Current Employment Statistics by Size Class using Base-size Definitions October 2015

Nicholas Fett, Brenda Loya
U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212

Abstract
The Bureau of Labor Statistics is examining the feasibility of publishing employment statistics by size of firm, where the firm is defined at the EIN level. Estimates by size class are one of the most requested items to be published based from the Current Employment Statistics program. Producing monthly estimates by size class will allow data users to analyze the contribution to monthly employment numbers by small, medium, and large businesses. The estimates analyzed in this paper include the most recent recession and the subsequent recovery. This paper discusses an updated methodology chosen for creating estimates by size class, the rationale of using a base-sizing definition, limitations associated with constructing a time series of over-the-month changes, and the insights gained from examining the results.

Key Words: Size, employment, BLS, firm size class

1. Introduction

The Current Employment Statistics (CES) program developed experimental estimates by size of firm on February 2012[1]. The experimental monthly estimates track employment growth by keeping constant the size classification of the firm from the base of the period, or by using *base-sizing*. Base-sizing does not allow reclassification of the firm even if its growth exceeds the maximum employment level for the size class within the year. This method allows CES to track the employment change within each size classification. On an annual basis CES recalibrates sample based estimates to the population value, referred to as benchmarking the estimates, or the benchmark[2] process. The initial experimental estimates reclassified the firms based on the ending size of the firm during the benchmark process. The resulting time series merged two distinct methodologies, base-sizing for the monthly estimates and end of period sizing for the benchmarked estimates, making the benchmarked data inconsistent with the goal to track employment based on base-sizing. The methodology for benchmarking size of firm estimates and producing the net birth death factors used for monthly estimation have been re-examined. This paper discusses the updated methodology, the rationale of using a base-sizing, limitations associated with constructing a time series of over-the-month changes, and the insights gained from examining the results.

[1] The original experimental firm size estimates are available at http://www.bls.gov/ces/cessizeclass.htm.
[2] More information about CES benchmark methodology is available on the CES technical notes under the "Benchmarks" section at http://www.bls.gov/web/empsit/cestn.htm.

2. Methodology

In the new methodology examined, the annual benchmark process is different from the normal benchmark process since it involves two population values. One population value maintains the base-sizing and represents the ending employment for each size class, the *end benchmark level*: this allows CES to capture the growth within each size class from the beginning of the period. The second population value allows reclassification of the firm's size class, the *start benchmark level*: this allows CES to begin monthly estimation with an updated snapshot of the population's size class distribution. In this paper, employment estimates by size of firm are referred to as size class estimates.

CES monthly estimation procedures for size class estimates differs from normal monthly estimation in the following ways:

- Estimation occurs at the supersector[3] level instead of more detailed industry levels.
- Estimation cells are grouped by ownership, NAICS, and size class instead of just by ownership and NAICS.
- Size class estimates, both not seasonally adjusted (NSA) and seasonally adjusted (SA) are ratio adjusted to equal the industry estimates.

The net birth death factors are produced by size class using similar base-sizing methodology (section 2.4 describes the process for the creating the net birth death factors).

Three size classes are defined as small (size class 1 with 0 to 49 employees), medium (size class 2 with 50 to 499 employees), and large (size class 3 with 500+ employees). Using the updated methodology, experimental size class estimates, seasonally and not seasonally adjusted, have been produced at the supersector level from March 2006 to March 2014 for this analysis.

2.1 Benchmark process

The CES program recalibrates sample based estimates on an annual basis. Every January, the difference between the previous March population value (March is the ***benchmark month***) and the CES March estimate is distributed with a linear "wedge back" procedure to the previous April, the ***wedge period***. Estimates for April through October following the March benchmark, the ***post-benchmark*** period, are recalculated by applying the sample-based over the month change link to the new benchmark level and updating the net birth death factor. For example, on January 2015, the difference between the March 2014 population value and the CES March 2014 estimate was wedged back to April 2013. The period from April 2013 through March 2014 is referred to as the ***wedge period***. Estimates for April 2014 through October 2014 were recalculated by applying the sample-based over the month change link and updating the net birth death factor.[4] The benchmark process applies to not seasonally adjusted estimates. Once the benchmark processing is complete, the benchmark series is seasonally adjusted and five years of data are revised with new seasonal factors.

