Occupational Employment and Wage Statistics

The Occupational Employment Statistics (OES) program produces employment and wage estimates annually for over 800 occupations. These estimates are available for the nation as a whole, for individual states, and for metropolitan and nonmetropolitan areas; national occupational estimates for specific industries are also available.

Table of Contents

Quick Facts: Occupational Employment and Wage Statistics	
Subject areas	Employment, Pay
Key measures	Employment Occupational wages
How the data are obtained	Survey of businesses and governments
Classification system	Geography, Industry, Occupation
Periodicity of data availability	Annual
Geographic detail	Metro area, National, Nonmetropolitan areas, State
Scope	Government, Private sector
Key products	Occupational Employment Statistics news release OES Data Tables
Program webpage	www.bls.gov/oes

Concepts

The Occupational Employment Statistics (OES) program is the only comprehensive source of regularly produced occupational employment and wage rate information for the U.S. economy. The scope comprises 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 OES program.

Unit of observation

The OES 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 OES survey uses the Standard Occupational Classification system to classify workers into occupations and the North American Industry Classification System to classify establishments into industries.

The occupational coding system

The OES survey categorizes workers into occupations based on the Office of Management and Budget's Standard Occupational Classification (SOC) 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. Employees are assigned to an occupation based on the work they perform 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 website at www.bls.gov/soc/.

The industry coding system

The OES estimates use the North American Industry Classification System (NAICS). The NAICS is used throughout the federal government to group establishments into industries based on the goods or services they produce. The NAICS has a hierarchical structure with several levels of industry detail: by broad industrial sectors,

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 website at https://www.bls.gov/bls/naics.htm.

Ownership

OES classifies most government-owned establishments differently from the standard NAICS. The NAICS classifies government establishments according to their primary function and includes detailed industries within sector 92 Public Administration. Under the standard NAICS system, government establishments in NAICS sector 92 generally oversee governmental programs and activities that are not performed by private establishments, while government establishments producing goods and services similar to those typically identified with private-sector establishments should be classified in the same industry as private-sector establishments engaged in similar activities.

The OES program classifies some government establishments differently and therefore does not use NAICS sector 92. Instead, the OES 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 OES state and local government data (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 including schools, hospitals, and local government gambling establishments and casino hotels are also available as part of the cross-industry ownership estimates.

OES federal government data include the federal executive branch, U.S. Postal Service (USPS) and (since 2010) the Tennessee Valley Authority only. The judicial and legislative branches of the federal government are not covered. OES federal government data (NAICS 9991) consist of the federal executive branch and TVA; USPS data are in NAICS 491100 Postal Service. Data for the federal executive branch, USPS, and TVA combined are also available in the cross-industry ownership estimates.

Area definitions

The OES program uses the metropolitan statistical area (MSA) definitions provided by the Office of Management and Budget (OMB). For the New England states, OES uses the New England City and Town Area (NECTA) definitions instead of the MSA definitions. Nonmetropolitan areas are specific to the OES 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.

Key concepts and definitions

Employment represents the estimated number of full- and part-time jobs in an occupation. The OES 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, household workers, and unpaid family workers are not covered by the OES survey.

Wages are money that is paid or received for work or services performed in a specified period. Wages for the OES survey are straight-time, gross pay, exclusive of premium pay. 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.

OES receives individual wage rate data for the federal government, the U.S. Postal Service, and most state governments. For the remaining establishments, the OES survey data are placed into 12 intervals defined in the table below. Wage intervals are updated periodically based on the wages in the labor market. 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. The responding establishments are instructed to report the hourly rate for part-time workers, and to report annual rates for occupations that are typically paid at an annual rate but do not work 2,080 hours per year, such as teachers, pilots, and flight attendants. Other workers, such as some entertainment workers, are paid hourly rates, but generally do not work 40 hours per week, year round. For these workers, only an hourly wage is reported.

Table 1. OES wage intervals

Interval	Wages	
	Hourly	Annual
See footnotes at end of table.		
Range A	Under $9.25	Under $19,240
Range B	9.25 – 11.74	19,240 – 24,439
Range C	11.75 – 14.74	24,440 – 30,679
Range D	14.75 – 18.74	30,680 – 38,999
Range E	18.75 – 23.99	39,000 – 49,919
Range F	24.00 – 30.24	49,920 – 62,919
Range G	30.25 – 38.49	62,920 – 80,079
Range H	38.50 – 48.99	80,080 – 101,919
Range I	49.00 – 61.99	101,920 – 128,959
Range J	62.00 – 78.74	128,960 – 163,799
Range K	78.75 – 99.99	163,800 – 207,999
Range L	100.00 and over	208,000 and over

Source: U.S. Bureau of Labor Statistics, Occupational Employment Statistics.

Scope and exclusions

The OES 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 Occupational Employment Statistics (OES) survey is a semiannual survey measuring occupational employment and wage rates for wage and salary workers in nonfarm establishments in the United States. The OES data available from BLS include cross-industry occupational employment and wage estimates for the nation; over 580 areas, including states and the District of Columbia, metropolitan statistical areas (MSAs), nonmetropolitan areas, and territories; 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

The OES survey categorizes workers into occupations based on the Office of Management and Budget's Standard Occupational Classification (SOC) system. Together, these occupations make up 22 of the 23 SOC major occupational groups. Major group 55, Military Specific Occupations, is not included.

Ownership exclusions

With the exception of schools, hospitals, gambling establishments, and casino hotels, industry-specific estimates only include privately owned establishments. Schools and hospitals owned by state and local governments are included with privately owned schools and hospitals in the estimates for the appropriate NAICS code.

Industry coverage and exclusions

The OES 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. OES federal government data include the U.S. Postal Service and the federal executive branch only. All other industries, including state and local government, are covered by the survey. Industries that fall within the OES scope are shown in exhibit 1. The sectors shown in exhibit 1 are stratified into 302 industry groups at the 3-, 4-, 5-, or 6-digit NAICS level of detail.

Exhibit 1. NAICS industry sectors covered in OES

Industry code	Industry title
See footnotes at end of table.	
11	Logging (1133), support activities for crop production (1151), and support activities for animal production (1152) only
21	Mining

Exhibit 1. NAICS industry sectors covered in OES

Industry code	Industry title
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 [1]	Federal government executive branch
999200 [1]	State government
999300 [1]	Local government

[1] OES-defined industry code that is not part of the NAICS.
Source: U.S. Bureau of Labor Statistics, Occupational Employment Statistics.

