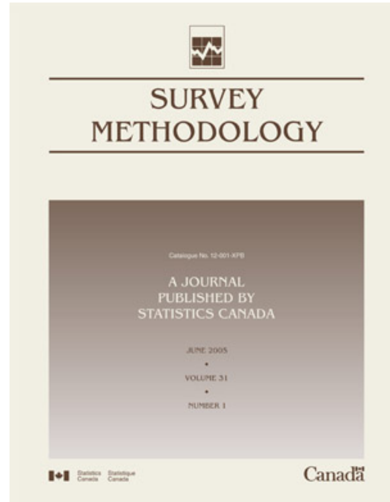
Catalogue no. 12-001-XIE

Survey Methodology

December 2004



Statistics Canada
Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada

Business Survey Methods Division

Survey Methodology

December 2004

Published by authority of the Minister responsible for Statistics Canada

April 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 30, No. 2, pp. 195-208
Statistics Canada, Catalogue No. 12-001

Robust Generalized Regression Estimation

Jean-François Beaumont and Asma Alavi [1]

Abstract

The Best Linear Unbiased (BLU) estimator (or predictor) of a population total is based on the following two assumptions: i) the estimation model underlying the BLU estimator is correctly specified and ii) the sampling design is ignorable with respect to the estimation model. In this context, an estimator is robust if it stays close to the BLU estimator when both assumptions hold and if it keeps good properties when one or both assumptions are not fully satisfied. Robustness with respect to deviations from assumption (i) is called model robustness while robustness with respect to deviations from assumption (ii) is called design robustness. The Generalized Regression (GREG) estimator is often viewed as being robust since its property of being Asymptotically Design Unbiased (ADU) is not dependent on assumptions (i) and (ii). However, if both assumptions hold, the GREG estimator may be far less efficient than the BLU estimator and, in that sense, it is not robust. The relative inefficiency of the GREG estimator as compared to the BLU estimator is caused by widely dispersed design weights. To obtain a design-robust estimator, we thus propose a compromise between the GREG and the BLU estimators. This compromise also provides some protection against deviations from assumption (i). However, it does not offer any protection against outliers, which can be viewed as a consequence of a model misspecification. To deal with outliers, we use the weighted generalized *M*-estimation technique to reduce the influence of units with large weighted population residuals. We propose two practical ways of implementing *M*-estimators for multipurpose surveys; either the weights of influential units are modified and a calibration approach is used to obtain a single set of robust estimation weights or the values of influential units are modified. Some properties of the proposed approach are evaluated in a simulation study using a skewed finite population created from real survey data.

Key Words: Design robustness; Model robustness; *M*-estimator; Outliers; Shrunk weights; Best linear unbiased predictor.

1. Introduction

In classical theory, sample data can be viewed as being randomly drawn from an infinite population and assumptions are made about the unknown distribution of the infinite population. In other words, a model is postulated and the interest lies in the estimation of model parameters. In this context, an estimator $\hat{\theta}$ of a model parameter θ is robust if it stays close to the maximum likelihood estimator of θ when the model assumptions hold and if it keeps good properties when the model assumptions are not fully satisfied. The unknown distribution of the infinite population is often assumed to be the normal distribution and, as a result, the maximum likelihood estimator reduces to the usual least-squares estimator.

The presence of outliers in the sample can be viewed as a consequence of a deviation from a model assumption. The majority of the sample could be assumed to come from the selected model but some units, called outliers, could be thought of as coming from a different model. Therefore, the presence of such outliers in the sample may introduce bias and increase the variance of the least-squares estimator of the selected model parameters. Outliers could also be the consequence of a highly skewed distribution. In this case, the least-squares estimator is not biased but may be highly inefficient due to a deviation from the usual normality assumption. The presence of outliers in the sample could also be the result of measurement errors. However, it is assumed in the rest of this paper that the data have been verified and corrected, if necessary, and that there is no measurement error left in the data. Outlier-robust estimation for infinite populations has been studied extensively (for a review, see Huber 1981; or Hampel, Ronchetti, Rousseeuw and Stahel 1986).

In survey sampling theory, the interest usually lies in the estimation of finite population parameters such as the total, $t_y = \sum_{k \in U} y_k$, of a variable of interest y for a finite population U of size N. Because it is usually not possible to observe the variable y for all population units, the usual practice consists of selecting from the finite population a random sample s of size n according to some probability sampling design $p(s \mid \mathbf{Z})$. The matrix of design information $\mathbf{Z}$ contains N rows with its k^{th} row equal to $\mathbf{z}'_\mathbf{k}$, and $\mathbf{z}$ is a vector of auxiliary variables available at the design stage. This does not preclude the finite population itself to be assumed to come from a model, as it is explicitly the case when it is chosen to make model-based inferences. Under this type of inference, Royall (1976) derived the Best Linear Unbiased (BLU) estimator (or predictor) $\hat{t}_y^B$ of t_y (see also Valliant, Dorfman and Royall 2000, Chapter 2). It is based

1. Jean-François Beaumont and Asma Alavi, Household Survey Methods Division, Statistics Canada, 16th floor, R.H. Coats Building, Ottawa, Ontario, Canada, K1A 0T6. E-mail: Jean-Francois.Beaumont@statcan.ca and Asma.Alavi@statcan.ca.

on the following two assumptions: i) the estimation model underlying the BLU estimator $\hat{t}_y^B$ is correctly specified and ii) the sampling design is ignorable with respect to the estimation model. In this context, an estimator $\hat{t}_y$ of the finite population total t_y is robust if it stays close to the BLU estimator $\hat{t}_y^B$ when both assumptions hold and if it keeps good properties when one or both assumptions are not fully satisfied. Robustness with respect to deviations from assumption (i) is called model robustness while robustness with respect to deviations from assumption (ii) is called design robustness.

Although we consider robust estimators that are constructed from a model-based viewpoint, we prefer evaluating their properties as much as possible with respect to the sampling design. This allows us to choose the constants on which robust estimators depend and to evaluate their quality without having to rely on a model and, more specifically, without having to rely on a model for the outliers. This also provides an objective framework for comparing estimators derived under different models. This preference of evaluating properties of model-based estimators with respect to the sampling design is also shared by Little (1983) who notes that design-based asymptotics may be more useful for assessing estimators than model-based asymptotics, particularly when the data set is large.

The Generalized Regression (GREG) estimator of t_y is often viewed as being robust since its property of being Asymptotically Design Unbiased (ADU) is not dependent on assumptions (i) and (ii); that is, the GREG estimator is bias-robust even though its form can be justified by an estimation model. However, if both assumptions hold, the GREG estimator may be far less efficient than the BLU estimator and, in that sense, it is not robust. The relative inefficiency of the GREG estimator as compared to the BLU estimator is caused by widely dispersed design weights. The fact that variable design weights may increase the variance of an estimator is well known (see, for example, Rao 1966; DuMouchel and Duncan 1983; Kish 1992; Pfeffermann 1993; Korn and Graubard 1999, Chapter 4; Elliott and Little 2000; and Kalton and Flores-Cervantes 2003) and is not uncommon in household surveys due to the presence of many weight adjustments before calibration (Kish 1992; and Kalton and Flores-Cervantes 2003). This problem is often treated by truncating the larger design weights (Potter 1988, 1990, 1993; and Stokes 1990).

To obtain a design-robust estimator when the design weights are highly variable, we propose a compromise between the GREG and the BLU estimators based on the weighted Least-Squares (LS) technique. This compromise estimator has a smaller design bias than the BLU estimator when the ignorability assumption is not satisfied and, at the same time, is more efficient than the GREG estimator when this assumption holds. It also provides some protection against deviations from model assumptions. Balanced sampling (Royall and Herson 1973) and nonparametric calibration (Chambers, Dorfman and Wehrly 1993) are other methods that provide protection against certain types of model misspecifications (see also Valliant, Dorfman and Royall 2000, Chapter 3, 4 and 11). However, none of these methods offer any protection against outliers, which can be viewed as a consequence of a model misspecification. In a model-based framework, the idea underlying the *M*-estimation technique has been proposed to develop outlier-robust alternatives to the BLU estimator (Chambers 1986; Lee 1991; and Welsh and Ronchetti 1998). In a design-based framework, the *M*-estimation technique has also been used to develop outlier-robust alternatives to the GREG estimator (Gwet and Rivest 1992; Hulliger 1995 1999; Duchesne 1999; and Zaslavsky, Schenker and Belin 2001). *M*-estimation is also discussed in the review paper by Lee (1995) and an empirical comparison of several outlier-robust estimators can be found in Gwet and Lee (2000).

Finite population parameters are often very sensitive to the presence of outliers in the population. This is to be contrasted to model (infinite population) parameters, which are usually insensitive to outliers. The problem of outlier robustness is therefore different for finite and infinite populations. As noted in Chambers (1986), it is the sampling error (or the prediction error in a model-based framework) of an estimator which must be insensitive to outliers in finite populations and not necessarily the estimator itself. For instance, when a simple random sampling design is used, the sample median is robust in the classical sense. As a result, its design variance is essentially unaffected by the presence of an outlier in the finite population, no matter how large is that outlier. However, the sampling error and the design bias of the sample median, when used as an estimator of the finite population mean, take an arbitrarily large value when one or more population unit takes an arbitrarily large value. This is explained by the fact that the finite population mean itself takes an arbitrarily large value in such a case. Unlike the sample median, the sample mean is design unbiased but it is not robust in the classical sense. The sampling error and the design variance of the sample mean can thus be very affected by the presence of an outlier in the finite population. This illustrates why outlier-robustness for finite populations is often viewed as a trade-off between bias and variance and why outliers must usually have an influence, at least to some extent, on estimators. The Mean Squared Error (MSE) is therefore a useful criterion for evaluating the quality of outlier-robust estimators of finite population parameters.

