

TIME-IN-SAMPLE BIASES IN THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Rameswar P. Chakrabarty and Todd R. Williams, Bureau of the Census *
Rameswar P. Chakrabarty, Washington, D.C. 20233

KEY WORDS: Panel Surveys; Non-sampling error;
Response bias.

1. Introduction

Sample households in each yearly panel of the Survey of Income and Program Participation (SIPP) are interviewed once every four months for 2½ years. To facilitate field operations, each panel is divided into four approximately equal subsamples called rotation groups; one rotation group is interviewed in a given month. Thus, one cycle or "wave" of interviewing generally takes four consecutive months. The reference period for the survey is the four months preceding the interview month. It has been reported by many different survey organizations [1] that the number of times respondents have been exposed to a survey seems to affect the data reported. For many household characteristics, different subsamples with different times-in-sample provide estimates for the same period that differ significantly. In the Current Population Survey (CPS), the sample in each month has eight rotation groups representing subsamples that are in the sample for the first, second ... and eighth time. The existence of bias due to rotation groups has been well-documented for the CPS by Bailar [1].

Time-in-sample biases are non-sampling errors and such biases occur because with time, sample composition changes and/or survey conditions that influence responses change in some significant ways. The ways in which the survey conditions in the CPS tend to change with number of times in sample have been identified by Hanson [3].

For the SIPP, it has been documented by Nelson, Bowie and Walker [6] that noninterview rate increases with waves. An analysis of sample attrition and response patterns of SIPP original sample persons 15 years old and over who survived the sample cut of the fifth/sixth wave is given by McArthur [4] for all 9 interviews of the 1984 Panel. Reasons for noninterviews and characteristics of persons who leave the sample are discussed by Short and McArthur [8] and McArthur [4]. Burkhead [2] shows that month-to-month changes in work status and income reciprocity status increase substantially between waves. Such changes in sample composition and/or noninterview rates are likely to contribute to time-in-sample biases. Various kinds of learning effects are another possible source of the time-in-sample bias in a panel survey like SIPP. For example, a respondent may begin participating in a government program which he or she learned about because of the survey. This would give an upward bias to program participation estimates. Also, a respondent may, after several interviews, avoid mentioning the receipt of an income source to avoid having to answer additional questions about that source of income. This would give a downward bias to reciprocity estimates.

In this paper, we evaluate the effects of repeated interviews on SIPP characteristics using the 1984 Panel data.

2. SIPP Data Structure and Time-in-Sample Effects

The sample for each yearly panel is divided into 4 subsamples called rotation groups and households in each rotation group are interviewed once every 4 months for the life of the panel which is about 2½ years. A new panel is introduced every year. This data collection system does not provide the same time-in-sample pattern for each month as in the CPS. Time-in-sample pattern varies from month to month and quarter to quarter as can be seen from the analysis of data structure for the 1984 Panel that follows.

Table 1 shows the number of rotation groups that provides estimates by reference month and time-in-sample for the 1984 Panel. Unlike CPS, rotation groups in the SIPP cannot be directly equated to time-in-sample effects. Waves of interviews generally indicate the time-in-sample for rotation groups. For the 1984 Panel, T_i , i th time-in-sample, and W_i , i th wave, are related as follows:

For rotation groups 1 & 2 $T_i = W_i$ $i=1, 2, \dots 9$

For rotation group 3 $\begin{cases} T_i = W_i, & i=1, 2, \dots 7 \\ T_8 = W_9 \end{cases}$

For rotation group 4 $\begin{cases} T_1 = W_1 \\ T_i = W_{i+1}, & i = 2, \dots 8 \end{cases}$

It can be seen from Table 1 that beginning with the 4th quarter of 1983, two estimates of a characteristic reflecting two time-in-sample effects can be obtained for most reference months. For example, for October 1983 three rotation groups that are first time-in-sample provide an estimate and one rotation group that is second time-in-sample provides another estimate. The two estimates from households that are first time-in-sample and the households that are second time-in-sample would have the same expected value (i.e., true value) if the time-in-sample has no effect on estimates. However, if the two estimates differ significantly, then we can infer that time-in-sample introduces a bias in an estimate.

It can be seen from Table 1 that although households in each rotation group are repeatedly interviewed (9 times for the 1984 Panel), monthly and quarterly estimates provide direct comparison of only two consecutive time-in-sample effects (time i and $i+1, i > 1$). For example, October 1983 has households interviewed only first and second times and February 1984 has households interviewed second and third times. Similarly, the first quarter of 1984 has

households interviewed second and third times. Only the fourth quarter of each year, 1983, 1984 and 1985, provides two estimates reflecting two time-in-sample effects for all three months in the quarter. Thus, unlike CPS, SIPP data does not provide direct comparisons of effects of all nine repeated interviews. The 1984 Panel data provide an annual estimate for 1984 from households interviewed 2nd, 3rd, 4th and 5th times, and an annual estimate for 1985 from households interviewed 5th to 8th times. Two or more annual estimates reflecting different time-in-sample effects cannot be obtained from the 1984 panel for the year 1984 or 1985.

