HOW DOES SMOOTHING ESTIMATED MONTHLY CONTROL TOTALS AFFECT SIPP ESTIMATION?

Lynn Weidman and Larry Bobbitt, Bureau of the Census *
Lynn Weidman, Washington, DC 20233

1. Introduction

For each of the Census Bureau's household surveys, calculating final person weights involves several stages of adjustment to the initial sampling weights. The number of stages and details of the procedures differ somewhat by survey, but they all have two stages in common: adjustment for household nonresponse and adjustment to monthly age x race x sex (demographic) totals. When the latter adjustment is carried out for the Survey of Income and Program Participation (SIPP), it is done simultaneously with adjustment to monthly controls for six household types. These controls are estimated from the Current Population Survey, whereas the demographic totals are projected from the previous decennial census using birth, death and immigration data.

Each survey uses these same demographic totals and they follow smooth long term trends. SIPP is the only survey to use household type controls which, being estimates, show quite a bit of month-to-month variation. It seems reasonable that since these controls are being treated as the "truth" for SIPP, it is desirable to fit to their "true" values rather than the "true" values plus monthly variability due to sampling. This study considers a smoothing of the household type controls by time series methods and compares properties of the original and smoothed controls, as well as properties of estimates derived from the two sets of controls. The next section gives an overview of the adjustment procedure being analyzed and the smoothing methodology used. The estimates are compared in section 3, and the final section discusses the results.

The investigation carried out is in the spirit of exploratory data analysis. No statistical tests have been performed or statistical claims made. We have compared the month-to-month variability for estimates using the original and smoothed weights, and tried to determine if there is a consistent pattern of variability reduction when the smoothed weights are applied. This analysis will help to determine the utility of smoothing household type controls before using them in the final adjustment procedure.

2. Final Adjustment and Smoothing Methodology

Figure 1 shows the two-way table used in the SIPP final stage adjustment of black males aged 15+. There is a similar table for each of eight age (0-14, 15+) x race (black, nonblack) x sex (male, female) combinations. The weighted cell counts in a table are alternately ratio adjusted to the row and column controls until each marginal count is within a specified closeness to its control. The overall adjustment procedure is quite complicated, including adjustments for Hispanic totals and swapping of husband and wife weights between male and female tables, and will not be further described. This study includes only persons 15+ because a separate adjustment program would have to be written for the four age 0-14 tables, but any pattern in the results should also hold when these additional persons are included.

Consider the household type controls that we want to smooth, one corresponding to each column in Figure 1. For each table there are six monthly time series of household types that are restricted to add up to the sum of the row totals (we will call this the table total) for each month, so the smoothed series must maintain this property. How smooth do we want the adjusted controls to be? Some of the factors, e.g., births, deaths, and marriages, that contribute to change in household type exhibit seasonal patterns of variation, so we don't want to eliminate seasonality from the adjusted controls. If we view each time series as the sum of three components, (trend + seasonality + variability), then the best we can do is smooth a series by removing its variability component.

Our ability to smooth these time series is limited since there were only 72 data points, June 1983 through May 1989, available for each series when this study commenced. Despite this, the Census Bureau's X-11 seasonal adjustment program (Shiskin, 1967) diagnostics, as well as spectrum plots, were used to determine which of these series were seasonal. The variability component determined by X-11 was removed from the seasonal series. Two series, household types 5 and 6 for black females, were found to not have a seasonal component. These series were viewed as the sum of only two components, (trend + variability). ARIMA models for these two series were identified, and the series were extended by one year of forecasts and one year of backcasts. The extended series were smoothed using LOESS, (Cleveland and Devlin, 1988, Cleveland, 1979). This procedure was employed to improve the quality of the smoothing at the ends of the

series.

The above smoothing procedures were carried out on all the series without the requirement that the monthly values of the six series in a set sum to the monthly table totals. Fortunately, these two totals are nearly equal. This is probably because when the six estimates were calculated from CPS the two totals were required to be equal, and so the total of the variability components of the series as estimated by X-11 is very close to zero. Because the monthly smoothed totals are so close to the table totals, a last simple monthly adjustment was made by multiplying each of the six smoothed controls in a set by the table total divided by the smoothed total. We call the resulting series the smoothed series.

Table 1 summarizes the effect of smoothing on the relative month-to-month changes in the controls, $(c_{i+1} - c_i)/c_i$, where c_i is a column total for month i, by comparing their mean absolute values for the original and smoothed series. An examination of these statistics shows that the smoothing has satisfied our desire to reduce month-to-month variability in the series.

3. Comparison of Estimates

So far we have smoothed the household type controls to be used in the final stage of adjusting person weights. Because of the iterative procedure used in the final stage, we cannot predict the effect it will have on final weights or on estimates made using them. What we expect is that the month-to-month variation in many estimates will be reduced, especially for variables closely related to household type. We will look at the eleven month-to-month change estimates available from the twelve months of 1988 and the 1987 SIPP panel, which are summarized by the mean absolute deviation (MAD) statistic. They are compared for final weights calculated using the original and smoothed controls. Recall that these are informal comparisons of the patterns of increase/decrease in computed change estimates, not statistical significance tests.