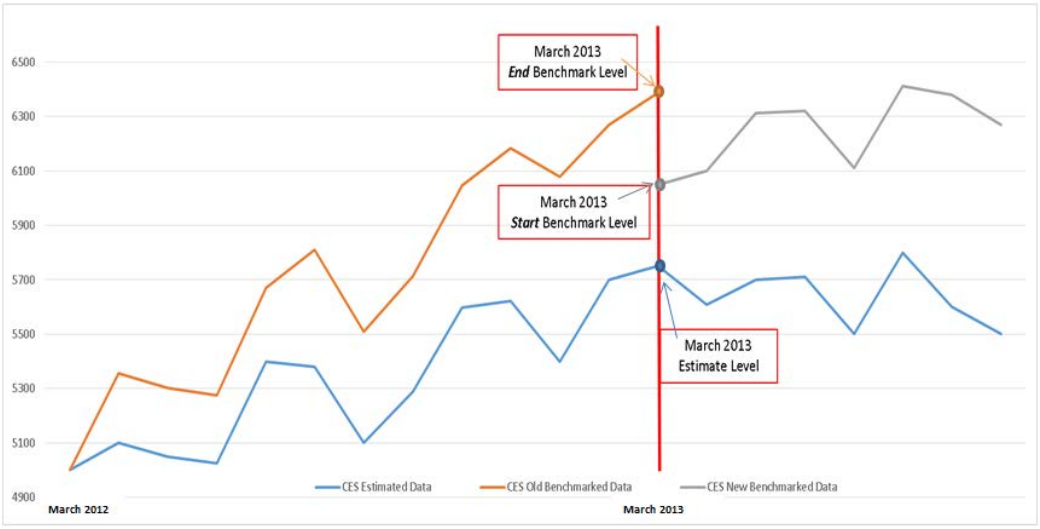
---

[3] Major industry and aggregate industry sectors are referred to as supersectors. The major industry and aggregate sectors are listed on the CES technical notes on table 6 and 7, respectively.

[4] More information about CES benchmark methodology is available on the CES technical notes under the "Benchmarks" section at http://www.bls.gov/web/empsit/cestn.htm.

For size class estimates, each benchmark, two separate benchmark levels of employment are established for each estimation cell: the *end benchmark level* and the *start benchmark level*. Both are derived by summing the universe employment levels[5] to the size class estimation cells. The *start benchmark level* (beginning of grey line in Figure 1) allows for size class reclassification and the *end benchmark level* (end of orange line in Figure 1) does not, hence, these values may not be equal. CES will not publish size class estimates levels since the difference between the end and start benchmark level can cause a series break (see Figure 1 below).

Figure 1: Size class benchmark process*



**Note: not actual data*

The *end benchmark level* is used for the wedge back procedure and the wedge is functionally the same for size class estimates as for CES industry estimates. However, once the wedge is performed on the size class estimates, the wedged size class estimates are ratio adjusted to the level of the industry estimates to ensure consistency across CES products. The size class post-benchmark period estimates are calculated differently than the normal CES industry estimates. Instead of using the sample-based over the month change link, the link is recalculated using the sample used in the monthly estimates but using the new *start benchmark level* sizing. The monthly estimates, which begin after the post-benchmark period, continue to use the base-size assigned in the post-benchmark period.

Base-sizing

The employment identification number (EIN) is the definitional base for the "firm". The size of the firm is determined by grouping all establishments, identified by the unemployment insurance (UI) number, associated with an EIN and aggregating the employment. The size class is assigned based on the maximum aggregate employment for the firm within the most recent wedge period, the twelve months April through March. The establishments associated with the firm are assigned the same size class as the firm.

[5] The BLS longitudinal database was used to create the population values.

Setting the End benchmark level

For generating the *end benchmark level* of employment, establishments maintain the size class assignment from the beginning of the period to remove sizing re-assignments from affecting the difference between the estimate and the benchmark level. Maintaining the initial sizing assignments makes the end benchmark level consistent with the sizing used for monthly estimation. The end benchmark level is used to calculate the difference between the estimate and benchmark, which will be distributed through the wedge period.

Maintaining the same size class assignment is straight-forward for continuing units and deaths since these can be directly mapped to the original EIN where these occurred. Three categories of "births" have been identified. The birth units' base-sizing for the end benchmark level are assigned as follows:

- *True birth*: a unit that is not associated with any EINs from the beginning of the period. When the birth unit's EIN does **not** match any EIN from the beginning of the period, the birth unit is assigned to the small size class (0 to 49).

- *Expansion unit from one EIN*: a new worksite opened by an EIN that existed at the beginning of the period. In cases where the birth unit's EIN can be traced back to an EIN that existed at the beginning of the period as a **one-to-one** relationship, the birth unit is assigned the size class of the EIN at the beginning of the period.

- *Expansion unit from multiple EINs*: a new worksite opened after two or more of the initial EINs complete a merger. When the birth unit's EIN matches more than one EIN from the beginning of the period in a **one-to-many** relationship, the birth unit's employment is distributed proportionally across the size classes assigned to the EINs at the beginning of the period.

