This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Calibration

Contents

General section

1. Summary

Weighting is a statistical technique commonly used and applied in practice to compensate for nonresponse and coverage error. It is also used to make weighted sample estimates conform to known population external totals. In recent years a lot of theoretical work has been done in the area of weighting and there has been a rise in the use of these methods in many statistical surveys conducted by National Statistical Offices around the world. This module describes in detail calibration as a method of adjusting initial weights in surveys based on sampling in order to estimate known population totals of all auxiliary variables perfectly. This method can also be used in surveys as a possible solution for treatment of unit nonresponse and enables gain on efficiency in term of variance when strong correlation between the variable of interest and auxiliary variables exists. It is worth noting that this is one of many weighting methods which can be used in practice. Others include weighting, poststratification, raking, GREG weighting, logistic regression weighting, mixture approach and logit weighting. A review of the weighting method with examples can be found in Kalton and Flores-Cervantes (2003). More information can also be found in "Weighting and Estimation –Main Module".

Calibration estimation, whereby sampling weights are adjusted to reproduce known population totals, is commonly used in survey sampling. The milestone was the article by Deville and Särndal (1992) in which calibration was described in details. Calibration can be treated as an important methodological instrument, especially in large-scale production of statistics. Many national statistical agencies have developed software designed to compute final weights, usually calibrated using auxiliary information available in administrative registers, censuses and other accurate sources. Calibration as a method of weighting has been described in detail in many articles. A full definition of calibration approach was formulated by Särndal (2007). According to Särndal, the calibration approach to estimation for finite populations consists of:

(a) the computation of weights that incorporate specified auxiliary information and are restrained by calibration equation(s);

(b) the use of these weights to compute linearly weighted estimates of totals and other finite population parameters: weight times variable value, summed over a set of observed units;

(c) satisfying an objective of obtaining nearly design unbiased estimates given that nonresponse and other non-sampling errors are absent.

2. General description of the method

We will assume that we are interested in computing the total value of variable Y (see formula 1). Let us assume that the whole population $U = \{1, \dots, N\}$ consists of N elements. From this population we draw, according to a certain sampling scheme, a sample $s \subseteq U$, which consists of n elements. Let π_i denote first order inclusion probability, i.e., $\pi_i = P(i \in s)$ and $d_i = \frac{1}{\pi_i}$ the design weight. Let $\pi_{ij} = P(i, j \in s)$ denote the second-order inclusion probability. We assume that our main goal is to estimate the total value of variable y:

$$\mathrm{Y}=\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{y}_{\mathrm{i}}, \tag{1}$$

where y_i denotes the value of variable y for i-th unit , $i = 1, \dots, N$.

One well known, classical estimator of the total value (1) is the Horvitz-Thompson estimator, which is given by the formula:

$$\hat{Y}_{HT} = \sum_s d_i y_i = \sum_{i=1}^{n} d_i y_i. \tag{2}$$

If, in addition to y_i, auxiliary variables $x_1, \dots, x_k$ are available from the sample and the population totals $\mathbf{X}_j = \sum_{i=1}^{N} x_{ij}, j = 1, \dots, k$ are known, it may occur that:

$$\sum_s d_i x_{ij} = \sum_{i=1}^{n} d_i x_{ij} \neq \mathbf{X}_j \tag{3}$$

where x_{ij} denotes the value of j-th auxiliary variable for the i-th unit.

Let $\mathbf{X}$ denote the known vector of population totals for the vector of auxiliary variables:

$$\mathbf{X} = \left(\sum_{i=1}^{N} x_{i1}, \sum_{i=1}^{N} x_{i2}, \dots, \sum_{i=1}^{N} x_{ik}\right)^T. \tag{4}$$

This vector is often called the vector of calibration totals or calibration benchmarks.

Let w_i denote calibration weight $i = 1, \dots, n$. Our main goal is to look for new weights w_i which are as close as possible to the design weights d_i and satisfy

$$\mathbf{X} = \widetilde{\mathbf{X}} \tag{5}$$

where

$$\widetilde{\mathbf{X}} = \left(\sum_{i=1}^{n} w_i x_{i1}, \sum_{i=1}^{n} w_i x_{i2}, \dots, \sum_{i=1}^{n} w_i x_{\mathrm{ik}}\right)^T. \tag{6}$$

The calibration estimator for totals (1) takes the form

$$\hat{Y}_{cal} = \sum_{i=1}^{n} w_i y_i, \tag{7}$$

and weights w_i fulfill the so called calibration equation given by formula (5).

