

# **Integrating a refreshment sample in a longitudinal survey: effects of different weighting methods over four waves**

**Keywords:** weighting, top-up sample, non-response, probability-based panel

## **1. INTRODUCTION**

Economic statistical analysis is more and more interested in the construction and use of longitudinal panels to monitor the social and economic phenomena. Longitudinal surveys typically provide both cross-sectional and longitudinal weights that facilitate population inference from the sample. The estimates obtained in the longitudinal context are affected by the different elements that influence the weights adopted in the estimates.

In the course of their life, longitudinal panel surveys often make use of refreshment samples, either to increase their sample size (overall or for specific groups), or to improve coverage/representativity. When refreshment samples are introduced, weights should be adjusted properly to account for that. The development of these weights is straightforward when the additional sample is from a new part of the population that was not included in the original sample. It is more difficult when the samples are drawn from the same population.

Further, following rules are implemented in the panel to ensure the sample mimics the changes in the population as much as possible. As a consequence, the sample changes over time. Also the population changes overtime, due to immigration, emigration, births, and deaths. We can think of the original sample augmented by the following rules and the new refreshment sample as representing overlapping but not identical populations.

Various ways can be adopted for incorporating refreshment samples into an ongoing longitudinal panel. [1] investigates how to best incorporate the refreshment sample into the cross-sectional weights. Particularly, [1] evaluates six options for integrating a refreshment sample with an ongoing sample using the HILDA Survey [2]. Another study that examined this issue in a longitudinal setting is [3] using a cohort study (the US National Longitudinal Survey of Youth), though both of their samples were selected at the same point in time.

Building on [1], the focus of this paper is to investigate effects of different weighting procedures for integrating a refreshment sample in an ongoing longitudinal panel. Particularly, we are interested in the effects on both cross-sectional and longitudinal weights and over different waves of the panel.

The empirical analysis is carried out with reference to the UK Understanding Society Innovation Panel (IP), a longitudinal panel of households running since 2008 [4]. The target population for the IP is all individuals aged 16 or over and living in England, Scotland, or Wales. The IP consists of an initial stratified clustered probability sample of 2760 households in Great Britain [5]. All members of selected households are included in the panel. Starting from the initial sample, all members of the current household of each sample person are re-approached for interview at each subsequent wave. 960 were added at Wave 4 in 2011 by means of a refreshment sample.

The Understanding Society Innovation Panel involves interviews at 12-month intervals. Interviews cover a wide range of topics, such as household dynamics, economic activity, income, health, housing, and political attitudes.

## **2. METHODS**

The methodology for creating estimates from samples drawn from two or more frames has been developed in the context of multiple frame surveys.

Two main ways to integrate independent surveys together have emerged. One is based on combination the estimates from each frame in such a way that the variance of the estimate is minimized. The other method is based on pooling the samples using the inclusion probabilities for the two frames.

We compare different methods with reference to the IP for waves 4 to 8. The evaluation takes into account several aspects: variability of the weights, bias of some key estimates, and root mean square error. Both socio-demographic variables and economic variables (such as gross pay, savings and investments, and political related variables) are considered for the bias analysis. Of course, in some cases, comparison to the population value is possible, in other cases only relative comparison is performed. Further, we investigate the effects of different weighting procedures over different waves of the panel.

## **3. RESULTS**

The comparison allows to highlight strengths and weaknesses of each combination method. Initial results show that the combination method has an effect over the estimates for several waves of the panel. Further, it seems that pooling samples is more efficient than combining estimates. This is in agreement with results in [1].

## **4. CONCLUSIONS**

We examine properties of alternative methods for integrating refreshment sample. This is important, as it is expected to have an impact on final economic estimates. We extend the evidence provided by [1] by analyzing the effects of different weighting strategies not only for cross-sectional weights but also for longitudinal weights and by investigating effects over multiple waves of the panel.

The deepening of the study of the weights in the longitudinal context can provide ideas for further studies on data integration methodologies from different sources as well.

## **REFERENCES**

- [1] N. Watson, Evaluation of weighting methods to integrate a top-up sample with an ongoing longitudinal sample, *Survey Research Methods* 8 (2014), 195-208.
- [2] N. Watson, and M. Wooden,. The household, income and labour dynamics in Australia (HILDA) Survey: wave 1 survey methodology. *HILDA Technical Paper Series*, 1(2) (2002).
- [3] C. O’Muircheartaigh and S. Pedlow, Combining samples vs. cumulating cases: a comparison of two weighting strategies in NLSY97. *American Statistical Association Proceedings of the Joint Statistical Meetings* (2002), 2557–2562.
- [4] S.C.N. Uhrig, Using experiments to guide decision making in Understanding Society: Introducing the Innovation Panel, chapter 13. In: McFall SL, Garrington C, editors. *Understanding Society: Early Findings from the First Wave of the UK’s Household*

Longitudinal Study. Colchester: University of Essex, 2011. Available from:  
<http://research.understandingsociety.org.uk/findings/early-findings>.

- [5] P. Lynn and O. Kaminska, Weighting strategy for Understanding Society, Understanding Society Working Paper Series, 2010-05.