

Occupation reports on wage records:

What we've learned so far from Nebraska and Indiana data

Elizabeth Weber Handwerker
Employment Research Staff

June 27, 2023



The basic idea

- Some states collect occupation information on wage records
 - ▶ Where this is voluntary, which employers send in occupation information?
 - ▶ How do occupation & wage information on wage records compare with information collected in the OEWS *for the same employers*
- OEWS = the Occupational Employment and Wage Statistics Survey of 400,000 establishments/year
- Any potential to use occupation information on wage records to reduce or re-target the OEWS sample?



Occupation data on wage records

State	Item	Years
Alaska	SOC Code or Job Title	Decades but state law prohibits sharing
Connecticut	SOC Code or Job Title	Not yet
Indiana	SOC Code	2019 - present
Louisiana	SOC Code or Job Title	??
Nebraska	Job Title	Recent years
South Carolina	SOC Code	??
Virgin Islands	Job Title	??
Washington	SOC Code	Voluntary in 2022, mandatory in 2023

Source: "An inventory of Employee-Specific Data Collected on Unemployment Insurance Wage Records: A Study Conducted for the BLS Labor Market Information Oversight Council and the LMI Institute" January 2022



Wage Record occupation data now at BLS

- We received **job title data** from **Nebraska** last August
 - ▶ Occupation collected on a **voluntary** basis.
 - ▶ Nebraska shared data with us from 2018 and 2021.
- We received **SOC data** (and other items) from **Indiana** in September
 - ▶ Items collected on a **voluntary** basis.
 - ▶ Indiana shared data with us from 2019 through early 2022.
- Note: Nebraska shared only wage records with job titles; Indiana shared all wage records with any supplemental data (such as months worked and full-time/part-time).

Great to have multiple states!

- Insights into what varies from state to state and how much differences matter
 - ▶ Job Titles vs SOC Codes
 - ▶ Support for respondents
 - ▶ Data processing
- Insights into the robustness of patterns
 - ▶ Patterns of voluntary data provision
 - ▶ Patterns of similarity with OEWS

Comparison Questions

- (1) When occupation reports on wage records are voluntary, which employers report this information?
- (2) How many of them also respond to the OEWS?
- (3) How do occupation reports compare *for the same employers*?
- (4) How do wage reports (by occupation) compare *for the same employers*?



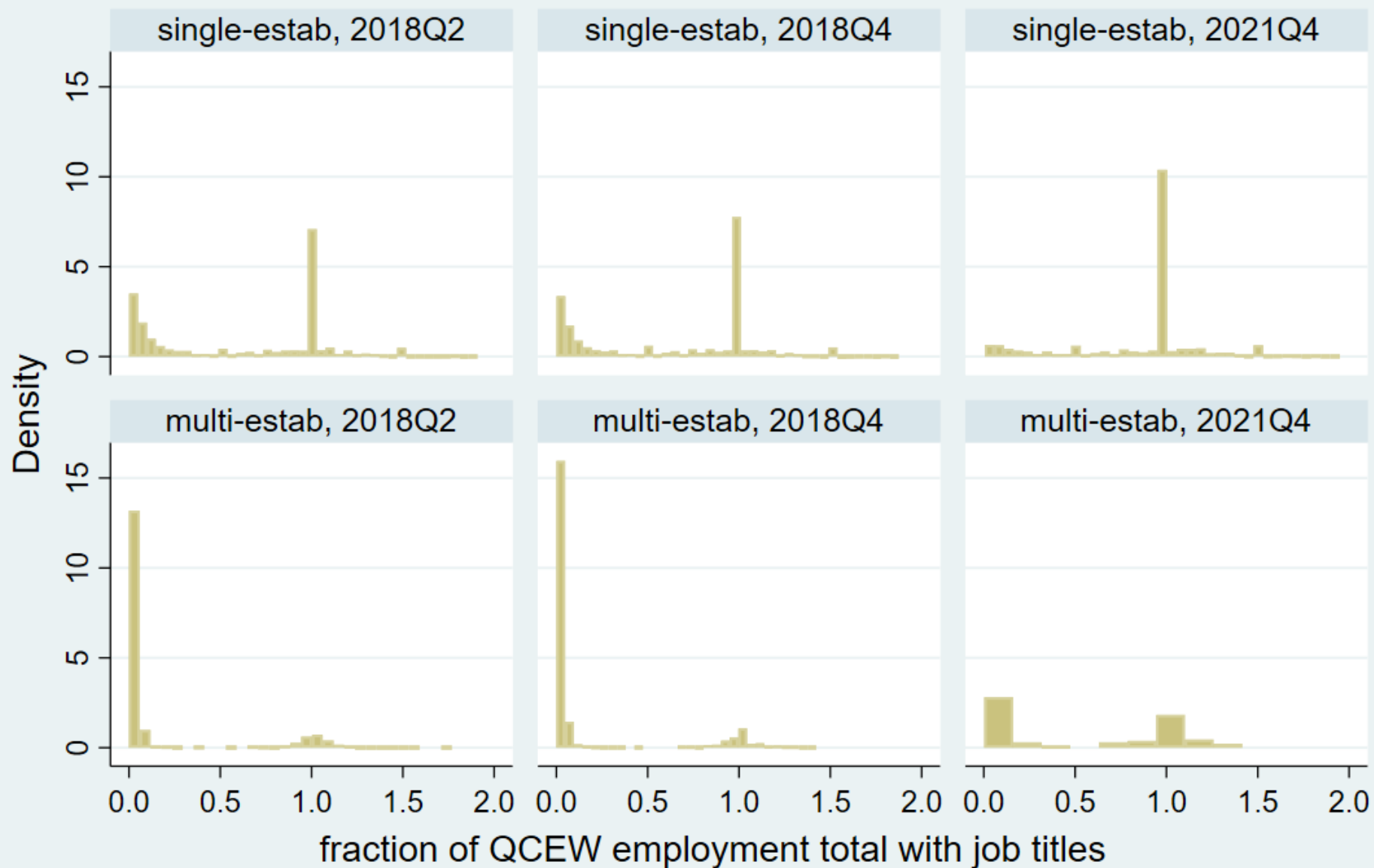
Question 1: Which employers voluntarily report occupations?

