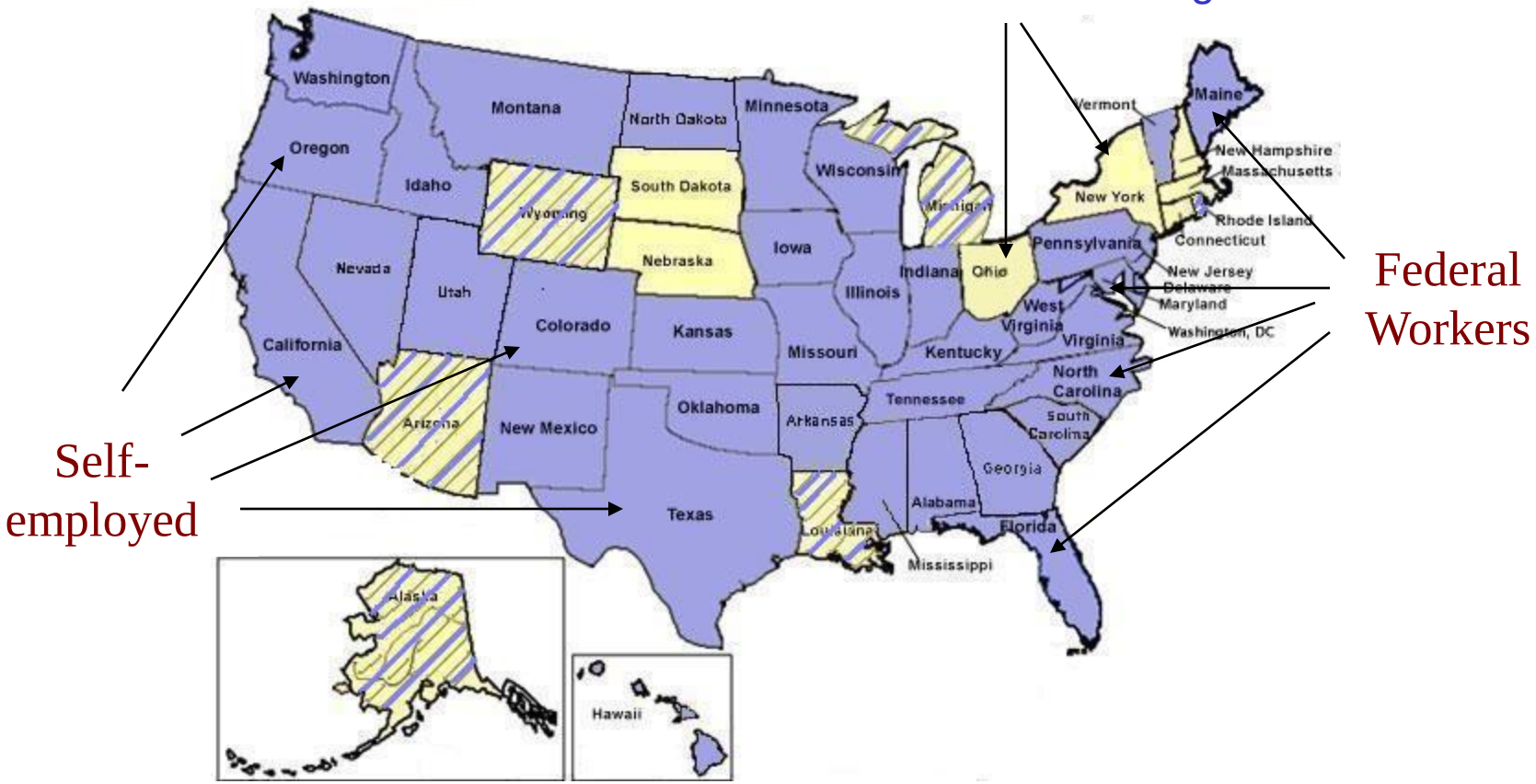
Looking Ahead:
Research and Development at LEHD

Erika McEntarfer
Kristin Sandusky
Fredrik Andersson
LEHD, U.S. Census Bureau

2007 State Workshop
Brookings Institution, Washington DC

National QWIs with Universal Coverage: Coming Soon.

