**The Calibration of Weights Using Calmar2 and Calif in the Practice of the Statistical Office of the Slovak Republic**

Helena Glaser-Opitzová, Ľudmila Ivančíková, Boris Frankovič

Statistical Office of the Slovak Republic

**Abstract**

The application of calibrated weights in the statistical practice results from the requirement of higher accuracy in estimates. In the conditions of the Statistical Office of the Slovak Republic, the implementation of the calibration of weights is associated, a. o., also with the EU SILC sample survey. For the period since 2005, we have gradually moved from intuitive methods to the use of sophisticated products – namely, to the SAS macro CALMAR2. The calculation of calibrated weights implemented across the whole system of sample surveys has encouraged the development of other methods of the calibration. The main reason are the costs. The commercial SAS with the IML module, which are the prerequisites of CALMAR2, are being replaced by the freely available program R. In this paper, we present the algorithmization of the tool with the working title of Calif, which, in comparison with CALMAR2, works in the full GUI environment, considers the stratification, and, most importantly, allows finding an approximate solution. At the same time, we analyze the results of the calibration of weights using both of the tools on the data from the Household Budget Survey and EU-SILC.

**Key words**

calibration, weights, survey, EU SILC, HBS

**1. Introduction**

The main scope of statistical offices is to undertake statistical surveys. In most cases, parameters derived from them are just the estimates of real values. Sampling weights play a crucial role, enabling outcomes of the whole population without a knowledge about it. However, some auxiliary variables, at least their total values, are often known for the whole population and these are a part of the survey design. An inferential step is then beneficial. The idea is to modify the sampling weights so that the population totals of auxiliary variables match exactly to those inferred using „new“ weights and this modification is minimal. This technique proposed by Devill and Särndal [[1]](#footnote-1)[1] is called calibration. As [[2]](#footnote-2)[2] states, “Calibration is a procedure that can be used to incorporate auxiliary data. This procedure adjusts the sampling weights by multipliers known as calibration factors, that make the estimates agree with known totals. The resulting weights are called calibration weights. These calibration weights will generally result in estimates that are design consistent, and that have a smaller variance than the Horvitz-Thompson estimator.” The main advantage of calibration is then to enhance estimates precision, especially when auxiliary variables are correlated with the study variable. The calibration brings consistency to the weight system, so that the population totals throughout the several surveys agree with each other and the additional improved accuracy could be achieved (via lower variance and reduced nonresponse bias).

**2. Calibration estimator**

Let us consider a population with units. The probability sampling of size is undertaken. Every unit in has its own design sampling weight and it is equal to where is the inclusion probability of unit . The objective is to estimate the population total of the study variable . The basic estimator is the Horwitz-Thompson unbiased estimator . However, when auxiliary information is available, it is rather inefficient.

Assume auxiliary variables either as population totals or value vectors known for every . The former case is usual in statistical production when totals are known from administrative sources and censuses or estimated totals from more precise surveys. The latter gives a very effective advantage but unfortunately is less common. Knowledge of every value of makes it possible to construct a derived auxiliary vector that correlates more with the study variable , which certainly enhances the accuracy of estimates. Even totals like could be taken into account, if appropriate.

Two approaches are usually considered. The first one is *generalized regression estimator (GREG)* and is probably more widespread among academics [2]. Its principle is a prediction of values, , for all population elements via an assisting model and auxiliary information (either population totals or known for every ). Predicted values serve to make a nearly design unbiased estimator of the population total

|  |  |  |
| --- | --- | --- |
|  |  | (2.1) |

as stated in [[3]](#footnote-3)[2].Often used in NSIs is *calibration approach*. Its main objective is, as was mentioned above, to reproduce new weights for each that confirm auxiliary totals and differ minimally from design weights . These weights are independent of , totals of many study variables could be estimated. In contrast to GREG, calibration approach doesn’t rely on a specific model; it only operates with information to calibrate on. As [[4]](#footnote-4)[3] point out, “The calibration approach has gained popularity in real applications because the resulting estimates are easy to interpret and to motivate, relying, as they do, on design weights and natural calibration constraints.” As [2] notes, “calibration approach seems to be transparent and natural, since the design weights are just slightly modified and unbiasedness only negligibly disturbed. Under this approach a unique weighting system is given, applicable to all study variables”.

