Quantifying Discontinuities in Time Series obtained with Repeated Surveys

Jan van den Brakel[1,2]

[1] Maastricht University, The Netherlands

[2] Statistics Netherlands, The Netherlands
e-mail: jbrl@cbs.nl

Abstract

Official statistics are often published repeatedly with the purpose of building consistent time series that describe the evolution of finite population parameters. A significant quality aspect of these surveys is the comparability over time of their estimates. This is a major reason to keep the underlying process of the survey unchanged as long as possible. It is inevitable, however, that adjustment or redesign of the process is needed from time to time, as and when the existing procedures become gradually outdated or more cost-effective methods are required. Recently most European national statistical institutes (NSIs) had to implement significant changes in the survey design of their Labor Force Survey (LFS) to meet the new Eurostat regulation for integrated European social statistics. Partly as a result of the COVID-19 pandemic but also from a cost perspective, many NSIs currently consider to move from uni-mode to mixed-mode data collection strategies.

Implementing such changes in a survey process generally affects measurement and selection bias in the responses of the survey, resulting in a systematic shock in the sample estimates. These shocks or discontinuities disturb comparability with figures published in the past. An important aspect of a survey redesign is to quantify the discontinuities in the main outcomes of the survey. In this way it can be avoided that discontinuities are incorrectly interpreted as real period-to-period changes of the population parameters of interest.

Collecting data under the old and new design in parallel for some period of time, time series modelling or a combination of both are established methods to quantify discontinuities. In this paper different approaches will be discussed and illustrated with real life examples at Statistics Netherlands. One example is a multilevel time series model used to quantify discontinuities due to three different survey redesigns in the Dutch Mobility Survey. Time series are modelled for a breakdown of the population parameter in about 700 domains. Predictions at higher aggregation levels are obtained by aggregation of the predictions of these 700 domains. This result in a numerically consistent set of estimates for all target variables, which are corrected for the different discontinuities. In another application it is shown how discontinuities due to a redesign of the Dutch Crime Victimization Survey are estimated on low regional level, using a small parallel run. With a cross-sectional multivariate Fay-Herriot model prediction for discontinuities at the most detailed regional level are obtained. Numerically consistent predictions for discontinuities at higher output levels are obtained by aggregation. In a third example discontinuities are estimated due to the implementation of the Eurostat 2021 regulations in the Dutch LFS. It is shown how a smooth transition in a rotating panel design is accomplished by integrating data from a parallel run with time series data in a multivariate state space model.

Keywords: multilevel time series models, multivariate Fay-Herriot models, survey redesigns, state space models.