University of Michigan, Institute for Social Research, Research Report Series.

Kalton, G., and D.S. Maligalig (1991). A Comparison of Methods of Weighting Adjustment for Nonresponse. (With discussion). *Bureau of the Census, 1991 Proceedings of the Annual Research Conference,* 409–447.

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

Kott, P.S. (1994). A Note on Handling Nonresponse in Simple Surveys. *Journal of the American Statistical Association,* Vol. 89 No. 426, 693–696.

Laaksonen, S. (1992). *Handling Household Survey Non-response Data.* The Finnish Statistical Society, Statistical Studies 13. Helsinki.

Lehtonen, R., and E. Pahkinen (1995). *Practical Methods for Design and Analysis of Complex Samples.* Chichester: John Wiley & Sons.

Lehtonen, R., and A. Veijanen (1996). The Use of Administrative Records as Auxiliary Information in the Estimation of Unemployment Figures. Paper presented at the 5th Independent Conference of the International Association for Official Statistics, 2–5 July 1996, Reykjavik Iceland.

Lehtonen, R., and A. Veijanen (1998). Logistic Generalized Regression Estimators, Survey Methodology, Vol. 24, 51–55.

Little, R.J.A. (1983a). The Ignorable Case, in *Incomplete Data in Sample Surveys,* eds. W.G. Madow, I. Olkin and D.B. Rubin, Vol 2., 341–382. New York: Academic Press.

Little, R.J.A. (1983b). The Nonignorable Case, in *Incomplete Data in Sample Surveys,* eds. W.G. Madow, I. Olkin and D.B. Rubin, Vol 2., 383–413. New York: Academic Press.

Little, R.J.A. (1986). Survey Nonresponse Adjustments for Estimates of Means. *International Statistical Review,* Vol. 54, 139–157.

Little, R.J.A., and D.B. Rubin (1987). *Statistical Analysis with Missing Data.* New York: John Wiley & Sons.

Lundström, S. (1997). *Calibration As a Standard Method for Treatment of Nonresponse.* Doctoral Dissertation, University of Stockholm, Department of Statistics. Stockholm: Akademitryck AB.

Measurement Errors in Surveys (1991). (Eds. P. Biemer, R. Groves, L. Lyberg, N.A. Mathiowetz and S. Sudman) . New York: John Wiley & Sons.

Oh, J.L., and. F. Scheuren (1983). Weighting Adjustment for Unit Nonresponse, in *Incomplete Data in Sample Surveys,* eds. W.G. Madow, I. Olkin and D.B. Rubin, Vol 2., 143–184. New York: Academic Press.

Rosenbaum, P.R. (1987). Model-Based Direct Adjustment. *Journal of the American Statistical Association,* Vol. 82, No 398, 387–394.

Rosenbaum, P.R., and D.B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika,* Vol. 40, 41–55.

Sautory, O. (1993). La macro CALMAR. Redressement d'un échantillon par calage sur marges. I.N.S.E.E., Série des documents de travail *n*° F 9310. Paris: I.N.S.E.E.

Smith, T.M.F. (1990). Post-stratification. *The Statistician,* 40, 315–323.

Särndal, C.-E., and B. Swensson (1985). Incorporating nonresponse modelling in a general ran randomisation theory approach. *Bulletin of the International Statistical Institute, 45th session, Invited Papers,* Vol. 1, Amsterdam, 243–257.

Särndal, C.-E., and B. Swensson (1987). A General View of Estimation for two Phases of Selection with Applications to Two-Phase Sampling and Nonresponse. *International Statistical Review,* Vol. 55, No. 3, 279–294.

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

Wolter, K. (1985). *Introduction to Variance Estimation.* New York: Springer-Verlag.

Appendix 1.

Estimators for domains

The presentation of various estimators above did not contain the description of domain estimators but we will give a short introduction here mainly following the presentation by Särndal et al. (1992, ch. 10).

