Survey Methodology, June 1995
Vol. 21, No. 1, pp. 79–87
Statistics Canada

Outlier Robust Horvitz-Thompson Estimators

BEAT HULLIGER[1]

ABSTRACT

The Horvitz-Thompson estimator (HT-estimator) is not robust against outliers. Outliers in the population may increase its variance though it remains unbiased. The HT-estimator is expressed as a least squares functional to robustify it through M-estimators. An approximate variance of the robustified HT-estimator is derived using a kind of influence function for sampling and an estimator of this variance is developed. An adaptive method to choose an M-estimator leads to minimum estimated risk estimators. These estimators and robustified HT-estimators are often more efficient than the HT-estimator when outliers occur.

KEY WORDS: Outlier; M-estimator; Adaption; Population mean; Sampling; Sensitivity curve.

1. INTRODUCTION

The mean of a variable over a finite population is an important indicator. Examples are the mean salary of employees in a branch of the economy or the mean yield of corn of the farms in a region. Due to its connection to the sum the mean cannot be easily replaced by other indicators. But the population mean is a sensitive characteristic because a single large observation may determine its value. The Horvitz-Thompson estimator (HT-estimator) is a natural estimator of the population mean if the sample design has unequal inclusion probabilities and is without replacement. It is the sample mean in simple random sampling. It is always unbiased whatever the population distribution of the investigated variable is. But the HT-estimator is not robust against outliers because it is linear in the observed values like its estimand, the population mean. Large observations together with small inclusion probabilities have a particularly large influence on the HT-estimator.

Suppose there is an outlier in a sample. The outlier may be a correct observation from the target population. Discarding such a correct outlier makes the HT-estimator biased. But keeping it with full weight makes the HT-estimator highly variable because typically the outlier would show up only in a few of the possible samples. Thus there is a tradeoff between bias and variance in this case, which in particular includes asymmetric distributions with one heavy tail.

The outlier may also be an incorrect observation, *e.g.*, due to a measurement or coding error or stemming from an element outside the target population. In that case keeping the outlier with full weight may entail a large bias of the HT-estimator in addition to high variability. Thus discarding incorrect outliers reduces both bias and variance.

Since it is often difficult to detect outliers and to decide whether it is correct or not one would like to have estimators that perform well in terms of bias and variance irrespective of the nature and the detection of possible outliers. HT-estimators which are robustified through M-estimators are promising candidates for this difficult task.

In the survey sampling literature the problem of outliers or aberrant values is often treated under the heading "skew populations". Kish (1965, sec. 11.4 B) describes the problem in economical surveys and surveys of individuals. He proposes the formation of separate strata for outliers if possible, truncation, transformation or modelling. The idea of forming a separate class for large units and combining the class means is investigated for example in (Glasser 1962) and (Hidiroglou and Srinath 1981).

The truncation idea is made more precise by the winsorized mean proposed by Searls (1966). Fuller (1991) proposed a preliminary-test-estimator which reduces the impact of the largest data values only when a test for extreme values is significant. Rivest (1993) studied the behavior of various winsorization schemes under simple random sampling. Shoemaker and Rosenberger (1983) derive exact formulae for the expected value and variance of the median and trimmed mean under simple random sampling without replacement. Oehlert (1985) proposes the random average mode estimator to estimate the mean of finite populations in an outlier robust way. Smith (1987) emphasises that it is as important to detect and treat influential observations if the inference is based on the randomisation provided by the sample design as if the observations are considered realisations of random variables. He proposes an influence measure for linear estimators based on case deletion, which involves both the variable of interest and its weight.

The prediction approach in sampling theory uses stochastic models for the population to predict the total of the present realisation. Linear models and (nonrobust) linear estimators are used. Aspects of the sensitivity and robustification against model misspecification are reviewed

[1] Beat Hulliger, Swiss Federal Statistical Office, Schwarztorstrasse 96, CH-3003 Bern, Switzerland.

in (Iachan 1984). Chambers (1986) develops an outlier-robustification of the prediction approach using M-estimators. He distinguishes representative and nonrepresentative outliers in a sample. Representative outliers must be included with full weight in an unbiased estimate of the population mean while nonrepresentative outliers should be downweighted or discarded.

Little and Smith (1987) treat outliers and missing data in certain positive multivariate continuous data by a robustified EM-algorithm. Gwet and Rivest (1992) investigate resistant ratio estimators under simple random sampling without replacement.

M-estimators form a class of flexible and simple robust estimators. An M-estimator T of location is defined implicitely by the estimating equation

$$\sum_{i=1}^{n} \psi(X_i - T) = 0$$

for a predetermined function ψ, *e.g.*, $\psi_{\text{Hub}}(x,k) = \max(-k,\min(k,x))$, where k is a tuning constant. An M-estimator may be written as a functional of the empirical distribution function. The influence function of an estimator is a functional derivative of the estimator (Hampel 1974). It describes the reaction of the estimator to a small contamination in the data. An M-estimator with bounded ψ-function usually has a bounded influence function such that outliers cannot disturb the estimator too much. For the estimation of the mean of asymmetric finite populations M-estimators must be adapted.

