Monthly Labor Review

November 2017

Benchmarking the Current Employment Statistics survey: perspectives on current research

In the sample-survey world, benchmarking is an alignment of a survey-based estimate with a population value. Aligning estimates to a high-quality control total improves the quality of the set of estimates for both the controlled and correlated characteristics. This article discusses benchmarking in the Current Employment Statistics (CES) survey, from its roots in 1935 to current benchmark procedures. The general practice of CES benchmarking is briefly described and is accompanied by recent quantitative results at national and state levels. Impediments to improvements in the methodology are discussed and illustrated. The article concludes with a brief overview of methods included in recent research and with an outline of future plans.

Kenneth W. Robertson
robertson.ken@bls.gov

Kenneth W. Robertson is the Assistant Commissioner for the Office of Industry Employment Statistics, U.S. Bureau of Labor Statistics.

What is benchmarking?

A benchmark is generally described as a standard of excellence against which other things are judged. In a statistical sense, a benchmark is a high-quality population value against which a survey estimate may be judged in order to assess the quality of the estimate.

For surveys, a benchmark is something obtained outside the survey. For example, a high-quality sampling frame—a list of all those within a population who can be sampled—usually includes a population measure or census count correlated to a major statistic to be collected. Many surveys may select a sample from a population list (at time *t*), take the requisite time to collect the survey data, and then align the survey estimates with population values from an updated population list (at time *t + 1*) to make the estimates as current, relevant, and accurate as possible when published. This is one type of survey benchmarking.

In the Current Employment Statistics (CES) survey, the U.S. Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW) serves as both the population list and the primary source of benchmark employment. The QCEW covers about 97 percent of the employment that is in scope for CES estimates.

CES background

The CES survey is a quick-response business survey that provides some of the earliest information about the state of the U.S. economy each month. BLS produces and publishes the initial estimates from the survey about 3 weeks after the reference period. The reference period is the payroll period that includes the 12th of the month. The survey produces employment, hours, and earnings data by detailed industry levels. About 2 weeks later, data are produced for states and metropolitan statistical areas. The national data are revised two times on the basis of additional data collected for the reference period, and the subnational data are revised once. BLS produces about 50,000 data series each month, using the business reports collected for the CES program.

The primary estimator for national employment data is referred to as a weighted link-relative estimator. It takes the ratio of current-month weighted employment to prior-month weighted employment (based on a matched sample of collected data) and multiplies that quotient by the prior-month employment level. An estimated adjustment value is then added to account for the net employment change in businesses not captured by the survey; this estimated adjustment value is called the net birth/death factor. The primary employment estimator,[1] shown just below, provides an efficient estimate of over-the-month (OTM) change and is also referred to as the "change ratio" for month *c*:

$$\hat{R}_c = \frac{\sum_i w_i e_{c,i}}{\sum_i w_i e_{p,i}} \text{ ,(1)}$$

where

i is an establishment that reported both this month and last month,

w_i is the sampling weight associated with establishment *i*,

c is the current month,

p is the prior month,

$e_{c,i}$ is the employment for current month *c* for establishment *i*, and

$e_{p,i}$ is the employment for prior month *p* for establishment *i*.

The weighted link-relative estimator, shown below, is composed of

$\hat{R}_c$ and two additional components.

$$\widehat{E_c} = \widehat{E_p}\hat{R}_c + b_c \text{ ,(2)}$$

where

$\widehat{E_c}$ is the employment estimate for current month *c*,

$\widehat{E_p}$ is the employment estimate for prior month *p*,

$\hat{R}_c$ is the change ratio for current month *c* (equation 1), and

b_c is the net birth/death factor for month *c*.

The primary estimator for state and metropolitan area employment estimates, a derivative of the weighted link-relative estimator, includes a robust procedure for identifying and reducing the influence of unusual reports (or outliers).

Although the link-relative estimator is an efficient estimator for employment change, errors may accumulate over time because each estimate is linked to the one before it. The hope, of course, is that the errors are random and offsetting—that is, some errors on the positive side are balanced by some errors on the negative side. To ensure that the accumulated error does not get too large in any direction, the CES program benchmarks the survey estimate level by using a lagged population value.

Benchmarking is used to align a survey estimate with a census value. This generalized statement is, however, overly simplified in some cases. It assumes that the census value used in the alignment is essentially without error. In reality, the CES sample estimates include sampling and nonsampling errors, and the QCEW census data include nonsampling errors.[2] Therefore, the CES benchmark can best be described as an alignment of CES estimates with an independently derived employment count. Nonetheless, both sources of data are subject to error.

This article describes some of the history associated with CES benchmarking, the current process used, and research efforts to improve that process.

A brief history of CES benchmarking

The CES survey was first envisioned as an employment index for a limited number of industries: boots and shoes, cotton goods, cotton finishing, hosiery and underwear, and iron and steel. BLS started the survey in 1915, and at that time there was no plan in place to benchmark the data against another source of employment information. In 1935, however, BLS implemented the first benchmark procedure after identifying a very large downward bias in the manufacturing index of about 12 percent over the period from 1923 to 1929.[3] This first benchmark incorporated approximated census levels for each of the years 1923, 1925, 1927, 1929, and 1931. The censuses of manufactures and the censuses of businesses provided the data used in this first benchmark.

Although this benchmark procedure improved the quality of the CES data, the procedure did not account for the full scope of the survey because the benchmark source data did not cover the full nonfarm business population. In 1935, Congress passed the Social Security Act, which established the unemployment insurance (UI) program; data became available to BLS from the new program in about 1940. These new UI data included industries not covered in the census of manufactures and the census of businesses, making the UI data the preferred source of benchmark data for the CES program. The UI data from each state are edited by states and collected by BLS through the QCEW program.

Originally, both the national and state data series were benchmarked only to March QCEW data.[4] The benchmark procedure essentially replaced the prior year CES March estimate with the QCEW March level and then used a linear wedge to distribute the correction amount to the 11 months before that March. The implicit assumption with this procedure is that errors accumulated at a steady rate over the year before the benchmark. The benchmark with linear wedge procedure is applied as follows:

$$\hat{E}_m^B = \hat{E}_m + \frac{m}{12}\left(P_M - \hat{E}_M\right), \quad (3)$$

where

$\hat{E}_m^B$ is the benchmarked all-employee level for month *m,*

$\hat{E}_m$ is the not-yet-benchmarked all-employee estimate for month *m,*

m is the month being benchmarked (1 is April, 2 is May, …, 11 is February, 12 is March),

P_M is the March "population" value, which is QCEW employment plus the number of jobs not covered by the UI system (about 3 percent of nonfarm jobs), and

$\hat{E}_M$ is the not-yet-benchmarked March all-employee estimate.

