
Current work on automatic multisource editing at Statistics Netherlands

Sander Scholtus, Arnout van Delden, Rob Willems, and Frank Aelen (Statistics Netherlands, The Netherlands)
s.scholtus@cbs.nl

Abstract

As part of a new integrated uniform production system for business statistics, Statistics Netherlands aims to develop a system for simultaneous editing of variables that are observed across different data sources. In this way inconsistencies between statistics could be identified and resolved as early as possible. In addition to top-down manual editing of the most influential inconsistencies, we aim to introduce automatic editing of other inconsistencies where possible. An initial approach and results for automatic multisource editing from a pilot study were previously presented in Scholtus et al. (2022).

Currently, there is a project at Statistics Netherlands to develop automatic multisource editing further for use in regular statistical production. This project is supported as part of a grant from Eurostat for timelier, more relevant and more integrated European business statistics. The project will run from January 2024 until June 2025.

Regarding automatic editing across statistics, the main aims of the project are, first, to improve the quality of automatically edited data by introducing better edit rules and incorporating more unit-specific information, and second, to evaluate the effects of automatic editing on the quality of statistical output. As part of the project, a probability sample of units will be edited manually outside of regular production, to provide information both for the development of automatic editing logic and for the evaluation of the quality of automatically edited data. While the project is still ongoing, we will present some ideas and initial developments.

Reference

S. Scholtus, W. de Jong, A. Vaasen-Otten and F. Aelen (2022), Towards a new integrated uniform production system for business statistics at Statistics Netherlands: automatic data editing with multiple data sources. Paper prepared for the UNECE Expert Meeting on Statistical Data Editing, 3-7 October 2022 (virtual).