A Better Benchmark Process for the U.S. Current Employment Statistics Survey December 2021

Kenneth W Robertson
U.S. Bureau of Labor Statistics

Abstract

The U.S. Bureau of Labor Statistics' Current Employment Statistics (CES) survey is a highly watched indicator of the U.S. economy. The survey sample of 689,000 worksites provides data published in about 50,000 time series each month. On an annual basis, two substantially different procedures are used to align these time series with population values. The procedure used for national series aligns the prior March value with population levels and wedges the adjustment back one year. The procedure used for state and area series replaces the CES estimates with population values. Neither of these procedures is optimal. The national procedure ignores information contained in quarterly population reports; therefore the wedging of the annual difference back one year may not always provide the most accurate historical data. The state and area procedure ignores the substantially different seasonality of the monthly population data, creating a historical series that is seasonally different from current estimates. Research has identified a candidate procedure which makes better use of the population data to solve both of these problems. This paper describes the background leading to this research, the new procedure, and work still to be done to fully develop this improved process.

Keywords: Current employment statistics, benchmarking, alignment

1. Background Leading to Research

The U.S. Bureau of Labor Statistics' Current Employment Statistics (CES) survey[i] is a very large monthly survey that collects and publishes data on employment, hours, and earnings by industry and geography. These data are among the first indicators of the health of the U.S. economy, and the national data are designated as a Principal Federal Economic Indicator as part of *The Employment Situation* news release.

The CES survey is a quick response, repeated survey, where respondents are asked to report each month for the pay period that includes the 12th day of the month, and estimates are initially published a few weeks following the reference period, usually on the first Friday of each month. Results are revised to reflect additional data collected for that reference period over the next two months. Smaller businesses are typically in the sample for three years, while larger businesses may be in the sample indefinitely.

Both the national[ii] and sub-national[iii] components of the CES program publish about 25,000 data series each. The national data series are broadly distributed across detailed industries and data types, while the sub-national series are broadly distributed across broader industries, states, and metropolitan area geographies. The sub-national data are published a few weeks following the national data.

Unlike many surveys, the CES survey has an administrative dataset produced at regular intervals that allows BLS to align the survey-based employment values to a near population of in-scope employment. The administrative data comes primarily from the Quarterly Census of Employment and Wages (QCEW)[iv]. The QCEW data are published about six months following the end of each calendar quarter. The alignment to a population value is referred to as benchmarking, and the population values are referred to as benchmark values, or benchmarks. Other data sources are used to develop benchmarks for the 3 percent of data in-scope for CES that are not in-scope for QCEW.

1.1 A brief history

The first CES benchmark took place in 1935[v], with revised data being published for 1923-1929. That first revision was quite large, at approximately 12 percent. Benchmarking was done periodically during these early years, until, in 1982, BLS began benchmarking the CES every year. Revisions over the past ten years have averaged less than 0.2 percent in absolute value.

In 1980, the source data for the QCEW – data from the Unemployment Insurance program – began to be available for all 12 months of the year. At that time, states began to move from an alignment to the QCEW's March employment value, with a 12-month linear wedge to distribute the error back in time, to a procedure that replaced all 12 months of CES estimates with population values. States began to change to this replacement procedure because BLS and state analysts believed that the error associated with the administrative data was smaller than the error associated with the survey estimates. This assumption seemed especially reasonable for smaller domains, while for larger domains the survey data with the linear wedge were deemed to be of high quality. Articles describing the current benchmark procedures for national[vi] and subnational[vii] data are available on the BLS website.

1.2 Identification of a problem

In 1993, Berger and Phillips[viii], two researchers at the Dallas Federal Reserve Bank, identified a substantial problem with the state CES data; the seasonal patterns in the population-replaced data did not match the seasonal patterns in the CES estimates. This resulted in very large over-the-month movements in some months, particularly in January. Further evidence of the difference in seasonality was documented by Groen[ix]. Berger and Philips also proposed a fix for this problem; seasonally adjust the population data, then seasonally adjust the survey data, and then splice the two seasonally adjusted series together. The spliced series would include population data from the point of its most recent availability back in time to the series start point, and survey data would be present from that end point forward to the most current estimates. In application, this means that after the benchmark data are released in March, population data form the data series from the inception point (1990 for most series) up to the most recent September, and survey data form the time series from October forward through January. Survey estimates then add to this time series until the next benchmark.