Around 1980, all 12 months of UI data became available to the QCEW program. Also at that time, individual states began transitioning from a March replacement with wedge procedure to a 12-month replacement procedure for benchmarking. Meanwhile, the national procedure remained unchanged. By the late 1980s, all states had switched to the replacement method. State and metropolitan area estimates changed to the replacement method because BLS and state analysts believed the error in the administrative data was preferable to the sampling error that remains in the benchmarked data for months other than March.[5] When states began replacing their CES data with QCEW data for all months, BLS and state analysts assumed that the administrative error in the QCEW was smaller than the sampling error in CES, especially for smaller domains. For the larger domains, specifically for national-level data series, a different assumption was made: the best estimate of over-the-month change was the CES estimate, and a level correction with a linear wedge applied once per year was sufficient to ensure high-quality historical estimates.

Current research is focused on answering two questions. First, is there a procedure that would allow for the use of more frequent alignment to the population data that consistently improves the quality of the CES data? And second, is there a procedure that would allow for more consistency between the benchmark procedures used at the national and state levels?

Current methodology and some qualitative results

The current methodology for national estimates differs substantially from that for state and area estimates.

National methodology

The national procedure takes the prior March QCEW level for each industry and adds employment not covered by the UI system. This noncovered employment level is mostly obtained by taking the difference between the QCEW data and older data from the U.S. Census Bureau *County Business Patterns* and then forecasting that difference forward to the benchmark period.[6] The Railroad Retirement Board is another source of noncovered employment data.[7] This census employment value for the prior March then replaces the CES survey estimate, and the difference between the original estimate and the benchmark value is wedged into the 11 months preceding March (see equation 3).

Estimates for the postbenchmark period must also be adjusted to the new March level.[8] This is done by taking the already estimated over-the-month change ratio and newly forecasted net birth/death factors for each month and applying them to the new level and repeating this process moving forward to the most recent estimate. That is, in equation 2, $\widehat{E_p}$ takes on a new value as you move the new levels forward until you get to the most recent month.

More on the current national methodology is available in an article by Chris Manning and John Stewart and in the national program's annual benchmark article.[9]

State and area methodology

The state and area procedure is simpler in concept than the national procedure, although it is also more resource intensive. This procedure develops a census employment value by following the same process used by the national procedure for March but does that for all 12 months. These census data then replace the CES survey estimates.

The postbenchmark period estimates also receive different treatment for the state and area data. Preliminary QCEW data are used to develop population values for the April-to-September period following the procedure used to develop the March values, and these values replace the CES estimates. For the October-to-December period, the values are updated by using procedures similar to those used for national estimates. That is, for this period, estimates are developed that use (1) the new benchmark level, (2) all data collected for the reference periods following the benchmark, and (3) newly forecasted net birth/death factors to update the estimates. More on the current state and area methodology is available in an article by Kirk Mueller and in the annual CES state-and-area benchmark article.[10]

Some quantitative results

One interpretation of the benchmark revision is that it serves as a proxy for total survey error. However, this interpretation does not consider error in the administrative data that is the source of the benchmark data. Taking errors in both data sources into account, BLS interprets the CES benchmark revision as the difference between two independently derived employment counts, each subject to its own sources of error. QCEW errors accrue from the usual errors associated with large administrative datasets: comprehension and collection errors, data entry and coding errors, imputation errors, and other processing errors. Both BLS and its state partners work hard to minimize those errors, but they still exist to some extent.

Some results for CES benchmark revisions are shown below. These are percentage benchmark revisions for national total nonfarm employment in March, 2005–15:

Year	Percent
2005	-0.1
2006	0.6
2007	-0.2
2008	-0.1
2009	-0.7
2010	-0.3
2011	0.1
2012	0.3
2013	-0.1
2014	0.0
2015	-0.1

National revisions generally are small. Over the period depicted, the 5-year average of the absolute values of the percentage revisions, hereafter called absolute average percentage revisions (AAPRs), has ranged from 0.4 percent in 2010 to 0.1 percent in 2015.

Here is the AAPR for states, for total nonfarm employment in March, 2005–15:

Year	AAPR
2005	0.5
2006	0.5
2007	0.4
2008	0.4
2009	0.9
2010	0.4
2011	0.5
2012	0.7
2013	0.4
2014	0.5
2015	0.4

As expected, average state revisions tend to be a bit larger in absolute value than national revisions.

For metropolitan statistical areas (MSAs), revisions are presented by MSA size as well as total. MSAs with smaller employment levels tend to have smaller sample sizes and hence a higher relative sampling error (see table 1).

Table 1. Absolute average percentage revisions for metropolitan statistical areas (MSAs), for total nonfarm employment, March, 2005–15

MSA employment size	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
All sizes	1.1	1.1	0.9	1.0	1.8	1.1	1.1	1.6	1.2	1.1	1.0
Less than 100,000	1.3	1.3	1.1	1.3	2.1	1.4	1.4	2.0	1.4	1.4	1.2
100,000–499,999	1.1	0.4	0.8	0.9	1.6	0.9	0.9	1.5	1.1	0.9	0.8
500,000–999,999	0.6	0.6	0.7	0.6	1.4	0.7	0.8	0.9	0.8	0.9	0.8
More than 1,000,000	0.7	0.5	0.6	0.5	0.9	0.5	0.7	0.7	1.1	0.7	0.4

Source: U.S. Bureau of Labor Statistics.

The largest MSAs generally have revisions that are near the size of the absolute average state revision. The smallest MSAs have revisions that are larger on average, and a few of the smallest MSAs each year can have large revisions. The range of percentage revisions, as shown in table 2, illustrates this.

