Improving the Business Register Through Data-Sharing:
Uses and Challenges

Brandy L. Yarbrough

United States™
Census
Bureau
Economic Statistics

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Overview

1. How has the Census Bureau used the BLS multi-unit (MU) data to improve its business register (BR)?
2. What are some of the challenges and limitations in using these data?

Part 1– Uses

- Clients of Professional Employer Organizations (PEOs):
  - New employer units -> improved coverage of output (receipts)
  - Verifying employer vs. non-employer status
- Improved distribution of payroll tax data for selected companies
- Better targeting of "splitters" in the Company Organization Survey (COS)
- Improved NAICS codes for sub-units of consolidated reporters
- Ad hoc research

Part 2– Challenges

- Data file completeness: missing states, different filing rules by state
- Timing
- EIN matching issues
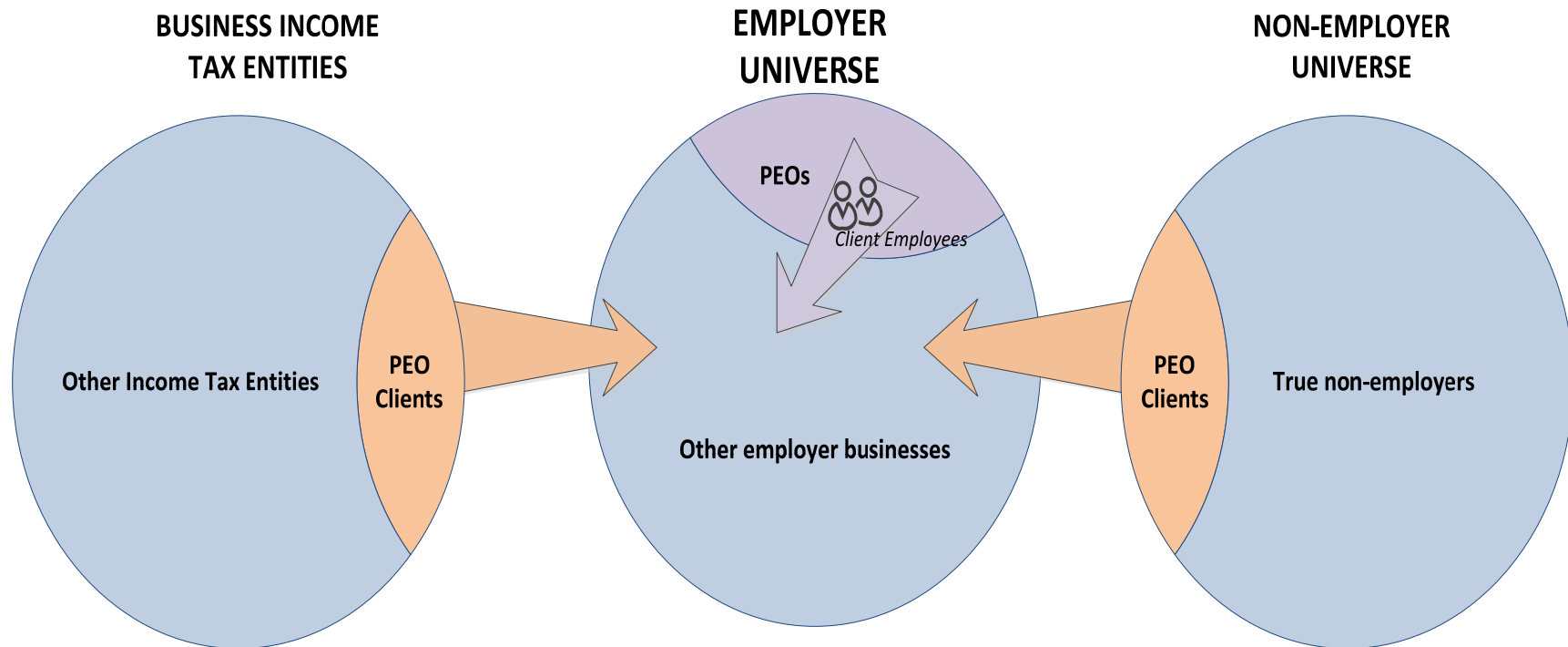
Part 1-- Uses

Adding PEO Clients to the BR

The Situation

- The foundation of the BR employer universe is IRS Form 941 payroll tax data
- The PEO is the "employer of record" for their clients
- Payroll tax data for clients is filed under the EIN of the PEO
- Consequently, PEO clients do not get counted in the employer universe
- PEO client output (receipts) is either not accounted for or it is wrongly included in the non-employer data

Adding PEO Clients to the BR

Goals:

- Add PEO clients to the employer universe.
- Remove them from the non-employer universe.
- Distribute employment to client industries and locations.

