Calibration Weighting in Survey Sampling (Based on Sample Socio-Demographic Survey)

Yuliia Halytsia

Taras Shevchenko National University of Kiev, e-mail: halytsya2013@gmail.com

Abstract

Today, one of the most pressing informational and statistical problems is the problem of ensuring the reliability of the results of sample population surveys. The solution to these problems is closely linked to the creation and use (including further adjustment) of the system of statistical weights of the sample survey. This contributed paper contains a short overview of calibration method and its applying to sample socio-demographic survey.
Keywords: calibration, household survey, statistical weights

1 Introduction

Large-scale population surveys are unique on the basis of an array of received primary data, a system of indicators, evaluated on the results of the survey, the principles of organization and conduct of the survey, data processing, etc. First of all, it concerns sample surveys of the population, because, firstly, the design of the sample is developed on the basis of existing actual data sources and is, in a sense, unique; and secondly, at many stages of the survey it is necessary to take into account the fact that not all the general population is examined, but a certain, specially selected part of it.

It is necessary to pay much attention to the problem of coordinating the results of sample surveys of the population with available high-quality external information to increase their representativeness and usefulness. Such coordination is most often appropriate at micro level - for individual units: individuals, households, etc. The expediency of coordinating the results of surveys with external information is conditioned, at least, by the fact that: the evaluation of the indicators based on the results of sample surveys is characterized by a certain error due to lack of observation, as well as non-sample mistakes; a sample population survey cannot provide estimates of certain characteristics of the general population (although they are obtained on the basis of the survey, but mainly reflect the parameters laid down in the formation of the sample); the main characteristics of the general population can change rapidly in time, so at the time of completion of the processing of the survey data - somewhat different from those at the time of the organization of the survey.

Ensuring the maximum quality and usefulness of the results of sample surveys in estimation of the target characteristics of the general population is the main purpose of the system of statistical weights. The most reasonable way of solving the problem of coordinating the results of a sample survey with external data is the corresponding adjustment of the system of statistical weights.

2 Calibration method

A method of calibration is the most theoretically developed and effective from the currently known generalized methods for adjusting the system of statistical weights in order to coordinate the results of the survey with several external distributions for different types of units. The method consists in solving a special task of minimizing the change in the value of weights in the process of coordinating the results of the survey with the external data. Usually, at the same time as the sample is formed, the basic statistical weights are calculated. On the following stages, which are implemented after the survey, the basic weights are adjusted precisely in order to take into account structural features of the general population.

The statement of the calibration problem as an optimization problem of minimizing the distance between design weights and calibrated (adjusted) weights, provided that in the household / population survey reliable external information on the total number of households and gender-age structure of the population is available, can be represented as follows:

$$\begin{cases} \sum\limits_{i=1}^{n} q \dfrac{\left(w_i^{(c)} - w_i^{(d)}\right)^2}{w_i^{(d)}} \to \min; \\ \sum\limits_{i=1}^{n} k_{ji}^{(f)} w_i^{(c)} = F(j); \;\; j = 1,2,...,J; \\ \sum\limits_{i=1}^{n} k_{li}^{(m)} w_i^{(c)} = M(l); \;\; l = 1,2,...,L; \\ \sum\limits_{i=1}^{n} w_i^{(c)} = H, \end{cases}$$

where:

n – the sample size of the population / households who participated in the survey;
q - parameter;

$w_i^{(d)}$ - statistical weight of i - th respondent / household, which needs to be adjusted;

$w_i^{(c)}$ - weight of i - th respondent / household after calibration;

$k_{ji}^{(f)}$ - the number of women in i - th household, which, according to the survey, belong to the same j - th gender-age group;

$F(j)$- the total number of women in j - th gender-age group according to external data;

J - number of gender-age groups for women;

$k_{li}^{(m)}$ - the number of men in i -th household, which, according to the survey, belong to the same l - th gender-age group;

$M(l)$- the total number of men in l - th gender-age group according to external data;

L - number of gender-age groups for men;

H - total population by external data.

The solution of the formulated problem can be accomplished using method of Lagrange multipliers.