This solution, which we call the CES two-step seasonal adjustment, provided an important correction to this problem for seasonally adjusted data. However, it only solved a small part of the problem. BLS only seasonally adjusts a small fraction of the state and metropolitan area data series. Therefore, whenever a data user analyzes a series that does not have a seasonally adjusted counterpart, or if they analyze the not-seasonally-adjusted series instead of the seasonally adjusted series, they are comparing apples and oranges across a span of time. The result is an analysis that includes an uncorrected and unanticipated seasonal component. Most data users wrongly infer that this is the result of survey error, and they may conclude that the survey data do not meet their needs. Many data users have told BLS that they just wait the half year to get the population data instead of using the timely survey data, because of these large survey "errors".

The problem above is limited to data in state and metropolitan area series. The national data are faced with another, perhaps less severe problem. While the state and metropolitan area benchmark replaces 12 months of data with population data, the national benchmark replaces only one month – March. Then, a linear wedge is used to distribute the March correction back into the prior 11 months. The assumption with this procedure is that the error accumulated in a consistent linear manner each month. We know that this assumption is almost certainly wrong, with errors of different sizes accruing to different months in the wedge time range. We do not know how much deviation there is from the linear assumption – and that deviation almost certainly changes from year to year. We can speculate that at times the deviation may be substantial, especially at economic turning points. The national benchmark procedure ignores potential data series improvements that reside in the population data that might correct this error component.

Using administrative population data for every month simply transfers administrative errors and foreign seasonal patterns into the CES data, while using the population data for only one month ignores potentially valuable information about how error accumulated over the prior year. Therefore, this is a bit of a Goldilocks story; is there a middle ground that is "just right"?

1.3 Researching alternative procedures

Following this realization that the survey and population data series were providing measures that were somewhat different, BLS initiated several research projects over time to identify an alternative benchmark procedure with better properties. However, each research project ended in an impasse. Each of these projects included analysts who focused on national data, and analysts who focused on state and area data. Analysts who produced the national data very firmly believed that the current procedure, replacing March with population values and applying a linear wedge to distribute the error back 12 months, was the best procedure for these data. The primary goal from their point of view was to preserve the survey-estimated over-the-month changes to the greatest extent possible, even though that meant ignoring potential information about how error accumulated over the past year in the months between successive March periods. The premise among this group was that CES provided the highest quality information on over-the-month change, and only needed to have the employment level corrected periodically. The state analysts firmly believed that the QCEW data was vastly superior in quality to survey estimates. The primary goal from the state analyst point of view was to remove survey error from the estimates to the greatest extent possible, even though that meant eliminating all CES information from the benchmarked series and accepting the non-sampling errors and foreign seasonality associated with the population data. The premise among this group was that QCEW provided the highest quality information on over-the-month change, and CES survey data was only useful as a forecast of the QCEW data.

2. Towards a solution

When I realized that a big part of the impasse was due to differing assumptions and goals, I convened a team that extended beyond the national and state analysts. This team also included senior economic and statistical research staff at the BLS. One objective for this team was to define a goal for benchmarking that was relevant to both national and state data. Another change for this team was to include research on procedures that were based on seasonally adjusted data; earlier research had eliminated from consideration other potential procedures but had not explored this one.

Making these changes to the composition of the research team, to the objective, and to the set of procedures to research led to a breakthrough[x]. A benchmark goal common to both national and state data was identified – to maximize the use of information contained in CES over-the-month changes, while simultaneously maximizing the best use of information contained in the end-of-quarter months in QCEW data.

