PANEL CONDITIONING EFFECTS IN THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Steven G. Pennell and James M. Lepkowski, Survey Research Center, University of Michigan
Steven G. Pennell, Survey Research Center, University of Michigan, Ann Arbor, MI 48104

Key Words: Time in Panel Bias, Panel Survey, Measurement Error

The Survey of Income and Program Participation (SIPP) was initiated in late 1983 (the start of the 1984 panel) by the U.S. Bureau of the Census with the principal objective to provide policy-makers with more accurate and comprehensive information on income and participation in government programs than were available through other data sources. The survey results were intended to inform policy in the areas of welfare and tax reform, and improvement to entitlement programs such as Social Security and Aid to Families with Dependent Children (AFDC).

Interviews of panel members are conducted every four months for seven or eight consecutive interviews. A new sample or panel is introduced each year. The survey may thus be thought of as a rotating panel survey. An overview of the SIPP program is found in Nelson, McMillen, and Kasprzyk (1985).

In a rotating panel design such as the SIPP, it is possible to compare estimates based on the responses given by two different panels for the same time period. In the absence of nonresponse, such a comparison will indicate the presence of panel conditioning effects.

The literature concerning panel conditioning has recently been strengthened by a number of carefully designed studies that have been able to identify changes in aggregate responses given by panel members as the number of interviews increases. The causes of the observed panel conditioning are not known. It may be due to a change in the behavior of the respondent as a result of panel membership or a change in how respondents answer the same question over time.

Whatever the mechanism at work, panel bias has been observed in a variety of settings. Kalton, Kasprzyk, and McMillen (1989) review a number of these, including studies of consumer expenditure (Neter and Waksberg, 1964, 1965; Silberstein and Jacobs, 1989), social attitudes (Lievesley and Waterton, 1985), medical care expenditures (Cohen and Burt, 1985), political attitudes (Traugott and Katosh, 1979), and labor force characteristics (Bailar, 1975). Despite this long list of studies where panel bias has been observed, there is no consensus about the inevitability of the effect, or its size. In the same panel surveys where panel conditioning has been found for some items, it is small or absent for others.

This present investigation was designed to examine the presence and nature of panel conditioning in the SIPP. Panel conditioning was assessed by comparing estimates from three successive SIPP panels: 1985, 1986, and 1987. Issues of attrition nonresponse were an important consideration in the design of the investigation. If nonresponse patterns differed across these panels, direct comparison of estimates might be confounded with bias due to attrition and other nonresponse. One way to assess the presence of this bias would be to obtain objective, record based information for panel members, and compare record based information for responding and nonresponding panel members. Such data are expensive to obtain and were not available for this investigation. Instead, other indirect methods were employed to control for this potentially confounding feature of any panel comparisons that might be made to assess panel conditioning.

The next section presents the comparison of wave nonresponse patterns for the three panels. Then comparison of 1986 calendar year estimates obtained from the 1985 and 1986 panels are presented, followed by an examination of panel conditioning for selected reference months for a number of characteristics using weighted monthly estimates for panel members responding about the reference month. Findings and conclusions for an identical investigation limited to persons who were full seven wave respondents for each panel are also presented. The report concludes with a brief review of implications of the investigations.

Wave Nonresponse Patterns for the 1985-7 SIPP Panels

A previous investigation of the 1984 SIPP panel (Lepkowski, Kalton, and Kasprzyk, 1988) produced eight wave patterns of nonresponse, controlling for removals from the population (e.g., deaths, joining the Armed Forces) as part of a larger examination of methods for compensating for panel nonresponse. That investigation showed that over eight waves of the 1984 panel over 70% of the persons to be interviewed at wave 1 provided data for all eight waves. The purpose of conducting a reexamination of wave response patterns in this research is two-fold. First, a comparison of wave response patterns across successive panels can indicate improvements or declines in SIPP field procedures. Second, changes in the relative frequency of wave response patterns could indicate potential problems for subsequent investigations of panel conditioning. Since panel conditioning analysis will necessarily involve comparisons of successive panels, different relative frequencies of wave response patterns across panels, and any wave nonresponse bias associated with those relative frequencies, could confound the examination of panel conditioning.

