

# FEDERAL GOVERNMENT SHUTDOWN: OPTIONS FOR CURRENT POPULATION SURVEY (CPS) DATA COLLECTION<sup>1</sup>

Shail Butani, Brian Kojetin, Bureau of Labor Statistics; Lawrence Cahoon, Bureau of the Census  
Shail Butani, 2 Massachusetts Ave., NE Room 4985, Washington, D.C. 20212

**KEY WORDS:** Recall bias, rotation group bias, noninterviews, composite estimation, and seasonal adjustment

## I. Introduction

The Current Population Survey (CPS) is a national labor force survey of over 50,000 households conducted monthly by the Bureau of the Census for the Bureau of Labor Statistics (BLS). In addition to being best known for the national monthly unemployment rate—one of the foremost closely watched economic indicators for the nation—the CPS is the source of thousands of labor force statistics at the national and sub-national level. Due to disagreements between Congress and the White House over FY1996 funding for several executive branch departments including the Departments of Labor and Commerce, both the BLS and Bureau of Census had their normal operations shutdown for a few days in November 1995 and again in December 1995, that extended into January 1996. These “Federal Government Shutdowns,” as they were commonly known, were unprecedented in duration, and their timing and length created extraordinary challenges for data collection activities pertaining to CPS.

In this paper, we discuss the various options that were considered by the officials at both agencies for the collection of November 1995 through January 1996 data and the estimation of labor force statistics from the data. In the next section, we give a brief overview of the CPS sample design, weighting, and estimation procedures. In Section III, we describe the circumstances surrounding each shutdown of these agencies and discuss the consequences of implementing different data collection options that were considered for the CPS. We report the results of research we did in examining the characteristics of early and late respondents in prior months. We also summarize the course of action taken for collecting data for each of the three months and the resulting response rates and unemployment rates. In Section IV, we discuss alternative estimation procedures that were considered for the December data and report on additional research conducted to examine the effect of using prior and future months data to improve estimates for

nonrespondents. Finally, we discuss what we learned about the impact of data collection procedures on CPS and our recommendations when faced with similar constraints.

## II. Overview of CPS Sample Design

The CPS sample is a multistage State-based probability sample of over 50,000 eligible households. Each sampled household is designated for interview for 4 consecutive months, taken out of the sample for the next 8 months, and is designated for interview again for the following 4 months. Under this 4-8-4 rotation, 75 percent of the sample is common from one month to the next and 50 percent for months one year apart. This rotation scheme reduces variance for estimates of monthly changes and over the year changes for the same month. Interviews are conducted beginning on Sunday during the week that includes the 19th of the month, with the reference period being the week containing the 12th. The first and fifth interviews are conducted by personal visits while those for the remaining months are usually done over the telephone either by the field interviewers from their homes or from one of three centralized facilities.

The initial weight for the CPS reflects the inverse of the probability of selection. There are three major adjustments made to the weights as part of the estimation procedures. First, the weights are adjusted for all interviewed households to account for noninterviewed households. Next, two ratio adjustments are done. The first-stage ratio estimation procedure reduces the portion of the variance that results from requiring sampled primary sample units (PSU's) in a State to represent nonsampled PSU's in the same State and is not applied to data from self-representing PSU's. This adjustment is made separately for two racial groups, blacks and nonblacks. This adjustment corrects for differences between the distribution of these two racial groups in the sample PSU's and the State that existed at the time of the 1990 census. The second-stage ratio estimation procedure adjusts the weights for sampled persons in specific demographic categories to be consistent with the independent current estimates of the civilian

<sup>1</sup> All views expressed in this paper are those of the authors and do not reflect the views or policies of the Bureau of Labor Statistics or the Bureau of the Census.

noninstitutional population. This adjustment further reduces variability of the estimates and at the same time corrects for undercoverage.