Table 2. Benchmark revisions for total nonfarm employment in metropolitan statistical areas (MSAs), March 2015

Item	Number of employees				
	All sizes	Less than 100,000	100,000–499,999	500,000–999,999	More than 1,000,000
Number of MSAs	387	188	147	22	30
Absolute average percentage revision	1.0	1.2	0.8	0.8	0.4
Range of revisions	-6.4 to 6.0	-6.4 to 6.0	-3.7 to 2.8	-1.1 to 1.6	-1.9 to 0.7

Source: U.S. Bureau of Labor Statistics.

Benchmarking by data users

Benchmarking of CES data used to be the sole province of BLS. The process formerly taxed the capabilities of mainframe systems and data processing budgets by processing thousands of data series and updating years of historical data. However, with today's powerful desktop and laptop computers, simple macro-level benchmarking is a process available to almost anyone with the technical savvy to download the appropriate datasets and apply a well-considered update process to the data.

Several of the Federal Reserve regional banks conduct a quarterly benchmarking process for state and area estimates before BLS issues the official annual benchmarks.[11] The process, in general, can be described as being made up of three parts. The first part is the oldest data, comprising data already benchmarked by BLS. The second part is data for which QCEW data are available but have not yet been incorporated into the time series. The third and most recent part is data for which QCEW data are not yet available. The process uses QCEW growth rates (similar to a change ratio) to replace the second part and CES growth rates to adjust the third part.

More specifically,

$$\hat{\ddot{E}}_v = \hat{\ddot{E}}_p \hat{\ddot{R}}_v ,$$

where

$\hat{\ddot{E}}_v$ is the early benchmarked CES value for month v,

$\hat{\ddot{E}}_p$ is the prior month level, perhaps produced as a result of early benchmarking or a CES value as originally published,

$$\hat{\ddot{R}}_v = \left\{ \begin{array}{l} \widehat{CES}_{E,v} / \widehat{CES}_{E,v-1} \text{; if part 1—data are already benchmarked by BLS—use this, or} \\ \widehat{QCEW}_{E,v} / \widehat{QCEW}_{E,v-1} \text{; if part 2—QCEW data are available but not incorporated—use this, or} \\ \widehat{CES}_{E,v} / \widehat{CES}_{E,v-1} \text{; if part 3—QCEW data are not yet available—use this.} \end{array} \right\},$$

$\widehat{CES}_{E,v}$ is the published CES all-employee estimate for month v, and

$\widehat{QCEW}_{E,v}$ is the published QCEW all-employee estimate for month v.

This procedure can also be described as follows:[12]

1. Begin with the CES state time series at some level of industry detail, and identify the parts of the series as described above.
2. Where QCEW data are available but not yet incorporated into CES data (i.e., for part 2), use the QCEW data to calculate over-the-month (OTM) change ratios or links. The calculation is link$_t$ = E_t / E_{t-1}. This calculation will serve as a proxy for the eventually benchmarked OTM changes.
3. Multiply the last BLS benchmarked value by the first QCEW OTM link to replace the CES estimate, then multiply the product times the next link, and so on, until you update through the most recently available QCEW data.

4. For data in part 3, create CES OTM links in the same manner as that used to create QCEW OTM links forward from the latest QCEW replacement value, and apply them from that point forward to the most recent estimate.
5. Calculate the seasonal factors for all three parts of the series described above. Using ARIMA X-12 or ARIMA X-13-SEATS,[13] seasonally adjust the three parts described, and then link the parts as described in (1) to (4). Calculating seasonal factors for the first two parts (the officially benchmarked series and the QCEW links) is straightforward because long time series are available for these data. The third part is more difficult. There are two options for this part: (a) create a time series consisting of historical values of nonbenchmarked CES data or (b) simply use the changes in the official seasonally adjusted BLS data to extrapolate the seasonally adjusted CES data past the early benchmark period. Option (b) is the easiest solution, but it limits potential industry detail to those industries that BLS seasonally adjusts, which is typically the broad industry level.

The intent of *early* benchmarking is to take QCEW data published every quarter and incorporate them into the CES data more often than once per year. The expectation is that this process will reduce the error in the portion of the linked CES data thus updated, providing a data series that will be closer to what BLS will eventually publish after it benchmarks the data. As noted above, some entities have been doing early benchmarking for years. Researchers at the Federal Reserve Bank of Dallas, for example, have been exploring methodologies and applying early benchmarking since at least 1993.[14] Since then, early benchmarking has been undertaken by researchers at other Federal Reserve banks and at several universities.

Impediments to improving the methodology—why is this hard?

Data users might well ask: “Why hasn’t BLS started benchmarking more often than once per year?” The answer is twofold. First, there are significant resource considerations. Benchmarking, while technically easy to describe and implement at a macrolevel, requires a considerable amount of BLS resources. BLS must carefully review the initial microlevel inputs for accuracy, develop and review forecasts for some inputs (e.g., *County Business Patterns*), and carefully review the results before publication. More than 50,000 CES series are reviewed as part of the benchmark process. BLS data are considered the gold standard of economic labor market data. This standard is maintained because the agency takes great pains to ensure that the data it releases are as accurate and complete as possible. BLS staffing is currently funded to conduct this “peak-load” benchmarking process once per year. To do this process more than once per year means more funding would be needed. Without additional funding, BLS would need to reduce the number of published data series in order to offset the increased workload.

The second consideration is significantly more technical in nature. BLS has known, at least since 1993 when two Dallas Federal Reserve Bank researchers documented and discussed the issue with BLS, that data from the QCEW and the CES survey have different seasonal patterns.[15] A Response Analysis Survey (RAS) was conducted several times to explore differences between CES and QCEW estimates.[16] The most recent RAS was conducted in 2008, and a number of papers and analyses were based on this survey.[17] The RAS results consistently highlight several common themes.

First, when exploring reports with *different* employment responses for the CES survey and the QCEW, the RAS results show that about one-third of the time the CES report was correct, one-third of the time the QCEW report was correct, and about one-third of the time neither was exactly correct.[18] There are many reasons a report might be in error. For example, different people in the business may respond to the two programs on the basis of data

from different internal databases. Also, a report to one requester might include employees that are out of scope (e.g., by including contract employees) while the correct scope is reported to another requester.

