A UNIFIED APPROACH TO ROBUST ESTIMATION IN FINITE POPULATION SAMPLING

Jean-François BEAUMONT (*), David HAZIZA (**), Anne RUIZ-GAZEN (***)

(*) Statistique Canada,
(**) Université de Montréal,
(***) Université de Toulouse, Département de mathématiques et de statistique

Abstract

The influence function is well known in classical statistics. It is often used in the context of outlier-robust estimation to study properties of estimators or derive robust estimators. There have been attempts to adapt the influence function to finite population sampling under the design-based approach to inference. Although many of these adaptations are quite useful for variance estimation, they are not entirely satisfactory for robust estimation; either they fail to fully take the sampling design into account or they are not easily generalizable to arbitrary sampling designs.

We show that, in classical statistics, the influence function has an approximate relationship with the conditional bias. This suggests that the conditional bias might be a useful measure of influence since it can be easily extended to any inferential framework unlike the influence function. We use the conditional bias to derive robust estimators in finite population sampling by downweighting the most influential sample units. Under the model-based approach to inference, our proposed robust estimator is closely related to the well-known estimator of Chambers (1986). Under the design-based approach, it possesses the desirable feature of being applicable with arbitrary sampling designs. For stratified simple random sampling, it is essentially equivalent to the estimator of Kokic and Bell (1994). The proposed robust estimator involves a ψ-function, which depends on a tuning constant. In this paper, we propose a method for determining the tuning constant, which consists of minimizing the maximum estimated conditional bias. A limited simulation study is conducted to investigate the performance of the proposed robust estimator in terms of relative bias and relative efficiency. Finally, we discuss a popular method used in practice that consists of setting equal to one the weight of units identified as influential. We show that this method can be linked to the concept of conditional bias.

1. Introduction

1.1. The influence function in classical statistics

The influence function (Hampel, 1974) is well known in classical statistics. It is often used in the context of outlier-robust estimation to study properties of estimators or derive robust estimators. To fix ideas, let Y_i, $i=1,...,n$, be *n* independent random variables having the same distribution *F* and suppose that we are interested in estimating $\theta = t(F)$. For a given fixed value *y*, the influence function is defined as

$$IF(y;t,F) = \lim_{\varepsilon \to 0} \frac{t\left((1-\varepsilon)F + \varepsilon\delta_y\right) - t(F)}{\varepsilon} \quad , \tag{1.1}$$

where δ_y is the degenerate distribution in *y*. If $\theta = t(F)$ is the mean of *F*, the influence function reduces to $IF(y;t,F) = y - \theta$, which is clearly unbounded. Robustness is typically achieved by choosing functionals $t(F)$ that have a bounded influence function such as the median of *F*.

Let us now consider the estimator $\hat{\theta} = t(\hat{F})$ of θ, where $\hat{F} = \sum_{i=1}^{n} \delta_{Y_i} \big/ n$ is the empirical distribution function. A well-known approximation (e.g., Hampel, Ronchetti, Rousseeuw, and Stahel, 1986, p. 85) that uses the influence function (1.1) is

$$\hat{\theta} \approx \theta + \frac{1}{n} \sum_{i=1}^{n} IF(Y_i;t,F), \tag{1.2}$$

with $E_F\left\{IF(Y_i;t,F)\right\} = 0$, $i = 1,...,n$. This approximation is sometimes used to estimate the variance of $\hat{\theta}$.

1.2. The conditional bias and its relationship with the influence function in classical statistics

In the context of classical statistics, Muñoz-Pichardo, Muñoz-Garcia, Moreno-Rebollo and Piño-Mejias (1995) proposed to use the conditional bias

$$B(y_i;\theta) = E_F\left(\hat{\theta} \mid Y_i = y_i\right) - \theta \tag{1.3}$$

as a measure of the influence of the *i*th observation. Using the approximation (1.2), with $E_F\left\{IF(Y_i;t,F)\right\} = 0$, it is straightforward to approximate the conditional bias (1.3) by

$$B(y_i;\theta) \approx \frac{1}{n} IF(y_i;t,F) \ . \tag{1.4}$$

The conditional bias is thus approximately proportional to the influence function $IF(y_i;t,F)$. This justifies why it can be viewed as an influence measure for the *i*th observation. From (1.2) and (1.4), the error $\hat{\theta} - \theta$ can be approximated as

$$\hat{\theta} - \theta \approx \sum_{i=1}^{n} B(Y_i;\theta). \tag{1.5}$$

The conditional bias can thus also be viewed as the contribution of the *i*th observation to the error $\hat{\theta} - \theta$. Robustness is achieved by curbing the influence (or contribution) of the largest observations on this error.

Example: Suppose that θ is the mean of distribution *F* and $\hat{\theta} = \sum_{i=1}^{n} Y_i \big/ n$ is the sample mean. The conditional bias is $B(y_i;\theta) = (y_i - \theta)/n$ and the estimator $\hat{\theta}$ is the solution in $\tilde{\theta}$ to the estimating equation $\sum_{i=1}^{n} B(Y_i;\tilde{\theta}) = 0$ or, alternatively, $\sum_{i=1}^{n} (Y_i - \tilde{\theta}) = 0$. It is well known that $\hat{\theta}$ is not robust. The M-estimator, $\hat{\theta}^M$, is a common robust alternative to $\hat{\theta}$. It is obtained as the solution in $\tilde{\theta}$ to the estimating equation $\sum_{i=1}^{n} \psi(Y_i - \tilde{\theta}) = 0$, where $\psi(\cdot)$ is any bounded function such that $\psi(z) \approx z$ when *z* is close to 0. A typical choice is the Huber function:

$$\psi(z;c) = \text{sign}(z) \times \min\left(|z|, c\right)$$

where *c* is a positive tuning constant and $\text{sign}(z) = 1$, $z \geq 0$, while $\text{sign}(z) = -1$, otherwise. A large choice of *c* yields the non-robust estimator $\hat{\theta}$ while choosing c close to zero yields the sample median, which is more robust than $\hat{\theta}$ but could potentially be biased as an estimator of θ and inefficient under the ideal parametric distribution *F*. The choice of the tuning constant *c* is usually determined by making a compromise between robustness and efficiency.

1.3. The influence function in finite population sampling

In finite population sampling, inference is usually made with respect to the known probability sampling design $P(S)$ that is used to select a random sample *S* from the finite population *U* of size *N*. This is often called the design-based approach to inference. In this approach, only the sample inclusion indicators are random; all other quantities are treated as being fixed. The influence function (1.1) seems irrelevant in the design-based approach because there is no underlying distribution *F* that is assumed to have generated the population values y_i, $i \in U$, at least for inference purposes. To circumvent this difficulty, one might view the sampling design *P*, which generates the sample inclusion indicators, as playing the role of the distribution *F* and continue using (1.1) or a similar definition. However, there does not seem to be any natural and unique definition of a functional $t(P) = \theta$, with θ being here some finite population parameter; e.g., the population mean $\theta = \sum_{i \in U} y_i \big/ N$.