New York & Ohio bring LEHD much closer to covering entire US.



National QWIs with Universal Coverage: Why it is important

Brings us closer to our goal of longitudinal data on entire US workforce:

- Examine employment flows across state lines.
- Account for flows of workers in and out of self-employment.
- Identify flows of workers in and out of the federal workforce.

Overview of this session

Two R&D projects in depth:

- Self-Employed Workers
- Daytime Population Estimates

Overview of Other Projects.

- Adapting the current production system to a national frame
- Employer-to-Employer Flows
- Integrating OPM federal worker file

Two Current Research and Development Projects:

- Integration of Self-Employed Workers into the QWI
- Estimating the Daytime Population of the U.S.

Expanding QWI Coverage: Including the Self-Employed

Why are the Self-Employed Important?

Large Group:

Over 9% of workforce has some ties to self-employment

Changing economy:

Increased numbers of Internet based businesses and contract workers

General Interest:

Universal measures of entrepreneurship needed

Finding Data on Self-Employed: From Business Data?

Business Register

Bureau's sample frame for censuses and surveys of businesses.

Data from tax reports to IRS for businesses

Data are annual.

Moving from Annual Business Data to Quarterly Worker Data:

First Key Challenge: Who is Self-Employed?

1. Partnerships and Corporations

Difficult to link to a set of individuals.

2. Sole Proprietors

Data from IRS 1040 Schedule C – Business ID is SSN

Not married and filing jointly - SSN is sole prop

Married joint filers - determine which spouse is self-employed using supplemental data from SSA

Moving from Annual Business Data to Quarterly Worker Data:

Second Key Challenge: Who are Active Sole Proprietors Each Quarter?

Must impute quarterly self-employment pattern

Many possible sources of data for imputation model among household surveys

Moving from Annual Business Data to Quarterly Worker Data:

Third Key Challenge: Measuring Earnings

1. Annual Self-employment earnings sometimes available

2. In other cases, we know annual receipts (after removing spending on payroll and costs of good sold)

3. In these cases, must impute quarterly self-employment earnings

An Early Look….

Select two years that have supplemental SSA data:
1992 and 1997 (economic census years)

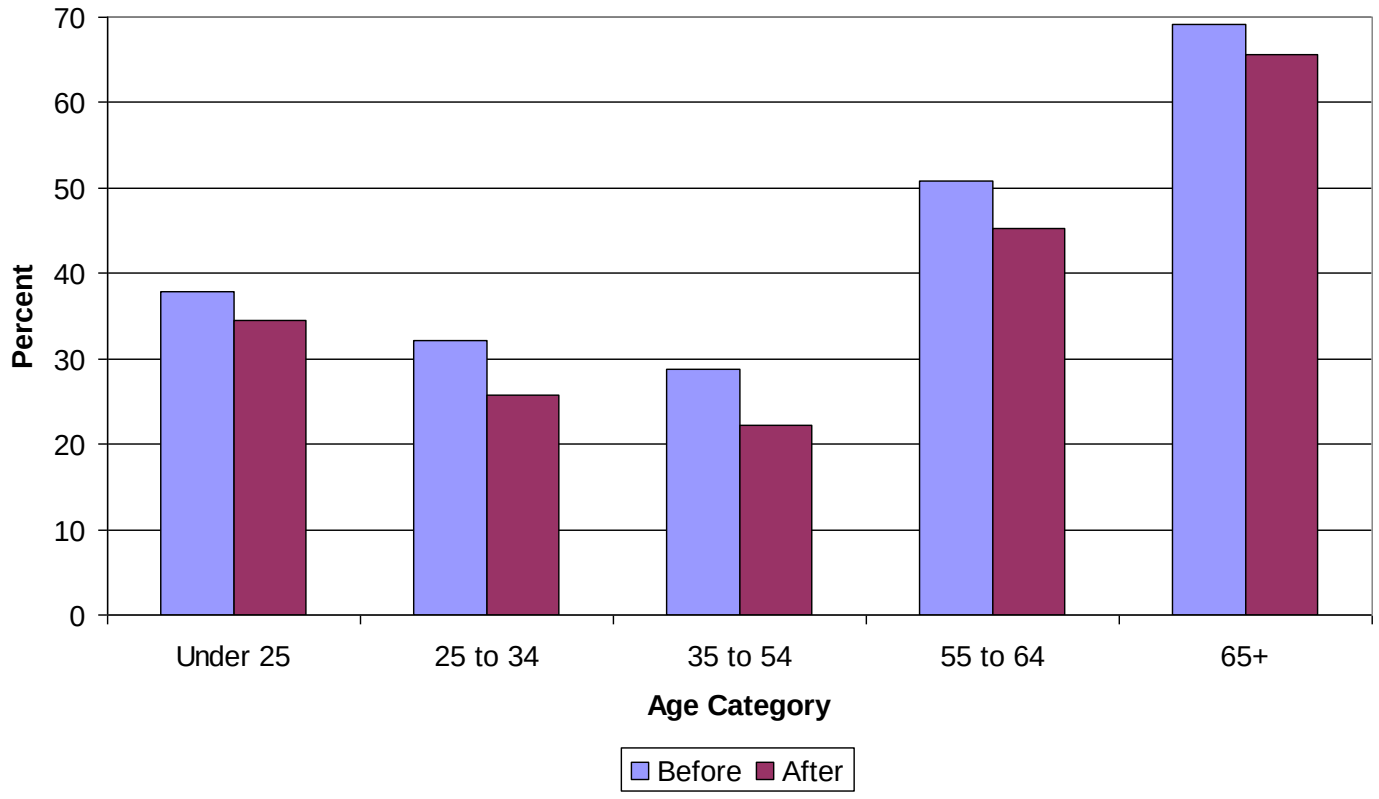
Link via SSN (or PIK) to annual wage and salary data for a few states.

