Estimating the Level of Underreporting of Expenditures among Expenditure Reporters: A Further Micro-Level Latent Class Analysis

(August 2005)

Clyde Tucker, Bureau of Labor Statistics
Paul Biemer, Research Triangle Institute
Brian Meekins, Bureau of Labor Statistics

1. Introduction

This paper makes estimates of the level of underreporting of consumer expenditures. The paper examines reporting in particular commodity categories and attempts only to make estimates of underreporting among those that report at least one expenditure in the category. The measure of the level of underreporting in a category by a particular responding unit is based on latent class analysis using characteristics of the respondent's reporting behavior. It is assumed that the level of underreporting is similar within a particular subpopulation defined by these characteristics. A demographic analysis will identify the types of responding units that have various levels of underreporting.

Latent class analysis, a theory for detecting unobserved variables, was developed by Paul Lazarsfeld (1950). According to Lazarsfeld, an unobservable variable (such as underreporting) could be constructed by taking into account the interrelationships between observed variables. The mathematics underlying this theory were extended by Lazarsfeld and Henry (1968) and Goodman (1974).

The paper begins with an introduction to the Consumer Expenditure Interview Survey (CEIS) sponsored by the Bureau of Labor Statistics (BLS) and conducted by the Census Bureau. Previous related work by the authors in this area is summarized and the design of this particular study is outlined. The following section presents the analytical results, and a final section is devoted to the discussion of the results and the consideration of additional avenues for research.

2. CEIS

The data used in this study consists of interviews collected in six years of the CEIS: 1996 through 2002. Each survey was designed to collect information on up to 95 percent of total household expenditures. We define a consumer unit (CU) as the members of a household who are related and/or pool their incomes to make joint expenditure decisions. In the CEIS, CU's are interviewed once every three months for five consecutive quarters to obtain the expenditures for 12 consecutive months. The initial interview for a CU is used as a bounding interview and these data are not used in the estimation. The survey is designed to collect data on major items of expense which respondents can be expected to recall for three months or longer. New panels are initiated every quarter of the year so that each quarter, 20 percent of the CU's are being interviewed for the first time. Only CU's completing and reporting an expense in wave 2 are used in this analysis, for a total of 14,877 respondents.

3. Previous Work

For panel surveys such as the CEIS, a related statistical method referred to as Markov latent class analysis (MLCA) is available, which essentially relaxes the requirement that the replicate measurements pertain to the same point. Thus, this method of analysis is feasible for analyzing repeated measurements of the same units at different time points available in panel surveys. MLCA requires a minimum of three measurements of the same units, as would be the case for a panel survey where units are interviewed on three occasions. The MLCA model then specifies parameters for both the period-to-period changes in the status of the item as well as the measurement error associated with measuring those changes.

Previous work by the authors used MLCA to make aggregate estimates of underreporting in a category only by respondents reporting no expenditures in that category. Biemer (2000) applied the MLCA methodology to the CEIS in order to determine whether useful information on the magnitudes and correlates of screening question reporting error can be extracted directly from the CEIS panel data. Biemer and Tucker (2001) extended the earlier analysis using data from four consecutive quarters of the CEIS by considering CU's that were interviewed four consecutive times beginning in the first quarter of 1996 and ending in the last quarter of 1998. This allowed the authors to consider a wider-range of models including second-order Markov models. First order Markov models assume that a purchase or non-purchase at quarter q is affected only by quarter q-1 purchases or non-purchases. A second order Markov model assumes that both quarters q-1 and q-2 affect purchasing behavior at quarter q. Their analysis provided evidence of second-order Markov effects and recommended that second-order terms be included in the models.

In Tucker, Biemer, and Vermunt (2002), model estimates with both unweighted and weighted data were compared. The results indicated that few differences were found between the two; therefore, given the ease of use, unweighted data were used in these analyses. A thorough examination of all explanatory variables considered in the previous studies was undertaken, and a reduced set of the most powerful ones was identified. A new diagnostic technique was developed and used to evaluate the validity of the models. In 2003, Tucker, Biemer, and Meekins developed methodology for estimating the amount of the missing expenditures.

Unlike the previous work, a micro-level approach incorporating measures specific to a given interview was used by Tucker, Biemer, Meekins, and Shields (2004) to examine underreporting for for total expenditures across approximately 20 commodity categories. In essence, a latent variable that adequately accounted for the shared varaince among a set of observed response error indicators was created. The observed varaibles were based on information collected from each CU during the interview. The latent variable was believed to be a better measure of underreporting than any of the observed varaibles taken individually. Each CU then was assigned to a particular class of the latent variable representing its hypothesized level of expendituire underreporting based on the CU's values on the observed variables. See Tucker (1992) for an earlier empirical example.

We used only second wave data[1]. We examined reporters of expenditures and ignored nonreporters. We wished to develop a model separate from covariates with only indicators of the quality of response. We began with the simplest identifiable model composed of three indicators (each with three classes) and a latent variable with three classes. From this point we ran all possible combinations of three indicators for a three class latent variable. The analysis was further extended by examining restricted models based on the hypothetical relationship of some of the indicators with the latent variable, thus ordering the latent classes in what we believed to be an interpretable manner. These "restricted" models were compared to the unrestricted models to aid in interpretability and choices of model fit. Some of the indicators are dichotomous. These were entered into the best three variable models along with other combinations to create four-indicator models. Our goal was to develop a latent variable (preferably ordered) that indicated the quality of responses, such that poor reporters could be easily identified.

