RISQ

Representativity Indicators for Survey Quality

Indicators for Representative Response Based on Population Totals

Work package 3
Deliverable 2.2

Natalie Shlomo & Chris Skinner
University of Southampton, United Kingdom

Barry Schouten, Vincent de Heij, Jelke Bethlehem & Pim Ouwehand
Centraal Bureau voor de Statistiek, Netherlands

August 28, 2009

SEVENTH FRAMEWORK PROGRAMME

The RISQ Project is financed by the 7th Framework Programme (FP7) of the European Union.
Cooperation Programme, Socio-economic Sciences and the Humanities, Provision for Underlying Statistics

Indicators for representative response based on population totals

1. Introduction

The RISQ (Representativity Indicators for Survey Quality) project, funded by the European 7th Framework Programme, is a joint effort of the NSI's of Norway, The Netherlands and Slovenia, and the Universities of Leuven and Southampton to develop quality indicators for survey response. These indicators may serve three goals (Bethlehem and Schouten 2008): to enable the comparison of different surveys or of a single survey in time, to assist in monitoring survey data collection, and most ambitiously to guide survey data collection. Indicators may also be of help in monitoring and evaluating the completion of a register. The application to registers is especially useful when registers have a time lag and fill gradually in time.

RISQ deliverable 2.1, Shlomo et al (2009), is devoted to the construction of indicators for representativity indicators, or R-indicators, and the evaluation of their statistical properties. They assume that the survey sample is available and can be linked to a set of external data sources like registers and administrative data. The variables from these data sources are known for both respondents and non-respondents to the survey, and function as auxiliary variables for the prediction of response behaviour. Shlomo et al (2009) propose two indicators; an indicator based on the variation in estimated response propensities and an indicator based on the variation in adjustment weights. The second indicator is based on work by Särndal and Lundström (2008). Both indicators are biased and have standard errors. In the paper analytic approximations to the bias and standard errors are given.

RISQ deliverable 3, Schouten et al (2009), applies the R-indicator based on the variation of response propensities, to a number of simulated and real data sets. The sample sizes and sets of auxiliary variables are varied in order to assess the dependence on the survey size and response model. Schouten et al (2009) conclude that it is imperative that R-indicator values are accompanied by the set of auxiliary variables and by confidence intervals. Confidence intervals can be rather wide for small samples but the size of these intervals is only mildly affected by the number of variables selected in the response models.

During the RISQ project it was recognized that the presumption of linked survey samples is in many settings not a valid one. While national statistical institutes often do have access to a number of government registrations, university and market researchers usually do not. For indicators to become useful for these researchers, they must be based on different forms of auxiliary information. The only form of auxiliary information that is generally accessible are the sets of statistics produced by the national statistical institutes. These institutes disseminate tables about a wide range of population statistics. This paper is about R-indicators that are based completely on such population statistics and that can be computed without any knowledge about the non-respondents. We do, however, assume that the survey questionnaire contains the questions that correspond to the population characteristics employed in the assessment of the representativity of survey response.

We adapt the two indicators developed in Shlomo et al (2009) to population tables and population counts. We replace sample covariances and sample means by population covariances and population means. We will call the resulting indicators population-based R-indicators. To our knowledge there is no record in the literature about models for response propensities that employ population information only. In this respect the current paper is innovative and may be applicable to other statistical areas as well. As a consequence, however, we feel this paper is just a first start. More research is likely to be necessary in order to refine estimators and estimation strategies.

Population-based R-indicators allow for weaker conclusions about the nature of response than their counterparts, the sample-based R-indicators. The population-based R-indicators will in general have a larger bias as they cannot completely discern sampling variation from response variation. Consequently it must be remarked that population-based indicators will not be very useful for small surveys. Nevertheless, the indicators may be helpful for large surveys.

In this paper, we propose a generic approach to the use of population statistics and will not discuss or investigate the quality of the survey questions and the population information themselves. However, there is an imminent risk of measurement errors when comparing the representativeness of survey questions to population statistics. It must be ascertained that the survey questions that are employed have the same definitions and classifications as the population tables. Furthermore, it is best to avoid questions that are prone to measurement errors, like questions that require a strong cognitive effort or that may lead to socially desirable answers. Finally, it is strongly recommendable to use population statistics that are based on registrations or administrative data. The population-based R-indicators can be used for population statistics that are based on survey, but these statistics may be biased themselves and may not reflect the true population distribution accurately. One would draw erroneous conclusions about the representativeness of the response when the population statistics are biased.

We apply the proposed population-based indicators to various country data sets. The data sets are documented in RISQ deliverable 1 (RISQ 2008). For comparison we also compute the sample-base R-indicators.

With the simulated data sets and the survey data sets we investigate three research questions:

- How to extend sample-based R-indicators to population-based R-indicators?
- What are the statistical properties of population-based R-indicators?
- Are the population-based R-indicators practicable in real survey settings?

In section 2 we briefly refresh the definitions and methodology behind R-indicators. In section 3 we then move to the estimation of R-indicators, i.e. the sample-based estimators and their population-based analogues. In section 4 we describe the data sets from the various countries that participate in RISQ. In section 5, we give an overview of the application of sample-based and population-based indicators to simulated and real data sets. In section 6 we end with a discussion.

2. R-indicators

In this section we briefly repeat the definition and concepts of R-indicators. Details can be found in Shlomo et al (2009).

2.1 General notation

We suppose that a sample survey is undertaken, where a sample s is selected from a finite population U. The sizes of s and U are denoted n and N, respectively. The units in U are labelled $i = 1, 2, \ldots, N$. The sample is assumed to be drawn by a probability sampling design $p(.)$, where the sample s is selected with probability $p(s)$. The first order inclusion probability of unit i is denoted π_i and $d_i = \pi_i^{-1}$ is the design weight. In some cases, we shall assume simple random sampling without replacement.

We suppose that the survey is subject to unit non-response. The set of responding units is denoted . Thus, we have $r \subset s \subset U$. We denote summation over the respondents, sample and population by Σ_r, Σ_s and Σ_U, respectively. We let R_i be the response indicator variable so that $R_i = 1$ if unit i responds and $R_i = 0$, otherwise. Hence, $r = \{i \in s; R_i = 1\}$. We shall suppose that the typical target of inference is a population mean $\bar{Y} = N^{-1} \sum_U y_i$ of a survey variable, taking value y_i for unit i.

We suppose that the data available for estimation purposes consists first of the values $\{y_i; i \in r\}$ of the survey variable, observed only for respondents. Secondly, we suppose that information is available on the values of $x_i = (x_{1,i}, x_{2,i}, \ldots, x_{K,i})^T$, a vector of auxiliary variables. We shall usually suppose each $x_{k,i}$ is a binary indicator variable, where x_i represents one or more categorical variables, since this will be the case in the applications we consider, but our presentation allows for general $x_{k,i}$ values. We assume that values of x_i are observed for all respondents.

We distinguish two settings. One in which x_i is known for all sample units, i.e. for both respondents and non-respondents, and one in which x_i is known only at the aggregate level, i.e. the population total $\sum_U x_i$ and/or the population cross-products $\sum_U x_i x_i^T$. We refer to the two types of information as *sample-based auxiliary information* and *aggregate population-based auxiliary information*. The first setting appears if the variables making up x_i are available on a register. However, in many countries and surveys the availability of auxiliary information on non-respondents may be very limited, e.g. because of the absence of a register. In those cases the second setting holds.

2.2 Response propensities and response influences

We define the response propensity as a conditional expectation of the response indicator variable R_i given the values of specified variables and survey conditions (Little, 1986, 1988): $\rho_X(x_i) = E(R_i \mid x_i)$, where the vector of auxiliary variables is defined as in section 2.1. For simplicity, we shall write $\rho_i = \rho_X(x_i)$ and hence denote the response propensity just by ρ_i. A detailed discussion of the notion of response propensities and their properties is presented in Shlomo, et al. (2009). In this discussion it was argued that it is desirable to select the auxiliary variables constituting x_i in such a way that the missing at random, denoted MAR (Little and Rubin, 2002) holds as closely as possible and that our definition of $\rho_i = \rho_X(x_i)$ relates to a specific choice of auxiliary variables x_i. A different choice would generally result in a different ρ_i. We note also that we define the response propensity conditional on the survey conditions that apply when the data are collected in order to be able to compare the representativeness of different surveys. Following Särndal and Lundström (2008), we refer to ρ_i^{-1} as the *response influence* and denote it ϕ_i.

2.3 R-indicators

Let $\boldsymbol{\rho} = (\rho_1, \rho_2, ..., \rho_N)'$ denote the vector of response propensities in the population. The representativity of the response mechanism may be measured by the variation between the ρ_i and in particular by the standard deviation of the response propensities given by:

$$S(\boldsymbol{\rho}) = \sqrt{\frac{1}{N-1}\sum_U (\rho_i - \bar{\rho}_U)^2}\ , \tag{1}$$

where $\bar{\rho}_U = \sum_U \rho_i / N$. It may be shown that: $S(\boldsymbol{\rho}) \le \sqrt{\bar{\rho}_U(1-\bar{\rho}_U)} \le \frac{1}{2}$. Hence, transforming $S(\boldsymbol{\rho})$ to:

$$R(\boldsymbol{\rho}) = 1 - 2S(\boldsymbol{\rho}) \tag{2}$$

ensures that $0 \le R(\boldsymbol{\rho}) \le 1$ and, as discussed by Schouten et al. (2009), $R(\boldsymbol{\rho})$ defines an *R-indicator* which takes values on the interval [0,1] with the value 1 indicating the most representative response, where the ρ_i display no variation, and the value 0 indicating the least representative response, where the ρ_i display maximum variation.

