THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION

THE SEAM EFFECT IN PANEL SURVEYS

No. 116

G. Kalton, D. Hill and M. Miller
University of Michigan

U. S. Department of Commerce BUREAU OF THE CENSUS

This report is the final report for Joint Statistical Agreement 87-5 between the Bureau of the Census and the Survey Research Center, University of Michigan.

TABLE OF CONTENTS

INTRODUCTION 1

1. THE SEAM EFFECT WITH SOCIAL SECURITY INCOME IN THE SURVEY OF INCOME AND PROGRAM PARTICIPATION
Graham Kalton and Michael E. Miller 3

2. RESPONSE ERRORS AROUND THE SEAM: ANALYSIS OF CHANGE IN A PANEL WITH OVERLAPPING REFERENCE PERIODS
Daniel H. Hill 17

3. A POISSON MODEL OF RESPONSE AND PROCEDURAL ERROR ANALYSIS OF SIPP REINTERVIEW DATA
Daniel H. Hill 37

Introduction

This report contains the findings of research conducted under a Joint Statistical Agreement between the Bureau of the Census and the Survey Research Center, University of Michigan. The Joint Statistical Agreement was entitled "Measuring Gross Change in Panel Surveys", and the research was conducted during the period 1987-88.

An important type of nonsampling error that has been identified in the Survey of Income and Program Participation (SIPP) is known as the seam effect. The SIPP is a panel survey with an interval of four months between waves, but 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 of most income types occur when the data are collected in different waves (e.g., between months 4 and 5, or 8 and 9) than when the data are collected in the same wave (e.g., months 1 and 2, 3 and 4, 6 and 7). This finding, termed the seam effect, can affect the measurement of gross change and measures of durations of spells on social programs. It has been a central focus of the research conducted under the JSA.

The report comprises three chapters. The first, by Kalton and Miller, examines the seam effect in relation to the monthly amounts of Social Security payments reported 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.

Chapter 2, by Hill, investigates the seam effect for several characteristics in the 1984 SIPP Panel, and also for characteristics in the Panel Study of Income Dynamics (PSID). The conclusion reached is that the seam effect in PSID is at least as severe as that in SIPP. The paper also reports findings on some of the correlates of the propensity to provide inconsistent reports which give rise to the seam effect.

As part of the SIPP quality control program, a small subsample of SIPP respondents is reinterviewed each month. The aim is to evaluate and to indicate when retraining is required. Although the reinterview program is not designed to provide evidence on

nonsampling errors, it has the potential to do so. Chapter 3, by Hill, explores the use of the reinterview data for investigating nonsampling errors in the SIPP. The chapter demonstrates that the reinterview data can be useful for this purpose, and suggests some changes to the program that would improve its utility for nonsampling error research.

Chapter 1

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

Graham Kalton and Michael E. Miller

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, 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.

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 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 *et al.* (1987), and Weidman (1986) in relation to SIPP, and

by Moore and Kasprzyk (1984) and Kalton *et al.* (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 *et al.*, 1989). Hill (1987) reports the occurrence of a similar seam effect with the Panel Study of Income Dynamics.

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 rounds 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. The first rotation group was interviewed for the first time in October 1983, and then reinterviewed in February 1984, June 1984, etc. The second rotation group was first interviewed in November 1983 and reinterviewed in March 1984, July 1984, etc. Similarly, the third and fourth rotation groups were first interviewed in December 1983, and January 1984, respectively, and then reinterviewed at four-monthly 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 1984 were collected in different waves for the first rotation group (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 system 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 fourth rotation group 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 (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. Also excluded are a small number of extreme amounts of $1500 or more of Social Security income in a single month.

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 $1500 or more and changes of more than $200 between months have been excluded (ten records in the subsample had amounts of $1500 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 3000

persons who reported amounts of Social Security income in both of the two months involved (excluding imputed values and extreme values as noted above).

Table 1
Cross-Month Correlations for Social Security Income Amounts

1	2	3	4	5	6	7	8	9	10	11
0.99										
1.00	0.99									
0.99	0.98	0.99								
0.92	0.92	0.92	0.92							
0.92	0.91	0.92	0.92	0.99						
0.92	0.91	0.92	0.92	0.99	0.99					
0.93	0.92	0.93	0.92	0.99	0.99	1.00				
0.92	0.91	0.92	0.92	0.94	0.93	0.94	0.94			
0.92	0.91	0.92	0.92	0.94	0.94	0.94	0.94	1.00		
0.92	0.91	0.92	0.92	0.94	0.93	0.94	0.94	0.99	1.00	
0.92	0.91	0.92	0.92	0.93	0.93	0.94	0.94	0.99	1.00	1.00

The correlations in Table 1 exhibit the same pattern that Kalton *et al.* (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 Panel (apart from the exclusions noted in 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 receiving Social Security income in that month and the preceding one.

Table 2
Percentage Change in Amount of Social Security
Income in Current Month Compared to Previous Month

Month	Within (W) or Between (B) wave	Percent change from previous month					Total	Sample size
		Reduction			Increase			
		More than 10%	10% or less	No change	10% or less	More than 10%		
September	W	0.2	0.1	99.1	0.3	0.3	100.0	4917
October	W	0.1	0.2	99.2	0.3	0.2	100.0	3285
	B	6.4	21.4	36.0	27.6	8.6	100.0	1510
November	W	0.2	0.9	97.8	0.7	0.4	100.0	3257
	B	5.9	23.0	29.5	34.2	7.4	100.0	1496
December	W	0.3	0.5	97.2	1.5	0.5	100.0	3221
	B	6.2	24.3	23.1	38.9	7.5	100.0	1491
January	W	0.3	1.0	36.6	60.5	1.6	100.0	4809
February	W	0.2	0.3	96.7	2.5	0.3	100.0	3191
	B	6.3	21.7	36.1	29.5	6.4	100.0	1475
March	W	0.1	1.2	97.6	0.9	0.2	100.0	3157
	B	5.4	20.4	41.7	26.3	6.2	100.0	1451
April	W	0.1	0.2	99.0	0.3	0.4	100.0	3113
	B	6.1	18.7	40.4	27.4	7.4	100.0	1440
May	W	0.3	0.2	99.1	0.1	0.3	100.0	4650
Average	W*	0.2	0.4	98.3	0.8	0.3	100.0	
	B	6.0	21.6	34.4	30.7	7.3	100.0	

*Excluding January

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 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

Table 3
Percent Change from December to January by Rotation Group

	Group 1 %	Group 2 %	Group 3 %
Reduction			
> 10%	0.3	0.4	0.2
≤ 10%	0.2	2.4	0.3
No change	29.4	35.8	45.0
Increase			
≤ 10%	68.4	59.6	53.0
> 10%	1.7	1.8	1.5
Total	100.0	100.0	100.0
Number of respondents	1626	1614	1569

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 2310 "correct" reporters, 1762 "incorrect" reporters (that is persons who reported no increase from December to January), and 737 reporters for whom it was 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 self reporter or proxy informant.

The logistic regression analyses employed the approach described by Koch *et al.* (1975) for 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 *et al.* (1969). Wald statistics were generated in GENCAT (Landis *et al.*, 1976) to test hypotheses about the relationship of the predictor variables to the logits.

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

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

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.

Table 4
Analysis of Variance for the Logistic Regression Model

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

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.

Table 5
Weighted Observed and Predicted Percentages of Correct Reports of the 3.5% January Increase in Social Security Payments

Rotation Group	Race	Interview Status	January Payment	Percentage Reporting the January increase	
				Observed	Predicted
				%	%
1	White	Self	Over $413	73	75
			$413 or less	69	70
		Proxy	Over $413	67	62
			$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

Given that a respondent failed to report 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.

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

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

*An increase of between 2.0 and 4.1%.

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 *et al.*, 1989). Another possible explanation is that the fewer changes within a wave are the result of a false consistency of within-wave 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 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 Bureau of the Census is engaged in various studies of the seam effect (Petroni *et al.*, 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. U.S. Bureau of the Census, Washington, D.C.

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

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. (1989). *Survey of Income and Program Participation (SIPP): Quality Profile*. U.S. Bureau of the Census, Washington, D.C.

Kalton, G., Kasprzyk, D. and McMillen, D.B. (1989). Nonsampling errors in panel surveys. In *Panel Surveys*, D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh (eds.), 249-270. Wiley, New York.

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

Kasprzyk, D. (1988). *The Survey of Income and Program Participation: An Overview and Discussion of Research Issues*. SIPP Working Paper No. 8830. U.S. Bureau of the Census, Washington, D.C.

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 Bureau of the Census Fifth Annual Research Conference*, 515-536.

Moore, J. and Kasprzyk, D. (1984). Month-to-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 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.

Chapter 2

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

Daniel H. Hill

1. Introduction

We have seen repeated evidence in the SIPP, and in its predecessor the ISDP, that between-wave change dominates within-wave change.[1] Most analysis, to date, has been largely descriptive 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 the 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 is how these errors might affect estimates of models intended to explain the dynamics of welfare participation and employment. Hill and Hill (1986) have found that in the context of a proportional hazards model of transitions to employment estimated with SIPP data, whether or not the week of the transition was a seam

---

[1]This research was sponsored, in part, by a Joint Statistical Agreement (JSA 87–5) between the United States Bureau of the Census and the Survey Research Center of the University of Michigan. The current paper is an extention of a similar paper presented and the 1987 American Statistical Association meetings. The author would like to thank Dan Kasprzyk, Graham Kalton and Charlie Brown for their helpful suggestions 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.

week was the single most important predictor. 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 the causes of reporting inconsistencies which tend to amplify or attenuate between- relative to within-wave transitions.

2. A Comparison of SIPP and PSID Recipiency Transitions

SIPP Methodology

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 was received (and how much) in the calendar month prior to the interview, then for the month prior to that, etc. until the reference period is complete.

SIPP Seams: Unemployment Compensation

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 Figure 1. To make comparisons 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 experienced their first 'seam' between December 1983 and January 1984. The figure shows a pronounced 'bulge' in reported exits from unemployment compensation programs during this seam period—approximately twice as many people exited at this time than at any other time. Previous analysis show this same pattern appears for all calendar months. The corresponding bulge for entrances, while still quite noticeable, is less dramatic.

PSID Methodology

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 has been collecting information from the same families (and the descendents of these families) annually since 1968. The interviewing is conducted in the spring and summer 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 we do not even try to obtain monthly amounts—only annual total amounts and monthly receipency 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 continuous spell of recipiency.

PSID Seams: Unemployment Compensation

Figure 2 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 Figure 2 are surprisingly close to the corresponding SIPP pattern of Figure 1. The PSID, in general, appears to have somewhat less within-wave transition and a markedly more pronounced bulge in exits from unemployment compensation at the seam than the SIPP. Otherwise, however, the patterns of monthly transitions from the two studies are quite comparable.

Seams in Foodstamp Receipency

This same general conclusion holds for Foodstamp recipiency, as examination of Figures 3 and 4 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 some 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.

