Occupational Employment and Wage Statistics May 2024 Survey Methods

Concepts

The Occupational Employment and Wage Statistics (OEWS) program is the only comprehensive source of regularly produced occupational employment and wage rate information for the U.S. economy. The scope includes the 50 states, the District of Columbia, Guam, Puerto Rico, the U.S. Virgin Islands, and metropolitan and nonmetropolitan areas covering the entire United States.

The following are definitions of the key concepts used in the OEWS program.

Unit of observation

The OEWS survey measures occupational employment and wage rates for wage and salary workers in nonfarm establishments in the United States. An *establishment* is generally a single physical location at which economic activity occurs (e.g., store, factory, restaurant, etc.). When a single physical location encompasses two or more distinct economic activities, it is treated as two or more separate establishments if separate payroll records are available and certain other criteria are met.

Classification systems

The OEWS survey uses the Office of Management and Budget's Standard Occupational Classification (SOC) system to classify jobs into occupational categories and the North American Industry Classification System (NAICS) to classify establishments into industries.

The occupational coding system

The SOC system is used by federal statistical agencies to classify workers and jobs into occupational categories for the purpose of collecting, calculating, analyzing, or disseminating data. Jobs are assigned to an occupation based on the work the employee performs and not on their education or training. For example, an employee trained as an engineer but working as a drafter is reported as a drafter. Employees who perform the duties of two or more occupations are reported in the occupation that requires the highest level of skill or in the occupation where the most time is spent if there is no measurable difference in skill requirements. Working supervisors (those spending 20 percent or more of their time doing work similar to that of the workers they supervise) are classified with the workers they supervise. Workers receiving on-the-job training, apprentices, and trainees are classified with the occupations for which they are being trained. For more information about the SOC system, please see the BLS SOC webpage.

Since May 2021, OEWS data are based on the 2018 SOC. OEWS publishes data for most 2018 SOC detailed occupations. For a few occupations, OEWS publishes data only to the SOC broad occupation level or OEWS-specific combinations of detailed occupations. Occupations are aggregated to improve data quality in cases where it is difficult to collect the information needed to code jobs accurately to the individual detailed occupations within the aggregation.

The industry coding system

The NAICS is used throughout the federal government to group business establishments into industries based on the primary goods or services they produce. The NAICS has a hierarchical structure with several levels of industry detail: by broad industrial sectors (2-digit NAICS levels), subsectors (3-digit NAICS levels), industry groups (4-digit NAICS levels), and NAICS industries (5- and 6-digit NAICS levels). For more information about NAICS, see the BLS NAICS webpage.

Since May 2022, OEWS data are based on the 2022 NAICS. OEWS publishes national industry-specific data to the 4-digit NAICS level of detail for most industries. For some industries, OEWS publishes data only to the 3-digit subsector level or for OEWS-specific combinations of 4-digit industry groups. Data at the 5- and 6-digit NAICS level are available for selected industries only.

Ownership

OEWS classifies most government-owned establishments differently from the standard NAICS system. The standard NAICS classifies government establishments based on their primary activity, not ownership. Under the standard NAICS, government establishments that oversee programs and activities not generally performed by private-sector establishments are classified in NAICS sector 92, the code for the public administration sector. Government establishments producing goods and services that can also be provided by private-sector establishments are classified in the same industry as private-sector establishments engaged in similar activities.

The OEWS program classifies most government establishments based on ownership rather than primary activity and therefore does not use NAICS sector 92. Instead, the OEWS survey produces occupational employment and wage estimates at the federal, state, and local government levels and denotes them with industry codes 9991, 9992, and 9993, respectively.

The OEWS industry-specific data for state and local government (NAICS 9992 and 9993) consist of all state and local government establishments, except schools, hospitals, and local government gambling establishments and casino hotels. State and local government schools and hospitals and local government gambling establishments and casino hotels are classified in their respective NAICS industries, along with similar private sector establishments. Estimates for schools and hospitals are available for private, state, and local government ownerships combined, as well as by individual ownership types. State and local government data that include schools, hospitals, and local government gambling establishments and casino hotels are also available as part of the OEWS cross-industry ownership estimates.

Within federal government, the OEWS survey covers the federal executive branch, U.S. Postal Service (USPS), and the Tennessee Valley Authority (TVA) only. The military, the federal legislative, and judicial branches are not included. OEWS industry-specific data for federal government (NAICS 9991) consist of the federal executive branch and TVA; in the industry-specific estimates, USPS is classified in NAICS 491100 - Postal Service. Data for the federal executive branch, TVA, and USPS combined are also available in the cross-industry ownership estimates.

Area definitions

The OEWS program uses the metropolitan statistical area (MSA) definitions provided by the Office of Management and Budget (OMB). The May 2024 OEWS estimates use the metropolitan area definitions delineated in OMB Bulletin 23-01. Nonmetropolitan areas are specific to the OEWS program and are set with guidance from our state program offices. The nonmetropolitan areas cover all counties that are not part of an OMB-defined metropolitan area. For detailed information on the current OEWS metropolitan and nonmetropolitan area definitions, see the downloadable Excel file of OEWS area definitions.

Key concepts and definitions

Employment represents the estimated number of full- and part-time jobs in an occupation. The OEWS survey covers full- and part-time wage and salary employees in nonfarm industries, including employees on paid vacations or other types of paid leave; salaried officers, executives, and staff members of incorporated firms; employees temporarily assigned to other units; and noncontract employees for whom the reporting unit is their permanent duty station, regardless of whether that unit prepares their paychecks.

Self-employed workers, owners and partners in unincorporated firms, employees of private households, and unpaid family workers are not covered by the OEWS survey.

Wages are money that is paid or received for work or services performed in a specified period. Wages for the OEWS survey are straight-time, gross pay, excluding premium pay, such as overtime. Base rate pay; cost-of-living allowances; guaranteed pay; hazardous-duty pay; incentive pay, including commissions and production bonuses; and tips are included. Excluded are back pay, jury duty pay, overtime pay, severance pay, shift differentials, nonproduction bonuses, employer cost for supplementary benefits, and tuition reimbursements.

OEWS obtains two types of wage data. OEWS receives exact wage rates for federal government, USPS, TVA, and most employees in state government, local government, and private sector establishments. For a small percentage of records for which exact wage rates are not available, the wage data are processed in 12 wage intervals. The intervals are defined both as hourly rates and the corresponding annual rates, where the annual rate for an occupation is calculated by multiplying the hourly wage rate by a typical work year of 2,080 hours. Wage intervals are updated periodically based on the wages in the labor market. The current OEWS wage intervals are shown in table 1.

Table 1: Wage intervals for the November 2021–May 2024 survey panels

Interval	Wages	
	Hourly wages	Annual wages
Range A	Under $9.25	Under $19,240
Range B	$9.25–$11.99	$19,240–$24,959
Range C	$12.00–$15.49	$24,960–$32,239
Range D	$15.50–$19.74	$32,240–$41,079
Range E	$19.75–$25.49	$41,080–$53,039
Range F	$25.50–$32.74	$53,040–$68,119
Range G	$32.75–$41.99	$68,120–$87,359
Range H	$42.00–$53.99	$87,360–$112,319
Range I	$54.00–$69.49	$112,320–$144,559
Range J	$69.50–$89.49	$144,560–$186,159
Range K	$89.50–$114.99	$186,160–$239,199
Range L	$115.00 and over	$239,200 and over

Responding establishments are instructed to report the hourly rate for part-time workers and to report either hourly rates or annual salaries for full-time workers, depending on how that worker is paid.

For most occupations, the OEWS program publishes both hourly and annual wage estimates, using a standard work year of 2,080 hours to convert between hourly and annual wage data. Workers in some occupations, such as teachers, pilots, and flight attendants, are typically paid at an annual rate but do not work 2,080 hours per year. For these occupations, OEWS collects and publishes only annual wages. Other occupations, such as actors or musicians and singers, are paid hourly rates but generally do not work 40 hours per week, year-round. For these occupations, OEWS collects and publishes only hourly wages.

Scope and exclusions

The OEWS survey measures employment and wages by occupation for wage and salary employees in nonfarm establishments. The survey excludes most of the agricultural sector, private household employers, and the self-employed.

Scope of the survey

The OEWS program annually publishes employment and wage estimates by occupation for wage and salary jobs in nonfarm establishments in the United States. The OEWS data available from the Bureau of Labor Statistics (BLS) include cross-industry occupational employment and wage estimates for the nation; states, the District of Columbia, and Guam, Puerto Rico, and the Virgin Islands; and approximately 530 metropolitan statistical areas (MSAs) and nonmetropolitan areas. BLS also publishes national industry-specific estimates at the NAICS sector, 3-digit, most 4-digit, and selected 5- and 6-digit industry levels; and national estimates by ownership across all industries and for schools and hospitals.

Occupation exclusions

OEWS produces data for approximately 830 occupational categories based on the Office of Management and Budget's 2018 Standard Occupational Classification (SOC) system.
These occupational categories make up 22 of the 23 SOC major occupational groups.
Major group 55, which refers to Military Specific Occupations, is not included.

Industry coverage and exclusions

The OEWS survey excludes the majority of the agricultural sector, with the exception of logging (NAICS 113310), support activities for crop production (NAICS 1151), and support activities for animal production (NAICS 1152). Private households (NAICS 814) also are excluded. OEWS federal government data include the federal executive branch, U.S. Postal Service, and TVA only. All other industries, including state and local government, are covered by the survey. Industries that fall within the OEWS scope are shown in table 2.

Table 2: NAICS industry sectors covered in OEWS

Industry code	Industry title
11	Logging (1133), support activities for crop production (1151), and support activities for animal production (1152) only
21	Mining, quarrying, and oil and gas extraction
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale trade
44-45	Retail trade
48-49	Transportation and warehousing
51	Information
52	Finance and insurance
53	Real estate and rental and leasing
54	Professional, scientific, and technical services
55	Management of companies and enterprises
56	Administrative and support and waste management and remediation services
61	Educational services
62	Healthcare and social assistance
71	Arts, entertainment, and recreation
72	Accommodation and food services
81	Other services, except public administration [private households (814) are excluded]
999100*	Federal government executive branch
999200*	State government, excluding schools and hospitals
999300*	Local government, excluding schools, hospitals, gambling establishments, and casino hotels

Note: “*” indicates an OEWS-defined industry code that is not part of the NAICS industry classification.