The rationale behind the R-indicator defined by (2) is the following definition of representative response.

Definition: The response to a survey is representative with respect to X when the response propensity function $\rho_X(x)$ is constant in x.

It can easily be shown that the standard deviation of response propensities corresponds to the Euclidean distance function which is a natural choice in measuring deviations from representative response.

$R(\rho)$ can be related to the maximal absolute non-response bias of the response mean of any target variable Y. Shlomo et al (2009) show that

$$|B(\hat{\bar{y}}_r)| = \frac{\mathrm{cov}(Y, \rho_Y)}{\bar{\rho}_Y} \le \frac{S(Y)S(\rho_Y)}{\bar{\rho}_Y} = \frac{S(Y)(1 - R(\rho_Y))}{2\bar{\rho}_Y}. \tag{3}$$

The upper bound in (3) is unknown, but we may use as a surrogate

$$\frac{S(Y)(1 - R(\rho_X))}{2\bar{\rho}_X} \tag{4}$$

and divide by $S(Y)$ to make it independent of any specific Y.

Särndal and Lundström (2008) define the following measure for the selection of weighting variables in calibration:

$$Q^2(\boldsymbol{\rho}) = [\sum_U \rho_i]^{-1}[\sum_U \rho_i(\phi_{Ui} - \bar{\phi}_{\rho U})^2] \tag{5}$$

where $\bar{\phi}_{\rho U}$ is the ρ_i- weighted mean of the ϕ_{Ui} given by

$$\bar{\phi}_{\rho U} = (\sum_U \rho_i)^{-1}(\sum_U \rho_i \phi_{Ui}) . \tag{6}$$

This measure is, as Särndal and Lundström argue, the proportion of the relative nearbias (see Särndal and Lundström 2005) that is independent of survey target variables. Measure (5) is large whenever a candidate variable leads to strongly deviating calibration weights. Although the original motivation and objective of the measure is different, we regard it as a candidate R-indicator.

This quantity is a weighted variance of the approximate response influences. We may expect its magnitude to be inversely related to the magnitude of $R(\boldsymbol{\rho})$. Thus, in very rough terms, we expect $R(\boldsymbol{\rho})$ to decrease and $Q^2(\boldsymbol{\rho})$ to increase as the variability of the ρ_i increases.

3. The estimation of R-indicators based on population totals

3.1 Sample-based R-indicators

Let us for the moment assume we have estimators $\hat{\rho}_i$ and $\hat{\phi}_i$ of the response probability ρ_i and the response influence ϕ_i, respectively, based on sample-based auxiliary information. $\hat{\rho}_i$ and $\hat{\phi}_i$ may be computed for each $i \in s$.

We first look at the estimation of indicator $R(\rho)$. An estimator of $\bar{\rho}_U$ is given by

$$\hat{\bar{\rho}}_U = (\sum_s d_i \hat{\rho}_i)/N , \tag{7}$$

and we estimate $R(\boldsymbol{\rho})$ by

$$\hat{R}(\boldsymbol{\rho}) = 1 - 2\sqrt{\frac{1}{N-1}\sum_s d_i(\hat{\rho}_i - \hat{\bar{\rho}}_U)^2} \ . \tag{8}$$

In (7) and (8) we could replace N by $\sum_s d_i$. Hence, the estimated R-indicator is based on the design-weighted sample variance of estimated response propensities.

Next we turn to the estimation of $Q^2(\boldsymbol{\rho})$ in (5). Särndal and Lundström (2008) propose the following estimator:

$$q^2 = [\sum_r d_i]^{-1}[\sum_r d_i(\hat{\phi}_i - \bar{\phi}_r)^2], \qquad (9)$$

where $\bar{\phi}_r = (\sum_r d_i\hat{\phi}_i)/(\sum_r d_i)$. In words, we estimate the indicator by the design-weighted response variance of estimated response influences. $\bar{\phi}$ can be re-expressed as

$$\bar{\phi}_r = (\sum_s d_i)/(\sum_r d_i). \qquad (10)$$

In the sample-based setting, the response propensities and influences may be modelled by generalized linear models. Shlomo et al (2009) use a logistic link function for the response propensities and a reciprocal link function for the response influences. Schouten et al (2009) compare logistic and identity link functions for the response propensities. They conclude that the resulting R-indicators depend only very mildly on the type of link function.

In section 4 we will show that it is straightforward to extend the identity link function to the population-based setting. We, therefore, restrict ourselves to the identity link function for the sample-based R-indicators. The identity link function leads to the 'linear probability model'

$$\rho_i = x_i'\beta. \qquad (11)$$

Following Särndal and Lundström (2008) we employ the reciprocal link function

$$\rho_i^{-1} \equiv \phi_i = x_i'\lambda. \qquad (12)$$

They assume that the vector x_i is defined in such a way that there exists a constant vector c such that $c'x_i = 1$ for all $i \in U$. This restriction will in most practical situations be met and is effectively equivalent to assuming that a constant intercept term is included in the auxiliary information. Särndal and Lundström (2008) view (12) as a hypothetical model which will not hold in practice and they instead focus on a finite population approximation to this model. This approximation is obtained by first defining a value λ_U of λ which achieves the best fit of model (12) in the finite population. For mathematical convenience, they define the fit as the weighted sum of squared differences $\sum_U \rho_i(\rho_i^{-1} - x_i'\lambda)^2$ and this is minimised when

$$\lambda_U = (\sum_U \rho_i x_i x_i')^{-1} \sum_U x_i, \qquad (13)$$

provided x_i is defined so that the inverted matrix in (13) is non-singular. This implies that a finite population approximation to ϕ_i is given by:

$$\phi_{Ui} = x_i'(\sum_U \rho_i x_i x_i')^{-1} \sum_U x_i. \qquad (14)$$

The linear probability model in (11) can be estimated in closed form by ordinary least squares or by weighted least squares, where the weights are the design weights. For the reciprocal link function model in (12), Särndal and Lundström (2008) approximate the model by $\rho_i^{-1} \approx x_i'\lambda_U$, where λ_U is defined in (13) and estimate this approximate model by estimating λ_U from the sample data by:

$$\hat{\lambda}_U = (\sum_r d_i x_i x_i')^{-1} \sum_s d_i x_i \ . \tag{15}$$

In the case of the linear probability model in (11), if β is estimated by (design-) weighted least squares, the implied estimator of ρ_i is given by

$$\hat{\rho}_i^{OLS} = x_i'(\sum_s d_i x_i x_i')^{-1} \sum_s d_i x_i R_i \ , \tag{16}$$

which may also be expressed as

$$\hat{\rho}_i^{OLS} = x_i'(\sum_s d_i x_i x_i')^{-1} \sum_r d_i x_i \ . \tag{17}$$

In the approach of Särndal and Lundström (2008), with the reciprocal link function, ϕ_i is estimated by

$$\hat{\phi}_i = x_i'\hat{\lambda}_U \ , \tag{18}$$

where $\hat{\lambda}_U$ is defined in (12), so that:

$$\hat{\phi}_i = x_i'(\sum_r d_i x_i x_i')^{-1} \sum_s d_i x_i \tag{19}$$

and the resulting estimator of ρ_i is $\hat{\phi}_i^{-1}$.

Shlomo, et. al. (2009) describe bias adjustments for $\hat{R}(\boldsymbol{\rho})$ and q^2 as well as confidence intervals for the two *R-indicators*. We do not repeat them here.

3.2 Population-based R-indicators

In the population-based setting we have x_i for respondents and there is population aggregated auxiliary information. We distinguish two types:

1. *Full aggregate population-based auxiliary information (type 1);* the population cross-products are available, i.e. $\sum_U x_i x_i^T$ or $\sum_U (x_i - \bar{x}_U)(x_i^T - \bar{x}_U^T)$.
2. *Marginal aggregate population-based auxiliary information (type 2);* the population marginal counts are available, i.e. $\sum_U x_i$.

The first type implies that we have available of all two by two tables, e.g. age times gender, age times marital status and age times marital status. The second type is much more restrictive as we have only the frequency counts, e.g. age, gender, marital status, without any knowledge about the interactions. For the estimation of the R-indicators the two types do not make a difference. However, for the estimation of response propensities the two types lead to different estimators with very different properties.

In the population-based setting $\hat{\rho}_i$ is available only for respondents ($i \in$). A natural candidate estimator of $R(\boldsymbol{\rho})$ that makes use of the estimated propensities is

$$\hat{R}_r(\boldsymbol{\rho}) = 1 - 2\sqrt{\frac{1}{N-1}\sum_r d_i \hat{\rho}_i^{-1}(\hat{\rho}_i - \hat{\bar{\rho}}_r)^2} \tag{20}$$

where $\hat{\bar{\rho}}_r = (\sum_r d_i)/N$. The propensity-weighting by $\hat{\rho}_i^{-1}$ adjusts for the non-response bias. As for standard non-response weighting, the validity of this correction depends on the validity of the estimates $\hat{\rho}_i$.

We like to remark that any adjustment technique for non-response can be applied to construct estimators for $R(\boldsymbol{\rho})$, e.g. calibration estimators like linear or multiplicative weighting (Särndal and Lundström 2005) or weighting class adjustments (Little 1986). It is generally known that propensity weighting may lead to large standard errors. It may, therefore, be more efficient to use parsimonious models to estimate the R-indicator. This can for instance be done by stratifying on response propensity classes. We did, however, not explore such estimators, and restricted ourselves to the propensity-weighted estimator (20). This is a topic for future research.