BUSINESS INCOME TAX ENTITIES

Other Income Tax Entities

PEO Clients

EMPLOYER UNIVERSE

PEOs

Client Employees

Other employer businesses

NON-EMPLOYER UNIVERSE

PEO Clients

True non-employers

United States™ Census Bureau
Economic Statistics

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Adding PEO Clients to the BR

Using the BLS PEO data file

- Evaluated clients with >= 2 million in annual payroll: ~5,000 cases
- ~2,600 new establishments added from 2011 and 2012 files
- Automatic exclusion from the non-employer universe
- Account for ~130,000 employees and ~$9 billion in annual payroll
- Some mailed in the 2012 census in order to capture output
- ~$40 billion in output (including imputes for non-response)
- Wholesale sector (NAICS 42) most significant in terms of output
- Custom computer services (NAICS 541511) largest in employment

Unresolved issue:
"Double-counting" of payroll and employment in the PEO client industries and in NAICS 561330

Distributing Payroll Tax Data

The Situation

- On the BR, non-response or non-mailed MU companies need to have payroll and employment data for their establishments imputed
- This is done to provide updated measures of size for sampling operations and also to provide complete source data for the County Business Patterns program
- This imputation is done by allocating current year tax data in proportion with prior year establishment values
- If a company has not responded or has not been mailed for several years, the allocation may not be accurate

Distributing Payroll Tax Data

Using the BLS MU data file

- The BLS data are a year behind the BR cycle when Census received the files…
- …but tax data allocation are based on prior year values anyway
- The BLS payroll and employment data are assumed to be "equivalent quality" to Census survey reported values
- For selected companies, BLS data could be substituted for prior year imputes
- Desired outcome: a more accurate spread of the tax data when BR imputation occurs.

Distributing Payroll Tax Data

Determining the "selected companies"
- No response or not mailed in 2013 COS
- Totally imputed in 2012
- Simple match to BLS MU file on EIN x address
- All locations in a company must match
- Payroll tax data within 5% of BLS data at the EIN level

Results
- Not productive– only ~250 locations met criteria
- Not done for 2013 survey year
- 2014 may provide a better opportunity:
  - Larger universe of candidates
  - More refined matching

Targeting "Splitters"

The Situation

- Splitters are establishments that are on the BR as single units (SU)
- Exhibit characteristics of multi-establishment companies
- Mainly identified through tax and other administrative data
- Small number are sampled and mailed in the COS to capture location information
- BLS Multi Establishment Employer Indicator Code (MEEI) already used in sampling– i.e., MEEI=3 (multi-unit employer)
- Budget and resources limit mailing to ~5,000 cases
- Success rate is often less than 40%

Targeting "Splitters"

Using the BLS MU data file

- Many EINs on the BLS MU data file are on the BR as SU and have MEEI=1 (single establishment unit)
- This disqualifies them from inclusion in the splitter sampling frame under current rules
- For the 2014 COS:
    - A sample of 100 cases will be mailed
    - Success rate will be compared to other cases
- May result in modifications to splitter sampling procedures
- Evidence to the contrary? (Discussed more in Part 2)

NAICS Codes for Selected Establishments

The Situation

- In the Economic Census (EC) most MU companies are mailed detailed questionnaires for each of their establishments
- Responses to EC questionnaires are used to verify/update NAICS codes through complex algorithms
- For some industries, establishment-based output measures are not practical so an alternative reporting unit (ARU) is offered instead
- The ARU consolidates multiple establishments operating in the same industry or related group of industries. Output is reported only at the ARU level
- Only payroll and employment are reported at the establishment level
- NAICS codes for establishments that are part of an ARU are only updated by explicit action on the part of the respondent– i.e., writing in a new business description

NAICS Codes for Selected Establishments

Interagency Meeting
- On March 20, 2014 representatives from the Census Bureau and BLS met to conduct a comprehensive review of 40 example cases where the two agencies had assigned different NAICS codes
- Several of the cases were ARU establishments or cases out-of-scope to the Economic Census
- ~200,000 such establishments on the BR

Using the BLS MU data file
- Provides a source for verifying/updating NAICS codes for ARU and out-of-scope MU establishments
- Review and update largest establishments on the BR
- Start after the final 2012 EC NAICS codes have been fed back to the BR

Ad Hoc Research

- The BLS MU data files have been used extensively as a resource for resolving issues with individual companies
- Used to augment or supplement data reported in the EC or the COS
- Analysts use information provided to manually update the BR via standard interactive review and correction tools

Examples:
- Improved location lists for historically difficult EC respondents
- Explaining variances between payroll tax data and response data
- Verifying addresses and NAICS codes for newly reported locations (births, acquisitions)

Part 2 -- Challenges

Completeness of the files

- BLS files that Census receives are incomplete
- Some states will not allow data sharing with Census
- In some states, the BR has significantly more establishments than the BLS file

States not provided

State	2012 County Business Patterns Data	
	Establishments	Employees
MA	171,278	3,035,897
NH	37,213	548,985
NY	527,001	7,556,521
WY	20,635	214,241
Total	756,127	11,355,644

