The Seam Effect with Social Security Income in the Survey of Income and Program Participation

Graham Kalton¹ and Michael E. Miller²

Abstract: The Survey of Income and Program Participation (SIPP) is a panel survey with an interval of four months between waves of data collection, and with information on many income sources being collected on a monthly basis. A common finding has been that more month-to-month changes in recipiency from most income sources, and the amounts received if any, have occurred when the data are collected in different waves than when they are collected in the same wave. This finding has been termed the seam effect. This paper examines

the seam effect in relation to the monthly amounts of Social Security payments received in the first twelve months of the 1984 SIPP panel. The analyses take advantage of a known 3.5% increase in Social Security payments that occurred in January 1984 to compare the characteristics of recipients who reported an increase that month with those of recipients who failed to do so

Key words: panel survey; measurement error.

1. Introduction

This paper is concerned with a type of measurement error encountered in panel surveys that has become known as the seam effect. This effect has been found to be pervasive in the Survey of Income and Program Participation (SIPP), a household panel survey program of the U.S. Bureau of the Census. In order to describe the seam effect it is first necessary to give some basic details of the SIPP design.

¹ Survey Research Center, University of Michigan, Ann Arbor, MI 48106-1248, U.S.A.

Acknowledgements: The research reported here was supported by Joint Statistical Agreement 87-5 between the U.S. Bureau of the Census and the Survey Research Center, University of Michigan, U.S.A. The views expressed and the findings reported in this article are those of the authors. They are not necessarily endorsed by the Bureau of the Census.

The SIPP is an ongoing survey program with a new panel being introduced each year. Each panel collects detailed information on the economic resources and participation in welfare programs of sample members by means of interviews conducted every four months for a period of 32 months. At each wave of a SIPP panel sample members are asked whether they received any income from a wide range of income sources and transfer programs (e.g., Social Security, Federal Supplemental Security Income, Aid to Families with Dependent Children, Food Stamps) during the preceding four months. For each source, they are asked for each of the preceding four months in turn, starting with last month and working back to four months ago, first whether they received any income from that source and then, if so, how much was received. Merging the data

² Division of Biostatistics, Indiana University, Indianapolis, IN 46202-5200, U.S.A.

collected in the individual waves of the panel for each sample member thus creates a continuous monthly history of recipiency or non-recipiency of each income source, and of the amounts received, if any, for the 32-month life of the panel.

Analyses of the month-to-month variation in recipiency of the various income and transfer program sources and in the amounts received from the individual sources has uncovered the common pattern that changes in recipiency status and in amounts received occur much more frequently between months for which the data are collected in different waves (months 4 and 5. 8 and 9, 12 and 13, etc.) than between months for which the data are collected in the same wave (months 1 and 2, 2 and 3, 3 and 4, 5, and 6, etc.) of the panel. Since changes occur more frequently at the seam between two waves of data collection, this pattern has become known as the "seam effect." Findings on the seam effect are reported by Burkhead and Coder (1985), Coder, Burkhead, Feldman-Harkins, and McNeil (1987), and Weidman (1986) in relation to SIPP, and by Moore and Kasprzyk (1984) and Kalton, Lepkowski, and Lin (1985) in relation to the Income Survey Development Program (ISDP) 1979 panel, a pilot survey for the SIPP. Marquis and Moore (1989) report on a study of the seam effect based on a comparison of survey reports with administrative records. Further references are given by Kasprzyk (1988) and in the SIPP quality profile (Jabine, King, and Petroni 1990). Hill (1987) reports the occurrence of a similar seam effect with the Panel Study of Income Dynamics, an ongoing panel survey that started with a nationally representative sample of the members of 5,000 U.S. families in 1968, and that has collected data annually since that time on the original sample members and the members of their families at each wave.

This paper examines the seam effect in relation to the monthly amounts of Social Security payments reported in the 1984 SIPP panel. As preparation for the analyses that follow, Section 2 below provides some necessary background on the 1984 SIPP panel and describes the data set used for the analysis. Section 3 then presents the results of some analyses that document the magnitude of the seam effect for Social Security income. Section 4 takes advantage of a 3.5% increase in Social Security payments that was introduced in January 1984 to compare the characteristics of recipients who reported an increase in that month with those of recipients who failed to do so. The final section of the paper discusses the findings.

