JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **20**: 563–569 (2005)
Published online 7 April 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.794

I DIDN'T TELL, AND I WON'T TELL: DYNAMIC RESPONSE ERROR IN THE SIPP

CHRISTOPHER R. BOLLINGER[a]* AND MARTIN H. DAVID[b]
[a] *Department of Economics, University of Kentucky, USA*
[b] *Joint Program for Survey Methodology, University of Maryland, USA*

SUMMARY

Using state administrative records matched to the *1984 Survey of Income and Program Participation*, we examine intertemporal relationships in response errors. False negative errors in reporting food stamps are highly correlated across interviews for the same household. Hypotheses that the error process can be explained by learning behaviour are not supported. Bivariate probit of response error in two periods reveals that responses to covariates are stable over time and the latent error terms are positively correlated. These findings support the hypothesis that respondents have a latent tendency to cooperate (or not cooperate) with surveys.

1. INTRODUCTION

Theoretical treatment of the impact of measurement error on estimation (Bound *et al*., 2002; Fuller, 1987) demonstrates that some information about the structure of response error is necessary for consistent estimation. Kalton *et al*. (1989, pp. 265–266) enumerate multiple sources of measurement error, while Griliches and Hausman (1986) describe the importance of information about the structure of the response error in panel data. Validation studies, such as this one, match survey data to some highly accurate measure of the variables of interest to provide information about response error structure. In this paper, administrative records of food stamp participation are matched to a subsample of the 1984 panel of the *Survey of Income and Program Participation* (SIPP). Other examples of this approach are Bound *et al*. (1990), Ferber *et al*. (1969a,b), Greenberg and Halsey (1983), Hill (1993), Lansing *et al*. (1961) and Mathiowetz and Duncan (1988). A few studies (Bollinger, 1998; Bound and Krueger, 1991; Marquis and Moore, 1990) measure multi-period response error.

This paper addresses two questions about the structure of response errors in a multi-interview panel. Is the response error structure stable across interviews? Response error structure may change due to learning. Respondents improve response accuracy through increased comprehension, willingness to access records, or building trust with the interviewer (often referred to as conditioning effects; Kalton *et al*., 1989). Alternatively, respondents may engage in strategic behaviour to reduce the length of the interview. The SIPP questionnaire asks a screening question about food stamp participation, and follows a 'yes' response with more detailed questions. Respondents may learn to give negative answers to avoid the detailed questions.

* Correspondence to: Christopher R. Bollinger, Department of Economics, University of Kentucky, Lexington, KY 40506, USA. E-mail: crboll@uky.edu
Contract/grant sponsor: National Science Foundation; Contract/grant numbers: SBR-9 511 935; SBR-952 174.

Received 24 March 2000
Revised 24 December 2003

Second, are response errors correlated over time? This second question extends Bollinger and David (1997, 2001), who find support for the *cooperator hypothesis*: respondents are predisposed to provide accurate answers or are predisposed to provide poor data. They find that response errors are concentrated in the screening question rather than on recall of details in specific months, proxy reports are not less accurate than self reports, and response errors are correlated with sample attrition. The cooperator hypothesis also suggests that response errors should be correlated across interviews.

The results fail to support any learning behaviour. The results provide strong support for positive correlation of errors across interviews, which is consistent with the cooperator hypothesis. Correlation implies that i.i.d. error structure is not applicable to modelling response error in panel data. Models that hypothesize i.i.d. structure are misspecified and estimation that relies upon that maintained assumption is likely to be biased. Section 2 describes the data. Section 3 presents the models and estimates. Section 4 concludes and provides direction for further research.

2. 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 (Bollinger and David, 1997 and Marquis and Moore, 1990 give details of that process). Administrative records for the Food Stamp Program are highly 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 (Marquis and Moore, 1990). Matching errors are reduced by validating self-reported Social Security numbers. The Social Security Administration confirms that reported Social Security numbers are consistent with other identifying information reported in the survey.

