Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

The use of estimating equations to perform a calibration on complex parameters

by Éric Lesage

June 2011

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2011
Vol. 37, No. 1, pp. 103-108
Statistics Canada, Catalogue No. 12-001-X

The use of estimating equations to perform a calibration on complex parameters

Éric Lesage [1]

Abstract

In the calibration method proposed by Deville and Särndal (1992), the calibration equations take only exact estimates of auxiliary variable totals into account. This article examines other parameters besides totals for calibration. Parameters that are considered complex include the ratio, median or variance of auxiliary variables.

Key Words: Calibration; Complex parameter; Estimating equation; Calibration weight.

1. Introduction

In survey statistics, two main approaches are used in the estimation phase: "model-assisted" estimators (such as the regression estimator or the ratio estimator) and calibration estimators (such as the raking ratio), proposed by Deville and Särndal (1992). The two approaches are somewhat similar, as shown by the regression estimator, which is the same as the calibration estimator with the χ^2 distance ("linear" calibration method).

The purpose of this article is to expand the family of calibration estimators. With the current method, calibration can be performed on totals. The idea is to be able to take into account the calibration constraints of complex parameters or statistics such as a ratio, a median or a geometric mean. The reason for doing this is that auxiliary information may consist of a complex statistic rather than totals. For example, a ratio relative to the total population might be known, but not the total in the numerator or denominator.

The issue of complex parameters in calibrations has been discussed in the literature. Särndal (2007) reviewed a number of them, in particular the work of Harms and Duchesne (2006) on the calibration estimation of quantiles, and the work of Krapavickaite and Plikusas (2005) on calibration estimators of certain functions of totals.

The originality of the approach in this article is that it reduces calibration on a complex parameter to calibration on a total for a new *ad hoc* auxiliary variable. The advantage of this approach is that current calibration tools can be used and that there is no need to solve a complex optimization program.

In section 2 of the article, we review how the calibration method works, define calibration on complex parameters and describe simple cases in which calibration on a complex parameter can be reduced to calibration on a total. In section 3, we focus on parameters that can be defined as a solution to an estimating equation (Godambe and Thompson 1986). We introduce the concept of calibration on a complex parameter defined by an estimating equation and show that the resulting calibration equation can be replaced with an equation for calibration on a total.

2. A complex parameter defined as a function of totals

2.1 Review of calibration on totals

Let U be a finite population of size N. The statistical units of the population are indexed by a label k, where $k \in \{1, \ldots, N\}$. A sample s is selected using sample plan $p(s)$. Its size is denoted n and may be random. Let π_k be the probability that k is included in sample s, and let $d_k = 1 / \pi_k$ be its sampling weight.

For any variable z that takes the values z_k for the units in U indexed by k, the sum $t_z = \sum_{k \in U} z_k$ is referred to as the total of z over U.

Let $y^{(1)}, \ldots, y^{(Q)}$ be Q variables of interest, whose values are known only for sample s, and let θ_y be the parameter of interest that is a function of the totals $t_{y^{(1)}}, \ldots, t_{y^{(Q)}}$:

$$\theta_{\mathbf{y}} = f(t_{y^{(1)}}, \ldots, t_{y^{(Q)}}).$$

The estimator of $\theta_{\mathbf{y}}$ is

$$\hat{\theta}_{\mathbf{y},\pi} = f(\hat{t}_{y^{(1)},\pi}, \ldots, \hat{t}_{y^{(Q)},\pi}).$$

It is simply the function $f(\cdot, \ldots, \cdot)$ with totals $t_{y^{(q)}}$ replaced by their Horvitz-Thompson estimator $\hat{t}_{y^{(q)},\pi} = \sum_{k \in s} d_k y_k^{(q)}$ (Särndal, Swensson and Wretman 1992). This estimator can be described as a substitution estimator.

Let $x^{(1)}, \ldots, x^{(P)}$ be P auxiliary variables known on s, and let $t_{x^{(1)}}, \ldots, t_{x^{(P)}}$ be the totals on U for those auxiliary variables, also known. For an individual k, the vector of values taken by the auxiliary variables on k is denoted $\mathbf{x}'_k = (x_k^{(1)}, \ldots, x_k^{(P)})$.