To increase acceptance of the proposed procedure among our state partners in the program, I began to frame the problem as one that didn't portray CES as high quality and QCEW as low quality. To our state partners, I began to describe the QCEW as a high-quality data source, but one with some reporting errors related to the accuracy of reference periods that led to the data not strictly following CES reporting guidelines. From the perspective of measuring employment levels at a point in time, some slippage in the reference period does not substantially degrade the quality of the data; the data are accurate for the period reported even if it is not exactly the reference period requested. However, for the purpose of measuring over-the-month change the reference period is extremely important. In a seasonal hiring month, moving the average reporting

period[1] even a few days can have highly significant effects on the measured over-the-month change. This difference in reference period reporting is evident and easy to see in the data – leading to differences in seasonality that are stark and highly significant.

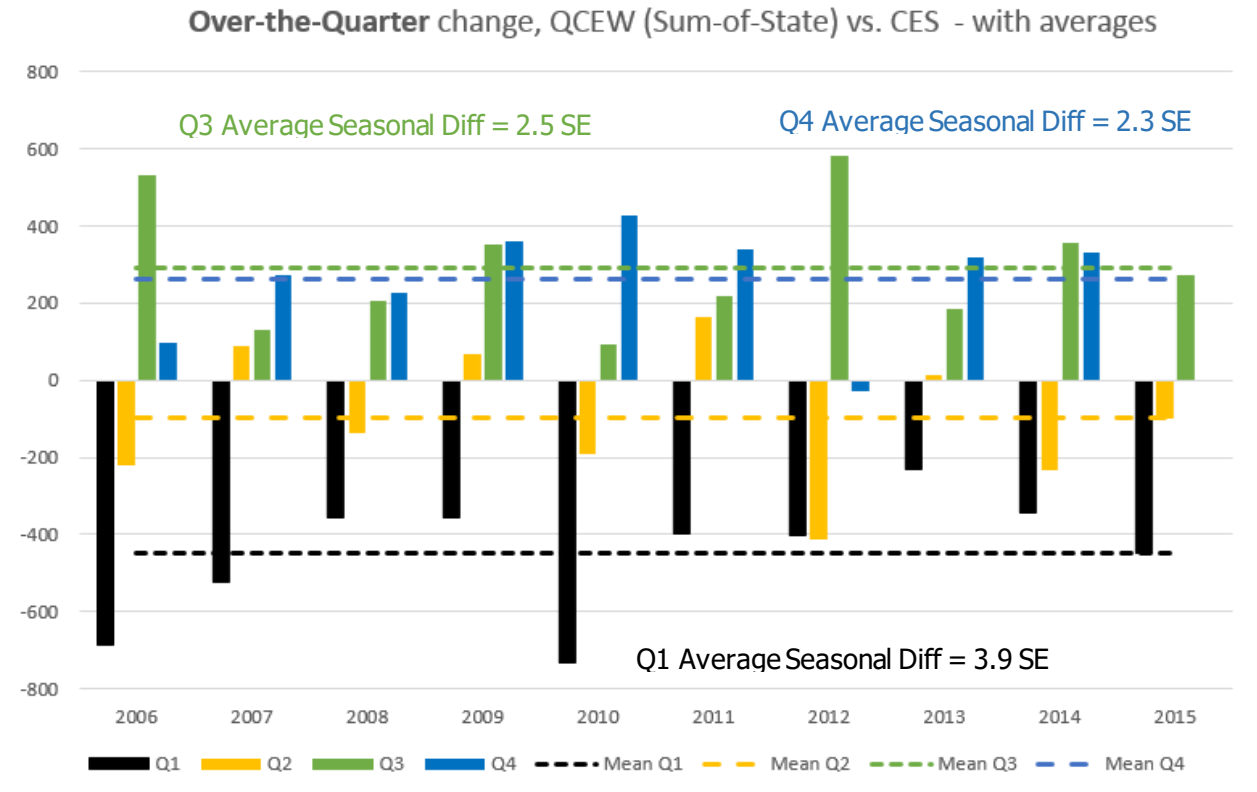Chart 1 below provides an illustration of the size of this problem.

Chart 1.

Table 1.