In the SIPP each monthly interview collects data for four previous months and thus the recall period for any reference month ranges from 1 to 4 months. Table 2 presents a reconstruction of time-in-sample and length of recall patterns for six reference months (Oct., Nov. and Dec. 1983 and Feb., March, April 1984). It will be seen that the two times-in-sample have different recall lengths. For example, the average length of recall is 2 and 4 months respectively for 1st and 2nd time-in-sample estimates for October 1983. For November 1983, estimates have average recall periods 1.5 and 3.5 months respectively. The recall pattern is the same for each time-in-sample comparison (i and $i + 1$, $i = 1, 2 \dots 8$). Thus, time-in-sample comparisons are confounded with recall biases such as telescoping and memory loss if such biases are present. However, note that in the SIPP, information for previous 4 months are obtained at the same interview and consequently, tend to be very similar. An analysis of September 1983 data by Singh [9] did not reveal any statistically significant recall biases. We assume, therefore, that the time-in-sample comparisons for SIPP monthly estimates are not likely to be affected by recall biases.

3. Estimates of Time-in-Sample Bias

The 1984 SIPP 9-wave data file was used for this research. Two estimates for each reference month reflecting two time-in-sample effects (as shown in Table 1) were computed for selected characteristics. An overview of weighting procedures for SIPP is given in Nelson, McMillen and Kasprzyk [5]. Variances of these monthly estimates were computed by the replicate weighting method developed by Fay and documented in Roebuck [7]. The difference between two estimates for a given month was tested for statistical significance at the 5% level. A statistically significant difference would indicate the existence of bias due to time-in-sample effects.

As mentioned earlier, direct comparison of only two consecutive interviews (i and $i + 1$, $i = 1, 2, \dots 8$) could be made. The results are given in the next section.

4. Results

Two estimates of totals of selected person characteristics for October 1983 are given in Table 3. One estimate is based on 3 rotation groups that are first time-in-sample and the second estimate is based on one rotation group

that is second time-in-sample (see Table 1). The difference between two estimates and the standard error (S.E.) of the difference for each characteristic are also given in Table 3. It will be seen that two estimates by two times-in-sample are significantly different at the 5% level for only 3 labor force activity items. These three items are number of persons 16 years and over who were (1) absent 1 or more weeks without pay, (2) with job part of the month, and (3) spent time looking for work or on layoff. The three items are not independent. It must be noted that time-in-sample has no statistically significant effect on estimates for 12 earnings and benefit reciprocity items considered in this study. It should be mentioned that questionnaire designs and interviewing techniques are different for labor force activity items and income items. Labor force questions are asked independently at each interview (without reference to past responses), whereas, all sources of income are updated through an income update roster which relies on responses from previous interviews. This difference in interviewing methods may explain why some labor force activity items (but no earnings and benefit reciprocity items) show a significant difference.

Overall, only 3 out of 26 person characteristics considered in this study were significant at the 5% level. And all 26 items are not independent; some labor force items are correlated in the sense that a person could have more than one labor force characteristics during a month. Three tests out of 26 such tests at the 5% level could be significant by chance, when there is no time-in-sample bias. Thus, we conclude that SIPP monthly estimates for person characteristics were not significantly affected by time-in-sample biases.

Monthly estimates of eight selected household characteristics for October 1983 are given in Table 4 for all households. Only one estimate, number of households receiving public or subsidized rental housing, shows a significant effect of time-in-sample bias.

We have compared estimates from all consecutive interviews (i & $i + 1$, $i = 1, 2, \dots 8$). Results regarding the effects of time-in-sample for 2nd & 3rd, 3rd & 4th, ... 8th & 9th interviews are similar to results for 1st and 2nd interviews provided in Tables 3 & 4. Therefore, a summary of results from all consecutive times-in-sample (time i & $i + 1$, $i = 1, 2, \dots 8$) is given in Table 5. Note that a comparison of effects of two consecutive times-in-sample on a characteristic can be obtained from three months. For example, October, November and December of 1983 provide estimates of time 1 and time 2 (see Table 1). Monthly estimates and differences were computed for 26 person characteristics and 8 household characteristics (see tables 3 and 4). The average number of significant differences per month, and % significant differences are given in Table 5. It can be seen from Table 5 that for person characteristics for both sexes, 7.7% differences are significant between times 1 and 2, 12.7% between times 2 & 3 and between times 3 & 4, 2.5% between times 7 & 8 and 10.3% between times 8 & 9. For all eight time-in-sample

