Comparing Current Population Survey Estimates Computed Using Different Composite Estimators

Janice Lent, Stephen Miller, Martha Duff, and Patrick J. Cantwell
Janice Lent, Bureau of Labor Statistics, 2 Mass. Ave. NE, Room 4915, Washington, DC 20212

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

KEY WORDS: AK composite estimator, generalized replication

1. Introduction. In this paper, we use a modeling approach to estimate the effect of a new composite estimation method on estimates from the Current Population Survey (CPS). The CPS, a household survey sponsored by the Bureau of Labor Statistics (BLS) and conducted by the Census Bureau, is the primary source of information on U.S. unemployment levels and rates. The survey has undergone a series of changes during the 1990's. In 1994, computer assisted data collection was introduced along with a redesigned questionnaire. Two years later--due to budgetary changes--a substantial sample cut took place, requiring modest changes in the sample design.

January 1998 saw the introduction of "composite weights" into the CPS. The new micro-data weights incorporate the effect of a composite estimation step which previously had to be performed at an aggregate level. In the next section, we briefly describe the method and compare its properties with those of the "macro-level" composite estimator it replaced. In the remaining sections, we present a method of computing additive and multiplicative adjustments which may be applied to allow comparison of CPS estimates computed by the old and new composite estimation methods. The adjustment factors given here are intended for use by economic analysts who must compare labor force estimates computed before and after the implementation of composite weighting in the CPS.

2. CPS Composite Estimation. The CPS estimation process involves computing an adjusted sampling weight for each person residing in a sample household. The original sampling weight, representing the inverse of the household selection probability, is modified through a series of ratio adjustments to be consistent with population controls derived from the decennial census and other administrative data. Each monthly sample consists of eight panels or "rotation groups," which enter and leave the sample according to a "4-8-4" rotation pattern (i.e., four months in, eight months out, followed by four months in). Due to the composite estimation step that follows the CPS weighting adjustments, the final adjustments to population controls are performed separately for each rotation--or month-in-sample--group.

For data months prior to January 1998, the CPS *AK* composite estimator was applied directly for each estimated total (e.g., number of persons unemployed), according to the following formula.

$$Y_t'' = (1-K)Y_t + K\left(Y_{t-1}'' + \Delta_t\right) + Ab_t\,,$$

where

$x_{t,i}$ is the estimator for month t, based on a sum of weights for sample persons completing their ith monthly interview in month t;

$$Y_t = \frac{1}{8}\sum_{i=1}^{8} x_{t,i}\,;$$

$$\Delta_t = \frac{1}{6}\sum_{i\in S}\left(x_{t,i} - x_{t-1,i-1}\right);$$

$$b_t = \frac{1}{8}\left\{\sum_{i\notin S} x_{t,i} - \frac{1}{3}\sum_{i\in S} x_{t,i}\right\};$$

$S = \{2,3,4,6,7,8\}$;

$K = 0.4$; and

$A = 0.2$.

To ensure consistency, the *A* and *K* coefficients given above were used for all estimates, though they were selected to be optimal only for unemployment totals. (See Breau and Ernst 1983.) Composite estimates of ratios and means are computed as functions of composite estimates of totals.

The new "composite weighting" method, suggested by Fuller (1990), involves two main steps:

1. Compute composite estimates for the main labor force (LF) categories, classified by important demographic characteristics. The LF categories are (a) employed, (b) unemployed, and (c) not in the labor force.
2. Through a series of ratio adjustments, adjust the micro data weights to agree with these composite estimates.

The new technique provides increased operational simplicity for micro-data users: They may now compute composite estimates simply by adding composite weights for the current month, whereas data from several months are needed to compute composite estimates using the old estimator.

Further, the new estimator allows optimization of compositing coefficients for different labor force categories. We selected coefficients for unemployment and employment, the two categories of greatest interest to CPS data users. Our prior research involved testing three different sets of coefficients: the old coefficients (given above), coefficients selected to reduce variance, and coefficients selected to reduce both variance and "month-in-sample" bias. For reasons explained in Lent et al. (1996), BLS decided to implement composite weighting with parameters optimized with respect to variance. The new parameters for unemployment are K=0.4 and A=0.3. For employment, the values K=0.7 and A=0.4 were selected. To ensure additivity of estimates, the estimates of the numbers of persons not in the labor force (computed in step (1) above) are calculated as residuals from the population controls. That is, the sum of the estimates for employed and unemployed is subtracted from the population estimate to obtain an estimate of the number of persons not in the labor force. For a description of the ratio adjustments performed in the second step, see Lent et al. (1996).