The next stage of the CPS estimation process is the use of a composite estimation procedure. The main reason for applying this procedure is to further reduce the variance, especially on estimates of month-to-month change. The current composite estimator adjusts for month-in-sample bias (Bailar 1975). The expected value of an estimate of the unemployment rate is higher for new participants (month-in-sample one) than the average unemployment rate for the entire sample. Similarly, those reentering the survey, month-in-sample five, have a higher unemployment rate than the average of participants for month-in-sample six, seven, and eight. The composite estimator differentially weights the month-in-sample data in order to adjust for month-in-sample bias. Finally, seasonal adjustment procedures are applied to the final estimates. A more detailed description of the sample design and estimation process is given in U.S. Bureau of the Census (in preparation).

### III. Options Considered for Changing Data Collection Procedures

November Data Collection--The first Federal Government Shutdown covered the period from November 14 through November 19, 1995. Normally, during the days covered by this shutdown, a good deal of preprocessing would have occurred with the Census Bureau interviewers receiving their case assignments as well as updates to the instrument for the month of November. Interviewers would have started interviewing on Sunday, November 19th. When the Federal Government opened on Monday, November 20th, the interviewers could not immediately begin collecting data because this preprocessing had not yet been performed. In effect, two critical days for data collection were lost. An additional factor further complicated the data collection with the Thanksgiving holiday falling during the data collection week. In view of these conditions, the following options were considered for CPS data collection: eliminating the November data collection, changing the reference week, delaying the data collection by one week, and collecting data as soon as possible with an extended collection period.

A case was made for not collecting November data at all because if the unemployment rate changed, then it would be impossible to separate the effect of the changes in data collection procedures from real economic changes--in particular the effect of the shutdown. To attempt to gauge the effects of altering

the data collection procedures, we examined several prior months data by the day of interview. This research revealed that the labor force participation rate is higher and the unemployment rate is lower for persons responding on the first Sunday of the data collection period relative to the average of all days. There was uncertainty whether this group of respondents could be reached for the November data collection period, especially because of the Thanksgiving holiday. Ultimately, this option was dismissed because it would not only leave a hole in the time series but also have an adverse effect on rotation group bias for future months, composite estimation, year-to-year changes, and quarterly and annual averages.

Another option considered was a change in the reference week to the week containing the 19th of the month with data collection beginning one week later. The main drawback to this option was the appropriateness of seasonal adjustment factors especially in light of the Thanksgiving holiday. It was agreed that a better choice would be to keep the reference period as the week containing the 12th even if the data collection was delayed by as much as one week. Delaying the data collection, however, would increase the effect of recall bias on responses.

In the end, a decision was made to take whatever action necessary to begin an accelerated data collection and to extend the data collection period by one day. The interviewers were instructed that they would be paid to work overtime and to adjust their schedules as they saw fit to complete this task. In spite of these initiatives, the final nonresponse rate for November was 7.2 percent, significantly higher than an average of 6.6 percent for the first ten months in 1995. The national monthly unemployment rate, seasonally adjusted, was 5.5 percent in October and 5.6 percent in November. There were no apparent effects from these altered procedures on the data for November.

December Data Collection--The second Federal Government Shutdown covered the period from December 16, 1995 through January 5, 1996. Although the first shutdown caught both agencies quite by surprise, we were able to anticipate the possibility of the second shutdown and its possible effects. Compared to the November shutdown, a shutdown on December 16th would curtail data collection on CPS rather than delay its start. The normal data collection period would have been December 10 through December 19. Although the week of the 12th is used as the reference week for every other month, it is quite common for the reference period to be the week containing the 5th of the month for

December; this is done so that the data collection does not interfere with Christmas holidays. Thus, the seasonal factors already account for this change in the reference period.

Because the December shutdown would permit data collection for the first six days (i.e., December 10 - 15), we analyzed previous months data to examine the characteristics of later respondents. This research indicated that the percentage of cases that were completed in days 7-10 of the collection period varied by month-in-sample (see Table 1). Specifically, about 20 percent of the month-in-sample 1 and 5 cases are completed in days 7-10 of the collection period as compared to an average of 11 percent for all eight months. We also found that a differential percentage of cases were completed in the same period across the 12 Census Regional Offices. For example, in the New York Regional Office about 17 percent of the cases were completed in days 7-10 as compared to about 9 percent in the Kansas City Regional Office. (Thus, we knew that there were some differences between later and earlier respondents to the CPS.)