Quarter	Mean value of Over-the-quarter change, Benchmarked CES – QCEW (population)
1	-447,500 (3.9 SE)
2	-96,200 (0.8 SE)
3	+292,000 (2.5 SE)
4	+260,700 (2.3 SE)

Table 1 presents the means from that chart. These data are the mean values of the difference in over-the-quarter (OTQ) change from benchmarked national CES data and benchmarked state and area CES data. The latter serve as a proxy for the QCEW population data, and they are labeled as such in the chart and table. In an ideal world, these series would be identical, portraying the same March value, the same OTQ changes, and therefore displaying no differences. However, what we

---

[1] An average reporting period here is intended to mean the average, across all reports, of the period reported in the data, rather than the period requested. The period requested for all respondents is the pay period that includes the 12th of each month. Many QCEW respondents provide data for this pay period, but others provide data for the end of the month, or for the period most proximate to the date they prepare the report.

see are differences that, when presented as standard errors of survey-based OTQ change, are very large and highly significant.

We would normally expect that, after benchmarking, we have removed all error from the employment levels and what remains is some trivial level of residual random noise. We can see that this is not the case. There are clear and very large quarterly artifacts in these data after benchmarking. This result supports prior evidence that there are different seasonal patterns in these two datasets. Therefore, a direct replacement of one with the other is not advised. A more nuanced approach will give a better result.

2.1 A New Procedure

The team identified and documented a procedure that seasonally adjusts both CES and QCEW, and then adjusts the not-seasonally-adjusted CES data based on the difference in the two seasonally adjusted series. This procedure proved to have excellent properties from both a theoretical basis and empirically, at an aggregate series level.

To be more specific, the procedure would seasonally adjust a data series for CES and QCEW for the third month of each quarter. While QCEW does capture employment for each month, the data are obtained in the weeks following the end of the quarter. Evidence suggests that the most recent month reported in the administrative data has the highest quality. The March CES not-seasonally-adjusted employment level would be replaced each year with not-seasonally-adjusted QCEW data. This will ensure that seasonal adjustment anomalies don't impact the long run accuracy of the data series. The other three quarters would be benchmarked to the difference in the seasonally adjusted series.

The first two months of each calendar quarter would be adjusted to the new benchmarked level by applying a linear wedge.

The new procedure can also be described as follows:

[6] $\widehat{AE}_t^B = \widehat{AE}_t + \left(QCEW_t^{SA} - \widehat{AE}_t^{SA} \right)$

Where:

$\widehat{AE}_t^B$ is the benchmarked not-seasonally-adjusted employment level for month t, where t is (June, September, or December)

$\widehat{AE}_t$ is the not yet benchmarked not-seasonally-adjusted employment estimate for month t

$QCEW_t^{SA}$ is the seasonally adjusted QCEW data (with non-covered employment included) for month t, and

$\widehat{AE}_t^{SA}$ is the not yet benchmarked seasonally adjusted CES employment estimate for month t.

This procedure captures the best qualities of both CES and QCEW. From CES we capture and maintain the seasonality in the over-the-month employment changes that are collected from respondents who have been carefully instructed about the reference period to report for. From QCEW we capture and bring into CES auxiliary information about the population level of employment each quarter. By combining these two data sources and maximizing their strengths, we create a CES dataset that is more accurate from a historical perspective, leading to an enhanced understanding of the context in which we are analyzing labor market activity.