State estab. counts: BR > BLS

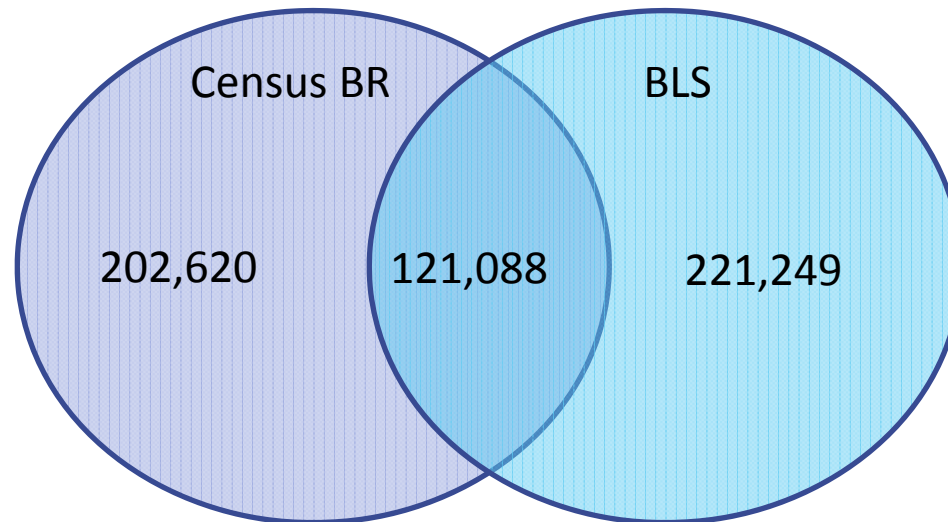| State *	Estabs: BR > BLS
AZ	4,000
DE	2,200
IL	10,000
MI	10,000
TN	3,200

Timing

- BLS files are currently a year behind the COS cycle– e.g., 2013 files will arrive during the 2014 survey reference year

- Good for "retro-active" research and analyst decision making

- Less good for full-scale operational use in the current survey year

- More current delivery in the future? Quarterly?

EIN matching: Census vs. BLS

- Comparison of MU EINs– 2012 BR vs. 2012 BLS MU file
- In total, Census has only about 6% fewer EINs than BLS
- But, there is not much overlap between the two sets of EINs

EINs linked to MU companies in 2012

Census BR   BLS

202,620   121,088   221,249

EIN matching: What is a "multi-unit"?

Prior to the 2012 EC, nearly ~1.4 million (44%) of the locations on the 2011 BLS MU file had EINs that were on the BR as SU

Exercise:
Match EINs in 2012 BLS MU file to the BR <u>after</u> the 2012 EC frame updates were completed

Why?:
- Over 2 million SU cases were mailed in the EC
- Most had the opportunity to report operations at multiple locations (i.e., splitters)
- Would significantly more BLS EINs be linked to MU companies on the BR after the EC splitter results had been applied?
- High agreement suggests that the BLS file might be a good non-survey alternative source for creating splitters on the BR

Matching: What is a "multi-unit"?

<u>Outcome:</u>

- After the EC, the EINs of ~1.2 million (36%) of the locations on the 2012 BLS file were still linked to SUs on the BR
- Disposition of these SUs in the EC:

Description	Count
EINs with multiple locations on the BLS file but linked to an SU on the BR	178,499
Mailed as SU in the 2012 EC	134,238
Responded in the 2012 EC	105,381
Reported operating in multiple locations *	4,787

    * None of these were set up on the BR as MU due to being below size thresholds

- Less than 3% even reported multiple locations in the EC
- Suggests inherent differences in what constitutes a "multi-unit"
- Automated creation of MUs on the BR? Unlikely
- Is "Targeting Splitters" worthwhile?

EIN matching: Where are they?

- More than 200,000 (~63%) of the MU EINs that are on the BR are not on the BLS file
- Only about 10% of these are explained by the states that are missing (by design) from the BLS files (MA, NH, NY, and WY)
- These "missing" EINS account for nearly 450,000 establishments and 14 million employees across over 100,000 distinct companies
- Over 90% of these companies are below the certainty cutoff for the COS...
- ...theoretically making them ideal candidates for updating with the BLS data...
- ...but without the EIN connection there are concerns
- Are these "missing" EINs cases that BLS has as SU?
- Are the 450,000 establishments in the BLS file under a different EIN? If so, why?

Summary

- BLS MU data very valuable as a research and validation tool for BR analysts

- Useful for making incremental improvements to the BR, COS, and non-employer program

- Not ready to operationalize

- More interactive tools designed around the data

- More time, resources, and evaluation needed for other applications

- Periodic technical meetings

Discussion Questions

- What are the potential causes of the low degree of agreement between the BR and the BLS MU file on an EIN basis?

- How should the Census Bureau deal with PEOs and their clients in it's survey programs and on the BR?

- What impact is the "global LEI" expected to have in facilitating data-sharing and in statistical systems in general?

United States™
Census
Bureau

Economic Statistics

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU