Federal Government Wage Indexes*

Travis Cyronek Ted To

FCSM Research & Policy Conference

23 October 2024

∗: The views and opinions expressed herein are those of the authors and do not necessarily reflect any official positions, policies, or products of the US Bureau of Labor Statistics.

Employer Cost Index (ECI)

- One of seven BLS programs producing **PFEI**s
- Fed by establishment data collected in the **National Compensation Survey**
- Measures the change in hourly costs to employ a fixed basket of labor services
- Currently covers *Private* and *State & Local Government* ownership
  - but no *Federal Government*...

This Paper

Can OPM salary data be used to construct wage indexes for Federal workers?

- The *Federal Government* is the largest US employer (3.2 million)
- It would be useful for researchers and policy makers
- Can be accomplished without major changes to current procedures
  - utilizing administrative data that OPM already has
  - guided by ECI, but avoids having to expand NCS collection efforts

Challenges

- The ECI needs occupation (SOC) and industry (NAICS) for cell aggregation
- OPM has their own codes...
  - occupations: BLS has a crosswalk (easy)
  - industries: machine-learning tools (more difficult)
- **Goal**: OPM department and agency + QCEW $\longrightarrow$ agency-to-NAICS concordance

OPM Salary Data

- Individual Federal employees; Q2 for 2020, 2021, and 2022
- Annual (full-time) salary
- OPM occupation
- Full-time / part-time status
- Grade
- Agency
- City and state

Summary Stats (Occupation)

	2020-Q2		2021-Q2		2022-Q2	
	N	%	N	%	N	%
Occupation Group						
Management, business, and financial occupations	1,588,381	49.64	1,608,050	49.48	1,604,617	49.97
Professional and related occupations	924,123	28.88	949,506	29.21	940,476	29.28
Office and administrative support occupations	301,965	9.44	304,487	9.37	293,667	9.14
Service occupations	251,655	7.86	255,296	7.86	244,730	7.62
Transportation and material moving occupations	75,143	2.35	74,427	2.29	71,750	2.23
Construction, extraction, farming, fishing and forestry	17,075	0.53	16,680	0.51	16,408	0.51
Installation, maintenance and repair occupations	18,298	0.57	18,862	0.58	18,380	0.57
Production occupations	12,549	0.39	12,496	0.38	12,115	0.38
Sales and related occupations	10,908	0.34	10,279	0.32	9,330	0.29
Total	3,200,097	1.00	3,250,083	1.00	3,211,473	1.00

Summary Stats (Industry)

	2020-Q2		2021-Q2		2022-Q2	
	N	%	N	%	N	%
Industry Group						
Public administration	3,005,275	93.91	3,052,558	93.92	3,005,394	93.58
Rest of Services	80,381	2.51	82,526	2.54	83,502	2.6
Hospitals	39,908	1.25	40,945	1.26	49,224	1.53
Wholesale and Retail Trade	37,199	1.16	36,203	1.11	35,752	1.11
Goods Producing	11,076	0.35	11,176	0.34	11,205	0.35
Elementary and secondary schools	9,012	0.28	9,176	0.28	9,299	0.29
Transportation and warehousing	7,123	0.22	7,117	0.22	7,345	0.23
Rest of Health Services	5,975	0.19	6,174	0.19	6,126	0.19
Colleges, universities, and professional schools	3,273	0.1	3,330	0.1	2,774	0.09
Nursing and residential care facilities	875	0.03	878	0.03	852	0.03
Total	3,200,097	1.00	3,250,083	1.00	3,211,473	1.00

Summary Stats (Work Schedule)

	2020-Q2		2021-Q2		2022-Q2	
	N	%	N	%	N	%
Work Schedule						
full time	3,097,080	96.78	3,147,790	96.85	3,115,654	97.02
part time	103,017	3.22	102,293	3.15	95,819	2.98
Total	3,200,097	1.00	3,250,083	1.00	3,211,473	1.00

QCEW Data

- Serves as the sampling frame for the NCS
- Comes from state UI accounting systems
- NAICS industry
- Establishment size
- Employment counts

Step 1: Some Cleaning . . .

- Drop military and postal service from both datasets
- Convert OPM salaries to hourly (divide by 2,087 hours)
- Aggregate OPM data to the "establishment" level (like the NCS / QCEW)
  - imperfect proxy if agency has more than one location in a city
- Use OPM-SOC crosswalk for occupations

Step 2: Vectorize OPM Agencies / QCEW Establishments

- Use **TF–IDF** to create measures of the importance of "terms"
- Multiple options to handle descriptions . . .
  - Bag-of-words unigrams: unordered list of the words
  - **Character n-grams**: contiguous sequence of n characters from some text
  - We choose the latter; more robust to misspellings / variations

Step 3: Matching

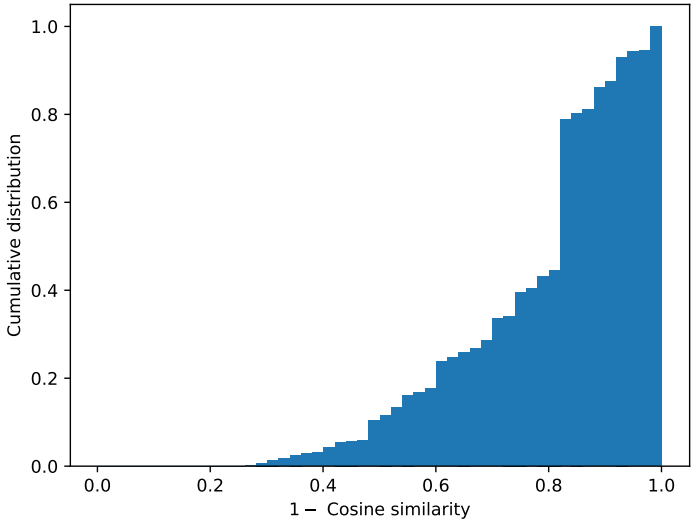
- For each possible match, calculate **cosine similarities**
- Construct **weighted** cosine similarity:

$$WeightedCosSim = CosSim \times E_j^{qcew} \times EmpPenalty_j$$

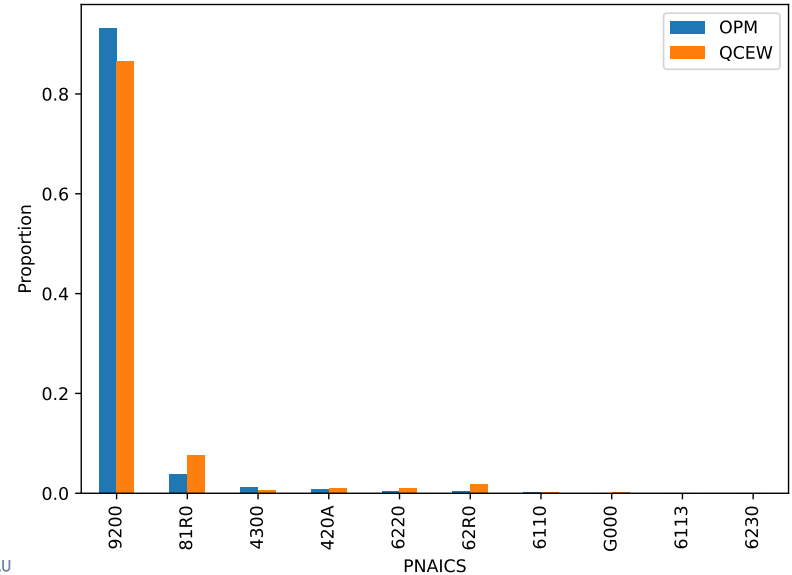
$$EmpPenalty_j = 1 - \frac{\left| E_j^{opm} - E_j^{qcew} \right|}{E_j^{qcew}}$$

- Selection Rule: choose match with the best $WeightedCosSim$

Weighted Distribution of Cosine Similarity Scores



Comparison of the Distribution of PNAICS Codes



Wage Index Formula

$$\underbrace{I_L = \frac{\sum w_1 * e_0}{\sum w_0 * e_1}}_{Laspeyres} \qquad \underbrace{I_P = \frac{\sum w_1 * e_1}{\sum w_0 * e_1}}_{Paasche} \qquad \underbrace{I_F = I_L^{1/2} * I_P^{1/2}}_{Fisher}$$

- $\sum$: over cells (industry-by-occupation)
- w_t: cell average wage in t
- e_t: cell weight in t

Results (Aggregate)

	Laspeyres	Paasche	Fisher
2020Q2–2021Q2	1.0131	1.0131	1.0131
2021Q2–2022Q2	1.0342	1.0341	1.0341

Results (Work Schedule)

		Laspeyres	Paasche	Fisher
Full-time	2020Q2-2021Q2	1.0130	1.0129	1.0129
	2021Q2-2022Q2	1.0337	1.0337	1.0337
Part-time	2020Q2-2021Q2	1.0366	1.0361	1.0363
	2021Q2-2022Q2	1.0427	1.0425	1.0426

Results (Census Region)

		Laspeyres	Paasche	Fisher
Northeast	2020Q2-2021Q2	1.0139	1.0139	1.0139
	2021Q2-2022Q2	1.0388	1.0388	1.0388
South	2020Q2-2021Q2	1.0155	1.0155	1.0155
	2021Q2-2022Q2	1.0361	1.0360	1.0360
Midwest	2020Q2-2021Q2	1.0111	1.0111	1.0111
	2021Q2-2022Q2	1.0322	1.0322	1.0322
West	2020Q2-2021Q2	1.0180	1.0179	1.0180
	2021Q2-2022Q2	1.0412	1.0411	1.0412

Results (Comparison to ECI)

	Private	State & Local	Federal (exploratory)
2020Q2–2021Q2	1.0356	1.0162	1.0131
2021Q2–2022Q2	1.0567	1.0323	1.0342

Concluding Remarks

- While annually announced federal pay increases provide some information about federal employment cost growth, it is imprecise.
- Actual growth depends on ...
  - the flow of employees into and out of federal service
  - the mix of employee tenures
- We demonstrate the practicality of using OPM data to compute federal government wage indexes in the spirit of the ECI
- If benefit-cost and hours data are obtainable, other series could also be computed (total compensation, total benefits)

Contact Information

Travis Cyronek
cyronek.travis@bls.gov

Ted To
to.theodore@bls.gov