The real goal of this paper is to find a robust alternative to the commonly-used GREG estimator of t_y. However, it is more natural to discuss robustness issues by first introducing the optimal (BLU) estimator. Therefore, the assumptions underlying the BLU estimator are discussed in section 2. We also give additional conditions under which the BLU estimator has a negligible asymptotic design bias. Section 3 deals with design robustness and the weighted LS estimator is introduced. In section 4, model robustness (more specifically, outlier robustness) is discussed and the weighted generalized *M*-estimation technique is suggested to reduce the influence of units with large weighted population residuals. The proposed estimator is census-consistent in the sense that it is equal to the finite population total t_y when a census is conducted. We propose two practical ways of implementing *M*-estimators for multipurpose surveys; either the weights of influential units are modified and a calibration approach is used to obtain a single set of robust estimation weights or the values of influential units are modified. Mean Squared Error (MSE) estimation is discussed in section 5. In section 6, some properties of the proposed approach are evaluated in a simulation study using a skewed finite population created from real survey data. Finally, some concluding remarks are made in the last section.

2. The Best Linear Unbiased Estimator

Let us assume that we have a vector of auxiliary variables $\mathbf{x}$ available for all units of the sample s and for which population totals, $\mathbf{t_x} = \sum_{k\in U} \mathbf{x_k}$, are known. Let us also denote by $\mathbf{X}$, the matrix containing N rows with its k^{th} row equal to $\mathbf{x'_k}$. The vector $\mathbf{x}$ may or may not contain some variables in the vector $\mathbf{z}$ of design variables. Before discussing robustness, we first describe the two assumptions (see A1 and A2 below) with respect to which robustness is desired. Then, we briefly explain how to validate them.

A1) The following estimation model *m* holds: y_k given $\mathbf{X}$, for $k \in U$, are independently distributed with mean $\mathrm{E}_m(y_k \mid \mathbf{X}) = \mathbf{x'_k}\boldsymbol{\beta}$ and variance $\mathrm{V}_m(y_k \mid \mathbf{X}) = \sigma^2 v_k$, where $\boldsymbol{\beta}$ and σ^2 are unknown model parameters, $v_k = \mathbf{x'_k}\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}$ is a vector of known constants. The subscript "*m*" indicates that expectations and variances are evaluated with respect to model *m*.

A2) The sampling design is independent of $\mathbf{y}$ after conditioning on $\mathbf{X}$; that is, $p(s \mid \mathbf{y}, \mathbf{X}) = p(s \mid \mathbf{X})$, where $\mathbf{y}$ is a vector containing N elements with its k^{th} element equal to y_k.

Assumption (A1) describes the estimation model *m*, which specifies the distribution of $\mathbf{y}$ conditional on $\mathbf{X}$. Standard techniques can be used to validate this model (see, for example, Draper and Smith 1980, Chapter 3). The linearity assumption $\mathrm{E}_m(y_k \mid \mathbf{X}) = \mathbf{x'_k}\boldsymbol{\beta}$ is an important assumption underlying the estimation model *m*. There are many ways of assessing the validity of this assumption. A graph of residuals $e_k = y_k - \mathbf{x'_k}\hat{\boldsymbol{\beta}}$ versus $\mathbf{x'_k}\hat{\boldsymbol{\beta}}$, for some *m*-unbiased estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, is often suggested for this purpose. Any trend in this graph is an indication that the relationship between $\mathbf{y}$ and $\mathbf{x}$ is not linear. To obtain robustness against a deviation from the linearity assumption, a poststratification model can be used when it is possible to partition the population into homogeneous and mutually exclusive groups. An example of the importance of careful modeling in sample surveys can be found in Hedlin, Falvey, Chambers and Kokic (2001).

Assumption (A2) is a sufficient condition for the ignorability (Rubin 1976) of the sampling design with respect to the distribution of $\mathbf{y}$ conditional on $\mathbf{X}$. In other words, it means that the distribution of $\mathbf{y}$ is independent of s after conditioning on $\mathbf{X}$. Using assumption (A1), $\mathbf{y}$ can be split into a fixed term $\mathbf{X}\boldsymbol{\beta}$ and a random error term $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. Consequently, if the sampling design is independent of $\boldsymbol{\varepsilon}$ after conditioning on $\mathbf{X}$; that is, if $p(s \mid \boldsymbol{\varepsilon}, \mathbf{X}) = p(s \mid \mathbf{X})$, then assumption (A2) is satisfied and the sampling design is ignorable. Since we only consider sampling designs of the form $p(s \mid \mathbf{Z})$, an obvious way to make the sampling design ignorable is achieved by including all design variables $\mathbf{z}$ into the estimation model. Examples of such design variables may include the variables used to form the strata, the variable used as a size measure if probability-proportional-to-size sampling is used and so on. The design weights may also provide a useful summary of the design information. Note that it may not be necessary to include all design variables into the estimation model (see Sugden and Smith 1984). Design variables that are independent of $\mathbf{y}$ (or $\boldsymbol{\varepsilon}$) after conditioning on $\mathbf{X}$ should not be included. To assess the validity of assumption (A2), a graph of the residuals, $e_k = y_k - \mathbf{x'_k}\hat{\boldsymbol{\beta}}$, versus design weights w_k (or any design variable) may be useful (see Pfeffermann 1993). Any trend in this graph suggests that the design weights are correlated with the random error $\boldsymbol{\varepsilon}$ and that the sampling design is not ignorable with respect to the estimation model. More formal tests can also be performed to assess the validity of this assumption (see, for example, DuMouchel and Duncan 1983; Graubard and Korn 1993; and, for more references on this topic, Pfeffermann 1993).

Under the estimation model *m* and the ignorability assumption (A2), it is easy to show that the BLU estimator (Royall 1976) $\hat{t}_y^B$ of t_y takes the simple projection form $\hat{t}_y^B = \mathbf{t'_x}\hat{\mathbf{B}}^{\mathbf{B}}$, where $\hat{\mathbf{B}}^{\mathbf{B}}$ is implicitly defined by the equation

$$\sum_{k\in s} (y_k - \mathbf{x'_k}\hat{\mathbf{B}}^{\mathbf{B}}) \frac{\mathbf{x_k}}{v_k} = \mathbf{0}. \tag{2.1}$$

The BLU estimator can also be written as $\hat{t}_y^B = \sum_{k\in s} w_k^B y_k$, where the BLU estimation weights w_k^B are given by

$$w_k^B = \frac{\mathbf{x}_\mathbf{k}'}{v_k}\left(\sum_{k\in s}\frac{\mathbf{x}_\mathbf{k}\mathbf{x}_\mathbf{k}'}{v_k}\right)^{-1}\mathbf{t}_\mathbf{x}. \qquad (2.2)$$

The model variance $\mathrm{V}_m\{(\hat{t}_y^B - t_y) \mid s, \mathbf{X}\}$ of $\hat{t}_y^B$ is the smallest for every possible sample among all linear m-unbiased estimators of t_y. A direct consequence of this result is that the anticipated variance $\mathrm{E}_m\{\mathrm{E}_p(\hat{t}_y^B - t_y)^2 \mid \mathbf{X}\}$ of $\hat{t}_y^B$ is also the smallest among all linear m-unbiased estimators of t_y, where the subscript p indicates that the expectation is evaluated with respect to the sampling design. Under the additional assumption that y_k given $\mathbf{X}$ follows a normal distribution, $\hat{\mathbf{B}}^\mathbf{B}$ is also the maximum likelihood estimator of the vector of model parameters $\boldsymbol{\beta}$.

In general, the BLU estimator $\hat{t}_y^B$ is not ADU. However, under the estimation model *m*, the ignorability assumption (A2) and the following additional assumption (A3), the BLU estimator has the property of being Asymptotically Design Unbiased in Probability (ADUP) in the sense that its relative design bias $\mathrm{E}_p(\hat{t}_y^B - t_y)/t_y$ converges in probability to 0 as *n* and *N* increase without bound.

A3) $\sum_{k\in U}\mathrm{E}_p\{(w_k^B)^2 I_k\}\sigma_k^2 = O(N)$, $\sum_{k\in U}\mathbf{x}_\mathbf{k}'\boldsymbol{\beta} = O(N)$ and $\sum_{k\in U}\sigma_k^2 = O(N)$, where $\sigma_k^2 = \sigma^2 v_k$ and I_k is a dummy random variable indicating whether unit k is selected in the sample ($I_k = 1$) or not ($I_k = 0$).

Assumption (A3) describes the asymptotic behaviour of three population quantities. In particular, requiring that $\sum_{k\in U}\mathrm{E}_p\{(w_k^B)^2 I_k\}\sigma_k^2 = O(N)$ essentially means that none of the BLU estimation weights becomes too large as the sample size and the population size increase. For instance, if $\mathbf{x}_\mathbf{k} = v_k = 1$ and if a sampling design of fixed size *n* is used, then condition $\sum_{k\in U}\mathrm{E}_p\{(w_k^B)^2 I_k\}\sigma_k^2 = O(N)$ is equivalent to assuming that the weights $w_k^B = N/n$ remain bounded as both *n* and *N* grow. The proof that $\hat{t}_y^B$ is ADUP is given in the appendix and does not require that $v_k = \mathbf{x}_\mathbf{k}'\boldsymbol{\lambda}$. As a result, the BLU estimator is ADUP even when the model variance $\mathrm{V}_m(y_k \mid \mathbf{X})$ is misspecified.