Setting the Start benchmark level

The *start benchmark level* is generated using the updated size class assignments based on the maximum size of the EIN in the 12 month period corresponding to the latest wedge period. The updated size class assignments are used for estimation moving forward off of the March benchmark through post-benchmark into monthly estimation.

2.3 Base-sizing and dynamic-sizing

The Bureau of Labor Statistics Business Employment Dynamics (BED) program produces data by size class. BED uses the Quarterly Census of Employment and Wages (QCEW), universe data, which is lagged administrative data to derive size class estimates. BED data are lagged two quarters. CES also uses QCEW data, adjusted to CES scope, as the basis for the benchmark level and the birth death factors, but produces more timely employment estimates on monthly basis using probability sample (for all industries except government).

The main difference of the BED estimation procedure is that the BED produces estimates using a quarterly dynamic-sizing method and publishes estimates quarterly on a lagged basis. For this experimental data, CES uses annual base-sizing method but recreates estimates on a monthly basis. BED is published quarterly and uses dynamic sizing estimates to allow the reclassification of establishments each quarter and distributes the

employment change into the corresponding size class when the size class threshold is crossed. To simplify the following example, assume BED uses the same size classes as CES experimental size class estimates. For example, a firm is assigned to size class 1 at the beginning of a quarter based on the employment reported of 45 employees. The next quarter the firm reports employment of 55, which crosses into size class 2. The employment change of 10, would be distributed by adding 4 to size class 1 the growth from 45 to 49) and 6 to size class 2 (the growth from 49-55). On the other hand, annual base- sizing method does not allow employment to cross size classes and CES would have captured the entire change of employment of 10 in size class 1. Annual base-sizing tracks employment growth in each size class without allowing for a reclassification until the next year when new sizes are given to each establishment. This annual base-sizing method will thus show a larger proportion of growth in sizes class 1 since all new firms are placed into the smallest size class (zero employment in previous year). For BED however, estimates are resized quarterly so establishments can grow out or shrink into different size classes every quarter. BED also allows for birth units to be dynamically sized in the same way that firms would grow. For example, if a new business has 100 employees in its first month, BED would show a growth of 50 in size class 1 and 50 in size class 2. CES however would show all 100 new employees in size class 1. In addition, if that unit then grew to 150 the following quarter, CES would still have that business as size class 1 whereas BED would have it in size class 2.

A full comparison of the sizing methods can be found in the 2006 Monthly Labor Review article: *Business employment dynamics: tabulations by employer size.*[6]

2.4 Birth/death model

Business deaths are difficult to capture since it is problematic to discern survey non-respondents and business deaths. Births do not exist in the universe at the time the sample is chosen. Births and deaths are eventually observed on a lagged basis in the population. CES uses a model based adjustment in conjunction with an implicit imputation to account for births and deaths that cannot be captured by the sample in a timely manner. The birth/death model is a two-step process. First, by using only continuous units to calculate the monthly rate of growth, CES implicitly imputes the same rate of growth to business deaths and non-respondent units. The second step is a model based adjustment created by modeling the residual of the birth and death employment change, which research has shown is relatively stable. The birth/death model forecast uses five 24-month long spans of input data, or ***frames***, representing historical net births and deaths. To create the net births and deaths, simulated monthly probability estimates are created, by explicitly imputing deaths and non-respondent units with the continuous units' rate of change. Then the level differences are calculated between the population and the simulated monthly probability estimates. The level differences are then converted into over the month changes, the net births and deaths. The net birth/death factors are forecasted using an Auto-Regressive Integrated Moving Average (ARIMA). The factors are used for monthly estimation and updated on a quarterly basis with the latest available population information. During the benchmark process the post-benchmark estimates are updated with the latest forecasted net birth/death factors. [7]

---

[6] http://www.bls.gov/opub/mlr/2006/02/art1full.pdf?origin=publication_detail

[7] More information about the birth death model is available in the technical notes under the birth/death model section at http://www.bls.gov/web/empsit/cestn.htm#section5c.

The net birth death factors for the CES industry estimates and the size class estimates are created using the same methodology, but are sliced a different way. The frames used to create net birth death factor used in the CES industry estimates are grouped by ownership and NAICS. The size class net birth death frames are grouped by ownership, NAICS, and size class. The size class is assigned to establishments based on the maximum employment level of the EIN in the 12 month period immediately prior to each of the five frames used in the forecast. Once the size class is assigned for the frame, it remains unchanged through the frame, for consistency with the base-sizing used in size class estimates. Additionally, the same birth unit categories and classifications as the *end benchmark level,* are used. Hence, size class one was assigned the *true birth* units, and any *expansion units* associated with the small size class. Size classes two and three are assigned only expansion units. Continuous and death units are mapped to their starting period size class.