2. The 1984 SIPP Panel

The analyses reported in this paper relate to the first three waves of data collection for the 1984 SIPP panel. That panel started with about 20,000 interviewed households. The sample was made up of four subsamples, called rotation groups, of approximately equal size, with one rotation group being interviewed each month to collect data for the preceding four months, and with most of the income data collected on a monthly basis. One rotation group was interviewed for the first time in October 1983, and then reinterviewed in February 1984, June 1984, etc. Another rotation group was first interviewed in November 1983 and reinterviewed in March 1984, July 1984, etc. Similarly, the other two rotation groups were first interviewed in December 1983, and January 1984, respectively, and then reinterviewed at fourmonth intervals. As a consequence of this data collection procedure, data for two adjacent months were collected in different waves for one rotation group but in the same wave for the other rotation groups. Thus, for instance, data for September and October 1983 were collected in different waves for the rotation group that was first interviewed in October 1983 (the first wave for September and the second wave for October) but in the same wave (the first wave) for the other three rotation groups. For the present analyses, this rotation scheme has the benefit of providing the opportunity to compare the change between two adjacent calendar months when the data were collected in different waves with the corresponding change when the data were collected in the same wave.

All persons aged 15 and over in the approximately 20,000 households sampled at the first wave of the 1984 SIPP panel became panel members who were followed even if they changed addresses or moved out of their sampled households. Children under 15 in sampled households became panel members at later waves after reaching the age of 15, provided that they were still living with a panel member at that time. Persons who were not in the initial sample but who subsequently resided with panel members – termed associated persons – were included in the survey while they continued to live with panel members.

The data set used for this study was constructed by merging the public use files for the first three waves of the 1984 SIPP panel. A number of exclusions were then made from the merged file. First, the rotation group first interviewed in January 1984 has been excluded because data were not collected from this group in the second wave. Second, all associated persons have been excluded. Third, all children aged under 15 at the first interview have been excluded. Fourth, all panel members leaving the survey population during the first three waves (e.g., through death, entering an institution, or emigration) have been excluded. Fifth, all sample persons who were nonrespondents on one or more of the first three waves have been excluded. The study is thus confined to panel members aged 15 and over at the first wave who responded on each of the first three waves of the 1984 SIPP panel.

A final set of exclusions has been made on the basis of the variable under study, the monthly amounts of Social Security income. These amounts were subject to some item nonresponse. When this occurred, an imputation procedure was used to assign values for the missing amounts. Since imputations are likely to distort measures of individual monthly changes, imputed amounts have been treated as missing values in the analyses that follow. As such, they have been excluded from the analyses. In addition, a small number of extreme amounts of \$1,500 or more of Social Security income in a single month have been excluded on the grounds that these unusual payments could also distort the analyses.

3. The Seam Effect with Social Security Income

One way to illustrate the seam effect with the amount of Social Security received is to correlate the amounts received in different panel months. Table 1 presents the correlation matrix for the monthly amounts of Social Security received in each of the twelve panel months covered by the first three waves of the 1984 SIPP panel. This correlation matrix is computed for a subsample of the panel. Extreme values of monthly amounts of \$1,500 or more and changes of more than \$200 between months have been excluded (ten records in the subsample had amounts of \$1,500 or more for one or more months and six records had changes of more than \$200 between months). Each of the correlations is based on a subsample of about 3,000 persons who reported amounts of Social Security income in both of the two

9 10 11
1.00
0.99 1.00
0.99 1.00 1.0
)

Table 1. Cross-month correlations for Social Security income amounts¹

¹The stepped line separates the correlations into those for which the two monthly amounts were collected in the same wave (above the line) and those for which the two monthly amounts were collected in different waves (below the line).

months involved (excluding imputed values and extreme values as noted above).

