



Catalogue no. 11-522-XIE

**Statistics Canada International Symposium
Series - Proceedings**

**Symposium 2003: Challenges
in Survey Taking for the Next
Decade**

2003



Statistics
Canada

Statistique
Canada

Canada

Proceedings of Statistics Canada's Symposium 2003
Challenges in Survey Taking for the Next Decade

CALMAR 2: A NEW VERSION OF THE CALMAR CALIBRATION ADJUSTMENT PROGRAM

Olivier Sautory¹

ABSTRACT

Calmar 2 is the new version of the Calmar calibration adjustment program. It contains two major developments.

When survey data are collected at different levels (e.g., households and individuals), simultaneous calibration of the samples helps maintain consistency in the statistics produced from the samples.

Where there is total non-response, generalized calibration makes it possible to rewrite the calibration equations with two sets of variables: the actual calibration variables and the non-response explanatory variables. This corrects for non-response even when the variables that explain it are unknown for the sample non-respondents.

KEYWORDS: Calibration; Generalized Calibration; Non-Response; Simultaneous Calibration.

1. THE CALMAR MACROS

1.1 Background

Calmar is a SAS macro program that implements the calibration methods developed by Deville and Särndal (1992). The program adjusts samples, through reweighting of individuals, using auxiliary information available from a number of variables referred to as calibration variables. The weights produced by this method are used to calibrate the sample on known population totals in the case of quantitative variables and on known category frequencies in the case of qualitative variables.

Calmar is an acronym for CALibration on MARgins, an adjustment technique which adjusts the margins (estimated from a sample) of a contingency table of two or more qualitative variables to the known population margins. However, the program is more general than mere "calibration on margins," since it also calibrates on the totals of quantitative variables.

Calmar was developed in 1990 at France's Institut National de la Statistique et des Études Économiques (INSEE), where it is used regularly to adjust survey data. It is also used by many other statistics agencies in France and other countries.

The new version, Calmar 2, developed in 2003, offers the user new resources for performing calibrations and implements the generalized calibration method of handling non-response proposed by Deville (1998).

Calmar can be downloaded from INSEE's Web site (www.insee.fr), and Calmar 2 will also be available on the site sometime in 2004.

1.2 Calmar's calibration methods

It is worth restating the principle underlying the calibration methods implemented by Calmar (see also Deville *et al.*, 1993).

¹ Olivier Sautory, Cepe-Insee, 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France, sautory@ensae.fr

Consider a population U of individuals, from which a probabilistic sample s has been selected. Let Y be a variable of interest, for which we want to estimate the total in the population: $Y = \sum_{k \in U} y_k$.

The usual estimator of Y is the Horvitz-Thompson estimator:

$$\hat{Y}_{HT} = \sum_{k \in s} \frac{1}{\pi_k} y_k = \sum_{k \in s} d_k y_k.$$

Assume that we know the population totals for J auxiliary variables² $X_1 \dots X_j \dots X_J$ available in the sample:

$$X_j = \sum_{k \in U} x_{jk}.$$

We will look for new “calibration weights” w_k that are as close as possible (as determined by a certain distance function) to the initial weights d_k (these are usually the “sampling weights,” equal to the inverses of the probabilities of inclusion π_k). These w_k are calibrated on the totals of the X_j variables; in other words, they verify the **calibration equations**:

$$\forall j = 1 \dots J \quad \sum_{k \in s} w_k x_{jk} = X_j. \quad (1)$$

The solution to this problem is given by $w_k = d_k F(x'_k \lambda)$ where $x'_k = (x_{1k} \dots x_{Jk})$, λ is a vector of J Lagrange multipliers associated with the constraints (1), and F is a function – the calibration function – whose terms depend on the distance function that is used.

Vector λ is determined by the solution to the non-linear system of J equations in J unknowns resulting from the calibration equations:

$$\sum_{k \in s} d_k F(x'_k \lambda) x_k = X.$$

The estimator of the total for a variable of interest will be the “calibrated” estimator $\hat{Y}_w = \sum_{k \in s} w_k y_k$.

