Appendix B. Survey Methods and Reliability of the 1999 Occupational Employment Statistics Estimates

The Occupational Employment Statistics (OES) survey samples approximately 400,000 establishments each year and, over a 3-year period, contacts approximately 1.2 million establishments. Each single-year sample represents one-third of both the certainty and noncertainty strata for the full 3-year sample plan. While estimates can be made from a single year of data, the OES survey has been designed to produce estimates using a full 3 years of data. The sample allows the production of estimates at detailed area, industry, and occupational levels. Estimates using any one year of data are subject to a higher sampling error (due to the smaller sample size) and the limitations associated with including reports from only one-third of the certainty units.

Occupational and industrial classification

New occupational classification standards for 1999. In 1999, the OES survey began using the new Standard Occupational Classification System (SOC) authorized by the U.S. Office of Management and Budget (OMB). (See appendix A for a detailed description of the SOC.) The SOC system is the first OMB-mandated occupational classification system for Federal agencies. The OES survey uses 22 major occupational groups from the SOC to categorize workers in 1 of almost 770 detailed occupations. Data for previous years were "crosswalked" to the new classification system when possible, and used in producing wage estimates for these occupations. Wages for 374 of the matched occupations are estimated using data from the 1997, 1998, and 1999 surveys. The remaining occupations either are new SOC occupations or are slightly different from similar occupations in the old OES structure; wages for these occupations are estimated from a single year of data only. In order to maintain employment additivity, all occupational employment estimates are based only on the data collected in the 1999 survey.

The industrial classification system. The industrial classification system used in this survey is described in the *1987 Standard Industrial Classification Manual* (SIC), wherein reporting establishments are classified into industries on the basis of major product or activity. The OES program produces estimates by both two-digit and three-digit SIC codes and across all industries.

Scope of the survey

Occupational employment data by wage interval are used to produce the 1999 national, State, and area occupational employment and wage estimates by industry. This is the fourth year for which the OES program has collected both occupational employment and wage data for all nonfarm industries, except private households. The survey covers establishments in SIC codes 07, 10 through 42, 44 through 87, 89, and State and local governments. In addition, data for the U.S. Postal Service and for the Federal Government are universe counts obtained from the U.S. Office of Personnel Management. Occupational employment and wage estimates at the national level were produced by the Bureau of Labor Statistics (BLS) using employment and wage data from the 50 States and the District of Columbia. Guam, Puerto Rico, and the Virgin Islands were surveyed; however, data from these territories are not included in the production of national estimates.

The OES survey requests that employers provide occupational data for a particular reference date. The reference date of the 1999 survey is the pay period that included October 12th, November 12th, or December 12th of 1999. The pay period including the 12th day of the reference month is standard for Federal agencies collecting employment data. The reference date for any particular establishment in this survey was dependent on its SIC code. (See table below.)

Reference date	*SIC codes of industries surveyed*
October 12	07, 15, 16, 17, 241, 472, 50, 51, 52, 53, 541, 542, 543, 545, 546, 549, 55, 56, 57, 58, 59, 60, 61, 62, 637, 655, 672, 673, 679, 70, 722, 731, 732, 733, 734, 736, 738, 792, 793, 794, 799, and 84.
November 12	26, 27, 28, 29, 351, 352, 353, 354, 355, 356, 358, 359, 37, 386, 40, 41, 42, 44, 45, 46, 473, 474, 478, 48, 631, 632, 633, 635, 636, 639, 64, 651, 653, 654, 671, 735, 737, 751, 753, 754, 76, 78, 80, 81, 83, 86, 87, and 89.
December 12	10, 12, 13, 14, 20, 21, 22, 23, 242, 243, 244, 245, 249, 25, 30, 31, 32, 33, 34, 357, 36, 381, 382, 384, 385, 387, 39, 49, 544, 721, 723, 724, 725, 726, 729, 752, 791, 82, and State and local governments.

The employment estimates have been adjusted to the full universe counts of the 1999 survey reference period based on the Covered Employment and Wages program. The 1997 and 1998 wage data have been adjusted to the 1999 reference period by using the national over-the-year fourth-quarter rate of change in wages for nine major occupational groups obtained from the BLS national Employment Cost Index.

