

AUXILIARY INFORMATION IN DATA COLLECTION AND ESTIMATION STAGE

Kaur Lumiste

University of Tartu, Estonia, kaur.lumiste@ut.ee

Large non-response has become an almost unavoidable part of every survey, leading to biased results and questionable inference. There is extensive literature on how to reduce non-response bias of estimates in the estimation stage of the survey, but corrective actions can and should be taken earlier – in planning and data collection phases. There is a growing number of research being done on *responsive survey designs* (Särndal 2011), where the goal is to get a well representative set of respondents through planning and appropriate intervention in the data collection process. In current work we aspire to a “representative” set of respondents through balance of the response set with respect to a given set of auxiliary variables – means of auxiliary variables have to be approximately the same in the sample and the response set. We assess balance with an imbalance measure discussed by Särndal (2011) using auxiliary information.

The same auxiliary variables can also be used in the estimation stage to improve our estimates, but assume that we have access to more auxiliary variables in the estimation stage than we did in the data collection stage. Is the effect of additional explanation power affected by balancing? Finding an answer to this question brings another one - should we emphasise on acquiring more auxiliary variables for the estimation stage or should we focus more on balancing the response? Which would have a larger effect on the bias and/or accuracy of the final estimates? Our goal is to determine the effects of using auxiliary information both in the data collection stage and post-weighting.

References

Särndal, C.-E. (2011). The 2010 Morris Hansen lecture dealing with survey nonresponse in data collection, in estimation. *Journal of Official Statistics* 27, 1–21.