SHORT-TERM CHANGES TO THE CPS COMPOSITE ESTIMATOR IN JANUARY 1994

Patrick J. Cantwell, Lawrence R. Ernst, Bureau of the Census
Patrick J. Cantwell, Bureau of the Census, Washington, DC 20233*

KEY WORDS: New questionnaire, variance, month-in-sample bias.

1. Introduction

The Current Population Survey (CPS), sponsored by the Bureau of Labor Statistics (BLS) and conducted by the Census Bureau, measures labor force status in the United States. Presently, about 9% of interviews are assigned to one of two centralized computer assisted telephone interviewing (CATI) facilities. This portion may increase in the 1990s. Until the end of 1993, the remaining 91% are conducted with a paper-and-pencil instrument in the field regions.

In January 1994, BLS and the Census Bureau plan to eliminate all paper-and-pencil questionnaires in favor of computer assisted personal or telephone interviewing (CAPI) on laptop computers using a revised questionnaire. The new instrument will elicit better responses and try to improve measurement of labor force concepts. The change in procedures may effect measurements of labor force status in various ways.

To help prepare for this event, the Census Bureau has been interviewing an experimental set of households, called the CATI/CAPI Overlap (CCO) Panel, since the middle of 1992. This panel will operate through December 1993. Except for necessary modifications, the CCO's instrument is the same as will be used in regular CPS in January 1994, and subsequently. Further, the CCO's 15,000 sample households (per month) will be interviewed in the same rotation pattern as CPS. (However, the full rotation design will not be completely phased in until the last three months of the panel's duration.) It should be noted that, during its duration, the CCO Panel is independent of the regular CPS; none of the data from the CCO are used in estimates or publications, nor will any respondents continue in sample in 1994.

The final stage in estimation for a nonseasonally adjusted estimator in CPS is the composite estimator, described in Section 2. Because these one-time changes (which will occur in January 1994) will affect the composite estimator, temporary adjustments to this estimator are being investigated. The currently used composite estimator combines data from the current and the previous months. In January 1994, such an estimator would mix the new and old procedures, creating certain operational difficulties and affecting the estimates. We studied two alternatives which do not mix the old and new procedures: (i) not compositing in January 1994 (but resuming in February 1994), and (ii) compositing with survey data from the CCO.

As we will discuss, this last alternative was discarded because of large variances, especially when measuring month-to-month change. We detected little difference between the current composite estimator and the first alternative based on the resulting variances and biases as projected for 1994. Therefore, weighing operational aspects, it has been recommended that we start a new series in 1994: do not composite in January, but resume compositing in February and later months.

Because of space limitations here, some details and results of the research have been omitted. They are included in the full-length version of the paper which can be obtained from the authors.

2. Composite Estimation in CPS

In the CPS, sample households are interviewed for four consecutive months, rotated out of sample for the next eight months, and finally returned to the sample for four more months. The overlap in households between consecutive months--75% of the housing units--reduces the variance of estimates of month-to-month change due to the positive correlation between estimates from the same rotation group one or more months apart (U.S. Bureau of the Census, Technical Paper 40, 1978).

To take advantage of the rotation overlap, the CPS uses a composite estimator. After the data are collected, the basic weights are run through several stages of ratio adjustment. We call the estimator incorporating all these adjustments the ratio estimator.

For a specified characteristic, let $x_{h,i}$ be the estimator of total for month h arising from the rotation group which is interviewed for the ith time in month h. In each month, there are eight such estimators. The ratio estimator takes the form $Y_h = (1/8)\ \Sigma_i\, x_{h,i}$.

The AK composite estimator is defined as

$$Y_h' = (1\text{-}K)\, Y_h + K\,(Y_{h-1}' + \Delta_h) + A\,\beta, \quad \text{where}$$

$$\Delta_h = (1/6) \{ \Sigma x_{h,i} - \Sigma x_{h-1,i-1} \},$$

with i = 2,3,4,6,7,8 (the sums are over the rotation groups which are common to the two months); and

$$\beta = (1/8) \{ \Sigma x_{h,i} - (1/3) \Sigma x_{h,j} \},$$

with i = 1,5, and j = 2,3,4,6,7,8. The Census Bureau adopted this estimator with A = .2 and K = .4 in the mid-1980s.

