Survey of Income and Program Participation

[redacted]

8921 98

Nathan Young
The Urban Institute

September 1989

ACKNOWLEDGEMENTS

Funding for the research presented in this paper was provided by the Bureau of the Census. Opinions expressed in this paper are those of the author and do not necessarily represent the views of The Urban Institute or its sponsors. The author wishes to acknowledge the assistance given by readers Daniel Kasprzyk, Sharon Long, and Patricia Ruggles, and by programmer Taube Wilson.

TABLE OF CONTENTS

TABLES

INTRODUCTION

A seam is the point at which two separate pieces of material abut. In manufacturing processes, the quality of the product is often dictated by the quality of seams. Longitudinal surveys also have seams, and as with manufactured products, these seams are critical elements in determining the overall quality of the survey. The seams in a longitudinal survey occur in places where data that cover two different interview periods (or waves) are brought together. The SIPP because it gathers information for four months at a time, at four-month intervals, has many "wave-seams." [redacted] The dependency of responses on the change in waves is called the wave-seam effect.

The wave-seam effect is made clearer by the following specific example of households interviewed in February and June of 1985. In February, interviewers would collect personal earnings information for the months of October, November, December, and January. In June, interviewers would collect personal earnings information for the months February, March, April, and May. Reported personal earnings would be more than twice as likely to change between January and February than they would be between any other pair of consecutive months. This is because data for January is collected during one interview, and data for February is collected during the next interview.

Where wave-seam effects are large, questions are raised about whether statistical inferences based on the SIPP are misleading. These questions are not unique to the SIPP, as repeated measurements in surveys often generate different responses. The problem of the SIPP is not necessarily that there is more intrinsic response error. Rather the problem of the SIPP is that there are more repeated measurements.

This study addresses data-quality in longitudinal surveys from a new perspective. In studies of cross-sectional data quality in the SIPP (King et. al, 1987; Dalrymple and Carlson, 1986) the SIPP has been shown to concur with other cross-sectional surveys. In studies of rates of change in the SIPP (King et. al, 1987) mean rates of change have been shown to exhibit large wave-seam effects. In this study of the SIPP another question is addressed. Given that response behavior in the SIPP is influenced by timing of the interviews, it is important to ask: How valid are inferences based on intertemporal associations in the SIPP? In answering this question we show that dynamic analysis in the SIPP, while not immune to the wave-seam effect, is not crippled by it.

This paper is divided into two parts. In the first part we discuss our direct findings from a descriptive study of the wave-seam effect. The first section demonstrates directly that while absolute rates of change are strongly influenced by the wave-seam effect, that the correlations of the rates of change are not. In

the second part of we discuss some of the most plausible sources of the wave-seam effect in the SIPP and their implications for analysis. In particular, we discuss the constant wave response, a form of response bias (measurement error) that may be present in many retrospective surveys, but which will be revealed only in surveys like the SIPP that are both retrospective and longitudinal.

TRANSITION FLAGS

To conduct dynamic analysis of different data types (ratio, interval, ordinal, and categorical) the data was transformed into transition flags. A transition is any reported change in the same variable between two time periods that exceeds a pre-determined critical value. Transition flags can be created for any data measured at two points in time. We follow the convention of assigning 1 to the transition flag when the variable's value changes significantly and assigning 0 to the transition flag when the variable's value did not change or changed by an insignificant amount.

SAMPLE

The source of data in this paper is the SIPP 1984 Full Panel Longitudinal Research File. The Full Panel Longitudinal Research File contains entries for 32 months for every individual, of any age, who was a member of an interviewed household anytime between June 1983 and April 1986. The variables on this file ~~are~~ refer primarily to individual behavior or characteristics. To identify family income of various types it was necessary to sum together the amounts reported under individual family members for each month. This task involved identifying key variables on the record, splitting the person records up into person-month records, sorting the person-month records into families, doing the aggregation to create the family level variables and then remerging the data back onto the individual person month records.

The analysis sample was restricted to a random subsample of the individuals who were 18 years or older at the time of the first month of data, and for whom information was collected in all eight waves (32 months of data on 5134 individuals).