In this article we develop design-based M-estimators for samples with unequal inclusion probabilities. The simple linear model which implicitely is the basis of the Horvitz-Thompson strategy is made explicit and the HT-estimator is expressed as a functional of an empirical distribution function which accounts for the complex sample design. This establishes the link to classical robust statistics and allows a straightforward robustification of the HT-estimator (Section 2). We define an influence function for sampling which clarifies the outlier-sensitivity of the HT-estimator and leads to an approximation of the sampling variance of the robustified HT-estimator. An estimator of this variance is presented. In Section (3) we briefly comment on stratification, domains, robust designs and one-step estimators. In Section (4) an adaptive robustification of the HT-estimator is developed. The method chooses from a class of robustified HT-estimators the one which minimizes an estimate of the mean squared error. The resulting estimator is called minimum estimated risk estimator (MER-estimator). A Monte-Carlo simulation is presented in Section 5. Robustified HT-estimators and MER-estimators outperform the HT-estimator in many outlier situations. The premium to pay is a moderate loss of efficiency in situations where the HT-estimator is optimal.

2. ROBUSTIFICATION OF HORVITZ-THOMPSON ESTIMATORS

2.1 The HT-Estimator as a Least Squares Functional

A finite population $U = \{1, \ldots, N\}$ of $0 < N < \infty$ distinct elements is sampled. We are interested in a variable y which takes the values y_i for $i \in U$. The sample design $p(S)$ on the space of samples S of fixed size n has inclusion probabilities $\pi_i = \mathrm{P}[i \in S] = \sum_{S \ni i} p(S)$. These π_i are proportional to some known positive auxiliary variable $x_i\,(i \in U)$. Such sample designs are called IPPS designs (inclusion probability proportional to size) because often x_i is some size measure. Denote by π_{ij} the joint inclusion probability $\mathrm{P}[i \in S, j \in S]$ $(i, j \in U)$. The vector of all y-values is denoted $y_U := (y_1, \ldots, y_N)^T$ and x_U is defined in an analogous way. The vector of the y-values of a sample S is denoted $y_S := (y_{i_1}, \ldots, y_{i_n})^T$ $(i_k \in S)$. The goal is to estimate the population mean of the variable $y: \bar{y}_U := \sum_{i \in U} y_i / N$.

The HT-estimator for $\bar{y}_U$ is $T_{HT} := \sum_{i \in S} y_i / (N\pi_i)$. The variance of T_{HT} is estimated by the well known estimator

$$v_{HT}(T_{HT}) =$$

$$\frac{1}{N^2}\left[\sum_{i \in S} (1 - \pi_i)\frac{y_i^2}{\pi_i^2} + \sum_{i \neq j \in S} (1 - \pi_i\pi_j/\pi_{ij})\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}\right], \tag{1}$$

which is due to Horvitz and Thompson or by the variance estimator due to Yates, Grundy and Sen (see Cochran 1977, p. 261).

The rationale behind the HT-estimator given in the survey sampling literature is that it has sampling variance zero if the inclusion probabilities π_i are exactly proportional to y_i. Then $T_{HT}(y_S) = \bar{y}_U$ for every sample S. The HT-estimator is bias-robust but not variance-robust with respect to deviations from proportionality between y_i and π_i (*cf.* Rao 1966).

How can the HT-estimator be formulated in a way which allows the derivation of an influence function analogue and a variance estimator? The key idea is to express the HT-estimator as a least squares (LS) functional of an estimate of the population distribution function in such a way that the design is incorporated in the estimator of the population distribution function while the proportionality of y_i and x_i is taken up by the LS-functional.

The joint population distribution function of two variables (x_i, y_i) is defined as $F_U(r,t) = \sum_{i \in U} \mathbf{1}\{x_i \leq r\}\,\mathbf{1}\{y_i \leq t\}/N$, where $\mathbf{1}\{y_i \leq t\} = 1$ if $y_i \leq t$ and 0 elsewhere. There are various possibilities to estimate F_U but the easiest and most generally applicable estimator is the sample distribution function

$$F_S(r,t) = \sum_{i \in S} \frac{1}{\pi_i} \mathbf{1}\{x_i \le r\} \mathbf{1}\{y_i \le t\} \Big/ \sum_{i \in S} \frac{1}{\pi_i}. \qquad (2)$$

The estimator F_S is a distribution function itself.

To derive a LS-functional the following superpopulation model for the proportionality between y_i and x_i is used: We assume that y_U is a vector of realisations of independent random variables Y_i with expectation βx_i and variance $\sigma^2 x_i$.

Definition 1. The LS-estimator $\beta_{LS}(F_S)$ of β in the above model with respect to the sampling distribution function F_S of (x_i,y_i) $(i \in S)$ minimizes $\int (y - \beta x)^2 / x \, dF_S(x,y)$ or equivalently solves

$$\sum_{i \in S} \frac{1}{\pi_i} \left(\frac{y_i - \beta x_i}{\sqrt{x_i}} \right) \frac{x_i}{\sqrt{x_i}} = 0. \qquad (3)$$

The following statement is well known and its proof is easy. If S is a sample drawn according to an IPPS sample design with inclusion probabilities $\pi_i = n x_i / \sum_{i \in U} x_i$ $(i \in U)$ then the HT-estimator is $T_{HT} = \bar{x}_U \beta_{LS}(F_S)$, where $\beta_{LS}(F_S)$, the LS-estimator defined by (3):, is given by

$$\beta_{LS}(F_S) = \frac{\sum_{i \in S} y_i / \pi_i}{\sum_{i \in S} x_i / \pi_i}.$$

Note that the expression $T_{HT} = \bar{x}_U \beta_{LS}(F_S) = \bar{x}_U (\sum_{i \in S} y_i / \pi_i) / (\sum_{i \in S} x_i / \pi_i)$ does not depend on the superpopulation model. However the superpopulation model clarifies the role of the HT-estimator: The slope $\beta_{LS}(F_S)$ involved in the HT-estimator is a weighted least squares estimator that incorporates the information in the design through F_S as well as the information in the auxiliary variable through the regression.