It is obvious that

Let denote the calibration weight of element . The calibration estimator of total is

|  |  |  |
| --- | --- | --- |
|  |  | (2.2) |

so that constraints are fulfilled

for all . According to [2], comparing calibration estimator with unbiased Horwitz-Thompson estimator implies

so the objective of near design unbiasedness requires . Following calibration aims, small values for are sought after. The distance between design and calibration weights is expressed via *distance function*. Let denote the quotient of these weights. Then the distance function is a nonnegative convex function of such that , , . As stated in [[5]](#footnote-5)[4], to find calibration weights we have to find a minimum of the equation

where , , , is a vector of Lagrange multipliers and is a matrix of auxiliary variables. More precisely,

By taking partial derivatives of we get

|  |  |  |
| --- | --- | --- |
|  | whereas | (2.3) |

where is the inverse function to derivative of . This gives

|  |  |  |
| --- | --- | --- |
|  |  | (2.4) |

This system can be solved by several optimization methods taking as starting values.

**3. Distance functions**

Several functions are used among software tools for measuring the distance between design and calibration weights. We consider 4 of them here that are most frequently used.

* *linear* – analogical to linear GREG estimator. This function is often used due to its ability to find exact solution of (2.3) or (2.4) (if the solution exists). If no solution is found it is worthless to try other functions. On the other hand, resulting weights could be negative, which seems to be inconvenient for statistical production purposes. The function itself is defined as
* *raking ratio* – nonlinear distance function that circumvents the „negative weights“ problem. Not to be so optimistic, also raking ratio brings some difficulties, because weights less than 1 could appear.
* *logit* – bounded version of raking ratio. User is able to enter lower and upper bounds for quotient , differences between design and calibration weights as well as the condition that weights are not less than 1 can be controlled. It gives

User must be aware of range allowed for calibration weights, tense bounds often lead to unsolvable system. The goal is to seek an appropriate balance between distance applied and precision of The function is defined as

* *bounded linear* – as the name invokes, it is the bounded version of the linear method. User has to specify the lower and upper bounds for

These functions proposed by [[6]](#footnote-6)[5] are discussed in more detail in [[7]](#footnote-7)[4].

**4. Calif**

Several software tools deal with calibration. They differ in options and environment where they run. Most used are:

* *CALMAR2* – SAS macro developed by INSEE (2000)
* *g-CALIB 2* – written in the SPSS language by Statistics Belgium
* *GES* – SAS based application with interface, developed by Statistics Canada
* *Bascula* – tool developed by Statistics Netherlands in the Delphi language
* *Caljack* – extension of SAS macro CALMAR made by Statistics Canada
* *CALWGT* – free program written by Li-Chun Zhang in S-Plus for Unix
* *CLAN97* – developed by Statistics Sweden
* *calibrate* – function that is a part of *survey* package developed in the R language
* *calib* – function that is a part of *sampling* package, also runs under R

Macro CALMAR2 has been used by Statistical Office of the Slovak Republic (SO SR) for calibration for past years but it has met financial difficulties. SAS/IML is necessary to run CALMAR2 in SAS. Some cheaper and easy-to-use solution has been sought after. R [[8]](#footnote-8)[6] seems to be most eligible environment for statistical computing considering its licence, abilities and online help. SO SR has prepared a free R code called *Calif* that combines various calibration aspects and offers a user-friendly graphical user interface.

The package *fgui* [[9]](#footnote-9)[7] was used for creating the GUI. Calif covers all four distance functions mentioned in section 3. To solve calibration equations (2.4) or (2.3), various optimization functions are implemented. Functions *BBsolve* and *dfsane* from *BB* package [[10]](#footnote-10)[8] and the function *nleqslv* from *nleqslv* package [[11]](#footnote-11)[9] are able to solve nonlinear systems of equations. The former package uses Barzilai-Borwein spectral methods, while the latter relies on Broyden or Newton method. To make Calif more utilizable, function *calib* from *sampling* package [[12]](#footnote-12)[10] is incorporated, which calculates the Moore-Penrose generalized matrix inverses.

