Assessing the data quality of public use tabulations produced from synthetic data: Synthetic Business Dynamics Statistics

Javier Miranda[1] and Lars Vilhuber[2]

[1] U.S. Bureau of the Census, Washington, DC, USA javier.miranda@census.gov
[2] Cornell University, Ithaca, NY, USA, lars.vilhuber@cornell.edu, Corresponding author

Abstract. We describe and analyze a method that blends records from both observed and synthetic microdata into public-use tabulations on establishment statistics. The resulting tables use synthetic data only in potentially sensitive cells. We describe different algorithms, and present preliminary results when applied to the Census Bureau's Business Dynamics Statistics and Synthetic Longitudinal Business Database, highlighting accuracy and protection afforded by the method when compared to existing public-use tabulations (with suppressions).

This version: $Id: synbds-noise-synthetic-jos2015.tex 1657 2015-02-01 00:28:47Z lv39 $