Estimation Procedures for the Bureau of Labor Statistics Current Employment Statistics Program

Shail Butani, Rachel Harter, and Kirk Wolter

Key Words: Ratio estimator, raked ratio estimator, generalized regression estimator, composite estimation, variance estimation, seasonal adjustment

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

I. Introduction

For the first time in decades, the Bureau of Labor Statistics (BLS) is planning a major sample redesign of the Current Employment Statistics (CES) Survey. An overview of the redesign including the transition and implementation phase is given in Werking (1997), while the goals, sample design parameters, features, and characteristics are described in Butani, Stamas, and Brick (1997).

The CES is a Federal/State cooperative program. The BLS produces national estimates at detailed industry levels, and the states produce industry estimates for state totals and major metropolitan statistical areas (MSAs); the level of state industry detail varies from state to state and by MSAs. The statistics estimated at the national level include all employees, production and nonsupervisory workers, and women workers. For production and nonsupervisory workers, the CES program also estimates average weekly hours and average hourly earnings. Additionally, BLS produces estimated average overtime hours for the manufacturing sector. The states publish estimates for all employees and some combination of the other statistics. The exact combination varies by state.

The primary advantage of the CES estimates is their timeliness. The preliminary estimates, generally released on the first Friday of every month for employment levels and for trends from the preceding month, are among the first economic indicators available. Several months later, more complete information becomes available from the BLS' Covered Employment and Wages Program, commonly known as the ES-202 Program. The ES-202 Program is an administrative file based primarily on each states' unemployment insurance programs in which virtually all employers participate. Under this program, the employers file <u>quarterly</u> contribution reports on employment for each month during the quarter and total quarterly wages for each of their U.I. accounts; a U.I. account may consist of more than one establishment (worksite or location). These establishments are known collectively as the Business Establishment List (BEL). Although there are minor differences between the CES and ES-202 in scope, coverage, and reporting procedures, the ES-202 data can be, and generally are, considered "truth" for employment for most industries. Today's CES employment estimates are benchmarked (aligned or ratio adjusted) once per year to the ES-202 figures for March, and the other statistics are recomputed accordingly. In the redesigned CES survey, the ES-202 data will play a major role in estimation as well as control totals for benchmarking employment.

In this paper, we summarize the research and recommendations with respect to the estimation process and identify known issues that remain to be researched. We begin by giving a brief overview of the steps in estimation after the collection of sample data. The new estimation procedures include: (1) editing procedures for erroneous data, (2) adjustment for misaligned data, (3) adjustments for outliers, (4) imputation for missing data, (5) simple unbiased estimation, (6) estimation incorporating ES-202 employment data as an auxiliary variable, (7) estimation for births/deaths, (8)composite estimation, (9) seasonal adjustment, (10) variance estimation, and (11) benchmarking or realignment to ES-202 employment data. Before describing these steps, we define selection weights and make distinctions between sampling units and establishments and between estimation and model cells.

II. Definitions and Terms

<u>Selection weights</u>—These are defined as inverse of the probability of selection.

<u>Sampling units and establishments</u>--As discussed by Butani, Stamas, and Brick (1997), the sampling unit is the unemployment insurance (UI) account. During the estimation process, however, all the reported data as well as information with respect to industry and MSA of each establishment will be utilized.

<u>Estimation and Modeling Cells</u>--The new sample design and estimation process make use of different types of cells, or groupings of units. There are allocation cells for determining sample sizes, selection cells, imputation cells, model cells for estimating parameters, and estimation cells.

For state estimates, model cells are MSA by industry groupings, while for national estimates they are by industry groupings only; the model cells are used for estimating parameters. In terms of industry detail, the nation and each state will have its own set of model industries since the importance of industries varies geographically (e.g., automobile manufacturing in Michigan, meat packing in Iowa). Within a state, model industries will also vary from major industry division (e.g., mining in Iowa) to 4-digit industry (e.g. meat packing in Iowa), taking into account benchmark employment, the number of establishments, and the percentage of employment in the sample; at the national level, they will vary from 2 to 4-digit SIC (Standard Industrial Classification) level.

Model industries also play an important role during imputation and one type of adjustment for outliers. The estimation cells can be at any level specified by the users; the estimation level may be at a higher or lower level than the model cell or model industry.

