“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 error is highly correlated for these interviews. Hypotheses that the error process can be explained by learning behaviors are rejected. Bivariate probits of response error, including income and household characteristics as covariates, are stable across the two periods and show that autocorrelation in the errors is not attributable to readily available attributes. These findings support the cooperator hypothesis previously forwarded by Bollinger and David (1999a).

11 November 1999

Acknowledgement 1 *Thanks to Kent Marquis and the Bureau of the Census for access to the data and logistic support and encouragement. 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.*

.

1 Introduction

Groves (1989), Morganstern (1963), Fuller (1987), and others explore the importance of measurement error in estimating levels of economic activity and models of economic behavior. They advocate primary research into the extent and structure of measurement error in survey data. Some research has begun to address this important issue for widely used surveys – for example, 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). Marquis and Moore (1990) find that response errors in reporting participation in transfer programs are often asymmetric (false negative error rates exceed positive error rates). Error rates vary dramatically across programs. Furthermore, a substantial seam effect exists: exits and entrance to programs are reported more frequently across the seam between interviews than within the reference period of four months. Rodgers and Herzog (1987) and Bound et al (1990) find multivariate correlation in response errors on different measures in the same interview. Bollinger (1998) and Bound and Krueger (1991) find temporal correlation of errors in economic variables within a panel .

Repeated exposure of subjects to the same questions in a panel survey affords the opportunity to estimate temporal correlation in response errors, as well as the multivariate correlations that might characterize variables measured at a single point in time. This opportunity is exploited here. Our prior work offers a clear starting point for this investigation. Bollinger and David (1997, 1999a), show that response errors in measuring participation in the Food Stamp Program are related to income and household size and to wave nonresponse in the *1984 Survey of Income and Program Participation* panel. Asymmetry of errors is confirmed, when response error is conditioned on those covariates. The central hypothesis of both papers is the *cooperator hypothesis* – A distribution of propensities to tell truthfully (and to err) is the result of omitted covariates or a nuisance parameter. The hypothesis implies that response errors are correlated over time whenever omitted covariates are stable or when the nuisance parameter is a fixed effect.

We undertake to confirm previous work that demonstrates substantial correlation between response errors in successive interviews (Bollinger 1998, Bound and Krueger 1991, Traugott and Katosh 1979, Ferber 1964). Our investigation extends the models of response errors in the previous work to include errors measured at two points in time. We can test the cooperator hypothesis and alternative hypotheses of positive or negative learning in successive interviews.

The work here has an important relationship to earlier work on conditioning in panel surveys.

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).[1]

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 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.

[1] 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).

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 and as interviewers leave the organization.

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

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, and 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 at the end of the core questions. 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.

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, or wave, 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 waves 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 wave 1 and wave 2 was discarded from the sample, leaving 2597 households in the dynamic analysis.[2]

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.[3] Change in residence of household members and change in the membership of the

[2]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.

[3]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).

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 wave. Every wave 2 household is uniquely linked to a single wave 1 household. For all households wave 1 household attributes are imputed to the wave 2 household for the first time period. That is, household earnings and household size in wave 1 would be attributed to any member who relocates to another address in wave 2. In our construction, the demographic information about the first household is used for both of the new households in wave two.

The analysis here focuses on errors of omission: false negative survey responses. Errors of omission are much more likely than errors of commission: reporting participation when administrative records fail to indicate participation or fail to match (Marquis and Moore 1990 and Bollinger and David 1997, 1999a).[4] Bollinger and David (1999a) 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 who participated in both wave 1 and wave 2. Only those households can exhibit identifiable learning behavior. Households

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

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. As can be seen from the row totals which represent the administrative records, 2370 households did not participate in the Food Stamp Program at any time during the first two waves. There are 28 households who participated during wave 1 only (exiters) and 22 households who participated during only wave 2 only (entrants). We focus on the 177 households who participate in the Food Stamp Program during both periods.

