MONTHLY LABOR REVIEW

October 2016

Reconstruction of CES time series: implementing the 2010 OMB metropolitan area delineations

With the release of January 2015 data, the Current Employment Statistics program at the U.S. Bureau of Labor Statistics incorporated new area delineations from the Office of Management and Budget. Taking into account population and commuting data from the 2010 census, the program added 34 new areas, dropped 15 previously published areas, and changed the geographical scope of 129 areas. Throughout the revisions, the chief aim was to maintain the integrity of series in the redefined areas.

Steven M. Mance
mance.steven@bls.gov

Steven M. Mance is a supervisory economist in the Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics.

John R. Stewart
stewart.john@bls.gov

John R. Stewart is a supervisory economist in the Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics.

The Current Employment Statistics (CES) program is a federal–state cooperative program between the U.S. Bureau of Labor Statistics (BLS) and State Workforce Agencies. Through the CES survey, the program produces data on employment, hours, and earnings at the national level for the 50 states, the District of Columbia, Puerto Rico, the Virgin Islands, and more than 400 metropolitan areas. The program produces some of the timeliest economic indicators each month—usually available 3 to 5 weeks after the reference period—by surveying approximately 146,000 businesses and government agencies that represent about 623,000 individual worksites.

Each year, CES sample-based estimates are benchmarked to universe counts derived primarily from state unemployment insurance (UI) tax records compiled by the BLS Quarterly Census of Employment and Wages (QCEW) program. At the state and metropolitan area levels, the CES program replaces sample-based estimates with a version of QCEW data adjusted to CES definitions and corrected for noneconomic breaks in time series.[1] At the level of total nonfarm employment, CES time series go back to at least 1990 for all metropolitan areas and to 1939 for all states except Alaska and Hawaii. At the national level, most detailed industry time series go back to 1990, although many go back further.

Defining CES areas

The U.S. Office of Management and Budget (OMB) provides federal statistical agencies with common delineations of geographic areas consisting of urban clusters economically integrated with surrounding communities. These delineations in turn afford data users a needed commonality across agencies and databases. The OMB delineations are based primarily upon the concept of a Core Based Statistical Area (CBSA), which is made up of adjacent counties (or equivalent jurisdictions, such as boroughs in Alaska and parishes in Louisiana) having at least one core population of 10,000 or more.[2] Commuting patterns between the urban core and surrounding counties are used to quantify the economic integration of the region, and qualifying adjacent counties are included in the CBSA. Counties can be in only one CBSA, and CBSAs may merge or split over time.

CBSAs fall into two categories: Metropolitan Statistical Areas (MSAs)—that is, urban areas having a population of at least 50,000—and Micropolitan Statistical Areas—that is, urban clusters between 10,000 and 50,000. OMB also defines New England City and Township Areas (NECTAs), using almost identical methodology as that for CBSAs, except with cities and towns instead of counties as the core population. For very large areas containing an urban area of at least 2.5 million, OMB divides MSAs and NECTAs into Metropolitan Divisions (MDs) and NECTA Divisions.

The CES program produces estimates of employment, hours, and earnings for all MSAs and MDs in the nation, except for New England, where the program produces estimates for NECTAs and NECTA Divisions. Although the program does not produce estimates for Micropolitan Statistical Areas, it does provide estimates of employment, hours, and earnings for some nonstandard areas that are not based on OMB definitions.[3] Nonstandard areas can be large municipalities, individual state pieces of cross-state areas, or the residual portions of MSAs not elsewhere defined. Because of their economic importance, and owing to demand from data users, CES maintains data on these areas. Nonstandard areas published by the CES program in 2014 and 2015 are shown in table 1.

