

Estimates by Firm Size Using the CES Survey¹ October 2012

Nathan Clausen and Chris Manning

U.S. Bureau of Labor Statistics, 2 Massachusetts Ave NE, Washington, DC 20212

Abstract

The Current Employment Statistics (CES) program of the Bureau of Labor Statistics (BLS), surveys approximately 490,000 establishments in order to provide detailed industry data on employment, hours, and earnings of workers on nonfarm payrolls. The authors have been researching the use of CES survey data to produce industry estimates by firm size.

This paper first details research to develop methodologies for producing CES estimates by firm size. This includes defining size classes, determining firm size, additivity across size classes to the industry estimates, forecasting net birth/death residuals, seasonal adjustment, and the creation of a historical series of CES size class estimates. Next, the paper covers several issues addressed by the developed methodology. CES considered the adequacy of the sample for size class estimates along with movements of firms between size classes to understand how this might impact estimate reliability. Finally, the paper analyzes several quality measures of the time series produced in the research. The research suggests that the introduction of size class estimates would provide a valuable and reliable addition to what is currently published by the CES program.

Key Words: Current Employment Statistics (CES), CES methodology, Size-class estimates

1. Background

The Current Employment Statistics (CES)² Survey, conducted by the U.S. Bureau of Labor Statistics (BLS) in cooperation with State Employment Security Agencies (SESAs) collects payroll data each month on employment, hours, and earnings from a sample of nonagricultural establishments. The current CES sample includes about 140,000 businesses and government agencies, representing approximately 490,000 individual worksites. From these data, a large number of employment, hours, and earnings series are prepared and published each month with industry and geographic detail.

Each month, CES publishes preliminary estimates just three weeks after the close of the survey reference period (the pay period including the 12th of the month). This compressed timeframe allows CES estimates to be among the first economic indicators that are released. But it also means that a portion of the sample does not always have time to respond before the estimation process begins. Taking this into

¹ Any opinions expressed in this paper are those of the authors and do not constitute policy of the U.S. Bureau of Labor Statistics.

² For more information on the Current Employment Statistics (CES) Program, see <http://www.bls.gov/ces/home.htm>.

account, CES publishes estimates with a reduced level of detail, aggregating collected sample into larger estimating cells to reduce volatility. Table 1 shows the total number of published employment, hours, and earnings series, the number that are published as first preliminary estimates. Currently, CES publishes just 12% of series with the first preliminary release.

Table 1: Counts of Published Series

Datatype	Seasonal Adjustment	Total Published	First Preliminary	Size-class
Employment	Unadjusted	2,420	214	45
	Adjusted	225	214	45
Hours and Earnings	Unadjusted	3,222	168	90
	Adjusted	126	126	90
Total		5,993	722	270

The table also shows the number of series that would be added to the first preliminary release if size-class estimates were introduced. Adding size-class estimates would increase first preliminary published series by almost 40%. More importantly, adding size-class estimates to CES would provide an additional dimension of valuable employment data that can help understand the current limited first preliminary CES estimates.

2. Size-class Estimation Methodology

Size-class methodology has been designed to overlap with existing CES methodology (for example, the use of matched sample, formulae for calculating over-the-month changes, annual benchmarking, and incorporating birth/death residuals)³. When current methodology was expanded to account for peculiarities of size-class estimation, an attempt was made to maintain parallels with existing methodology. These guidelines, established at the start of the size-class estimation research project, prioritized the compatibility of industry and size-class estimates and the use of existing survey data for estimating size-class series. This section covers the expanded CES methodologies and definitions used to estimate size-class employment, hours, and earnings.

2.1 Sizing Basis

CES relies on the federal Employer Identification Number (EIN) as reported with Unemployment Insurance (UI) filings to define a firm and to determine its size. Use of the EIN is consistent with other BLS data that are based on firm size and allows comparability with other size based statistics. The definition for identifying firms by EIN differs from official CES sample stratification methodology, which defines businesses by UI numbers because CES is a state-based design. However, the EIN allows BLS to identify firms across state lines, while UI numbers are unique to each state where a firm does business. The EIN is also more economically meaningful.

³ An overview of CES estimation methodology is available at <http://www.bls.gov/bls/empsitquickguide.htm>.