In figure 4.1 the main Calif window is shown. Let’s make a short description of particular items. Data for calibration as well as auxiliary totals can be loaded into Calif in .csv and .txt format. Columns of table of totals refer to separate auxiliary variables in data. In case of categorical variable, a total for each category has to be defined in the pertaining column. Rows of the table of totals refer to separate strata. The first column consists of their identification number.

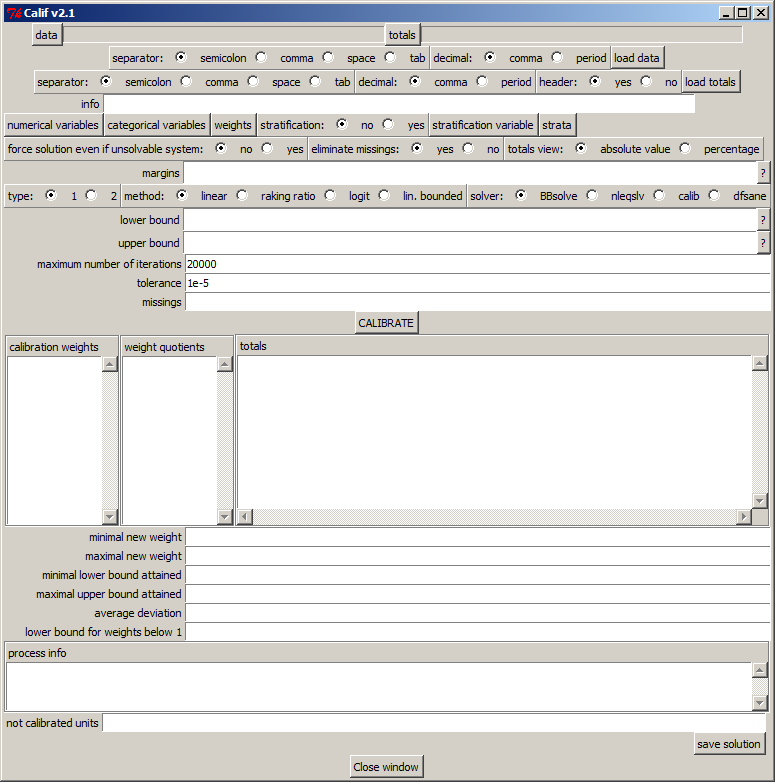


Figure 4.1. *The main Calif window*

If no strata is present in the data, auxiliary population totals can be entered either by a table of totals (with values in either one row or one column) or directly in the „margins“ entry in the main window. Unlike the table of totals, data structure is completely free, the only necessity is to describe it via the graphical user interface. By pressing the „numerical variables“ button, the list of all variables indicated in the data appears. User is then free to mark variables that have to be deemed as numerical. Analogical procedure is applied for categorical variables, design weights column and, if stratification takes part, also for marking variable indicating classification of strata. Calibration can be processed in selected strata first or in the whole dataset at once. Several ways how to find a feasible solution are implemented in Calif, which make it a promising tool. Option „type“ reflects the equations (2.4) and (2.3), one of these systems has to be chosen. If logit or linear bounded method is performed, lower and upper bounds for must be entered. One of the solver will solve the equation system, if calib is chosen to do the job, option „type“ is not taken into account. Specified missing values can be omitted from calibration. If a nonlinear system cannot be algebraically solved, user is able to force at least an approximate solution. Once the calibration is finished, new weights, quotients , totals attained and some other outputs show up. It is one’s favour to choose between absolute value and percentage display of totals obtained. Average deviation of calibration weights is computed simply by which acts just as a first view on distortion applied to design weights. If inconvenient weights result from the process, “lower bound for weights below 1“ will keep them over 1. When pressing „save solution“ button, input data with added column (calibration weights) is saved.

Main advantages of Calif are: a free environment, a possibility to find an approximate solution, a stratification aspect, a graphical user interface, several optimization tools, cope with complex and large tables and last but not least, easy handling and transparent outputs. On the other hand, disadvantages could be summarized as: missing GREG estimator, extended computing time when using BBsolve optimizer and no multi-stage calibration at the moment.

