Comments on "Improving Business Lists Through Data Sharing"

By

John Haltiwanger, University of Maryland

Overview

- One of the challenges for building key national indicators (including output, productivity, and inflation for producers) is that the source data are collected by different agencies with different core business lists.
- CIPSEA permits data sharing not involving FTI.
- FESAC on record as strongly endorsing enabling legislation.
- But good news is that agencies are working on what they can share under existing laws.
- Today's presentations are mostly about the latter.
- Main Comment: Congratulate the Statistical Agencies for addressing these challenges in creative ways.
- Most of my comments are about issues raised in BLS and Census presentations about challenges of business list differences.
    - BEA presentation does a nice job of highlighting importance of the issue and challenges BEA faces in reconciliation. BEA (and all of us) will benefit in reducing discrepancies.

Establishments vs. "Firms" vs. Enterprises

- Most enterprises are single unit establishments
- But most economic activity is at enterprises
- EIN defines an entity that is not associated with an economicly meaningful concept.
  - Establishments (physical location for activity) and enterprises (operational control) are economicly meaningful concepts.
  - MU firms can have as many EINs as useful.
  - Likelihood of using multiple EINs varies by sector – more likely for sectors dominated by firms with operations in multiple states.
- Statistics should focus on establishment vs. enterprises if possible
  - Useful of course to know how much the distinction between EIN firms and enterprises matters

Establishments vs. "Firms" vs. Enterprises

- Job flow patterns show what we would expect:
  - Much higher rate at establishment than firm/enterprise (reminds us that there is much within firm reallocation).
  - Many establishment openings and closings are not new or exiting firms/enterprises.
  - Useful to show rates as well as levels – decline in pace of reallocation more apparent in rates.
- Smaller gaps for firms vs. enterprises.
  - This is telling us not much reallocation across EINs within the same firm.
- Some caution since based only on MU EINs that match to BLS
- Interesting but does not imply there are not more substantial differences likely on other dimensions:
  - Size distribution (BDS has 50 percent of employment in enterprises>500, BED has 44 percent of employment in "firms">500)
  - Size distribution and age distribution in specific sectors like Retail Trade
    - EIN based firms unlikely to capture fully shift to large, national chains.
    - Will also likely influence gross flows (expansion across state lines within firms).

Challenges of Business List Integration

- Very different source information (UI tax system vs. payroll taxes).
- Reporting units are different
  - UI reporting units are activity of business within state. Establishments broken out for MU by MWR.
  - EIN is administrative data unit for payroll taxes. Census breaks out MU with variety of sources (Economic Censuses and COS).
    - Large role for Economic Censuses yields spikes in splitters in micro data in Census years. One of the areas that could be improved.
- Matching of MU files only is its own challenge:
  - Difference concept of MU (not sure what was done about this in project):
    - BLS: multiple units with 10 or more employees per site in given reporting unit.
    - Census: multiple establishments under same national enterprise.
    - Should inherently be some MU for Census that are SU for BLS (operate in only one state) if use original definition. What did this project do?
  - Useful to review the results from the BLS and Census comparison project as background.

Table 2 from Elvery, Foster, Krizan and Talan (JSM, 2006)

Table 2: Summary Statistics for Single-unit versus Multi-unit Classifications

Restricted to the joint sample

	EINs	BLS Data		BOC Data	
		Employment	Establishments	Employment	Establishments
MM	110,166	48,263,307	1,417,721	48,177,843	1,174,587
SS	3,403,011	29,375,863	3,403,011	31,077,750	3,403,619
MS	174,503	9,521,799	843,101	10,409,060	176,463
SM	106,770	6,267,855	106,770	6,900,720	204,804
M_	22,822	3,658,970	167,172		
S_	247,373	2,287,081	247,373		
_M	22,212			2,576,229	60,340
_S	439,837			5,780,457	439,904
Total	4,526,694	99,374,875	6,185,148	104,922,059	5,459,717

Figure 1 from Elvery, Foster, Krizan and Talan (JSM, 2006)