A more frequent way of defining the influence function in finite population sampling consists of replacing *F* by $F_N = \sum_{i \in U} \delta_{y_i} \big/ N$ so that the finite population parameter θ can be expressed as $\theta = t(F_N)$. Unfortunately, this leads to an influence function that fails to account for the sampling design. For instance, if $\theta = t(F_N)$ is the population mean of F_N, i.e. $\theta = \sum_{i \in U} y_i \big/ N$, the influence function is again $IF(y;t,F_N) = y - \theta$. The quantity $IF(y_i;t,F_N) = y_i - \theta$ is not a good measure of the influence of unit *i* because it ignores the sampling design. As an example, suppose that unit *i* is selected with certainty in the sample. It would seem intuitively appealing to consider an influence measure that is equal to 0 for this unit. This is unfortunately not the case of $IF(y_i;t,F_N) = y_i - \theta$ because it does not involve the sampling design.

Approximations analogous to (1.2) have also been developed within the design-based approach (e.g., Campbell, 1980; Gwet and Rivest, 1992; Deville, 1999; and Demnati and

Rao, 2004). All these methods are quite useful for variance estimation, but they lead to influence measures that again do not account for the sampling design. Approximations similar to (1.2) may thus not provide a useful starting point for determining an influence measure for unit *i* that can then be used to derive robust estimators and study their properties.

Assuming with-replacement sampling, and thus independent and identically distributed observations, Zaslavsky, Schenker and Belin (2001) extended the definition (1.1) to finite population sampling under the design-based approach. Hulliger (1995) defined a sensitivity curve for probability proportional to size sampling. He took the sampling design into account by expressing the finite population parameter θ as a function of the population values of the size variable. Although the sampling design is involved in both methods, they again yield a non-zero influence measure for a unit *i* selected with certainty in the sample. Also, it does not seem easy to generalize these methods to arbitrary sampling designs with possibly large sampling fractions.

We have pointed out above that the extension of the influence function to the design-based approach to inference in finite population sampling is not trivial. In section 1.2, we have also shown the relationship between the influence function and the conditional bias in classical statistics. This suggests that the conditional bias might be a useful measure of influence since it can be easily extended to any inferential framework unlike the influence function. In the next two sections, we use the conditional bias to derive robust estimators in finite population sampling by downweighting the most influential sample units. We first consider the model-based approach to inference in section 2. Our robust estimator is closely related to the well-known estimator of Chambers (1986). We then consider the design-based approach to inference in sections 3, 4 and 5. In this approach, the conditional bias has been earlier studied by Moreno-Rebollo, Muñoz-Reyes and Muñoz-Pichardo (1999) and Moreno-Rebollo, Muñoz-Reyes, Jimenez-Gamero and Muñoz-Pichardo (2002). Our proposed robust estimator possesses the desirable feature of being applicable with arbitrary sampling designs. For stratified simple random sampling, it is essentially equivalent to the estimator of Kokic and Bell (1994).

2. Inference for finite populations : the model-based approach

In the model-based approach to inference in finite population sampling (e.g., Valliant, Dorfman and Royall, 2000), the *Y*-values of the *N* population units are assumed to be generated by some model. We denote by **X**, the known *N*-row matrix containing the vector of explanatory variables $\mathbf{x}_i'$ in its *i*th row. Often, the following linear model is considered:

Model: $(Y_i - \mathbf{x}_i'\boldsymbol{\beta})/\sigma_i$ given **X**, $i \in U$, are mutually independent and have all the same distribution *F*, where $\boldsymbol{\beta}$ is a vector of unknown model parameters and σ_i is usually assumed to be known up to a constant factor. Furthermore, the distribution *F* has a mean of 0 and a variance of 1.

A non-informative sample *s* is selected from the finite population *U* and is treated as fixed when making inferences. The interest is in the prediction of a function of the population *Y*-values through the sample *Y*-values. To fix ideas, we will assume that we are interested in predicting the random population total $\theta = \sum_{i \in U} Y_i$. Royall (1976) proposed the Best Linear Unbiased Predictor (BLUP) of $\sum_{i \in U} Y_i$, which can be expressed as

$$\hat{\theta} = \sum_{i \in s} w_i Y_i \qquad (2.1)$$

with the weights

$$w_i = 1 + \frac{\mathbf{x}_i'}{\sigma_i^2}\left(\sum_{i\in s}\frac{\mathbf{x}_i\mathbf{x}_i'}{\sigma_i^2}\right)^{-1}\left(\sum_{i\in U-s}\mathbf{x}_i\right) . \qquad (2.2)$$

In this context, the conditional bias attached to unit *i* is given by

$$B_i(y_i;\boldsymbol{\beta}) = E_F\left(\hat{\theta}-\theta \mid s, Y_i = y_i\right) . \qquad (2.3)$$

Note that definition (2.3) is slightly different than definition (1.3) to account for the fact that θ is a random variable in this section. Using (2.1) and noting from (2.2) that $\sum_{i\in s} w_i\mathbf{x}_i = \sum_{i\in U}\mathbf{x}_i$, the conditional bias (2.3) can be expressed as

$$B_i(y_i;\boldsymbol{\beta}) = \begin{cases}(w_i-1)(y_i-\mathbf{x}_i'\boldsymbol{\beta}) & ,i\in s\\ -(y_i-\mathbf{x}_i'\boldsymbol{\beta}) & ,i\in U-s.\end{cases} \qquad (2.4)$$

This expression highlights that the conditional bias takes a different form depending on whether unit *i* has been selected in the sample or not. The prediction error of the BLUP can be written as

$$\hat{\theta}-\theta = \sum_{i\in U} B_i(Y_i;\boldsymbol{\beta}) . \qquad (2.5)$$

Therefore, the conditional bias $B_i(Y_i;\boldsymbol{\beta})$ can be interpreted as the contribution of unit *i* to the prediction error of the BLUP. Although they did not consider the concept of the conditional bias, Beaumont and Rivest (2009) showed that this decomposition of the prediction error holds for any weighted estimator that satisfies the calibration equation $\sum_{i\in s} w_i\mathbf{x}_i = \sum_{i\in U}\mathbf{x}_i$.