2.2 The Robustified HT-Estimator

After the separation of design and auxiliary information and its expression as a LS-functional the robustification of the HT-estimator is analogous to the robustification of LS-estimators in linear models for infinite populations through M-estimators (*cf.* Hampel *et al.* 1986, Chapter 6): The estimating equation (3) now involves some function η which depends on the standardized residuals $(y_i - \beta x_i) / x_i^{1/2}$ and on x_i. For ease of notation denote by a prime the division by $x^{1/2}$ and let $r'(\beta) = (y - \beta x) / x^{1/2}$.

Definition 2. Let $\beta(F_S,\eta)$ be a solution of the equation

$$\sum_{i \in S} \frac{1}{\pi_i} \eta(x_i', r_i'(\beta)) \, x_i' = 0. \qquad (4)$$

The robustified HT-estimator (RHT-estimator) is

$$T_{RHT}(F_S) := \bar{x}_U \beta(F_S,\eta).$$

$\beta(F_S,\eta)$ is called the slope of the RHT-estimator.

In general useful choices of η are of the form $\eta(x,r) = w(x)\psi(r \cdot u(x))$, where $w(x)$ and $u(x)$ are two weighting functions and ψ is a defining function for a location M-estimator (*cf.* Hampel *et al.* 1986, p. 315). In the following we use the so-called Mallows form, which sets $u(x) \equiv 1$. Mallows-type regression downweights outlying x-values and outlying residuals independently. A well-known example, which also sets $w(x) \equiv 1$, is the Huber-function $\eta(x,r) = \psi_{\mathrm{Hub}}(r,k) = \max(-k, \min(k,r))$ for some constant k. The RHT-estimator with defining function $\eta(x,r) \equiv r \; \forall \; x$ is the HT-estimator. Thus by adjusting the tuning constant k in the Huber-function a smooth transition of estimators from the HT-estimator to more and more robust estimators is possible.

Scale estimates are needed in $w(x)$ and $\psi(r)$ to make $\beta(F_S,\eta)$ scale equivariant. While for the weighting function $w(x_i')$ preliminary scale estimators are available, *e.g.*, the median of the x_i', the scale of the residuals must be estimated simultaneously with the slope β. The median of the absolute residuals may be used. In the following theoretical development (Sections 2.3 to 4) scale is assumed known to simplify the treatment.

The RHT-estimator is a nonparametric estimator. The model $\mathbf{E}y = \beta x$ is merely used to motivate the expression of the HT-estimator as a least squares functional. Neither the HT-estimator nor the RHT-estimator need this model or symmetry of errors with variance proportional to x in order to be applied.

Other formulations of the HT-estimator as least squares functionals may be appropriate in certain conditions. Suppose that in spite of the IPPS-design y_i is not correlated with π_i. Then one would probably choose the unweighted sample mean $\bar{y}_S = \sum_{i \in S} y_i / n$ as an estimator of the population mean (*cf.* Rao 1966). A robustification of $\bar{y}_S$ could be a solution $\hat{\mu}$ of $\sum_{i \in S} \psi(y_i - \mu) = 0$. This is a location M-estimator. If the HT-estimator is in fact appropriate due to the correlation between y_i and π_i then this robustification is not efficient.

A third robustification would assume y_i proportional to x_i but with variance proportional to the square of x_i. This is in fact the situation where the HT-estimator is optimal. The corresponding robustification would be a solution $\hat{\beta}$ of $\sum_{i \in S} \eta(x_i, y_i / x_i - \beta) = 0$. Obviously this robustification does not account for the IPPS-sample design. If the design is put back into the estimating equation by solving $\sum_{i \in S} \eta(x_i, y_i / x_i - \beta) / \pi_i = 0$ then we do not get back the HT-estimator when $\eta(x,r) \equiv r$.

One may argue that in fact the HT-estimator is never used in its pure form for estimating population means. The usual estimator is $(\sum_{i \in S} y_i / \pi_i) / (\sum_{i \in S} 1 / \pi_i)$, sometimes called the Hájek-estimator. The estimating equation of the Hájek-estimator, $\sum_{i \in S} (y_i - T) / \pi_i = 0$, makes obvious that the Hájek estimator is not robust against outliers in y.

But the residual $y_i - T$ does not involve the auxiliary variable x_i. Therefore the Hájek-estimator does not suffer from a possible combined effect of large y_i together with small x_i, which may be a leverage point for the regression model underlying the HT-estimator.

2.3 A Sampling Sensitivity Curve

The derivation of an approximate sampling variance of the RHT-estimator (see Section 2.4) uses a finite population analogue to the influence function for infinite populations (Hampel 1974). For finite population sampling with design based inference it is appropriate to develop a sensitivity curve (SC) (*cf.* Hampel *et al.* 1986, p. 93) for $\beta(F,\eta)$ at the population distribution function F_U. In other words, the slope of the RHT-functional is linearized around F_U. Denote by $U+$ the population U augmented by a unit with characteristic (x,y). Denote by $\lambda(\beta,F_U)$ the function $\sum_{i\in U}\eta(x_i',r_i'(\beta))x_i'/N$, such that the defining equation for $\beta(F_U,\eta)$, the M-estimator at the population distribution function, is $\lambda(\beta,F_U) = 0$. Clearly

$$(N+1)[\lambda(\beta(F_{U+},\eta),F_{U+}) - \lambda(\beta(F_U,\eta),F_U)] = 0.$$

Using a linear approximation to $\eta(x,\cdot)$ and neglecting terms in $1/N$ the sensitivity curve of $\beta(F_U,\eta)$ can be isolated from this equation:

$$(N+1)(\beta(F_{U+},\eta) - \beta(F_U,\eta)) \approx \frac{\eta(x',r')x'}{\sum_{i\in U}\eta_2(x_i',r_i')x_i'^2/N} =: \mathrm{SC}(x,y,F_U,\eta), \quad (5)$$

where $\eta_2(x,r) = \partial\eta(x,r)/\partial r$ and both r' and r_i' are evaluated at $\beta(F_U,\eta)$. This SC may be extended to the case of a p-dimensional explanatory variable (*cf.* Hampel *et al.* 1986, p. 316 and Hulliger 1991, p. 183).