Table 2 provides descriptive statistics for these participating households. The average household size (HHsize1 and HHsize2) shows little change between waves. 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 (SingFem1 and SingFem2) dominate the sample. Household earnings (HHearn1 and HHearn2) relate to the month prior to the interview month. Household earnings aggregates wages and salaries from all jobs, and net income from sole proprietorships, over all adults in the household. Income from other sources is excluded. Household income (HHinc1 and HHinc2) includes earnings and income from all other sources, including food stamps and other government transfer programs.

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, tabulation and probit estimation. Positive learning is indicated by error rates that fall between wave 1 and wave 2. Negative learning is indicated by error rates that rise between wave 1 and wave 2. The cooperator hypothesis is indicated by correlation of error rates over both waves: the probability of responding correctly in one period should be higher if we know that a correct response was given in the other period.

Although cross-tabulation is informative, the bivariate probit is more general. 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 wave 1 data. The distribution of demographic characteristics of a household tends to be stable over the two-wave period (see Table 2). Observed persistence in cross-tabulation may be due to stability of the underlying response function. That persistence is different than the latent variable mechanism implicit in 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 for wave 1 and wave 2 below.

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 (cost - benefit of an accurate report). The covariates X_{it} may change from period to period. We specify a saturated model for the covariates.

Two competing models for the unobserved term ε_{it} are random effects and autocorrelation. Random effects are defined by

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

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

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

The propensity to respond is determined by last period's propensity plus an additive 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 distinguish the two models of the stochastic process. All that can be estimated is the correlation, ρ_{w1-w2}, between ε_{i1} and ε_{i2}. Thus our model and estimation approach focuses on that correlation. If the random effects model is correct, then $\rho_{w1-w2} = \frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2}$, the percentage of variance in ε_{it} that is due to u_i. If the autocorrelation model is correct, then $\rho_{w1-w2} = \phi$, the percentage of variance in ε_{it} that is due to ε_{it-1}. We simply focus on ρ_{w1-w2} as a measure of intertemporal correlation in unobservable characteristics determining response behavior.

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.[5] 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.

5 Results

First, consider Table 1. The cells with zeros are striking. These indicate that households that terminated the Food Stamp Program in wave 1 did not later report participating in wave 2. Similarly, households that entered the program during wave 2 did not report that they participated in wave 1. This strongly 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 wave 1 occurs in the month *after* the reference period for wave 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 wave 1 reference period. Yet none do.

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 wave 1. 34 or 16.5% of wave 1 participants fail to report participation. Similarly, adding the third and fourth rows of the table, we find 199 households who participate in wave 2. 31 or 15.6% of wave 2 participants fail to report

[5] 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.

participation. A simple test for the null hypothesis that the two proportions are equal yields the test statistic 0.42909.[6] 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 wave, even though they participate in both periods. If positive learning is occurring, we expect that the number of households that misreport only in wave 1 would be larger than the number of households that misreport only in wave 2. If negative learning is occurring, we would expect the opposite. The data provide no evidence for either of these hypothesis, since there are 11 households who only misreport in wave 2, and 12 households who only misreport in wave 1.

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 test statistic is 45.16. The chi-squared critical value (at the 1% alpha level) is 11.3. The null hypothesis is rejected; response error is not independent across periods. Inspection of Table 1 shows that fully 78% of the households report participation in both periods, and 9% of the households fail to report in both periods. The preponderant division of households into all good reports or all bad reports supports the cooperator hypothesis.

As noted in the introduction, it has already been established (Bollinger and David 1997) that response error is related to demographic and economic variables. Estimation of the bivariate probit model described in the previous section allows a more

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

refined approach to testing these hypotheses. The results of that estimation are presented in Table 3. We present two specifications for covariates. Model 1 repeats the specification used in the one-period analysis of Bollinger and David (1997). Model 2 incorporates more demographic information; it is particularly important that households headed by a married person afford the opportunity for enumerators to get information on Food Stamp participation from at least two adults, whereas many households headed by single parents include no other adults. That feature, and the looser parameterization of earnings and household size effects, leads us to a slight preference for model 2.