To construct a robust version of the BLUP, we first express it as:

$$\hat{\theta} = \left(\hat{\theta}-\sum_{i\in U} B_i(Y_i;\boldsymbol{\beta})\right) + \sum_{i\in U} B_i(Y_i;\boldsymbol{\beta}) . \qquad (2.6)$$

From (2.5), the first term on the right-hand side of (2.6) is equal to θ and is thus not affected at all by influential units. The error comes entirely from the second term. To obtain robustness, it would be desirable to reduce the contribution of the largest $B_i(Y_i;\boldsymbol{\beta})$ on this second term. It is not possible for nonsample units as their conditional bias is not observed. Thus, nothing can be done at the estimation stage to obtain protection against influential units in the nonsample portion of the population. Protection can be achieved against the occurrence of influential sample units by downweighting their contribution in the second term on the right-hand side of (2.6). Using (2.4), this leads to the robust estimator:

$$\begin{aligned}\hat{\theta}^R(\boldsymbol{\beta}) &= \left(\hat{\theta}-\sum_{i\in U} B_i(Y_i;\boldsymbol{\beta})\right) + \sum_{i\in s}\psi\left\{B_i(Y_i;\boldsymbol{\beta})\right\} + \sum_{i\in U-s} B_i(Y_i;\boldsymbol{\beta})\\ &= \left(\hat{\theta}-\sum_{i\in s} B_i(Y_i;\boldsymbol{\beta})\right) + \sum_{i\in s}\psi\left\{B_i(Y_i;\boldsymbol{\beta})\right\} \qquad (2.7)\\ &= \sum_{i\in s} Y_i + \sum_{i\in U-s}\mathbf{x}_i'\boldsymbol{\beta} + \sum_{i\in s}\psi\left\{(w_i-1)(Y_i-\mathbf{x}_i'\boldsymbol{\beta})\right\} .\end{aligned}$$

In general, the vector $\boldsymbol{\beta}$ is unknown and must be replaced by some estimator $\hat{\boldsymbol{\beta}}$, which yields the robust estimator $\hat{\theta}^R(\hat{\boldsymbol{\beta}})$. The estimator $\hat{\boldsymbol{\beta}}$ could be any robust estimator developed in classical statistics, as in Chambers (1986), or could be obtained using an independent source of data, as in Kokic and Bell (1994). Other non-robust estimators should be avoided, as prescribed in the literature. The function $\psi(\cdot)$ can be the Huber function. If the underlying tuning constant *c* is large, the estimator $\hat{\theta}^R(\hat{\boldsymbol{\beta}})$ reduces to the BLUP for every choice of $\hat{\boldsymbol{\beta}}$.

It is interesting to note that the robust estimator $\hat{\theta}^R(\hat{\boldsymbol{\beta}})$, obtained from (2.7), is closely related to the estimator developed by Chambers (1986) using other arguments not involving the conditional bias. Chambers' estimator is:

$$\hat{\theta}^C(\hat{\boldsymbol{\beta}}) = \sum\nolimits_{i\in s} Y_i + \sum\nolimits_{i\in U-s} \mathbf{x}_i'\hat{\boldsymbol{\beta}} + \sum\nolimits_{i\in s} (w_i - 1)\hat{\sigma}_i \psi\left\{(Y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})/\hat{\sigma}_i\right\} , \qquad (2.8)$$

where $\hat{\sigma}_i$ is a robust estimator of σ_i. It reduces the impact of large standardized residuals $(Y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})/\hat{\sigma}_i$ but, unlike $\hat{\theta}^R(\hat{\boldsymbol{\beta}})$, does not address the combined influence of weights $(w_i - 1)\hat{\sigma}_i$ and standardized residuals. Both $\hat{\theta}^R(\hat{\boldsymbol{\beta}})$ and $\hat{\theta}^C(\hat{\boldsymbol{\beta}})$ are special cases of a slightly more general estimator given in equation (6) of Beaumont and Rivest (2009). In an empirical study, they have shown a slight superiority of $\hat{\theta}^R(\hat{\boldsymbol{\beta}})$ over $\hat{\theta}^C(\hat{\boldsymbol{\beta}})$.

From the above discussion, it seems that the conditional bias is useful for deriving robust estimators in the model-based approach to inference. Next, we apply this general methodology to the design-based approach to inference.

3. Inference for finite populations : the design-based approach

As pointed out in section 1.3, the *y*-values of the *N* population units are treated as fixed in the design-based approach to inference and a random sample *S* is selected from the finite population *U* according to a probability sampling design $P(S)$. We denote by I_i, $i \in U$, the *N* sample inclusion indicators such that $I_i = 1$, if $i \in S$, and $I_i = 0$, otherwise. Let us suppose again that we are interested in estimating the finite population total $\theta = \sum\nolimits_{i\in U} y_i$ and that we consider the Horvitz-Thompson estimator $\hat{\theta}^{HT} = \sum\nolimits_{i\in S} d_i y_i$, where $d_i = 1/\pi_i$ is the design weight attached to unit *i* and $\pi_i = \Pr(I_i = 1)$ is its first-order inclusion probability. The Horvitz-Thompson estimator is design-unbiased; i.e., $E_P(\hat{\theta}^{HT}) = \theta$, where the subscript *P* indicates that the expectation is evaluated with respect to the sampling design. For a sample unit, the conditional bias of the Horvitz-Thompson estimator is defined as

$$B_i^{HT}(I_i = 1) = E_P\left(\hat{\theta}^{HT} \mid I_i = 1\right) - \theta = \sum_{j\in U}\left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right) y_j , \qquad (3.1)$$

where $\pi_{ij} = \Pr(I_i = 1, I_j = 1)$ denotes the second-order inclusion probability of units *i* and *j*. Using $E_P(\hat{\theta}^{HT}) = \theta$, it is not difficult to show that the conditional bias for a nonsample unit is

$$B_i^{HT}(I_i = 0) = E_P\left(\hat{\theta}^{HT} \mid I_i = 0\right) - \theta = -B_i^{HT}(I_i = 1)/(d_i - 1) . \qquad (3.2)$$

See also Moreno-Rebollo et al. (1999) and Moreno-Rebollo et al. (2002). Note that the conditional bias $B_i^{HT}(I_i = 1)$ is equal to 0 when $\pi_i = 1$. In other words, the conditional bias is 0 for any unit selected with certainty in the sample. This is an intuitively-appealing property. Note also that the design variance of the Horvitz-Thompson estimator can be expressed as

$$\text{var}_P\left(\hat{\theta}^{HT}\right) = E_P\left(\hat{\theta}^{HT} - \theta\right)^2 = \sum_{i \in U} B_i^{HT}(I_i = 1)\, y_i \ .$$

Therefore, the design variance of $\hat{\theta}^{HT}$ is identically equal to 0 if and only if $B_i^{HT}(I_i = 1) = 0$ for all $i \in U$. In other words, the design variance is zero when no unit has an influence. For instance, this occurs when the sample size is fixed and $y_i / \pi_i = a,$ for some constant *a.* This also occurs when a census is conducted.

