A Comparison of Person-Reported Industry to Employer-Reported Industry in Survey and Administrative Data

Emily Isenberg
Center for Economic Studies, U.S. Census Bureau
Liana Christin Landivar
Industry and Occupation Statistics Branch, U.S. Census Bureau
Esther Mezey
Axioma, Inc.

Presented at the Federal Committee on
Statistical Methodology Research Conference
November 6, 2013

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Industry Statistics

Industry statistics describe the kind of business conducted by a business establishment.

- All federal agencies use the North American Industry Classification System (NAICS) to classify business establishments.
- The NAICS uses a six-digit hierarchical coding system to classify all economic activities into 20 major sectors. These sectors can be further disaggregated into 1175 industries (NAICS 2007).
- The Census Bureau's Economic Directorate uses the NAICS to provide industry statistics.
- The Census Bureau's Demographic Directorate bases its industry classification on the NAICS, but aggregates industry statistics into 13 major sectors and 269 detailed industry categories.

Sources of Industry Statistics

- Person-reported: Data directly provided by household survey respondents (e.g., American Community Survey, Current Population Survey)
- Employer-reported: Data directly provided by employers (e.g., Economic Census)
- Administrative records: Data obtained from local, state, or federal government agencies (e.g., Internal Revenue Service, Unemployment Insurance)

Industry in Public-Use Data Tools

Person-reported data

- Public Use Microdata Statistics (PUMS)
- ACS American FactFinder tables (e.g., median earnings by industry, block-level industry data, state ranking of % employed in manufacturing)

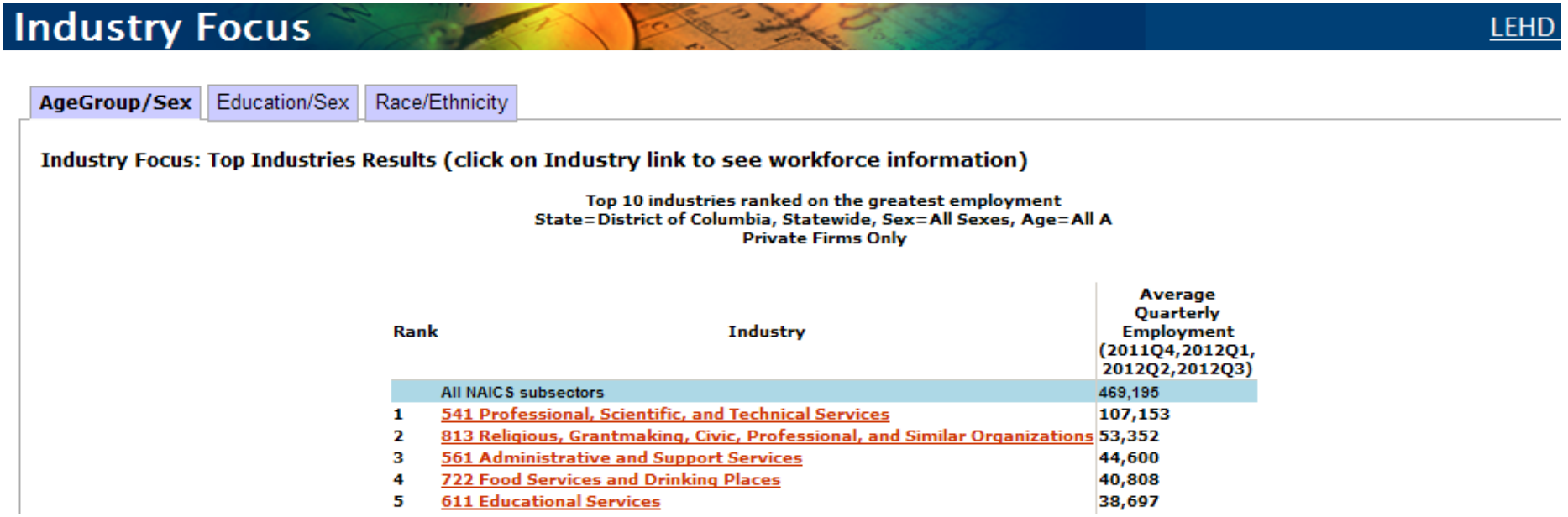
Rank	Geographical Area	Percent	Margin of Error
	United States	10.5	+/-0.1
1	Indiana	19.0	+/-0.4
2	Wisconsin	18.4	+/-0.3
3	Michigan	17.2	+/-0.3

1 - 53 of 53

Industry in Public-Use Data Tools, Cont.