2.2 Defining Size Groupings

The Office of Management and Budget defines twelve size classes for use with establishment surveys⁴. Following OMB recommendations, CES groups these twelve into three firm size classes as shown in Table 2. These three sizes were selected as the best balance between a need for meaningful analytical detail and a need to ensure adequate sample coverage across all industries. While there may be good reason for additional size groupings, many users of the experimentally released data series have actually combined the smaller classes to compare just large and small businesses.

Table 2. Size Class Definitions

Size Class	Employees
Small	1 - 49
Medium	50 – 499
Large	500 +

2.3 Determining Firm Size

At the time of sample selection, the firm size of each selected sample unit is measured and retained by the sample unit throughout estimation. The size of a firm is determined using the same 12-month maximum employment criteria as is currently used for sample stratification, and is attributed to all establishments in a firm. For example, a ski resort employing 75 people during the peak ski season, but only 12 during the summer, would be classified as a medium sized firm. It is important to note, however, that not all establishments in a firm may be selected as part of the CES sample, since sample is selected at the Unemployment Insurance (UI) level which is limited to the portions of a firm that lie within a single state.

The combination of the limited number of size classes and the maximum employment measure used to measure firm size results in very stable size groupings. This stability is addressed below in section 4.2.

2.4 Industry Detail and Sample Allocation

In order to provide industry detail for analysis while ensuring adequate sample coverage, CES researched producing estimates by size class for major industry sectors (generally 2-digit NAICS groupings). Table 3 below shows the industry detail for which size-class estimates have been analyzed.

Table 3. Industry Detail for Size-class Estimation

Industry Title	NAICS Industries
Total private	(1)
Mining and logging	1133, 21
Construction	23

⁴ Office of Management and Budget. 1982. "New Statistical Standard on Comparability of Statistics on Business Size." Federal Register 47 (96; May 18), 21362-21363.

Industry Title	NAICS Industries
Manufacturing	(1)
Durable goods	32x, 33
Nondurable goods	31, 32x
Trade, transportation, and utilities	(1)
Wholesale trade	42
Retail trade	44, 45
Transportation and warehousing	48, 49
Utilities	22
Information	51
Financial activities	52, 53
Professional and business services	54, 55, 56
Education and health services	61, 62
Leisure and hospitality	71, 72
Other services	81
1. Aggregate Sector	

Maintaining consistency with CES industry sample allocation methodology, each reporting establishment is used for estimation in the industry of the establishment. So, for example, a large firm that manufactures and sells widgets could have one portion of its employment represented in the manufacturing large size class, while the rest is in the large retail trade size class.

2.5 Benchmark Levels

On an annual basis, BLS realigns the CES sample-based employment estimates for March of each year with UI-based population counts or benchmarks for March. Employment by size must be further reconciled to account for employment in establishments not classified by industry and in establishments uncovered by UI. Ratios of industry employment by size to total industry employment in all three size classes are calculated from Quarterly Census of Employment and Wages (QCEW) data and multiplied against the official CES industry employment that includes non-covered and unclassified employment.