Concepts

An *establishment* is an economic unit that produces goods or services. It generally is found at a single physical location and is engaged predominantly in one type of economic activity. Where a single physical location encompasses two or more distinct activities, these are treated as separate establishments if separate payroll records are available and certain other criteria are met.

Employment includes full- and part-time workers; workers on paid vacations or other types of leave; workers on unpaid short-term absences (such as those due to illness, bad weather, temporary layoff, or jury duty); salaried officers, executives, and staff of incorporated firms; employees temporarily assigned to other units; and employees for whom the unit is their permanent (home) duty station, regardless of whether the unit prepares their paycheck. Among those excluded from coverage are most proprietors (owners and partners of unincorporated firms), self-employed workers, and unpaid family workers.

Occupation refers to the occupation in which employees are working, rather than the occupation for which they may have been trained. 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 in which 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 performed by workers under their supervision) are reported in the occupation most closely related to their work.

Part-time workers, workers receiving on-the-job training, and apprentices are reported in the occupation in which they ordinarily work.

A *wage* is money that is paid or received for work or services performed in a specified period. Included in wages for this survey are: Base rate; cost-of-living allowance; guaranteed pay; hazardous duty pay; incentive pay, including commissions; piece rates; production bonuses; length-of-service allowance (longevity pay); oncall pay; portal-to-portal pay; and tips. Not included are: Backpay; overtime pay; severance pay; shift differentials; jury-duty pay; vacation pay; premium pay for work on holidays or weekends; attendance bonuses; holiday bonuses; meal and lodging allowances; merchandise discounts; nonproduction bonuses; profit-sharing distributions; relocation allowances; stock bonuses; tool allowances; tuition reimbursements; or uniform allowances.

Survey procedures

The survey is based on a probability sample, stratified by area, industry, and size of establishment, and is designed to represent the total or "universe" of establishments covered by the survey. The survey is conducted over a 3-year cycle. Each year, one-third of the sample units are included in the survey. To the extent possible, units selected in 1 year are not included in the sample the following 2 years.

Employers are asked to classify each of their workers in an occupation and wage range. There are 12 wage ranges, on both an hourly and annual basis, as follows:

Interval	Wages	
	Hourly	Annual
Range A	Under $6.75	Under $14,040
Range B	$6.75 to $8.49	$14,040 to $17,679
Range C	$8.50 to $10.74	$17,680 to $22,359
Range D	$10.75 to $13.49	$22,360 to $28,079
Range E	$13.50 to $16.99	$28,080 to $35,359
Range F	$17.00 to $21.49	$35,360 to $44,719
Range G	$21.50 to $27.24	$44,720 to $56,679
Range H	$27.25 to $34.49	$56,680 to $71,759
Range I	$34.50 to $43.74	$71,760 to $90,999
Range J	$43.75 to $55.49	$91,000 to $115,439
Range K	$55.50 to $69.99	$115,440 to $145,599
Range L	$70.00 and over	$145,600 and over

Method of collection

Survey schedules are initially mailed out to almost all sampled establishments; personal visits are made to some of the larger establishments.

Two additional mailings are sent to nonrespondents at approximately 3-week intervals. Telephone or personal visit followups are made for those nonrespondents considered critical to the survey because of their size.

Sampling procedures

The sampling frame for this survey was the establishments in the two- and three-digit SIC codes listed above that reported to the State Employment Security Agencies for Unemployment Insurance (UI) purposes. Each quarter, BLS combines the lists from all States into a single file, called the Longitudinal Database (LDB). The LDB is a compilation of State unemployment insurance reports. Virtually all businesses are required to file such a report within the State in which they are located. For the 1997 survey, the sample frame was the LDB file from the third quarter of 1996; for the 1998 survey, it was from the second quarter of 1997; and for the 1999 survey, it was from the second quarter of 1998. This frame was supplemented with a list supplying establishment information on railroads (SIC 401).

A census is taken of Federal Government establishments

each year. Data representing Federal Government employment and wages are obtained from the Office of Personnel Management at the end of the survey process.