As pointed out above, the BLU estimator is efficient when the estimation model *m* and the normality assumption hold as well as the ignorability assumption (A2). Under these assumptions and the additional assumption (A3), the BLU estimator is also ADUP. Consequently, a first step towards robustness consists of selecting and validating an estimation model such that these assumptions are satisfied as much as possible. However, they are rarely fully satisfied in practice. For example, one can be reluctant to include all strata identifiers into the estimation model when the number of strata is very large. In such a case, the ignorability assumption might not fully hold. Also, the estimation model, including the normality assumption, may not hold for every variable of interest. Consequently, the non-critical use of the BLU estimator $\hat{t}_y^B$ of t_y is not always appropriate and robust estimators may be needed.

3. Design Robustness

Using the fact that $v_k = \mathbf{x}_\mathbf{k}'\boldsymbol{\lambda}$, it can be easily shown (see Särndal, Swensson and Wretman, 1992, page 231) that t_y can be expressed as $t_y = \mathbf{t}_\mathbf{x}'\mathbf{B}$, where **B** is implicitly defined by the equation

$$\sum_{k\in U}(y_k - \mathbf{x}_\mathbf{k}'\mathbf{B})\,\frac{\mathbf{x}_\mathbf{k}}{v_k} = \mathbf{0}. \qquad (3.1)$$

The vector **B** would be the LS estimator of $\boldsymbol{\beta}$, under the estimation model *m*, if a census could be conducted. Since $\mathbf{t}_\mathbf{x}$ is known, the objective of finding an estimator of the population total t_y is thus equivalent to finding an estimator of **B**. In the design-based theory, a natural estimator $\hat{\mathbf{B}}^\mathbf{G}$ of **B** is implicitly defined by the equation

$$\sum_{k\in s} w_k(y_k - \mathbf{x}_\mathbf{k}'\hat{\mathbf{B}}^\mathbf{G})\,\frac{\mathbf{x}_\mathbf{k}}{v_k} = \mathbf{0}, \qquad (3.2)$$

where w_k, the design weight of unit *k*, equals to the inverse of the selection probability π_k. The use of $\hat{\mathbf{B}}^\mathbf{G}$ leads to the GREG estimator $\hat{t}_y^G = \mathbf{t}_\mathbf{x}'\hat{\mathbf{B}}^\mathbf{G}$ of t_y. The GREG estimator $\hat{t}_y^G$ takes a simple projection form because $v_k = \mathbf{x}_\mathbf{k}'\boldsymbol{\lambda}$ (see Särndal *et al.* 1992, page 231). It can also be written as $\hat{t}_y^G = \sum_{k\in s} w_k^G y_k$, where the GREG estimation weights w_k^G are given by

$$w_k^G = w_k\frac{\mathbf{x}_\mathbf{k}'}{v_k}\left(\sum_{k\in s} w_k\frac{\mathbf{x}_\mathbf{k}\mathbf{x}_\mathbf{k}'}{v_k}\right)^{-1}\mathbf{t}_\mathbf{x}. \qquad (3.3)$$

As pointed out in the introduction, the GREG estimator is bias-robust since its property of being ADU is not dependent on the validity of the estimation model *m* and the ignorability assumption. However, the GREG estimator is not variance-robust since it may be far less efficient than the BLU estimator when both assumptions hold. The inefficiency of the GREG estimator is due to widely dispersed design weights. In household surveys, this situation is not uncommon because of many weight adjustments before calibration. Also, practical considerations for the choice of a sampling design combined with limited information available at the design stage often lead to sampling designs that are approximately ignorable. In household surveys, for instance, geographic information is often the main auxiliary information available to construct the strata. Unless the number of strata is very large, such information is usually weakly correlated with quantitative variables of interest, such as *expenditures* or *income*, and their corresponding population residual variable $E = y - \mathbf{x}'\mathbf{B}$. As a result, the design weight variable w is also weakly correlated with E. This suggests that the ignorability assumption may approximately hold. This also suggests that the design weights act more or less as a random noise when estimating **B** using (3.2) and that their influence could be significantly reduced. To obtain a design-robust estimator when the design weights are highly variable, we thus propose to shrink the design weights towards their mean and to use the LS estimator $\hat{t}_y^{\text{LS}} = \mathbf{t}'_\mathbf{x}\hat{\mathbf{B}}^{\mathbf{LS}}$, where $\hat{\mathbf{B}}^{\mathbf{LS}}$ is implicitly defined by

$$\sum_{k\in s} \tilde{w}_k (y_k - \mathbf{x}'_\mathbf{k}\hat{\mathbf{B}}^{\mathbf{LS}}) \frac{\mathbf{x}_\mathbf{k}}{v_k} = \mathbf{0} \tag{3.4}$$

and where $\tilde{w}_k$ is the shrunk weight of unit k given by

$$\tilde{w}_k = \left(\frac{\sum_{k\in s} w_k}{\sum_{k\in s} g(w_k;\alpha)} \right) g(w_k;\alpha). \tag{3.5}$$

The reason for the ratio in the right side of (3.5) is simply to ensure that $\sum_{k\in s} \tilde{w}_k = \sum_{k\in s} w_k$ and the role of the function $g(w_k;\alpha)$ is to obtain shrunk weights $\tilde{w}_k$ that are less variable than the design weights w_k. This function is assumed to be monotone in the constant α, with $1 \le g(w_k;\alpha) \le w_k$. The BLU and GREG estimators are therefore extreme special cases of the LS estimator obtained when α is such that $g(w_k;\alpha) = 1$ and $g(w_k;\alpha) = w_k$ respectively. To obtain a simple compromise between these two extreme estimators, we suggest using $g(w_k;\alpha) = w_k^\alpha$, with $0 \le \alpha \le 1$. The choice $\alpha = 0$ leads to the BLU estimator while the choice $\alpha = 1$ leads to the GREG estimator. In fact, this suggestion was proposed by Kish (1992, page 198). Other functions $g(w_k;\alpha)$ and other ways of reducing the variability of design weights can be found in the literature (see, for example, Elliott and Little 2000). Truncating large design weights ($g(w_k;\alpha) = \min(w_k,\alpha)$, with $\alpha > 0$) is a common approach that deals with this problem. This approach may be useful when assumptions (A1) and (A2) are not fully satisfied and when there are some abnormally large design weights. A better approach may be to truncate large weighted residuals. The weighted generalized *M*-estimation technique discussed in the next section can be used for this purpose.

The LS estimator $\hat{t}_y^{\text{LS}}$ can also be written as $\hat{t}_y^{\text{LS}} = \sum_{k\in s} w_k^{\text{LS}} y_k$, where the LS estimation weights w_k^{LS} are given by

$$w_k^{\text{LS}} = \tilde{w}_k \frac{\mathbf{x}'_\mathbf{k}}{v_k} \left(\sum_{k\in s} \tilde{w}_k \frac{\mathbf{x}_\mathbf{k}\mathbf{x}'_\mathbf{k}}{v_k} \right)^{-1} \mathbf{t}_\mathbf{x}. \tag{3.6}$$

Note that the estimation weights w_k^{LS}, including w_k^B and w_k^G as special cases, are calibrated on the known population totals $\mathbf{t}_\mathbf{x}$ in the sense that they satisfy the calibration equation $\sum_{k\in s} w_k^{\text{LS}} \mathbf{x}_\mathbf{k} = \mathbf{t}_\mathbf{x}$ (see Deville and Särndal 1992).

4. Model (Outlier) Robustness

As pointed out in the introduction, the LS estimator $\hat{t}_y^{\text{LS}}$ provides some protection against deviations from the ignorability assumption and also against deviations from model assumptions. However, it does not offer any protection against outliers, which can be viewed as a consequence of a model misspecification, including a deviation from the normality assumption. For instance, the GREG estimator is ADU no matter the validity of the estimation model. However, its design variance may be very large in the presence of outliers in the finite population because they may greatly influence its sampling error when they are selected in the sample. This problem may be amplified when the design weights are widely dispersed. For the Horvitz-Thompson estimator, this was well illustrated in the circus example of Basu (1971). Of course, the use of efficient auxiliary variables at the estimation stage can control the impact of outliers on estimates. However, such auxiliary variables are often not available and outlier-robust estimators may provide significant gains over the LS estimator.

Using the Taylor linearization technique (see, for example, Särndal *et al.* 1992, page 235) and given that $t_y = \mathbf{t}'_\mathbf{x}\mathbf{B}$, it is well known and easy to show that the sampling error of the GREG estimator can be approximated as follows: $\hat{t}_y^G - t_y \approx \sum_{k\in s} w_k E_k$, where $E_k = y_k - \mathbf{x}'_\mathbf{k}\mathbf{B}$ is the population residual for unit k. As a result, a large design weight associated with a large population residual (or outlier) may have a substantial impact on the quality of the GREG estimator. Moreover, it is straightforward to show that the sampling error of the LS estimator can be expressed as $\hat{t}_y^{\text{LS}} - t_y = \sum_{k\in s} w_k^{\text{LS}} E_k$. Therefore, a large estimation weight associated with a large population residual may greatly influence the sampling error and the quality of the LS estimator. To deal with this problem, we use the Schweppe version (Hampel *et al.* 1986, pages 315 – 316) of the weighted generalized *M*-estimation technique to reduce the influence of units with large weighted population residuals. This leads to the *M*-estimator $\hat{\mathbf{B}}^{\mathbf{M}}$ of $\mathbf{B}$, which is implicitly defined by

$$\sum_{k\in s} \tilde{w}_k \frac{1}{h_k} \psi\left(\frac{h_k \tilde{E}_k(\hat{\mathbf{B}}^{\mathbf{M}})}{Q} \right) \frac{\mathbf{x}_\mathbf{k}}{\sqrt{v_k}} = \mathbf{0}, \tag{4.1}$$

