Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods

Daniel H. Hill, The University of Michigan
Survey Research Center, University of Michigan, Ann Arbor, Michigan, 48106–1248

Introduction

We have seen repeated evidence in the Survey of Income and Program Participation (SIPP) that between-wave change dominates within-wave change.[1] Most analysis, to date, has been largely descriptive analysis of recipiency data (see e.g., Moore and Kasprzyk, 1984, Burkhead and Coder, 1985, Coder, 1986, Rascavage, 1986, and Weideman, 1986) and has resulted estimated between- to (average) within-wave transition ratios in the range of three to nine.[2] Since the same problem appears regardless of when the seam month occurs in calendar time, it is suggestive of substantial response error in reporting of monthly recipiency.

Whether or not this type of error is peculiar to studies employing the SIPP methodology of sequential-retrospective reporting for months in the reference period is a question of some considerable practical importance which has not yet been addressed. In the present paper we provide some evidence on this by comparing the between- and within-wave transitions observed for the SIPP with those observed in another study, the Panel Study of Income Dynamics (PSID), which employs a different methodology in collecting monthly data. We ask the very specific question of whether there is any evidence that the PSID methodology results in fewer between- relative to within-wave transitions than the SIPP methodology. While, in general, we would need to compute complex sampling errors and conduct formal tests to answer this question, in the present case these statistics are not necessary.

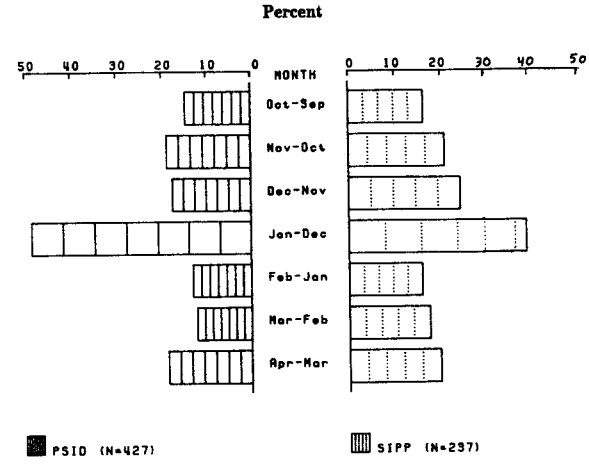
Another question of considerable concern which has not been adequately addressed is how these errors might affect estimates of models intended to explain the dynamics of welfare participation and employment.[3] If the response errors leading to exaggerated between- relative to within-wave transitions are systematically associated with either employment status or its determinants then it may result in serious biases in behavioral models. Using data from the PSID's 1984 and 1985 interviewing waves which incorporated an overlapping seam design, we will attempt to answer the question of whether there are significant associations of response errors around the seam to factors which might be viewed as determinants of behavior. We will also attempt to isolate some of causes of reporting inconsistencies which tend to amplify or attenuate between-relative to within-wave transitions.

A Comparison of SIPP and PSID Recipiency Transitions

As noted above, the methodology employed in the SIPP to obtain monthly recipiency and amounts data is sequential and retrospective. Early in the questionnaire, the respondent is asked about the receipt of income from an exhaustive list of possible sources. In addition, after Wave 1, respondents were reminded of the income sources they reported during the prior wave and asked if they continued to receive that income in the current reference period. Once the individuals income recipiency 'roster' is completed for the period, the respondent is asked about the timing of receipt within the four-month reference period. This questioning is sequential. For each income type listed in the roster the respondent is asked about whether it (and how much) was received in the calendar month prior to the interview, then for the month prior to that, etc. until the reference period is complete.