Data Sources

Data collection

The OEWS survey is a cooperative effort between the Bureau of Labor Statistics (BLS) and the state workforce agencies (SWAs). BLS funds the survey and provides the procedures and technical support, while the SWAs collect most of the data. OEWS estimates are constructed from a probability sample of about 1.1 million establishments. Each year, 2 semiannual panels of approximately 186,000 to 189,000 sampled establishments are contacted, one panel in May and the other in November. Responses are obtained online or by mail, email, telephone, or personal visit. For a given panel, most sampled establishments initially receive a letter or email with instructions for reporting their data electronically. At approximately 4-week intervals, nonrespondents receive up to three additional mailings of a survey questionnaire or letter with instructions for reporting electronically. Nonrespondents may also be contacted by phone or email. For more information about OEWS data collection, see the OEWS survey respondents webpage.

Confidentiality

BLS has a strict confidentiality policy that ensures the survey sample composition, lists of reporters, and names of respondents will be kept confidential. Additionally, the policy assures respondents that published figures will not reveal the identity of any specific respondent and will not allow the data of any specific respondent to be inferred. The most relevant statute that governs BLS confidentiality is the Confidential Information Protection and Statistical Efficiency Act (CIPSEA). Each published estimate is screened to ensure that it meets these confidentiality requirements. To further protect the confidentiality of the data, the specific screening criteria are not listed in this publication. For additional information, please visit the BLS confidentiality pledge and laws webpage.

Quality control

The OEWS survey is a federal–state cooperative effort that enables states to conduct their own surveys. A major concern in a cooperative program such as OEWS is accommodating the needs of BLS and other federal agencies, as well as state-specific publication needs, with limited resources while simultaneously standardizing survey procedures across all 50 states, the District of Columbia, Guam, Puerto Rico, and the Virgin Islands. Controlling sources of nonsampling error in this decentralized environment can be difficult. One important quality control tool used by the OEWS survey is a computerized survey data management system, which was developed to provide a consistent and automated framework for survey processing.

To ensure standardized sampling methods in all areas, the sample is drawn in the BLS national office. Standardizing data processing activities, such as validating the sampling frame, allocating and selecting the sample, refining mailing addresses, addressing envelopes and mailers, editing and updating questionnaires, conducting electronic review, producing management reports, and calculating estimates, have resulted in the overall standardization of the OEWS survey methodology. This has reduced the number of errors in the data files as well as the time needed to review them.

Several editing and quality control procedures are used to reduce nonsampling error:

- Follow-up email, mail, and telephone solicitation of nonrespondents, especially critical or large nonrespondents
- Review of collected data to verify its accuracy and reasonableness
- Adjustments for atypical reporting units on the data file
- Validation of unit matching and donor profiles
- Quality review of estimates before publication

Additional data sources

Although most data are collected through the process outlined above, additional data sources are used for both the collection and processing of the data.

Data collected through a census

A census of the executive branch of the federal government and the U.S. Postal Service (USPS) is collected annually from the U.S. Office of Personnel Management (OPM), the Tennessee Valley Authority, and USPS. Data from only the most recent year are retained for use in OEWS estimates.

A census of state government establishments, except for schools and hospitals, is collected annually every November. Data from only the most recent year are retained for use in OEWS estimates.

A census of Hawaii’s local government is conducted each November. With the exception of schools and hospitals, all local-government-owned establishments in Hawaii are included. A census of public- and private-owned hospitals is taken over a 3-year period.

BLS data sources

Data from the BLS Quarterly Census of Employment and Wages (QCEW), the database of businesses reporting to the state unemployment insurance programs, are used to adjust the employment estimates to represent the entire population covered by the OEWS survey. Population employment for each in-scope establishment—including establishments that receive modeled data—is set to equal the average of its May and November QCEW employment for the two most recent survey panels used in the estimates.

Design

The OEWS survey is based on a probability sample drawn from a universe of about 8.7 million in-scope business establishments stratified by geography, industry, size, and ownership. The sample is designed to represent all nonfarm establishments in the United States.

The full OEWS survey sample is collected over a 6-panel (or 3-year) cycle in order to provide adequate geographic, industry, and occupational coverage. Each year, the OEWS program collects 2 samples of survey data, each consisting of approximately 186,000 to 189,000 establishments. These semiannual samples are referred to as “panels.” Data are collected semiannually to help reduce seasonal effects. Respondents are asked to provide data as of their payroll that includes May 12 or November 12, depending on the panel in which they are sampled.

Over the course of a full 6-panel cycle, approximately 1.1 million establishments are sampled. For example, data collected in May 2024 were combined with data collected in November 2023, May 2023, November 2022, May 2022, and November 2021 to produce the May 2024 OEWS estimates, for a total sample size of approximately 1.1 million units.

Of the approximately 1.1 million establishments in the 50 states and the District of Columbia in the May 2024 combined initial sample, approximately 1,063,000 were viable establishments (that is, establishments that are not outside the scope or out of business). Of the viable establishments, approximately 698,000 responded and 365,000 did not, yielding a 65.7-percent response rate. The response rate in terms of weighted sample employment is 65.9 percent.

A probability sample is taken of private sector establishments, local government establishments, and state government schools and hospitals. OEWS receives an annual census of employees in the federal executive branch, U.S. Postal Service (USPS), Tennessee Valley Authority (TVA), and state government establishments (excluding state government schools and hospitals). A census of Hawaii’s local government establishments, excluding schools and hospitals, is also conducted each November. For these establishments for which census data are received, only the most recent census is used in each year’s estimates. Units from older panels are deleted to avoid double counting.

Frame construction

The sampling frame, or universe, is a list of about 8.7 million in-scope nonfarm establishments that file unemployment insurance (UI) reports to the state workforce agencies. Employers are required by law to file these reports to the state where each establishment is located. Every quarter, the U.S. Bureau of Labor Statistics (BLS) creates a national sampling frame by combining the administrative lists of unemployment insurance reports from all of the states into a single database called the Quarterly Census of Employment and Wages (QCEW). Every 6 months, OEWS extracts the administrative data of establishments that are in scope for the OEWS survey from the most current QCEW. QCEW files were supplemented with frame files covering establishments in Guam and the rail transportation industry (NAICS 4821) because these are outside the UI program’s scope.

Construction of the sampling frame includes a process in which establishments that are linked together into multiunit companies are assigned to either the May or November sample. This prevents BLS from contacting multiunit companies more than once per year for this survey. Furthermore, the frame is matched to the five prior sample panels, and units that have been previously selected in the five prior panels are marked as ineligible for sampling for the current panel.

Stratification

Establishments in the sampling frame are stratified by geographic area, industry group, ownership, and size. Stratification is done at the state; metropolitan or nonmetropolitan area; 3-, 4-, 5-, or 6-digit NAICS; and ownership level.

Geography

There are over 580 metropolitan statistical areas (MSAs) and nonmetropolitan or balance-of-state (BOS) areas specified. MSAs are defined and mandated by the Office of Management and Budget. Each officially defined metropolitan area within a state is specified as a substate area. MSAs that cross state borders have a separate portion for each state contributing to that MSA. In addition, states may have up to six residual nonmetropolitan areas that together cover the remaining non-MSA portion of their state.

Industry

There are about 300 industry groups defined at the NAICS 3-, 4-, 5-, or 6-digit level.

Ownership

Schools and hospitals are stratified by state government, local government, or private ownership. Local government casinos and gambling establishments are sampled separately from the rest of local government.

Size

Sampled establishments are separated into two types based on employment size: certainty units and noncertainty units. Large employers are selected into the sample with certainty over the 3-year sample cycle. A probability sample is taken of smaller establishments with employment below the certainty size cutoff.

At any given time, there are about 145,000 nonempty strata on the frame. When comparing nonempty strata between frames, there may be substantial frame-to-frame differences. The differences are primarily due to normal establishment birth and death processes and normal establishment growth and shrinkage. Other differences are caused by changes in establishments' NAICS classifications or geographic locations.

A small number of establishments provide the state in which their employees are located, but do not provide the specific county in which they are located. These establishments are also sampled and used in the calculation of the statewide and national estimates. They are not included in the estimates of any substate area. Therefore, the sum of the employment in the MSAs and nonmetropolitan areas within a state may be less than the statewide employment.

Sample size

The combined 6-panel sample allocations, excluding all federal government census data, for the May 2024 estimates are as follows:

- 188,236 establishments for May 2024
- 186,410 establishments for November 2023
- 188,045 establishments for May 2023
- 186,064 establishments for November 2022
- 186,911 establishments for May 2022
- 187,215 establishments for November 2021

The May 2024 data include a census of about 6,300 federal executive branch, TVA, and USPS units. The combined sample size for the May 2024 estimates is approximately 1.1 million establishments, which includes only the most recent data for federal and state government.

Allocation methods

The sampling frame is stratified into approximately 145,000 nonempty strata. Each time a sample is selected, a 6-panel allocation of the 1.1 million sample units among these strata is performed.

The largest establishments are removed from the allocation because they will be selected with certainty once during the 6-panel cycle. Once the certainty units have been removed, the probability sample of noncertainty units is allocated across strata. For the remaining nonempty strata, a set of minimum sample size requirements based on the number of establishments in each cell is used to ensure sufficient coverage by industry and geographic area. For each stratum, a sample allocation is also calculated using a power Neyman allocation.[1] The actual 6-panel sample allocation is the larger of the minimum sample allocation and the power Neyman allocation. To determine the current single panel allocation, the 6-panel allocation is divided by 6, and the resulting quotient is randomly rounded.