The birth death factors for the monthly and benchmark size class estimates are ratio adjusted to aggregate to the CES industry estimates birth death factors.

For the monthly size class estimates, the net birth/death forecasts are used according to normal production rules. However, for the benchmark estimates the net birth/death derived from the population file are used rather than the forecast, referred to as ***actuals***. The reason for breaking from traditional practice is twofold; 1) using the actuals is more accurate than the forecast and 2) the birth death factor ratio adjustment, that ensures aggregation to the CES industry estimates birth death factor, is larger than the difference between the forecasted and the actual value observed in the population file.

2.5 Seasonal Adjustment

CES uses 10 years of historical data as inputs to seasonal adjustment. The input data is created by removing the strikes and buildup of employment associated with the decennial census from the not seasonally adjusted estimates. These non-seasonal movements are removed to avoid an impact on the magnitude of seasonal adjustment. To seasonally adjust the size class estimates, a series that goes back at least 5 years is needed. Due to the resizing of the data each year, seasonal adjustment of the size class estimates would not be possible due to the breaks in levels and problems arising from changes resulting in these breaks. If no adjustment was made, the break in the start benchmark and end benchmark levels would appear as seasonality and thus distort our estimates. To correct this problem and create a consistent level series, the over-the-month changes are linked backwards from the final benchmark level (i.e. March 2014 for the final series). The series is then seasonally adjusted and the over-the-month changes derived from this series are the published seasonally adjusted size class data.

CES will not publish levels for seasonally or not seasonally adjusted size class estimates since the difference between the end and start benchmark level can cause a series break on the not seasonally adjusted data. Only over the month change for seasonally and not seasonally adjusted size class estimates would be published.

3. Results

The original monthly and benchmarked size class estimates created for this analysis produced two separate sets of estimates for over the month changes. Although only over-the-month changes will be published, this paper provides detailed analyses of levels along with information on the birth death factors.

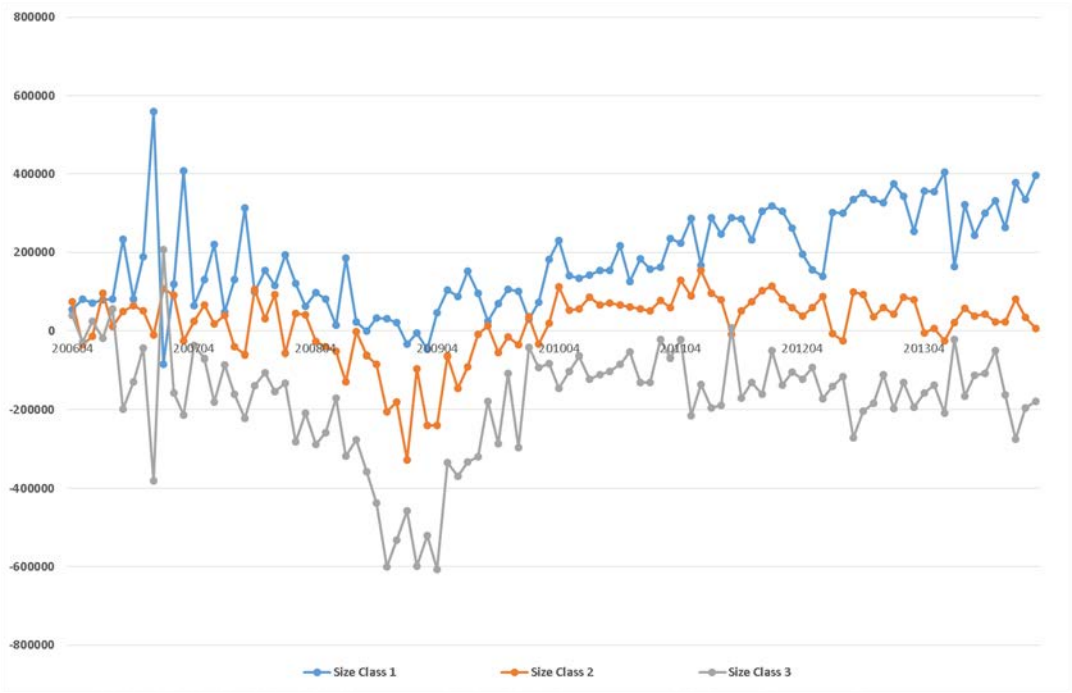
A sample of estimates to be published with the employment situation can be seen in Table 1. The over-the-month changes for December 2012 refer to the growth within each size from November to December.

