"I didn't tell, and I won't tell": Dynamic response error in the SIPP

Christopher R. Bollinger

Martin H. David

Using state administrative records matched to two interviews of the *1984 Survey of Income and Program Participation* panel, we examine intertemporal relationships in response errors for participation in the Food Stamp Program. Response errors are highly correlated for these interviews taken four months apart. The error process can not be explained by respondent learning. We control for income and household characteristics in bivariate probit analysis and find that response to these characteristics is stable and that the error terms are highly correlated. The stability of the model across periods gives no evidence of learning. The high, positive correlation of the error term across periods supports the hypothesis that respondents have a latent tendency to cooperate (or not cooperate) in giving truthful answers. Cooperators provide accurate answers throughout the panel, while non-cooperators provide false answers and may not fully participate throughout the panel.

21 February 2000

Acknowledgement 1 *Thanks to Kent Marquis for many helpful discussions and to the Bureau of the Census for access to the data and logistic support and encouragement. Thanks also to Dan Black and Mark Berger for helpful comments. We acknowledge support from the National Science Foundation, grants SBR-9511935 and SBR-952174. The views expressed here are those of the authors and do not in any way reflect the views of the Census Bureau or any employees therof.*

"I didn't tell, and I won't tell": Dynamic response error in the SIPP 1

.

1 Introduction

False negative responses to surveys confound estimates of participation in income transfer programs and estimates of models of behavior that leads to participation. (Doyle 1990, Goudreau, Oberheu, and Vaughn 1984). False negative responses doubly confound estimates of change in participation status and models of entry and exit from program use. Furthermore the extent of the problem of serial errors depends on the correlation of false response over time. In this paper we examine false negative errors in reports of Food Stamp Program participation for the first two interviews *1984 Survey of Income and Program Participation.* We search for evidence that reports become more accurate (consistent with positive learning described below), or less accurate (consistent with negative learning described below). We find no evidence that there is any improvement or deterioration of accuracy. We also find that there is high positive correlation in response error between the two interviews. These findings suggest that response accuracy is partially determined by a common latent variable that determines propensity to cooperate. The finding confirms models of the error process in Bollinger and David (1997, 1999). Some respondents have high values, and thus are consistently good reporters, while others have low values and thus consistently give evasive answers.

Analysis of estimators appropriate to modeling that data include measurement errors (for example, Fuller (1987)) reveals that information about the probabilistic process that induces response error is essential to consistent estimators. Models may be fit on the basis of maintained hypotheses about that process, but all conclusions are conditional on those untested assumptions. Clearly direct observation of the error process through validation studies supplies more information than assumptions. Validation studies match survey data to some "true" measure of the variables of interest. Typically administrative records validate responses stimulated by requests for economic characteristics (Biemer and Fecso 1995, Biemer and Stokes 1991).

Grilliches and Hausman (1986) examine estimation of a linear model in a panel setting where regressors are measured with error. They assume no prior information about the error process. They find that the temporal correlogram of the measurement error process strongly influences the analytical expression for consistent estimators. For that reason understanding dynamic properties of response errors is essential to estimating models from panel data. They also determine that many repeated measurements are necessary to obtain sufficient instruments for consistent estimators when two or more regressors are measured with error. Overidentification is typical. This outcome is discouraging on two counts: More replications of measurements are required and more than one consistent estimator may be available. Grilliches and Hausman (1986) conclude that understanding the dynamics of response errors is the key to estimating models from panel data.

Using information from estimated models of the error process may reduce data collection costs and the social costs of decisions made on models that incorporate no information about measurement error processes. Knowing that the response error process is stable validates adjustments for response error in previous research (for example, Bollinger and David 1997, Bollinger and David 1999, Bound and Krueger 1991, Mathiowetz and Duncan 1988). Knowledge about dynamics of response error also can be used to reduce total survey error (Groves 1989). Our evidence implies that survey methodologists need not design to anticipate negative learning, nor can they expect benefits from positive learning on the part of respondents. The high autocorrelation of response error can be interpreted as the consequence of a fixed effect. That is, some respondents have a high propensity to cooperate, and are willing and able to provide accurate data. Others have a low propensity to cooperate and provide false responses or no response at all. This interpretation (which can not be rejected by our evidence) makes validation of survey data imperative – to provide a model that dominates fixed effects.

Several features of the analysis that follows make findings about error processes more important than in the Grilliches and Hausman context. First, the variable in error is binary. Second, parameters being estimated pertain to a non-linear equation system. Third, existing studies fail to focus on the correlogram of response errors in repeated measurements.[1] Fourth, learning behavior associated with the survey process has not been adequately identified.

The paper continues with a discussion of the sources of response error and the possibility of identifying respondent behavior in a context in which the question sequence, interviewers and the match of interviewers to respondents also contribute to measurement error. Section 3 discusses the data. Section 4 presents our methodology for assessing gross correlations obtained from the panel of households and the meaning of the bivariate probit model. Section 5 findings from tabulation and the bivariate probit model of false positive errors. Conclusions and implications follow

[1]Bollinger and David (1999a), Bollinger (1998), Bollinger and David (1997), Hill (1993), Bound and Krueger (1991), Bound et al. (1990), Marquis and Moore (1990), Mathiowetz and Duncan(1988), Greenberg and Halsey (1983), Ferber et al. (1969a and 1969b), and Lansing et al. (1961) provide examples of information about error processes obtained by validation studies. These studies typically demonstrate that a process with predictable components generates response errors. The process also includes a random component. None of these studies focus on dynamic aspects of response error. Only Bollinger (1998), Bound and Krueger (1991) and Marquis and Moore (1990) measure any multi-period response error. None of these studies examine learning behavior or correlation associated with latent variables.

in section 6.

