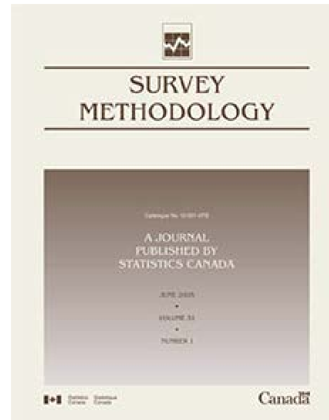
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Criteria for choosing between calibration weighting and survey weighting



by Mohammed El Haj Tirari and Boutaina Hdioud

Release date: December 20, 2018

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. 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.gc.ca under "Contact us" > "Standards of service to the public."

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 co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2018
Vol. 44, No. 2, pp. 339-349
Statistics Canada, Catalogue No. 12-001-X

Criteria for choosing between calibration weighting and survey weighting

Mohammed El Haj Tirari and Boutaina Hdioud[1]

Abstract

Based on auxiliary information, calibration is often used to improve the precision of estimates. However, calibration weighting may not be appropriate for all variables of interest of the survey, particularly those not related to the auxiliary variables used in calibration. In this paper, we propose a criterion to assess, for any variable of interest, the impact of calibration weighting on the precision of the estimated total. This criterion can be used to decide on the weights associated with each survey variable of interest and determine the variables for which calibration weighting is appropriate.

Key Words: Estimation of a total; calibration estimator; superpopulation model; model-based approach; weighting impact.

1 Introduction

When estimating population parameters, adjustment techniques are often used to reduce variance or correct non-response. When there is auxiliary information, calibration is an adjustment technique often used in practice. The weight of the calibration estimator is used to adjust the sample so that it reflects the known population totals for a set of auxiliary variables (Deville and Särndal, 1992). The improved accuracy by the calibration estimator depends on the auxiliary variables used in calibration. The variance of the calibration estimator is low when the calibration variables are strongly linked to the variable of interest.

In practice, once the calibration weights are calculated, they replace the survey weights for the production of parameter estimates of all survey variables of interest. However, using calibration weighting can lead to an increase in the mean square error (MSE) for some variables of interest, particularly those not linked to calibration variables. Therefore, calibration weights cannot be used systematically to estimate population parameters for any variable of interest, particularly in the case of multi-purpose surveys covering different subjects. That is why it is necessary to develop a criterion to assess the impact of calibration weighting on the precision of estimates for each variable of interest.

To develop this type of criterion, we can refer to a comparison of the precision of calibration estimators with the Horvitz-Thompson (HT) estimator. Several inferential approaches can be used to measure the precision of these estimators. In this paper, we will consider a sample design- and model-based approach. This approach was chosen because it is the only one with which we can develop a measurement of the MSE of the calibration estimator in order to account for bias due to the use of calibration weights, as well as variance, which depends on the quality of the model. In other approaches (design-based or model-assisted), it is extremely difficult to calculate the MSE of the calibration estimator, and the estimates do not take into account the bias introduced by the use of calibration weights.

1. Mohammed El Haj Tirari, Institut National de Statistique et d’Économie Appliquée, Rabat, Morocco. E-mail: mtirari@hotmail.fr; Boutaina Hdioud, École Nationale Supérieure d’Informatique et d’Analyse des Systèmes, Rabat, Morocco. E-mail: hdioud.boutaina@hotmail.fr.

Using the design- and model-based approach allows us to develop a criterion with the advantage of approaching a situation where the loss in bias increase for the calibration estimator exceeds the gain in the reduction of variance obtained when there is a link between the variable of interest and the calibration variables. This is a case where the calibration estimator must not be used.

In this paper, we propose a new criterion that measures the impact of using calibration weighting. The proposed criterion takes into account the degree of the existing link between the variable of interest and the calibration variables. Furthermore, it is simple to calculate for each survey variable of interest so that the best sets of weights to use can be identified.

It should be noted that the impact of using calibration weights was studied previously, but only in the context of measuring the design effect (Deff) used to assess the relative increase or decrease in the variance of an estimator compared with simple random sampling. For example, in the model-assisted approach, Henry and Valliant (2015) proposed a Deff measurement that translated the joint impact of an unequal probability sample design and an adjustment of sampling weights compared with simple random sampling.

