Recent changes in the State and Metropolitan Area CES survey

Like its national counterpart, the State and Metropolitan Area Current Employment Statistics survey has converted to the North American Industry Classification System (NAICS); two other changes are a minimum guaranteed publication structure and the use of a Small Domain Model for guaranteed series that do not meet publication criteria

Molly E. Barth

Molly E. Barth is an economist in the Office of Employment and Unemployment Statistics, Bureau of Labor Statistics. E-mail: Barth_M@bls.gov

With the release of the January 2003 preliminary State and Area data on March 20, 2003, the Current Employment Statistics (CES) survey introduced a number of important changes. Most notable is the conversion from the 1987 Standard Industrial Classification (SIC) system to the 2002 North American Industry Classification System (NAICS). All State and Area CES series, including historical data as well as current estimates, are now reported on a NAICS basis; SIC data will no longer be published. The CES program also completed a redesign of its sample methodology, moving from a quota-based to a probability-based sampling technique.

The new changes apply to State and Area, as well as national, CES series. The national CES program implemented NAICS and the final phase of the CES redesign with the release of the preliminary May 2003 estimates on June 6, 2003.[1] Other important State and Area CES changes include the development of a minimum guaranteed publication structure for all States and metropolitan areas and the CES Small Domain Model for guaranteed series that do not meet publication criteria under sampling methodology. These changes and others are described in this article.

Overview of the CES program

The CES program is a Federal-State cooperative program that produces monthly estimates of employment, hours, and earnings based on nonfarm establishment payrolls for the Nation, the 50 States, the District of Columbia, Puerto Rico, the Virgin Islands, and more than 270 Metropolitan Statistical Areas (MSA's). Information for these estimates is derived from a sample of 300,000-plus business establishments. CES estimates are closely followed and widely used economic indicators, offering timely data with an abundance of industry and geographic detail.

State and Area CES estimates are derived by the States, as provided for in the Federal-State cooperative agreement; sample and estimation procedures are designed to produce accurate data for each State. The Bureau of Labor Statistics estimates national employment series independently, using sample data from all States and the District of Columbia. State estimates are not forced to sum to national totals and vice versa.

Estimates from the CES survey are benchmarked once a year to universe counts from State Unemployment Insurance (UI) tax records, provided by the Covered Employment and Wages program. For State and Area CES series, the most recent UI universe counts available are used to benchmark employment estimates; the monthly estimates are replaced with universe data back to the previous benchmark. Estimates for months subsequent to the most currently available UI universe counts are recomputed by

using sample-based links applied to the new UI benchmark levels.

NAICS replaces SIC

After 60 years of use by U.S. statistical agencies, the SIC system was retired, and NAICS was adopted as the standard means of industry classification. NAICS was devised by the statistical agencies of the United States, Mexico, and Canada to enhance comparability of economic data across the North American Free Trade Association (NAFTA) trade region.[2]

Conversion of State and Area series to NAICS

The CES program has reconstructed State and Area all-employees series on a NAICS basis back to January 1990 for most industries. Converting the data from SIC to NAICS involved several steps. First, the CES reporting units were assigned NAICS codes. Then, publication standards were developed, estimating cells were defined, and historical data were reconstructed.

Reconstruction methodology. Before the CES program could begin the reconstruction of historical employment series, NAICS codes needed to be assigned to all the records of the BLS Longitudinal Database (LDB).[3] The NAICS and ownership codes for each establishment in the LDB as of the first quarter of 2001 were then carried back through the entire history of the database. For statewide series, the microdata for all establishments within the scope of the CES survey were first added to six-digit NAICS levels. These totals were then aggregated to all official NAICS-based statewide CES publication levels. For MSA series, the LDB microdata were added to six-digit NAICS employment levels on the basis of the county code as of the first quarter of each year. These levels were then aggregated to all official NAICS-based CES publication levels in a process to be described shortly.

One of the strengths of this reconstruction methodology is its universal applicability: because series were summed from information on individual establishments, all the industry and geographic detail needed to re-create employment data for even small MSA's was available.[4]

Keeping the NAICS codes constant over the historical data of the LDB gave consistency to the data by eliminating all non-economic code changes caused by coding error. However, any true economic code changes caused by a change in primary business practice were removed as well. Similarly, the ownership codes remained constant. This again eliminated any non-economic changes, but also erased all true changes in ownership, such as the privatization of an establishment. The location of establishments was not held constant; rather, the county code from the first quarter of each year was used to assign a given firm to an MSA. Consequently, both economic and noneconomic location changes were present in the data.

