
Quarterly Census of Employment and Wages

Data Compilation and Estimates Process

Introduction

This report describes the processing steps used to prepare the Quarterly Census of Employment and Wages (QCEW) annual employment and wage data from the Bureau of Labor Statistics (BLS) for use by the EPS application. These annual, county-level data are processed at the 6-digit NAICS industry level for 2001 through the latest published year. The steps include downloading the data from the BLS web site, calculating estimates of employment and wages when data are not disclosed by the BLS, and importing the reported and estimated data into a standard format used by the EPS application. The purpose of this report is to describe the estimation method and the resulting accuracy.

Reporting and nondisclosure criteria

Due to the government obligation to not report data that can be used to identify individual firms, the BLS suppresses observations in QCEW that do not meet their reporting criteria. The reporting criteria require that all disclosed data must consist of at least three observations and no one observation can account for more than 80% of the total. If these criteria are not met, the specific observation will not be disclosed.

For example, if there is only one firm that is in NAICS classification code 212114 in Richland County, Montana, then the BLS will not report employment or wages for that specific NAICS code in that county. However, the BLS will report totals at higher levels of sectoral and/or geographic aggregation that include the nondisclosed employment and wage values. For example, if there are enough firms in NAICS 212114 in the state of Montana and no firm represents more than 80% of the total, then the BLS will disclose the total employment and wages in NAICS 212114 for the state of Montana and that total will include observations that were undisclosable at lower levels of aggregation.

Likewise, if the disclosure criteria are met in Richland County, Montana, at higher levels of sectoral aggregation, then the BLS will disclose those employment and wage data. For example, while NAICS 212114 may be subject to nondisclosure in Richland County, Montana, there may be enough observations at the NAICS 212 level of aggregation to meet BLS reporting criteria. In that case the BLS can report the employment and wages for the NAICS 212 level of aggregation, which will include the nondisclosed value for NAICS 212114.

The estimation method

The above example demonstrates that there are two dimensions in which control totals can be obtained: sectoral and geographical. Across all geographies sector totals must sum to observed control totals, and across all sectors geographic totals must sum to observed control totals. This is the basis for the hierarchical proportional fitting model that is used to estimate the nondisclosed observations of 6-digit employment and wage QCEW data for every county in the United States. The method was created and scripted by Greg Alward of the Alward Institute

and University of Idaho Professor Philip Watson and provided to Headwaters Economics for integration with the EPS application.

The estimated data are generated through a set of steps and algorithms that start with a standard downloadable file of QCEW data and then fill in missing values for employment and wages that are due to data reporting and nondisclosure criteria on the part of the BLS. The estimation method first determines control totals for all data at the state level 2-digit NAICS aggregation. This is the level of geographic and sectoral aggregation where QCEW discloses all values. These control totals become the constraint to which all lower-level data must conform. For example, all NAICS codes in NAICS 21 (211120, 212114, 212115, etc.) across all the counties in Montana must sum to the observed state-level total for NAICS 21 reported in the QCEW data. The estimation method iteratively adjusts possible values for nondisclosed data to increase their agreement with related values within their industry and geographic hierarchies.

Summary of steps in the estimation method

For each year (2001 to latest available):

- 1) Download full QCEW datafile (“CSVs By Industry” – “Annual Averages”) from the BLS. The data is available for download at <https://www.bls.gov/cew/downloadable-data-files.htm>.
- 2) Fix totals for employment and wage data in geographical and industry hierarchies.
 - a. This ensures that employment and wage data that are disclosed by the BLS will not be changed in the estimation algorithm.
 - i. Note that there are some infeasible totals in the source QCEW dataset. This is due to purposeful random variation (or “noise”) which the BLS adds to the data.
- 3) Calculate total, minimum, and maximum employment and wage values for nondisclosed values.
 - a. Within any hierarchy (geographic or sectoral) there is a total for the disclosed values and a total given at a higher hierarchy. The difference between the sum of the disclosed values and the reported value of the higher hierarchy represents the total value that must be allocated to nondisclosed observations.
 - i. The minimum employment of an industry is its parent industry’s minimum minus the summed maximum of its sibling industries. Its maximum is its parent’s maximum minus the summed minimum of all its siblings. For industries that are not suppressed, the minimum and maximum are both set to be equal to actual disclosed employment.
 - ii. The minimum employment of a nondisclosed industry is the sum of its children’s minimum values, and its maximum employment is the sum of its children’s maximum values.
 - iii. The above two calculations can also be done for the geographical hierarchies (county, state, national).
 - b. All the above relationships are processed iteratively since they are interdependent. Because we are calculating new minimums and maximums and these are then used for calculating the minimum and maximum of other industries and at other levels, these calculations are interdependent and need to be performed repeatedly in multiple iterations. Each iteration uses the largest minimum and the smallest maximum identified in all previous steps. The above calculations are repeated iteratively until no further adjustments can be made.
- 4) Initiate the estimation algorithm

