

Data Quality Issues

John M. Abowd
John Haltiwanger
Bryce Stephens

Overview

- Confidentiality suppressions in the detailed industry sub-state QWIs
- Differences between the QWIs and the QCEW
- Geographic and industry fluctuations in the QWIs
- Longitudinal establishment linkages in the LEHD infrastructure files
- Next steps

Confidentiality Suppressions

- Values are suppressed more often in the detailed tables (status codes 2, 3, and 4)
- These suppressions make the creation of custom aggregates and time series difficult
- Particularly true for status code 3

Research

- LEHD staff have developed a sophisticated fix for this problem that is being programmed now
- The fix replaces the missing data element with a good estimate of its expected value
- Suppressed counts and ratios are fixed
- No results to report yet
- Entire procedure uses the released data and does not require any changes to the approved confidentiality protections for the QWIs

Differences between the QWIs and the QCEW Data

- Important to assure that the correct universe is used in the comparison
- LEHD staff are assembling a comprehensive comparison of the employment and total payroll series in QWIs and QCEWs
- Some differences are expected
 - Slight differences in employment concepts
 - Differences in the establishment geocoding and industry coding procedures
- Major differences will be catalogued and documented

Geographic and Industry Fluctuations in QWIs

- Quarterly quality assurance of the QWIs identifies unusual fluctuations in employment at sub-state geography or detailed industry
- These are traced to problem micro-data
- Algorithmic deficiencies are corrected when possible
- State is alerted to issue

Research

- An experienced evaluation team (Bryce Stephens and Lars Vilhuber) work with the production staff (primarily Lou Schwarz and Pete Welbrock) to detect and correct these problems
- Future modifications include more automatic detection of micro-data issues
 - New states = new data issues
- Algorithms are constantly being improved

Longitudinal Establishment Linkages in the LEHD Infrastructure Files

- Currently use many good methods to establish these linkages
- Births, deaths, mergers, divestitures and complex recombinations are all problematic
- Continuing efforts to find other data sources to improve these linkages

Research

- Some comparisons of the longitudinal linkages have been done using the LEHD Business Register Bridge
 - connects ES-202 establishments to Census establishments
- These provide alternative/additional industry and geography data to help with the longitudinal linkages

Next Steps

- Release the suppression fix for testing and evaluation by 2005:3
- Set priorities for R&D efforts required to move research into QWIs