Changes and special procedures in the May 2019 estimates

With the May 2019 estimates, the OES program began implementing the revised 2018 Standard Occupational Classification (SOC) system. For more information on the 2018 SOC, see www.bls.gov/soc/2018/home.htm. Because the May 2019 estimates are based on a combination of survey data collected using the 2010 SOC and survey data collected using the 2018 SOC, these estimates use a hybrid of the two classification systems that contains some combinations of occupations that are not found in either the 2010 or 2018 SOC. The May 2021 estimates, to be published in spring 2022, will be the first OES estimates based entirely on survey data collected using the 2018 SOC.

The November 2018 and May 2019 survey panels were collected using the 2018 SOC. The May 2017 through May 2018 panel data collection used a slightly modified version of the 2010 SOC in which 21 detailed occupational codes were replaced with 10 aggregations of those occupations. More information on these SOC aggregations is available in table 1 at www.bls.gov/oes/changes_2017.htm. The November 2016 panel was collected using the 2010 SOC system. More information about the 2010 SOC system can be found at www.bls.gov/soc/2010.

Almost all occupations in the May 2019 OES hybrid system are the same as 2018 SOC occupations. However, in some cases, survey data for more than one 2018 and/or 2010 SOC occupation were combined to form a hybrid occupation. Some of these combinations are equivalent to a standard 2018 SOC broad occupation and are published under the 2018 SOC broad occupation code and title. Other combinations are equivalent to an

occupation in the 2010 SOC and are published under their 2010 SOC code and title. Finally, some combinations are not found in either the 2010 or 2018 SOC structure. These are published under an OES-specific hybrid code and a composite title that indicates the content of the hybrid.

Because some of the hybrid codes combine occupations from more than one 2018 SOC minor group or broad occupation, the May 2019 estimates do not contain data for some 2018 SOC minor groups and broad occupations.

The OES program has also replaced some 2018 SOC detailed occupations with SOC broad occupations or OES-specific aggregations. These include home health aides and personal care aides, for which OES will publish only the 2018 SOC broad occupation 31-1120 Home Health and Personal Care Aides.

For more information on the occupational classification system used in the May 2019 OES estimates, please see www.bls.gov/oes/soc_2018.htm and www.bls.gov/oes/oes_ques.htm#qf10.

The May 2019 OES estimates use the metropolitan area definitions delineated in Office of Management and Budget (OMB) Bulletin 17-01. These definitions add a new Twin Falls, Idaho, metropolitan area, consisting of Jerome County and Twin Falls County, Idaho. These counties have been removed from the definition for the OES Southeast-Central Idaho nonmetropolitan area.

The 2012 NAICS was used to define sampling cells and the 2017 NAICS was used to define estimation cells for panel data for November 2016 and May 2017. Sample weights for these panels were recomputed to reflect the update to 2017 NAICS codes. Sampling and data collection for the November 2017 through May 2019 panels used the 2017 NAICS.

Last Modified Date: August 31, 2020

Data Sources

Data collection

The Occupational Employment Statistics (OES) 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 State Workforce Agencies collect most of the data. OES estimates are constructed from a probability sample of about 1.1 million establishments. Each year, two semiannual panels of approximately 180,000 to 200,000 sampled establishments are contacted, one panel in May and the other in November. Responses are obtained by mail, online, email, telephone, or personal visit. For a given panel, most sampled establishments initially receive either a survey questionnaire or instructions for reporting their data electronically. Nonrespondents receive up to three additional mailings and may be contacted by phone or email. A sample survey form is available at https://www.bls.gov/respondents/oes/pdf/forms/uuuuuu_fillable.pdf. Instructions for electronic submission are available at https://www.bls.gov/respondents/oes/instructions.htm.

Confidentiality

BLS has a strict confidentiality policy that ensures that 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 which 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 regarding confidentiality, please visit the BLS website at www.bls.gov/bls/confidentiality.htm.

Quality control

The OES survey is a Federal-State cooperative effort that enables states to conduct their own surveys. A major concern regarding a cooperative program such as OES is the accommodation of 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 computerized quality control tool used by the OES survey is the Occupation and Wage Data Network (OWDN) system. It was developed to provide a consistent and automated framework for survey processing and to reduce the workload for analysts at the State, regional, and national levels.

To ensure standardized sampling methods in all areas, the sample is drawn in the 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 employment estimates, have resulted in the overall

standardization of the OES survey methodology. This has reduced the number of errors in the data files as well as the time needed to review them.

Several edit and quality control procedures are used to reduce nonsampling error. Follow-up mailings, emails, and phone calls are sent out to nonresponding establishments to improve the survey response rate, especially those that are critical or large. Completed survey questionnaires are checked for data consistency and reviewed to verify the accuracy and reasonableness of the reported data.

Other quality control measures used in the OES survey include:

- follow-up mail and telephone solicitations of nonrespondents, especially critical or large nonrespondents
- review of data during collection to verify its accuracy and reasonableness
- adjustments for atypical reporting units on the data file
- validation of the benchmark employment figures and the benchmark factors
- validation of the analytical tables of estimates at the NAICS 3/4/5/6 level
- response analysis studies conducted to assess respondents' comprehension of the questionnaire

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 the U.S. Postal Service. Data from only the most recent year are retained for use in OES estimates.

In each area, 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 OES 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) are used in occupational employment benchmarking. A ratio estimator is used to calculate estimates of occupational employment. The auxiliary variable for the estimator is the average of the latest May and November employment totals from the QCEW.

Data from the BLS National Compensation Survey (NCS) are used to compute mean wage rates for all workers in any given wage interval. Although smaller than the OES survey in terms of sample size, the NCS program, unlike OES, collects individual wage data. The mean hourly wage rate for interval *L* (the upper, open-ended wage interval) is calculated without wage data for pilots. This occupation is excluded because pilots work fewer hours than workers in other occupations. Consequently, their hourly wage rates are much higher.

Data from the Bureau's Employment Cost Index (ECI) survey are used to develop wage-aging factors. The ECI survey measures the rate of change in compensation for 11 major occupational groups on a quarterly basis. Aging factors are used to adjust OES wage data from prior survey reference periods to the current survey reference period.