The effect of the new estimation procedure on CPS estimates may be summarized as follows:

(a) A slight increase in the reliability of estimates for some major labor force categories, especially those related to employment. Estimates of employment totals for consecutive months tend to be more strongly correlated than the corresponding unemployment estimates; the new compositing parameters improve reliability by exploiting this stronger correlation.

(b) A slight decrease in the reliability of some estimates not directly related to labor force status, e.g., educational attainment. These estimates are no longer directly composited.

(c) Decreases in estimates of employment totals and less pronounced increases in estimates of unemployment totals. Our previous research results indicated that these changes would not substantially affect estimates of national unemployment rates.

3. Method of Computing Adjustment Factors. Using both additive and multiplicative models, we can calculate adjustment factors that allow comparison of labor force estimates computed using the old and new composite estimators. In this section, we describe the models and the data used in the research.

Data used. For the months of May 1996 through December 1997, we have estimates for 2489 characteristics computed using both the new and old composite estimators. Thus we have 20 observations for each of the two methods--a total of 40 for each characteristic of interest. The CPS variance estimation system employs a generalized replication technique to compute variance estimates for CPS data. For each of the 2489 estimates, we therefore have 160 replicate estimates from which to estimate variances and covariances for the 40 observations.

Models. Let y_{it} be a CPS estimator of a particular characteristic in month t, where i=1 denotes the old composite estimator and i=2 denotes the new composite estimator. The additive factor model is given by

$$y_{it} = m_t + l_i + e_{it}, \tag{1}$$

where

m_t = true mean for month t;
l_i = effect of composite estimator i; and
e_{it} = sampling error for estimator i and month t.

Similarly, the multiplicative model is

$$\ln y_{it} = \ln m_t^* + \ln l_i^* + u_{it}, \tag{2}$$

where the parameters are defined by analogy with those in (1). For the additive model, we assume that $l_2 = 0$; for the multiplicative model, we likewise assume that $l_2^* = 1$. Thus our estimates of l_1 and l_1^* may be used to adjust CPS composite estimates computed under the old procedure, making them comparable--assuming the model holds--to those computed by the new method. We assume that the sampling error terms in the models are approximately normally distributed with zero means.

Using the full sample estimate and the 160 replicate estimates, we can estimate l_1 and l_1^* for each characteristic by generalized least squares. The estimation process for the additive model proceeds as follows: Let $\boldsymbol{y}_1$ be the vector of 20 monthly observations (from May 1996 through December 1997) computed using the old composite estimator. Let $\boldsymbol{y}_2$ be the corresponding vector computed using the new composite estimator, and let $\boldsymbol{y}' = (\boldsymbol{y}_1', \boldsymbol{y}_2')$. Let $\boldsymbol{X}$ denote the 40x21 design matrix associated with the model (1). Let $\boldsymbol{b}$ be the 21x1 vector of free parameters, i.e., the 20 monthly means m_t and the effect l_1 of the old composite estimator. Then

$$\text{(3)} \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e} \quad ,$$

where $\boldsymbol{e} \sim MVN_{40}(\boldsymbol{0}, \boldsymbol{V})$, and $\boldsymbol{V}$ is the 40x40 covariance matrix of the vector of (old and new) composite estimates $\boldsymbol{y}$. We construct an estimator $\hat{\boldsymbol{V}}$ of the covariance matrix $\boldsymbol{V}$ using the 160 replicate estimates.

The vector of parameters is then estimated by the method of estimated generalized least squares:

$$\text{(4)} \qquad \hat{\boldsymbol{b}} = \left(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y} \, ,$$

where the estimated covariance matrix of the parameter estimates is given by

$$\hat{\boldsymbol{V}}(\hat{\boldsymbol{b}}) = \left(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1} .$$

Under our assumptions, $\hat{\boldsymbol{b}}$ is a consistent estimator of $\boldsymbol{b}$ and is approximately normally distributed. In addition, $\hat{\boldsymbol{V}}(\hat{\boldsymbol{b}})$ is a consistent estimator of the covariance matrix of $\hat{\boldsymbol{b}}$. We also construct the goodness of fit statistic

$$G^2 = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}\right)' \hat{\boldsymbol{V}}^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}\right)$$

which is approximately distributed as a chi-square random variable with 19 degrees of freedom if the model is true. Comparing the statistic G^2 to the distribution of a chi-square random variable with 19 degrees of freedom yields a test of model adequacy. Rejection of the model would indicate differences between the two methods that are not constant across time.