Table 1. Nonstandard areas published by CES, 2014 and 2015

2014		2015	
Area code	**Area name**	**Area code**	**Area name**
92581	Baltimore City, MD	92581	Baltimore City, MD
92811	Kansas City, MO	92811	Kansas City, MO
92812	Kansas City, KS	92812	Kansas City, KS
93561	New York City, NY	93561	New York City, NY
93562	Putnam–Rockland–Westchester, NY	93562	Orange–Rockland–Westchester, NY
93563	Bergen–Hudson–Passaic, NJ	93563	Bergen–Hudson–Passaic, NJ
94781	Calvert–Charles–Prince George's, MD	93565	Middlesex–Monmouth–Ocean, NJ
94783	Northern Virginia, VA	94781	Calvert–Charles–Prince George's, MD
97961	Philadelphia City, PA	94783	Northern Virginia, VA
…	…	97961	Philadelphia City, PA
…	…	97962	Delaware County, PA

Source: U.S. Bureau of Labor Statistics.

2010 updates

Because of changes in economic and demographic trends, the delineations of metropolitan areas need to be reassessed frequently in order for them to maintain economic relevance. Areas expand or contract for a number of reasons, all of which can affect the quality of data and information provided by the CES program. Each year, OMB evaluates its definitions of metropolitan areas. Annual updates have been negligible, often not affecting data published by the CES program. However, with each decennial census, OMB receives substantive data updates on population distributions and commuting patterns from the U.S. Census Bureau, prompting a more extensive reassessment of the delineation of metropolitan areas.

In February 2013, OMB incorporated data from the 2010 census and released updates to area delineations.[4] The effect of these delineations on CES areas is shown in table 2, where the categories “changed” and “unchanged” denote areas that were changed or unchanged *geographically*. (Areas that had only administrative changes to their titles or area codes were considered unchanged but were not included in the table.)

Table 2. Impact of 2013 OMB delineations on CES areas

Type of area	Number added	Number dropped	Number unchanged	Number changed
Total	32	12	286	125
Metropolitan Statistical Area (MSA)	24	10	267	82
Metropolitan Division (MD)	5	1	16	7
New England City and Township Area (NECTA)	0	1	0	21
NECTA Division	1	0	0	9
Nonstandard	2	0	3	6

Source: U.S. Bureau of Labor Statistics.

Reconstructing time series by using administrative data

The primary source for reconstructing employment time series is the BLS Longitudinal Database (LDB), which consists of establishment-level microdata from the QCEW and represents all employment covered by the UI system. The LDB contains the state, county, township, ownership (private industry; or federal, state, or local government), and industry codes from the 2012 North American Industry Classification System (NAICS) that were assigned to each establishment in a given quarter. The LDB also contains monthly employment values and other information. The LDB has data on approximately 29 million establishments from 1990 to 2013, including data on business births and deaths. The number of active establishments reporting employment has grown from about 5 million per quarter in 1990 to about 8 million in 2013. The LDB connects businesses reporting to the UI system across time in two ways that aid in reconstructing employment time series. First, establishments that changed UI account numbers but represent the same business location are linked together with a common identifier (a unique “LDB number” for each establishment). Second, the LDB tracks more complicated predecessor–successor relationships where changes in reporting may be administrative rather than economic in nature. These kinds of relationships may exist when old and new UI reporting units share some physical assets but do not represent the exact same worksites. An example is a firm that changes from reporting all of its jobs in one report to reporting separately about individual worksites. The establishments newly reported

on do not represent actual business births, so it would be reasonable to impute some of the predecessor's employment data onto them prior to the date of the administrative change. For time-series reconstruction, each establishment involved in a predecessor–successor transaction was given an adjustment value based on its most recent relationship. For example, if a worksite represented 10 percent of its firm's employment when reporting was broken out in detail, then, prior to that point in time, 10 percent of its firm's reported employment would have been imputed to that worksite. The process was made possible by improvements to the LDB linkage file in the years following the previous reconstruction of OMB areas.