comparisons, 9.5% of the differences are significant for both sexes and 7.1% for males and 7.4% for females. Note that differences are significant mostly for labor force activity items which are not independent. For persons with earnings or receiving noncash benefits only about 2% of the differences are significant. For household characteristics, 6.2% of the differences are significant. Overall 9.5% of the 624 t-tests on time-in-sample differences for person characteristics were found to be significant. Given that the labor force characteristics are not independent and also, the eight successive time-in-sample differences are not independent, these results could occur by chance when no time-in-sample differences exist. We, therefore, conclude that SIPP monthly estimates for the 1984 Panel were not affected significantly by time-in-sample biases.

5. Summary

(1) Our analysis of the 1984 Panel data show that SIPP monthly estimates were not generally affected by time-in-sample biases. Only a few correlated labor force activity items were affected significantly.

(2) These preliminary results are encouraging but should be treated as tentative being based on only one panel data. Analysis of data from 1985+ panels would be necessary to evaluate the effects of time-in-sample bias on SIPP estimates.

(3) No direct comparison of effects of all nine interviews could be made from the 1984 Panel. Such comparisons can be made in the future using other administrative records e.g., IRS data and/or data from two or three SIPP panels.

(4) As SIPP data accumulate, studies of various ways in which survey conditions change with time and the impact of these changes on sample estimates would help in understanding and hopefully, controlling time-in-sample biases.

Acknowledgement: The authors would like to thank Dan Kasprzyk, Lynn Weidman, Larry Cahoon and Don Luery for their helpful comments, and Pat Ellis for typing the paper.

* 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.

References

- [1] Bailer, B.A. "The Effects of Rotation Group Bias on Estimates from Panel Surveys." *Journal of the American Statistical Association*, 1975, vol.70, p. 23-30.
- [2] Burkhead, D. "SIPP Waves 1,2 and 3 (partial): Month to Month Changes of Work Status and Income Reciprocity Status." Memorandum to John Coder, March 12, 1985.
- [3] Hanson, R.H. "The Current Population Survey Design and Methodology," Technical Paper 40. U.S. Department of Commerce, Bureau of the Census, January 1978.
- [4] McArthur, E. "Measurement of Attrition Through the Completed SIPP 1984 Panel; Preliminary Results" Memorandum to Daniel Kasprzyk. March 4, 1988.
- [5] Nelson, D., McMillen, D. and Kasprzyk, D., "An Overview of the Survey of Income and Program participation." SIPP Working Paper No. 8401, Washington, D.C.: U.S. Bureau of the Census, 1985.
- [6] Nelson, D.D., Bowie, C. and Walker, A. "Survey of Income and Program Participation (SIPP) Sample Loss and the Efforts to Reduce It". SIPP Working paper No. 8709. September 1987.
- [7] Roebuck, M.J., "SIPP Variance Estimation for 3rd Quarter 1983" Memorandum for Documentation, May 28, 1985. U.S. Bureau of the Census.
- [8] Short, K. and McArthur, E. "Life Events and Sample Attrition in the Survey of Income and Program Participation". Proceedings of the American Statistical Association, Social Statistics Section. 1986.
- [9] Singh, R.P. "SIPP: Recall Effect Study" Memorandum to Shapiro, October 2, 1986, U.S. Bureau of the Census.

TABLE 1. Number of Rotation Groups that Provides Estimates
by Month and Time-in-Sample: 1984 Panel

		Number of Rotation Groups								
Reference	Time-in-Sample									
Month	1	2	3	4	5	6	7	8	9	
<u>1983</u>										
June	1	0								
July	2	0								
August	3	0								
Sept.	4	0								
Oct.	3	1								
Nov.	2	2								
Dec.	1	3								
<u>1984</u>										
Jan.		4	0							
Feb.		3	1							
March		2	2							
April		1	3							
May			4	0						
June			3	1						
July			2	2						
August			1	3						
Sept.				4	0					
Oct.				3	1					
Nov.				2	2					
Dec.				1	3					
<u>1985</u>										
Jan.					4	0				
Feb.					3	1				
March					2	2				
April					1	3				
May						4	0			
June						3	1			
July						2	2			
August						1	3			
Sept.							4	0		
Oct.							3	1		
Nov.							2	2		
Dec.							1	3		
<u>1986</u>										
Jan.								4	0	
Feb.								3	1	
March								2	2	
April								1	2	
May								0	3	
June								0	2	
July								0	1	