For some sampling designs, the sampling error of the Horvitz-Thompson estimator can be written as

$$\hat{\theta}^{HT} - \theta = \sum\nolimits_{i \in S} B_i^{HT}(I_i = 1) + \sum\nolimits_{i \in U-S} B_i^{HT}(I_i = 0). \qquad (3.3)$$

It can be shown that (3.3) holds if

$$\sum\nolimits_{i \in U} \left(I_i - \pi_i\right) a_i = 0, \qquad (3.4)$$

where $a_i = (1-\pi_i)^{-1}\left\{B_i^{HT}(I_i = 1) - (d_i - 1) y_i\right\}.$ For fixed size sampling designs, the condition (3.4) is satisfied if the coefficient a_i is independent of *i*, noting that $\sum\nolimits_{i \in U}\left(I_i - \pi_i\right) = 0.$ In Section 3.1, we observe that (3.3) holds exactly for Poisson sampling, whereas it holds only approximately for stratified simple random sampling and fixed-size high-entropy sampling with varying first-order inclusion probabilities.

Provided that (3.3) holds, the conditional bias can again be interpreted as the contribution of unit *i* to the sampling error of the Horvitz-Thompson estimator. Using an argument similar to the one that led to (2.7), we obtain the following robust alternative to the Horvitz-Thompson estimator:

$$\hat{\theta}^{RHT} = \left(\hat{\theta}^{HT} - \sum\nolimits_{i \in S} B_i^{HT}(I_i = 1)\right) + \sum\nolimits_{i \in S} \psi\left\{B_i^{HT}(I_i = 1)\right\} \ . \qquad (3.5)$$

The conditional bias $B_i^{HT}(I_i = 1)$ in (3.4) depends in general on unknown population parameters that should be estimated robustly or using an independent source of data. The resulting estimated conditional bias is denoted by $\hat{B}_i^{HT}(I_i = 1)$ and replaces $B_i^{HT}(I_i = 1)$ in (3.5) when it is unknown.

3.1. Examples

Poisson sampling

For Poisson sampling, $\pi_{ij} = \pi_i \pi_j$, $i \neq j,$ and the conditional bias in (3.1) reduces to

$$B_i^{HT}(I_i = 1) = (d_i - 1)\, y_i. \quad (3.6)$$

Unit *i* has a large influence if its design weight, d_i, is large and/or if its *y*-value, y_i, is large. Note that the conditional bias is known for all the sample units; that is, it does not need to be estimated. Also, it follows from (3.6) that (3.4) is satisfied since $a_i = 0$ for all *i* for Poisson sampling. As a result, the decomposition (3.3) holds exactly.

Now, using (3.6) in (3.5), we obtain the following robust version of the Horvitz-Thompson estimator:

$$\begin{aligned}\hat{\theta}^{RHT} &= \sum_{i\in S} y_i + \sum_{i\in S} \psi\left[(d_i - 1) y_i\right] \\ &= \sum_{i\in S} \tilde{w}_i y_i,\end{aligned} \quad (3.7)$$

where

$$\tilde{w}_i = 1 + \frac{\psi\left[(d_i - 1) y_i\right]}{(d_i - 1) y_i}(d_i - 1).$$

Noting that $0 \le \psi(t)/t \le 1,$ it follows that $1 \le \tilde{w}_i \le d_i.$ That is, if the conditional bias for unit *i* is small, its weight $\tilde{w}_i$ will reduce to its original weight, d_i. In other words, the weight of the non-influential units is not modified. Inversely, the weights of influential units (i.e., the units with a large conditional bias) is reduced. Note that the weights $\tilde{w}_i$ cannot be smaller than 1, which is intuitively appealing.

Stratified simple random sampling

For without-replacement stratified simple random sampling, it is straightforward to show that

$$B_i^{HT}(I_i = 1) = \frac{N_h}{(N_h - 1)}\left(\frac{N_h}{n_h} - 1\right)\left(y_{hi} - \bar{Y}_h\right), \quad (3.8)$$

where N_h is the population size in stratum *h*, n_h is the sample size in stratum *h*, $\bar{Y}_h = N_h^{-1}\sum_{i\in U_h} y_{hi}$ and U_h is the set of population units in stratum *h*. From (3.8), an observation in stratum *h* has a large influence when it is far from the stratum mean, $\bar{Y}_h$. As discussed above, the conditional bias (3.8) can be estimated by estimating the unknown population mean in stratum *h*, $\bar{Y}_h$ (e.g., using the median of the sample *y*-values in stratum *h* or using an independent source of data). Ignoring the factor $N_h/(N_h - 1)$ in (3.8) and considering the one-sided Huber function $\psi(z;c) = \min(z,c)$, the resulting robust estimator (3.5) becomes equivalent to the Winsorized estimator of Kokic and Bell (1994). Also, ignoring the factor $N_h/(N_h - 1)$, the condition (3.4) is satisfied, noting that $a_{hi} = -\left(N_h/n_h\right)\bar{Y}_h$, which is independent of *i* within each stratum *h*. As a result, the decomposition (3.3) holds approximately.

High entropy sampling designs

In this section, we consider the family of high entropy sampling designs. In the case of the maximum entropy fixed size sampling design (often called Conditional Poisson sampling), Hajek (1981) proposed the following approximation of π_{ij} :

$$\pi_{ij} \approx \pi_i \pi_j \left[1 - D^{-1}\left(1-\pi_i\right)\left(1-\pi_j\right)\right], \tag{3.9}$$

where $D = \sum_{l \in U} \pi_l \left(1-\pi_l\right)$, assuming that $n \to \infty$ and $N \to \infty$. Berger (1998) showed that this approximation can be used for a larger class of highly randomized and high entropy sampling designs, including the Rao-Sampford design (Rao, 1965; Sampford, 1967) and the Chao procedure (Chao, 1982). Using (3.9) in (3.1), we can approximate the conditional bias attached to unit *i* by

$$B_i^{HT}(I_i = 1) \approx \left(d_i - 1\right)\left\{\left[1 + \frac{\pi_i\left(1-\pi_i\right)}{D}\right]\left(y_i - B\pi_i\right)\right\}, \tag{3.10}$$

where $B = \left[\sum_{j \in U}\left(1-\pi_j\right)\pi_j\right]^{-1} \sum_{j \in U}\left(1-\pi_j\right)y_j$. Under mild regularity conditions, the term $D^{-1}\pi_i\left(1-\pi_i\right)$ in (3.10) is $O\left(N^{-1}\right)$. Assuming that the population size *N* is large, we can neglect this term from (3.10), which leads to

$$B_i^{HT}(I_i = 1) \approx \left(d_i - 1\right)\left(y_i - B\pi_i\right). \tag{3.11}$$

Note that B in (3.11) can be seen as a census regression coefficient obtained by fitting a ratio model with *y* as the dependent variable and the inclusion probability of a unit as the independent variable. From (3.11), it is clear that unit *i* has a large influence if its weight, d_i, is large or if its census residual, $y_i - B\pi_i$, is large. Also, it follows from (3.11) that (3.4) is satisfied since $a_i = -B$, which is independent of *i*. As a result, the decomposition (3.3) holds approximately.