Industry, area, and ownership code changes, which may be for either economic or noneconomic reasons, also occur in the LDB. Economic code changes represent a change in business activity that was denoted in the quarter it occurred. These changes are included in the time series because they are, indeed, economic changes. Often, such changes are large and abrupt. Examples are a business moving its physical location to another county, a factory changing its main product, and the privatization of a hospital. If the changes are found at a time other than when they occurred, they are considered noneconomic. Noneconomic code changes also include fixes for codes that had been assigned erroneously and the initial assignment of codes for establishments that had been unassigned.[5] Unlike economic code changes, noneconomic code changes are administrative in nature and therefore are adjusted before their inclusion in a time series, in order to eliminate series breaks. With the aim of reducing the number of noneconomic breaks, the LDB was adjusted so that each establishment was given its final (i.e., most recently assigned) codes at the same time that a list of economic code changes was compiled.

The sum of LDB employment—adjusted for predecessor–successor transactions—was then tallied for each industry, county, township, and ownership level. To these totals, employment data for LDB records with unclassified county or town codes were distributed on the basis of the proportion of employment in each county and town, for every NAICS and ownership code. Employment data associated with unassigned NAICS codes were distributed proportionally to other industries within a county or town. Records that lacked NAICS and county or town codes were distributed to counties and towns on the basis of their proportion of total CES-assigned employment within the state and then distributed proportionally to all industries.

Employment not covered by the LDB

The scope of employment covered by the UI system and by the CES definition of nonfarm payroll overlap broadly but not entirely: employment found in the LDB accounts for about 98 percent of nonfarm payroll employment and includes some agricultural and household workers who are covered by the unemployment insurance system but do not fit within the CES scope. Unemployment insurance laws vary by state, but examples of the 2 percent of workers who are in the scope of the CES survey but often are not covered by UI laws include employees of religious organizations, elected officials, commissioned insurance sales agents, corporate officers, student employees of colleges and universities, and workers covered under the Railroad Retirement Act. The CES program works with states each year to review UI laws and determine an appropriate noncovered employment (NCE) value for each industry and area.

To determine initial NCE values for reconstructions, the most recent year's NCE values were examined and ratios of noncovered-to-covered employment were derived for every industry and ownership classification. It

was then assumed that similar ratios of noncovered-to-covered employment would hold in substate jurisdictions (counties and townships). Finally, total LDB employment was multiplied by the noncovered-to-covered employment ratios to derive a noncovered-employment level for each industry–area–ownership[6] (IAO) cell. This method accounts for the fact that the distribution of noncovered employment can vary significantly across geographic areas. For example, a small county with a large university would be expected to have more noncovered student workers than a large county without a university.[7]

Regular faculty members with contracts of at least 1 year at primary and secondary schools, colleges, and universities are counted as employed for the entire year in the CES survey, whether or not they receive pay year round. Many school faculty members do not get paid during summer breaks and are not counted under QCEW employment definitions, creating an additional difference in scope that required adjustment.

Noncovered employment totals and summer faculty adjustments were added to the sum of LDB employment, creating employment totals for every possible IAO cell in the country and forming a basis for reconstruction.

Scope of reconstructions

The reconstruction process used existing benchmarked time series as much as possible. Resources for the reconstruction were limited, and a thorough examination of the millions of LDB establishment records within new and changing areas spanning more than 23 years was not practical. Therefore, only jurisdictions (counties or townships) whose area definition changed were directly analyzed.

New areas had to be constructed “from the ground up”: employment at establishments within the area’s boundaries were summed together and adjusted for any noneconomic breaks, to get a grand total of employment in the area. For existing areas that changed, employment in the changing jurisdictions were added to or subtracted from the area’s time series. Some areas merged. When this happened, published histories from the original areas were used to the fullest extent possible and then were adjusted for any additional definitional changes as needed. For example, the Grand Rapids, MI, MSA absorbed the Holland, MI, MSA, which consisted of Ottawa County. In that case, benchmarked employment counts for industries that the two areas had in common were added together, and when an industry was not published for Holland, adjusted LDB data were added to Grand Rapids to account for that industry’s not being published for Holland. In addition to merging, Grand Rapids added one other county and subtracted two counties. In this case, employment in these counties was added to or subtracted from the Grand Rapids MSA in the same fashion as with any other changing area.[8]