The correlations in Table 1 exhibit the same pattern that Kalton, Lepkowski, and Lin (1985) found with the ISDP 1979 panel: For a given difference in panel months, the correlations when both amounts are collected in the same wave are appreciably higher than when they are obtained in different waves. In particular, the leading diagonal, which gives the correlations of amounts from adjacent months, shows the drop in correlation between months 4 and 5 and months 8 and 9. The correlation matrix in Table 1 in fact partitions into two parts: the correlations between amounts for months within a wave (above the stepped line in the table) are on average about 0.99 whereas those between amounts in different waves (below the stepped line) are on average about 0.92.

The correlations in Table 1 relate to panel months, which represent different calendar months for the different rotation groups. Table 2 provides another way of illustrating

the seam effect, this time relating to calendar months. The table, which relates to the full sample of persons who responded on each of the first three waves of the 1984 SIPP panel (apart from the exclusions noted in the last paragraph of the previous section), gives the distributions of the percentage changes in the amount of Social Security income received from one calendar month to the next. Separate distributions are given for the situation where the data for both the current month and the preceding month are collected in the same wave - the within-wave distributions (W) - and the situation where the data for the two months are collected in different waves - the between-wave distributions (B). The results for each month are based on persons reporting receipt of Social Security income in that month and the preceding one.

Inspection of the within-wave distributions in Table 2 shows that they are very similar for each of the months, with very little change reported. The only exception is January 1984, when a 3.5% increase in

Table 2. Percentage change in amount of Social Security income in current month compared to previous month

	Within	Percent change from previous month						
(W)		Reduction			Increase		÷	
Month	between (B) wave	More than 10%	10% or less	No change	10% or less	More than 10%	Total	Sample size
September	W	0.2	0.1	99.1	0.3	0.3	100.0	4,917
October	W B	0.1 6.4	0.2 21.4	99.2 36.0	0.3 27.6	0.2 8.6	100.0 100.0	3,285 1,510
November	W B	0.2 5.9	0.9 23.0	97.8 29.5	0.7 34.2	0.4 7.4	100.0 100.0	3,257 1,496
December	W B	0.3 6.2	0.5 24.3	97.2 23.1	1.5 38.9	0.5 7.5	100.0 100.0	3,221 1,491
January	W	0.3	1.0	36.6	60.5	1.6	100.0	4,809
February	W B	0.2 6.3	0.3 21.7	96.7 36.1	2.5 29.5	0.3 6.4	100.0 100.0	3,191 1,475
March	W B	0.1 5.4	1.2 20.4	97.6 41.7	0.9 26.3	0.2 6.2	100.0 100.0	3,157 1,451
April	W B	0.1 6.1	0.2 18.7	99.0 40.4	0.3 27.4	0.4 7.4	100.0 100.0	3,113 1,440
May	W	0.3	0.2	99.1	0.1	0.3	100.0	4,650
Average	W* B	0.2 6.0	0.4 21.6	98.3 34.4	0.8 30.7	0.3 7.3	100.0 100.0	

^{*}Excluding January.

Social Security payments was introduced. The average within-wave percentage change distribution for all months excluding January is given at the bottom of the table. The between-wave distributions are also very similar for each of the months, and their average is given at the bottom of the table. As these average distributions show, 98.3% of amounts show no change from the last month when the amounts for both months were collected in the same wave whereas only 34.4% of amounts show no change from the last month when the amount for the last

month was collected in the previous wave. The marked contrast between the average within-wave and between-wave distributions of percentage change clearly demonstrates the magnitude of the seam effect.

4. The December to January Change

The December to January change in Social Security amounts was measured as a within-wave change for the three rotation groups analyzed in this study. As noted above, the percentage change distribution from December to January differs markedly from

the within-wave percentage change distributions for other adjacent months. This difference can be explained by the 3.5% increase in Social Security payments that began in January 1984. As can be seen from Table 2. three-fifths of the respondents reported an increase of under 10% for January. However, over one-third did not report an increase at that time. While it is conceivable that some Social Security recipients experienced a drop in their payments in January that exactly counterbalanced the 3.5% increase, this eventuality seems improbable. In the following analysis we assume that those who reported that they received the same payments in December and January have failed to report the increase.

