Draft: May 10, 1996

Federal Government Shutdown: Options for CPS Data Collection

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

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

I. Introduction

The Federal Government shutdowns in November 1995 and again in December 1995 that extended into January 1996 created extraordinary challenges for data collection activities pertaining to the Current Population Survey (CPS). The CPS is a national labor force survey of over 50,000 households conducted monthly by the Bureau of Census for the Bureau of Labor Statistics (BLS). In addition to being best known for the national monthly unemployment rate--one of the leading economic indicators for the nation--CPS is the source of thousands of labor force statistics at the national and sub-national level. In this paper, we discuss the various options that were considered by the officials at both agencies for collection of November 1995 through January 1996 data and the potential effects associated with each option.

The options considered were: not collecting data for a particular month, changing the reference week, delaying collection, starting collection early, extending the collection period, and using accelerated (e.g., overtime) collection procedures. Changes in data collection procedures for CPS could, among other things, result in recall bias, different rotation group bias, and distortion of compositing estimation and seasonal adjustment factors. We examined the effect on estimates because of an increase in noninterviews resulting from closing down the survey prior to completion of normal interviewing. The research conducted on several months of prior data indicated that the characteristics of the late respondents differed from the early respondents. In particular, the expected values of some key labor force estimates were different. The data collection decisions that were made attempted to minimize the potential biases we uncovered.

In this paper, we first give a brief overview of the CPS sample design; discuss the consequences of implementing each option under consideration; compare characteristics of late respondents with those of early respondents; describe alternative estimation procedures to compensate for truncated data collection; present results from these procedures; summarize the course of action taken for collecting data for each of the three months and the achieved final response rates.

II. Overview of CPS Sample Design

The CPS is a multistage state based probability sample of over 50,000 eligible households with the reference period being the week containing the 12th of the month and the data collection beginning on Sunday during the week that includes the 19th. 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 from year to year, thereby reducing variance for estimates of monthly changes and over the year changes for the same month. The interviews in the first and fifth month-in-samples are conducted by personal visits while those for the remaining months are usually done over the telephone either by the field interviewers from their home or from one of the three centralized facilities.

The initial weight for the CPS reflects the inverse of the probability of selection. The estimation procedures involves first adjusting the weights for all interviewed households to account for noninterviewed households; next, the 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 self-representing PSU's. This adjustment is made by cells of black and nonblack and corrects for differences that existed at the time of the 1990 census between the race distribution of the population in the sample PSU's and the known race distribution of the State. The second stage ratio estimation procedure adjusts the sample proportions of persons in specific demographic categories to the distribution of independent current estimates of the civilian noninstitutional population. This adjustment further reduces variability of the estimates and at the same time corrects for undercoverage.

In the CPS estimation process, a last step before seasonal adjustment 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.

A detailed description of the sample design is given in the U.S. Bureau of the Census, *Current Population Survey; Design and Methodology*, Technical Paper No. 63, forthcoming.

III Options Considered for Changing Data Collection Procedures

November Data Collection--The first Federal Government shutdown was due to budget and it covered period from November 14 through November 19, 1995. Normally, during the days covered by this shutdown, the Census Bureau interviewers would have received their case assignments as well as updates to the instrument for the month of November and they 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 preprocessing function had yet to be performed. In effect, two critical days for data collection were lost. An additional factor complicating the data

collection was Thanksgiving Holiday falling during the data collection week. Given that, the following options were considered: not collecting the November data, 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 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. Research on prior data showed that labor force participation rate is higher and unemployment rate 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 November data collection period, especially because of the Thanksgiving Holiday. Although there was concern about bias in the estimates due to missing respondents who are normally contacted the first Sunday of the data collection period, 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, over the year changes, and quarterly and annual averages.

Another option considered was to change the reference week to the week containing the 19th of the month and begin the data collection one week later. The main drawback to this option was the appropriateness of seasonal 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. The main concern, however, with this option was the effect of recall bias.

In the end, a decision was made to take whatever action necessary to collect the data as soon as possible and to extend the data collection period by one day. The interviewers were instructed to work overtime and to adjust their schedules as they saw fit to complete this task. Due to these initiatives, the final nonresponse rate for November was 7.2 percent as compared to an average of 6.6 percent for the first ten months in 1995. The national monthly unemployment rate, seasonally adjusted, changed from 5.5 percent in October to 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 was again due to budget and it covered the period from December 16, 1995 through January 5, 1996. 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. The normal data collection period would have been December 10 through December 19.

Unlike the November shutdown, both BLS and Census staff were better prepared for this anticipated shutdown. Data from previous months indicated that percentage of cases that were completed in days 7-10 of collection period varied by month-in-sample (Table 1). In particular, about 20 percent of the month-in-sample 1 and 5 cases are completed in

days 7-10 of collection period as compared to an average of 11 percent for all months. The data also showed that a differential percentage of cases were completed in the same period across the 12 Census Regional Offices. For example, in New York Regional Office about 17 percent of the cases were completed in days 7-10 as compared to about 9 percent in Kansas City Regional Office. We had also examined the effect of treating 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 (Table 2). These data indicate that the effect of designating all units responding after the sixth day of collection as nonrespondents is marginally significant for all three major labor force characteristics. Because of this potential bias, especially on the unemployment rate, there were 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 shutdown; third, to perform normal procedures and model the existing data in case of shutdown. The option to begin collection on Saturday was quickly rejected because the effect it would have on "actual hours worked" series for respondents working that Saturday.

Of the two remaining options, the one on aggressively pursue interviewing appeared more appealing since it would take time to develop a suitable model. 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. Interviewers were also instructed to collect data more evenly across all eight month-in-samples cases. Additionally, Regional Offices were provided with the table that displayed the distribution of completed cases for first six days and days 7-10 for all Regions; some Regions 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. 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.

January Data Collection--The three factors that affected the data collection for January were: 1) the extraordinary 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) Federal Government shutdown due to snow on January 8, 9, 10, and 12, 1996. A major factor easing the data collection for January was that Federal Government was open on Thursday, January 11th, setting the stage for planning. The issues related to January data collection were much simpler compared to the data collection for November and December. For this month 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 for January was 7.7

percent as compared to an average of 6.6 percent and the, seasonally adjusted, national unemployment rate was 5.8 percent compared to 5.6 for December.

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 month as well as current month and future month? 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 prior to the second shutdown on a special data set constructed using the CPS data for August 1995. The following three imputation procedures were performed on 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 MIS 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-samples 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 given 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. 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 tremendous amount of research that was done in connection with the shutdowns. 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 also 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 of anecdotal nature.

Finally, it is our recommendation that accelerated data collection procedures be utilized in situation like the ones mentioned above because there is no substitute for high response rates. In surveys like CPS, consequences of higher than normal nonresponse tend to affect the rates for future months as shown below.

Month	Nonresponse Rate in %
August 95	6.3
September	6.6
October	6.5
November	**7.2**
December	**9.2**
January 96	**7.7**
February	8.0
March	7.8

V. 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, *The Current Population Survey; Design and Methodology, Technical Paper No. 63*, U.S. Government Printing Office, Washing, D.C., forthcoming.