Use linked data to answer some interesting questions

Example 1:

Will adding data on self-employment spells impact our measures of labor market transitions?

What Fraction Leave Labor Force?
With and Without Self-Employment



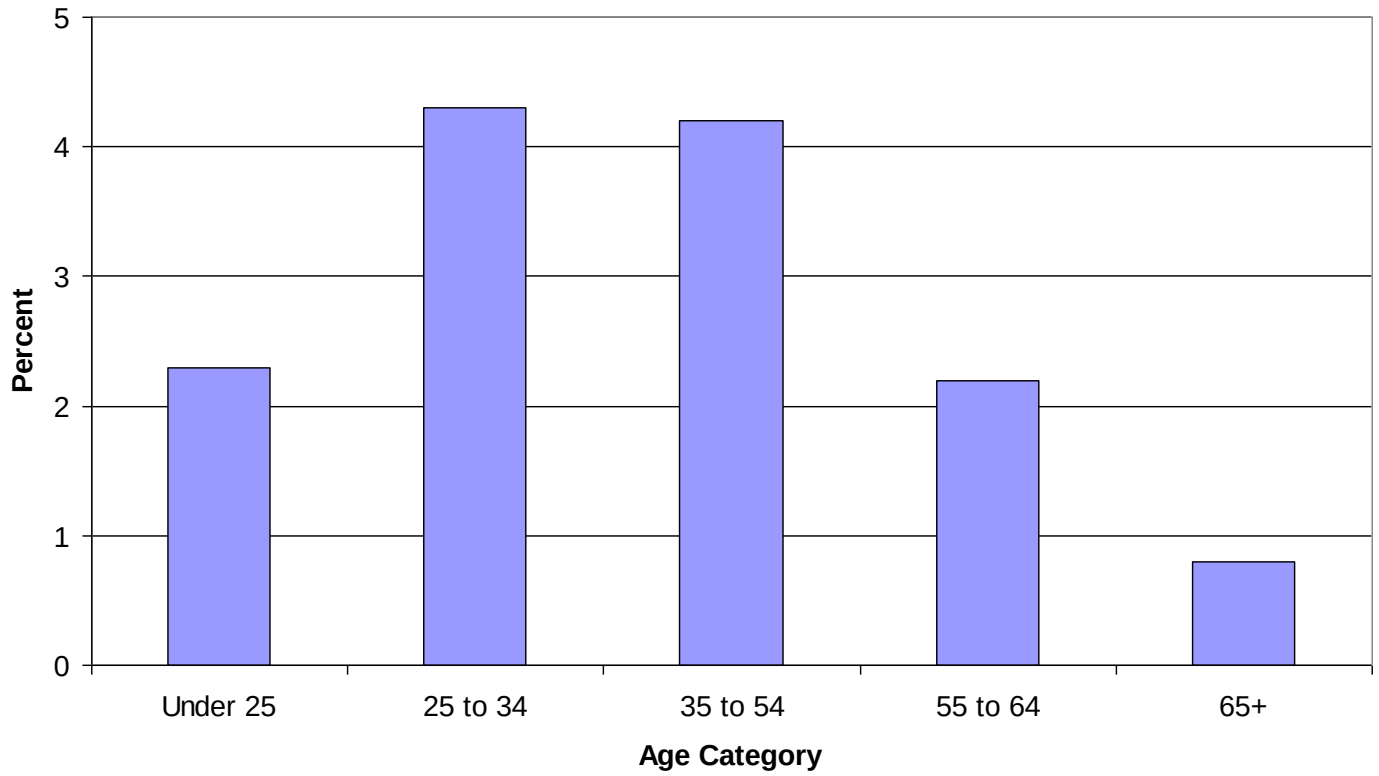
Example 2:

Among wage and salary workers, what fraction "try out" self employment?

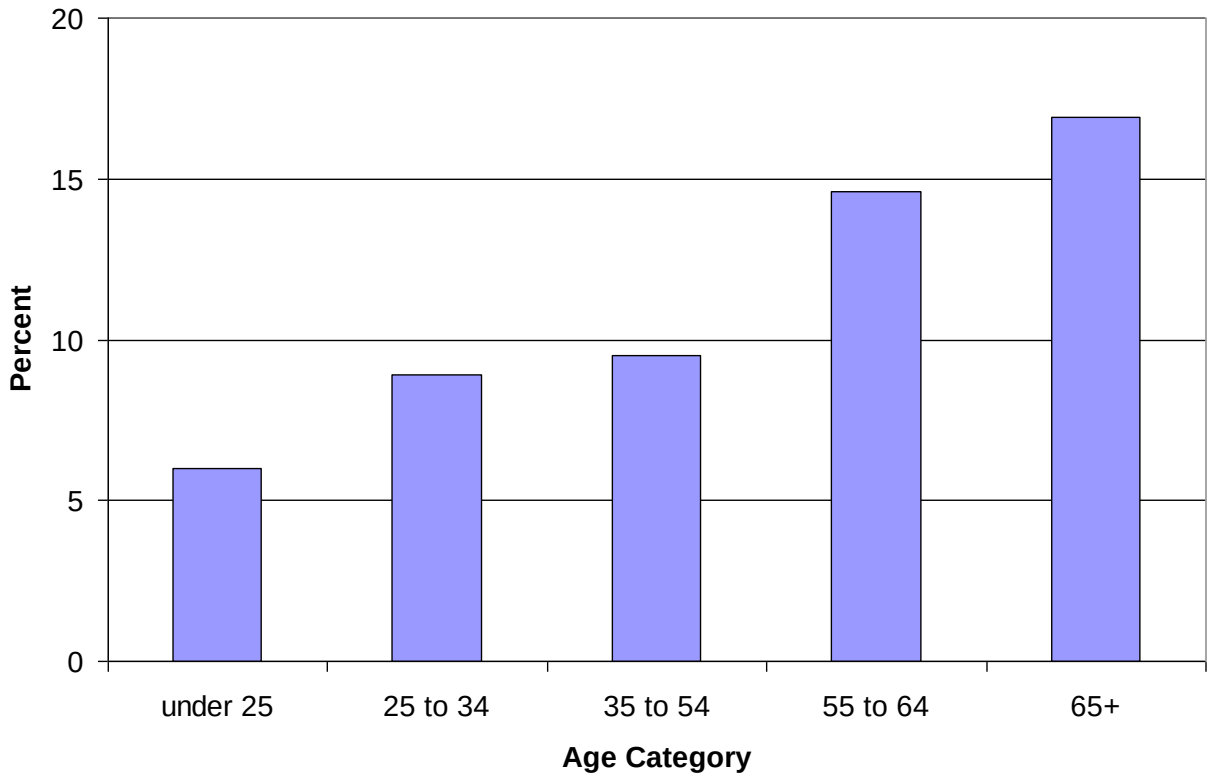
Among workers “trying out” self-employment, what fraction transition to full self-employment?