- a. Assign the midpoint of the calculated minimum and maximum as the initial value for each nondisclosed value. Tests using the minimum or zero as the initial estimate resulted in minor differences in the final output, although the process completed faster when using the midpoint.
 - b. Iteratively adjust initial estimates to increase their agreement with related (parent - children) values in the industry and geographic hierarchies. Within each iteration, adjustments are made based on the following comparisons:
 - i. National and the total of all states.
 - ii. State and the total of all counties.
 - iii. Parent and children industries at the national level.
 - iv. Parent and children industries at the state level.
 - v. Parent and children industries at the county level.
 - c. Adjust parent and children values simultaneously. Within each of the comparisons above, adjustments are scheduled and not applied until the end of the step. Since many values can be both a parent and a child depending on the comparison made, there may be multiple adjustments applied to the same value. These adjustments are combined into a single adjustment by taking their average.
 - d. Adjust estimates by calculating the difference between the parent and the sum of the children. Then for each child and the parent, calculate the amount of adjustment possible within the constraints of its minimum and maximum. These adjustments are summed to find the total of all possible adjustments. We then calculate for each child and the parent, the proportion of their possible adjustment to the total of all possible adjustments. Adjustments are made by multiplying this proportion by the adjustment needed.
- 5) Finalize estimate using post-processing adjustments
- a. Adjust estimates to ensure the sum of all 6-digit NAICS values for a given county equals the corresponding 1-digit NAICS value for the same county. These adjustments are only done to 6-digit NAICS values when the corresponding 2-digit NAICS value is not disclosed. Adjustments are done proportionally to the estimated values so that estimated values that are larger get a higher proportion of the needed total adjustment.
 - b. Adjust estimates to ensure the sum of all 6-digit NAICS values equals the owner code zero value at the 1-digit NAICS level within the county data. In QCEW, owner codes differentiate between private and government employment. These adjustments are only made at the 6-digit NAICS level and only if the parent 1-digit NAICS is not disclosed for the specific ownership type. If all 1-digit NAICS are disclosed for the county, we would expect these to add up to the owner code zero value at the 1-digit NAICS level. However, there may be a slight difference because some of the 1-digit NAICS values may have been adjusted so that they fit the total of their children's values for that ownership. To accommodate for this case, when all 1-digit NAICS are disclosed for the county, adjustments are made to all nondisclosed 6-digit NAICS values.

Accuracy

The algorithm performs well in creating a consistent and precise estimate of QCEW employment and wage data.

One measure of accuracy can be assessed by comparing the sum of 6-digit NAICS employment across all counties to the reported total employment in the United States in any given year. The median percent difference between estimated employment (calculated as the sum of all 6-digit NAICS employment across all counties) and the reported employment in the United States is 2.3% at the 2-digit NAICS level. For all 3-digit NAICS codes used in the Economic Profile System, the median percent difference between estimated and reported U.S.

employment is 2.0%. And for all 4-digit NAICS codes used in the Economic Profile System, the median percent difference between estimated and reported U.S. employment is 5.5%. See “Appendix A: Estimation Accuracy” for more detail.

There are two primary reasons the estimation method results cannot be perfectly accurate. First, as mentioned earlier, the BLS intentionally inserts random variation (or “noise”) into the disclosed QCEW estimates.¹ Second, the estimation method is less accurate when there are multiple levels of nondisclosed data within a given geography. This occurs most often in small geographies where for any given 6-digit NAICS code, the closest level for which there is a disclosed value is at the 3-digit level. This skipping of multiple hierarchy levels makes it more difficult to estimate the nondisclosed values in a manner that preserves the geographic and sectoral constraints.

While there is no objective solution to the intentional noise that the BLS inserts into the QCEW data, and this noise interacts with the other sources of error in the estimates, the overall results from the estimation method are highly accurate for most counties.