CES review of reconstructed time series

A manual review by BLS and State Workforce Agency analysts followed the automated process of summing LDB records, adjusting for predecessor–successor transactions, estimating noncovered employment, and distributing unclassified employment. X-13ARIMA-SEATS, a seasonal adjustment and time-series modeling program maintained by the Census Bureau,[9] was used to scan for additive (point) outliers and level shifts in the net change in the series. Level shifts sometimes represent series breaks—for example, a large predecessor–successor transaction that was not accounted for by the automated process—but may also represent an economic event, such as a strike or the opening or closing of a large business. Additive outliers could be caused

as well by noneconomic events, such as data entry errors. All anomalies were investigated and, if they were shown to be noneconomic, were adjusted by analysts.

During annual CES production, states are responsible for providing noncovered-employment values—for instance, by conducting supplemental surveys or using other resources, such as County Business Patterns, a Census Bureau series that provides subnational economic data by industry.[10] In many cases, states were able to provide better values of noncovered employment for area reconstruction purposes than the initially derived noncovered-employment values. Further, states had previously benchmarked and published data on a number of the Micropolitan Statistical Areas that were reclassified as MSAs. In these cases, BLS used state-published histories to the fullest extent possible. Many of these series did not go back to 1990, however, so they had to be spliced into series derived from the LDB.

Seasonal adjustment

A number of methodological constraints associated with the CES two-step seasonal adjustment process limited the ability of the CES program to provide seasonally adjusted all-employee updates for areas being redelineated and for new areas. Research from the Dallas Federal Reserve has shown that CES benchmarked population data exhibit a seasonal pattern different from that of the sample-based estimates.[11] The benchmarked population data are used in the two-step process to seasonally adjust from the benchmark point back. By contrast, the previously published sample-based estimates are used as input to forecast seasonal factors for the upcoming estimation year. This process of independently adjusting benchmarked data and sample-based data accounts for seasonal differences between the two series and allows for a better seasonal adjustment of sample data in the coming year. The two series are independently adjusted and then spliced together at the benchmark month (in this case, September 2014).[12] However, as with the population reconstructions, areas being redelineated will show breaks in their historical sample-based estimates while new areas will have no historical sample-based estimates. Once the redefined population data were reconstructed, the CES program utilized several statistical techniques to examine differences in seasonality within the population data across area delineations. The aim of such an examination was to learn whether these differences could serve as a proxy for breaks in the sample-based component of a given series. The examination found that areas with greater changes in levels performed more poorly with regard to the test statistics. Therefore, a threshold was set to identify areas that CES analysts could be confident would not experience seasonal breaks due to the new delineations: areas whose geographic compositional change was less than an absolute percent change of 4 percent (as of March 2013) remained eligible to be published on a seasonally adjusted basis, while areas whose change was greater than 4 percent would not be seasonally adjusted in 2015. As a result, the CES program was able to publish 57 of the affected areas on a seasonally adjusted basis. Currently, the program does not provide seasonally adjusted data for 91 areas (59 that are compositionally changing and 32 that are new).[13]

Non–all-employee data

Two methods were developed to reconstruct non–all-employee (non-AE) data: one for new areas and one for preexisting areas. Reconstruction was necessary for preexisting areas that had new geographical compositions under the new delineations. Non-AE data series were not reconstructed for areas for which only a title or code

change occurred. Instead, this type of administrative information was updated, and the previously published data were published for the updated area.

Because the availability of microdata was limited, all new areas were assigned a start year of January 2011. The non-AE microdata were mapped to the new area delineations for the reconstructions. A weighted-link-and-taper estimator was used to create the non-AE time series.[14] This estimator accounts for the over-the-month change in the sampled units, but also includes a tapering feature used to keep the estimates close to the overall sample average over time.

Areas that existed prior to the revised delineation maintained the same publication structures and start dates. Like the new areas, the preexisting areas used the existing microdata and the weighted-link-and-taper estimator from January 2011 forward. Monthly average ratios of the reconstructed series to the previously published series from January 2011 to September 2014 were created and then applied to the previously published history to develop reconstructed histories beginning with December 2010.

