UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Editing
7-9 October 2024, Vienna Austria

12 July 2024

Selective editing for the production of new Services Producer Price Indices (SPPIs) from indirect data sources.

Simona Rosati, Diego Bellisai, Marco Lattanzio, & Tiziana Pichiorri (Istat, Italy)
sirosati@istat.it

Abstract

To comply with the Regulation EBS 2152/2019 on European business statistics, the Italian National Institute of Statistics (Istat) is gearing up to generate quarterly producer price indices for services pertaining to statistical units primarily engaged in divisions 74 and 82 of the Statistical Classification of Economic Activities, Nace Rev. 2. These indices aim to measure quarterly fluctuations in business-to-business prices for services, focusing exclusively on transactions between businesses, thereby excluding sales to consumers of goods and services.

The estimation of these indices relies on hourly labor cost data from the OROS short term survey, which in turn takes these data from the Italian National Social Security Institute (INPS) administrative archives.

However, administrative as well as survey data may contain errors, such as measurement errors, that can lead to biased estimates when data are used for statistical purposes. Given the nature of the variables involved (labor cost, regularly paid hours worked) and the required timeliness between data availability and their dissemination, we adopted a selective editing approach to identify outliers and influential errors. Specifically, we employed a method based on contamination normal models. This method is implemented in the R package SeleMix (Guarnera and Buglielli, 2013), which was developed at Istat. SeleMix aims to detect units with the most influential values, i.e., potential errors with the highest impact on the target estimates. Suspicious outliers and influential errors are then flagged for manual review by subject matter experts.

The application of the SeleMix method to target estimates, which here are Laspeyres indices, represents the innovative aspect of this work.

A comparison between the two time series of the indices, one representing raw data and the other representing correct data, revealed a more regular behavior in the latter ones.