The process of constructing calibration weights depends on the properly chosen so called distance function G, which measures the difference between initial weights d_i and final weights w_i. This function must satisfy the following regularity conditions:

- $G(\cdot)$ is strictly convex and twice continuously differentiable,
- $G(\cdot) \geq 0$,
- $G(1) = 0$,
- $G'(1) = 0$,
- $G''(1) = 1$.

The calibration problem involves searching for new weights for a given sample *s* which are as close as possible to the initial weights and satisfy calibration equations and possibly the boundary constraints. This problem can be formulated as a non-linear optimisation problem, see Vanderhoeft (2001):

(C1) Minimise the distance:

$$D(\boldsymbol{w},\boldsymbol{d}) = \sum_{i=1}^{n} d_i \mathrm{G}\left(\frac{w_i}{d_i}\right) \rightarrow min \tag{8}$$

(C2) subject to k calibration equations:

$$\sum_{i=1}^{n} w_i x_{ij} = \mathbf{X}_j, \quad j = 1, \dots, k, \tag{9}$$

(C3) subject to boundary constraints:

$$L \leq \frac{w_i}{d_i} \leq U, \text{ where } 0 \leq L \leq 1 \leq U, \; i = 1, \dots, n. \tag{10}$$

The first constraint (C1) says that calibration weights w_i should be as close as possible to initial weights d_i in terms of distance function G, which measures the difference between both weights. It means that the ratio between final weights and initial weights should not be very different from one. In a special situation, where $w_i = d_i$, no correction is required. The second constraint (C2) is fundamental and constitutes the essence of the calibration approach. According to this constraint, calibration weights must perfectly estimate the totals of all auxiliary variables taken into account in the calibration procedure. This means that the totals of all auxiliary variables are estimated with zero variance using calibration weights. The third constraint (C3) is optional and it may be added whenever calibration weights are negative or extreme. In such a situation, the ratio between final and initial weights should be limited to a carefully specified range.

There is also some freedom in choosing the function G, i.e., this function can be chosen conveniently. The following functions are the most commonly used in practice

$$G_1(x) = \frac{1}{2}(x-1)^2, \tag{11}$$

$$G_2(x) = \frac{(x-1)^2}{x}, \tag{12}$$

$$G_3(x) = x(\log x - 1) + 1, \tag{13}$$

$$G_4(x) = 2x - 4\sqrt{x} + 2, \tag{14}$$

$$G_5(x) = \frac{1}{2\alpha}\int_1^x \sinh\left[\alpha\left(t - \frac{1}{t}\right)\right] dt, \tag{15}$$

where α is a positive parameter, which is used to control the degree of dispersion of calibrated weights in relation to initial weights and sinh denotes the hyperbolic sinus function.

In many statistical packages the problem of finding calibration weights is implemented using different G functions. For example, in CALMAR, which is a macro written in 4GL in SAS four distance functions were implemented, i.e.:

- the linear method, which is based on formula (11),
- the raking ratio method, which is based on the distance function given by (13),
- the logit method, which provides lower limits L and upper limits U on the weight ratios w_i/d_i. In this case, the G function can be expressed as follows:

$$G(x) = \left[(x-L)\log\frac{x-L}{1-L} + (U-x)\log\frac{U-x}{U-1}\right]\frac{1}{A}, \tag{16}$$

  where

$$A = \frac{U-L}{(1-L)(U-1)}, \tag{17}$$

- the truncated linear method, which is based on formula (11), but constraints on the weight ratiosw_i/d_i are imposed, i.e., $L \leq \frac{w_i}{d_i} \leq U$.

In CALMAR 2, which is a later version of CALMAR, the distance function (15) is also implemented. The method expressed by the formula (16) and the truncated linear method are used to control the range of weight ratios. They are used when negative or large weights occur, which may happen when the linear method is taken into account.

The linear method is often used in practice because negative or extreme weights do not occur. This is also the fastest procedure, because it does not need an iterative approach to the problem of finding calibration weights. It can be proved that estimators based on this method are equal to *generalised regression* (GREG) *estimators* (Deville and Särndal 1992). More information about GREG estimators can also be found in Cassel, Särndal and Wretman (1976) and in the module "Weighting and Estimation – Generalised Regression Estimator".