Employer-reported data

- Quarterly Workforce Indicators (QWI) Online
- OnTheMap: Home/Work Area Profile Analysis
- Industry Focus



Industry Focus

LEHD

AgeGroup/Sex | Education/Sex | Race/Ethnicity

Industry Focus: Top Industries Results (click on Industry link to see workforce information)

Top 10 industries ranked on the greatest employment
State=District of Columbia, Statewide, Sex=All Sexes, Age=All A
Private Firms Only

Rank	Industry	Average Quarterly Employment (2011Q4,2012Q1, 2012Q2,2012Q3)
	All NAICS subsectors	469,195
1	541 Professional, Scientific, and Technical Services	107,153
2	813 Religious, Grantmaking, Civic, Professional, and Similar Organizations	53,352
3	561 Administrative and Support Services	44,600
4	722 Food Services and Drinking Places	40,808
5	611 Educational Services	38,697

Study Motivation

- Compare person-reported industry to employer-reported industry. Each source of industry statistics has potential advantages and disadvantages (e.g., universe coverage, respondent reporting accuracy, coding and editing expertise, level of detail provided).

  How similar are the resulting industry estimates?

  Can employer-based administrative data accurately supplement missing person-reported data?

Data Sources

American Community Survey (ACS) 2009

- Largest U.S. household survey. About 3.2 million addresses sampled each year.
- Mandatory survey with high response rate.
- Collects data via mail, telephone or personal interview, and online (starting in 2013).

Longitudinal Employer-Household Dynamics (LEHD) Data

- Administrative data integrated by the Census Bureau.
- Unemployment Insurance (UI) earnings data. Person can have earnings from more than one employer.
- UI excludes self-employed, independent contractors, military, federal government, most agriculture and railroad workers.
- Quarterly Census of Employment and Wages (QCEW) provides firm data (e.g., geographic location, industry).

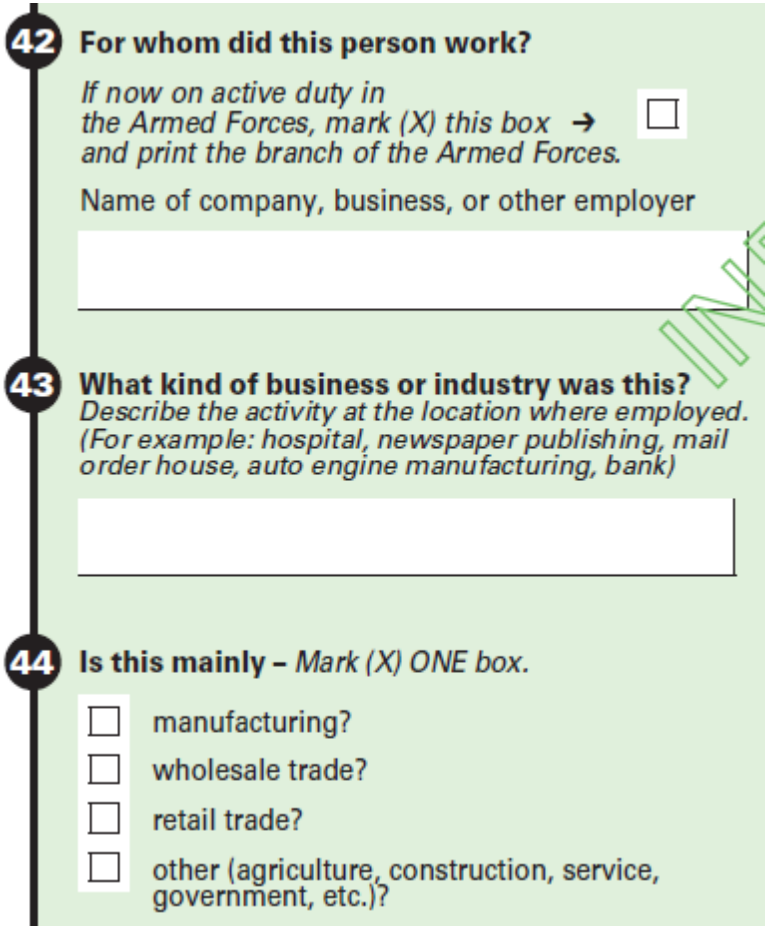
ACS Industry Data

- Data on the primary job of all respondents aged 16 and over who are employed or were employed within the past 5 years.