Following the introduction, which identifies the issue examined in this paper, Section 2 presents the inferential approach adopted in this paper and the criterion used to measure the precision of estimators, while determining its expression for a calibration estimator and an HT estimator. In Section 3, we present the proposed new criterion for assessing the impact of using calibration weights. Section 4 evaluates the proposed criterion using simulations. The purpose of this evaluation is to verify that this criterion identifies situations where a set of calibration weights should be used. In Section 5, we conclude with a discussion of the advantages of the proposed criterion.

2 Estimator of a variable of interest total

$U = \{1, \ldots, N\}$ for a population size N from which sample s of size n is selected based on survey design $p(s)$. S designates a random variable such as $p(s) = P(S = s)$, and π_k and π_{kl} respectively designate the first and second probabilities of inclusion in survey design $p(s)$. We are interested in a variable of interest $Y = (y_1, \ldots, y_k, \ldots, y_N)'$, with the objective of estimating its total $t_y = \sum_{k \in U} y_k$. To do that, we consider the category of linear estimators $\hat{t}_{yw} = \sum_{k \in S} w_{kS} y_k$ where w_{kS} are the weights that can depend on sample S and the auxiliary variables available. The basic weights used are the sampling weights generated by $d_k = 1/\pi_k$. They correspond to the Horvitz-Thompson estimator $\hat{t}_{y\pi}$ (1952).

It is assumed that we have p auxiliary variables $X_1, \ldots, X_p$, for which the values may be represented by vectors $\mathbf{x}_k = (x_{k1}, \ldots, x_{kp})'$ and for which the vector of their totals $t_{\mathbf{x}} = \sum_{k \in U} \mathbf{x}_k$ is known. The category of calibration estimators is defined by $\hat{t}_{yC} = \sum_{k \in S} w_{kS,C} y_k$ where $w_{kS,C}$, referred to as calibration weights, verify the calibration equation given by

$$\sum_{k \in S} w_{kS,C} \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k. \tag{2.1}$$

Calibration helps to reduce the variance of a total estimator, particularly for variables of interest that are linked to the auxiliary variables used in calibration. However, calibration results in an estimator with a bias other than zero. That is why the calibration weights are determined so that they are as close as possible to the sampling weights in order to manage bias.

2.1 Precision of a linear total estimator

In order to measure the precision of a linear total estimator, we will consider the design and model-based approach. In addition to the design distribution, this approach consists of assuming that values $y_1, \ldots, y_k, \ldots, y_N$ for the variable of interest Y are the product of a random vector $(Y_1, \ldots, Y_k, \ldots, Y_N)'$ whose joint probability distribution is given by the *Superpopulation* model ξ defined by:

$$Y_k = \mathbf{x}'_k \boldsymbol{\beta} + \varepsilon_k \tag{2.2}$$

with

$$E_\xi(\varepsilon_k) = 0, \quad \mathrm{Var}_\xi(\varepsilon_k) = \sigma_k^2 \quad \text{and} \quad \mathrm{Cov}_\xi(\varepsilon_k, \varepsilon_l) = 0$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, $\sigma_k^2\ (k \in U)$ are unknown parameters. E_ξ, Var_ξ and Cov_ξ represent respectively the expectation, variance and covariance for the model. Vector estimator $\boldsymbol{\beta}$ for the regression coefficients is produced by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'_S \boldsymbol{\Pi}_S^{-1} \mathbf{V}_S^{-1} \mathbf{X}_S\right)^{-1} \mathbf{X}'_S \boldsymbol{\Pi}_S^{-1} \mathbf{V}_S^{-1} \mathbf{Y}_S$$

where $\mathbf{X}'_S$ is the matrix of $\mathbf{x}'_k$ values for $k \in S$, $\boldsymbol{\Pi}_S = \mathrm{diag}(\pi_k)_{k \in S}$ and $\mathbf{V}_S = \mathrm{diag}(\sigma_k^2)_{k \in S}$. Under the the design and model-based approach, the criterion used to measure the precision of a linear total estimator is

$$\mathrm{MSE}_{p\xi}(\hat{t}_{yw}) = E_p E_\xi (\hat{t}_{yw} - t_y)^2 \tag{2.3}$$

which corresponds to the mean square error (MSE) for the design and model, also referred to as the *anticipated mean square error* (AMSE). This is based on the assumption that the design is not informative. We can then show that the AMSE for linear estimator $\hat{t}_{yw}$ is (Nedyalkova and Tillé, 2008):