Since units usually are not independently included into an IPPS sample, the reaction of the RHT-slope to a particular observation must be investigated by conditioning on a particular sample. The deviation of the estimator $\beta(F_S,\eta)$ at a particular sample S from $\beta(F_U,\eta)$ may be approximated by integrating the SC of $\beta(F,\eta)$ with respect to the sampling distribution function F_S (*cf.* Hampel *et al.* 1986, p. 85):

$$\beta(F_S,\eta) - \beta(F_U,\eta) \approx \int \mathrm{SC}(x,y,F_U,\eta)\,dF_S. \quad (6)$$

The influence of unit i in sample S may then be defined as the contribution of the unit i to the deviation due to the sample S, *i.e.*,

$$\mathrm{SC}((x_i,\pi_i,y_i) \mid S,F_U,\eta) = \frac{\eta(x_i',r_i')x_i'/\pi_i}{(\sum_{j\in S}1/\pi_j)\sum_{j\in U}\eta_2(x_j',r_j')x_j'^2/N}. \quad (7)$$

The SC may be studied theoretically to discuss the properties of the RHT-estimator and to choose a good η-function. And it may be estimated by replacing the standardization factor $N/(\sum_{j\in U}\eta_2(x_j',r_j')x_j'^2)$ by an appropriate estimator. The estimated SC may be used as a tool for outlier detection.

The influence of unit i in sample S on the HT-estimator is

$$\bar{x}_U \mathrm{SC}((x_i,\pi_i,y_i) \mid S,F_U,\eta \equiv r) = (y_i - \beta_{LS}(F_U)x_i)\Big/\Big(1 + \pi_i \sum_{j\in S\setminus i} 1/\pi_j\Big).$$

This SC is unbounded in y_i such that the HT-estimator is not robust against outlying y_i. The y_i influences the HT-estimator through the residual $y_i - \beta_{LS}(F_U)x_i$. This makes clear why a large y_i combined with a small x_i (or small π_i) has a large influence. If π_i is directly proportional to x_i, as the IPPS design in principle requires, then the SC of the HT-estimator is bounded in x_i. In other words the HT-estimator is robust against outlying x_i. However the bound may be quantitatively too high to be efficient and further downweighting of outlying x_i may be necessary.

2.4 Approximate Expectation and Variance

Along the lines of the proof of proposition 2.1 in Gwet and Rivest (1992) it can be shown that $\beta(F_S,\eta)$ is consistent for $\beta(F_U,\eta)$ in the sense that for a growing and nested sequence of populations and IPPS samples $\lim_{N,n\to\infty}P[\,|\beta(F_S,\eta) - \beta(F_U,\eta)| < \epsilon] = 1\;\forall\epsilon > 0$.

Due to the consistency of $\beta(F_S,\eta)$ the sampling expectation $\mathrm{E}_S\beta(F_S,\eta)$ is approximately $\beta(F_U,\eta)$. Of course $\bar{x}_U\beta(F_U,\eta)$ may be different from the population mean and then $\bar{x}_U\beta(F_S,\eta)$ has a bias as an estimator of $\bar{y}_U$. In particular if the population distribution is not symmetric then $\bar{x}_U\beta(F_S,\eta)$ is in general a biased estimator for $\bar{y}_U$ but nevertheless consistent for $\bar{x}_U\beta(F_U,\eta)$. The important question then is how large is the bias of $\bar{x}_U\beta(F_S,\eta)$, in particular when compared with the variance.

The SC (5) may be used to derive a variance approximation. The derivation is analogous to the case of independent identically distributed random variables with the influence function replaced by the sampling SC. Taking the expectation of the square of (6) one gets after some approximations

$$\begin{aligned}\mathrm{Var}_S\beta(F_S,\eta) &\approx \mathrm{E}_S[(\beta(F_S,\eta) - \beta(F_U,\eta))^2] \\ &\approx \frac{\mathrm{Var}_S(\sum_S\eta(x_i',r_i')x_i'/\pi_i)}{(\sum_{i\in U}\eta_2(x_i',r_i')x_i'^2)^2}\end{aligned}$$

$$\approx \frac{\sum_{i \in U}\left(\frac{1}{\pi_i} - 1\right)\eta(x_i', r_i')^2 x_i'^2 + \sum_{i \neq j \in U}\left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right)\eta(x_i', r_i')x_i'\eta(x_j', r_j')x_j'}{\sum_{i \in U}\eta_2(x_i', r_i')^2 x_i'^4 + \sum_{i \neq j \in U}\eta_2(x_i', r_i')x_i'^2\eta_2(x_j', r_j')x_j'^2}, \tag{8}$$

where r_i' is evaluated at $\beta(F_U, \eta)$. Denote this approximate variance by V_r. An important difference to the case of the asymptotic variance of an M-estimator with independent identically distributed random variables is that the cross-product terms in the numerator of V_r do not vanish. If $\eta(x,r) \equiv r$ then V_r yields the correct variance of the HT-estimator.