where $\tilde{E}_k(\hat{\mathbf{B}}^{\mathbf{M}}) = (y_k - \mathbf{x}'_\mathbf{k}\hat{\mathbf{B}}^{\mathbf{M}})/\sqrt{v_k}$, Q is a positive population scale parameter and h_k is a weight that may depend not only on $\mathbf{x}_\mathbf{k}$ but also on $\mathbf{z}_\mathbf{k}$. The role of the function $\psi(.)$ consists of reducing the influence of units with a large $h_k\tilde{E}_k(\mathbf{B})$. From the above considerations, $h_k = w_k^{\text{LS}}\sqrt{v_k}$ or $h_k = \tilde{w}_k\sqrt{v_k}$ is a natural choice. In the former case, the influence of large $w_k^{\text{LS}} E_k$ is reduced while, in the latter case, the influence of large $\tilde{w}_k E_k$ is reduced. The choice $h_k = w_k^{\text{LS}}\sqrt{v_k}$ may be preferred to $h_k = \tilde{w}_k\sqrt{v_k}$ when there are outliers in the auxiliary variables $\mathbf{x}$ or when α is not close to 1 (assuming $g(w_k;\alpha) = w_k^\alpha$). The main point here is that h_k should depend on survey weights w_k^{LS} or $\tilde{w}_k$ and that both choices suggested above should perform better than simpler choices that do not take into account the auxiliary variables $\mathbf{z}$ such as $h_k = \sqrt{v_k}$ or $h_k = 1$, which reduce the influence of large unweighted residuals. Also, it should again be noted that the interest is in finding a robust estimator for the vector of population parameters $\mathbf{B}$ and not for the vector of model parameters $\boldsymbol{\beta}$. In fact, $\mathbf{B}$ is itself not robust (in the classical sense) for $\boldsymbol{\beta}$ since it may be highly affected by the presence of outliers in the finite population. As a result, outliers must have a certain influence on $\hat{\mathbf{B}}^{\mathbf{M}}$.

Equation (4.1) can be written in the weighted linear regression form:

$$\sum_{k\in s} \tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, Q)\ (y_k - \mathbf{x}_{\mathbf{k}}'\hat{\mathbf{B}}^{\mathbf{M}})\ \frac{\mathbf{x}_{\mathbf{k}}}{v_k} = 0\,, \qquad (4.2)$$

where

$$\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, Q) = \tilde{w}_k \frac{\psi(r_k)}{r_k}$$

and

$$r_k = \frac{h_k \tilde{E}_k(\hat{\mathbf{B}}^{\mathbf{M}})}{Q}.$$

We propose the following modification of the popular function $\psi(.)$ of Huber (1964) that makes the adjusted weights $\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, Q)$ always greater than or equal to 1: $\psi(r_k) = r_k$, if $|r_k| \le \varphi$, and $\psi(r_k) = \text{sign}(r_k)\ \max(|r_k|/\tilde{w}_k, \varphi)$, otherwise, where φ is a positive constant. This leads to adjusted weights

$$\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, Q) = \begin{cases} \tilde{w}_k\,, & \text{if } |r_k| \le \varphi, \\ \max\left(1,\ \tilde{w}_k \dfrac{\varphi}{|r_k|}\right), & \text{otherwise}. \end{cases} \qquad (4.3)$$

The Iteratively Reweighted Least-Squares (IRLS) algorithm (Beaton and Tukey 1974) is often used to solve (4.2) and (4.3). At a given iteration i, the adjusted weights $\tilde{w}_k^*(\mathbf{B}^{(i-1)}, Q^{(i-1)})$ are first calculated using (4.3) and then $\mathbf{B}^{(i)}$ is obtained by solving (4.2) with $\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, Q)$ and $\hat{\mathbf{B}}^{\mathbf{M}}$ replaced by $\tilde{w}_k^*(\mathbf{B}^{(i-1)}, Q^{(i-1)})$ and $\mathbf{B}^{(i)}$ respectively. To obtain $\mathbf{B}^{(i)}$, an estimate of Q is usually calculated at each iteration of the IRLS algorithm. In the simulation study of section 6, we have used

$$\begin{aligned} Q^{(i-1)} = 1.483 &\\ \times \text{ weighted sample}&\\ \text{median of}\left(\left|h_k \tilde{E}_k\left(\mathbf{B}^{(i-1)}\right)\right|\ ; k \in s\right), &\qquad (4.4)\end{aligned}$$

where the weighted sample median is calculated using the weights $\tilde{w}_k / h_k$. Equation (4.4) reduces to the proposal of Hulliger (1999) when $h_k = 1$ and $g(w_k; \alpha) = w_k$. We suggest using $\mathbf{B}^{(0)} = \hat{\mathbf{B}}^{\mathbf{LS}}$ as the vector of starting values since $\hat{\mathbf{B}}^{\mathbf{LS}}$ is easy to obtain. The iterative procedure is normally repeated until convergence is reached. To reduce computer time, especially if a resampling method is used for MSE estimation, a single iteration of the IRLS algorithm can be performed. In section 6, it is shown empirically that performing a single iteration yields an estimator of the population total that has properties similar to the fully-iterated estimator. This point has also been noted by Lee (1991).

The M-estimator of t_y is given by $\hat{t}_y^M = \mathbf{t}_{\mathbf{x}}'\hat{\mathbf{B}}^{\mathbf{M}}$. With the restriction that $\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q}) \ge 1$, where $\hat{Q}$ is an estimator of Q, the estimators $\hat{\mathbf{B}}^{\mathbf{M}}$ and $\hat{t}_y^M$ are census-consistent in the sense that they are exactly equal to $\mathbf{B}$ and t_y respectively, no matter the value of φ and α, when a census is conducted ($\pi_k = 1$, for $k \in U$). This restriction might be useful for controlling the design bias of $\hat{t}_y^M$ when there are shrunk weights $\tilde{w}_k$ close to 1. Note that the estimators $\hat{\mathbf{B}}^{\mathbf{M}}$ and $\hat{t}_y^M$ reduce to $\hat{\mathbf{B}}^{\mathbf{LS}}$ and $\hat{t}_y^{\text{LS}}$ respectively when $\varphi = \infty$ ($\psi(r_k) = r_k$). The M-estimator $\hat{t}_y^M$ can also be expressed as $\hat{t}_y^M = \sum_{k\in s} w_k^M y_k$, where the M-estimation weights w_k^M are given by

$$w_k^M = \tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q}) \frac{\mathbf{x}_{\mathbf{k}}'}{v_k} \left(\sum_{k\in s} \tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q}) \frac{\mathbf{x}_{\mathbf{k}}\mathbf{x}_{\mathbf{k}}'}{v_k} \right)^{-1} \mathbf{t}_{\mathbf{x}}. \qquad (4.5)$$

The estimation weights w_k^M are still calibrated on the known population totals $\mathbf{t}_{\mathbf{x}}$ ($\sum_{k\in s} w_k^M \mathbf{x}_{\mathbf{k}} = \mathbf{t}_{\mathbf{x}}$).

In order to determine appropriate values for α and φ, the MSE of the M-estimator $\hat{t}_y^M$ can be estimated for different choices of α and φ using past or current sample data. Then, the values of α and φ that give the smallest estimated MSE can be chosen. Estimation of MSE is discussed in section 5. As noted in Hulliger (1995), choosing adaptively α and φ by minimizing the estimated MSE with current sample data leads to an estimator $\hat{t}_y^M$ that does not require estimating the scale parameter Q. Also, this procedure controls the magnitude of the design bias of $\hat{t}_y^M$ without requiring the use of additional constants. However, it is likely to provide less efficiency than using the optimal (although unknown) values of α and φ.

In multipurpose surveys, different values of α and φ are likely to be obtained for different variables of interest. If multiple sets of weights are to be avoided, some form of compromise is needed. As a first step towards a compromise, a common value of α, satisfactory for the most important variables of interest, can be determined. Then, we propose two practical ways of implementing the M-estimator $\hat{t}_y^M$ without having to find a compromise value for φ; either the weights of influential units are modified and a calibration approach is used to obtain a single set of robust estimation weights or the values of influential units are modified. The former is discussed in section 4.1 while the latter is discussed in section 4.2.

4.1 Modification of the Weights of Influential Units

Let us now assume that it is desired to estimate the population totals of a vector of q variables of interest $\mathbf{y} = (y_1, y_2, ..., y_q)'$. A vector of q M-estimators $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathbf{M}} = (\hat{t}_{y_1}^M, \hat{t}_{y_2}^M, ..., \hat{t}_{y_q}^M)'$ of $\mathbf{t}_{\mathbf{y}} = \sum_{k \in U} \mathbf{y}_{\mathbf{k}}$ can be obtained, with potentially different values of φ for different variables. To simplify the notation, we denote the adjusted weights associated with variable y_i by $\tilde{w}_k^*(y_i)$, for $i = 1, 2, ..., q$. Since the adjusted weights $\tilde{w}_k^*(y_i)$ depend on the variable of interest y_i, we obtain q sets of weights, even if a common value of φ is chosen.