Contact

Patty Hernandez | 406-599-7425 | patty@headwaterseconomics.org

Scott Story | 406-202-1424 | scott@headwaterseconomics.org

About Headwaters Economics

Headwaters Economics is an independent, nonprofit research group whose mission is to improve community development and land management decisions. <https://headwaterseconomics.org/>

¹ For more information on this topic, see: Yang MY, Pramanik S, Hiles D, Buso M, & Butani SJ. (2011). An evaluation of BLS noise research for the Quarterly Census of Employment and Wages. Washington, DC: U.S. Bureau of Labor Statistics. <https://www.bls.gov/osmr/research-papers/2011/st110120.htm>; and Carpenter CW, VanSandt A, & Loveridge S. (2021). Measurement error in U.S. regional economic data. *Journal of Regional Science*, 62(1), 57-80. <https://onlinelibrary.wiley.com/doi/abs/10.1111/jors.12551>

Appendix A. Estimation Accuracy

Table 1. Accuracy of QCEW employment estimates for 2020 two-digit NAICS codes used in the Economic Profile System.

NAICS Code	Disclosed Total	Estimated Total	Difference	Count	Percent Difference
11 Agriculture, forestry, fishing and hunting	1,247,796	1,243,841	-3,955	30,182	0.32%
21 Mining, quarrying, and oil and gas extraction	540,591	531,545	-9,046	7,363	1.69%
22 Utilities	804,872	900,829	95,957	11,662	11.25%
23 Construction	7,380,421	7,181,322	-199,099	80,716	2.73%
31-33 Manufacturing	12,136,380	11,907,137	-229,243	126,222	1.91%
42 Wholesale trade	5,620,782	5,102,113	-518,669	70,424	9.67%
44-45 Retail trade	14,818,498	14,610,016	-208,482	109,648	1.42%
48-49 Transportation and warehousing	6,498,364	6,593,736	95,372	44,431	1.46%
51 Information	2,844,966	2,736,150	-108,816	26,367	3.9%
52 Finance and insurance	6,079,981	5,827,088	-252,893	39,166	4.25%
53 Real estate and rental and leasing	2,194,966	2,172,692	-22,274	31,255	1.02%
54 Professional and technical services	9,569,702	8,955,044	-614,658	65,723	6.64%
55 Management of companies and enterprises	2,318,008	2,230,033	-87,975	3,825	3.87%
56 Administrative and waste services	8,628,861	8,102,767	-526,094	50,613	6.29%
61 Educational services	12,083,284	11,366,211	-717,073	21,956	6.12%
62 Health care and social assistance	21,544,346	21,623,919	79,573	71,464	0.37%
71 Arts, entertainment, and recreation	2,019,105	2,121,073	101,968	25,341	4.93%
72 Accommodation and food services	11,170,329	11,205,431	35,102	25,484	0.31%
81 Other services, except public administration	3,984,430	3,994,231	9,801	66,671	0.25%
92 Public administration	7,480,297	7,337,825	-142,472	56,810	1.92%

Table 2. Accuracy of QCEW employment estimates for 2020 three-digit NAICS codes used in the Economic Profile System.

NAICS Code	Disclosed Total	Estimated Total	Difference	Count	Percent Difference
111 Crop production	540,971	540,697	-274	12,242	0.05%
112 Animal production and aquaculture	267,136	268,385	1,249	8,162	0.47%
113 Forestry and logging	55,817	53,261	-2,556	2,208	4.69%
114 Fishing, hunting and trapping	8,142	8,612	470	790	5.61%
115 Agriculture and forestry support activities	375,658	372,886	-2,772	6,780	0.74%
211 Oil and gas extraction	126,045	120,393	-5,652	1,176	4.59%
212 Mining, except oil and gas	176,114	184,141	8,027	3,793	4.46%
213 Support activities for mining	238,153	227,011	-11,142	2,394	4.79%
221 Utilities	804,872	900,829	95,957	11,662	11.25%
236 Construction of buildings	1,597,171	1,572,050	-25,121	11,656	1.59%
237 Heavy and civil engineering construction	1,209,325	1,162,286	-47,039	12,930	3.97%
238 Specialty trade contractors	4,573,663	4,446,986	-126,677	56,130	2.81%
311 Food manufacturing	1,600,921	1,555,050	-45,871	14,040	2.91%
312 Beverage and tobacco product manufacturing	273,578	269,196	-4,382	4,106	1.61%
313 Textile mills	95,124	92,158	-2,966	1,433	3.17%
314 Textile product mills	100,592	98,960	-1,632	2,486	1.64%
315 Apparel manufacturing	88,544	87,500	-1,044	1,779	1.19%
316 Leather and allied product manufacturing	24,161	23,730	-431	819	1.80%
321 Wood product manufacturing	395,098	377,710	-17,388	7,407	4.50%
322 Paper manufacturing	352,086	342,558	-9,528	2,597	2.74%
323 Printing and related support activities	377,701	373,518	-4,183	4,042	1.11%