**5. Comparison of softwares**

In [[13]](#footnote-13)[11] discussion of CALMAR2, g-Calib 2 and Bascula 4.0 are presented. Let’s add another player.

* *partial nonresponse and missing values* – has to be solved before using any of 4 softwares concerned
* *global nonresponse –* just CALMAR2 uses GREG estimator to compensate global non-response by using instrumental variables
* *installation and running* – like other softwares, Calif can be simply run (after installation of necessary packages mentioned in section 4) by executing calif()
* *data input* – just .csv and .txt formats are supported by Calif, which is limited
* *stratification* – is part of g-Calib2 and Calif as well. In case of bounded methods, lower and upper bound are the same for each stratum
* *one-stage sampling* – it is part of each software
* *two*-*stage sampling* – calibration on totals over primary and secondary sampling units is available in CALMAR2 and g-Calib2, it is also planned for Calif. Calibration on 3 stages is possible only with CALMAR2
* *distance functions* – all 4 functions mentioned in section 3 are part of the softwares, Bascula 4.0 doesn’t contain the logit function. CALMAR2 uses also the hyperbolic sinus function
* *approximate solutions* – in case of unsolvable system (mainly due to strict bounds for quotients), in contrast to CALMAR2 and Bascula 4.0, Calif is able to bring approximate solution. This is shown in the following section.

**6. Case study**

EU-SILC and Household Budget Survey are part of the most requested data in Europe. The EU-Statistics on Income and Living Conditions (EU-SILC) instrument is the EU reference source for comparative statistics on income distribution and social inclusion at the European level. The reference population in EU-SILC includes all private households and their current members residing in the territory of the countries at the time of data collection [[14]](#footnote-14)[12]. The Household Budget Survey (HBS) describes the expenditure structure of different types of households. This information is mainly used at EU level in the context of Consumer Protection Policy [[15]](#footnote-15)[13]. All household members are surveyed, but only those aged 16 and more are interviewed.

EU-SILC as well as HBS is calibrated at two levels. The first calibration is performed at the household level and the second at the individual level. Condition that members of the household should have equal weight than household itself (integrated weights) imply the need of simultaneous calibration. Inferential step has to come. The sample of individuals is turned into a database of households such that auxiliary variables on individual level are summed within particular households one by one. To be specific, as [[16]](#footnote-16)[14] presents, if is a sample of households, a sample of individuals, are design weights, are auxiliary population totals for household level and auxiliary totals in individual population, we compute totals of the individual variables for each household , i.e. . After this step, on household level there are auxiliary variables (household calibration) and (calibration of individuals). Resulting weights are and calibration is correct.