Two factors influence the power Neyman allocation. The first is the square root of the employment size of each stratum. With a Neyman allocation, strata with higher levels of employment generally are allocated more sample units than strata with lower levels of employment. Using the square root within the Neyman allocation softens this effect. The second factor is a measure of the occupational variability of the industry based on prior OEWS survey data. The occupational variability of an industry is measured by computing the coefficient of variation (CV) for each occupation (excluding the least common occupations in each industry), averaging those CVs, and then calculating the standard error from that average CV. Using this measure, industries that tend to have greater occupational variability will get more sample units than industries that are more occupationally homogeneous.

---

[1] The power Neyman allocation is a statistical method of balancing the efficiency of the overall estimate with the efficiency of subnational estimates. For more information, see Michael D. Bankier, “Power allocations: Determining sample sizes for subnational areas,” *The American Statistician*, vol. 42, no. 3 (August, 1988), pp. 174–177.

Sample selection

To provide the most occupational coverage, sample selection within strata is approximately proportional to size. Some of the largest establishments are selected with certainty. Among the noncertainty units, establishments with higher employment are more likely to be selected than those with lower employment. The unweighted employment of sampled establishments makes up approximately 55 percent of total employment in May 2024.

Permanent random numbers (PRNs) are used in the sample selection process. To minimize sample overlap between the OEWS survey and other large surveys conducted by BLS, each establishment is assigned a PRN. For each stratum, a specific PRN value is designated as the “starting” point to select a sample. From this starting point, we sequentially select the first ‘*n’* eligible establishments in the frame into the sample, where *n* denotes the number of establishments to be sampled.

Sampling weights

Sampling weights are computed so that each panel will roughly represent the entire universe of establishments.

Federal government, USPS, TVA, and state government units are assigned a sample weight of 1. Other sampled establishments are assigned a design-based sample weight, which reflects the inverse of the probability of selection.

Calculation

The OEWS program uses a model-based estimation method called "MB3" to produce occupational employment and wage estimates from the collected OEWS survey data. Each set of estimates is produced by combining data from six semiannual survey panels collected over a 3-year period, for a total sample of approximately 1.1 million establishments. To produce the May 2024 estimates, data collected in the May 2024 survey panel were combined with data collected in the November 2023, May 2023, November 2022, May 2022, and November 2021 survey panels. Federal and state government data are collected by annual census, with only the most recent year of data used in the estimates.

All establishments in the population covered by the OEWS survey are represented in the estimates, either by their reported survey data or by data modeled from similar responding establishments. Establishments that responded to the survey and met certain stability criteria are represented in the estimates by their reported survey data, while all other in-scope establishments in the population receive modeled data. These may be sampled units that responded but did not meet stability criteria, sampled units that did not respond, or in-scope establishments that were not sampled to participate in the OEWS survey. For the occupational employment estimates to sum to total population employment, each in-scope establishment's employment is set to the average of its May and November Quarterly Census of Employment and Wages (QCEW) employment for the two most recent survey panels used in the estimates.

MB3 estimation methodology

Under MB3, occupational employment and wage estimates are calculated directly from a population containing both establishments with response data and establishments with modeled data. For most industries, the OEWS population consists of in-scope establishments present in QCEW as of the survey reference period. Each establishment in the population is defined by characteristics including industry, size, ownership, and location, which are known to be strong predictors of occupational employment and wages. OEWS survey response data provide occupational employment distributions, known as staffing patterns, and wage information for a portion of the population. For the remaining establishments in the population, OEWS response data from the current panel and five previous panels are used to predict staffing patterns and wages.

Matching population units to respondent data

The prediction framework splits establishments in the population into two categories: "observed units" and "unobserved units." Observed units are stable establishments with response data from the previous three years. Stability is determined by comparing the values of several variables reported to the OEWS survey with recent QCEW values for the same establishment, as described in the following section. Observed units are represented in the estimates by their reported survey data.

Unobserved units may be nonsampled units, nonresponding units, or responding units that do not meet stability criteria. For any given unobserved unit, occupational employment and wages are predicted using modeled data from similar responding establishments. Responding units that do not satisfy stability criteria can still be used as donors for unobserved units.

Direct matching—observed units

The stability criteria for observed units require that the 6-digit NAICS industry, ownership, and metropolitan or nonmetropolitan area reported to the OEWS survey exactly match the establishment's QCEW values for the May reference panel. In addition, the establishment's reported employment must be similar to its population employment, defined as the average of the establishment's most recent May and November QCEW employment for a given survey reference period. An establishment's employment is considered stable if its reported employment is within 50 percent or 5 jobs of its population employment. That is, for population employment E_P and respondent employment E_R, the establishment fulfills either condition:

$$\frac{|E_R - E_P|}{E_P} < 0.5 \quad \text{or} \quad |E_R - E_P| < 5$$

During data collection, states can correct the QCEW NAICS, metropolitan and nonmetropolitan area, and previous employment values if they are incorrect. These corrections are applied to the population file prior to the stability calculations so that incorrect population data do not automatically prevent a unit from being classified as stable.

For observed units, the unit's reported staffing pattern is scaled up or down to match the unit's population employment, and wages collected in earlier survey panels are adjusted to reflect wage levels as of the reference date. A small percentage of respondents provide complete staffing patterns, but do not provide complete wage data. For these partial respondents, missing wage data are imputed as described in the "Imputing nonrespondents for wage modeling" section below, and the units are then treated as respondents. They will be tested for stability, and if stable, will be used as observed units. They can also be used as donors for unobserved units.

Prediction—unobserved units

The staffing patterns and wages of unobserved units in the population are predicted using data from nearest neighbor respondents. Responding units that do not pass stability criteria are not representative of the population cell for which they were sampled, but they may be used to predict units in the cell represented by the characteristics they reported to the OEWS survey. A pool of 10 nearest neighbor responding units is typically used to predict each unobserved unit. Unobserved units with identical characteristics are predicted as a group and receive the same donors, so the predicted staffing pattern and wages of any unit of a given size, location, ownership group, and industry will be identical.

Donor Scoring

Potential donors are assigned a series of five scores that measure the similarity of the donor to the unobserved unit in terms of characteristics like industry and geographic location. Each score can take on values between 0 to 1, with 1 indicating a perfect match between donor and recipient. The five individual scores are then multiplied to produce a single overall score for the potential donor. For a given unobserved unit, a set of (typically) 10 responding units with the highest scores is used for the prediction.[2]

---

[2] There is not technically an upper bound on the number of donors. If several donors have the same score, they will all be used, which may put the total number of donors at more than 10. For example, if there are 15 donors with the 10th highest score, all will be used, resulting in a total of 24 donors. For the

The scoring function for each predictive factor aims to assign a score value based on the relative importance of that factor. Industry and establishment size as measured by total employment are the strongest predictors of staffing patterns. Therefore, differences in either of these characteristics result in large reductions in match scores. Time and location are also important predictors, but differences in either of these dimensions result in relatively smaller score penalties than are given for industry, size, and ownership differences. The specific score values used in the MB3 system were evaluated using simulation studies. Various proposed scoring functions were tested to generate estimates and the best performing of these were used.

Where establishment *a* is an unobserved unit and establishment *b* is a potential donor, each component of the score function accounts for differences between *a* and *b*. The overall match score of the potential donor is:

$$\mathrm{S}(a,b) = \mathrm{S_E}(a,b) \cdot \mathrm{S_T}(a,b) \cdot \mathrm{S_I}(a,b) \cdot \mathrm{S_O}(a,b) \cdot \mathrm{S_A}(a,b) \text{ where:}$$

$\mathrm{S}(a,b) \leq 1$
$\mathrm{S_E}(a,b)$ – Score for difference in total employment between *a* and *b*
$\mathrm{S_T}(a,b)$ – Score for difference in time between *b* and the most recent panel
$\mathrm{S_I}(a,b)$ – Score for difference in six-digit industry between *a* and *b*
$\mathrm{S_O}(a,b)$ – Score for difference in ownership between *a* and *b*
$\mathrm{S_A}(a,b)$ – Score for difference in detailed area between *a* and *b*

The employment component is $\mathrm{S_E}(a,b) = \left(1 - \frac{|E_a - E_b|}{E_a + E_b}\right)$, where E_a and E_b are the employment totals for the respective units. For a potential donor with 20 employees and a unit to be predicted with 15 employees, this works out to $\mathrm{S_E}(a,b) = \left(1 - \frac{|15-20|}{15+20}\right) = 0.857$.

Recently collected data are favored over data collected in previous panels. The time score component that reflects this is $\mathrm{S_T}(a,b) = 1 - \frac{p_b}{6}$, where p_b is the number of panels between the collection of data for potential donor b and the reference period. A donor unit observed in the current panel would have $\mathrm{S_T}(a,b) = 1$ and a donor unit sampled 5 panels previously would have $\mathrm{S_T}(a,b) = 1 - \frac{5}{6} = \frac{1}{6}$.

Donors would ideally be in the same industry or ownership group as the unit to be predicted, but they could be in a similar industry or different ownership.

The score component reflecting differences in industry at the detailed, 6-digit NAICS level is

$$\mathrm{S_I}(a,b) = \begin{cases} 1 & \text{if industry matches} \\ 0.25 & \text{if industry mismatches} \end{cases}$$

---

sake of explanation, we will be assuming 10 donors, which is typically the case. However, if 10 donors cannot be found, a minimum of 5 donors can be used.

Although any difference in industry is given the same penalty, donors are chosen according to a hierarchy and therefore more similar industry matches will be used before more different matches.

The score component reflecting differences in ownership is

$$\mathrm{S_O}(a,b) = \begin{cases} 1 & if\ ownership\ matches \\ 0.5 & if\ ownership\ mismatches \end{cases}$$

For example, if an unobserved unit is a private school, a private school donor would have an ownership score of 1, while a public school donor would have an ownership score of 0.5.