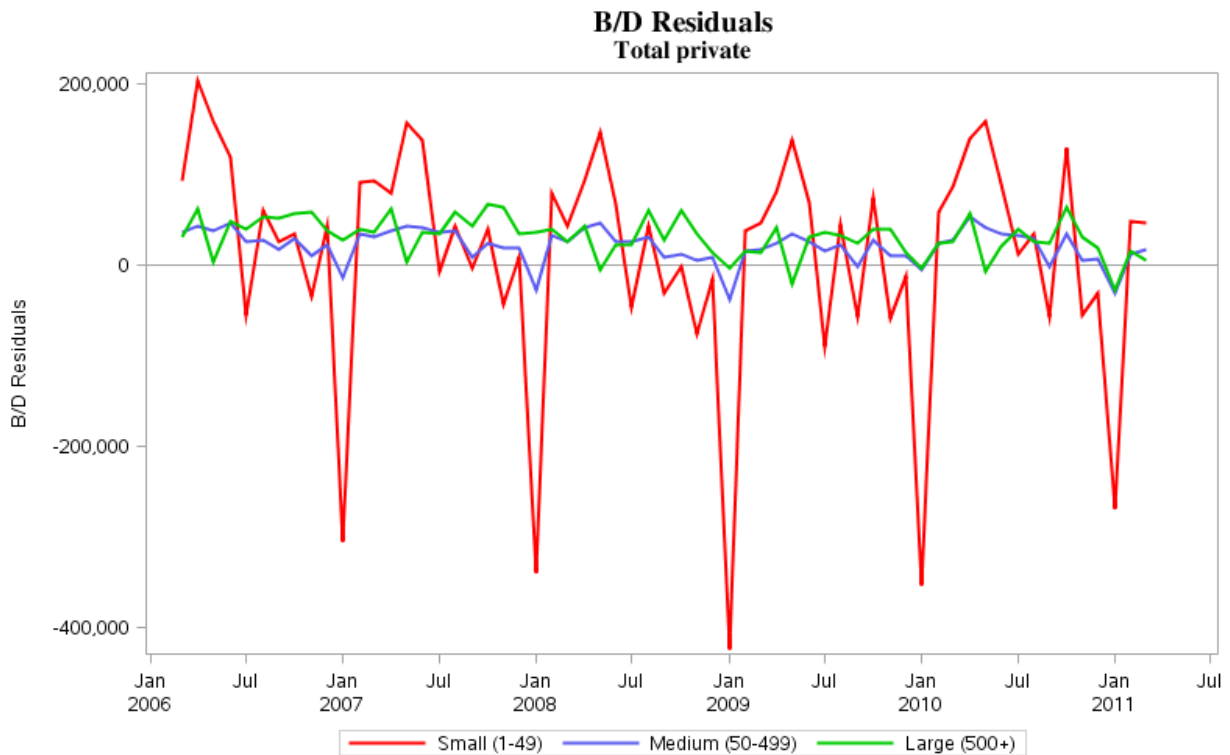
2.6 Birth/Death residual modeling

CES uses a birth/death model to adjust monthly employment estimates by adding a net birth/death residual to the estimate. This reduces a primary source of non-sampling error which is the inability of the sample to capture, on a timely basis, employment growth generated by new business formations.

A key tenet of the CES research on size-class estimation is that the size-class series are economically differentiated. A critical component of this difference is the disproportionate share of the industry birth/death residual present in each of the size classes. While it is intuitive to think that the size-class residuals would be different, it is interesting to see in what ways. As shown in Figure 1, the most striking is the contrast in the range of values of the small size class compared with both the large and medium. The former ranges from 200 to -400 thousand, while the latter are between 100 and -50 thousand. This is

a reflection both of the number of firms present in the small size class as well as the volatility of firm births/deaths observed. The small size class also shows a very strong seasonal pattern that is less visible in the other size classes. A second interesting observation is that the larger size classes generally have positive residuals, meaning that larger firms generally gain more jobs from business births than they lose from deaths (e.g, employment from openings at new locations tends to be larger than employment losses from existing locations that closed). Additional research is needed to further clarify the exact nature of difference, whether it is due to large firms “downsizing” before going out of business, large firms being “born” through spinoffs, or some other reason. However, these differences highlight the different behaviours demonstrated across firm size.

Figure 1: Total private birth/death residual by firm size



2.7 Calibration

CES industry estimates are made at a detailed level and aggregated to higher sector levels. Size-class estimates are made at an industry sector level. Because different groupings of sample have differing sample variances, the resulting summary levels of estimates by size class may not be exactly the same as the total industry estimate. To ensure that the sum of the three size-class pieces is equal to the industry total, size-class estimates are calibrated on a monthly basis to the industry level.

Controlling size-class estimates is done using an employment ratio adjustment. Table 4 shows an example of how the size-class estimates for an industry are adjusted to match the industry estimate. The sum of the size-class estimates are 4 less than the industry estimate. In this case, the two smaller industries would add 1 each, and largest would add 2 for a total of 4.

Table 4. Calibration Adjustment

	Estimated Level	Calibration Adjustment	Calibrated Level
Small	200	1	201
Medium	410	1	411
Large	620	2	622
Sum	1,230	4	1,234
Sector Level	1,234		

An analysis of calibration adjustments made to the size-class estimates shows that the average adjustments are less than 0.1% of the total series' employment level, and between 6% and 10% of the over-the-month change in absolute terms.