The calibration estimator of $\theta_{\mathbf{y}}$ is

1. Éric Lesage, CREST(ENSAI).and IRMAR(UEB). Ker Lann Campus, F-35172 BRUZ, France. E-mail: eric.lesage@ensai.fr.

$$\hat{\theta}_{\mathbf{y},\mathrm{CAL}} = f(\hat{t}_{y^{(1)},\mathrm{CAL}}, \ldots, \hat{t}_{y^{(Q)},\mathrm{CAL}})$$

with $\hat{t}_{y^{(q)},\mathrm{CAL}} = \sum_{k\in s} w_k\, y_k^{(q)}$, and a series of weights $\{w_k\}_{(k\in s)}$, known as calibration weights (which should be denoted $w_k(s)$, since they depend on the sampling), obtained by solving the following optimization program:

$$\min_{\{w_k\}_{(k\in s)}} \sum_{k\in s} d(w_k, d_k)$$

under constraints

$$\begin{cases} \hat{t}_{x^{(1)},\mathrm{CAL}} = t_{x^{(1)}} \\ \ldots \\ \hat{t}_{x^{(P)},\mathrm{CAL}} = t_{x^{(P)}} \end{cases}$$

$d(\cdot,\cdot)$ is a pseudo-distance, *i.e.*, a function that measures the difference between the calibration weight and the sampling weight (unlike a difference, a pseudo-distance is not necessarily symmetrical on its two arguments). The program is solved with a Lagrangian. When the distance used is the χ^2 distance (*i.e.*, $d(w_k, d_k) = (1/2)(w_k - d_k)^2/d_k$), the solution is $w_k = d_k(1 + \mathbf{x}_k'\lambda)$ (where λ is a P-vector of Lagrange multipliers).

2.2 Calibration on a complex parameter $\eta_{\mathbf{x}}$

Definition 1: *Let* $x^{(1)}, \ldots, x^{(P)}$ *be* P *auxiliary variables known on* s, *and let* $\eta_{\mathbf{x}} = g(t_{x^{(1)}}, \ldots, t_{x^{(P)}})$ *be a complex parameter*, *a function of the totals of those auxiliary variables*, *also known.*

In the case of calibration on the complex parameter $\eta_{\mathbf{x}}$, *the calibration weights are obtained by solving the following optimization program*:

$$\min_{\{w_k\}_{(k\in s)}} \sum_{k\in s} d(w_k, d_k)$$

under constraints

$$\hat{\eta}_{\mathbf{x},\mathrm{CAL}} = g(\hat{t}_{x^{(1)},\mathrm{CAL}}, \ldots, \hat{t}_{x^{(P)},\mathrm{CAL}}) = \eta_{\mathbf{x}}.$$

The totals $t_{x^{(q)}}$ do not have to be known, but the complex parameter $\eta_{\mathbf{x}}$ does.

Consider the example of the ratio

$$R_{\mathbf{x}} = \frac{t_{x^{(1)}}}{t_{x^{(2)}}} = \frac{\sum_{k\in U} x_k^{(1)}}{\sum_{k\in U} x_k^{(2)}}.$$

The calibration estimator of $R_{\mathbf{x}}$ is of the form

$$\hat{R}_{\mathbf{x},\mathrm{CAL}} = \frac{\sum_{k\in s} w_k x_k^{(1)}}{\sum_{k\in s} w_k x_k^{(2)}}.$$

The calibration equation in the case of calibration on a ratio is

$$\hat{R}_{\mathbf{x},\mathrm{CAL}} = \frac{\sum_{k\in s} w_k x_k^{(1)}}{\sum_{k\in s} w_k x_k^{(2)}} = R_{\mathbf{x}}$$

$R_{\mathbf{x}}$ is known auxiliary information, as the total of the auxiliary variables usually is. This scenario may occur when we have proportions that are well known and stable over time, for example, but the specific totals in the numerator and denominator are not known.

We described the case of calibration on a single complex parameter, but it is clearly a simple matter to calibrate on more than one complex parameter. In that case, there are as many constraints as calibration parameters.