Donors in the same geographic area are preferred and are treated at 4 different matching levels. Units in the same state and MSA or nonmetropolitan area have an area match score of 1. Units receive an area match score of 0.75 if they are in two different areas of the same state, but both areas have the same area status—that is, both are MSAs or both are nonmetropolitan areas. Units receive a score of 0.5 if they are in two different areas of the same state, and one area is an MSA and the other is a nonmetropolitan area. Finally, units from different states receive a score of 0.25. The score component reflecting differences in area is:

$$\mathrm{S_A}(a,b) = \begin{cases} 1 & if\ same\ state\ and\ same\ MSA\ or\ nonmetropolitan\ area \\ 0.75 & if\ same\ state, different\ area, but\ same\ area\ status \\ 0.5 & if\ same\ state, different\ area\ status \\ 0.25 & otherwise \end{cases}$$

A potential donor from the most recent survey panel and the same MSA or nonmetropolitan area, detailed industry, ownership, and employment level as an unobserved unit will receive a score of 1, the maximum possible score. For a potential donor with 20 employees that would predict a unit of 15 employees ($\mathrm{S_E}(a,b) = 0.857$) from 2 panels previous to the reference period ($\mathrm{S_T}(a,b) = 0.667$), with a mismatching industry ($\mathrm{S_I}(a,b) = 0.25$), where both are privately owned ($\mathrm{S_O}(a,b) = 1$), and a matching state but differing MSA/nonmetropolitan area and area status ($\mathrm{S_A}(a,b) = 0.5$), the match score is:

$$\mathrm{S}(a,b) = 0.857 \cdot 0.667 \cdot 0.25 \cdot 1 \cdot 0.5 = 0.0715$$

Depending on the match scores of other potential donors, the unit may or may not be used in prediction.

Using match scores to select donors

Potential matches are found by a hierarchical nearest neighbor search detailed in table 3 below. All establishments with the same employment, NAICS, ownership category, state, and MSA or nonmetropolitan area will be predicted using the same set of donors.

An employment criterion is defined for each level such that the donor's employment must be within a certain percentage of the unobserved unit's. For example, in the first hierarchical level, the donor must be in the same state, NAICS, ownership category, and MSA or nonmetropolitan area of the unobserved unit to be predicted while having employment within ± 10 percent of the unobserved unit's employment. Broader industry groups use the most detailed industry level at which OEWS publishes estimates. For most industries, this is the 4-digit NAICS level. For a minority of industries, the published

OEWS estimates are defined at the 3-, 5-, or 6-digit NAICS level, or as OEWS-specific combinations of 4-digit industries.

If fewer than 10 potential donors are found at the first level of the hierarchy, the search proceeds through subsequent levels of the hierarchy, stopping when at least 10 suitable donors are found. If fewer than 10 donor units are available at hierarchy level 10, prediction will still proceed if at least 5 donor units are found when the search reaches the highest level. The matches with the highest scores are used for prediction.

Table 3: Hierarchical levels of donor matches

Hierarchy level	Characteristics that must match between the prediction cell and responder (donor)	Employment criterion
1	State – NAICS – Ownership – MSA or nonmetropolitan area	10%
2	State – NAICS – Ownership – MSA or nonmetropolitan area	20%
3	State – NAICS – Ownership	10%
4	State – NAICS – Ownership	20%
5	State – NAICS group – Ownership	None
6	State – NAICS group	None
7	NAICS – Ownership	10%
8	NAICS – Ownership	20%
9	NAICS	None
10	NAICS group	None

Note: “NAICS” is 6-digit NAICS. “NAICS group” is the most detailed NAICS level for which OEWS publishes estimates, generally the 4-digit NAICS level.

Staffing pattern donors

Staffing pattern and wage data for the 10 closest matches are used to predict the staffing pattern and wages of each unobserved unit. If the closest matches include several donors with the same match score, they will all be used, which may result in more than 10 donors. The total number of jobs predicted for each unobserved unit is represented by the unit’s population employment, defined as the average of its May and November QCEW employment for the two most recent survey panels. The unobserved unit’s occupational staffing pattern is calculated as a weighted average of the donor staffing patterns.

Within a given set of selected donors, donors that are closer matches contribute more to the prediction than donors that do not match as closely. Each selected donor’s contribution to the staffing pattern prediction is proportional to its relative match score, defined as the donor’s individual match score divided by the total combined scores of all the selected donors. The relative match score W_{b_i} of any match b_i among 10 matches, $b_1, b_2, \ldots, b_{10}$, is:

$$W_{b_i} = \frac{S(a, b_i)}{\sum_{i=1}^{10} S(a, b_i)}$$

For a given unobserved unit *U* and set of matches $b_i \; in \; (b_1, b_2, \ldots, b_{10})$, the predicted employment *E* for occupation *O* in wage interval *M* will be:

$$E_{UOM} = \sum_{i=1}^{10} W_{b_i} \cdot E_U \cdot \frac{E_{b_iOM}}{E_{b_i}}$$

where E_{b_iOM} is the employment in wage interval M for the occupation O in establishment b_i, E_{b_i} is the employment for establishment b_i, E_U is the employment of the unit to be predicted, and W_{b_i} is the relative match score of match b_i.

Modeling wages

Wages are predicted separately for each occupation in the unobserved unit's staffing pattern, based on the subset of staffing pattern donors reporting that specific occupation. For establishment b_i, w_{b_iOM} represents the wage for occupation O in interval *M.* For responding establishments, w_{b_iOM} is the wage value that will represent them as observed units and as donors to unobserved units that receive modeled data. The establishment's reported wage rates are used as w_{b_iOM} whenever wage rate data are available for all the establishment's employment in the occupation. If any interval wage data are available for a given establishment and occupation, then employees are assigned values of w_{b_iOM} sampled from a wage distribution. Donor wage values are scaled using a wage adjustment factor if a match differs from the unobserved unit in industry, ownership, area, employment, or survey panel. Methods for processing interval wage data and modeling wage adjustment factors are discussed in the following sections.

A random subset of donor wages is used to predict wages for each wage interval within each occupation for unobserved units. If employment is reported for an occupation and wage interval for at least 5 donors, then it is expected that 5 donor wages will be used, but at minimum one donor wage will be used. If fewer than 5 donors are available for an occupation and wage interval, it is likely that all donor wages will be used. For each wage interval, donor wages are sampled using Poisson sampling, with a target of 5 wages for each wage interval. A systematic sample of a single unit is also taken for use in the case where no wage is sampled using Poisson sampling. The probability of selection for a given donor wage within a given wage interval in the Poisson sample is:

$$p_{UOM} = 5\frac{W_{b_i} \cdot E_{b_iOM}/E_{b_i}}{\sum(W_{b_i} \cdot E_{b_iOM}/E_{b_i})}$$

The probability of selection for systematic sampling is one fifth of the probability for Poisson sampling.

The occupational wage W predicted for unit U for wage interval *M* is derived from a weighted composite of the occupational employment of the donor units. Assuming 10 donors, this is given by:

$$W_{UOM} = \sum_{i=1}^{10} \frac{W_{b_i} \cdot A_O(U, b_i) \cdot w_{b_iOM} \cdot \mathrm{I}\big(E_{b_iOM} \neq 0\big) \cdot \mathrm{I}(i \in S_{OM})}{\sum_{i=1}^{10} W_{b_i} \cdot \mathrm{I}\big(E_{b_iOM} \neq 0\big) \cdot \mathrm{I}(i \in S_{OM})}$$

Here, $\mathrm{A}_O(U, b_i)$ is the wage adjustment factor discussed below under "Model-based adjustments." The function $\mathrm{I}\big(E_{b_iOM} \neq 0\big)$ equals 1 when the establishment's occupational employment is nonzero for a wage interval and equals 0 otherwise. The function $\mathrm{I}(i \in S_{OM})$ equals 1 when the wage of establishment *i* has been sampled and equals 0 otherwise.

To illustrate the prediction of an unobserved unit, suppose the unobserved unit U is a jewelry store in a medium-sized MSA. The staffing pattern is predicted for each wage interval of each occupation based on the nearest donors. Detailed below is the prediction for retail salespersons in wage interval C. Suppose that of the 10 nearest respondents available, eight had employment of retail salespersons in wage interval C. Of those eight, four are other jewelry stores of similar size in the same MSA and most recent survey panel. Three are other jewelry stores in the same MSA with larger differences in size. Of those three, one is from one panel back. The last unit is in the same state and most recent panel, of a similar size, but in a different industry and MSA. To predict the employment and wage for retail salespersons in unobserved unit U, wage interval C, we use the example data and calculations in table 4.

Table 4: Example data for predicting employment and wages

Matching unit: i	1	2	3	4	5	6	7	8	9	10
Unit i relative match score: W_{b_i}	0.116	0.116	0.116	0.113	0.113	0.110	0.107	0.097	0.088	0.022
Unit i employment in interval C, occupation O: $E_{b_i O\mathrm{C}}$	4	3	0	2	3	7	8	0	3	11
Unit i total employment: E_{b_i}	21	21	21	22	20	19	18	21	25	21
Unobserved unit total employment: E_U	21	21	21	21	21	21	21	21	21	21
Unit i occupational employment ratio in interval C, occupation O: $\frac{E_{b_i O\mathrm{C}}}{E_{b_i}}$	0.190	0.143	0	0.091	0.150	0.368	0.444	0	0.120	0.524
Wage adjustment factor: $\mathrm{A}_O(U, b_i)$	1	1	1	1	1	0.9	1	1.03	1.03	1.1
Unit i wage in interval C, occupation O: $w_{b_i O\mathrm{C}}$	13.25	13.25	N.A.	13.25	13.25	13.25	13.25	N.A	13.05	13.25
Poisson Sampling Indicator: $\mathrm{I}(i \in S_{O\mathrm{C}})$	0	1	0	0	1	0	1	0	1	0
Computed unit i employment share in wage interval C: $W_{b_i} \cdot E_U \cdot \frac{E_{b_i O\mathrm{C}}}{E_{b_i}}$	0.463	0.348	0	0.216	0.356	0.850	0.998	0	0.222	0.242
Computed unit i wage share in wage interval C: $\frac{W_{b_i} \cdot \mathrm{A}_O(U, b_i) \cdot w_{b_i O\mathrm{C}} \cdot \mathrm{I}(E_{b_i O\mathrm{C}} \neq 0) \cdot \mathrm{I}(i \in S_{O\mathrm{C}})}{\sum_{i=1}^{10} W_{b_i} \cdot \mathrm{I}(E_{b_i O\mathrm{C}} \neq 0) \cdot \mathrm{I}(i \in S_{O\mathrm{C}})}$	0	3.625	0	0	3.531	0	3.344	0	2.790	0

Note: Data do not correspond to existing establishments or weights.