For the estimation of response propensities, types 1 and 2 are very different. In the case of population-based auxiliary information, we first note that $\sum_s d_i x_i$ and $\sum_s d_i x_i x_i'$ are unbiased for $\sum_U x_i$ and $\sum_U x_i x_i'$, respectively and that in large samples we may expect that $\sum_s d_i x_i \approx \sum_U x_i$ and $\sum_s d_i x_i x_i' \approx \sum_U x_i x_i'$. It follows from (17) that in type 1, we may approximate $\hat{\rho}_i^{OLS}$ by

$$\tilde{\rho}_i^{OLS} = x_i'(\sum_U x_i x_i')^{-1}\sum_r d_i x_i \,. \tag{21}$$

The estimator in (21) requires knowledge of the population sums of squares and cross-products $\sum_U x_i x_i'$ of the elements of x_i.

In type 2, the cross-products are unknown. We can estimate $\sum_s d_i x_i x_i'$ in (17) by rewriting

$$\sum_s d_i x_i x_i' = \sum_s d_i (x_i - \bar{x}_s)(x_i - \bar{x}_s)' + N\bar{x}_s\bar{x}_s', \tag{22}$$

where $\bar{x}_s = \sum_s x_i / n$. $\bar{x}_s$ can be replaced by $\bar{X}_U$. The covariance matrix

$$S_{xx} = N^{-1}\sum_s d_i (x_i - \bar{x}_s)(x_i - \bar{x}_s)' \tag{23}$$

may be replaced by the observed covariance matrix

$$S_{xxr} = (\sum_s d_i R_i)^{-1}\sum_s d_i R_i (x_i - \bar{x}_r)(x_i - \bar{x}_r)', \tag{24}$$

where $\bar{x}_r = (\sum_s R_i x_i)/(\sum_s R_i)$. The estimator in (24) only requires knowledge of the population total of each of the elements of x_i. Combining (21), (22) and (24) we get the following estimator for type 2

$$\tilde{\rho}_i^{OLS} = x_i^{'}(NS_{xxr} + N\bar{x}_s\bar{x}_s^{'})^{-1}\sum_r d_i x_i \ . \quad (25)$$

Remarkably for the $Q^2(\boldsymbol{\rho})$ indicator it does not make a difference if type 1 or 2 holds. The estimator in (9) is based only upon respondent data, and the $\hat{\phi}_i$ in (19) depend only on $\sum_s d_i x_i$. Hence, it is sufficient to have the population totals. We replace $\hat{\phi}_i$ by

$$\tilde{\phi}_i = x_i'(\sum_r d_i x_i x_i')^{-1}\sum_U x_i \ . \quad (26)$$

3.3 Bias and standard error of population-based R-indicators

Shlomo et al (2009) derive analytic approximations for the bias and standard errors of the sample-based estimators (8) and (9). The bias in these estimators arises mostly from plugging in estimated response propensities and response influences in the sample variances. A much smaller and usually negligible contribution to the bias, originates from using sample means rather than population means. Even if the response would be representative, i.e. has equal response propensities, some variation in estimated response propensities is found. The bias is inversely proportional to the sample size. The larger the sample, the smaller the bias. Schouten et al (2009) investigated the bias for different sample sizes. From their analyses it follows that the bias is relatively small for usual sample sizes with respect to the standard error of the R-indicators. Also, the bias adjustment is successful in removing the bias.

The statistical properties of population-based R-indicators are, however, quite different from their sample-based counterparts. As these estimators use less information, we first expect larger standard errors. We will show in section 5 that the increase in standard errors is modest in many cases. However, the standard errors are larger which leads to weaker conclusions about the representativeness of the response.

Second the bias of the population-based estimators is relatively much larger. The estimators have additional bias that stems from the estimation of the sample means and covariances and from the restriction to (propensity-weighted) response means in case of $\hat{R}_r(\hat{\rho})$. As no knowledge is available about the non-respondents, the estimators cannot discern sampling variation from response variation. For example, if the population consists of 50% males and 50% females and the response of 45% males and 55% females, part (or all) of the difference in gender distribution may be the result of sampling variation. Hence, even if the response is representative, the estimated response propensities will show some variation that is attributable to sampling. Clearly, this bias will diminish when the sample size is increased as the sample distribution will grow relatively closer to the population distribution. The response variance of the response propensities is another potential cause of bias because of the selectiveness of the response. For this reason the response means are propensity-weighted. However, the propensity-weighting may not be effective in removing the bias. Note that when response is representative, the propensity-weighted response means of estimated propensities will have more bias than the sample means of estimated propensities because of the non-linear nature of the weights that tends to attenuate the quadratic differences between propensities and their mean.

In the previous subsection, we distinguished two types of population-based R-indicators $\hat{R}_r(\hat{\rho})$; one based on population cross-products and one based on population counts. Clearly, the second type of estimators may provide very poor estimators for the response propensities when there are strong interactions between the components of the auxiliary vector x_i.

In appendix A we give an analytic approximation of the bias of the type 1 $\hat{R}_r(\hat{\rho})$ estimator.

In this paper we will, however, restrain from bias adjustments for the type 2 estimators. The results in section 4 show that most of the bias for these types of estimators comes from the approximation with population counts. Type 2 estimators may be considerably more biased than type 1 estimators. Bias-adjustment of the type 2 estimators is not possible without any knowledge of the population (or sample) interactions.

Analytic standard errors for $\hat{R}_r(\hat{\rho})$ are provided in Appendix B. However, in section 4, we will construct confidence intervals based on resampling methods. More specifically we employ bootstrap methods (Efron and Tibshirani 1993, Wolter 2007).

4. Data

We use three types of data: 1) survey data sets from each of the participating RISQ countries, 2) simulated survey data sets based on the 1995 Israel Census Sample, and 3) samples drawn from the 2007 Dutch VAT records. We will describe each of the data sets before we move to the calculation of the sample-based and population-based indicators.

4.1 Country data sets

Extensive descriptions and documentation of the RISQ data sets can be found in RISQ (2008) and is available at www.R-indicator.eu. We apply indicators to the following selection of data sets:

- *Health survey 2005 (CBS-HS)*: The Consumer Sentiments Survey is a continuous survey of households with questions about general economic development, and the financial situation of the household. The survey is meant to provide insight into short term economic development, and early indicators of differences in consumer trends. The number of cases in the file is 17,908. The response rate was 66.9%.
- *Norwegian Level of Living Survey 2004 (SSB-LLS)*: The survey of living conditions has two main purposes. One is to throw light on the main aspects of the living conditions in general and for various groups of people. Another purpose is to monitor development in living conditions, both level and distribution. Over a three-year period the cross-sectional survey of living conditions will cover all main areas of the living conditions. The survey topics change during a three-year cycle. Housing conditions, participation in organisations, leisure activities, offences and fear of crime were topics in 2004. It is a survey of individuals. The number of cases in the file is 4,837. The response rate was 69.1%.
- *Norwegian European Social Survey Norway 2006 (SSB-ESS)*: ESS is a biennial multi-country survey of individuals covering over 30 nations. It is an academically-driven social survey designed to chart and explain the interaction between Europe's changing

institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. The data set only contains the survey data of Norway. The number of cases in the file is 2,673. The response rate was 65.5%.

- *Belgian European Social Survey 2006 (KUL-ESS)*: As described for the Norwegian dataset, the ESS is an EU harmonized social survey. The data set contains the survey data of Belgium. The number of cases in the file is 2,927. The response rate was 61.4%.
- *Slovenian Labour Force Survey 2007 (SURS-LFS)*: The Slovenian Labour Force Survey is an EU harmonized rotating panel survey conducted continuously through the year. The data contains employment related characteristics and demographic characteristics of all individuals 15 years or older living in selected households. The number of households varies between 7,010 and 7,160 households which is around 16,900 responding individuals. The response rate is around 80%.

To all country data sets a minimal set of auxiliary variables was linked comprising of age, gender and region/degree of urbanisation. The R-indicators are computed for these minimal sets. Each country additionally also linked auxiliary variables that are available in their sampling frame or registrations.

4.2 Simulations based on the 1995 Israel Census Sample

The 1995 20% Israel Census Sample contains 753,711 individuals aged 15 and over in 322,411 households. The sample design is similar to a standard household survey carried out at National Statistical Institutes. The sample units are households and all persons over the age of 15 in the sampled households are interviewed. Typically a proxy questionnaire is used and therefore there is no individual non-response within the household. In this study, we assume that every household has an equal probability to be included in the sample.

We carry out a two-step design to define response probabilities in the Census data. In the first step, we determine probabilities of response based on explanatory variables that typically lead to differential non- response based on our experiences of working with survey data collection. A response indicator was then generated for each unit in the Census from these probabilities. In the second step, we fit a logistic regression model to these Census data and thus determined a 'true' response propensity for each unit as predicted by this model fitted to the population. The dependent variable of the model is the response indicator and the independent variables of the model the explanatory variables used in the first step. This two-step design ensures that we have a known model generating the response propensities and therefore we can assess model misspecification besides the sampling properties of the indicators.

The explanatory variables used to generate the response probabilities are Type of locality (3 categories), number of persons in household (1,2,3,4,5,6+), children in the household indicator (yes, no). Samples of size n are drawn from the Census population of size N at different sampling fractions 1:50, 1:100, and 1:200. For each sample drawn, a sample response indicator is generated from the 'true' population response probability. The overall response rate is 82%. Response propensities and R-indicators are subsequently estimated from the sample.