2 Dynamics of the propensity to err

Kalton et al. (1989, 265-266) enumerate 11 sources of error that distort estimates of gross change (flows) in status over successive interviews in a panel. The 11 sources relate to four dimensions: respondents, interviewers, questionnaires, and the mode through which data are captured. A specific propensity to err is induced by the interactions of design and implementation through these four dimensions. Repeating measurement under a fixed questionnaire and mode may alter the prior propensity to err as the behaviors of the interviewer and the respondent adapt to the ongoing mode of collection. This phenomenon is loosely referred to as conditioning, without clearly identifying the source of behavior that alters the propensity to err. Clearly, respondents react to the questions posed, the particular interviewer, and the manner of presentation of the question (face-to-face interviewing or telephone interviewing versus paper-and-pencil or electronic self-reporting).[2]

Dynamics in propensity to err in a panel thus include five possibilities of importance in this study:

1. Change of status may require adjustment to new questions by the respondent.

2. Continuing status for a respondent may induce change in the respondent's behavior, leading to lower propensity to error. Respondents' comprehension of questions and reference periods may increase; willingness to access accurate records may rise. They may learn to trust the interviewer and judge that it is appropriate

[2] Compare this enumeration of causal factors to more common models that decompose the variable of interest observed into recall-period and rotation-group effects that reference neither the respondent, nor the interviewer, nor the potential interaction between respondent and interviewer over time (Holt 1989, 342).

to reveal sensitive information. These behaviors can be termed a "positive learning" response, leading to more accurate responses with time-in-sample.

3. Continuing status for a respondent may lead to strategic behaviors that reduce the burden of participation. Respondents learn which responses lead to a longer set of questions and avoid those responses. For example, respondents to *SIPP* are asked a screener question, "Did you receive any food stamps during the previous four months?" If the respondent answers "yes," she is asked questions about each of the four months in the reference period. If she answers "no," those questions can be avoided. Respondents may learn that answering "yes" leads to longer interviews, and reduce their effort by giving false negative responses. This behavior can be termed a "negative learning" response, leading to fewer accurate responses with time-in-sample.

4. Interviewers improve their ability to route the respondent through the complexities of the questionnaire.

5. The respondent-interviewer match may change, because proxies may offer information or interviewers assigned to a particular respondent change as respondents move or, alternatively, as interviewers leave their assignment to the survey.

Because we are able to measure the actual food stamp participation status of respondents through the administrative records, we are able to identify change in respondents' behavior. We model the propensity to omit food stamp use on respondents' attributes. The resulting estimates can be interpreted as respondent behavior if change of interviewer and change in a particular interviewer's ability to navigate through the interview occur at random in our sample. That will be our maintained hypothesis. Few of the households move in ways that lead to new interviewers. As the majority (60%) of respondents report for themselves over the entire *SIPP* panel, the same interviewer-respondent pairing will be the dominant situation for both the first and second contacts. However, systematic changes in interviewer-respondent interaction are induced by the *SIPP* scientific design. The text table below makes the conditioning of interviewers and respondents clear in relation to the order of interviewing of rotation groups on successive questionnaires:

Contact	Rotation	Instrument	Experience	
			Interviewer	Respondent
1	1	1	**None**	**None**
	2	1	Experienced	**None**
	3	1	Experienced	**None**
	4	1	Experienced	**None**
2	1	2	**New** precoding	**New** conditioning on prior interview
	2	2	Experienced	**New** conditioning on prior interview
	3	2	Experienced	**New** conditioning on prior interview
	4	3	**New** topical Q's	**New** conditioning **and topical Q's**

Bold indicates a source of response error greater than that associated with an on-going panel.

The table shows that all respondents who were interviewed twice could learn from repetition of the screening questions for all types of income, including the food stamp screener. At the second contact all respondents were reminded of their response to the first interview by precoded verification questions. Those questions bound answers. Persons who changed food stamp recipiency are reminded of their prior report. Persons giving false reports on the first contact are reminded of their past behavior and are forced to reiterate the false report or admit to making an error. The conditioning questions, plus the experience of the first interview either leads the respondent to correct past responses and possibly to report more truthfully in the present, or it leads to reiteration of a false report and, most likely, a continuation of false reporting behavior in the present.

These respondent effects could be confounded through two possible mechanisms. First, learning by interviewers will induce differences between rotation group 1 and the rest of the sample. Rotation group 1 was exposed to the problems that interviewers had with an entirely new and complex panel on the first contact. Rotation group 1was exposed to interviewer inexperience with a burdensome set of questions that asked for verification of past answers on the second contact. In contrast, rotation groups 2-3 were always contacted by interviewers seasoned in the protocol for interviewing. Thus an interviewer effect may differentially affect independent subsamples. The size of the available validation sample made it unlikely that we could detect such effects.