We argue that mismatching is rare and that differences between the administrative record and the survey response can be attributed to response error in the survey. If matching, or administrative record error, were large, one would expect errors of omission and errors of commission to be relatively symmetric. Both the probability and the count of errors of commission are far smaller than errors of omission. Errors of commission can be used as an upper bound on errors due to matching or administrative error: Table I shows errors of commission are less than 1% in each wave (0.3% in wave 1 and 0.6% in wave 2).

Data were aggregated to household units. As validation data are specific to states, any household that moved out-of-state between wave 1 and wave 2 was discarded from the sample, leaving 2597 households for the dynamic analysis. Change in residence of household members and change in the membership of the household pose a serious conceptual problem for longitudinal analysis (Citro and Hernandez, 1986). We resolve the problem by using the household units that exist at the time of the second interview as the unit of analysis. For most households, there are no changes between the first and second interview. Households that split are treated as two separate

Table I. Joint wave 1 and wave 2 screener level responses

Administrative record	No FS	Food stamps in			Row totals
		Wave 1	Wave 2	Both	
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					
Wave 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%)
Wave 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 waves	**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%)	

households throughout our analysis. We focus on households that received food stamps during both wave 1 and wave 2 of the 1984 SIPP. We focus on false negative survey responses as these households can exhibit learning behaviour. We cannot control for change in interviewer, which is relatively rare, and conjecture that interviewer effects are second-order to the incentive effects on respondents created by the administration of welfare programmes (Hotz and Scholz, 2002).

Table I compares household responses to the screening question and administrative records matched to the reference period. The totals in the second to fourth rows count food stamp indicators derived from administrative records. Our analysis pertains to the fourth row—177 households who participated in the Food Stamp Program during both periods. The remaining rows place the population studied in the context of a representative sample that does not move across state lines in the four months following the initial interview.

Table II provides descriptive statistics for the 177 participating households. The average household size shows little change between waves. Since the time interval is short (four months), stability is not surprising. Approximately 30% of the households were headed by a married couple. Single, female-headed households dominate the sample. Household earnings relate to the month prior to the interview month. Household earnings aggregate wages and salaries from all jobs, and

Table II. Descriptive statistics for food stamp participants

Variable	Wave 1		Wave 2	
	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			

net income from sole proprietorships, over all adults in the household. Income from other sources is excluded.

3. RESULTS

We use two approaches to address the questions posed. First we examine indicators of response error, without conditioning on covariates. If positive (negative) learning occurs, then the proportion of errors of omission should fall (rise) between the first and second wave. The cooperator hypothesis implies that errors of omission should be positively correlated between waves of interviews.

Bollinger and David (1997) establish that response errors are related to demographic and economic characteristics. Cross-tabulations do not control for these characteristics. To control for known covariates, models of response error conditional on economic and demographic characteristics are estimated over both periods. The null hypothesis for learning is that the coefficients of the models are constant across the two interviews. The cooperator hypothesis implies that temporal correlation exists for the unobserved latent variable in the model.

3.1. Indicators of Response Error

The indicator analysis focuses on the fourth row of Table I. Failing to report receiving food stamps in the first wave is the sum of the first and third columns (those who fail to report in both waves and those who only report in the second wave): 28/177. Those failing to report in only the second wave are the first and second column: 27/177. The difference is not statistically significant. There is no evidence of positive or negative learning behaviour.

The simplest measure of dependence in response error between the two periods is the correlation of the indicator variables for an error of omission across waves 1 and 2. The Pearson correlation coefficient is $r = 0.504$. This estimate is significantly positive at the 1% level. Other approaches, based on chi-squared tests of independence, yield similar results. The positive correlation supports our cooperator hypothesis.