Let us divide our population U into D mutually disjoint subpopulations which we call domains:

$$U = \bigcup_{d=1}^{D} U_d, \text{ subject to } N = \sum_{d=1}^{D} N_d \tag{A1.1}$$

The ordinary expansion estimator for a total in domain d is simply

$$\hat{t}_{d\pi} = \sum_{s_d} \frac{y_k}{\pi_{k,}} \quad k \in s_d \tag{A1.2}$$

In general one has to consider that also population will be unknown and has to be estimated by

$$\hat{N}_{d\pi} = \sum_{s_d} \frac{1}{\pi_{k,}} \quad k \in s_d \tag{A1.3}$$

and the corresponding variance estimator is

$$\hat{V}(\hat{t}_d) = \sum \sum_{s_d} (\pi_k \pi_l - \pi_{kl}) \frac{y_l}{\pi_k} \frac{y_k}{\pi_l} \tag{A1.4}$$

In the case of SRSWOR design the variance estimator is

$$\hat{V}_{SRSWOR}(\hat{t}_d) \doteq N^2 \frac{1 - n/N}{n} \frac{n_d}{n} \left(\frac{1}{n_d - 1} \sum_{s_d} (y_k - \bar{y}_d^2) + (1 - \frac{n_d}{n}) \bar{y}_d^{-2} \right) \tag{A1.5}$$

(Särndal et al. 1992, 393). When population sizes in domains are known the variance will be smaller.

The domain estimators for other designs are solved in the similar manner. We present, however, the domain estimators for the GREG estimator. Here we assume in the general case where the auxiliary information vector and the domains of interest do not completely coincide but cross each other:

$$\hat{t}_{GREG,d} = \sum_{p=1}^{P} \sum_{s_d} d_k g_k y_k$$

$$\text{where} \quad g_k = 1 + q_k(\mathbf{x}_d - \hat{\mathbf{x}}_{d\pi})' \left(\sum_{s_d} d_k q_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \mathbf{x}_k, \quad \text{and } k \in d \tag{A1.6}$$

Here the index p denotes for the P-dimensional vector of parameters in the generalized regression equation. We have to assume that the total of vector $\mathbf{X}_p$ is known. The variance estimator looks quite the same as in the case of grand total. However, there are terms which inflate the variance always when the domain d crosses the parameter p:

$$\hat{V}(\hat{t}_{GREG,d}) = \sum \sum_{s_d} \frac{\pi_k \pi_l - \pi_{kl}}{\pi_{kl}} (w_k e_k)(w_l e_l)$$

$$\text{where } e_{dk} = \begin{cases} y_k - \mathbf{x}_{pk} \hat{\mathrm{B}}_{dp}, & \text{when } k \in s_d,\ k \in U_d \\ -\mathbf{x}_{pk} \hat{\mathrm{B}}_{dp}, & \text{otherwise} \end{cases} \tag{A1.7}$$

(Estevao et al. 1995, 194).

One has to note that similar inflation in variance takes place also in stratification estimator for such domains which cross the strata and do not contain the elements in the strata completely. In our case the model group and the domain will coincide. Therefore we cannot expect to see very large sampling variances.

Appendix 2.

A logit model for explaining response and nonresponse in the data set of the study

Variable	Parameter Estimate	Standard Error
Intercept	–5.497	0.203
Sex = female	–0.394	0.044
Region 1	0.440	0.068
2	0.462	0.083
3	0.160	0.072
4	–0.238	0.071
5	0.216	0.085
Municipality: big town	0.287	0.055
rural	–0.217	0.057
Mother tongue Swedish	–0.073	0.095
other than Finnish or Swedish	0.979	0.161
Marital status = married	–0.627	0.048
Age	0.142	0.011
Age squared	–0.002	0.000
Profession 1991 Farmer	–0.055	0.072
Entrepreneur or unknown	0.528	0.082
Type of pension 1991	0.156	0.067
Unemployed job seeker	0.286	0.062
Interviewing in April	0.098	0.054
May	0.202	0.053

	Intercept	Model	^{2}for covariates
Akaike's Information Criterion	17671	16684	–
– 2 log Likelihood	17669	16644	1025 (p<.0001)
McFadden's ρ^2 (likelihood ratio index for the fitted logistic model)		0.058	

Appendix 3.

