

Historical Reconstruction of Employment Time Series for Metropolitan Statistical Areas in the CES State and Area Program

August 2005

Brian Dahlin, Molly Garber
U.S. Bureau of Labor Statistics

ABSTRACT

The Current Employment Statistics (CES) program at the Bureau of Labor Statistics implemented revised Metropolitan Statistical Area (MSA) definitions with the release of its monthly employment estimates effective with the January 2005 release. MSA geographical boundaries are determined by the Office of Management and Budget (OMB) and are updated as new population data and information regarding commuting patterns become available. The conversion of existing time series for MSAs that added or lost counties under the new definitions, as well as the creation of historical time series for areas that did not exist under the previous set of MSA definitions, required significant analysis to ensure the integrity of the reconstructed data. The methods used in the reconstruction process are discussed in detail, as are limitations that were faced and techniques innovated to address them. The methods shown here form the basis of the reconstruction process to be used in future years to keep CES state and area time series current, on a near real-time basis, with future MSA revisions issued by OMB.

Keywords: BLS, CES, MSA, Reconstruction

CES Program Overview

The CES survey 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, and more than 400 metropolitan areas and divisions. Information for these estimates is derived from a sample of over 300,000 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 Quarterly Census of Employment and Wages program. For State and Area 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 monthly links applied to the new UI benchmark levels.

Metropolitan Area Concepts

The Office of Management and Budget (OMB) defines metropolitan and micropolitan statistical areas for purposes of collecting, tabulating, and publishing Federal data. Metropolitan and micropolitan statistical area definitions result from applying published standards to data from the U.S. Census Bureau.

The general concept of a metropolitan or micropolitan statistical area is that of a core area containing a substantial population nucleus, known as an *urban cluster*, together with adjacent communities having a high degree of social and economic integration with that core. Metropolitan and micropolitan statistical areas comprise one or more entire counties. A conceptually similar set of areas is defined in New England using cities and towns and is referred to as New England city and town areas (NECTAs).

Under the current standards, the county (or counties) in which at least 50 percent of the population resides within urban areas of 10,000 or more population, or that contain at least 5,000 people residing within a single urban area of 10,000 or more population, is identified as a "central county" (counties). Additional "outlying counties" are included in the core-based statistical area (CBSA) if they meet specified requirements of commuting to or from the central

counties. Counties or equivalent entities form the geographic “building blocks” for metropolitan and micropolitan statistical areas throughout the United States and Puerto Rico. In New England, the CBSAs are defined in terms of cities and towns rather than counties.

A micropolitan area contains a population core of at least 10,000 people but fewer than 50,000, the minimum threshold for a metropolitan statistical area (MSA). If specified criteria are met, an MSA containing a single core with a population of 2.5 million or more may be subdivided to form smaller groupings of counties referred to as “metropolitan divisions.”

As of June 6, 2000, there are 362 metropolitan statistical areas and 560 micropolitan statistical areas in the United States. In addition, there are 8 metropolitan statistical areas and 5 micropolitan statistical areas in Puerto Rico. CES publishes all of the metropolitan statistical areas and 31 metropolitan divisions. Note: CES does not publish the Lawrence-Methuen-Salem, Mass. or Taunton-Norton-Raynham, Mass. divisions of the Boston-Cambridge-Quincy metropolitan NECTA, nor does CES publish estimates for micropolitan statistical areas.

In addition, CES publishes employment, hours, and earnings series for nine non-standard MSAs. When metropolitan areas cross state boundaries, the state with the largest share of employment is referred to as the *controlling state* for the MSA; that is, it collects the relevant area data from the other state(s) and sums it to produce the final estimate value for the MSA. There are some MSAs and Metropolitan Divisions where the non-controlling state’s part represents a very significant percentage of the state’s total statewide employment. In these cases the non-controlling states were interested in having a separate estimate for its portion because of its importance to the state. If the intrastate portion has an employment size of at least 250,000 and represents at least 15% of total statewide employment, acceptable estimates intrastate parts can be made.

Besides these intrastate portions of MSAs and Metropolitan Divisions, CES is continuing to publish estimates for Baltimore City, Md.; New York City, N.Y.; and Philadelphia City, Pa., as it did under the previous MSA definitions.

Time Series Reconstruction Methodology

The CES program reconstructed the metropolitan area all-employees series back to January 1990 for most industries. Converting the series to the new MSA definitions involved a number of steps. First, estimating cells were defined. Then, the reconstruction methodology for each area was determined according to how different it was from its previous definition. The series were then constructed based on county level microdata. Finally, series breaks between the historical reconstructions and new benchmark levels were corrected to produce continuous time series.

Publication Criteria and Guaranteed Series

The CES program has established a minimum amount of published industry detail for all Statewide and MSA series in order to make employment data consistent and comparable across all areas. Because of extremely small populations and samples, MSAs with less than 100,000 employment could have a smaller minimum guaranteed publication structure than MSAs larger than 100,000 and States.

Beyond the minimum guaranteed series, additional industry detail is published only where there is sufficient sample to produce a reliable estimate. CES uses the following criteria to determine the whether a series warrants publication: either (1) a minimum sample size of 30 selected UI units or (2) a universe employment level of at least 3,000 and a minimum sample coverage of 50 percent. Variance is also considered, but as a secondary condition.