2.8 Seasonal Adjustment

Seasonal adjustment is the process of estimating and removing periodic fluctuations caused by events such as seasonal weather patterns, holidays, and the beginning and ending of the school year. Seasonal adjustment makes it easier to observe fundamental changes in the level of the series, particularly those associated with general economic expansions and contractions.⁵ Size class estimates, after they have been calibrated, are seasonally adjusted using X-12 ARIMA. The seasonally adjusted values are then calibrated to the seasonally adjusted sector level.

CES incorporates adjustments for several calendar effects when adjusting industry hours and earnings series. These same adjustments are also used for adjusting size-class estimates. If an adjustment for the length of pay period is significant then it is applied⁶. For series where a pay period adjustment is not significant, the series is adjusted for a 4 vs. 5 week effect. In addition, an adjustment for floating holidays (Easter and Labor Day) is used for hours series where it is significant.

2.8 Hours and earnings estimates

Estimates of average weekly hours and average hourly earnings are made after employment has been estimated following exactly the same methodology used for estimating industry hours and earnings. The estimates are calibrated to industry hours and earnings estimates by ratio adjusting total hours and gross earnings using a methodology equivalent to what is used for calibrating size-class employment levels. As with industry estimates, the average weekly hours and average hourly earnings series are not directly benchmarked, but small revisions may arise when controlling to benchmarked industry data.

⁵ More information on CES seasonal adjustment methodology is available at http://www.bls.gov/web/emp/sit/cessa_oview.htm.

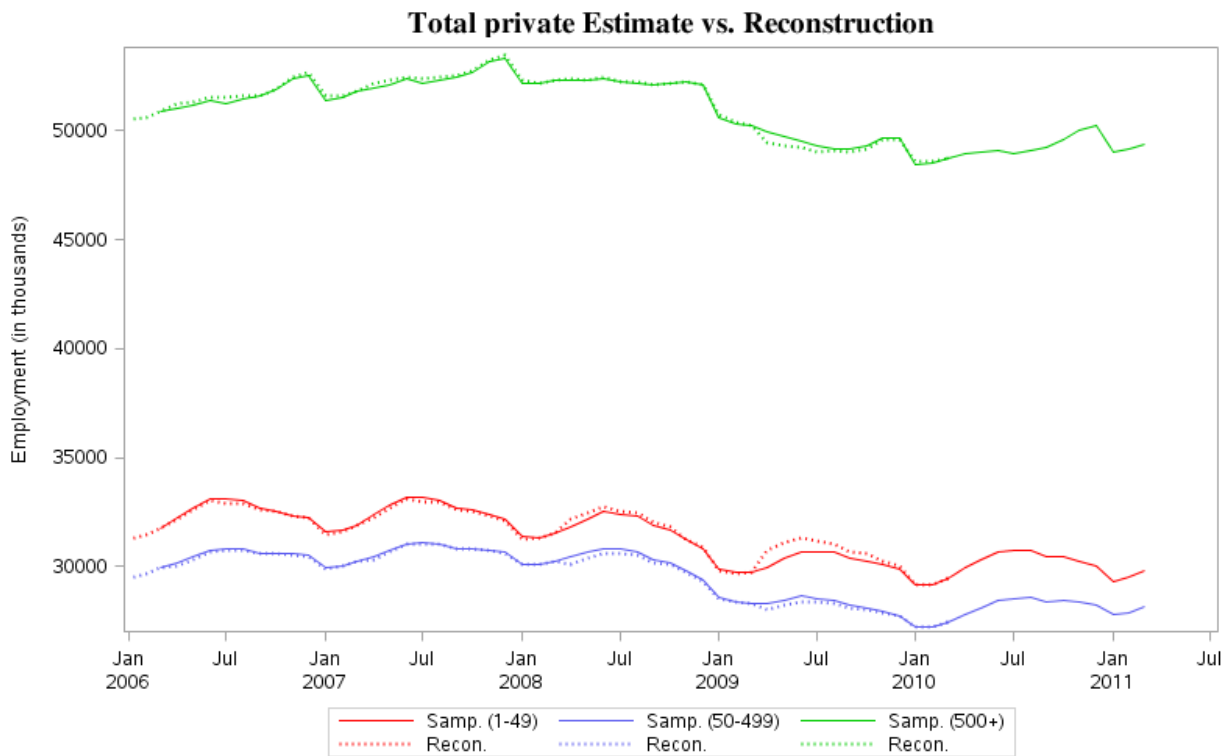
⁶ Significance is determined by a T-statistic greater than 2.