NAICS Code	Disclosed Total	Estimated Total	Difference	Count	Percent Difference
324 Petroleum and coal products manufacturing	108,893	107,500	-1,393	1,327	1.29%
325 Chemical manufacturing	844,162	829,369	-14,793	8,264	1.77%
326 Plastics and rubber products manufacturing	692,365	679,644	-12,721	5,498	1.85%
327 Nonmetallic mineral product manufacturing	396,955	388,094	-8,861	7,597	2.26%
331 Primary metal manufacturing	350,578	343,372	-7,206	3,378	2.08%
332 Fabricated metal product manufacturing	1,374,820	1,351,856	-22,964	16,940	1.68%
333 Machinery manufacturing	1,045,080	1,024,465	-20,615	12,898	1.99%
334 Computer and electronic product manufacturing	1,058,381	1,052,779	-5,602	6,637	0.53%
335 Electrical equipment and appliance mfg.	380,807	374,932	-5,875	4,050	1.55%
336 Transportation equipment manufacturing	1,634,724	1,612,484	-22,240	7,228	1.37%
337 Furniture and related product manufacturing	359,625	353,889	-5,736	5,577	1.61%
339 Miscellaneous manufacturing	581,230	568,373	-12,857	8,119	2.24%
423 Merchant wholesalers, durable goods	3,080,096	2,877,457	-202,639	38,934	6.80%
424 Merchant wholesalers, nondurable goods	2,049,171	1,883,581	-165,590	28,350	8.42%
425 Electronic markets and agents and brokers	491,418	341,075	-150,343	3,140	36.12%
441 Motor vehicle and parts dealers	1,887,914	1,868,110	-19,804	13,893	1.05%
442 Furniture and home furnishings stores	408,131	403,827	-4,304	5,976	1.06%
443 Electronics and appliance stores	435,949	422,501	-13,448	3,662	3.13%
444 Building material and garden supply stores	1,354,864	1,333,676	-21,188	12,478	1.58%
445 Food and beverage stores	3,112,184	3,096,224	-15,960	13,593	0.51%
446 Health and personal care stores	973,584	953,506	-20,078	8,246	2.08%
447 Gasoline stations	921,746	902,054	-19,692	5,147	2.16%

NAICS Code	Disclosed Total	Estimated Total	Difference	Count	Percent Difference
448 Clothing and clothing accessories stores	967,043	954,068	-12,975	11,470	1.35%
451 Sports, hobby, music instrument, book stores	471,652	466,892	-4,760	7,320	1.01%
452 General merchandise stores	2,964,233	2,923,898	-40,335	6,555	1.37%
453 Miscellaneous store retailers	736,407	727,198	-9,209	14,675	1.26%
454 Nonstore retailers	584,634	558,062	-26,572	6,633	4.65%
481 Air transportation	462,510	458,715	-3,795	1,775	0.82%
482 Rail transportation	409	3,207	2,798	93	154.76%
483 Water transportation	60,816	58,397	-2,419	844	4.06%
484 Truck transportation	1,471,069	1,415,395	-55,674	13,057	3.86%
485 Transit and ground passenger transportation	615,326	740,462	125,136	6,481	18.46%
486 Pipeline transportation	51,818	51,223	-595	1,558	1.15%
487 Scenic and sightseeing transportation	20,065	26,911	6,846	929	29.15%
488 Support activities for transportation	759,540	807,620	48,080	9,862	6.14%
491 Postal service	609,679	615,314	5,635	3,670	0.92%
492 Couriers and messengers	966,384	956,368	-10,016	2,564	1.04%
493 Warehousing and storage	1,480,102	1,460,124	-19,978	3,598	1.36%
511 Publishing industries, except internet	765,032	724,086	-40,946	6,506	5.50%
512 Motion picture and sound recording industries	295,637	279,543	-16,094	4,512	5.60%
515 Broadcasting, except internet	245,755	248,791	3,036	3,277	1.23%
517 Telecommunications	689,899	657,591	-32,308	5,832	4.80%
518 Data processing, hosting and related services	366,475	334,849	-31,626	1,573	9.02%
519 Other information services	482,002	491,290	9,288	4,667	1.91%