3. Which Estimator in January 1994?

Section 1 described several one-time changes in the CPS to be implemented in January 1994. During the switch to the new questionnaire, we hope to disrupt the time series of estimators as little as possible, bearing in mind that there may be a "jump" in January 1994, due to the new procedures.

Three estimators are investigated in this section. They are derived as follows:

(U1) Composite as usual in 1/94 and in succeeding months, with A = .2 and K = .4.

(U2) Start with the ratio estimator in 1/94. Composite 2/94 with 1/94 (A = .2, K = .4); composite 3/94 with 2/94; etc.

(U3) Composite 1/94 with the (independent) 12/93 composite estimator obtained from the CCO Panel, with A = .2 and K = .4; composite 2/94 with 1/94 (A = .2, K = .4); composite 3/94 with 2/94; etc. To estimate change in 1/94, subtract the 12/93 CCO estimator from this 1/94 estimator.

When evaluating estimators for 1994, several factors may have an effect.

▸ First, the old and new questionnaires are not completely compatible. When compositing January 1994 with previous months to measure a specific characteristic, we need comparable data items. Some questions will be introduced in the new instrument, and the choice of responses will be changed in others. This question of compatibility is important if we use a composite estimator which combines data from 1993 with 1994, and when we consider month-to-month change in January 1994.

▸ Second, there is evidence to imply that, with the new procedures, the patterns of month-in-sample bias will change. Bailar (1975) defines and discusses the concept of month-in-sample effects caused by panel conditioning in the CPS. Briefly, for any given month and characteristic to be estimated, the expected values of the eight rotation group estimators are generally not equal, but reflect the number of previous interviews or other influences. Using the notation in Section 2, the bias *index* for the *i*th month in sample can be defined as $E(x_{h,i}) / E(\Sigma_j x_{h,j}/8)$, so that an index greater than 1 implies an overestimate in that month relative to all eight months.

Recent analysis by Adams (1991) on CPS responses from 1980 to 1987 provides estimates of the bias indices for unemployed (UE) and civilian labor force (CLF) under the current procedures. Yet computations from the CATI (Computer Assisted Telephone Interviewing) phase-in study (Shoemaker 1993) yield significantly different indices for UE. Because all interviewing will be computer assisted in 1994, the new indices may well be closer to those obtained under CATI.

In addition to these two sets of bias indices, we obtained a third set using the following motivation. Before 1970, certain questions were asked of discouraged workers in months in sample 1 and 5 only. It is thought that these questions tend to increase the level of UE slightly, and thus affect the month-in-sample bias factors. Since 1970, the questions have been asked only in months 4 and 8. However, in the new instrument, they will be asked in all eight months in sample.

We'd like to estimate the effect of these questions on month-in-sample bias, and to project what the indices might look like after 1/94, when the questions will be asked in each month. We assume a bias factor f_i applies to month in sample i when the discouraged worker questions are *not* asked, and multiply by an effect d when the questions are asked. Bailar (1975) provides estimates of the bias indices (f_i in some months, $f_i \times d$ in others) before and after 1970, when the questions are asked in different months. This allows us to estimate d and each f_i, and thereby derive the indices (which we call "DWQ indices," for discouraged worker questions). Table 1 displays for the eight months in sample the current, CATI, and DWQ bias indices for UE.

▸ Third, as usual, 75% of sample households from the regular CPS in December 1993 will continue in sample in January 1994. An estimator such as U1 would take advantage of the correlated estimates from common rotation groups. On the other hand, the experimental CATI/CAPI Overlap (CCO) Panel, operating through December 1993, will have no households in sample in 1994.

To make a decision for January 1994, managers at the Bureaus of Labor Statistics and the Census considered three aspects of the competing estimators: operational circumstances, the variances of the competing estimators, and their "biases." In our work, we actually measured, instead of bias (which is unknown), *the deviation from the expected value of the steady-state composite estimator* (that is, the state of the usual composite estimator reached after new effects are fully phased in).

In January, there may be a shift in the value of labor force characteristics due to the new procedures. Because the actual level will never be known, our hope is to suffer this jump at once (January), and then proceed rapidly to the "steady-state." We want an estimator to reach the bias level of the long-run composite estimator as quickly as possible. (This use of "bias" or "deviation" should not be confused with the concept of month-in-sample bias.) Note, however, that there *will* be a one-time mix of month-in-sample bias effects for respondents early in 1994. In January, for example, six rotation groups will have been interviewed in December with the old instrument and then in January with the new one.