The previous examples show that the conditional bias (3.1) is a measure of influence that fully account for the sampling design. A given observation may be highly influential under a given sampling design and have no influence under another sampling design. To illustrate this point, consider a population consisting of *N* population values $y_1,...,y_N$, such that $y_1 = 0$, $y_2 = ... = y_{N-1} = 500$ and $y_N = 1000$ so that the population mean is equal to 500. Under simple random sampling without replacement, both y_1 and y_N are influential because they are far from the population mean (see expression (3.8)), whereas y_1 has no influence under Poisson sampling (see expression (3.6)).

3.2. Choice of the tuning constant

The ψ-function (e.g., the Huber function) in (3.5) usually depends on a tuning constant *c*. A suitable value for *c* is sometimes determined by minimizing an estimator of the mean square error of the robust estimator (e.g., Hulliger, 1995; Kokic and Bell, 1994; and Beaumont and Rivest, 2009). We propose an alternative approach. It consists of finding the value of *c* that minimizes

$$\max\left\{ \left|\hat{B}_i^{RHT}(I_i=1;c)\right| ; i \in S \right\} , \tag{3.12}$$

where $\hat{B}_i^{RHT}(I_i=1;c)$ is an estimator of the conditional bias of the robust Horvitz-Thompson $\hat{\theta}^{RHT}$. From (3.5), this conditional bias can be written as

$$\begin{aligned} B_i^{RHT}(I_i=1;c) &= E_P\left(\hat{\theta}^{RHT} \mid I_i=1\right)-\theta \\ &= B_i^{HT}(I_i=1)+E_P\left(\sum_{i\in S}\left[\psi\left\{\hat{B}_i^{HT}(I_i=1);c\right\}-\hat{B}_i^{HT}(I_i=1)\right]\middle| I_i=1\right) \end{aligned} \tag{3.13}$$

As pointed out above, the conditional bias $B_i^{HT}(I_i=1)$, if unknown, can be estimated by $\hat{B}_i^{HT}(I_i=1)$. A conditionally unbiased estimator of the last conditional expectation in (3.13) is simply $\sum_{i\in S}\left[\psi\left\{\hat{B}_i^{HT}(I_i=1);c\right\}-\hat{B}_i^{HT}(I_i=1)\right]$. This yields the estimator

$$\hat{B}_i^{RHT}(I_i=1;c)=\hat{B}_i^{HT}(I_i=1)+n\bar{\Delta}(c) , \tag{3.14}$$

where

$$\bar{\Delta}(c)=\frac{1}{n}\sum_{i\in S}\left[\psi\left\{\hat{B}_i^{HT}(I_i=1);c\right\}-\hat{B}_i^{HT}(I_i=1)\right] . \tag{3.15}$$

Alternatively, one can replace $\bar{\Delta}(c)$ in (3.14) by $\bar{\Delta}^R(c)$, some robust alternative to the average in (3.14). For instance, the once-Winsorized mean (e.g., Rivest, 1994) could be used. Note that the median may not be suitable as it leads to a value of *c* that may require downweighting more than half of the sample units; otherwise, $\bar{\Delta}^R(c)=0$ and $\hat{B}_i^{RHT}(I_i=1;c)=\hat{B}_i^{HT}(I_i=1)$ for all sample units.

To minimize (3.12) using (3.14), one can first try several values of *c*. We suggest $c=0$ and $c=\left|\hat{B}_i^{HT}(I_i=1)\right|$, for $i\in S$. Then, (3.12) can be minimized among those values of *c*. For a given value of *c*, note that the maximum in (3.12) is either $\left|\max\left(\hat{B}_i^{HT}(I_i=1) ; i\in S\right)+n\bar{\Delta}(c)\right|$ or $\left|\min\left(\hat{B}_i^{HT}(I_i=1) ; i\in S\right)+n\bar{\Delta}(c)\right|$. The global minimum of (3.12) can be found by noting that $\bar{\Delta}(c)$ is piecewise linear in *c*.

3.3. Robust version of the generalized regression estimator

In this section, we consider the case of the generalized regression estimator (GREG) of $\theta = \sum_{i \in U} y_i$. We assume that a vector of auxiliary variables $\mathbf{x}_i$ is available for all $i \in S$ and that the population total of the $\mathbf{x}$-vector, $\sum_{i \in U} \mathbf{x}_i$, is known. The GREG estimator of θ is given by

$$\hat{\theta}^G = \sum_{i \in S} d_i y_i + \left(\sum_{i \in S} d_i \mathbf{x}_i - \sum_{i \in U} \mathbf{x}_i \right)' \hat{\boldsymbol{\gamma}},$$

where $\hat{\boldsymbol{\gamma}} = \left(\sum_{j \in S} d_i v_i^{-1} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{j \in S} d_i v_i^{-1} \mathbf{x}_i y_i$ and v_i is a known constant attached to unit *i*.

Using a first-order Taylor expansion and neglecting the higher order terms, we can write

$$\hat{\theta}^G - \theta \approx \sum_{i \in S} d_i E_i,$$

where $E_i = \left(y_i - \mathbf{x}_i' \boldsymbol{\gamma} \right)$ denotes the census residual attached to unit *i* with

$$\boldsymbol{\gamma} = \left(\sum_{j \in U} v_i^{-1} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{j \in U} v_i^{-1} \mathbf{x}_i y_i.$$

The asymptotic conditional bias attached to unit *i* with respect to the GREG estimator is given by

$$B_i^G(I_i = 1) = E_P \left(\hat{\theta}^G \mid I_i = 1 \right) - \theta \approx \sum_{j \in U} \frac{\pi_{ij}}{\pi_i \pi_j} E_j.$$

If $v_i = \boldsymbol{\lambda}' \mathbf{x}_i$ for a vector of known constant $\boldsymbol{\lambda}$, we have $\sum_{j \in U} E_j = 0$ and the conditional bias $B_i^G(I_i = 1)$ can be approximated by (3.1) with y_i replaced by E_i. In this case, for Poisson sampling and stratified simple random sampling, the asymptotic conditional bias attached to unit *i* is obtained from (3.6) and (3.8), respectively, by replacing y_i with E_i. For these two sampling designs, a unit has a large influence if its design weight is large and/or its census residual is large.