Relative Frequency of Seams

While there is no evidence in these data to suggest that the PSID methodology results in any better quality data than the SIPP methodology, there are some differences in the

Figure 1
Unemployment Compensation Transitions
(Percent by Month)
SIPP

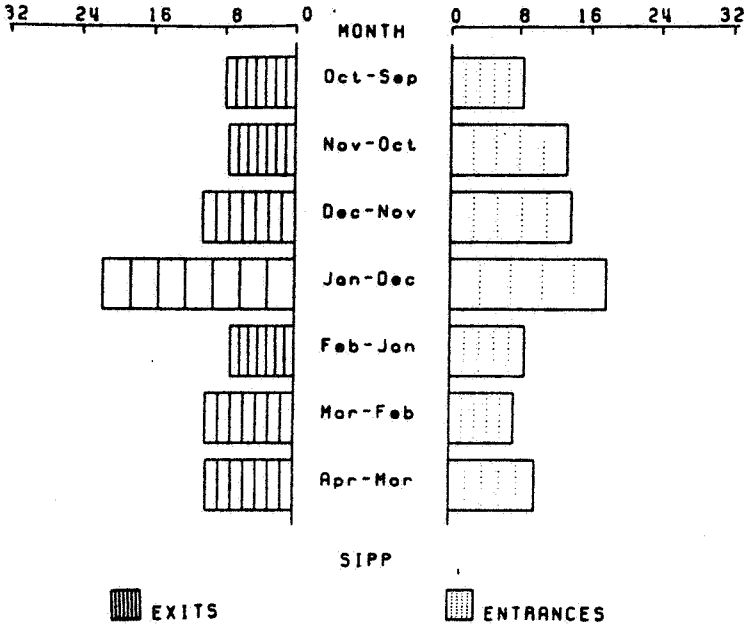


Figure 3
Food Stamp Recipiency Transitions
(Percent by Month)
SIPP

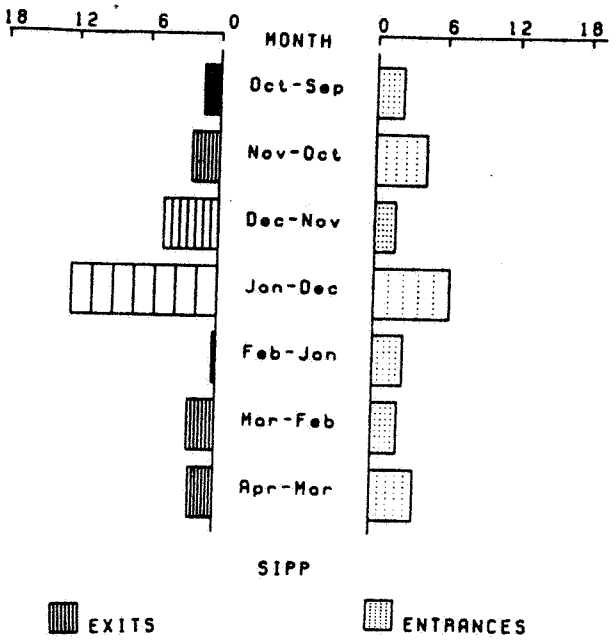


Figure 2
Unemployment Compensation Transitions
(Percnet by Month)
PSID

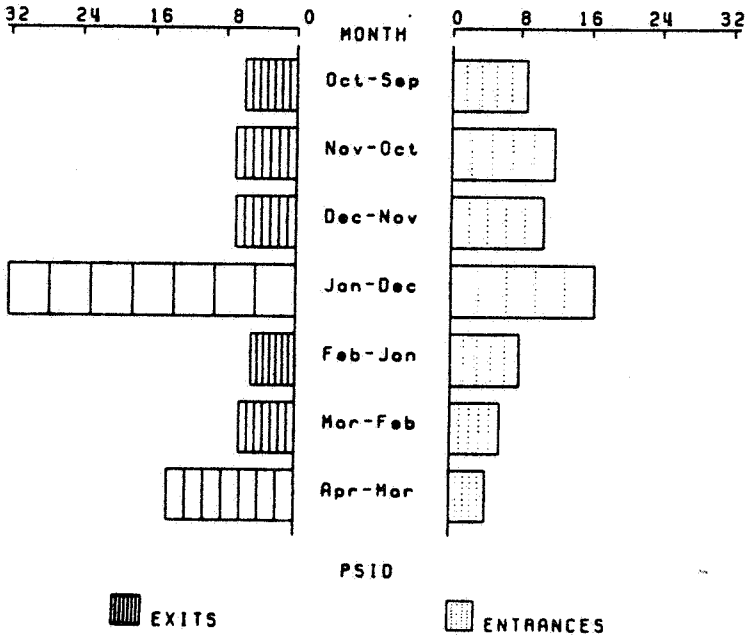
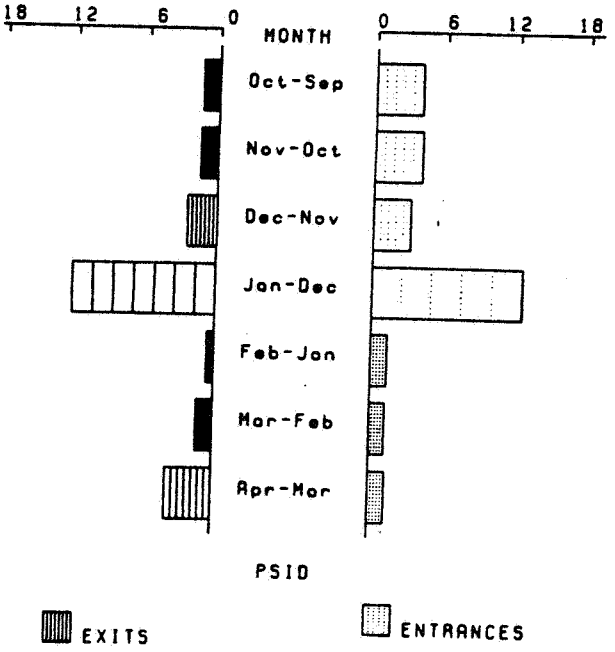


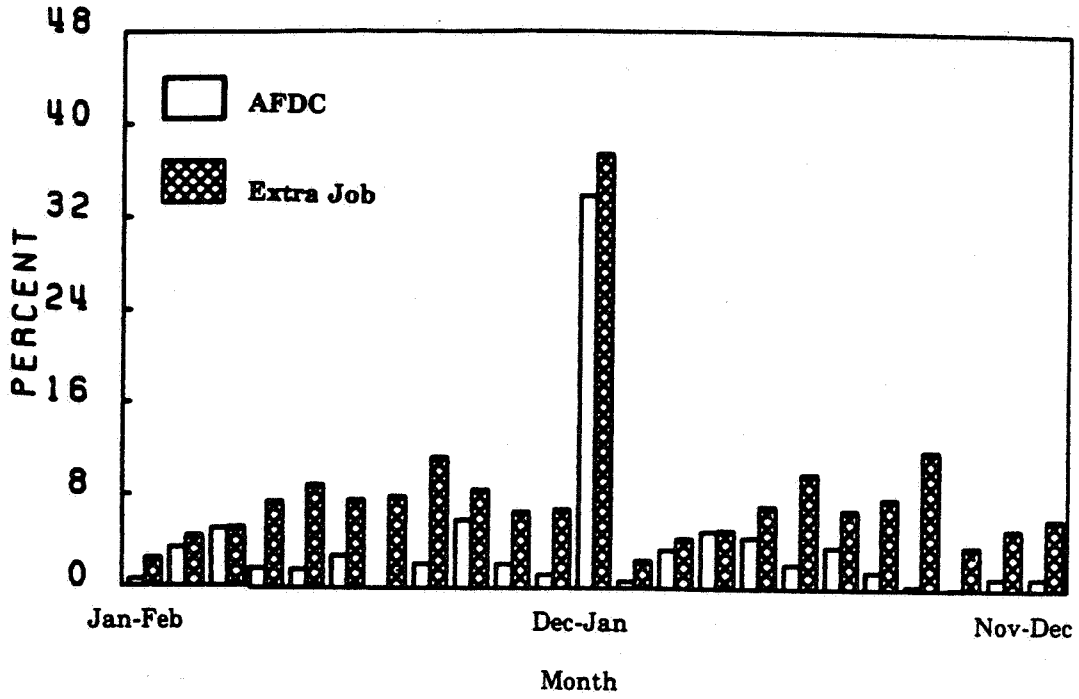
Figure 4
Food Stamp Recipiency Transitions
(Percent by Month)
PSID



importance of between- and within-wave transitions between the two studies. The most important is that, by design, the PSID has fewer seams than the SIPP. This can be seen quite clearly in Figure 5 which plots the percent of the PSID individuals reporting transitions in secondary employment and in AFDC receipt, by month.[3] There are twenty-two within-wave transitions for each income source, and one between-wave transition occurring between December 1983 and January 1984. Again, we see that the seam problem is less severe for the short-term individual specific measure (extra jobs) than for the longer term family level variable (AFDC).

If the source of the 'seam problem' is an exorbitant amount of between-wave change, then the PSID methodology may be superior. There is some evidence, however, that at least part of the cause of the seam problem is too little within-wave change.[4] In this case the more frequent interview schedule of the SIPP may be a definite advantage.

Figure 5
Within- versus Between-Wave Transitions in PSID
(Percent by Month)



[3]The populations of inference for Figure 3 are, for the extra jobs figures, all individuals who had some secondary or 'extra' job in the two year period January 1983-December 1984 and, for the AFDC figures, all individuals receiving AFDC in at least one month during the same two-year period. significant under-reporting of the January 1984 increase in Social Security benefit levels within waves for rotation groups 1–3 of the 1984 SIPP panel.

[4]Kalton and Miller (1987) find significant under-reporting of the January 1984 increase in Social Security benefit levels within waves for rotation groups 1–3 of the 1984 SIPP panel.

Conclusion

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. 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.

3. Correlates of Reporting Inconsistencies Leading to Seam Transitions

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.

Dual Employment Status Reports

Data on employment status in this latter month were collected during both the 1984 and 1985 interviewing years. 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.[5] 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.[6]

Overall, the figures in Table 1 suggest substantial agreement in reports from the two interviewing years. The simple response variance 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 1984 that they were employed in January, also reported that they were employed in January of 1984 when asked about it in

---

[5]Because the study began in 1968 we originally used the now archaic and admittedly sexist 1960 Census definition of Head of Household in our original design. Furthermore, since the PSID is a panel study, we cannot deviate from our original design if we wish to maintain its longitudinal value.

[6]More precisely anyone either retired, permanently disabled, keeping house, or a student, and who was not working at least ten hours a week at the time of the 1984 interview was 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).

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

1985 Report	1984 Report			
	Employed	Unem-ployed	Out of L.F.	Mixed
Employed	6,039	207	40	5
Unemployed	94	286	28	7
Out of L. F.	88	75	64	5
Mixed	5	12	2	1

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.

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.

Identifying and Modeling Inconsistent Seam Transitions

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 (1985),[7] since spurious between wave transitions and response variance are closely related. A more direct approach, however, involves concentrating only on those cases reporting transitions (either within a wave or between waves) and examining the extent of agreement in between and within-wave transitions.

[7]O'Muircheartaigh employed CPS interview/reinterview data in his analysis of response variance in reports of employment status.

There are three possible outcomes in this case. These are illustrated in Table 2. 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' which has most concerned analysts in the past, are nevertheless, reflective of poor response quality.

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.

Empirical Model

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 <0,1,2> 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 are composed of systematic and stochastic components. The systematic portion of the response propensity consists of the effects of a series of exogenous measured factors X while the stochastic portion, denoted μ, reflects the effects unmeasured excluded factors and chance. As a first order approximation, the response propensities can be expressed as:

Table 2
Patterns of Inconsistency in Overlapping Reports
Heads and Wives with Either Between- or Within-Wave Transitions

	December 1983	January 1984	Reporting Year
Consistent Reports		(n= 425)	
	1	0	84
		0	85
	0	1	84
		1	85
Seam-Amplifying Inconsistencies		(n=454)	
	1	1	84
		0	85
	0	0	84
		1	85
Seam-Attenuating Inconsistencies		(n=143)	
	1	0	84
		1	85
	0	1	84
		0	85

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

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:

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

The factors (X) assumed to affect the propensity to provide reports of varying quality are of four types. The first type are factors which affect the difficulty of the recall task the

respondent is being asked to perform. These 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. 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 second 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. Both of these factors are thought to raise the propensity of the respondent providing faulty information. The third 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-completion 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 final set of factors are included as controls and consists of basic demographic measures. These are race (whether Black), education, age, gender (whether male), and income.

