LED Data Quality: Selected Issues from State Data Submissions

Stephen Tibbets
CES/LEHD
U.S. Census Bureau

Nelse Grundvig
Employment Security
Commission of North Carolina

Selected LED Data Quality Issues

- State data inputs are the critical pieces that make the LED partnership possible.

- This discussion will look at two common issues that can have a significant impact on LED data quality.
  - Completeness of QCEW and UI wage data submissions: preliminary|second
  - Concordance between QCEW and UI wage data

How State Data Are Combined in LED Processing

Quarterly Census of Employment and Wages:

Firm and Establishment Level (Single/Multi-unit)

Geography
Industry
Ownership

UI Account Number:

Firm Level

(SEIN)

Unemployment Insurance Wage Records:

Firm-Worker Level *(Usually)*

Wages
Job history
Link to demography

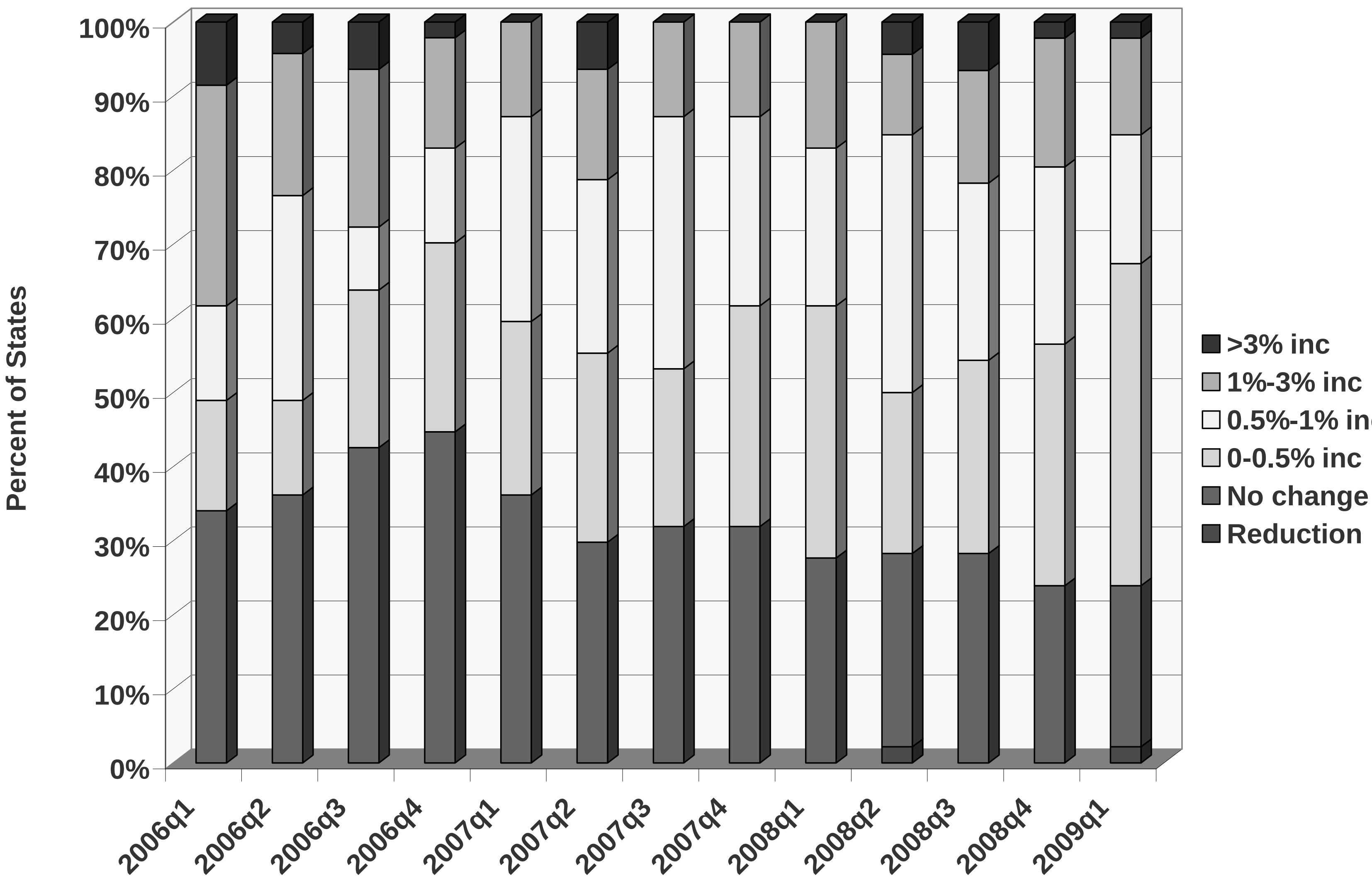
Timing of Data Submissions

- LED Standard Operating Procedure
  (Both QCEW and UI wage data)

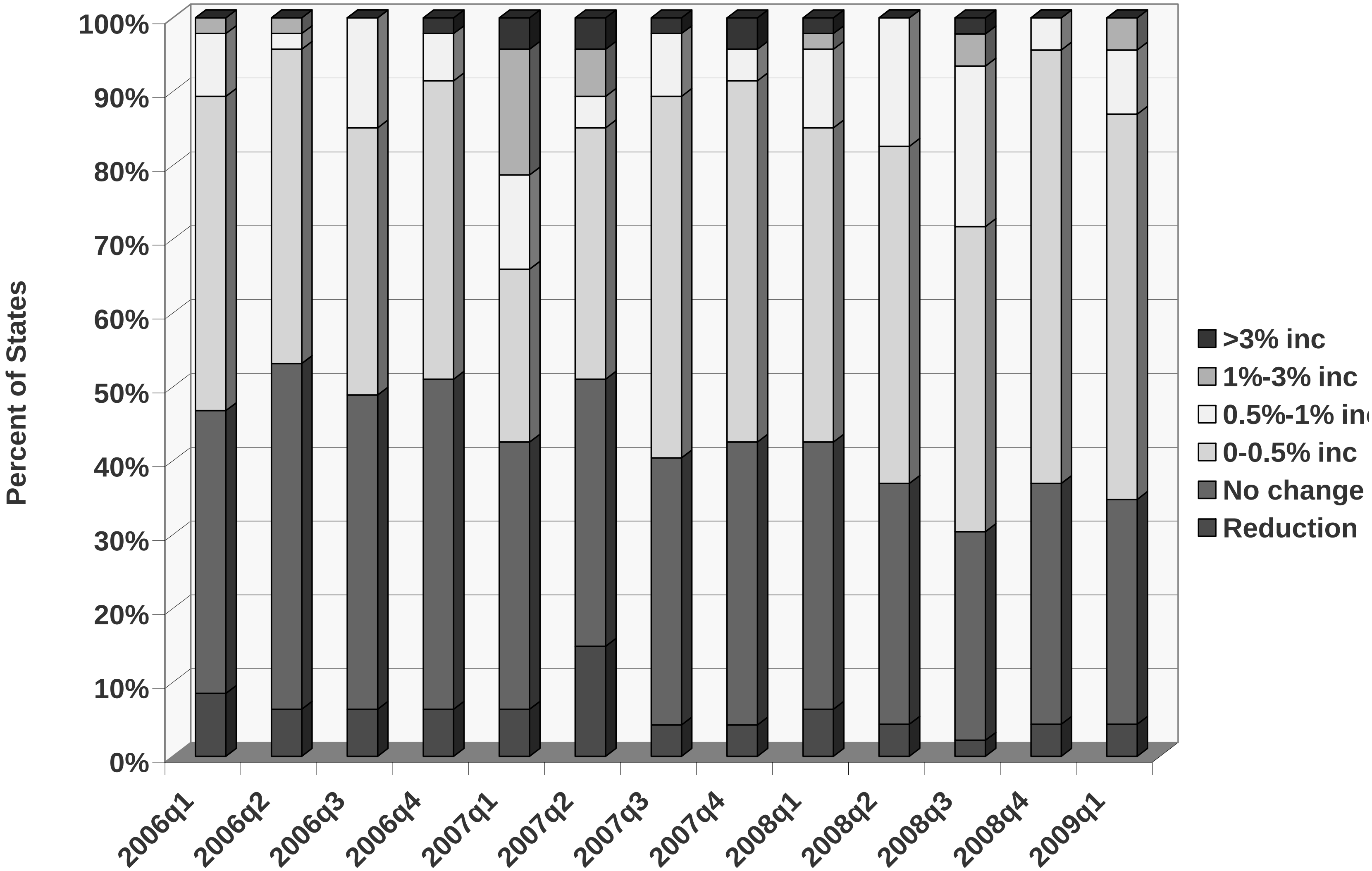
- Preliminary submission – 6 months after completion of quarter
  – On (or about) April 1, 2010 states will submit data for 2009 Q3.

