Improving Estimates of Employment in Expanding and Contracting Businesses

Kenneth W. Robertson, Joshua Duffin, and Jennifer Kim
U.S. Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Washington, D.C. 20212

Abstract
The Current Employment Statistics (CES) program produces estimates of employment, hours, and earnings by industry on a monthly basis for the non-agricultural economy. Recent research has suggested that the CES respondent data can be used to produce estimates of employment change in businesses with expanding and contracting employment. These research estimates have included a small overestimate because we have not accounted for the prior month employment of establishments going out of business. In this paper we will describe research to quantify, and to potentially account for, this issue. Data from the Quarterly Census of Employment and Wages program will be used with CES data to assess the size of the overestimate. ARIMA-X12 will be used to develop forecasts of factors that we will use to adjust the estimates to account for this error. The forecast factors will be compared to the actual factors to assess the feasibility of modeling these values.

Key Words: ARIMA-X12, Current Employment Statistics, business employment dynamics, Quarterly Census of Employment and Wages

Improving Estimates of Employment in Expanding and Contracting Businesses

Kenneth W. Robertson, Joshua Duffin, and Jennifer Kim
U.S. Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Washington, D.C. 20212

Background

The U.S. Bureau of Labor Statistics' (BLS) Current Employment Statistics (CES) survey is a monthly business survey that produces timely estimates of employment, hours, and earnings. The estimates are produced by industry for the nation, states, and metropolitan areas. These data are among the first indicators of the health of the U.S. economy. In-sample responding businesses provide these data each month for the pay period that includes the 12th day of the month. Preliminary estimates are published about 3 weeks after the reference period, with revised estimates published the following two months.

Another program of the BLS is the Quarterly Census of Employment and Wages (QCEW). This program collects information based upon the Unemployment Insurance (UI) program. The UI program collects quarterly data on taxable wages and monthly employment data from most businesses. These data are published quarterly about 7 to 9 months after the reference period. Among the products tabulated from these administrative records is the Business Employment Dynamics (BED) data. These data disaggregate the total quarterly employment change into four component parts: employment change in expanding businesses; employment change in contracting businesses; employment change in units that went out of business; and employment change in newly opened businesses. Previous research (Robertson and Roosma, 2009) has shown that the CES data can be used to produce timely monthly estimates of employment change in expanding and contracting businesses.

Problem

The Current Employment Statistics Business Employment Dynamics (CES-BED) estimates are restricted to measures from continuous businesses. That is, we can measure the employment change from businesses whose employment grew or expanded over the month, and from businesses whose employment declined or contracted over the month. CES data cannot, however, be used to accurately measure employment change from business startups or business closings.

In previous work we defined expansions and contractions from CES data as follows

[1] $$\hat{E}_t = \hat{X}_{t-1}\left(\frac{\sum_{i\in q} w_i\left(x_{i,t} - x_{i,t-1}\right)}{\sum_{i\in m} w_i x_{i,t-1}}\right)$$

[2] $$\hat{C}_t = \hat{X}_{t-1}\left(\frac{\sum_{i\in r} w_i\left(x_{i,t-1} - x_{i,t}\right)}{\sum_{i\in m} w_i x_{i,t-1}}\right)$$

Where:

t=current month
t-1=previous month
$i \in m$ = linked business "i" reported for both the current and the prior month

w_i = sample weight for business "i"
x_i = employment for business "i"
$\hat{X}_{t-1}$ = published employment estimate for previous month
$\hat{E}_t$ = net change in employment from expanding businesses
$\hat{C}_t$ = net change in employment from contracting businesses
q is the subset of units whose employment grew over the month
r is the subset of units whose employment declined over the month

The quantity $\hat{X}_{t-1}$ includes employment from businesses that remained open over the month, and the last month of employment for businesses that closed over the month. The latter quantity is small, but if we don't account for it then we include a small positive bias in each estimate of $\hat{E}_t$ and $\hat{C}_t$. This paper will describe research to determine how large this bias is, and to determine if we can remove it from these estimates.

Procedures
The primary statistic of interest in this research is the adjustment needed to modify the prior month employment value so that it excludes employment from businesses that closed over the month. This will allow the expansion and contraction statistics to appropriately reflect employment change restricted to those businesses that remained open over the month.

[3] $$\hat{X}^{c}_{t-1} = \hat{X}_{t-1}\left(1 - \frac{\hat{d}_p}{\hat{X}_p}\right)$$

Where:
p = period: *p* may represent either a month or a calendar quarter
$\hat{d}_p$ = prior month employment estimate for businesses that closed over the month

There are a number of procedures to evaluate with respect to this error component. First, we need to determine how large the unaddressed bias is, that is, how large is the proportion d_p / X_p. It would also be useful to understand how this error component affects the analysis of these data. Next, we need to evaluate several procedures to mitigate this error.

First, we want to understand how this error component affects the analysis of the data. Obviously, at a given point in time, the impact is to slightly overestimate the level of expansions and contractions. The impact on the level at a point in time will be evaluated empirically later. For now, let's briefly turn our attention to the impact of the error component on an analysis of change in levels over some period. We start by showing the comparison of estimates of employment in expanding businesses at two points in time.

$$\hat{E}_{t-1} - \hat{E}_{t-2} = \left[\hat{X}_{t-1}\left(1 - \frac{\hat{d}_{t-1}}{\hat{X}_{t-1}}\right)\left(\frac{\sum_{i \in q} w_i \left(x_{i,t} - x_{i,t-1}\right)}{\sum_{i \in m} w_i x_{i,t-1}}\right)\right] - \left[\hat{X}_{t-2}\left(1 - \frac{\hat{d}_{t-2}}{\hat{X}_{t-2}}\right)\left(\frac{\sum_{i \in q} w_i \left(x_{i,t-1} - x_{i,t-2}\right)}{\sum_{i \in m} w_i x_{i,t-2}}\right)\right]$$

Now, let's assume that each quantity in the comparison is equal except for the error components. This assumption gives us the following.

$$\hat{E}_{t-1} - \hat{E}_{t-2} = \left[\hat{X}\left(1 - \frac{\hat{d}_{t-1}}{\hat{X}_{t-1}}\right)\hat{Q} - \hat{X}\left(1 - \frac{\hat{d}_{t-2}}{\hat{X}_{t-2}}\right)\hat{Q} \right] = \hat{X}\hat{Q}\left[\frac{\hat{d}_{t-2}}{\hat{X}_{t-2}} - \frac{\hat{d}_{t-1}}{\hat{X}_{t-1}}\right]$$

As the error components $\hat{d}/\hat{X}$ from the two periods approach each other, the impact on the analysis approaches 0. It is clear, however, that if these error components differ substantively then the impact on an analysis may not be negligible.

The measures that we will explore in this research to mitigate the error are the following:

- Develop quarterly averages from several years of QCEW-BED closings data. Divide this estimate by 3 to develop a monthly average for each quarter. For example, we could take three recent 1^{st} Quarter measures and use their average as a forecast for this quantity in the current 1^{st} Quarter. Note that this measure is not sensitive to intra-quarter changes in the normal monthly pattern of business closings.
- Develop a history of empirical data on business employment in the month before the business closes. We will develop forecasts of these data using SAS™ Proc Forecast. Note that the purpose of this research is not to develop models to produce the best forecast available, but rather to evaluate if a reasonable forecast can mitigate this error component. Later research can evaluate and identify the best forecasting models.

Both alternatives will be reviewed across several years to evaluate the effectiveness of each in eliminating the bias in the expansion and contraction statistics.

The effectiveness for each measure will be evaluated using three calculations. The first is an average absolute error measure.

[4] $$\hat{\delta} = \frac{\sum_{p \in P}\left|\frac{\hat{d}_p}{\hat{X}_p} - \frac{d_p}{X_p}\right|}{n_P}$$

Where:
P = the set of monthly periods to evaluate
n_P = the number of months in P

This measures the average absolute deviation of each estimated proportion from its true value. This statistic is used to evaluate how close the estimate is, on average, to the true value of the statistic. That is, it informs us about the average magnitude of the error. In a comparison the smaller statistic indicates the better estimator.

The second calculation is an average error measure.

[5] $$\hat{\varepsilon} = \frac{\sum_{p \in P} \left(\frac{\hat{d}_p}{\hat{X}_p} - \frac{d_p}{X_p} \right)}{n_P}$$

This measures the average error of the estimate from its true value. This statistic is used to evaluate the effectiveness of the estimate in removing bias. A statistic close to zero indicates an unbiased estimator.

The third calculation is a standard deviation measure σ. This statistic will measure the dispersion of the difference between the estimator and the true value. A small standard error value tells us that the estimate is usually close to the true value, while a larger value tells us that the estimate has a larger dispersion about that value. In comparisons between estimators, a smaller standard error value is better.