The type of seam problem that has been of such concern in past analysis of the SIPP is clearly evident in the reported transitions in unemployment compensation presented in the right half of Figure 1. To make things completely comparable with the PSID we limit out attention here to Rotation Group 4, Waves 1 and 3, of the 1984 SIPP Panel. The members of this subsample had their first 'seam' occurring between December 1983 and January 1984. The figure shows a pronounced 'bulge' in unemployment compensation transitions during this seam period—approximately twice as many people experienced a change at this time than at any other time. Since the same pattern appears for other rotation groups having their seams in other months, the excess of between- relative to within-wave change is <u>not</u> a result of true seasonal patterns.

Figure 1
Unemployment Compensation Transitions by Month

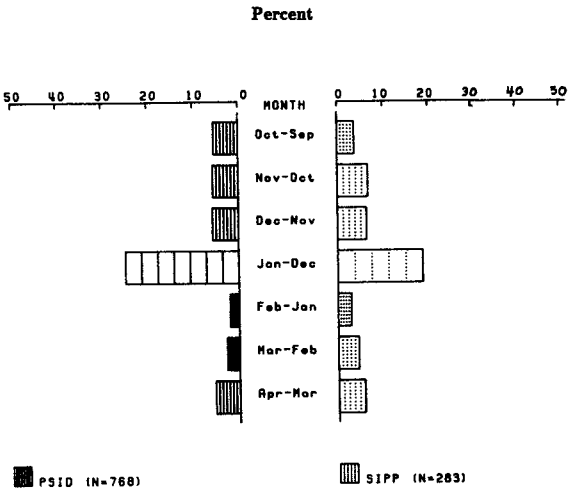


With respect to recipiency measures such as for unemployment compensation, the PSID methodology differs in three major respects from the SIPP. First, the PSID has a longer recall period. The PSID interviewing is conducted in the Spring and Summer of each year since 1968 with the reference period being the prior calendar year. Thus, the reference period requires recall of at least fifteen months and for some respondents, who are not interviewed until the end of the Summer, as much as twenty-one months. The second major difference in PSID methodology, is that no attempt is made to obtain monthly amounts—only annual total amounts are recorded. Finally, rather than ask about each month retrospectively and sequentially, the PSID asks the respondent to give the beginning and ending months for each spell of recipiency. All of these differences, particularly the first and third, should lead to more reporting error in the PSID than in the SIPP.

The left half of Figure 1 presents the monthly transitions in unemployment compensation derived from the seventeenth

(1984) and eighteenth (1985) waves of the PSID. Given the rather drastic differences in methodology, the patterns in the left half of Figure 1 are surprisingly close to the corresponding SIPP pattern on the right. The PSID, in general, appears to have somewhat less within-wave transition and a markedly more pronounced bulge at the seam than the SIPP. Otherwise, however, the patterns of monthly transitions from the two studies are quite comparable. This same general conclusion holds for Foodstamp recipiency, as examination of Figure 2 will confirm. With Foodstamps, however, the dominance of seam transitions over within-wave transitions is even more pronounced than with unemployment compensation in both studies. Unlike unemployment compensation, Foodstamps are not necessarily individual specific, but are provided to recipiency units which are either individuals, families, or subfamilies. Part of the large amount of between-wave change may be due to changes in the composition of households between waves, coupled with confusion regarding who is in the recipiency unit. Also, unemployment compensation tends to be a shorter duration phenomenon than foodstamp receipt, and true transition may be more common.[4]

Figure 2
Foodstamp Recipiency Transitions by Month



In conclusion, there is no evidence to suggest that the excess of between-wave relative to within-wave transitions is peculiar to the SIPP. The same patterns appear for the PSID which employs a radically different collection methodology. If anything, the PSID's longer reference and recall period may lead to more pronounced seam problems.[5] One common element to the design of both studies which may be responsible for this problem is simply that the time-unit of measurement, the month, is shorter than the reference and recall period.

