

Revisions to the Current Employment Statistics State and Area Estimates Effective January 2003

Molly E. Barth

With the release of estimates for January 2003, State and area employment, hours, and earnings data produced from the Current Employment Statistics (CES) program incorporate several important changes that affect data analysis and comparison. These changes include:

- Incorporation of March 2002 benchmarks,
- Completion of the CES sample redesign,
- Conversion to the 2002 North American Industry Classification System (NAICS), and
- Modification of seasonal adjustment methodology.

This article summarizes the nature of these changes and provides an indication of their effect on published estimates.

Background

The CES program is a Federal-State cooperative program that produces monthly estimates of employment, hours, and earnings based on nonagricultural establishment payrolls for the Nation, the 50 States, the District of Columbia, Puerto Rico, the Virgin Islands, and more than 270 Metropolitan Statistical Areas (MSAs). Information for these estimates is derived from a sample of more than 300,000 business establishments. CES estimates are some of the most closely watched and widely used economic indicators among public and private policymakers alike. The CES program offers several important attributes to its users: Timely release of data, an abundance of industry and geographic detail, and an annual benchmark to full population counts from State unemployment insurance (UI) tax records, which helps to maintain overall survey accuracy.

Incorporation of March 2002 benchmarks

With the release of data for January 2003, the State and area sample-based estimates have been adjusted to new benchmarks—comprehensive counts of employment—based primarily on UI reports filed by employers with State Employment Security Agencies. The revisions incorporate March 2002 benchmarks and affect data from 2001 forward. The size of employment revisions to March 2002 estimates is shown in table 1. Additional information on benchmark

revisions will be published in the May issue of this publication. Over the last 5 years, the average absolute benchmark revision to State estimates of total nonfarm employment ranged from 0.4 to 0.7 percent.

Completion of the CES sample redesign

Background. Historically, the CES was based on a quota-based sample design whose inception over 50 years ago predates the introduction of probability sampling as the internationally recognized standard for sample surveys. Quota-based samples are known to be at risk for potentially significant biases; the large sample size and annual benchmark to population counts only partially mitigate the risk of bias inherent in a quota design.

In June 1995, the Bureau of Labor Statistics (BLS) announced plans for a comprehensive sample redesign of its monthly payroll survey. The primary goal of the redesign was to develop probability-based sampling and estimating techniques, which would in turn enable CES to produce standard survey accuracy measures and confidence intervals, and improve upon methods for estimating business births and deaths. Probability sampling was to be enhanced by improved sample solicitation techniques.

Methodology. The CES probability sample is a State-based design that minimizes variance on statewide total private nonfarm employment estimates. It is a simple random sample, stratified by industry and size, clustered by unemployment insurance (UI) report number. (UI records are the basis for the CES sampling frame and are a benchmark source.) The sample frame and the sample itself are updated twice a year, but on a lagged basis, as new quarters of UI records become available. Because of the lag in the sample frame, CES is using an ARIMA (Auto-Regressive Integrated Moving Average) time series model to estimate the net of business births and deaths not measurable on a current basis by the sample.

The primary strength of the birth/death models is their basis in a 10-year time series from the BLS longitudinal UI database (LDB). The database stores firm-level information on business births, deaths, and employment levels. Reliance on historical trend, however, will somewhat limit the models' sensitivity to economic turning points.

The initial research phase for the CES sample redesign was completed in 1997, and the BLS launched a production

Molly E. Barth is an economist in the Division of Current Employment Statistics, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics. Telephone: (202) 691-6559; e-mail: barth_m@bls.gov

Table 1. Percent differences resulting from nonfarm payroll employment benchmarks by State, March 2002

State	Percent revision	State	Percent revision	State	Percent revision
Alabama	-0.8	Kentucky	-2.0	North Dakota	-1.1
Alaska	1.0	Louisiana	-1.9	Ohio	-1.5
Arizona	.5	Maine	-.8	Oklahoma	-1.8
Arkansas	-.6	Maryland	.9	Oregon	-.7
California	-1.2	Massachusetts	-1.4	Pennsylvania	.0
Colorado	-.6	Michigan	-2.0	Rhode Island	-.5
Connecticut	-.1	Minnesota	-.5	South Carolina	-1.6
Delaware	-1.2	Mississippi	-.8	South Dakota	-1.0
District of Columbia	2.1	Missouri	.6	Tennessee	-2.1
Florida	-.3	Montana	-.2	Texas	-.2
Georgia	1.0	Nebraska	-.6	Utah	-.1
Hawaii	.3	Nevada	-2.1	Vermont	.6
Idaho	-1.2	New Hampshire	-1.2	Virginia	-.3
Illinois	-.9	New Jersey	-.2	Washington	-.2
Indiana	-.8	New Mexico	.1	West Virginia	-.1
Iowa	-1.2	New York	-.9	Wisconsin	-1.4
Kansas	-2.1	North Carolina	-.9	Wyoming	-.5