Let us assume that the distance function is expressed by the formula (11). In this situation we have:

$$D(\mathbf{w}, \mathbf{d}) = \sum_{i=1}^{n} d_i G\left(\frac{w_i}{d_i}\right) = \sum_{i=1}^{n} d_i \frac{1}{2}\left(\frac{w_i}{d_i} - 1\right)^2 = \frac{1}{2}\sum_{i=1}^{n} \frac{(w_i - d_i)^2}{d_i}. \tag{18}$$

This kind of formula allows us to find calibration weights in an explicit form. We can prove that if the matrix $\sum_{i=1}^{n} d_i \mathbf{x}_i \mathbf{x}_i^T$ is nonsingular, then the solution of the minimisation problem (8), subject to the calibration constraint (9) is a vector of calibration weights $\mathbf{w} = (w_1, \dots, w_n)^T$, whose elements are described by the formula:

$$w_i = d_i + d_i\left(\mathbf{X} - \widehat{\mathbf{X}}\right)^T \left(\sum_{i=1}^{n} d_i \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \mathbf{x}_i, \tag{19}$$

where

$$\widehat{\mathbf{X}} = \left(\sum_{i=1}^{n} d_i x_{i1}, \sum_{i=1}^{n} d_i x_{i2}, \dots, \sum_{i=1}^{n} d_i x_{ik}\right)^T, \tag{20}$$

and

$$\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^T, \tag{21}$$

is the vector consisting of values of all auxiliary variables for the *i*-th unit in the sample $i = 1, \dots, n$.

All of calibrated estimators $\widehat{Y}_{cal}$ have the same asymptotical precision, regardless of the distance function G used. It was proven that the family of calibration estimators $\widehat{Y}_{cal}$ is asymptotically equivalent to the GREG-estimator (see Deville and Särndal, 1992). From this point of view, the variance of any calibration estimator $\widehat{Y}_{cal}$ can be estimated using the following formula for estimating the variance of the GREG estimator (see Deville and Särndal, 1992):

$$\widehat{V}(\widehat{Y}_{cal}) = \sum_{i \in s} \sum_{j \in s} \left(1 - \frac{\pi_i \pi_j}{\pi_{ij}}\right) (w_i e_i)\left(w_j e_j\right) \tag{22}$$

where e_i are residuals, which are calculated from a sample using weighted linear regression of y on calibration variables $x_1, \dots, x_k$, i.e.,

$$e_i = y_i - \mathbf{x}_i' \mathbf{B}_s, \tag{23}$$

$$\mathbf{B}_s = (\sum_{i \in s} w_i \mathbf{x}_i \mathbf{x}_i')^{-1} (\sum_{i \in s} w_i \mathbf{x}_i y_i). \tag{24}$$

For any distance function, as stated above, the variance is similar to that of the generalised regression estimator. This variance is given by residuals of a regression of the target variable y on auxiliary variables $x_1, \dots, x_k$. If the variable of interest is strongly correlated with the auxiliary variables the gain on precision will be noticeable.

3. Preparatory phase

4. Examples – not tool specific

Examples of how to use calibration can be found in a paper written by McCormack and Sautory (2003). Examples relate to the CALMAR/CALMAR2 macro written in 4GL in SAS language. A function with examples for applying calibration in the R environment can be found in Lumley (2012) and in SPSS software in Vanderhoeft (2001, 2002).

We refer to Wallgren and Wallgren (2007) for examples of applying calibration estimators in register-based statistics. In this book the method of determining calibration weights is presented step by step using operations on matrices.

One example of using calibration with the CALMAR2 macro to determine final weights was also described in detail in a section on tools in this module.

5. Examples – tool specific

Presented below is a detailed description of how to use the CALMAR2 macro to determine calibration weights.

We consider an artificial population of enterprises of size N=1000 from which a simple random sample of size n=20 is drawn. Hence design (initial) weights are equal, N/n=1000/20=50. We also consider a numerical variable x_1 (for instance, monthly revenue of enterprise) and one categorical variable x_2 (for instance, enterprise size, i.e., large - L and medium - M). In this example it will only be shown how calibration weights should be computed. We do not take into account the variable of interest y which is not necessary to compute calibration weights and would be necessary to calculate the variance of the estimator. Monthly revenue of enterprise and enterprise size are chosen as auxiliary variables.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Enterprise size	M	M	M	L	L	L	M	M	L	M	M	M	L	M	M	M	M	M	L	M
Monthly revenue	18	14	16	35	30	10	15	23	23	12	18	16	22	15	15	10	18	18	35	16