Correlates of Reporting Inconsistencies Leading to Seam Problems

Having established the dominance of between-wave transitions in the PSID, as well as the SIPP, we now turn to capitalizing on the overlapping design of the PSID in isolating factors affecting inconsistent within- and between-wave transitions. The measure we will concentrate on is employment status and we will be especially concerned with transitions between December of 1983 and January of 1984. Data on employment status in this latter month were collected during both the 1984 and 1985 interviewing waves. Table 1 presents a cross classification of the two January reports for all respondents who were either a 'head of household' or a wife of the head of household in each year.[6] Because the 1984 questions upon which these reports are based were not asked of individuals who were not in the labor force as of the time of the 1984 interview, such individuals are eliminated from the analysis.[7]

Overall, the figures in Table 1 suggest substantial agreement in reports from the two interviewing years. The simple response variance (which, for a dichotomous variable, is simply one-half the ratio of inconsistent to consistent reports) indicated by the numbers in Table 1 is only .045. Most of this agreement, however, is the result of consistent reports of employment in the two years. Ninety-seven percent of those reporting in the Spring of 1984 that they were employed in January, also reported that they were employed in January of 1984 when they were asked about it in the Spring of 1985. For those who reported in 1984 being unemployed, in contrast, only forty-nine percent provided a consistent report one year later. Most of the others said in 1985 that they were employed in January of 1984—suggesting that they had forgotten all about the unemployment they reported a year earlier. Most of the individuals providing consistent 'employed' responses are people who had continuous employment throughout the reporting period and the reporting task for these people is orders of magnitude less difficult than for those experiencing a variety of employment situations.

Table 1
1984 and 1985 Reports of Employment Status in January 1984
Panel Study of Income Dynamics*

1985 Report	1984 Report		
	Employed	Unemployed	Out of Labor Force
Employed	6,039	207	40
Unemployed	94	286	28
Out of Labor Force	88	75	64

*Those retired, permanently disable, keeping house, or full-time students who were not working at the time of the 1984 interview have been eliminated from the analysis. Nineteen cases with combinations of employment, unemployment and out of the labor force reports in January 1984 are excluded from the table.

Given the type of data in Table 1 along with reports on employment status in December of 1983, there are several ways we could proceed in isolating factors associated with erroneous seam transitions. We could, for instance, analyze the simple response variance directly as has O'Muircheartaigh (1986), since spurious between-wave transitions and response variance are closely related. A more direct approach, however, involves concentrating only on those observations which involve measured transitions either within a wave or between waves and examining the extent of agreement in between- and within-wave transitions. There are three possible outcomes in this case. These are illustrated in Figure 3. First, both the between- and within-wave measures may indicate the same employment status transition between December 1983 and January 1984. Such consistent reports of *change* would be indicative of very good reporting on the part of respondents. Since they will occur if, and only if, the two January reports

are the same, they are inversely related to gross-difference rates and simple-response variance. Second, comparison of the 1985 report of January 1984 employment with the 1984 report of December 1983 (i.e. the between-wave measure) might indicate change whereas there is no corresponding change indicated by the 1984 reports (i.e. the within-wave measure). These inconsistencies would tend to amplify the ratio of between- to within-wave transitions and are the types of errors which seem most likely to be causing the seam problem. The third and final possibility is that the within-wave measure indicates change which disappears when one examines the between-wave measure. Such reports, while tending to attenuate the 'seam problem', are, nevertheless, reflective of poor response quality.

Figure 3
Patterns of Inconsistency in Overlapping Reports

	December 1983	January 1984	Reporting Year
Consistent Reports	1	0	84
		0	85
or	0	1	84
		1	85
Seam-Amplifying	1	1	84
Inconsistencies		0	85
or	0	0	84
		1	85
Seam-Attenuating	1	0	84
Inconsistencies		1	85
or	0	1	84
		0	85

The advantages of this approach over the analysis of simple response variances are largely interpretational and analytic. The interpretational advantage is that we can see directly the effects of factors on the likelihood of reporting inconsistencies which both exaggerate and attenuate measured between-wave change. The analytic advantage becomes apparent once we note that any observed effect on the simple response variance may come about either via an effect on the probability of actually being in a stable employment situation or via a true effect on the error variance of the response. By ignoring those individuals with stable employment situations, we are, in effect, controlling for stable employment situations and we can more directly attribute any observed effects to true response quality.