DEFINITION OF VARIABLES

This paper focuses on the following seven variables that were reported each month:

- Marital status
- Employment status
- Amount of personal earnings
- Amount of total family income
- Amount of individual social security income
- Receipt of AFDC (welfare) income
- Receipt of family food stamp benefits

The seven variables were chosen to represent the wide variety of variables available on the SIPP. Because of the primary objectives of the Survey of Income and Program Participation, five of the variables chosen represented responses to income and program participation questions. The actual variables were not examined, instead the analysis used created transition flags alone. For marital status, employment status, social security income, AFDC receipt and food stamp benefits, transition flags were generated by any change in consecutive months. For total family income and personal earnings, transition flags were generated only if in two consecutive months the higher income exceeded the lower income by more than a ratio of three to two, where the ratio of three to two was chosen to accommodate three paychecks in one month and two in the other.

WAVE-SEAM EFFECTS IN TRANSITION RATES

In Table 1, the magnitude of the wave-seam is revealed by comparisons of the transition rates occurring in the seam month with those occurring in the remaining months. The seam month transitions refer to changes that are reported for consecutive months within two separate interview periods, the on-seam transitions. The remaining months refer to changes that are reported for consecutive months sharing the same interview period, i.e.,the off-seam transitions.

Transition rates display large wave-seam effects. For all the variables in table 1, the proportion of transitions occurring in the seam month is over 40 percent, for social security income transitions it is nearly 70 percent. [redacted] month [redacted].

Table 1
Mean Transition Rates, by Wave Month

	Month 1 (%)	Month 2 (%)	Month 3 (%)	Month 4 (%)
Variables				
Marital status	0.7	0.3	0.3	0.3
Employment status	10.2	4.6	5.0	5.4
Personal earnings	16.3	5.5	6.3	6.4
Total family income	17.9	4.9	5.4	5.7
Individual social security	12.0	1.9	1.6	1.8
Family AFDC receipt	0.3	0.1	0.1	0.1
Family food stamp receipt	0.9	0.2	0.4	0.2

source: SIPP 1984 Full Panel Longitudinal Research Panel

WAVE-SEAM EFFECTS IN TRANSITION CORRELATIONS

In addition to the transition rates in table 1 we constructed twos correlation matrices, shown in table 2. Measures of correlation are key elements in multivariate statistical analysis, used by themselves and as computational building blocks for more complex procedures such as multiple regression, analysis of variance, discriminant analysis, and factor analysis. Correlation may be used directly to construct causitive models; for example, correlations may be used as support for a model where marital separation is a cause for individuals entering welfare. Correlation may also be used indirectly to compute regression coefficients in multivariate models; for example when multiple causes for welfare entries are considered. Correlations between variables will be lower when response error is present, except in the case where response errors are themselves associated.

[redacted]

off-seam correlations are small, then one or the other or both are wrong. Depending on the hypothesized response error mechanism, one can further suggest that identical correlations are both necessary and sufficient to show that the wave-seam effect does not bias correlation estimates. As is demonstrated in table 2, the extent of the bias is significant statistically. Pairs of variables sharing large transition correlations in the same interview period share large transition correlations across interview periods. Pairs of variables sharing small transition correlations in the same interview period share small transition correlations across interview periods.

Comparison of correlations for transitions reported in wave month one and those reported in wave months two through four reveal that transition correlations are less systematically biased than are the transition means. While there is a tendency for on-seam correlations to be lower than the off-seam correlations, this tendency is not universal. For example, the correlation between employment status and personal earnings is .380 for on-seam months and .486 for off-seam months but the correlation between employment status with total family income is .172 for on-seam months and only .114 for the off-seam months.

Correlations between transitions are normally positive. Changes tend to occur together. Negative correlations are uncommon except, as in the case of social security income and all other variables, where the population with a high likelihood of one type of transition is substantially different than the population with a high likelihood of the other.

[redacted]
The on-seam correlation for these variables is no more than 0.02 outside the range of the correlations observed for the other months.

Variables with large mean transition rates -- employment status, personal earnings, and total family income -- also display large differences in on-seam and off-seam correlations. This behavior is consistent with at least two of the most plausible sources of the wave-seam response discussed below.

Overall, the correlations in table 2 clearly display some wave-seam effects though the severity of the effects is slight. Collecting information at two points in time should result in different responses than collecting information retrospectively. In survey research response errors are unavoidable and the SIPP is no different in this respect than any other survey. What is striking about comparisons of the on-seam and off-seam correlations is that the difference is as small as it is. Focusing on the difference in the transition rates alone, it would appear that response error dominates the true process. This is because transition rates are exactly the aspect of the distribution most affected by the wave-seam effect. [redacted]

FALSE PRECISION

One applied principle of measurement that this study reconfirms is that false precision in measurement may lead to weak and biased results. An operational definition of transition flags is dependent on what is the magnitude of the change criteria. Any difference in response errors greater than the transition criteria can generate a transition flag. If changes in response errors are common but of small magnitude, then a smaller percentage of transition flags will be generated by response error alone if the critical change required to set the transition flag is larger. Table 3 demonstrates that both the size of the correlations is larger, and the degree of wave-seam effect is smaller when changes in income are measured with stricter change criteria. Using a small critical change criteria is equivalent to analyzing data with false precision.