Last Modified Date: August 31, 2020

Design

The Occupational Employment Statistics (OES) survey is based on a probability sample drawn from a universe of about 7.9 million in-scope establishments stratified by geography, industry, size, and ownership. The sample is designed to represent all nonfarm establishments in the United States.

Semiannual samples are referred to as panels. The survey is conducted over a rolling 6-panel (or 3-year) cycle. This is done to provide adequate geographic, industrial, and occupational coverage. Over the course of a 6-panel (or 3-year) cycle, approximately 1.1 million establishments are sampled. To the extent possible, private sector units selected in any one panel are not sampled again in the next five panels. For example, in a cycle, data collected in May 2019 are combined with data collected in November 2018, May 2018, November 2017, May 2017, and November 2016.

A probability sample is taken of local government establishments, private sector establishments, and state schools and hospitals.

Frame construction

The sampling frame, or universe, is a list of about 7.9 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, OES extracts the administrative data of establishments that are in scope for the OES 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 5 prior sample panels, and units that have been previously selected in the 5 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.

Geography. There are 606 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. Cross-state MSAs 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 302 industry groups defined at the NAICS 3-, 4-, 5-, or 6-digit level.

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

Size. Establishments are divided into certainty and noncertainty size classes.

At any given time, there are about 150,000 nonempty State/MSA-BOS/NAICS 3-, 4-, 5-, 6-digit/ownership strata on the frame. When comparing nonempty strata between frames, there may be substantial frame-to-frame differences. The differences are due primarily to normal establishment birth and death processes and normal establishment growth and shrinkage. Other differences are due to establishment NAICS reclassification and changes in geographic location.

A small number of establishments indicate the state in which their employees are located, but do not indicate 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 sample for the May 2019 survey is the equivalent of six panels. The sample was reduced in recent panels. To the extent possible, private sector units selected in any one panel are not sampled again in the next five panels. The sample allocations, excluding federal government and U.S. Postal Service (USPS), for the panels in this cycle are:

182,809 establishments for May 2019

186,679 establishments for November 2018

186,125 establishments for May 2018

185,450 establishments for November 2017

195,117 establishments for May 2017

201,952 establishments for November 2016

The May 2019 data include a census of 8,094 federal and USPS units. The combined sample size for the May 2019 estimates is approximately 1.1 million establishments, which includes only the most recent data for federal and state government. Federal and state government units from older panels are deleted to avoid double counting.

Allocation methods

The sampling frame is stratified into approximately 150,000 nonempty State/MSA-BOS/NAICS 3-, 4-, 5-, 6-digit/ownership 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. For the remaining noncertainty strata, a set of minimum sample size requirements based on the number of establishments in each cell is used to ensure coverage for industries and MSAs. For each State/MSA-BOS/NAICS 3-, 4-, 5-, 6-digit/ownership stratum, a sample allocation is calculated using a power Neyman allocation.[1] The actual 6-panel sample allocation is the larger of the minimum sample allocation and the power 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. One 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 samples than strata with lower levels of employment. Using the square root within the Neyman allocation softens this effect. The other is a measure of the occupational variability of the industry based on prior OES survey data. The occupational variability of an industry is measured by computing the coefficient of variation (CV) for each occupation within the 90th percentile of occupational employment in a given 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 than industries that are more occupationally homogeneous.

Sample selection

Sample selection within strata is approximately proportional to size. In order to provide the most occupational coverage, establishments with higher employment are more likely to be selected than those with lower employment; some of the largest establishments are selected with certainty. The unweighted employment of sampled establishments makes up approximately 57.2 percent of total employment.

Permanent random numbers (PRNs) are used in the sample selection process. To minimize sample overlap between the OES 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, and state government units are assigned a panel weight of 1. Other sampled establishments are assigned a design-based panel weight, which reflects the inverse of the probability of selection.

NOTES

[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 "Power Allocations: Determining Sample Sizes for Subnational Areas," Michael D. Bankier, *The American Statistician*, vol. 42, no. 3 (Aug., 1988), pp. 174–177.

Last Modified Date: August 31, 2020

Calculation

Each semiannual sample represents roughly one-sixth of the establishments for the full 6-panel sample plan. Each sample is used in conjunction with the previous five semiannual samples in order to create a combined sample of approximately 1.1 million establishments. This includes only the most recent data for federal and state government. In this cycle, data collected in May 2019 are combined with data collected in November 2018, May 2018, November 2017, May 2017, and November 2016.

Of the approximately 1.1 million establishments in the 50 states and the District of Columbia in the combined initial sample, approximately 1,067,000 were viable establishments (that is, establishments that are not outside the scope or out of business). Of the viable establishments, approximately 756,000 responded and 311,000 did not—a 70.9-percent response rate. The response rate in terms of weighted sample employment is 67.7 percent.

Preparing data for estimation

Sample data must be correctly prepared prior to computation of occupational employment and wage estimates and estimates of their variance. Data for sampled nonrespondents are imputed and benchmarking factors are computed before estimation. This is necessary for sampled data from the current panel to be reweighted to correctly reflect industrial employment levels recorded in the U.S. Bureau of Labor Statistics Quarterly Census of Employment and Wages (QCEW).

Nonresponse

Nonresponse is a chronic problem in virtually all large-scale surveys because it may introduce a bias in estimates if the nonrespondents tend to differ from respondents in terms of the characteristic being measured. To partially compensate for nonresponse, the missing data for each nonrespondent are imputed using plausible data from responding units with similar characteristics.

Preparing data for estimation

Data for sampled nonrespondents are imputed and benchmarking factors are computed before estimation. This is necessary for sampled data from the current panel to be reweighted to correctly reflect industrial employment levels recorded in the U.S. Bureau of Labor Statistics Quarterly Census of Employment and Wages (QCEW).

Establishments that do not report occupational employment data are called “unit” nonrespondents. Establishments that report employment data but fail to report some or all of the corresponding wages are called “partial” nonrespondents. Missing data for unit nonrespondents are imputed through a two-step imputation process. Missing data for partial nonrespondents are imputed through the second step of the process only.