4.3 Simulations based on the 2007 Dutch VAT records

Statistics Netherlands is considering to replace the Short Term Statistics (STS or KS in Dutch) surveys by the VAT records collected by the Dutch Tax Board as the main data source for statistics about business turnover. The VAT records are submitted to the Tax Offices by all business on an annual, quarterly or monthly basis and contain reported turnover and VAT. Reporting is obligatory by the Dutch law. The frequency of reporting is determined by the size of the business in the previous calendar year. Large businesses need to report every month, while very small businesses have to submit annual reports only. Mid-size businesses report every quarter.

STS provides monthly and quarterly statistics. Because of the high frequency of the STS statistics, responses need to be available ultimately 25 days after the end of the reference month. Statistics are produced approximately 30 days after the end of the reference month. As a consequence there is only a very small time span for the processing of the responses.

The VAT reports are collected for very different purposes and do not share the 30 day deadlines of the STS surveys. This holds especially true for the quarterly and annual reporting businesses. Nonetheless, part of the reports is available before the 30 day deadline and may form the input to the STS. The availability of reports depends strongly on the month. For the final months of each quarter part of the quarterly reports is available, and for December additionally part of the annual reports becomes accessible. Hence, we expect strong differences between the representativeness of submitted reports for each month of the year.

In the analysis of section 5 we will look at the reports for all businesses in retail for the months January, June and December and derive their responses for 25 days and 30 days after the end of the month (January 31st, June 30th and December 31st). The deadline of 25 days provides the ideal time point, while the deadline of 30 days corresponds to the ultimate time point.

Quarterly VAT reports contain the quarterly turnover and annual VAT reports have the total annual turnover. To be useful in the production of statistics, the quarterly and annual turnover need to be divide over, respectively, three and 12 months. In this paper we take a simple approach and make equal distributions over the months. As such we ignore seasonal influences in retail.

As auxiliary vector x_i we have the type of business according to the Tax Board taxonomy of businesses, the VAT, if reported, of the previous month, the VAT, if reported of the same month in the previous year, and the total wages reported by the business in the same month. Businesses also need to report their monthly labour costs and wages to the Tax Board. These reports are available much faster than the turnover and may provide good predictors. Rather than using the actual turnover or wages, we recode all variables into categorical variables.

A complicating factor is that businesses do not have to report wages and turnover under the same Tax Number and at the same aggregated level. As a consequence the total sum of wages may not be attributable at the desired level. For these cases we classify the total wages as missing/unknown.

The list of all businesses that have to report in a specific month form the population. From month to month there are small differences in the population frame, because of businesses that are starting or finishing activities. When a business was non-existent in the previous month or the same month of the previous year, we classify the corresponding turnover as non-active. The population size is approximately equal to 125,000.

From the January, June and December Tax Board population frames we took 1:2 (n = 60,000), 1:8 (n = 15,000) and 1:18 (n = 6,700) simple random samples without replacement.

5. Results

In this section we describe the results for each type of data sets separately.

5.1 Simulations based on the 1995 Israel Census Sample

Throughout the simulation we examine the sampling properties of the R-indicators as well as the impact of model misspecification on their properties. Because smaller sample sizes generally lead to the selection of a less complex model, we shall consider that misspecification is represented by a simpler model.

In table 5.1, we examine samples drawn at different sampling rates, estimate response propensities for each sample and calculate the measure $\hat{R}(\boldsymbol{\rho})$ for both sample and population based auxiliary variables. We present results for both the true model, number of persons + locality type + child indicator, and a simple model with number of persons only.

Table 5.1: Simulation means of $\hat{R}(\boldsymbol{\rho})$ for sample and population-based auxiliary variables for 500 samples, ('true' R-indicator = 0.8780).

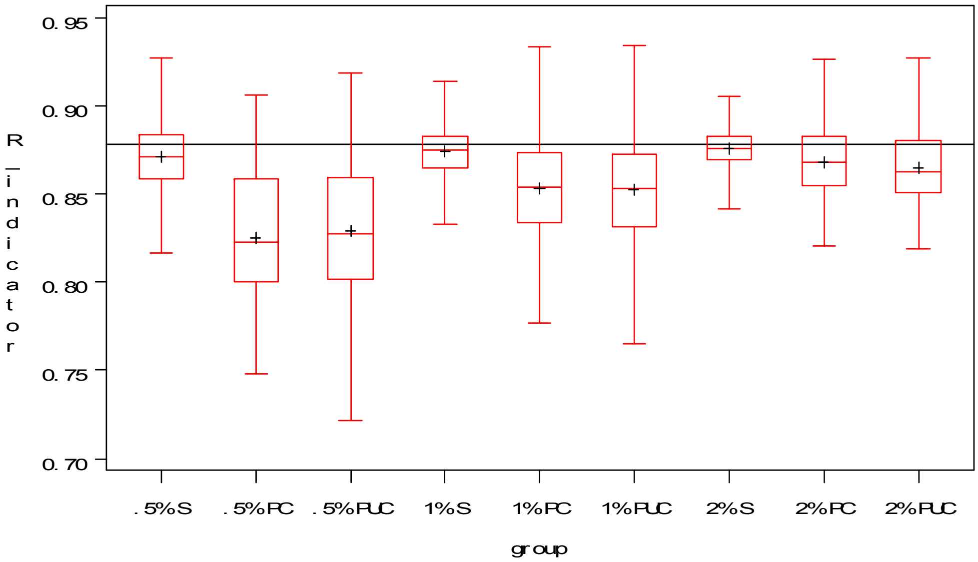
	$\hat{R}(\boldsymbol{\rho})$					
	Model = number of persons + locality type, + child indicator			*Model = number of persons*		
sampling fraction		*population based*			*population based*	
	Sample based	*Type 1*	*Type 2*	*Sample based*	*Type 1*	*Type 2*
1:200 (n=1,612)	0.8714	0.8251	0.8290	0.8791	0.8477	0.8456
1:100 (n=3,224)	0.8741	0.8531	0.8524	0.8801	0.8684	0.8652
1:50 (n=6,448)	0.8761	0.8683	0.8652	0.8810	0.8729	0.8680

Table 5.1 shows that the sample based estimator $\hat{R}(\boldsymbol{\rho})$ increases as the sample size increases. If the specified model is correct, there is some downward bias and this tends to increase as the sample size increases. This is as expected. Sampling error tends to lead to overestimation of the variability of the estimated response propensities and this leads to underestimation of the R-indicator. The degree of underestimation is, however, small. Under the less complex model, estimation of response propensities results in a 'smoothing' of the propensities and hence an overestimation in $\hat{R}(\boldsymbol{\rho})$. Results of $\hat{R}(\boldsymbol{\rho})$ with the bias correction (not shown here) have produced a more stabilized indicator across different sample sizes under the correct

model, but for the less complex model with overestimation of $\hat{R}(\boldsymbol{\rho})$, the bias correction can exacerbate the overestimation. Comparing $\hat{R}(\boldsymbol{\rho})$ when using sample based auxiliary variables and population based auxiliary variables, Table 5.1 shows that the R-indicator is underestimated when using population based auxiliary variables. The variation of response propensities is larger than the variation under sample based auxiliary variables. Since we used simple random samples and assumptions of *MAR*, there seems to be little difference between the two population level indicators of $\hat{R}(\boldsymbol{\rho})$ based on a known population covariance matrix and an estimated covariance matrix from the response set.

Figure 5.1: Boxplots for 500 estimated R-indicators for 0,5%, 1% and 2% samples for a) the (true) model = number of persons + locality type, + child indicator, and for b) model = number of persons. (S) is the sample-based R-indicator, (PC) the type 1 population-based R-indicator, and (PUC) the type 2 population-based R-indicator.

a)



b)

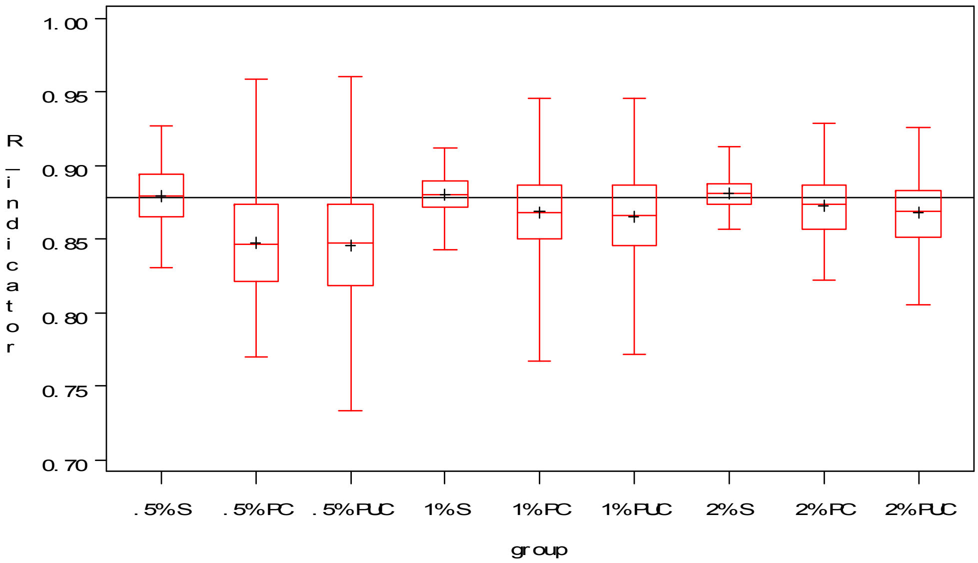


Figure 5.1 shows the boxplots for the sample-based and population-based R-indicators based on 500 samples of 0,5%, 1% and 2%. The boxplots are given for the true model and the reduced model. The boxplots clearly show the increase in bias and standard errors when the sample proportion is smaller.