1. Link wage records with their UI account information in the QCEW and estimate $\frac{\textit{Occupation reports}}{\textit{Average QCEW Employment}}$
 2. See how this fraction varies by employer size, number of establishments, industry, geographic area, average wage levels, employer age, etc.
 3. Use multi-variate regressions to show that employer size does NOT explain all other patterns.
- Geographic and industry patterns of which employers report occupations differ between Nebraska and Indiana

Nebraska

- In 2018, 17% of UI accounts provided at least some job titles
- In 2021, Nebraska changed their reporting system, and afterwards gave us almost no job titles from UI accounts with 50+ employees
 - ▶ Nebraska staff have investigated this: job titles from file uploads of wage reports (instead of direct entry) were not initially linked with other data in a way LMI could access
 - ▶ We got job titles from only 10% of UI accounts
 - ▶ Among UI accounts for which we got job titles, the number of job titles provided is generally closer to QCEW employment in 2021 than in 2018

UI Accounts with some job title reporting



Graphs by multi_date

Indiana

- Indiana started collecting SOC codes on a voluntary basis in 2019.
 - ▶ They got them from 25% of employers in 2019Q1
 - ▶ Increased to 52% of employers by 2021Q4
- Many Indiana employers also report months worked per quarter and full-time/part-time work and Indiana shared these variables
 - ▶ We can use these variables to weight SOC codes (work in 1 month = 1/3 weight of work in 3 months)

Comparing Occupation Reporting Patterns in Nebraska & Indiana

	Nebraska	Indiana
By employment in UI	↑ for less employment	↑ for less employment
By establishment count	↑ for fewer estabs	↑ for fewer estabs
By average quarterly wage for the UI Account	↑ for accounts with lower average quarterly wages	↑ for accounts with middle average quarterly wages
By UI Account Age	↑ for older accounts	↑ for older accounts
By dominant Industry of the UI Account	↑ in Public Admin, Extraction, Other Services, Transportation ↓ in Management, Health care, Accommodation	↑ in Agriculture, Public Admin, Extraction, Finance, Education ↓ in Management, Wholesaling
By dominant Geography of the UI Account	↑ in non-MSA areas , Sioux City MSA ↓ in Omaha-Council Bluffs MSA, Lincoln MSA	↑ Bloomington MSA, Elkhart-Goshen MSA, Kokomo MSA ↓ in Lafayette MSA, non-MSA areas

Can we predict which employers voluntarily provide occupations on wage records?

- Unfortunately—NO.
 - ▶ R^2 from regressions of occupation reporting on observable characteristics have values of about .05
 - ▶ More sophisticated data science methods, such as regression trees, reach same conclusion
- Past provision of these data does predict current provision of data

Question 2: How many employers overlap between wage records and OEWS?

- Remember, wage records contain UI accounts, *not exact establishments*, and the OEWS is an *establishment* survey
 - ▶ Single-establishment UI accounts—no problem
 - ▶ We only count multi-establishment UI accounts as overlapping if we have similar employment (within 50%) to compare

Nebraska

- 281 overlaps in 2018Q2
- 206 overlaps in 2018Q4
- 133 overlaps in 2021Q4

Indiana

- 1,210 overlaps in 2019Q2
- 1,432 overlaps in 2019Q4
- similar in 2020, 2021...
- 1,496 overlaps in 2021Q4

Question 3: How similar are occupation reports in wage records and OEWS responses?