Summing the second-to-last line ("Computed unit *i* employment share") of table 4 yields predicted employment of retail salespersons O in wage interval C for establishment *U*:

$$E_{UO\mathrm{C}} = \sum_{i=1}^{10} W_{b_i} \cdot E_U \cdot \frac{E_{b_i O\mathrm{C}}}{E_{b_i}}$$

$= 0.463 + 0.348 + 0 + 0.216 + 0.356 + 0.850 + 0.998 + 0 + 0.222 + 0.242$
$= 3.695$

Although this does not add up to a whole number, for estimation purposes it is reasonable. Summing the last line of table 4 yields the predicted wage of retail salespersons *O* in wage interval C for establishment *U*:

$$W_{UO\mathrm{C}} = \sum_{i=1}^{10} \frac{W_{b_i} \cdot A_O(U, b_i) \cdot \ w_{b_i O\mathrm{C}} \ \cdot \ \mathrm{I}\left(E_{b_i O\mathrm{C}} \neq 0\right) \ \cdot \ \ \mathrm{I}(i \ \in S_{O\mathrm{C}})}{\sum_{i=1}^{10} W_{b_i} \cdot \ \mathrm{I}\left(E_{b_i O\mathrm{C}} \neq 0\right) \ \cdot \ \ \mathrm{I}(i \ \in S_{O\mathrm{C}})}$$

$= 0 + 3.625 + 0 + 0 + 3.531 + 0 + 3.344 + 0 + 2.790 + 0$
$= \$13.29$

When this process is completed for all occupations and wage interval levels observed in the donor units, the predicted wage and employment profile of establishment *U* can be used for estimation. If the predicted wage is less than the state or federal minimum wage, the predicted wage will be set to the state or federal minimum wage (whichever is higher).

Wage processing

Wage parameters

Wage data for three types of units require additional adjustments before being used to calculate wage estimates: units with interval wage data, observed units from earlier survey panels, and unobserved units to be predicted. This wage data processing uses both wage rates and the wage interval groups shown in table 1 in the "Concepts" section.

Using interval data to compute mean wage estimates requires that a wage value be assigned to each employee. MB3 wage estimates use sampled wage rates that are computed using log-normal models fit to each panel of OEWS wage data, aggregated by occupation group and area group.

Predicting unobserved units also requires adjusting wages in the donor units to current local dollars for the unobserved units. For example, suppose an interior design firm (NAICS 541410) in a large metropolitan area and surveyed in a previous survey panel contributes to the wage prediction for an industrial design firm (NAICS 541420) in a small metropolitan area. Occupational wages will differ between these firms due to geography, industry, and time effects. Thus, wages from the first unit must be adjusted with these factors in mind to give a reasonable prediction of the second unit. A fixed effect linear regression model, fit to observed unit data, is the basis for these adjustments. Wages for observed units collected in earlier survey panels are also updated to the reference date using a regression model.

Imputing nonrespondents for wage modeling

The wage distribution and wage adjustment models are derived from the full OEWS survey sample, which includes both responding and nonresponding establishments. The wage modeling process uses hot deck imputation to impute missing staffing patterns and wages for nonresponding units. For complete nonrespondents that did not provide either staffing patterns or wages, a single “nearest neighbor” donor is used to impute the entire occupational staffing pattern. This “nearest neighbor” donor is selected based on industry, state, size class, and, for some industries, ownership.

For each occupation in the imputed staffing pattern, a wage distribution is imputed from a pool of similar respondents reporting that occupation. Partial respondents that reported complete staffing patterns, but did not report complete wage data for some or all of their occupations, also receive imputed wages for any occupations in their staffing patterns that do not have complete wage data. The wage donor search initially looks for donors from the same survey panel, MSA/nonmetropolitan area, 4-digit NAICS, size class, and, for selected industries, same ownership as the recipient. If there are not enough donors to provide wage distributions for the nonrespondents that need them, then the search criteria are loosened and the search repeated. Once a sufficiently large donor pool is found, the donor pool’s wage distribution is used to prorate the recipient’s reported employment in the occupation across the 12 wage intervals outlined in table 1 in the “Concepts” section of this document.

Once the wage modeling process is complete, the hot deck imputed data for complete nonrespondents are discarded. These units will be treated as unobserved units for estimation and will receive predicted employment and wage data using the MB3 modeling process described earlier. Partial respondents retain their imputed wage distributions (along with their reported staffing patterns) and are treated as respondents thereafter. The interval wage distributions imputed for partial respondents are assigned specific wage values for estimation, as described in “Processing interval wage data” below.

Benchmarking for wage modeling

The wage distribution model and wage adjustment model both use weighted least squares regression to estimate model parameters. Benchmarked sample weights are used in this process, such that weighted employment totals for the current panel will equal QCEW frame values for each industry, state, MSA, and size subgroup.

In MB3, benchmarking factors are used only to adjust data for the purposes of model fitting and are not used directly for estimation. For the May 2024 OEWS estimates, benchmarking uses the average of May 2024 and November 2023 QCEW employment to adjust the weighted reported occupational employment and improve the accuracy of the sampled wage rates and wage adjustment models. The ratio estimation process is carried out through a series of four hierarchical employment ratio adjustments. The ratio adjustments are also known as benchmark factors (BMFs). The BMFs are calculated for the cells defined at each of the following hierarchy levels:

Table 5: Hierarchy of benchmark factors

Level	Area	Industry	Size	Ownership
1	MSA/BOS	NAICS 3/4/5/6 digits	1-19, 20-49, 50-249, 250+	
2	State	NAICS 3/4/5/6 digits		
3	State	NAICS 3 digits		For hospitals, schools, gambling establishments, and casino hotels
4	State	NAICS 2 digits		

For each establishment, a BMF is generally calculated by finding the ratio of QCEW employment (average of May 2024 and November 2023) to weighted cell OEWS employment for the hierarchy level. There is a universal maximum and minimum BMF value to which the BMF will be set if it is higher than the maximum or lower than the minimum. The second, third, and fourth BMF hierarchy levels are computed to account for inadequate coverage of the universe employment—for example, if an establishment is in a first-level hierarchy cell with no other establishments, other factors will be calculated at the other hierarchy levels to accommodate coverage. The BMFs are dependent upon the establishment's previous hierarchy levels BMFs. A final benchmark factor is calculated for each establishment as the product of its four hierarchical benchmark factors. A benchmark weight value is then calculated as the product of the establishment's six-panel combined sample weight and final benchmark factor.

Processing interval wage data

For respondents with interval wage data, the interval data must be replaced by specific wage rate values for use in estimation and to provide donor wages to the modeling process for unobserved units. Wage rate values assigned to interval data are derived from modeled wage distributions. Wage distributions are modeled for each panel using only weighted data from that panel to represent the population. Occupation and geographic area are the strongest predictors of wages and may cause substantial differences in wage levels between establishments. To provide greater homogeneity within the data, occupations and areas with similar median wages are aggregated into groups.

We assign occupation group codes to all occupations with median wages in a given wage interval, and likewise assign group codes to all geographic areas with similar median wages. Then, all data with a given ownership status, occupation group, and area group are pooled together for modeling a wage distribution function. The units within an occupation-area-ownership group are not necessarily related in any way other than the wage interval that the median falls into.

Wage distribution modeling incorporates reported wage rate data (specific wages of each employee) from private and local government establishments. The wage distributions for each group are modeled by a log normal model fit using a log-likelihood expression that incorporates both wage rate and wage interval data.

The assignment of occupation wage groups uses single panel sample weights and reported employment levels within wage intervals to compute the national wage distribution for each detailed occupation, and then determine into which interval the median wage for that occupation falls. This determines the wage occupation group for every six-digit occupation. To be specific, we calculate occupation-specific employment in each of the twelve wage intervals in panel p:

$$\hat{E}_{ob_pp} = \sum_{e \in R_p} w_{ep} \times E_{ob_pep}$$

where R_p represents the set of panel p OEWS sampled units, w_{ep} is the sample weight for establishment e in panel p, and E_{ob_pep} is the reported level of employment in occupation o at establishment e in wage interval b_p and panel p.[3] We then calculate total occupation-specific employment:

$$\hat{E}_{op} = \sum_{b_p} \hat{E}_{ob_pp}$$

and compute the relative employment shares by wage interval:

$$\hat{s}_{b_p|o,p} = \hat{E}_{ob_pp} \Big/ \hat{E}_{op}$$

We then compute cumulative employment shares:

$$\pi_{b_p|o,p} = \sum_{b \le b_p} \hat{s}_{b_p|o,p}$$

The detailed occupation is mapped into the aggregate occupation *O* in the lowest wage interval that contains at least 50 percent of the detailed occupation's cumulative employment:

$$\pi_{O-1|o,p} < 0.5 \le \pi_{o|o,p}$$

so that each aggregate occupation corresponds to a wage interval.

Typically, there are either 11 or 12 aggregate occupations corresponding to the various wage intervals.[4] For example, if tax preparers, substitute teachers, and fast food cooks all have median wages in interval C, they would be grouped together in occupation group C, and if architectural and civil drafters, actors, and construction and building inspectors all have median wages in interval F, they would be grouped together into occupation group F.