Table 1. December 2012, Over the Month Changes by Size Class, in thousands

Industry	NAICS	Size 1	Size 2	Size 3	Total
Total Private	05	305	41	-123	223
Logging and Mining	10	5	3	0	7
Construction	20	43	-4	-1	38
Manufacturing	30	27	2	-16	13
Nondurable Manufacturing	31	16	1	-6	11
Durable goods manufacturing	32	11	1	-10	2
Trade, transportation, and utilities	40	16	-2	34	48
Wholesale Trade	4142	9	5	-7	6
Retail Trade	42	0	0	6	6
Transportation and warehousing	43	5	-7	37	35
Utilities	4422	2	0	-2	1
Information	50	5	-3	-11	-9
Financial Activities	55	26	11	-27	9
Professional and business services	60	69	6	-40	35
Education and health services	65	57	25	-45	36
Leisure and hospitality	70	50	3	-14	40
Other services	80	7	1	-2	6

Table 1 can be seen in historical context within Figure 2. As both the table and graph show, size class 1 grows almost every month of the analysis period. The over the month change for each industry is ratio adjusted to aggregate to the CES industry estimates. The difference between the total size class estimate for the industry and the CES official industry total is distributed proportionally to each size class to ensure additivity. Publishing the over the month change by size class can provide an additional dimension to the monthly CES employment estimates. However, tracking employment growth within each size class using base-sizing creates a series break when the data is benchmarked, making it difficult to analyze the data as a time series even if only over the month changes are published. Creating a time series from different end and start benchmark levels has led to data that appears to tell conflicting stories.

Figure 2. Seasonally Adjusted, over-the-month change, total private, benchmark series



Size class 1 consistently accounts for the majority of growth since all births are assigned to size class 1, yet deaths can occur in any size class. This effect is due to the annual base-sizing methodology where firms cannot move from the initially assigned size class. On Figure 2, despite over-the-month changes appearing to come from one March level, these numbers are based on different sample composition and levels that stem from the previous March level. Size class 3 appears to show a decrease in employment share from 2006 to 2012. However, by examining the start benchmark levels, which allow resizing, size class 3 actually contains a larger share of total employment in 2012 than in 2006 as shown in Table 2. In general, each size class appears to maintain a consistent share of employment over time, in Table 2.

Table 2. Total private, distribution by size class at starting March benchmark level.

	Size 1	Size 2	Size 3
2006	27%	25%	48%
2007	26%	25%	49%
2008	26%	25%	49%
2009	26%	25%	49%
2010	26%	25%	49%
2011	26%	24%	50%
2012	25%	24%	50%
2013	26%	25%	50%

In terms of levels, Table 3 shows the levels for each of the supersectors in March of 2012. Although levels will not be published monthly, a start benchmark period such as this could be an option to publish for users to have a baseline for tracking over-the-month changes.

Table 3. March 2012, Industry supersectors by Size Class, not seasonally adjusted, in thousands

Industry	NAICS	Size 1	Size 2	Size 3	Total
Total Private	05	27,886	26,930	55,341	110,157
Logging and Mining	10	158	218	460	836
Construction	20	2,654	1,666	994	5,313
Manufacturing	30	1,965	3,758	6,098	11,822
Nondurable Manufacturing	31	1,311	2,401	3,703	7,415
Durable goods manufacturing	32	654	1,357	2,396	4,407
Trade, transportation, and utilities	40	5,711	4,756	14,616	25,082
Wholesale Trade	4142	1,669	1,669	1,669	1,669
Retail Trade	42	3,175	2,024	9,376	14,574
Transportation and warehousing	43	826	925	2,597	4,348
Utilities	4422	41	41	41	41
Information	50	334	526	1,813	2,672
Financial Activities	55	1,613	1,513	4,599	7,726
Professional and business services	60	3,967	3,938	9,697	17,601
Education and health services	65	4,038	5,586	10,754	20,377
Leisure and hospitality	70	4,403	3,738	5,193	13,334
Other services	80	3,044	1,232	1,118	5,394

Figure 3 shows the seasonally adjusted level series for the research period. Note that each series is created by applying the over the month change link to the latest not seasonally adjusted value (the latest benchmark level) back in time to create a level series for input for seasonal adjustment. Since size class 1 grows most months, it appears as though size class one grew very quickly throughout the period. In reality, the final start benchmark level is correct, but the levels before that are not accurate as they are not readjusted to account for breaks in size class.