Data preparation needed for both states!

- Nebraska: We auto-coded Job Titles to SOC using “Suggestion 1” of the OEWS auto-coder
- Indiana: convert the SOC codes employers reported to the same SOC as the OEWS

Using the OEWS Auto-coder for Nebraska

■ Nebraska collects Job Titles

▶ OEWS staff auto-coded Job Titles to SOC codes

- We need to add employer NAICS codes (from the LDB)
- This gave us a most-likely SOC code for each job title, plus an estimated probability of accurate coding.
 - The OEWS autocoder is confident about how to code a “registered nurse.” Confidence in the SOC code for a “field representative” or “laborer” varies by industry.
- The OEWS auto-coder was trained on OEWS data, not wage data. This affects how well it works on these data
 - In coding OEWS data, the autocoder’s “Suggestion #1” is only chosen 78% of the time

Recoding SOC data from Indiana

- Indiana asks employers to report SOC codes
- We converted many SOC codes employers reported in Indiana to the same SOC version OEWS used
 - ▶ The OEWS began collecting data using SOC 2018 beginning with the November 2018 panel
 - ▶ Indiana accepts *any* SOC code that was *ever* valid.
 - As of late 2022, the <https://www.hoosierdata.in.gov/coder/> SOC autocoder was providing SOC 2010 codes
- Many Indiana employers reported a mix of broad and detailed SOC codes

Question 3: How similar are occupation reports in wage records and OEWS responses?

We use two methods to compare SOC codes

1. What fraction of occupation reports are in the same MAJOR (2-digit) SOC code?
 - ▶ Easy to calculate, but a very crude measure of similarity
 - ▶ Average is **.354** for Nebraska UIs and **.577** for Indiana UIs
2. Estimate an index of overall occupational closeness, selecting the “most similar” of all possible combinations of non-overlapping occupations for each pair of observations
 - ▶ Based on similarity of pairs of detailed SOC codes
 - ▶ Allows a score between 0 and 1 for pairs of detailed SOC codes that are not identical
 - ▶ Hard to calculate for larger employers, but more precise



How to compute similarity index

1. Start with a matrix (from Atalay, Sotelo, and Tannenbaum) of the similarity of each pair of occupations (based on online job descriptions)
2. Where occupations match, similarity = 1
3. Where occupations don't match, try every combination of non-matching occupations. Pick the combination with the highest overall score, weighted by employment.
 - ▶ This can handle a mix of SOC systems (2000, 2010, 2018), but *not* a mix of broad and detailed SOCs for an employer
 - ▶ Checking all possible combinations of non-matching occupations gets computationally intensive as the number of non-matching occupations ↑
 - ▶ Average is **.337** for Nebraska UIs and **.587** for Indiana UIs

Patterns of similarity scores

Nebraska:

- Unsurprisingly, in the Nebraska data, occupations are more similar for UI Accounts when the auto-coder was more confident about the SOCs (on average)
- Similarity was higher in Nebraska in 2021 than 2018
- The correlation between the two measures is .85
- Similarity *patterns* are very similar for the two methods of comparing occupations

Patterns of similarity scores

Indiana:

- We can estimate the fraction of Major SOCs that match for a LOT more Indiana employers than we can estimate the Occupational Similarity Score
 - ▶ This is mostly due to many employers reported a mix of broad and detailed SOC codes on wage records (not so much due to the computational intensity of large numbers of occupations that don't match between the OEWS and wage records)
- The correlation between the measures is .52
- Similarity *patterns* are very similar for the two methods of comparing occupations

Comparing Occupational Similarity Patterns in Nebraska & Indiana

	Nebraska	Indiana
By employment in UI	↑ for less employment	↑ for less employment (fraction of Major SOCs only)
By average quarterly wage for the UI Account	↑ for accounts with middle average quarterly wages	Little pattern
By UI Account Age	Little pattern	↑ for older accounts
By dominant Industry of the UI Account	↑ in Finance, Construction, Other Services, Real Estate ↓ in Public Admin, Education, Accommodation	↑ in Construction, Transportation, Finance ↓ in Management, Agriculture, Arts
By dominant Geography of the UI Account	↑ in non-MSA areas , Sioux City MSA ↓ in Omaha-Council Bluffs MSA	↑ Bloomington MSA, Cincinnati MSA, Columbus MSA, Louisville MSA ↓ in Evansville MSA, Indianapolis MSA, Lafayette MSA, South Bend MSA

Can we predict which employers have more similar occupation scores?