Results

The results of the multinomial logit analysis of within-and between-wave transition inconsistency for employment status are presented in Table 3. The coefficients for consistent transition reports are normalized to zero, and the coefficients presented for the two types of inconsistencies can be thought of as deviations from the effects of the factors on consistent reports. Positive coefficients, therefore, represent adverse effects of the corresponding factor on data quality.

Demographic Factors

The first pair of columns in Table 3 correspond to the model which includes demographic factors only. Whether or not there are systematic associations of these factors with the propensity to provide erroneous transition reports is particularly important because most event history analyses will include these factors as predictors. Indeed, in many cases the major motivation in analyzing the micro-dynamics is to understand better the reasons for persistent differences in experiences of various demographic subgroups of the population. 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 transition exaggerating inconsistent reports. Blacks are far more likely to provide inconsistent reports which serve to amplify between-wave transitions than are non-Blacks.[8] 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 events into memory as well as the efficiency with which they retrieve data 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 simple response variances in reinterview data for the CPS than younger respondents. This is in direct conflict with our findings presented in Table 3. There are several differences between his analysis and ours which might account for the conflicting results. Perhaps most 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

[8]See Hill and Hill (1986) for a comparison of proportional hazards models of re-employment transitions estimated on SIPP data with seam transitions and PSID data without. Race effects were found to be much strong in the latter study. Thus true racial differences in re-employment probabilities may be being obscured in the SIPP by the erroneous seam transition reports.

casual employment situations and this type of employment might be particularly prone to mis-reporting.

Similarly, 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.

Difficulty of Task and Other Controls

While the differences between O'Muircheartaigh's and our age effects are easily explained by differences in procedures, they also suggest what is a difficult problem in studies such as ours where there is no independent validating data. This is that without such data we cannot tell to what extent older respondents, for instance, are providing better reports (as O'Muircheartaigh suggests) or that they just have nothing of interest on a particular topic to report. If they do not, then their reporting task is trivial unless they happen to be in some nebulous transitory or casual employment situation.

If age and race are truly responsible for lower quality of data for older and Black respondents, then the estimated coefficients on these variables should not be greatly reduced when 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 columns three and four of Table 3 indicate that this is the case. Indeed, the effects of age and race are slightly increased by the inclusion of measures intended to capture the effects of task, and cognitive ability on response quality. On the other hand, 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 3). 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

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

Variable	Demographic Controls Only		Demographic & Task Controls		Whether Inconsistent
	Attenuating Incon-sistencies	Amplifying Incon-sistencies	Attenuating Incon-sistencies	Amplifying Incon-sistencies	
Constant	−.723 (.717)	−.567 (.513)	.344 (1.11)	.334 (.776)	.930 (726)
Demographics					
Whether Black	.050 (.214)	.577** (.148)	.163 (.235)	.631** (.017)	.528** (.153)
Age (decades)	−.072 (.093)	.167** (.062)	−.168 (.115)	.024** (.008)	.014* (.007)
Whether Male	.327+ (.197)	−.133 (.138)	−.569* (.230)	.004 (.159)	−.138 (.148)
Education	−.027 (.048)	.001 (.033)	−.032 (.051)	−.002 ((.036)	−.011 (.033)
Income ($1,000)	.016* (.007)	−.009 (.006)	.016* (.008)	−.008 (.006)	−.009 (.054)
Difficulty of Task					
Extra Jobs			.272 (.271)	.117 (.208)	.167 (.191)
Months with Current Employer			−.062 (.053)	−.153** (.037)	−.132** (.034)
Construction			−.057 (.048)	−.002 (.034)	−.012 .031)
Self Employed			.590+ (.323)	−.014 (.252)	.147 (.234)
Self Reports			−.219 (.214)	−.254+ (.155)	−.236 .144
Length of Recall			−.085 (.101)	−.068 (.069)	−.072 (.065)

Table 3 (Continued)

Variable	Demographic Controls Only		Demographic & Task Controls		Whether Inconsistent
	Attenuating Incon-sistencies	Amplifying Incon-sistencies	Attenuating Incon-sistencies	Amplifying Incon-sistencies	
# Intervening Transitions			−.024 (.073)	−.064 (.051)	−.056 (.048)
Interview Characteristics					
Initial Refusal			.355 (.625)	−.178 (.513)	−.027 (.463)
Length of Interview			.061 (.056)	−.018 (.039)	−.001 (.036)
Cognitive Ability					
Test-Score			.025 (.052)	−.054+ (.037)	−.032 (.035)
Test-Score x Not '72 Respondent			−.038 (.033)	.036+ (.025)	.017 (.023)
χ^2 (d.f.)	49.7 (10)		92.4 (32)		45.1 (16)

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

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.

Cognitive Factors

There is 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,[9] 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.[10]

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 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.

Structural Dissimilarity of Inconsistency Types

Before concluding our discussion it is worthwhile to consider the question of whether there is a common structure to the determinants of the two types of inconsistencies we have identified. If they do share a common structure, then their effects may tend to cancel each other out in structural analysis, and their net effect may only be to reduce measures of goodness-of-fit. Casual comparisons of the coefficients in columns three and four are not very useful. Some factors appear to affect the propensities in opposite directions (e.g. race) which would imply that these errors would reinforce each other in biasing structural parameter

---

[9]The remaining eighty percent of 1985 respondents are composed primarily of children or spouses of the 1972 respondent.

[10]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.

estimates. Other factors appear to affect the response propensities in the same direction (e.g. tenure with current employer)—something which would attenuate their net effects on structural parameter estimates.

A formal test of whether the two types of response errors are reinforcing or off-setting is possible. Specifically, we can constrain the effects to be equal by re-analyzing the model using a dummy dependent variable for whether either type of inconsistency occurs and compare its goodness-of-fit with that from the unconstrained model. If there is a significant reduction in the χ-square statistic then the joint structure hypothesis can be rejected. The results of this analysis are presented in the final column of figures in Table 3.[11] The χ-square statistic declines significantly from 92.4 with 32 degrees of freedom to 45.1 with 16 degrees of freedom. Thus we can be confident that the effects of the two types of response errors on structural model estimates (e.g. those of a proportional hazards model of unemployment) will not be offsetting.

Summary

In order to summarize our findings, and to provide the reader with a more intuitive appreciation of the size of effects we do find, Table 4 presents the results of simulations based on the coefficients presented in columns three and four of Table 3 for selected predictors. Because the model is non-linear we perform these simulations by calculating predicted values of the probability of each response separately for each respondent and then averaging these probabilities across respondents. The simulations are performed first using the actual values of the X's and then adding to each X separately an amount corresponding to (1/100th) of a standard deviation. The resulting change in the predicted probabilities is then scaled by (1/100th) of the standard deviation of the dependent variable. The resulting coefficients presented in Table 4 are therefore analogous to standardized regression coefficients.[12]

Using these measures, we see that the most important predictor of the propensity of respondents to provide seam transition attenuating reports is income, with a one standard deviation ($14,472) increase resulting in a .0967 standard deviation, or 3.35 (=.0967*.3469) percentage point, increase in the probability of providing such a report. This effect is followed very closely by gender with standardized coefficient for males being −.0953. For the

[11]Such a model is analogous to an analysis of simple response variance or gross-difference rates for the population of individuals reporting a transition (either between or within waves) between December 1983 and January 1984.

[12]Unlike standardized regression coefficients, however, there is not a one-to-one relationship between the relative size of these coefficients and other measures of predictor importance such as t-ratios.

Table 4
Simulated Standardized* Effects of Various Factors on
Between-Wave Attenuating and Amplifying Inconsistencies in Reported Employment Status

Variable	Standard Deviation	Consistent Reports	Attenuating Inconsistencies	Amplifying Inconsistencies
Base Probabilities (Standard Dev.)		.4159 (.4929)	.1399 (.3469)	.4442 (.4969)
Whether Black	.4988	−.1205	−.0248	.1368
Age	12.02	−.0783	−.1154	.1582
Whether Male	.5002	.0332	−.0953	.0336
Income	14,472	.0131	.0967	−.0805
Length of Employment	2.15	.1315	.0097	−.1372
Self-Employment	.3162	−.0207	.0632	−.0236
Self-Report	.4803	.0549	−.0153	−.0438
Test Score	2.448	.0387	.0425	−.0681

*Estimated effects are analogous to standardized regression coefficients (i.e. they reflect the number of standard deviations the dependent variable changes in response to a one standard deviation increase in the independent variable). Raw score effects can be obtained by multiplying the above coefficients by the ratio of the standard deviations of the dependent to the independent variable.

propensity of respondents providing seam-transition amplifying reports, on the other hand, age, tenure of employment, and race are the most important factors. A one standard deviation increase in age (12.02 years) results in a .1582 standard deviation (i.e. a 7.86 percentage point) increase in the probability of seam amplifying reports. Employment tenure and race are of roughly equal power with Blacks and short-term employees having the highest propensities. All remaining predictors are of only tertiary importance in predicting response quality.

4. Summary

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 and presumably more directive 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. In addition, some hypotheses were noted without being tested. For one thing, the data suggest that the '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.

Nevertheless, the simple fact that there are systematic associations between various demographic factors and the propensity of respondents to provide inconsistent reports leading to seam problems is important. It means that micro-dynamic analyses such as those based on event history models are not justified in ignoring response errors. It also means that improved data collection methodologies need to be sought and tested.

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.

Chapter 3

A Poisson Model of Response and Procedural Error Analysis of SIPP Reinterview Data

Daniel H. Hill[1]

1. Introduction

As part of its ongoing quality control program the Field Division of the Census Bureau conducts reinterviews monthly with small samples of the Survey of Income and Program Participation (SIPP) respondents. The purpose of this reinterview program is to evaluate individual interviewer[2] performance to determine if retraining or dismissal is necessary. In addition to ascertaining whether the interview was actually conducted with the correct unit and whether the proper procedures were employed, the reinterview contains a small set of questions of substantive content. While it was never the intent of the reinterview program designers, the existence of the reinterview data makes estimation and analysis of nonsampling error in the SIPP possible. Such analysis is potentially important because it is quite apparent[3] that data from the SIPP are far from perfect.

The purpose of the present research is assess this potential by merging the reinterview data with public release data and analyzing the combined data. The paper is organized in three sections. In Section 2 the SIPP reinterview program is described in some detail. Section 3 presents a question-by-question description of response procedural and overall interview/reinterview discrepancies. Finally, in Section 4, two classes of multivariate models are developed and estimated.

---

[1]The author would like to thank Dan Kasprzyk, Fred Cavanaugh and Chet Bowie of the Census Bureau for making the data available and Laura Klem of the Survey Research Center for merging the reinterview data with the public release files. The author would also like to thank Dan Kasprzyk, Irv Schreiner, Vicki Stout, Gary Shapiro, and Jeff Moore of the Census Bureau and Jim Lepkowski and Graham Kalton of the Survey Research Center for their helpful comments on a preliminary draft of this report.

[2]As of September 1, 1989 Census Bureau interviewers are officially referred to as "Field Representatives." Throughout this report, however, the more functionally descriptive term 'interviewer' will be used to facilitate distinctions between them and 'reinterviewers.'

[3]This is not to say that SIPP data are in any sense more error prone than other survey data. The error that exists, however, is more easily seen because of the longitudinal nature of the data.

