

Comparison of 3 calibration softwares

GUILLAUME CHAUVET
JEAN-CLAUDE DEVILLE
MOHAMMED EL HAJ TIRARI
JOSIANE LE GUENNEC

CREST-ENSAI, France

Plan of the presentation

1. Presentation of the work
2. Achievement of calibration
3. Facility of use
4. Application with our data

I - PRESENTATION OF THE WORK

I - Presentation of the work

1. Purpose of the study
2. Presentation of the softwares

I.1 - Purpose of the study

This work is the result of a command of Statistics Belgium.

The aim was

- to appraise the G-Calib 2 software
- to compare its efficiency with other calibration softwares
- to propose improvements

I.2 - Presentation of the softwares

G-Calib 2 has been developed in SPSS language by Statistics Belgium (Vanderhoeft, 2003).

Calmar 2 has been developed in SAS language by Insee (Sautory and Le Guennec, 2003).

Bascula 4.0 has been developed in Delphi language by Statistics Netherlands (Nieuwenbroek and Boonstra, 2002) and can be used with Blaise.

II- ACHIEVEMENT OF CALIBRATION

II- Achievement of calibration

1. Preliminary treatment of the sample
2. Installation of the softwares
3. Entry of the data in the software
4. Parameters concerning the calibration technique
5. Following of the calibration
6. Recovery of the table of weightings

II.1 - Preliminary treatment of the sample

The data collected on the sample generally suffer from partial non-response and total non-response.

The problem of partial non-response or missing values must be solved before using any of the 3 softwares.

Global non-response can be treated by Calmar 2, and partially by G-Calib 2. This point will be examined later.

II.2 - Installation of the softwares

G-Calib 2 has been tested with the Windows NT system, and SPSS version 12. The installation has been quite difficult, for two main reasons :

- It seemed impossible to use G-Calib 2 as a shared software
- G-Calib 2 seemed more adapted for the english version of SPSS, what occurred a few problems for running jobs

Despite these problems, its pleasant interface made it easy to use.

II.2 - Installation of the softwares

Calmar 2 is a SAS macro, that can be easily called at the beginning of the calibration program (but needs basics knowledges on SAS Macro). The interface CALMAR2 GUIDE can also be used.

Bascula is a component of Blaise, a system of treatment of survey data. Its installation didn't occur particular problems.

II.3 - Entry of the data in the softwares

The input data correspond to the information collected on the sample, and the margins known over the whole population on which the calibration will be made.

The input data for G-Calib 2 must be put in a SPSS table. This software enables to create a SPSS table starting from various formats (Excel, SAS, text files, ...), what facilitates the seizure.

II.3 - Entry of the data in the softwares

The input data for Calmar 2 must be put in a SAS table. SAS also enables to create tables starting from various formats.

On the other hand, Bascula only uses data bases generated with Blaise, or of Ascii format. That can occur an important work for preparing the Input data.

II.4 - Parameters concerning the calibration technique

The writing of the different parameters is quite easy for any of the 3 softwares.

Yet, forgetting a compulsory parameter in G-Calib 2 doesn't lead to an error message, while both Calmar 2 and G-Calib 2 perform tests on the validity and coherence of parameters.

II.4 - Parameters concerning the calibration technique

Choice of the calibration level

A calibration can be performed on different levels. The basic technique concerns a one-stage sampling ; calibration is made for only one level of observation.

This technique is available for any of the 3 softwares.

II.4 - Parameters concerning the calibration technique

In case of cluster sampling, there exists specific techniques of calibration :

- A calibration imposing the equality of weights for individuals belonging to the same cluster. G-Calib 2 and Bascula 4 use the Lemaître-Dufour technique, contrary to Calmar 2.
- A calibration on totals over clusters (primary units) and individuals (secondary units) ; this technique is only available in Calmar 2 and G-Calib 2

II.4 - Parameters concerning the calibration technique

In case of two-stage sampling, there also exists specific techniques of calibration :

- A calibration on totals over clusters (primary units) and individuals (secondary units) ; this technique is only available in Calmar 2 and G-Calib 2
- A calibration on totals over clusters (primary units) and individuals (secondary units) with 3 levels of observation ; this technique is only available in Calmar 2

II.4 - Parameters concerning the calibration technique

In case of stratified sampling, it is also possible to perform a stratum by stratum calibration, that is to calibrate on totals over the different strata.

This technique is only available in G-Calib 2.

In case of using a restricted distance function, the boundaries are the same for each stratum.

II.4 - Parameters concerning the calibration technique

Choice of the distance function

4 distance functions had been initially proposed in the original article about calibration (Deville, Särndal and Sautory, 1993). They are : the linear function, the exponential function, the truncated linear function and the logit function.

The two last functions enable to control the ratios of weights.

II.4 - Parameters concerning the calibration technique

All these functions are available in both G-Calib 2 and Calmar 2. Bascula 4 doesn't contain the last function, and impose the distance function or particular calibration techniques (such as the ratio estimation or the poststratification, that correspond to particular uses of the linear function).

Another distance function has been developed recently, only in Calmar 2 : the hyperbolic sinus function.