$$\mathrm{MSE}_{p\xi}(\hat{t}_{yw}) = E_p \left(\sum_{k \in s} w_{kS} \mathbf{x}'_k \boldsymbol{\beta} - \sum_{k \in U} \mathbf{x}'_k \boldsymbol{\beta} \right)^2 + \sum_{k \in U} \sigma_k^2 \left[\mathrm{var}_p (w_{kS} I_k) + (R_{kS} - 1)^2 \right] \tag{2.4}$$

where

$$R_{kS} = \frac{E(w_{kS} \mid I_k = 1)}{d_k}$$

with $d_k = 1/\pi_k$ (sampling weight) and $I_k = 1$ for $k \in S$ and $I_k = 0$ otherwise. Ratio R_{kS} equals 1 when linear estimator $\hat{t}_{yw}$ is unbiased according to the design.

2.2 AMSE for the calibration estimator

For the calibration estimator, verifying the calibration equation renders it unbiased under the model:

$$E_{\xi}\left(\hat{t}_{yC}-t_{y}\right)=\sum_{k\in S} w_{kS,C}\mathbf{x}_{k}^{\prime}\boldsymbol{\beta}-\sum_{k\in U}\mathbf{x}_{k}^{\prime}\boldsymbol{\beta}=0.$$

Consequently, the AMSE is expressed as:

$$\begin{aligned}\mathrm{MSE}_{p\xi}\left(\hat{t}_{yC}\right) &= \sum_{k\in U}\sigma_{k}^{2}\left[\mathrm{var}_{p}\left(w_{kS,C}I_{k}\right)+\left(R_{k}-1\right)^{2}\right]\\ &= \sum_{k\in U}\sigma_{k}^{2}\left[\frac{V_{k}}{d_{k}}+R_{k}^{2}\left(d_{k}-1\right)+\left(R_{k}-1\right)^{2}\right] \qquad (2.5)\end{aligned}$$

where $V_{k}=\mathrm{var}_{p}\left(w_{kS,C}\mid I_{k}=1\right)$ and $R_{k}=E_{p}\left(w_{kS,C}\mid I_{k}=1\right)/d_{k}$.

Giving

$$\begin{aligned}\mathrm{var}_{p}\left(w_{kS,C}I_{k}\right) &= E_{p}\left[\mathrm{var}_{p}\left(w_{kS,C}I_{k}\mid I_{k}\right)\right]+\mathrm{var}_{p}\left[E_{p}\left(w_{kS,C}I_{k}\mid I_{k}\right)\right]\\ &= \pi_{k}\,\mathrm{var}_{p}\left(w_{kS,C}\mid I_{k}=1\right)+\pi_{k}\left[E_{p}\left(w_{kS,C}\mid I_{k}=1\right)\right]^{2}-\left[E_{p}\left(w_{kS,C}I_{k}\right)\right]^{2}\\ &= \frac{V_{k}}{d_{k}}+R_{k}^{2}\left(d_{k}-1\right). \qquad (2.6)\end{aligned}$$

Note that the expression (2.5) of $\mathrm{MSE}_{p\xi}\left(\hat{t}_{yC}\right)$ makes it possible to underscore the two criteria that determine the accuracy of calibration estimator $\hat{t}_{yC}$. The first corresponds to *Superpopulation model* ξ through its residual variance σ_{k}^{2}, which decreases when the variable of interest and the calibration variables are correlated (variance reduction $\hat{t}_{yC}$). The second criterion is represented by weight ratios R_{k}, which become important when the calibration weights are very different from the sampling weights (bias increase $\hat{t}_{yC}$).

2.3 AMSE for the HT estimator

In order to develop our criterion for choosing between calibration weighting and sample weighting, we need to determine the expression of the AMSE for the HT estimator. Since the latter is unbiased under the design $(R_{kS}=1)$, its AMSE is given by:

$$\begin{aligned}\mathrm{MSE}_{p\xi}\left(\hat{t}_{y\pi}\right) &= \mathrm{var}_{p}\left(\sum_{k\in s} d_{k}\mathbf{x}_{k}^{\prime}\boldsymbol{\beta}\right)+\sum_{k\in U}\sigma_{k}^{2}d_{k}\left(1-\pi_{k}\right)\\ &= \sum_{k\in U}\sum_{l\in U}\left(\pi_{kl}-\pi_{k}\pi_{l}\right)d_{k}\mathbf{x}_{k}^{\prime}\boldsymbol{\beta}d_{l}\mathbf{x}_{l}^{\prime}\boldsymbol{\beta}+\sum_{k\in U}\sigma_{k}^{2}d_{k}\left(1-\pi_{k}\right). \qquad (2.7)\end{aligned}$$