III. Editing and Adjustments for Erroneous Data

Some of the current edits are logical, while others are empirical that can be applied both at micro and macro levels. An example of a logical edit is that production workers must be equal to or less than all employees. An example of an empirical edit is to require a reporter's employment to fall within an expected range based on that reporter's or industry's past history. At the macro level, edits are typically based on observations of over-the-month trends in the context of historical levels and trends. The current edit procedures, however, are not standardized among the states. A workgroup is in the process of finalizing a standardized set of edits that are to be used by all states and by BLS. Additionally, these edits will be consistent with the ES-202 edits [ES-202 Manual] developed by BLS that have a proven track record over a number of years. A brief description of the ES-202 edits is given below.

The edit for employment data of continuing units is a sequential six step edit; once a record passes one of the steps, it proceeds to the total quarterly wage edit. In other words, a record must fail all six steps in order to be flagged for verification. In the first three steps, the comparison of the current month data is made to the previous month data; then to account for seasonality, the three steps are repeated and the comparison is made to the same month a year ago. In the first step, the test is on absolute difference; this test is designed to bypass reports that have small absolute change in employment, especially those that have been reporting constant employment in the past (e.g., 3, 3, 3, 3 to 5). It is estimated that over 50 percent of the records are bypassed with this edit. The second test is designed to account for small percentage changes (e.g., 500 to 540). The parameters for the first two tests are based on historical data and vary by employment size but not by industry. The third is a t-test that takes into account a record's variability. For new units, the employment must be less than some prespecified level, which is also based on historical data.

Wage data are edited in two sequential steps. In the first edit, a record is flagged if the absolute difference between the current and previous quarter wages exceeds some parameter (currently $10,000). Records that fails this edit are subject to an interquartile range edit based on the Hoaglin, Iglewizc, and Tukey test. Their work uses the upper and lower quartiles (F_U & F_L) to define outliers. To begin, the average quarterly wages are ranked from smallest to largest, with $X_{(1)}$ being the smallest value and $X_{(n)}$ being the largest value. The quartiles are then determined by the formulas $X_{(f)}$ and $X_{(n+1-f)}$ where $f = .5[\ (n+3)/2]$. Outliers are then defined as all observations falling:

below $F_L - k^* (F_U - F_L)$ or above $F_U + k^*(F_U - F_L)$, where k is a parameter.

A major consideration that was given in developing the ES-202 edits is their efficiency in terms of manual review (flagging the right records), computer processing, and timeliness. With seven million records (establishments) to process each quarter, the editing must be on a flow basis in order to be efficient. This also means the parameters must be based on historical data rather than on cell distributions at a current point in time which may change as more data are received and will certainly be different for state vs. national cells. The same thought process is being given to processing of CES data because of its vast sample size and general two week turn-around time between data collection and release of estimates. Editing on a flow basis is especially suited for CES because of its automated data collection environment (Clayton 1997); with online editing, it saves a call back to the respondents. The editing of the data will be performed at the U.I. account level (sampling unit). For employment data, each unit's history will be brought forward from the frame (ES-202). For other data series, historical data from the current CES will be used to set the parameters (e.g., ratio of production workers to all employee, and average weekly hours) until a unit establishes its own history. In order to prioritize the cases, a score will be assigned to each unit, and to improve the editing process over time, an audit trail is also being programmed.

IV. Outliers and Atypical Adjustments

The new design is being programmed to have atypical adjustments for three types of outliers. The identification of the first two types of outliers will be automated while the third type will be based on analyst judgment. To begin the atypical adjustment process, all units are assigned atypical weights equal to selection weights.

At the time of estimation, a sampling unit (U.I. Account) is considered a Type I outlier if it has grown in employment by three or more size classes or by more than 500 employees from where it was sampled---allocation and sampling size classification is determined by a unit's maximum monthly employment over the 12 month period on the frame. Type I outliers are termed "representative outliers" in that they are representative of the population as a whole at the state or allocation industry level on a probability basis but not of small domains like MSA or 4-digit SIC. A sampling unit that is identified as Type I gets an atypical weight that is equal to one-half of its selection weight, provided the selection weight is greater than or equal to 2.000. The atypical weights of the other sampling units in the state/allocation industry/size class are increased accordingly to account for the <u>excess weighted employment</u> of the Type I atypical unit [Business Survey Methods, Chapter 26].