In order to understand the effects of factors on the observed between- and within-wave transitions it is necessary to develop a model. Specifically, we assume that each individual i has a propensity R to provide reports of type j (where j <1,2,3> corresponds to consistent transition reports, inconsistent reports which tend to attenuate between-wave transitions, and inconsistent reports which tend to amplify between-wave transition measures). These response propensities contain systematic and stochastic components. The systematic portion consists of the effects of a series of exogenous measured factors X while the stochastic portion, denoted μ, reflects the effects unmeasured excluded factors. As a first order approximation, the response propensities can be expressed as:

$$R_{ij} = X_i\beta_j + \mu_{ij} \qquad 1)$$

The individual is assumed to provide the response j with the highest propensity score R. If the error term μ follows a Type I extreme value, or log-Weibull distribution with density $\phi = \exp(-\mu)\exp[-\exp(-\mu)]$ then we can model the response process according to the multinomial logit model. In this case, the probability that individual i will fall into response class j is:

$$P_{ij} = \exp((X_i\beta_j) \div \sum_{k=1}^{3} \exp(X_i\beta_k) \qquad 2)$$

The factors (X) assumed to affect the propensity to provide reports of varying quality can be broken into four groups—demographics, difficulty of task measures, interview characteristics, and respondent cognitive ability. The demographic measures (race, age, gender, education, and income) are included both because they may have direct impacts on the quality of data and because they are often used as controls in structural analysis of other behaviors. Any systematic association of these factors with response quality will mean that parameter estimates from these latter models will be biased.

The difficulty of the recall task measures consist of the length of the recall period, the number of intervening transitions in employment status, the length of time the person has been employed by the employer as of the time of the 1985 interview, whether or not the respondent is the reference person, the person's industry of employment, whether or not he is self employed, and whether or not he has extra jobs in addition to his 'main job'. Cognitive psychologist have made quite a lot of the first two of these factors. Both length of recall (via its effect on telescoping and omissions) and number of intervening transitions (via interference phenomenon) are thought to adversely affect the quality of recall. While the direction of causation is somewhat questionable, length of employment is thought to have a positive effect on observed data quality because being employed with one employer for a long period should reduce the recall task. The ambiguity of employment status for the self employed, those with extra jobs, and those in the construction industry should result in reduced data quality as should the respondent having to report on the labor force behavior of some other individual.

The third set of factors thought to affect the propensity of respondents to provide data of varying quality have to do with the interview itself. It is comprised of two measures—a dummy variable indicating the the respondent initially refused the 1985 interview, and the length of time the interview took to complete. Having refused the interview is an indication of low motivation and should raise the propensity of the respondent providing faulty information. The effect of length of recall is more ambiguous. On the one hand, the more complicated the respondents situation, the longer the interview will take, and the greater temptation to provide simplified or stylized reports. On the other hand, the more careful the respondent is in formulating responses the longer the interview will take. Which of these influences dominates is view as an empirical question. The final factor assumed to affect response propensities is a measure of the individual's cognitive skill's in Standard American English. This measure is derived from a sentence complete test administered to the respondent in 1972. Scores on this test have been found to be highly correlated with more rigorous 'IQ' tests but are also highly culturally dependent. Since the PSID questionnaire is written in Standard American English, however, it is reasonable to assume that both cognitive and language skills on the part of the respondent will affect the quality of the data derived from it.

The figures in Table 2 correspond to the model which includes demographic factors only. The only demographic factor having a significant effect on the propensity of respondents to provide inconsistent reports which attenuate

between-wave transitions is total family income. Each thousand dollars of such income has the effect of raising the log of the odds of such an inconsistent report being given by .016 points. Even this effect is only marginally significant. In contrast, both race and age have strongly significant effects on the propensity of respondents to provide seam-transition exaggerating inconsistent reports. Blacks are far more likely to provide inconsistent reports which serve to amplify between-wave transitions than are non-Blacks. Similarly, the older the respondent, the more likely he is to provide seam-transition amplifying reports.