It should be noted that the expression of the AMSE for $\hat{t}_{y\pi}$ depends on probabilities π_{kl}, which are generally unknown and difficult to calculate for unequal probability sampling designs. Several approximations for these probabilities have been proposed in literature, enabling us to obtain several possible estimators for the variance of the HT estimator. However, Matei and Tillé (2005) showed, through a series of simulations, that these estimators are almost equivalent and allow us to effectively estimate the exact expression of the variance under design $\hat{t}_{y\pi}$.

An approximation of $\operatorname{var}_p\left(\sum_{k\in s} d_k \mathbf{x}_k' \boldsymbol{\beta}\right)$ can be obtained by considering the one proposed by Hájek (1981) for the variance of the HT estimator, produced by:

$$V_{\text{Approx}} = \sum_{k\in U} c_k \left(d_k \mathbf{x}_k' \boldsymbol{\beta}\right)^2 - \frac{1}{h}\left(\sum_{k\in U} c_k d_k \mathbf{x}_k' \boldsymbol{\beta}\right)^2 \tag{2.8}$$

where $h = \sum_{k\in U} c_k$ and $c_k = N\pi_k (1-\pi_k)/(N-1)$. The latter is obtained from the following approximation of probabilities π_{kl} (see Deville and Tillé, 2005; Tirari, 2003):

$$\pi_{kl} - \pi_k \pi_l \approx \begin{cases} c_k - \dfrac{c_k^2}{h} & \text{if } k = l \\ -\dfrac{c_k c_l}{h} & \text{if } k \neq l. \end{cases} \tag{2.9}$$

Consequently, the AMSE for $\hat{t}_{y\pi}$ can be approximated by:

$$\widetilde{\text{MSE}}_{p\xi}\left(\hat{t}_{y\pi}\right) = V_{\text{Approx}} + \sum_{k\in U} \sigma_k^2 d_k (1-\pi_k). \tag{2.10}$$

It should be noted that for simple designs, such as Poisson design or simple stratified random design, joint probability can be calculated precisely without the need for an approximation. In the next section, we will be basing calibration and HT estimators on the AMSE to develop a new *measurement* of the impact of using calibration weights.

3 Proposed criterion for measuring the impact of using calibration weights

Calibration weights are used to improve the precision of estimates for survey parameters of interest. This improvement depends largely on how strongly the variable of interest is linked to the calibration variables. To assess the impact of using calibration weights, we can compare the AMSE for estimators $\hat{t}_{yC}$ and $\hat{t}_{y\pi}$ given respectively by (2.5) and (2.10). The impact of using calibration weights can then be measured through the following criterion:

$$\text{Weff} = \frac{\sum_{k\in U} \sigma_k^2 \left[\frac{V_k}{d_k} + R_k^2 (d_k - 1) + (R_k - 1)^2\right]}{V_{\text{Approx}} + \sum_{k\in U} \sigma_k^2 d_k (1-\pi_k)} \tag{3.1}$$

where calibration weights are chosen in cases where the Weff value is less than 1. Note that the Weff expression (3.1) depends on the population and must be estimated. Furthermore, for any $k \in U, V_k$ represents the variance of calibration weight $w_{kS,C}$, considering the s set of samples containing unit k. Variance V_k is generally not zero since the $w_{kS,C}$ weights depend on the calibration variables and the s

sample selected. In order to take variance V_k into account in measuring the impact of using calibration weights $w_{kS,C}$, we propose estimating the quantity

$$V_w = \sum_{k\in U} \sigma_k^2 \frac{V_k}{d_k} \tag{3.2}$$

by

$$\hat{V}_w = \sum_{k\in S} \hat{\sigma}_k^2 \left(w_{kS,C} - d_k\right)^2 \tag{3.3}$$