- Second submission – 9 months after completion of quarter
  – On (or about) April 1, 2010, states will resubmit updated data for 2009 Q2.

Change in Record Count, First to Final Submission: UI Wage data



Change in Record Count, First to Final Submission: QCEW data



Comments on Submission Charts

- Increases represent the percent change in record count between the initial and final submissions of data to LED.

- “No Change” can reflect both instances when identical files are submitted, as well as when only one submission was received.

- File size does increase significantly with second submission.
  - There is typically more growth in UI wage record files than QCEW between submissions.

Importance of Maximizing Completeness of Input Data

- A small overall percentage of missing records may be concentrated in certain industries or geographies
  - Even small changes to overall counts can have very significant impacts on smaller cells.
- The impact on employment flow measures can be magnified several times.
  - Jobs at nonreporting firms will be reflected in separations, which has a much smaller base than total employment.
  - Again, in the smaller cells these issues become even more significant.

Concordance Between QCEW and UI Wage Records

- There is always some degree of mismatch between data sources.

  – QCEW only
    - Especially, Federal employment
    - Can be 5-10% of total employment

  – UI wage only
    - Much lower, <2% of total employment
    - Unclear when records “should” be UI-only, frequently suggests reporting irregularities

What may cause QCEW-UI mismatch?

- Firm did not report to one data source (or reported late).
- Firm did not report at all, and the state imputes QCEW record.
- Successor-predecessor issues:
  - Firm account number may change on one data source one (or many) quarters before the other.
- Public sector (especially state governments) pose particular challenges.
  - Some states report using alternate account numbers and different levels of aggregation to each system.