- Written responses coded into 269 Census Industry Code (CIC) categories by clerical staff at the National Processing Center in Jeffersonville, IN.

- Standardized coding and editing procedures:
  - Checked for consistency with occupation, class of worker (private, government, self-employed), and other variables
  - If missing, industry code imputed from a “similar” person

ACS Industry Questions



42 **For whom did this person work?**

If now on active duty in the Armed Forces, mark (X) this box → ☐ *and print the branch of the Armed Forces.*

Name of company, business, or other employer

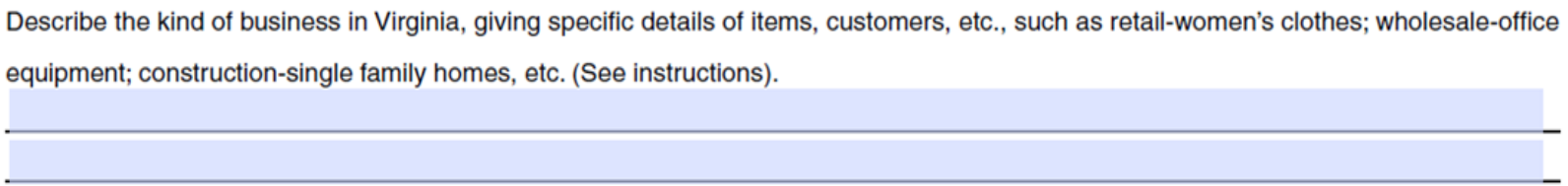
43 **What kind of business or industry was this?**
Describe the activity at the location where employed. (For example: hospital, newspaper publishing, mail order house, auto engine manufacturing, bank)

44 **Is this mainly –** *Mark (X) ONE box.*

- ☐ manufacturing?
- ☐ wholesale trade?
- ☐ retail trade?
- ☐ other (agriculture, construction, service, government, etc.)?

LEHD Industry Data (QCEW)

- Initial industry information from UI liability forms
  - E.g. Virginia Report to Determine Liability for State Unemployment Tax (VEC FC-27)


11. Describe the kind of business in Virginia, giving specific details of items, customers, etc., such as retail-women's clothes; wholesale-office equipment; construction-single family homes, etc. (See instructions).

  - Coded into 6-digit NAICS codes
  - Missing data filled in longitudinally, if possible. Otherwise imputed.
- In some states QCEW industry used for initial UI tax rates

Linking ACS and LEHD Job Data

- Individuals in LEHD data are identified with Protected Identification Keys (PIKs). These are assigned by the Census Bureau based on personal identifying information in the Unemployment Insurance system.
- About 92 percent of the ACS sample had a PIK that matched to LEHD.
- We link this subsample to the LEHD earnings history data and then select the LEHD job that is most likely to correspond to the job referenced in the ACS.
- Likely LEHD job matches are found using ACS response date. The response date (RDATE) is based on questionnaire “today’s date” or when survey is marked complete if finished via telephone or personal interview. RDATE is used to match to jobs in the appropriate LEHD quarter.
- Of individuals with at least one plausible LEHD main job, 87 percent have one plausible main job and 13 percent have multiple plausible jobs.