Following the approach described above, we define a robust version of $\hat{\theta}^G$ as

$$\hat{\theta}^{RG} = \hat{\theta}^G - \sum_{i \in S} B_i^G(I_i = 1) + \sum_{i \in S} \psi \left\{ B_i^G(I_i = 1) \right\}. \qquad (3.16)$$

Once again, the conditional bias $B_i^G(I_i = 1)$ in (3.16) depends in general on unknown population parameters that should be estimated robustly or using an independent source of data. The resulting estimated conditional bias is denoted by $\hat{B}_i^G(I_i = 1)$ and replaces $B_i^G(I_i = 1)$ in (3.16).

4. Empirical study

We conducted a limited simulation to investigate the performance of the proposed robust estimator in terms of relative bias and relative efficiency. We generated three populations of size $N = 500$, each consisting of an auxiliary variable *x* and a variable of interest *y*. First, the *x*-values were generated from a Gamma distribution with mean 100 and standard deviation 50. Then, for the non-outlier portion of the populations, the *y*-values were generated according to the ratio model

$$y_i = 2x_i + \varepsilon_i,$$

where the error terms ε_i were generated from a normal distribution with mean 0 and variance σ^2, whose value was set to lead to a coefficient of determination (R^2) approximately equal to 0.8. Note that the outliers were manually added in the population.

From each population, we selected *R* = 10, 000 samples according to Poisson sampling with inclusion probabilities, π_i, proportional to x_i; that is, $\pi_i = nx_i / \sum_{i \in U} x_i$. The expected sample size was set to 10, 25, 50 and 100.

Figures 1a)-1c) show the relationship between the variables *x* and *y*. Note that Population 1 contained no influential value, whereas Population 2 and Population 3 contained 4 and 10 influential values, respectively. For the latter populations, the observations in the upper left side of the plots (Figures 1b) and 1c)) are highly influential under Poisson sampling (see expression (3.6)) because they exhibit both a large design weight and a large *y*-value.

In each sample, we computed three estimators: (i) the Horvitz-Thompson estimator given by $\hat{\theta}^{HT} = \sum_{i \in S} d_i y_i$. (ii) The robust estimator (3.4), where the tuning constant *c* was chosen so that (3.11) was minimized among the set of values $c = 0$ and $c = \left|B_i^{HT}(I_i = 1)\right| = \left|(d_i - 1)y_i\right|$ for $i \in S$. We denoted the resulting estimator by $\hat{\theta}_{cb}^{RHT}$. (iii) The robust estimator (3.5), where the tuning constant *c* was chosen so that its estimated mean square error was minimized; e.g., Hulliger (1995) and Beaumont and Alavi (2004). We denoted the resulting estimator by $\hat{\theta}_{mse}^{RHT}$. Note that the estimated mean square error of $\hat{\theta}^{RHT}$ in (3.5) under Poisson sampling is given by

$$mse\left(\hat{\theta}^{RHT}\right) = \sum_{i \in S} (1 - \pi_i)\left[y_i + \psi\left\{B_i^{HT}(I_i = 1)\right\}\right]^2$$
$$+ \max\left(0, \left(\hat{\theta}^{RHT} - \hat{\theta}^{HT}\right)^2 - \sum_{i \in S} (1 - \pi_i)\left[\psi\left\{B_i^{HT}(I_i = 1)\right\} - B_i^{HT}(I_i = 1)\right]^2\right);$$

e.g., Gwet and Rivest (1992).

For comparisons of estimators, we computed the Monte Carlo percent Relative Bias (RB) given by

$$RB_{MC}\left(\hat{\theta}\right)=100\times\frac{E_{MC}\left(\hat{\theta}\right)-\theta}{\theta},$$

where $E_{MC}\left(\hat{\theta}\right)=R^{-1}\sum_{r=1}^{R}\hat{\theta}^{(r)}$ with $\hat{\theta}^{(r)}$ denoting the estimator $\hat{\theta}$ (either $\hat{\theta}^{HT}$, $\hat{\theta}_{cb}^{RHT}$ or $\hat{\theta}_{mse}^{RHT}$), in the *r*-th simulated sample, $r=1,...,R.$ We also computed the Monte Relative Efficiency (RE), using the Horvitz-Thompson estimator, $\hat{\theta}^{HT}$, as the reference:

$$RE_{MC}\left(\hat{\theta}\right)=100\times\frac{MSE_{MC}\left(\hat{\theta}\right)}{MSE_{MC}\left(\hat{\theta}^{HT}\right)},$$

where

$$MSE_{MC}\left(\hat{\theta}\right)=R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)}-\theta\right)^{2}.$$

Table 1 shows the values of RB and RE for the three estimators described above. Finally, in each sample, we computed the percent Absolute Relative Error (ARE) given by

$$ARE\left(\hat{\theta}^{(r)}\right)=100\times\left|\frac{\hat{\theta}^{(r)}-\theta}{\theta}\right|.$$

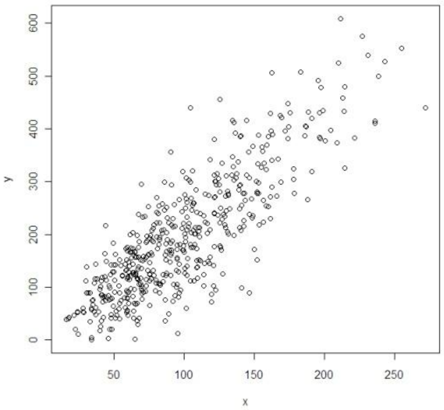
We obtained the 90-th, 95-th and 99-th percentiles of the ARE values. The results are shown in Table 2.

From Table 1, we note that the Horvitz-Thompson estimator, $\hat{\theta}^{HT}$, showed a negligible bias in all the scenarios, as expected. The robust estimators $\hat{\theta}_{cb}^{RHT}$ and $\hat{\theta}_{mse}^{RHT}$ were slightly or moderately biased, as expected. We note that the RB decreased as the expected sample size increased. For Population 1 (that did not contain any influential observation), we note that both robust estimators were slightly less efficient than the Horvitz-Thompson estimator with a value of RE ranging from 100 to 108. For Population 2 (that contained 4 influential units), both $\hat{\theta}_{cb}^{RHT}$ and $\hat{\theta}_{mse}^{RHT}$ were significantly more efficient than the Horvitz-Thompson estimator. For example, with an expected sample size equal to 10, the value of RE corresponding to $\hat{\theta}_{cb}^{RHT}$ was equal to 67%, whereas it was equal to 69% for $\hat{\theta}_{mse}^{RHT}$. Also, we note that the value of RE approached 100% as the expected sample size increased. This seems to suggest that our robust estimators are consistent. However, the formal proof of consistency is beyond the scope of this paper. Observations similar to those about Population 2 can be made for Population 3. Comparing $\hat{\theta}_{cb}^{RHT}$ and $\hat{\theta}_{mse}^{RHT}$, we note that they performed similarly in terms of both RB and RE with a slight advantage for $\hat{\theta}_{cb}^{RHT}$. These results suggest that minimizing (3.12) can be viewed as a good alternative to minimizing the estimated mean square error.