Within each State, establishments in the universe were stratified by Metropolitan Statistical Area (MSA), three-digit SIC code, and size of firm. An establishment's size class is determined by its employment as reported on the sampling frame. Establishments in smaller size classes were selected based on a probability sample. Establishments in larger size classes are sampled with virtual certainty across the 3-year cycle of the survey. The targeted sample size of 1.2 million establishments per 3-year cycle was allocated in a manner that equalized the expected relative standard error of the typical occupational employment within the cell for each MSA and three-digit SIC. Within each of these cells, the sample was allocated across size classes in a manner that minimized the variance of the average typical occupational employment estimate.

The OES survey uses permanent random numbers (PRNs) in its sample selection methodology. The purpose of the PRN is to limit, to the extent possible, overlap of respondents between the OES survey and other BLS surveys. These numbers are placed on the frame and are retained by establishments across time. A sample selection using PRNs can be done in several ways. For example, a range of PRNs can be used to select a portion of the universe within each stratum. Alternatively, a specific PRN value can be used as a "start" point within a stratum. Within a stratum sorted by PRN value, n_h establishments are selected sequentially, beginning with this "start" point (where n_h is the number of sample units allocated to stratum h). This latter method is the one used for the OES sample selection. In the OES sample selection, a stratum is defined by State/MSA/three-digit SIC/employment size class. Approximately one-third of the allocated units are selected within each MSA/SIC/size class each year.

The above allocation method resulted in initial sample sizes of 408,805, 400,404, and 402,636 establishments for 1997, 1998, and 1999, respectively, for a combined initial sample size of 1,199,393 establishments. Note that the sum of samples across the 3 years does not equal the combined sample size because only the current year's Federal and State government establishments are included.

Response

Of the 383,861 eligible units from the 1997 sample, usable responses were obtained from 301,671, producing a response rate of 78.6 percent based on units. Of the 363,267 eligible units from the 1998 sample, usable responses were obtained from 284,159, producing a response rate of 78.2 percent based on units. Of the 369,694 eligible units from the 1999 sample, usable responses were obtained from 286,903, producing a response rate of 77.6 percent based on units.

Nonresponse

Nonresponding establishments are accounted for in the OES survey by a two-step imputation process. First, the staffing pattern is imputed using a "hot-deck," "nearest-neighbor" imputation method. "Hot-deck" procedures utilize data from the current period to impute for missing data (from the current period). The "nearest-neighbor" method searches the responding establishments within a defined cell and finds the responding establishment that most closely matches the nonresponding establishment for key classification values (area/SIC/size class). The staffing pattern, or employment distribution, of the responding establishment is used as the staffing pattern of the nonresponding establishment. The second step is to impute a wage distribution for each occupation of the imputed staffing pattern. This imputation procedure replaces the missing data by determining the distribution of the reported occupational wage data across wage intervals in the current area/SIC/size class. If there are sufficient data at this level, the procedure uses this reported wage distribution to allocate the nonrespondent's imputed occupational employment across the wage intervals. If there are not enough data, the pool of donors is expanded to include adjacent size classes, industries, and areas until a distribution can be determined.

Occasionally, a responding establishment may provide employment information but refuse to provide wage distribution information for selected occupations. The OES survey uses the "distribution within a cell" procedure described above to impute the missing data for this partial report.

Combining and benchmarking multiyear data

Whenever possible, data from the 1997, 1998, and 1999 surveys have been combined. Survey data from 1997, 1998, and 1999 were used to produce the wage estimates for 374 occupations. The remaining occupational wage estimates and all of the employment estimates were produced using only 1999 data. Each year's sample is weighted to represent the sample as it appeared at the time the sample was selected. In order to combine the data, each unit's weight is modified so that the aggregate sample represents the universe. This is done via a fairly simple procedure. Each unit's weight is divided by the number of years for which sample units were selected for that stratum.

A ratio estimator is used to develop estimates of occupational employment. The auxiliary variable used was the 1999 reference-month population value of total employment. In order to balance the States' needs for estimates at different levels of geographic and industrial aggregation, the ratio adjustment process was applied as a hierarchical series of ratio adjustment, or "benchmark," factors.

