

## **The Synthetic Longitudinal Business Database (SynLBD): Access to Data and Remote Processing**

Ron S. Jarmin, Thomas A Louis\*, Javier Miranda, and Lars Vilhuber

**ABSTRACT:** In most countries, statistical agencies do not release business microdata due to the risk of re-identifying information collected from businesses under a pledge of confidentiality. This practice severely limits the range of analyses that can be conducted with public-use business data compared to what's possible with public-use household data. One potential approach for overcoming this shortcoming, while ensuring confidentiality protection is to release synthetic data. Here, establishment data are simulated from statistical models designed to mimic the distributions of the real, underlying microdata. The US Census Bureau's Center for Economic Studies in collaboration with Duke University, the National Institute of Statistical Sciences, and Cornell University makes available a synthetic public use file for the Longitudinal Business Database (LBD), an annual economic census of establishments in the United States comprising more than 20 million records dating back to 1976. This paper reviews SynLBD access protocols and discusses options for automated remote processing; a form of access that limits Census Bureau interaction to proposal review and output disclosure. We also discuss the use of cross country synthetic data access.

**KEYWORDS:** synthetic data, statistical modeling, business microdata, confidentiality protection.

\* thomas.arthur.louis@census.gov, Associate Director for Research and Methodology and Chief Scientist, U.S. Census Bureau, Washington D.C., USA.