Second, a consistent RAS result over time is that the QCEW has a larger proportion than the CES survey of respondents reporting for pay periods that occur after the reference period. That is, some reports may be for the pay period that follows the one that includes the 12th day of the month, some may be an end-of-month count, etc.[19] A side effect of this pay-period misreporting is that the QCEW would show more employment gain during a seasonal hiring month (such as December) and more employment loss during a seasonal layoff month (such as January). This effect is, in fact, what we see in the QCEW. For example, the QCEW shows more employment gain than the CES survey in December and more employment loss in January.

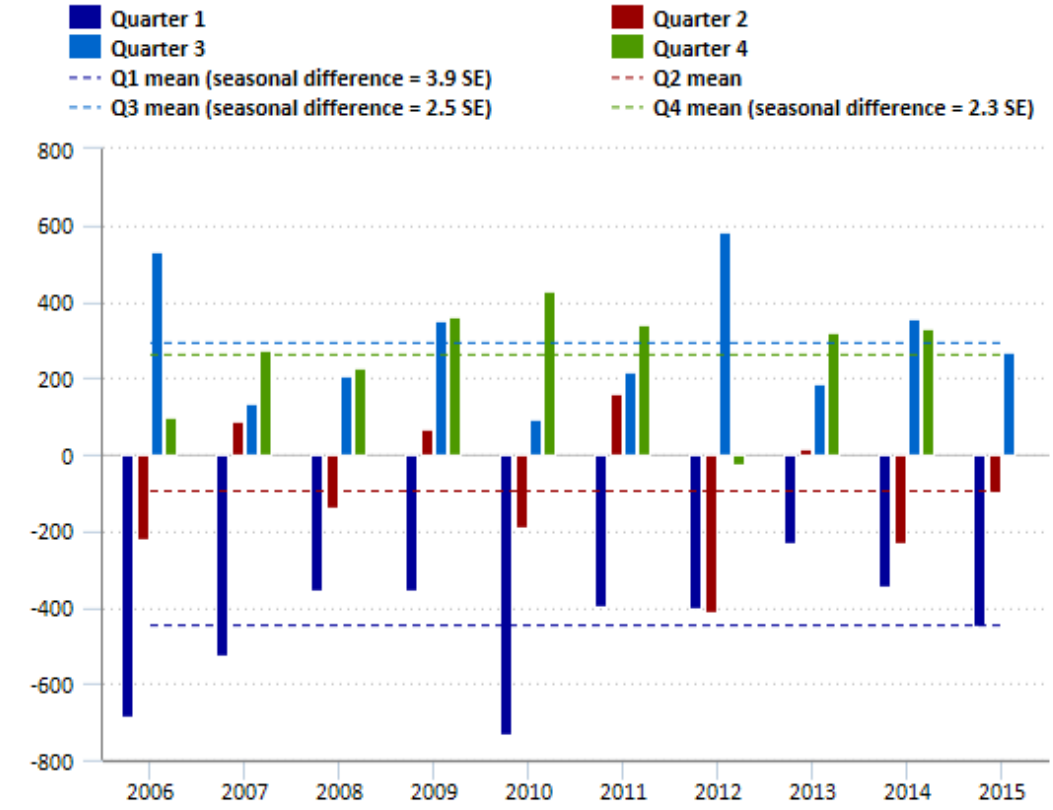
I note here that this is a minor issue for the QCEW. The goal of the QCEW is to report total employment and wages each quarter. The primary purpose for this data collection is to administer state unemployment insurance (UI) programs. State UI staff collect the data quarterly and provide them to state Labor Market Information (LMI) staff. These data are not intended to be a monthly report of employment change; they are primarily intended to support the collection of UI taxes. After supporting their primary purpose, these administrative data are leveraged to support additional uses. For example, the data can provide additional value for the LMI community by serving as a comprehensive source of employment and wage data by industry and county. The state LMI staff work with BLS to review these administrative data and correct errors and omissions. If we assumed the difference between the CES survey and the QCEW to be pay-period reporting error by the QCEW, then the QCEW is still 99.7 percent correct or better in accuracy in its worst month.[20]

Another interpretation of this is that, for the QCEW, the reports are nearly 100 percent accurate with some reporting-period error. That is, the QCEW reporter is reporting accurate employment monthly counts, just sometimes not for the requested pay period. However, for the CES survey, whose goal is to measure the month-to-month change in employment (for the week that includes the 12th of the month), a 0.3-percent error can lead to huge adjustments to the CES data. For context, the average CES over-the-month change is about 0.6 percent of the employment level. A 0.3-percentage-point error on a 0.6-percent average change suggests that as much as a 50-percent adjustment in the OTM change (not seasonally adjusted) is required for some months to account for the seasonal differences in these data sources. Perception is important, and, historically, the small seasonal errors in the QCEW employment levels have been perceived by some state data users as very large errors in the CES employment estimates even though the QCEW errors actually are small.

Figure 1 summarizes these differences between QCEW and CES data. The differences are depicted on a quarterly basis because the benchmarking procedures being considered use only the third month of each quarter of QCEW data. It is important to note that the data shown in this figure present differences between *benchmarked* CES national data and QCEW data; that is, the CES data series have already been corrected for errors in the level of employment. Therefore, what this figure shows are the differences in CES and QCEW data after the level correction has taken place. Further note that the data labeled as “QCEW (sum-of-state)” are the sum-of-the-states benchmarked CES data. Because this is, by design, QCEW data plus noncovered employment, these data serve as a high-quality proxy for over-the-quarter employment change in QCEW data. The figure graphs the difference in over-the-quarter change between QCEW and benchmarked CES levels. It presents the different quarters highlighted by different colors (bars), as well as the average difference for each quarter over time (shown by the dashed lines). As the figure shows, the average over-the-quarter difference between QCEW and CES data for the

first quarter is about −447,000. That is, the QCEW regularly shows a loss of employment that is much larger than the CES loss of employment from December to March. This difference is about 3.9 standard errors (for a CES 3-month change). Given that this large and significant discrepancy is a regular difference between QCEW and CES levels, we interpret this as a normal seasonal difference between these data series, not as sampling error for the CES data. Large seasonal differences occur for other quarters. For example, the third quarter difference is 292,000 (or 2.5 standard errors), and the fourth quarter difference is 261,000 (or 2.3 standard errors).