2.3 Simple cases where calibration on a complex parameter can be reduced to calibration on a total

It is not easy to determine from the outset whether an equation for calibration on a complex parameter can be written in the form of an equation for calibration on a total. In other words, it is not always a trivial matter to find a "new" auxiliary variable z, associated with the complex parameter, on whose total we can calibrate.

For example, that is quite straightforward for all moments of an auxiliary variable x (it is assumed that under the sampling plan, the population size N can be estimated exactly). If $\mu_{x^m} = N^{-1}\sum_{k\in U} x_k^m$ is auxiliary information, we can simply take $z_k = x_k^m / N$ and calibrate on μ_{x^m}: $\sum_{k\in s} w_k\, x_k^m / N = \mu_{x^m}$.

If we want to calibrate on the variance and the mean of variable x with μ_x and σ_x^2 as auxiliary information, we can use the two new auxiliary variables

$$z_k^{(1)} = \frac{x_k}{N}$$

and

$$z_k^{(2)} = \frac{(x_k - \mu_x)^2}{N}.$$

On the other hand, if we do not know μ_x, but we have σ_x^2 in the auxiliary information and we want to calibrate on that variance, things become more complicated. We can see this if we write the substitution estimator of σ_x^2 (where the sampling plan allows the population size N to be estimated exactly):

$$\hat{\sigma}^2_{x,\mathrm{CAL}} = \frac{1}{N}\sum_{k\in s} w_k \left(x_k - \left(\frac{\sum_{l\in s} w_l x_l}{N} \right) \right)^2.$$

Finding a new auxiliary variable z is not straightforward, since the initial calibration equation is not linear relative to the weight vector. We will return to the variance case in section 3.3 below.

Ratio example

Proposition 1: *Calibration on a ratio is equivalent to calibration on the total of the new auxiliary variable*: $z_k = x_k^{(1)} - R_{\mathbf{x}} x_k^{(2)}$.

The calibration equation is written

$$\hat{t}_{z,\mathrm{CAL}} = t_z = 0.$$

Proof:

$$\hat{t}_{z,\mathrm{CAL}} = t_z$$

$$\Leftrightarrow \sum_{k\in s} w_k (x_k^{(1)} - R_{\mathbf{x}} x_k^{(2)}) = \sum_{k\in U} (x_k^{(1)} - R_{\mathbf{x}} x_k^{(2)})$$

$$\Leftrightarrow \hat{t}_{x^{(1)},\mathrm{CAL}} - R_{\mathbf{x}}\, \hat{t}_{x^{(2)},\mathrm{CAL}} = t_{x^{(1)},\mathrm{CAL}} - R_{\mathbf{x}}\, t_{x^{(2)},\mathrm{CAL}} = 0$$

$$\Leftrightarrow \frac{\hat{t}_{x^{(1)},\mathrm{CAL}}}{\hat{t}_{x^{(2)},\mathrm{CAL}}} = R_{\mathbf{x}}$$

i.e., $\hat{R}_{\mathbf{x},\mathrm{CAL}} = R_{\mathbf{x}}$.

Function of a ratio of linear combinations of totals

Let $\eta_{\mathbf{x}}$ be a complex parameter that is a bijective function of a ratio of linear combinations of totals:

$$\eta_{\mathbf{x}} = h\left(\frac{\alpha' \cdot \mathbf{t}_{\mathbf{x}}}{\beta' \cdot \mathbf{t}_{\mathbf{x}}}\right) \quad (1)$$

with $\alpha' = (\alpha_1, ..., \alpha_P)$ and $\beta' = (\beta_1, ..., \beta_P)$ being vectors of real coefficients of size P, and $\mathbf{t}'_{\mathbf{x}} = (t_{x^{(1)}}, \ldots, t_{x^{(P)}})$.