Data

QCEW-BED

In order to assess the size of the error caused by including employment from closing establishments in the calculation of employment change in expanding and contracting establishments, data from the Bureau's QCEW-BED program was used. These data were also used to verify that tabulations of closing employment based on QCEW-microdata were reasonable. QCEW-BED quarterly employment closing data from 2003 to 2010 were obtained for natural resources and mining, construction, manufacturing, wholesale trade, retail trade, transportation and warehousing, utilities, information, financial activities, professional and business services, education and health services, leisure and hospitality services, and other services (excluding public administration). The range of employment closing within each super-sector was tabulated.

The not-seasonally adjusted quarterly level and rate of employment associated with closing establishments was obtained for each super-sector. From this, a moving average for each Quarter's employment level and rate were computed within each super-sector. For example, the quarterly moving average forecast for 2009 Quarter 1 would be the average of the Quarter 1 data for 2006, 2007, and 2008. The quarterly forecast rate was then by divided by 3 to create a monthly forecast rate. Although this measure is not sensitive to intra-quarter changes, it does quantify the adjustment due to closing businesses and provides an approximation for the size of this error component on estimates of employment changes in businesses with expanding and contracting employment.

QCEW-microdata

To create a time series for employment lost in closing establishments, we used longitudinally linked data from the Bureau's Longitudinal Database (LDB); these linked data are derived from the QCEW program. We used LDB data starting in Quarter 1 of

2000 and going through Quarter 4 of 2010. Starting with the full universe of QCEW linked data, we selected establishments that had at least one employee at some point during that time-span, and whose employment had gone to zero (and stayed there) or gone out of business by December 2010. We then excluded those establishments that had a "successor" listed (i.e. the establishment had not actually closed, but was now reporting under a new identifier). We then added up the final positive employment for each establishment that closed by super-sector, resulting in a time series of last-month closing-employment for each month from January 2000 through November 2010 for each major industry.

Results

The results of these procedures met our expectations. In general, the range of *True Values* tabulated from QCEW microdata were generally near the range of data published by the QCEW-BED program. Deviations were expected, as we are comparing monthly tabulations with quarterly tabulations adjusted to a monthly value. Within a quarter there can be both accumulations and offsetting movements for establishment closings. The range of data forecast using the *True Values* falls within an expected range for a forecasting procedure that reduces the impact of extreme value in a time series. The ranges of the closing rates are presented below, for the True Values, the QCEW-BED averaged forecast, and the forecast using the *True Values*.

Ranges of Closing Rates by Super-Sector

Super-Sector	Range of $\hat{d}_p / \hat{X}_p$ for *True Values*	Range of $\hat{d}_p / \hat{X}_p$ for p = Quarterly / 3	Range of $\hat{d}_p / \hat{X}_p$ for p = Monthly
Mining and Logging	0.0% - 1.5%	0.5% - 1.0%	0.1% - 0.6%
Construction	0.1% - 1.3%	0.6% - 1.2%	0.1% - 1.0%
Manufacturing	0.0% - 0.8%	0.1% - 0.3%	0.0% - 0.3%
Wholesale Trade	0.1% - 1.2%	0.3% - 0.5%	0.1% - 0.6%
Retail Trade	0.1% - 0.7%	0.3% - 0.4%	0.1% - 0.5%
Transportation and Warehousing	0.0% - 0.8%	0.3% - 0.5%	0.0% - 0.4%
Utilities	0.0% - 0.4%	0.1% - 0.2%	0.0% - 0.2%
Information	0.0% - 2.3%	0.2% - 0.6%	0.0% - 0.5%
Financial Activities	0.1% - 1.4%	0.3% - 0.6%	0.1% - 0.8%
Professional and Business Services	0.1% - 1.7%	0.4% - 0.8%	0.1% - 0.8%
Education and Health Services	0.0% - 0.4%	0.2% - 0.3%	0.0% - 0.2%
Leisure and Hospitality	0.1% - 1.1%	0.4% - 0.7%	0.1% - 0.7%
Other Services	0.1% - 1.3%	0.5% - 0.7%	0.2% - 1.0%

The next set of statistics we looked at were measures of the magnitude of the error. The first column presents the magnitude of the error if nothing is done to reduce it. The second column is the error if we were to use the QCEW-BED based forecast, and the third column is a simple forecast based on the time series of *True Values*. Surprisingly, the QCEW-BED based forecast does not reduce this error, but rather on average returns a value that is about the equivalent of not addressing the error. This may be because the monthly *True Values* data do not show results that approximate one-third of the quarterly value each month, but rather show a very large value in the third month of each quarter and much smaller values for the first two months of each quarter. This may be, at least in part, an administrative artifact of these data which are reported quarterly. As expected, the forecast based on the time series of *True Values* significantly reduces the magnitude of the error for each Super-Sector as compared to the *True Value* data series.