2.5 Estimation of the Variance

The numerator of V_r is the variance of $\sum_{i \in S}\eta(x_i', r_i')(\beta(F_U, \eta))x_i'/\pi_i$ which is a HT-estimator apart from the unknown $r_i'(\beta(F_U, \eta))$. Therefore the variance estimator (1) for the HT-estimator may be used. After replacing $\beta(F_U, \eta)$ by the estimator $\beta(F_S, \eta)$, the estimator of the variance of the RHT-estimator becomes

$$v_{rHT} = -\bar{x}_U^2 \frac{\sum_{i \in S}\frac{1}{\pi_i}\eta(x_i', r_i')^2 x_i'^2 + \sum_{i \neq j \in S}\frac{1}{\pi_{ij}}\eta(x_i', r_i')x_i'\eta(x_j', r_j')x_j'}{\sum_{i \in S}\frac{1}{\pi_i}\eta_2(x_i', r_i')^2 x_i'^4 + \sum_{i \neq j \in S}\frac{1}{\pi_{ij}}\eta^2(x_i', r_i')x_i'^2\eta_2(x_j', r_j')x_j'^2}. \tag{9}$$

The minus sign in (9) is in order. The (negative) cross-product terms in the numerator usually dominate. Nevertheless v_{rHT} may become negative as can the HT-variance estimator (1) itself (*cf.* Cochran 1977, p. 261). The variance estimator v_{rHT} does not yield the variance estimator (1) if $\eta(x,r) \equiv r$. Of course the Yates-Grundy-Sen estimator may be used to estimate the numerator of V_r. A third variance estimator may be derived by writing the RHT-estimator as a weighted least squares estimator whose weights depend on the estimate (*cf.* Hulliger 1991, p. 166). Since the MER-estimators (*cf.* Section 4) performed slightly better with v_{rHT} than with the other variance estimators the simulations of Section 5 were done with v_{rHT}.

3. EXTENSIONS

3.1 Stratification and Domains

The stratified mean under stratified random sampling is a HT-estimator. The stratified mean may be written as the mean of predicted values under a one-way analysis of variance model. The corresponding robustification is straightforward. It amounts to the separate robustification of the stratum means (Hulliger 1991). However, if the stratum sample size is 1 or 2 no outlier can be down-weighted without the help of further assumptions. Furthermore the biases of the robustified stratum means may add up to a large overall bias (*cf.* Rivest 1993, Section 4). Therefore different robustifications may be appropriate for estimating stratum means and overall means.

This is a general problem for robust estimation in sub-populations (domains) since the definition of an outlier depends on the reference population. An observation may be an outlier in a particular subpopulation but may be harmless in another one. Thus a robust estimator may be suited for one subpopulation but perform poorly in another subpopulation. Often no robustification is needed or wanted for overall means but subpopulation means need to be robustified because of outliers that turn up. Luckily the sample size is often considerably smaller in a subpopulation than in the whole population and then the bias component of the MSE of a robust estimator is often smaller than the variance component. Thus robust estimators may be more efficient than the HT-estimator when used in domain estimation.

3.2 Hansen-Hurwitz Strategy

When sampling is done with replacement and with unequal drawing probabilities the Hansen-Hurwitz estimator is used instead of the HT-estimator. The Hansen-Hurwitz estimator may be robustified analogously to the HT-estimator (see Hulliger 1991, section 4.4) since the underlying model is the same. The variance approximation for the robustified HH-estimator is simpler than for the RHT-estimator because the crossproduct terms vanish due to the drawing with replacement of the Hansen-Hurwitz design.

3.3 Robustified IPPS Design

The ratios y_i/π_i in the HT-estimator act like the summands of an arithmetic mean. Small π_i together with large y_i inflate the HT-estimator. To robustify the design against very large and very small inclusion probabilities we may put $\tilde{\pi}_i = n\tilde{x}_i/\sum_U \tilde{x}_i$, where $\tilde{x}_i = \bar{x}_U + \psi_{\text{Hub}}(x_i - \bar{x}_U, k)$. Thus the auxiliary variable x_i is "Huberised" from its mean to prevent too high and too low values. Now an IPPS sample is drawn with inclusion probabilities $\tilde{\pi}_i$. The HT-estimator is still $T_{HT} = (1/N)\sum_S y_i/\tilde{\pi}_i$ and it is still unbiased. Of course it is not robust against outliers in y and it may loose efficiency if the expectation of the y_i is not proportional to $\tilde{\pi}_i$. The weighted LS-estimator under the superpopulation model for the HT-estimator (see Section 2.1) with inclusion probabilities $\tilde{\pi}_i$ and unmodified auxiliary variable x_i is

$$\beta_{LS}(F_S) = \frac{\sum_S y_i/\tilde{\pi}_i}{\sum_S x_i/\tilde{\pi}_i}, \tag{10}$$

with corresponding estimator for the population mean $\bar{x}_U \beta_{LS}(F_S)$. This β_{LS} may be robustified against outliers in y_i like the HT-estimator. Ratio estimators in IPPS samples are of the same form with the original π_i instead of $\tilde{\pi}_i$. Thus ratio estimators may be robustified analogously to HT-estimators, too (*cf.* Gwet and Rivest 1992).

3.4 One-step Estimators

It is not advisable to express robust estimators as weighted means with fixed weights attached to the observations because the notion and the effect of an outlier depend on the particular domain and variable to be analysed. However, so-called one-step estimators, which are expressed as weighted means, reduce the computational complexity of robust estimators. The one-step RHT-estimator is

$$\bar{x}_U \frac{\sum_{i \in S} w_i y_i' x_i' / \pi_i}{\sum_{i \in S} w_i x_i'^2 / \pi_i}, \tag{11}$$

with weights $w_i = \eta(x_i', y_i' - \beta_{LS} x_i')/(y_i' - \beta_{LS} x_i')$. In fact this is the result of the first step of the iteratively reweighted least squares algorithm, which is often used to calculate M-estimators. The one-step RHT-estimator inherits much of the good properties of the fully iterated RHT-estimator and is simpler to implement and faster to compute.