Summary

Every 10 years, metropolitan areas in the United States undergo a major redelineation.[15] Subsequently, the CES program must construct employment, hours, and earnings data for new and changed areas to provide users with a time series. The most recent major change came in 2013, and CES data were published in accordance with the new delineations in March 2015.

Employment data were reconstructed back to January 1990, primarily with the use of administrative data adjusted to remove noneconomic breaks. Existing data were used when possible, to further minimize error. Because of different seasonal patterns in the CES survey and administrative data, and because of a lack of survey-based employment histories, seasonally adjusted employment data are currently not published for new areas and areas that have changed substantially. The CES program is currently evaluating the resumption of seasonal adjustment for the latter areas.

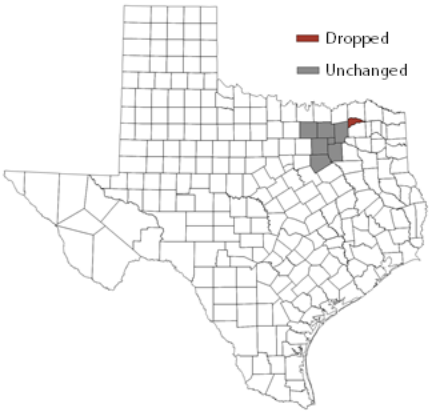
Hours and earnings data for new and changed areas were reconstructed with the use of CES survey data because there were no available administrative data. For hours and earnings series in changed areas, existing histories were spliced together; new hours and earnings could be reconstructed only back to 2011.

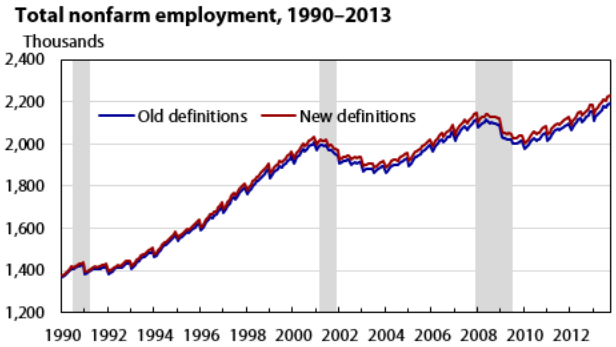
Metropolitan areas will continue to undergo periodic changes in delineation. When that happens, CES data will need to be periodically reconstructed in order to maintain their relevance.

Appendix 1

Appendix 1: Examples of redefined areas

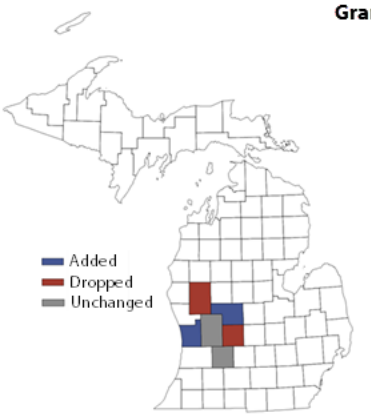
Dallas-Plano-Irving, TX

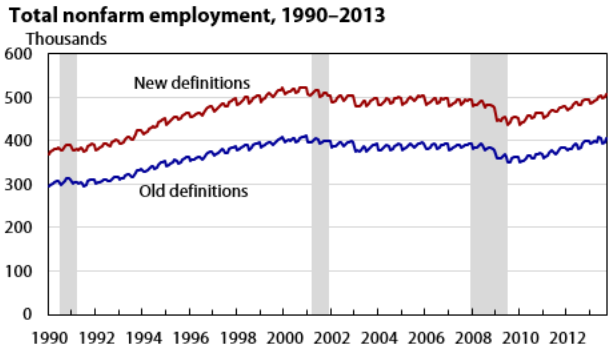




Note: Shaded areas represent recessions as defined by NBER.