ACS-LEHD Job Match Challenges

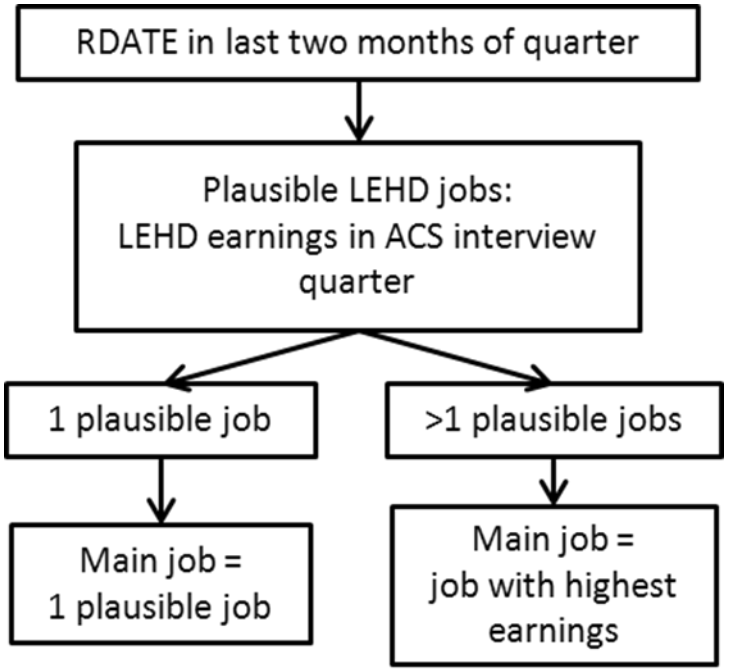
- **Sample**

  Select ACS sample likely to be in LEHD data:
  - Keep currently employed
  - Exclude federal, self-employed, working in MA
- **Timing**
  - Select LEHD quarter(s) to match to reported ACS job
  - ACS RDATE not exact; ACS job may link to multiple LEHD quarters.
- **Employer Identification**
  - ACS respondent may match to multiple LEHD jobs (State Employer Identification Numbers - SEIN) if worked multiple jobs or changed jobs within a quarter.
  - SEINs can have multiple establishments with different industry codes. For most states, we do not know which establishment an individual works in.