test of the new sample design at that time. In March 2001, the first State and area estimates from the redesign were published, for the wholesale trade major industry division. In March 2002, the next phase was implemented with the publication of redesign estimates for the mining, construction, and manufacturing divisions. The completion of the phase-in for the redesign, in March 2003, for the remaining industries coincides with the conversion of all State and area CES series from industry coding based on the 1987 Standard Industrial Classification (SIC) system to industry coding based on the 2002 version of the North American Industry Classification System (NAICS).¹

Conversion to the 2002 NAICS

Background. With the release of the January 2003 data, the CES program begins publishing its State and area data series under NAICS. NAICS replaces the SIC system used by U.S. statistical agencies for 60 years. Created in 1997, NAICS is a collaborative effort by Canada, Mexico, and the United States to standardize their economic data series.

In addition to providing data comparability among the North American countries, NAICS conversion carries several other positive implications. First, NAICS increases the number of industry classifications to 1,170, 15 percent more than were previously recognized under SIC. Most of this expansion comes in service-related industries, with new industries such as HMO medical centers, warehouse clubs and superstores, and bed-and-breakfast inns. Other important additions are numerous “high-tech” industries, such as cellular telecommunications and software reproduction.

Second, from an organizational standpoint, NAICS is more consistent than the SIC because all industries are classified according to a single principle: similarity of production processes. The SIC had no unifying standard by which industries were categorized. Third, NAICS will be reviewed and updated every 5 years by all three North American countries. This will ensure that the classification system always captures the newest of our economy’s ever-evolving industries.

NAICS doubles the amount of top-level industry groupings, known as sectors. There are 20 broad sectors, compared with only 10 major industry divisions under SIC. In addition to these sectors, BLS and its U.S. NAICS partners (the Bureau of Economic Analysis and the U.S. Census Bureau) have further aggregated industry classifications into 11 *supersectors*. Above the supersectors are two more aggregations, the commonly known domains of *goods-producing* and *service-providing*.²

The resulting changes are fundamental. Manufacturing, which is a NAICS sector as well as a BLS supersector, is structurally different from manufacturing under SIC. The sector has 79 new industries, including computer and electronic product manufacturing. Also notable under NAICS, publishing has been moved from manufacturing to information, and logging has been reclassified into natural resources and mining.

Another important change is the creation of an information sector. Also a BLS supersector, information contains 34 industries that produce, disseminate, or process information and cultural products. Twenty of these industries are new, including Internet service providers,

¹ For more information on CES redesign methodology and implementation, see the *Update on the BLS Sample Redesign for the Current Employment Statistics Survey* at <http://www.bls.gov/web/cesred.htm>.

² See <http://www.bls.gov/sae/saenaics.htm> for complete information on the State and area CES conversion to NAICS. An overview of NAICS classification and its differences from SIC can be found at <http://www.bls.gov/sae/saewhatis.htm>.

sound recording studios, and cable program distributors. The remainder reflects reorganization of SIC industries, such as broadcasting and publishing.

Management of companies and enterprises (sector 55) also is the result of major structural SIC reorganization. Under SIC, corporate headquarters were treated as “auxiliary” establishments and classified in the same industry as their parent companies. NAICS treats headquarters as separate establishments and classifies them within sector 55, which is part of the BLS supersector of professional and business services.

Publication levels and data availability. NAICS conversion ultimately results in more meaningful economic data, but it presents the initial problem of discontinuity with currently published SIC series. There are 544 SIC industries that are either discontinued or unidentifiable within NAICS, compared with only 38 that are identical to a given NAICS industry. Moreover, there are 320 NAICS industries that are completely new. While the NAICS coding structure provides many new breakouts not available in the SIC system, the amount of CES published detail is constrained by the overall CES sample size.