Proposition 2: *Performing a calibration on complex parameter* $\eta_{\mathbf{x}}$ *defined by function* (1) *is equivalent to calibrating on the total of the new auxiliary variable*:

$$z_k = (\alpha' - h^{-1}(\eta_{\mathbf{x}})\beta') \cdot \mathbf{x}_{\mathbf{k}}$$

with calibration equation

$$\hat{t}_{z,\mathrm{CAL}} = \sum_{k\in s} w_k z_k = t_z = 0.$$

Proof:

$$\hat{\eta}_{\mathbf{x},\mathrm{CAL}} = \eta_{\mathbf{x}} \Leftrightarrow h\left(\frac{\alpha' \cdot \hat{\mathbf{t}}_{\mathbf{x,CAL}}}{\beta' \cdot \hat{\mathbf{t}}_{\mathbf{x,CAL}}}\right) = \eta_{\mathbf{x}}$$

$$\Leftrightarrow \frac{\alpha' \cdot \hat{\mathbf{t}}_{\mathbf{x,CAL}}}{\beta' \cdot \hat{\mathbf{t}}_{\mathbf{x,CAL}}} = h^{-1}(\eta_{\mathbf{x}})$$

$$\Leftrightarrow (\alpha' - h^{-1}(\eta_{\mathbf{x}})\, \beta') \cdot \hat{\mathbf{t}}_{\mathbf{x,CAL}} = 0$$

$$\Leftrightarrow \sum_{k\in s} w_k (\alpha' - h^{-1}(\eta_{\mathbf{x}})\, \beta') \cdot \mathbf{x}_{\mathbf{k}} = 0.$$

Consider the example of the geometric mean:

$$\mu_{\mathrm{Geo},\mathbf{x}} = \left(\prod_{k\in U} x_k\right)^{1/N}.$$

This expression can be rewritten as

$$\mu_{\mathrm{Geo},\mathbf{x}} = \exp\left(\frac{\sum_{k\in U} \ln(x_k)}{\sum_{k\in U} 1}\right).$$

We denote $\mathbf{x}'_k = (x_k^{(1)}, x_k^{(2)}) = (\ln(x_k), 1)$, $\alpha' = (1, 0)$, $\beta' = (0, 1)$ and $h^{-1}(u) = \exp^{-1}(u) = \ln(u)$.

Hence, the new auxiliary variable is

$$z_k = \ln(x_k) - \ln(\mu_{\mathrm{Geo},\mathbf{x}}) \cdot 1.$$

We will see later in the article that the estimating equations method provides another approach to displaying the new auxiliary variable(s) **z**.

3. Parameter defined by an estimating equation

3.1 Estimating with an estimating equation

Certain parameters θ_y are defined, or can be defined, as the solution to an implicit function known as the *estimating equation on U* (Godambe and Thompson 1986), *i.e.*:

$$\sum_{k\in U} \Phi(\theta_{\mathbf{y}}, \mathbf{y}_k) = 0$$

with $\mathbf{y}'_k = (y_k^{(1)}, ..., y_k^{(Q)})$ being the vector of values taken by the variables of interest for individual k.

In this context, an estimator of $\theta_{\mathbf{y}}$ is defined for sample s, denoted $\hat{\theta}_{\mathbf{y},ee,\pi}$, which is the solution of the *estimating equation on s* (see in particular Hidiroglou, Rao and Yung 2002):

$$\sum_{k\in s} d_k \Phi(\hat{\theta}_{\mathbf{y},ee,\pi}, \mathbf{y}_k) = 0.$$

Table 1
Examples of parameters defined by estimating equations on *U*

Parameter	$\Phi(\theta_{\mathbf{y}}, \mathbf{y}_{\mathbf{k}})$	Estimating equation on *U*
mean μ	$(\mathbf{y}_k - \mu)$	$\sum_{k\in U} (y_k - \mu) = 0$
ratio $R = \mu_1 / \mu_2$	$(y_k^{(1)} - R\, y_k^{(2)})$	$\sum_{k\in U} (y_k^{(1)} - R y_k^{(2)}) = 0$
median m	$(1_{y_k \le m} - 1/2)$	$\sum_{k\in U} (1_{y_k \le m} - 1/2) = 0$

Consider also the example of the coefficient of a logistic regression. Let $y^{(1)}$ be a dichotomous variable that takes the values 0 and 1 on U, and let $y^{(2)}$ be a quantitative variable. The value $y_k^{(1)}$ taken by $y^{(1)}$ for unit k is assumed to be an instance of the random variable $Y_k^{(1)}$, which has a Bernoulli distribution

$$\mathfrak{B}\left(1,\ p_k = \frac{1}{1 + \exp(-\beta_0 y_k^{(2)})}\right).$$

We have limited the number of parameters to one, but it would be just as simple to consider the multidimensional case. However, we should provide a definition of the estimating equations that take the case of the vector parameters into account.