Type II outliers are identified through comment codes that are assigned at the time of data collection; a common example of this type of outlier pertains to strikes by workers. Type II outliers are termed "nonrepresentative outliers" or "self-representative" units in that they are not representative of the population as a whole. A Type II outlier is given an atypical weight of 1.000, and the atypical weights of the remaining sampling units in the state/allocation industry/size class are increased accordingly to account for the <u>excess weight</u> (i.e. selection weight minus 1.000) of the Type II atypical unit. This way the sum of the atypical weights in an allocation cell is equal to the number of units in the population [Business Survey Methods, Chapter 26]. During the review of the estimates, the analysts will have the option to override both Type I and II outliers.

Type III outliers are those establishments that the analysts, during their review of the estimates, consider to be "self representative". These establishments are treated in the same manner as the Type II units; that is, they receive an atypical weight of 1.000. For operational reasons, the atypical weights of the remaining units in the state/model industry(not allocation industry)/size class are increased.

The reason that weights of the non-atypical units are being adjusted is that in CES even a very small underestimation bias is a major concern. One of the goals of the redesign is to keep the annual benchmark revision to within 0.2 percentage point.

V. Imputation

Explicit imputation for missing data is a new feature in the redesign. Because missing values will be imputed, weighting for nonresponse adjustment will not be necessary. For the most part, imputation for all employees is theoretically comparable to past practice. The new design also incorporates several alternative options, like the establishment trend of a year ago, and the analysts will be able to override an imputed value if better information from some independent source is available. Imputation is performed at the state/model industry/size class level. The four size classes for imputation purposes are 1-9, 10-49, 50-249, and 250 or more employees. For nonparticipants, the initial month employment will be imputed as the latest available ES-202 data times the sample trend of the model industry/imputation size class. Additionally, the imputed values will be replaced each time an updated ES-202 data become available. Employment is imputed as:

$$y_{AEit} = y_{AEi(t-1)} \left[\frac{\sum w_{jt}^{sel} y_{AEjt}}{\sum w_{jt}^{sel} y_{AEj(t-1)}} \right]$$

where, w_{jt}^{sel} is selection weight for unit j, and Σ indicates the sum across <u>all units j that were non-atypical and that responded for both months t and t-1</u>. It is important to note that the selection weight in the denominator is for time period t and not t-1 because of matched respondents, sample rotation, and updates to the sample due to frame maintenance.

Four different imputation methods for unit nonresponse and the most common combinations of item nonresponse among production workers, hours, and payroll were tested. The methods are: composite estimation (λ=0.5, and λ=1.0); proportional-to-cell average for month t; and monthly trend (same procedure as for employment except different characteristics like production workers are substituted). The test was performed on current CES respondents for March 1994 to March 1995. In this simulation, imputation cells were at the 2-digit industry by 4 size classes as defined above.

Composite estimation--The imputed value of production workers (PW) is

$$y_{PWit} = \lambda\left[y_{AEit}\frac{y_{PWi(t-1)}}{y_{AEi(t-1)}}\right]+ (1-\lambda)y_{PWi(t-1)}\left[\frac{\sum w_{jt}^{sel} y_{AEjt}}{\sum w_{jt}^{sel} y_{AEj(t-1)}}\right],$$

where y_{AEit}, $y_{AEi(t-1)}$, and $y_{PWi(t-1)}$ may be reported or imputed and $\sum$ is as defined above. In the first method tested, $\lambda=.5$. In the second method, $\lambda=1.0$. The imputed value of hours (H) is

$$y_{Hit} = y_{PWit}\frac{y_{Hi(t-1)}}{y_{PWi(t-1)}}\left[\frac{\left(\dfrac{\sum w_{jt}^{sel} y_{Hjt}}{\sum w_{jt}^{sel} y_{PWjt}}\right)}{\left(\dfrac{\sum w_{jt}^{sel} y_{Hj(t-1)}}{\sum w_{jt}^{sel} y_{PWj(t-1)}}\right)}\right].$$

Finally, the imputed value of payroll ($) is

$$y_{\$it} = y_{Hit}\frac{y_{\$i(t-1)}}{y_{Hi(t-1)}}\left[\frac{\left(\dfrac{\sum w_{jt}^{sel} y_{\$jt}}{\sum w_{jt}^{sel} y_{Hjt}}\right)}{\left(\dfrac{\sum w_{jt}^{sel} y_{\$j(t-1)}}{\sum w_{jt}^{sel} y_{Hj(t-1)}}\right)}\right].$$