To accommodate data users’ needs for time series data, CES has reconstructed State and area all-employees series coded for NAICS back to January 1990 for most industries. Total nonfarm and total government data for each State and metropolitan area are still available for the entire history of the series. Although the logging industry was reclassified under agriculture, forestry, fishing and hunting under

NAICS, it has been retained within the scope of the CES definition of nonfarm employment to preserve historical consistency.

For the purpose of comparison, the CES program has established a minimum amount of industry detail for every State and MSA. This publication structure is shown in table 2. Beyond the minimum guaranteed number of series, additional NAICS employment series are published where there is sufficient sample. The guaranteed published series add up to goods-producing, service-providing, total private, and total nonfarm employment.

Guaranteed employment series not meeting the minimum publication criteria for direct sample-based estimation are estimated using an econometric modeling technique known as the CES Small Domain Model (SDM). The SDM is a weighted least-squares-regression model based on the following inputs: (1) An estimate based on the available CES sample for that series, (2) a time series projection based on the trend from 10 years of historical data, and (3) for an MSA, an estimate borrowed from the statewide series for the same industry. The weights for each of the inputs are recalculated monthly. Most of the models are applied to smaller industries in small MSAs.

NAICS conversion has resulted in a net expansion in the number of employment-series available. There are 15,293 all-employees series published under NAICS, compared with 13,064 under the SIC. (See table 3.) The majority of CES hours and earnings series published are in manufacturing industries. The historical data were not reconstructed for NAICS hours and earnings series. The

Table 2. Minimum CES State and area publication detail

BLS code	NAICS code	Industry
00-000000	–	Total nonfarm
05-000000	–	Total private (total nonfarm less government)
06-000000	10, 20, 30	Goods-producing
10-000000	1133 (logging), 21	Natural resources and mining ¹
20-000000	23	Construction
30-000000	31-33	Manufacturing
07-000000	40-90	Service-providing
40-000000	42, 44-45, 48-49, 22	Trade, transportation, and utilities
41-000000	42	Wholesale trade
42-000000	44-45	Retail trade
43-000000	48-49, 22	Transportation and warehousing and utilities
50-000000	51	Information
55-000000	52-53	Financial activities
60-000000	54-56	Professional and business services
65-000000	61-62	Educational services and health services
70-000000	71-72	Leisure and hospitality
80-000000	81	Other services, except public administration ²
90-000000	–	Government (defined by ownership) ³
90-910000	–	Federal
90-920000	–	State
90-930000	–	Local

¹ Natural resources and mining may be combined with construction if there is insufficient sample to publish mining alone.

² Some MSAs do not publish all the government publication detail due to small sample sizes.

Table 3. **CES State and area employment series: Publication structure**

Area employment	Number of areas	SIC published series	NAICS published series	Difference
Statewide	51	4,020	4,989	969
MSA>1,000,000	29	1,880	2,300	420
500,000-999,999 ..	35	1,491	1,936	445
250,000-499,999 ..	41	1,411	1,490	79
100,000-249,999 ..	97	2,533	2,719	186
MSA<100,000	85	1,729	1,859	130
Total	338	13,064	15,293	2,229

series will start with January 2001 levels set from the probability sample averages.

Reconstruction of all-employees time series data. The primary input to the reconstructed series is the LDB, which contains monthly employment levels for roughly 10 million establishments since 1990. The LDB received its first 2002 NAICS-coded data for the first quarter of 2001.

For those establishments that did not have NAICS codes on file, or that went out of business prior to March 2001, codes were imputed. If an SIC industry had a direct match in NAICS, the NAICS code was applied to the uncoded establishment. However, if an establishment's SIC did not have a direct match within NAICS, a NAICS code was assigned through an imputation procedure known as the "nearest neighbor" method. This process required use of a table that summarizes microdata across all States by ownership for possible SIC-to-NAICS combinations. An uncoded establishment was matched with a record that shared the same SIC and ownership codes, and whose average employment was closest to its own. This process thus assumed that employment is similar among establishments in the same NAICS industry. When there was more than one match, the uncoded establishment was randomly assigned a NAICS code based on the proportion of records assigned each NAICS code in a given SIC/NAICS group. For establishments that are part of a multiunit reporter, meaning that several establishments report employment and earnings data under a single UI account number, a nearest neighbor was first sought within the UI account.³