3 Application of calibration method

For the practical part of the work we used the results of a sample survey "Social Inequalities: Perceptions by Ukrainian Society", conducted by the Center "Social Monitoring" in 2017.

As a subject of research and evaluation for the general population, we chose nominal variables that reflect the perceptions of the population about the minimum necessary and sufficient level of income, as well as the level at which a household can be considered poor.

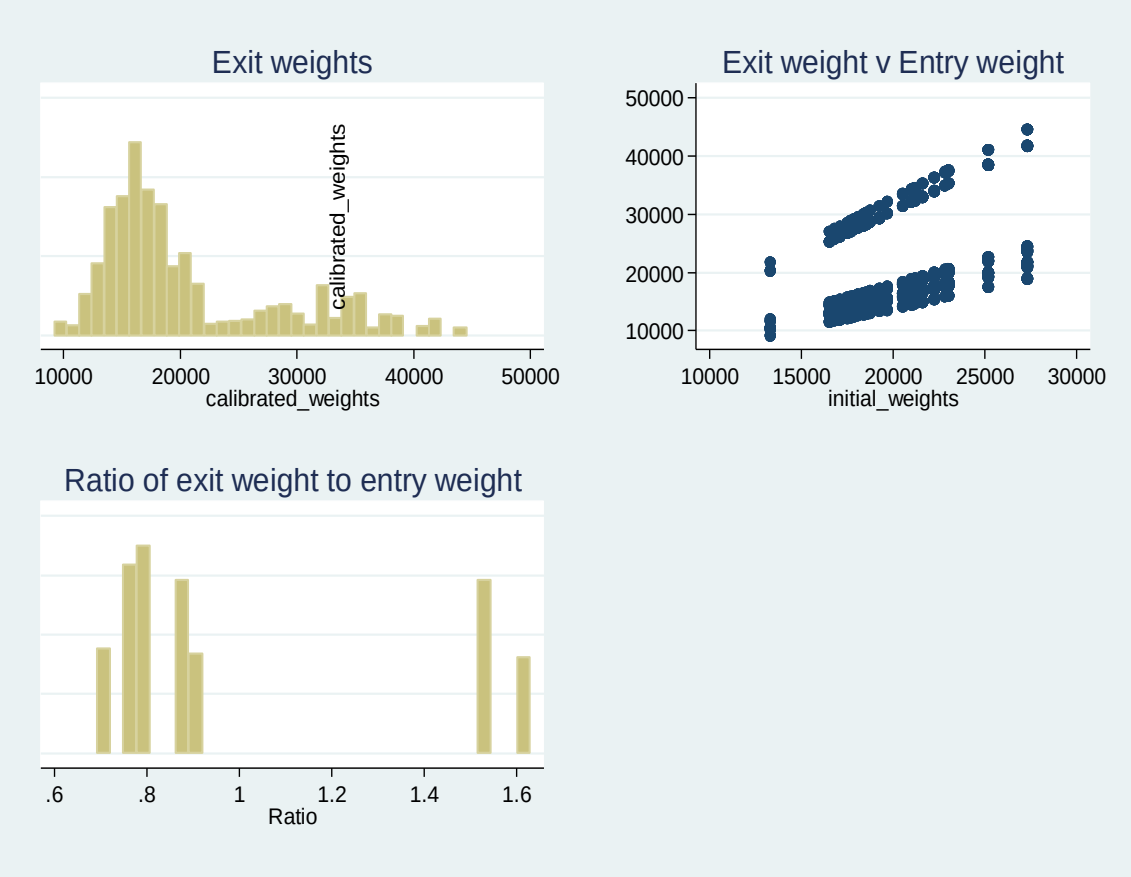
Base on the survey data, were calculated average sample values of: *monthly average per capita income per family member, which the respondent considers to be sufficient for normal life* and *monthly average per capita income, which, according to the respondent's opinion, provides the living wage*, as well as *monthly average per capita income for which family can be considered poor*. These indicators are to be estimated for the general population after calibration of statistical weights.