There are a variety of reasons why we might see such race and age effects. Cognitive psychologists have often argued that age reduces the efficiency with which people encode and retrieve events into and from memory. If this is the case, then the age effect may simply be reflecting less accurate recall. Past empirical evidence, however, has not consistently shown such a relationship. Indeed, O'Muirchearchtaigh (1986) finds that older respondents have lower gross-difference rates in reinterview data for the CPS than younger respondents. This is in direct conflict with our findings presented in Table 2. There are several differences between his analysis and ours which might account for the conflicting results. Perhaps most

Table 2
Multinomial Logit Estimates for Between-Wave Attenuating and Amplifying Inconsistencies in Reported Employment Status

	Demographic Controls Only	
	Attenuating Inconsistencies	Amplifying Inconsistencies
Constant	−.723	−.567
	(.717)	(.513)
Whether Black	.050	.577**
	(.214)	(.148)
Age	−.072	.167**
(decades)	(.093)	(.062)
Whether Male	.327+	−.133
	(.197)	(.138)
Education	−.027	.001
	(.048)	(.033)
Income	.016*	−.009
($1,000)	(.007)	(.006)
Log Likelihood	−997	

+ Some evidence of effect. * Significant at p = .05.
** Significant at p = .01.

important is our exclusion of those reporting being either retired, a student, a housewife, or permanently disabled and who did not work at the time of the 1984 interview. These people are disproportionately located at both extremes of the age distribution and O'Muircheartaigh's age effects are most pronounced at these extremes. The only elderly people left in our sample are those working at least ten hours a week at the time of the 1984 interview. Many of these people are likely in part-time or casual employment situations and this type of employment might be particularly prone to mis-reporting.

The difficulty of the recall task may also explain the race effect. Blacks are far more likely than non-Blacks to experience labor force disruptions and the difficulty of their recall task is likely to be far greater. Additionally, the reporting task is probably made more difficult for some Blacks because they are less facile in standard American English than are non-Blacks.

If these factors are responsible for the lower quality of data for older and Black respondents, then the estimated coefficients on these variables should be greatly reduced once we control for factors reflecting the difficulty of the reporting task, the cognitive and language skills of the respondent, and the nature of the interview situation itself. The figures presented in Table 3, however, indicate that this is not the case. If anything, the effects of age and race are increased by the inclusion of measures intended to capture the effects of task, and cognitive ability on response quality. The only such measure to have a truly significant effect on response propensities is tenure with the employer of record at the time of the 1985 interview. While not entirely tautological—people can and do experience periods of unemployment and absences from the labor force in the midst of a period of employment with a single employer—there is a strong definitional component to this effect.

The inclusion of task and cognitive factors also has the effect of increasing the power of the demographic effects on the propensity to provide inconsistent reports which attenuate between- to within-wave transition ratios (column 1). Specifically, both gender and income now become significant—with males and low income respondents being significantly less likely to provide such reports.

Although none of the other variables are significant at conventional levels, a couple of factors do have sufficiently large estimated effects relative to their estimated standard errors to be worth noting. Specifically there is some evidence to suggest that self-employed respondents are more prone to providing between-wave attenuating responses than are respondents who work only for others. Similarly, the reports from respondents who answered for themselves in both 1984 and 1985 are less likely to lead to seam transition amplifying inconsistencies than are reports involving proxy respondents. There is also some evidence that cognitive ability in standard American English as measured by the sentence completion test administered in 1972 does reduce the propensity to provide inconsistent seam transition amplifying reports. Since only twenty percent of the 1985 respondents were respondents in 1972,[8] it was necessary to include two measures—the test score of the 1972 respondent, and an interaction of this score and a dummy variable indicating a change in respondent between 1972 and 1985. The estimated effect of one's own test score is given by the coefficient on the first of these variables, while the effect of the test score of the 1972 respondent on some other 1985 respondent is given by the sum of the coefficients on the two variables. The former effect on the propensity to provide seam amplifying effects is negative, while the latter effect is virtually zero. To test the significance of the cognitive skills/language test-score it is necessary to remove both measures from the analysis and perform a likelihood ratio test. The results of this test is a reduction in the log-likelihood value of 3.07 which implies a χ-square of 6.14 with 4 degrees of freedom.[9]