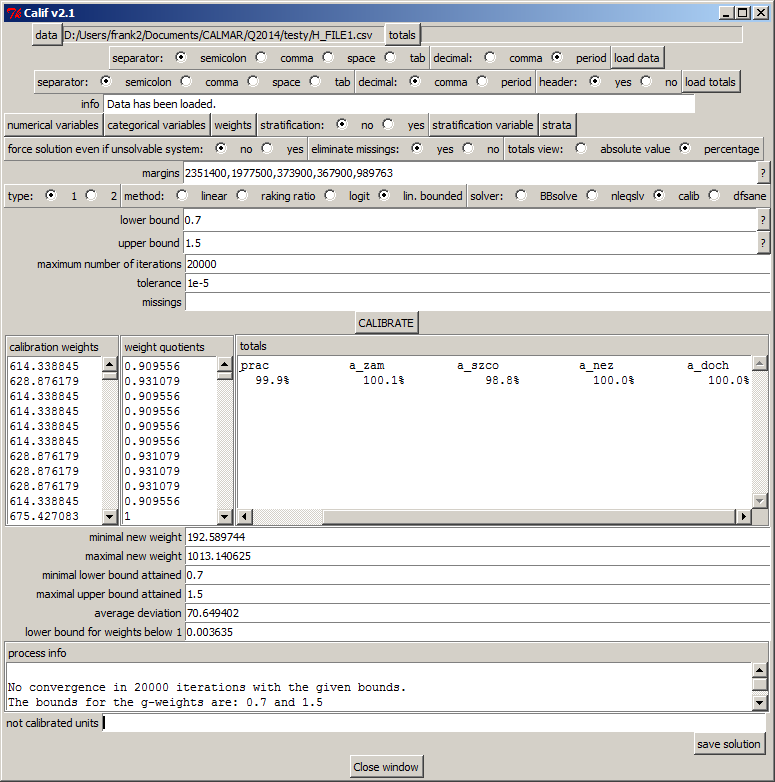
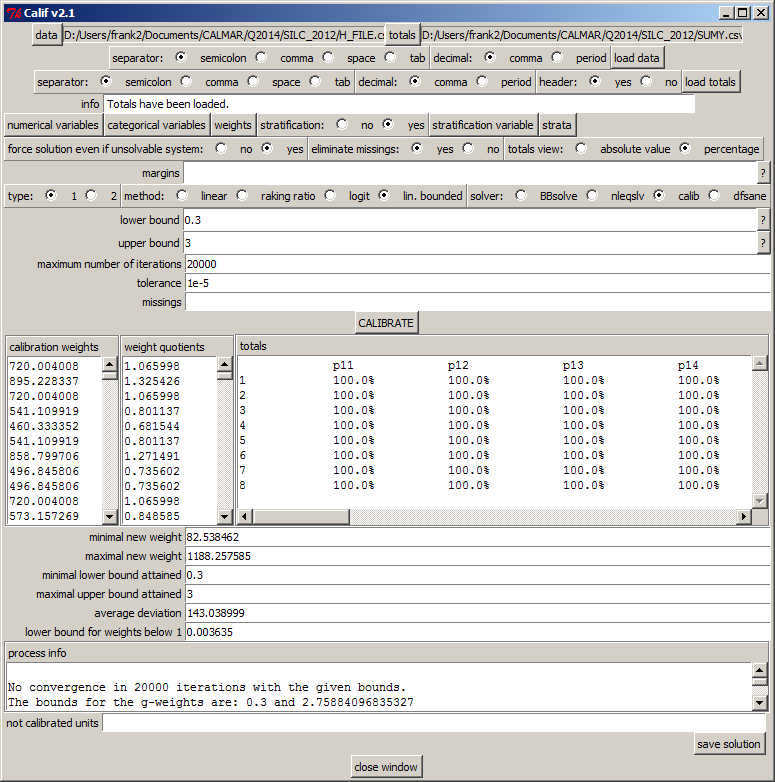
Calibration of EU-SILC is performed on a number of household members (5 categories), sex + age groups (2\*6 categories) and additional auxiliary variables – number of working persons, employees, unemployed, self-employed and pensioners, which constitute 22 variables to calibrate on. Stratification by NUTS3 level is considered (8 strata). Up to now, calibration has been done by CALMAR2 but it was a little bit of drudgery. An exact simultaneous solution has never been found, so the calibration process ran by iterative manner. The variables (see above) were calibrated first, resulting weights were taken as design weights and after that the calibration of variables was executed. This was done several times, separately for each stratum. The result was some kind of an approximate solution, closer to the exact one after each iteration. Therefore, CALMAR2 must have been run for over 100 times.

Figure 6.1. *Result of calibration*

Size of Slovakian EU-SILC 2012 is 5291 households with 15463 members, which represents about 660 households in each stratum on average. Just to test Calif versus CALMAR2, table with 3 auxiliary variables was taken, specifically number of males, females and members of household. Both CALMAR2 and Calif found the same exact solution for each distance function (in particular strata). The second table taken was calibrated on 5 additional variables mentioned above without stratification. CALMAR2 came with an error message, Calif found an exact solution using linear method (with bounds attained 0,54 and 3,06). When linear bounded method with , and *calib* function (from package *sampling*) was used, an approximate solution within these bounds was found, as shown in figure 6.1.

Iterative process of calibration by CALMAR2 brought an approximate solution with bounds 0,34 and 2,72. Just the number of individuals by sex and the number of households (24 totals) were calibrated exactly in each stratum. Precision of other totals varied between 75,4% and 126,9%. Calif gave a nice result within 3 minutes on one mouseclick. Bounds 0,3 and 3 and function *calib* were applied. 153 out of 176 totals were 100,0% (all individual variables), others varied between 96,3% and 101,3%.