The parameter of interest to us is the estimator of β_0, denoted β, calculated on the finite population by the maximum likelihood method. The estimating equation of β on U will be the maximum likelihood equation. The log-likelihood in the case of Bernoulli variables is

$$\mathbf{L}(\beta) = \sum_{k \in U} y_k^{(1)} \ln(p_k) + \sum_{k \in U} (1 - y_k^{(1)}) \ln(1 - p_k).$$

It is easy to derive the estimating equation of β on U:

$$\sum_{k \in U} y_k^{(2)} \left(y_k^{(1)} - \frac{1}{1 + \exp(-\beta y_k^{(2)})} \right) = 0.$$

The estimating equation on s which defines the estimator $\hat{\beta}_{ee,\pi}$ on the basis of the sampling weights is

$$\sum_{k \in s} d_k y_k^{(2)} \left(y_k^{(1)} - \frac{1}{1 + \exp(-\hat{\beta}_{ee,\pi} y_k^{(2)})} \right) = 0.$$

The estimating equation is not linear in the parameter; $\hat{\beta}_{ee,\pi}$ cannot be expressed as a simple function of the observations.

The logistic regression example is very interesting because it shows that we do not need to know $\hat{\beta}_{ee,\pi}$ to perform the calibration. We will see in the next subsection that we only need to know the generic term of the estimating equation on

$$U,\ \Phi(\beta, \mathbf{y}_k) = y_k^{(2)} \left(y_k^{(1)} - \frac{1}{1 + \exp(-\beta y_k^{(2)})} \right),$$

for all $k \in s$.

3.2 Calibration in the case of parameters defined by estimating equations

Let $\mathbf{x}'_k = (x^{(1)}, \ldots, x^{(P)})$ be the vector of P known auxiliary variables on s, and let $\eta_{\mathbf{x}}$ be a complex parameter, also known, defined by the estimating equation

$$\sum_{k \in U} \Psi(\eta_{\mathbf{x}}, \mathbf{x}_k) = 0.$$

Definition 2: *In the case of calibration on the complex parameter* $\eta_{\mathbf{x}}$, *the calibration weights are obtained by solving the following optimization program*:

$$\min_{\{w_k\}_{(k \in s)}} \sum_{k \in s} d(w_k, d_k)$$

under constraints

$$\sum_{k \in s} w_k \Psi(\eta_{\mathbf{x}}, \mathbf{x}_k) = 0.$$

Proposition 3: *Calibration on a complex parameter* $\eta_{\mathbf{x}}$, *defined by an estimating equation*, *is equivalent to a calibration on the total of the new auxiliary variable*: $z_k = \Psi(\eta_{\mathbf{x}}, \mathbf{x}_k)$, *with the calibration constraint* $\sum_{k \in s} w_k z_k = 0$.

Definition 3: *A calibration estimator of the parameter of interest* $\theta_{\mathbf{y}}$, *denoted* $\hat{\theta}_{\mathbf{y},ee,\mathrm{CAL}}$, *is a solution to the estimating equation on s weighted by the calibration weights* $\{w_k\}_{(k \in s)}$:

$$\sum_{k \in s} w_k \Phi(\hat{\theta}_{\mathbf{y},ee,\mathrm{CAL}}, \mathbf{y}_k) = 0.$$

In most cases, the solution to the estimating equation is unique. The median is an example of a parameter for which there may be more than one solution. In this case, the infimum is often used as an estimator.