Second, additional "topical module" questions asked of rotation group 4 during the second contact, contributed to a different problem for interviewers and respondents. The additional questions created new challenges for the interviewers and additional burdens for the respondent. A small proportion of respondents terminated their interview just before answering the topical module questions (and after completing the "core" questions asked in the first interview). It is not clear how this "new" questionnaire could induce a change in response behavior on the core questions. Problems with the additional questions and concern for respondent burden may have been communicated to the respondent by interviewers who were uncertain about problems that would arise with the topical module questions. Again, the small sample implies that we could not detect an effect of interviewer behavior, or of interviewer-respondent interactions.

This paper examines three hypotheses that can be tested using the repeated measurement of food stamp recipiency:

1. Respondents receiving food stamps show positive learning in their answers to the survey.
2. Respondents receiving food stamps show negative learning in their answers to the survey.

3. Respondent behavior is consistent with the *cooperator* hypothesis. Namely, respondents have a propensity to provide accurate answers. If the propensity is high, answers will be complete and correct. If the propensity is low, answers will be incomplete or false.

We find that response error rates are remarkably stable over the first two interviews of the SIPP. Estimated models of error give no support to positive or negative learning. Substantial correlation between response error in the first interview and response error in the second interview exists after systematic effects have been modeled. This finding supports the *cooperator* hypothesis: respondents who are willing and able to provide accurate responses to survey questions do so over repeated interviews, while respondents who are unwilling or unable to provide accurate responses at one time will have high rates of error in subsequent interviews.

3 Data

The data used here derive from the 1984 panel of the *Survey of Income and Program Participation (SIPP)*. The 1984 panel began interviewing households in October 1983. Each interview repeats the same set of core questions for all adults in the household. Demographic information is also collected about minor children. Each interview elicits information about events in the previous four months. Thus an interview occurring in December 1983 would ask questions about a reference period pertaining to the months of August, September, October, and November of 1983.

Researchers at the Census Bureau compiled a census of state administrative records for the Food Stamp Program in Florida, Pennsylvania, and Wisconsin. Those records were matched by the Census Bureau to individuals in the SIPP sample for interviews 1 and 2. The match was based on name, social security number, address, and demographic information. Detailed information on the match can be found in Marquis and Moore (1990). The matched data are referred to as the validation sample. Since the validation data are specific to the state, any household that moved out of state between interview 1 and interview 2 was discarded from the sample, leaving 2597 households in the dynamic analysis.[3]

Data used for analysis in this paper augment the validation sample with public use data from *SIPP*. The augmented individual records were aggregated into household records. The household is the relevant unit of analysis for food stamp participation.[4] Change in residence of household members and change in the membership of the household pose a serious conceptual problem for longitudinal analysis (Citro and Herandez 1986). We resolve the problem by using the household units that exist at the time of the second interview as the unit of analysis. Hence, we examine the reporting behavior of an identical number of household units over time. For most households, there are no changes between the first and second interview. Households which split are followed and appear as two separate households in the second interview. Every interview 2 household is uniquely linked to a single interview 1 household. For all households interview 1 household attributes are imputed to the

[3] Administrative records for the Food Stamp Program are nearly accurate measures of true Food Stamp participation. Federal auditing provides an incentive for each state agency to keep accurate, machine-readable records. The matching procedure used by the Census is based on multiple levels of information and has a high success rate. Therefore disparity between the administrative record and the survey response can be attributed to response error.

[4] The Food and Nutrition Service defines the food stamp unit to be all individuals sharing cooking facilities or eating together. In order for individuals to be considered as a separate food stamp unit they would generally have to have separate kitchen facilities. Since housing units with multiple kitchen facilities are extremely rare, each household is assumed to correspond to a single food stamp unit (Bollinger and David 1997).

Some modifications to the Food Stamp unit were adopted in 1993 that are now used by the Food and Nutrition Service to impute units to groups of persons within the household. See USDA/FNS 1997.

interview 2 household for the first time period. That is, household earnings and household size in interview 1 would be attributed to any member who relocates to another address in interview 2. In our construction, the demographic information about the first household is used for both of the new households in interview two.

The analysis here focuses on errors of omission: false negative survey responses. Errors of omission are much more likely than errors of commission, that is, reporting participation when administrative records fail to indicate participation or fail to match (Marquis and Moore 1990 and Bollinger and David 1997).[5] Bollinger and David (1999) note that response error in reporting food stamps is concentrated at the screener question. That is, over 90% of the individuals who fail to report food stamps for a particular month in the reference period fail to report participation for *any* month in the reference period. Errors of omission are primarily a total denial of participation at any time in the reference period, rather than a simple timing error. The main focus of this paper will be on households at risk to answer the food stamp questions in both periods: households that participated in both interview 1 and interview 2. Only those households can exhibit identifiable learning behavior. (Households who participate only during the second reference period may have learned from the structure of the questionnaire, but we cannot distinguish positive learning from truth-telling, nor can we distinguish negative learning from other forms of misinformation.)