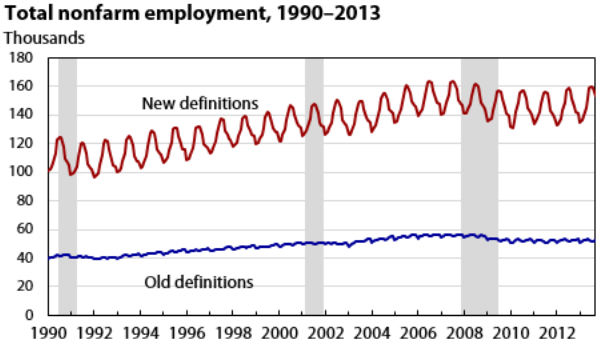
Grand Rapids-Wyoming, MI





Note: Shaded areas represent recessions as defined by NBER.

Salisbury, MD-DE



Note: Shaded areas represent recessions as defined by NBER.

Note: The Dallas–Plano–Irving, TX, Metropolitan Division dropped one outlying county. Employment in this area was little changed by redefinition. Grand Rapids, MI, absorbed the Holland Metropolitan Statistical Area (Ottawa County), dropped two counties, and added one other county. Employment in this area is somewhat lower than the sum of employment in the Grand Rapids area and employment in the Holland area under the old definitions. The Salisbury, MD–DE, Metropolitan Statistical Area added two counties along the Atlantic coast. After redefinition, employment in this area more than doubled, and took on a seasonal pattern characteristic of resort areas.

Source: U.S. Bureau of Labor Statistics.

Appendix 2

Comparison of methodologies from January 2005 redefinition to January 2015 redefinition

Task	January 2005 redefinition	January 2015 redefinition
Holding industry– area– ownership codes constant	Industry and ownership were held constant for establishments.	Industry, ownership, county, and township were held constant for establishments; economic code changes were reintroduced manually.
Construction method for changing areas	Areas with one county or town added or subtracted had employment for that county or town added or subtracted; areas with other changes were completely reconstructed from microdata.	Net changes in changing counties or towns were added or subtracted from the areas; only new areas were constructed completely from microdata.
Series breaks	X-12-ARIMA was used to identify level shifts; if State analysts believed that the shifts were noneconomic, the series was adjusted by the amount of the shift.	X-13ARIMA-SEATS was used to scan for anomalies in the net series changes; national office and state analysts investigated the microdata in these series and made adjustments when necessary.
Predecessor– successor transactions	Predecessor–successor transactions were smoothed primarily with the use of level shifts identified by X-12-ARIMA.	Linkage files were used to smooth predecessor– successor transactions when possible; complex transactions sometimes required manual adjustments.
Estimation of noncovered employment	Counties and townships were given estimates of noncovered employment that were proportional to their total employment in the state.	Counties and townships were assumed to have the same ratio of covered-to-noncovered employment as the statewide ratio at the NAICS or ownership level, except for NAICS 482 (rail transportation) and 813 (religious, grantmaking, civic, professional, and similar organizations). For these industries, counties and towns were given estimates of noncovered employment that were proportional to their total employment in the state.
Seasonal adjustment	Substate areas were not seasonally adjusted at this time. Areas that did not change with this redefinition were seasonally adjusted starting in 2007.	Seasonal adjustment of unchanged areas and areas that changed by less than ±4 percent as of March 2013 would continue to be adjusted; areas that exceeded this threshold and new areas were not seasonally adjusted in 2015.
Hours and earnings	There was no reconstruction of hours and earnings series. Hours and earnings series were not produced for new areas. Hours and earnings series for areas that split into multiple areas were evaluated individually.	Two methods were used for reconstructing non–all-employee estimates, one for new areas and one for preexisting areas.

Source: U.S. Bureau of Labor Statistics.

NOTES

1 The CES national benchmarking method differs from the method used for states and areas. Instead of replacing each month's data with an adjusted QCEW value, national data are benchmarked with the use of a wedge method described in chapter 2 of the *BLS Handbook of Methods* (U.S. Bureau of Labor Statistics), http://www.bls.gov/opub/hom/pdf/homch2.pdf.