Chart 2 below portrays data similar to Chart 1, but this time showing the difference between the datasets after implementation of the proposed procedure.
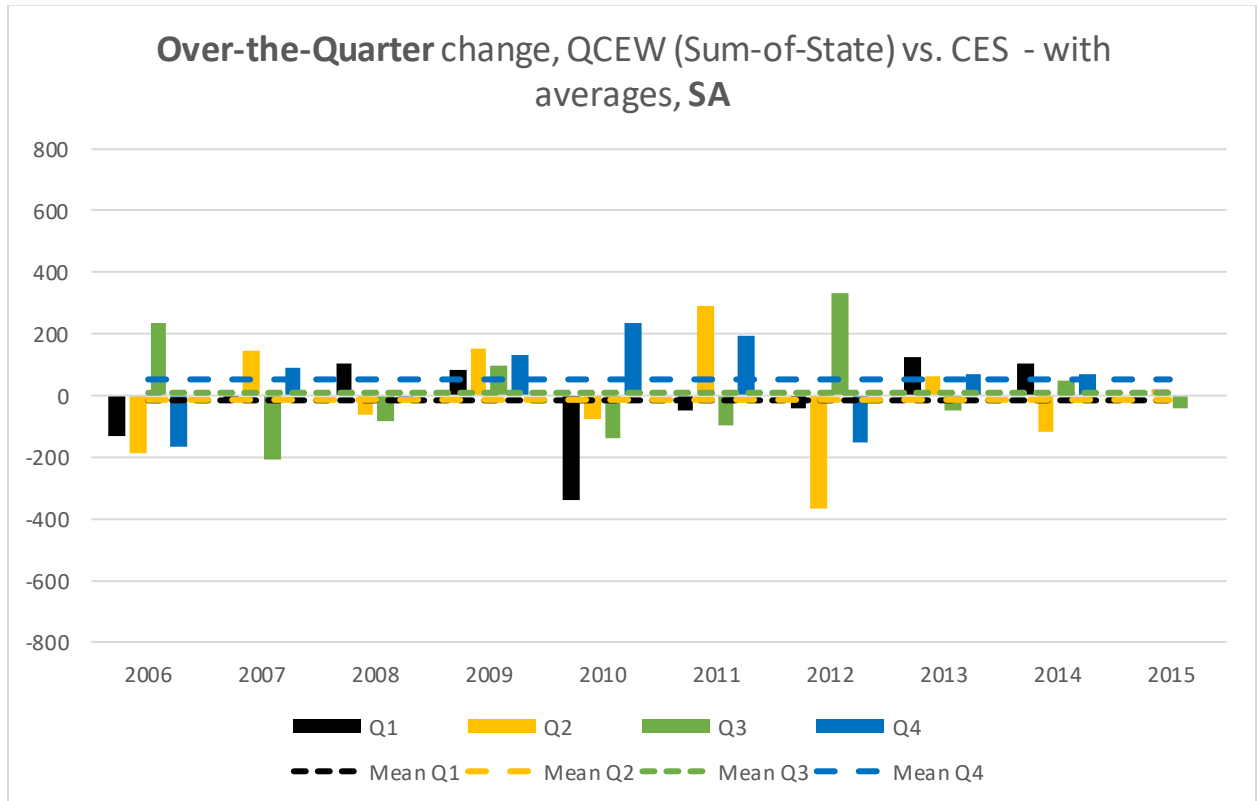
Chart 2.

Over-the-Quarter change, QCEW (Sum-of-State) vs. CES - with averages, **SA**

Legend: Q1, Q2, Q3, Q4, Mean Q1, Mean Q2, Mean Q3, Mean Q4

Table 2.

Quarter	Mean value of Over-the-quarter change, Benchmarked CES – QCEW (population) Using simulated data following proposed procedure
1	-15,500 (0.1 SE)
2	-13,900 (0.1 SE)
3	+9,200 (0.1 SE)
4	+52,400 (0.5 SE)

Table 2 presents the means values from that chart, as well as those values converted into standard errors of the over-the-quarter (OTQ) change in survey data.

So, what is this telling us? After benchmarking using the proposed solution, we have removed error from the employment levels and now, as expected, what remains is some trivial level of residual random noise. When looking at the OTQ employment change in the benchmarked data we can see that there are no longer any large quarterly artifacts remaining in these data. Using a procedure that incorporates seasonal adjustment as part of the solution has, as expected, removed the seasonal artifacts from the result. This solution (1) accounts for the seasonal differences in the two data series, (2) incorporates population data into the survey benchmark at four points each year, and (3) retains the high-quality survey data on over-the-month change. Additional context on this problem and solution can be found by Robertson, on the BLS website[xi].