Similarly, we compute the wage distribution for each detailed geographic area (across all occupations) and then determine in which interval the median wage for that area would fall. This determines the aggregate area for every detailed MSA or BOS area. To be specific, we calculate area-specific employment in each of the twelve wage intervals b_p in the current panel:

$$\hat{E}_{vb_pp} = \sum_{o} \sum_{e \in R_{vp}} w_{ep} \times E_{ob_pep}$$

[3] The wage interval is indexed by *p*.
[4] There are 12 wage intervals, but each wage interval is not necessarily assigned an aggregate occupation.

where R_{vp} represents the set of panel *p* OEWS sampled units in area v, w_{ep} represents the sampling weight for establishment *e* in panel *p*, and E_{ob_pep} is the reported level of employment in occupation o at establishment e in wage interval b_p and panel *p*.[5] We then calculate total area-specific employment:

$$\hat{E}_{vp} = \sum_{b_p} \hat{E}_{vb_pp}$$

and compute the relative employment shares by wage interval:

$$\hat{s}_{b_p|v,p} = \hat{E}_{vb_pp} \Big/ \hat{E}_{vp}$$

We then compute cumulative employment shares:

$$\pi_{b_p|v,p} = \sum_{b \leq b_p} \hat{s}_{b_p|v,p}$$

The detailed area *v* is mapped into the aggregate area *V* into the lowest wage interval that contains at least 50 percent of the detailed area's cumulative employment:

$$\pi_{V-1|v,p} < 0.5 \leq \pi_{V|v,p}$$

so that each aggregate area corresponds to a wage interval.

Typically, there are only three or four aggregate areas, corresponding to interval C, D, E, or F. For example, if the median wages in San Francisco, CA, and Boston, MA, fall into wage interval F, then these areas will be grouped together in area group F, while if the median wages in Chicago, IL, and Atlanta, GA, fall into wage interval E, these areas will be grouped together in area group E.

Now that we have calculated the separate aggregate occupations and aggregate areas, we combine them to create aggregated area and occupation groups within a wage interval. These aggregate occupation-areas are necessary to correctly adjust the parameters of the log-normal model and subsequently predict local sampled wage rates.

For every possible aggregate occupation-area, denoted as OV, we compute the single panel sample-weighted employment levels for each wage interval:

$$\hat{E}_{OVb_pp} = \sum_{o \in O} \sum_{e \in R_{Vp}} w_{ep} \times E_{oeb_pp}$$

where R_{Vp} is the panel p sample in aggregate area V, w_{ep} represents the sampling weight for establishment *e* in panel *p*, and E_{ob_pep} is the reported level of employment in occupation o at establishment e in wage interval b_p and panel *p*.[6] In general, there will be a limited number of aggregate occupation-area groups, typically between 33 and 48.

[5] The wage interval is indexed by *p*.

[6] The wage interval is indexed by *p*.

For example, suppose nurses and paralegals are in occupation group D for a given panel, while doctors and lawyers happen to be in occupation group G. Their employers, a hospital and a law firm, are in different metropolitan areas, but both areas are in area group C. Both employers are also privately owned with ownership code 5. The data for nurses from the hospital and paralegals from the law office are pooled with other data from the same occupation-area-ownership combination to estimate wage group DC5, while the data for doctors in the hospital and lawyers in the law firm are pooled with other data to estimate wage group GC5.

A log-normal model is fit to these aggregated-occupation-by-aggregated-area cells. A maximum likelihood estimator and the sample-weighted employment sums from the current sample are used to estimate the two parameters of the lognormal model for wage *w*, occupation *O*, and area *V,* which falls into a wage interval:

$$\ln(w_{VO}) \sim \mathrm{N}(\mu_{VO}, \sigma^2_{VO})$$

Local sampled wage rates are predicted using these wage distribution parameter estimates. All data from occupation-area-ownership group *OV* are used to fit a log-normal model. We then sample a wage for the appropriate wage interval from the wage distribution model. For example, say a paralegal's reported wage falls into wage interval E, while their occupation and geographic location fall into occupation-area-ownership group DC5. This paralegal will be assigned a wage sampled from the log-normal model within the specific occupation-area-ownership aggregate group. Each paralegal in this area reported in interval E will be independently assigned an interval E wage rate that was sampled from the distribution modeled for occupation-area-ownership group DC5.

This process converts all wage interval data to wage rate values. These sampled wage rates, along with usable reported wage rates, directly define wages for all respondents. Respondent wages are adjusted, if needed, to define wages for unobserved units.

Modeling donor wage adjustments

Using observed units to predict unobserved units relies on similarity between the units. Wage adjustment is necessary if the unobserved unit differs from donor units in industry, size, location, or time of data collection. Sample response data from the current and previous two years are used to fit fixed effect linear regression models for wage adjustments. Coefficients are determined using maximum likelihood estimation over data from the six panels. For a given occupation O, the model is of the form:

$$\ln(w_{O\mathrm{T}}) \sim \beta_{\mathrm{O}} + \beta_{\mathrm{IH}} + \beta_{\mathrm{V}} + \beta_{\mathrm{S}} \cdot \mathrm{E} + \beta_{\mathrm{T}} + \epsilon$$

w – wage
T – time between the panel of collection and the current panel
β_{O} – occupation effect across about 850 detailed occupations
β_{IH} – industry-ownership combined effects across about 1,100 detailed NAICS and ownership combinations
β_{V} – area effect across about 480 detailed areas
β_{S} – size effect linear coefficient
E – total establishment employment
β_{T} – time effect computed independently for each of 22 major occupational groups

Model-based adjustments: wage aging and cell-level adjustments

Aging factors, which provide adjustments for changes to occupational wages over time, and locality adjustments are both computed directly using wage regression model parameters. All direct match units are separately aged according to a factor based on the combination of year and SOC major group. All donors, including unstable units, are independently adjusted to account for the year and SOC major group combination, detailed occupation, industry, ownership, and size of the unit to be predicted.

Suppose unit *b* is selected as a donor for unit *a.* Unit *b* is one of typically 10 donors for unit *a* and might come from one or more panels back, in which case unit *b*'s wage data are adjusted to match local current dollars for unit *a*. The adjusted donor wage for occupation *O* in unit *a* based on adjusted unit *b* data is:

$$\widetilde{w}_{bO} = w_{bO} \cdot A_O(a,b) \cdot \beta_{\mathrm{T}}$$

Where:

$$A_O(a,b) = \frac{\exp\,(\beta_{\mathrm{V}}(a) + \beta_{\mathrm{IH}}(\mathrm{a}) + \beta_{\mathrm{S}} \cdot \mathrm{E_a})}{\exp\,(\beta_{\mathrm{V}}(b) + \beta_{\mathrm{IH}}(\mathrm{b}) + \beta_{\mathrm{S}} \cdot \mathrm{E_b})}$$

Estimates

Occupational employment and wage estimates are computed using observed data and predicted data for the population of about 8.7 million units. After all modeling and wage adjustment is completed, every unit in the population will have either reported or predicted data for occupational employment and wage rates. Employment estimates are computed by summing employment within an estimation cell, while mean wage estimates are computed by dividing summed wages by total employment for an estimation cell. Because all units also are linked to their establishment information, OEWS can also calculate employment, mean wage, and percentile estimates for a mix of industry, ownership, and area levels.

Occupational employment estimates

Estimates of occupational employment totals are computed by summing all employment counts of a given occupation over the modeled population data. Estimates are made over area, industry, and ownership. For occupation *o*, where unit *i* is any establishment in cell *c*, the occupational employment estimate is:

$$\hat{X}_{o,c} = \sum_{i \in o,c} x_{i,o}$$

Hourly wage rate estimates

Mean hourly wages are calculated by summing the hourly wages—reported or predicted—for all employees in the estimation cell and dividing by the total employment in the cell. Employees *E* in a given occupation and wage interval at a single establishment will all have the same predicted wage *w*. For establishment *i*, wage range *r*, and occupation *o* in cell *c*, the computation is as follows:

$$\hat{w}_{c,o} = \frac{\sum_{i \in c,o} \sum_r x_{iro} \cdot w_{iro}}{\sum_{i \in c,o} x_{iro}}$$

Percentile wage rate estimates are computed directly from the predicted population using the empirical distribution function with averaging, which is available in many statistical packages.

Annual wage rate estimates

For most occupations, annual wage estimates are calculated by multiplying mean or percentile hourly wage estimates by a "year-round, full time" figure of 2,080 hours (52 weeks x 40 hours) per year. These estimates, however, may not represent annual pay should the workers work more or less than 2,080 hours per year.

Although OEWS publishes both annual and hourly wage estimates for most occupations, there are some occupations for which only annual or only hourly wages are published. For example, some workers such as teachers, pilots, and flight attendants are typically paid annual salaries, but work less than the usual 2,080 hours per year. Because the survey does not collect the actual number of hours worked, hourly wage rates cannot be derived from annual wage rates with any reasonable degree of confidence. Therefore, only annual wage estimates are published for these occupations. On the other hand, full-time, year-round work may not be typical in some occupations that are usually paid on an hourly basis, such as actors or musicians and singers. For these workers, only hourly wage estimates are published.

Variance estimation

Variances for both mean wage estimates and occupational employment estimates are computed using the "bootstrap" replication technique. Many weights may be associated with a given respondent in MB3 estimates because that respondent may be used to predict multiple unobserved units. This presents problems for many approaches to computing sampling variances. However, bootstrap sample replication is amenable to this design because the full MB3 estimation system may be applied to each replicate sample. Studies that were performed using simulated data informed decisions on the specifics of the bootstrapping approach used here and the number of replicates needed for estimates to converge.

The MB3 variances are computed over 300 bootstrap sample replicates. Each set of replicate estimates is based on a subsample of the full sample and includes model fitting as well as population prediction based on this subsample. The subsample is drawn from the full sample using a stratified simple random sample with replacement design, where the size of the subsample is equal to the size of the full sample. By sampling with replacement, we are up-weighting some sampled units by including them more than once in the subsample and down-weighting others by not including them at all. MB3 selects six independent subsamples, one from each of the six semiannual survey panel samples. The stratification plan is the same used for drawing the full sample, where strata are defined by state, MSAs and nonmetropolitan areas, aggregate NAICS industry, and ownership for schools and hospitals.