Source: artificial data set

The weighted sum of variable x_1 is equal to 18950. Number of medium and large enterprises according to this survey is equal to 700 (14 medium enterprises times 50) and 300 (6 large enterprises times 50), respectively. The exact population total of monthly revenue is known and equals to 19000 and the real number of medium and large enterprises are equal to 720 and 280, respectively. We would like to calibrate the design weights in such a way that known auxiliary totals will be reproduced. In other words, we would like to slightly modify the initial weights so that the sum of x_1 based on the new weights is equal to 19000 and weighted sum of medium and large enterprises is equal to 720 and 280, respectively. We will use the CALMAR2 code in SAS to solve this problem. The SAS code for creating the preliminary datasets and recalling the macro CALMAR2 command is given below.

```
/******************Library containing CALMAR*********************/
libname calm 'D:\Calibration';
optionsmstoredsasmstore=calm;

/************Creation of input dataset with drawn units***********/
data sample;
input enterprise $ size $ revenue weight;
cards;
ent01 M         18    50
ent02 M         14    50
ent03 M         16    50
ent04 L         35    50
ent05 L         30    50
ent06 L         10    50
ent07 M          15    50
ent08 M          23    50
ent09 L          23    50
ent10 M          12    50
ent11 M          18    50
ent12 M          16    50
ent13 L          22    50
ent14 M          15    50
ent15 M          15    50
ent16 M          10    50
ent17 M          18    50
ent18 M          18    50
ent19 L          35    50
ent20 M          16    50
;
run;

/*******Creation dataset with known population totals************/
data totals;
inputvar $ n mar1 mar2;
cards;
size 2 280 720
revenue 0 19000 .
;
run;

/*******************Call to CALMAR*******************************/
```

```
%CALMAR2(DATAMEN=sample, POIDS=weight, IDENT=enterprise,
MARMEN=totals, M=1,DATAPOI=wcal, POIDSFIN=cal_weights )

/*******************Printing final result***********************/
procprintdata=wcalnoobs;
run;
```

The following dataset, with the final weights, is printed:

enterprise	cal_weights	enterprise	cal_weights
ent01	52.2750	ent11	52.2750
ent02	50.5821	ent12	51.4286
ent03	51.4286	ent13	45.0443
ent04	50.5462	ent14	51.0054
ent05	48.4301	ent15	51.0054
ent06	39.9657	ent16	48.8893
ent07	51.0054	ent17	52.2750
ent08	54.3911	ent18	52.2750
ent09	45.4675	ent19	50.5462
ent10	49.7357	ent20	51.4286

CALMAR2 changed the design weights so that the weighted total of variable x_1 is equal to 19000 and weighted number of medium and large enterprises is equal to 720 and 280, respectively. In this example a linear method was used (M=1; 1 – linear, 2 – raking ratio, 3 – logit, 4 – truncated linear, 5 – sinus hyperbolic). The macro parameter DATAMEN contains information about input dataset, POIDS contains information about design weights, IDENT contains the name of an identifying variable for the units in the sample dataset, MARMEN stores information about known totals of all auxiliary variables, M is the identifier of the calibration method that was used, DATAPOI is the name of a new dataset which will be created and will contain calibration weights.

6. Glossary

For definitions of terms used in this module, please refer to the separate “Glossary” provided as part of the handbook.

7. References

Andersson, C. and Nordberg, L. (2000), *CLAN – A SAS Program for Computation of Point and Standard Error Estimates in Sample Surveys*. Statistics Sweden.

Casciano, M.C., Giorgi, V., Oropallo, F., andSiesto, G. (2012), Estimation of Structural Business Statistics for Small Firms by Using Administrative Data.*Rivista Di StatisticaUfficiale*, N. 2-3.

Cassel, C.M., Särndal, C-E., andWretman, J. (1976), Some Results on Generalized difference estimation and Generalized Regression Estimation for Finite Populations.*Biometrika***63**, 615–620.

Cassel, C., Lundquist, P., and Selén, J. (2002), *Model-based calibration for survey estimation, with an example from expenditure analysis*. R&D Report, Research-Methods-Development, Statistics Sweden.

Central Statistical Office in Poland (2011),*Incomes and Living Conditions of the Population in Poland (report from the EU-SILC survey of 2009)*. Statistical Information and Elaborations, Warsaw 2011.

Deville, J.-C. andSärndal, C.-E. (1992), Calibration Estimators in Survey Sampling.*Journal of the American Statistical Association***87**, 376–382.

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

Éltetö, Ö. andMihályffi, L. (2002), Household Surveys in Hungary.*Statistics in Transition***5**, 521–540.