In table 5.2, we examine the properties of $q^2(\hat{\rho})$ based on the variance of the estimated response influences. For this indicator, we expect low values to reflect good quality and small non-response bias. We compare the full set of explanatory variables in the model used in this simulation to a less complex model as before.

Table 5.2: Simulation means of $q^2(\hat{\rho})$ for sample and population based auxiliary variables for 500 samples, ('true' indicator is 0.0087).

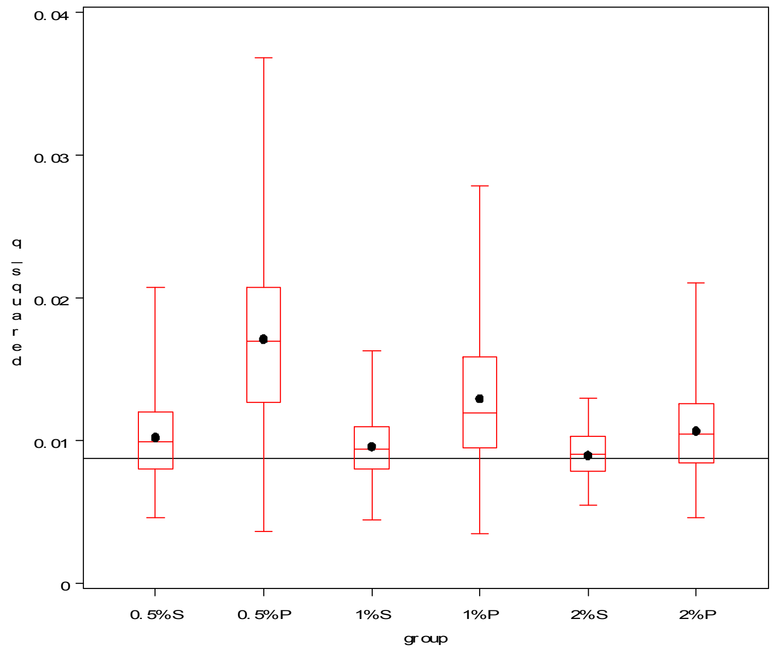
	$q^2(\hat{\rho})$			
	Model = number of persons + locality type, + child indicator		*Model = number of persons*	
Sampling fraction	*Sample-based*	*Population-based*	*Sample-based*	*Population-based*
1:200 (n=1,612)	0.0102	0.0171	0.0092	0.0132
1:100 (n=3,224)	0.0096	0.0129	0.0088	0.0109
1:50 (n=6,448)	0.0089	0.0107	0.0083	0.0095

Results from table 5.2 show the decrease in $q^2(\hat{\rho})$ as the sample sizes increase. Shlomo, et al. (2009) discuss a bias correction for this indicator as well. The less complex model produces 'smoother' estimated influences and hence less variation. This is expected since the indicator was developed to assess the effectiveness of auxiliary variables on bias reduction. Using population based auxiliary variables have increased the variation of the estimated influences.

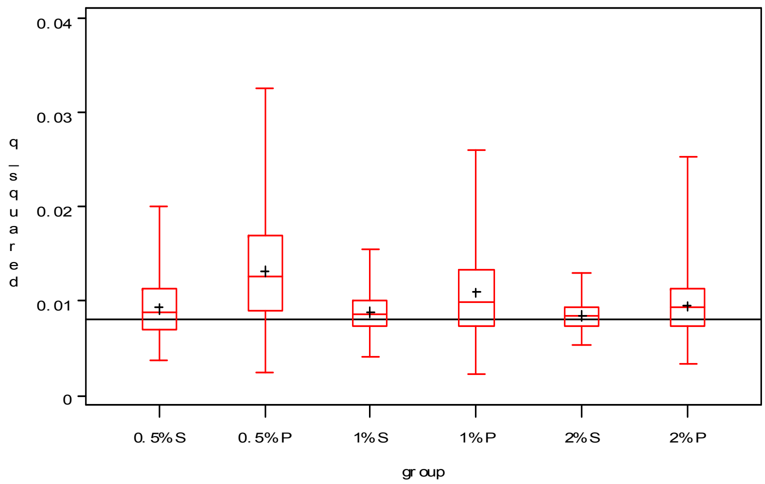
Figure 5.2 again gives boxplots based on 500 samples of different sample proportion, 0,5%, 1% and 2%. The boxplots again emphasize the dependence of bias and standard errors on sample size.

Figure 5.2: Boxplots for 500 estimated $q^2(\hat{\rho})$ for 0,5%, 1% and 2% samples for a) the(true) model = number of persons + locality type, + child indicator, and for b) model = number of persons. (S) is the sample-based $q^2(\hat{\rho})$, (P) population-based $q^2(\hat{\rho})$.

a)



b)



5.2 Simulations based on the 2007 VAT records

We compute the R-indicator estimates for VAT records that were submitted within 25 and 30 days after the end of the reference month. The true values for $R(\rho)$ and $Q^2(\rho)$ are given in table 5.3 with respect to the model VAT month t-12 + wages month t. Also given are the completion rates for the three months after 25 and 30 days. Note that the completion rate for

January is extremely low, only 20% of the businesses has submitted a tax report after 25 days. For June and December these rates are much higher. After 30 days more than 85% of the businesses has reported for December.

There are remarkable differences between the true $Q^2(\rho)$ values for each of the three months. There seems to be a strong dependence on the completion rate. This effect has not yet been investigated thoroughly in other papers and needs more elaborate analysis.

Table 5.3: The true values for $R(\rho)$ and $Q^2(\rho)$ for January, June and December after 25 and 30 days of data collection.

Indicator	*January*		*June*		*December*	
	25 days	*30 days*	*25 days*	*30 days*	*25 days*	*30 days*
$R(\rho)$	82,1%	77,2%	85,4%	83,3%	91,5%	86,1%
$Q^2(\rho)$	13,77	6,86	0,11	0,05	0,12	0,03
Completion	19,7%	26,2%	64,1%	81,5%	48,1%	86,1%

Table 5.4: The unadjusted and adjusted sample-based $\hat{R}(\hat{\rho})$ and the unadjusted $q^2(\hat{\rho})$ for January, June and December 2007 after 25 days and 30 days for the 1:2 subsample. 95% confidence intervals are given.

Sample	$\hat{R}(\hat{\rho})$ *sample-based*				$q^2(\hat{\rho})$ *sample-based*	
	25 days		*30 days*			
	raw	*adj*	*raw*	*adj*	*25 days*	*30 days*
January	82,1%	82,1%	77,2%	77,2%	13,57	6,86
June	85,8%	85,8%	83,3%	83,3%	0,12	0,05
December	91,3%	91,3%	85,9%	85,9%	0,12	0,03

Table 5.4 contains the unadjusted and adjusted sample-based R-indicators. Tables 5.5 to 5.7 give the various R-indicators for, respectively, January, June and December. For each month we give the indicators after 25 and after 30 days of data collection and for the 1:2, 1:9 and 1:18 simple random samples.

Table 5.4 reconfirms the conclusions of Shlomo et al (2009) that the sample-based R-indicators give estimates that are close to the true value for both indicators. Bias corrections for the indicators are small and in the right direction.

Table 5.5: The R-indicators $\hat{R}(\hat{\rho})$ based on, population cross-products and population frequency counts, and the R-indicator $q^2(\hat{\rho})$ for January 2007. Values are given for 25 days and 30 days after the end of the reference month and for the 1:2, 1:9 and 1:18 subsamples. 95% confidence intervals are given.

Sample	$\hat{R}(\hat{\rho})$ - *cross-products* *Type 1*		$\hat{R}(\hat{\rho})$ - *frequency counts* *Type 2*		$q^2(\hat{\rho})$	
	25 days	30 days	25 days	30 days	25 days	30 days
1:2	69,2%	61,5%	62,6%	59,5%	13,51	6,83
1:9	68,8%	61,1%		59,4%	13,51	6,66
1:18	68,6%	83,4%	60,8%	47,6%	15,46	7,28

Table 5.6: The R-indicators $\hat{R}(\hat{\rho})$ based on, population cross-products and population frequency counts, and the R-indicator $q^2(\hat{\rho})$ for June 2007. Values are given for 25 days and 30 days after the end of the reference month and for the 1:2, 1:9 and 1:18 subsamples. 95% confidence intervals are given.

Sample	$\hat{R}(\hat{\rho})$ - *cross-products* *Type 1*		$\hat{R}(\hat{\rho})$ - *frequency counts* *Type 2*		$q^2(\hat{\rho})$	
	25 days	30 days	25 days	30 days	25 days	30 days
1:2	74,0%	72,1%	69,4%	68,3%	0,12	0,05
1:9	74,2%	72,6%	70,0%	69,1%	0,12	0,05
1:18	74,5%	70,1%	71,5%	65,2%	0,11	0,06

Table 5.7: The R-indicators $\hat{R}(\hat{\rho})$ based on, population cross-products and population frequency counts, and the R-indicator $q^2(\hat{\rho})$ for December 2007. Values are given for 25 days and 30 days after the end of the reference month and for the 1:2, 1:9 and 1:18 subsamples. 95% confidence intervals are given.

