THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

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

No. 90

Daniel Hill
The University of Michigan

May 1989

U.S. Department of Commerce U.S. CENSUS BUREAU

Survey of Income and Program Participation

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

8913 90

Daniel H. Hill
The University of Michigan

SIPP Working Paper #90

May 1989

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. The views expressed are those of the authors and do not necessarily reflect those of the Census Bureau.

TABLE OF CONTENTS

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

Daniel H. Hill[1],
Survey Research Center
Institute for Social Research
The University of Michigan
Ann Arbor, Mi. 48106

March 29, 1989

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 1 the SIPP reinterview program is described in some detail. Section 2 presents a question-by-question description of response, procedural and overall interview/ reinterview discrepancies. Finally, in Section 3, 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.

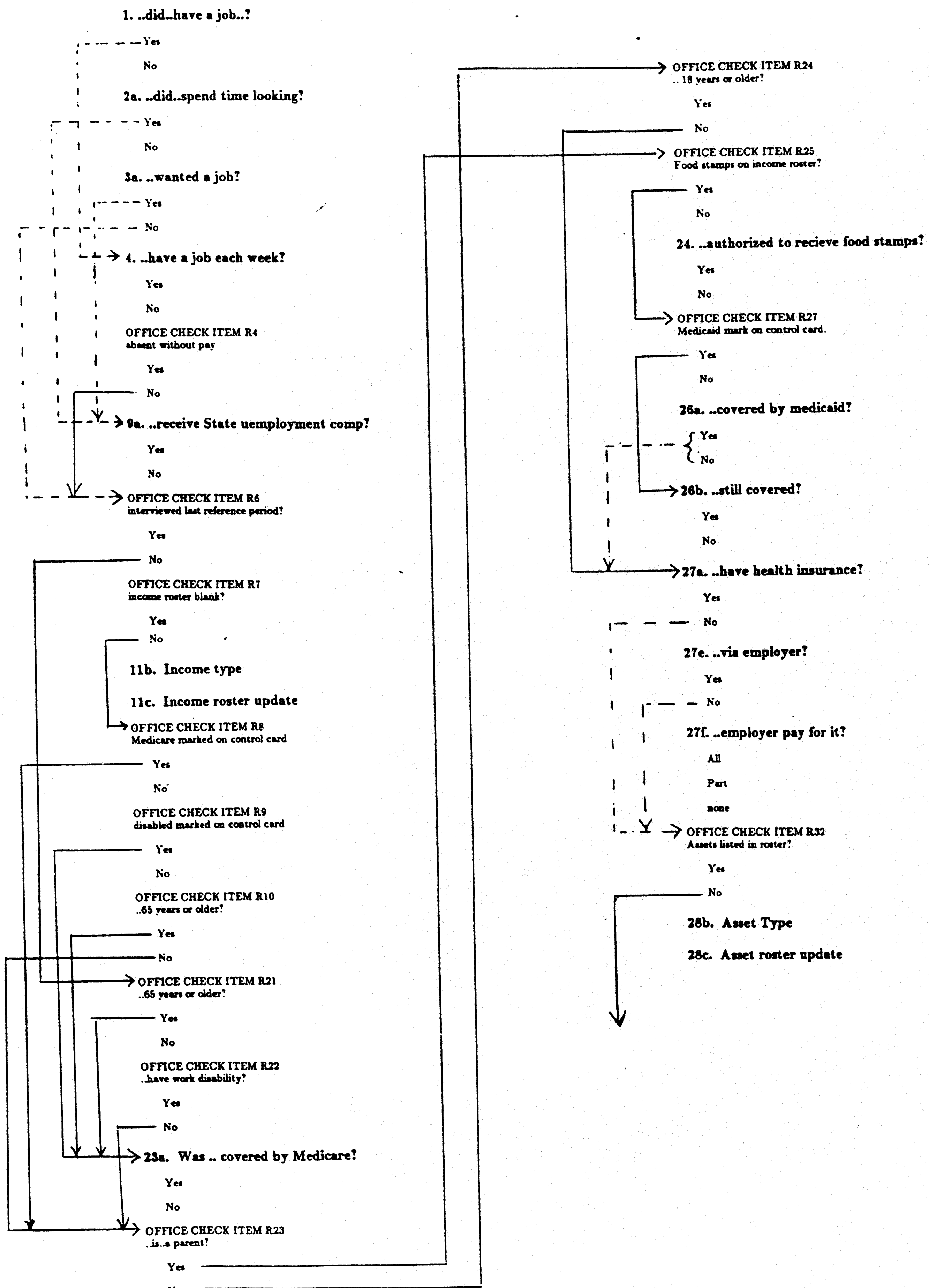
Figure 1 illustrates the question flow for Section 2 of the Reinterview Questionnaire. The questions actually asked of the respondent in both the interview and reinterview are printed in bold, while the Office Check Items which are transcribed to the Reinterview Questionnaire from the original appear 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.