3. Historical Data

For several methodological reasons, BLS was not able to identify the size class of reports collected before March 2006. In order to provide a longer history of size-class employment levels, an alternate method was applied. This method the size-class benchmark methodology by applying relative employment ratios based on QCEW data to the CES employment level for all months prior to March 2006. As with establishing benchmark levels, this method accounts for scope differences between the QCEW and CES, while retaining CES industry seasonality.

A comparison of employment level estimates and the reconstructed historical data can be made for years when both series are available, as shown in Figure 2. Small divergences are observed in some months, but overall the economic usefulness of interpretation is identical. Data for 2009 was more problematic in terms of matching levels, with several industries showing noticeable divergences between the two methods. But overall there is no cause to think the reconstructed data is incongruous with the estimated data.

Figure 2. Estimated employment level and reconstructed employment level comparison



4. Evaluation of Methodology

4.1 Sample Adequacy

The CES sample is stratified by UI account, grouped by size of employment, which represents all of a firm's employment within a particular state. Without redefining CES sample selection methodology, it is not possible to assure that each firm (at least as defined by EIN) is included in the selected sample.

However, while it is not optimized for firm-based estimation, the sample is adequate for size-class estimation. There is a high correlation between UI size and EIN size which facilitates a sufficient sample within each sector for size-class estimation. Table 5 shows the number of EINs that are represented in the CES sample, on average from 2006 to 2011, and the coverage of all firms. Just over 58% of the sample is in the small size class, and the 14% of sample in the large size class represents over 62% of total firms in that class.

Table 5. CES Sample Coverage of Firm

Size Class	Represented EINs	Universe EINs	Coverage
Small (1-49)	58,124	5,168,967	1.1%
Medium (50-499)	27,559	256,939	10.7%
Large (500+)	13,844	22,055	62.8%
All	99,527	5,447,961	1.8%

4.2 Effect of Sample Movements between Size Classes

From the time of the original sample reference period until the sample is actually used in estimation, at least 21 months pass. For example, CES sample used for January 2012 estimates was selected from QCEW data through March 2010, and updated for births with data through September 2010. Because of this lag, large movements of sample between size classes could negatively impact the accuracy of size-class estimates. However, several factors work together to minimize this source of error. First, the fact that there are only three size groups means that there are only two size-class boundaries to cross.

Table 6. Sampled Employment Movements between Size Classes

Movement Direction	Large (500+)	Medium (50-499)	Small (1-49)	All
Constant	81.7%	10.9%	2.8%	95.4%
Into Size Class	1.0%	0.8%	0.7%	2.5%
From Size Class	0.6%	1.0%	0.5%	2.1%

Second, the maximum employment over the last 12 months measure adds stability, especially due to seasonal movements. As shown in Table 6, the vast majority (95.4 %) of establishments in the CES sample do not migrate across firm size-class boundaries. Further, the number of establishments moving into a size class roughly equals the number moving out of that size class. Because of this stability, it is reasonable to expect that the CES sample would produce reliable size-class estimates.

As an additional validation of this finding, a set of research estimates were run using the actual size of the selected sample units over the estimation period (information that would not be available in practice, since 12-month maximum employment cannot be calculated for future months). These “actual size” estimates did not differ significantly from the estimates using the standard sizing methodology.

5. Conclusion

The research suggests that the introduction of size class estimates would provide a valuable and reliable addition to what is currently published by the CES program. The available sample in the existing CES establishment survey, though not designed or optimized for estimating employment by firm size, is more than sufficient to make reliable estimates of the number of employees in firms by size class. In addition to this research, feedback from users of this research data has confirmed that the data is indeed a valuable supplement to the monthly industry employment numbers.