To remedy these disadvantages, each State reviewed the reconstructed NAICS-based series thoroughly. Once the States approved their reconstructed histories, the Bureau set monthly NAICS total nonfarm employment to previously published SIC levels by adjusting the amounts in each estimating cell accordingly. For example, if the NAICS total nonfarm employment in an MSA for May 1995 equaled 1,010,000, but the SIC value for the same month was 1,000,000, each estimating cell was revised downward by 1 percent for May 1995.

Publication criteria and guaranteed levels for all-employees series. To bring consistency and comparability to State and Area employment series, the CES program established a minimum amount of published industry detail for all States and MSA's. The new publication structure is shown in exhibit 1. In addition, NAICS two-digit levels, or sectors, are considered guaranteed published levels for Statewide estimates, except in the case of States with very small employment totals in any given sector; these additional guaranteed levels are shown in exhibit 2.

Beyond the minimum guaranteed levels, additional NAICS employment series are published only where there is a sufficient sample. The following criteria were applied to determine whether a NAICS series warranted publication: there must be a minimum sample size of 30 UI units selected, or a minimum universe employment level of at least 3,000 *and* a minimum sample coverage of at least 50 percent. Variance was used as a secondary criterion.

State and Area series under NAICS

The CES State and Area series under NAICS consist of data for the guaranteed publication levels in all States and MSA's, as well as data for (two-digit) sectors, (three-digit) subsectors, (four-digit) industry groups, and selected industries at the five- and six-digit levels when a sufficient sample is present.

Change in the number of CES series. The number of CES series under NAICS is smaller than the number of series that were produced under the SIC system. The decrease is due to a reduction in the number of hours and earnings series. Reallocating the CES sample under the probability redesign resulted in a more representative distribution of establishments in each industry. Most of the hours and earnings series are published in manufacturing industries, which have been declining in recent years. Consequently, fewer industries met the minimum publication criteria for hours and earnings series (See table 1.)

However, the number of all-employees series has *increased*. States and large MSA's gained relatively more publication lines than did smaller areas. The increase is again due to the reallocation of the CES sample, which shifted into larger areas.[5] Still, the guaranteed publication structure kept small MSA's from

Exhibit 1.	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
90–000000	–	Government (defined by ownership)[2]
90–910000	–	Federal
90–920000	–	State
90–930000	–	Local

[1] Natural resources and mining may be combined with construction if the sample is insufficient to publish mining alone.
[2] Because of small sample sizes, some MSAs do not publish the entire amount of government publication detail.

NOTE: Dash indicates no NAICS code for category.

Exhibit 2.	Additional CES guaranteed publication levels for statewide employment estimates	
BLS code	**NAICS code**	**Industry**
43–220000	22	Utilities
43–400089	48, 49	Transportation and warehousing
55–520000	52	Finance and insurance
55–530000	53	Real estate and rental and leasing
60–540000	54	Professional, scientific, and technical services
60–550000	55	Management of companies and enterprises
60–560000	56	Administration and support and waste management and remediation services
65–610000	61	Education services
65–620000	62	Health services and social assistance
70–710000	71	Arts, entertainment, and recreation
70–720000	72	Accommodation and food services

posting a net loss of series. The guaranteed publication structure across all States is made feasible by the development of a new small-domain modeling technique applicable to series with small sample sizes; the model is described shortly. Table 2 shows the increase in the number of all-employees series by the size of nonfarm employment, both statewide and for the MSA's.

Amount of history. Nearly all of the all-employees series for the guaranteed published industries have reconstructed historical data going back to 1990. The only exception is in government: small MSA's did not previously publish detail below the total government level, so Federal, State, and local government series begin in 2001. More detailed industries also have reconstructed historical data back to 1990. However, each State had the option to suppress the historical data of nonguaranteed series from publication if the reconstructed

historical data were questionable. The nonguaranteed series begin in 2001, when the first benchmark NAICS data were available. Total nonfarm employment is available back to the original starting date of each series—as far back as 1939.

