Weighting adjustments using micro and macro auxiliary variables

**Keywords:** Response propensity weights, calibrated weights, confidentiality

# Introduction

Weights in sample surveys are necessary to create so that the parameters of the target population are possible to estimate without biases as well as possible. We here are considering both point estimates and interval estimates. The bias is common for several reasons, due to nonresponse in particular. It is also possible that the frame or the frames are low quality. Currently estimates are often becoming more biased as earlier times since nonresponse has been worsening in all developed countries.

This paper is focused on survey weights that can be created with several methods. Usually, the basic weights are first created following the principles of the sampling design of the survey. Since there is a problem of unit nonresponse, it is needed to assume something about its characteristics. It is enough first to assume that unit nonresponse is ignorable. This assumption can be made over the whole data or separately by strata for example. This latter assumption leads to less biased estimates but it is good to continue towards better adjusted weights.

There are many strategies to adjust the weights. The auxiliary variables have a big role in this work. They should be available both for the respondents and for the non-respondents (gross sample). This paper presents the basic strategy for using auxiliary variables. To do it well, it is necessary to classify auxiliary variables into the two main groups: (i) macro auxiliary variables are aggregates and (ii) micro auxiliary variables are for individuals of the statistical units. Naturally, the categories of aggregates should be available in the complete gross sample.

It is not common to exploit both types of auxiliary variables in Europe, as the case of the European Social Survey (ESS) shows. Such variables are understood to be important, but the traditions of many countries are too far from this. On the other hand, the countries with good registers can easily collect both micro and macro auxiliary variables. A big question is confidentiality when getting auxiliary variables for research purposes including weighting adjustments. This is more demanding for micro auxiliary variables. The permission to use such variables is not possible to get from the non-respondents. Hence, the too strict data owner cannot give such necessary information for quality or weighting adjustment purposes. It would mean that the macro auxiliary variables only might be tried. It is not still automatic, since macro information may still concern small groups and individuals can be recognised respectively. Our aim to show how good are some survey estimates when using following auxiliary variables.

1. No auxiliary variables that is basic weights only
2. Macro auxiliary variables only, that is, for calibrated weights
3. Micro auxiliary variables, that is, response propensity weights
4. Macro auxiliary variables using response propensity weights
5. Micro auxiliary variables using response propensity weights followed by macro auxiliary variables using calibration.

Our exercise is based on empirical survey data that resembles the European Social Survey but it is here artificial created by the author. We describe our data in section 2. The following section includes our weighting methodologies. Then we present the results for the weights and estimates for one variable, that is, for happiness. The results for other variables have similar profiles but the room of this paper does not give opportunity to present them. The last session concludes.

# Test Sample Data

The gross sample data of the European Social Survey (ESS) is confidential for the outsiders but the author of this paper was able to use such data to create the synthetic data file ([1]). Its gross sample size is 4598 and the net sample respectively 2178 meaning that the unit nonresponse rate is 47.3 per cent. This is quite close to the average rate of the ESS, and thus rather low. The target population size is nearly 20 million which size of the country does not exist. Our data thus are used for testing weightings and it is not necessary to know the aggregate characteristics of the target population.

The sample design of the test data is also some type of a compromise of the ESS. We call it two-domain design that is used in some countries so that stratified simple random sampling is used in urban areas where the interviewers have short distances from one place to next to contact potential respondents. Respectively, the sampling of the rural is stratified two-stage cluster design that is common in many countries. We do not here concentrate on the sampling design in details. The basic weights were created so that nonresponse varies by eight strata but is random within each stratum (ignorability).

Our gross sample data include many micro auxiliary variables: gender, age (we used 10-year and 5-year age groups), education (6 categories), income (ten deciles), region, household size, household members below 15 years, citizenship, marital status; stratum can be considered as an auxiliary variable as well. All these thus are micro auxiliary variables but we created from those the following macro auxiliary variables too:

1. Target population sizes by stratum, gender, age group, education, income. These macro variables were used as calibration margins. Note that stratum is important to include there, otherwise it is not guaranteed that these will be remained correct after calibration. We tested various categorizations for these variables but we do not present all results. For example, since the age groups with 5 year intervals and 10 year intervals give rather similar results, we present the 10 year groups only. On the other hand, we found that income and education groups lead to quite similar results: hence our calibration margins are only for education since its impact is more significant. It means that our calibration margins are as follows: 8 strata, 2 genders, 8 age groups and 6 education groups.
2. The second macro auxiliary variable type in our exercise is created from the same variables as group a) but now the relative frequencies of each category were used (e.g. by gender or education). These auxiliary variables were used in response propensity weighting. We thus use these same variables as both micro auxiliary variables and macro variables but their form is different.

