Using Administrative Data for Program Evaluation and Research in Education-Labor Market Nexus

Thomas Bailey
Community College Research Center
Teachers College, Columbia University

USDA-ERS Conference: November 1-2, 2016
Outline

• Why is administrative data needed for education-employment nexus?
• What is CAPSEE?
• Results from CAPSEE research with big data
• Evaluating the usefulness of big data
WHY IS ADMINISTRATIVE DATA NEEDED FOR EDUCATION-EMPLOYMENT NEXUS?
Student Progress is Complex

- Pre-college work
- High school
- College
- In-college work
- Post-college work
- More college
State Administrative College Data

• Begin with college data: link across, forward, back

• These data are different from longitudinal surveys:
  • Created for basic administration and compliance purposes
  • Variation by state in quality, comprehensiveness, history

• Coverage issues are very important:
  • Often limited to public sector within one state
  • University systems typically hold own data; community college districts or systems typically hold own data
  • Centralized states (collect and hold data across all publics) and decentralized states (data available college-by-college basis)
Datasets Linkable to College Data

• National Student Clearinghouse data on where students transfer to, how long they persist, award earned
  o Merge on name and birthday
  o High match rate: NSC coverage is very full (includes all Title IV colleges)

• State high school data with full transcript information
  o Merge on name/birth/birthday/ID
  o Low match rate: student mobility and lagged/delayed college enrollment and enrollment out of publics or out of state

• College-level data from IPEDS or other sources; census data
  o Merge on geocode or college name
Linkable Labor Market Data

• BLS has local labor market conditions (also use ACS)
  o Merge on geocode or area
  o High match rate: but may confound with college fixed effects

• Unemployment Insurance data for individual student earnings
  o Merge college and UI data using SSN
  o Moderate match rate: coverage of employment data may not be complete

• Labor market data may differ from national surveys
  o Quarterly employment and income data from all formal jobs
  o Sometimes hours worked and occupation
Advantages with Linked Data (1)
Better Map Complex Pathways

• Longitudinal
• Population analysis (reducing bias from attrition)
• Large sample sizes allow for subgroup analysis
  – Colleges, programs, courses
  – Demographic groups
• Address a lot of questions for education policy
• More precise, accurate, and various measures of educational attainment – student pathways
Student Pathways: Intensity

Many pathways over 6 years in community college

These are the most popular:

- Part-time for 1 semester: 19%
- Full-time for 1 semester: 9%
- PT for 2 semesters: 5%
- FT for 2 semesters: 4%
- FT then PT in 2 semesters: 2%
- FT-FT-summer break-FT-FT: 1%
- FT-FT-summer break-FT: 1%
Student Pathways: Transfers

According to NSC:

• One-third of all students transfer
• 14% of students who start at 4-year college transfer to 2-year college
• >50% of these transfers never return to 4-year college
• Co-enrollment
Student Pathways: Course-taking

Students take many different courses:

- Remedial classes
- General education core
- Surplus courses
- Subject-specific courses
- Courses required for awards
Advantages (2)
Better Specifications

• Many pre-college controls
  – Time-varying controls
  – Ability measures
  – Proxies for non-cognitive attributes (e.g. credits accumulated in school for effort)
• Help reduce and test for omitted variable bias
• Some opportunities to test for selection bias (variations in college practices, changes over time, compare students to themselves in other classes)
Advantages (3)
Better Construct Validity for Earnings

• More precise and accurate measures of earnings/income:
  o Self-reports less reliable at lower earnings (overstate low income): compress the education-earnings premium
  o Self-reports more measurement error for the less educated (low education persons misstating their income): reduce precision
  o More educated persons have multiple jobs (bonuses/commissions)
  o No non-response missing data (CPS is 20-30%)

• Data on income over time, including before and during college, and quarterly (not annual)
① What are the employment/earnings benefits of CC?
② What institutional programs and public policies improve student outcomes?

Big data:
- Transcripts: state-wide CC systems for FTIC cohorts
- Transfers: National Student Clearinghouse
- Earnings: UI wage records pre-, in-, post-college
- AR, CA, MI, NC, NY, OH, VA, WA in 2000s
RESULTS FROM OUR WORK
Results from Big Data

- We have some important results for returns to AA degrees, certificates, and credits
Quarterly Earnings Gains:
AA over No Award 8 Years Post-FTIC (Fixed Effects)

<table>
<thead>
<tr>
<th>State</th>
<th>Male Mean</th>
<th>Female Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>WA</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>VA</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>NC</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>MI</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>KY</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>AR</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>WA</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>VA</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>NC</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>MI</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
<tr>
<td>KY</td>
<td>$1,040</td>
<td>$1,800</td>
</tr>
</tbody>
</table>

CENTER FOR ANALYSIS OF POSTSECONDARY EDUCATION AND EMPLOYMENT
Quarterly Earnings Gains: AA over No Award 8 Years Post-FTIC (Fixed Effects)

- **Male**: $1,040
- **Female**: $1,800

PV gain for AA (30 years d=3%)
- Male $75,000
- Female $135,000
Gains are: modest, temporary, non-robust (except in health).
But certificates vary substantially in credits.
Quarterly Earnings Gain over Zero Credits for Non-Completers
(8 Years After FTIC in VA/NC)
Quarterly Earnings by Gender
Degree Completers (AR, KY, MI, OH)

College entry 18 Quarters later
EVALUATING BIG DATA
Evaluating Big Data

• Useful, but less useful than we might hope
• Why?
  ① Time/resources for analysis
  ② Obvious or outdated conclusions
  ③ Barriers to change
1. Time/Resources for Analysis

A. Lots of data curating
B. Lots of questions
C. Identification problems
1A. Data Curating is a Lot of Work

Requirements:
• Individual-level data (not college-level or program-level)
• Long “windows” per student

Tasks:
• Collating data longitudinally and across systems
• Cleaning data for missings (transfers/earnings)
• Coding data from flat files over courses/colleges
Practicalities of Using Linked Data

• Need links with state system officials and UI data-holders
  – Personal relationships to persuade data-owners that research is useful
  – Work with many agencies; some have good mutual relationships
• This is not a priority for state officers; may take time
• Cannot ask repeatedly for more information
  – Need to know exactly how much data you need
  – Data-owners typically do not mind if ask for more years if data is all in same format
1C. Identification Problems

- Hard to causally identify impacts because no random assignment in administrative datasets
- SES missing (use occupation, geocode, financial aid)
- Attitudinal data usually not available
- UI data does not cover everyone and sample truncation or censoring may be endogenous
  - Students who move across state lines, self-employed, military, some federal workers
  - Cannot be sure that missing earnings is zero
2. Conclusions from Big Data

A. Obvious
B. Old news
C. Context-specific
2A. Confirm Known Conclusions

Get students to:

• accumulate more credits
• complete programs
• transfer onward
2B. Old News

Now

Students need ~4 years in labor market 2016
Students need ~3 years to complete AA 2012
Students need ~6 years to complete BA 2009

Ten years out of date
Labor market change over a decade…
Actionable Narrow Questions

There are still plenty of questions left for research. What is the labor market gain from:

- Summer session courses?
- First-semester course-loads (momentum)?
- Transfer with an Associate degree or as fast as possible?
- Higher instructor quality?
- Smaller class size?
Visit us on the web at capseecenter.org
We’re also on Facebook and Twitter.

Center for Analysis of Postsecondary Education and Employment
Teachers College, Columbia University
525 West 120th Street, Box 174, New York, NY 10027
capsee@columbia.edu
212.678.3091

CAPSEE is funded through a grant (R305C110011) from the Institute of Education Sciences, U.S. Department of Education.