How do these fractions vary by age?

What Fraction "Try Out" Self-Employment?



Among Workers Trying Out Self-Employment, What Fraction Move to Full Self-Employment?



What Next?

Work in progress but key steps:

1. Identify who is self-employed each year
2. Impute quarterly self-employment pattern
3. For these, impute quarterly self-employment earnings when not available
4. Integrate information into existing data on employment histories

Final Note:

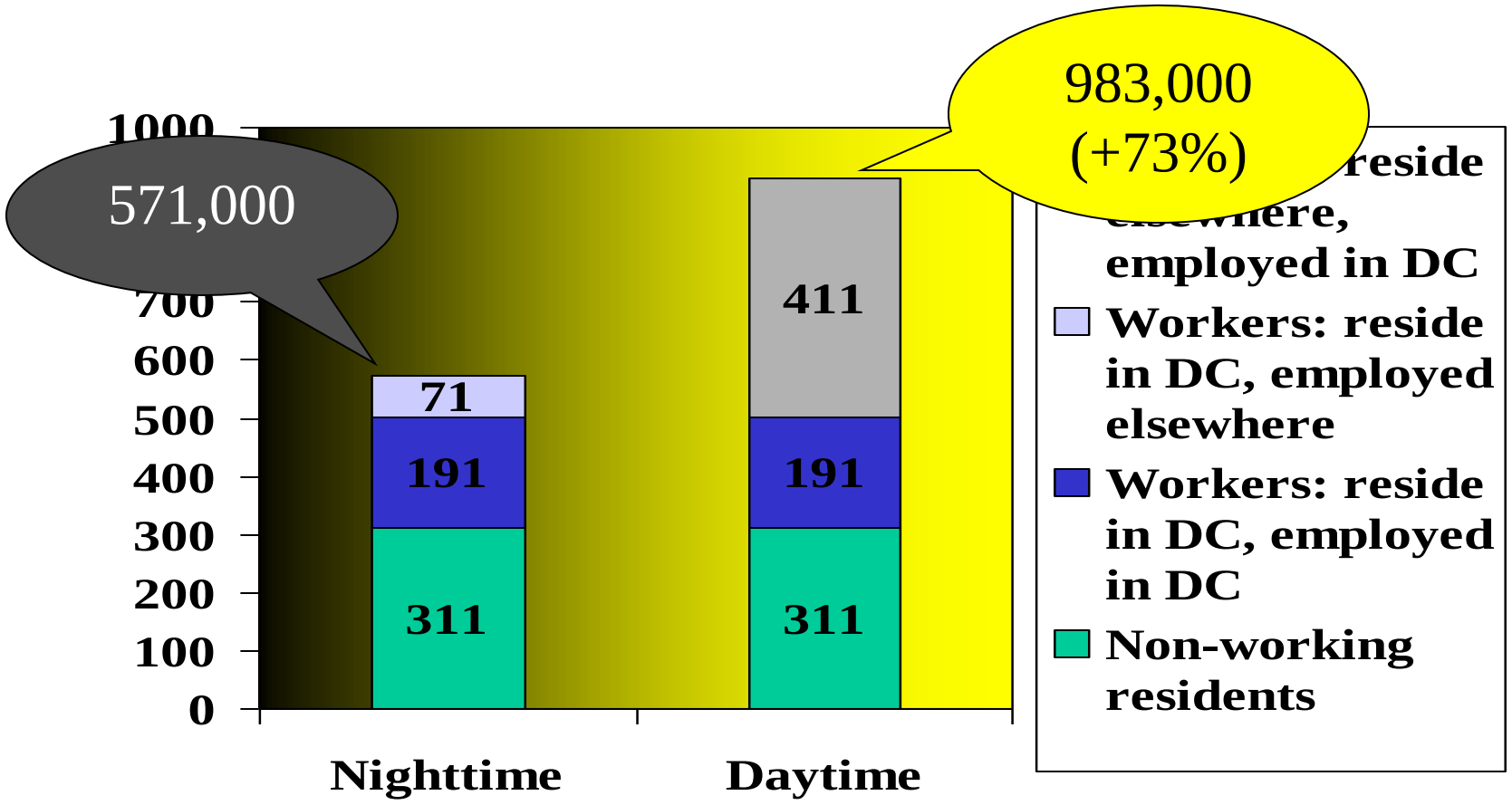
Adding Self-employed to QWIs may provide clearer picture of workforce dynamics, especially for certain groups.