The primary component of this procedure is a ratio adjustment at the State, MSA, three-digit SIC, employment size class level. If these ratio adjustment values are out of range, they are set at predetermined maximum or minimum values. This adjustment can be described as follows**:**

Define:

h = State/MSA/three-digit SIC
H = State/three-digit SIC

s = 1 of 4 employment size classes {1-19, 20-49, 50 - 249, 250+}

S = 1 of 2 aggregate employment size classes {1- 49, 50+}

M = 1999 reference month population value of total employment

i = establishment

w_i = adjusted sample weight for establishment i

p_i = total establishment employment

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

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

$$\beta_{hs} = \left(M_{hs} \Big/ \sum_{i \in hs} w_i p_i\right), \quad \beta_{hS} = \left(M_{hS} \Big/ \sum_{i \in hS} w_i p_i\right), \quad \beta_h = \left(M_h \Big/ \sum_{i \in h} w_i p_i\right), \text{ then}$$

$$BMF_{1,\,hs} = \begin{cases} \beta_{hs}, \text{ if all } \beta_{hs} \text{ within h are bounded by } (BMF_{min}, BMF_{max}), \\ \beta_{hS}, \text{ if all } \beta_{hS} \text{ within h are bounded by } (BMF_{min}, BMF_{max}), \\ BMF_{min}, \text{ if } \beta_h < BMF_{min}, \\ BMF_{max}, \text{ if } \beta_h > BMF_{max}, \\ \beta_h \text{ otherwise} \end{cases}$$

The next component in the procedure is a ratio adjustment at the State, three-digit SIC level using the product of the adjusted sampling weight and the first ratio adjustment as a final weight value. If these ratio adjustment values are out of range, they are set at predetermined maximum or minimum values. This ratio adjustment accounts for weighted, ratio-adjusted sample employment that does not adequately represent the universe within one or more of the State, MSA, three-digit SIC strata. This adjustment is calculated as follows:

Define:

$$\beta_H = \left(M_H \Big/ \sum_{hs \in H} \sum_{i \in hs} w_i p_i BMF_{1,\,hs}\right), \quad \text{then}$$

$$BMF_{2,\,H} = \begin{cases} BMF_{min}, \text{ if } \beta_H < BMF_{min}, \\ BMF_{max}, \text{ if } \beta_H > BMF_{max}, \\ \beta_H \text{ otherwise} \end{cases}$$

The procedure then calculates a ratio adjustment at the State, two-digit SIC level using the product of the adjusted sampling weight, the first ratio adjustment, and the second ratio adjustment as a final weight value. If these ratio adjustment values are out of range, they are set at predetermined maximum or minimum values. This ratio adjustment accounts for weighted, ratio-adjusted sample employment that does not adequately represent the universe within one or more of the State, three-digit SIC strata. This adjustment is calculated similarly to $BMF_{2,H}$.

Finally, the procedure calculates a ratio adjustment at the State, industry-division level using the product of the adjusted sampling weight, the first ratio adjustment, the second ratio adjustment, and the third ratio adjustment as a final weight value. If these ratio adjustment values are out of range, they are set at predetermined maximum or minimum values. This ratio adjustment accounts for weighted, ratio-adjusted sample employment that does not adequately represent the universe within one or more of the State, two-digit SIC strata. This adjustment also is calculated similarly to $BMF_{2,H}$.

A final ratio adjustment factor, BMF_k, is calculated as the product of the four hierarchical ratio adjustment factors. That is, $BMF_k = BMF_1 * BMF_2 * BMF_3 * BMF_4$. A final weight value is then calculated as the product of the adjusted sample weight and the final ratio adjustment factor. Note that the population values of total employment (M_{hs}) are obtained from the BLS Longitudinal Data Base (LDB) file.

Estimation methodology

Producing estimates using the 3 years of sample data provides additional occupational detail and sampling error reductions (particularly for small geographic areas and occupations). However, this procedure also has some quality limitations because it requires the adjustment of earlier years' data to the current reference period—a procedure referred to as "wage updating."

The 1997 OES survey estimates were from the second year of OES estimates and were developed using both the 1996 and 1997 surveys. The 1997 estimates also were the first to be developed using a "wage-updating" methodology. In addition to the wage-updating procedure, the 1997 estimates used an improved estimation methodology, which uses a "nearest neighbor" imputation approach for nonrespondents and applies employment benchmarks at a detailed MSA by three-digit industry and broad size class level. A variant of the imputation procedure also is used to account for item nonresponse. It should be noted that, because of the difference in estimation methods for these first 2 years of OES estimates, the estimates for 1997 are not strictly comparable with those published for 1996.