2. The SIPP Reinterview Program

The SIPP reinterview program is an ongoing systematic operation which is intended to monitor data quality by checking the interviewers' work. The sample to be reinterviewed each month is a multistage probability sample of current SIPP respondents. The sample selections are made monthly at the Regional Offices with instructions from the National Field Office in Suitland, Md. The first stage of sampling consists of partitioning the interviewers into twelve groups two of which are selected for reinterview each month. The selections are made by the national field office in Suitland. The second stage consists of randomly selecting a sample of the selected interviewers' sampling units. This is accomplished by selecting every 'n^{th}' unit from the Interviewer's Assignment and Control form beginning with the 'k^{th}' unit. If fewer than five units are selected subsequent passes through the listing are conducted until five units are selected. Both the selection interval 'n' and the random start number 'k' are determined by the national field staff and transmitted monthly to the Regional Offices. The final stage of the reinterview sample selection is to select one individual per unit for reinterviewing. This is accomplished by determining the number of individuals interviewed in the unit and using a random selection table to choose which of these individuals is to be interviewed.

The result of this sample selection procedure is that each individual interviewed in the main SIPP study has a probability of:

$$P_{it} = (1/6) * P_{it}^{U} * (1/fs_{it}^{U}) \qquad 1)$$

where P_{it}^{U} is the probability that individual i's unit, U, was selected in month t, given that his interviewer was, and fs_{it}^{U} is the number of individuals interviewed in that individual's unit in month t. P_{it}^{U} can vary from 1/3 to 1 depending on the number of units assigned to the interviewer.

The implication of equation 1) is that if inferences are to be made from the reinterview sample to the SIPP sample as a whole, the analyst will need to know a) the number of units assigned to each interviewer and b) the number of individuals interviewed by the interviewer in the selected unit. While it is theoretically possible to obtain measures from the public release data, they could not be obtained with complete accuracy and it would be quite

expensive.[4] Thus, *it would be helpful if these numbers could be transcribed to the Reinterview Questionnaire at the Regional Office.*

Once individuals are selected for reinterview, the Reinterview Questionnaire and Reconciliation Record (RQRR) is prepared. This is done by anyone familiar with the SIPP at the Regional Office other than the reinterviewer. This restriction is imposed so as to maintain the independence of the interview and reinterview responses. The preparation consists first of transcribing the identification codes and names of the individual to be reinterviewed, the interviewer, and the original respondent. Second, the "Office Check Items" are transcribed from the unedited original interview to the RQRR's Section 2. These items determine the question flow in both original interview and reinterview questionnaires.

Figure 1 illustrates the question flow for Section 2 of the Reinterview Questionnaire. The questions actually asked of the respondent in both the interview and reinterview are printed in bold, while the Office Check Items which are transcribed to the Reinterview Questionnaire from the original appear in normal print. Unless otherwise indicated, questions are asked in sequence. In most cases, however, respondents are skipped around certain questions and these skips are indicated in the figure by lines and arrows. If, in response to question 1, for instance, the respondent said he had a job for at least part of the reference period ('yes' on item 1.), he is skipped around the questions about whether he spent any time looking for a job (2a.), or whether he wanted a job (3a.), and is asked about whether he had a job each week of the reference period instead (4.). In Figure 1, a skip such as this which results from a response to a question asked in the reinterview study is depicted with a dotted line. Skips from Office Check Items, being automatic from the reinterviewer's point of view, are depicted as solid lines.

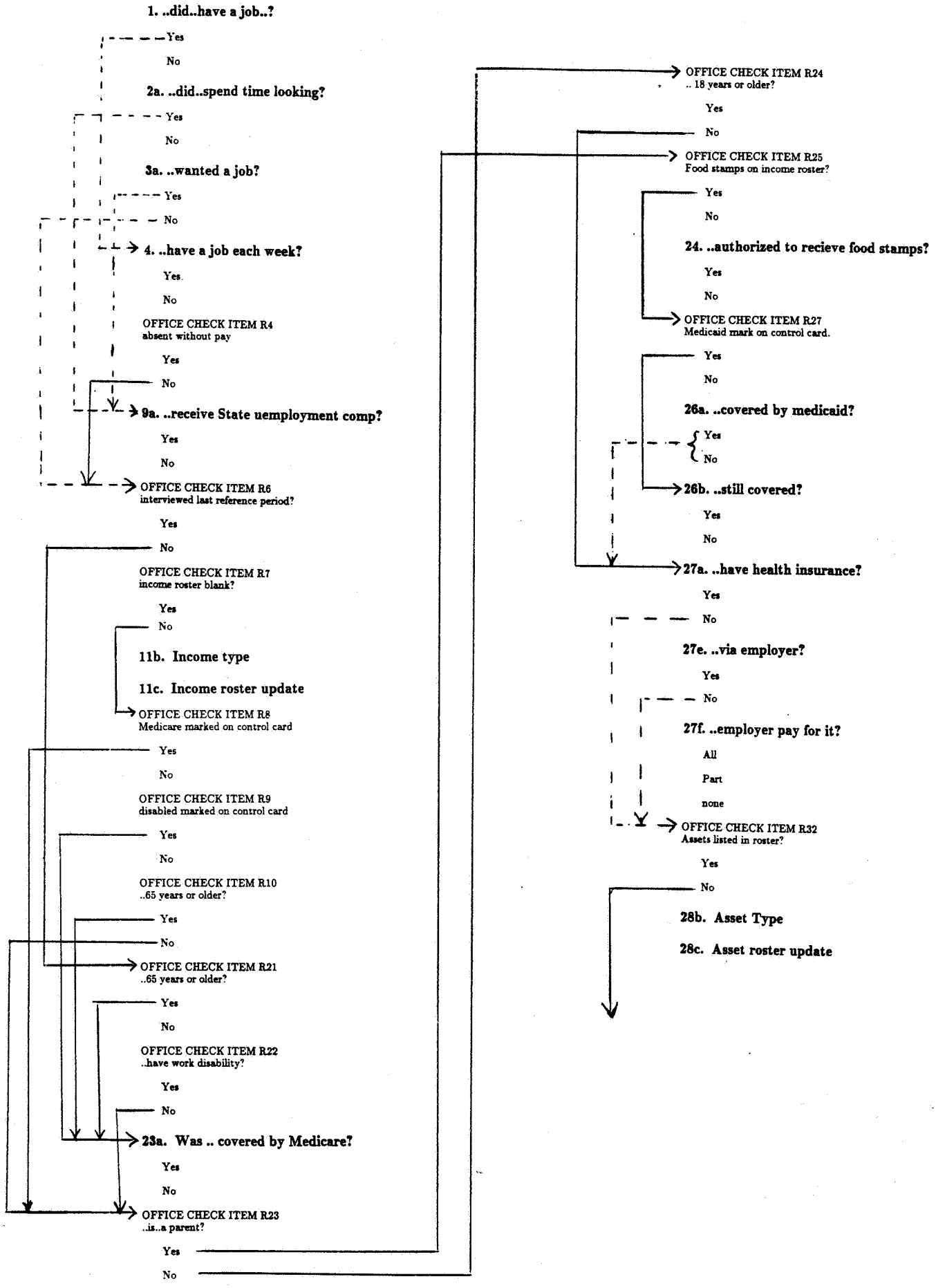
It does not take a great deal of study of Figure 1 to see that the skip sequences employed in the SIPP can be quite complicated. Indeed, a major goal of the reinterview program is to see if individual interviewers are following these skip sequences properly. It is important to note that the Office Check Items are transcribed from the original questionnaire before it is edited by the Regional Office staff. This is done so that the question flow employed by the

---

[4]One could obtain an estimate of the interviewer's assigned workload by sorting the sample unit file by interviewer ID and counting. Similarly an estimate of the number of individuals interviewed by the interviewer within the sample unit could be obtained by subtracting the number of children less than fourteen from the household size variable on the public release file. While sampling rates from these estimates would be preferable to those based on, for instance, average workloads within regional offices (since even within RO's workloads vary greatly), it would be far better if the actual numbers used in the reinterview selection procedure were recorded and passed on to the analyst.

Figure 1

SIPP Reinterview Questionnaire Flow



reinterviewer is the same as that which the interviewer used. Quite often Regional Office editing uncovers errors in the Check Items and consequent skip sequences. If these are sufficiently serious, the original interview is returned to the field so that missed questions can be asked of the respondent. These editing changes and 'send-backs' are done after the reinterview is completed.

The final task in preparing the reinterview questionnaire is to transcribe the original question responses to the 'reconciliation' portion (section 3) of the questionnaire. To help insure independence between the interview and reinterview responses, the reinterviewer is instructed not to look at these answers until after the questions have been re-asked.

When the materials are prepared the reinterview is assigned to the reinterviewer and is conducted by telephone. Once a respondent is contacted the reinterviewer records the time, date, mode, and person number of the reinterview respondent. Next the Control Card items for the selected sample individual are verified. First, and in many respects most importantly, the reinterviewer determines if the proper sample unit was actually visited by the original interviewer. Second, the reinterviewer ascertains if the living quarters, household composition, relationship to reference person, household membership status and birth date are properly recorded on the (photo-copy of the) Control Card.

Next, the reinterviewer begins the Labor Force and Recipiency portion of the reinterview (Section 2) which is as depicted in Figure 1. Only when this is completed does the reinterviewer turn to the Reconciliation section. At this point, the answers just obtained are transcribed by the reinterviewer to reconciliation section and are compared with the original responses. The respondent is then asked to help reconcile any discrepancies, and the reinterviewer records which of the two reports is judged to be correct.

After the reinterview is completed it is returned to the Regional Office where a summary report for each reinterviewer is compiled. On the basis of these reports reinterviewers are either congratulated, counselled, retrained or dismissed.

In the normal course of the reinterview program a summary report is prepared and these are analyzed on an annual basis by the Field Division. A special keying operation was conducted during the summer of 1987 to prepare the data from the 1984 panel's reinterview questionnaires for the analysis which follows.

3. Inconsistency Rates and Simple Response Variance Estimates

With the two independent observations provided by the interview and reinterview responses it is possible to estimate the simple response variance for the various questions.[5] To do so, we first confine our attention to that portion of the reinterview sample where a) the reinterview was successfully conducted and b) it was determined that the interviewer had visited the proper sample unit in conducting the original interview. We also eliminate from our sample those cases where the date of the original interview as recorded in the interview failed to match the date coded in the public release files,[6] and those few cases where, even though the reinterview was conducted, no substantive questions were re-asked. These restrictions leave us with a sample of 1,559 cases of interview/reinterview data for waves 2 and 3 of the 1984 panel.

In comparing interview and reinterview data we have a choice of using the pre-edited original interview information which was transcribed to Section 3 of the RQRR or the post-edited data which is available from the public release files. Evidently, however, not all the information from the original interview is transcribed to Section 3. Transcriptions are made only if a discrepancy is encountered. How discrepancies resulting from a question being skipped in one interview and not the other are treated is not clear. Thus we use instead original reports as recorded on the public release files and recognize that some of the discrepancies between interview and reinterview reports are due to edits and imputations performed subsequent to the original interview.

We can distinguish two distinct types of inconsistencies when the interview and reinterview reports do not agree—response inconsistencies and procedural or 'skip' inconsistencies. Response inconsistencies have been studied extensively in the CPS and quite elegant models of response variance have been developed (see e.g. O'Muircheartaigh, 1986). The underlying response model most commonly employed can be expressed as:

---

[5]To the extent that the respondent's reinterview response is affected by their memory of their response in the interview response errors in the two will tend to be positively correlated rather than independent. Thus, to this extent, the estimated response variances presented in the present analysis will tend to be conservative.

[6]Several hundred reinterviews for waves 2 and 3 of the 1984 panel were found to match on the basis of wave and entry identification numbers, but were found to have original interview dates which differed by roughly a multiple of four months. Apparently, the wrong reinterview schedule was employed for some subsequent waves of the reinterview program. While the content of the reinterview schedule remained the same throughout the panel, the form number changes each wave, and this form number is used as the wave identifier.

$$y_{it} = y_i + \beta_i + \epsilon_{it}$$

where y_{it} is the report provided by the ith respondent during the tth measurement (t=1 for interview, 2 for reinterview), y_i is the true value of y, β_i is the bias in individual i's reports, and ϵ_{it} is the random component to individual i's reports. The simple response variance is simply the variance of the ϵ_{it} across t. With categorical data such as we will be examining, response variance can be estimated as one-half the fraction of responses to a given question which differ between the interview and reinterview reports (*i.e.* one-half the gross difference rate).