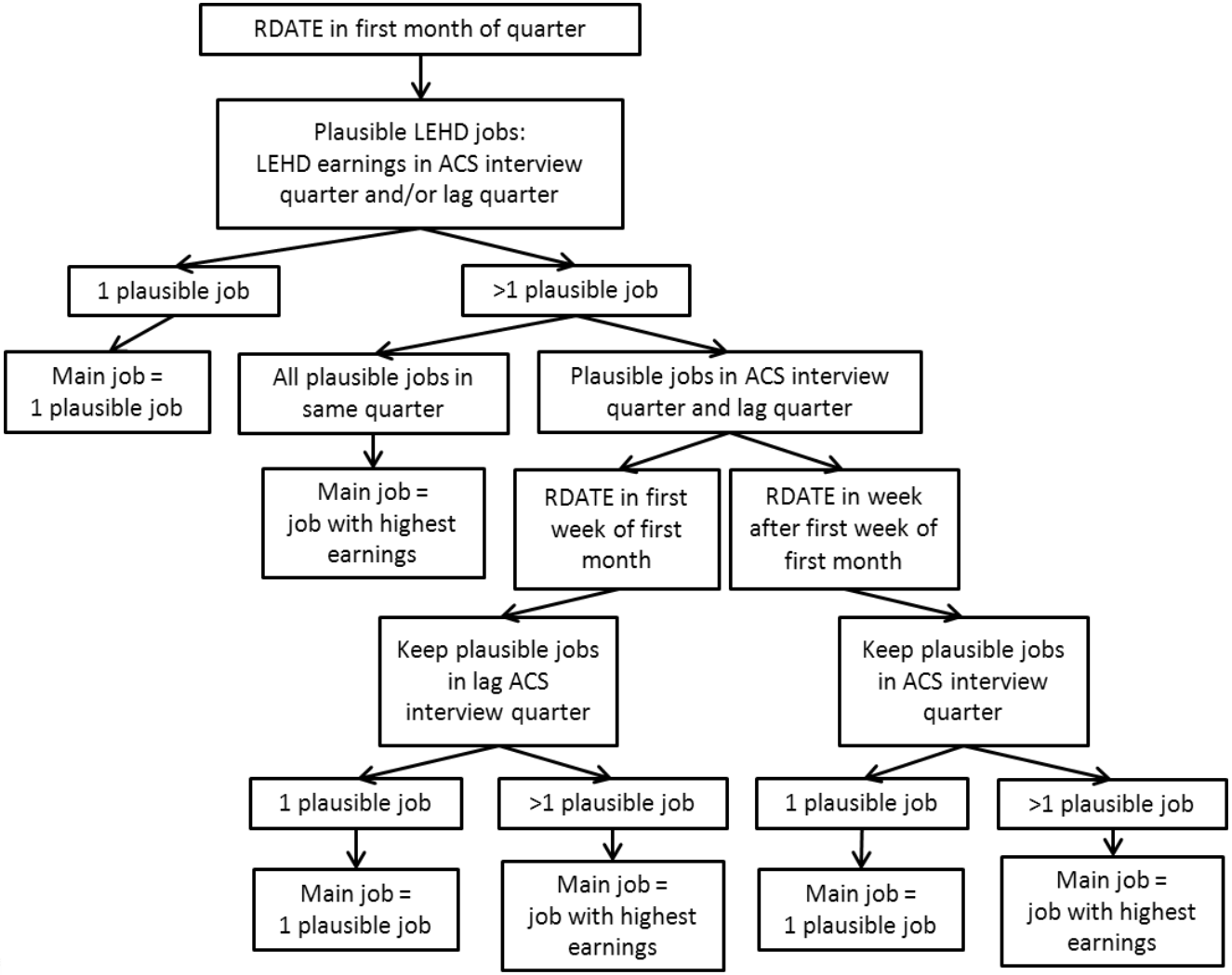
Sample Selection

	n
2009 ACS Population File	4,534,632
Populations Excluded from Analysis Sample	
NOT employed, at work	2,532,851
Federal government employee	84,063
Self-employed not incorporated	190,700
Self-employed incorporated	94,188
Unpaid family workers	7,934
Work in Massachusetts	43,996