The original version of Calmar offered four calibration methods, corresponding to four different distance functions. These methods are characterized by the form of function F :

- the *linear* method: the calibrated estimator is the generalized regression estimator:

$$\hat{Y}_{reg} = \hat{Y}_{HT} + (X - \hat{X}_{HT}) \hat{B}_s, \text{ where } \hat{B}_s = \left(\sum_{k \in s} d_k x_k x'_k \right)^{-1} \left(\sum_{k \in s} d_k x_k y_k \right);$$

- the *exponential* method: where all the calibration variables are qualitative, this is the raking ratio method (Deming and Stephan, 1940);
- the *logit* method: this method provides lower limits L and upper limits U on the weight ratios w_k / d_k ;
- the *truncated linear* method, very similar to the logit method.

² Quantitative variables or indicators associated with the response categories of qualitative variables.

The last two methods are used to control the range of the distribution of weight ratios. The logit method is used more often because it avoids excessively large weights, which can compromise the robustness of the estimates, and excessively small or even negative weights, which can be produced by the linear method.

Precision

All of the Y_w calibrated estimators have the same precision (asymptotically), regardless of the method used: the approximate variance of Y_w is therefore equal to that of the regression estimator \hat{Y}_{reg} :

$$AV(\hat{Y}_w) = \sum_U \sum \Delta_{k\ell} (d_k E_k)(d_\ell E_\ell),$$

where $E_k = y_k - x'_k B$, $B = \left(\sum_{k \in U} x_k x'_k \right)^{-1} \left(\sum_{k \in U} x_k y_k \right)$ and E_k the residual of the regression of Y on the X_j in the population U .

This variance is especially small if the variable of interest Y and calibration variables $X_1 \dots X_j \dots X_J$ are strongly correlated.

A variance estimator is given by $\hat{V}(\hat{Y}_w) = \sum_s \sum \frac{\Delta_{k\ell}}{\pi_{k\ell}} (d_k e_k)(d_\ell e_\ell)$,

where $e_k = y_k - x'_k B_s$, $B_s = \left(\sum_{k \in s} d_k x_k x'_k \right)^{-1} \left(\sum_{k \in s} d_k x_k y_k \right)$ and e_k is the residual of the regression (weighted by the d_k) of Y on the X_j in sample s .

1.3 What's new in Calmar 2

In addition to the four calibration methods mentioned above, Calmar 2 (Le Guennec and Sautory, 2003) offers the following:

- simultaneous calibration for different levels in a survey;
- adjustment for total non-response using generalized calibration.

These two features will be described in detail below.

Calmar 2 offers a solution to the problem of collinearity between calibration variables: it uses generalized inverse matrices to compute the calibration weights. The original version of Calmar produced an error message in such cases.

Calmar 2 also offers a new distance function, the *generalized hyperbolic sine* function, which depends on a parameter α . Like the exponential method, this method always yields positive weights, but the distribution of weights at the high end is narrower. In addition, the method reduces the range of the weight distribution, as do the logit and truncated linear methods, but it does so with only one parameter, α (Roy *et al.*, 2001).

Finally, the program is more user-friendly, especially in two respects:

- users can enter qualitative calibration variables without prior recoding to obtain sequential response categories;

- users have the option of entering parameters interactively using capture screens that guide them in their choices.

2. SIMULTANEOUS CALIBRATIONS

2.1 The problem

In some surveys, data are collected at different levels:

- INSEE's continuing survey of household living conditions includes questions about the household (type of dwelling, number of persons, occupation of the head of the household, etc.), each member of the household (sex, age, occupation, etc.) and usually a specific set of questions for an individual selected at random from the "eligible" members of the household (often those aged 15 and over), referred to as the "Kish individual";
- the French industry ministry's annual business survey contains questions on each firm's overall activities and a section on each of its establishments.

When the survey data are adjusted, either independent calibrations can be performed for the various levels or simultaneous ("combined") calibrations can be carried out. Simultaneous calibration produces the same weights for all members of a household provided they were all surveyed and ensures consistency in the statistics obtained from the various data files. For example, when independent calibrations are performed on the sample of households and on the sample of household members, the number of one-person households estimated from the former sample cannot be expected to match the number of persons belonging to one-person households estimated from the latter sample.

2.2 The method

More generally, the situations described above relate to surveys that involve cluster sampling or multi-stage sampling, where there is auxiliary information about the clusters (or primary units) and the secondary units, and where the survey's variables of interest concern both the clusters (or PUs) and the SUs.

The simultaneous calibration method was proposed by Sautory (1996). It is more general than the method proposed by Lemaître and Dufour (1987). It consists in performing a single calibration at the PU level. Estimates of the totals for the calibration variables defined at the SU level are computed and then used in the PU calibration, which includes both PU and SU variables.

Thus, if X is a calibration variable for the SUs, the estimate $\hat{X}_m = \sum_{k \in m} \pi_{k/m} x_k$ is calculated for each PU m , where $\pi_{k/m}$ denotes the probability of inclusion of SU k when PU m has been selected. Hence, the calibration equation for variable X can be written $\sum_{m \in s_M} w_m \hat{X}_m = X$, where s_M denotes the PU sample.

2.3 An example

Suppose we have a survey in which a sample of households s_M was selected and some data on the sample households were collected. All members of the selected households were surveyed, forming a sample s_I . In addition, an individual k_m (referred to as the Kish individual) was chosen in each selected household m by simple random sampling without replacement among the e_m eligible members of the household (e.g., individuals aged 15 and over) and surveyed with a specific questionnaire.

Note that:

x_m is the vector of known auxiliary variables for each household m in household sample s_M ,

$X = \sum_{m \in U_M} x_m$ is the vector of the totals for these variables, which totals are known for the population of households U_M

$z_{m,i}$ is the vector of known auxiliary variables for each individual i in household m ,

$Z = \sum_{i \in U_I} z_i$ is the vector of the totals for these variables whose totals are known for the population of individuals U_I ,

v_{k_m} is the vector of known auxiliary variables for each Kish individual k_m in household m and

$V = \sum_{i \in U_I^e} v_i$ is the vector of the totals for these variables whose totals are known for the population of eligible individuals U_I^e .

The probabilities of inclusion of households m are denoted π_m , and we let $d_m = 1/\pi_m$. The probabilities of inclusion of individuals (m,i) when household m has been selected are 1. The probability of inclusion of Kish individual k_m when household m has been selected is $1/e_m$.

The method involves performing a single calibration at the household level, calculating for each household m the totals of the calibration variables for individuals $Z_m = \sum_{(m,i) \in men_m} z_{m,i}$ and the estimated totals of the calibration variables for Kish individuals $\hat{V}_m = e_m v_{k_m}$.

The calibration variables vector for household m becomes (x_m, Z_m, \hat{V}_m) , and the totals vector (X, Z, V) . The calibration equations are written as follows:

$$\sum_{m \in s_M} d_m F(x'_m \lambda + Z'_m \mu + \hat{V}'_m \gamma) (x_m, Z_m, \hat{V}_m) = (X, Z, V)$$

λ , μ and γ denote components of the Lagrange multipliers vector.

The solutions $w_m = d_m F(x'_m \lambda + Z'_m \mu + \hat{V}'_m \gamma)$ of these equations are the new household weights. Thus, the weight $w_{m,i}$ assigned to individual i of household m in the sample of individuals is equal to the weight w_m of household m . The weight w_{k_m} assigned to the Kish individual of household m is equal to $e_m w_m$. It can be verified that with these weights, the various samples are correctly calibrated on totals X , Z and V :

$$\begin{aligned} \sum_{i \in s_I} w_{m,i} z_{m,i} &= \sum_{m \in s_M} w_m \left(\sum_{(m,i) \in men_m} z_{m,i} \right) = \sum_{m \in s_M} w_m Z_m = Z \\ \sum_{k_m \in s_K} w_{k_m} v_{k_m} &= \sum_{k_m \in s_K} w_m e_m v_{k_m} = \sum_{k_m \in s_K} w_m \hat{V}_m = V \end{aligned}$$

This method could be used with Calmar (see Caron and Sautory, 2004), but some SAS programming would be required. Calmar 2 performs all the operations necessary to reduce the process to a single calibration. The user must provide the entry tables for the various levels and the totals for the calibration variables.

Estevao and Särndal (2003) compare several calibration methods for two-stage sample designs, including the method described below.

3. GENERALIZED CALIBRATION

3.1 The underlying principle

While calibration is usually presented using functions of distance between weights, Deville (2002, for example) states the calibration equations directly, with calibration functions defined in a very general form:

$F_k : \lambda \in \mathbb{R}^J \rightarrow F_k(\lambda) \in \mathbb{R}$, with $F_k(0) = 1$, where λ is a vector of J adjustment parameters.

The generalized calibration equations are written $\sum_{k \in s} d_k F_k(\lambda) x_k = X$, where x_k , as previously, denotes the vector of the J calibration variables. Solving this system for λ yields the new weights $w_k = d_k F_k(\lambda)$.

Basic result

Let $\text{grad } F_k(0) = z_k$, vectors that will be referred to as “instruments” (see below). We can show that calibrated estimators based on the same instruments and the same calibration variables are all asymptotically equivalent.

We can rewrite the calibration equations $X = \sum_{k \in s} d_k (1 + z'_k \lambda + O\|\lambda\|^2) x_k$.

This yields $X - \hat{X}_{HT} = \left[\sum_{k \in s} d_k x_k z'_k \right] \lambda + \sum_{k \in s} d_k x_k O\|\lambda\|^2$, or

$$\lambda_s = (T'_{sZX})^{-1} (X - \hat{X}_{HT}) + O\|X - \hat{X}_{HT}\|^2,$$

if we let $T_{sZX} = \sum_{k \in s} d_k z_k z'_k$ which is assumed to be of full rank.

A calibrated estimator $\hat{Y}_w = \sum_{k \in s} w_k y_k$ is therefore asymptotically equivalent to

$$\begin{aligned} \sum_{k \in s} d_k \left[1 + z'_k (T'_{sZX})^{-1} (X - \hat{X}_{HT}) \right] y_k &= \hat{Y}_{HT} + (X - \hat{X}_{HT})' T_{sZX}^{-1} \sum_{k \in s} d_k z_k y_k \\ &= \hat{Y}_{HT} + (X - \hat{X}_{HT})' \hat{B}_{sZX} = \hat{Y}_{reg i} \end{aligned}$$

\hat{B}_{sZX} verifies $\sum_{k \in s} d_k z_k y_k = \left(\sum_{k \in s} d_k z_k x'_k \right) \hat{B}_{sZX}$: it is the vector of the coefficients of the instrumental variable regression (weighted by the d_k) of Y on the $X_1 \dots X_j \dots X_J$ variables; the variables that make up the z_k vectors are the “instruments” (for example, see Fuller, 1987). By analogy with the generalized regression estimator, the estimator $\hat{Y}_{reg i}$ is referred to as the instrumental variable regression estimator.

3.2 Standard form of the calibration functions

In practice, calibration functions F_k are generally of the form $F_k(\lambda) = F(z'_k \lambda)$, where z_k is a vector of J variables Z_j known for sample s , and F is a function of \mathbb{R} in \mathbb{R} such that $F(0) = 1$ and $F'(0) = 1$ (and hence $\text{grad } F_k(0) = z_k$).

The calibration equations are $\sum_{k \in s} d_k F(z'_k \lambda) x_k = X$.

When F is a linear function $F(z'_k \lambda) = 1 + z'_k \lambda$, the calibrated estimator is the instrumental variable regression estimator $\hat{Y}_{reg i}$, since we have $\lambda_s = (T'_{sZX})^{-1} (X - \hat{X}_{HT})$.

3.3 Precision

Through proofs similar to the ones used in conventional calibration, we obtain the following results.

The approximate variance of the calibrated estimator can be written $AV(\hat{Y}_w) = \sum_U \sum_{k\ell} \Delta_{k\ell} (d_k E_k)(d_\ell E_\ell)$,

where $E_k = y_k - x'_k B_{ZX}$ and $B_{ZX} = (\sum_{k \in U} z_k x'_k)^{-1} (\sum_{k \in U} z_k y_k)$ is the residual of the regression of Y on the X_j in U , with instrumental variables Z_j .

A variance estimator is given by $\hat{V}(\hat{Y}_w) = \sum_s \sum_{k\ell} \frac{\Delta_{k\ell}}{\pi_{k\ell}} (d_k e_k)(d_\ell e_\ell)$,

where $e_k = y_k - x'_k B_{sZX}$ and $B_{sZX} = (\sum_{k \in s} d_k z_k x'_k)^{-1} (\sum_{k \in s} d_k z_k y_k)$, is the residual of the regression (weighted by the d_k of Y on the X_j in sample s , with instrumental variables Z_j).

4. CALIBRATION IN THE CASE OF TOTAL NON-RESPONSE

4.1 Standard methods of correcting for total non-response

Total non-response is usually accounted for by reweighting the respondent units. Reweighting techniques are based on models of the response mechanism. This mechanism is similar to random selection of a sample r (of size n_r) from sample s . This selection can be viewed as a supplementary phase added to the original sample design, defined by a pseudo sample design denoted $q(r/s)$. Associated with this design are the individual response probabilities $p_k = P(k \in r / k \in s)$.

If these probabilities were known, the total Y for a variable of interest would be estimated without bias by $\hat{Y}_{exp} = \sum_{k \in r} y_k / (\pi_k p_k)$, known as the expansion estimator. In fact, though, the design $q(r/s)$ and therefore the probabilities p_k are unknown. They must therefore be estimated, substituting a model for the response mechanism and using an estimation method (maximum likelihood, moments, etc.).

A logical choice is the Poisson model: $q(r/s) = \prod_{k \in r} p_k \prod_{k \in s \setminus r} (1 - p_k)$. To fully specify this model, we must provide the form of the probabilities p_k . Three conventional models of the non-response mechanism are described below.

Uniform response model

We assume that each individual has the same probability of response: $p_k = p \quad \forall k \in U$. The maximum likelihood method produces the estimate $\hat{p} = n_r / n$ = the observed response rate.

Homogeneous response groups

Population U is split into H groups that are assumed homogeneous with respect to non-response. All individuals in group h have the same probability of response, denoted p_h . The maximum likelihood method produces the estimates $\hat{p}_h = n_{rh} / n_h$, where n_h (resp. n_{rh}) is the number of individuals in group h who are in sample s (resp. sample r). \hat{p}_h is therefore the observed response rate in group h .

Generalized linear model

The probability of response is a function of a vector z_k of non-response explanatory variables Z_j and an unknown parameter β : $p_k = 1/H(z'_k \beta)$, where H is a function defined on \mathbb{R} with values in $[1, +\infty]$ (in principle). To estimate β and therefore the p_k , the Z_j variables must be known for both respondents and non-respondents.

It is possible to use an even more general model of the form $p_k = 1/H_k(\beta)$, where β is a vector of J adjustment parameters and H_k is a function dependent on individual k .

We will now examine various calibration strategies for cases where there is total non-response.

4.2 Calibration after correcting for non-response

Suppose we have corrected for total non-response, for example with one of the methods described above. Thus, we can perform a conventional calibration starting with the weights corrected for non-response $d_k^* = d_k / \hat{p}_k$. The calibration equations are written $\sum_{k \in r} d_k^* F^*(x'_k \lambda) x_k = X$, where F^* is one of the usual calibration functions.

4.3 Direct conventional calibration

Another strategy is to perform a calibration directly, without prior correction for non-response. The calibration equations are $\sum_{k \in r} d_k F(x'_k \lambda) x_k = X$.

If one of the calibration variables is the constant variable equal to 1, or at least a qualitative variable, the d_k can be multiplied by a constant, with no effect on the $w_k = d_k F(x'_k \lambda)$. Consequently, the calibration equations can be

rewritten $\sum_{k \in r} d_k \frac{1}{n_r / n} F(x'_k \lambda) x_k = X$, which shows that this strategy is equivalent to the previous one with a

non-response correction using a uniform response model.