Figure 1. The difference in the over-the-quarter change between Quarterly Census of Employment and Wages data and Current Employment Statistics data, not seasonally adjusted



Source: U.S. Bureau of Labor Statistics.

One impact of these large seasonal differences is evident in the state CES data because of the replacement of CES data with QCEW data during each annual benchmark. This process results in two different seasonal patterns in every state CES not seasonally adjusted series: a QCEW seasonal pattern from the benchmark point back and a CES seasonal pattern from the benchmark point forward. This issue is corrected in seasonally adjusted data by the use of the two-step seasonal adjustment process.[21] A less obvious but important impact of these large seasonal differences lies with the interpretation of the over-the-year changes. If seasonally adjusted data are available, then an analyst may appropriately interpret the over-the-year change by using the 12-month employment difference in the seasonally adjusted series. However, if a seasonally adjusted series is not available, then the analyst must use caution when interpreting over-the-year change if the analysis crosses the splice

point.[22] Here, the value calculated is actually the sum of the over-the-year change plus a value associated with the difference in seasonality between the CES and QCEW components.

These large, regular seasonal differences stymied historical research efforts to define a benchmark procedure that would address the differences in an acceptable way. Basically, historical CES research efforts focused on the exploration of alternative procedures to adjust the not seasonally adjusted CES data to not seasonally adjusted QCEW data. Each of those efforts ended with a recommendation not to implement those methods because CES data quality did not improve. Only after those straightforward efforts failed did we turn to an examination of procedures that more directly address the seasonal differences.

Recent research

BLS has conducted periodic research on CES benchmarking for several decades. Much of that research has explored alternative benchmark procedures that focus on not seasonally adjusted data. For example, research has applied the state replacement method at the national level and similarly has applied the national wedge procedure for states to explore the errors that are incorporated into the CES estimates. After several iterations of this research, we have concluded that neither of these methods is optimal.

What makes a benchmark procedure optimal or what features of a benchmark procedure are most important are useful questions to consider. Originally, the focus was simply on which method was “correct” without defining what that meant. More recently, I have identified a set of features (listed below) that I consider important in determining the usefulness of a new procedure.

1. The procedure must align to the March QCEW level.
2. The procedure must have a well-defined mathematical relationship to the third month of each quarter of the QCEW.
3. The procedure must align the CES data to QCEW data more than once per year.
4. The procedure must address the different seasonal patterns in the data and maintain the CES seasonal pattern.
5. The procedure must address the small-domain problem.
6. The procedure must be explainable from a statistical and economic perspective.

Element 1 requires the alignment of the CES level to the March QCEW level. This requirement is consistent with both the national and state benchmarking procedures, and I consider this a critical requirement for any future procedure. Element 2 refers to a set of procedures that create a relationship between CES and QCEW data for other quarters. One such relationship is the direct alignment to the third month of each quarter that is consistent with the current state procedure. Element 3 is self-explanatory; the procedure must align the data more than once, which means either twice per year or four times per year. Element 4 indicates that the procedure must account for the different seasonal patterns between these data sources and maintain the CES seasonal pattern where the domain is large enough to support its identification. Element 5 indicates that the procedure must account for the small-domain problem. This is the problem that led to the use of the current replacement method for states. This problem is focused on small domains with the smallest samples. Small domains that also have very small samples may have very large relative sampling errors. The problem is that aligning the data once per year or even for the third month of each quarter in these domains can leave very large and uncorrected sampling errors in place for

other months. A new procedure must explicitly include a method to mitigate these errors. And finally, element 6 notes that a new procedure must make sense and be relatively easy to explain to data users from both a statistical and economic perspective. While not a criterion for evaluation, a new procedure will be most useful if it can be applied consistently to both national and state data.

We can evaluate alternative procedures against these criteria to see how well they achieve this set of goals. To frame that evaluation, I start with an evaluation of the current procedures.

This is an evaluation of the national procedure:

- ☑ Aligns the CES data to the QCEW data in March
- ☒ Has a well-defined mathematical relationship with the third month of other quarters
- ☒ Aligns CES and QCEW employment more than once per year
- ☑ Maintains the CES seasonal pattern
- ☒ Addresses the small-domain problem
- ☑ Is easily explained

Therefore, the national procedure meets only three of our six criteria.

This is an evaluation of the state procedure:

- ☑ Aligns the CES data to the QCEW in March
- ☑ Has a well-defined mathematical relationship with the third month of other quarters
- ☒ Aligns CES and QCEW employment more than once per year
- ☒ Maintains the CES seasonal pattern
- ☑ Addresses the small-domain problem
- ☑ Is easily explained

The state procedure meets only four of our six criteria.

As stated earlier, neither the national nor state procedures currently used is optimal, and with this set of defined criteria we can evaluate the usefulness of proposed alternatives. I now turn to a brief description and evaluation of procedures reviewed during our most recent round of research. Each of the new procedures is quarterly, and each procedure aligns directly to the March QCEW level, so they meet at least two of our six criteria by definition.

The procedures proposed are (1) Over-the-Year-Change Difference,[23] (2) Weighted Model, and (3) Seasonally Adjusted Difference in Employment Method. These procedures are described below.

The Over-the-Year-Change Difference procedure takes the difference in over-the-year change for the third month of each quarter in CES and in QCEW levels and aligns the CES over-the-year change to the QCEW over-the-year change. It was initially thought that this method created an appropriate relationship between CES and QCEW levels for each quarter. However, further research identified a problem: the seasonal pattern of the first year is

"baked in" to the CES survey because this procedure does not account for evolving seasonality. "Baked in" means the seasonality in the start period—that is, whatever year is selected to start the over-the-year difference calculation—becomes the expected seasonality for both series. If CES or QCEW seasonality evolves from what was present in that start period, then there may be residual seasonality errors not accounted for by this method.

The basic equation describing this procedure is shown below. Note that this depiction can be extended to show the linkage to previous adjustments, which ultimately lead to the starting-point issue described above.

$$\hat{E}_t^B = \hat{E}_t + \left((\mathrm{QCEW}_t - \mathrm{QCEW}_{t-12}) - (\hat{E}_t - \hat{E}_{t-12}^B)\right), \quad (4)$$

where

$\hat{E}_t^B$ is the benchmarked all-employee level for month *t*, where *t* is June, September, or December,

$\hat{E}_t$ is the not-yet-benchmarked all-employee estimate for month *t*, and

QCEW_t are the data from the QCEW (with noncovered employment added) for month *t*.