The data for all of the investigations reported here come from the Full Panel Microdata Research File for each panel year. Definitions and procedures used for the 1984 panel investigation

Table 1
Frequency of panel and non-panel wave nonresponse patterns in the 1985-7 SIPP panels

Pattern Group	1985 Panel		1986 Panel		1987 Panel	
	No.	%	No.	%	No.	%
PANEL MEMBERS	23300	76.1	17937	77.5	24448	79.5
A. Responded All Seven Waves	22568	73.7	17299	74.7	23653	76.9
B. Some Nonresponse Due to Ineligibility	732	2.4	638	2.8	795	2.6
NON-PANEL MEMBERS	7305	23.9	5220	22.5	6321	20.5
Attrition Nonresponse Only	4835	15.8	3328	14.4	3887	12.6
C. At Wave Two	1553	5.1	1150	5.0	1453	4.7
D. At Wave Three	1033	3.4	672	2.9	713	2.3
E. At Wave Four	855	2.8	493	2.1	571	1.9
F. At Wave Five	562	1.8	463	2.0	480	1.6
G. At Wave Six	404	1.3	256	1.1	338	1.1
H. At Wave Seven	428	1.4	294	1.3	332	1.1
Interim Nonresponse Only	1737	5.7	1329	5.7	714	5.6
I. Single Wave	1376	4.5	1068	4.6	1323	4.3
J. Two Waves	274	0.9	216	0.9	298	1.0
K. Three to Five Waves	87	0.3	45	0.2	93	0.3
L. Other Patterns of Nonresponse	733	2.4	563	2.4	720	2.3
TOTAL	30605	100.0	23157	100.0	30769	100.0

were altered for the 1985-87 panel research. These altered definitions prevent direct comparisons across the first four panels of the SIPP. In these analyses all 100-level persons listed in the household at wave one comprise the sample for each panel, including type Z noninterviews at wave one. Excluded from both the 1984 panel investigation and the present investigation were a small group of 100 level persons who were apparently in the household at wave one, but mistakenly not listed. These unlisted persons appear as 100 level persons in later waves but they have no defined interview status code at wave one.

In order to establish comparability across panels, only selected data could be used from several of the panels. Wave eight of the 1985 panel could not be used since the 1986 and 1987 panels had only seven waves. In addition, the 1985 panel was subject to a sample size reduction in February, 1986. We have dropped from this analysis all persons in the sample segments who were cut from the sample in the 1986 sample reduction. Finally, rotation group one of the 1986 panel was also dropped from these comparisons. By design, this rotation group was to be interviewed only six times. Rather than reduce the comparisons to six wave patterns to accommodate this one rotation group in one panel, we decided to drop the rotation group from the 1986 panel.

Patterns of nonresponse for each wave are outlined in Table 1. The three panels have similar proportions of panel members who became ineligible because they died, entered an institution or live in armed forces barracks, moved abroad or were ineligible for some other reason. What differs across the panels is the relative frequency of nonresponse, and hence response in all seven waves. There is a steady increase in the proportion with seven waves of response from 1985 to 1987, indicating steady improvement in the ability of SIPP interviewers to retain original sample members in the panel. The 1987 panel is the first panel to have over three quarters (76.9%) of the original sample members respond at all seven waves.

Turning to the non-panel members (i.e., original sample members who had one or more nonresponding waves; groups C through L), over sixty percent had attrition patterns. This proportion declined from 67% in the 1985 panel to 62% in the 1987 panel reflecting better retention of original sample members in the panel. More than thirty percent of the attrition nonrespondents began their non-interviews at wave two. While the proportion of attrition nonresponse has steadily decreased from the 1985 to 1987 panels, the concentration of

attrition nonresponse at wave two has increased from 32% in 1985 to 37% in 1987. Thus, the improvements in panel retention have come through efforts to reduce panel attrition at later waves of the panel.