Table 3 presents a breakdown of the percentage change distribution for January by rotation group. The table shows that the proportion of Social Security recipients failing to report the January increase differs appreciably by rotation group, being lowest for rotation group 1 and highest for rotation group 3. In interpreting this finding, it should be noted that rotation group 1 was interviewed in February about the October to January period, rotation group 2 was

interviewed in March about the November to February period, and rotation group 3 was interviewed in April about the December to March period. Thus, the proportion failing to report the increase rises the longer the interval between the occurrence of the increase and the interview date.

The next step in our analysis is to compare the characteristics of persons who reported the 3.5% increase in January with those of persons who failed to do so. For this purpose, we needed to identify those who reported the 3.5% increase. A histogram of the percentage increases from December to January showed that a sizeable number of cases fell in the neighborhood of 3.5%, but that there were no clearcut boundaries to distinguish those reporting 3.5% increases from others. Based on a review of the histogram, we chose to classify those reporting January increases between 2.0% and 4.1% as correctly reporting the 3.5% increase. This classification is necessarily imperfect, but we believe it should suffice for the following analyses. This classification yielded 2,310 "correct" reporters, 1,762 "incorrect" reporters (that is persons who reported no increase from December to January), and 737 reporters for whom it was

Table 3. Percent change from December to January by rotation group

	Group 1	Group 2	Group 3
Reduction > 10% ≤ 10%	0.3 0.2	0.4 2.4	0.2 0.3
No change	29.4	35.8	45.0
Increase ≤ 10% > 10%	68.4 1.7	59.6 1.8	53.0 1.5
Total	100.0	100.0	100.0
Number of respondents	1,626	1,614	1,569

uncertain whether or not they had reported the 3.5% increase. The last group is excluded from the following analysis.

A logistic regression modelling exercise was conducted to find a combination of explanatory variables to predict correct reporting of the January increase. The variables examined as potential explanatory variables were: rotation group (three groups); interview status (self-reporter, proxy informant); highest grade of education attended (0-8, 9-12, over 12); gender; marital status (married and living together, other); race (white, non-white); age (above median age, below median age); receipt of pension (yes, no); January household income (above median income, below median income); and January Social Security payment (above median payment, below median payment). Five "correct" and 66 "incorrect" reporters were excluded from these analyses since they were coded as a category other than selfreporter or proxy informant.

The logistic regression analyses employed the approach described by Koch, Freeman, and Freeman (1975) or the analysis of complex survey data. Weighted proportions and a corresponding covariance matrix were computed for the contingency table defined by the cross-classification of the potential explanatory variables and the response variable using the OSIRIS IV Statistical Software System (Computer Support Group 1984). The weighted proportions were transformed into logits, and the logits were modelled relative to the complex sample covariance matrix using the weighted least squares approach described in Grizzle, Starmer, and Koch (1969). Wald statistics were generated in GENCAT (Landis, Stanish. Freeman, and Koch 1976) to test hypotheses about the relationships of the predictor variables to the logits.

After examining several competing models, the following model was chosen as

the most appropriate:

$$\log \left[\hat{\rho} / (1 - \hat{p}) \right] = 0.014 + 0.45R_1$$
$$+ 0.0003R_2 + 0.29S + 0.23W - 0.098P$$

where \hat{p} = the predicted proportion giving correct responses

 $R_1 = 1$ for rotation group 1, 0 for rotation group 2, -1 for rotation group 3

 $R_2 = 0$ for rotation group 1, 1 for rotation group 2, -1 for rotation group 3

S = 1 for self-reporter, -1 for proxy informant

W = 1 for white, -1 for non-white

P = 1 if the January Social Security payment is the median payment of \$413 or less, -1 if it is greater than \$413.

The analysis of variance for this model is given in Table 4. According to this model, there is a clear linear trend by rotation group (as observed in Table 3), and self-reporters, whites, and persons receiving larger Social Security payments are more likely to report the January increase than their counterparts.

The logistic model can be used to predict the percentage of correct reports in each of the cells of the crosstabulation of the explanatory variables involved. These predicted percentages are presented along with the observed percentages of correct reports for each of the cells in Table 5. As can be seen from that table, the predicted percentages of correct reports range from a high of 75% (rotation group 1, white, self-reporter, with a January Social Security payment of over \$413) to a low of 26% (rotation group 3, non-white, proxy informant, with a January Social Security payment of \$413 or less). The observed percentages are generally close to the predicted percentages.