The BLS Longitudinal Database (LDB) provided the source for reconstructing the historical MSA data. The database stores firm-level information back to 1990 for all establishments covered by Unemployment Insurance, including business births, deaths, relocations, and employment levels. These data were summed from the individual firm records to county (town) level series for all levels of industry detail.

Before reconstructing the series, CES classified all MSAs into four categories:

- *Unchanged*: area did not change from its previous definition.
- *Slightly changed*: area gained or lost fewer than two counties or towns from its previous definition.
- *Greatly changed*: area changed by more than two counties or towns, or split into two or more individual areas.
- *New*: never before published by CES

The pre-existing data series were left intact for unchanged areas. For slightly changed MSAs, CES added or subtracted the relevant counties (towns) from the pre-existing data series. The data for new and greatly changed MSAs were derived entirely from the county level LDB summations.

To obtain these summations, CES held the NAICS industry and ownership for each establishment as of the first quarter of 2001 constant for the entire span of the time series. Industries out of CES scope, such as agriculture and private households, were removed from the files. The microdata were then tabulated to six-digit NAICS employment levels on the basis of the county (township) code as of the first quarter of each year. These levels were in turn aggregated to all official CES publication levels as determined by the publication structure review process described above.

Accounting for Employment Outside the LDB

Presumed non-covered (PNC) employees - those who work for churches, railroads, and other non-profit organizations that do not pay unemployment insurance - are not included in the LDB data. Inclusion of PNC employment in the reconstruction of MSA employment time series posed several challenges. Estimates of PNC employment existed at the statewide level only. Further, these estimates were available for December 2000 and January 2003 only.

To construct a time series of PNC employment at the MSA level, CES first needed to account for the missing PNC data from January 2001 through December 2002. These values were approximated for each state by applying the over-the-month employment change trends from the total nonfarm level to the PNC series. These extended series were then augmented with the known PNC values for January 2003.

Including the known January 2003 data points created issues with series breaks for many series. This was addressed by wedging the breaks across the January 2001 to December 2002 period. The PNC series were then distributed to MSAs by employment ratios. Each MSA received a percentage of PNC employment equivalent to its percentage of statewide total nonfarm employment.

Accounting for Unclassified Counties (Townships) and Unclassified Industries

Not all LDB records contain precise information as to the location or industry of firms. Some records indicate the state, but not the county in which firms operate. Others may accurately note the state and

county but not the industry. In some cases, records indicate only the state. These firms must be counted when reconstructing MSA employment; however, several questions must be addressed. Namely, it must be decided to which MSAs and to which industries to assign the employment. State offices were given the option of doing their own distribution based on their knowledge of their state, or having the National office perform a standard distribution.

Most states elected to have the National office distribute these employment records. The methodology is as follows. Records that indicated an industry but no county were spread proportionally across the same industry in all MSAs within the state. The distribution was done according to the proportion of the industry's statewide employment contained within each MSA.

Records that indicated a county but no industry were distributed across industries within the MSA that contains the county. The distribution was done according to the percentage of the MSA's total nonfarm employment contained within an individual industry.

In cases where the county and industry indicators were missing, a special distribution was applied. For a given industry within an MSA, the amount of known unclassified employment, divided by the statewide total nonfarm unclassified employment serves as the distribution ratio. This ratio is applied to each month's statewide total of employment from firms with unknown industries and locations and distributed to each industry according to this ratio.

Handling Series Breaks

Potential series breaks in the reconstructed series were identified using the X-12 ARIMA software. Series flagged by X-12 were presented to state Labor Market Information (LMI) offices for review; if state analysts agreed that a series contained a break, then X-12 was used to eliminate the break by applying a level shift to the data. State analysts were also asked to notify the national office regarding series that may contain breaks but were not identified as such by X-12. In such cases, the national office intervened to apply level shifts using X-12 to smooth the time series.

Upon completion of the reconstruction process, state LMI offices were presented with the reconstructed series and given the opportunity to review them for accuracy. State LMI offices were permitted to edit the reconstructed series according to their own knowledge of individual labor markets. Such reviews are

necessary as the distribution of PNC and unclassified employment may not accurately characterize labor markets in all cases - the automated distributions of PNC and unclassified data are “best guess” solutions.

Methodological Consistency with Previous Reconstructions

This methodology is strong in its universal applicability and its consistency. Because information on individual establishments was used to sum the county series, all the industry and geographic detail needed to re-create employment data for MSAs of all sizes was available. Furthermore, this was the methodology applied to the series reconstruction performed when CES converted to the NAICS industry coding system in 2003. This means that the basis for the historical data is the same for all MSAs, even those whose boundaries did not change with the Census 2000 redefinition.

Holding NAICS and ownership codes constant over the span of the historical time series helped keep the data consistent by eliminating all non-economic changes caused by coding error. However, any true economic industry code changes caused by a change in primary business practice were also removed, as were true economic changes in ownership. Since the location of establishments was not held constant, both economic and non-economic location changes were present in the data.

Future Reconstruction Methodology

It is anticipated that future reconstructions will be necessary as OMB continues to revise MSA definitions to match demographic trends in future years. In order to keep CES data relevant, CES will use the above reconstruction methodology to maintain currency with the official MSA definitions.