Sample	$\hat{R}(\hat{\rho})$ - *cross-products* *Type 1*		$\hat{R}(\hat{\rho})$ - *frequency counts* *Type 2*		$q^2(\hat{\rho})$	
	25 days	30 days	25 days	30 days	25 days	30 days
1:2	84,2%	77,0%	83,0%	74,9%	0,12	0,03
1:9	83,0%	75,5%	82,4%	74,4%	0,14	0,03
1:18	80,9%	80,9%	75,5%	75,8%	0,17	0,03

Table 5.5 to 5.7 shows clear patterns for the bias of the indicators. In general the population-based $\hat{R}(\hat{\rho})$ are much smaller than the sample-based $\hat{R}(\hat{\rho})$. The bias of the population-based indicators is large. The bias of the type 1 estimators is in most cases smaller than the bias of the type 2 estimators. There are, however, a few exceptions. There is a clear relation between the size of the bias and the completion rate. The January estimates are considerably more biased than the June and December estimates. Also, the bias is larger for the smaller samples. The differences between the sample-based and population-based $q^2(\hat{\rho})$ are much smaller. The same patterns are visible, i.e. dependence on sample size and completion rate, but to a much smaller extent.

5.3 Comparisons for survey data sets

Table 5.8 gives the various R-indicators for the selected country data sets. The R-indicators are computed for models that contain gender, age and a coding of region or urbanisation of the dwelling of the sample units. The R-indicators for CBS-HS are also computed for a model containing additionally the variables ethnicity, type of household, house value, marital status and job status.

Table 5.8: The unadjusted sample-based R-indicator $\hat{R}(\hat{\rho})$, the unadjusted population-based R-indicators $\hat{R}(\hat{\rho})$ using population cross-products and population frequency counts, and the unadjusted sample-based and population-based R-indicator $q^2(\hat{\rho})$ for the country data sets.

Data set		$\hat{R}(\hat{\rho})$			$q^2(\hat{\rho})$	
		Sample-based	*Pop-based Cross-products Type 1*	*Pop-based Frequencies Type 2*	*Sample*	*Pop*
CBS-HS	Small	89,2%	79,2%	73,7%	0,03	0,06
	Extended	86,4%	72,9%	54,6%	0,05	0,12
SSB-LLS		91,2%	83,9%	92,4%	0,01	0,03
SSB-ESS		88,7%	81,3%	89,2%	0,02	0,05
KUL-ESS		83,1%	80,1%	60,0%	0,11	0,14
SURS-LFS		89.5%	89,7%	90,2%	0,01	0,01

In all cases the type 1 estimators are smaller than the sample-based R-indicators. The picture for the type 2 indicators is somewhat mixed. For the SSB and SURS data sets the type 2 estimators produce values that are similar to the sample-based indicators. For the CBS and KUL data sets the type 2 estimators are smaller than the type 1 estimators and indeed seem to be outliers for the extended CBS-HS model and KUL_ESS model. The population-based $q^2(\hat{\rho})$ are all larger than their sample-based counterparts.

6. Discussion

In this paper we have constructed population-based estimators for the two R-indicators selected in Shlomo et al (2009), i.e. the R-indicator based on the variation of response propensities and the R-indicator based on the variation in response weights. The estimators are applied to simulated data from the 1995 Israel Census Data, the samples from the 2007 Dutch VAT records, and to a number of selected survey data sets from the countries that participate in the RISQ project.

We posed three research questions at the beginning of the paper: How to extend sample-based R-indicators to population-based R-indicators?, What are the statistical properties of population-based R-indicators?, and Are the population-based R-indicators practicable in real survey settings? We consider each question separately.

The extension of sample-based to population-based estimators comprises of two steps: 1) the estimation of response propensities and response influences, and 2) the estimation of the R-indicators based on these propensities and influences.

The population-based estimation of response propensities and response influences proved to be straight forward when linear models are assumed for response propensities and response influences. The sample-based estimators contain sample covariance matrices and sample frequencies that can be replaced by population covariance matrices or population frequencies. We identified two types: either population cross-products are available or auxiliary information is restricted to marginal population counts. We labelled the corresponding estimators as type 1 and type 2 estimators, respectively. The type 2 setting is much more restrictive than the type 1 setting.

We did not consider population-based estimation for other types of models like logistic or probit regression. This would be a welcome extension to the theory of R-indicators as these models are often used in practice and avoid propensities outside the [0, 1] interval. In our analysis we encountered a number of cases where some of the estimated propensities were negative or above 1.

There are many options for the estimation of R-indicators based on the response to the survey. We used propensity weighted response means as the propensities are available. However, any calibration method can be used like linear weighting or adjustment classes. In fact, the set of auxiliary variables used for the estimation of the R-indicators may be a subset of the auxiliary variables used for the estimation of propensities and influences. Parsimonious models may prove to be more efficient as it is known that propensity-weighting may seriously affect the precision of the estimators. Again this is a topic for future research.

We looked at the bias and standard errors of the proposed population-based R-indicators. As expected the bias and standard errors are dependent on the size of the sample and the type of auxiliary information available. The smaller the sample, the larger the bias and the standard error. This is not at all surprising. When samples are smaller, it becomes more difficult to distinguish sampling variation from response variation. Clearly, the confidence intervals become larger as there is less information in small samples.

Type 2 estimators (population marginal counts) have a stronger bias than type 1 estimators (population cross-products). Again this is not surprising as the type 2 estimators provide no information about interactions between auxiliary variables and make it more difficult to discern sampling variation from response variation. Part of the sampling variation is attributed to response variation.

The simulations and the application to real survey data sets show that the bias of the population-based estimators can be considerable. It is not yet clear whether bias corrections are capable of removing this bias, but given the size of the bias it is likely that at least some part of the bias will remain.

From the analyses it becomes apparent that the bias depends also on the number of auxiliary variables. When detailed models are used, containing many variables, in the estimation of response propensities, then the bias may increase considerably. The rationale behind this is that detailed models allow for more sampling variation to be picked up as bias.

Somewhat suprisingly, the bias of population-based estimators also depends on the response rate. The lower the response rate, the more bias.

The research into population-based indicators is still in its infancy and it is too early to draw strong conclusions about the feasibility and practicability of R-indicators based on aggregate population auxiliary information. Nonetheless, the bias of the population-based indicators is considerable and may seriously hamper comparisons between the response to different surveys or even a single survey in time. This may be true especially for the type 2 estimators. Future research needs to find out whether bias corrections are feasible and whether parsimonious models for the estimation of the indicators themselves are fruitful in improving the properties of the indicators.

Acknowledgements: We thank the members of the RISQ project: Katja Rutar from Statistični Urad Republike Slovenije, Geert Loosveldt and Koen Beullens from Katholieke Universiteit, Leuven, Øyvin Kleven, Johan Fosen and Li-Chun Zhang from Statistisk Sentralbyrå, Norway, Ana Marujo from the University of Southampton, UK and Paul Knottnerus, Centraal Bureau voor de Statistiek, for their valuable input.

References

Bethlehem, J., Schouten, B. (2008), Representativity Indicators for Survey Quality (RISQ), collaborative project, 7th Framework Programme FP7SSH20071, Socioeconomic sciences and the humanities Part 8, CBS, Voorburg, The Netherlands.

Efron, B. and Tibshirani, R.J. (1993), An Introduction to the Bootstrap. New York: Chapman and Hall.

Little, R.J.A. (1986) Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 139-157

Little, R.J.A. (1988) Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics*, 6, 287-301

Little, R.J.A. and Rubin,D.B. (2002) *Statistical Analysis with Missing Data*, Hoboken, New Jersey: Wiley

RISQ (2008), RISQ Data set documentations, Deliverable 1, available at www.r-indicator.eu .

Särndal, C-E and Lundström, S. (2005), Estimation in Surveys with Nonresponse, New York, Wiley.

Särndal, C-E and Lundström, S. (2008), Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator, Journal of Official Statistics, 24, 167-191.

Schouten, B., Cobben, F. and Bethlehem, J. (2009), Indicators for the Representativeness of Survey Response. Survey Methodology (to appear)

Schouten, B., Morren, M., Bethlehem, J., Shlomo, N. and Skinner, C. (2009), How to use R-indicators?. Work Package 4, Deliverable 3, RISQ Project, 7th Framework Programme (FP7) of the European Union, available at www.r-indicator.eu

Shlomo, N., Skinner, C.J., Schouten, B., Bethlehem, J., Zhang, L.C. (2009), The statistical properties of R-indicators, RISQ deliverable 2.1, available at www.r-indicator.eu .

Shlomo, N., Skinner, C.J., Schouten, B., Carolina, N., Morren, M. (2009), Partial indicators for representative response, RISQ deliverable 4, available at www.r-indicator.eu .

Wolter, K.M. (2007), Introduction to Variance Estimation. 2nd Ed. New York: Springer.

Appendix A: Analytic approximation to the bias of type 1 $\hat{R}_r(\hat{\rho})$ estimators

In the following assume that $\rho_i > 0$ for all units.