Step 1, Impute an occupational employment staffing pattern

For each unit nonrespondent, a staffing pattern is imputed using a nearest-neighbor “hot deck” imputation method. The procedure links a responding donor establishment to each nonrespondent. The nearest-neighbor hot deck procedure searches within defined cells for a donor that most closely resembles the nonrespondent by geographic area, industry, and employment size. Ownership is also used in the hospital, education, gambling, and casino hotel industries. The procedure initially searches for a donor whose reported employment is approximately the same as the nonrespondent’s frame employment within the same 5- or 6-digit NAICS (North American Industry Classification System) or NAICS aggregation, state, and ownership. If more than one otherwise equally qualified donor is found, a donor from a more recent panel will be selected over a donor from an older panel. If the search is unsuccessful, the pool of donors is enlarged in incremental steps by expanding geographic area and industry until a suitable donor is found. Limits are placed on the number of times a donor can be used.

After a donor has been found, its occupational staffing pattern is used to prorate the nonrespondent’s frame employment by occupation. The prorated employment is the nonrespondent’s imputed occupational employment.

Step 2, Impute an employment distribution across wage intervals

For each “unit” nonrespondent in step 1 or for each “partial” nonrespondent, impute an employment distribution across wage intervals for occupations without complete wage data. This distribution, called the wage employment distribution, is imputed as follows:

- Identify the imputation cell for each of the nonrespondent’s occupations. Imputation cells are initially defined by MSA (Metroplitan Statistical Area) / BOS (Balance of State), NAICS 5/6 or NAICS aggregation, and size class from the most recent panel only. For schools, hospitals, gambling establishments, and casino hotels, cells are further divided by ownership.

- Determine if the imputation cell has enough respondents to compute wage employment distributions. If not, incrementally enlarge the cell until there are enough respondents.

- Use the distributions above to prorate the nonrespondent’s imputed occupational employment across wage intervals. (Or, for partial respondents, use the distributions above to prorate the reported occupational employment across wage intervals.)

Special procedures

Within the past 3-year cycle, the OES had critical nonrespondents that could not be imputed using current OES methods. The OES employed special imputation procedures that used nonrespondents’ prior staffing patterns. The occupational employment was benchmarked to the current year and the wage distribution was imputed using procedures very similar to the current partial imputation method.

Reweighting for the combined sample

Employment and wage rate estimates are computed using a rolling 6-panel (3-year) sample. Establishments from each panel's sample are initially assigned weights as if one panel were being used to represent the entire population. When the samples are combined, each sampled establishment must be reweighted so that the aggregated sample across 6 panels represents the entire population. Establishments selected with certainty in the 6-panel cycle are given a weight equal to 1. Noncertainty units are reweighted stratum by stratum. This revised weight is called the 6-panel combined sample weight. The original single-panel sampling weights are computed so that responses in a stratum could be weighted to represent the entire stratum population. In one common scenario, 6-panel samples are combined, and all 6 panels have sample units for a particular stratum. A summation of the single-panel weights would over-represent the population by a factor of six. Because we do not want to over-represent the stratum population, the 6-panel combined sample weight of each establishment is set equal to 1/K times its single-panel sampling weight. In general, when 6-panel samples are combined, a count of the number of panels with at least one unit selected for a given stratum is assigned to K.

Benchmarking to QCEW employment

A sum of ratio-adjusted weighted reported occupational employment is used to calculate estimates of occupational employment. The auxiliary variable for the estimator is the average of the latest May and November employment totals from the BLS Quarterly Census of Employment and Wages (QCEW). For the May 2019 estimates, the auxiliary variable is the average of May 2019 and November 2018 employment. To balance the states' need for estimates at differing levels of geography and industry, 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 first of the hierarchical benchmark factors is calculated for cells defined by state, MSA/BOS, NAICS 3/4/5/6, and employment size class (4 size classes: 1-19, 20-49, 50-249, 250+). For establishments in the hospital and education industries (NAICS 622 and 611), the first hierarchical factor is calculated for cells defined by state, MSA/BOS, NAICS 3/4/5/6, employment size class (4 size classes: 1-19, 20-49, 50-249, 250+), and ownership (state government, local government, or privately owned). If a first-level BMF is out of range, it is reset to a maximum (ceiling) or minimum (floor) value. First-level BMFs are calculated as follows:

h = MSA/BOS by NAICS 3/4/5/6

H = state by NAICS 3/4/5/6

s = employment size classes (1-19, 20-49, 50-249, 250+)

S = aggregated employment size classes (1-49, 50+)

o = ownership (state government, local government, or privately owned)

M = average of May and November QCEW employment

w_i = six-panel combined sample weight for establishment i

x_i = total establishment employment

BMF_{min} = a parameter, the lowest value allowed for BMF

BMF_{max} = a parameter, the highest value allowed for BMF

$w_i,\ x_i,\ \mathrm{BMF}_{\min}$

$\mathrm{BMF}_{\max},\ \beta_{hs} = (M_{hs}/\sum_{i\in hs} w_i x_i),\quad \beta_{hS} = (M_{hS}/\sum_{i\in hS} w_i x_i),\quad \beta_h = (M_h/\sum_{i\in h} w_i x_i),$
$\beta_{hso} = (M_{hso}/\sum_{i\in hso} w_i x_i), \beta_{hSo} = (M_{hSo}/\sum_{i\in hSo} w_i x_i),\ \beta_{ho} = (M_{ho}/\sum_{i\in ho} w_i x_i), \text{then},$ then

$$\mathrm{BMF}_{1,\mathrm{hs}} = \begin{cases} \beta_{hso}, \text{if all } \beta_{hso} \text{ within } h \text{ are bounded by } (\mathrm{BMF}_{\min}, \mathrm{BMF}_{\max}), \\ \beta_{hs}, \text{if all } \beta_{hs} \text{ within } h \text{ are bounded by } (\mathrm{BMF}_{\min}, \mathrm{BMF}_{\max}), \\ \beta_{hSo}, \text{if all } \beta_{hSo} \text{ within } h \text{ are bounded by } (\mathrm{BMF}_{\min}, \mathrm{BMF}_{\max}), \\ \beta_{hS}, \text{if all } \beta_{hS} \text{ within } h \text{ are bounded by } (\mathrm{BMF}_{\min}, \mathrm{BMF}_{\max}), \\ \beta_{ho}, \text{if all } \beta_{ho} \text{ within } h \text{ are bounded by } (\mathrm{BMF}_{\min}, \mathrm{BMF}_{\max}), \\ \beta_{h}, \text{if all } \beta_{h} \text{ within } h \text{ are bounded by } (\mathrm{BMF}_{\min}, \mathrm{BMF}_{\max}), \\ \mathrm{BMF}_{\min}, \text{if } \beta_h < \mathrm{BMF}_{\min}, \\ \mathrm{BMF}_{\max}, \text{if } \beta_h > \mathrm{BMF}_{\max} \end{cases}$$