Magnitude Measures [smaller is better]

Super-Sector	$\hat{\delta}$ for d Not Estimated	$\hat{\delta}$ for p = Quarterly / 3	$\hat{\delta}$ for p = Monthly
Mining and Logging	0.26%	0.52%	0.04% *
Construction	0.37%	0.49%	0.05% *
Manufacturing	0.13%	0.12%	0.02% *
Wholesale Trade	0.26%	0.25%	0.04% *
Retail Trade	0.20%	0.19%	0.04% *
Transportation and Warehousing	0.17%	0.21%	0.02% *
Utilities	0.05%	0.08%	0.02% *
Information	0.24%	0.24%	0.05% *
Financial Activities	0.27%	0.29%	0.04% *
Professional and Business Services	0.30%	0.36%	0.04% *
Education and Health Services	0.08%	0.17%	0.01% *
Leisure and Hospitality	0.29%	0.32%	0.04% *
Other Services	0.44%	0.37%	0.06% *

* Smallest/best value in row

The next set of statistics we looked at was a bias measure. The results show that using a forecast based on the time series of *True Values* provides an essentially unbiased estimate of the *True Values* themselves to use to reduce the error.

Bias Measures [closer to 0 is better]

Super-Sector	$\hat{\varepsilon}$ for d Not Estimated	$\hat{\varepsilon}$ for p = Quarterly / 3	$\hat{\varepsilon}$ for p = Monthly
Mining and Logging	-0.26%	0.52%	-0.02% *
Construction	-0.37%	0.42%	-0.00% *
Manufacturing	-0.13%	0.06%	-0.01% *
Wholesale Trade	-0.26%	0.17%	-0.00% *
Retail Trade	-0.20%	0.14%	0.00% *
Transportation and Warehousing	-0.17%	0.19%	-0.00% *
Utilities	-0.05%	0.07%	0.00% *
Information	-0.24%	0.18%	-0.03% *
Financial Activities	-0.27%	0.19%	0.00% *
Professional and Business Services	-0.30%	0.28%	-0.02% *
Education and Health	-0.08%	0.17%	0.00% *

Services			
Leisure and Hospitality	-0.29%	0.30%	0.02% *
Other Services	-0.44%	0.12%	0.01% *

* Value closest to 0 in row

The final measures we looked at were standard errors of the forecast statistics. A smaller statistic indicates that forecasts are generally close to the *True Value*. The forecasts based on the *True Values* generally have small standard errors.

Standard Error Measures [smaller is better]

Super-Sector	$\sigma/\sqrt{n}$ for d Not Estimated	$\sigma/\sqrt{n}$ for p = Quarterly / 3	$\sigma/\sqrt{n}$ for p = Monthly
Mining and Logging	0.026	0.035	0.011 *
Construction	0.031	0.050	0.013 *
Manufacturing	0.015	0.017	0.004 *
Wholesale Trade	0.025	0.032	0.008 *
Retail Trade	0.017	0.022	0.008 *
Transportation and Warehousing	0.016	0.020	0.005 *
Utilities	0.007	0.008	0.005 *
Information	0.033	0.031	0.014 *
Financial Activities	0.028	0.038	0.012 *
Professional and Business Services	0.032	0.041	0.010 *
Education and Health Services	0.008	0.012	0.003 *
Leisure and Hospitality	0.024	0.031	0.008 *
Other Services	0.034	0.050	0.012 *

* Smallest value in row

Conclusions

This research has shown that it is possible to mitigate the error caused by inclusion of employment from closing establishments in estimates of employment in expanding and contracting establishments. The results of the empirical work point towards a forecast based on closings data developed from the Bureau's LDB as a reasonable way to reduce this error.

Future Research

There are several issues that remain to be researched with respect to these research series. We need to explore more efficient forecasting models, and evaluate them on a series by series basis. Another issue we need to explore a bit more is the potential impact on error reduction of what may be administrative artifacts in the LDB closings data.

References

Using Current Employment Statistics survey data to estimate employment in expanding and contracting establishments: preliminary results and issues. Kenneth W. Robertson and Mike Roosma, U.S. Bureau of Labor Statistics. Proceedings of the Joint Statistical Meetings, 2009.

Aligning Survey Estimates of Employment in Expanding and Contracting Establishments with Population Counts: Preliminary Issues and Methods. Kenneth W. Robertson, Mike Roosma, and James R. Spletzer, U.S. Bureau of Labor Statistics. Proceedings of the Joint Statistical Meetings, 2010.

All opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.