The proposed population-based estimator equals

$$\hat{R} = 1 - 2\sqrt{\frac{1}{N-1}\sum_r d_i \hat{\rho}_i^{-1}(\hat{\rho}_i - \hat{\bar{\rho}}_r)^2}\ , \tag{A1}$$

with $\hat{\rho}_i$ an estimator for the response propensity based on population tables, and $\hat{\bar{\rho}}_r$ the estimated mean of the response propensities

$$\hat{\bar{\rho}}_r = \frac{1}{N}\sum_r d_i \hat{\rho}_i^{-1}\hat{\rho}_i = \frac{1}{N}\sum_r d_i\ . \tag{A2}$$

If we for the moment ignore the form of the estimator $\hat{\rho}_i$ and assume that $E_r(\hat{\rho}_i \mid s) = \rho_i$, i.e. the response expectation of the estimated propensity given the sample is the true propensity, then it follows that

$$E_r(\hat{\bar{\rho}}_r \mid s) = \frac{1}{N}\sum_U E(s_i r_i d_i \mid s) = \frac{1}{N}\sum_U s_i \rho_i d_i = \bar{\rho}_s\ ,$$

i.e. the expectation of $\hat{\bar{\rho}}_r$ is the weighted sample mean of the true propensities.

In section 3.2 we have identified two estimators $\hat{\rho}_i$; one based on all 2 x 2 population tables and one on all marginal population totals. They have the following forms

$$\hat{\rho}_{1,i} = x_i^T\left(\sum_U x_j x_j^T\right)^{-1}\sum_s d_k x_k r_k \tag{A3a}$$

$$\hat{\rho}_{2,i} = x_i^T\left(\frac{N}{\sum_s d_k r_k}\sum_r d_j (x_j - \bar{x}_r)(x_j - \bar{x}_r)^T + N\bar{x}_U \bar{x}_U^T\right)^{-1}\sum_s d_k x_k r_k \tag{A3b}$$

The difference between (A3a) and (A3b) lies in the estimation of the sample covariance matrix of the selected auxiliary variables. We restrict ourselves to the bias of (A3a).

For the moment let us derive the bias $B(\hat{R})$ for an arbitrary estimator $\hat{\rho}_i$. As for the sample-base R-indicator we first look at the bias of the estimated variance of the response propensities $\hat{S}^2(\hat{\rho}_i)$

$$B(\hat{S}^2(\hat{\rho}_i)) = \frac{1}{N-1}E_s E_r \sum_s d_i \hat{\rho}_i^{-1} r_i (\hat{\rho}_i - \hat{\bar{\rho}}_r)^2 - \frac{1}{N-1}\sum_U (\rho_i - \bar{\rho}_U)^2\ . \tag{A4}$$

The expectation in (A4) is with respect to the response probability mechanism and the sampling distribution. We start with the response distribution and linearize $E_r(r_i \hat{\rho}_i^{-1}(\hat{\rho}_i - \hat{\bar{\rho}}_r)^2 \mid s)$. It holds that

$$E\frac{YX^2}{Z} \cong \frac{EY(EX)^2}{EZ} + \mathrm{var}(X)\frac{EY}{EZ} + \mathrm{var}(Z)\frac{(EX)^2\,EY}{(EZ)^3} - 2\,\mathrm{cov}(X,Z)\frac{EXEY}{(EZ)^2} + 2\,\mathrm{cov}(X,Y)\frac{EX}{EZ} - \mathrm{cov}(Z,Y)\left(\frac{EX}{EZ}\right)^2$$

So if we let $X = \hat{\rho}_i - \hat{\bar{\rho}}_r$, $Y = r_i$ and $Z = \hat{\rho}_i$, then we get $E(X\,|\,s) = \rho_i - \bar{\rho}_s$, $E(Y\,|\,s) = \rho_i$, $E(Z\,|\,s) = \rho_i$, and

$$\begin{aligned} E_r(r_i\hat{\rho}_i^{-1}(\hat{\rho}_i - \hat{\bar{\rho}}_r)^2\,|\,s) \cong{}& (\rho_i - \bar{\rho}_s)^2 + \mathrm{var}(\hat{\rho}_i - \hat{\bar{\rho}}_r\,|\,s) + \mathrm{var}(\hat{\rho}_i\,|\,s)\left(\frac{\rho_i - \bar{\rho}_s}{\rho_i}\right)^2 \\ & - 2\,\mathrm{cov}(\hat{\rho}_i, \hat{\rho}_i - \hat{\bar{\rho}}_r\,|\,s)\frac{\rho_i - \bar{\rho}_s}{\rho_i} + 2\,\mathrm{cov}(r_i, \hat{\rho}_i - \hat{\bar{\rho}}_r\,|\,s)\frac{\rho_i - \bar{\rho}_s}{\rho_i} - \mathrm{cov}(r_i, \hat{\rho}_i\,|\,s)\left(\frac{\rho_i - \bar{\rho}_s}{\rho_i}\right)^2 \end{aligned} \tag{A5}$$

After rewriting and combining the variance and covariance terms in (A5) we get

$$\begin{aligned} E_r(r_i\hat{\rho}_i^{-1}(\hat{\rho}_i - \hat{\bar{\rho}}_r)^2\,|\,s) \cong{}& (\rho_i - \bar{\rho}_s)^2 + \mathrm{var}(\frac{\bar{\rho}_s}{\rho_i}\hat{\rho}_i - \hat{\bar{\rho}}_r\,|\,s) \\ & + \mathrm{cov}(r_i, \hat{\rho}_i\,|\,s)\left(1 - \left(\frac{\bar{\rho}_s}{\rho_i}\right)^2\right) - 2\,\mathrm{cov}(r_i, \hat{\bar{\rho}}_r\,|\,s)\frac{\rho_i - \bar{\rho}_s}{\rho_i} \end{aligned} \tag{A6}$$

The first term in (A6) can be expanded analogous to the derivation of bias for the sample-based R-indicator

$$(\rho_i - \bar{\rho}_s)^2 = (\rho_i - \bar{\rho}_U)^2 + (\bar{\rho}_s - \bar{\rho}_U)^2 - 2(\rho_i - \bar{\rho}_U)(\bar{\rho}_s - \bar{\rho}_U)\,. \tag{A7}$$

Now combining (A5) and (A6), we get for the bias

$$\begin{aligned} B(\hat{S}^2(\hat{\rho}_i)) \cong{}& \frac{1}{N-1}E_s\hat{N}_s(\bar{\rho}_s - \bar{\rho}_U)^2 - \frac{2}{N-1}E_s((\bar{\rho}_s - \bar{\rho}_U)(N\bar{\rho}_s - \hat{N}_s\bar{\rho}_U) \\ & + \frac{1}{N-1}E_s\sum_s d_i\,\mathrm{var}(\frac{\bar{\rho}_s}{\rho_i}\hat{\rho}_i - \hat{\bar{\rho}}_r\,|\,s) \\ & + \frac{1}{N-1}E_s\sum_s d_i\,\mathrm{cov}(r_i, \hat{\rho}_i\,|\,s)\left(1 - \left(\frac{\bar{\rho}_s}{\rho_i}\right)^2\right) - \frac{2}{N-1}E_s\sum_s d_i\,\mathrm{cov}(r_i, \hat{\bar{\rho}}_r\,|\,s)\frac{\rho_i - \bar{\rho}_s}{\rho_i} \end{aligned} \tag{A8}$$

Without specifying the sampling distribution and the estimators for $\hat{\rho}_i$, we cannot simplify (A8). We will, however, decompose the third term for convenience

$$\begin{aligned} B(\hat{S}^2(\hat{\rho}_i)) \cong{}& \frac{1}{N-1}E_s\hat{N}_s(\bar{\rho}_s - \bar{\rho}_U)^2 - \frac{2}{N-1}E_s((\bar{\rho}_s - \bar{\rho}_U)(N\bar{\rho}_s - \hat{N}_s\bar{\rho}_U) \\ & + \frac{1}{N-1}E_s\sum_s d_i\left(\frac{\bar{\rho}_s}{\rho_i}\right)^2\mathrm{var}(\hat{\rho}_i\,|\,s) + \frac{1}{N-1}E_s\hat{N}_s\,\mathrm{var}(\hat{\bar{\rho}}_r\,|\,s) - \frac{2}{N-1}E_s\sum_s d_i\left(\frac{\bar{\rho}_s}{\rho_i}\right)\mathrm{cov}(\hat{\rho}_i, \hat{\bar{\rho}}_r\,|\,s) \\ & + \frac{1}{N-1}E_s\sum_s d_i\,\mathrm{cov}(r_i, \hat{\rho}_i\,|\,s)\left(1 - \left(\frac{\bar{\rho}_s}{\rho_i}\right)^2\right) - \frac{2}{N-1}E_s\sum_s d_i\,\mathrm{cov}(r_i, \hat{\bar{\rho}}_r\,|\,s)\frac{\rho_i - \bar{\rho}_s}{\rho_i} \end{aligned} \tag{A9}$$

In (A8) and (A9) $\hat{N}_s = \sum_d d_i$ is the sum of the sample design weights.