The 1998 OES survey estimates are developed from the full 3 years of the OES sample. The combined 1996, 1997, and 1998 data cover approximately 1.2 million sample units. The 1998 estimates employ the wage-updating methodology introduced in 1997, which uses the over-the-year fourth-quarter rate of change in wages for nine major occupational groups from the BLS Employment Cost Index to adjust prior years' wage data to the current year's reference period. In addition, the 1998 estimates employ the imputation methodology introduced in 1997, which uses a "nearest neighbor" approach for nonrespondents and applies employment benchmarks at a detailed MSA by three-digit industry and broad size class level.

The 1999 OES survey wage estimates for some occupations are developed from the full 3 years of the OES sample, while the remaining occupational wage estimates and all of the employment estimates are from 1 year of data. The

combined 1997, 1998, and 1999 data cover approximately 1.2 million sample units. The 1999 estimates also use the wage-updating and estimation methodology introduced in 1997.

Estimated employment

As mentioned above, a ratio estimator is used to develop estimates of occupational employment. The auxiliary variable is the population value of total employment obtained from the refined UI files for the 1999 reference month. Within each MSA, the estimated employment for an occupation at the reported three-digit SIC level was calculated by multiplying the weighted employment by its ratio factor. The estimated employment for an occupation at the all-industry level was obtained by summing the occupational employment estimates across all industries within an MSA reporting that occupation. The employment and wage data for Federal Government workers in each occupation were added to the survey-derived data.

First, within each MSA, the estimated employment for an occupation at the reported three-digit SIC i level was calculated using the following equation:

$$\hat{P}_{oi} = \sum_j \sum_k \left(w_{ijk} \, p_{oijk} \right) * BMF_k$$

where
- o = occupation
- i = reported three-digit SIC
- j = reported size class
- k = establishment
- w_{ijk} = adjusted sample weight for establishment k
- p_{oijk} = reported employment for occupation o in establishment k within SIC i and size class j
- $\hat{P}_{oi}$ = estimated employment for occupation o in SIC i

The estimated employment for an occupation at the all-industry level was obtained by summing the occupational employment estimate $\hat{P}_{oi}$ across all industries within an MSA that reported the occupation. See the formula below:

$$\hat{P}_o = \sum_{i=1}^{L_i} \hat{P}_{oi}$$

where L_i is the number of industries reporting that occupation within the MSA.

Estimated wage rates

Occupational wage data in the OES survey are collected as the number of workers in an occupation who are paid wages within each of 12 contiguous wage intervals. For example, an establishment may report that it employs 10 secretaries: 2 in wage interval B, paid wages between $6.75 and $8.49 per hour; 6 in wage interval D, paid wages between $10.75 and $13.49 per hour; and 2 in wage interval E, paid wages between $13.50 and $16.99 per hour. As a result, individual wage rates of workers are not collected. Conventional arithmetic mean formulas are not applicable in this situation. Because wage data are collected within an interval matrix, the particular wage rate of all employees within an interval is approximated by a mean wage rate value for the interval for each of the first 11 wage intervals. Data from the BLS National Compensation Survey (NCS) are used to calculate these mean wage rate values. The mean wage value for the upper open-ended wage interval was set at that interval's starting point. Occupational wage rates are calculated by summing a weighted estimate of total occupational wages, and dividing that by a weighted estimate of total occupational employment ($\hat{P}_o$).

Wage updating process. Because data from 3 years were used to produce the 1999 OES wage estimates for some occupations, a process was used to update prior year information so that it would be representative of the 1999 reference period. This was done by adjusting most 1997 and 1998 wage data by a factor developed from the BLS Employment Cost Index (ECI) program. The ECI program provides a rate of change in wages from fourth-quarter 1997 to fourth-quarter 1999 and from fourth-quarter 1998 to fourth-quarter 1999 for nine major occupational groups. Each OES occupation belongs to one of these major occupational groups. These rates of change were used to update the 1997 and 1998 OES data to the 1999 reference period.