Approximately one quarter of nonrespondents are interim nonresponders. As the proportion of non-panel members with attrition patterns has decreased the proportion with interim non-interviews has increased from 24% in 1985 to 27% in 1987. The percentage of interim nonresponders who were non-interviews for one wave, and one wave only, has fluctuated at around 80%. Approximately fifteen percent of interim nonresponders are non-interviews at exactly two waves and only three to six percent are interim nonrespondents in three or more waves.

Comparison of Calendar Year Estimates

Weighted 1985 and 1986 calendar year estimates were compared for income recipiency, income amounts, health insurance coverage, and labor force participation for all sample persons, blacks, non-blacks, males, females, and persons ages 15-24, 25-44, 45-64, and 65 and older. Estimates from each panel were compared using a two sample t-test, assuming independence between panels.

For each income characteristic, recipiency in any month, recipiency in all 12 months, mean amount received by recipients, and the distribution of the amount received were compared. Recipiency for all months, but not for any month, was examined for income from asset types 100-103 and 104-107, and means and distributions were also compared for these two income types. Health insurance coverage in any and all months was compared as was labor force participation in all months.

Approximately 25,000 individuals from each panel with positive 1986 calendar year weights in the 1985 and 1986 Full Panel Microdata Research Files were used for this investigation. In cases where values were imputed, the imputed values were used in estimates. Standard errors for each estimate were computed taking into account the complex sampling design using the stratum and half-sample code provided with the Full Panel Research files. Sampling error estimation used a Taylor series approximation for the means and proportions.

Few differences were observed between panels. In fact, given the large number of tests conducted, it was surprising so few p-values less than 0.05 were observed. No highly significant ($p < 0.001$) differences were detected. Only a handful (eight in total) of the more than 110 comparisons were statistically significant at the .05 level.

The conclusion is that, at least at the level of aggregate estimates, there is little evidence that the 1985 and 1986 SIPP panels are providing different estimates for 1986. This occurs despite the fact that the 1985 panel was being interviewed for the 4th, 5th, 6th, and 7th times, while the 1986 panel was being interviewed for the 1st, 2nd, 3rd, and 4th times. There is little evidence of panel conditioning present when calendar year estimates are compared.

Panel Conditioning in the 1985-7 SIPP Panels

Panel conditioning analyses were also conducted for income recipiency, income amounts, health insurance coverage, and labor force participation. Two sets of panel comparisons were made. The 1985 and 1986 panels were compared for January, May, and September in 1986 when the 1985 panel respondents were being interviewed for the 4th, 5th, and 6th times, respectively, and the 1986 panel respondents were being interviewed for the 1st, 2nd, and 3rd times, respectively. Similarly, the 1986 and 1987 panels were compared for January, May, and September in 1987 when the 1986 panel respondents were being interviewed for the 4th, 5th, and 6th times, respectively, and the 1987 panel respondents were being interviewed for the 1st, 2nd, and 3rd times, respectively.

Approximately 25,000 original sample persons (i.e., 100 level persons) with positive monthly cross-sectional weights were included in the analyses. Comparison of monthly aggregate responses allowed the estimation of differences between panel reports. Panel conditioning was studied by comparing two sets of overlapping panels, the 1985 and 1986 panels and the 1986 and 1987 panels.

A linear model $X\beta$ was fitted to the observed 8 X 1 vector of logits, F, across the four months and two panels. For recipiency, the model for any given reference month is of the form

$$\lambda_{ij} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij}$$

where λ_{ij} denotes the logit of persons receiving a particular type of income, α_i denotes the effect of the ith month (i = 1, 2, 3, or 4), γ_j denotes the effect of the jth panel (j = 85, 86, or j = 86, 87), and $(\alpha\gamma)_{ij}$ denotes the interaction between month and panel. Computation of an estimate of the parameter vector β used weighted least squares estimation:
$b = (X' V^{-1} X)^{-1} X' V^{-1} F$. Hypothesis tests about the parameters in β were then formulated as $H_o : C\beta = 0$. A quadratic form was used to compute a test statistic with an asymptotically chi-square distribution, with degrees of freedom corresponding to the rank of the matrix C. Finally, based on results of the hypothesis testing, a reduced model Xb was fit to the observed vector of proportions F. The lack of fit statistic
$Q = (F - Xb)' (X' V^{-1} X)^{-1} (F - Xb)$
has an asymptotically chi-square distribution, and was used to test the statistical significance of the fit of the reduced model to the data.