We will reserve the term response variance or 'response inconsistencies' for estimates involving cases where the question was actually asked of the respondent in both the interview and reinterview and where a response was recorded. Given the complicated skip sequences employed, it should not be surprising that there are differences between the two reports not just in responses, but in whether or not the question was asked each time. Discrepancies between the interview and reinterview arising because a question was skipped in one and not the other will be referred to as 'procedural discrepancies'.[7],[8]

An example may be useful in clarifying these distinctions. Table 1 presents the recorded responses for the interview and reinterview for Item 4.—the question regarding receipt of state unemployment compensation. Actual responses in both interviews were recorded for only some thirteen percent (=100*207/1559) of the cases. Of these 2.9% (=100*(3+3)/207) of the reports were different. The simple response variance for this question is, therefore, .0145, or half the gross difference rate amongst those respondents who answered the question in both the interview and reinterview. We will define the procedural discrepancy rate as the simple gross difference rate for whether the question was skipped. For the unemployment compensation question results in Table 1, the procedural discrepancy rate is 6.54 percent (= 100*(7+59+7+29)/1559). The overall discrepancy rate is simply the fraction of the entire sample for which the interview and reinterview reports differ. It is equal to the sum of the procedural discrepancy rate and the response discrepancy rate weighted by the fraction of the sample with valid responses in both interviews. That is, for each question j:

---

[7]Neither of the terms 'response' or 'procedural' in referring to discrepancies should be taken too literally. Response discrepancies can come about, for instance, because the interviewer marked the wrong answer (a procedural error), and procedural discrepancies can appear because of a discrepancy in an earlier answer provided by a respondent.

[8]It would be interesting to see to what extent changes in respondent could account for these discrepancies. Unfortunately, respondent identifiers from the reinterview form were not keyed.

$$ODR_j = PDR_j + RDR_j * DR_j$$

where ODR is the overall discrepancy rate, PDR is the procedural discrepancy rate, RDR is the response discrepancy rate, and DR is the dual response rate.

Table 1
Whether Received State Unemployment Compensation
As Recorded in the Reinterview by How Recorded in Original Interview

	Reinterview			
Original Interview	Blank	1 'Yes'	2 'No'	Total
Blank	1,250	7	59	1,316
1 (Yes)	7	29	3	39
2 No	29	3	172	204
Total	1,286	39	234	1,559

Table 2 presents these discrepancy rates and the dual response rates for each of twelve substantive questions asked in the SIPP reinterview.[9] There is considerable variation in the overall discrepancy rates for these questions ranging from less than two percent for questions on employment during the reference period (1) and continued Medicaid coverage (26b) to about seven percent for the Health Insurance coverage (27a) and the employer's contribution to Health Insurance (27f) questions.[10] This pattern is quite similar to that reported by the Census Bureau's Reinterview Evaluation Section (see e.g. Smith, 1987). While it does vary from question to question, the majority of the discrepancies in the data as a whole are procedural rather than response discrepancies. Given the skip patterns depicted in Figure 1,

[9]The questions asked in connection with the update of the income and asset rosters are excluded from the present analysis.

[10]The overall discrepancy rate over all items was 3.82% which is only moderately higher than the 3.07% reported by St. Clair (1985) for Waves 2–4 of the 1984 Panel. Most of this difference is probably due to differences in the definitions of difference rates. It is also likely, given the results of Section 4 below, that our rate would have been lower had we included wave 4 in our analysis.

it is not surprising that virtually all of the discrepancies on the Medicare coverage question were procedural in nature—i.e. the result of the question being skipped in one interview and not in the other. There are, after all, three distinct ways in which a respondent can be routed around question 23a and four ways in which he could be routed to it.[11]

Table 2
Discrepancy Rates for the Substantive Reinterview Questions

Question	Discrepancy Rates (percent)			
	Overall	Procedural	Response	Dual Response
1. Have job?	1.89	0.26	1.63	99.7
2a. Look for job?	2.20	1.28	2.55	36.18
3a. Want job?	2.81	1.54	4.04	31.37
4. Each week?	3.84	1.92	3.11	61.58
9a. U.I. Comp?	3.15	2.76	2.99	13.28
23a. Medicare?	5.13	5.07	*	4.75
24. Food Stamps?	1.93	1.67	.28	91.02
26a. Mcaid now?	3.08	2.44	0.71	90.89
26b. Mcaid B4?	1.68	1.48	*	3.40
27a. Health Ins?	6.78	0.77	6.03	99.23
27e. Via emplyr?	6.00	4.68	2.35	55.93
27f. Emplyr pay?	7.35	4.11	7.62	42.59

*Rate suppressed due to the small number of cases in the denominator.

[11]The respondent is routed around 23a if either 1) R6=N, R21=N and R22=N, 2) R6=Y, and R8=Y, or 3) R6=Y, R8=N, R9=N and R10=N. The respondent is to be asked question 23a if either 1) R6=N and R21=Y, 2) R6=N, R21=N and R22=Y, 3) R6=Y, R8=N, and R9=Y, or 4) R6=Y, R8=N, R9=N, and R10=Y.

Procedural discrepancies also accounted for most of the overall discrepancies in all the remaining questions except for the initial employment and health insurance questions. That these are the initial questions in a sequence which all respondents are to be asked is significant and points to the fact that some of the procedural inconsistencies are the result of response inconsistencies in earlier portions of the interview.

Response inconsistencies also vary widely from a low of less that three-tenths of one-percent for the Foodstamp authorization question to more than seven and a half percent for the employer health insurance contribution question. The high response variances of health insurance coverage and employer contribution of .03 (=.5*6.03/100) and .038 (=.5*7.62/100), respectively, would suggest that there is something wrong with these questions. The full health insurance coverage question reads:

> 27a)**"During the 4-month period, did ... have group or individual health insurance in ...'s own name?"**

While the problem with this question is quite likely that 'whose name the insurance is in' is not particularly salient or important to the respondent, it would be interesting to know how many respondents are giving either "group" or "individual" as their initial response. Similarly, from the respondent's point of view, reasonable responses to the question:

> 27f)**"Did the employer or union (former employer or pension plan) pay for all or part of the cost of this plan?"**

could be 'employer', 'union', 'all', 'part', 'no', or 'yes'. The allowed responses are 'all', 'part' and 'none'. Thus, it is quite likely that the interviewer is having to probe for the 'all', 'part', or 'none' responses in a large number of cases when the respondent's answer is 'yes', 'employer' or 'union'. Part of the response variance may be due to variance in how and whether these probes are being made.

While response variance is most troublesome for the health insurance questions, it is also quite high for discouraged worker question. In this case, the question seems rather unambiguously worded and it would seem that the problem must lie in the ambiguity of the concept itself.

Before leaving our discussion of the extent of interview/reinterview discrepancies it should be noted that independent analyses of the reinterview data by Bureau staff revealed the same pattern of results for the health and discouraged worker questions. As a result the

health questions have been substantially modified, while the discouraged worker question has been dropped.

In summary, simple comparisons of interview and reinterview reports from the reinterview data are sufficient to highlight some questions and procedures that are particularly problematic in the current SIPP instrument. Considerable error is probably being introduced to the data, for instance, because the skip sequences are sometimes quite complex and may not always be successfully followed. Additional errors occur because not all the questions are as clearly worded as we would like, and the reinterview data reflect these glitches in the form of high response variance.

4. Correlates of Inconsistency

If the procedural and response variability is the same for all respondents, then its existence is relatively benign. In multivariate analysis its existence in dependent variables will only reduce the model's goodness of fit and in independent variables will (predictably) bias the estimated coefficients toward zero.[12] If, on the other hand, the extent of response or procedural variance differs systematically from one respondent to the next, all manner of problems can be expected to arise in bivariate or multivariate analysis. The purpose of this section is to explore the extent to which response and procedural variance differs systematically with characteristics of respondents and interviewers.

Traditionally, analysts have chosen some form of logit model (see e.g. O'Muircheartaigh and Wiggins, 1981) in investigating the association of respondent and interviewer characteristics with response discrepancies. Such analyses are done on a question by question basis. In a preliminary investigation of such a model with the current data, the author found that, given the rarity of response discrepancies and the relatively small size of the SIPP reinterview program, there were too few cases of response discrepancies to analyze effectively in this manner.

An alternative modeling approach is to analyze the reinterview data, not on a question-by-question basis, but as single experiment in which the outcome is the number of discrepancies occurring in the course of the reinterview. Each question asked in the reinterview can be thought of as a Bernoulli trial with a 'success' being defined as a report being given which differs from that provided in the original interview. If we assume that these

---

[12]It will also exacerbate the seam problem (see Moore and Kasprzyk, 1984, Burkhead and Coder, 1985, or Kalton, Lepkowski and Lin, 1985, for information on this problem in the SIPP).

trials are independent,[13] then the reinterview process itself would be a series of Q_i Bernoulli trials where Q_i is the total number of questions put to the ith respondent. Furthermore, the total number of inconsistencies, n_i, in Q_i trials would be binomially distributed and if Q_i where sufficiently large, we could treat the distribution of n_i, conditional on a set of exogenous variables, as $N(Q_ip,Q_ip(1-p))$ where p is the probability of a response inconsistency. In other words, if each respondent were asked a very large number of questions (say 1000) then we could treat the number of inconsistencies observed as a continuous variable and apply ordinary least squares to determine the relationship of response variance to a set of exogenous factors (X_i).

As Figure 2 indicates, however, the probability of an inconsistency on any one question is so low that the distribution of the sum of inconsistencies is highly skewed—so highly skewed that Q_i would have to be extremely large for the central limit theorem to apply. In such cases, the Poisson distribution is often a useful approximating distribution to the binomial (see, e.g. Lindgren, 1976),[14] and as we shall see below, has some particularly attractive features in the present application. According to the Poisson distribution, the probability of exactly n inconsistencies occurring is:

$$p(n) = \exp(-\lambda)\, \lambda^n/n! \qquad 4.1)$$

where λ is the mean number of inconsistencies observed (i.e. $\lambda = Qp$). Both the mean and variance of the Poisson distribution are λ. Figure 2 presents, in addition to the actual distribution of response errors in the SIPP reinterview data, the theoretical distribution obtained from the Poisson using the sample average number of response inconsistencies of .171