There are three types of estimates included in this study: demographic characteristics (household type and marital status), income-related (number of persons in poverty and mean income), and program benefits recipiency (social security, unemployment compensation, AFDC, food stamps and child support). They are calculated for various demographic combinations as given in the Appendix. Table 2 includes comparisons of month-to-month change, quarter-to-quarter change, and annual estimates derived using the original and smoothed controls. Because of the large number of estimates examined, selected comparisons that are representative of the general patterns found are shown.

For all household types there is a large reduction in the

month-to-month changes when using the smoothed weights. (Only one example is given in Table 2.) We expect the observed pattern of variability reduction for blacks and nonblacks because their household type totals have been directly smoothed. This pattern does not old for Hispanics, where the changes in MADs are in both directions. This differing Hispanic behavior apparently occurs because in the final stage adjustment they are removed after several iterations and have a separate adjustment to Hispanic controls performed on them. The results for Hispanics follow this same lack of pattern for other estimates, so they will not be discussed further.

We expect that the effect on other types of estimates will be less noticeable, since they haven't been directly smoothed. Single and married, spouse present show a pattern very similar to those for household type. Excluding Hispanics, there are no cases for which the MAD is increased by using the smoothed weights. The changes for married, spouse absent and widowed are not as marked, and in a few cases the MAD increases. Divorced males have increased MADs in all cases, while divorced females have changes in both directions. The different patterns evident in marital status reflect our general expectations, with apparent reduction in variation dependent on how close the categories are to household types. For example, married, spouse present persons are mostly from the married couple family type, and we expect the effect on MAD to be similar for the two. Divorced persons fall into both other family and nonfamily households, so the estimates are a combination of estimates from the two types, and do not closely follow the pattern of either.

The poverty and program recipiency estimates in the lower half of Table 2 show a complete lack of reduction pattern. The MADs for numbers of persons in poverty show little change in either direction, even for household types. One factor possibly contributing to this is that there is a lot of movement of households into and out of poverty each month, which is in no way affected by the smoothing. Another cause is probably that these estimates cut across separate male and female smoothing more than did many of the estimates for marital status.

We have computed similar tables for marital status and household type cross-classified by age groups. The results are closer to what we see in the lower part of Table 2 than in the upper. For household type the MADs are usually marginally smaller for the adjusted, but not noticeably so, while for marital status there are small changes in both directions. The sum over age groups was smoothed, not the individual groups themselves, which is the likely reason for this result.

We also compared the quarter-to-quarter MADs and the annual totals for the original and smoothed weights,

in order to determine the effects of the smoothing on longer-term estimates. Before doing the study, we thought that the differences between the two weightings would decrease as the time period of an estimate increased. This turned out to be the case for annual estimates, as there is rarely as much as a .1% difference between the two. This is due to the trend component of the original series being changed little in the smoothing process. However, no overall pattern emerged for the quarterly estimates. This is because the short-term trends of the smoothed series are still very irregular due to the presence of seasonality.

4. Discussion

This study was initiated with the idea that smoothing the monthly household type controls would reduce month-to-month variation in person weights, which would lead to a reduction in month-to-month change estimates for many variables. This could be the case for SIPP because of the relative constancy of sample persons constituting a SIPP panel, but not for other Census Bureau household surveys because of their rotating samples.

The results were as expected for household type estimates and some marital statuses that are closely related to individual household types. For the other variables examined there was no general reduction in monthly variability. After rethinking the study, two possible reasons for the latter result were proposed. First, monthly variablity for a characteristic will most noticeably be reduced only if the group of people having the characteristic remains largely unchanged from month to month. If the individuals in the group are changing, so are the weights of those who make up the group, and the estimates will also change. Second, the smoothed household type controls still have too much variation to give a general reduction in the month-to-month variability of person weights. The validity of these reasons could be examined by in further studies by, respectively, finding estimates that are derived from an almost unchanging cohort, and making the household type series much smoother. The latter could be accomplished by, for example, eliminating all seasonality.

What we have found out to this point is that we can reduce the monthly variablity in household type estimates and several related marital statuses via the proposed smoothing approach. Is it worthwhile to go through the smoothing process to accomplish this alone? Not with the available results. A necessary next step is the comparison of these estimates to benchmarks to determine if the original or smoothed estimates of characteristics related to changes in household types, such as marriages, births and deaths, are closer to the estimates from administrative records. If the smoothed estimates are closer, then a change in procedures would be recommended.

References

Cleveland, W.S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots", JASA, 74, 829-836.

Cleveland, W.S., and Devlin, S.J. (1988), "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting", JASA, 83, 596-610.

Shiskin, J., Young, A.H., and Musgrave, J.C. (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program", Bureau of the Census Technical Paper No. 15.