Both models confirm that response error is positively related to earnings (either per capita or total). This finding suggests that stigma or threat may cause some response error. In both models, household structure is also important. Larger households are less likely to have error. This may be due to the fact that larger households have more chances to be interviewed and "get it right." Finally, households headed by a single male are more likely to commit an error of omission than other types of households.

Learning can be tested by examining differences in the intercept between periods. If positive learning occurs, the intercept should fall from period 1 to period 2. If negative learning occurs, the intercept will rise. Since the intercept rises, 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 wave 1 and wave 2. No evidence for learning emerges.

As none of the covariates are significantly different from each other, the deterministic model appears to be stable over time.

The estimated correlation coefficients support persistence. The correlation is indistinguishable between the two specifications and $\rho_{w1-w2} = 0.727$ in model 2. This estimate is significantly different from zero, and qualitatively high. Hence, after controlling for the dominant observable characteristics, strong evidence emerges that response error is correlated over time. This is consistent with the cooperator hypothesis: individuals who are willing and able to cooperate with the survey will provide more accurate answers throughout the survey.

6 Conclusions

This paper has examined a number of hypotheses concerning the structure and causes of response error in reporting Food Stamp Program participation. These properties are likely to extend to other programs. 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 one finding is good news for survey designers and for data users. Survey designers can concentrate on the cognitive aspects of interviews, one-at-a-time. 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. Further, this result, and the results of Bollinger and David (1999a), suggest that as attrition occurs, responses should improve. Individuals with high ε_{i1}(propensity to commit an error) are likely to have higher than average ε_{it}. Bollinger and David (1999a) find that response error in the first period correlates to interview nonresponse in later periods (significant at +0.4 from wave 1 to wave 2). Tentatively, researchers might conclude that some improvement in data quality is achieved by censoring data from households that fail to respond later in a panel.

Last, the remarkable stability of the omission error model across time periods confirms that the approaches suggested by Bollinger and David (1997, 1999a) are appropriate. Using predictions based on estimates from the cross-sectional model of data in period 1 are appropriate. The structure of response error does not change over time.

References

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. 1999a. 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.

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.

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.

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

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. Nonsampling 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.

7 Tables

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

Administrative record	SIPP Food Stamp Screener Response				
	No FS	Food Stamps in			Row
		Wave 1	Wave 2	Both	totals
No food stamps	2356	3	10	1	2370
Food stamps in					
Wave 1 only	6	19	0	3	28
Wave 2 only	4	0	17	1	22
Both Waves	16	11	12	138	177
Column Totals	2382	33	39	143	2597

Table 2: Descriptive Statistics for Food Stamp Participants

Variable	Definition	Mean	Std. Deviation
QFS1	Reported FS Participation, Wave 1	0.842	0.366
QFS2	Reported FS Participation, Wave 2	0.847	0.361
HHSize1	Persons in Household, Wave 1	3.311	1.834
HHSize2	Persons in Household, Wave 2	3.412	1.890
HHearn1	Household Monthly Earnings, Wave 1	421.06	833.02
HHearn2	Household Monthly Earnings, Wave 2	386.68	639.58
Percapearn1	Per Capita HH earnings, Wave 1	132.14	317.97
Percapearn2	Per Capita HH earnings, Wave 2	123.25	239.57
SingFem1	Single Female Head of HH, Wave 1	0.548	0.499
SingFem2	Single Female Head of HH, Wave 2	0.548	0.499
SingMale1	Single Male Head of HH, Wave 1	0.136	0.343
SingMale2	Single Male Head of HH, Wave 2	0.164	0.371
HHinc1	Household Monthly Income, Wave 1	731.24	619.85
HHinc2	Household Monthly Income, Wave 2	734.02	800.65
	Sample	177	

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

Variable	**Model 1**		**Model 2**	
	Wave 1	**Wave 2**	**Wave 1**	**Wave 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_{w1-w2}	(0.109)		(0.120)	
Likelihood	-0.599		-0.557	
N	177		177	