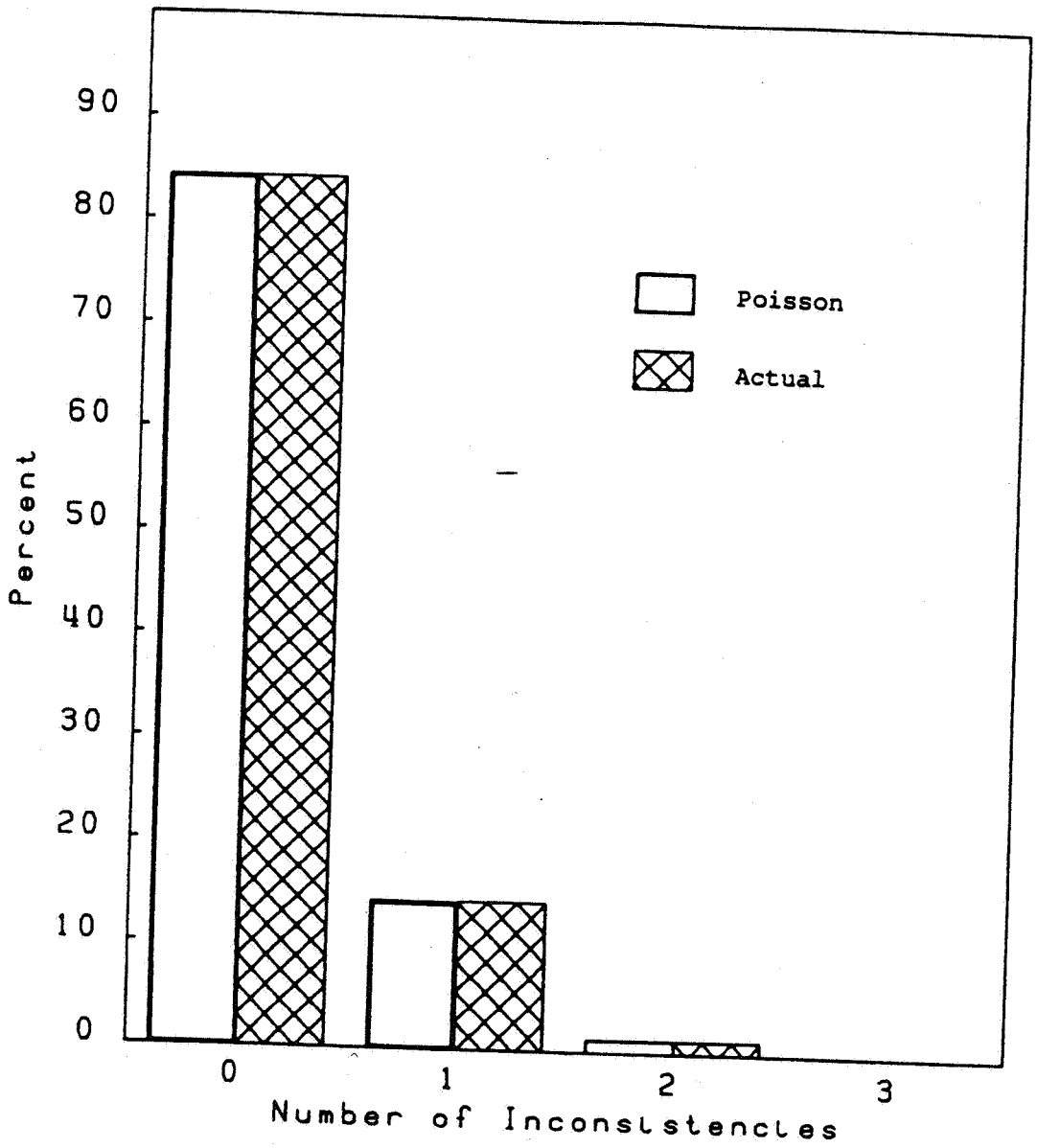
---

[13]Note that this independence assumption represents the null hypothesis to be tested. It is not a maintained assumption of the model. Indeed, one of the most important findings of our analysis will be that the independence hypothesis can not be rejected when we restrict our attention to response inconsistencies, but must be rejected when we add in procedural inconsistencies. Thus, the questionnaire sequencing acts as a strong correlating influence on the errors from one question to the next.

[14]We have a choice here in how we conceptualize the response process. We can consider the Poisson as merely an approximation to a binomial process which is useful for rare events, or we can consider the response process itself Poisson. Each question 'q' could, in theory, be presented to each respondent 'i' a very large number of times and we could count the number of times the responses are inconsistent (n_{iq}). If these inconsistencies occur randomly and independently in time (sequence), then n_{iq} would be Poisson with a mean of λ_{iq}. Furthermore, the <u>sum</u> of these counts over a sequence of questions ($q \in <1, .., Q_i>$) will also be Poisson with mean $\lambda_i = \sum_q \lambda_{iq}$.

FIGURE 2

Actual and Theoretical Probabilities of Counts of Response Discrepancies



per reinterview. While a χ^2 statistic for testing the goodness of fit of this model to the empirical distribution is easily constructed, it is not necessary in the present case—the theoretical distribution fits the data like a glove. The mean *and variance* of the observed data are .171, which is yet further confirmation of the extremely good fit of the Poisson to the response inconsistency data. Since respondents were asked, on average 6.3 questions per reinterview, this would imply an average response discrepancy rate of 2.7% (=(.171/6.3)*100) and an average response variance of .0135.

Conceptually, the nearly perfect fit of the response inconsistency data to the Poisson suggests that if respondents were asked a reinterview question repeatedly (and their memories of their previous responses were wiped clean) inconsistent reports would appear infrequently, randomly and independently in time. Indeed, the Poisson can be shown to be the maximum entropy or disorder process. One might think that given the skip sequences used in the SIPP that errors in one variable would lead to errors in subsequent ones, and the independence aspect would not be accurate. This would be the case for procedural or overall inconsistencies, but is not for response inconsistencies—any subsequent inconsistencies resulting from a response error are, by construction, procedural and are not counted in the response discrepancy rate.[15]

While all this is interesting and reassuring, it may not be entirely obvious that the fit of the unconditional distribution is particularly relevant in developing a multivariate model. As it turns out however, if the mean number of inconsistencies (λ_i) given by individual i over a number of independent trials is related to a set of individual characteristics X_i according to:[16]

$$\lambda_i = Q_i \exp(X_i \beta) \qquad 4.2)$$

and if n_i follows a Poisson distribution, then

$$E(n_i \quad X_i) = \lambda_i \qquad 4.3)$$

Expressions 4.1) – 4.3) form the basis of what is sometimes referred to as Poisson Regression (see Maddala, 1984). The likelihood of observing a sample of N cases

---

[15]This does mean that the number of questions from which the response discrepancy counts are derived vary from one respondent to the next. This complication is easily handled as shown in equation 4.2).

[16]In the parlance of collective risk theory, where Poisson models are used extensively, the term Q_i in equation 4.2), the number of questions asked of the ith individual, is his 'exposure'.

$$L(n_1, ..n_N | X_1, .., X_N; \beta) = \prod_{i=1}^{N} P(n_i) \qquad 4.4)$$

$P(n_i)$ can be obtained by substituting 4.2) into 4.1). That is:

$$P(n_i) = [\exp(-Q_i \exp(X_i\beta))] \quad \frac{Q_i^{n_i} \exp(n_i X_i \beta)}{n!} \qquad 4.5)$$

Substituting 4.5) into 4.4), taking logs, and collecting terms yields the following log likelihood function:

$$\mathbf{L} = \sum_{i=1}^{N} [-\exp(X_i\beta + \ln(Q_i)) + n_i(X_i\beta + \ln(Q_i)) - \ln(n_i!)] \qquad 4.6)$$

It can be shown that so long as the X's are not perfectly colinear (and so long as $\exp((X_i\beta)) > 0$ for some i) this log-likelihood function is globally concave in the β's.[17] This means that efficient and consistent estimates of the proportionate effects of exogenous factors on inconsistency rates can be obtained quickly by any one of a number of maximization routines. In the present analysis we employ the Davidson-Fletcher-Goldfarb-Shanno version of the David-Fletcher-Powell algorithm to maximize 4.6) and obtain our estimates of β.[18] Estimated standard errors are constructed from the diagonal elements of the inverse-Hessian matrix.[19]

There are several attractive features of Poisson regression in analyzing response discrepancies. First, the effects of change independent variables are easily interpretable and

---

[17]See Hausman, Hall, and Griliches (1984).

[18]The algorithm we employ is written in Pascal by the author using sub-routines described in Press, Flannery, Teukolsky and Vetterling (1986). The programs were compiled on a Zenith 20286 micro-computer using the TURBO-PASCAL 4.0 compiler and a 20287 numeric coprocessor. The extended precision real number type provided by this compiler allows the computation of very precise numeric derivatives and thereby reduced programming time considerably.

[19]The estimated standard errors, therefore, are based on the assumption of simple random sampling. If we define the population of inference as the full SIPP sample, then we should have weighted the data by the inverse of the selection probabilities discussed in Section 2 and computed complex sampling errors using some form of replication. Unfortunately the number of units assigned to interviewers and the numbers of eligible persons in these units, necessary to the construction of the weights, were not available and we are forced to abandon finite population inferences.

can be readily compared with the results of other analyses in the literature. To see this first note that:

$$\lambda_i = \sum_{q=1}^{Q_i} GDR_q/Q_i = 2 \sum_{q=1}^{Q_i} SRV_q/Q_i$$

where GDR_q and SRV_q are, respectively, the gross-difference rate and simple response variances for the qth question. Second, note that taking logs of equation 4.2) yields:

$$\ln(\lambda_i(X_i)) = X_i\beta + \ln(Q_i)$$

Therefore, a unit change in X_{ji} will result in a *proportionate*[20] change in the mean discrepancy rate of β_j and in the estimate simple-response variance of $\beta_j/2$. The second advantage of the Poisson regression is avoids the limited dependent variable problems which would arise if one attempted to apply the central limit theorem and analyze the data under the normality assumption. The Poisson distribution is a natural counting distribution in which zero is a legitimate outcome. The discrete ('lumpy') nature of the dependent variable is also automatically handled by the Poisson regression model. The third and final advantage of the Poisson regression model is that it is consistent with a very reasonable view of the response process itself—response errors are like accidents of other types. They happen relatively infrequently and at random. But as with other types of accidents, some types of individuals may be more prone making errors than others and the Poisson regression model allows us to test for significant correlates of error-proneness.

The independent variables we employ in our analysis are of two types—those intended to capture (at least some of) the effects variability in interviewing process, and those characteristics of respondents which might affect response variability. The first of the interviewing process variables is simply the calendar month in which the original interview was taken. Since the data are taken from the second and third waves of the 1984 panel, the study was still quite new to the interviewers at the beginning of our observation period. We would expect more inconsistencies in these months. By the end of our observation period, on the other hand, most interviewers had been administering the study monthly for a full year, and we would expect their error rates to have settled down. Because we would expect declining marginal improvements with additional months of experience, we include the natural logarithm of the interview month rather than the month itself in our empirical specification.

[20]Recall that, for $f(x) > 0$, $\partial\ln(f(x))/\partial x = (\partial f(x)/\partial x)/f(x)$ and thus the change in $\ln(f(x))$ resulting from a change in x is proportionate to the size of f(x).

The second interviewing process variable is a scale based on the overall performance of interviewers in the various Regional Offices. The underlying rationale for this scale is that an unknown portion of the observed variation between these offices is due to differences in interviewers and in local procedures and the remainder is due to differences in the characteristics of the respondents. If all of the individual-to-individual variability is due to these Regional Office factors, then a scale constructed from the Regional Office rates should bear a one-to-one relationship with the individual discrepancy rates, and should explain all of the variance in them. That is, if interviewers and regional office characteristics determine the individual's response variance then:

$$\lambda_i = \mathrm{ROR}_i \exp(\alpha)$$

where ROR_i is the Regional Office discrepancy rate for the ith individual's region, and α is a constant. If on the other hand, the reason Regional Offices differ is that the characteristics of their respondents differ then the one-to-one relationship between the Regional Office rate and the individual rates should disappear once the individual factors are controlled. That is, in:

$$\lambda_i = \gamma \mathrm{ROR}_i \exp(X_i\beta) = \exp(X_i\beta + \gamma \ln(\mathrm{ROR}_i))$$

γ should be significantly less than unity and should not explain a significant portion of the variance.

The third and forth interviewing process variables included are the relationship of the individual to the household reference person, and a dummy variable for whether a proxy informant was used in the original interview. The relationship to reference person measure is also a dummy variable equaling 1 if the individual is some one other than the reference person or his/her spouse (*e.g.* child, parent, aunt, *etc.*).

The individual characteristics included in our empirical specification are the same ones thought to affect market productivity in the human-capital model of earnings. These consist of age (and its square), education, race, and gender. We also include income itself in some of our specifications.