Table 2 summarizes the analyses of panel conditioning in the SIPP for three of the sixteen characteristics examined. For each comparison of two panels, tests were conducted for 3 overlapping

Table 2
Summary of p-values Less Than 0.10 for Panel (P) and Month (M) Effect on Selected Characteristics for January, May, and September 1986 and 1987, 100-Level Persons in the 1985 - 1987 SIPP Panels

Type of effect	1985/1986			1986/1987		
	Jan	May	Sep	Jan	May	Sep
Wages and Salary						
P	.026	.058	...	.095	...	...
M	...	...	...	...	...	...
P × M	...	...	...	...	...	...
Self Employment						
P	...	...	...	...	...	...
M	.047	...	...	...	...	...
P × M	.067	...	...	...	...	...
State Unemployment						
P	...	...	...	.005	.029	.025
M	...	.003	...	...	...	...
P × M	...	...	...	...	...	...
In the Labor Force						
P	.002	.079	...	...	.063	...
M	...	...	...	...	...	...
P × M	...	.066	...	...	...	...
Covered by Health Insurance						
P	...	...	...	...	...	...
M	...	...	...	...	...	...
P × M	...	...	...	...	...	...

reference months: January, May, and September. For each reference month, a linear model was fit to the monthly aggregate percentage or mean which included a factor to distinguish the panel [P] (1985 or 1986 for one pairing, 1986 or 1987 for the other), the month [M] of interview (e.g., February, March, April, or May for the January reference month), and an interaction between panel and month [P × M].

The results shown for selected characteristics in Table 2 are similar to the results for the other characteristics which are not shown. Given the large number of hypothesis tests conducted (over 1,000), caution should be exercised in interpreting the nature of the effects identified. By chance alone we would expect 1 in 10 tests to be statistically significant at the 0.10 level, but fewer than expected were. If only the panel effect hypothesis tests are examined, there is a surprising lack of statistically significant findings. There is some limited evidence of panel conditioning for a few items for a single panel comparison. For example, state unemployment compensation recipiency is higher for the 1987 panel than the 1986 panel in all three reference months. However, there is no statistically significant difference between the recipiency estimates for the 1985 and 1986 panels. If the 1986/1987 comparison is indicative of a panel conditioning effect, it appears elusive when the other panel comparison is examined.

A summary table for mean income amounts (not shown here) had results that are similar to those in Table 2. The general impression of these summary results is an almost total absence of any panel differences.

One significant methodological deficiency in these comparisons is the control of attrition nonresponse. The monthly weights employed in all comparisons have an implicit adjustment for

attrition nonresponse. However, the effectiveness of the adjustment for attrition nonresponse is unknown. Failure to detect differences between panels could be due to differences in panel attrition. If persons lost to nonresponse are different from those who remain with respect to the characteristics examined, and if the weights are not effective in compensating for these differences, nonresponse bias could hide a panel conditioning effect.

Two alternative approaches which address the nonresponse bias issue were followed for a subsample of original sample persons consisting of those who gave data for the characteristic of interest in each of the 4th, 8th, 12th, 16th, 20th, and 24th months of each panel. That is, we investigated the presence of panel bias among those persons who responded at all seven waves of the panel. To the extent that the same types of persons (and characteristics) are full seven wave respondents in two panels, the comparisons given here control for any panel attrition bias through sample selection. Analyses were conducted with and without the full panel longitudinal weight which compensates this group for nonresponse. Overall, the comparisons among seven wave respondents and those with positive monthly cross-sectional weights were similar. In almost every analysis repeated for the seven wave respondents, the panel conditioning effect is the same or weaker than for the previous analysis. Further, the results are virtually identical when the panel weight is not used.