Subsampling occurs only for the noncertainty sample units. All certainty units from the full sample are used in every replicate's bootstrap sample. Some strata may only contain a single noncertainty unit, for which a variance cannot be computed. These are referred to as 1-PSU strata. A collapsing algorithm combines these 1-PSU strata with other like strata to ensure that two or more noncertainty sample units are present in a particular stratum. The collapsing is by the hierarchy detailed in table 6.

Table 6: Hierarchical definitions for collapsing 1-PSU strata

Hierarchy level	Collapse
1	Panels*: (0,1), (2,3), and (4,5)
2	Panels*: (0,1,2), and (3,4,5)
3	Panels*: (0,1,2,3,4,5)
4	MSAs
5	Allocation NAICS (A_NAICS)
6	Nationally

Note: "*" Panels are labeled 0 to 5, where 0 corresponds to the May 2024 panel

Sampling variance estimates obtained through these methods do not use the same probabilities used in selection of the full sample, which presents a possible source of error. Analysis indicates that these estimates perform well in estimating sampling variance despite this potential for error.

The six replicate subsamples are combined for calculating MB3 replicate estimates. Single-panel sample weights for the most recent panel are retained for computation of wage distribution model parameters and the wage adjustment factors in each replicate. All matching, wage parameter, and estimation methods described previously are used with each 6-panel bootstrap subsample to create occupational employment and wage replicate estimates for every estimation cell. This process is repeated to create 300 sets of replicate estimates. For every estimation cell in which OEWS calculates an estimate, there are occupational employment and mean wage estimates based on the full OEWS sample, as well as 300 occupational employment and mean wage replicate estimates each based on a different bootstrap subsample. The variance estimates for the occupational estimates based on the full sample are calculated by finding the variability across the occupational replicate estimates. The bootstrap variance estimates are calculated as:

$$v_{BS}\left(\hat{\theta}_{j,D}\right) = \frac{1}{(300-1)} \sum_{b=1}^{300} \left(\hat{\theta}_{j,D}^{(b)} - \hat{\theta}_{j,D}\right)^2$$

where

$\hat{\theta}_{j,D}$ = occupational estimate (employment or mean wage) for occupation j, within estimation domain D, based on **full** sample

$\hat{\theta}_{j,D}^{(b)}$ = occupational replicate estimates (employment or mean wage) for occupation j, within estimation domain D, based on the bootstrap subsample for replicate b

Changes and special procedures for the May 2024 estimates

With the May 2024 OEWS release, the metropolitan area estimates have been updated to reflect definitions based on the 2020 decennial census and delineated in OMB Bulletin 23-01. As a result of this update, new MSAs were added to publication, some existing MSAs had name and/or compositional changes, and other MSAs were removed from publication. Some nonmetropolitan areas were also added, removed, or modified due to the new MSA delineations. For the six New England states, New England City and Town Areas (NECTAs) have been discontinued, and OEWS now publishes the MSAs and nonmetropolitan areas for these states.

On November 20, 2024, QCEW suspended publication of industry and substate data for Colorado because of data quality concerns caused by issues with the modernization of the state's unemployment insurance (UI) system. Because the QCEW microdata are fundamentally a byproduct of state UI systems, QCEW data quality is sensitive to changes in these systems. By February 19, 2025, the data quality concerns for Colorado had been sufficiently addressed to resume QCEW publication.

OEWS uses QCEW employment data—which provide comprehensive counts of nonfarm payroll employment—to adjust estimates to represent all employment that is in scope for the OEWS survey. The updated Colorado employment data were not available in time to be used in this OEWS release. As a result, data in this release do not include substate data for Colorado and its areas. Therefore, this release only contains data for about 520 metropolitan and nonmetropolitan areas instead of the full count of approximately 530 areas. The OEWS national data may be marginally impacted by the quality issues in Colorado.

Presentation

The OEWS program publishes cross-industry occupational data for the United States as a whole, for individual states, and for metropolitan and nonmetropolitan areas, along with U.S. industry-specific estimates by 2-, 3-, most 4-, and some 5- and 6-digit NAICS levels. Public–private sector ownership data are available for all industries combined and for schools and hospitals. OEWS publishes employment and wage estimates aggregated by typical entry-level educational requirements and for science, technology, engineering, and mathematics (STEM) occupations. OEWS also publishes a research dataset of estimates by state and industry.

Available data elements include estimates of employment, hourly and annual mean wages, and hourly and annual percentile wages by occupation, as well as relative standard errors (RSEs) for the employment and mean wage estimates.

OEWS data are updated on an annual basis. When updated estimates become available, we announce featured data highlights in a news release.

Accessing OEWS data

OEWS data are available in several formats on the OEWS home page. The OEWS database search tool allows customers to create customized data tables using the most recent OEWS estimates. The OEWS profile application and mapping application allow data users to create occupational profiles and state and area maps using the most recent data. OEWS data can be downloaded as zipped XLSX files from the main OEWS data page. Additional OEWS datasets for STEM occupations and by typical entry-level educational requirements and research datasets by state and industry are also available. BLS does not publish OEWS estimates by metropolitan or nonmetropolitan area and industry, but these data may be available from individual state workforce agencies.

Reliability of the estimates

Estimates developed from a sample will differ from the results of a census. An estimate based on a sample survey is subject to two types of error: sampling and nonsampling error. An estimate based on a census is subject only to nonsampling error.

The OEWS model-based estimation method (MB3) relies on a statistical model that uses information from the population and survey data to predict occupational employment and wage information needed for the OEWS estimates. Under this kind of prediction-based estimation approach, the variance of the estimates is derived from the uncertainty of the model used for prediction, rather than from the sample design. The variability of the estimates is a function of how well the model fits the data. This is measured by the model error term (residuals), which is equal to the difference between the predicted and actual outcomes. This differs from design-based estimation, where variance estimates reflect the variability in the population that arises due to the sample design and how the sample is selected.

Nonsampling error

This type of error is attributable to several causes, such as:

- Errors in the sampling frame

- Inability to obtain information for all establishments in the sample
- Differences in respondents' interpretation of a survey question
- Inability or unwillingness of the respondents to provide correct information
- Errors made in recording, coding, or processing the data
- Errors made in imputing values for missing data

Explicit measures of the effects of nonsampling error are not available.

Sampling error

When a sample, rather than an entire population, is surveyed, estimates differ from the true population values that they represent. This difference, the sampling error, occurs by chance and its variability is measured by the variance of the estimate or the standard error of the estimate (square root of the variance). The relative standard error is the ratio of the standard error to the estimate itself.

Estimates of the sampling error for occupational employment and mean wage rates are provided for all employment and mean wage estimates to allow data users to determine if those statistics are reliable enough for their needs. Only a probability-based sample can be used to calculate estimates of sampling error. The formulas used to estimate OEWS variances are adaptations of formulas appropriate for the survey design used.

The sample used in the OEWS survey is one of many possible samples of the same size that could have been selected using the same sample design. Sample estimates from a given design are said to be unbiased when an average of the estimates from all possible samples yields the true population value. In this case, the sample estimate and its standard error can be used to construct confidence intervals, or ranges of values that include the true population value with known probabilities.

To illustrate, if the process of selecting a sample from the population were repeated many times, if each sample were surveyed under essentially the same unbiased conditions, and if an estimate and a suitable estimate of its standard error were made from each sample, then the following calculations would be accurate:

1. Approximately 68 percent of the intervals from one standard error below to one standard error above the estimate would include the true population value. This interval is called a 68-percent confidence interval.
2. Approximately 90 percent of the intervals from 1.6 standard errors below to 1.6 standard errors above the estimate would include the true population value. This interval is called a 90-percent confidence interval.
3. Approximately 95 percent of the intervals from 2 standard errors below to 2 standard errors above the estimate would include the true population value. This interval is called the 95-percent confidence interval.
4. Almost all (99.7 percent) of the intervals from 3 standard errors below to 3 standard errors above the estimate would include the true population value.

For example, suppose that an estimated occupational employment total is 5,000 with an associated estimate of relative standard error of 2.0 percent. Based on these data, the standard error of the estimate is 100 (2 percent of 5,000 occupational employment). To construct a 90-percent confidence interval, add and subtract 160 (1.6 times the standard error multiplied by the standard error of 100) from the estimate to produce a confidence interval of 4,840 to 5,160. Approximately 90 percent of the intervals constructed in this manner will include the true occupational employment if survey methods are nearly unbiased.

Estimated standard errors should be taken to indicate the magnitude of sampling error only. They are not intended to measure nonsampling error, including any biases in the data. Particular care should be exercised in the interpretation of small estimates or of small differences between estimates when the sampling error is relatively large or the magnitude of the bias is unknown.

Data correction

If an error is found in a published OEWS data product (news release, data table, etc.), the product is corrected and republished, or incorrect data products are removed. A record of the error is added to the list of BLS errata, a notice describing the error is posted on the OEWS website, and data users who have signed up to receive notifications from the OEWS program are alerted via email.

Uses

The OEWS survey is a source of detailed occupational employment data for many data users, including individuals and organizations engaged in planning vocational education programs, higher education programs, and employment and training programs. OEWS data also are used to prepare information for career counseling, for job placement activities performed by state workforce agencies, and for personnel planning and market research conducted by private enterprises.

Occupational employment data are used to develop information regarding current and projected employment needs and job opportunities. This information is used in the production of state education and workforce development plans. These data enable users to analyze the occupational composition of different industries and to compare occupational composition across states and local areas, including analysis for economic development purposes. OEWS employment estimates also are used as job placement aids by helping to identify industries that employ the skills gained by enrollees in career-technical training programs. In addition, OEWS survey data serve as primary inputs into occupational information systems designed for those who are exploring career opportunities or assisting others in career decision making.