Dupont (1996) compared the two strategies on the basis of theoretical considerations and simulations. His study led to the following findings:

- If the non-response correction is performed by a generalized linear model, where the H function is one of the usual calibration functions F , and if the calibration variables X_j contain the non-response explanatory variables Z_j , then the two strategies produce very similar results.
- Furthermore, if the calibration variables X_j are identical with the non-response explanatory variables Z_j , the following two strategies are equivalent:
 - performing a non-response correction using a generalized linear model with the exponential function as the H function, then performing a calibration using the corrected weights with the exponential function as the calibration function F^* ;
 - performing a direct calibration using the initial weights with the exponential function as the calibration function F .

The same is true if we perform a non-response correction using a homogeneous response group model and then a post-stratification, where the groups and the post-strata are identical. This is equivalent to performing a direct "formal" post-stratification on the respondent sample.

The advantage of direct calibration is that it does not require explicit modelling of the response mechanism. Lundström and Särndal (1999) also studied the properties of direct calibration and in particular, they proposed variance estimators that take sampling variance and non-response variance into account.

4.4 Direct generalized calibration

Let's start with a system of calibration equations on the respondent sample, of the form $\sum_{k \in r} d_k H_k(\beta) x_k = X$. These equations can be interpreted as indicated below.

Let there be a response model of the form $p_k = \frac{1}{H_k(\beta_0)}$, where β_0 denotes the actual value of the model's parameter. The calibration equations can be rewritten as follows, where $\hat{\beta}$ denotes the solution to the system:

$$X = \sum_{k \in r} d_k H_k(\beta_0) \frac{H_k(\hat{\beta})}{H_k(\beta_0)} x_k = \sum_{k \in r} d_k H_k(\beta_0) \frac{H_k(\beta_0 + \lambda)}{H_k(\beta_0)} x_k = \sum_{k \in r} d_k \frac{1}{p_k} F_k(\lambda) x_k$$

with $\hat{\beta} = \beta_0 + \lambda$ and $F_k(\lambda) = \frac{H_k(\beta_0 + \lambda)}{H_k(\beta_0)}$.

Hence, these equations take the form of generalized calibration equations, where the initial weights are the d_k / p_k , i.e., the sampling weights corrected for non-response and the functions F_k , which verify $F_k(0) = 1$, are the calibration functions. The instruments are $z_k^* = \text{grad } F_k(0) = \frac{1}{H_k(\beta_0)} \text{grad } H_k(\beta_0)$. Solving this system is equivalent to performing a non-response correction and a generalized calibration at the same time.

The results in section 3.3 can be used to calculate the precision of the estimators calibrated by this method.

The approximate variance $AV(\hat{Y}_w)$ uses the residuals of the instrumental variable regression in the population $E_k = y_k - x'_k B_{z^*x}$.

The variance estimator $\hat{V}(\hat{Y}_w)$ uses the residuals of the instrumental variable regression in the respondent sample, weighted by the $d_k H_k(\beta_0)$: $e_k = y_k - x'_k B_{rz^*x0}$, where $\sum_{k \in r} d_k H_k(\beta_0) z_k^* (y_k - x'_k B_{rz^*x0}) = 0$. B_{rz^*x0} is the estimator of B_{z^*x} , which would be calculated if the response probabilities $H_k^{-1}(\beta_0)$ were known. These probabilities are unknown because of β_0 ; they are estimated by replacing β_0 with $\hat{\beta}$. The residuals become $e_k = y_k - x'_k \hat{B}_{rz^*x}$ where $\sum_{k \in r} d_k H_k(\hat{\beta}) z_k^* (y_k - x'_k \hat{B}_{rz^*x}) = 0$, which is an instrumental variable regression in sample r , weighted by the calibration weights $w_k = d_k H_k(\hat{\beta})$.

Note: The estimated variance $\hat{V}(\hat{Y}_w)$ is written in the form $Q_1(e_k) + Q_2(e_k)$, where the quadratic form $Q_1(e_k)$ denotes the phase 1 (selection of sample s) variance estimate and $Q_2(e_k)$ denotes the phase 2 ("selection" of sample r) variance estimate.

