Appendix B. Survey Methods and Reliability of the 2000 Occupational Employment Statistics Estimates

The Occupational Employment Statistics (OES) program 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. (Certainty strata consist of establishments that are included in the sample because of their large employment size.) While estimates can be made from data for 1 or 2 years, the OES survey has been designed to produce estimates using a full 3 years' worth of data. The sample allows the production of estimates for 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 to the limitations associated with being based on data from only one-third of the certainty units. Estimates from the 2000 survey are based on only 2 years' worth of data because of the conversion to the Standard Occupational Classification (SOC) system in 1999.

Occupational and industrial classification

The occupational classification system. In 1999, the OES survey began using the U.S. Office of Management and Budget's (OMB) new occupational classification system—the **Standard Occupational Classification System** (SOC). (See appendix A for a detailed description of the SOC.) The SOC is the first OMB-required occupational classification system for Federal statistical agencies. The OES survey uses 22 major occupational groups from the SOC to categorize workers in 1 of almost 770 detailed occupations.

The industrial classification system. The industrial classification system used in this survey is described in the *1987 Standard Industrial Classification Manual* (SIC), whereby 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 2000 national, State, and area occupational employment and wage estimates by industry. This is the fifth 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, and 89, and State and local governments. In addition, data for the U.S. Postal Service and 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.

Employers participating in the OES survey are asked to provide occupational data for a particular reference period. The reference period of the 2000 survey is the pay period that included October 12th, November 12th, or December 12th of the year. The pay period including the 12th day of the reference month is standard for Federal agencies collecting employment data. The reference period for any particular establishment in this survey was dependent on the establishment's 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 2000 survey reference period based on information from the Covered Employment and Wages program. The 1999 wage data have been adjusted to the 2000 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 reporting 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 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 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 questionnaires, or "schedules," initially are mailed out to almost all sampled establishments; OES State personnel make personal visits 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 list of 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, the lists from all States are compiled into a single file at BLS. This comprehensive file is called the Longitudinal Database (LDB), and is a compilation of State unemployment insurance reports. Virtually all businesses are required to file these reports within the State in which they are located. For the 1999 survey, the sample frame was the LDB file from the second quarter of 1998; for the 2000 survey, it was the file from the second quarter of 1999. 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 employ-

ment and wages are obtained at the end of the survey process from the U.S. Office of Personnel Management.

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 using a probability-based sampling procedure. Establishments in the largest size class 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 each MSA/three-digit SIC cell. 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 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. For purposes of the 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 stratum each year.

The above allocation method resulted in initial sample sizes of 402,636 and 406,876 establishments for 1999 and 2000, respectively, for a combined initial sample size of 805,037 establishments. Note that the sum of samples across the 2 years does not equal the combined sample size because only the current year's State and Federal Government establishments are included.

Response

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. Of the 375,387 eligible units from the 2000 sample, usable responses were obtained from 293,450, producing a response rate of 78.2 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 provides employment information, but refuses 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

Survey data from 1999 and 2000 were used to produce the wage and employment estimates for almost all of the occupations. The exception is Physicians and Surgeons, All Other, for which estimates were produced using only the 2000 data. Each year's sample is weighted to represent the universe 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 2000 reference-month population value of total employment. In order to balance the States' need 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 = 2000 reference month population value of total employment

i = establishment

w_i = adjusted 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, 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:

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

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

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, 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 sample data for 3 years 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 1999 OES survey wage estimates for some occupations were developed from data for the full 3 years of the OES sample, while the remaining occupational wage estimates and all of the employment estimates were from 1 year's worth of data. The combined 1997, 1998, and 1999 data were obtained from approximately 1.2 million sample units. The 1999 estimates were derived using the wage-updating and estimation methodology introduced in 1997, which employs a "nearest neighbor" approach for nonrespondents and applies employment benchmarks at a detailed MSA/three-digit industry/broad size class level.

The 2000 OES survey estimates are developed from data for 2 years of the OES sample. The combined 1999 and 2000 data were obtained from approximately 800,000 sample units. The 2000 estimates also are based on the wage-updating and estimation methodology introduced in 1997.

Estimated employment

As discussed 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 2000 reference month. For each MSA, the estimated employment for an occupation at the reported three-digit SIC level was calculated by summing the product of the weighted employment and the ratio factor for each sampled establishment in the MSA/three-digit SIC. 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 h level was calcu-

lated using the following equation:

$$\hat{X}_{ho} = \sum_{i \in h} \left(w_i \; BMF_i \; x_{io} \right)$$

where o = occupation;
h = reported three-digit SIC within an MSA;
i = establishment;
w_i = adjusted sample weight for establishment I;
BMF_I = the hierarchical benchmark factor applied to establishment i;
x_{io} = reported employment for occupation o in establishment i;
$\hat{X}_{ho}$ = estimated employment for occupation o in SIC h within an MSA.

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

$$\hat{X}_o = \sum_{h=1}^{L_h} \hat{X}_{ho}$$

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 might 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 $16.99 per hour. As a result, individual wage rates of workers are not collected. Conventional arithmetic mean formulas are not applicable in this situation. Because wage data are collected within an interval matrix, the particular wage rate of all employees within an interval is approximated by a mean wage rate value for the interval for each of the first 11 wage intervals. Data from the BLS National Compensation Survey (NCS) are used to calculate these mean wage rate values. The mean wage value for the upper open-ended wage interval is set at that interval's starting point. Occupational wage rates are calculated by developing a weighted estimate of total occupational wages, and dividing that by a weighted estimate of total occupational employment ($\hat{X}_o$).