Proportional-to-cell average--This is the current method.

$$y_{PWit} = AEit\left[\sum w_{jt}^{sel} y_{PWjt} \Big/ \sum w_{jt}^{sel} AE_{jt}\right]$$

$$y_{Hit} = y_{PWit}\left[\sum w_{jt}^{sel} y_{Hjt} \Big/ \sum w_{jt}^{sel} y_{PWjt}\right]$$

$$y_{\$it} = y_{Hit}\left[\sum w_{jt}^{sel} y_{\$jt} \Big/ \sum w_{jt}^{sel} y_{Hjt}\right]$$

In the first three methods, the initial month values were set at proportional-to-cell average; while for the fourth method, the initial values were set proportional-to-employment (e.g., total hours/total employment).

The details of the simulations on hours and earnings imputation is given in Grden (1997). No method was consistently best across major industry divisions and characteristics (e.g., average weekly hours, average weekly earnings) in terms of level and month-to-month change. Overall, the two composite methods performed more consistently for all industries, with the results differing only slightly between the two methods. Statistically, proportional-to-cell-average will yield larger variability from month-to-month at the microdata level, and monthly trend can yield inconsistent results (e.g., production workers greater than all employee). Between the two composite methods, the one where λ=1.0 consistently yields smaller average absolute errors at the microdata level in estimating production workers, and so based on this criteria, it is a better method. Additionally, this method is simpler than the one with λ=0.5. In the new design, therefore, the composite estimation method with λ=1.0 will be used. Technically, with λ=1.0, it is no longer a composite.

VI. Estimators

Employment--To begin the research on estimators, a multi-year simulation study was conducted using ES-202 data for the period April 1989-September 1994 from Iowa. In this study, ten random samples, according to the sample design, were selected and estimates were tabulated for four different estimators with some built-in nonresponse pattern. They are: 1) current link relative (Butani, Stamas, Brick, 1997), this estimator does not use sampling weights; 2) weighted link relative; 3) ratio estimator $\left[\frac{\hat{Y}}{\hat{X}}X\right]$ which is equivalent to generalized regression estimator (GRE) with no intercept; and 4) GRE with intercept. These estimators were evaluated in terms of: mean and maximum percentage benchmark revision for March 1991-94, and relative standard errors on the mean revisions; mean and maximum error in estimated month-to-month percent change; and mean and maximum revisions between preliminary and final estimates. While no estimator consistently performed the best across industries and time, the current link relative estimator produced biased results at certain time periods and for certain industries. Overall, the ratio estimator was more robust and easiest to implement.

At this point, employment estimator for state estimates was established as: $\sum w_{it}^{aty} * w_{it}^{BMF} * y_{it}$ where, w_{it}^{aty} is the atypical weight (selection weight adjusted for outliers), and w_t^{BMF} calculated at the state/model industry level and attached to each unit i within the state/model industry.

$$w_t^{BMF} = \frac{X}{\hat{X}} = \frac{\text{total universe employment of the noncertainty units in the model industry at benchmark month (t=0)}}{\sum w_i^{sel} x_i},$$

where, x_i is the benchmark employment of the ith in the sample $\sum$ is over all the non-certainty units that are in sample for model industry at time t. Since w_{it}^{BMF} is accounting for sampling variability at benchmark month (t=0), all certainty units receive a w_{it}^{BMF} =1.000.

Note: The use of w_{it}^{aty} instead of w_{it}^{sel} in the denominator leads to an underestimation bias. In the simulations performed for the 11 states, the mean percentage revision based on ten samples for each state were tabulated for March 1991, 92, 93, and 94. The results indicated a consistent underestimation (see table). As mentioned elsewhere, for CES, bias is more of a problem than variance because of large sample size.

Since the w_{it}^{BMF} are calculated without regard to MSA data, the use of ratio estimator gave large relative standard errors for MSA estimates. In order to produce reliable estimates for both industry (statewide) and MSA, the GRE without intercept and raked ratio estimator were considered. For very small domains, borth estimators are problematic. GRE can produce negative employment values, while raked ratio estimator generates estimates of zero employment for subdomains with no sample units. See Harter (1997) for estimators for small domains. For operational reasons, raked ratio estimator was selected for industry and MSA estimates. The raked ratio estimator utilizes the benchmark weight concept and adjusts estimates in not one, but two directions—MSAs and model industries both within a major industry division. These factors are calculated by iteratively adjusting benchmark weighted employment for MSAs by model industries to MSA major industry division total and to state model industries total. This iterative adjustment process continues until the marginal sample totals are within a tolerance limit of the fixed benchmark levels; currently, the tolerance level is 100 employees.