4. MINIMUM ESTIMATED RISK ESTIMATORS

The RHT-estimator is in general biased. A convenient performance criterium is the sampling mean squared error (MSE) $\mathbf{E}_S[(\bar{x}_U \beta(F_S, \eta) - \bar{y}_U)^2]$. For small to moderate samples the gains of RHT-estimators over the HT-estimator are not very sensitive to the particular robustification chosen if there are outliers in the sample (*cf.* Hulliger 1991, Chapter 3). But with well-behaved data or for moderate to large samples the losses in MSE of certain RHT-estimators may be considerable. The question arises how to choose a good RHT-estimator. Minimum estimated risk estimators (MER-estimators), which adapt the tuning constant of a RHT-estimator to the sample, are a possibility. MER-estimators for the expectation of a univariate random variable are investigated in Hulliger (1991, Chapter 2). The idea is to take a simple M-estimator like a Huber M-estimator, to estimate its MSE for a set of tuning constants k, and to choose the tuning constant with least estimated MSE.

Huber's (1964, p. 97) proposal 3 and Jaeckels (1971) adaptive trimmed mean aim at symmetric random variables and therefore use a variance estimate instead of an estimate of the MSE. MER-estimators are similar but their aim is to estimate the mean of asymmetric distributions.

Here we introduce MER-estimators for IPPS designs. Consider a parametric set of functions $\{\eta_k(x,r) : k \in K\}$, where $K \subset \mathbf{R}_+^p$ is the set of parameters. Usually $p = 1$ or 2 to make minimization feasible and to keep the efficiency loss due to the estimation of the nuisance parameter k low. We do not call k a parameter but a tuning constant to avoid any confusion with the concept of parameters in probability distributions. A suitable set of η-functions induces a set $\mathcal{B} := \{\beta(F_S, \eta_k) : k \in K\}$, where $\beta(F_S, \eta_k)$ is the slope of an RHT-estimator. To ensure consistency of the MER-estimator let $\lim_{k \to \infty} \eta_k(x,r) = r \,\forall (x,r)$ such that the HT-estimator is an element of $\mathcal{B}$. The MSE of $\beta(F_S, \eta_k)$ may be estimated by

$$r(F_S,k) = \max(v_r(F_S,k),0) + (\beta(F_S,k) - \beta_{LS}(F_S))^2, \tag{12}$$

where $v_r(F_S,k)$ is the variance estimator (9) or some other estimator of the variance of $\beta(F_S, \eta_k)$. We use $\max(v_r, 0)$ in $r(F_S,k)$ because the variance estimator (9) may become negative. Typically the function $r(F_S,k)$ with $k \in \mathbf{R}_+$ has a maximum at or close to $k = 0$ which stems from a large bias. Then it drops to a minimum where bias and variance are both small. For large tuning constants $r(F_S,k)$ approaches the variance of the HT-estimator, usually from below.

Definition 3. Suppose $r(F_S,.)$ has a global minimum at $k_m(F_S) \in K$. Then the MER-estimator of the population mean is $M(F_S) = \bar{x}_U \beta(F_S, \eta_{k_m})$.

MER-estimators with suitable defining functions are scale equivariant and do not need a scale estimator. MER-estimators are in general consistent estimators of the population mean. A proof of the strong consistency of MER-estimators of the expectation of a random variable is in Hulliger (1991, Chapter 2).

Problems with nonuniqueness of the minimum or when the minimum is not attained on K are easily resolved in practice by inspection of the function $r(F_S,k)$. (If there are several global minima choose the one with smallest tuning constant to obtain more robustness.) The bias part of $r(F_S,k)$ involves the slope $\beta_{LS}(F_S)$ of the HT-estimator. By this term the sensitivity of the HT-estimator is transferred to MER-estimators and thus the robustness of RHT-estimators is lost again. But if the MER-estimator should be consistent for the population mean there is no way around a consistent and therefore nonrobust estimator in the bias part of the risk estimator. Nevertheless MER-estimators are quantitatively less sensitive to outliers and more efficient than the HT-estimator if outliers occur (see Section 5).

It is even possible to bound the influence of outliers on the MER-estimator for finite samples without loosing its (asymptotic) consistency. This is achieved by downweighting the bias part in the estimated risk of the HT-estimator in an appropriate way (MER2-estimators, (Hulliger 1991, Paragraph 2.4.1)).

MER-estimators may be more efficient than the HT-estimator because their bias is more than compensated by the variance reduction due to the downweighting of outliers. How much can be gained quantitatively is explored in Section 5.

5. A SMALL SIMULATION STUDY

Simulations with populations of size $N = 128$ and with samples of size $n = 16$ are presented here. The sample design in Dey and Srivastava (1987) is used (Note that there is a factor 2 missing in their formula (2.3)). Dey and Srivastava propose to form $m > n/2$ groups. The group totals $\sum_{U_j} x_i (j = 1, \ldots, m)$ must fulfill the inequality $\sum_{U_j} x_i / \sum_U x_i > (n - 2)/(n(m - 1))$. Thus the group totals are allowed only little variability and the groups are difficult to form in particular for larger samples (Hulliger 1991, p. 179).