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 i^{th} respondent during the t^{th} 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 i^{th} 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 $N(Q_i p, Q_i p(1-p))$ 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:

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

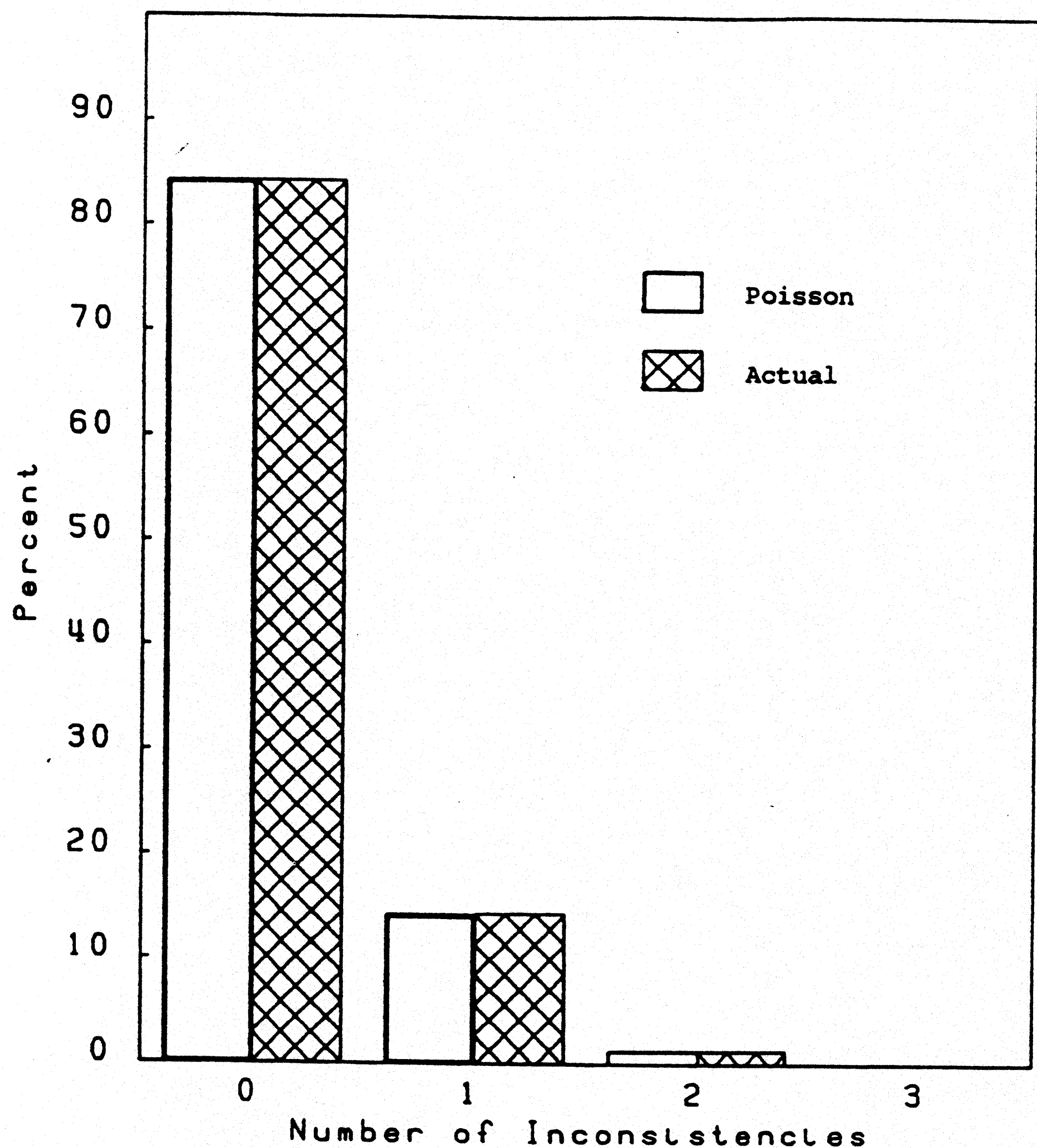
where λ is the mean number of inconsistencies observed (i.e. $\lambda = Qp$). Both the mean and variance of the Poisson distribution are λ. Figure 2 presents, in addition to the actual distribution of response errors in the SIPP reinterview data, the theoretical distribution obtained from the Poisson 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 sum of these counts over a sequence of questions ($q \in \langle 1, .., Q_i \rangle$) 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

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

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 i^{th} 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 q^{th} 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 i^{th} 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.