where $\hat{\sigma}_k^2$ is the White estimator for σ_k^2 defined by $n\hat{\varepsilon}_k^2/(n-p)$ with $\hat{\varepsilon}_k = Y_k - \mathbf{x}_k'\hat{\boldsymbol{\beta}}$. The estimator (3.3) is obtained by replacing V_k by $\left(w_{kS,C} - d_k\right)^2$, which can be viewed as a first-order approximation of V_k. For any unit $k \in U$, the use of calibration produces weight $w_{kS,C}$, which varies from one sample to another, but for which the design-based expectation can be approximated by sampling weight d_k. The simulations discussed in Section 4 show that $\hat{V}_w$ is a good V_w estimator since it helps to deduct an effective estimator of the Weff criterion. The Weff criterion that we propose for choosing between calibration weights $w_{kS,C}$ and sampling weights d_k can be estimated by

$$\widehat{\text{Weff}}_S = \frac{\sum_{k\in S} d_k \hat{\sigma}_k^2 \left[\frac{(w_{kS,C}-d_k)^2}{d_k} + \hat{R}_{kS}^2 \left(d_k - 1\right) + \left(\hat{R}_{kS} - 1\right)^2\right]}{\hat{V}_{\text{Approx},S} + \sum_{k\in S} d_k \hat{\sigma}_k^2 \left(d_k - 1\right)} \tag{3.4}$$

where $\hat{R}_{kS} = w_{kS}/d_k$ and $\hat{V}_{\text{Approx},S}$ is an estimator for $\text{var}_p\left(\sum_{k\in S} d_k \mathbf{x}_k'\boldsymbol{\beta}\right)$ resulting from the approximation (2.8). It is produced by:

$$\hat{V}_{\text{Approx},S} = \sum_{k\in S} \tilde{c}_k \left(d_k \mathbf{x}_k'\hat{\boldsymbol{\beta}}\right)^2 - \frac{1}{\hat{h}}\left(\sum_{k\in S} \tilde{c}_k d_k \mathbf{x}_k'\hat{\boldsymbol{\beta}}\right)^2 \tag{3.5}$$

with $\tilde{c}_k = n(1-\pi_k)/(n-1)$ and $\hat{h} = \sum_{k\in S} \tilde{c}_k$. The proposed $\widehat{\text{Weff}}_S$ criterion has the benefit of considering bias due to the use of calibration weights, through $\hat{R}_{kS}$, as well as the quality of the linear regression model representing the link between the variable of interest and the calibration variables, through variance $\hat{\sigma}_k^2$. For some survey designs, the weighting traditionally used for estimates effectively leads to an unbiased estimator for the design, but it is not necessarily the HT estimator. This is the case, for example, with a two-stage design where the second stage design depends on the sample from the first stage and the weighting used is the product of the sampling weights for each stage. It is important to note that the $\widehat{\text{Weff}}_S$ criterion proposed in this paper is not linked to the HT estimator, since it enables us to compare the calibration estimator with any other estimator using the sampling weights once it is unbiased.

4 Simulation study

In order to evaluate the $\widehat{\text{Weff}}_S$ criterion (3.4), so that we can determine whether to use calibration weights or sampling weights, we conducted a series of simulations using data observed for a population of

5,800 cottage-industry units. We considered six calibration variables, from which several variables of interest Y_i were generated, with consideration for linear regression models, while accounting for the strength of the link between the variables of interest and the calibration variables through the choice of residual variance in the regression models. Furthermore, to study the impact of the heteroskedasticity of the model residuals on the results obtained for criterion $\widehat{\mathrm{Weff}}_S$, we also considered the case where the variables of interest are generated using models with heteroskedastic residuals.

For the purposes of these simulations, we selected 10,000 samples using a simple random sampling design (SRSD), with three sample sizes: 100, 200 and 400 cottage-industry units, to study the impact of the sample size on the results obtained. Across the 10,000 samples selected, we calculated the following indicators:

- $\mathrm{MSE}_{\mathrm{Cal}}$: the AMSE for the calibration estimator, the expression of which is given by (2.5) and where $E\left(w_{kS,C} \mid I_k = 1\right)$ and V_k are determined respectively by the mean and the variance of weights $w_{kS,C}$ considering all of the selected samples containing unit k.
- $\widetilde{\mathrm{MSE}}_{\mathrm{HT}}$: approximation (2.10) of the AMSE for the HT estimator. $\widetilde{\mathrm{MSE}}_{\mathrm{HT}}$ corresponds to $\mathrm{MSE}_{\mathrm{HT}}$ (AMSE (2.7) for the HT estimator) that we were able to calculate in these simulations since the samples were selected using SRSD.
- Weff: the theoretical value of the Weff calculated using (3.1) and defined by the ratio of $\mathrm{MSE}_{\mathrm{Cal}}$ and $\widetilde{\mathrm{MSE}}_{\mathrm{HT}}$.
- $\overline{\widehat{\mathrm{MSE}}}_{\mathrm{Cal}}$: the simulation mean for the $\widehat{\mathrm{MSE}}_{\mathrm{Cal}}$ estimator of $\mathrm{MSE}_{\mathrm{Cal}}$ where