The Construction of Daytime Employment Estimates Using LEHD Data

Fredrik Andersson

February 2, 2007

The Composition of the Night and Daytime Population in D.C.



The Variation Matters!

The expansion and contraction of the population during the 24 hours of a day has important implications for many planning purposes:

- Transportation
- Community
- Economic
- Disaster and relief operations

Purpose of Presentation

"To demonstrate how data from the LEHD program can be integrated with survey data to provide estimates of high geographic resolution of the daytime working population"

The methodology addresses two issues:

- How to identify daytime employment in LEHD Data
- How to deal with non-covered sectors in LEHD data

Some preliminary empirical results

The Census Bureau's Daytime Population Project

- Census Bureau produced estimates of the 2000 daytime population at the county and place level
  - Daytime population = Residents + incommuters – out commuters
- The Daytime Population Project's aim is to enhance existing estimates in several ways:
  - Broadening the scope of the daytime population
  - Creating estimates at the tract level
  - Produce estimates in inter-decennial years
  - Improve estimates of the daytime working population
- Initial demonstration for four states: Illinois, Missouri, North Carolina and South Carolina

The Daytime Working Population

Workers constitute a key component of the daytime population:

- in terms of its sheer size
- in terms of its intra-daily geographic mobility

No single ideal data

Decennial too infrequent

ACS do not provide estimates by detailed geography

Administrative LEHD data do not identify *when* workers are employed

Definition: *Worked last week, arriving at work between hours of 5 am to noon, who are 16+ years old and not enrolled in school or college level*

Methodology Overview

- Integrate 2000 Decennial Long Form and LEHD data
- Use statistical relationships between daytime employment status and observable LEHD characteristics in 2000 to impute daytime employment status for the broader LEHD population in different periods

  Adjust daytime employment estimates for non-covered sectors using survey data benchmarks

Combining LEHD data with Decennial Information

For approximately one in six of all workers with positive earnings in LEHD 2000Q1 we can obtain additional information from the Decennial

- About 80-85% are "Decennial Employed"
- About 60-65% are "Decennial Daytime Employed"

Substantial variation in Daytime Employment

	Illinois	South Carolina
Age		
- 16-20	10.14	12.84
- 21-25	45.26	45.83
- 26-35	64.93	65.80
- 36-55	72.67	73.33
- 56-high	68.43	69.79
Earnings		
- below 33rd percentile	40.81	41.58
- between 33rd-66th percentile	68.68	69.13
- above 66th percentile	77.94	78.70
Employment Status		
- Full-Quarter Employed	67.46	69.43
- Only Beginning-of-Q1 Employed	52.02	53.22
- Only End-of-Q1 Employed	39.65	41.35
- Only Q1 Employed	31.16	34.52
Selected NAICS Sectors		
-Utilities	77.16	82.10
-Finance and Insurance	77.02	80.71
-Arts, Entertainment, and Recreation	48.34	51.84
-Accommodation and Food Services	35.28	36.39