The conclusions from these extensive tests are almost identical for those performed on a sample with attrition nonresponse: there are almost no statistically significant panel effects present. The expected number of statistically significant tests at the $\alpha = 0.10$ level is higher than the actual number of statistically significant results in the table. There is slight evidence for a panel effect in Social Security, Supplemental Security Income (operating through the interaction between month and panel), WIC, and G2 100 income recipiency for the 1985 to 1986 panel comparisons. However, this effect was not present for the 1986 to 1987 comparison. We find this puzzling, and believe that the effects observed in the 1985 to 1986 panel comparison are not true panel conditioning effects.

For the sake of brevity, we do not present a table of the p-values for mean amount comparisons. The analysis was similar to that reported in the previous section. There is slight evidence of a panel conditioning effect only for wage and salary and WIC income amounts. As was the case for recipiency, those results are significant only for the 1985 to 1986 panel comparison, not for the 1986 to 1987 comparison.

Implications for the SIPP

Panel conditioning was investigated through three examinations of the differences between estimates for two successive panels: 1986 calendar year estimates, single reference month estimates computed using cross-sectional weights, and single reference month estimates for individuals who responded on all seven waves of the study. In the comparison of calendar year estimates, the criterion for establishing panel conditioning effects was the presence of a difference between the two estimates. In the latter investigations of monthly estimates, the criteria for establishing panel conditioning effects was the presence of a statistically significant difference across reference months and for both sets of comparisons. One of the primary reasons for the overlapping panel feature of the SIPP is that attrition nonresponse and panel conditioning bias over repeated waves of interviewing may effect panel results. To the extent that there are detectable differences between panel estimates for the same time point, panel data combined from overlapping panels will be less subject to the bias due to attrition and panel conditioning.

There is another feature of the SIPP design that is driven by the desire to reduce the effects of attrition nonresponse and panel conditioning: the number of panel waves. The more waves of interviewing, the more attrition nonresponse reduces sample size, and, potentially, the greater the effects of panel conditioning on interview reports.

The results of this investigation indicate that there is no evidence that attrition nonresponse and panel conditioning create panel differences in the SIPP under the current seven wave design. Three different types of comparisons showed few statistically significant differences between panel estimates. The few differences found were not consistent across panel comparisons. That is, statistically significant results for the 1985 and 1986 panel comparisons could not be replicated for the 1986 and 1987 panel comparisons.

We do not conclude from this that panel conditioning is totally absent. Given the sample sizes in each panel comparison, the power to detect very small relative differences in recipiency rates or income amounts is limited. With a large enough sample size, and differences as large as observed for some of the characteristics examined here, one might be able to establish the presence of a panel conditioning effect. However, power considerations are not important to this investigation. This was not an experiment in which differences of a fixed magnitude were specified in advance. The observational nature of the study only allows us to conclude that there are no statistically significant differences between panels.

This conclusion leads us to a recommendation about the redesign of the SIPP. If there are no detectable differences between panels at current sample sizes, then there is no statistical reason for maintaining overlapping panels at present sample sizes. The 1985 panel provides estimates for January, May, and September, 1986, or for the 1986 calendar year that are statistically indistinguishable from estimates obtained for the same time periods using the 1986 panel.

Differences in the number of previous waves have no apparent effect on the estimates obtained.

This recommendation is not without it limitations, though. If the elimination of overlap means the introduction of a single panel with twice the sample size of these current panels, and if one were to conduct a methodological study with a test panel of the same doubled sample size, more panel differences will be found to be statistically significant simply because of the increased sample size. Completely abandoning panel overlap would eliminate the possibility of being able to test for panel differences in the future. Some overlap, particularly at later waves of a panel, should be considered in order to provide future opportunities to examine panel conditioning under larger sample sizes than available for this investigation.