3.2. Bivariate Probit Models

We generalize the model used for errors of omission in Bollinger and David (1997) by estimating a bivariate probit model for omission errors in two waves. The learning hypothesis is tested by comparing the estimated coefficients for each interview. The correlation between the latent variables ε_{i1} and ε_{i2} is used to test the cooperator hypothesis. When observations on a particular household are available for only two periods, it is impossible to differentiate between a random effects model and an autocorrelation model in the error terms, only the correlation can be estimated.

We estimate a bivariate probit model on the 177 households that received food stamps in both waves and participated in both wave 1 and wave 2 interviews. The model uses the covariates specified in the one-period analysis by Bollinger and David (1997). Estimates are presented in Table III. We also examine the robustness of these results to outliers in the income variable using truncated (and winsorized) samples. The results appear to be robust, not driven by a few outliers. (Additional one-period models were explored in Bollinger and David, 1997, 2001.)

Response error is positively related to earnings (either per capita or total), and negatively related to household size. Households headed by a single male are more likely to commit an error

Table III. Coefficients of joint omission error model: participants in both waves (standard errors in parentheses)

Variable	Wave 1	Wave 2
Intercept	−1.208	−0.760
	(0.425)	(0.427)
HH earnings	0.000 987	0.000 869
	(0.000 200)	(0.000 211)
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	0.727	
Correlation ρ_{w1-w2}	(0.120)	
Mean log-likelihood	−0.557	
N	177	

of omission than married couples. Households headed by a single female are less likely to commit an error than married couples, although this coefficient is not significant at conventional levels. Similar results were reported and examined in detail in Bollinger and David (1997, 2001). The new finding here is evidence about learning and persistence.

The learning model implies differences in the slope coefficients between the two periods. There is no evidence that the error-generating model is changing across waves. The asymptotic chi-squared test statistic for our model is 4.28, while the critical value (d.f. = 5) is 11.07. Tests on specific coefficients, including the intercept, produce similar results. Caution should be used in extrapolating this result to subsequent interviews, as attrition in waves 3 and 4 is correlated to wave 1 response errors (Bollinger and David, 2001).

The estimate of ρ supports persistence of response errors: $\rho_{w1-w2} = 0.727$. ρ_{w1-w2} is significantly positive and larger than the simple correlation calculated for Table I. The increase is due to the smaller variance in the unobservable ε_{it}. $\rho_{w1-w2} > 0$ supports our cooperator hypothesis: individuals in households who are willing and able to cooperate with the survey will provide more accurate answers throughout the survey. Other households are not cooperators: they are unwilling or unable to provide accurate data.

4. CONCLUSIONS

This paper has examined two important hypotheses concerning the structure of response error in reporting Food Stamp Program participation: response errors persist and learning effects are significant. Significant persistence in response error for particular households confirms our cooperator hypothesis, and supports previous results by Bollinger and David (2001). Further research that identifies root causes of persistence is needed. Similarity between coefficients of the two-interview model and the cross-sectional model demonstrates stability in the model specified for response error.

We find no evidence for learning, either positive or negative. While the sample is small and extends over a limited time period, this negative finding is good news for survey designers and for

data users. Survey designers can concentrate on overcoming cognitive problems and sensitivity that interferes with giving accurate responses. For data users, it sounds a hopeful note: holding other factors constant, response accuracy does not decline with time. Quality of estimates at the second wave of a panel appears to be about the same as the first. Research is needed to determine whether models of response error estimated here generalize to other variables, especially variables measured for individuals rather than households.

The most critical finding here is that the correlation of response errors over time is large and significant. The often-used assumption that measurement errors are uncorrelated over time is not supported here. The corollary to that assumption, that prior reports of a particular variable can be used as instruments when measurement error is present, is also not supported. Further work on dynamics of errors is needed to identify the possibility that persistence emanates from autocorrelation of response errors. Griliches and Hausman (1986) correctly suggest that the autocorrelation structure of response error coefficients be used to adjust model estimates. The work here provides one estimate of the size of that correlation. Further work on dynamics of errors is needed to identify the possibility that persistence emanates from autocorrelation of response errors.