Empirical and predicted response probabilities according to some demographic characteristics

		LFS Basic data	Predicted data
Sex	male	92.6	92.6
	female	95.1	95.0
Age category	15 – 19	97.4	99.6
	20 – 24	94.5	95.3
	25 – 29	93.1	92.8
	30 – 34	92.7	91.8
	35 – 39	92.0	91.4
	40 – 44	92.6	91.0
	45 – 49	92.5	90.6
	50 – 54	93.1	91.5
	55 – 59	93.3	93.4
	60 – 64	94.6	96.2
	65 – 69	97.1	98.9
	70 – 74	97.0	99.8
Province	Greater Helsinki Area	90.3	86.0
	Other Uusimaa	92.8	90.9
	Turku and Pori	93.9	94.0
	Häme	94.8	94.0
	Kymi	94.8	96.1
	St. Michels	94.3	96.5
	Pohjois-Karjala	95.9	96.6
	Kuopio	95.8	98.0
	Keski-Suomi	96.5	97.1
	Vasa	95.5	98.1
	Oulu	93.5	93.2
	Lapland	95.2	94.5
	Åland	93.1	94.0

TUTKIMUKSIA-SARJA
RESEARCH REPORTS SERIES

Tilastokeskus on julkaissut Tutkimuksia v. 1966 alkaen, v. 1990 lähtien ovat ilmestyneet seuraavat:

164. **Henry Takala,** Kunnat ja kuntainliitot kansantalouden tilinpidossa. Tammikuu 1990. 60 s.

165. **Jarmo Hyrkkö,** Palkansaajien ansiotasoindeksi 1985=100. Tammikuu 1990. 66 s.

166. **Pekka Rytkönen,** Siivouspalvelu, ympäristöhuolto ja pesulapalvelu 1980-luvulla. Tammikuu 1990. 70 s.

167. **Jukka Muukkonen,** Luonnonvaratilinpito kestävän kehityksen kuvaajana. 1990. 119 s.

168. **Juha-Pekka Ollila,** Tieliikenteen tavarankuljetus 1980-luvulla. Helmikuu 1990. 45 s.

169. **Tuovi Allén – Seppo Laaksonen – Päivi Keinänen – Seija Ilmakunnas,** Palkkaa työstä ja sukupuolesta. Huhtikuu 1990. 90 s.

170. **Ari Tyrkkö,** Asuinolotiedot väestölaskennassa ja kotitaloustiedustelussa. Huhtikuu 1990. 63 s.

171. **Hannu Isoaho – Osmo Kivinen – Risto Rinne,** Nuorten koulutus ja kotitausta. Toukokuu 1990. 115 s.

171b. **Hannu Isoaho – Osmo Kivinen – Risto Rinne,** Education and the family background of the young in Finland. 1990. 115 pp.

172. **Tapani Valkonen – Tuija Martelin – Arja Rimpelä,** Eriarvoisuus kuoleman edessä. Sosioekonomiset kuolleisuuserot Suomessa 1971–85. Kesäkuu 1990. 145 s.

173. **Jukka Muukkonen,** Sustainable development and natural resource accounting. August 1990. 96 pp.

174. **Iiris Niemi – Hannu Pääkkönen,** Time use changes in Finland in the 1980s. August 1990. 118 pp.

175. **Väinö Kannisto,** Mortality of the elderly in late 19th and early 20th century Finland. August 1990. 50 pp.

176. **Tapani Valkonen – Tuija Martelin – Arja Rimpelä,** Socio-economic mortality differences in Finland 1971–85. December 1990. 108 pp.

177. **Jaana Lähteenmaa – Lasse Siurala,** Nuoret ja muutos. Tammikuu 1991. 211 s.

178. **Tuomo Martikainen – Risto Yrjönen,** Vaalit, puolueet ja yhteiskunnan muutos. Maaliskuu 1991. 120 s.

179. **Seppo Laaksonen,** Comparative Adjustments for Missingness in Short-term Panels. April 1991. 74 pp.

180. **Ágnes Babarczy – István Harcsa – Hannu Pääkkönen,** Time use trends in Finland and in Hungary. April 1991. 72 pp.

181. **Timo Matala,** Asumisen tuki 1988. Kesäkuu 1991. 64 s.

182. **Iiris Niemi – Parsla Eglite – Algimantas Mitrikas – V.D. Patrushev – Hannu Pääkkönen,** Time Use in Finland, Latvia, Lithuania and Russia. July 1991. 80 pp.

183. **Iiris Niemi – Hannu Pääkkönen,** Vuotuinen ajankäyttö. Joulukuu 1992. 83 s.

183b. **Iiris Niemi – Hannu Pääkkönen – Veli Rajaniemi – Seppo Laaksonen – Jarmo Lauri,** Vuotuinen ajankäyttö. Ajankäyttötutkimuksen 1987–88 taulukot. Elokuu 1991. 116 s.

184. **Ari Leppälahti – Mikael Åkerblom,** Industrial Innovation in Finland. August 1991. 82 pp.

