Current and Potential Uses Of Big Data at the Bureau of Labor Statistics

Remarks by Steven J. Davis
faculty.chicagobooth.edu/steven.davis/

BLS Technical Advisory Committee
Washington, DC
20 June 2014
Potential Gains from Linking

Some Easy(?) Stuff

1. Produce statistics on job flows (BED), employment growth (CES), hires and separations (JOLTS), job openings (JOLTS), wage growth (NCS), etc. that distinguish firms with publicly traded equity from privately held firms. Both types of firms are hugely important, but their economic behavior differs sharply in many respects. See, e.g., Davis et al. (2007).

2. Among publicly traded firms, produce statistics sorted by market cap, beta, P-E ratio, book-to-market ratio, multinational vs. domestic-only firms, etc. There would be **HUGE** interests in these stats, much of it from finance economists.
Potential Gains from Linking

Some Easy(?) Stuff – continued

Method:

- Identify publicly traded firms using COMPUSTAT and other sources.
- Link publicly traded firms to BLS sample frames by matching on business name, location, industry and activity levels. Supplement with manual linking.
- Incorporate information on publicly traded/privately held status into the QCEW sample frame for the BED, CES, JOLTS, NCS, etc.
- Produce tabs – just a different sort on establishment-level data that BLS already has and routinely uses.
Potential Gains from Linking

So how easy is it?

• Davis et al. (2007) did the linking to a Census-side universe database (Census Business Register, LBD) using COMPUSTAT for information on publicly traded status.

• They achieved a 92% match rate (for both number of firms and as a percent of employment).

• Institutional resources could push the match rate higher, perhaps to near 100%.
Potential Gains from Linking

So how easy is it?

• Can BLS access data items in Census Business Register that say which firms own which establishments? If so, BLS can borrow and build directly on Davis et al. (2007) and a BR-QCEW cross-walk.

• If BLS cannot access these items in the Census Business Register, it would need to integrate data on which firms are publicly traded directly into the QCEW. Feasibility and difficulty depends on accuracy and completeness of BLS linkages of establishments to parent firms.
Potential Gains from Linking

Question: How do key outcomes vary with local economic conditions? Some examples:

• Price inflation (CPI) and wage growth (NCS)
• Job vacancy rates (JOLTS)
• Job-finding rates (CPS)

Method:
1. Use QCEW data on employment to sort local areas into growth rate quantiles. (Or quantiles of growth rates deviated about own past local area mean.)
2. Calculate average inflation, wage growth, vacancy, job-finding, and employment growth rates by local area quantiles.
Potential Gains from Linking

Feasibility

• These stats are also in the “easy” category. Even easier – all data elements are internal to BLS.

• Straightforward to develop mechanical rules for setting quantile boundaries and sorting units into quantile groups.

• This approach to local area stats differs modestly from the nature of (most) BLS tabs in that:
  • Quantile boundaries vary over time
  • Local areas migrate across quantiles over time
  • Perhaps somewhat less timely
Potential Gains from Linking

Why Produce These Local Area Stats?

• Intense interest in better understanding of:
  – How economic growth relates to inflationary pressures for wages and prices. See, for example, Hilsenrath and McGrane (2014).
  – The relationship between labor market tightness/slack and wage pressures is a fundamental one for monetary policy, macroeconomics and labor economics.
  – Characteristics of local economic growth is of huge interest to regional and urban economists, policymakers, etc.
Potential Gains from Linking

Why Produce These Local Area Stats?

• Nobody else can produce these types of local area stats on a recurring basis, because they require continual access to confidential BLS micro data.

• Census could produce similar stats – but mostly using annual outcome measures. That would be much less interesting for understanding cyclical phenomena and for informing policy makers in a timely manner about cyclical developments.
On Potential Gains from Linking

Reforming CIPSEA to let BLS access Census datasets that draw on IRS data would dramatically increase the scope for fruitful linking.

Example: New BED Statistics on Job Flows by:

- Capital intensity
- Investment rates
- Energy intensity (fuels, electricity)
- Technology use, technology adoption
- Degree of product specialization

See Davis, Haltiwanger and Schuh (1996), Davis and Haltiwanger (2001), and Davis et al. (2007) for examples using Census sources with annual data or limited sectoral coverage.
On Web Scraping

1. It makes sense to apply web-scraping methods to complement and augment existing surveys and administrative record sources.
   1. Build on existing BLS expertise and infrastructure
   2. Exploit access to confidential data
   3. Third-party willingness to cooperate with BLS

2. Not a good use of BLS resources to duplicate or create statistical products, no matter how cool, that can be done just as well outside BLS.

3. Not a panacea – must evaluate case by case
On Web Scraping

Example of a Complements Approach:

• JOLTS draws on a stratified random sample of establishments to produce stats on worker flows, job openings, and employment – a very limited set of outcomes

• “A Million Job Openings” project that web scrapes to attach characteristics of job openings (occupation, qualifications, offer wage, etc.) to JOLTS micro data.

• Much need for reweighting. Some loss of representativeness for scraped outcomes.
References