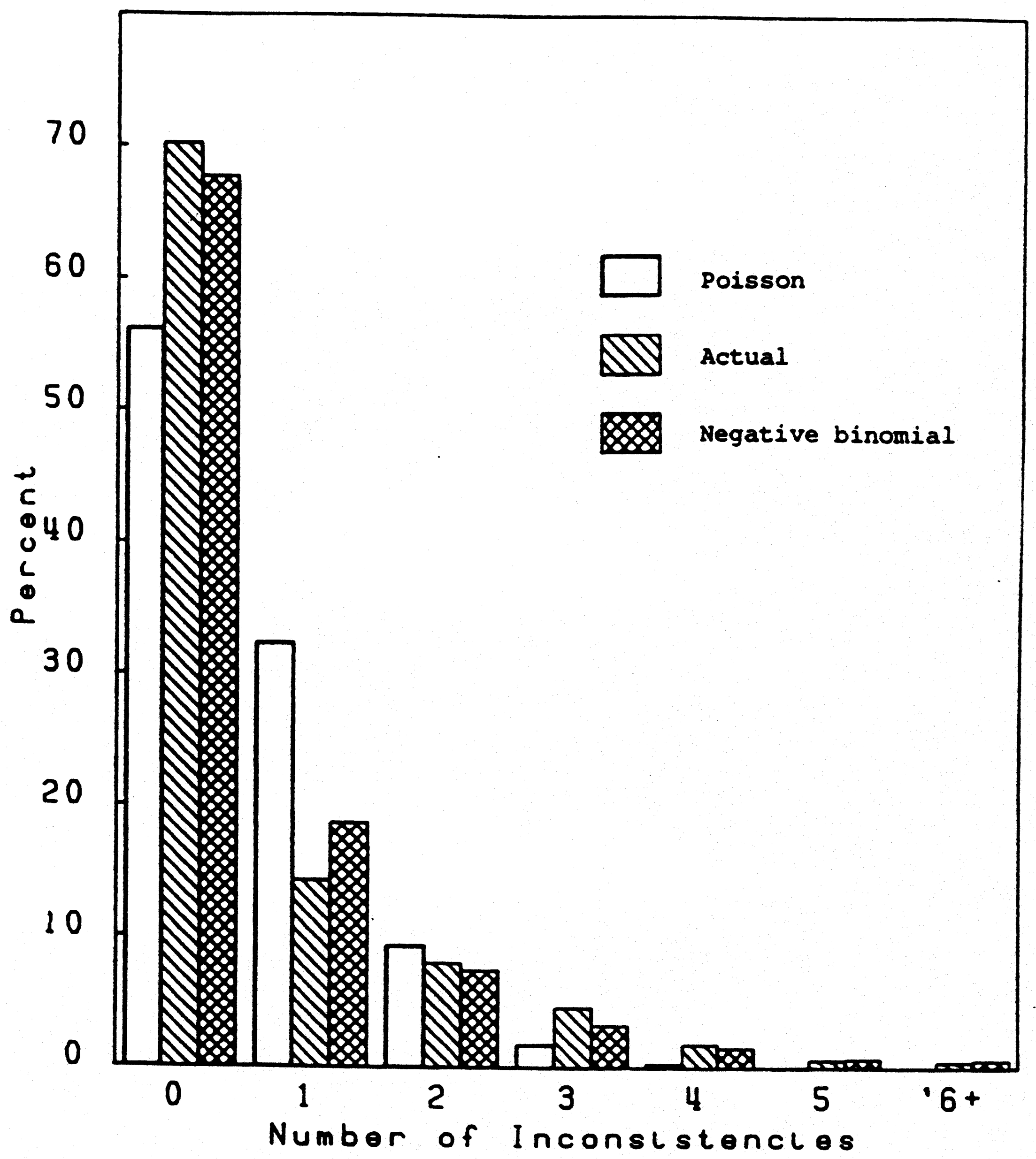
Total Inconsistency Rates

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:

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

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:

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

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).