Finally, given their prominence in psychological discussions of recall accuracy, two variables should be noted for their lack of apparent effect on our measures of response consistency. These are length of recall and the number of transitions intervening between the time of the 1985 interview and the period being reported. Both of these were expected to adversely affect response quality, but the estimated effects are so small relative to their standard errors as to preclude our rejecting the hypothesis of zero effect. If anything, the point estimates suggest that both factors are associated with higher quality recall. With respect to the reported number of intervening transitions this may be the result of respondent heterogeneity with those reporting any within-wave change providing better data than those who do not. With respect to

Table 3
Multinomial Logit Estimates for Between-Wave Attenuating and Amplifying Inconsistencies in Reported Employment Status

	Attenuating Inconsistencies	Amplifying Inconsistencies
Constant	.344	.340
	(1.11)	(.776)
Demographics		
Whether Black	.163	.631**
	(.235)	(.168)
Age	−.168	.024**
(decades)	(.115)	(.008)
Whether Male	−.569*	.004
	(.230)	(.159)
Education	−.032	−.002
	(.051)	(.036)
Income	.016*	−.008
($1,000)	(.008)	(.006)
Difficulty of Task		
Extra Jobs	.272	.117
	(.271)	(.208)
Months with	−.062	−.153**
Current Employer	(.053)	(.037)
Construction	−.057	−.002
	(.048)	(.034)
Self Employed	.590+	−.014
	(.322)	(.252)
Self Reports	−.219	−.251+
	(.214)	(.155)
Length of Recall	−.085	−.068
	(.101)	(.069)
# Intervening	−.024	−.064
Transitions	(.073)	(.051)
Interview Characteristics		
Initial Refusal	.355	−.178
	(.625)	(.513)
Length of Interview	.061	−.018
	(.056)	(.039)
Cognitive Ability		
Test-Score	.025	−.054+
	(.052)	(.037)
Test-Score x	−.038	.036+
Not '72 Respondent	(.033)	(.025)
Log Likelihood	−976	

+ Some evidence of effect. * Significant at p = .05.
**Significant at p = .01.

length of recall, our results may be consistent with very rapid memory decay—so rapid that the difference between nine and nineteen months recall is irrelevant.

Conclusions

In this paper we have employed monthly data from the 1984 and 1985 waves of the Panel Study of Income Dynamics to investigate the extent and determinants of excessive measured change between waves relative to measured change within waves of panel surveys. We find that, in spite of different question sequences, the dominance of between-wave change in the PSID is at least as severe as in the SIPP. If anything the PSID data are worse in this regard than the SIPP. Furthermore, the data suggest that this type of 'seam problem' may be more severe for measures that are tied to groups of individuals (e.g. Foodstamps) rather than to a specific individual (unemployment compensation). There is also the suggestion in the data that the average duration of receipt of income sources may positively affect the severity of the seam problem.

Our attempt to understand the determinants of seam problems using overlapping reports of employment status from the last two waves of the PSID was only partly successful. We did identify significant correlates of the propensity to provide inconsistent reports which amplified between- to within-wave transition ratios, but we failed to identify their causes. Blacks and older respondents were found to be significantly more likely to provide seam transition amplifying reports, but none of the measures intended to explain why this might be the case (with the exception of employment tenure) had the expected significant effects. There was some weak evidence that cognitive ability and facility in standard American English enhanced the quality of reports, but no evidence of the much touted effects of length of recall and interference of like events was found. Similar inexplicable effects of gender and income for the propensity of providing inconsistent reports which tended to attenuate between-wave changes were also found.