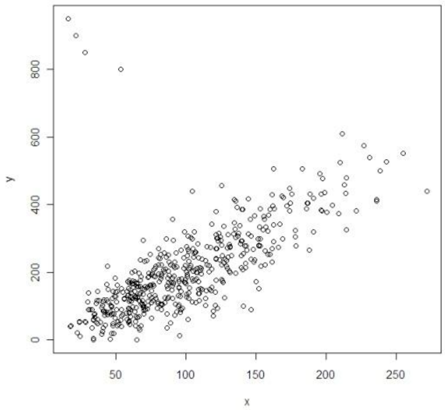
Table 2 shows that both robust estimators performed similarly in terms of the percentiles of the AREs and much better than the Horvitz-Thompson estimator for Population 2 and Population 3. For Population 1, which contains no influential observation, all three estimators performed similarly. The benefits of using a robust estimator over the Horvitz-Thompson estimator are much more apparent when the sample size is smaller.

Figure 1: Relationship between *x* and *y*

(a) Population 1



(b) Population 2



(c) Population 3

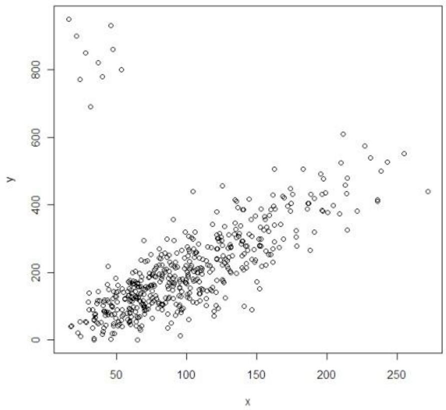


Table 1: Monte Carlo percent Relative Bias and Relative Efficiency (in parentheses) of three estimators

N	$\hat{\theta}^{HT}$	$\hat{\theta}_{cb}^{RHT}$	$\hat{\theta}_{mse}^{RHT}$
Population 1			
10	-0.2 (100)	-9.8 (106)	-12.0 (108)
25	-0.2 (100)	-3.7 (102)	-5.1 (104)
50	0.1 (100)	-1.8 (101)	-2.6 (102)
100	0.0 (100)	-0.9 (100)	-1.3 (101)
Population 2			
10	-0.1 (100)	-12.3 (67)	-14.3 (69)
25	-0.1 (100)	-6.6 (69)	-7.8 (71)
50	0.0 (100)	-4.4 (69)	-5.1 (72)
100	0.0 (100)	-3.3 (72)	-3.6 (74)
Population 3			
10	0.2 (100)	-14.9 (59)	-16.9 (61)
25	0.5 (100)	-8.7 (65)	-10.0 (65)
50	-0.1 (100)	-6.5 (75)	-7.5 (74)
100	0.1 (100)	-4.4 (85)	-5.4 (83)

Table 2: Percentiles for the Absolute Relative Error of three estimators

	$\hat{\theta}^{HT}$			$\hat{\theta}_{cb}^{RHT}$			$\hat{\theta}_{mse}^{RHT}$		
	90%	95%	99%	90%	95%	99%	90%	95%	99%
n	Population 1								
10	52	62	84	55	64	80	55	64	80
25	33	40	53	34	40	52	35	41	52
50	23	27	36	23	27	36	23	27	36
100	15	18	23	15	18	23	15	18	23
	Population 2								
10	55	67	163	55	64	79	56	65	80
25	37	46	95	35	40	53	35	41	53
50	26	33	58	24	28	36	24	28	37
100	19	24	36	16	19	26	17	20	26
	Population 3								
10	61	83	170	56	64	81	57	66	81
25	42	55	104	36	44	60	37	44	57
50	30	37	60	27	32	42	27	32	41
100	31	26	38	20	23	31	19	23	30

5. A method frequently used in practice

A popular method used in practice is to reduce the weight of units that have been identified as influential. Most often, the weight of these units is set equal to one, while the outstanding weight is redistributed among the other units. This method would only be appropriate if the unit identified as influential is unique in the population. In this section, we show that this method can be expressed in terms of the conditional bias. Suppose for now that a single unit was identified as influential, say *k*.

Let $\hat{B}_i^{HT}(I_i = 1)$ be a conditionally design-unbiased estimator of the conditional bias $B_i^{HT}(I_i = 1)$ in (3.1). We have

$$\hat{B}_i^{HT}(I_i = 1) = (d_i - 1)\, y_i + \sum_{\substack{j \in S \\ j \neq i}} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_j \pi_{ij}} \right) y_j, \qquad (5.1)$$

provided $\pi_{ij} > 0$ for all $j \in U.$ That is, $E_p\left(\hat{B}_i^{HT}(I_i = 1) - B_i^{HT}(I_i = 1) \mid I_i = 1 \right) = 0.$ Furthermore, consider the following ψ-function:

$$\psi(z;c) = \begin{cases} 0 & \text{if } z \geq c \\ z & \text{if } -c \leq z \leq c \\ 0 & \text{if } z \leq -c \end{cases} \qquad (5.2)$$

The ψ-function in (5.2) can be viewed as the extreme case of a redescending ψ-function. Then, replacing $B_i^{HT}(I_i = 1)$ in (3.5) with $\hat{B}_i^{HT}(I_i = 1)$ given by (5.1) and using (5.2) leads to

$$\begin{aligned} \hat{\theta}_*^{RHT} &= \hat{\theta}^{HT} - \hat{B}_k^{HT}(I_k = 1) \\ &= y_k + \sum_{\substack{j \in S \\ j \neq k}} \frac{\pi_k}{\pi_{kj}} y_j \end{aligned} \qquad (5.3)$$

noting that $\psi\left\{ \hat{B}_i^{HT}(I_i = 1) \right\} = \hat{B}_i^{HT}(I_i = 1),$ except for $i = k.$ From (5.3), we note that unit *k* has now a weight equal to 1, whereas the weight of unit *j* is equal to π_k / π_{kj} for $j \neq k.$ The weight π_k / π_{kj} represents the inverse of the inclusion probability of unit *j* under the conditional sampling design $P(S \mid I_k = 1).$ For example, in the case of simple random sampling without replacement, we have $\pi_k / \pi_{kj} = (N - 1)/(n - 1)$ for $j \neq k,$ and the sum of weights remains equal to *N*. This situation was discussed in Rao (1971).