Figure 3. Seasonally Adjusted, Total private levels

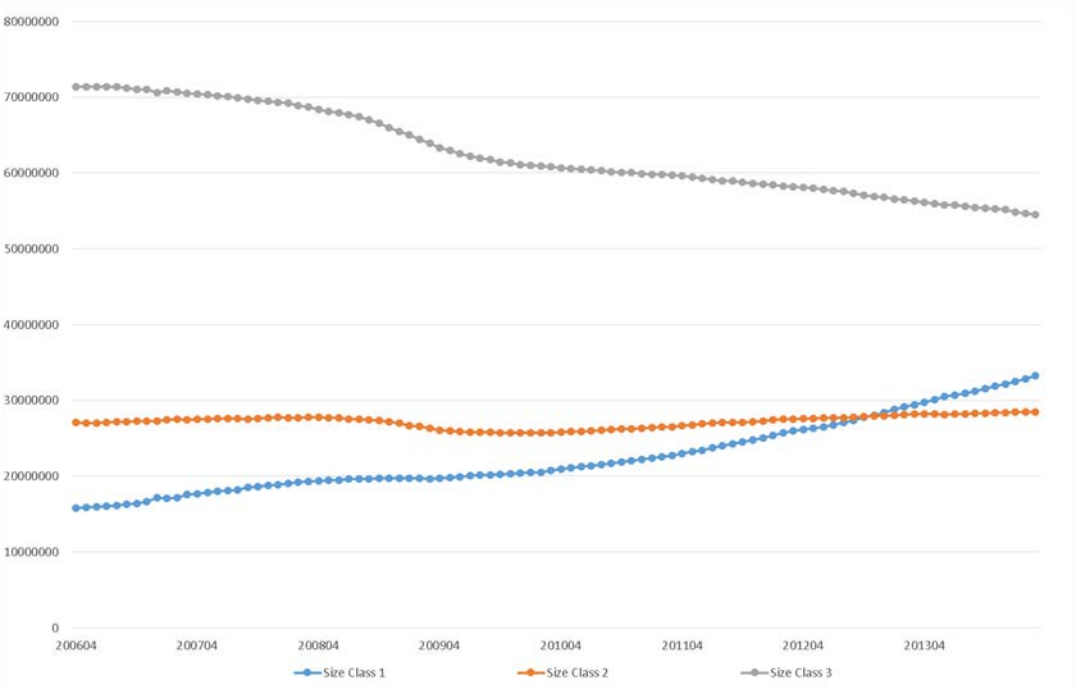


Figure 4 shows the not seasonally adjusted level series for total private size class 1. It shows the level series of not seasonally adjusted values from which the over-the-month change links are created from to produce the input data for the seasonal adjustment procedure. The large jumps downward every March are the result of resizing. Since establishments cannot grow out of size class one during the year, they grow every year and are then resized to a lower level. The breaks, shown in Figure 4, are the reason why seasonal adjustment of this series is not feasible without creating the smoother series from Figure 3, why CES has decided to publish only over the month changes, and why two benchmark levels are needed. On Figure 4, comparing March levels one can see that from 2006 to 2013, size class 1 actually decreased in employment share but Figure 3 shows an increase.

Figure 4. Not Seasonally Adjusted, Total private, Size Class 1

The results in Figure 4, seem to indicate that firms who are less than 50 employees at the beginning of the year are likely to experience higher growth than firms with 50 or more employees due to the addition of births. Figure 4 also appears to indicate that our birth-death model is understating births as the benchmark figure is consistently upwards spare 2009. This consistent upward benchmark of size class 1 stems from under-counting of births in size class 1 due to the ratio adjustment that ensures aggregation to the CES industry estimates birth death factor and the from having an upward bias in terms of employment change (zero lower limit to employment, but no upper limit). More research is needed on the modeling of the birth death estimate for smaller sized units.

Comparison to BED

Figure 5 shows the BED size class estimates and the CES size class estimates. BED estimates are aggregated to create the same size classes used by CES. CES size class 2 is notably close to the BED size class 2. Size classes 1 and 3, however, show a large divergence. The discrepancy is due to the difference in methodology used by CES and BED. CES size class 1 is attributed all the true birth units when setting the ***start benchmark level*** and for modeling the net birth death factors, resulting in large positive factors. CES size class 3 is allowed expansion units and deaths when setting the ***start benchmark level*** and for modeling the net birth death factors, resulting in large negative factors. Net birth death factors are modeled using an autoregressive model, hence, the factors are highly dependent on the history of the input series. By design, size class 1 and size class 3 are expected to have highly positive and highly negative net birth death factors, respectively.

BED does not need to model births and deaths because they use universe data. However, since CES produces a sample based estimate it is important that births and deaths be incorporated with the birth death model. Figure 6 shows CES data after removing the net birth death factors from the CES estimates. CES estimates without the net birth death factors applied yield time series that more closely resemble BED size class data.