$$\overline{\widehat{\mathrm{MSE}}}_{\mathrm{Cal}} = \frac{1}{10{,}000} \sum_{s=1}^{10{,}000} \left(\sum_{k \in s} d_k \hat{\sigma}_k^2 \left[\frac{\left(w_{ks,C} - d_k\right)^2}{d_k} + \hat{R}_{ks}^2 \left(d_k - 1\right) + \left(\hat{R}_{ks} - 1\right)^2 \right] \right).$$

- $\overline{\widehat{\widetilde{\mathrm{MSE}}}}_{\mathrm{HT}}$: the simulation mean for the $\widehat{\widetilde{\mathrm{MSE}}}_{\mathrm{HT}}$ estimator of $\widetilde{\mathrm{MSE}}_{\mathrm{HT}}$ where

$$\overline{\widehat{\widetilde{\mathrm{MSE}}}}_{\mathrm{HT}} = \frac{1}{10{,}000} \sum_{s=1}^{10{,}000} \left(\hat{V}_{\mathrm{Approx},\, s} + \sum_{k \in s} d_k \hat{\sigma}_k^2 \left(d_k - 1\right) \right).$$

- $\overline{\widehat{\mathrm{Weff}}}$: the simulation mean for the $\widehat{\mathrm{Weff}}_S$ estimator (3.4) of Weff.
- $\mathrm{MSE}\left(\widehat{\mathrm{Weff}}_S\right)$: the MSE of $\widehat{\mathrm{Weff}}_S$ simulations defined by

$$\mathrm{MSE}\left(\widehat{\mathrm{Weff}}_S\right) = \frac{1}{10{,}000} \sum_{s=1}^{10{,}000} \left(\widehat{\mathrm{Weff}}_s - \mathrm{Weff}\right)^2.$$

The simulation results for heteroskedastic regression models are presented in Table 4.1 below, while the results for homoskedastic models are given in Table A.1 in the appendix.

Table 4.1
(*Heteroskedastic populations*): Simulation results for the $\widehat{\text{Weff}}$ criterion, by sample size and degree of the link between the variables of interest and the calibration variables

		Variables of interest					
		Y_1 ($R^2 = 0.01$)	Y_2 ($R^2 = 0.10$)	Y_3 ($R^2 = 0.20$)	Y_4 ($R^2 = 0.50$)	Y_5 ($R^2 = 0.75$)	Y_6 ($R^2 = 0.98$)
$n = 100$	MSE_{Cal} (10^7)	12,301.13	9,334.81	1,860.23	173.61	59.47	3.07
	MSE_{HT} (10^7)	11,285.46	8,643.37	1,841.84	323.46	212.69	160.35
	$\widetilde{\text{MSE}}_{\text{HT}}$ (10^7)	11,285.44	8,643.34	1,841.81	323.43	212.66	160.32
	Weff	1.09	1.08	1.01	0.54	0.28	0.02
	$\overline{\widehat{\text{MSE}}}_{\text{Cal}}$ (10^7)	12,463.22	9,484.87	1,984.51	180.37	62.07	3.21
	$\overline{\widehat{\text{MSE}}}_{\text{HT}}$ (10^7)	11,856.45	9,068.99	1,929.87	330.59	215.13	160.07
	$\overline{\widehat{\text{Weff}}}$	1.08	1.07	1.00	0.55	0.30	0.02
	$\text{MSE}\left(\widehat{\text{Weff}}\right)$	0.030	0.034	0.030	0.02	0.008	0.00005
$n = 200$	MSE_{Cal} (10^7)	5,931.78	4,500.60	905.42	81.86	27.99	1.41
	MSE_{HT} (10^7)	5,543.74	4,245.87	904.76	158.89	104.48	78.77
	$\widetilde{\text{MSE}}_{\text{HT}}$ (10^7)	5,543.72	4,245.85	904.75	158.88	104.46	78.75
	Weff	1.07	1.06	1.00	0.52	0.27	0.02
	$\overline{\widehat{\text{MSE}}}_{\text{Cal}}$ (10^7)	5,770.29	4,382.31	969.57	83.81	28.68	1.48
	$\overline{\widehat{\text{MSE}}}_{\text{HT}}$ (10^7)	5,673.08	4,341.19	924.64	160.71	105.06	78.71
	$\overline{\widehat{\text{Weff}}}$	1.05	1.05	1.01	0.53	0.28	0.02
	$\text{MSE}\left(\widehat{\text{Weff}}\right)$	0.008	0.008	0.007	0.006	0.002	0.00005
$n = 400$	MSE_{Cal} (10^7)	3,847.61	2,919.12	589.97	53.05	18.13	0.94
	MSE_{HT} (10^7)	3,629.83	2,780.03	592.40	104.04	68.41	51.57
	$\widetilde{\text{MSE}}_{\text{HT}}$ (10^7)	3,629.82	2,780.02	592.39	104.03	68.40	51.56
	Weff	1.06	1.05	0.99	0.51	0.27	0.02
	$\overline{\widehat{\text{MSE}}}_{\text{Cal}}$ (10^7)	3,718.79	2,889.81	594.01	53.89	18.44	0.95
	$\overline{\widehat{\text{MSE}}}_{\text{HT}}$ (10^7)	3,687.44	2,821.34	602.39	104.83	68.68	51.60
	$\overline{\widehat{\text{Weff}}}$	1.04	1.04	0.98	0.52	0.27	0.02
	$\text{MSE}\left(\widehat{\text{Weff}}\right)$	0.004	0.005	0.004	0.003	0.001	0.00001

