

STATISTICS PRODUCTION SYSTEM 4.0

Thomas Laitila

Örebro University, Sweden
e-mail: thomas.laitila@oru.se

Abstract

Surveys based on the randomization theory have been and are still today a major source for the production of statistics at National Statistical Offices (NSOs). One major problem in the application of the theory is nonresponse. While the theory presupposes full response applications at NSOs turn out with some sample units not responding. The problem of nonresponse has not yet been satisfactorily resolved (e.g. Brick, 2013) and is today a major threat to validity of sample survey statistics. Partly because of this problem and increasing costs of sample surveys, nontraditional data sources are considered by NSOs as an alternative for production of statistics. Such data may also offer interesting features (e.g. Daas et al., 2015; Japac et al., 2015).

There are many challenging problems in using nontraditional data sources in production of official statistics. Issues around accessibility, involving legal and technical aspects, and sustainability are obvious. Validity of statistics considering variable and measurements are also. The most critical problem, the validity, objectivity, and interpretability of produced statistics, are seldom addressed, however. boyd and Crawford (2012) discuss these problems in the context of using Big Data in research.

To my knowledge, there are yet no example of nontraditional data sources replacing traditional ones in an official statistics product. This is remarkable in relation to the world wide interest to do so, and research conducted for at least two decades on how to use nontraditional data sources.

In this paper I suggest the long-time failures of incorporating nontraditional data sources depends on the focus on replacing traditional sample surveys. Instead, the question must be how to integrate nontraditional data sources in the existing production system to improve quality of statistics. This idea is expressed in the conference paper SCB (2017) and is here further elaborated. It summarizes into the Statistics Production system (SP) 4.0.

Keywords: official statistics, selectivity, model based inference, survey design.

References

- boyd, D. and Crawford K. (2012). Critical Questions for Big Data, *Information, Communication & Society*, 15:5, 662-679.
- Brick, J.M. (2013). Unit Nonresponse and Weighting Adjustments: A Critical Review, *Journal of Official Statistics*, 29:3, 329-353.
- Daas, P.J.H., Puts, M.J., Buelens, B. and P.A.M. van den Hurk (2015). Big Data as a Source for Official Statistics, *Journal of Official Statistics*, 31:2, 249-262.
- Japac, L., Kreuter, F., Berg, M., Biemer, P., et al. (2015). Big Data in Survey Research: AAPOR Task Force Report, *Public Opinion Quarterly*, 79:4, 839-880.
- SCB (2017). Statistics productions system 4.0. Contribution to Conference of European Statisticians, Economic Commission for Europe, Geneva, 19-21 June 2017