185. **Maarit Säynevirta,** Indeksiteoria ja ansiotasoindeksi. Lokakuu 1991. 95 s.

186. **Ari Tyrkkö,** Ahtaasti asuvat. Syyskuu 1991. 134 s.

187. **Tuomo Martikainen – Risto Yrjönen,** Voting, parties and social change in Finland. October 1991. 108 pp.

188. **Timo Kolu,** Työelämän laatu 1977–1990. Työn ja hyvinvoinnin koettuja muutoksia. Tammikuu 1992. 194 s.

189. **Anna-Maija Lehto,** Työelämän laatu ja tasa-arvo. Tammikuu 1992. 196 s.

190. **Tuovi Allén – Päivi Keinänen – Seppo Laaksonen – Seija Ilmakunnas,** Wage from Work and Gender. A Study on Wage Differentials in Finland in 1985. 88 pp.

191. **Kirsti Ahlqvist,** Kodinomistajaksi velalla. Maaliskuu 1992. 98 s.

192. **Matti Simpanen – Irja Blomqvist,** Aikuiskoulutukseen osallistuminen. Aikuiskoulutustutkimus 1990. Toukokuu 1992. 135 s.

193. **Leena M. Kirjavainen – Bistra Anachkova – Seppo Laaksonen – Iiris Niemi – Hannu Pääkkönen – Zahari Staikov,** Housework Time in Bulgaria and Finland. June 1992. 131 pp.

194. **Pekka Haapala – Seppo Kouvonen,** Kuntasektorin työvoimakustannukset. Kesäkuu 1992. 70 s.

195. **Pirkko Aulin-Ahmavaara,** The Productivity of a Nation. November 1992. 72 pp.

196. **Tuula Melkas,** Valtion ja markkinoiden tuolla puolen. Joulukuu 1992. 150 s.

197. **Fjalar Finnäs,** Formation of unions and families in Finnish cohorts born 1938–67. April 1993. 58 pp.

198. **Antti Siikanen – Ari Tyrkkö,** Koti – Talous – Asuntomarkkinat. Kesäkuu 1993. 167 s.

199. **Timo Matala,** Asumisen tuki ja aravavuokralaiset. Kesäkuu 1993. 84 s.

200. **Arja Kinnunen,** Kuluttajahintaindeksi 1990=100. Menetelmät ja käytäntö. Elokuu 1993. 89 s.

201. **Matti Simpanen,** Aikuiskoulutus ja työelämä. Aikuiskoulutustutkimus 1990. Syyskuu 1993. 150 s.

202. **Martti Puohiniemi,** Suomalaisten arvot ja tulevaisuus. Lokakuu 1993. 100 s.

203. **Juha Kivinen – Ari Mäkinen,** Suomen elintarvike- ja metallituoteteollisuuden rakenteen, kannattavuuden ja suhdannevaihteluiden yhteys; ekonometrinen analyysi vuosilta 1974 – 1990. Marraskuu 1993. 92 s.

204. **Juha Nurmela,** Kotitalouksien energian kokonaiskulutus 1990. Marraskuu 1993. 108 s.

205a. **Georg Luther,** Suomen tilastotoimen historia vuoteen 1970. Joulukuu 1993. 382 s.

205b. **Georg Luther,** Statistikens historia i Finland till 1970. December 1993. 380 s.

206. **Riitta Harala – Eva Hänninen-Salmelin – Kaisa Kauppinen-Toropainen – Päivi Keinänen – Tuulikki Petäjäniemi – Sinikka Vanhala,** Naiset huipulla. Huhtikuu 1994. 64 s.

207. **Wangqiu Song,** Hedoninen regressioanalyysi kuluttajahintaindeksissä. Huhtikuu 1994. 100 s.

208. **Anne Koponen,** Työolot ja ammattillinen aikuiskoulutus 1990. Toukokuu 1994. 118 s.

209. **Fjalar Finnäs,** Language Shifts and Migration. May 1994. 37 pp.

210. **Erkki Pahkinen – Veijo Ritola,** Suhdannekäänne ja taloudelliset aikasarjat. Kesäkuu 1994. 200 s.

211. **Riitta Harala – Eva Hänninen-Salmelin – Kaisa Kauppinen-Toropainen – Päivi Keinänen – Tuulikki Petäjäniemi – Sinikka Vanhala,** Women at the Top. July 1994. 66 pp.