Table 1 tabulates household responses to the screener questions and the matched administrative record. The rows represent classification according to administrative records, the columns represent classification according to SIPP. The first row reveals that 2370 households did not participate in the Food Stamp Program at any time

[5] Table 1 shows errors of commission. Inferring behavior from a few cases is suspect.

during the first two interviews. 28 households participated during interview 1 only (exited FSP) and 22 households participated during only interview 2 only (entered FSP). We focus on the 177 households who participated in the Food Stamp Program during both interviews.

Table 2 provides descriptive statistics for these participating households. The average household size shows little change between interviews. Since the time interval is short (4 months) the result is not surprising. Approximately 70% of the households were headed by a single individual. Single-female-headed households dominate the sample. Household earnings relate to the month prior to the interview month. Household earnings aggregates wages and salaries from all jobs, and net income from sole proprietorship, over all adults in the household. Income from other sources is excluded.

4 Methodology

Recall our hypotheses:

1. Respondents receiving food stamps show positive learning in their answers to the survey.

2. Respondents receiving food stamps show negative learning in their answers to the survey.

3. Respondent behavior is consistent with the *cooperator* hypothesis.

Two methods are used to examine these hypotheses, tests based on the cross tabulation of the two replications of measures and tests based on estimation of a bivariate probit model.

4.1 Tests based on cross tabulation

The three hypothesis can be tested by cross tabulating responses and the true status. If positive learning is occurring the proportion of errors in reporting food stamp program participation should fall from the first interview to the second. More precisely, the error rate should fall for respondents that participate in both interviews. The difference in the proportion of respondents who fail to report participation in the two interviews is the basis for a test of the learning hypotheses. (The standard error of that difference will take into account any correlation across interviews, as the samples are not independent and constitute an unbalanced panel.) The hypothesis of negative learning is tested by the same statistic. The learning hypotheses also imply that the difference in the proportions should be more pronounced for households that participate in both interviews: They *are* the source of learning effects. Tests will be done on both the overall proportion of errors (including households who only participate in one interview) and on the subset of households who participate in both interviews.

The hypothesis that response errors are correlated is also tested by cross-tabulation. Independence across interviews can be assessed by the chi-squared statistic. The test is applied to households that participate in both interviews. The chi-squared test uses the four cells that describe survey responses for the balanced panel ($n = 177$). By comparing the cell counts to the expected cell counts if the processes were independent (determined by using marginal error rates), we can discern if there is any evidence for independence of response across interviews. We also estimate the correlation of errors of omission between interview 1 and interview two for those households who participate in both interviews.

4.2 Bivariate Model

Tests for the three hypotheses are more definitive when systematic variation in response error is modeled in a bivariate probit. Bollinger and David (1997) establish that demographic variables correlate to response error and estimate a single-index response function for errors of omission on the interview 1 data.[6] The distribution of household demographic characteristics tends to be stable over the two interviews (see Table 2[7]). Observed persistence in response error may be due to stability of the distribution of underlying population characteristics. That persistence differs from persistence of a latent variable posited by the cooperator hypothesis. The need to distinguish persistence associated with measured covariates from persistence associated with unmeasured variables motivates the bivariate model that we estimate on the balanced sample.

Individual households choose to give accurate responses if the cost (psychic or otherwise) is less than the benefit. This leads to a reduced form single index model of omission errors for any particular period:

$$O_{it} = \begin{cases} 1 & \text{if } X_{it}\beta_t + \varepsilon_{it} > 0 \\ 0 & \text{otherwise.} \end{cases} \quad t = 1, 2; j \quad \epsilon \text{ balanced panel}$$

A particular household, i, commits an error of omission, $O_{it} = 1$, in period t if an index exceeds a threshold (i.e., cost is less than subjective benefit of an accurate

[6] The index included per capita income, marital status and gender. In later work (Bollinger and David 1999) we determine that earned income per capita is more appropriate as it excludes the value of transfer programs.

[7] To Chris:

Table 2 does not show the distribution. It would be more satisfactory to show the quartiles of the distribution:

	1st	2nd
1/4		
median		
3/4		

report). Covariates, X_{it}, may change from period to period. We specify a saturated model that allows coefficients of the covariates to vary by interview.

Two processes generate the unobserved term ε_{it}– fixed effects and autocorrelation. Fixed effects relate to an unmeasured and unchanging attribute of the respondent; they are defined by

$$\varepsilon_{it} \equiv u_i + \nu_{it}.$$

The u_i is a random variable, independent from v_{it}, and constant for each household over time. $u_i > 0$ represents a time-invariant predisposition to response error ($u_i < 0$ a predisposition to accuracy). Autocorrelation is defined by

$$\varepsilon_{it} \equiv \phi \varepsilon_{it-1} + \eta_{it}.$$

The propensity to respond is a linear combination of last period's propensity and a random term. $\phi > 0$ reflects persistence in response error. $\phi < 0$ induces cyclical variation in response propensity – that is, good reporters in the current period, become bad reporters in the next period.