Estimated mean wage rate. Mean wage is the estimated total wages for an occupation, divided by the occupation's weighted survey employment. An estimate of the mean wage rate was calculated using a standard interval-based estimation formula, modified to account for the wage-updating process. See the formula below:

$$\hat{R}_o = \frac{\sum_{yr} \sum_{k \in yr} w_k \hat{y}_{k\,o}}{\hat{P}_o} \;, \; where \;\; \hat{y}_{k\,o} = uf_{yr,o} \sum_r p_{k,\,o,\,r} \, c_{yr,\,r} \;; (k \in yr)$$

where
- o = occupation
- w_k = weight for establishment k
- yr = year
- r = wage interval
- $P_{k,\,o,\,r}$ = reported employment for occupation o in establishment k in wage interval r. (Note that establishment k reported data for 1 year (yr))
- $\hat{y}_{k\,o}$ = unweighted total wage estimate for occupation o in establishment k

$\hat{P}_o$ = estimated employment for occupation o

$uf_{yr,o}$ = ECI updating factor for year yr and occupation o

$c_{yr,r}$ = see below

In this formula, $c_{yr,r}$ represents the mean of interval r for year yr. This mean was determined empirically using data from the BLS NCS survey. Research is conducted at periodic intervals to verify the continued utility of this updating procedure.

Median wage. The median wage is the estimated 50th percentile of the distribution of wages; 50 percent of workers in an occupation earn wages below, and 50 percent earn wages above, the median wage. The wage interval containing the median wage is located using a cumulative frequency count of employment across wage intervals. After the targeted wage interval is identified, the median wage rate is then estimated by a linear interpolation procedure.

Variance of estimates

Estimates of sampling error are calculated to allow the users to determine if occupational employment estimates are reliable enough for their needs. Only a probability-based sample can be used to calculate estimates of sampling error from the sample itself.

The formula used to estimate variances (a common measure of sampling error) is based on the survey's sample design and method of estimation. The OES survey used a subsample replication technique called the "jackknife random group" to estimate variances of occupational employment. In this technique, each sampled establishment is assigned to one of G random groups. Using the data in these groups, G subsamples are formed from the parent sample. Next, G estimates of total occupational employment ($\hat{P}_{ijg}$) are calculated, one employment estimate per subsample. Afterwards, the variability of these G employment estimates is calculated to obtain the estimated occupational employment variance.

The occupational employment variance estimate at the reported three-digit SIC i/reported size class j level is calculated using the following equation:

$$s^2(\hat{P}_{ij}) = \left[\sum_{g=1}^{G} (\hat{P}_{ijg} - \overline{\hat{P}}_{ij})^2\right] \Big/ [G(G-1)]$$

where $s^2(\hat{P}_{ij})$ = estimated variance of $\hat{P}_{ij}$

G = number of random groups

$\hat{P}_{ij}$ = estimated occupational employment in SIC i and size class j

$\hat{P}_{ijg}$ = estimated occupational employment in SIC i, size class j, and subsample g

$\overline{\hat{P}}_{ij}$ = estimated mean occupational employment in SIC i and size class j based on the G subsamples

(Note: a finite population correction factor is applied to the terms $\hat{P}_{ijg}$ and $\overline{\hat{P}}_{ij}$.)

The variance for an occupational employment estimate at the reported three-digit SIC i level was obtained by summing the variance $s^2(\hat{P}_{ij})$ across all reported size classes j in SIC i.

Similarly, the variance for an occupational employment estimate at the reported two-digit SIC h level is obtained by summing the variance $s^2(\hat{P}_i)$ across all reported three-digit SICs i within SIC h.

Reliability of the estimates

Estimates developed from a sample may differ from the results of a census. Two types of error, sampling and nonsampling, can occur in estimates calculated from a sample. *Sampling error* occurs because our observations are based on a sample, not on the entire population. *Nonsampling error* occurs because of response and operational errors in the survey. Unlike sampling error, this form of error can also occur in a census.

Sampling errors

The particular sample used in this 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. Estimates derived from different samples would tend to differ from one another. As indicated above, the variance of a survey estimate is a measure of the variation among the estimates from all possible samples. The standard error of a survey estimate is the square root of its variance; the relative standard error is the ratio of the standard error to the estimate itself.

The sample estimate and its standard error allow the user to construct an interval estimate with a prescribed level of confidence that the interval will include the mean value of the estimate from all possible samples.