Given the peculiarities of the sample design and the used system of primary weights of the socio-demographic survey "Social Inequalities: Perceptions by Ukrainian Society", we consider it expedient to perform the calibration procedure using a combination of such variables as gender and age (gender-age structure of the population). Within this work, calibration was be carried out only at the national level, without considering the type of settlement or regions.

Table 1: Primary and calibrated statistical weights

Weights	Number of observations	Average	Standard deviation	Minimum value	Maximum value
Primary	2 046	20 730.7	3 135.8	13 313.9	27 317.5
Calibrated	2 046	20 730.7	7 814.7	9 217.1	44 516.9

Figure 1: Results of calibration of statistical weights



The obtained results of calculations of the mean values of the indicators for the sample and estimates for the general population (that is, after the calibration procedure) are given below:

- Monthly income per family member, which is considered sufficient for normal life:

Primary weights

```
Survey: Mean estimation

Number of strata =      1        Number of obs   =      2,046
Number of PSUs   =  2,046        Population size =  42,414,905
                                  Design df       =      2,045

                 |             Linearized
                 |       Mean   Std. Err.     [95% Conf. Interval]
-----------------+------------------------------------------------
             f02 |   9043.391   297.785       8459.398   9627.385
```

Calibrated weights

```
Survey: Mean estimation

Number of strata =      1        Number of obs   =      2,046
Number of PSUs   =  2,046        Population size =  42,414,905
                                  Design df       =      2,045

                 |             Linearized
                 |       Mean   Std. Err.     [95% Conf. Interval]
-----------------+------------------------------------------------
             f02 |   8663.918   280.934       8112.972   9214.865
```

- Monthly income per person, which provides a living wage for today:

Primary weights

```
Survey: Mean estimation

Number of strata =      1        Number of obs   =      2,046
Number of PSUs   =  2,046        Population size =  42,414,905
                                 Design df       =      2,045

                    Linearized
             Mean   Std. Err.     [95% Conf. Interval]
-------------------------------------------------------------
f03      4699.368   79.73994      4542.988    4855.748
```

Calibrated weights

```
Survey: Mean estimation

Number of strata =      1        Number of obs   =      2,046
Number of PSUs   =  2,046        Population size =  42,414,905
                                 Design df       =      2,045

                    Linearized
             Mean   Std. Err.     [95% Conf. Interval]
-------------------------------------------------------------
f03      4570.809   74.74176      4424.231    4717.387
```

- ***Family with such monthly income per capita can be considered poor:***

Primary weights

```
Survey: Mean estimation

Number of strata =      1        Number of obs   =      2,046
Number of PSUs   =  2,046        Population size =  42,414,905
                                 Design df       =      2,045

                    Linearized
             Mean   Std. Err.     [95% Conf. Interval]
-------------------------------------------------------------
f04      2466.568   48.82644      2370.814    2562.323
```

Calibrated weights

```
Survey: Mean estimation

Number of strata =      1        Number of obs   =
Number of PSUs   =  2,046        Population size =  42,4
                                 Design df       =

                    Linearized
             Mean   Std. Err.     [95% Conf. Int
-------------------------------------------------------------
f04      2395.195   46.65419      2303.701    2
```

Conclusion

Average values for all investigated nominal variables after calibration were lower than before the calibration procedure for statistical weights was performed. We believe that this is due to consideration of a high level of demographic aging in Ukraine in estimating indicators for the general population, since elderly people are usually distinguished by comparatively lower financial needs and showing lower self-esteem levels of necessary and sufficient income.

References

Sarioglo V.G. (2012) Estimation of socio-economic indicators: applied aspects of the application of indirect methods. *Ptoukha Institute for Demography and Social Studies of the National Academy of Sciences of Ukraine*, 136

D'Souza, J. (2010). Calibrate: a Stata Program for Calibration Weighting.- *London: Stata User Group.*

Kish L. Survey sampling / Kish L. – *New York : John Wiley & Sons, 1995*. – 643 p.

Rozora I., Lukovych O. Mean Estimation with Robust Calibrated Estimators// *Baltic-Nordic Summer School on Survey Statistics.- Kyiv, 2016.*