Second-level BMFs are calculated for cells defined at the state, NAICS 3/4/5/6 level by summing the product of combined 6-panel weight and first-level BMF for each establishment in the cell. For establishments in the hospital, education, gambling, and casino hotel industries (NAICS 622, 611, 7132 and 72112), the first hierarchical of the second-level BMK factor is calculated at the state, NAICS 3/4/5/6, and ownership level. Second-level BMFs account for the portion of universe employment that is not adequately covered by weighted employment in first-level benchmarking. Inadequate coverage occurs when “MSA/BOS | NAICS 3/4/5/6 | size class” cells have no sample data or when a floor or ceiling is imposed on first-level BMFs. Second-level benchmarks are calculated as follows:

$$\beta_{Ho} = \left(M_{Ho} \Big/ \sum_{hs\in H} \sum_{i\in hs} w_i x_i \mathrm{BMF}_{1,\mathrm{hs}} \right)$$

$\beta_H = \left(M_H \Big/ \sum_{hs\in H} \sum_{i\in hs} w_i x_i \mathrm{BMF}_{1,\mathrm{hs}} \right),\ \text{then},$ then

$$\mathrm{BMF}_{2,\mathrm{H}} = \begin{cases} \beta_{Ho}, \text{if all } \beta_{Ho} \text{ within H are bounded by } (\mathrm{BMF}_{\min}, \mathrm{BMF}_{\max}), \\ \beta_{H}, \text{if all } \beta_{H} \text{ within H are bounded by } (\mathrm{BMF}_{\min}, \mathrm{BMF}_{\max}), \\ \mathrm{BMF}_{\min}, \text{if } \beta_{H} < \mathrm{BMF}_{\min}, \\ \mathrm{BMF}_{\max}, \text{if } \beta_{H} > \mathrm{BMF}_{\max} \end{cases}$$

Third-level BMFs ($BMF_{3,H}$) are calculated at the “State | 3-digit NAICS” cell level by summing the product of combined 6-panel weight, first-level BMF, and second-level BMF for each establishment in the cell. The third-level BMF also benchmarks by ownership for the hospital, education, gambling, and casino hotel industries. Fourth-level BMFs ($BMF_{4,H}$) are calculated at the “State | 2-digit NAICS” cell level by summing the product of final weight, first-level BMF, second-level BMF, and third-level BMF for each establishment in the cell. The fourth-level BMK factor does not benchmark by ownership. As with second-level BMFs, third- and fourth-level BMFs are computed to account for inadequate coverage of the universe employment.

A final benchmark factor, BMFi, is calculated for each establishment as the product of its four hierarchical benchmark factors ($BMF_i = BMF_1 * BMF_2 * BMF_3 * BMF_4$). A benchmark weight value is then calculated as the product of the establishment’s six-panel combined sample weight and final benchmark factor.

Estimation methodology

OES produces estimates of occupational employment totals, mean wage rates, and wage rate percentiles. Variance estimates are produced via jackknife random group and Taylor series linearization methods.

Occupational employment estimates

Benchmark factors and the combined 6-panel weights are used to compute estimates of occupational employment. Estimates are produced for cells defined by geographic area and industry group. The total employment for an occupation in a cell is estimated by taking the product of the reported occupational employment, the 6-panel combined sample weight, and the final benchmark factor for each establishment in the cell, and summing the product across all establishments in the cell. This sum is the estimate of total occupational employment in the cell.

The equation below is used to calculate occupational employment estimates for an estimation cell defined by geographic area, industry group, and size class.

$$(BMF_{3,H})$$

where

o = occupation

h = estimation cell

w_i = six-panel combined sample weight for establishment *i*

BMF_i = final benchmark factor for establishment *i*

x_{io} = employment for occupation *o* in establishment *i*

$(BMF_{4,H})$ = estimated employment for occupation *o* in cell *h*

Wage rate estimation

Two externally derived parameters are used to calculate wage rate estimates. They are:

- the mean wage rates for each of the 12 wage intervals and
- wage updating factors (also known as aging factors)

Wage rates of workers are converted to 1 of 12 consecutive, nonoverlapping wage bands. Individual wage rates are used for federal government and U.S. Postal Service workers. State governments may report their data as either individual wage rates or interval wage rates.

An illustration

An establishment employs 10 secretaries at the following wage rates:

$9/hour	1 secretary
$10/hour	1 secretary
$12/hour	2 secretaries
$13/hour	2 secretaries
$14/hour	2 secretaries
$16/hour	1 secretary
$17/hour	1 secretary

Wage rates for secretaries, however, are used in the OES survey as follows:

Wage interval A (under $9.25/hour)	1 secretary
Wage interval B ($9.25-$11.74/hour)	1 secretary

Wage interval C ($11.75-$14.74/hour)	6 secretaries
Wage interval D ($14.75-$18.74/hour)	2 secretaries

The remaining wage intervals have 0 secretaries.

Because wage rates are grouped into intervals, we must use grouped data formulas to calculate estimates of mean and percentile wage rates. Assumptions are made when using grouped data formulas. For the mean wage rate formula, we assume that we can calculate the average wage rate for workers in each interval. For the percentile wage rate formula, we assume that workers are evenly distributed in each interval.

Wage data from the May 2019, November 2018, May 2018, November 2017, May 2017, and November 2016 panels were used to calculate May 2019 wage rate estimates. Wage data from different panels, however, are not equivalent in real-dollar terms due to inflation and changing compensation costs. Consequently, wage data collected prior to the current survey reference period have to be updated or aged to approximate that period.