Hence, the simulation results show that the Weff criterion proposed to measure the impact of using calibration weights helps us to identify situations where calibration weighting should not be used, i.e., when the variable of interest is weakly correlated with the calibration variables $(R^2 < 0.20)$. Furthermore, the $\widehat{\text{Weff}}_S$ estimator (3.4) proposed to estimate the Weff criterion proved to be an effective estimator, recording the same performances, regardless of the strength of the link between the variable of interest and the calibration variables. Heteroskedastic residuals for regression models, representing the link between the variable of interest and the calibration variables, had little impact on the performances of the Weff criterion and the $\widehat{\text{Weff}}_S$ estimator. We also noted a lack of impact in using approximation (2.8) for the variance under design $\sum_{k \in S} d_k \mathbf{x}_k' \boldsymbol{\beta}$ since the impact of the deviation between the AMSE for the HT estimator (MSE_{HT}) and its approximation $\widetilde{\text{MSE}}_{\text{HT}}$ (2.10) was negligible in the results for the Weff criterion. This was predictable since the design being considered was a SRSD.

5 Conclusion

In this paper, we have proposed a new criterion for measuring the impact of using calibration weights to estimate the total for a variable of interest. This criterion can be calculated for each variable of interest to determine whether it is better to use a set of calibration weights or sampling weights to estimate the total for the variable. The proposed criterion has the benefit of taking into account the two main aspects that influence the precision of a total estimator: bias due to the use of calibration weights and the quality of the linear regression model that represents the link between the variable of interest and the calibration variables. Therefore, this criterion can be seen as a measurement of the threshold where the gain in the variance obtained with the calibration estimator exceeds the loss in bias due to the use of calibration weights rather than sampling weights. The simulations conducted to evaluate the proposed criterion showed that this criterion does indeed identify, for a given variable of interest, situations where it is best to use calibration weights, i.e., when the variable of interest is sufficiently correlated with the calibration variables.

It is important to note that the role of this criterion is not to introduce a new weighting system to replace calibration weighting or sample weighting. It is used solely to identify which of the two weighting systems would be best to use for a given variable of interest, which is very useful for practitioners, particularly in the case of surveys that cover different subjects, such as omnibus surveys. However, it would be interesting to study the possibility of producing a unique new weighting system for all survey variables, based on this criterion, while taking into account the advantages of both calibration weights and sampling weights. Finally, it should be noted that the proposed criterion requires a linear relationship between the variables of interest and the calibration variables, and the robustness of the criterion is worth investigating.

Acknowledgements

We would like to thank the reviewers for their thorough work, which helped to improve the results presented in this paper.