This is an evaluation of the Over-the-Year-Change Difference procedure:

☑ Aligns the CES data to the QCEW in March

☒ Has a well-defined mathematical relationship with the third month of other quarters

☑ Aligns CES and QCEW employment more than once per year

☒ Maintains the CES seasonal pattern

☑ Addresses the small-domain problem

☑ Is easily explained

Therefore, the Over-the-Year-Change Difference procedure meets only four of our six criteria. Unfortunately, this procedure fails to appropriately account for the seasonal differences in the data.

The Weighted Model procedure develops a weighted function of QCEW and CES levels for the third month of each quarter. The weighting for March is 100 percent of the QCEW level, and the weighting for the other quarters could be determined on the basis of objective criteria using historical data.

The basic equation describing this procedure is as follows:

$$\hat{E}_t^B = (w_{1,t}\hat{E}_t) + (w_{2,t}\mathrm{QCEW}_t), \quad (5)$$

where

$\hat{E}_t^B$ is the benchmarked all-employee level for month *t*, where *t* is June, September, or December,

$\hat{E}_t$ is the not-yet-benchmarked all-employee estimate for month *t*,

QCEW_t are the data from the QCEW (with noncovered employment added) for month *t*,

$w_{1,t}$ is the weight applied to the CES data for month *t*, where *w* is between 0 and 1, and

$w_{2,t}$ is the weight applied to the QCEW data, where $w_{2,t} = 1 - w_{1,t}$.

Note that there are many functions that could be used to generate the weights.

This is an evaluation of the Weighted Model procedure:

- ☑ Aligns the CES data to the QCEW in March
- ☑ Has a well-defined mathematical relationship with the third month of other quarters
- ☑ Aligns CES and QCEW employment more than once per year
- ☑ Maintains the CES seasonal pattern
- ☑ Addresses the small-domain problem
- ☒ Is easily explained

Therefore, the Weighted Model procedure meets five of our six criteria. As such, this procedure likely would address the seasonal differences in the data sources and allow for maintaining the CES seasonal pattern. However, we did not pursue this method with extensive research because of another consideration: we were unsure how to explain the results from an economic perspective. That is, what would the result mean? It would not be the CES linked to the QCEW in an easily explained economic perspective. Also, the explanation that the function would be selected to reduce the error in the CES level and to limit change in the CES seasonal pattern was a less clear explanation than we would hope to have for these highly scrutinized data.

The final method included in recent research is the Seasonally Adjusted Difference in Employment Method. This method seasonally adjusts the QCEW and CES levels. It then takes the difference for each of these at the third month of the quarter after aligning the not seasonally adjusted CES level to the March not seasonally adjusted QCEW level.[24]

The basic equation describing the Seasonally Adjusted Difference in Employment Method is as follows:

$$\hat{E}_t^B = \hat{E}_t + \left(\mathrm{QCEW}_t^{SA} - \hat{E}_t^{SA}\right), \quad (6)$$

where

$\hat{E}_t^B$ is the benchmarked all-employee not seasonally adjusted level for month *t*, where *t* is June, September, or December,

$\hat{E}_t$ is the not-yet-benchmarked not seasonally adjusted all-employee estimate for month *t*,

QCEW_t^{SA} is the seasonally adjusted QCEW data (with noncovered employment included) for month *t*, and

$\hat{E}_t^{SA}$ is the CES seasonally adjusted all-employee estimate for month *t*.

This is an evaluation of the Seasonally Adjusted Difference in Employment Method:

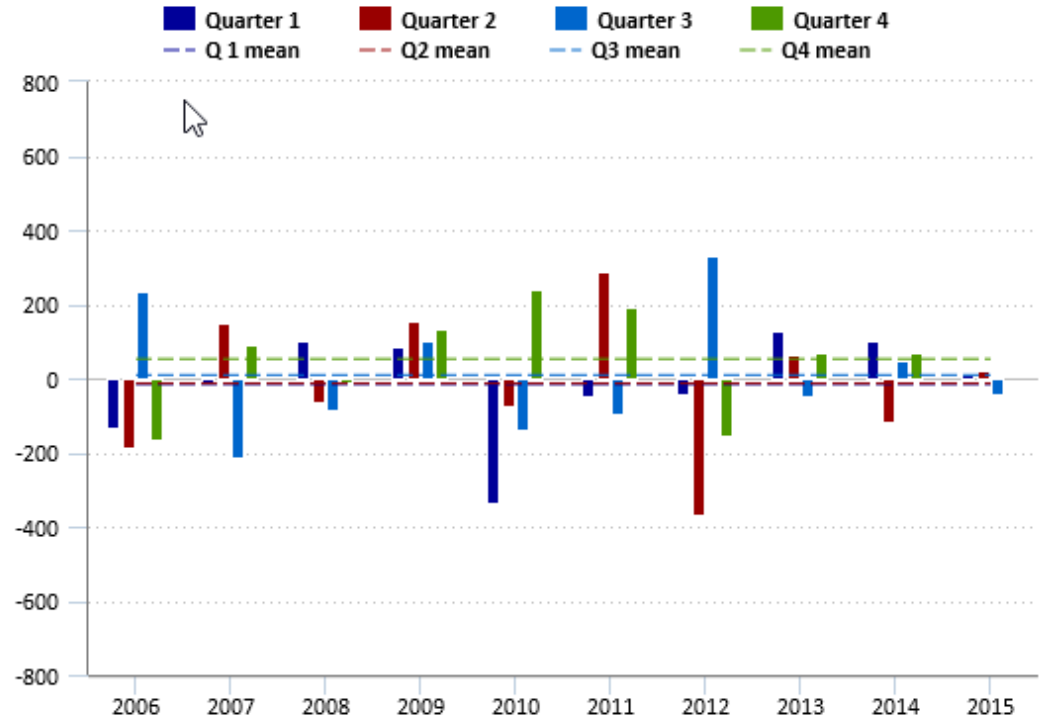
- ☑ Aligns the CES data to the QCEW level in March
- ☑ Has a well-defined mathematical relationship with the third month of other quarters

- ☑ Aligns CES and QCEW employment more than once per year
- ☑ Maintains the CES seasonal pattern
- ☑ Addresses the small-domain problem
- ☑ Is easily explained

The Seasonally Adjusted Difference in Employment Method directly addresses and appropriately accounts for the seasonal differences in the source data. This procedure may not directly address the issue of a small domain with small sample, but it does allow that problem to be solved through several potential supplemental procedures. This method also is relatively easy to explain. Therefore, this procedure meets at least five of our six criteria and potentially meets all six. Preliminary results strongly suggest that this method is viable and that it leads to improvements in the quality of CES data. More detail on this method is available in an article by Matthew Dey and Mark Loewenstein, and in a technical paper by the same authors.[25] Furthermore, this method appears able to be developed in a manner that allows a consistent application for both national and state data.