ACS subsample	1,692,601
ACS subsample with PIKs	1,550,845
ACS subsample with at least one plausible LEHD job	1,532,355

Job Match Algorithm, Part 1



Job Match Algorithm, Part 2



Results: Overall Agreement Rates

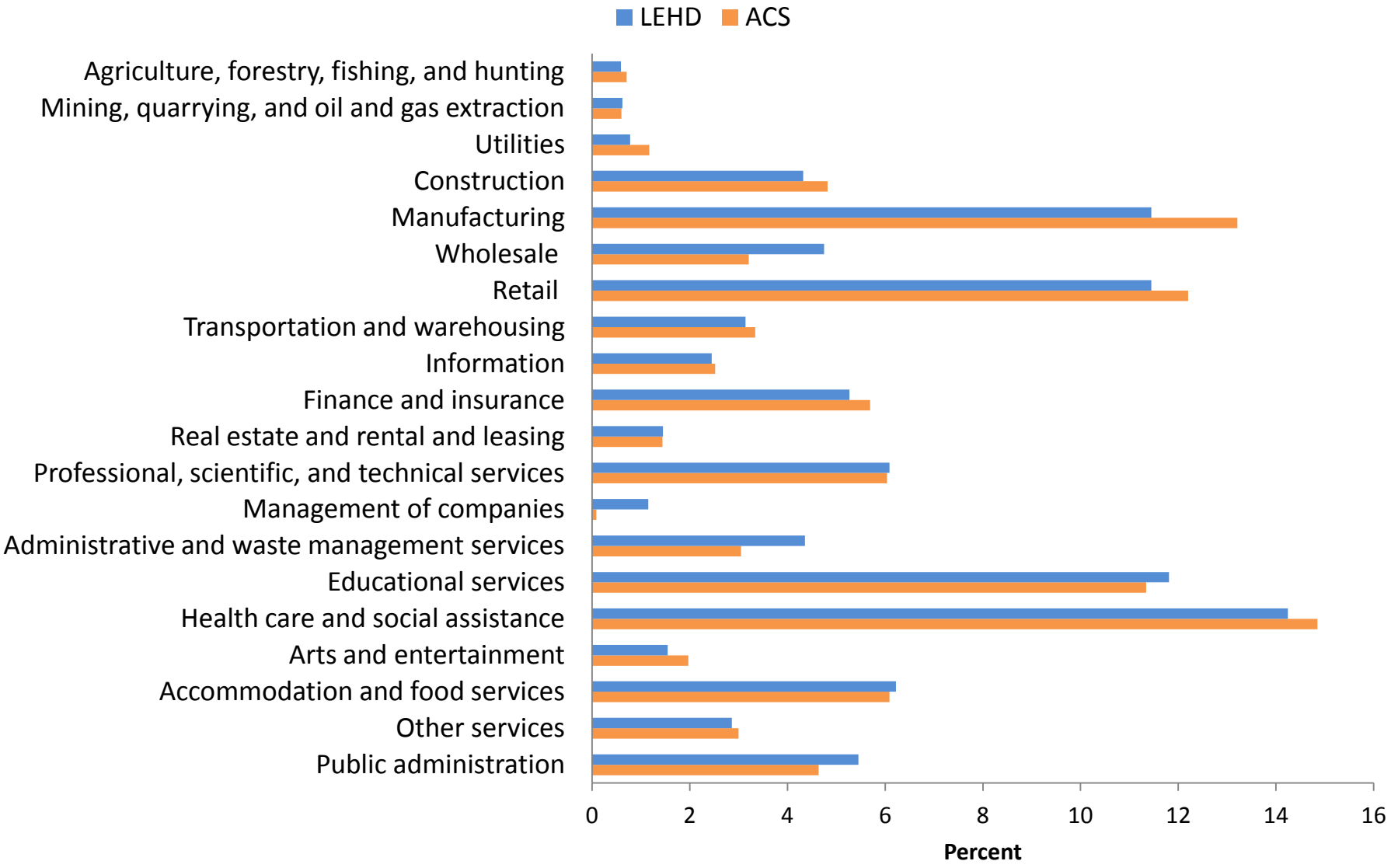
Industry agreement = industry is the same in ACS and LEHD