When observations on a particular household are available for only two periods, it is impossible to identify the underlying stochastic process. However, as Grilliches and Hausman (1986) point out, the main information needed for consistent estimation is the autocorrelation of the response errors. We would like to know the entire autocorrelation structure – the correlogram, but that can not be obtained from two replicates of the measurement. All that can be estimated is the correlation, ρ_{12}, between ε_{i1} and ε_{i2}. If the random effects model is correct, then $\rho_{12} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}$, the percentage of variance in ε_{it} that is due to u_i. If the autocorrelation model is correct, then $\rho_{12} = \phi$, the percentage of variance in ε_{it} that is due to ε_{it-1}. We report ρ_{12} without taking a stance on the underlying structure. For example, an unmeasured variable, such as intoxication at the time of the interview, might account for the error correlation yet could be altered at a subsequent interview, should the respondent take the sobriety oath. In this case autocorrelation would be more appropriate than fixed effect, as some alcoholics fail to adhere to the sobriety oath.

To estimate the two-period model, we assume that ε_{it} is independent of $X_{is} \quad (s = 1, 2)$ and that the vector $(\varepsilon_{i1}, \varepsilon_{i2})$ is distributed according to the joint standard normal distribution with a correlation ρ_{12}.[8] The parameters in the model are defined by the assumption that ε_{it} is independent of the X_{is} variables. The unobservable term ε_{it} is by definition the unpredictable behavior. Given the model and the joint normality assumption, maximum likelihood can be used to estimate the parameters β_1, β_2, and ρ_{12}.

The three hypotheses can then be tested through the parameters estimated for the model. Differences in the intercepts reveal differences in the level of error in the two interviews. Positive learning implies that the intercept should be lower in interview two than in interview 1, while negative learning implies that the intercept should be higher in interview two than in interview 1. Similarly, differences in response coefficients reveal a shift in behavior that is mediated by covariates. The learning hypotheses do not imply a change in these coefficients; however, absence of learning behaviors implies identical response coefficients for both interviews.

[8]Much literature (Hsiao, 1986) has considered the problems associated with estimating models where the random (or fixed) effect is correlated with the regressors. As this model represents a statistical model, rather than a structural model, we are unconcerned by this possibility.

5 Results

5.1 Cross-Tabulation Results

Table 1 displays a cross tabulation of the responses to two periods of questions concerning food stamp program participation. The columns represent the four possible pairs of responses in the questionnaire, while the rows represent the four possible true states in which the household could be. The cell in the upper left corner demonstrates that 2356 households correctly reported not participating in the food stamp program. There were 2370 households who actually did not participate, so 99.4% correctly reported not participating. The row total, for "Both interviews" shows that 177 households (6.8%) of the 2597 in the sample, participated in the Food Stamp Program during both reference periods. Of those 177 households, 138 (78%) reported participating in both interviews, while 16 (9%) failed to report participation in either interview.

The two cells with zeros are striking: These indicate that households that terminated the Food Stamp Program in interview 1 did not later report participating in interview 2. Similarly, households that entered the program during interview 2 did not report that they participated in interview 1. This suggests that timing is not a major cause of response error (supporting previous work of Marquis and Moore 1990, and Bollinger and David 1997). As the interview for interview 1 occurs in the month *after* the reference period for interview 1, a real risk exists that households who begin participating in the Food Stamp Program during the interview month will report food stamp use for the interview 1 reference period. Yet none do.

As can be seen from table 1, errors of commission – reporting FS participation when none occurs – are rare: Less than 1% of households not participating report participation; only 0.6% give false positive responses. However, errors of omission are more common, 9% of the 177 households who participate during both reference periods fail to report any participation, while and additional 13% fail to report in at least one period. Of those 50 households who participate in only 1 interview, 20% fail to report that participation. Clearly, bias will result more from omission than commission.

To test the two learning hypotheses, we examine whether response error rates change between the two periods. We compare the marginal rate for errors of omission in the two periods. Adding the second and fourth rows of the table, we find that 205 households participate in interview 1. 34 or 16.5% of interview 1 participants fail to report participation. Similarly, adding the third and fourth rows of the table, we find 199 households who participate in interview 2. 31 or 15.6% of interview 2 participants fail to report participation. A simple test for the null hypothesis that the two proportions are equal yields the test statistic 0.43.[9] The null hypothesis that the two proportions are estimating the same population proportion (the population probability of an error of omission) is accepted at all conventional levels. No evidence supports learning.

A second approach to testing for positive and negative learning is to examine the proportion of people who fail to report participation in only one interview, even though they participate in both periods. If positive learning is occurring, we expect that the number of households that misreport only in interview 1 would be larger than the number of households that misreport only in interview 2. If negative learning is occurring, we would expect the opposite. The data provide no evidence for either

[9]The two estimators are correlated. The standard error of the difference in these proportions is
$V\left(\widehat{P}_1\right)+V\left(\widehat{P}_2\right)-2Cov\left(\widehat{P}_1,\widehat{P}_2\right).$

of these hypotheses, since there are 11 households who only misreport in interview 2, and 12 households who only misreport in interview 1. The formal test statistic is -0.24: the null hypothesis is again accepted at all conventional levels. There is no evidence for a systematic learning effect.

