

## **Replicating the Synthetic LBD with German Establishment Data**

Jörg Drechsler\*

Institute for Employment Research, Nuremberg, Germany, joerg.drechsler@iab.de

Lars Vilhuber

Cornell University, Ithaca, NY, USA, lars.vilhuber@cornell.edu

One major criticism against the use of synthetic data has been that the efforts necessary to generate useful synthetic data are so intense that many statistical agencies cannot afford them. However, we argue in this paper that the field is still evolving and many lessons that have been learned in the early years of synthetic data generation can now be used in the development of new synthetic data products considerably reducing the required investments.

To this extent we evaluate in this paper whether synthetic data algorithms that have been developed in the U.S. to generate a synthetic version of the Longitudinal Business Database (LBD) can easily be transferred to generate a similar data product for other countries. We construct a German data product with information comparable to the LBD – the German Longitudinal Business Database (GLBD) – that is generated from different administrative sources at the Institute for Employment Research, Germany. In a second stage, the algorithms developed for the synthesis of the LBD will be applied to the GLBD. Extensive evaluations will illustrate whether the algorithms provide useful synthetic data without further adjustment.

The final goal of the project is to provide access to multiple synthetic datasets similar to the SynLBD at Cornell to enable comparative studies between countries. The GLBD is a first step towards that goal.

Keywords: confidentiality, comparative studies, German Longitudinal Business Database, synthetic data