Gwet and Rivest (1992), Duchesne (1999) and Hulliger (1999) suggested using the adjusted weights $\tilde{w}_k^*(\mathbf{y}) = \min(\tilde{w}_k^*(y_1), \tilde{w}_k^*(y_2), ..., \tilde{w}_k^*(y_q))$ to obtain a unique set of weights. Then, estimation weights $w_k^M(\mathbf{y})$ are calculated by replacing $\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q})$ by $\tilde{w}_k^*(\mathbf{y})$ in (4.5) and $\mathbf{t}_{\mathbf{y}}$ is estimated by $\sum_{k \in s} w_k^M(\mathbf{y})\, \mathbf{y}_{\mathbf{k}}$. Although the estimation weights $w_k^M(\mathbf{y})$ are calibrated on the known population totals $\mathbf{t}_{\mathbf{x}}$, they are not calibrated on the vector of estimates $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathbf{M}}$, which are believed to be our best estimates in the sense of minimizing the estimated MSE. Moreover, the use of $\sum_{k \in s} w_k^M(\mathbf{y})\, \mathbf{y}_{\mathbf{k}}$ likely leads to a larger design bias than $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathbf{M}}$ although it controls the design variance. To cope with these issues, we propose computing the estimation weights $w_k^{M,A}(\mathbf{y})$ by replacing $\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q})$ by the adjusted weights $\tilde{w}_k^*(\mathbf{y})$ in (4.5), and by augmenting the vector of auxiliary variables $\mathbf{x}$ and the known population totals $\mathbf{t}_{\mathbf{x}}$ using $\mathbf{y}$ and $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathbf{M}}$ respectively. As a result, the estimation weights $w_k^{M,A}(\mathbf{y})$ are calibrated on $\mathbf{t}_{\mathbf{x}}$ and $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathbf{M}}$, and $\mathbf{t}_{\mathbf{y}}$ is estimated by $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathbf{M}} = \sum_{k \in s} w_k^{M,A}(\mathbf{y})\, \mathbf{y}_{\mathbf{k}}$. Of course, there may be a limit on the number of variables that can be used for calibration purposes. This may somewhat restrict the applicability of this method when q is very large.

4.2 Modification of the Values of Influential Units

Another way of implementing the M-estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathbf{M}}$ in practice consists of modifying the values of the variables of interest $\mathbf{y}$ and using the LS estimation weights w_k^{LS} for all variables. This can be done separately for each variable of interest, so we return to the case of only one variable of interest in this section.

Let us first denote by s_o the random set of all sample units k for which $\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q}) \neq \tilde{w}_k$. In other words, s_o is the random set of units that have been detected as being influential. Let also $\hat{\mathbf{B}}^{\mathbf{M}*}$ be implicitly defined by the equation

$$\sum_{k \in s} \tilde{w}_k (y_{\bullet k} - \mathbf{x}_{\mathbf{k}}' \hat{\mathbf{B}}^{\mathbf{M}*}) \frac{\mathbf{x}_{\mathbf{k}}}{v_k} = \mathbf{0}, \tag{4.6}$$

where $y_{\bullet k} = y_k$, if $k \in s - s_o$, and $y_{\bullet k} = y_k^*$, otherwise. The quantity y_k^* is a modified value for the influential unit k that is used to replace y_k. Note that $\hat{\mathbf{B}}^{\mathbf{M}*} = \hat{\mathbf{B}}^{\mathbf{LS}}$ if $y_{\bullet k} = y_k$, for $k \in s$. The population total t_y can then be estimated by $\hat{t}_y^{M*} = \mathbf{t}_{\mathbf{x}}' \hat{\mathbf{B}}^{\mathbf{M}*}$. It is also easy to show that $\hat{t}_y^{M*} = \sum_{k \in s} w_k^{LS} y_{\bullet k}$.

The idea here consists of finding modified values y_k^*, for $k \in s_o$, as close as possible to the original values y_k and that satisfy the constraint $\hat{\mathbf{B}}^{\mathbf{M}*} = \hat{\mathbf{B}}^{\mathbf{M}}$. Under this constraint, it is obvious that $\hat{t}_y^{M*} = \hat{t}_y^M$. A possible implementation of this idea is obtained by minimizing the distance function $\sum_{k \in s_o} \tilde{w}_k (y_k - y_k^*)^2 / v_k$ subject to the constraint $\hat{\mathbf{B}}^{\mathbf{M}*} = \hat{\mathbf{B}}^{\mathbf{M}}$. This leads to the modified values

$$y_k^* = y_k + \mathbf{x}_{\mathbf{k}}' \left(\sum_{k \in s_o} \frac{\tilde{w}_k}{v_k} \mathbf{x}_{\mathbf{k}} \mathbf{x}_{\mathbf{k}}' \right)^{-1} \left(\sum_{k \in s} \frac{\tilde{w}_k}{v_k} \mathbf{x}_{\mathbf{k}} \mathbf{x}_{\mathbf{k}}' \right) \left(\hat{\mathbf{B}}^{\mathbf{M}} - \hat{\mathbf{B}}^{\mathbf{LS}} \right). \tag{4.7}$$

This idea is essentially equivalent to reverse calibration proposed by Ren and Chambers (2002), except that these authors used the constraint $\hat{t}_y^{M*} = \hat{t}_y^M$ instead of $\hat{\mathbf{B}}^{\mathbf{M}*} = \hat{\mathbf{B}}^{\mathbf{M}}$. We prefer the latter since it leads to modified values that better preserve the relationships between the variable of interest y and the auxiliary variables $\mathbf{x}$.

Other ways of determining modified values that satisfy the constraint $\hat{\mathbf{B}}^{\mathbf{M}*} = \hat{\mathbf{B}}^{\mathbf{M}}$ can be found. For example, it is straightforward to show that this constraint is satisfied when the following modified values are used:

$$y_k^* = a_k y_k + (1 - a_k) \mathbf{x}_{\mathbf{k}}' \hat{\mathbf{B}}^{\mathbf{M}}, \tag{4.8}$$

where $a_k = \tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q}) / \tilde{w}_k$. The modified values in equation (4.8) have a simple interpretation: they are a weighted average of the robust prediction $\mathbf{x}_{\mathbf{k}}' \hat{\mathbf{B}}^{\mathbf{M}}$ and the observed value y_k. Less weight is given to the observed value y_k when it has a smaller value of a_k and, therefore, when it is highly influential.

5. Mean Squared Error Estimation

Estimation of the MSE of $\hat{t}_y^M$ can be used for three different purposes: i) finding appropriate values for α and φ using past or current sample data, ii) evaluating the quality of estimates and iii) making inferences about unknown population quantities. Using the fact that $\mathrm{E}_p(\hat{t}_y^G) \approx t_y$, it can be easily shown that the MSE of $\hat{t}_y^M$ can be approximated by

$$\mathrm{MSE}_p(\hat{t}_y^M) \approx \mathrm{V}_p(\hat{t}_y^M) + \mathrm{E}_p(\hat{t}_y^M - \hat{t}_y^G)^2 - \mathrm{V}_p(\hat{t}_y^M - \hat{t}_y^G). \tag{5.1}$$

The last two terms of (5.1) are equal to $[\mathrm{E}_p(\hat{t}_y^M - \hat{t}_y^G)]^2$. They represent the square of the design bias of $\hat{t}_y^M$. As suggested in Gwet and Rivest (1992), a potential estimator of $\mathrm{MSE}_p(\hat{t}_y^M)$ is given by

$$\mathrm{mse}_p(\hat{t}_y^M) = \hat{\mathrm{V}}_p(\hat{t}_y^M) + \max\left(0\ ,\ (\hat{t}_y^M - \hat{t}_y^G)^2 - \hat{\mathrm{V}}_p(\hat{t}_y^M - \hat{t}_y^G)\right), \qquad (5.2)$$

where $\hat{\mathrm{V}}_p(\hat{t}_y^M)$ and $\hat{\mathrm{V}}_p(\hat{t}_y^M - \hat{t}_y^G)$ are estimators of $\mathrm{V}_p(\hat{t}_y^M)$ and $\mathrm{V}_p(\hat{t}_y^M - \hat{t}_y^G)$ respectively.

Since the estimator $\hat{t}_y^M$ has a complex structure, re-sampling variance estimation methods provide a convenient way of estimating $\mathrm{V}_p(\hat{t}_y^M)$ and $\mathrm{V}_p(\hat{t}_y^M - \hat{t}_y^G)$. The jack-knife, the bootstrap and the balanced repeated replications methods are described and evaluated in Rao, Wu and Yue (1992) for stratified multistage sampling designs, where the primary sampling units are assumed to have been selected with replacement. They have shown in an empirical study that the jackknife variance estimator can have a large bias when estimating the variance of a non-smooth estimator, such as the sample median. Therefore, the jackknife variance estimator might be more biased for estimating the variance of the *M*-estimator than the balanced repeated replication or the bootstrap method when, at each iteration of the IRLS algorithm, Q is estimated using a non-smooth estimator such as (4.4). Gwet and Lee (2000) studied empirically the performance of the jackknife and the bootstrap methods for some robust estimators. In general, they found encouraging results. It is important to note that the estimator $\hat{t}_y^M$ should be recomputed for each resample. This includes repeating the procedure used to estimate α and φ if they are estimated using current sample data.

When the goal of MSE estimation is only to find appropriate values for α and φ, it may be convenient to consider simplified MSE estimators in order to reduce computer time. We now propose four different ways of simplifying MSE estimation:

i) Only a single iteration of the IRLS algorithm could be done for each resample even if a fully-iterated *M*-estimator is used. This might yield reasonable variance estimates since the singly-iterated and fully-iterated *M*-estimators seem to have similar properties (see section 6.4).

ii) Some quantities could be assumed fixed (not random) for MSE estimation. This is likely to lead to an underestimation of the MSE but it may be useful if the goal of MSE estimation is only to find appropriate values for α and φ. For example, the adjusted weights $\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q})$ could be assumed fixed. This approximation was in fact suggested in Hulliger (1999). Alternatively, if the *M*-estimator is implemented using the methodology in section (4.2), the modified values in (4.7) or (4.8) could be treated as true values for MSE estimation.

iii) The term $\hat{\mathrm{V}}_p(\hat{t}_y^M - \hat{t}_y^G)$ in (5.2) could be omitted. This would lead to the MSE estimator: $\mathrm{mse}_p(\hat{t}_y^M) = \hat{\mathrm{V}}_p(\hat{t}_y^M) + (\hat{t}_y^M - \hat{t}_y^G)^2$. Note that this approach leads to an overestimation of the MSE.

iv) A combination of two of the above three propositions could be considered. For example, the adjusted weights $\tilde{w}_k^*(\hat{\mathbf{B}}^{\mathbf{M}}, \hat{Q})$ could be assumed fixed and the term $\hat{\mathrm{V}}_p(\hat{t}_y^M - \hat{t}_y^G)$ in (5.2) could be omitted. In such a case, an estimator for $\mathrm{V}_p(\hat{t}_y^M)$ could be obtained by noting that $\mathrm{V}_p(\hat{t}_y^M) = \mathbf{t}'_{\mathbf{x}} \mathbf{V}_p(\hat{\mathbf{B}}^{\mathbf{M}}) \mathbf{t}_{\mathbf{x}}$ and by using the well known Taylor linearization technique of Binder (1983) to estimate $\mathbf{V}_p(\hat{\mathbf{B}}^{\mathbf{M}})$. After some straightforward algebra, we obtain the MSE estimator

$$\mathrm{mse}_p(\hat{t}_y^M) = \sum_{k\in s}\sum_{l\in s} \frac{(\pi_{kl} - \pi_k \pi_l)}{\pi_{kl}} w_k^M (y_k - \mathbf{x}'_{\mathbf{k}} \hat{\mathbf{B}}^{\mathbf{M}}) w_l^M (y_l - \mathbf{x}'_{\mathbf{l}} \hat{\mathbf{B}}^{\mathbf{M}}) + (\hat{t}_y^M - \hat{t}_y^G)^2, \qquad (5.3)$$

where π_{kl} is the joint probability of selection of units k and l.