# Methodologies

We already explained the creation of the basic weights. It is possible to continue from these weights straightforwardly to calibration ([2], [3]). This strategy means that any micro auxiliary variables are not used, and if these macro variables are confidentially done, we can forget such problems. It is not however sure that the estimates are best possible ones.

The better weights can be expected to get if the response propensity weights, either real micro auxiliary variables or macro variables (group b) above), are used before calibration ([1], [4}). Our conclusion earlier is that calibration is good to keep as the last method in weighting but not necessarily as the only adjusted method. It is not however clear which calibration margins and how many are good to use. A problem is also the calibration method itself. It has been found that linear calibration may lead to negative or other undesirable weights and hence this method cannot be used always. Such bad weights are found to be more common if too many calibration categories are tried. Our recommendation is to select a few calibration margins such that are well known (e.g. gender, age group). If they do not lead to correct estimates, the whole survey results might not be trusted. The response propensity weights can be made using alternative link functions. We use in our exercise the probit link function but the results are not much different with logit.

# Results

The results of this section are summaries. We constructed many other weights and estimates. A logit link function is more used than probit but the basic results are not dramatically different since we here are focused the different chains of the weights. We thus start from the basic weights, then go to the least problematic weights due to confidentiality, that is, using macro auxiliary variables. These are first calibration margins, and then the same things are used as response propensity weights. The yariation of the weights will increase respectively. Our last weights use the real micro auxiliary variables but finally we use the calibration margins as the final stage.

Table 1 shows the results first for the weights in that order. The same table includes also averages and standard errors using the same weights of happiness. As we see while the variation of the weights (CV) increases, the happiness average decreases. The reason is that the weights take more advantage of auxiliary variables, meaning that the bias is reducing as well. The happiness using the basic weights only is too high (7.60) since the respondents are often happier than the non-respondents. The minimum and least biased average is 7.33 when using the full number of micro auxiliary variables in the response propensity model and calibration margins after that. We do not believe that this result is unbiased since our pattern of the auxiliary variables could be better. Note that happiness results are estimated using complex sample tools since our data covers strata and clusters.

The calibration in Table 1 is based on linear if any negative values are not got. Otherwise, we have used either a raking ratio option or the sinus hyperbolicus option. The French CALMAR SAS Macro is applied for calibration [[5])

# Conclusions

The weighting is necessary to use for sample survey estimation. This paper presents the basic principles of the simple and more advanced weights and some examples. All weights require auxiliary data. We here present these as two main categories, either macro or micro. Our recommendation is to use both meaning that these need to be obtained. Two obstacles are met: One is a general understanding of their importance. The second might to confidentiality if the data owner does not release such data. This might happen also due to the new EU regulation GDPR if not solved correctly.

Table 1. Core results for basic and adjusted weights and happiness. The average of all weights is equal to 9140. Micro 1 = Macro variables used micro, Micro 2 = the full number of micro auxiliary variables (see the text) used as micro.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | **Characteristics of weights** | **Happiness** |
| Weight |  Res-ponse propen-sity model | Cali-bration margins | Coeff of Vari-ation | Mini-mum | Maxi-mum | Mean | Std error |
| Basic weight | No | No | 64,0 | 3332 | 22838 | 7.60 | 0.04 |
| Basic weight and linear calibration | No | Yes | 69,0 | 2349 | 380306 | 7.59 | 0.04 |
| Simple adjustment | Micro 1 | No | 66,0 | 2655 | 26677 | 7.57 | 0.04 |
| Basic adjustment | Micro 1 | Yes | 82,2 | 1348 | 54540 | 7.40 | 0.05 |
| Basic adjustment | Micro 1 | No | 78,3 | 1903 | 51170 | 7.39 | 0.05 |
| Basic adjustment | Micro 1 | Yes | 84,4 | 1562 | 66121 | 7.36 | 0.06 |
| Large adjustment | Micro 2 | No | 84,2 | 1860 | 72644 | 7.39 | 0.05 |
| Large adjustment | Micro 2 | Yes | 91,0 | 1511 | 50404 | 7.33 | 0.06 |

# References

1. S Laaksonen (2018). *Survey Methodology and Missing Data. Methods and Tools for Practitioners.* Springer.
2. J-C Deville & C-E Särndal (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*. 376-382.
3. J-C Deville, C-E Särndal & O. Sautory (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*. 1013-1020.
4. S Laaksonen & M Heiskanen (2014). Comparison of Three Modes for a Crime Victimization Survey, *Journal of Survey Statistics and Methodology 2 (4): 459-483 doi:10.1093/jssam/smu018*
5. O Sautory (2003). CALMAR 2: A NEW VERSION OF THE CALMAR CALIBRATION ADJUSTMENT PROGRAM. Proceedings of Statistics Canada’s Symposium: Challenges in Survey Taking for the Next Decade