The data from BED and CES show that all size classes experienced slower or negative growth during the 2008-2009 recession. The same trend was observed for CES with and without the net birth death factors applied.

Figure 5. BED vs CES over the month changes

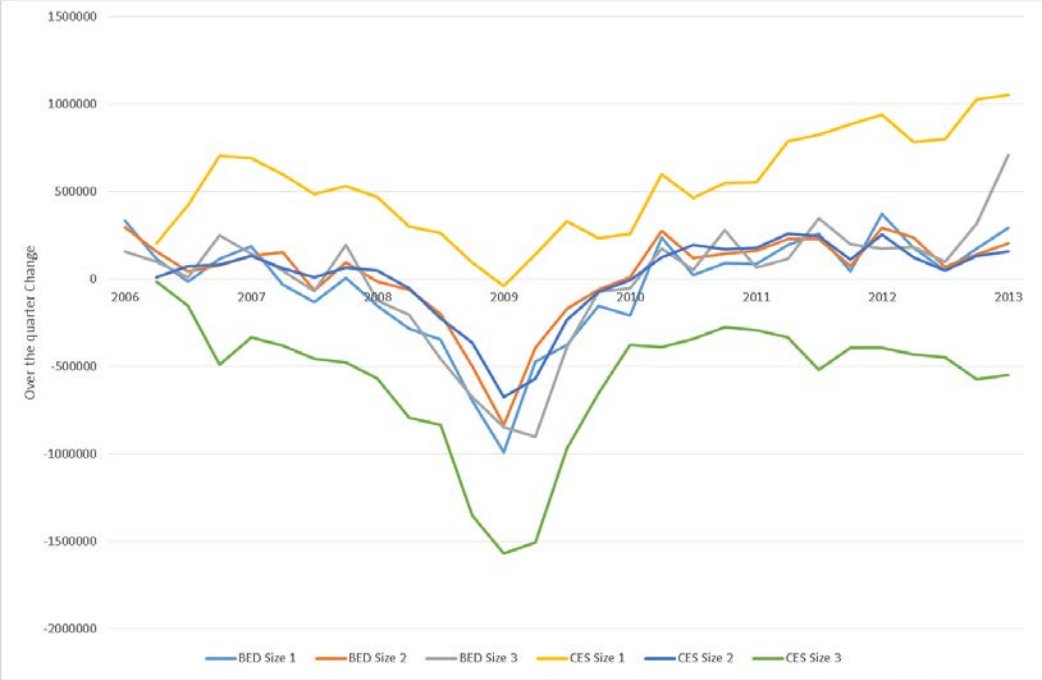
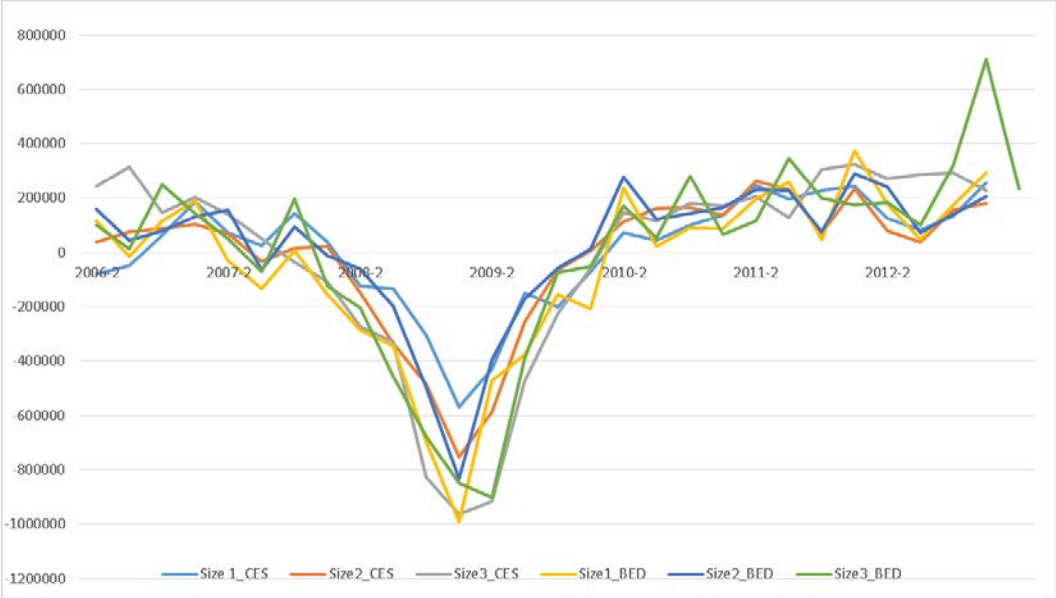


