

Calibrating on Principal Components in the presence of Multiple Auxiliary Variables for Nonresponse Adjustment

Bernardo Rota
Örebro University, Sweden, bernardo.rota@oru.se

Thomas Laitila
Örebro University, Sweden, thomas.laitila@oru.se

Abstract

A prerequisite for valid estimation in surveys with nonresponse is access to appropriate auxiliary information. When a large set of auxiliary variables is available, estimating on all of these may result in inefficient estimators, especially if the set contains duplications or variables highly correlated. Thus, a subset of available auxiliary variables has to be selected. This selection has to be made with care avoiding exclusion of auxiliary variables bringing information on the estimation problem at hand. In this paper the principal components method is suggested for dimension reduction. The effectiveness of using principal components in two different calibration schemes is studied: the linear calibration that uses no explicit response function and the propensity calibration which is based on an explicit functional form. Furthermore, a principal component retention criteria based on the canonical correlation between the principal components and the model variables is suggested. Simulation results illustrate that the properties of the estimators are improved by using principal components of the auxiliary variables.

References

- Bardsley, P. and Chambers, R. L. (1984). Multipurpose estimation from unbalanced samples. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 3:33, 290-299.
- Beaumont, J. F. (2006). On the use of data collection process information for the treatment of unit nonresponse through weight adjustment. *Survey Methodology*, **31**, 227-231.
- Bethlehem, J. and B. Schouten (2004). Nonresponse adjustment in household surveys, *Discussion paper 04007. Statistics Netherlands, Voorburg/Heerlen, The Netherlands*.
- Bilen, C., Khan, A. and Yadav, O. P. (2010). Principal components regression control for multivariate autocorrelated cascade process. *Int. J. Quality Engineering and Technology*, **1:3**
- Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, **51**, 279-292.
- Brick, J. M. (2013). Unit Nonresponse and Weighting Adjustments: A Critical Review. *Journal of Official Statistics*, **29:3**, 329-353
- Borga, M. (2001). *Canonical Correlation a Tutorial*. Retrieved from https://www.cs.cmu.edu/~tom/10701_sp11/slides/CCA_tutorial.pdf
- Cadima, J. and Jolliffe, I. T. (1995). Loading and correlations in the interpretation of principle components. *Journal of Applied Statistics*, **22:2**, 203-214, DOI: 10.1080/757584614
- Cardot, H., Goga, C. and Shehzad, M.-A. (2014). Calibration and Partial Calibration on Principal Components when the Number of Auxiliary Variables is Large. *Xiv:1406.7686v2 [stat.ME]*
- Chang, T. and Kott, P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model, *Biometrika*, **95:3**, 555-571.
- Deville, J.C. and Särndal, C.E. (1992). Calibration estimators in survey sampling, *Journal of the American Statistical Association*, **87**, 376-382.
- Deville, J.C. and Särndal, C. E. and Sautory, O. (1993). Generalized raking procedures in survey sampling, *Journal of the American Statistical Association*, **88:423**, 1013-1020.