Determining a mean wage rate for each interval

The mean hourly wage rate for all workers in any given wage interval cannot be computed using grouped data collected by the OES survey. This value is calculated externally using data from the BLS National Compensation Survey (NCS). With the exception of the highest wage interval, mean wage rates for each panel are calculated using the most recent NCS data available. The hourly mean wage rate of the highest wage interval is calculated differently from the others. A weighted average of the previous 3 years' means is used, instead of just the current year's mean. Note that the mean hourly wage rate for interval *L* (the upper, open-ended wage interval) is calculated without wage data for pilots. This occupation is excluded because pilots work fewer hours than workers in other occupations.

Wage aging process

Aging factors are developed from the Bureau's Employment Cost Index (ECI) survey. The ECI survey measures the rate of change in wages and salaries for ten major occupational groups on a quarterly basis. Aging factors are used to adjust OES wage data from past survey reference periods to the current survey reference period. The procedure assumes that there are no major differences by geography, industry, or detailed occupation within the occupational division. The twelfth, open-ended, interval is not aged.

Mean hourly wage rate estimates

For data from private sector, local government, and certain state government establishments, the mean hourly wage is calculated as the total weighted hourly wages for an occupation divided by its weighted survey employment. Estimates of mean hourly wages are calculated using a standard grouped data formula that was modified to use ECI aging factors as:

BMF_i

where

$(\mathrm{BMF}_i = \mathrm{BMF}_1 * \mathrm{BMF}_2 * \mathrm{BMF}_3 * \mathrm{BMF}_4 \qquad \hat{X}_{ho} = \sum_{i \in h}(w_i \mathrm{BMF}_i x_{io}), \qquad \boldsymbol{w_i}$

o = occupation

x_{io} = mean hourly wage rate for occupation o

z = panel (or year)

t = current panel

w_i = six-panel combined sample weight for establishment *i*

BMF_i = final benchmark factor applied to establishment *i*

$(\mathrm{BMF}_i = \mathrm{BMF}_1 * \mathrm{BMF}_2 * \mathrm{BMF}_3 * \mathrm{BMF}_4$ = unweighted total hourly wage estimate for occupation o in establishment *i*

r = wage interval

$\hat{X}_{ho}$ = estimated employment for occupation *o*

$x_{i\,o\,r}$ = reported employment for occupation *o* in establishment *i* in wage interval *r* (note that establishment *i* reports data

for only one panel z or one year z)

$u_{z\,o}$ = ECI aging factor for panel (or year) *z* and occupation *o*

$c_{z\,r}$ = mean hourly wage for interval *r* in panel (or year) *z*

In this formula, $c_{z\,r}$ represents the mean hourly wage of interval *r* in panel (or year) *z*. The mean is computed externally using data from the Bureau's NCS survey.

For wage rate data from federal and certain state government establishments, the hourly wages for an occupation within an establishment are summed to get total wages. Employment for that occupation within that establishment

is also summed to get total employment. The total wages and total employment across all establishments in the occupation for the estimation level of interest are summed.

$$\widehat{R}_o = \frac{\sum_{z=t-5}^{t}(\sum_{i\in z} w_i \mathrm{BMF}_i\, \hat{y}_{io})}{\hat{X}_o}$$

Percentile hourly wage rate estimates

The *p*-th percentile hourly wage rate for an occupation is the wage where *p* percent of all workers earn that amount or less and where (100-*p*) percent of all workers earn that amount or more. The wage interval containing the p-th percentile hourly wage rate is located using a cumulative frequency count of estimated employment across all wage intervals. After the targeted wage interval is identified, the *p*-th percentile wage rate is then estimated using a linear interpolation procedure. This statistic is calculated by first distributing federal, state, local government, and private sector workers inside each wage interval. Federal and certain state government workers are distributed throughout the wage intervals according to their wage rates, while certain state government, local government, and private sector workers are distributed uniformly within each wage interval. Next, workers are ranked from lowest paid to highest paid. Finally, the product of the total employment for the occupation and the desired percentile is calculated to determine the worker that earns the p-th percentile wage rate.

$$\widehat{R}_o$$

where

$\hat{y}_{io}$ = *p*-th percentile hourly wage rate for occupation *o*

r = wage interval that encompasses $\hat{y}_{io}$

$i = u_{zo} \sum_r c_{zr}\,,\ (i \in z)$ = lower bound of wage interval *r*

$\hat{X}_o$, = upper bound of wage interval *r*

x_{ior} = number of workers in interval *r*

j = difference between the number of workers needed to reach the *p*-th percentile wage rate and the number of workers needed to reach the L_r wage rate

Annual wage rate estimates

These estimates are calculated by multiplying mean or percentile hourly wage rate 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 mean annual pay should the workers work more or less than 2,080 hours per year.

Alternatively, some workers are paid on an annual basis but do not work the usual 2,080 hours per year. For these workers, survey respondents report annual wages. Hourly wage rates cannot be derived from annual wage rates with any reasonable degree of confidence because the OES survey does not collect the actual number of hours worked. Only annual wages are reported for some occupations.

Occupational employment variance estimation

A subsample replication technique called the “jackknife random group” is used to estimate variances of occupational employment. In this technique, each sampled establishment is assigned to one of *G* random groups. *G* subsamples are created from the *G* random groups. Each subsample is reweighted to represent the universe.

G estimates of total occupational employment (u_{zo} (one estimate per subsample) are calculated. The variability among the *G* employment estimates is a good variance estimate for occupational employment. The two formulas below are used to estimate the variance of occupational employment for an estimation cell defined by geographic area and industry group.

c_{zr}

where

h = estimation cell defined by geographic area and industry group

j = employment size class (1-19, 20-49, 50-249, 250+)

o = occupation

$Mean\ Wage = \frac{\text{Total Interval Wages + Total Individual Wages}}{\text{Total Interval Employment + Total Individual Employment}}$ = estimated variance of $pR_o = L_r + \frac{j}{f_r}(U_r - L_r)$

G = number of random groups

pR_o = estimated employment of occupation *o* in cell *h* and size class *j*

pR_o = estimated employment of occupation o in cell *h*, size class *j*, and subsample *g*

L_r = estimated mean employment for occupation *o* in cell *h* and size class *j* based on the

G subsamples (Note: a finite population correction factor is applied to the terms pR_o and L_r)

The variance for an occupational employment estimate in cell *h* is obtained by the equation:

$$U_r$$

This sums the variances f_r across all size classes *j* in the cell.