We then examined the effect of treating these late respondents as nonrespondents on three key labor force statistics--unemployment rate, employment-to-population ratio, and labor force participation rate at the national level for five months of data (see Table 2). In general, a change of 0.2 percentage point in the unemployment rate is considered statistically significant at the 90 percent confidence level; for employment-to-population ratio and labor force participation rate, a change of 0.3 percentage point is significant. These data indicate that the effect of designating all units responding after the sixth day of collection as nonrespondents results in biased estimates for all three major labor force characteristics.

Because of this potential bias, especially on the unemployment rate, we considered three options for data collection: first, begin the collection one day early; second, to aggressively accelerate interviewing based on the assumption that there would be a shutdown; third, to perform normal procedures and model the existing data in case of a shutdown. The option to begin collection on Saturday was quickly rejected because of the effect it would have on "actual hours worked" estimates for respondents working that Saturday.

Of the two remaining options, the one to aggressively pursue interviewing appeared more appealing since it would take time to develop a suitable model required by the last option. Additionally, the model would have its

own biases. Accelerating data collection required authorization for overtime, telephone interviews for month-in-sample 5 cases that are normally done by personal visits, and a delay in conducting reinterviews. Because of our research findings, interviewers were instructed to collect data that would be representative of all eight months-in-sample groups. That is, to collect about an equal percentage of cases for all eight groups. Additionally, Regional Offices were provided with information about their number of completed cases for the first six days and were encouraged to boost their response rates. These accelerated procedures were put in operation but were later withdrawn because there were indications of a temporary resolution to the Federal budget disagreements. This change in policy caused confusion for the interviewers and some of them did proceed to collect data on December 16th and 17th. In the end, about 3,000 cases were still left incomplete in the field because of the shutdown. This resulted in a nonresponse rate of 9.2 percent and an unemployment rate of 5.6 percent for December (see Section IV).

January Data Collection—A continuing resolution was passed that allowed normal operations to resume at the Bureau of Labor Statistics and the Bureau of the Census on January 8th; however, the Federal Government in Washington, DC was shutdown due to a major snowstorm the week of January 8th, except for the 11th. January data collection was thus affected by: 1) the extraordinarily high number (3,000) of incomplete cases from December that still had to be processed; 2) higher than average nonresponse rates for November and December data collection; and 3) a shutdown in Washington, DC. A major factor easing the data collection for January was that the Federal Government in Washington was open on Thursday, January 11th, setting the stage for planning. The issues related to January data collection were much simpler than the data collection issues for November and December. For this month of data collection, the best option was to use accelerated procedures. By working on Saturday and Sunday, the Census Bureau was able to complete necessary processing of December materials and prepare for the January data in time to start limited collection from the centralized facilities by Sunday evening. Full field collection was in operation on Monday, January 15th; in effect, one day of data collection was lost. The nonresponse rate of 7.7 percent for January was significantly higher than an average of 6.6 percent and the seasonally adjusted national unemployment rate was 5.8 percent.

#### IV. Alternative Estimation Procedures Considered for December Data

In terms of estimation, the December data posed the most significant potential problems. Considering the premature termination of data collection, the December nonresponse rate of 9.2 percent was not bad. Since the average nonresponse rate was 6.6 percent, there was, however, concern about the reliability of the estimates. Was it possible to improve the reliability of the estimates by taking advantage of the 75 percent sample overlap between previous and current months as well as current and future months? In other words, was it possible to develop a model that would improve the reliability of the labor force estimates by imputing for nonrespondents?

To answer these questions, simulations were done on a special data set constructed using the CPS data for August 1995. The following three imputation procedures were performed for sampled cases that were not completed by the sixth day of collection.