References

Burkhead, D. and J. Coder (1985). "Gross Changes in Income Recipiency from the Survey of Income and Program Participation", *Proceedings of the Social Statistics Section, American Statistical Association*, 351–356.

Coder, J. (1986). "Monthly Transitions from the SIPP Longitudinal Research File", Bureau of the Census Memorandum to Paula J. Schneider, May 20, 1986.

Hill, M.S. and D.H. Hill, (1986) "Labor Force Transitions: A Comparison of Estimates from Two Longitudinal Surveys", *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 220–225.

Kalton, G., J. Lepkowski, and T. K. Lin (1985). "Compensating for Wave Nonresponse in the 1979 ISDP Research Panel", *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 147–164.

Kalton G., and M. E. Miller (1987). "Errors in Reporting the Annual Increase in Social Security Payments in SIPP", Internal working paper, Survey Research Center, University of Michigan.

Moore, J. C. and D. Kasprzyk (1984). "Month-to-month Recipiency Turnover in the ISDP", *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 726–731.

O'Muircheartaigh, C. A. (1986). "Correlates of Reinterview Response Inconsistency in the Current Population Survey",

Bureau of the Census: Second Annual Research Conference 208–234.

Rascavage, P. and A. Feldman-Harkins (1986). "Work Experience Data from the SIPP", paper presented at the Allied Social Science Association meetings, New Orleans, December 28.

Weideman, L. (1986). "Investigation of Gross Changes in Income Recipiency from the Survey of Income and Program Participation", *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 231–236.

Footnotes

[1]This research was sponsored, in part, by a Joint Statistical Agreement (JSA 87–5) between the Survey Research Center of the University of Michigan and the United States Bureau of the Census. The author would like to thank Graham Kalton and Charlie Brown for their helpful suggestions on earlier versions of the paper and Judy Connors for her SIPP data management assistance. Any errors are the responsibility of the author.

[2]There is evidence that the 'seam problem' is not confined to discrete data. Kalton, Lepkowski, and Lin (1985) find similar patterns for changes in income.

[3]Hill and Hill (1986) did find that seam transitions had powerful effects in a proportional hazards model of re-employment estimated with SIPP data.

[4]Additional evidence in support of these propositions is provided by an examination of month-to-month transitions in the receipt of AFDC benefits (long duration support for families) and secondary employment (individual specific and short in duration).

[5]This also means, of course, that there are fewer seams in the PSID. Whether this results in better overall data quality depends on whether the seam problem is the result of too much measured change between waves or too little measured change within.

[6]Because the study began in 1968 we originally used the now archaic and admittedly sexist 1960 Census definitions of Head of Household in our original design.

[7]More precisely anyone either retired, permanently disabled, a keeping house, or a student, and who were not working at least ten hours a week at the time of the 1984 interview were skipped out of the employment work history sequence in 1984. This is an unfortunate restriction because it reduces the variance in both the outcome measure we are interested in (response error) and in a potentially important predictor (initial employment status).

[8]The eighty percent of 1985 respondents who were not respondents in 1972 are composed primarily of children or spouses of the 1972 respondent.

[9]When test-score is included as the sole predictor we find it to be significant at the 95% level of significance. When race is added, however, the effect of test-score becomes insignificant.