Wage updating process. Because data from 2 years were used to produce the 2000 OES wage estimates for most occupations, a process was used to update prior year information so that it would be representative of the 2000 reference period. This was done by adjusting 1999 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 1999 to fourth-quarter 2000 for nine major occupational groups. Each OES occupation belongs to one of these major occupational groups.

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

$$\hat{R}_o = \frac{\sum_{z=t-2}^{t} \left(\sum_{i \in z} w_i \; BMF_i \; \hat{y}_{io} \right)}{\hat{X}_o}, \text{ where } \hat{y}_{io} = u_{zo} \sum_r x_{ior} c_{zr} \; ; (i \in z)$$

and where o = occupation;
$\hat{R}_o$ = mean wage rate for occupation o;
z = year;
t = current reference year;
w_i = adjusted sampling weight for estatablishment i;
$\hat{y}_{io}$ = unweighted total wage estimate for occupation o in establishment i;
r = wage interval;
$\hat{X}_o$ = estimated employment for occupation o;
$x_{io\,r}$ = reported employment for occupation o in establishment i in wage interval r (note that establishment i reported data for 1 year (z));
u_{zo} = ECI updating factor for year z and occupation o; and
c_{zr} = see below.

In this formula, c_{zr} represents the mean wage of interval r for year z. 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 using a linear interpolation procedure.

Variance of estimates

Occupational employment variance estimates. Estimates of sampling error are calculated to allow data 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) for the occupational employment estimates is based on the survey's sample design and method of estimation. The OES program employs 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 ($\hat{X}_{hjog}$) 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 *h*/reported size class *j* level is calculated using the following equation:

$$v(\hat{X}_{hjo}) = \frac{\sum_{g=1}^{G} (\hat{X}_{hjog} - \bar{\hat{X}}_{hjo})^2}{G(G-1)}$$

where $v(\hat{X}_{hjo})$ = estimated variance of $\hat{X}_{hjo}$;
G = number of random groups;
$\hat{X}_{hjo}$ = estimated employment of occupation *o* in SIC *h* and size class *j*;
$\hat{X}_{hjog}$ = estimated employment of occupation *o* in SIC *h*, size class *j*, and subsample *g*; and
$\bar{\hat{X}}_{hjo}$ = estimated mean employment for occupation *o* in SIC *h* and size class *j* based on the G subsamples.

(Note that a finite population correction factor is applied to the terms $\hat{X}_{hjog}$ and $\bar{\hat{X}}_{hjo}$.)

The variance for an occupational employment estimate at the reported three-digit SIC *h* level was obtained by summing the variance $v(\hat{X}_{hjo})$ across all reported size classes *j* in SIC *h*.

$$v(\hat{X}_{ho}) = \sum_{j \in h} v(\hat{X}_{hjo})$$

Similarly, the variance for an occupational employment estimate at the reported two-digit SIC level *H* is obtained by summing the variance $v(\hat{X}_{ho})$ across all reported three-digit SICs *h* within the two-digit SIC.

$$v(\hat{X}_{Ho}) = \sum_{h \in H} v(\hat{X}_{ho})$$

Occupational mean wage variance estimates. The formula used to estimate occupational mean wage variances also is based on the survey's sample design and method of estimation. Because the OES wage data are collected in intervals, we do not capture the exact wage of each worker. Therefore, some components of the wage variance are approximated using factors developed from NCS data. A *Taylor Linearization* technique was 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:

$$v(\hat{R}_o) = \begin{pmatrix} \frac{1}{\hat{X}_o^2}\left(\sum_h \left\{\frac{n_{ho}(1-f_{ho})}{n_{ho}-1}\right\}\left\{\sum_{i \in h} w_i^2 (q_{io} - \bar{q}_{ho})^2\right\}\right) + \\ \sum_r \theta_{or}^2 \sigma_{cr}^2 + \frac{1}{\hat{X}_o^2}\sum_r \left(\sum_{i=1}^{n_{or}} (w_i x_{ior})^2\right)\sigma_{er}^2 + \frac{1}{\hat{X}_o}\sum_r \theta_{or}\sigma_{\omega r}^2 \end{pmatrix}$$

where $\hat{R}_o$ = the estimated mean wage for occupation *o*;
$v(\hat{R}_o)$ = the estimated variance of $\hat{R}_o$;
$\hat{X}_o$ = the estimated occupational employment for occupation *o*;
h = a stratum (area / industry / establishment employment size);
f_{ho} = the sampling fraction for occupation *o* in stratum *h*;
n_{ho} = the number of sampled establishments that reported occupation *o* in stratum *h*;
i = an establishment;
w_i = the sampling weight for establishment *i*;
q_{io} = the quantity $\left(\hat{y}_{io} - \hat{R}_o x_{io}\right)$ for occupation *o* in establishment *I*;
$\hat{y}_{io}$ = the estimated total occupational wage in establishment *i* for occupation *o*;
x_{io} = the reported employment in establishment *i* for occupation *o*;
$\bar{q}_{ho}$ = the mean of the q_{io} quantities

for occupation o in stratum h;

r = a wage interval;

θ_{or} = the proportion of employment within interval r for occupation o;

x_{ior} = he reported employment in establishment i within wage interval r for occupation o; and

$\left(\sigma_{cr}^2, \sigma_{er}^2, and\ \sigma_{\omega r}^2\right)$ respectively represent—within wage interval r—the variability of the wage value imputed to each worker, the variability of wages across establishments, and the variability of wages within establishments. These quantities are estimated using data from the BLS NCS.

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 also can 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 data 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, and if each of these were 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 and mean wage 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 and phone calls 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 such as 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:

- Follow-up 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; and
- 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. 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) program 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 and metropolitan areas each year. Prior to 1996, the OES survey collected data from specified industries in 1of 3 years in the survey round, as indicated in the table below.

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