Appendix

ESTIMATES COMPARED

Demographic Characteristics

Household type
Total, Race of householder
Hispanic origin of householder
Age of householder

Marital status by sex
Total, Race, Hispanic origin, Age

Poverty and Income

Number of persons in poverty
Total, Household type, Hispanic origin
Race x sex, Race x age

Median income
Household type, Hispanic origin, Race x sex

Program Recipiency

Number of persons receiving benefits:
social security, unemployment compensation,
AFDC, food stamps, child support
Hispanic origin, Race x sex

FIGURE 1. SECOND STAGE CELLS FOR BLACK MALES (15+ YEARS OF AGE)

Age	Persons in Households that contain a Primary Family or a Subfamily				Persons not in Household containing a Primary Family or a Subfamily	
	Husband of Primary Family	Male Householder No Spouse Present	Other Household Members		Householder	Not Householder or Persons in Group Quarters
			Husband of Subfamily	Not a Husband		
15 16-17 18-19 20-21 22-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70+	column 1	column 2	column 3	column 4	column 5	column 6

TABLE 1. MEAN ABSOLUTE MONTH-TO-MONTH CHANGES FOR ORIGINAL AND SMOOTHED HOUSEHOLD TYPE CONTROLS

HH Type Control	Black Males	Black Females	Nonblack Males	Nonblack Females
1 Original	1.15%	1.12%	0.25%	0.25%
1 Smoothed	0.74%	0.77%	0.19%	0.16%
2 Original	4.23%	1.54%	2.43%	0.99%
2 Smoothed	2.58%	0.98%	1.25%	0.79%
3 Original	11.32%	11.46%	3.55%	3.54%
3 Smoothed	7.87%	8.03%	2.64%	1.58%
4 Original	2.02%	1.25%	0.97%	0.64%
4 Smoothed	1.27%	0.86%	0.63%	0.45%
5 Original	2.81%	1.82%	0.98%	0.59%
5 Smoothed	2.20%	0.75%	0.75%	0.39%
6 Original	9.12%	8.53%	3.49%	2.82%
6 Smoothed	6.25%	1.64%	2.46%	2.17%

TABLE 2. COMPARISONS OF ESTIMATES FROM ORIGINAL AND SMOOTHED CONTROLS

Characteristic	Change in month-to-month differences	Change in quarter-to-quarter differences	Change in estimates of annual totals
Married Couple Family Households:			
Nonblack	-65.0%	4.3%	-0.02%
Black	-45.4%	-20.01%	-0.08%
Single Males:			
Nonblack	-29.7%	-20.4%	0.00%
Black	-35.8%	-3.1%	-0.06%
Single Females:			
Nonblack	-37.0%	-75.5%	-0.01%
Black	-32.0%	-25.4%	0.08%
Married Male:			
Spouse Present			
Nonblack	-61.5%	3.8%	-0.02%
Black	-45.5%	-25.4%	0.08%
Spouse Absent			
Nonblack	-14.3%	3.2%	.10%
Black	34.6%	-17.4%	-.01%
Married Female:			
Spouse Present			
Nonblack	-63.5%	-6.1%	-0.025
Black	-55.3%	-29.4%	-0.09%
Spouse Absent			
Nonblack	-3.5%	-6.8%	0.06%
Black	2.8%	-0.2%	-0.09%
In Poverty:			
Married Couple Family Household	-2.0%	0.9%	-0.02%
Nonblack Male	1.2%	-3.3%	0.03%
Nonblack Female	-0.8%	-4.6%	-0.02%
Black Male	1.8%	9.2%	-0.07%
Black Female	-0.2%	1.6%	0.05%
Social Security:			
Nonblack Male	7.9%	12.7%	0.00%
Nonblack Female	-1.1%	-1.1%	0.00%
Black Male	-1.7%	-5.3%	0.02%
Black Female	7.7%	-0.6%	-0.04%
Unemployment:			
Nonblack Male	0.1%	-0.6%	0.02%
Nonblack Female	0.1%	0.1%	-0.03%
Black Male	-1.6%	0.1%	0.11%
Black Female	-1.1%	0.6%	-0.03%
Receives AFDC Benefits:			
Nonblack Male	0.2%	0.0%	0.08%
Nonblack Female	-8.1%	5.6%	0.07%
Black Male	-0.3%	-2.1%	-0.09%
Black Female	-5.5%	6.3%	-0.12%
Receives Food Stamps:			
Nonblack Male	.7%	1.8%	0.07%
Nonblack Female	14.3%	-16.3%	0.07%
Black Male	2.4%	1.0%	-0.02%
Black Female	12.4%	8.0%	0.14%
Receives Child Support:			
Nonblack Male	-5.1%	2.3%	0.02%
Nonblack Female	-15.8%	-4.9%	0.05%
Black Male *			
Black Female	-4.2%	-1.5%	-0.11%

    * = no cases