1. Treat all incomplete cases as noninterviews (same procedure as in Table 2).
2. Treat all incomplete cases in month-in-sample one and five as noninterviews; for the remaining month-in-sample groups, use previous and current month's data (in this case, July and August 1995) to impute labor force status.
3. Use the previous and current month's data to impute labor force status for the incomplete cases in all month-in-sample groups except one and five; use the current and following month's data (in this case, August and September 1995) to impute labor force status for incomplete cases in month-in-sample one and five.

After each edit procedure, all the usual CPS weighting adjustments were performed on the data. Note that implementation of the third edit procedure for December would have required waiting until January data were available. In this situation, the delay in release of December data was acceptable. The results of this simulation are shown in Table 3. These data indicate that neither edit procedure number two or three provided estimates uniformly closer to those from the full CPS sample than the first method (designating incomplete cases as noninterviews). The data also indicate that the effects of the alternative edit procedures vary by demographic groups. Using data from three monthly samples results in substantial

improvement only for large groups; using data from the current and previous months does not appear to provide substantial overall improvement. For smaller demographic groups such as blacks and Hispanics, use of additional monthly data does not appear to provide overall improvement in the estimates.

Based on the results of this research, a decision was made to produce and publish the December estimates according to the normal CPS estimation procedures. As noted earlier, the seasonally adjusted, unemployment rate of 5.6 percent for December was the same as for November.

#### V. Summary

In summary, we learned a great deal from the research that was conducted to examine and anticipate the effect of the shutdowns of our agencies on the quality of the data from CPS. Specifically, we learned about: the characteristics of the early vs. late respondents with respect to labor force status; the variation in the percentage of completed cases in the first six days of collection by month-in-sample and by Regional Offices; and the effectiveness of our estimation procedures with respect to noninterview adjustment, and first- and second-stage ratio adjustments. From our simulations, we learned that imputing for responses obtained during the last days of data collection by using prior and future months data does not appear to provide substantial overall improvement compared to the normal estimation procedures. Before this research, much of the information that existed was anecdotal in nature.

Finally, it is our recommendation that accelerated data collection procedures be utilized in situations like those mentioned above because there is no substitute for high response rates. In panel surveys like the CPS, consequences of higher than normal nonresponse tend to affect the rates for future months. The monthly CPS nonresponse rates in 1996 remained higher than the 1995 rates for several months following these shutdowns.

#### VI. References

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

U.S. Bureau of Census (in preparation), *The Current Population Survey; Design and Methodology, Technical Paper No. 63*, U.S. Government Printing Office, Washington, D.C.

**Table 1. Percentage of Cases Completed in Day 7-10 of Data Collection Period**

<b>Month in Sample</b>	<b>December '94</b>	<b>October '95</b>	<b>September '95</b>	<b>August '95</b>
<b>Month 1</b>				
total cases	6576	6351	6403	6624
% of cases days 7-10	20.76	19.86	23.61	18.27
<b>Month 2</b>				
total cases	6803	6611	6612	6789
% of cases days 7-10	8.86	8.77	13.25	9.53
<b>Month 3</b>				
total cases	6915	6602	6574	6838
% of cases days 7-10	8.52	8.15	13.54	8.85
<b>Month 4</b>				
total cases	6778	6611	6570	6912
% of cases days 7-10	8.17	8.59	13.38	9.36
<b>Month 5</b>				
total cases	6706	6465	6390	6681
% of cases days 7-10	18.39	17.05	22.72	18.23
<b>Month 6</b>				
total cases	6716	6510	6602	6764
% of cases days 7-10	8.26	9.11	13.63	10.07
<b>Month 7</b>				
total cases	6751	6668	6558	6857
% of cases days 7-10	7.55	7.83	13.69	8.78
<b>Month 8</b>				
total cases	6757	6598	6639	6620
% of cases days 7-10	7.77	7.84	12.31	8.66
<b>Total across months 1-8</b>				
total cases	54002	52416	52348	54085
% of cases days 7-10	10.99	10.84	15.71	11.43