6. Simulation Study

We performed a simulation study to evaluate some properties of the LS estimator and the *M*-estimator for a skewed finite population. In particular, we compared a version of the *M*-estimator that reduces the influence of large weighted population residuals to another one that reduces the influence of large unweighted population residuals. We also compared the performance of the singly- and fully-iterated *M*-estimators. Section 6.1 describes the population and the sampling design, and sections 6.2 to 6.4 discuss results from the simulation.

6.1 Population and Sampling Design

The data from Statistics Canada's 1998 Survey of Household Spending (SHS) are used to serve as the population. This survey uses a stratified multi-stage design and contains information about 15,457 households on several variables. The variable *Renovation/Repair* is chosen as the variable of interest y. This variable is considered for its greater potential of having very large values. A vector $\mathbf{x}$ of three binary auxiliary variables have been created by dividing the variable *Income* into three categories (*Income* ≤ 30,000, 30,000 < *Income* ≤ 60,000 and *Income* > 60,000) and we have chosen $v_k = 1$, for all $k \in U$. In other words, we have considered a poststratification estimation model, which should give us robustness against deviations from the linearity assumption. The population coefficient of determination (R^2) for this estimation model is 0.13. This is a typical R^2 in household surveys.

From this population, 5,000 samples of expected sample size 300 have been selected using Poisson sampling. We wanted to give households quite dispersed probabilities of selection resulting in variable design weights. We thus assigned probabilities of selection such that they were proportional to the inverse of the SHS design weights (which include a nonresponse adjustment factor). The selection probabilities are thus given by $\pi_k = (300/\sum_{k\in U}\pi_k^*)\,\pi_k^*$, where π_k^*, for $k \in U$, is the reciprocal of the design weight (including a nonresponse adjustment factor) from the SHS data.

Table 6.1 gives some summary statistics for this population. We note that the population residuals are very skewed and that the skewness increases when the residuals are multiplied by the design weights. Figure 6.1 shows a graph of the population residuals versus the design weights. First, we note that there is a clear outlier with a residual greater than 50,000 and with a design weight not close to 1. Fortunately, the most extreme design weights are not associated with large population residuals. Also, although this graph may be misleading because of the huge number of points that are overlapping, there does not seem to be any clear relationship between the population residuals and the design weights. In fact, the coefficient of correlation between the design weights and the population residuals is 0.0049. Such a small coefficient of correlation is not atypical in household surveys, for reasons discussed in section 3, and suggests that the ignorability assumption may hold approximately.

Table 6.1
Summary Statistics about the Population

Variable	Mean	Standard Deviation	Skewness
Renovation/Repair	367	1,124	12.6
Population Residual	0	1,104	12.8
Design Weight	177	170	1.8
Weighted Population Residual	922	295,685	15.0

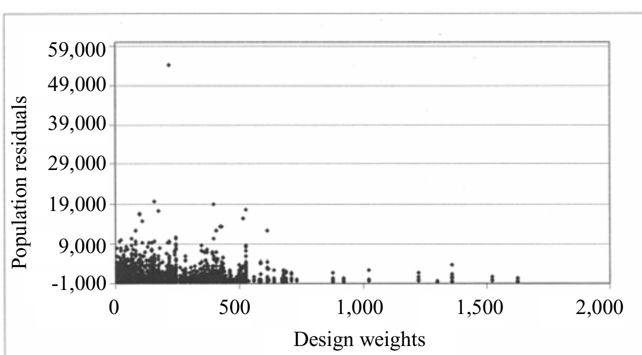


Figure 6.1. Graph of the population residuals versus the design weights.

For each of the 5,000 samples, estimates of the population total for the *Renovation/Repair* variable have been calculated for both the LS estimator and two versions of the *M*-estimator; one that reduces the influence of large weighted population residuals ($h_k = \tilde{w}_k$) and another one that reduces the influence of large unweighted population residuals ($h_k = 1$). For the i^{th} sample, the relative error in percentage of any estimate $\hat{t}_{yi}$ of t_y is defined as $\Delta_i = 100\% \times (\hat{t}_{yi} - t_y)/t_y$. The Relative Bias (RB) and the Relative Root Mean Squared Error (RRMSE) of any estimator $\hat{t}_y$, expressed as a percentage of the population total, can thus be estimated by RB $= \sum_{i=1}^{5,000} \Delta_i / 5,000$ and RRMSE $= \sqrt{\sum_{i=1}^{5,000} \Delta_i^2 / 5,000}$ respectively. Another measure of interest is the Maximum Absolute Relative Error (MARE) in percentage given by MARE $= \max(|\Delta_i|;\ i = 1, 2, ..., 5,000)$. This measure may be useful to assess the sensitivity of an estimator to the presence of influential units in the sample.

6.2 The LS Estimator: Design Robustness

In this section, we evaluate the properties of the LS estimator. Figure 6.2 illustrates the RB, RRMSE and MARE of the LS estimator for 11 values of α ($\alpha = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1) when $g(w_k;\alpha) = w_k^{\alpha}$. On the one hand, the BLU estimator ($\alpha = 0$) has an RRMSE close to the minimum and the smallest MARE among these 11 values of α but, as expected, leads to the largest RB (in absolute value). Its RB is equal to -13.05%, which is not negligible. Given that a poststratification model is used, this suggests that the ignorability assumption is not fully satisfied even though the correlation between the design weights and the population residuals is small. On the other hand, the GREG estimator ($\alpha = 1$) has a very small RB but has the largest RRMSE and MARE due to the variability of the design weights. When $\alpha = 0.2$, the LS estimator is biased, with an RB of –9.11%, but has a value of MARE relatively close to the smallest value and has the smallest RRMSE (17.94%) among the values of α considered. This is a substantial reduction in comparison with the RRMSE of the GREG estimator (34.77%). In general, values of α between 0.2 and 0.5 provide a reasonable compromise estimator with respect to RB, RRMSE, and MARE. Note that, for larger expected sample sizes, we expect that the minimum MSE be reached for larger values of α because the bias of the LS estimator may dominate its variance.

We have also considered the LS estimator obtained by choosing adaptively, for each selected sample, the value of α that leads to the smallest estimated MSE among the set of 11 values of α considered above. The MSE has been estimated using equation (5.3). The average value of α over the 5,000 selected samples is 0.43. This is slightly larger than the value of α (0.2) that leads to the smallest MSE (see figure 6.2). This may be due to the simplification made to obtain (5.3), which omits a component of the square design bias when estimating the MSE. Nevertheless, this LS estimator shows a significant improvement over the GREG estimator in terms of RRMSE (26.05%) and MARE (217.99%). This LS estimator shows also a significant improvement over the BLU estimator in terms of RB (-6.24%). Therefore, it seems that choosing adaptively the value of α leads to a useful compromise between the GREG and BLU estimators. However, there is a price to pay in terms of RRMSE by estimating α instead of using the optimal (although unknown) value of α.

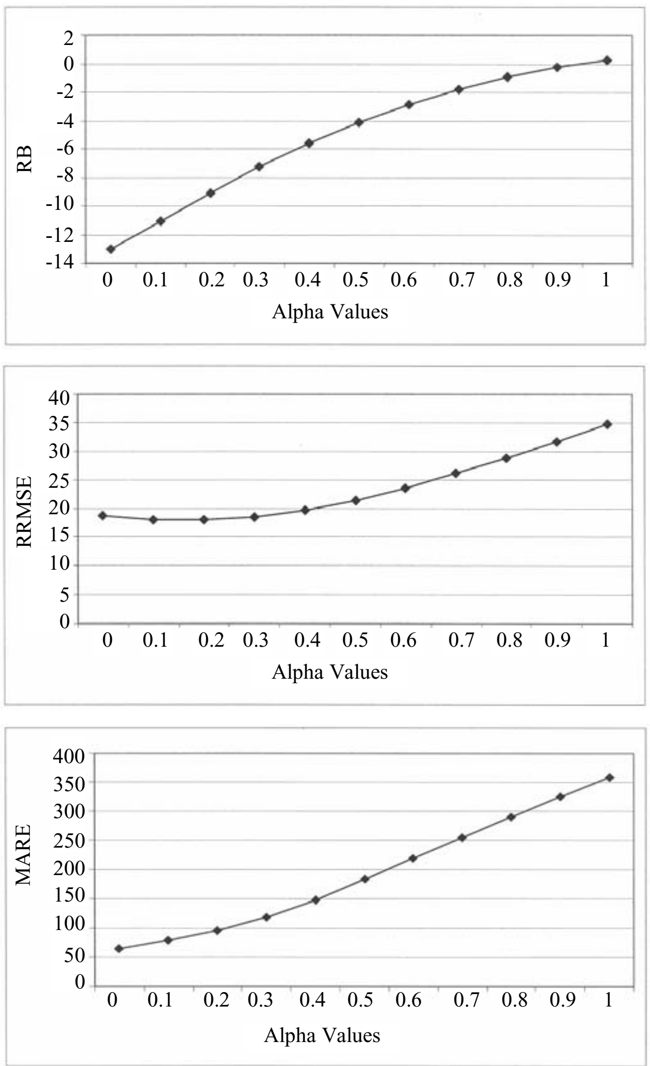


Figure 6.2. RB, RRMSE and MARE of the LS estimator.