EXAMINATION OF RELATIONSHIPS BETWEEN ACTUAL AND REPORTED CHANGES IN THE SIPP

Lynn Weidman, U.S. Bureau of the Census[1]

1. INTRODUCTION

One set of variables that is important in SIPP and has been examined closely is the receipt of benefits from various state and federal government programs. Burkhead and Coder (1985) showed that a large proportion of the changes between receipt and nonreceipt (hereafter referred to as transitions) were reported as occurring between the last month of a reference period and the first month of the following one. As an example of this pattern see Table 1. It shows the distribution of transitions and non-transitions in food stamp recipiency for the first 16 months of all rotation groups combined in the SIPP 1984 panel. (Only individuals who responded to the first four interviews and received food stamps at some time during this period are included.) Transitions between R = receipt and N = nonreceipt are recorded in the rows labeled RN and NR. The months noting adjoining reference periods are 4th to 5th, 8th to 9th and 12th to 13th, and they have much larger RN and NR counts than the other months.

Weidman (1986, 1987) explored possible relationships of this pattern to demographic characteristics, interview status and imputed receipt. It was determined that for food stamps and social security there is some difference between the reporting of receipt by self-and proxy-respondents, and, in general, a larger proportion of between reference period transitions occurs when at least one of the two months is imputed. However, these effects are of a scale too small to register the magnitude of the observed pattern. A natural conclusion from these results is that frequently people simply report the same receipt status for all months of a reference period, regardless of when a change actually occurred.

As long as the actual number of transitions remains constant, the difficulty in determining this number is due solely to errors in reporting the months of occurrence. Without the aid of auxiliary data it is impossible to estimate the error structure due to reporting for four consecutive months simultaneously, unless assumptions are made to oversimplify the structure. A discussion of the bias in the estimation of transition levels and an approach to measuring it was given by Hubble and Judkins (1986). Currently an administrative record study is being carried out at the Census Bureau in order to estimate the SIPP error structure in reporting receipt of benefits from nine government transfer programs (see Moore (1986)).

Another item of interest is the period of time a person receives benefits from a given program, which will be called a spell. Agencies responsible for benefit programs and people studying the effectiveness of them want to track changes in spell lengths in order to monitor costs, determine effects of changes in qualifications for benefit eligibility, examine the effect of government policies on benefit receipt, etc. To this point, not much effort has been put into the estimation and examination of spell lengths as reported in SIPP. One of the reasons for this is undoubtedly that they require longitudinal estimation through the matching of several waves of data as compared to the simpler cross-sectional nature of transitions. (A wave is a set of reference periods, usually one per rotation group, whose data are processed together.) Also, the error structure of reported lengths is complicated because both the start and end of spell lengths have their own response errors. Estimating a distribution of spell lengths is more complicated than estimating a number of transitions. The possible patterns of changes in these distributions are more complicated than numbers of transitions increasing or decreasing.

AS for transitions, when the distribution of spell lengths is constant, estimation is complicated solely by errors in reporting the starting and ending months. When the number of transitions and distributions of spell lengths change, the estimation problems become more difficult because each month must be allowed to have a different set of parameters associated with it. Our interest is in being able to identify the types of changes that take place and to recognize them as soon as data will allow.

In this paper we present linear models that attempt to represent the relationships between the observed and actual transitions and spell lengths. The parameter vector of unknown values to be estimated is the mean number of a specified transition actually occurring for a rotation group in a specified set of months, or the mean reported number of spells with specified start and end months. These models are obviously oversimplified because they use only the reported data and no other information source. However, since our purpose is to estimate changes and not actual levels, the estimates may not be unreasonable. In the future we hope to combine this approach with the results of the administrative records study of Moore (1986) and the work of Hubble and Judkins (1986) to get improved estimates of the error structure. Estimation of these models will not be pursued here, but we will take a brief graphical look at the comparison of observed and actual values (Figure 2).

In the next section the relationship between reported and actual transitions is discussed and a basic model for reporting transitions is presented. The reporting probabilities are discussed in section 3. Section 4 examines the error structure for the transition models and section 5 extends the modeling approach to reported spell lengths.

2. MODELING TRANSITIONS

A transition between receipt states for "source A" is said to occur in a given month when the answer to the question about whether or not a person received benefits from "source A" is different than it was in the previous