212. **Olavi Lehtoranta,** Teollisuuden tuottavuuskehityksen mittaaminen toimialatasolla. Tammikuu 1995. 73 s.

213. **Kristiina Manderbacka,** Terveydentilan mittarit. Syyskuu 1995. 121 s.

214. **Andres Vikat,** Perheellistyminen Virossa ja Suomessa. Joulukuu 1995. 52 s.

215. **Mika Maliranta,** Suomen tehdasteollisuuden tuottavuus. Helmikuu 1996. 189 s.

216. **Juha Nurmela,** Kotitaloudet ja energia vuonna 2015. Huhtikuu 1996. 285 s.

217. **Rauno Sairinen,** Suomalaiset ja ympäristöpolitiikka. Elokuu 1996. 179 s.

218. **Johanna Moisander,** Attitudes and Ecologically Responsible Consumption. August 1996. 159 pp.

219. **Seppo Laaksonen** (ed.), International Perspectives on Nonresponse. Proceedings of the Sixth International Workshop on Household Survey Nonresponse. December 1996. 240 pp.

220. **Jukka Hoffrén,** Metsien ekologisen laadun mittaaminen. Elokuu 1996. 79 s.

221. **Jarmo Rusanen – Arvo Naukkarinen** – Alfred Colpaert – Toivo Muilu, Differences in the Spatial Structure of the Population Between Finland and Sweden in 1995 – a GIS viewpoint. March 1997. 46 pp.

222. **Anna-Maija Lehto,** Työolot tutkimuskohteena. Marraskuu 1996. 289 s.

223. **Seppo Laaksonen** (ed.), The Evolution of Firms and Industries. June 1997. 505 pp.

224. **Jukka Hoffrén,** Finnish Forest Resource Accounting and Ecological Sustainability. June 1997. 132 pp.

225. **Eero Tanskanen**, Suomalaiset ja ympäristö kansainvälisestä näkökulmasta. Elokuu 1997. 153 s.

226. **Jukka Hoffrén**, Talous hyvinvoinnin ja ympäristöhaittojen tuottajana –Suomen ekotehokkuuden mittaaminen. Toukokuu 1999. 154 s.

227. **Sirpa Kolehmainen,** Naisten ja miesten työt. Työmarkkinoiden segregoituminen Suomessa 1970–1990. Lokakuu 1999. 321 s.

228. **Seppo Paananen**, Suomalaisuuden armoilla. Ulkomaalaisten työnhakijoiden luokittelu. Lokakuu 1999. 152 s.

229. **Jukka Hoffrén,** Measuring the Eco-efficiency of the Finnish Economy. October 1999. 80 pp.

230. **Anna-Maija Lehto – Noora Järnefelt** (toim.), Jaksaen ja joustaen. Artikkeleita työolotutkimuksesta. Joulukuu 2000. 264 s.

231. **Kari Djerf**, Properties of some estimators under unit nonresponse. January 2001. 76 s.

The Research Report series describe the Finnish society in the light of up-to-date research results. Scientific studies that are carried out at Statistics Finland or that are based on the data sets of Statistics Finland are published in the series. The series also includes collections of scientific articles such as conference proceedings.

This study reviews some reweighting methodology widely used in statistical agencies to adjust for unit nonresponse in sample surveys. As an academic thesis it covers many theoretical sections, but the ultimate goal was to find out which methods perform best in simulation where a real data set was used. The results confirmed some earlier empirical findings on the usefulness of modern calibration estimators. Thus, the decision to replace traditional types of estimators by calibration estimators in many of Statistical Finland´s surveys was justified from the nonresponse adjustment point of view as well.

9 789517 278577

Tilastokeskus, myyntipalvelu
PL 4V
00022 TILASTOKESKUS
puh. (09) 1734 2011
faksi (09) 1734 2500
myynti.tilastokeskus@tilastokeskus.fi
www.tilastokeskus.fi

Statistikcentralen, försäljningstjänsten
PB 4V
00022 STATISTIKCENTRALEN
tfn (09) 1734 2011
fax (09) 1734 2500
myynti.tilastokeskus@stat.fi
www.stat.fi

Statistics Finland, Sales Services
P.O.Box 4V
FIN-00022 STATISTICS FINLAND
Tel. +358 9 1734 2011
Fax +358 9 1734 2500
myynti.tilastokeskus@stat.fi
www.stat.fi

ISSN 0355–2071
=Tutkimuksia
ISBN 951–727–857-
Tuotenumero 9453
AGO