Proposition 4: *If there is only one solution to the equation* $\sum_{k \in s} w_k \Psi(\hat{\eta}_{\mathbf{x},ee,\mathrm{CAL}}, \mathbf{x}_k) = 0$, *then*

$$\hat{\eta}_{\mathbf{x},ee,\mathrm{CAL}} = \eta_{\mathbf{x}}.$$

Proof: $\eta_{\mathbf{x}}$ *is a solution to the estimating equation that defines* $\hat{\eta}_{\mathbf{x},ee,\mathrm{CAL}}$. *Since there is a unique solution*, *we have* $\hat{\eta}_{\mathbf{x},ee,\mathrm{CAL}} = \eta_{\mathbf{x}}$.

3.3 Calibration on a variance

In this section, we examine calibration on variance σ_x^2, which is a more complicated complex parameter than those discussed above. We will show that when the variance is the only auxiliary information we have, we can perform an approximate calibration that produces calibration weights that have better properties than the sampling weights.

Back to the variance case. The mean μ_x and the variance σ_x^2 on U of auxiliary variable x can be defined by two estimating equations on U:

$$\begin{cases} \sum_{k \in U} (x_k - \mu_x) = 0 & (2) \\ \sum_{k \in U} ((x_k - \mu_x)^2 - \sigma_x^2) = 0. & (3) \end{cases}$$

If we know the two parameters, calibrating on them is easy, since we merely have to calibrate on the totals of the two new auxiliary variables $z^{(1)} = x - \mu_x$ and $z^{(2)} = (x - \mu_x)^2 - \sigma_x^2$.

On the other hand, if we consider the textbook case where the mean μ_x is not known, the parameter σ_x^2 cannot be defined by a unique estimating equation. If we replace μ_x with its explicit definition

$$\mu_x = \frac{\sum_{l \in U} x_l}{\sum_{j \in U} 1}$$

in equation (3), we obtain the equation

$$\sum_{k\in U}\left(\left(x_k-\frac{\sum_{l\in U}x_l}{\sum_{j\in U}1}\right)^2-\sigma_x^2\right)=0,$$

which cannot be written in the form of an estimating equation: $\sum_{k\in U}\Psi(\sigma_x^2, x_k) = 0$.

μ_x thus becomes a nuisance parameter (Binder 1991). To overcome this difficulty, we can replace it in equation (3) with its substitution estimator: $\hat{\mu}_{x,\pi} = \hat{t}_{x,\pi}/\hat{N}_\pi$, with $\hat{N}_\pi = \sum_{k\in s} d_k 1$ being the Horvitz-Thompson estimator of the size of population U. This leads to the "approximate" calibration equation

$$\sum_{k\in s}w_k\left(\left(x_k-\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}\right)^2-\sigma_x^2\right)=0. \quad (4)$$

Proposition 5: *With estimating equation* (4), *calibration on the variance is not perfect*, *and we have*

$$\hat{\sigma}^2_{x,ee,\mathrm{CAL}}=\sigma_x^2-\left(\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}-\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}\right)^2. \quad (5)$$

Proof:

- The "approximate" calibration equation is equation (4).
- The definition of the parameters' calibration estimators:

$$\begin{cases}\sum_{k\in s}w_k(x_k-\hat{\mu}_{x,ee,\mathrm{CAL}}) = 0\\ \sum_{k\in s}w_k((x_k-\hat{\mu}_{x,ee,\mathrm{CAL}})^2-\hat{\sigma}^2_{x,ee,\mathrm{CAL}}) = 0.\end{cases}$$

  This can be rewritten

$$\begin{cases}\hat{\mu}_{x,ee,\mathrm{CAL}}=\dfrac{\sum_{k\in s}w_kx_k}{\sum_{k\in s}w_k}=\dfrac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}\\ \sum_{k\in s}w_k\left(\left(x_k-\dfrac{\hat{t}_{x,CAL}}{\hat{N}_{\mathrm{CAL}}}\right)^2-\hat{\sigma}^2_{x,ee,\mathrm{CAL}}\right)=0.\end{cases}$$

- If we subtract the second estimating equation from the approximate calibration equation, we get

$$\sum_{k\in s}w_k\left(\left(x_k-\frac{\hat{t}_{x,\pi}}{\hat{N}}\right)^2-\left(x_k-\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}\right)^2-\sigma_x^2+\hat{\sigma}^2_{x,ee,\mathrm{CAL}}\right)=0.$$