→ Calls for a more sophisticated imputation methodology

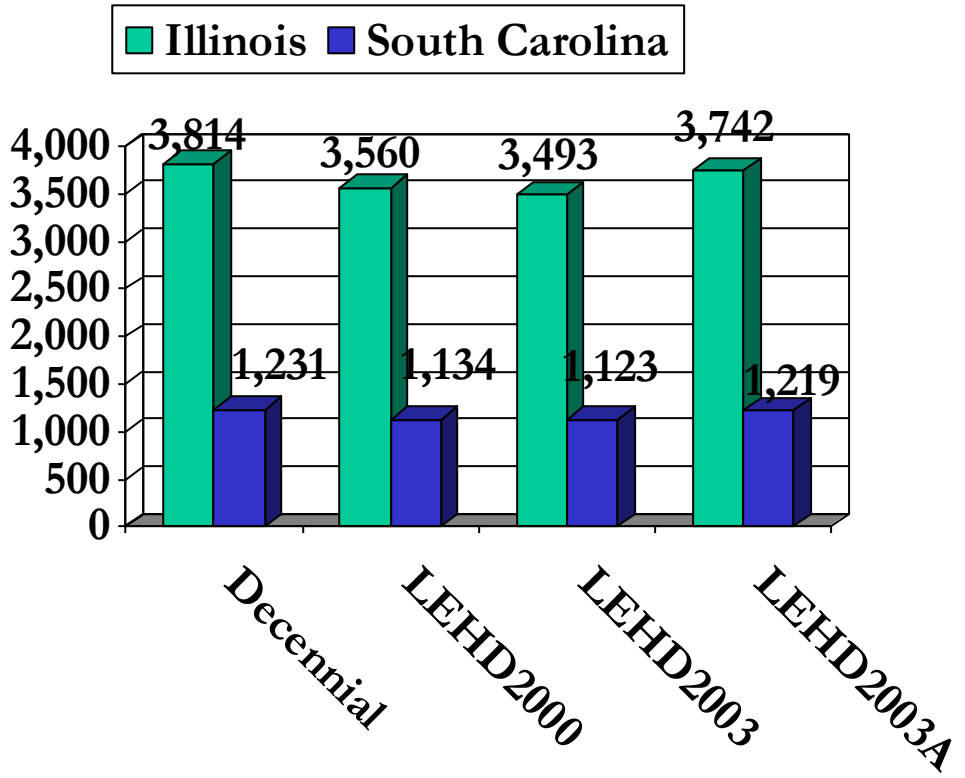
Imputation model

Statistical model to impute daytime employment status at the person level conditional on:

- Age categories
- Earnings categories
- Gender
- County of employment
- Employment status
- Industry

Possible to aggregate daytime population estimates to any geographical level

Preliminary Daytime Worker Estimates (in thousands)



Next Steps

- From prototype to production
  - Phase 1: Developing proof of concept
  - Phase 2: Development of prototype estimates
  - Phase 3: Production?
- Methodology development
  - Flow approach
  - Extending the daytime population definition
- Data dissemination tools
  - Integration of daytime worker estimates in the QWIs/OnTheMap?

Other Current Data Innovations and Research at LEHD

- Integration of federal workers in the LEHD data infrastructure
- Employer-to-Employer Flows
- Development of national frame for QWI production

Adding Federal Workers

LEHD will soon receive employment and wage data on federal workers from the U.S. Office of Personnel Management (OPM).

- This is a transactional file of personnel actions (hires, promotions, transfers, terminations, etc.) for federal workers.

Integrating OPM file into QWIs.

- Develop a data infrastructure from OPM data that mimics that of the underlying UI wage records and ES202 firm data.
- Geography information on file is limited.

Employer-to-Employer (EE) Flows

QWIs currently provide separations and hires.

Clearly many of these flows are movements between jobs.

Would like to characterize job paths of workers from job to job.

What we find:

- Almost 4% of workers experience an EE flow every quarter.
- About 30% of separations are direct flows into another job.
- EE flows are highly procyclical.
- EE flows are associated with strong earnings gains, particularly for younger workers but for mid-age workers as well.

Next steps:

Research possible integration of EE flows into QWI.

Redesign of Production System to National Frame

Current production system is based on state-based data universe.

Moving to a national frame poses a number of R&D challenges.

- Tracking of workers across states - OnTheMap.
- Tracking of workers across UI and non-UI employment universes.
- Increased computational burden – more workers, more establishments.

Conclusions and Next steps:

Addition of new partner states and new data sources will bring us much closer to our goal of constructing a national longitudinal database on the US workforce.

Next steps toward that goal:

- Integration of federal and self-employed workers into production.
- Adapting production system to national data frame.