Anne Konrad (University of Trier)

Consistent estimation at person-level and household-level

Topic 2 – Learning more from what we already know

Keywords: official statistics production, integrated weighting, general regression estimators

Introduction

Many household surveys are based on cluster sampling which permits estimation at both, the person- and household-level. This raises the question how to guarantee consistent estimates at both levels. In practice, integrated weighting, which substitutes individual auxiliary variables by aggregated or mean values, is often used. Eurostat recommends integrated weighting for producing cross-sectional weights for EU-SILC (European Commission 2014).

By assigning unique weights to all persons within a household as well as to the household itself, integrated weighting automatically produces consistent estimates. However, we expect some loss of information caused by the irreversible aggregation process which induces the usage of constructed instead of original auxiliary variables. Hence, our research questions are twofold:

1. Is there a price to pay for consistent estimates due to the restriction of unique weights?

2. Does an alternative weighting strategy exist which is capable of both, ensuring consistent estimates at both levels and allowing for individual weights for persons within the same household?

Methods / Problem statement

First, we compare the performance of a usual person-level GREG and an integrated GREG proposed by Lemaître and Dufour (1987) in order to emphasize the consequence of a restriction of unique weights. In a simulation study based on the AMELIA dataset we compare the distribution of the resulting weights and the quality estimates of person as well as household characteristics.

Secondly, we suggest an alternative weighting method related to modified regression which has three main features: it is based on all available information at both levels, ensures internal as well as external consistency and allows for different weights for persons within the same household.

The underlying idea of our alternative modified regression estimator is that internal consistency is only required for the common variables at both levels. We firstly estimate the unknown totals of the common variables and thereafter use these estimated totals as additional auxiliary variables in the calibration. Again, in a simulation study, we compare the resulting distribution of weights and the accuracy of this estimator with those of the usual and integrated GREG estimator mentioned above.

Results / Proposed solution

Although the usual person-level GREG and the integrated GREG both utilize the same auxiliary variables, the integrated weights exhibit a greater dispersion. This greater dispersion is not justified by a different composition of the sample and the population. For most person-level characteristics the person-level GREG performed slightly better, but the reverse holds for household-level characteristics.

The weight dispersion of our modified GREG is similar to the one of the integrated GREG, while showing significantly smaller standard errors. Comparing all three GREG estimators, our modified GREG performs best for most characteristics.

Conclusions

We have shown that there is a price to be paid to enforce consistent estimates due to the restriction of unique weights within a household. In contrast, our modified GREG estimator represents an alternative weighting approach without the restriction of unique weights. The advantage of allowing for different weights within a household is reflected by improved estimation quality, while the weights do not spread more than the weights of an integrated GREG.