Case of a generalized linear model

In practice, the functions $H_k(\beta)$ are of the form $H(z'_k \beta)$, where z_k is a vector of non-response explanatory variables Z_j . The calibration equations are

$$\sum_{k \in r} d_k H(z'_k \beta) x_k = X. \quad (E)$$

The instruments are $z_k^* = z_k \frac{H'(z'_k \beta_0)}{H(z'_k \beta_0)}$. These are "estimated" by $z_k \frac{H'(z'_k \hat{\beta})}{H(z'_k \hat{\beta})}$ and are equal to the z_k when H is the exponential function.

Properties of the method

The dissociation in a system of calibration equations (E) between the Z_j non-response explanatory variables and the X_j calibration variables results in a lower non-response bias (courtesy of the Z_j) and a smaller variance (thanks to the X_j).

The method requires that the number of Z_j variables (quantitative variables and indicators of quantitative variable response categories) be equal to the number of X_j (calibration) variables. In addition, the method is effective only if the correlations between the Z_j and the X_j are sufficiently strong.

Unlike the standard non-response adjustment methods, this method works even when the variables that cause the non-response are known only for respondents. In particular, it handles situations where the non-response factors are variables of interest ("non-ignorable" response mechanism).

Calmar 2 makes it possible to use this method, with the H functions being the usual calibration functions. Le Guennec (2004) provides an example of how the method can be applied to survey data.

REFERENCES

- Caron, N. et Sautory, O. (2004). "Calages simultanés pour différentes unités d'une même enquête", *Document de travail Méthodologie statistique n° 0403, INSEE*.
- Deming, W.E. and Stephan, F.F. (1940). "On a least squares adjustment of a sampled frequency table when the exact totals are known", *Annals of Mathematical Statistics*, 11, pp. 427-444.
- Deville, J.-C. and Särndal, C.-E. (1992). "Calibration estimation in survey sampling", *Journal of the American Statistical Association*, 87, n°418, pp. 375-382.
- Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993). "Generalized raking procedures in survey sampling", *Journal of the American Statistical Association*, 88, n°423, pp. 1013-1020.
- Deville, J.-C. (1998). "La correction de la non-réponse par calage ou par échantillonnage équilibré", Actes du colloque de la Société Statistique du Canada, Sherbrooke, Canada.
- Deville, J.-C. (2004). "La correction de la non-réponse par calage généralisé", *Actes des journées de méthodologie statistique, 16 et 17 décembre 2002, INSEE-Méthodes à paraître*.
- Dupont, F. (1996). "Calage et redressement de la non-réponse totale", *Actes des journées de méthodologie statistique, 15 et 16 décembre 1993, INSEE-Méthodes n°56-57-58*.
- Estevao, V. and Särndal, C.-E. (2003). "Calibration estimation in sample surveys: an overview and recent developments", article présenté au Joint Statistical Meetings de l'ASA, San Francisco.
- Fuller, W. (1987). "Measurement Error Models", New York, Wiley.

- Lemaître, G. and Dufour, J. (1987). "An integrated method for weighting persons and families", *Survey Methodology*, 13, pp. 199-207.
- Le Guennec, J. et Sautory, O. (2003). "La macro Calmar2, manuel d'utilisation", document interne INSEE.
- Le Guennec, J. (2004). "Correction de la non-réponse par calage généralisé: une expérimentation", *Actes des journées de méthodologie statistique, 16 et 17 décembre 2002, INSEE-Méthodes à paraître*.
- Lundström, S. and Särndal, C.-E. (1999). "Calibration as a standard method for treatment of nonresponse", *Journal of Official Statistics*, 15, pp. 305-327.
- Roy, G. et Vanheuverzwyn, A. (2001). "Redressement par la macro CALMAR: applications et pistes d'amélioration", in *Traitements des fichiers d'enquête*. Presses Universitaires de Grenoble, pp. 31-46.
- Sautory, O (1996). "Calage sur des échantillons de ménages, d'individus, d'individus-Kish, issus d'une même enquête", communication invitée aux Journées de Statistique de l'ASU, Québec, Canada.