Given that a respondent failed to report

Parameters	d.f.	Wald statistic	P-value
Intercept	1	0.04	0.851
Rotation groups	2	108.76	< 0.001
Race	1	20.04	< 0.001
Interview status	1	80.08	< 0.001
January Social Security payment	1	9.24	0.002
Lack of fit	18	18.64	0.414

Table 4. Analysis of variance for the logistic regression model

Table 5. Weighted observed and predicted percentages of correct reports of the 3.5% January increase in Social Security payments

Rotation		Interview	January	Percentage the Januar		
group	Race	Observed	Predicted			
				%	%	
1	White	Self	Over \$413	73	75 7 5	
		Decree	\$413 or less Over \$413	69 67	70 62	
		Proxy	\$413 or less	52	57	
	Non-white	Self	Over \$413	57	65	
			\$413 or less	63	61	
		Proxy	Over \$413	60	51	
			\$413 or less	42	47	
2	White	Self	Over \$413	64	65	
			\$413 or less	60	61	
		Proxy	Over \$413	51	51	
			\$413 or less	44	46	
	Non-white	Self	Over \$413	49	54	
			\$413 or less	52	50	
		Proxy	Over \$413	48	40	
			\$413 or less	26	36	
3	White	Self	Over \$413	55	54	
			\$413 or less	55	49	
		Proxy	Over \$413	45	40	
			\$413 or less	32	35	
	Non-white	Self	Over \$413	43	43	
			\$413 or less	31	39	
		Proxy	Over \$413	0	30	
			\$413 or less	16	26	

the 3.5% increase in January, the question arises as to whether that increase appears at some other time, such as the preceding or succeeding seam. Table 6 presents evidence on that issue. The table gives for each rotation group the percentage of respondents who reported an increase of around 3.5% in some other month among those who failed to report an increase in January (columns (a)) and it also gives comparable percentages for those who did report an increase of around 3.5% in January (columns (b)). Overall, of those who reported the 3.5% increase in January, some 9% also reported an increase of this magnitude at the previous seam and some 7% also reported such an increase at the subsequent seam. For those who failed to report the increase in January, the corresponding percentages are appreciably larger at 27% and 11%. It appears that a sizeable number of the January increases are appearing at an adjacent seam, mainly the previous one. The percentages reporting an increase at the preceding seam among those failing to report the January increase differ markedly by rotation group, ranging from 19% for rotation group 1 to 35% for rotation group 3. Shifting the change to the previous seam thus appears to be more likely the greater the time interval between the occurrence of the change and the date of interview. Another finding in Table 6 is that some 7% of those who failed to report the increase in January reported an increase of around 3.5% at some other time within the second wave, but none of those who reported the January increase did so. It therefore seems likely that some of those who failed to report the increase in January misplaced the date of the increase within the wave.

5. Discussion

The causes of the seam effect have not been clearly identified. One possible explanation is that the excess changes at the seam are a manifestation of the general problem of measuring gross changes in panel surveys. Measures of gross changes between waves of a panel survey are generally overstated because of changes in measurement errors between the waves (Kalton, Kasprzyk, and McMillen 1989). Another possible explanation is that the fewer changes within a wave are the result of a false consistency of withinwave reporting. Respondents may give the same answers for each month because they have forgotten that a change occurred during the four-month reference period or simply

Table 6. Percentages of respondents reporting increases of around 3.5%* in Social Security payments at various months for (a) those reporting no increase in January and (b) those reporting an increase of around 3.5% in January, by rotation group

	Rotation group							
Panel months	1		nonths 1 2		3		Total	
4-5 Within the second wave (excluding Dec./Jan.) 8-9	(a) % 18.6 5.0 14.9	(b) % 8.3 0.0 7.1	(a) % 25.6 9.9 8.7	(b) % 8.5 0.0 7.3	(a) % 34.6 5.2 10.2	(b) % 10.2 0.0 6.6	(a) % 27.3 6.7 11.0	(b) % 8.9 0.0 7.0
Number of respondents	478	919	578	756	706	635	1,762	2,310

^{*}An increase of between 2.0% and 4.1%.

because repeating the same answer requires less effort. Based on their record check study, Marquis and Moore (1989) conclude that both these explanations operate, that is, that there is both an overstatement of changes between waves and an understatement of changes within waves.