3. Work Still to be Done

It is readily apparent that the proposed solution is an improvement over either of the procedures currently used to benchmark CES data. However, the solution as presented here is prepared at an aggregate level – while CES benchmarks at the most detailed levels, and then sums those levels to

arrive at an aggregate result. The work being done now is focused on understanding where problems might be at those detailed levels. We can certainly imagine that a seasonal signal might be overshadowed by noise in small area cells for the survey, but apparent in the population data. What do we do in this case when we know the seasonal patterns are different? We can also look ahead at classification system changes that have us restructure historical microdata responses into new series. We can clearly do that for the population data. However, it may be difficult to create historical survey data according to a different classification. And that history under the new classification is required for this proposed procedure to work. A third issue, perhaps less daunting, is that the national program does not have a history of quarterly population data grouped into the series classifications we publish. This can be done, but it will require significant review of those thousands of data series to ensure that any anomalies have been accounted for.

These are three of the biggest challenges this project faces. We believe we have an important improvement to the benchmarking procedure that will improve the quality of national and subnational data. This will also align the processes for both national and subnational data, making data from these two parts of the program more comparable. Difficult work remains to get this to implementation.

4.0 Final Thoughts

This notion of improving the CES benchmark process is not new. Projects to explore this have come and gone several times over the years. Each of the earlier projects failed because they did not create a clear solution that offered improvements for both national and subnational data. This solution does provide improvements to both sets of data, at least at an aggregate level. Time will tell if solutions can be found for the remaining problems at the more detailed levels of application. I hope so, because having better data, and having a more unified procedure for both sets of data, will benefit economic analyses of the labor market by both BLS economists and by the many other users of these vital economic statistics.

Acknowledgements

I would like to note that this paper builds on the work of others who have done significant work in this area. Principals among these others are Kirk Mueller, Christopher Manning, John Stewart, and Steven Mance, all employees at BLS.

Disclaimer
Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

References

[i] BLS Handbook of Methods, this resource provides substantial detail on the content and operations of the survey. See BLS Handbook of Methods: Chapter 2. Employment, Hours, and Earnings from the Establishment Survey

[ii] For information about the national CES program, see Current Employment Statistics - CES (National) : U.S. Bureau of Labor Statistics (bls.gov)

[iii] For information about the subnational CES program, see SAE Home Page : U.S. Bureau of Labor Statistics (bls.gov)

[iv] For information about the Quarterly Census of Employment and Wages, see Quarterly Census of Employment and Wages : U.S. Bureau of Labor Statistics (bls.gov)

[v] For a history of CES, and more about early benchmarking of the survey, see one-hundred-years-of-current-employment-statistics-an-overview-of-survey-advancements.pdf (bls.gov)

[vi] Manning, C., and Stewart, J., describe the national benchmark, see Benchmarking the Current Employment Statistics national estimates : Monthly Labor Review: U.S. Bureau of Labor Statistics (bls.gov)

[vii] Mueller, K., describes the subnational benchmark procedure, see Benchmarking the Current Employment Statistics state and area estimates : Monthly Labor Review: U.S. Bureau of Labor Statistics (bls.gov)

[viii] Berger, F., and Phillips, K., describe early benchmarking on a quarterly basis, and a solution to a seasonal difference between CES and QCEW data, see
http://www.dallasfed.org/assets/documents/research/swe/1993/swe9304a.pdf

[ix] Groen, J., Seasonal Differences in Employment between Survey and Administrative Data, BLS Working Papers, Seasonal Differences in Employment between Survey and Administrative Data (bls.gov)

[x] Loewenstein, M., and Dey, M., describe a new benchmark procedure for the CES data, see A quarterly benchmarking procedure for the Current Employment Statistics program : Monthly Labor Review: U.S. Bureau of Labor Statistics (bls.gov)

[xi] Robertson, K., provides additional context on both the history of CES benchmark research and on this solution, Benchmarking the Current Employment Statistics survey: perspectives on current research : Monthly Labor Review: U.S. Bureau of Labor Statistics (bls.gov)