Figure 1: Sector Match Rates for SS EINs (employment weighted)
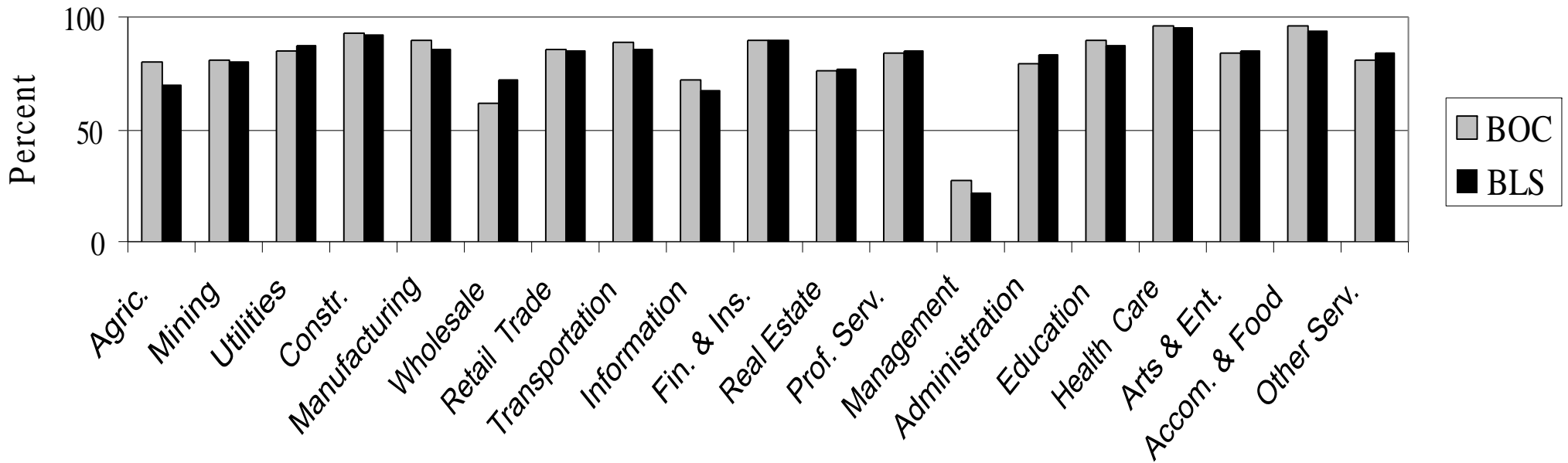
Even at broad sector level, substantial discrepancies in NAICS codes for single units. Discrepancies much larger at detailed NAICS code level (about 70 percent match rate).

Comments on Challenges

- Large number of SU (BLS) – MU (Census) and MU (BLS) – SU (Census) cases
  - This helps to explain some aspects of discrepancies reported in Census presentation. Shows challenges of MU sharing only.
- Census has fewer establishments but more employment than BLS.
  - Greater employment especially in SU, SU-MU and MU-SU cases.

- NAICS discrepancies are large even for Single Units (so it is not only or even primarily a discrepancy in MUs that is at issue) .
  - Industry discrepancies are a core aspect of the problem. Reconciliation should be a high priority.

Limitations of EIN as an ID

- EINs are limited as business identifiers
  - Represent a taxpayer ID. Not an economic concept of either enterprise or establishment.
  - Change with ownership/LFO/other reorganization activity
  - Are used in very different ways:
    - Payroll EINs
    - EINs for filing consolidated income taxes
    - EINs are not primary reporting unit or ID for UI system.
  - Multiple uses causes difficulties in matching across BLS and Census and even within Census
    - Matching of revenue and payroll data .
    - Relevant for employer vs. non-employer statistics and PEOs.

Can we move to a common system of establishment and firm identifiers?

- A common system of firm *and* establishment identifiers would be an enormous help across and within agencies.
  - Both BLS and Census have invested substantially in developing longitudinally consistent establishment identifiers (but separately).
  - These identifiers have enabled the rich new longitudinally based products.
- Can Big Data help?
  - Establishments are physical locations where economic activity is occurring. Many different possible sources of such information from Big Data.
- If enabling legislation is passed, moving towards common identifiers should have high priority.   But even before we should think about moving towards common identifiers.

Business List Reconciliation: Thinking Big (Data and IT)

- Priorities should be:
    - Large discrepancies in the share of activity by industry and state.
    - This impacts the key national indicators and industry/state level analysis

- One approach:
    - Take data as given. Creatively try to reconcile. This is what is being done now and I applaud and encourage such efforts.

- Alternatively:
    - Can we change how the data are generated/harvested?
    - **Information/Big Data age:**
        - Can information obtained from businesses at key points (application for EIN, application for UI account) be sufficiently synchronized that respondents will provide the same information to all? "One stop" IT based shop?
        - Can UI and payroll tax data be harvested from same business data systems?
        - Can similar harvesting yield core information for Economic Censuses?
        - Can such a synchronized system be used to enable common identifiers from the start as well as consistent information about business activity (payroll, employment, location, industry)?

What more can be done?

- It would be helpful to create a comprehensive list all of the data types and information that could be currently shared under CIPSEA.

- Interagency teams (and advisory committees like FESAC) could brainstorm about how the information could be used for addressing core measurement challenges.

- Such brainstorming will also help make the case for enabling legislation and help establish priorities for **WHEN** that happens…