SOURCE OF THE WAVE-SEAM EFFECT

What causes the wave-seam effect in the SIPP? Three explanations are plausible. First, respondents may give an answer for earlier months in an interview period, identical with the answer they give for the most recent month or their current state. For example, a person may just report their current activities in the reference

month of April, for the entire wave December through March. This type of response is called a constant wave response. It can be shown that a constant wave response will tend to increase measured transition correlations both on and off the seam by a small amount, with the greatest impact on variables with large numbers of overall transitions and variables with low correlations.

A second possible source for the wave-seam effect is the variation in how respondents understand a question, i.e. simple response bias. Changes in understanding of questions from interview to interview will generate spurious transitions. As long as the variation in understanding is not correlated with family attributes or events, simple response error will tend to reduce correlations of transitions on the seam.

A third possible source for the wave-seam effect is underreporting. For example, if food stamp recipients accurately report that they receive food stamps for one interview period but in the subsequent interview period neglect to report food stamp receipt a spurious transition is generated. Once again overall numbers of transitions may be overestimated. If the underreporting is uncorrelated across variables than the correlations of transitions on the seam will be underestimated. If the underreporting is itself correlated than the correlations of transitions off the seam may be overestimated

THE CONSTANT WAVE RESPONSE

The constant wave response mechanism induces some interesting biases in dynamic analysis. A simple theoretical model of this response mechanism can be constructed using the following assumptions:

- A proportion, π, of the SIPP respondents report that no change occurred in the entire retrospective period in the wave. This is true simultaneously for all variables.
- The reported value for all variables reflects what the respondent would normally report as their current state.
- The proportion π is independent of household characteristics and dynamics.

Some of the consequences of this response error mechanism are obvious. Average monthly transition rates in an interview period are reduced in proportion with π. Average monthly transition rates between interview periods are increased by a factor that is composed of the proportion π times the difference in four month and one month net transition rates.

The theoretical relationship between the constant wave response mechanism and the relative bias in transition rates can be used to construct estimates of π. These estimates assume that the only

reason for the seam effect is a constant wave response satisfying the three assumptions discussed above. It is easy to show that under these assumptions that estimates of π should be identical for different variables. The estimates of π presented in the first column of table 4 are not. Five of the estimates of π lie in the range of 44 to 54 percent. The two estimates of π associated with marital status and social security most assuredly do not. The constant wave response is an incomplete explanation for the wave-seam effect.

Even though the constant wave response is an incomplete explanation for the wave-seam effect, we thought it worthwhile to explore the consequences of the constant wave response on measurements of mean transition rates. In column two of table 4, we present the percentage that mean transition rates are biased under a constant wave response scenario, using the estimates of π that were reported in column one. Because measured transition rates should be lower within an interview period and measured transition rates should be higher between interview periods, the biases in transition rates for the whole SIPP sample are partially offsetting. Therefore under a constant wave response hypothesis the proportion of bias in transition rates ranges from only 5 percent for marital status to 40 percent for social security.

TRANSITION RATES AND THE CONSTANT WAVE RESPONSE PHENOMENON

The properties result from randomly mixing a dynamic process sampled at one-month and four-month intervals. Sample statistics for both on-seam and off-seam transition samples thus reflect four-month net transitions as well as true one-month transitions.

First consider how mean transition rates are affected. Observed transition rates for the seam months (m_s) and for the off-seam months (m_o) are determined by three factors: underlying one-month transition rates μ_1, underlying four-month transition rates μ_4, and the probability that a respondent responds with answers identical to their current or most recent status across all earlier months in a wave (π).

In these circumstances π proportion of off-seam months will have no transitions because of the constant wave-seam response, and $(1-\pi)$ proportion of off-seam months will report their normal transition rate μ_1. Averaging across constant wave seam respondents and regular respondents measured off-seam transition rates would be:

$$m_o = \pi*0 + (1-\pi)\mu_1 \quad \text{or} \quad m_o = (1-\pi)\mu_1 \qquad (1)$$

Similarly π proportion of on-seam months will have four month transitions because for that proportion of respondents the first month in their second wave will actually reflect their fourth month's state. For the remaining $(1-\pi)$ respondents, the measured

transitions will reflect actual behavior. Therefore measured on-seam transition rates would be:

$$m_s = \pi * \mu_4 + \pi * \mu_1 \qquad (2)$$

From observed values of m_o, m_s and μ_4 we can solve for π for different variables and that was the method used to derive table 4.