OEWS data are used by several other BLS and government programs, such as the BLS *Occupational Outlook Handbook*, Employment Projections program, Modeled Wage Estimates, and the U.S. Department of Labor Employment and Training Administration (ETA). OEWS data are used to establish the fixed employment weights for the Employment Cost Index and in the calculation of occupational rates for the Survey of Occupational Injuries and Illnesses. The Department of Labor Foreign Labor Certification (FLC) program also uses OEWS data in administering visa programs. OEWS data are used as an input into federal locality pay recommendations for the President’s Pay Agent and used for setting prevailing wages for federal contracting. OEWS employment and wage data are used in ETA's CareerOneStop.

Many OEWS data users rely on data provided by the state labor market information programs. OEWS data are used by workforce investment boards and economic development programs to attract businesses. The data provide information on labor availability by occupation as well as wages. Occupational wage data are also used by jobseekers and employers to gather wage and salary information for different occupations in different locations or different industries.

History

Key developments

- **1977:** Occupational Employment Statistics (OES) data collection begins in every state and the District of Columbia
- **1988:** A new OES data collection method begins with the compilation of employment data by industry in a 3-year cycle
- **1991:** 15 states begin to collect wage information along with occupational employment information
- **1996:** OES program begins collecting occupational employment and wage data from an annual sample of 400,000 business establishments
- **1997:** First OES estimates published
- **1999:** OES switches to the Standard Occupational Classification (SOC) system
- **2002:** OES switches to the North American Industry Classification System (NAICS)
- **2002:** OES switches to semiannual data collection
- **2003–04:** OES publishes data semiannually
- **2004:** Estimates for residual (“all other”) occupations are published for the first time
- **2005:** OES returns to annual publication (but retains semiannual data collection)
- **2005:** OES adopts new metropolitan area definitions based on the 2000 decennial census
- **2006:** Estimates for nonmetropolitan areas are published for the first time
- **2008:** OES switches from the 2002 NAICS to the 2007 NAICS
- **2009:** National estimates by public/private sector ownership are added
- **2010:** Tennessee Valley Authority (TVA) added to federal government coverage
- **2010–12:** OES transitions from the 2000 SOC to the 2010 SOC
- **2012:** OES switches from the 2007 NAICS to the 2012 NAICS
- **2012:** National estimates for SOC minor groups and broad occupations are added
- **2014:** Gambling establishments and casino hotels are reclassified in NAICS
- **2015:** OES adopts metropolitan area definitions based on the 2010 decennial census
- **2017:** OES aggregates some occupations and industries
- **2017:** Scope increased to cover some establishments previously classified in private households
- **2017:** OES switches from the 2012 NAICS to the 2017 NAICS
- **2017:** OES sample reduced
- **2018:** OES reduces some geographic detail
- **2019–21:** OES transitions from the 2010 SOC to the 2018 SOC
- **2021:** Name changed to Occupational Employment and Wage Statistics (OEWS)
- **2021:** OEWS switches to a new model-based estimation method called MB3
- **2022:** Additional changes made to the MB3 wage estimation methodology
- **2022:** OEWS switches from the 2017 NAICS to the 2022 NAICS
- **2024:** OEWS adopts metropolitan area definitions based on the 2020 decennial census

The OEWS program in its current form dates to 1996 and began publishing occupational employment and wage estimates in 1997. Since 1997, the OEWS data have undergone several changes, including changes to the estimation method; changes to the occupational and industry classification systems used; changes to the metropolitan and nonmetropolitan area definitions; and changes to the sample size, survey coverage, and survey reference dates.

Changes to the estimation method

With the May 2021 estimates, the OEWS program switched to a new model-based estimation method called MB3. Additional changes to the wage estimation methodology were made with the May 2022 estimates. These wage processing changes include using reported wage rates, if available, to represent private sector and local government employers, instead of placing the wage data into 12 wage intervals.

Research shows that MB3 produces better quality estimates than the previous estimation method, as described in the *Monthly Labor Review* article "Model-Based Estimates for the Occupational Employment Statistics program." Details of the MB3 estimation method as currently implemented are discussed in the “Calculation” section of this document.

Changes in occupational classification

The 1997 and 1998 OEWS estimates used an occupational classification system that was specific to the OEWS program. In 1999, the OEWS program adopted the federal Standard Occupational Classification (SOC) system. The 1999–May 2009 estimates were based on the 2000 version of the SOC.

Between May 2010 and May 2012, the OEWS program transitioned to the 2010 SOC. Because each set of OEWS estimates is produced by combining 3 years of survey data, the OEWS program requires 3 years to fully implement changes to the SOC. The May 2010 and May 2011 estimates were based on a combination of newer survey panels collected using the 2010 SOC and older survey panels collected using the 2000 SOC, and used a hybrid of the 2000 and 2010 SOC systems that included some OEWS-specific combinations of occupations. For more information, see the OEWS frequently asked questions. The May 2012–May 2018 estimates were based on the 2010 SOC.

Beginning with the May 2017 estimates, the OEWS program replaced 21 SOC detailed occupations with SOC broad occupations or OEWS-specific combinations of detailed occupations. These changes were made to improve data quality in cases where occupations are similar and it is difficult to obtain the information needed to code accurately to the detailed occupational level. For more information, see the May 2017 occupational and industry aggregations.

Between May 2019 and May 2021, the OEWS program implemented the 2018 SOC. The May 2019 and May 2020 estimates used a hybrid of the 2010 and 2018 SOC systems that includes some combinations of occupations not found in either version of the system. For more information, see the OEWS 2018 SOC implementation page and frequently asked questions. OEWS estimates for May 2021 and later are based fully on the 2018 SOC. To improve data quality, the OEWS program continues to aggregate some occupations to the SOC broad occupation level or as OEWS-specific combinations of 2018 SOC detailed occupations.

Changes in industry classification and survey scope

The 1997–2001 OEWS estimates used the Standard Industrial Classification (SIC) system. In 2002, the OEWS program switched from the SIC to the 2002 North American Industry Classification System

(NAICS). Updates to the NAICS system were adopted in the May 2008 estimates (2007 NAICS), May 2012 estimates (2012 NAICS), May 2017 estimates (2017 NAICS), and May 2022 estimates (2022 NAICS).

OEWS federal government coverage was expanded to include the Tennessee Valley Authority (TVA) in 2010.

Beginning with the May 2014 estimates, gambling establishments and casino hotels owned by local governments were moved from the OEWS local government industry (9993) to NAICS 7132 Gambling Industries and 72112 Casino Hotels, respectively.

The May 2017 estimates included for the first time some establishments that were reclassified from NAICS 814 Private Households, which is out of scope for the OEWS survey, to NAICS 624120 Services for the Elderly and Persons with Disabilities, which is in scope. As a result, the May 2017 estimates may show increased employment in occupations that are common in NAICS 624120.

Changes to area definitions

The OEWS program uses standard metropolitan area definitions from the U.S. Office of Management and Budget (OMB). The OEWS nonmetropolitan areas use definitions that are specific to the OEWS program and are developed in cooperation with the state workforce agencies.

The OEWS program implemented major revisions to the area definitions in the May 2005, May 2015, and May 2024 estimates. The May 2005 estimates introduced revised OMB area definitions based on the results of the 2000 census. The May 2015 and May 2024 estimates introduced revised definitions based on the 2010 and 2020 censuses, respectively. In addition to these major revisions, smaller revisions were implemented in other years. Because the OEWS nonmetropolitan areas cover the remainder of each state outside of the OMB-defined metropolitan areas, changes to the metropolitan area definitions may also affect the nonmetropolitan area definitions.

Before May 2024, OEWS metropolitan area data for the New England states were based on New England City and Town Area (NECTA) definitions rather than on the county-based Metropolitan Statistical Area (MSA) definitions. With the publication of the May 2024 estimates, NECTA definitions were discontinued in the New England states and replaced with MSA definitions.

With the May 2018 estimates, the OEWS program reduced the level of geographic detail available in some areas. For the 11 large metropolitan areas that are further broken down into metropolitan divisions, OEWS no longer publishes data for the divisions. Data for these 11 areas are now available at the metropolitan area level only. In addition, some smaller nonmetropolitan areas were combined to form larger nonmetropolitan areas. For more information, see changes to metropolitan and nonmetropolitan data.

Changes to sample size and reference period

Before 2002, the OEWS program collected data from 400,000 business establishments annually with a 4th quarter reference date. Survey respondents were asked to provide data as of an October, November, or December payroll, depending on the specific respondent.

In 2002, OEWS switched to semiannual data collection to reduce seasonal effects. Data were collected in two semiannual survey panels of approximately 200,000 business establishments each, with reference dates of May 12 and November 12.

The OEWS program also published estimates semiannually in 2003 and 2004. In 2005, the OEWS program returned to publishing data annually, but retained semiannual data collection.

The OEWS sample was reduced in recent years. Prior to May 2017, each survey panel contained approximately 200,000 establishments. Information on current sample sizes is available in the "Design" section of this document.

Program name change

In the spring of 2021, the Occupational Employment Statistics (OES) program began using the name Occupational Employment and Wage Statistics (OEWS) to better reflect the range of data available from the program. May 2021 and later data products reflect the new program name. Data collection materials, including forms, emails, and letters, were updated beginning with the May 2021 survey panel. Webpages, publications, and other materials associated with previous data releases retain the Occupational Employment Statistics name.

Data before 1997

Data from the immediate predecessor to the current OEWS program are available at the bottom of the main OEWS data page. These data cover the period 1988–95 and are not directly comparable to more recent OEWS data. The 1988–95 data consist only of national occupational employment estimates by 2- and 3-digit SIC industry, with data for each industry available only once every three years. These estimates do not contain wage data or state and area data. Because data are not available for all industries in a given year, it is not possible to calculate total national employment in an occupation from these estimates.

More information

Further information on the OEWS program can be found through the OEWS webpages.

OEWS estimates and information are available at multiple locations on the BLS website:

- Main OEWS data page
- Query tool (contains only the most recent OEWS data)
- OEWS FAQs page
- Current area definitions
- Standard Occupational Classification (SOC) homepage
- Documentation (contains technical notes, survey methods and reliability statements, occupational definitions, and other documentation)
- OEWS publications