Before displaying computational results, let us briefly examine the three estimators defined earlier. U1 continues to composite as usual, combining data from 1993 with January 1994 and later months. As it takes advantage of the correlations among common rotation groups, we expect it to yield the lowest variances in most cases. Nevertheless, combining data from different questionnaires can be operationally complex. In addition, there would be different month-in-sample bias effects--due to the different procedures--for several months into 1994. Even if there is no "real" change during these months, this could lead to an estimator of month-to-month change with nonzero expectation due entirely to the different bias effects.

The second estimator, U2, is operationally simple. By starting 1994 with the ratio estimator, and subsequently compositing only with earlier months in 1994, there is a clean break with the bias effects from the old system. Still, by ignoring rotation groups which were in sample in late 1993, we lose the potential variance reduction. Moreover, for all three sets of bias patterns we consider, the new bias effects will take several months to reach those of the steady-state composite estimator.

The third estimator, U3, by compositing 1/94 with the CCO Panel, uses the same instrument before and after 1/94. Thus, there is no problem with the compatibility of the questionnaires. In addition, the steady-state bias effects should be reached almost immediately. Unfortunately, U3 suffers from two serious problems: (i) the CCO Panel has about one-fourth as many households as regular CPS, and (ii) there is no overlap between the CCO in 1993 and CPS in 1994. The result is a serious increase in the relevant variances. Indeed, its variance for January 1994 is about 35% (50%) greater than that of the steady-state composite when estimating monthly level of UE (CLF), and about 4 (7) times greater when estimating month-to-month change (Cantwell 1992). In light of these results, we continued the investigation *comparing only U1 and U2.*

4. Variances and Mean Squared Deviations

In our analyses, we consider the total number of unemployed people (UE). We repeated the analyses for the number in the civilian labor force (CLF), but have omitted the results here, due to space limitations. The month-in-sample bias indices for UE used in this study are given in Table 1. For months before 1/94, we used the current indices; for 1/94 and subsequent months, we used the three patterns in the table.

To evaluate U1 and U2 in the first six months of 1994, we computed their variances, and their deviations from the expected value of the steady-state composite estimator. In this section, the results are computed and presented *before the application of the CPS seasonal adjustment*, the only stage of adjustment after composite estimation. Using the formulae in Cantwell (1990), we assumed that the variance of $x_{h,i}$ (the estimator from the rotation group interviewed for the *i*th time in month *h*) is constant for all *h* and *i*. Where necessary, modifications for changing variances, correlations, and bias effects--as described in Section 3--were made to the formulae.

Table 2 displays the variances and the mean squared deviations (MSD--variance plus squared deviation), each divided by the variance of the steady-state composite estimator. The first two columns portray monthly level, the last two month-to-month change. For ease of comparison, in each cell, the value for U1 ("continue compositing as usual") is placed directly above that for U2 (ratio estimator in 1/94).

As CPS introduces a new questionnaire and a new mode of interviewing in January 1994, there may be a "jump" in the level of UE or CLF from 12/93 to 1/94 due to the new procedures beyond that due merely to random change. Any such shift can affect the bias or variance of estimators. To get a broad view of how this shift would affect the estimators, we investigated increases and decreases of 0%, 5%, 10%, and 15% in the level of UE, and, for CLF, increases and decreases of 0%, 1%, 3%, and 5%. No further changes in the levels were assumed for later months in 1994.

Table 2 provides a comparison between U1 and U2 under the DWQ bias effects, when estimating *UE with a 10% increase* from 12/93 to 1/94. One can see that, in January 1994, U1 realizes an increase in variance of monthly level of 1.5% compared to the steady-state composite. For U2, this increase is 6.7%--greater than U1, but small nonetheless. In both cases, the increase tapers off quickly over the subsequent months. If the DWQ biases are accurate, the mean squared deviations (MSDs) for monthly level in 1/94 are more than double that of the steady-state composite estimator, and U1 again is slightly better than U2.

The two estimators fare comparably when measuring month-to-month change, with U1 doing slightly better. The variances reach the steady state very soon. (The steady-state change from month to month is assumed to be 0.) For both estimators, however, the deviations from the steady-state composite are quite large in January, reflecting the 10% jump and the change in biases factors. Still, for the six months, the values are similar for U1 and U2.