To illustrate, if all possible samples were selected, each of them surveyed under essentially the same conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from 1 standard error below to 1 standard error above the derived estimate would include the average value of the estimates from all possible samples. 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 derived estimate would include the average value of the estimates from all possible samples. 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 derived estimate would include the average value of the estimates from all possible samples. 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 derived estimate would include the average value of the estimates from all possible samples.

For example, suppose that an estimated occupational employment total is 5,000, with an associated relative standard error of 2.0 percent. Based on these data, the standard error of the estimate is 100 (2 percent of 5,000). A 68-percent confidence interval for the employment estimate is (5,000 +/- 100) or from 4,900 to 5,100. Approximately 68 percent of the intervals constructed in this manner will include the mean of all possible employment estimates as computed from all possible samples. A 95-percent confidence interval for the employment estimate is (5,000 +/- 200) or from 4,800 to 5,200. Approximately 95 percent of the intervals constructed in this manner will include the mean of all possible employment estimates as computed from all possible samples. Estimates of sampling errors for occupational employment estimates are provided with this publication.

Nonsampling error

This type of error is attributable to several causes, such as: An inability to obtain information for all establishments in the sample; differences in the respondents' interpretation of the 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.

Several edit and quality control procedures are used to reduce nonsampling error. For example, completed survey questionnaires are checked for data consistency. Followup mailings are sent out to nonresponding establishments to improve the survey response rate. Response analysis studies are conducted to assess the respondents' comprehension of the questionnaire. (See the section below for additional information on the quality control procedures used by the OES survey.) The relative standard error indicates the magnitude of the sampling error. It does not 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.

Quality control measures

The OES survey is a Federal-State cooperative effort that enables States to conduct their own surveys. A major concern with a cooperative program like OES is to accommodate 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, and the U.S. territories. Controlling sources of nonsampling error in this decentralized environment can be difficult. One important computerized quality control measure used by the OES survey is the Survey Processing and Management (SPAM) 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 on the data files, as well as the time needed to review them.

Other quality control measures used in the OES survey include:

- Followup solicitations of nonrespondents (especially critical nonrespondents)
- Review of schedules to verify the accuracy and reasonableness of the reported data
- Adjustments for atypical reporting units on the data file
- Validation of the benchmark employment figures and of the benchmark factors
- Validation of the analytical tables of estimates (at the two- and three-digit SIC levels)

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 imputed from the published figures. Each published estimate is screened to ensure that it meets these confidentiality requirements. The specific screening criteria are not listed in this publication to further protect the confidentiality of the data.

Appendix C. Availability of Historical Occupational Employment Statistics Survey Data Nationally and from State Agencies

The Occupational Employment Statistics (OES) survey first collected employment and wage data in 1996 to produce occupational estimates for all industries and also to produce occupational wage data for all States. Prior to 1996, the OES survey collected data from specified industries in 1 of 3 years in the survey round, as indicated in the table below. Now, the survey sample is designed to collect data from establishments in all industries each year.

States also produce occupational estimates by industry. Prior to 1983, not all States participated in the OES program in all survey years. Starting with the 1991 OES survey, certain States also collected wage data. In 1996, all States began collecting wage data. Check with the State Employment Security Agencies listed on the inside back cover of this publication regarding the availability of State data on occupational employment and wages.

Industry	*1987 SIC code*	*Years collected*
Agricultural services	07	1992, 1995
Mining	10-14	1978, 1981, 1984, 1987, 1990, 1993
Construction	15-17	1978, 1981, 1984, 1987, 1990, 1993
Manufacturing	20-39	1977, 1980, 1983, 1986, 1989, 1992, 1995
Transportation and public utilities	40-49	1979, 1982, 1985, 1988, 1991, 1994
Wholesale trade	50-51	1979, 1982, 1985, 1988, 1991, 1994
Retail trade	52-59	1979, 1982, 1985, 1988, 1991, 1994
Finance, insurance, and real estate	60-67	1978, 1981, 1984, 1987, 1990, 1993
Services	70-87, 89	1978, 1981, 1984, 1987, 1990, 1993
Hospitals	806	1980, 1983, 1986, 1989, 1992, 1995
Educational services	82	1978, 1981, 1985, 1988, 1991, 1994
State government		1979, 1982, 1985, 1988, 1991, 1994
Local government		1979, 1982, 1985, 1988, 1991, 1994