The analyses of the reporting of the January 1984 increases in Social Security payments presented in Section 4 lend support to false consistency within a wave as a partial explanation of the seam effect for this variable. Over one-third of Social Security recipients failed to report the increase as taking place in January, and the extent of the failure to report the increase rose with the interval between January and the month of interview. A fair proportion of those who failed to report the increase in January did, however, report an increase of around 3.5% at one of the adjacent seams, mostly the earlier one. These findings are consistent with a reporting behavior of giving the amount for the latest month, and then reporting the same amount for the preceding three months. Such behavior would produce stable reports within the wave and lead to excess changes being reported at the preceding seam.

Determining the causes of the seam effect is important in order to guide the search for a solution. If false consistency is indeed a major cause, then some form of dependent interviewing may be a remedy. One form of dependent interviewing would be to first ask the respondent for data relating to the latest month of the current wave, and then to provide the respondent with the data reported for the last month of the previous wave. Armed with these fixed endpoints, the respondent may then be asked to provide the data for the intervening months. The U.S. Bureau of the Census is engaged in various studies of the seam effect (Petroni, Huggins, and Carmody

1989), one of which involves the use of dependent interviewing.

References

Burkhead, D. and Coder, J. (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., Burkhead, D., Feldman-Harkins, A., and McNeil, J. (1987). Preliminary Data from the SIPP 1983–84 Longitudinal Research File. SIPP Working Paper No. 8702. Washington, D.C.: U.S. Bureau of the Census.

Computer Support Group (1984). OSIRIS IV User Guide, 7th edition. Ann Arbor, Michigan: Institute for Social Research.

Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). Analysis of Categorical Data by Linear Models. Biometrics, 25, 489-503.

Hill, D. (1987). Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods. Proceedings of the Section on Survey Research Methods, American Statistical Association, 210–215.

Jabine, T.B., King, K.E., and Petroni, R.J. (1990). Survey of Income and Program Participation (SIPP): Quality Profile. Washington, D.C.: U.S. Bureau of the Census.

Kalton, G., Kasprzyk, D., and McMillen,
D.B. (1989). Nonsampling Errors in
Panel Surveys. In Panel Surveys, eds. D.
Kasprzyk, G. Duncan, G. Kalton, and
M.P. Singh, 249–270. New York: Wiley.

Kalton, G., Lepkowski, J., and Lin, T.-K. (1985). Compensating for Wave Non-response in the 1979 ISDP Research Panel. Proceedings of the Section on Survey Research Methods, American Statistical Association, 372–377.

- Kasprzyk, D. (1988). Research Issues in the Survey of Income and Program Participation. Survey Methodology, 14, 45–58.
- Koch, G.G., Freeman, D.H., and Freeman,
 J.L. (1975). Strategies in the Multivariate
 Analysis of Data from Complex Surveys.
 International Statistical Review, 43,
 59-78.
- Landis, J.R., Stanish, W.M., Freeman, J.L., and Koch, G.G. (1976). A Computer Program for the Generalized Chi-Square Analysis of Categorical Data Using Weighted Least Squares (GENCAT). Computer Programs in Biomedicine, 6, 196–231.
- Marquis, K. and Moore, J. (1989). Response Errors in SIPP: Preliminary Results. Proceedings of the U.S. Bureau of the Census Fifth Annual Research Conference, 515-536.

- Moore, J. and Kasprzyk, D. (1984). Monthto-Month Recipiency Turnover in the ISDP. Proceedings of the Section on Survey Research Methods, American Statistical Association, 726–731.
- Petroni, R., Huggins, V., and Carmody, T. (1989). Research and Evaluation Conducted on SIPP. Proceedings of the U.S. Bureau of the Census Fifth Annual Research Conference, 308–338.
- Weidman, 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.

Received December 1990 Revised April 1991