To test that there is correlation across time in the response error, we consider the households participating in both periods. This can be tested using a chi-squared test for independence. The four cells in the fourth row of the table, pertaining to the 177 households who participate in both interviews, can be examined as a 2×2 contingency table. The chi-squared test statistic of independence in these four cells of the table is 45.2. The null hypothesis is rejected at all standard significance levels; response error is not independent across interviews. Inspection of Table 1 shows that fully 78% of the households report participation in both interviews, and 9% of the households fail to report in both interviews. The preponderant division of households into accurate reports in both interviews or complete denial in either interview supports the cooperator hypothesis. Further, the correlation of the indicator for an error of omission is 0.54. A simple test of the significance of this correlation is $z = 7.18$. The null that the correlation is zero is rejected at all conventional levels. The value indicates a strong positive correlation between response error in the first interview and response error in the second interview.

5.2 Bivariate Probit Results

As noted in the introduction, Bollinger and David (1997) determine that response error is related to demographic and economic variables for the first interview. The bivariate probit model described in the previous section allows a more extensive test of covariates as well as a refined approach to testing the learning hypotheses. The estimated probit appears in Table 3. We present two specifications for covariates. Model 1 repeats the specification used in Bollinger and David (1997). Model 2 incorporates more demographic information. Households headed by a married person are questioned about Food Stamp participation for at least two adults, while many households headed by single parents are questioned only once (as they include no other adults). The level of response error associated with that feature of the survey design can be captured in coefficients estimated separately for married couples, single males and single females (incorporated in the intercept). Looser parameterization of earnings and household size effects is also incorporated in Model 2. The expectation that survey design matters and doubts about the nature of income-household-size interactions lead us to prefer model 2.

Both models confirm that response error is positively related to earnings (either per capita or total). This finding could be induced by threat. Households with greater earning capacity include those who may wish to preserve their social standing by keeping use of welfare programs secret from others in their community. Anecdotal studies indicate that sharing information about Food Stamp Program availability is much more prevalent in households living in urban ghettoes, who we guess have lower earning capacity. Thus differences in earning capacity could be associated with differential rates of omission error.

Household structure also determines level of omission errors. Larger households are less likely to have error. This may be due to the larger number of interviews taken in larger households that can lead to correct responses as interviews are aggregated. Finally, households headed by single males are more likely to commit an error of omission than others. These results are consistent with Bollinger and David (1997). Now we focus on results unique to the two-interview model.

Learning can be tested by examining differences in the intercept between interviews. If positive learning occurs, the intercept will fall from interview 1 to interview 2. If negative learning occurs, the intercept will rise. Since the intercept rises from interview 1 to interview 2, there is no evidence for positive learning. Testing the null hypothesis that the intercepts are equal in model 2 yields the test statistic 0.9205. We accept the null hypothesis that intercepts are the same in interview 1 and interview 2. No evidence for learning emerges[10].

As none of the coefficients differ significantly between the first and second interview, the model is stable. Learning does not give a clear hypothesis about how slopes might change, but no change implies no learning when intercepts are unchanged. Using model 2, we calculate the Chi-squared statistic for the difference between the slope coefficients of the two time interviews as 4.28. The critical value is 11.07 – We cannot reject the null hypothesis that the model is identical for the two interviews. The stability of the response error model across two interviews confirms the one-period models and conclusions of Bollinger and David (1997, 1999). Furthermore, it appears reasonable to apply the cross-sectional model of response errors to later interviews of the 1984 SIPP. Lastly, stability leaves hope that a similar model applies to other panels.

The correlation coefficient implies persistence of response errors. The correlation is indistinguishable between the two specifications at $\rho_{12} = 0.73$. This estimate is significantly different from zero, and qualitatively high. The correlation is higher than the gross correlation calculated from cross tabulation. Hence, controlling for observable characteristics, reveals more positive autocorrelation in the stochastic process underlying response error than an unconditioned estimate. This finding is consistent

[10] The covariance between the estimated coefficients is 0.063.

with the *cooperator* hypothesis: individuals who are willing and able to cooperate with the survey will provide more accurate answers throughout a panel.

6 Conclusions

This paper has examined three important hypotheses concerning the structure of response error in reporting Food Stamp Program participation. We find no evidence for learning, either positive or negative. Respondents do not learn to be more accurate participants over two interviews. Neither do they learn to avoid the more burdensome detailed questions. This finding is good news for survey designers and for data users. Survey designers can concentrate on the cognitive aspects of single interviews without concern for time-in-sample effects on respondents. For data users, it suggests that, holding other factors constant, response accuracy is not getting worse over time. Estimates at the beginning or subsequent measurement in a panel appear to be about the same.

Combining the correlation in response errors and the correlation between first period response error and failure to respond to later interviews (Bollinger and David 1999), gives an enticing conclusion: Namely, as attrition occurs, the level of response error should decline. In this paper we determine that individuals with high ε_{i1}(propensity to commit an error) are likely to have higher than average ε_{it}. Our 1999 work shows that response error in the first period correlates to interview nonresponse in later periods (significant at +0.4 from interview 1 to interview 2). We can conclude that households that made errors in the first interview fail to give later interviews to a greater extent than households who made no errors in the first interview. Furthermore, the two correlations taken together imply that estimates can be improved in several ways:

- households with missing interviews can be given reduced weight

- all households can be given weight proportional to their propensity to omit food stamp recipiency, and

- combinations of the two.