6.3 The *M*-estimator: Outlier robustness

We have compared two versions of the *M*-estimator; one that reduces the influence of large weighted population residuals $(h_k = \tilde{w}_k)$ and another one that reduces the influence of large unweighted population residuals $(h_k = 1)$. For the weighted version, we chose 7 values of φ (φ = 10, 25, 50, 100, 150, 200, ∞) and for the unweighted version, we chose 9 values of φ (φ = 2, 5, 8, 11, 14, 17, 20, 30, ∞). We have only considered the case $\alpha = 1$, as we did not want to confound the effects of changing the constant α with the effect of changing the constant φ. Of course, a more efficient estimator could be found by an appropriate choice of both constants. It is to be noted that the results are based on a single iteration of the IRLS algorithm using $\mathbf{B}^{(0)} = \hat{\mathbf{B}}^{\mathbf{G}}$ as the vector of starting values.

It can be seen from figures 6.3 and 6.4 that the weighted version $(h_k = \tilde{w}_k)$ has a better potential for reducing the RRMSE and the MARE of *M*-estimators than the unweighted version $(h_k = 1)$. Both graphs of RRMSE present a *U*-shaped curve. The RRMSE curve for $h_k = \tilde{w}_k$ shows that a value of φ between 50 and 150 leads to an RRMSE between 25% and about 27%, while the RRMSE of the GREG estimator (last point on the graphs) is equal to 34.77%. The RRMSE curve for $h_k = 1$ shows that the RRMSE is around 30% for values of φ between 8 and 20. In the area where the RRMSE is close to its minimum value, the MARE is smaller when $h_k = \tilde{w}_k$. This suggests that $h_k = \tilde{w}_k$ may control influential units better than $h_k = 1$. As expected, the RB in both figures decreases as φ increases.

We have also considered the weighted and unweighted versions of the *M*-estimator obtained by choosing adaptively, for each selected sample, the value of φ that leads to the smallest estimated MSE (using equation 5.3) among the sets of values of φ considered above. The average value of φ over the selected samples is 72.34 for the weighted version and 10.58 for the unweighted version. Calculation of these averages excludes samples for which $\varphi = \infty$ (13 samples for $h_k = \tilde{w}_k$ and 1 sample for $h_k = 1$). Both averages are close to the optimal values of φ found in figures 6.3 and 6.4 (100 for $h_k = \tilde{w}_k$, and 11 for $h_k = 1$). The weighted version of the *M*-estimator has an RB of -10.24%, RRMSE of 28.07% and MARE of 197.86%. The unweighted version of the *M*-estimator has an RB of -8.26%, RRMSE of 28.18% and MARE of 232.57%. Therefore, both versions of the *M*-estimator lead to a significant improvement over the GREG estimator in terms of RRMSE and MARE at the expense of an increase in RB (around -10%). The MARE is smaller for the weighted version, which again indicates that it controls influential units better than the unweighted version. However, the difference in the RRMSE between these two estimators is very small. Curiously, it seems that there is no increase in MSE due to estimating φ instead of using the optimal value when the unweighted version is used. This observation is somewhat difficult to explain.

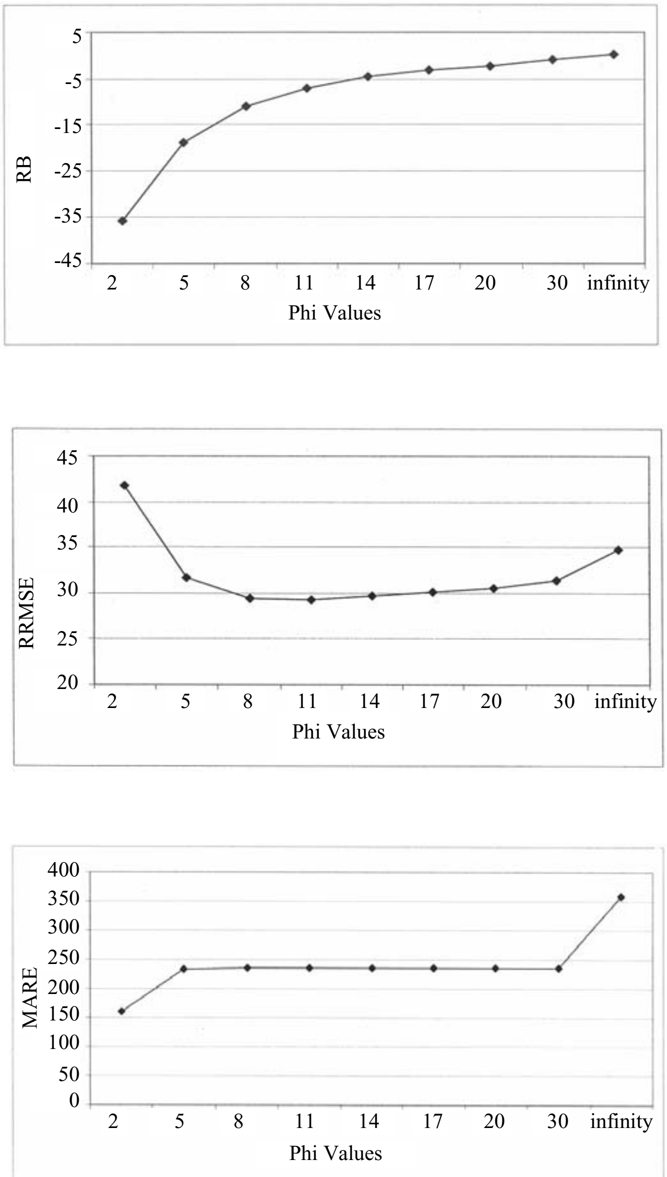


Figure 6.3. RB, RRMSE and MARE of the *M*-estimator when $h_k = 1$.

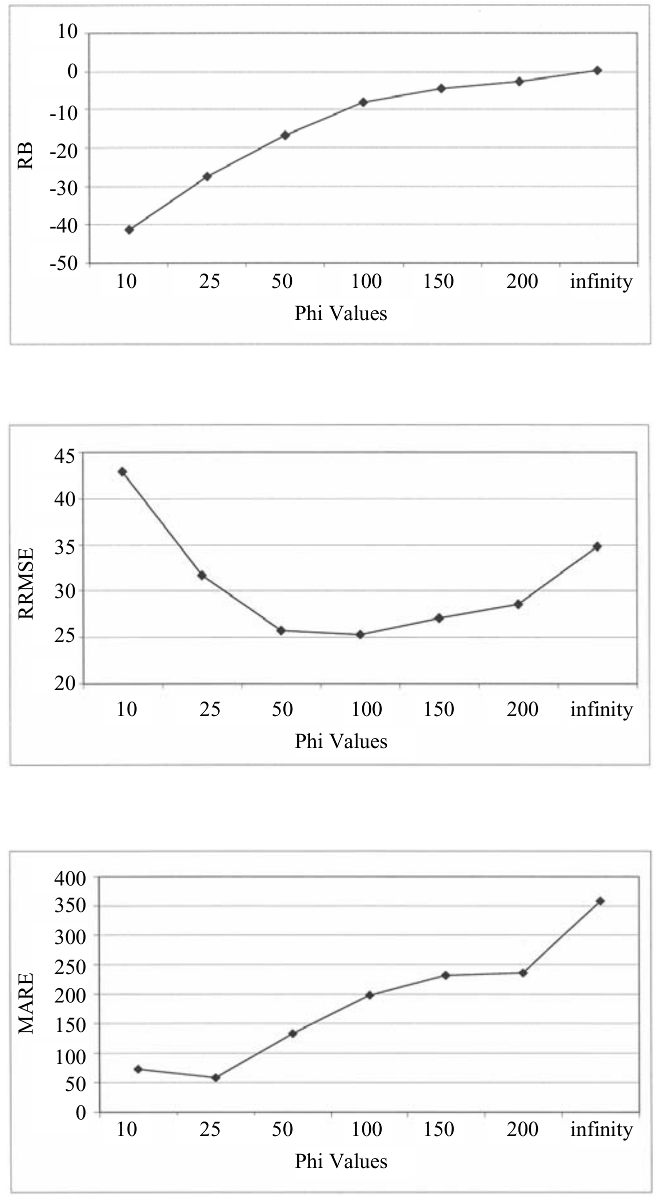


Figure 6.4. RB, RRMSE and MARE of the *M*-estimator when $h_k = \tilde{w}_k$.