Table 2. Time-in-Sample and Length of Recall for Reference Months

Reference Month	Rotation Group Number	Length of Recall in Months	Rotation Group Number	Length of Recall in Months
	First Time-in-Sample		Second Time-in-Sample	
Oct. '83	2	1	1	4
	3	2		
	4	3		
Nov. '83	3	1	1	3
	4	2	2	4
Dec. '83	4	1	1	2
			2	3
			3	4
	Second Time-in-Sample		Third Time-in-Sample	
Feb. '84	2	1	1	4
	3	2		
	4	3		
March '84	3	1	1	3
	4	2	2	4
April '84	4	1	1	2
			2	3
			3	4

Table 3. Monthly Estimates of Selected Person Characteristics by Time-in-Sample for October 1983

Person Characteristics	Time-in-Sample		Difference 1-2	S.E. of diff.
	1	2		
(Numbers in Thousands)				
BOTH SEXES				
Labor Force Activity Status:				
Total 16 years and over	175 880	176 290	-410	3 640
With some Labor Force Act.	115 060	112 710	2 350	2 427
With Job Entire Month	101 370	100 750	620	2 221
Worked Each Week	98 844	99 128	-284	2 258
Full-Time Worker	79 278	78 285	993	1 928
Part-Time Worker	19 566	20 844	-1 278	741
Absent 1 or more weeks w/o pay	2 530	1 625	905 *	238
Spent time on layoff	390	258	132	81
With Job Part of Month	4 013	2 277	1 736 *	246
Time looking for work or on layoff	2 251	1 130	1 121 *	205
No Job During Month	9 677	9 677	-0-	564
Looking for work/on layoff all mo.	8 659	8 599	60	487
Looking for work or on layoff part of month	1 018	1 087	-69	184
With no Labor force Activity	60 813	63 580	-2 767	2 190
Persons With Earnings:				
16 years and over	101 110	98 081	3 029	2 253
16 to 19 years	6 077	5 612	465	432
20 to 24 years	14 207	13 408	799	747
25 to 34 years	29 605	29 326	279	1 098
35 to 44 years	21 913	22 258	-345	791
45 to 54 years	15 360	14 708	652	696
55 to 64 years	11 260	10 546	714	516
65 years and over	2 692	2 222	470	254
Received Selected Sources of Noncash benefits:				
Medicaid:				
Total 16 years and over	10 421	11 187	-766	776
65 years and over	2 368	2 000	368	236
Medicare:				
Total 16 years or over	27 635	27 833	-198	1 049
65 years and over	25 297	25 630	-333	1 018

* Significant at 5% level.

Table 4. Monthly Estimates of Selected Household Characteristics by Time-in-Sample for October 1983

Household Characteristics	Time-in-Sample		Difference 1-2	S.E. of Diff.
	1	2		
(Numbers in Thousands)				
ALL HOUSEHOLDS				
Received Selected Sources of Noncash Benefits:				
Food Stamps	5 845	5 345	500	430
WIC	578	720	-142	158
Public or subsidized rental housing	2 382	1 633	749 *	319
Received Selected Source of money transfer payments:				
AFDC	2 043	2 023	20	277
Supplemental Security Income	1 982	1 927	55	210
Social Security or Railroad retirement Income	21 168	21 323	-155	718
VA Compensation or pensions	2 947	3 274	-327	263
Unemployment compensation	2 889	2 974	-85	249

* Significant at 5% level

Table 5. Significant Differences in monthly estimates due to Time-in-Sample Effects

CHARACTERISTICS	Time-in-Sample								
	1 & 2	2 & 3	3 & 4	4 & 5	5 & 6	6 & 7	7 & 8	8 & 9	All
PERSON CHARACTERISTICS									
BOTH SEXES									
Av. number of significant differences	2.0	3.3	3.3	2.0	2.3	3.0	0.7	2.7	2.5
% Significant differences	7.7	12.7	12.7	7.7	8.8	11.5	2.5	10.3	9.5
MALE									
Av. number of significant differences	3.3	2.7	2.0	0.3	1.0	2.3	1.0	2.0	1.8
% Significant Differences	12.7	10.3	7.7	1.2	3.8	8.8	3.8	7.7	7.1
FEMALES									
Av. number of significant differences	2.0	1.3	2.7	2.0	0.7	2.3	1.0	2.3	1.9
% Significant Differences	7.7	5.0	10.3	7.7	2.6	8.8	3.8	8.8	7.4
HOUSEHOLD CHARACTERISTICS									
All HOUSEHOLDS									
Av. number of significant differences	0.3	0.3	0.7	1.0	1.0	0.3	0.3	0.0	0.5
% Significant differences	4.2	4.2	8.3	12.5	12.5	4.2	4.2	0.0	6.2