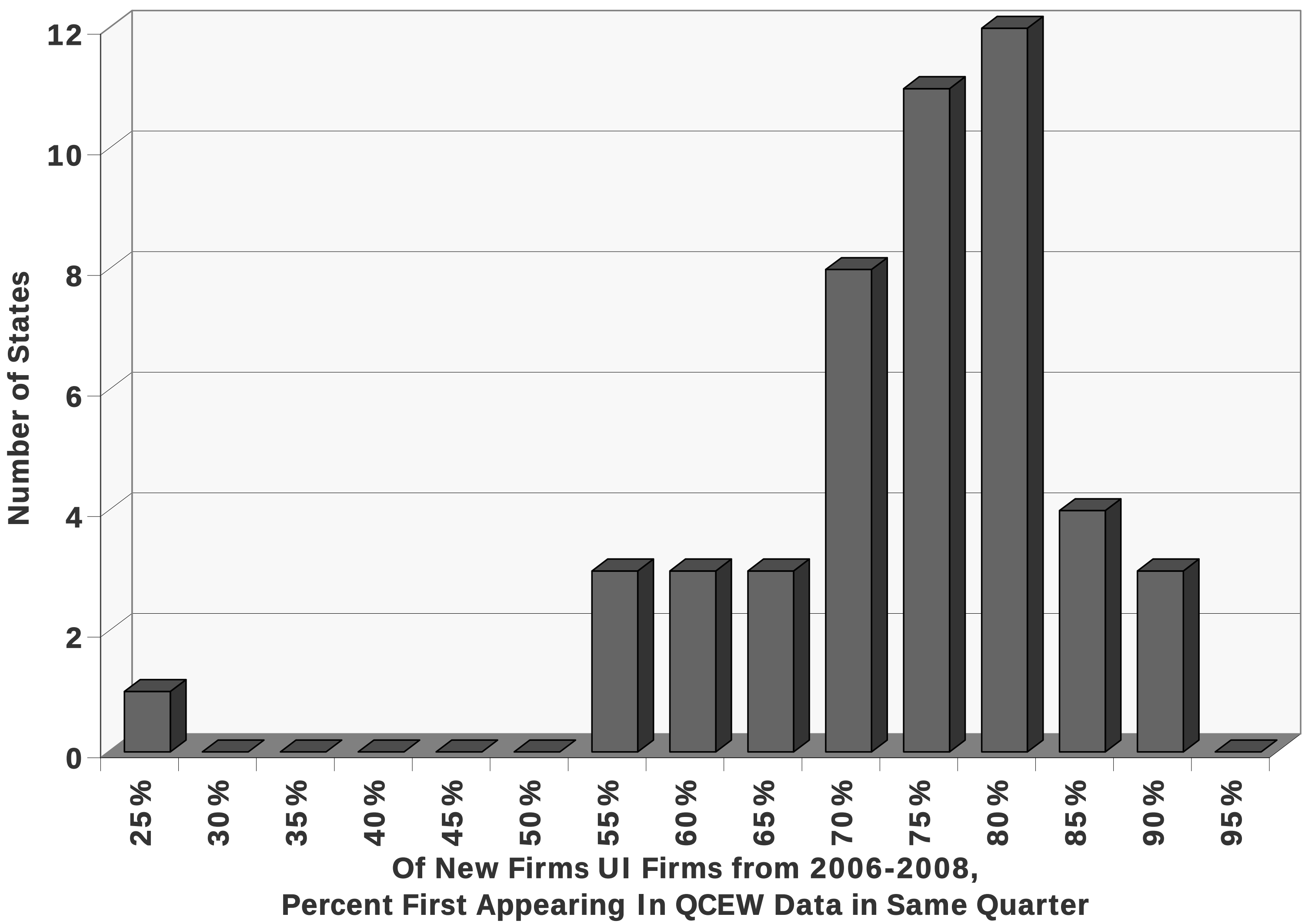
What happens when there is a mismatch?

- UI-only firms
  - If QCEW data is never available, firm data is dropped from QWI measures.
  - If QCEW data is available in another quarter, various automatic edits/imputations are applied to generate necessary establishment-level information.
    - Multi-unit firms can become difficult,especially when a long-term mismatch is involved.

- QCEW-only firms
  - Firm-level data, with no wage record information, are always ultimately dropped. The LEHD infrastructure is built from wage records.