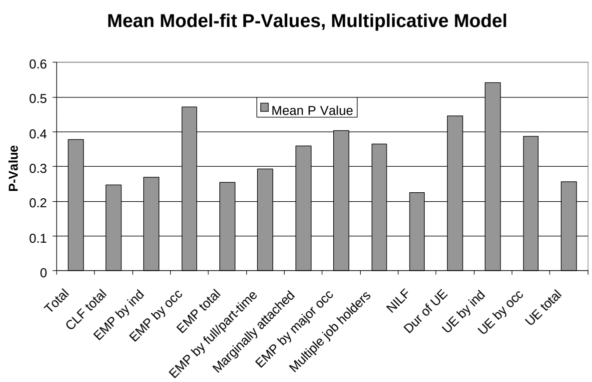
3. Results. The tables below give the resulting adjustment factors for major labor force estimates (Table 1) and for several estimates significantly affected by the new composite estimator (Table 2). As expected, we must adjust the old estimates for unemployment levels upward to render them comparable to the new estimates; by contrast, the old estimates for employment must be adjusted downward. The effect on the major labor force estimates, however, appears relatively slight: all the multiplicative factors in Table 1 represent changes of less than 0.5%.

Reliability of Estimates Shown in Table 2. For the estimates appearing in Table 2, the effect of the new composite estimator is more pronounced than for the major labor force categories. The new estimator is known to be more appropriate for some of these estimates. The estimated number of persons in a particular “duration of unemployment” category in a given month is necessarily weakly--or even negatively--correlated with the same estimate for the previous month, given a continuing sample. A person unemployed for zero to four weeks in one month, for example, cannot normally remain in the same category in the following month. Under the old composite estimation method, these estimates were directly composited in order to ensure additivity. Weakly or negatively correlated estimates, however, do not generally benefit from the application of the *AK* estimator. Implementation of the new composite estimator therefore improved the reliability of most of the duration of unemployment estimates.

The remaining groups of estimates shown in Table 2, however, suffered small decreases in reliability with the change in the composite estimator. These estimates are no longer directly composited under the new method, and characteristics having strong month-to-month correlations tend to benefit from direct compositing. The coefficient of variation (CV) for the estimated number of employed persons holding multiple jobs, for example, went from about 1.86% to roughly 1.93%--one of the largest increases in CV due to the new estimator. The effect of the new method on many estimates of employment by educational attainment was insignificant, but these estimates also suffered a decrease in reliability. For estimates of employment by hours worked, the reliability of estimates produced under the two methods proved about the same; the CV’s of some estimates in this category increased slightly while the CV’s of others decreased.

Applying the Adjustment Factors. The four plots on the next page illustrate the effect of using the adjustment factors for two key national rate estimates produced from CPS data: the unemployment rate (the top two figures) and the labor force participation rate (the bottom two figures). The figures on the left-hand side show the old composite estimates for the 20-month period from May 1996 to December 1997, with 95% confidence bounds. Also plotted are the additively and multiplicatively adjusted series. The two lines representing the adjusted data follow one another so closely that they are indistinguishable in all four plots. The right-hand plots show the composite estimates computed under the new method along with the adjusted estimates. The adjustment appears to work well for both series: in both of the right-hand plots, the adjusted estimates appear very close to the new estimates. For the unemployment rates, the adjustment may not be important for economic analysis: the adjusted and unadjusted estimates are about the same. This is to be expected, since the change in coefficients in the composite estimator for unemployment was very minor. The labor force participation rate, however, is more closely correlated with employment level than with unemployment level. The adjusted estimates for this rate are noticeably below the unadjusted estimates, so the adjustment factors may prove an important tool in the analysis of this series.

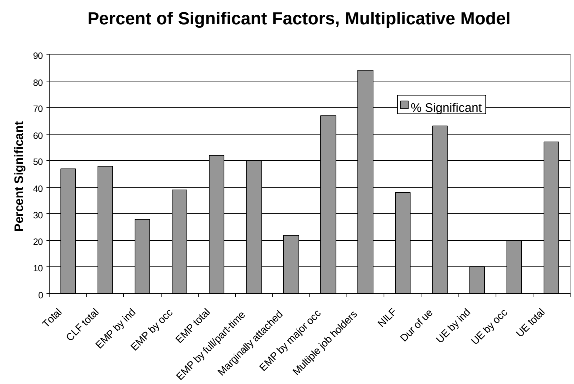
Goodness-of-Fit Tests. The preceding charts indicate that the models fit the data reasonably well. The goodness-of-fit statistics we computed for the additive and multiplicative models were very similar, so preference for additive or multiplicative factors will be a matter of user convenience. For each



estimate, we calculated the "model-fit p-value" as the probability of drawing a value larger than the model's goodness-of-fit statistic from a chi-square distribution with 19 degrees of freedom. The bar chart above shows the average model-fit p-values, for the multiplicative model, for some types of estimated totals. Most of the averages lie between 0.2 and 0.5, indicating that the multiplicative models fit sufficiently well to justify use of the adjustment factors. For all estimates studied, the model-fit p-values for the additive model were very close to those for the multiplicative model. These results indicate that the effect of the new composite estimator is approximately constant across the 20-month period examined.