The raked ratio estimator improved the reliability of the MSA estimates without sacrificing the reliability of the industry estimates at the state level.

Mean % Benchmark Revisions (March) and Mean Relative RMSEs
Averaged Over 10 Samples, All CES Industries, and 4 Years

	Mean % Benchmark Revision			Rel Root Mean Squared Error		
	Raked Ratio	Ratio $X/\Sigma w_i^{sel} x_i$	Ratio $X/\Sigma w_i^{ATY} x_i$	Raked Ratio	Ratio $X/\Sigma w_i^{sel} x_i$	Ratio $X/\Sigma w_i^{ATY} x_i$
California	-.01	.03	.35	.38	.38	.46
Connecticut	.09	.09	.25	.58	.55	.51
Florida	.03	.07	.56	.74	.67	.77
Illinois	-.02	-.01	.25	.43	.39	.37
Iowa	.04	.00	.13	.66	.67	.58
Massachusetts	-.04	-.05	.11	.53	.52	.48
Michigan	.07	.06	.21	.46	.45	.44
New Jersey	-.06	-.02	.28	.71	.62	.57
New York	.14	.12	.31	.39	.36	.44
Pennsylvania	.09	.10	.46	.55	.52	.61
Texas	-.16	-.08	.53	.57	.52	.71

The estimator for state estimates, including MSAs and MSA by industry, is: $\sum w_{it}^{aty} * w_{it}^{BMF} * y_{it}$; where, w_t^{BMF} is the raked ratio estimator. For national estimates, it is essentially the same estimator. The only difference being that w_t^{BMF} is calculated at the national model industries level with no geography component; hence, the term ratio estimator. Since the sample size is quite large at higher levels of aggregation, the sum of the state estimates are virtually the same as independently produced national estimates. In the empirical study conducted on the simulations of the 50 states plus District of Columbia, the difference at the total private employment level was less than 0.05 percentage point. This difference is much smaller than the difference arising from performing independent seasonal adjustments to national and states data.

<u>Production Workers, Hours, and Earnings</u>—For these characteristics, independent research was conducted by West, Kratzke, and Grden (1997), Harter (1997), and

Grden (1997); they tested numerous estimators. The decision is based on Grden (1997) research that tested three estimators. They are: ratio estimator, weighted link-relative, and the current link and taper estimator. The same estimator as the one for employment is being adopted for them as well. Estimates of average weekly hours are derived by taking the ratio of total hours to total production workers; similarly average hourly earnings is the ratio of total payroll to the total hours.

Three Different Estimators--In the new design, three sets of estimates will be produced for each characteristic at the national, state, and MSA level. They are: the simple unbiased , $\sum w_{it}^{sel} * y_{it}$; raked ratio (state) or ratio (national) with selection weights, $\sum w_{it}^{sel} * w_{it}^{BMF} * y_{it}$; and raked ratio or ratio with atypical adjusted weights, $\sum w_{it}^{aty} * w_{it}^{BMF} * y_{it}$, the official estimates. By comparing the first two estimates, the effect of the benchmark factors can be gauged. Also, by comparing the latter two estimates, the effect of atypical adjustments can be gauged.

VII. Estimation for Births/Deaths

Direct estimation of births in establishment surveys has always been problematic due to lack of a comprehensive and timely sampling frame [Stamas, Goldenberg, Cantor, 1997]. In theory, direct measurement of deaths or out-of-businesses should be possible from the sample; in practice, however, it is very hard to make a distinction between nonrespondent, out-of-business, and ownership changes in CES because of its vast sample size and short data collection period. Estimation procedures for this major and complex topic are described in Butani, Kratzke, and Shierholz (1997) and in Getz and Kropf (1997).

VIII. Summary and Future Issues

Of the many steps involved in estimation, so far the focus has been primarily on edits, adjustments for outliers, imputation, estimators, and birth/death issues; work has also begun on variance estimation. Work in the area of composite estimation, if necessary; seasonal adjustment; benchmarking; sample rotation; and further improvements will begin once the data from the new design become available.