Frequency First Appearance in QCEW and UI Occurs in Same Quarter

New Firms in UI Wage Data, 2006-2008

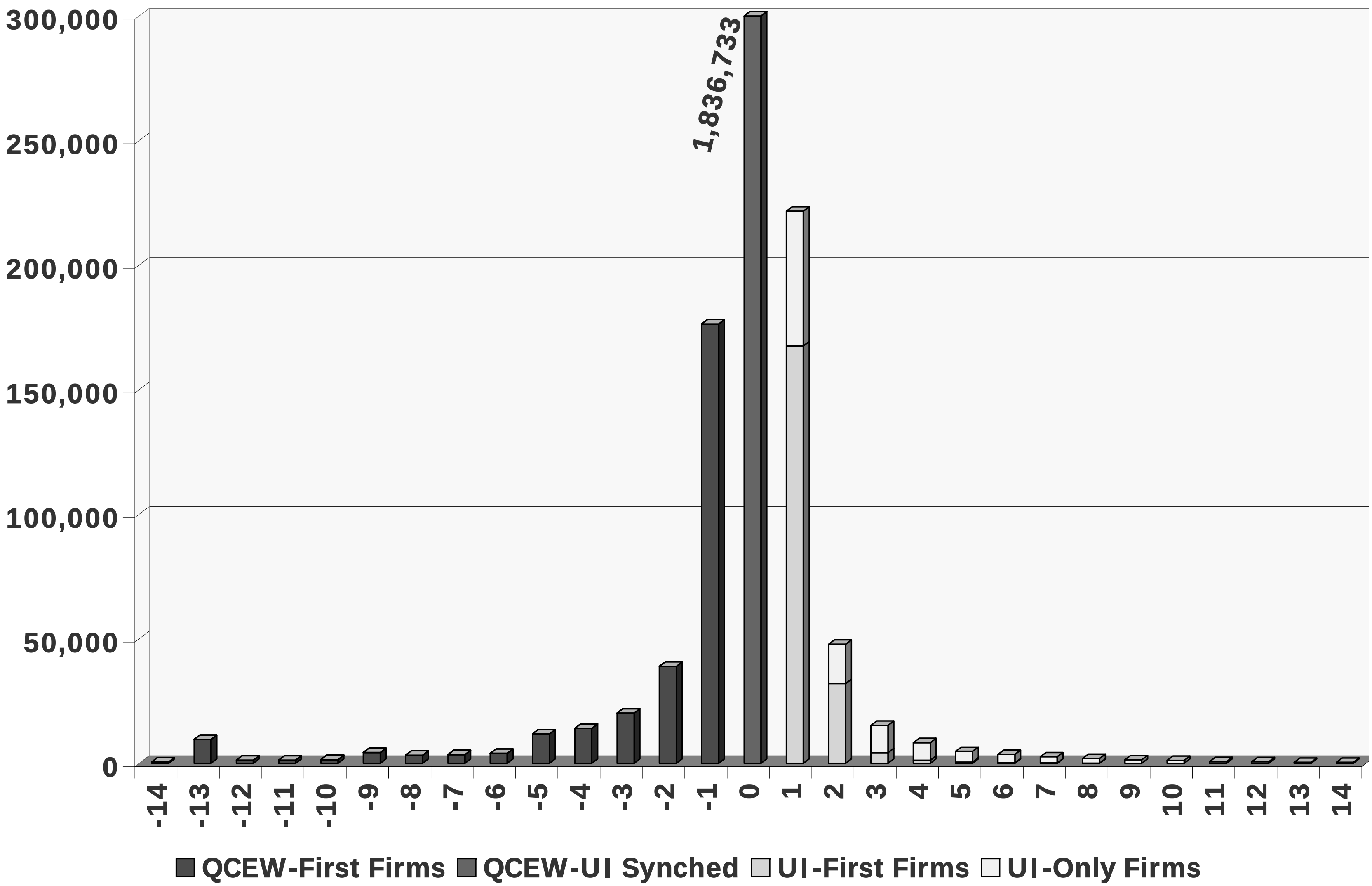


Comments on QCEW-UI Match

- In an ideal world, 100% of firms that appear in the UI wage data would also first appear in the QCEW in the same quarter.
- The greatest numbers of states are in synch for only 70-85% of new UI firms.
- The QCEW data series for a mismatched firm may have started before, after, or never.

Quarters Between First Appearance of Firm in QCEW and UI Wage Data *(QCEW Lag)*

New Firms in UI Wage Data, 2006-2008



Comments on QCEW Lag Chart

- Negative values (purple) indicate firms that appear in the QCEW before UI.
- The zero column, (blue) has been truncated for readability. The actual value is displayed.
- The green series indicates firms for which UI data appeared some number of quarters before QCEW.
- The yellow series indicate firms that appeared in the UI data for one or several quarters, but never in the QCEW.

Comments on QCEW Lag Chart

- Most firms which do not begin in the same quarter on both sources do appear within a year before or after first UI appearance (or cease reporting altogether).