Significance of Factors. The adjustment factors, however, need not be used for estimates not significantly affected by the change in composite estimators. Overall, about 55% of the adjustment factors we calculated indicated an insignificant difference; that is, the additive factors did not differ significantly from 0 nor the multiplicative factors from 1. The bar chart below indicates which types of estimates will likely benefit most from adjustment. For estimates related to multiple job holders, over eighty percent of the adjustment factors were significant**--**the largest percentage among the types of estimates we examined. As expected, a high percentage of estimates related to employment, especially employment by occupational category, had significant adjustment factors. More than half of the estimates of unemployment**--**totals and "duration of unemployment" categories**--**were also significantly affected.

It should be noted, however, that the adjustment factors were computed using 20 months of CPS data. A significant adjustment factor does not necessarily indicate a need for adjustment in economic analysis.



The factors for the national unemployment rate, for example, are statistically significant but may make little difference in the analysis of the series.

References

Breau, P. and Ernst, L. (1983). "Alternative Estimators to the Current Composite Estimator." *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 397-402.

Fuller, W. (1990). "Analysis of Repeated Surveys." *Survey Methodology,* **16**, **2**, pp. 167-180.

Lent, J., Miller, S., and Cantwell, P. (1996). "Effect of Composite Weights on Some Estimates from the Current Population Survey." *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 130-139.

Note: This paper reports results of research undertaken by staff of the Bureau of Labor Statistics and the Bureau of the Census. The views expressed are attributable to the authors and do not constituted policy of these agencies.

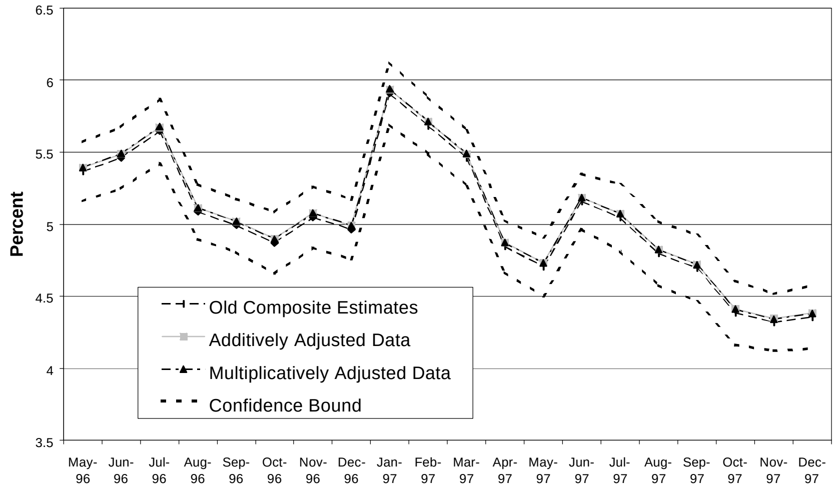
Table 1. Adjustment Factors for Major Labor Force Levels
(Estimates of level, and the corresponding additive factors, are given in thousands.)

	Average Estimate (New)	Average Estimate (Old)	Additive Factor	Multiplicative Factor
Unemployment				
Total	6,861	6,837	26	1.0037
Men 20+	2,887	2,879	7	1.0022
Women 20+	2,674	2,665	10	1.0037
Total 16-19	1,300	1,294	7	1.0047
White	4,936	4,919	16	1.0033
Black	1,583	1,578	6	1.0038
Employment				
Total	128,606	128,877	-285	0.9978
Men 20+	65,871	65,961	-106	0.9984
Women 20+	56,080	56,192	-95	0.9983
Total 16-19	6,655	6,725	-72	0.9888
White	109,179	109,398	-230	0.9979
Black	13,815	13,861	-50	0.9963