Occupational mean wage variance estimates

Because the OES wage data are placed into intervals (grouped), the exact wage of each worker is not used. Therefore, some components of the wage variance are approximated using factors developed from NCS data. A Taylor Series Linearization technique is used to develop a variance estimator appropriate for OES mean wage estimates. The primary component of the mean wage variance, which accounts for the variability of the observed sample data, is estimated using the standard estimator of variance for a ratio estimate. This component is the first term in the formula given below:

$$\widehat{(X_{hjog})}$$

where

$v(\hat{X}_{hjo}) = \frac{\sum_{g=1}^{G}(\hat{X}_{hjog} - \hat{\bar{X}}_{hjo})^2}{G(G-1)}$ = estimated mean wage for occupation *o*

$v(\hat{X}_{hjo})$ = estimated variance of x_{io}

$\hat{X}_{ho}$ = estimated occupational employment for occupation *o*

h = stratum (area/industry/size class)

$\hat{X}_{hjo}$ = sampling fraction for occupation o in stratum *h*

$\hat{X}_{hjog}$ = number of sampled establishments that reported occupation *o* in stratum *h*

$\hat{\bar{X}}_{hjo}$ = six-panel combined sample weight for establishment *i*

BMF_i = final benchmark factor applied to establishment *i*

$v(\hat{X}_{ho}) = \sum_{j \in h} v(\hat{X}_{hjo})$ = $v(\hat{X}_{hjo})$ for occupation *o* in establishment *i*

$(\mathrm{BMF}_i = \mathrm{BMF}_1 * \mathrm{BMF}_2 * \mathrm{BMF}_3 * \mathrm{BMF}_4$ = estimated total occupational wage in establishment *i* for occupation *o*

$v(\hat{R}_o) = \left(\frac{1}{\hat{X}_o^2} \left(\sum_h \left[\frac{n_{ho}(1-f_{ho})}{n_{ho}-1} \right] (\sum_{i \in h} (\mathrm{BMF}_i w_i)^2 \ (q_{io} - \bar{q}_{ho})^2) \right) + \sum_r \theta_{or}^2 \sigma_{or}^2 + \frac{1}{\hat{X}_o^2} \sum_r (\sum_{i=1}^{n_r} (\mathrm{BMF}_i w_i x_{iorr})^2) \sigma_{or}^2 + \frac{1}{\hat{X}_o^2} \sum_r \theta_{or} \sigma_{orr}^2 \right)$ = reported employment in establishment *i* for occupation *o*

$v(\hat{R}_o) = \left(\frac{1}{\hat{X}_o^2} \left(\sum_h \left[\frac{n_{ho}(1-f_{ho})}{n_{ho}-1} \right] (\sum_{i \in h} (\mathrm{BMF}_i w_i)^2 \ (q_{io} - \bar{q}_{ho})^2) \right) + \sum_r \theta_{or}^2 \sigma_{or}^2 + \frac{1}{\hat{X}_o^2} \sum_r (\sum_{i=1}^{n_r} (\mathrm{BMF}_i w_i x_{iorr})^2) \sigma_{or}^2 + \frac{1}{\hat{X}_o^2} \sum_r \theta_{or} \sigma_{orr}^2 \right)$ = mean of the q_{io} quantities for occupation *o* in stratum *h*

$v(\hat{R}_o)$ = proportion of employment within interval *r* for occupation *o*

$v(\hat{R}_o) = \left(\frac{1}{\hat{X}_o^2} \left(\sum_h \left[\frac{n_{ho}(1-f_{ho})}{n_{ho}-1} \right] (\sum_{i \in h} (\mathrm{BMF}_i w_i)^2 \ (q_{io} - \bar{q}_{ho})^2) \right) + \sum_r \theta_{or}^2 \sigma_{or}^2 + \frac{1}{\hat{X}_o^2} \sum_r (\sum_{i=1}^{n_r} (\mathrm{BMF}_i w_i x_{iorr})^2) \sigma_{or}^2 + \frac{1}{\hat{X}_o^2} \sum_r \theta_{or} \sigma_{orr}^2 \right)$ = reported employment in establishment *i* within wage interval *r* for occupation *o*

n_{ho} = Within wage interval *r*, these are estimated using the NCS and, respectively, represent the variability of

the wage value imputed to each worker, the variability of wages across establishments, and the variability

of wages within establishments.

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.

Nonsampling error

This type of error is attributable to several causes, such as errors in the sampling frame; an inability to obtain information for all establishments in the sample; differences in respondents' interpretation of a survey question; an inability or unwillingness of the respondents to provide correct information; errors made in recording, coding, or processing the data; and 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 OES variances are adaptations of formulas appropriate for the survey design used.

The particular sample used in the OES survey is one of a large number 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:

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). To construct a 90-percent confidence interval, add and subtract 160 (1.6 times the standard error) from the estimate: (4,840; 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.

Last Modified Date: August 31, 2020

Presentation

The Occupational Employment Statistics (OES) program publishes cross-industry occupational data for the United States as a whole, for individual U.S. 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. OES publishes employment and wage estimates aggregated by typical entry-level education requirements and by occupations categorized as Science, Technology, Engineering and Math (STEM). OES 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.

OES data are updated on an annual basis. When updated estimates become available, a BLS news release makes an announcement featuring highlights from the data.

Uses

For many years, the OES survey has been a major source of detailed occupational employment data for the nation, states, and areas, and by industry at the national level. This survey provides information for many data users, including individuals and organizations engaged in planning vocational education programs, higher education programs, and employment and training programs. OES data also are used to prepare information for career counseling, for job placement activities performed at 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 the comparison of occupational composition across states and local areas, including analysis for economic development purposes. OES 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, OES 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.

OES data are used by several other BLS and government programs, such as the BLS Employment Projections program and the Employment and Training Administration (ETA). OES 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. OES data also are used by the Department of Labor's Foreign Labor Certification (FLC) program, which sets the rate at which workers on work visas in the United States must be paid.

Employment and wage data for detailed science, engineering, mathematical, and other occupations are provided to the National Science Foundation, along with the complete staffing patterns for all industries.

Occupational wage data are used by jobseekers and employers to determine salary ranges for different occupations in different locations and in different industries. OES employment and wage data also can be found in ETA's CareerOneStop.