From our computations with other sets of parameters on UE and CLF, several trends are worth noting. Although the variances and MSDs for U2 are generally larger than those for U1, the difference is usually not serious. Also, regardless of the size of the variance or MSD in January, the values in February are typically much closer to the steady-state composite. By March or April, the values are generally only 1% or 2% away. If we speculate that there will likely be a more serious increase in other nonsampling errors through this transitional period, the increases due to the estimators may be minor by comparison.

Applying other shifts in level did not seriously affect the variance comparison between U1 and U2. We also examined MSDs using the same biases in 1994 that the current CPS instrument experiences. The monthly-level deviations are close to 0 in this case. The month-to-month change deviations for 1/94 are only slightly smaller here, however, because of the strong influence of the shift in level.

5. Expected Values of the Estimators

In this section, we try to evaluate the expected values of the competing estimators *and the resulting unemployment rates* from 12/93 through 6/94, assuming a designated "real change" in the characteristic from 12/93 to 1/94, but no further real change in subsequent months. To start, it is necessary to specify what is meant by real change, since there are several levels of adjustment performed on the survey data.

The estimate released by the Bureau of Labor Statistics is seasonally adjusted (SAE) to partially compensate for the explainable variability among the twelve months of the year. In this analysis, we considered "real change" to be a difference in the value of the characteristic after the seasonal adjustment. Because the CPS seasonal adjustment is quite complex, we used a rough approximation. For the *m*th month of the year, we averaged SAE_m for UE for that month in 1990, 1991, and 1992, and did the same for the ratio estimates, RE_m. Then $r_m = SAE_m / RE_m$ provides a rough factor for converting UE from the ratio to the seasonally adjusted estimate in month *m*. Conversion factors for CLF are obtained the same way. Finally, with CE_m representing the composite estimate in month *m*, the ratios SAE_m / CE_m (one set each for UE and CLF) provide analogous factors for converting from composite to seasonally adjusted estimates in month *m*.

For each of UE and CLF, we started with levels in 12/93 which were already seasonally adjusted. Here, UE was selected to be 7.0% of CLF. Then each of these seasonally adjusted levels (12/93) was increased or decreased a specific percentage (e.g., 5% or 10%) to study the effect of various changes possible in 1/94. The following months in 1994 were then assigned the same changed level (as 1/94) after seasonal adjustment. For each of these months, the underlying "expected value" of the RE was obtained by dividing the SAE by r_m. From these numbers, the expected values of U1 and U2 were then computed.

Not surprisingly, the expected values of U1 and U2 in 1994 fluctuate--even with no real change in the characteristics--because of the seasonal effects. To simulate the final estimates we would expect to see, we then converted these back to seasonally adjusted. Note that U1 and U2 are composite estimators (except for U1 in 1/94). Even in 1/94, the seasonal adjustment applied will likely be based on that currently in use for converting composite estimators. Therefore, we applied CE-to-SAE factors to U1 and U2 to observe the expected final estimates.

At this point--separately for the nonseasonally adjusted and seasonally adjusted estimates--we divided the expected values for UE by those for CLF to picture what the unemployment rates might look like under various conditions. The results are presented in Table 3 for increases in the level of UE of 5% and 10% under DWQ biases. (CLF was kept constant). In each case the seasonally adjusted UE rate in 12/93 is 7.0%.

Several observations can be made from the results in Table 3 and our analyses with other parameters and bias indices:

- As we consider each additional month, the seasonally adjusted UE rates seem to converge. This apparent convergence is not monotonic in the tables, perhaps because the conversion factors we used are random (subject to error), based on CPS data from 1990-92. Further, the factors for converting from RE to SAE are not quite the same as those for converting from CE to SAE.

- UE rates for U1 and U2 approach the same number. This is to be expected because, as more months of 1994 are included in the two estimators, they approach the common steady-state composite estimator.

- The DWQ bias factors exert a moderate affect on the limiting "expected" value of the UE rate. With any amount of increase in UE or CLF we tried, the limiting UE rates are about 2% below the "true value" (that is, the ratio of the expected levels of UE and CLF after seasonal adjustment). On the other hand, under CATI biases, the limiting UE ratios are virtually unbiased (to .01%) for the "true values," whatever the increases in the levels of UE or CLF.

- Under DWQ biases, the seasonally adjusted values of U1 appear to be slightly less in the early months of 1994 than those of U2. The converse holds under CATI biases.