- Industry 20 sector agreement rate: 75%
- Census 4-digit industry agreement rate: 61%
  - Similar rates in prior work (Stinson, Gathright, and Skog 2012) using Survey of Income and Program Participation and the Census Bureau Business Register (75% sector and 56%-63% 4-digit industry agreement)
  - Industry agreement rates using 2 business data sources (QCEW and BR) is 87% when including businesses on both lists (Elvery et al. 2006)
- Low agreement rate for ACS imputed cases
  - 14% at the sector level
  - 5% at the 4-digit industry level

Results: Agreement Rates by Industry

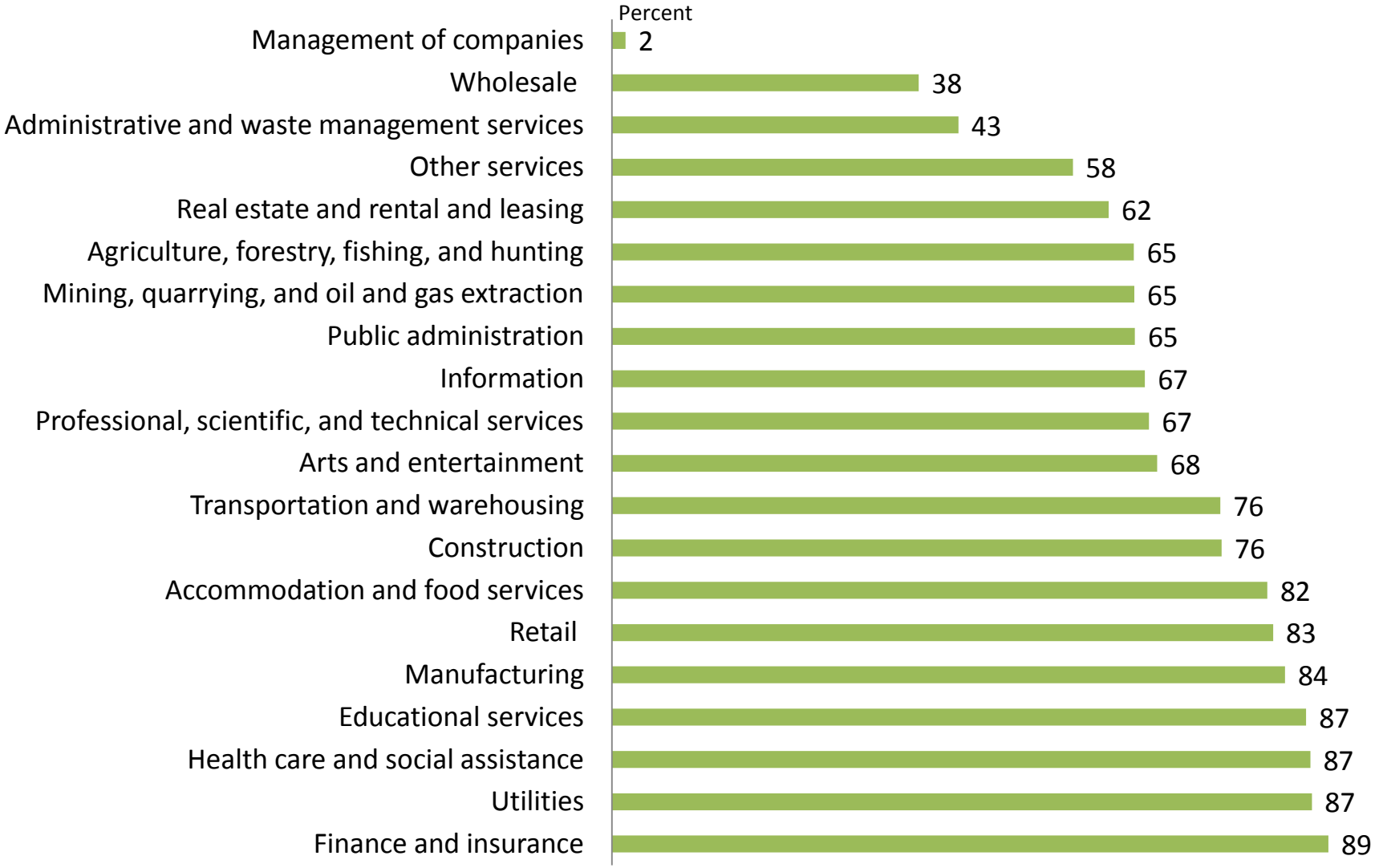
- Industry distributions in ACS and LEHD are similar.
  - More manufacturing and retail employment in ACS
  - More wholesale, management of companies, and public administration employment in LEHD
- Industry agreement rates vary by data source:
  - ACS as baseline
  - LEHD as baseline
- Industry agreement rates vary by sector:
  - Highest in finance and insurance, educational services, health care and social assistance, and utilities
  - Lowest in management of companies

Industry Sector Distribution



Industry Agreement Rate

(LEHD as baseline)



Industry Sector Mismatches

Management of companies

- Using LEHD as baseline, 2% agreement
- In ACS classified as:
  - Manufacturing: 28%
  - Finance and insurance: 13%

Educational services

- Using LEHD as baseline, 87% agreement
- In ACS classified as:
  - Health care and social assistance: 4%
  - Public administration: 3%

Data Source Tradeoffs

ACS

- Lacks data on multiple jobs**.**
- Proxy reporting may be less accurate for some respondents.
- Write-in responses lack precision to reliably code some industries (e.g., management of companies).

LEHD

- Do not know worker's establishment. About 42% of employers have multiple establishments, 61% with industry variation.
- Lacks data on occupation and hours worked.
- Imputation rate varies by state.
- Decentralized data collection operation. State and local training, coding, and editing procedures difficult to assess.

Conclusion

- Management of companies and wholesale trade sectors have low agreement rates; Educational services, health care and social assistance have high agreement rates.

- While aggregate industry distributions are similar even when using imputed data, ACS imputed industry matches poorly to LEHD industry at the person level.
  - Use imputed values with caution; multivariate analyses will differ

- Using LEHD to supplement ACS missing industry data would be difficult:
  - Those missing ACS industry data are less likely to be currently employed: 21% unemployed and 31% not in the labor force
  - Current or most recent job may not be in universe for LEHD
  - No occupation data in LEHD

Contact

Liana Christin Landivar
Industry and Occupation Statistics Branch
Social, Economic, and Housing Statistics Division
U.S. Census Bureau
301-763-5878
liana.christin.landivar@census.gov

Paper available at:
http://www.census.gov/people/io/publications/wp_posters.html