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

Using hidden Markov and macro integration models for combining data from different sources

Nino Mushkudiani, Jeroen Pannekoek and Sander Scholtus (Statistics Netherlands, The Netherlands)

n.mushkudiani@cbs.nl

Abstract

Statistical agencies increasingly need to devise methodologies to make data from different sources consistent. When the results from different sources on the same statistics are (systematically) inconsistent with each other, the question arises whether we understand the reasons for those differences and how we can arrive at one consistent estimate.

In this article we limit ourselves to an experiment on the integration of data from a sample survey (the Labor Force Survey) and a register (the Employment Register). Both data sources contain information on labor market variables with different definitions, different population coverage and different frequencies. In previous research data were analyzed using macro integration and latent class analysis with hidden Markov models. Even though two different models did not lead to very different estimates of the time series of temporary (or permanent) employment contracts, still there are clear differences in the outcomes of the two approaches. Motivated by the need to better understand these differences, we use synthetic data (inspired by the real data), where the (systematic and random) measurement errors are known. Using this synthetic data differences in outcomes of the two approaches are tested under different circumstances. Here we present the results of this simulation study.