Estevao, V., Hidiroglou, M.A., and Särndal, C.-E. (1995), Methodological Principles for a Generalized Estimation System at Statistics Canada.*Journal of Official Statistics***11**, 181–204.

Eurostat (2004), *Description of target variables: Cross-sectional and Longitudinal*. EU-SILC 065/04.

Gambino, J. (1999), *Discussion of Issues in weighting household and business surveys*. Statistics Canada, Household Survey Methods Division.

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

Lumley, T. (2012), *Package survey*.Version 3.28 of this package is available at the websitehttp://cran.r-project.org/web/packages/survey/survey.pdf.

McCormack, K. (2011), *The calibration software CALMAR – What is it?* Central Statistics Office Ireland.

MEETS (2011), *Use of Administrative Data for Business Statistics*. Final Report, Poznań (Poland).

Nieuwenbroek, N. and Boonstra, H.(2001), *Bascula 4.0 Reference Manual*. Technical paper 3554-99-RSM, Statistics Netherlands.

Sautory, O. (2003), Calmar 2: A New Version of the Calmar Calibration Adjustment Program. Program of Statistics Canada's Symposium 2003, Challenges in Survey Taking for the Next Decade.

Särndal, C.-E. andEstevao, V. M. (2000), A Functional Form Approach to Calibration.*Journal of Official Statistics***16**, 379–399.

Särndal, C.-E. andLundström, S. (2005), *Estimation in Surveys with Nonresponse*. John Wiley & Sons.

Särndal, C.-E. (2007),The Calibration Approach in Survey Theory and Practice.*Survey Methodology***33**, 99–119.

Vanderhoeft, C. (2001), *Generalised Calibration at Statistics Belgium. SPSS Module g-CALIB-S and Current Practises*. Statistics Belgium.

Vanderhoeft, C. (2002), *g-Calib Generalised Calibration Under SPSS*. Statistics Belgium.

Wallgren, A. andWallgren, B. (2007), *Register-based Statistics: Administrative Data for Statistical Purposes*. Wiley.

Zhang, L.-C. (1998), Post-Stratification and Calibration – A Synthesis. Statistics Norway, Research Department.

Specific section

8. Purpose of the method

The main purpose of the method is to adjust, using auxiliary variables, initial weights and construct final weights (calibration weights), which estimate perfectly the totals of all auxiliary variables taken into account in the calibration process in such a way that the final weights are as close as possible to the initial weights in terms of the distance function used. One of the main reasons why calibration should be used in survey sampling is the efficiency of estimates, which can be achieved by exploiting external information and can lead to a small variance of estimators which are based on calibration weights. As a result of calibration, potential improvements in the precision of estimates can be expected. Other reasons for using calibration and purposes of this method include, see Gambino (1999):

- balance, which can be understood to mean that following calibration, the sample ''looks'' like the population,
- consistency of estimates – after calibration each unit of the sample has a unique final weight, which ensures consistency in the sense that when weights are applied to auxiliary variables, they will conform to (will be consistent with) known aggregates for the same auxiliary variables, i.e., weighted parts will add up to totals and mutual consistency between estimated tables will be guaranteed,
- convenience and transparency – this is a particularly important purpose of calibration from the user's point of view, since the resulting estimates are easy to interpret and calibration based on known totals is natural and leads to slightly modified design weights, which can reproduce in a transparent way known benchmarks,

and, see Deville and Särndal (1992), Särndal and Lundström (2005), Särndal (2007):

- potential reduction in bias in the presence of nonresponse and coverage error,
- potential improvements to the precision of estimates,
- coherent estimates based on data coming from different sources.

9. Recommended use of the method

1. Missing data are one of the major types of non-random errors in statistical surveys. They produce significantly biased results and can considerably affect the survey quality. As a rule, this problem is evident in all kinds of surveys conducted by statistical offices of many countries where the lack of response to certain survey questions is quite normal, although definitely undesirable from the point of view of estimation. In view of the above, recent years have seen a growing interest in various methods, which are designed to offset the negative effect of missing data. One of these methods is calibration, which is successfully used by statistical offices of many countries and recommended in many articles and books as a method to handle unit nonresponse. For details on how to use calibration as a method of estimation in surveys with missing data, see Särndal and Lundström (2005).