Table 3 presents both bivariate and multivariate estimates of the Poisson regression model for response discrepancies obtained by maximizing 4.6) with respect to the β. The first column of figures, labeled 'Bivariate Parameters', are obtained when the Poisson Regression model is estimated with only a constant and the variable listed to the left of the coefficient included as predictors. As hypothesized, response inconsistencies decline significantly with interview month. Since the month is included as a proxy for interviewer and respondent experience with the SIPP, and since the logarithm of month is used, the coefficient of $-.275$ is

Table 3
Maximum Likelihood Poisson Regression Estimates of Response Inconsistencies
(Asymptotic SRS Standard Errors in Parentheses)

	Bivariate		Multivariate	
	Parameter	Log-likelihood	without Income	with income
Constant	−3.609** (.037)	−775.4	−1.455* (.623)	−1.724** (.603)
Interview Month	−.275* (.132)	−773.2	−.251+ (.132)	−.235 (.130)
Regional Office Discrepancy Rate	.935** (.322)	−770.8	.980** (.313)	.962** (.313)
Proxy Respondent	.175 (.132)	−774.5	.107 (.146)	.113 (.146)
Odd Relationship to Reference Person	.383* (.161)	−772.7	.109 (.199)	.030 (.203)
Age (decades)	−.485** (.176)	−771.4	−.369+ (.207)	−.215 (.197)
Age-squared (decades-squared)	.470** (.175)		.345+ (.202)	.205 (.197)
Education	−.044* (.019)	−772.7	−.042* (.021)	−.020 (.022)
Whthr Female	.098 (.123)	−775.0	.080 (.128)	−.071 (.137)
Whthr Black	.162 (.203)	−775.0	.140 (.209)	.128 (.203)
Income ($100's)	−.827** (.211)	−766.5		−.701** (.241)
ln(likelihood) (d.f.)			−761.5 (10)	−757.8 (11)

+significant at the 10% level.
*significant at the 5% level.
**significant at the 1% level.

interpretable as the experience elasticity of experience—a one percent increase in experience is associated with a .275 decrease in response inconsistency rates. This result is encouraging because it indicates that progress was being made in improving response quality early in the SIPP program.

The fact that the coefficient on the log on the Regional Office inconsistency rate is so close to unity, and is highly significant means that differences in something at the regional level are important, but the bivariate results can provide no clue as to what it might be. While the effect of the original interview having been taken with a proxy respondent is to increase response inconsistency, the effect is not sufficiently strong to attain statistical significance. The positive coefficient for the relationship to reference person dummy variable indicates that the response consistency for reference persons and their spouses is higher (by about 38.3 percent) than that obtained from other persons in the household.

The effects of age on response inconsistency rates is highly non-linear. The coefficients of $-.485$ and .47 on age and age square, respectively, suggest that response quality increases with age at a decreasing rate until age 51 where it attains its maximum.[21] For respondents much older or younger than this, response quality is significantly lower. While in the present case it is clear from the individual coefficient's standard errors that the age effects are significant, in general, one would need to test the change in the goodness of fit when age and its square are dropped out of the analysis as a set. This can be accomplished by means of a likelihood-ratio test constructed from the log likelihood values present in the second column of figures. In the present case the χ^2 associated with the hypothesis that age (and its square) are not associated with response quality is 8 $(= 2^*(-771.4 - (-775.4)))$, and has 2 degrees of freedom. Thus, the null hypothesis of no age effect can be rejected soundly.

The final two variables with significant bivariate associations with response inconsistencies are education and income. Each one-year increase in educational attainment is associated with a 4.4 percent decrease in the response inconsistency rate.[22] The extremely significant coefficient of $-.827$ on income, similarly, is interpreted as indicating that a dollar increase in monthly personal income is associated with a .83 percent decrease in the response inconsistency rate. Monthly personal income is the most powerful predictor of response

---

[21] To see this simply differentiate $\ln\lambda = -.485^*\text{age} - .47^*(\text{age})^2$ with respect to age and set the result equal to zero. Solving the result for the age yields $51.06 = 10^*(-.485/(2^*.47))$—the age at which $\ln\lambda$ attains its minimum.

[22] The interpretation of the coefficients from the Poisson regression is best seen by noting that, for education, $\ln(\lambda) = -.044^*\text{Ed}$. Differentiating this w.r.t. Ed yields $d\lambda/\lambda = -.044$—thus the coefficient for variables which enter the X matrix linearly is interpretable as the *proportionate* change in the mean inconsistency rate associated with a one unit increase in the independent variable.

inconsistencies included in our analyses. Conceivably some of this effect may be a reflection a tendency for fewer imputations being made for relatively complete interviewers and these interviewers tend to be interviews with people who have some income to report.

The bivariate results just discussed are analogous to simple correlations in linear models. The multivariate results presented in the last two columns of Table 3, in contrast, are analogous to multiple correlation coefficients. These coefficients are, therefore, interpretable as the net effects of the various factors on response inconsistency one obtains when the effects of other factors are controlled. Thus, it is not surprising that these multivariate effects are, in general, weaker than their bivariate counterparts. Indeed, with the single exception of the Regional Office inconsistency index, all the coefficients in column 3—the specification which includes everything but income—are of the same sign as those in column 1, but are smaller in absolute value. The estimated standard errors are also, in general, larger in the multivariate analyses—a second indication that the various predictors are correlated with each other. The decreased size of the estimated effects and their increased estimated variance combine to decrease the significance of almost all predictors in the multivariate analysis which excludes income. The only predictor to go from statistical significance to insignificance, however, is relationship to reference person. This indicates that most of the observed bivariate effect of not being the reference person (or his/her spouse) is, perhaps, due to the fact that most of these other individuals are children and children are younger, less educated and less likely to have income to report than their parents. Once the effects of these correlated factors are controlled, these individuals have response inconsistencies which are insignificantly different from those of reference persons (and spouses of reference persons). The combined effect of age and age-squared, by the way, remains significant even though the individual coefficients are not.

When income is added to the multivariate specification of the response inconsistency Poisson regression, every other individual characteristic becomes insignificant. Taken literally, this result would suggest that all of the effects of age and education on response quality discussed up to this point are the result of the correlation of these factors with income. We find this result hard to believe. Why income, itself, should have a positive effect on response quality is a mystery.[23]

Before moving on to our analysis of total inconsistencies, two further aspects of the multivariate Poisson regression estimates of response inconsistencies should be noted. First,

---

[23]One possibility is that the focus of the SIPP is income and transfer program participation and neither the respondent nor the interviewer may be taking the interview as seriously when the individual has 'nothing to report', than when individual income is substantial.

the overall goodness of fit of both versions of the multivariate model is highly significant. The χ^2 under the null hypothesis of no association for the model presented in column 3 is 27.8 with 10 degrees of freedom and that for the model in column 4 is 35.2 with 11 degrees of freedom. Second, and of more substantial interest, the coefficient on the Regional Office inconsistency index was unaffected by the inclusion of respondent characteristics. In fact, this coefficient increased slightly when the other factors were controlled. This suggests that the source of the regional differences in response inconsistencies is something other than regional differences in the characteristics of respondents. One possibility is that the quality of interviewer training or selection varies by region. Alternatively, it may be that the care given to the reinterview program varies from one Regional Office to the next. In either event, future analysis of the reinterview data with data on interviewer characteristics, would seem worthwhile.

Total Inconsistency Rates

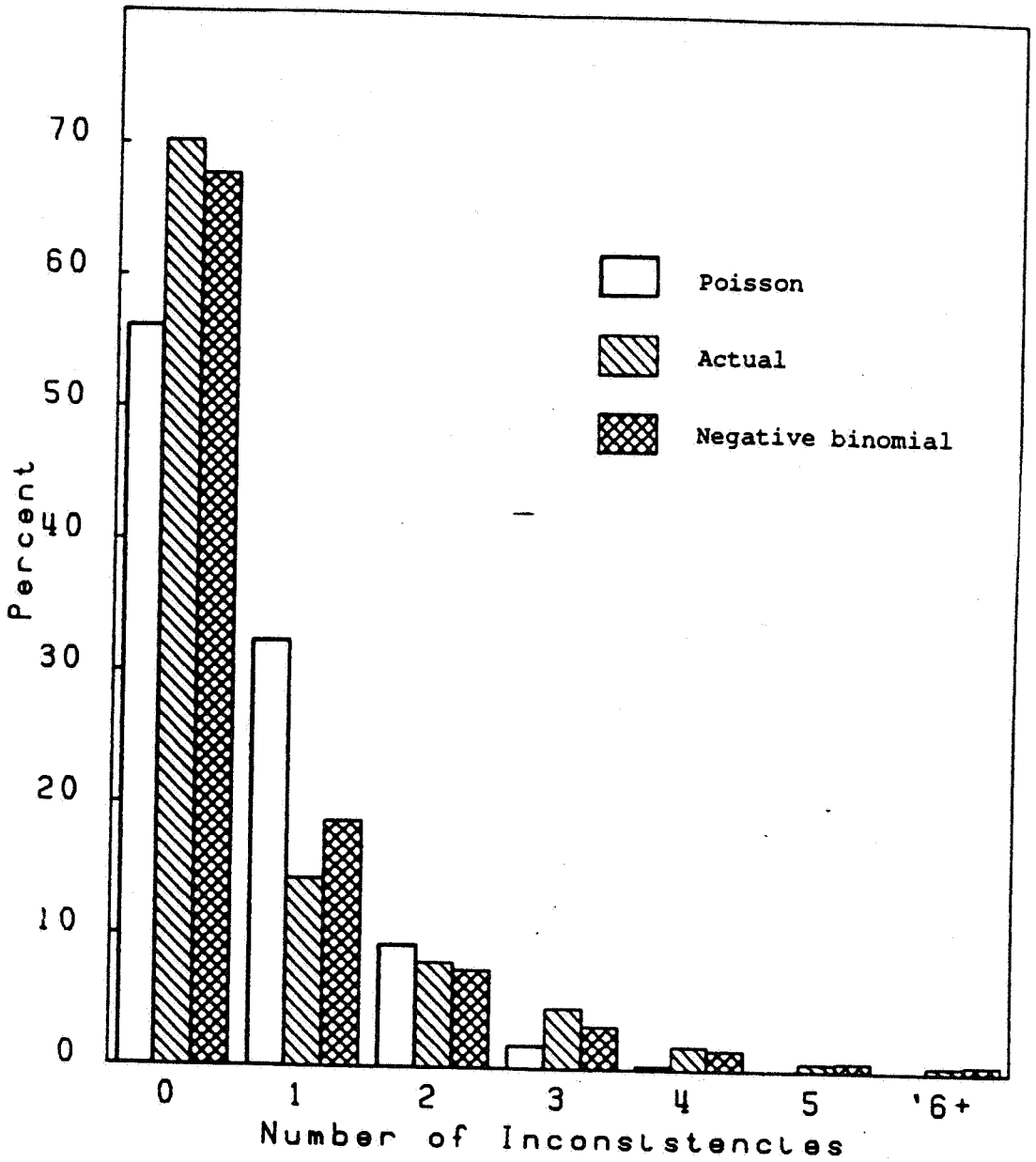
Response inconsistencies are relevant when one is trying to understand the response process itself, but in many respects a better measure of the reliability of survey items is the total inconsistency rate. This is simply the sum of the procedural rate and the response inconsistency rate weighted by the portion of the sample asked the question in both the interview and reinterview. Unlike the response inconsistency rate, the Poisson distribution is not a good choice for describing or modeling total inconsistencies. Figure 3 presents the a histogram of the actual inconsistency counts from the SIPP reinterview data, along side those implied by Poisson and Negative-binomial distributions constructed using sample moments. The probabilities predicted by the Poisson, based on the sample mean of .572 per reinterview, grossly under estimate the fraction of clean cases ($n = 0$) as well as of very dirty cases ($n \geq 3$). The problem is that there is more variability in the data than is implied by the Poisson distribution. If total inconsistencies were following a Poisson process, then their variance should equal their mean. In fact, it is more than twice (1.16/.572) as large.