An additional recommendation concerns the number of waves in a panel. There are two pieces of evidence presented here that suggest that more waves, and thus longer observational periods, may be statistically warranted for the SIPP. First, there is no apparent effect of attrition nonresponse or panel conditioning on estimates when the sixth or seventh wave of one panel is compared to the third or fourth wave the next panel, or even the first wave of yet a third panel. Projecting this observed similarity between panels beyond the seventh wave is going well beyond the available data. Nonetheless, the observed similarity with the sample sizes available suggests that additional waves of data collection may not lead to significant differences due to attrition nonresponse or panel conditioning.

Secondly, we also observed declining rates of attrition nonresponse in this study. Supposing that the 15% rate of attrition nonresponse in the 1985 panel after seven waves is "acceptable," then the 1987 panel with its 12% rate could possibly have run for another two to three waves before the "acceptable" rate of 15% attrition nonresponse were observed, provided that the rate of attrition nonresponse on subsequent waves remained at about the 1% level. Of course, the attrition nonresponse rate for the seventh wave of the 1985 panel may not be "acceptable," and efforts were made to improve the rate. Regressing to this rate may therefore be unacceptable. What would be of interest is to know whether the attrition nonresponse rate continued to decline with subsequent panels. If so, there may be a point at which additional attrition nonresponse because of additional waves of data collection will be unacceptable for the SIPP.

Both of these arguments favor a longer panel tenure for the SIPP. Both of these arguments make assumptions that go beyond the available data. It would be unwise to increase the number of waves based solely on these arguments. The cross-sectional objectives of SIPP estimation must be balanced against the longitudinal objectives. On the other hand, additional waves of data collection will provide longer follow-up periods for analysts to study length of spells of program participation and income recipiency.

Finally, from one point of view it is unfortunate that statistically significant panel differences were not observed. An investigation of the possible causes of panel conditioning from these results is impossible with the weak evidence for panel conditioning effect presented here.

References

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

Cohen SB and Burt VL (1985). Data collection frequency effect in the National Medical Care Expenditure Survey, Journal of Economic and Social Measurement, **13**: 125-151.

Computer Support Group (1988). OSIRIS IV User Guide, 8th edition. Ann Arbor, Michigan: Institute for Social Research.

Kalton G, Kasprzyk D, and McMillen DB (1989). Nonsampling errors in panel surveys, in Panel Surveys, D Kasprzyk, G Duncan, G Kalton, and MP Singh, ed. New York: J Wiley and Sons, Inc.

Lepkowski, J, Kalton, G, and Kasprzyk, D. "Weighting Adjustments for Partial Nonresponse in the 1984 SIPP Panel," Proceedings of the Survey Research Methods Section, American Statistical Association, 1988.

Lievesley D and Waterton J (1985). Measuring individual attitude change, in British Social Attitudes: The 1985 Report, R Jowell and S Witherspoon, ed. Aldershot, England: Gower Publishing Company, pp. 171-194.

Nelson D, McMillen DB, and Kasprzyk D (1985). An Overview of the Survey of Income and Program Participation: Update 1, SIPP Working Paper Series No. 8401, US Bureau of the Census, Washington DC.

Neter J and Waksberg J (1964). A study of response errors in expenditures data from household surveys, Journal of the American Statistical Association, **59**: 18-55.

Neter J and Waksberg J (1965). Response Errors in Collection of Expenditures Data by Household Interviews: An Experimental Study, Bureau of the Census Technical Report No. 11. Washington, D.C.: U.S. Government Printing Office.

Silberstein AR and Jacobs CA (1989). Symptoms of repeated interview effects in the Consumer Expenditure Survey, in Panel Surveys, D Kasprzyk, G Duncan, G Kalton, and MP Singh, ed. New York: J Wiley and Sons, Inc., pp. 289-303.

Traugott MW and Katosh JP (1979). Response validity in surveys of voting behavior, Public Opinion Quarterly, **43**: 359-377.