Hours and earnings series. As mentioned earlier, the majority of published State and Area CES hours and earnings series are in manufacturing industries. The historical data for these series were not reconstructed; instead, NAICS-based hours and earnings series begin in January 2001, with levels set from probability sample averages. Candidates for hours and earnings publication series were required to publish employment data and have at least 40 UI units selected for the CES sample. The higher UI threshold was to compensate for potential nonresponse to questions about hours and earnings data. There are 3,986 NAICS-based State and Area CES hours and earnings series.

Table 1.	Change in number of published CES series for all data types				
Area size by employment	Number of areas	Published series under SIC	Published series under NAICS	Difference	
Total	344	23,992	19,421	−4,571	
Statewide	53	9,497	7,285	−2,212	
MSA:					
More than 1,000,000 workers	27	3,293	2,830	−463	
500,000 to 999,999 workers	40	2,927	2,489	−438	
250,000 to 499,999 workers	39	2,099	1,760	−339	
100,000 to 249,999 workers	94	3,630	2,970	−660	
Less than100,000 workers	91	2,546	2,087	−459	

Table 2.	Change in number of published CES all-employees series				
Area size by employment	Number of areas	Published series under SIC	Published series under NAICS	Difference	
Total	344	13,283	15,396	2,113	
Statewide	53	4,116	4,961	845	
MSA:					
More than 1,000,000 workers	27	1,776	2,234	458	
500,000 to 999,999 workers	40	1,719	2,062	343	
250,000 to 499,999 workers	39	1,334	1,490	156	
100,000 to 249,999 workers	94	2,467	2,646	179	
Less than 100,000 workers	91	1,871	2,003	132	

New series code structure in LABSTAT. Having a new set of State and Area CES series also means adopting a new series code structure in LABSTAT, the BLS data Web site. As indicated in table 2, the prefix, or first two letters of the identifier, has changed to SM, to identify these series as State and Metropolitan Area NAICS series. The SIC series, with prefix SA, will remain in the LABSTAT database as discontinued series and will not be updated. The industry code has been updated to accommodate the longer NAICS code; all other fields remain the same. (See exhibit 3.)

Other changes released with NAICS

Along with the conversion to NAICS, the State and Area CES program completed the transition to a probability-based sample, adopted a model for estimating small-domain series, and implemented temporary changes to its seasonal adjustment methodology for State and Area series.

Completion of the probability sample redesign. With the release of the January 2003 preliminary data in March 2003, the CES program completed the implementation of its probability sample redesign for all private State and area industries. The national CES program finished its transition to the probability sample redesign with the release of preliminary May 2003 data in June 2003.[6]

Previously, the CES program used a quota-based sampling technique, which was subject to potentially significant biases. The transition from quota-based to probability-based sampling was a multiyear process completed in three phases. In March 2001, the first State and Area estimates from the redesign were published for the wholesale trade industry division. The second

phase, implemented in March 2002, added the mining, construction, and manufacturing divisions. The final phase of the redesign, in March 2003, phased in the remaining private-industry divisions, recoded for NAICS.

The primary goals of the redesign were to develop probability-based sampling and estimating techniques, to produce standard accuracy measures and confidence intervals for the survey, and to improve upon methods for estimating business births and deaths.[7] The introduction of CES probability sampling was enhanced by improved sample solicitation techniques.[8]

The CES probability sample is a State-based design that minimizes the variance on statewide total private nonfarm employment estimates. It is a simple random sample, stratified by industry and size, and clustered by UI report number. 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, the CES survey uses an ARIMA time series model to estimate the net of business births and deaths that are not measurable on a current basis by the sample. While this reliance on a historical trend will limit the model's sensitivity to economic turning points somewhat, the information is based on 5 years of firm-specific data on business births, deaths, and employment levels from the LDB.

The CES Small Domain Model. Guaranteed published employment series that do not meet the sample-based publication criteria discussed earlier are estimated with a regression model known as the CES Small Domain Model (SDM), developed by the Bureau. The SDM is a weighted-least-squares model consisting

Positions	Value	Field Name
	Exhibit 3. Example of series code format for the State and Area series under NAICS: series = SMS080000432200001	
1–2	SM	Prefix (SM refers to State and Metropolitan Area series)
3	S	Seasonal adjustment code (S is for seasonally adjusted, U for unadjusted)
4–5	08	State code
6–9	0000	Area code
10–11	43	Supersector code
12–17	220000	Industry code
18	1	Data type code

of the following inputs: (1) an estimate derived from the available CES sample for the cell in question, (2) an ARIMA projection of a universe trend based on 10 years of historical ES-202 data, and (3) for MSA's, an estimate borrowed from the statewide trend for the industry in question. The weights for each of the inputs are calculated monthly. In addition to the guaranteed employment series, the SDM was approved for use in estimating statewide two-digit NAICS industries whose sample is insufficient for direct estimates.