Such problems of excessive variability are often encountered in fitting data to counting distributions. In the Poisson, all of the variability is due to the fact that the n_{iq} are determined by a Poisson process—the λ_i are deterministic functions of the X_i. If we assume instead that the λ_i are themselves random variables, and that they follow a Gamma distribution with parameters exp((xB) and δ then it can be shown that:[24]

[24]See Hausman, Hall, and Griliches (1984), pp 916–922.

FIGURE 3

Actual and Theoretical Probabilities
of Counts of Total Discrepancies



$$P(n_i) = \frac{\Gamma(\exp(X_i\beta) + n_i)}{\Gamma(\exp(X_i\beta))\,\Gamma(n_i+1)} [\delta/(1+\delta)]^{\exp(X_i\beta)} [1+\delta]^{-n_i} \qquad 4.7)$$

where $\Gamma(\cdot)$ is the Gamma function:

$$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t)\, dt.$$

This rather intimidating function is in fact a negative-binomial and can be simplified considerably by defining $p \equiv \delta/(1+\delta)$, and $q = 1/(1+\delta)$ and by noting that:

$$\Gamma(k+1) = k\,\Gamma(k) \text{ and } n! = \Gamma(n+1)$$

Once we make these substitutions and perform the recursions we obtain for 4.7):

$$p(n_i) = \frac{\prod_{j=1}^{n_i} (\exp(X_i\beta) + j)}{n!} p^{\exp(X_i\beta)} q^{n_i} \qquad 4.8)$$

The mean and variance of n_i for the negative binomial are:

$$\exp(X_i\beta)/\delta, \quad \text{and} \quad \exp(X_i\beta)(1+\delta)/\delta^2 \qquad 4.9)$$

respectively.

Figure 3 includes the predicted probabilities for this negative binomial distribution with p set equal to the sample mean divided by the variance ($\bar{n}/v(n)$), and $\exp((X_i\beta)$ set to the square of the sample mean divided by the variance minus the mean ($\bar{n}^2/(v(n)-\bar{n})$.[25] Clearly the negative binomial fits the unconditional distribution significantly better than does the Poisson.

It is still not a perfect fit by any means. The chi-square obtained for the test that the unconditional distribution of total inconsistencies is a negative-binomial is 36.3 with six degrees of freedom. Part of the reason that the negative-binomial does not fit the data better is that it ignores the dependency of procedural errors in one question on procedural or response errors in preceding questions. This is a difficult problem and one which we will defer for future research.

---

[25]This is the method of moments technique for fitting the data to the distribution. One can easily verify these formulas using the expressions for the mean and variance of the negative binomial provided above.

The fit of the negative binomial, however, is sufficiently better than that of the Poisson that it seems preferable to use it as the basis of our multivariate model of total inconsistencies. The log-likelihood function can be obtained by substituting equation 4.9) into 4.4) and taking logs. This yields, for a sample of size N:

$$\mathbf{L} = \sum_{i=1}^{N} [\, \exp(X_i\beta)\ln(p_i) + n_i \ln(q_i) + \{ \sum_{j=1}^{n_i} \ln(\exp(X_i\beta) + j)\} - \ln(n_i!)] \qquad 4.10)$$

Maximization of 4.10) with respect to the β was accomplished using the same DFGS algorithm employed in our earlier estimation of the Poisson regression model. The results of this estimation are presented in Table 4.

The results of the maximum likelihood negative-binomial analysis of total inconsistencies (Table 4) look very much like those obtained for response inconsistencies using Poisson regression (Table 3). The interpretation of these coefficients is the same as that of the Poisson regression coefficients—for those variables entering linearly (e.g. education), a one unit increase is associated with a proportionate change in the inconsistency rate of β (a 5.2% decrease for education in the bivariate model). The only real difference between the Poisson regression coefficients for response inconsistencies and those of the negative-binomial for total inconsistencies is that the latter are generally larger in absolute value and have lower estimated variances. The same substantive results hold.

As was the case for response inconsistencies, the total inconsistency rate declines significantly with time, and there remains a one-to-one relationship between regional office inconsistency rates and individual rates. Reference persons (and their spouses) have significantly lower inconsistency rates than do more distantly related individuals in the sampling unit, but this is evidently due to their higher income and education and to the fact that they are more apt to be 'middle aged'. Inconsistency rates decline with age until attaining a minimum at age 44 and increase thereafter. Higher educated individuals have lower total inconsistency rates, although this effect disappears if one controls for income (i.e. it is not significant in the multivariate model).

Unlike the Poisson results for response inconsistencies, race is a significant correlate of total inconsistencies. Blacks have total inconsistency rates some twenty-eight percent higher than non-Blacks, and this effect does not appear to be merely a reflection of their lower average educations and incomes. Evidently interviewers are 'hitting the check points' less consistently for Black respondents than they do for non-Black respondents.

Table 4
Maximum Likelihood Negative-Binomial Regression Estimates Total Inconsistencies
(Asymptotic Standard Errors in Parentheses)

	Bivariate		Multivariate	
	Parameter	Log-likelihood	without Income	with income
Constant	−.872** (.094)	−1577.9	.807+ (.434)	.592 (.455)
δ	.726** (.076)		.779** (.065)	.797** (.062)
Interview Month	−.198* (.097)	−1576.8	−.191* (.090)	−.180* (.090)
Regional Office Discrepancy Rate	.985** (.236)	−1569.5	1.043** (.209)	1.022** (.221)
Proxy Respondent	.149 (.108)	−1577.7	.262+ (.142)	.197 (.143)
Odd Relationship to Reference Person	.382** (.121):	−1573.8	.109 (.100)	.121 (.103)
Age (decades)	−.511** (.125)	−1565.8	−.326** (.140)	−.189 (.143)
Age-squared (decades-squared)	.5745** (.123)		.400** (.136)	.271* (.138)
Education	−.052** (.013)	−1572.1	−.028* (.014)	−.011 (.015)
Whthr Female	.011 (.094)	−1578.8	.019 (.085)	−.149+ (.091)
Whthr Black	.284* (.143)	−1576.9	.275* (.135)	.258+ (.134)
Income ($100's)	−.698** (.135)	−1567.2		−.588** (.162)
ln(likelihood) (d.f.)			−1546.4 (11)	−1541.7 (12)

The bivariate results are obtained by estimating the model with the variable interest and the constant and shape parameter (δ) only.
+significant at the 10% level.
*significant at the 5% level.
**significant at the 1% level.

Finally, as was the case of response inconsistencies, monthly personal income is the strongest predictor of total inconsistency rates, and when it is included in the multivariate model along with the other predictors, absorbs most of their effects.

In sum, given the strong similarity of the results of the Poisson regression model of response inconsistencies and the negative-binomial model of total inconsistencies, we are lead to suspect that response and procedural inconsistencies share a common causal structure. Whatever this structure is, it evidently involves characteristics of both the respondent and the interviewer (or at least of the Regional Office).

Before closing out our discussion of the negative-binomial regression results it is useful to explore briefly the implications of the fact that response errors are well described as a Poisson process whereas procedural errors are not. What it means is that, abstracting from skip sequence effects, the occurrence of a response error in one question has no effect on the probability of a response error in a subsequent question. One can easily imagine mechanisms which would result in this not being the case. If a respondent realizes that he made a mistake, for instance, and 'got away with it' on one question, then he might be less careful with subsequent answers. But the close fit of the Poisson to the response error process indicates that there is no net effect of any such mechanisms.

That the inclusion of procedural errors destroys the fit of the Poisson model to the data suggests that the sequencing processes itself acts as a correlating influence on the inconsistency probabilities from one question to the next. This raises the possibility that more sequencing is being done in studies like the SIPP than is optimal. This potential problem is analogous to the problem of optimal interviewer workloads when the interviewer acts as a correlating influence for response errors. The trade-off in that case is that training costs decrease with work load while response variance increases. In the present case, the overall interview length can be reduced by skipping entire classes of respondents around questions based on their responses to earlier questions. The resulting interviewing time savings come at a cost of increased response (broadly defined) variance and therefore decreased question reliability. As is the case with interviewer workloads, this cost is generally unknown and is often ignored in the survey design process,[26] with the result that sequencing may be over utilized just as work loads are often too high.

[26]Decreased question reliability is not the only cost of extreme sequencing. Bias may also be introduced. Take, for instance, the employment sequence of Items 1–4 in Figure 1. Those answering yes to item 1. (that they had a job) were not asked if they spent time looking for a job. Many people may have a job, at least for a few days, and may also have spent time looking or even collecting unemployment compensation. Thus total estimates of the number of people seeking jobs would be biased downward by the sequencing.

5. Conclusions and Recommendations

In this paper we have analyzed data from the SIPP reinterview program to see if it can be of value in understanding nonsampling error issues. We concluded that it can, indeed, be very valuable in several ways. First, it allows us to appreciate the fact that not all inconsistencies in the data are due to respondents providing unreliable reports. A goodly portion of the discrepancies between interview and reinterview reports is due to inconsistencies in the interview procedures. The skip sequences used in the SIPP are complex and are not always successfully followed by the interviewers. Second, the reinterview data has proven valuable in identifying particular questions with unusually high response variances. This is important not just for analyst who may wish to correct for question reliability, but for future redesigns of the SIPP questionnaire. Third, we have shown with the reinterview data that data quality does vary systematically from one type of respondent to the next. Data quality appears to be significantly lower for low income, Black, and either very young or very old respondents. Finally, while there are significant effects of things which can only be attributed to the interviewing procedure or the interviewer her or himself, the quality of SIPP data apparently improved significantly between February and August of 1984.

While the SIPP reinterview program is useful in furthering our understanding of response errors, there are a number of changes which would make the program even more useful. Some of these changes are relatively minor. These include:

1) Keying the person number of both the original and reinterview respondents (items g and o); and

2) Transcribing to the reinterview form the information necessary for the construction of reinterview sampling weights (i.e. the number of units assigned to the interviewer during the wave in question and the number of reinterviews taken).

Other improvements are more difficult and costly, but might have substantial pay-offs and should probably be considered. These include:

3) Rotating content to cover the SIPP questionnaire more completely (The present analysis shows that as little as two waves of reinterviews at the present reinterview sample size are sufficient to uncover the most serious problems in questions. Therefore, four times as much content could be usefully covered without increasing the size of the reinterview program.); and

4) Randomizing the assignment of reinterviewers.

Finally, the results of the present analysis lead to one recommendation for the future redesign of the main SIPP instrument itself. This is that the rather baroque skip sequences currently being used be simplified—they are causing relatively minor response errors to be amplified into much more serious problems.

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.*

Hausman, J., B.H. Hall, and Z. Griliches, 1984, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica*, 52:4, pages 909–938.

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

Lindgren, B.W., 1976, *Statistical Theory, Third Edition* (New York: MacMillan Publishing Co.).

Maddala, G.S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics*, (Cambridge University Press). pages 51–55.

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

O'Muircheartaigh, Colm A., 1986, "Correlates of Reinterview Response Inconsistency in the Current Population Survey", *Proceedings of the Second Annual Research Conference*, (U.S. Department of Commerce: Bureau of the Census) pages 208–235.

O'Muircheartaigh C. A. and R.D. Wiggins, 1981, "The Impact of Interviewer Variability in an Epidemiological Survey", *Psychological Medicine*, 11, pages 817–824.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986, *Numerical Recipes*, (Cambridge University Press).

Smith, R., 1987. "The SIPP Results of the Reinterview Program for 1985", (Washington, D.C.: Bureau of the Census, June 18, 1987).

St. Clair, J., 1985. "SIPP-1984 Panel: Results of the Reinterview Program for Waves 2 through 4", (Washington, D.C.: Bureau of the Census, July 2, 1985).