In the following we will assume that the sampling design is a simple random sample without replacement and that $N-1 \approx N$. The bias simplifies to

$$\begin{aligned} B(\hat{S}^2(\hat{\rho}_i)) &\cong -\operatorname{var}(\bar{\rho}_s) \\ &+\frac{1}{n} E_s \sum_s \left(\frac{\bar{\rho}_s}{\rho_i}\right)^2 \operatorname{var}(\hat{\rho}_i \mid s) + E_s \operatorname{var}(\hat{\bar{\rho}}_r \mid s) - \frac{2}{n} E_s \sum_s \left(\frac{\bar{\rho}_s}{\rho_i}\right) \operatorname{cov}(\hat{\rho}_i, \hat{\bar{\rho}}_r \mid s) \\ &+\frac{1}{n} E_s \sum_s \operatorname{cov}(r_i, \hat{\rho}_i \mid s)\left(1-\left(\frac{\bar{\rho}_s}{\rho_i}\right)^2\right) - \frac{2}{n} E_s \sum_s \operatorname{cov}(r_i, \hat{\bar{\rho}}_r \mid s) \frac{\rho_i - \bar{\rho}_s}{\rho_i} \end{aligned} \tag{A10}$$

and

$$\operatorname{var}(\bar{\rho}_s) = \frac{1}{n}(1-\frac{n}{N}) S^2(\rho).$$

We drop the index for convenience. First we look at $\operatorname{var}(\hat{\rho}_i \mid s)$, $\operatorname{var}(\hat{\bar{\rho}}_r \mid s)$, $\operatorname{cov}(\hat{\rho}_i, \hat{\bar{\rho}}_r \mid s)$, $\operatorname{cov}(r_i, \hat{\rho}_i \mid s)$ and $\operatorname{cov}(r_i, \hat{\bar{\rho}}_r \mid s)$.

$$\operatorname{var}(\hat{\rho}_i \mid s) = \frac{N^2}{n^2} \sum_{k \in s} \left(x_i^T \left(\sum_U x_j x_j^T\right)^{-1} x_k \right)^2 \rho_k (1-\rho_k) \tag{A11a}$$

$$\operatorname{var}(\hat{\bar{\rho}}_r \mid s) = \frac{1}{n^2} \sum_s \rho_k (1-\rho_k) \tag{A11b}$$

$$\operatorname{cov}(\hat{\rho}_i, \hat{\bar{\rho}}_r \mid s) = \frac{N}{n^2} \sum_{k \in s} x_i^T \left(\sum_U x_j x_j^T\right)^{-1} x_k \rho_k (1-\rho_k) \tag{A11c}$$

$$\operatorname{cov}(r_i, \hat{\rho}_i \mid s) = \frac{N}{n} x_i^T \left(\sum_U x_j x_j^T\right)^{-1} x_i \rho_i (1-\rho_i) \tag{A11d}$$

$$\operatorname{cov}(r_i, \hat{\bar{\rho}}_r \mid s) = \frac{1}{n} \rho_i (1-\rho_i). \tag{A11e}$$

Combining (A10) and (A11a) to (A11e) we get

$$\begin{aligned} B(\hat{S}^2(\hat{\rho}_i)) &\cong -\frac{1}{n}(1-\frac{n}{N}) S^2(\rho) \\ &+\frac{N^2}{n^3} E_s \sum_{i \in s} \left(\frac{\bar{\rho}_s}{\rho_i}\right)^2 \sum_{k \in s} \left(x_i^T \left(\sum_U x_j x_j^T\right)^{-1} x_k \right)^2 \rho_k (1-\rho_k) + \frac{1}{n^2} E_s \sum_{k \in s} \rho_k (1-\rho_k) \\ &-\frac{2N}{n^3} E_s \sum_{i \in s} \left(\frac{\bar{\rho}_s}{\rho_i}\right) \sum_{k \in s} x_i^T \left(\sum_U x_j x_j^T\right)^{-1} x_k \rho_k (1-\rho_k) \\ &+\frac{N}{n^2} E_s \sum_{i \in s} x_i^T \left(\sum_U x_j x_j^T\right)^{-1} x_i \rho_i (1-\rho_i) \left(1-\left(\frac{\bar{\rho}_s}{\rho_i}\right)^2\right) - \frac{2}{n^2} E_s \sum_{i \in s} (1-\rho_i)(\rho_i - \bar{\rho}_s) \end{aligned} \tag{A12}$$

The dominant terms in (A12) are the second and fifth term. We replace $\bar{\rho}_s$ by $E_s \bar{\rho}_s = \bar{\rho}_U$ and ignore higher order terms. Furthermore, we simplify to a simple random sample with replacement, i.e. $E_s s_i s_j = n^2 / N^2$. As a consequence (A12) reduces to

$$\begin{aligned} B(\hat{S}^2(\hat{\rho}_i)) &\cong -\frac{1}{n}(1-\frac{n}{N})S^2(\rho) \\ &+\frac{1}{n}\sum_{i\in U}\left(\frac{\bar{\rho}_U}{\rho_i}\right)^2 \sum_{k\in U}\left(x_i^T\left(\sum_U x_j x_j^T\right)^{-1} x_k\right)^2 \rho_k(1-\rho_k)+\frac{1}{nN}\sum_U \rho_k(1-\rho_k) \\ &-\frac{2}{nN}\sum_{i\in U}\left(\frac{\bar{\rho}_U}{\rho_i}\right)\sum_{k\in U} x_i^T\left(\sum_U x_j x_j^T\right)^{-1} x_k \rho_k(1-\rho_k) \\ &+\frac{1}{n}\sum_{i\in U} x_i^T\left(\sum_U x_j x_j^T\right)^{-1} x_i \rho_i(1-\rho_i)\left(1-\left(\frac{\bar{\rho}_U}{\rho_i}\right)^2\right)-\frac{2}{nN}\sum_{i\in U}(1-\rho_i)(\rho_i-\bar{\rho}_U) \end{aligned} \tag{A13}$$

We estimate (A13) by replacing $\bar{\rho}_U$ by $\hat{\bar{\rho}}_r$, by replacing ρ_i by $\hat{\rho}_i$, and taking weighted sums over respondents instead of sums over population units. The estimator has the following form

$$\begin{aligned} \hat{B}(\hat{S}^2(\hat{\rho}_i)) &\cong -\frac{1}{n}(1-\frac{n}{N})\hat{S}^2(\hat{\rho}) \\ &+\frac{N^2}{n^3}\sum_{i\in r}\left(\frac{\hat{\bar{\rho}}_r^2}{\hat{\rho}_i^3}\right)\sum_{k\in r}\left(x_i^T\left(\sum_U x_j x_j^T\right)^{-1} x_k\right)^2 (1-\hat{\rho}_k)+\frac{1}{n^2}\sum_r (1-\hat{\rho}_k) \\ &-\frac{2N}{n^3}\sum_{i\in r}\left(\frac{\hat{\bar{\rho}}_r}{\hat{\rho}_i^2}\right)\sum_{k\in r} x_i^T\left(\sum_U x_j x_j^T\right)^{-1} x_k (1-\hat{\rho}_k) \\ &+\frac{N}{n^2}\sum_{i\in r} x_i^T\left(\sum_U x_j x_j^T\right)^{-1} x_i (1-\hat{\rho}_i)\left(1-\left(\frac{\hat{\bar{\rho}}_r}{\hat{\rho}_i}\right)^2\right)-\frac{2}{n^2}\sum_{i\in r}(1-\hat{\rho}_i)(1-\frac{\hat{\bar{\rho}}_r}{\hat{\rho}_i}) \end{aligned} \tag{A14}$$

where

$$\hat{S}^2(\hat{\rho}) = \frac{1}{n}\sum_r \hat{\rho}_i^{-1}(\hat{\rho}_i - \hat{\bar{\rho}}_r)^2 \ . \tag{A15}$$

The population-based R-indicator in (A1) is replaced by a bias-adjusted indicator

$$\hat{\tilde{R}} = 1 - 2\sqrt{\hat{S}^2(\hat{\rho}) - \hat{B}(\hat{S}^2(\hat{\rho}))} \ . \tag{A16}$$

Notice that (A16) may still have a small bias that is due to the square root transformation

$$B(\hat{\tilde{R}}) \cong \frac{1}{2}\frac{\mathrm{var}(\hat{S}^2(\hat{\rho}))}{E\hat{S}(\hat{\rho})E\hat{S}^2(\hat{\rho})} \ . \tag{A17}$$

We will, however, ignore this bias as for the sample-based R-indicator.

Appendix B: Analytic approximation to the standard error of $\hat{R}_r(\hat{\rho})$

In the approximation of the standard error we ignore unequal design weights, but assume that $d_i = N/n$ for all population units. Furthermore, we assume that $N \approx N-1$. Let m be the number of responding units in the survey.

We write

$$\hat{R}_r(\hat{\rho}) = 1 - 2\sqrt{\frac{m}{n}\hat{\Delta}},$$

(B1)

where

$$\hat{\Delta} = \frac{1}{m}\sum_r \hat{\rho}_i^{-1}(\hat{\rho}_i - \hat{\bar{\rho}}_r)^2. \tag{B2}$$

As a linear approximation we have

$$\text{var}(\hat{R}_r(\hat{\rho})) \approx \frac{m}{n} E(\hat{\Delta})^{-1}\,\text{var}(\hat{\Delta}).$$

(B3)

We now take a simple approach. We assume that the response is a simple random sample with replacement. Let us first define

$$z_i = \frac{(\hat{\rho}_i - \hat{\bar{\rho}}_r)^2}{\hat{\rho}_i}, \tag{B4}$$

i.e. propensity-weighted quadratic difference between a propensity and the response mean propensity. Now given that $m << N$, we can use the approximation

$$\hat{v}(\hat{\Delta}) = \frac{1}{m(m-1)}\sum_r (z_i - \bar{z}_r)^2. \tag{B5}$$

Combining (B3) and (B5) we get the following estimator for the variance of the R-indicator

$$\hat{v}(\hat{R}_r(\hat{\rho})) = \frac{m}{n\hat{\Delta}}\hat{v}(\hat{\Delta}) = \frac{1}{n\hat{\Delta}}\frac{1}{m-1}\sum_r (z_i - \bar{z}_r)^2. \tag{B6}$$

.