Modeled series account for 1,085 of the 15,396 State and Area CES all-employees series—about 7 percent. Approximately 1.7 percent of total nonfarm employment is estimated by the SDM.

Interim changes to seasonal adjustment methodology. The focus of the CES program is to provide data users with information about month-to-month changes in industry employment. However, the program also conducts the benchmark revision, an annual level adjustment for quality control. In the CES benchmarking process for States and metropolitan areas, sample-based estimates are replaced with actual UI universe counts through the latest quarter for which data are available. Series are then estimated with the use of the sample, going forward from the latest benchmark quarter.

Historically, the UI universe data have exhibited a seasonal pattern different from that of the CES sample-based estimates. A hybrid series is utilized to accommodate this difference in seasonal patterns. The CES survey uses the original sample estimates for a span of 10 years and seasonally adjusts that series; similarly, a 10-year span of benchmark data is also seasonally adjusted. The two series are 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.

The difficulty in the conversion to NAICS was that historical

sample-based NAICS estimates do not exist. Thus, the entire time-series reconstruction was accomplished using only universe, or benchmark, data. To resolve this problem, the CES program applied statewide SIC-to-NAICS employment ratios (calculated using first-quarter 2001 LDB data) to the original two-digit SIC sample history to create a NAICS-based proxy. The estimates obtained were then aggregated to NAICS supersectors. While these ratio-based series have limitations, seasonal adjustment models place the most weight on recent years, so the effect of the proxy series will diminish over time.

The Bureau is publishing seasonally adjusted statewide all-employees series for supersectors and higher level aggregations for which enough seasonality and a sufficient ratio-based sample history exist. Series with a SIC sample history that was inadequate to create a NAICS proxy series could not be seasonally adjusted.

Under the SIC system, statewide seasonally adjusted total nonfarm employment was derived by summing the seasonally adjusted major industry divisions, which in turn had been summed from seasonally adjusted two-digit industries. This basic methodology remains the same under NAICS for most States, and seasonally adjusted total nonfarm employment will be summed from the adjusted supersectors. However, 22 States had few publishable seasonally adjusted supersectors, because of an insufficient sample history or a lack of seasonality. For these States, total nonfarm data were seasonally adjusted directly at the aggregate level.

The release of State and Area CES data in January 2003 marks the completion of the most profound changes to the CES survey in decades. The transition to NAICS, along with a probability-based sample and a small-domain estimation model, will improve the quality of CES statistics on employment, hours, and earnings for all States and metropolitan areas. □

Notes

[1] See Teresa L. Morisi, "Recent changes in the national Current Employment Statistics survey," this issue, pp. 3–13.

Morisi's article contains detailed background information on NAICS, a description of the conversion of the national CES series from SIC to NAICS, and a discussion of other important program changes implemented along with the conversion to NAICS.

[2] *See North American Industry Classification System, United States, 2002* (Executive Office of the President, Office of Management and Budget, 2002).

[3] See Morisi, this issue, pp. 3–13, for a description of the process of assigning the NAICS codes from CES codes.

[4] See Morisi, this issue, pp. 3–13, for details on the NAICS re-construction methodology for national CES series.

[5] For detailed background information on the CES sample design and selection, see *Technical Notes to Establishment Survey Data Published in Employment and Earnings*, on the Internet at **http://stats.bls.gov/web/cestn2.htm**.

[6] See Morisi, this issue, pp. 3–13, for more information.

[7] *Technical Notes*. See also **http://www.bls.gov/sae** for standard errors for State and Area CES estimates.

For information on estimating business births and deaths, see Jurgen Kropf, Sharon Strifas, and Monica Traetow, "Accounting for Business Births and Deaths in CES: Bias vs. Net Birth/Death Modeling," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 2002; on the Internet at **http://stats.bls.gov/ore/pdf/st020090.pdf**.

[8] For detailed background information on sample solicitation techniques under the CES redesign, see Richard Rosen, Louis Harrell, Christopher Manning, and Doug Skuta, "Data Collection Issues Related to Implementing the Redesigned Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 1999; on the Internet at **http://stats.bls.gov/ore/pdf/st990030.pdf**.

.