The $x_i (i = 1, \ldots, N)$ are independent realisations according to a 5%-scale contaminated exponential distribution with origin at 1, *i.e.*, $(X_i - 1) \sim 0.95\,\mathrm{Exp}(1) + 0.05\,\mathrm{Exp}(3)$, where $\mathrm{Exp}(\theta)$ denotes the exponential distribution function $1 - \exp(-x/\theta)$. The shift $+1$ is introduced to lower the probability of negative responses in the regression through the origin model with symmetric errors.

The first response $y_U^{(1)}$, with acronym GODA, is a realization of independent normal variables distributed as $Y_i \sim \mathcal{N}(100x_i, x_i^2)$. This is the model under which the HT-estimator is optimal (*cf.* Godambe 1955). The response $y_U^{(2)}$ (HTLS) is a realization of independent variables distributed as $Y_i \sim \mathcal{N}(2x_i, x_i/4)$. This is the ideal model that yields the HT-estimator as LS-estimator. A third response $y_U^{(3)}$ (HTG) is created by the model $Y_i \sim 0.95\mathcal{N}(2x_i, x_i/4) + 0.05\mathcal{N}(2x_i, 9x_i/4)$. The residual outliers have 3 times larger scale. The response $y_U^{(4)}$ (HTE) has asymmetric outliers which are not related to the x-variable. The bulk of the data (120 observations) stems from the distribution $Y_i \sim \mathcal{N}(2x_i, x_i/4)$ of $y_U^{(2)}$ but 8 randomly chosen observations stem from Exp(2.5). The population $y_U^{(5)}$ (HMT) stems from a distribution with expectation $0.4 + 0.25x_i$ and has a Gamma distribution with variance proportional to $x^{3/2}$. Thus the variable y has the distribution proposed in Hansen, Madow and Tepping (1983, p. 781). Finally a population $y_U^{(6)}$ (HMTE) is generated with 120 observations from the same distribution as $y_U^{(5)}$ but with 8 randomly chosen observations from the distribution Exp(2). The six populations above are chosen to be realistic. They all use the same population of x-values (see Figure 1).

The RHT estimator in the simulation uses

$$\eta(x_i', r_i') = w(x_i', k_x)\psi_{\mathrm{Hub}}(r_i', k_r \,\mathrm{med}_S \,|\, r_i' \,|),$$

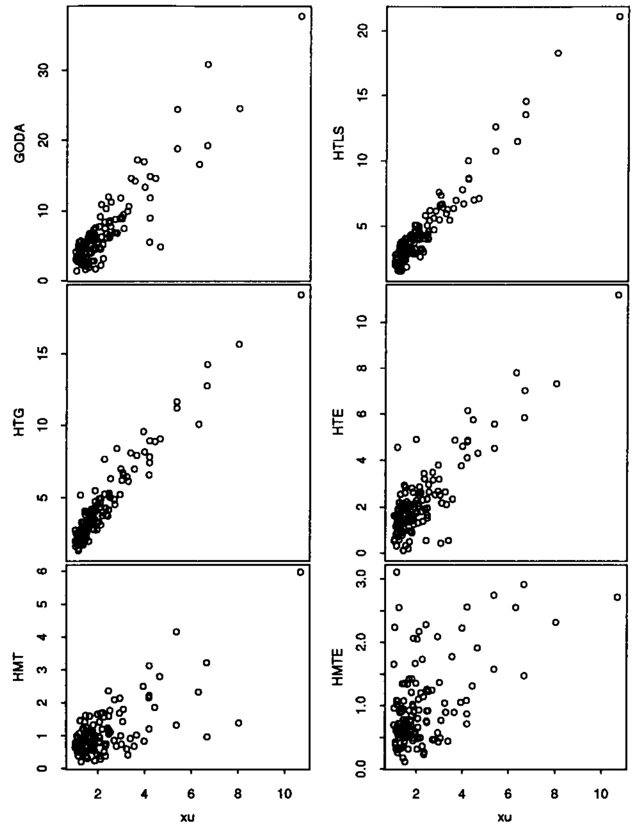

Figure 1. Populations of the Monte-Carlo Study.

with $w(x_i', k_x) = \min(1, k_x \,\mathrm{med}_U \,|\, x_i' \,| / |\, x_i' \,|)$ and $k_x = k_r = 2$. The weighting function $w(x_i', k_x)$ corresponds to an asymmetric Huber-function $\psi_{a\mathrm{Hub}} = \min(x_i', k_x)$, which downweights large x_i' only. The scale $\mathrm{med}_S \,|\, r_i' \,|$ is the median of the absolute residuals evaluated at the solution of the preceding iteration of the iteratively reweighted least squares algorithm. The MER-estimator uses the same η with tuning constants k_x, k_r evaluated at 20 points which lie on the diagonal of the range of k_x and k_r. S-PLUS functions for the calculation of the estimators may be obtained from the author.

For each of the populations a set of 400 samples was drawn to evaluate the estimators. The obtained precision is sufficient to draw conclusions (see the standard errors of the efficiencies in Table 1).

The results are presented in Table 1. The relative bias of the RHT-estimator is always larger than the relative bias of the MER-estimator. The biases of the two estimators have the same sign, except when they are very small. With the exception of populations HTE and HMTE the variance of the RHT-estimator is larger than the variance of the MER-estimator. While the RHT-estimator looses 9% efficiency at population GODA, where the HT-estimator should be optimal, the MER-estimator looses little. With population HTLS, where the HT-estimator is the least squares estimator, the RHT-estimator looses about 12%.

