

On the calibration approach and multiple imputation in the DG-1 business survey

Maciej Beręsewicz

Center for Small Area Estimation, Statistical Office in Poznań

Department of Statistics, Poznań University of Economics

maciej.beresewicz@ue.poznan.pl

The goal of the paper is to present a calibration approach (Deville and Särndal, 1992; Särndal and Lundström, 2005) and multiple imputation (Rubin, 1976, 2004) in the monthly Report on Economic Activity (DG-1) survey of big (over 49 employees) and medium-sized (10-49 employees) companies. A monthly sample for the DG-1 survey consists of the following: all of the big companies fill the questionnaire and a sample of medium-sized enterprises is selected. In addition, register data about these establishments are used to provide basic information. With the survey completed, sample and register data are linked (using deterministic linkage). The questionnaire covers basic information on the average number of employees, gross salaries and sales revenue and is the basis for monthly business statistics.

The practical goal of the paper is to provide estimates for out-of-sample companies based on information from sample and register data. The research goal is to conduct a comparative study of the calibration approach and multiple imputation with known auxiliary variables from register data. Monthly reports from the DG-1 survey from 2009 to 2014 will be used. Results of the study will be compared with the existing methodology to assess the possibility of implementing the approach in the monthly statistical production. All calculations will be made in **R** (R Core Team, 2014) with additional packages **laeken**, **Amelia**, **mi** and **mitools**.

Keywords: calibration, imputation, prediction, missing data, business survey, reporting.

References

- Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418):376–382.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.
- Särndal, C.-E. and Lundström, S. (2005). *Estimation in surveys with nonresponse*. John Wiley & Sons.