II.4 - Parameters concerning the calibration technique

Particularities of G-Calib 2

G-Calib 2 enables to normalize the calibration variables. The variables used for the calibration have then the same order of magnitude.

G-Calib 2 also enables to use the collapsing technique, to improve the time of execution. This option is essentially interesting if all calibration variables are categorial.

II.4 - Parameters concerning the calibration technique

Particularities of Calmar 2

Calmar 2 enables to perform the generalised calibration technique (Deville, 2002), to compensate global non-response by using instrumental variables.

This option is not available on G-Calib 2 and Bascula 2 ; yet, G-Calib 2 can perform a uniform treatment of non-response, using a scale parameter.

II.4 - Parameters concerning the calibration technique

Particularities of Bascula 4

Bascula 4 enables to perform a variance estimation, using the linearisation technique or the half-samples technique.

Yet, there is no mention of how this estimation takes into account the case of non-response, or the case of several levels of sampling. Moreover, the maximum number of simulations for the half-samples technique is quite weak.

II.5 - Following of the calibration

Statistics about the course of the calibration process are available for any of the software.

They are flexible for Calmar 2 only.

With G-Calib 2, the Log file isn't separated of the results file. That makes the reading of the realisation of the algorithm difficult.

II.6 - Recovery of the table of weightings

Calmar 2 enables to update an existing SAS table, what allows the user to calculate several vectors of weights (for example, to restrict as much as possible the ratios of weights) and to compare them easily.

This option is not available for G-Calib 2 and Bascula 4.

III- FACILITY OF USE

III- Facility of use

1. The user's manual
2. User-friendliness of the software

III.1 - The user's manual

The user's manuals of G-Calib 2 and Calmar 2 are very complete and give many examples of calibrations. Bascula's one is quite complete too.

In the manual of G-Calib 2, matricial notations are systematically used, what doesn't facilitate comprehension. The fusion of the manuals would also be good for comprehension.

In Bascula 4, not all specifications are mentioned in the manual and points are sometimes not clear.

III.2 - User-friendliness of the softwares

G-Calib 2

The interface for writing parameters is clear ; the main problem is the lack of control on these parameters.

A request can be saved, re-called and modified what appears to be very useful for calibrations to come.

III.2 - User-friendliness of the softwares

Calmar 2

The writing of parameters can be made :

- either directly by calling the CALMAR macro in the SAS Editor Window
- either with the CALMAR2 GUIDE macro that provides an interface for the writing

Both provide some coherence controls on parameters.
A request can also be easily saved and re-called.

III.2 - User-friendliness of the softwares

Bascula 4

Bascula provides an interface for writing parameters, and some controls on the coherence of parameters.

Unlike G-Calib 2 and Calmar 2, a request can't be saved and re-used.

IV - AN APPLICATION WITH THE SURVEY ABOUT EMPLOYMENT

IV - An application with the Survey about Employment

1. Data used for the comparison
2. The calibrations realised
3. The results

IV.1 - Data used for the comparison

Respondents of the 2001 survey about employment constitute the reference population (114 111 individuals for 58 189 households).

The first sample is constituted of 8 000 households, containing 15 696 individuals aged 15 years or more, selected by means of a simple random cluster sampling.

IV.1 - Data used for the comparison

In these 8 000 households, individuals have been selected by simple random sampling and constitute our second sample, obtained by two-stage sampling.

We also select a stratified sample with proportionnal allocation of about 8 000 households, that constitute our third test sample. Auxiliary variables relative to the individuals and to the households are available and are used for the calibration.

IV.2 - The calibrations realised

Simple calibration with auxiliary information on households

This calibration can be performed for any of the 3 softwares.

Simple calibration with auxiliary information on individuals

This calibration can be performed for any of the 3 softwares. Bascula 4 doesn't provide the raking ratio distance function (due to a continuous calibration variable).

IV.2 - The calibrations realised

Calibration with auxiliary information on individuals, imposing the equality of weights in each household

This calibration is available in Calmar 2, with a technique proposed by Sautory.

It is also available in G-Calib 2 and Bascula 4, with the Lemaître Dufour method. The weights are obtained on household level, and must be divided by the size of the household to be used on individual level.

IV.2 - The calibrations realised

Calibration with auxiliary information on individuals and households, in case of cluster sampling

This calibration can only be performed with G-Calib 2 and Calmar 2.

Calibration with auxiliary information on individuals and households, in case of two-stage sampling

This calibration can only be performed with Calmar 2.

IV.2 - The calibrations realised

Calibration by stratum

This calibration can be performed automatically with G-Calib 2 only.

With Calmar 2 and Bascula 4, the calibration must be performed separately on each stratum.

IV.3 - The results

In case of restricted distance functions, the boundaries obtained for the ratios of weights are generally very close.

The main differences are obtained in case of calibration imposing equality of weights for the individuals of the same household. Indeed, Calmar 2 uses a different technique than G-Calib 2 and Bascula 4.

IV.3 - The results

The calibration algorithm may not converge if the restrictions in the boundaries of the ratios of weights are too important.

In that case, the calibration is impossible and both Calmar 2 and Bascula 4 give an error message.

In G-Calib 2, the systematic use of generalised matrix inverses sometimes gives false solutions.