2. The calibration approach should be recommended and taken into account in all surveys based on sampling, because it can help to reduce bias due to unit nonresponse and variance of estimators. When auxiliary data are strongly correlated with variables of interest, calibration can allow an important gain in precision.
3. In many practical situations, especially involving economic surveys, the distribution of target variables is often asymmetric and some units might have extreme values compared to others (outliers). From one point of view a complete elimination of such units could lead to biased estimates. On the other hand, retaining them with their original weight could make the estimators used highly variable. Duchesne (1999) proposes robust calibration estimators in the case of outliers. This approach is an extension obtained by Deville and Särndal (1992) for the class of calibration estimators based on quantile regression technique, which are discussed in detail in this module. The approach could be extremely useful in business surveys, where distribution of variables, such as income or revenue is highly asymmetric. A broad discussion of the problem of outliers and their negative impact on final results can also be found in the module "Weighting and Estimation – Outlier Treatment".

10. Possible disadvantagesof the method

1. For some distance functions it is possible to receive quite large or negative calibration weights, which is very undesirable in terms of estimation. Such cases should be avoided, i.e., weights have to be positive and should lie within specific desirable limits in order to be as close as possible to original design weights. In any case, it is possible to fulfil this requirement by taking into account an appropriate chosen distance function which can exclude negative or large calibration weights while satisfying given calibration equations. For example, the function given by the formula (16) or (11) with constraints on the weight ratios can be a good remedy when large or negative weights occur.
2. When using the distance function, which helps to restrict the range of weights, it should be remembered that as a result of imposing too strong restrictions on calibration weights with respect to initial weights, the algorithm of finding adjusted weights may not converge.
3. The presence of outlying values in the auxiliary variables may produce extreme calibration weights, which differ a lot from original design weights. In such a situation calibration estimators can be highly variable.
4. In the presence of weak auxiliary information calibration may fail and lead to abnormally high or low weights and, as a consequence, can adversely affect the estimation process.
5. In the presence of some categorical auxiliary variables complete cross-classification may lead to small cells and, as a result, abnormal weights are possible.

11. Variants of the method

1. Variants of the method depend on the chosen distance function. All the calibrated estimators are asymptotically equivalent to the calibrated estimator obtained with the linear method. For more details, see Deville and Särndal (1992).

2. Final results depend on the availability and the choice of efficient auxiliary variables which, according to Särndal and Lundström (2005), should explain the response probability, the main study variable and identify the most important domains. If not, calibration may not be effective and may not bring any improvement or give inefficient or implausible estimators.

12. Input data

1. The input data generally corresponds to the information which is available in the sample and the margins known on the level of the population on which calibration will be done. The input data set usually contains some tables. For example, in CALMAR2, which is a macro written in 4GL in SAS, a table with sample data is required. This table should contain some important variables, e.g., initial weights for units in the sample, an identifying variable, values of the auxiliary variables. Another table should contain information with auxiliary variables, their names, the number of categories and associated margins.

13. Logical preconditions

1. Missing values
   1. When one wants to find calibration weights for a domain which is empty in the sample, it is impossible to create new adjusted weights or any linear estimator of the weighted form $\sum_{i=1}^{n} w_i y_i$. However, this can be done using over-weighting methods, e.g., the raking approach. When the problem of nonresponse concerns only some units in the domain, it is possible to apply calibration as a method of reducing bias and high variance of estimators. It can lead to reliable estimation provided that auxiliary information is used efficiently. For details, see Särndal and Lundström (2005).
2. Erroneous values
   1. Standard calibration methods do not take into consideration errors in variables. Possible misspecification of variables or corrections of variables are generally not taken into account. However, it is possible to construct robust calibration estimators, which can be very helpful in the presence of outliers and highly asymmetric distributions of variables under study. It can be especially important in business statistics, since in such surveys distributions are affected by extreme or erroneous values, e.g., monthly income of enterprises. For details, see Duchesne (1999).
3. Other quality related preconditions
   1.
4. Other types of preconditions
   1. When the sample is small, the linear approach to calibration may produce negative weights, which is undesirable; instead, restricted calibration methods based on iterative algorithms should be applied. For example, the function given by the formula (11) with additional constraints on weight ratios (lower and upper bounds) requires an iterative procedure of determining final calibration weights.

14. Tuning parameters

1. Parameters for the convergence of iterative methods used in the context of the calibration approach are: the maximum number of iterations, convergence criterion, choice of the method (distance function), choice of the lower and upper limits of the calibrated weights. Details of tuning parameters and how to establish them are described in detail in many publications; see, e.g., Lumley (2012), Nieuwenbroek and Boonstra (2001), Sautory (2003), Vanderhoeft (2002) and Zhang (1998).