2 A *core population* (alternatively, core county or core set of counties) (1) contains at least half of its population in urban areas of 10,000 or more or (2) has a population of at least 5,000 people within its boundaries, which are located in an urban area of at least 10,000. A full explanation of how OMB defined the areas used by the CES program can be found in "2010 standards for delineating Metropolitan and Micropolitan Statistical Areas," *Federal Register*, Vol. 75, No. 123, June 28, 2010, https://federalregister.gov/a/2010-15605.

3 In 2014, the CES program published data on 374 MSAs, 29 MDs, 22 NECTAs, 9 NECTA Divisions, and 9 nonstandard areas. In 2015, the program published data on 373 MSAs, 28 MDs, 21 NECTAs, 10 NECTA Divisions, and 11 nonstandard areas. In both years, the CES program published data on a total of 443 areas.

4 The new delineations are outlined in *Revised delineations of Metropolitan Statistical Areas, Micropolitan Statistical Areas, and Combined Statistical Areas, and guidance on uses of the delineations of these areas*, OMB Bulletin 13-01 (U.S. Office of Management and Budget, February 28, 2013), https://www.whitehouse.gov/sites/default/files/omb/bulletins/2013/b13-01.pdf.

5 Microdata in the LDB may be missing county, township, or industry codes.

6 Industry is determined by NAICS code; area by state, county and township. Ownership is broken down by private industry and federal, state, and local government.

7 Because NAICS 492 (rail transportation) and NAICS 813 (religious organizations) had no covered employment, the method just described was not used for them. Ratios of noncovered employment in those industries to total nonfarm employment were calculated at the statewide level and applied to substate jurisdictions.

8 See appendix 1 for detailed examples of changes to three areas, including the Grand Rapids MSA.

9 The X-13ARIMA-SEATS program is available for free from the U.S. Census Bureau. (See "X-13ARIMA-SEATS seasonal adjustment program" (U.S. Census Bureau), https://www.census.gov/srd/www/x13as/.

10 See *County Business Patterns (CBP)* (U.S. Census Bureau, last revised September 7, 2016), http://www.census.gov/programs-surveys/cbp/about.html.

11 See Franklin D. Berger and Keith R. Phillips, "Solving the mystery of the disappearing January blip in state employment data" (Federal Reserve Bank of Dallas, *Economic Review*, Second Quarter 1994, pp. 53–62, http://www.dallasfed.org/assets/documents/research/er/1994/er9402d.pdf.

12 The two-step seasonal adjustment process is explained in detail in Stuart Scott, George Stamas, Thomas J. Sullivan, and Paul Chester, *Seasonal adjustment of hybrid economic time series* (U.S. Bureau of Labor Statistics), http://www.bls.gov/osmr/pdf/st940350.pdf.

13 For more information, see Larry Akinyooye, Ryan Arbuckle, and Albert Kleine, *Revisions in state establishment-based employment estimates effective January 2015* (U.S. Bureau of Labor Statistics), http://www.bls.gov/sae/benchmark2015.pdf, especially "Seasonal adjustment," p. 5.

14 See *BLS Handbook of Methods*, chapter 2.

15 See appendix 2 for a comparison of the methodology described in this article with the methodology the CES program used to incorporate the previous set of area delineations in January 2005.

RELATED CONTENT

Related Articles

Current Employment Statistics data and their contributions as key economic indicators, *Monthly Labor Review*, March 2016.

Comparing with the original: a look at Current Employment Statistics vintage data, *Monthly Labor Review*, April 2015.

Analysis of the Current Employment Statistics program using customer outreach survey results, *Monthly Labor Review*, January 2015.

Current Employment Statistics seasonal adjustment and the 2007–2009 recession, *Monthly Labor Review*, October 2012.

Related Subjects

Employment | Statistical methods | Seasonal adjustment | Benchmarking | BLS Programs and surveys | Hours of work