From the results in Sections 4 and 5, there is little difference between U1 and U2 with respect to our statistical measures in January 1994, and even less in the following months. With these observations in mind, the decision between the two estimators was based mostly on operational or processing concerns, in particular, the compatibility of the new and old systems. It was recommended that we use U2, starting with a simple ratio estimator in January 1994. In subsequent months, the estimator would composite with sample responses back through January. As indicated earlier, additional results and tables are available in the full-length version of the paper.

ACKNOWLEDGMENTS

We thank Bob McIntyre and Greg Weyland for supplying estimates of labor force characteristics; Harland Shoemaker for providing month-in-sample bias factors; and Shail Butani, Larry Cahoon, Donna Kostanich, Elizabeth Huang, and Dave Hubble for offering helpful comments and suggestions.

REFERENCES

ADAMS, D. (1991). Memorandum for Documentation, "CPS Month-in-Sample (MIS) Bias Index Research," 10/21/91, U.S. Bureau of the Census, Washington, DC.

BAILAR, B. (1975). "The Effects of Rotation Group Bias on Estimates from Panel Surveys," *Journal of the American Statistical Association*, 70, 23-30.

CANTWELL, P. (1990). "Variance Formulae for Composite Estimators in Rotation Designs," *Survey Methodology, 16,* p. 153-163.

CANTWELL, P. (1992). Internal note comparing variances and MSEs of alternative estimators, 9/28/92, U.S. Bureau of the Census, Washington, DC.

SHOEMAKER, H. (1993). "Results From the Current Population Survey CATI Phase-In Project," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, to appear.

U.S. BUREAU OF THE CENSUS (1978). "The Current Population Survey: Design and Methodology," Technical Paper No. 40, Washington, DC: U.S. Government Printing Office (Department of Commerce).

* This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

Table 1. Bias Indices for Unemployed (UE) Used in Study of Competing Estimators

Month in Sample	1	2	3	4	5	6	7	8
Current Indices	1.072	1.001	.987	1.026	1.002	.963	.955	.996
CATI Indices	.961	1.102	1.052	1.001	.944	1.024	.968	.948
DWQ Indices	1.123	1.030	.993	.950	1.037	.987	.953	.927

Table 2. For *UE* with a *10%* Shift in Level, using *DWQ* Biases: Variances and Mean Squared Deviations (as Compared to the Steady-State Composite Estimator). Values for U1 Placed Above Those for U2.

Month	Monthly Level		Month-to-Month Change	
	Variance	Mean Squared Deviation	Variance	Mean Squared Deviation
1/94	1.0154 1.0673	2.1136 2.8033	1.0283 1.0592	22.33 24.66
2/94	1.0044 1.0246	1.1801 1.3023	1.0022 1.0028	1.3224 1.5090
3/94	1.0011 1.0087	1.0292 1.0531	1.0008 1.0014	1.0520 1.0824
4/94	1.0002 1.0025	1.0047 1.0096	1.0003 1.0012	1.0085 1.0142
5/94	1.0000 1.0004	1.0007 1.0015	1.0001 1.0007	1.0014 1.0028
6/94	1.0000 1.0000	1.0001 1.0002	1.0000 1.0001	1.0002 1.0005

Table 3. Under *DWQ* biases, "Expected" *UE Rates (as %)* for Various Increases in the Level of *UE*, with No Increase in CLF

		MONTH						
		12/93	1/94	2/94	3/94	4/94	5/94	6/94
Increase in UE: 5%; UE Rate: 7.0% → 7.35%								
BEFORE Seasonal Adjustment	U1	6.82	8.30	8.16	7.63	7.12	7.09	7.45
	U2	—	8.33	8.17	7.63	7.12	7.09	7.45
AFTER Seasonal Adjustment	U1	7.04	7.37	7.29	7.20	7.21	7.22	7.20
	U2	—	7.40	7.31	7.20	7.22	7.22	7.20
Increase in UE: 10%; UE Rate: 7.0% → 7.70%								
BEFORE Seasonal Adjustment	U1	6.82	8.70	8.55	7.99	7.45	7.43	7.81
	U2	—	8.73	8.56	8.00	7.46	7.43	7.81
AFTER Seasonal Adjustment	U1	7.04	7.72	7.64	7.54	7.56	7.56	7.54
	U2	—	7.75	7.65	7.55	7.56	7.56	7.54