15. Recommended use of the individual variants of the method

1. Since the linear method provides asymptotically the common linear approximation to all calibration estimators, in many cases it would be the best solution, because it does not need any iterative procedures and in this respect is the fastest one. Another reason why the linear method should be used first is the fact that in many surveys calibrated results often differ fairly little from one method to another, see Zhang (1998). It should be also underlined that other methods of calibration are widely used in practice (for instance, raking ratio) and give good results. Anyway when negative or extreme weights occur, other distance functions, which need iterative algorithms should be considered. In such cases, special attention should be paid to the choice of lower and upper limits of calibrated weights. Restricting the range of weights too much may prevent the algorithm of the calibration procedure from converging.

16. Output data

1. An output dataset depends on the program used and usually contains table(s) with the following information: number of iterations, number of negative weights after each iteration, termination criterion, information about the comparison between margins estimated from the sample (initial weights), using calibration weights and real margins in the population, a set of final (calibration) weights, information about the method used, coefficients of vector lambda of Lagrange multipliers after each iteration, ratios of weights (final weights/initial weights), statistics for ratios of weights, histograms with the distribution of initial and final weights, tables of estimates including estimates of standard errors.

17. Properties of the output data

1. The final output usually contains some tables written to separate files in the format compatible with input data sets (e.g., a file with calibrated weights) and information about the whole process of calibration written and exported to an appropriate file, e.g., pdf or html format. In this output one can find information about properties of calibration weights (number of iterations, number of negative weights, etc.). The user should check in detail the quality of estimates based on calibration weights and their knowledge of the investigated phenomenon, standard errors and bias of estimates.

18. Unit of input data suitable for the method

In order to compute calibration weights, information about initial weights and auxiliary variables should be available for all units in the sample (sample level). Unit level data are also necessary to compute variance estimation of the calibration estimator (input for the method).

19. User interaction - not tool specific

1. Select method of calibration (distance function). In the approach which needs limits for calibration weights establish the lower and upper limit of the range for ratio of initial and calibration weights.
2. Choose carefully potential auxiliary variables to be included into the calibration process.
3. Choose the right software and program to perform the process of calibration.
4. Establish tuning parameters (e.g., convergence criteria, number of iterations).
5. After the use of calibration, quality indicators should be checked and verified in order to evaluate the final results (existing negative or extreme weights, distribution of initial and final weights, correlation coefficient between initial and final weights, ratio of initial and final weights).

20. Logging indicators

1. Run time of the application.
2. Number of iterations to reach convergence in the calibration process.
3. Characteristics of input and output data.

21. Quality indicators of the output data

1. Information about negative or extreme calibration weights.
2. Tables of estimates including estimates of standard errors.
3. Basic statistics for ratios of weights (final weights/initial weights), e.g., mean, median, mode, standard deviation, variance, range, quantiles, interquartile interval.
4. Basic statistics for final weights, e.g., mean, median, mode, standard deviation, variance, range, quantiles, interquartile interval.
5. Histogram of distribution of initial and final weights.
6. Coefficient of correlation between initial and final weights.
7. Tables with margins estimated from the sample (initial weights), margins estimated using calibration weights and real margins in the population.

22. Actual use of the method

1. Calibration as a method of weighting is used by many statistical offices in many surveys. For instance, the Central Statistical Office in Sweden uses calibration in The Survey on Life and Health. This method was also used in Swedish household budget surveys to estimate average consumer expenditures. For details, see Särndal and Lundström (2005), Cassel, Lundquist and Selén (2002). The Hungarian Central Statistical Office (HCSO) adopted this approach in its Household Budget Survey in 1994 and in the Labour Survey in 1995 to compensate for nonresponse and for coverage deficiencies. HCSO uses this method in the form of the so called generalised iterative scaling (raking). For details, see Éltetö and Mihályffi (2002).