TRANSITION COVARIANCES AND THE CONSTANT WAVE RESPONSE PHENOMENON

Transition covariance structure under the constant wave response assumptions is only slightly more difficult to analyze. However now we consider not only the bias for mean transition rates but also the bias for mean transition cross-products. Let u and v be any random vectors with finite first and second moments, and w = u with probability π and w = v with probability $(1-\pi)$. Then $E(ww') = \pi E(uu') + (1-\pi)E(vv')$ and $E(w) = \pi E(u) + (1-\pi)E(v)$. Using the standard decomposition of variance into cross-product form:

$$V(w) = E(xx') - (Ex)(Ex)'$$

we have the result:

$$V(w) = \pi V(u) + (1-\pi)V(v) + \pi(1-\pi)(Eu-Ev)(Eu-Ev)' \qquad (3)$$

First consider the case where w represents a vector of measured off-seam transitions, w_0. Because of the constant wave response, π proportion of the respondents will answer with zero transitions, and $(1-\pi)$ will answer with normal one-month transition behavior. Letting u equal the zero transition vector, v equal the normal one-month transitions (w_1) we have:

$$V(w_o) = (1-\pi)V(w_1) + \pi(1-\pi)(Ew_1)(Ew_1)' \qquad (4)$$

Next consider the case where w represents a vector of measured on-seam transitions, w_s. Because of the constant wave response, π proportion of the respondents will answer with their four-month net transition behavior, and $(1-\pi)$ will answer with normal one-month transition behavior. Letting u equal the vector of four month transitions (w_4) and v equal the vector of one-month transitions (w_1) we have:

$$V(w_s) = \pi V(w_4) + (1-\pi)V(w_1) + \pi(1-\pi)(Ew_4-Ew_1)(Ew_4-Ew_1)' \qquad (5)$$

The observed on-seam covariance matrix will be a weighted average of four-month and one-month net transitions plus a bias term for the difference between four-month and one-month transition rates. This bias term reflects the additional variance introduced by the uncertainty of whether the respondents answer using a constant wave response or real values. If the covariance matrix of four-month net transitions exceeds the covariance matrix of one-month

transitions, $V(w_4)>V(w_1)$, then the covariance matrix of reported seam transitions will exceed the covariance matrix of the actual one-month transitions, $V(w_s)>V(w_1)$.

TRANSITION CORRELATIONS AND THE CONSTANT WAVE RESPONSE PHENOMENA

Equation (4) indicates that a constant wave response imposes a positive bias on off-seam transition correlations. Noting that the variance of a transition is equal to the product of its transition rate times one minus its transition rate, if we assume that two variables x and y have identical mean transition rates, m, then the following formula for observed correlations off-the-seam may be derived.

$$corr(x,y)_o = [cov(x_1,y_1)+\pi m^2]/[m(1-m)+\pi m^2] \qquad (6)$$

For transitions observed on-the-seam the situation is more complex. Under the restrictive assumptions that one month transition rates are identical and four month transition rates are identical it can be shown that also have a positive bias. The bias in on-seam correlations will be dependent upon the relative size of one-month and four-month transition rates and covariances, and may be either more or less than the bias in off-seam correlations.

CONCLUSION

In the analysis of longitudinal surveys like the SIPP, dynamic descriptive statistics will often reveal wave-seam effects. We believe that wave-seam effects stem largely from three basic causes, the constant wave response, underreporting of stigmatizing variables, and random misreporting. In this paper we have demonstrated that while wave-seam effects for univariate statistics are large, wave-seam effects for multivariate statistics are smaller. Measurement error research heretofore has focused on problems apparent in cross-sectional and retrospective research. The repeated measurements in longitudinal surveys, through providing a rich new source of data, brings to the surface new problems and new opportunities.