Table 1

Monte-Carlo simulations with RHT- and MER-estimator

	Populations					
	GODA	HTLS	HTG	HTE	HMT	HMTE
MC-mean of HT	6.996	4.531	4.483	2.271	1.068	0.991
Rel. bias of RHT	−0.002	−0.001	−0.009	−0.009	0.006	−0.052
Rel. bias of MER	0.000	−0.001	−0.007	−0.008	−0.002	−0.035
Rel. SE of HT	0.067	0.041	0.044	0.098	0.107	0.170
Rel. SE of RHT	0.070	0.044	0.040	0.087	0.117	0.144
Rel. SE of MER	0.068	0.042	0.040	0.091	0.107	0.146
Eff. of RHT	0.911	0.876	1.110	1.310	0.827	1.234
Eff. of MER	0.969	0.981	1.158	1.194	0.989	1.284
MC-SE of eff. RHT	0.020	0.017	0.073	0.009	0.018	0.001
MC-SE of eff. MER	0.003	0.009	0.037	0.002	0.013	0.002

NOTE: Relative bias and relative standard error (rel. SE) are biases and standard errors divided by the MC-mean of the HT-estimator. Efficiencies (Eff.) are MSE of the HT-estimator divided by the MSE of the estimator. Estimated standard errors of these Monte-Carlo estimates of efficiency are given in the last two lines.

The efficiency loss of the MER-estimator is once again small. Population HTG contains symmetric residual outliers. The RHT-estimator gains about 11% (but see the error of 7.3%) and the MER-estimator about 16%. Under the asymmetric outliers of population HTE the gain of the RHT-estimator is 31% while the MER-estimator gains 19%. If neither the regression through the origin, nor the symmetry of errors or the proportionality of their variance to the explanatory variable holds, *i.e.*, for population HMT, then the RHT-estimator looses 17% compared with the HT-estimator while the MER-estimator looses practically nothing. If in such a population a few asymmetric outliers turn up like in population HMTE then both robust estimators gain considerably against the HT-estimator, namely 23% and 28% respectively.

In conclusion from this limited simulation the MER-estimator looses little in terms of MSE, compared with the HT-estimator, when there are no outliers in the population. It gains moderately in populations with symmetric outliers and considerably when the outliers are asymmetric. The RHT-estimator looses more under ideal situations than the MER-estimator. The adaptivity of the MER-estimators pays off.

Extensive simulations with infinite populations in Hulliger (1991) confirm these conclusions and show that the gains of robust estimators may be very large for skew populations with outliers. However the possible efficiency gains with robust estimators vanish for large samples since then the bias dominates MSE. On the other hand if the outliers that turn up in a sample are not representative, *e.g.*, if they are uncorrected coding errors, then the robust estimators are much more efficient than the HT-estimator for all sample sizes.

ACKNOWLEDGEMENTS

This article is an outgrowth of the authors Ph.D. thesis at ETH Zürich. The author would like to thank Prof. F.R. Hampel and Prof. H.R. Künsch, for their advice. The author is grateful for the valuable comments of two anonymous referees on drafts of this paper.

REFERENCES

CHAMBERS, R.L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, 81, 1063-1069.

COCHRAN, W.G. (1977). *Sampling Techniques* (3rd. Ed.). New York: Wiley.

DEY, A., and SRIVASTAVA, A.K. (1987). A sampling procedure with inclusion probabilities proportional to size. *Survey Methodology*, 13, 85-92.

FULLER, W.A. (1991). Simple estimators of the mean of skewed populations. *Statistica Sinica*, 1, 137-158.

GLASSER, G.J. (1962). On the complete coverage of large units in a statistical study. *International Statistical Review*, 30, 28-32.

GODAMBE, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society*, Series B, 17, 269-278.

GWET, J.-P., and RIVEST, L.-P. (1992). Outlier resistant alternatives to the ratio estimator. *Journal of the American Statistical Association*, 87, 1174-1182.

HAMPEL, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69, 383-393.

HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J., and STAHEL, W.A. (1986). *Robust Statistics*. New York: Wiley.

HANSEN, M.H., MADOW, W.G., and TEPPING, B.J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78, 776-807.

HIDIROGLOU, M.A., and SRINATH, K.P. (1981). Some estimators of a population total from simple random samples containing large units. *Journal of the American Statistical Association*, 76, 690-695.

HUBER, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, 73-101.

HULLIGER, B. (1991). Nonparametric M-estimation of a population mean. Doctoral Dissertation ETH No. 9443, ETH Zürich.

IACHAN, R. (1984). Sampling strategies, robustness and efficiency: the state of the art. *International Statistical Review*, 52, 209-218.

JAECKEL, L.A. (1971). Robust estimates of location: symmetry and asymmetric contamination. *Annals of Mathematical Statistics*, 42, 1020-1034.

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

LITTLE, R.J.A., and SMITH, Ph.J.(1987). Editing and imputation for quantitative survey data. *Journal of the American Statistical Association*, 82, 58-68.

OEHLERT, G.W. (1985). The random average mode estimator. *Annals of Statistics*, 13, 1418-1431.

RAO, J.N.K. (1966). Alternative estimators in PPS sampling for multiple characteristics. *Sankhyā* A, 28, 47-60.

RIVEST, L.-P. (1993). Winsorization of survey data. *Proceedings of the 49th Session, International Statistical Institute.*

SEARLS, D.T. (1966). An estimator for a population mean which reduces the effect of large observations. *Journal of the American Statistical Association*, 61, 1200-1204.

SHOEMAKER, L.H., and ROSENBERGER, J.L. (1983). Moments and efficiency of the median and trimmed mean for finite populations. *Communications in Statistics, Simulations and Computations*, 12(4), 411-422.

SMITH, T.M.F. (1987). Influential observations in survey sampling. *Journal of Applied Statistics*, 14, 143-152.

STATISTICAL SCIENCES, INC. (1990). *S-PLUS Software*, Seattle: Statistical Science, Inc.