Figure 6.2. *Calibration of EU-SILC 2012*



In HBS, 12633 individuals are divided into 4705 households. Calibration process is the same as for EU-SILC. Iterative CALMAR2 procedure with linear bounded method gave us an approximate solution with bounds 0,09 and 4,5. As for EU-SILC, only the number of individuals by sex and the number of households (24 totals) were calibrated exactly in each stratum. Precision of others varied between 78,3% and 121,9%. Calif offered a result within bounds 0,1 and 4,5 on one click. Linear bounded method with *calib* as a solver made 163 out of 184 totals 100,0% precise, others were ranged from 97,9% to 105,7%.

**7. Conclusions**

We have introduced a free R based tool for calibration of weights in statistical surveys. It uses several functions to find a feasible solution, namely *calib* from package *sampling, BBsolve* and *dfsane* from package *BB* and *nleqslv* from package *nleqslv*. A graphical user interface is offered to the user. It can deal with complex and large surveys and contains various components for better comfort. Although it has some deficiences, especially handling of multistage samples and GREG estimator, we can entitle Calif as “promising”.

1. [1] DEVILLE, J.-C., SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, 87, 376-382

   [2] SÄRNDAL, C.-E. (2007). The calibration approach in survey theory and practice. Statistics Canada, Business Survey Methods Division. Catalogue no. 12-001-X, Vol. 33, No. 2, pp. 99-119 [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. [2] SÄRNDAL, C.-E. (2007). The calibration approach in survey theory and practice. Statistics Canada, Business Survey Methods Division. Catalogue no. 12-001-X, Vol. 33, No. 2, pp. 99-119 [↑](#footnote-ref-3)
4. [3] HARMS, T., DUCHENSE, P. (2006). On calibration estimation for quantiles. Survey Methodology, 32, 37-52 [↑](#footnote-ref-4)
5. [4] FRANKOVIČ, B. (2013). Calibration of weights of statistical surveys in R language. Bratislava: Forum Statisticum Slovacum 5/2013, p. 19-37 [↑](#footnote-ref-5)
6. [4] FRANKOVIČ, B. (2013). Calibration of weights of statistical surveys in R language. Bratislava: Forum Statisticum Slovacum 5/2013, p. 19-37 [↑](#footnote-ref-6)
7. [5] SAUTORY, O. (1993). La macro CALMAR. Paris: INSEE [↑](#footnote-ref-7)
8. [6] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for

   Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ [↑](#footnote-ref-8)
9. [7] Thomas J. Hoffmann, Nan M. Laird (2009). fgui: A Method for Automatically Creating Graphical User

   Interfaces for Command-Line R Packages. Journal of Statistical Software 30(2), 1-14. URL

   http://www.jstatsoft.org/v30/i02/ [↑](#footnote-ref-9)
10. [8] Ravi Varadhan, Paul Gilbert (2009). BB: An R Package for Solving a Large System of Nonlinear Equations

    and for Optimizing a High-Dimensional Nonlinear Objective Function. Journal of Statistical Software,

    32(4), 1-26. URL http://www.jstatsoft.org/v32/i04/ [↑](#footnote-ref-10)
11. [9] Berend Hasselman (2014). nleqslv: Solve systems of non linear equations. R package version 2.1.1.

    http://CRAN.R-project.org/package=nleqslv [↑](#footnote-ref-11)
12. [10] Yves Tillé and Alina Matei (2013). sampling: Survey Sampling. R package version 2.6.

    http://CRAN.R-project.org/package=sampling [↑](#footnote-ref-12)
13. [11] CHAUVET, G. et al. (2005). Comparison of 3 calibration softwares. Workshop Calibration tools for Survey Statisticians, CREST-ENSAI, France [↑](#footnote-ref-13)
14. [12] EUROSTAT. (2014). Income and Living Conditions [↑](#footnote-ref-14)
15. [13] EUROSTAT. (2010). Household Budget Surveys [↑](#footnote-ref-15)
16. [14] SAUTORY, O. (2003). A new version of the Calmar calibration adjustment program. Statistics Canada International Symposium Series - Proceedings [↑](#footnote-ref-16)