Clearly, these implications contradict conventional survey methods and selection models that assign increased weight to households from groups that are determined to have higher than average propensity to attrit.

The stability of the model across time periods is important in extending the results of Bollinger and David (1999, 1997): We believe that these models can be used throughout the 1984 panel. Fragmentary data from the 1990 SIPP panel suggests that the relation between per capita earnings and omission errors is consistent with Model 1 in Table 3, although the level may be lower. This leads us to suspect that the models already estimated could be applied in other SIPP panels, if the net aggregate rate of underreporting is used to adjust the intercept of the error function downwards (USDA/FNS 1997).

Knowledge gained here about the structure of the dynamic measurement process may foretell that measurement errors generally are correlated over time. We see no reason to think that Food Stamps is an idiosyncratic case Were measurement errors frequently correlated over time, the error structure assumed in some applications of Grilliches and Hausman (1986) is most likely wrong. The prior report of a variable being model is an inappropriate instrument when autocorrelated measurement errors are present. The autocorrelation coefficient of response errors must be introduced to correct to make the instrument orthogonal to the current value of the variable, as Grilliches and Houseman (1986) recommend.

Further, the evidence above implies that entire spells of participation are likely to go unreported. This will substantially bias estimation of Food Stamp participation spells via duration models and multivariate models of Food Stamp participation and labor force participation in which response error must be considered as an additional endogenous equation. Bollinger and David (1997, 1999) note that the important term in constructing likelihood functions which account for response error is the probability of response error given the true state. The models estimated there can be generalized to models of the dynamics of participation over several periods.

References

Biemer, Paul P. and Ronald S. Fecso. 1995. "Evaluating and controlling measurement error in business surveys". In Brenda Cox et al. *Business Survey Methods*. Wiley. 257-282.

Biemer, Paul and Lynne Stokes. "Approaches to the modeling of measurement errors." In Paul Biemer et al. *Measurement Errors in Surveys*.487-516

Bollinger, Christopher R. and Martin H. David. 1997. Modeling discrete choice with response error: Food Stamp participation. *J. of the American Statistical Association*, Vol. 92, no. 439 (September), pp. 827-835.

Bollinger, Christopher R. and Martin H. David. 1999. Estimation with response error and nonresponse: Food Stamp participation in the SIPP. SSRI Working Paper 9920. (http://www.ssc.wisc.edu/econ/archive/author.htm#d).

Bollinger, Christopher R. 1998. Measurement error in the *Current Population Survey*: A nonparametric look. *Journal of Labor Economics* Vol. 16, no. 3 (July), pp. 576 - 594.

Bound, John; Brown, Charles; Duncan, Greg J.; and Rodgers, Willard L. 1990. Measurement error in cross sectional and longitudinal labor market surveys: Results from two validation studies. *Panel Data and*

Labor Market Studies, edited by J. Hartog, G. Ridder and J. Theeuwes, pp. 1-19. New York: Elsevier, 1990.

Bound, John and Kreuger, Alan B. (1991) The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *J. of Labor Economics,* Vol. 9 no. 1 (January 1991), pp. 1-24.

Citro, Constance C. and Donald Hernandez. 1986. Longitudinal household concepts in SIPP. *Proceedings of the Social Statistics Section of the American Statistical Association*, pp. 361-365.

Doyle, Patricia. 1990. "Food Stamp Program Participation Rates, August 1985" *Current Perspectives on Food Stamp Program Participation Rates*

Ferber, Robert F. 1964. Does a panel operation increase the reliability of survey data: The case of consumer savings. *Proceedings of the Social Statistics Section of the American Statistical Association*, pp. 210-216.

Ferber, R.J. Forsythe, H. W. Guthrie, and E. S. Maynes. 1969a. Validation of a national survey of consumer financial characteristics: Savings accounts. *Review of Economics and Statistics*, Vol. 51 no. 4 (November), pp. 436-444.

Ferber, R., J. Forsythe, H. W. Guthrie, and E. S. Maynes. 1969b. Validation of a national survey of consumer financial characteristics: Common stocks. *J. of the American Statistical Association*, Vol. 64 no 326 (June), pp. 415-432.

Fuller, Wayne A. 1987. *Measurement Error Models.* New York, NY: John Wiley & Sons.

Gooderau, Karen, Howard Oberheu, and Denton Vaughn. 1988. "An assessment of the quality of survey reports of income from the Aid to Families With Dependent Children (AFDC)" *J. Business and Economic Statistics. v.2/2.*

Greenberg, David and Halsey, Harlan. 1983. Systematic misreporting and effects of income maintenance experiments on work effort: Evidence from the Seattle-Denver Experiment. *J. of Labor Economics,* Vol. 1 no. 4 (October), pp. 380-407.

Griliches, Zvi and Hausman, Jerry. 1986. Errors in Variables in Panel Data. *J. Econometrics,* Vol 31, pp.*93-118.*

Groves, Robert M. 1989. *Survey Errors and Survey Costs.* New York, NY: John Wiley. Chapter 1.

Hill, Daniel H. 1993. Response and sequencing errors in surveys: A discrete contagious regression analysis. *J. of American Statistical Association*, Vol. 88 no. 423 (September), pp. 775-781.

