LED Data Enhancements

Introducing New Quarterly Workforce Indicators by Worker Race, Ethnicity, and Education

LED Partnership Workshop
Arlington, VA
March 2011

Erika McEntarfer
LEHD Economic Research Group
Center for Economic Studies, US Census Bureau
Overview

- As part of the LED budget initiative, LEHD staff have been working on multiple enhancements to the microdata and public use data products
  - Enhancing core microdata with additional information on workers and firms
  - Expanding data coverage (federal jobs, self-employed)
  - New product development (job-to-job flows)
- Today will introduce the first deliverable from the budget initiative and discuss upcoming enhancements
New LED data w/race, ethnicity, and education now being released

• First set of deliverables from the LED budget initiative
  – New QWI tabulations by race & ethnicity, education by sex
  – New race, ethnicity, & education data in OnTheMap

• Enormous effort over the last year by the LEHD staff to get the new variables in the microdata and into the production code
New race & ethnicity variables in QWI and OnTheMap:

• Detailed employment indicators by six race variables
  – White alone
  – Black or African American Alone
  – Asian Alone
  – Native Hawaiian or Pacific Islander Alone
  – American Indian or Alaska Native Alone
  – Two or More Races

• And two ethnicity variables
  • Hispanic or Latino
  • Not Hispanic or Latino

• Full cross-tabulations for QWI, some categories collapsed for OTM
# Adding race and ethnicity to the LEHD microdata

## Individual Characteristics File (ICF)

<table>
<thead>
<tr>
<th>PIK</th>
<th>Decennial Race</th>
<th>Decennial Ethnicity</th>
<th>SSA Race</th>
<th>QWI Race</th>
<th>QWI Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person1</td>
<td>White alone</td>
<td>Not Hispanic</td>
<td>White</td>
<td>White alone</td>
<td>Not Hisp</td>
</tr>
<tr>
<td>Person2</td>
<td>Black alone</td>
<td>Not Hispanic</td>
<td>Black</td>
<td>Black alone</td>
<td>Not Hisp</td>
</tr>
<tr>
<td>Person3</td>
<td>.</td>
<td>.</td>
<td>Hispanic</td>
<td>White alone</td>
<td>Hispanic</td>
</tr>
</tbody>
</table>

When no Decennial response, use Census Numident race as conditioning variable for OMB compliant variable impute.

Use Decennial race & ethnicity when available.
QWI Example: Has the recession impacted different workers differently?

Impact of Recession on Employment in Ohio by Race & Ethnicity

- White, Non-Hispanic
- Black, Non-Hispanic
- White Hispanic

Seasonality dominates Hispanic trends

African-American employment 89% of 2007 level
Education in QWI and OnTheMap

- New release has detailed industry and geography employment indicators by 4 education classes
  - Less than a high school diploma
  - High school graduate, no college
  - Some college or associates degree
  - Bachelor’s degree or greater

- Education data are for workers age 25 and older only
  - Education status is dynamic, particularly before age 25
  - Needed a static education variable given limited source data
  - Older cut-off for OnTheMap
Why only four education categories?

- No education registry or census captures educational attainment for the entire U.S. population
  - Limited survey data on educational attainment (Long Form, ACS, SIPP, etc)
  - Use state of the art imputation methods to capture small-area estimation properties
  - Need to keep set of education categories at a higher level of aggregation to make estimation feasible
QWI Example: Recession impact on employment by education.

Impact of Recession on Employment in Ohio by Education Category

Employment of high-school graduates in 2009:3 93% of 2007 level
QWI Example: What are the highest paying industries for workers with college degrees in my state?
QWI Example: Comparison of average wages for skilled workers across states.

Comparison of Annual Wages in Selected Industries for College Educated Men in Ohio and New Jersey

Comparison diagram with industries and wage comparisons.
Imputing education when no survey response can be found

- Use information from workers with complete education data
  - Classify or assign each worker to a given cell (a particular combination of a set of known characteristics such as sex, age, industry, firm size, etc.)
  - Within each cell we estimate the distribution of education using the workers with complete information
- The resulting distribution is then used in our sampler to generate implicates
Checking the quality of the education impute

• Great care was taken in developing the imputation model and stratifying variables
• Will discuss in detail one test of quality of result:
  — Use sample with reported education. (‘D Sample’)
  — Compare key QWI statistics for this sample using reported education and compare to those using imputed education
  — Next slides show a sample of key results
Illinois: Distribution of Employment by Education

<table>
<thead>
<tr>
<th>Education Level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than High School</td>
<td></td>
</tr>
<tr>
<td>High School Graduate</td>
<td></td>
</tr>
<tr>
<td>Some College or Associates Degree</td>
<td></td>
</tr>
<tr>
<td>College Graduate or Greater</td>
<td></td>
</tr>
</tbody>
</table>

New Jersey: Distribution of Employment by Education

<table>
<thead>
<tr>
<th>Education Level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than High School</td>
<td></td>
</tr>
<tr>
<td>High School Graduate</td>
<td></td>
</tr>
<tr>
<td>Some College or Associates Degree</td>
<td></td>
</tr>
<tr>
<td>College Graduate or Greater</td>
<td></td>
</tr>
</tbody>
</table>

Delaware: Distribution of Employment by Education

<table>
<thead>
<tr>
<th>Education Level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than High School</td>
<td></td>
</tr>
<tr>
<td>High School Graduate</td>
<td></td>
</tr>
<tr>
<td>Some College or Associates Degree</td>
<td></td>
</tr>
<tr>
<td>College Graduate or Greater</td>
<td></td>
</tr>
</tbody>
</table>
Delaware: Wages by Sex and Education

Average wage by reported education

Average wage by imputed education (average over implicates)
Summary of results

- Comparisons generally show close correspondence between estimates using reported education and imputed education.
- At the statewide level:
  - The difference between average full-quarter wages ranges from -6.7% to +6.4% with the smallest difference being 0.2%.
  - The share of beginning of quarter employment varies by a range of -4.9 to 6.4 percentage points with the smallest difference being -0.1 percentage points.
- Results for full-quarter employment generally mirror those for beginning of quarter employment.
- Differences in wages and employment by gender and geography are captured in estimates using the imputed variable.
Upcoming enhancements

• Federal Workers in QWI and OTM
  – Very soon
  – Integrating OPM data
  – Release of DC QWI and OTM

• New LED data on firm age and size
  – Age of firm (new firm, old firm)
  – Size of firm (small firm, large firm)
  – Due fall 2011.
Upcoming enhancements

• Integration of self-employment data
  – Add to OnTheMap to see where self-employed workers are
  – Due to frequency and timing issues, planning only to release in OnTheMap

• New product: job-to-job flows
  – Identify flows of workers between firms, industries, geographies
  – Better understand how labor markets adjust to trade, demand shocks, where new workers come from etc.
Contact Us

Comments/Questions
Erika.McEntarfer@census.gov

Center for Economic Studies
http://www.ces.census.gov/

Longitudinal Employer-Household Dynamics
http://lehd.did.census.gov