Table 2
Correlation Between Occurence of Monthly Transitions, by Wave Month

Variables	Month	Marital Status	Employ-ment Status	Personal Earnings	Total Family Income	Social Security	Family AFDC Receipt	Family Food Stamp Receipt
Marital Staus	1	1.000	0.023	0.027*	0.027	-0.014*	-0.004	0.012
	2,3,4	1.000	0.014	0.009	0.033	-0.001	0.004	0.011
Employment Status	1	0.023	1.000	0.380*	0.172*	-0.058*	0.048*	0.073*
	2,3,4	0.014	1.000	0.486	0.114	-0.004	0.029	0.039
Personal Earnings	1	0.027*	0.380*	1.000	0.463*	-0.097*	0.018	0.032
	2,3,4	0.009	0.486	1.000	0.497	-0.006	0.022	0.034
Family Income	1	0.027	0.172*	0.463*	1.000	-0.055*	0.054	0.050
	2,3,4	0.033	0.114	0.497	1.000	0.008	0.063	0.052
Social Security	1	-0.014*	-0.058*	-0.097*	-0.055*	1.000	-0.015*	-0.009*
	2,3,4	-0.001	-0.004	-0.006	0.008	1.000	-0.001	-0.004
AFDC Receipt	1	-0.004	0.048*	0.018	0.054	-0.015*	1.000	0.162*
	2,3,4	0.004	0.029	0.022	0.063	-0.001	1.000	0.200
Food Stamps	1	0.012	0.073*	0.032	0.050	-0.009*	0.162*	1.000
	2,3,4	0.011	0.039	0.034	0.052	-0.004	0.200	1.000

source: SIPP 1984 Full Panel Longitudinal Research Panel
(*) indicates significant difference, for a .01 two-tailed probability level.

Table 3
Correlation Between Occurence of Monthly Transitions, by Wave Month and Transition Criteria

		Marital Status	Employ-ment Status	Personal Earnings	Total Family Income	Social Security	Family AFDC Income	Family Food Stamp Income
	Month	Any Change	Any Change	Big Change	Big Change	Any Change	On/off	On/off
Personal earnings								
Any change	1	0.021	0.100	0.302	0.126	-0.314	-0.013	-0.016
Any change	2,3,4	0.014	0.227	0.393	0.184	-0.058	0.004	0.007
Big change	1	0.027	0.380	1.000	0.463	-0.097	0.018	0.032
Big change	2,3,4	0.009	0.486	1.000	0.776	-0.006	0.022	0.034
On/off	1	0.023	0.290	0.569	0.247	-0.040	0.018	0.035
On/off	2,3,4	0.007	0.459	0.641	0.315	0.001	0.020	0.034
Total family income								
Any change	1	-0.009	0.021	0.115	0.122	0.081	0.011	-0.022
Any change	2,3,4	0.019	0.150	0.231	0.213	0.122	0.026	0.014
Big change	1	0.027	0.172	0.463	1.000	-0.055	0.054	0.050
Big change	2,3,4	0.033	0.227	0.497	1.000	0.008	0.063	0.052
On/off	1	0.011	0.133	0.106	0.241	-0.037	0.094	0.076
On/off	2,3,4	0.009	0.103	0.128	0.281	-0.003	0.078	0.036
Social security								
Any change	1	-0.014	-0.058	-0.097	-0.055	1.000	-0.015	-0.009
Any change	2,3,4	-0.001	-0.004	-0.006	0.008	1.000	-0.001	-0.004
Big change	1	0.004	-0.007	-0.013	0.070	0.285	-0.005	0.012
Big change	2,3,4	0.010	0.019	0.010	0.073	0.317	0.007	0.002
On/off	1	-0.001	-0.001	0.002	0.041	0.183	-0.003	0.011
On/off	2,3,4	-0.002	0.020	0.013	0.063	0.274	0.007	0.003

source: SIPP 1984 Full Panel Longitudinal Research Panel

Table 4
Implied Incidence of the Constant Wave Responses

	π (%)	Proportion of Bias in Sample Transition Rates (%)
Marital status	31%	5%
Employment status	44	28
Personal earnings	54	34
Total family income	54	33
Individual social security	76	40
Family AFDC receipt	47	17
Family food stamp receipt	51	21

source: SIPP 1984 Full Panel Longitudinal Research Panel

REFERENCES

DALRYMPLE, R. and S. Carlson (1986), "Food Stamp Participation: A Comparison of SIPP with Administrative Records", **Proceedings of the U.S. Bureau of the Census Second Annual Research Conference**, U.S. Department of Commerce, Bureau of the Census.

KING, K., R. Petroni and R. Singh (1987), "Quality Profile of the Survey of Income and Program Participation", **Survey of Income and Program Participation Working Paper No. 8708.** U.S. Department of Commerce, Bureau of the Census.