Now, suppose that $m < n$ units were identified as influential, say the first *m* units, $1, ..., m.$ In practice, the weight of these units is set equal to one and the outstanding weight is redistributed among the remaining units. We define a measure of the joint influence of units $1, ..., m,$ on $\hat{\theta}^{HT}$ as

$$B_{1\ldots m}^{HT}(I_1=1,\ldots,I_m=1)=E_p\left(\hat{\theta}^{HT}\left|I_1=1,\ldots,I_m=1\right.\right)-\theta$$

$$=\sum_{i\in S_m}(d_i-1)y_i+\sum_{\substack{j\in U\\ j\neq 1,\ldots,m}}\left(\frac{\pi_{j1\ldots m}}{\pi_{1\ldots m}\pi_j}-1\right)y_j, \qquad (5.4)$$

where S_m denotes the subset consisting of the *m* influential units, $\pi_{1\ldots m}=P\left(I_1=1,\ldots,I_m=1\right)$ denotes the joint inclusion probability of unit 1,…,*m* and $\pi_{j1\ldots m}$ is defined similarly. Once again, a conditionally *p*-unbiased estimator of $B_{1\ldots m}^{HT}(I_1=1,\ldots,I_m=1)$ in (5.4) is given by

$$\hat{B}_{1\ldots m}^{HT}(I_1=1,\ldots,I_m=1)=\sum_{k\in S_m}(d_k-1)y_k+\sum_{\substack{j\in S\\ j\neq 1,\ldots,m}}\left(\frac{\pi_{j1\ldots m}-\pi_{1\ldots m}\pi_j}{\pi_{j1\ldots m}\pi_j}\right)y_j. \qquad (5.5)$$

We consider a robust estimator of the form

$$\hat{\theta}_*^{RHT}=\hat{\theta}^{HT}-\hat{B}_{1\ldots m}^{HT}(I_1=1,\ldots,I_m=1)$$

$$=\sum_{k\in S_m}y_k+\sum_{j\in S-S_m}\frac{\pi_{1\ldots m}}{\pi_{j1\ldots m}}y_j, \qquad (5.6)$$

where $\hat{B}_{1\ldots m}^{HT}(I_1=1,\ldots,I_m=1)$ is given by (5.5). From (5.6), we note that the weight of each of the *m* influential units is now equal to 1, whereas the weight of the remaining units is equal to $\frac{\pi_{1\ldots m}}{\pi_{j1\ldots m}}$, $j\neq 1,\ldots,m$. In the case of simple random sampling without replacement, the latter weight reduces to $(N-m)/(n-m)$. When $m\geq 2$, it is not possible to express (5.6) as a special case of (3.5) (as in the case of single influential unit). An estimator similar to (3.5) could be developed but the latter would require higher order inclusion probabilities, which is not practical.

The rationale underlying (5.3) or (5.6) is to reduce the impact of units that have been identified as influential to the extent that they only represent themselves in the final estimate. This approach is appropriate only if these units are unique, which is not generally the case. As a result, this type of estimators tends to exhibit large biases.

6. Conclusion

We have shown that the conditional bias can be useful for deriving robust estimators in finite population sampling. Implementation of these estimators can be done by modifying either the *y*-values or the design weights of sample units. Estimating their variance is a topic that requires more investigation. Replication methods seem to be attractive in this context. Finally, the conditional bias could also be used as a diagnostic tool for choosing among alternative sampling designs those that offer some degree of protection against the potential occurrence of influential sample units (Bocci and Beaumont, 2006).

Bibliographie

Beaumont, J.-F. and Alavi, A. (2004). Robust generalized regression estimation. *Survey Methodology*, 30, 195-208.

Beaumont, J.-F. and Rivest, L.-P. (2009). Dealing with outliers with survey data. *Handbook of Statistics, Volume 29, Chapter 11, Sample Surveys: Theory Methods and Inference, Editors: C.R. Rao and D. Pfeffermann,* 247-279.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279-292.

Bocci, C. and Beaumont, J.-F. (2006). Dealing with the problem of combined reports at the sampling design stage for the Workplace and Employee Survey. *Proceedings of the Survey Methods Section*, Statistical Society of Canada.

Berger, Y.G. (1998). Rate of convergence for asymptotic variance for the Horvitz-Thompson estimator. *Journal of Statistical Planning and Inference,* 74, 149-168.

Campbell, C. (1980). A different view of finite population estimation. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 319-324.

Chao, M.T. (1982). A general purpose unequal probability sampling plan. *Biometrika*, 69, 653-656.

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

Chambers, R.L., Kokic, P., Smith, P. and Cruddas, M. (2000). Winsorization for identifying and treating outliers in business surveys. *Proceedings of the Second International Conference on Establishment Surveys*, American Statistical Association, Alexandria, Virginia, 717-726.

Demnati, A. and Rao, J.N.K (2004). Linearization variance estimators for survey data. *Survey Methodology*, 30, 17-26.

Deville, J.-C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203.

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

Hájek, J. (1981). *Sampling from a finite population*. Marcel Dekker. New York.

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: the Approach Based on Influence Functions*. John Wiley and Sons Inc., New-York.

Hulliger, B. (1995). Outlier Robust Horvitz-Thompson estimators. *Survey Methodology.* 21, 79-87.

Kokic, P.N. and Bell, P.A. (1994). Optimal winzorizing cutoffs for a stratified finite population estimator. *Journal of Official Statistics*, 10, 419-435.

Moreno-Rebollo, J.L., Muñoz-Reyez, A.M. and Muñoz-Pichardo, J.M. (1999). Influence diagnostics in survey sampling: conditional bias. *Biometrika*, 86, 923-968.

Moreno-Rebollo, J.L., Muñoz-Reyez, A.M., Jimenez-Gamero, M.D. and Muñoz-Pichardo, J. (2002). Influence diagnostics in survey sampling: estimating the conditional bias. *Metrika*, 55, 209-214.

Muñoz-Pichardo, J., Muñoz-Garcia, J., Moreno-Rebollo, J.L. and Piño-Mejias, R. (1995). A new approach to influence analysis in linear models. *Sankhya*, Series A, 57, 393-409.

Rao, C.R. (1971). Some aspects of statistical inference of sampling from finite populations. In *Foundations of Statistical Inference*, eds V.P. Godambe and D.A. Sprott. Toronto: Holt, Rinehart & Winston.

Rao, J.N.K. (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association*, 3, 173-180.

Rivest, L.-P. (1994). Statistical properties of winsorized means for skewed distributions. *Biometrika*, 81, 373-383.

Royall, R.M. (1976). The linear least-square prediction approach to two-stage sampling. *Journal of The American Statistical Association*, 71, 657-664.

Sampford, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, 54, 499-513.

Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference: a Prediction Approach*. John Wiley and Sons Inc., New-York.

Zaslavsky, A.M., Schenker, N. and Belin, T.R. (2001). Downweighting influential clusters in surveys: application to the 1990 post enumeration survey. *Journal of the American Statistical Association*, 96, 858-869.