2. In Poland the calibration approach is also used by the Central Statistical Office. For instance, the surveys which make use of calibration to compensate for the high percentage of nonresponse are the European Survey on Income and Living Conditions (EU-SILC) and the National Census of Population and Housing 2011. For details, see the Central Statistical Office in Poland (2011).
3. It is also worth noting that in many surveys calibration as a method of weighting and adjusting initial weights in order to reconstruct the known totals of auxiliary variables is recommended by Eurostat. This recommendation concerns primarily the European Union Survey on Income and Living Conditions (EU-SILC), where Eurostat recommends the method of integrated calibration. The idea of this approach is to use auxiliary variables defined at both household and individuals levels in such a way as to ensure consistency between households and individual estimates. After calibration households members will have the same household cross-sectional weight as the personal cross-sectional weight. This approach is used by many statistical offices in practice in EU-SILC. For details see Eurostat (2004).
4. In business statistics calibration is also used in practice. This method was used, for instance, by ISTAT, in the survey of Structural Business Statistics for small-medium enterprises. For more details see Casciano, Giorgi, Oropallo and Siesto (2012). Calibration was also used as a weighting technique for the Structural Business Survey on enterprises at Statistics Belgium. For details, see Vanderhoeft (2001). As a method of treating nonresponse, calibration was used in the MEETS project in a simulation study aimed at checking how it could improve the process of estimation for business data. For details, see MEETS (2011).

Interconnections with other modules

23. Themes that refer explicitly to this module

1. Weighting and Estimation – Main Module
2. Weighting and Estimation – Design of Estimation – Some Practical Issues
3. Weighting and Estimation – Small Area Estimation

24. Related methods described in other modules

1. Weighting and Estimation – Generalised Regression Estimator
2. Weighting and Estimation – Outlier Treatment

25. Mathematical techniques used by the method described in this module

1. Advanced knowledge of linear algebra (including operations on matrices) and differential calculus is required. To find calibration weights the method of Lagrange multipliers is required. In many cases, when calibration weights should be bounded, optimisation algorithms, e.g., the Newton-Raphson approach should be used.

26. GSBPM phases where the method described in this module is used

1. 5.6 Calculate weights

2. 5.7 Calculate aggregates

27. Tools that implement the method described in this module

Calibration as a method of weighting is implemented in many statistical programs and described in details in many articles. Presented below is a short description of the most popular software devoted to calibration and an example of how to use CALMAR2 to determine calibration weights.

1. **Bascula 4.0** – the statistical tool developed in the Delphi language by Statistics Netherlands for the calculation of estimates of population totals, means and ratios. This program uses the so called Balanced Repeated Replication method to adjust weights and Taylor series methods for variance estimation. For details, see Nieuwenbroek and Boonstra (2001).

2. **Caljack** – this is a SAS macro written and developed by Statistics Canada and is an extension of the Calmar macro. This macro provides all the calibration methods which are available in Calmar and is able to calculate variance for many statistics like totals, ratios etc.

3. **CALMAR/CALMAR 2** – the statistical software developed by INSEE. Calmar is a SAS macro program that implements the calibration approach and adjusts weights assigned to individuals using auxiliary variables. Calmar 2 is the newest version of this software and was developed in France in 2003. It implements the generalised calibration method of handling nonresponse. For details, see Sautory (2003).

4. **CALWGT** – this is a freely distributed program for calibration written by Li-Chun Zhang in S-plus for Unix. The user is given the possibility to choose one of the methods, i.e., linear or multiplicative with many options (unrestricted, truncated or restricted approach etc.).

5. **CLAN 97** – the statistical software designed to handle surveys in Statistics Sweden. This is a SAS program (macro) written in 4GL language which is designed to compute point and standard error of estimates in sample surveys. For details, see Andersson and Nordberg (2000).

6. **G-Calib 2** – the statistical software developed in the SPSS language by Statistics Belgium. For details on how to implement this program, see Vanderhoeft (2002).

7. **GES** – this is a SAS-based application with a Windows-like interface which was developed in SAS/AF by Statistics Canada. Details related to GES can be found in Estevao, Hidiroglou and Särndal (1995).

8. **R** – this is a free statistical software. The *calibrate* function, which can be found in the *survey* package, reweights the survey design weights and also adds additional information about estimated standard errors. For details, see Lumley (2012).

9. **ReGenesees System** – ReGenesees (R evolved Generalised software for sampling estimates and errors in surveys) – this is an R-based, full-fledged software system for design-based and model-assisted analysis of complex sample surveys with a user friendly interface which is very required especially by non R users. For details see web page https://joinup.ec.europa.eu/software/regenesees/description.

28. Process step performed by the method

Calibration of weights and estimation of parameters

Administrative section

29. Module code

Weighting and Estimation-M-Calibration

30. Version history

Version	Date	Description of changes	Author	Institute
0.1	30-06-2012	first version	Marcin Szymkowiak	GUS (Poland)
0.2	04-12-2012	second version	Marcin Szymkowiak	GUS (Poland)
0.3	18-03-2014	third version	Marcin Szymkowiak	GUS (Poland)
0.3.1	19-03-2014	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

31. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	26-3-2014 13:32