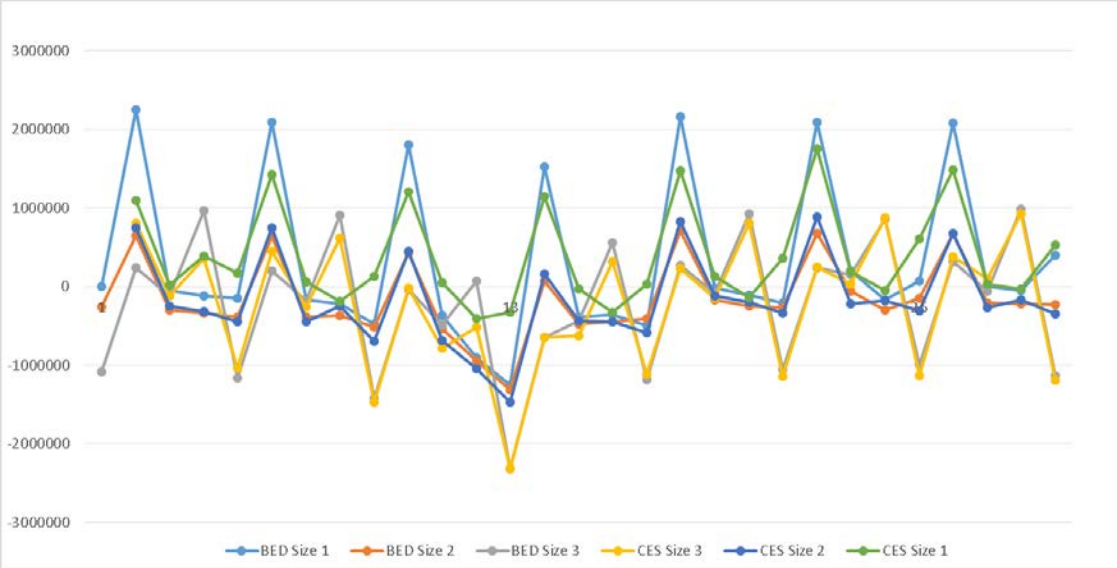
Figure 6. BED vs CES over the quarter change (CES excluding net birth death factors)



In addition to their published size class data, BED has produced experimental size estimates based on static sizing methodology. Figure 7 shows the comparison of the

experimental static BED data to the new CES series (quarterly averages). Even though their static series is resized every quarter (as opposed to annually for CES), using similar static methodology leads to very similar data across the size classes. When sized using a similar method as CES, the BED program also captures the large gains in size class 1 and the more negative series in size class 3. Note that this data is not seasonally adjusted and due to the non-time series nature of the data (Longitudinal Database maintained by the QCEW program), seasonality comparisons across size class are not possible.

Figure 7. Static BED series vs CES over the quarter change, not seasonally adjusted



Conclusion

Presenting the results produced by CES's annual base sizing methodology may be difficult for average users of the employment situation to understand. It appears on the surface that small businesses have had a revival over the past six years when looking at growth numbers. This is not true however since annual resizing has shown that small businesses have actually decreased in employment over the time period. The results are consistent with other base sizing methodologies in that the majority of growth comes from small firms. [8,9,10]

The data does provide users with information on size class, allowing users to look at the effect of the recession and other exogenous shocks to employment stratified by size class. The data provide insight into the nature of firm growth and provide the first monthly time series of employment by size class from the BLS. Future CES research will include estimation by age of firm, which can then be compared to size class data to analyze if new units are the true driving force behind the large size class 1 estimates.

[8] http://www.bls.gov/ore/pdf/st060020.pdf
[9] http://www.bls.gov/opub/mlr/2006/02/art1full.pdf?origin=publication_detail
[10] http://www.bls.gov/opub/mlr/2004/07/art1full.pdf

Acknowledgements

The authors would like to thank Kirk Mueller, BED, and ERPDS for their input and guidance in developing the concepts and methodology for consistent base-sized CES estimates.

References

Butani, Clayton, Kani, Spletzer, Talan, and Werking Jr. *Business employment dynamics: tabulations by employer size.* Monthly Labor Review. February 2006. http://www.bls.gov/opub/mlr/2006/02/art1full.pdf?origin=publication_detail

Butani, Grden, Kapani, Werking. *A Multi-Dimensional Analysis of Size Class Methodologies and Employment Changes—March 1993 to March 2003* 03/06/2006. http://www.bls.gov/ore/pdf/st060020.pdf

Okolie, Cordelia. *Why size class methodology matters in analyses of net and gross job flows* Monthly Labor Review. July 2004. http://www.bls.gov/opub/mlr/2004/07/art1full.pdf

Disclaimer

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