- Still, there are very long tails on either end.
  - A large firm in the long tail, especially, suggests a significant and persistent local data quality issue.

Case Study: North Carolina

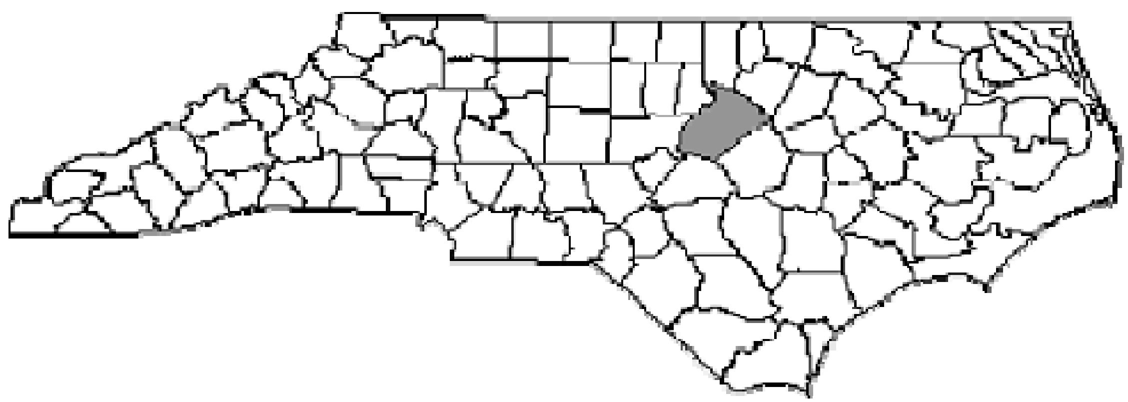
- In 2009, LEHD quality assurance tests noted an unusual jump in UI-only firms in recent submissions.
- This was referred to the state for investigation, and it was found to be a public sector issue.
- Because of the extent of the issues with the data, it was agreed that production would be halted until data could be corrected.

State Perspective: Systemic Issues

- North Carolina’s State Controller initiated a new payroll system.
  - The new system integrated several separate systems into a single system.
  - System was implemented with few checks and no review of downstream consequences
  - It destroyed the existing pattern of assigning state employees to the appropriate county for their work location used by the QCEW.
  - Made reporting by location by agency difficult, effectively eliminating our ability to report industrial employment patterns in state government

State Perspective: Consequence

- While QCEW statewide totals were correct, place of employment were now based out of Wake County for all state employees.
  - Disrupting our ability to report commuting patterns and substate employment



State Perspective: Possible Lessons

- Technology linking more databases may have unintended consequences
  - Administrators and others may not even be aware of who may be influenced by programmatic changes

- Using secondary data carries certain risks
  - The intent of the data must be remembered
  - Administrative changes can disrupt time series
  - Firms may only report what is legally required but it may not be accurate
  - Multiple forms, data entry and interventions can create error

- Improving coordination between QCEW and UI may require additional commitment and resources
  - The structure of the data may prevent a proactive response when editing or cleaning data
  - Responding reactively with the current process can be a difficult and time consuming activity.
  - Mechanisms used to report errors can make finding and repairing or reporting on issues problematic

What can be done to improve data completeness?

- Data submissions
  - Whenever possible, we continue to request that states make a second submission of both QCEW and UI wage data to capture late reporters.
    - Unless the UI database has been truly frozen, there are very likely new records that would improve the QWI's.
  - Are two submissions enough?

What can be done to improve concordance?

- State actions
  - Improve coordination between QCEW and UI wage systems moving forward
  - Correct and resubmit historic data, when possible

- LEHD actions
  - Feedback/quality loop
    - Would a state-specific report on data quality be useful?
  - If all else fails, develop crosswalks when necessary to ameliorate historic issues.
    - Labor intensive effort, requires state advice and consent.

Contact us:

- LEHD

  ces.local.employment.dynamics@census.gov

- Stephen Tibbets

  stephen.r.tibbets@census.gov

  (301) 763-2125