Appendix

Simulations results for homoskedastic residual models

Table A.1
(*Homoskedastic populations*): Simulation results for the $\widehat{\text{Weff}}$ criterion, by sample size and degree of the link between the variables of interest and the calibration variables

		Variables of interest					
		Y_1 ($R^2 = 0.01$)	Y_2 ($R^2 = 0.10$)	Y_3 ($R^2 = 0.20$)	Y_4 ($R^2 = 0.50$)	Y_5 ($R^2 = 0.75$)	Y_6 ($R^2 = 0.98$)
$n = 100$	MSE_{Cal} (10^7)	30,150.81	9,298.14	1,492.16	177.42	56.54	3.58
	MSE_{HT} (10^7)	27,162.87	8,530.43	1,477.41	326.93	207.72	160.37
	$\widetilde{\text{MSE}}_{HT}$ (10^7)	27,162.82	8,530.40	1,477.39	326.90	207.69	160.34
	Weff	1.11	1.09	1.01	0.54	0.27	0.02
	$\overline{\widehat{\text{MSE}}}_{\text{Cal}}$ (10^7)	31,523.63	9,775.29	1,565.31	192.17	61.49	3.90
	$\overline{\widehat{\text{MSE}}}_{HT}$ (10^7)	29,024.17	9,128.96	1,573.25	338.45	211.87	160.75
	$\overline{\widehat{\text{Weff}}}$	1.09	1.07	1.00	0.58	0.30	0.02
	$\text{MSE}\left(\widehat{\text{Weff}}\right)$	0.020	0.021	0.021	0.016	0.007	0.00008
$n = 200$	MSE_{Cal} (10^7)	14,277.16	4,441.79	732.99	83.44	26.59	1.68
	MSE_{HT} (10^7)	13,343.16	4,190.39	725.75	160.60	102.04	78.78
	$\widetilde{\text{MSE}}_{HT}$ (10^7)	13,343.14	4,190.37	725.73	160.58	102.02	78.77
	Weff	1.07	1.06	1.01	0.52	0.26	0.02
	$\overline{\widehat{\text{MSE}}}_{\text{Cal}}$ (10^7)	14,195.90	4,398.60	753.49	86.72	27.69	1.75
	$\overline{\widehat{\text{MSE}}}_{HT}$ (10^7)	13,795.17	4,336.28	748.77	163.53	102.90	78.84
	$\overline{\widehat{\text{Weff}}}$	1.06	1.05	1.01	0.53	0.27	0.02
	$\text{MSE}\left(\widehat{\text{Weff}}\right)$	0.003	0.003	0.004	0.005	0.002	0.00002
$n = 400$	MSE_{Cal} (10^7)	9,086.04	2,826.00	470.43	53.96	17.20	1.09
	MSE_{HT} (10^7)	8,736.60	2,743.71	475.19	105.15	66.81	51.58
	$\widetilde{\text{MSE}}_{HT}$ (10^7)	8,736.58	2,743.69	475.18	105.14	66.80	51.57
	Weff	1.04	1.03	0.99	0.51	0.26	0.02
	$\overline{\widehat{\text{MSE}}}_{\text{Cal}}$ (10^7)	9,178.88	2,894.26	478.67	55.38	17.65	1.12
	$\overline{\widehat{\text{MSE}}}_{HT}$ (10^7)	8,946.42	2,833.29	485.09	106.41	67.21	51.57
	$\overline{\widehat{\text{Weff}}}$	1.03	1.02	0.98	0.52	0.27	0.02
	$\text{MSE}\left(\widehat{\text{Weff}}\right)$	0.001	0.001	0.002	0.003	0.002	0.00001

References

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

Deville, J.-C., and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128, 411-425.

Hájek, J. (1981). *Sampling from a Finite Population*. New York: Marcel Dekker.

Henry, K.A., and Valliant, R. (2015). A design effect measure for calibration weighting in single-stage samples. *Survey Methodology*, 41, 2, 315-331. Paper available at https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2015002/article/14236-eng.pdf.

Horvitz, D., and Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

Matei, A., and Tillé, Y. (2005). Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *Journal of Official Statistics*, 21(4), 2005, 543-570.

Nedyalkova, D., and Tillé, Y. (2008). Optimal sampling and estimation strategies under linear model. *Biometrika*, 95, 521-537.

Tirari, M.H.T. (2003). Estimation d'un total pour les plans de sondage à taille fixe et équilibrés. Thesis report.