Comparing the difference between figure 1 and figure 2 is informative. As described earlier, figure 1 shows the differences, after benchmarking, between QCEW data and CES data. The differences are large and seasonal. Figure 2 shows differences in benchmarked CES and QCEW data that have been seasonally adjusted. The differences after seasonal adjustment are small and irregular. This simple change illustrates why we expect this method to be successful: it will correct for differences between the two data series after accounting for the different seasonal patterns. That is, this method extracts from the two series the differences that are primarily due to sampling and nonsampling error rather than deviations due to large seasonal differences. Hence, this method provides an appropriate correction to the CES data series.

Figure 2. The difference in the over-the-quarter change between Quarterly Census of Employment and Wages data and Current Employment Statistics data, seasonally adjusted



Note: The Quarterly Census of Employment and Wages data used here are the sum of the states' benchmarked Current Employment Statistics data.
Source: U.S. Bureau of Labor Statistics.

Future research

While the CES program staff (and other sophisticated CES data users) have long understood the substantial seasonal differences between QCEW and CES data, extensive research led us to conclude that there is no good way to address this issue on a not seasonally adjusted basis. Our research shows that a new benchmark procedure must directly address the differences in seasonality between the CES and QCEW levels by aligning data on a seasonally adjusted basis.

Future work on the Seasonally Adjusted Difference in Employment Method falls into two major categories. The first is additional work to verify that the applied results of this method meet the theoretical expectations. The rigorous evaluation of new methods in extensive simulations is the usual practice for BLS. Any changes to this important economic indicator would be made only after thorough research and evaluation has been completed.

The second category of future work is to catalog series in small domains with smaller sample sizes and then evaluate potential methods to produce a benchmarked series for them. The problem of small sample size is most obvious in smaller geographic domains but can also occur in some of the national series with the smallest employment. Two potential methods to explore are (1) a ratio adjustment of not seasonally adjusted data for very

small domains to a higher level seasonally adjusted domain for which the seasonally adjusted data meet selected quality characteristics and (2) a method based on a function of the reliability of the seasonal adjustment.

Regarding the first method, it should be relatively easy to design a procedure that replaces CES data with QCEW data for the smallest domains. The procedure would explicitly ratio adjust the QCEW data series and other data series so that they are additive to higher level aggregates because the seasonal adjustment process is more reliable for aggregates. The ratio adjustment would prorate the aggregate CES seasonality down to these smaller domains, with larger shares of that seasonality being allocated to data series with larger employment levels. This process would result in a dataset with the required additivity feature. The process would also highly weight the original QCEW data so that the sampling error component of the small-domain estimate is minimized. The second method is more complex. This method would require the development of a function that weights the seasonally adjusted CES and QCEW levels on the basis of the reliability of the two seasonally adjusted series. If the quality of both adjusted series is felt to be high, then the difference would be fully accepted. If the quality of the seasonal adjustment is low, then the difference might be discounted in a function that places more weight on the not seasonally adjusted QCEW data. An obvious problem with this method is that it will not have additivity to higher level aggregates as a built-in property. In this case, additivity must be provided through a second step, perhaps a ratio adjustment procedure similar to the one described earlier. Other methods for solving this problem likely exist and will be included in future research.

Concluding remarks

The CES survey has a long history, originating in 1915. Benchmarking for CES has nearly as long a history, first occurring in 1935. Over this history, and especially more recently, various user groups have expressed much interest in improving the quality of CES data by improving the timeliness of the benchmark process. As a result, the CES program has returned repeatedly to this research theme.

Earlier research efforts focused on the possibilities of finding an alternative procedure that aligned not seasonally adjusted CES data to not seasonally adjusted QCEW data. These efforts were ultimately unsuccessful because they did not account for the different seasonal patterns in CES and QCEW series, and a procedure could not be found that was convincingly an improvement. Recent research has reconfirmed that there is no obvious procedure to align not seasonally adjusted CES data to not seasonally adjusted QCEW data on a quarterly basis that adequately addresses the known seasonal differences between these independent data series.

The most recent research effort has focused on methods that do account for the different seasonal patterns. This research has found a method that appears to solve this longstanding roadblock by benchmarking to the difference in CES and QCEW levels after seasonally adjusting the two series. This procedure correctly accounts for the different seasonal patterns by removing the seasonality from the data and then benchmarking the not seasonally adjusted data to the difference that remains between the seasonally adjusted series.

This most recent round of research has focused on evaluating both national and state data at the North American Industry Classification System supersector level. Future research will focus on further evaluation of this method to ensure that it meets expectations at more detailed industry and geographic levels. Additionally, future research will evaluate alternative methods to solve the small-cell issues.

CES data are frequently used as an economic indicator. Moreover, these data are used as an input into other economic indicators. Therefore, research must be especially thorough and well documented, and it is typically shared with the economic community for comment before BLS makes significant changes in CES methodology. BLS plans to continue this research, and we will continue to share the results with the public via our advisory groups and through publication in various venues. [26]

ACKNOWLEDGEMENT: I would like to acknowledge that the success of this research is due to the work of a number of BLS employees who have participated in bringing new research ideas to the table. The team includes members from the CES program office, our statistical methods group, our economic research group, and our survey methods research group. The founding members of the team include Kenneth Robertson, Chris Manning, Kirk Mueller, Steven Mance, Greg Erkens, Larry Huff, Mark Loewenstein, Matthew Dey, Polly Phipps, Daniell Toth, and David Talan. Additional members have joined the team since its inception and contributed to these results.