Holt, D. 1989. Panel conditioning: Discussion. *Panel Surveys.* Edited by Daniel Kasprzyk, Greg Duncan, Graham Kalton, and M.P. Singh. New York: John Wiley and Sons. pp. 340-347.

Hsiao, Cheng 1986. *Analysis of Panel Data.* Econometric Society Monographs #11, Cambridge, MA: Cambridge University Press.

Kalton, Graham, Daniel Kasprzyk, and David McMillen. 1989. Non-sampling errors in panel surveys. *Panel Surveys.* Edited by Daniel Kasprzyk, Greg Duncan, Graham Kalton, and M.P. Singh. New York: John Wiley and Sons. pp. 249-270.

Keane, M. and R. Moffitt. 1998. A structural model of multiple welfare program participation and labor supply. *International Economic Review*, Vol. 39, no. 3 (August), pp. 553-589.

Lansing, John B., Gerald P. Ginsberg, and Kaisa Braaten. 1961. *An Investigation of Response Error.* Urbana, IL: Bureau of Economic and Business Research, University of Illinois.

Marquis, Kent and Jeffrey Moore. 1990. Measurement errors in the SIPP program reports. *Proceedings of the Bureau of the Census Annual Research Conference.* Washington, DC: Bureau of the Census, pp. 721-745.

Mathiowetz, Nancy A. and Greg J. Duncan (1988). Out of work, out of mind: Response errors in retrospective reports of unemployment. *J. of Business and Economic Statistics,* Vol. 6 no. 2 (April), pp. 221-229.

Morgenstern, Oskar. 1963. *On the Accuracy of Economic Observations.* Princeton, NJ: Princeton University Press. Census, pp. 721-745.

Rodgers, Willard L. Greg J. Duncan and Charles Brown. 1993. Errors in survey reports of earnings, hours worked and hourly wages. *J. American Statistical Association,* Vol. 88 no. 424 (December), pp. 1208-1218.

Rodgers, Willard L. and A. Regula Herzog. 1987. Covariances of Measurement Errors in Survey Responses, *J. of Official Statistics, Vol.* 3 no. 4, pp. 403-418.

Traugot, M.W. and J.P. Katosh. 1979. Response validity in surveys of voting behavior. *Public Opinion Quarterly* Vol. 43 pp. 359-377.

United States Department of Agriculture/ Food and Nutrition Service. 1997. "Trends in Food Stamp Program participation rates: Focus on August 1995" (http://www.fns.usda.gov/oane/menu/Published/fsp/fsp.htm)

7 Tables

Table 1: Joint interview 1 and interview 2 Screener Level Responses

	SIPP Food Stamp Screener Response				
Administrative record	**No FS**	**Food Stamps in**			**Row**
		interview 1	**interview 2**	**Both**	**totals**
No food stamps	2356	3	10	1	2370
(Row %)	(99.4%)	(0.1%)	(0.4%)	(0.04%)	
(column %)	(98.9%)	(9.1%)	(25.6%)	(0.7%)	(91.3%)
Food stamps in					
interview 1 only	6	19	0	3	28
(Row %)	(21.4%)	(67.9%)	(0%)	(10.7%)	
(column %)	(0.2%)	(57.6%)	(0%)	(2.1%)	(1.1%)
interview 2 only	4	0	17	1	22
(Row %)	(18.2%)	(0%)	(77.3%)	(4.5%)	
(column %)	(0.2%)	(0%)	(43.6%)	(0.7%)	(0.8%)
Both interviews	16	11	12	138	177
(Row %)	(9.0%)	(6.2%)	(6.8%)	(78.0%)	
(Column %)	(0.7%)	(33.3%)	(30.8%)	(96.5%)	(6.8%)
Column Totals	2382	33	39	143	2597
(Row %)	(91.7%)	(1.3%)	(1.5%)	(5.5%)	

Table 2: Descriptive Statistics for Food Stamp Participants

	interview 1		interview 2	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std.Dev**
Omission	0.158	0.366	0.153	0.361
Household Earnings/capita	132.14	317.97	123.25	239.57
Household Earnings	421.06	833.02	386.68	639.58
Household Size	3.311	1.834	3.412	1.890
Single Female Headed HH	0.548	0.499	0.548	0.499
Single Male Headed HH	0.136	0.343	0.164	0.371
Sample Size		177		

Table 3: Coefficients of Joint Omission Error Model
Participants in Both interviews (Standard errors in parentheses)

Variable	**Model 1**		**Model 2**	
	interview 1	**interview 2**	**interview 1**	**interview 2**
Intercept	-1.613	-1.501	-1.208	-0.760
	(0.171)	(0.164)	(0.425)	(0.427)
HH earnings/capita	0.00364	0.00294		
	(0.00063)	(0.00060)		
HH earnings			0.000987	0.000869
			(0.000200)	(0.000211)
HH size			-0.143	-0.265
			(0.0897)	(0.093)
Single Female			-0.235	-0.103
			(0.343)	(0.347)
Single Male			1.046	0.804
			(0.393)	(0.376)
Intertemporal	0.728		0.727	
Correlation ρ_{12}	(0.109)		(0.120)	
Likelihood	-0.599		-0.557	
N	177		177	