Many users of OES data use data provided by the State Labor Market Information programs. OES 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 average wages. OES is frequently cited as the most popular labor market information program within the United States.

Finally, employment and wage data are used by academic and government researchers to study labor markets and wage and employment trends. These data inform the so-called "good-jobs/bad-jobs" debate on how business cycles and structural economic change affect wages and employment across the range of occupations; and how many and what types of jobs are affected by off-shore outsourcing.

Data correction

If an error is found in a published OES 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 special notice describing the error is posted on the OES website, and data users who have signed up to receive notifications from the OES program are alerted via email.

Accessing OES data

OES data are available in several formats from the OES home page at www.bls.gov/oes/. The OES database search tool (www.bls.gov/oes/data.htm) allows customers to create customized HTML or Excel tables using the most recent OES estimates. OES data are also published as HTML tables or can be downloaded as zipped XLS files at www.bls.gov/oes/tables.htm.The research dataset of OES estimates by state and industry is available at www.bls.gov/oes/current/oes_research_estimates.htm. BLS does not publish OES estimates by metropolitan/nonmetropolitan area and industry, but these data may be available from individual state workforce agencies: www.bls.gov/bls/ofolist.htm.

For additional information, contact the OES staff at (202) 691-6569 or send e-mail to oesinfo@bls.gov.

Last Modified Date: August 31, 2020

History

Key developments

- **1977:** 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
- **2010–2012:** OES program 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
- **2017-19:** OES sample reduced
- **2018:** OES reduces some geographic detail
- **2019:** OES begins implementing the 2018 SOC

The OES program in its current form dates to 1996 and began publishing occupational employment and wage estimates in 1997. Since 1997, the OES data have undergone a number of changes, including changes to the occupational and industry classification systems used, changes to the metropolitan and nonmetropolitan area definitions, and changes to the sample size and survey reference dates.

Changes in occupational classification

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

Between May 2010 and May 2012, the OES program transitioned to the 2010 SOC. Because each set of OES estimates is produced by combining three years of survey data, 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. Therefore, these estimates used a hybrid of the 2000 and 2010 systems that included some OES-specific combinations of occupations. The May 2012 estimates were the first set of estimates based fully on the 2010 SOC. More information about the hybrid system used in the May 2010 and May 2011 estimates is available in the OES frequently asked questions.

Beginning with the May 2017 estimates, the OES program replaced 21 SOC detailed occupations with SOC broad occupations or OES-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. More information about these aggregations is available at www.bls.gov/oes/changes_2017.htm.

The OES program began implementing the 2018 SOC with the May 2019 estimates. Because of the OES 3-year methodology, the May 2019 and May 2020 estimates will use a hybrid of the 2010 and 2018 SOCs that includes some combinations of occupations that are not found in either version of the system. The May 2021 estimates will be the first estimates based entirely on survey data collected using the 2018 SOC. More information on the hybrid classification system used in the May 2019 estimates is available on the OES 2018 SOC implementation page and in the frequently asked questions.

The May 2019 estimates also introduced some new occupational aggregations designed to improve data quality, along with changes to some of the occupational aggregations introduced in May 2017.

Changes in industry classification and survey scope

The 1997–2001 OES estimates used the Standard Industrial Classification (SIC) system. In 2002, the OES 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), and May 2017 estimates (2017 NAICS).

Beginning with the May 2014 estimates, gambling establishments and casino hotels owned by local governments were moved from the OES 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 OES 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 OES program uses standard metropolitan area definitions from the U.S. Office of Management and Budget (OMB). For the New England states, OES uses the New England City and Town Area (NECTA) definitions rather than the Metropolitan Statistical Area (MSA) definitions. The OES nonmetropolitan areas use definitions that are specific to the OES program and are developed in cooperation with the state workforce agencies.

The OES program implemented major revisions to the area definitions in the May 2005 and May 2015 estimates. The May 2005 estimates introduced revised OMB area definitions based on the results of the 2000 decennial census. The May 2015 estimates introduced revised definitions based on the 2010 census. In addition to the major revisions in May 2005 and May 2015, smaller revisions were implemented in other years. Because the OES 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.

With the May 2018 estimates, the OES 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, OES no longer publishes data for the divisions. Data for these 11 areas are now available at the Metropolitan Statistical Area (MSA) or New England City and Town Area (NECTA) level only. In addition, some smaller nonmetropolitan areas were combined to form larger nonmetropolitan areas. More information on these area changes is available at www.bls.gov/oes/areas_2018.htm.

Changes to sample size and reference period

Before 2002, the OES 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, OES 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 OES program also published estimates semiannually in 2003 and 2004. In 2005, the OES program returned to publishing data annually, but retained semiannual data collection.

The OES sample has been reduced in recent survey panels. The May 2019 OES survey panel had a sample of approximately 183,000 establishments. The November 2017, May 2018, and November 2018 survey panels each had a sample of approximately 186,000 establishments. The May 2017 panel sample consisted of approximately 195,000 establishments.

Data before 1997

Data from the immediate predecessor to the current OES program are available at the bottom of the main OES data page. These data cover the period 1988–1995 and are not directly comparable to more recent OES data. The 1988-1995 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.

Last Modified Date: August 31, 2020

More Information

Further information on the Occupational Employment Statistics (OES) program can be found through the OES website: www.bls.gov/oes.

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

- Main OES data page: www.bls.gov/oes/tables.htm
- Query tool (contains only the most recent OES data): www.bls.gov/oes/data.htm
- OES FAQs page: www.bls.gov/oes/oes_ques.htm
- Current area definitions: www.bls.gov/oes/current/msa_def.htm
- Standard Occupational Classification (SOC) homepage: www.bls.gov/soc/home.htm
- Archived documentation: www.bls.gov/oes/oes_doc_arch.htm
- OES publications: https://www.bls.gov/oes/publications.htm

Model-based estimation methodology (MB3)

Economists in the BLS Office of Employment Research and Program Development have developed an alternative model-based estimation methodology (MB3) for the Occupational Employment Statistics program. Research estimates based on the MB3 methodology are available to the public for review and comment. The OES program is considering changing the estimation methodology to MB3 because of advantages over the existing methodology, as described in "Model-based estimates for the Occupational Employment Statistics program."

Last Modified Date: August 31, 2020