These conclusions must be qualified. Administrative data may contain errors. The maintained assumption throughout the paper is that the administrative data are correct and any differences between survey response and administrative record represent response error. The small sample limits confidence in the negative finding of no learning effects. Sampling variance should be captured in the estimates of standard errors. A larger sample, or a different sample, may overturn these results. These limitations suggest that further research on modelling response errors is warranted. Continuing programmes of matching survey data to administrative records will be essential to that research.

ACKNOWLEDGEMENT

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, Mark Berger and an anonymous referee 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 thereof.

REFERENCES

Bollinger CR. 1998. Measurement error in the Current Population Survey: a nonparametric look. *Journal of Labor Economics* **16**: 576–594.

Bollinger CR, David MH. 1997. Modeling discrete choice with response error: food stamp participation. *Journal of the American Statistical Association* **92**: 827–835.

Bollinger CR, David MH. 2001. Estimation with response error and nonresponse: food stamp participation in the SIPP. *Journal of Business and Economic Statistics* **19**: 129–141.

Bound J, Krueger AB. 1991. The extent of measurement error in longitudinal earnings data: do two wrongs make a right? *Journal of Labor Economics* **9**: 1–24.

Bound J, Brown C, Duncan GJ, Rodgers WL. 1990. Measurement error in cross sectional and longitudinal labor market surveys: results from two validation studies. In *Panel Data and Labor Market Studies*, Hartog J, Ridder G, Theeuwes J (eds). Elsevier: New York.

Bound J, Brown C, Mathiowetz N. 2002. Measurement error in survey data. In *Handbook of Econometrics*, Vol. 5, Heckman JJ, Leamer E (eds). Elsevier: New York.

Citro CC, Hernandez D. 1986. Longitudinal household concepts in SIPP. In *Proceedings of the Social Statistics Section of the American Statistical Association*; 361–365.
Ferber R, Forsythe J, Guthrie HW, Maynes ES. 1969a. Validation of a national survey of consumer financial characteristics: savings accounts. *Review of Economics and Statistics* **51**: 436–444.
Ferber R, Forsythe J, Guthrie HW, Maynes ES. 1969b. Validation of a national survey of consumer financial characteristics: common stocks. *Journal of the American Statistical Association* **64**: 415–432.
Fuller WA. 1987. *Measurement Error Models*. John Wiley & Sons: New York.
Greenberg D, Halsey H. 1983. Systematic misreporting and effects of income maintenance experiments on work effort: evidence from the Seattle–Denver experiment. *Journal of Labor Economics* **1**: 380–407.
Griliches Z, Hausman J. 1986. Errors in variables in panel data. *Journal of Econometrics* **31**: 93–118.
Hill DH. 1993. Response and sequencing errors in surveys: a discrete contagious regression analysis. *Journal of the American Statistical Association* **88**: 775–781.
Hotz VJ, Scholz JK. 2002. Measuring employment and income outcomes for low-income populations with administrative and survey data. In *Studies of Welfare Populations: Data Collection and Research Issues*. National Research Council/National Academy Press; 275–315.
Kalton G, Kasprzyk D, McMillen D. 1989. Nonsampling errors in panel surveys. In *Panel Surveys*, Kasprzyk D, Duncan G, Kalton G, Singh MP (eds). John Wiley & Sons: New York.
Lansing JB, Ginsberg GP, Braaten K. 1961. *An Investigation of Response Error*. Bureau of Economic and Business Research, University of Illinois.
Marquis K, Moore J. 1990. Measurement errors in the SIPP program reports. In *Proceedings of the Bureau of the Census Annual Research Conference*, Bureau of the Census, Washington, DC.
Mathiowetz NA, Duncan GJ. 1988. Out of work, out of mind: response errors in retrospective reports of unemployment. *Journal of Business and Economic Statistics* **6**: 221–229.