Once the LDB coding was accomplished, the establishment data were aggregated to produce statewide six-digit NAICS industry employment levels. These data were then further aggregated to all statewide publication levels. Similar logic was applied to MSA employment data; however, the data were aggregated based on the county code of each establishment as of the first quarter of each year. For

³ For a more detailed description of how NAICS codes were assigned to the LDB, see *Implementing the NAICS for Business Surveys at BLS* by Gordon Mikkelson, Teresa L. Morisi, and George Stamas, June 2000; available online at <http://www.bls.gov/ore/pdf/st000090.pdf>.

those industries that include jobs not covered by UI laws, such as religious organizations and railroads, the noncovered portion of employment was added to the summed LDB just as it is added to covered population counts for the CES benchmark.

The greatest strength of this reconstruction methodology is its precision, as the series were summed from information on individual establishments. Also, keeping the NAICS codes constant throughout the history of the LDB lends an element of consistency to the data. Any errors, and subsequent corrections, in industry coding over the 1990-2001 period were eliminated. The major limitation to this methodology, however, was that keeping the industry code constant removed any true economic industry-code changes. The aggregate NAICS total nonfarm employment levels were controlled to previously published SIC totals. Any difference between these two levels was distributed proportionally across NAICS industries.

Modification of seasonal adjustment methodology

Background. The primary purpose of the CES program is to provide users with month-to-month changes in industry employment. However, the program also conducts an annual level adjustment for quality control known as the benchmark process. The CES benchmarking process noted above is the replacement of CES sample-based estimates with UI universe counts through the latest quarter for which data are available. Series are estimated using the sample going forward from the latest benchmark quarter.

Because the CES sample-based estimates often exhibit a seasonal pattern that is different from the pattern demonstrated by the UI universe used as the benchmark replacement series, BLS uses a hybrid series for seasonal adjustment purposes. CES uses the original sample estimates for a span of 10 years and seasonally adjusts this series; similarly, a 10-year span of benchmarked data is also seasonally adjusted. The two series are then spliced together at the end of the benchmark quarter, and the seasonal factors generated from adjusting the sample series are applied to the next 12 months of sample estimates, until the next benchmark.⁴

Impact of NAICS on data availability. The difficulty NAICS conversion brings to the seasonal adjustment process is that there is no NAICS sample estimate history available. To resolve this problem, CES applied SIC-to-NAICS employment ratios to the original SIC sample history to create a NAICS-based proxy. These estimates were then aggregated up to the NAICS supersectors. These proxy series have limitations; however, seasonal adjustment models

⁴ For more detailed information on CES seasonal adjustment methodology, see *Seasonal Adjustment of Hybrid Economic Time Series* by Stuart Scott, George Stamas, Thomas J. Sullivan, and Paul Chester, 1994; available online at <http://www.bls.gov/ore/pdf/st940350.pdf>.

weight the most recent years more heavily, so the impact of the proxy series will diminish in future years. For series that altogether lacked the SIC sample history needed to create a NAICS proxy series, there will be no seasonally adjusted data published.

BLS is publishing seasonally adjusted statewide all employee series for supersectors and higher levels of aggregation, where sufficient seasonality and ratio-based sample history exist.

Statewide seasonally adjusted total nonfarm employment was derived under SIC by summing the seasonally adjusted major industry divisions, which in turn were summed from seasonally adjusted two-digit industries. This practice will continue for most States under NAICS; that is, seasonally adjusted total nonfarm employment will be obtained by adding the adjusted supersectors. However, for cases which a State had few publishable seasonally adjusted supersectors,

total nonfarm data were seasonally adjusted directly at the aggregate level. This affects only a small number of States.

Additional information

State and area employment, hours, and earnings data are available at <http://www.bls.gov/sae/> on the BLS Internet site. Users may access the data via various retrieval methods at this address. Any questions on how to access the data through the Internet should be directed to webmaster@bls.gov. Inquiries for additional information on the methods or estimates derived from the CES survey should be sent to: U.S. Department of Labor, Bureau of Labor Statistics, Room 4860, 2 Massachusetts Avenue, NE., Washington, DC 20212-0001. The telephone and fax numbers are (202) 691-6559 and (202) 691-6820, respectively. The e-mail address is sminfo@bls.gov.