NAICS Code	Disclosed Total	Estimated Total	Difference	Count	Percent Difference
521 Monetary authorities - central bank	20,306	20,271	-35	159	0.17%
522 Credit intermediation and related activities	2,669,104	2,600,464	-68,640	16,475	2.61%
523 Securities, commodity contracts, investments	947,198	931,357	-15,841	9,089	1.69%
524 Insurance carriers and related activities	2,423,831	2,257,855	-165,976	11,746	7.09%
525 Funds, trusts, and other financial vehicles	18,527	17,141	-1,386	1,697	7.77%
531 Real estate	1,690,084	1,682,331	-7,753	18,355	0.46%
532 Rental and leasing services	483,777	470,707	-13,070	12,262	2.74%
533 Lessors of nonfinancial intangible assets	21,106	19,654	-1,452	638	7.12%
541 Professional and technical services	9,569,702	8,955,044	-614,658	65,723	6.64%
551 Management of companies and enterprises	2,318,008	2,230,033	-87,975	3,825	3.87%
561 Administrative and support services	8,160,955	7,611,634	-549,321	40,466	6.97%
562 Waste management and remediation services	467,906	491,133	23,227	10,147	4.84%
611 Educational services	12,083,284	11,366,211	-717,073	21,956	6.12%
621 Ambulatory health care services	7,611,549	7,552,045	-59,504	36,218	0.78%
622 Hospitals	6,479,857	6,610,875	131,018	4,462	2.00%
623 Nursing and residential care facilities	3,373,398	3,356,941	-16,457	11,585	0.49%
624 Social assistance	4,079,543	4,104,058	24,515	19,199	0.60%
711 Performing arts and spectator sports	339,633	352,266	12,633	7,905	3.65%
712 Museums, historical sites, zoos, and parks	202,977	241,037	38,060	4,126	17.14%
713 Amusements, gambling, and recreation	1,476,496	1,527,770	51,274	13,310	3.41%
721 Accommodation	1,480,149	1,556,607	76,458	8,112	5.04%
722 Food services and drinking places	9,689,609	9,648,824	-40,785	17,372	0.42%

NAICS Code	Disclosed Total	Estimated Total	Difference	Count	Percent Difference
811 Repair and maintenance	1,281,526	1,263,798	-17,728	26,290	1.39%
812 Personal and laundry services	1,240,154	1,247,988	7,834	20,331	0.63%
813 Membership associations and organizations	1,245,604	1,269,469	23,865	17,287	1.90%
814 Private households	217,146	212,976	-4,170	2,763	1.94%
921 Executive, legislative and general government	3,005,357	2,933,275	-72,082	9,839	2.43%
922 Justice, public order, and safety activities	1,915,691	1,869,089	-46,602	14,312	2.46%
923 Administration of human resource programs	816,860	791,390	-25,470	7,175	3.17%
924 Administration of environmental programs	330,187	330,205	18	8,281	0.01%
925 Community and housing program administration	98,393	99,356	963	2,273	0.97%
926 Administration of economic programs	681,479	674,877	-6,602	13,213	0.97%
927 Space research and technology	17,536	17,436	-100	40	0.57%
928 National security and international affairs	614,689	622,197	7,508	1,677	1.21%

Table 3. Accuracy of QCEW employment estimates for 2020 four-digit NAICS codes used in the Economic Profile System.

NAICS Code	Disclosed Total	Estimated Total	Difference	Count	Percent Difference
1153 Support activities for forestry	18,070	18,700	630	1,091	3.43%
2121 Coal mining	40,109	37,242	-2,867	296	7.41%
2122 Metal ore mining	40,870	35,542	-5,328	204	13.95%
2123 Nonmetallic mineral mining and quarrying	95,135	111,357	16,222	3,293	15.71%
3211 Sawmills and wood preservation	88,655	83,506	-5,149	1,658	5.98%
3212 Plywood and engineered wood product mfg.	79,384	75,291	-4,093	1,162	5.29%
3219 Other wood product manufacturing	226,992	218,913	-8,079	4,587	3.62%
3221 Pulp, paper, and paperboard mills	91,952	87,044	-4,908	509	5.48%
3222 Converted paper product manufacturing	260,134	255,514	-4,620	2,088	1.79%