■ Unfortunately—NO.

- ▶ R^2 from regressions of occupational similarity on observable characteristics have values of
 - .17 for the fraction of major SOCs overlapping in NE
 - .22 for the occupational similarity score in NE
 - .06 for the fraction of major SOCs overlapping in IN
 - .07 for the occupational similarity score in IN
- ▶ More sophisticated data science methods, such as regression trees, reach same conclusion

Question 4: How similar are occupation-specific wages in wage records and OEWS?

Wage Records in Indiana

Wages are defined as cash or non-cash compensation received for services performed. Wages include salaries, bonuses, commissions, vacation pay, retroactive pay increases, and any other payments made by an employer unless specifically excluded by the Act. An example of compensation that is excluded by the Act is the value of Cafeteria (Section 125) Benefit plans.

We expect these wages to be larger

OEWS Wages

Wages for the OEWS survey are straight-time, gross pay, exclusive of premium pay. Base rate; cost-of-living allowances; guaranteed pay; hazardous-duty pay; incentive pay, including commissions and production bonuses; and tips are included. Excluded are overtime pay, severance pay, shift differentials, nonproduction bonuses, employer cost for supplementary benefits, and tuition reimbursements.



Question 4: How similar are occupation-specific wages in wage records and OEWS?

(We can only examine this for Indiana)

- For each UI account and each detailed SOC that appears in both datasets (about 25,000 combinations), estimate:

Average Quarterly wage in wage record – Average OEWS quarterly wage

Average Difference (Weighted by employment) = **\$623**

Average % difference (Weighted by employment) = **11%**

- Indiana asks if workers were employed in every month. So estimate this for full-quarter workers in UI Accounts with similar employment levels in wage reports and OEWS

Average Difference (Weighted by employment) = **\$1,118**

Average % difference (Weighted by employment) = **18%**

Similarity by OEWS collection type

- OEWS used to collect wage data in intervals
- New data collection system gathers point wages, either hourly (multiply by 520 to compare with UI) or annual (divide by 4)

Average Quarterly wage in wage record – Average OEWS quarterly wage =

Collection	All comparisons	Full-quarter +
Interval	\$1,754	\$2,553
Hourly	-\$192	\$105
Annual	\$459	\$1,024

Comparing Wage Difference Patterns

	All comparisons	Full-quarter +
By period	Negative in 2020Q2, otherwise positive, largest in 2021Q4	Positive in all quarters Largest in 2021Q4
By employment in UI	Negative for < 10 employees Positive for 10+ employees	Negative for sole proprietors and employment 6-9
By number of estabs	Smaller difference for multistarts	Smaller difference for multistarts
By average quarterly wage for the UI	Negative for avg wage < \$40K Positive for larger avg wages	Negative for avg wage < \$20k Positive for larger avg wages
By UI Account Age	Negative for age < 2 years	Positive for all ages
By dominant Industry of the UI Account	Negative for Arts, Health Care Big Positive for Extraction, Utilities, Wholesaling	Negative for Health Care Big Positive for Extraction, Utilities, Wholesaling
By Major SOC Category	Big Negative for Management, Computer & Math Big Positive for Business, Legal, Installation & Repair	Big Negative for Management, Computer & Math Big Positive for Business, Sales, Legal, Engineer, Install
By dominant Geography of the UI	Negative in Muncie MSA Big Positive in Lafayette MSA	Positive for all areas Biggest Diff for Lafayette MSA

Positive: Wage Record wage > OEWS wage

Negative: OEWS wage > Wage Record wage

Again, little power to predict which employers have more similar wages

- Low R^2 values from regressing occupation-specific wage differences on available employer characteristics: size, size², wage levels, wage², age, age², industry dummies, major occupation dummies, geography dummies...
- More sophisticated data science methods yield same conclusion: little predictive power

Summary findings

- Could we use information from enhanced wage records to
 - ▶ improve OEWS imputations?
 - ▶ increase the precision of OEWS estimates?
 - ▶ reduce OEWS sample size for employer types that provide similar information in wage records?
- Based on results shown today, we are not optimistic YET...but these data are worth watching for BLS

Advantages of these data

Benefits (aside from what the OEWS does now):

- The potential to measure seasonality of occupational employment
- The potential to observe mobility between occupations, within and between employers and locations

There is a big push now to improve UI systems

- Several groups are pushing to add these fields as part of overall UI system improvement
- How this is implemented matters!

Contact Information

Employment Research and Program Development Staff

<https://www.bls.gov/ers/contact.htm>

BLS Wage Record Staff

wagerecords@bls.gov

BLS Occupational Employment and Wage Statistics Staff

https://www.bls.gov/oes/oes_con.htm