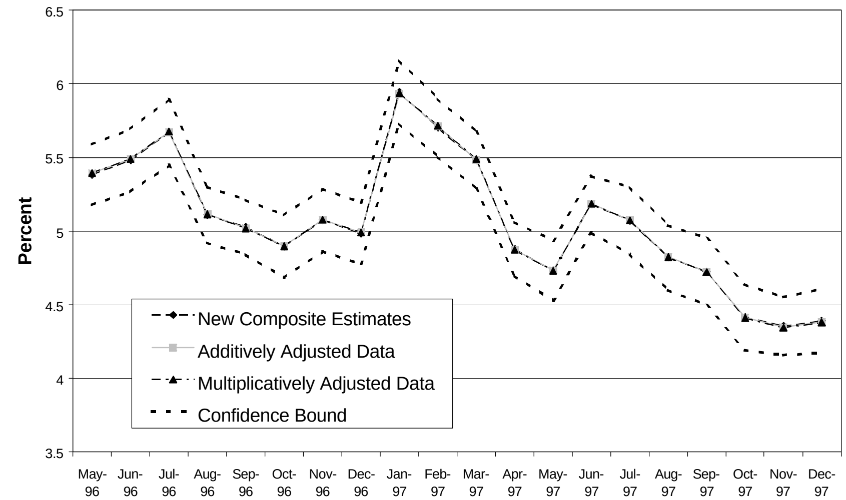
Table 2. Adjustment Factors for Some Estimates Significantly Affected by the New Method
(Estimates of level, and the corresponding additive factors, are given in thousands.)

	Average Estimate (New)	Average Estimate (Old)	Additive Factor	Multiplicative Factor
Employed, at Work				
Total	122,878	123,114	-248	0.9980
40 hours	43,752	44,568	-812	0.9819
41-48 hours	14,676	14,383	291	1.0203
49-59 hours	14,648	14,440	203	1.0140
MEAN hours	39.61	39.51	0.10	1.0026
MEAN hours nonagricultural	39.53	39.43	0.10	1.0026
Employed, at Multiple Jobs				
Total	8,163	7,953	204	1.0257
White	7,105	6,935	164	1.0237
Black	770	740	28	1.0392
Male ages 20+	4,225	4,100	126	1.0311
Female, ages 20+	3,583	3,505	79	1.0223
Full time first job, part time second job	4,559	4,452	107	1.0241
Duration of Unemployment				
0-4 weeks	2,691	2,583	108	1.0403
5-14 weeks	2,120	2,155	-29	0.9860
Male, 5-14 weeks	1,103	1,114	-8	0.9923
Female, 5-14 weeks	1,017	1,041	-23	0.9773
15+ weeks	2,050	2,099	-48	0.9782
MEAN (weeks)	15.95	16.10	-0.09	0.9943
Employed by Educational Attainment				
No high school diploma	16,566	16,663	-113	0.9932
High school, no college	41,739	41,735	16	1.0004
College, no degree	26,286	26,382	-104	0.9961
B.A.	22,744	22,775	-25	0.9989

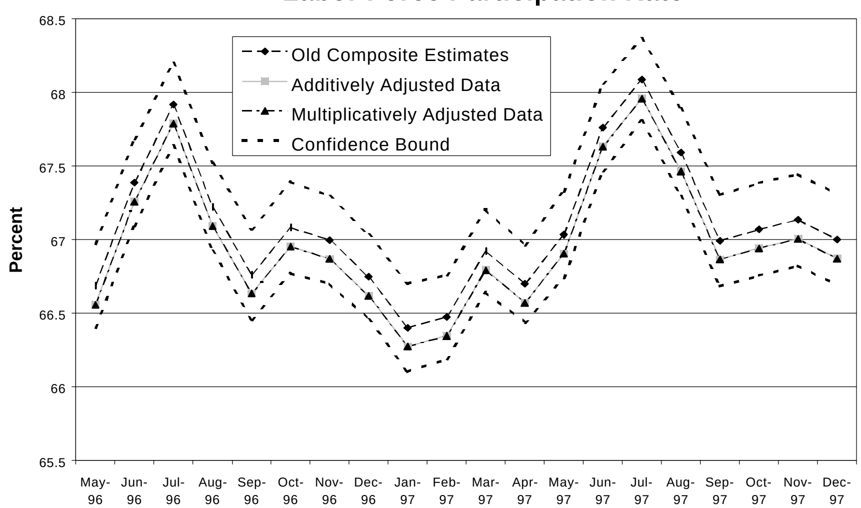
Old and Adjusted Old Unemployment Rate



New and Adjusted Old Unemployment Rate



Old and Adjusted Old Labor Force Participation Rate



New and Adjusted Old Labor Force Participation Rate