6.4 Comparison of the Singly-iterated and Fully-iterated *M*-estimators

We now compare the singly- and fully-iterated *M*-estimators when $\alpha = 1$. We only consider the following two cases: i) $h_k = 1$ and $\varphi = 11$; and ii) $h_k = \tilde{w}_k$ and $\varphi = 100$. Most of the time, the IRLS algorithm converged quickly in the fully-iterated case (average number of iterations for convergence is 7.53 for $h_k = 1$, and 7.29 for $h_k = \tilde{w}_k$), but in some of the 5,000 samples (64 for $h_k = 1$, and 75 for $h_k = \tilde{w}_k$) it did not converge. When this situation occurred, we kept the *M*-estimate from the last iteration of the IRLS algorithm. From table 6.2, it is evident that the RB, RRMSE and MARE of the singly- and fully-iterated *M*-estimators are very close to each other. A point worth noting is the slightly smaller RBs for singly-iterated *M*-estimators. This point has also been observed by Lee (1991) and is likely due to the fact that we used $\mathbf{B}^{(0)} = \hat{\mathbf{B}}^{\mathbf{G}}$ as the vector of starting values for the IRLS algorithm, which is ADU for **B**.

Table 6.2
Comparison of Singly- and Fully-iterated *M*-estimators

Estimator	Singly-iterated			Fully-iterated		
	RB	RRMSE	MARE	RB	RRMSE	MARE
M-estimator ($h_k = 1$, $\varphi = 11$)	−6.94%	29.28%	235.07%	−7.93%	29.27%	235.07%
M-estimator($h_k = \tilde{w}_k$, $\varphi = 100$)	−8.14%	25.36%	197.86%	−8.27%	25.33%	196.73%

7. Conclusion

In this paper, we considered robust alternatives to the optimal (BLU) estimator. We first proposed a compromise between the GREG and BLU estimators, the LS estimator, to deal with deviations from the ignorability assumption. The LS estimator is obtained by shrinking the design weights toward their mean. It is expected to be more stable than the GREG estimator when the ignorability assumption holds approximately and less biased than the BLU estimator when this assumption is not fully satisfied. This was confirmed in a simulation study using a population created from real survey data. The LS estimator also offers some protection against deviations from model assumptions.

To deal with outliers, we suggested using the weighted generalized *M*-estimation technique to reduce the influence of units with large weighted population residuals. We found in a simulation study that significant gains in MSE could be obtained with this method. We also found that an *M*-estimator obtained using a single iteration of the IRLS algorithm performed similarly to a fully-iterated *M*-estimator. Finally, we proposed implementing *M*-estimators for multipurpose surveys by modifying either the weights of influential units or their values. We believe that both approaches are useful and contribute to bridge a small gap between theory and practice.

Acknowledgements

The authors would like to sincerely thank the Associate Editor and three referees for their constructive remarks and suggestions. They would also like to thank Cynthia Bocci and Wesley Yung for their comments, which helped improve the clarity of the paper.

Appendix

In this proof, we remove the conditioning on **X** when taking expectations and variances with respect to model *m* in order to simplify the notation. Using Slutsky's theorem, to show that $\mathrm{E}_p(\hat{t}_y^B - t_y)/t_y$ converges in probability to 0, as the sample size *n* and the population size *N* tend to infinity, under assumptions (A1), (A2) and (A3), it suffices to show that:

a) $\mathrm{E}_p(t_y / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}) = t_y / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}$ converges in probability to 1 and

b) $\mathrm{E}_p(\hat{t}_y^B / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta})$ converges in probability to 1.

To show (a), note that

$$\mathrm{E}_m\left(\frac{t_y}{\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}}\right) = 1$$

and

$$\mathrm{V}_m\left(\frac{t_y}{\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}}\right) = \frac{1}{N}\ \frac{1}{(\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}/N)^2}\ \sum_{k\in U}\sigma_k^2 \Big/ N .$$

By Chebychev's inequality, $t_y / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}$ converges in probability to 1 under model *m*, as *N* increases, if $\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta} = O(N)$ and $\sum_{k\in U}\sigma_k^2 = O(N)$ (assumption A3).

To show (b), we first note that $\mathrm{E}_m\mathrm{E}_p(.) = \mathrm{E}_p\mathrm{E}_m(.\,|\,s)$ provided that the set of all possible samples does not depend on which population was generated by model *m*. Consequently, if assumption (A2) holds, it is straightforward to show that $\mathrm{E}_m\mathrm{E}_p(\hat{t}_y^B / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}) = 1$. Then, we note that

$$\mathrm{V}_{mp}\left(\frac{\hat{t}_y^B}{\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}}\right) = \mathrm{V}_m\mathrm{E}_p\left(\frac{\hat{t}_y^B}{\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}}\right) + \mathrm{E}_m\mathrm{V}_p\left(\frac{\hat{t}_y^B}{\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}}\right). \quad \text{(A.1)}$$

As a result, $\mathrm{V}_m\mathrm{E}_p(\hat{t}_y^B / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}) \le \mathrm{V}_{mp}(\hat{t}_y^B / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta})$ since the two terms on the right side of (A.1) are greater than or equal to 0. By the previous inequality and Chebychev's inequality, $\mathrm{E}_p(\hat{t}_y^B / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta})$ converges in probability to 1 under model *m*, as *n* and *N* increase, if $\lim_{n,N\to\infty}\mathrm{V}_{mp}(\hat{t}_y^B / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}) = 0$. Using assumption (A2), it is straightforward to show that

$$\mathrm{V}_{mp}\left(\frac{\hat{t}_y^B}{\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}}\right) = \frac{1}{N}\ \frac{1}{(\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}/N)^2}\ \sum_{k\in U}\mathrm{E}_p\left\{\left(w_k^B\right)^2 I_k\right\}\sigma_k^2 \Big/ N .$$

Consequently, $\lim_{n,N\to\infty}\mathrm{V}_{mp}(\hat{t}_y^B / \mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta}) = 0$ if $\mathbf{t}'_{\mathbf{x}}\boldsymbol{\beta} = O(N)$ and $\sum_{k\in U}\mathrm{E}_p\{(w_k^B)^2 I_k\}\ \sigma_k^2 = O(N)$ (assumption A3). This completes the proof.

References

Beaton, A.E., and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16, 147-185.

Basu, D. (1971). An essay on the logical foundations of survey sampling, part 1. In *Foundations of statistical inference*, (Eds. V.P. Godambe and D.A. Sprott), Toronto: Holt, Rinehart, and Winston, 203-233.

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

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

Chambers, R.L., Dorfman, A.H. and Wehrly, T.E. (1993). Bias robust estimation in finite populations using nonparametric calibration. *Journal of the American Statistical Association*, 88, 268-277.

Deville, J.-C., and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Draper, N., and Smith, H. (1980). *Applied regression analysis*, *second edition*. New-York, John Wiley & Sons, Inc.

Duchesne, P. (1999). Robust calibration estimators. *Survey Methodology*, 25, 43-56.

Dumouchel, W.H., and Duncan, G.J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association*, 78, 535-543.

Elliott, M.R., and Little, R.J.A. (2000). Model-based alternatives to trimming survey weights. *Journal of Official Statistics*, 16, 191-209.

Graubard, B.I., and Korn, E.L. (1993). Hypothesis testing with complex survey data: the use of classical quadratic test statistics with particular reference to regression problems. *Journal of the American Statistical Association*, 88, 629-641.

Gwet, J.-P., and Lee, H. (2000). An evaluation of outlier-resistant procedures in establishment surveys. In *The Second International Conference on Establishment Surveys*, American Statistical Association, Alexandria, Virginia, 707-716.

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., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: the Approach Based on Influence Functions*. New-York, John Wiley & Sons, Inc.

Hedlin, D., Falvey, H., Chambers, R. and Kokic, P. (2001). Does the model matter for GREG estimation? A business survey example. *Journal of Official Statistics*, 17, 527-544.

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

Huber, P.J. (1981). *Robust Statistics*. New-York, John Wiley & Sons, Inc.

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

Hulliger, B. (1999). Simple and robust estimators for sampling. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 54-63.

Kalton, G., and Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, 19, 81-97.

Kish, L. (1992). Weighting for unequal P_i. *Journal of Official Statistics*, 8, 183-200.

Korn, E.L., and Graubard, B.I. (1999). *Analysis of Health Surveys*. New-York, John Wiley & Sons, Inc.

Lee, H. (1991). Model-based estimators that are robust to outliers. In *Proceedings of the Annual Research Conference*, Washington, DC, U.S. Bureau of the Census, 178-202.

Lee, H. (1995). Outliers in business surveys. In *Business Survey Methods*, (Eds. B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge and P.S. Kott). Chapter 26, New-York, John Wiley & Sons, Inc.

Little, R.J.A. (1983). Estimating a finite population mean from unequal probability sampling. *Journal of the American Statistical Association*, 78, 596-604.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61, 317-337.

Potter, F. (1988). Survey of procedures to control extreme sampling weights. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 453-458.

Potter, F. (1990). A study of procedures to identify and trim extreme sampling weights. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 225-230.

Potter, F. (1993). The effect of weight trimming on nonlinear survey estimates. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 758-763.

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

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18, 209-217.

Ren, R., and Chambers, R.L. (2002). Outlier robust imputation of survey data via reverse calibration. Southampton Statistical Sciences Research Institute Methodology Working Paper M03/19, University of Southampton.

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

Royall, R.M., and Herson, J. (1973). Robust estimation in finite populations I. *Journal of the American Statistical Association,* 68, 880-889.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

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

Stokes, L. (1990). A comparison of truncation and shrinking of sampling weights. In *Proceedings of the* 1990 *Annual Research Conference*, Washington, DC: Bureau of the Census, 463-471.

Sugden, R.A., and Smith, T.M.F. (1984). Ignorable and informative designs in survey sampling inference. *Biometrika*, 71, 495-506.

Valliant, R., Dorfman, A. and Royall, R.M. (2000). *Finite population sampling: a prediction approach*. New-York, John Wiley & Sons, Inc.

Welsh, A.H., and Ronchetti, E. (1998). Bias-calibrated estimation from sample surveys containing outliers. *Journal of the Royal Statistical Society*, Series B, 60, 413-428.

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.