**Table 2. Effect of Late Respondents on Labor Force Estimates  
All Workers**

<b>Month</b>	<b>Unemployment Rate (%)</b>			<b>Employment-to-population Ratio (%)</b>			<b>Labor force participation Rate (%)</b>		
	<b>6 days</b>	<b>complete</b>	<b>difference</b>	<b>6 days</b>	<b>complete</b>	<b>difference</b>	<b>6 days</b>	<b>complete</b>	<b>difference</b>
October 1995	5.42	5.24	0.18	63.05	63.30	-0.25	66.66	66.80	-0.14
August 1995	5.77	5.66	0.11	63.13	63.36	-0.23	67.00	67.16	-0.16
July 1995	6.13	5.92	0.20	63.53	63.85	-0.33	67.68	67.88	-0.20
May 1995	5.71	5.53	0.18	62.46	62.92	-0.47	66.24	66.61	-0.37
January 1995	6.43	6.24	0.18	61.69	62.02	-0.33	65.93	66.15	-0.22

Note: All estimates are uncomposited and not seasonally adjusted

**Table 3. Effect of Late Respondents on Labor Force Estimates  
August 95**

Procedure	<u>Unemployment Rate (%)</u>				<u>Employment-to-Population Ratio(%)</u>				<u>Labor Force Participation Rate (%)</u>			
	<u>Aug</u> (1)	<u>Jul-Aug</u> (2)	<u>July-Sept</u> (3)	<u>Complete</u>	<u>Aug</u> (1)	<u>Jul-Aug</u> (2)	<u>July-Sept</u> (3)	<u>Complete</u>	<u>Aug</u> (1)	<u>July-Aug</u> (2)	<u>July-Sept</u> (3)	<u>Complete</u>
All workers difference*	5.77 (0.11)	5.75 (0.09)	5.68 (0.02)	5.66	63.13 (-0.23)	63.18 (-0.18)	63.27 (-0.09)	63.36	67.00 (-0.16)	67.03 (-0.13)	67.08 (-0.08)	67.16
Men, 16+ difference*	5.39 (0.08)	5.43 (0.12)	5.38 (0.07)	5.31	71.55 (-0.14)	71.43 (-0.26)	71.46 (-0.24)	71.69	75.63 (-0.09)	75.54 (-0.18)	75.52 (-0.19)	75.71
Women, 16+ difference*	6.22 (0.15)	6.11 (0.04)	6.03 (-0.04)	6.07	55.38 (-0.31)	55.59 (-0.11)	55.74 (0.05)	55.69	59.06 (-0.23)	59.20 (-0.09)	59.32 (0.03)	59.29
Teens difference*	15.11 (0.19)	14.44 (-0.49)	14.47 (-0.45)	14.92	52.18 (0.29)	52.53 (0.64)	52.23 (0.33)	51.90	61.48 (0.48)	61.39 (0.40)	61.06 (0.06)	61.00
White difference*	4.88 (0.11)	4.85 (0.08)	4.80 (0.03)	4.77	64.18 (-0.20)	64.21 (-0.17)	64.29 (-0.09)	64.38	67.47 (-0.13)	67.49 (-0.12)	67.54 (-0.07)	67.60
Black difference*	11.50 (0.02)	11.41 (-0.08)	11.22 (-0.26)	11.48	56.30 (-0.39)	56.45 (-0.25)	56.60 (-0.10)	56.70	63.62 (-0.43)	63.72 (-0.33)	63.76 (-0.30)	64.05
Hispanic difference*	9.86 (0.21)	9.65 (-0.00)	9.60 (-0.05)	9.65	60.19 (-0.03)	60.31 (0.09)	60.49 (0.26)	60.22	66.77 (0.12)	66.75 (0.09)	66.91 (0.25)	66.66

\* difference = simulated value - official value

Note: All estimates are uncomposited and not seasonally adjusted