  Using the identity $a^2-b^2=(a-b)(a+b)$, we have

$$\sum_{k\in s}w_k\left(\left(\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}-\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}\right)\left(2x_k-\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}-\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}\right)\right) - \hat{N}_{\mathrm{CAL}}(\sigma_x^2-\hat{\sigma}^2_{x,ee,\mathrm{CAL}}) = 0$$

$$\left(\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}-\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}\right)\sum_{k\in s}w_k\left(2x_k-\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}-\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}\right) - \hat{N}_{\mathrm{CAL}}(\sigma_x^2-\hat{\sigma}^2_{x,ee,\mathrm{CAL}}) = 0$$

$$\left(\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}-\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}\right)\left(2\hat{t}_{x,\mathrm{CAL}}-\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}\hat{N}_{\mathrm{CAL}}-\hat{t}_{x,\mathrm{CAL}}\right) - \hat{N}_{\mathrm{CAL}}(\sigma_x^2-\hat{\sigma}^2_{x,ee,\mathrm{CAL}}) = 0$$

$$\hat{N}_{\mathrm{CAL}}\left(\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}-\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}\right)^2 - \hat{N}_{\mathrm{CAL}}(\sigma_x^2-\hat{\sigma}^2_{x,ee,\mathrm{CAL}}) = 0.$$

  This is the same as the expression for $\hat{\sigma}^2_{x,ee,\mathrm{CAL}}$ in equation (5).

This result is interesting because, without an exact calibration, we have a calibration estimator of σ_x^2 that is asymptotically more precise than the substitution estimator $\hat{\sigma}^2_{x,\pi}$. That is, if we resort to the asymptotic framework typically used in surveys and employ linearization of complex estimators (Deville 1999), we have

$$\hat{\sigma}^2_{x,\pi}-\sigma_x^2=O_p\left(\frac{1}{\sqrt{n}}\right)$$

and

$$(\hat{\sigma}^2_{x,ee,\mathrm{CAL}}-\sigma_x^2)^{1/2}=\left(\frac{\hat{t}_{x,\pi}}{\hat{N}_\pi}-\frac{\hat{t}_{x,\mathrm{CAL}}}{\hat{N}_{\mathrm{CAL}}}\right)=O_p\left(\frac{1}{\sqrt{n}}\right).$$

This yields

$$\hat{\sigma}^2_{x,ee,\mathrm{CAL}}-\sigma_x^2=O_p\left(\frac{1}{n}\right).$$

4. Conclusion

In this article, we presented a simple method of performing a calibration in cases where the auxiliary information takes the form of a complex parameter. That method is based on the concept of the estimating equation. Its major advantage is that it can be used with current calibration software.

In future research, it would be interesting to determine the practical cases in which the use of complex parameters in the calibration improves the precision of the parameters of interest.

Acknowledgements

The author wishes to thank the journal's associate editor, the reviewers, Guillaume Chauvet, François Coquet and Jean-Claude Deville for their constructive comments on the preliminary versions of this paper.

References

Binder, D.A. (1991). Use of estimating functions for interval estimation from complex surveys. *Proceedings of the survey research methods section*, American Statistical Association, 34-42.

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

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

Godambe, V.P., and Thompson, M.E. (1986). Parameters of superpopulation and survey population: Their relationship and estimation. *International Statistical Review*, 54, 127-138.

Harms, T., and Duchesne, P. (2006). On calibration estimation for quantiles. *Survey Methodology*, 32, 37-52.

Hidiroglou, M., Rao, J.N.K. and Yung, W. (2002). Estimating equations for the analysis of survey data using poststratification information. *Sankhyā*, 64, 2, 364-378.

Krapavickaite, D., and Plikusas, A. (2005). Estimation of ratio in finite population. *Informatica*, 16, 347-364.

Plikusas, A. (2006). Non-linear calibration. Recueil du Colloque sur les méthodes de sondage, Venspils, Latvia. Riga: Central Statistical Bureau of Latvia.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New-York: Springer-Verlag, 162-163.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33, 99-119.