25th Meeting of the Wiesbaden Group on Business Registers
- International Roundtable on Business Survey Frames

Tokyo, 8 – 11 November 2016

David Talan, Kevin Cooksey, and Akbar Sadeghi
Bureau of Labor Statistics
Session No. _4

Quality and Coverage

Measuring Data Quality in the BLS Business Register: A Local Regression Model Approach

Abstract

Of the numerous concerns regarding large datasets, data integrity is primary.

The U.S. Bureau of Labor Statistics (BLS) maintains the Longitudinal Database (LDB) which serves as the Bureau's Quarterly Census of Employment and Wages (QCEW) establishment-based business register. The LDB is a relational database of over 9 million business establishments linked longitudinally. The database is built on microdata from Unemployment Insurance tax forms, submitted from the States on a quarterly basis. Data elements on these forms include information on monthly employment, quarterly wages, business name and addresses, industry classification, geocodes, and other administrative data. Every establishment on the database contains a unique identifier that allows for tracking of individual establishments at the micro-level across quarters for the United States.

The LDB serves three critical functions. First, the database allows for the publication of longitudinal Business Employment Dynamics (BED) statistics. Second, the LDB serves as a sampling frame for establishment-based surveys for BLS. Finally, the LDB serves as an invaluable resource for labor market research.

These data are used to generate high quality, high frequency, timely and historically consistent data on business and employment. As a result, BLS has developed measures to quantify the quality and monitor the integrity of these data.

This paper develops a statistical framework for measuring and monitoring data quality in the QCEW Business Register over time. The business register data quality, as a result of fiscal year 2014 budget cuts, will be examined. The paper first provides a brief history of the program and describes a recent budget cut. Next, the authors describe the various data quality measures and dimensions available for consideration. Then, the authors fit a series of local regressions to smooth the path of changes in a number of quality metrics including a composite index to measure the overall data quality. Finally the authors discuss results of the analysis and the lessons learned.