SUGGESTED CITATION

Kenneth W. Robertson, "Benchmarking the Current Employment Statistics survey: perspectives on current research," *Monthly Labor Review,* U.S. Bureau of Labor Statistics, November 2017, https://doi.org/10.21916/mlr.2017.27.

NOTES

1 The BLS *Handbook of methods* (https://www.bls.gov/opub/hom/) is a compendium of detailed information about how BLS collects and prepares its economic data. The second chapter of the handbook (https://www.bls.gov/opub/hom/pdf/homch2.pdf) is devoted to providing detailed information on the methods, procedures, and estimators used by the Current Employment Statistics survey. The handbook discusses in detail the estimator just described and many other aspects of the CES survey.

2 Sampling error occurs because a measurement is made from a sample rather than a population. If a probability-based sample design is used, then the sampling error can be estimated and a confidence interval can be calculated. The confidence interval describes the bounds that the true value of the measured characteristic is expected to fall in, with a corresponding level of confidence. Nonsampling errors include a number of different types of error, such as recall error, classification error, coding error, transcription error, processing error, etc.

3 John P. Mullins, “One hundred years of Current Employment Statistics—an overview of survey advancements,” *Monthly Labor Review*, August 2016, https://doi.org/10.21916/mlr.2016.2.

4 The term QCEW is used for two purposes in this article. When we talk about the program, QCEW refers specifically to the Quarterly Census of Employment and Wages program. When used in the context of CES benchmarking, the term is used to represent the CES benchmark population values, or QCEW plus noncovered employment.

5 In 2003, the CES program redesigned the survey to use modern probability sampling procedures, after which the usual size of revisions was smaller. Prior to the redesign, the state and metropolitan area benchmark revisions tended to be relatively larger than national revisions.

6 The Census Bureau’s *County Business Patterns* program produces data by county with a lag of about 2 years. Information about the program and its data can be found at https://www.census.gov/programs-surveys/cbp.html.

7 Information about the Railroad Retirement Board can be found at https://www.rrb.gov/.

8 The postbenchmark period is the months following the benchmark, April through October for national estimates and October through December for state estimates.

9 Christopher D. Manning and John R. Stewart, “Benchmarking the Current Employment Statistics national estimates,” *Monthly Labor Review*, October 2017, https://doi.org/10.21916/mlr.2017.25. National annual benchmark information can be found at https://www.bls.gov/ces/tables.htm#benchmark.

10 Kirk J. Mueller, “Benchmarking the Current Employment Statistics state and area estimates,” *Monthly Labor Review*, November 2017, https://doi.org/10.21916/mlr.2017.26. State and area benchmark information for 2016 can be found at https://www.bls.gov/sae/benchmark2016.pdf.

11 Early benchmarking on a quarterly basis is discussed in several articles by members of the Federal Reserve banks, including Keith Phillips (Dallas Federal Reserve Bank at San Antonio) and Thomas Walstrum (Chicago Federal Reserve Bank). See http://www.dallasfed.org/research/basics/benchmark.cfm and https://www.chicagofed.org/.

12 The author would like to thank Dr. Keith Phillips for reviewing and commenting on this section to ensure that the description captures the methodology appropriately.

13 ARIMA X-13 SEATS is publically available from the U.S. Census Bureau at https://www.census.gov/srd/www/x13as/.

14 Franklin D. Berger and Keith R. Phillips, “Reassessing Texas employment growth,” *The Southwest Economy*, July/August 1993, http://www.dallasfed.org/assets/documents/research/swe/1993/swe9304a.pdf.

15 Ibid.

16 Response-analysis surveys comparing CES and QCEW responses were conducted in 1994, 2001, and 2008.

17 Jeffrey A. Groen, “Seasonal differences in employment between survey and administrative data,” Working Paper 443 (U.S. Bureau of Labor Statistics, February 2011), https://www.bls.gov/osmr/research-papers/2011/pdf/ec110010.pdf.

18 By correct, I mean collecting the number of employees that meets the CES definition, not including employees who do not meet that definition, and reporting the data for the specified reference period rather than a different payroll period.

19 The reference period for the CES survey, QCEW, and most federal business surveys is the week or the pay period that includes the 12th day of the month.

20 Note that this assumption is for illustrative purposes only, as it ignores all other errors in the QCEW data and assumes that all error is related to the reporting period.

21 For a description of the two-step seasonal adjustment process, see Patricia Coil, Taylor Le, and TJ Lepoutre, "Revisions in state establishment-based employment estimates effective January 2016" (U.S. Bureau of Labor Statistics),p. 4, https://www.bls.gov/sae/benchmark2016.pdf.

22 The splice point is the point where benchmarked CES data, which is essentially QCEW data for states, meets data that have not been benchmarked, which are purely CES-based data.

23 An earlier, smaller study of several of these methods was conducted by Richard Valliant and Jill Dever in 2008. That research focused on the over-the-year change difference, a seasonally adjusted difference, and a Lowess estimator.

24 To be specific, what is seasonally adjusted on the QCEW side are QCEW data adjusted to meet the CES scope, which includes data for employment not covered by QCEW. These adjusted QCEW data are aggregated to match the industries and geographies published by CES, and those QCEW aggregates are seasonally adjusted in this procedure.

25 Mark A. Loewenstein and Matthew Dey, “A quarterly benchmarking procedure for the Current Employment Statistics program,” *Monthly Labor Review*, November 2017, https://doi.org/10.21916/mlr.2017.28.

26 BLS has two advisory groups involved with this issue, a Technical Advisory Committee (TAC, see https://www.bls.gov/advisory/taccharter.htm) and the Data User Advisory Committee (DUAC, https://www.bls.gov/advisory/duac.htm).

RELATED CONTENT

Related Articles

Benchmarking the Current Employment Statistics survey: the past, present, and future, *Monthly Labor Review*, October 2017.

Benchmarking the Current Employment Statistics national estimates, *Monthly Labor Review*, October 2017.

Benchmarking the Current Employment Statistics state and area estimates, *Monthly Labor Review*, November 2017.

A quarterly benchmarking procedure for the Current Employment Statistics program, *Monthly Labor Review*, November 2017.

Related Subjects

Survey methods | Benchmarking | Employment | Survey procedures | Industry studies