Models were estimated using ℓEM. Model selection was based on a number of objective and subjective measures. We primarily used the Bayesian Information Criteria (BIC), the L^2 test statistic, and the dissimilarity index. However, for each model we examined the conditional probabilities of the latent variable given each value of each indicator. In this way we assessed the relative influence of each indicator and the degree to which an indicator effectively differentiated the respondents with respect to the classes of the latent variable

Using these methods a "best" model was selected. Latent classes aligned with expenditure means as expected. Those with lower expenditure means had higher levels of underreporting. For example, those in the low underreporting class had a total expenditure mean of $10,625, while those in the high underreporting class had a mean of $6,948

4. Current Design

In this paper, we continue with a more in-depth exploration of micro-level measures of underreporting. Rather than considering just total expenditures, separate categories of expenditures are examined. A more refined set of indicators of underrepoting are utilized, and they are the following:

1. Number of contacts the interviewer made to complete the interview
2. The ratio of respondents to total number of household members
3. The ratio of household members earning an income to the total number of household members

[1] Wave 2 data are used because wave 1 is a bounding interview.

4. Whether the household was missing a response on the income question
5. The type and frequency of records used. This variable indicates whether a respondent used bills or their checkbook to answer questions, and how often they did so.
6. The percent of data requiring imputation or allocation.
7. The length of the interview
8. A ratio of expenditures reported for the last month of the 3 month reporting period to the total expenditures for the 3 months
9. And a combination of type of record used and the length of the interview.

5. Results

The LCM table gives the model fits for the commodity groups examined. All commodities had acceptable fits based on the statistical diagnostics. The dissimilarity indexes were less than .05 in all cases. The BIC was negative in each case, and the absolute size of the BIC depended on the degrees of freedom in each model –the larger the degrees of freedom, the greater the BIC.

For each commodity, cases were assigned values based on an appraisal of the loadings for each cell on three latent classes. These classes were labelled Poor Reporting (1), Fair Reporting, (2), and Good Reporting (3) based on theoretical expectations concerning the individual indicators in the models. The most influential indicators were number of contacts, missing income, length of the interview, and use of records.

The substantive diagnostic chosen for evaluating the models was the mean expenditure of the commodity for each laent class. The expenditure means table indicates that, for Kid's Clothing and Minor Vehicle Repairs, the expenditures are in the expected direction—increasing from Poor to Good. For Women's Clothing, Men's Clothing, and Kitchen Accessories, the Poor category, indeed, has the lowest mean; but the other two categories are indistinguishable. It is possible that a two-class model is more appropriate in these cases. The model for Furniture does not follow the expected pattern. Other models should be examined in this case. The means for Electricity also do not have the expected pattern, but earlier results indicated that electricity is reported very well by most respondents.

The proportional odds model for Minor Vehicle Repairs presents an example of the factors most important for determining respondents' scores on the latent variable. Odds ratios geater than 1.0 indicate a tendency toward better reporting, and those less than 1.0 indicate the opposite tendency. Poorer reporting is most common for young people, single-person households, those with less income,and renters.

6. Discussion

These results provide some validation for the latent class approach to modeling measurement error. Other commodity categories as well as total expenditure will be examined. Adding a Markov component that takes advantage of expenditure patterns over the four quarters might lead to improvements in the models. Eventually, estimates from both nonreporters and underreporters will be combined to produce overall estimates of underreporting by commodity category and respondent characteristics.

7. References

Bassi, Franesca, Jacques A. Hagenaars, Marcel A. Croon and Jeroen Vermunt. (2000). "Estimating True Changes When Categorical Panel Data Are Affected by Uncorrelated and Correlated Classification Errors: An Application to Unemploymnet Data." *Sociological Methods & Research* 29: 230-268.

Biemer, P.P. (2000). "An Application of Markov Latent Class Analysis for Evaluating the Screening Questions in the CE Interview Survey," Technical Report Submitted to BLS, June 6, 2000.

Biemer, P. P. and Tucker, C. (2001). "Estimation and Correction for Purchase Screening Errors in Expenditure Data: A Markov Latent Class Modeling Approach," *Proceedings of the International Statistics Institute*, Seoul, Korea.

Tucker, C. (1992). "The Estimation of Instrument Effects on Data Quality in the Consumer Expenditure Diary Survey." *Journal of Official Statistics.* 8: 41-61.

Tucker, C., Biemer, P., and Vermunt, J. (2002). "Estimation Error in Reports of Consumer Expenditures," Proceedings of the ASA,

Survey Research Methods Section, New York, NY.
Goodman, L. A. (1974), "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models," Biometrika, 61, 215-231.
Langeheine, R. and Van der Pol, F. (2002). “Latent Markov Chains,” in Hagenaars, J. and McCutcheon, A. (eds.) *Applied Latent Class Analysis*, Cambridge University Press, Cambridge, UK
Lazarsfeld, P.F. (1950). “The Logical and Mathematical Foundation of Latent Structure Analysis.” In S. Stauffer, E.A. Suchman, P.F. Lazarsfeld, S.A. Starr, and J. Clausen, Studies on Social Psychology in World War II, Vol. 4, Measurement and Prediction. Princeton: Princeton University Press.
Lazarsfeld, P.F. and Henry, N.W. (1968). Latent Structure Analysis. Boston: Houghton-Mifflin.
Tucker, C., Biemer, P., and Meekins, B. (2003),“Latent Class Modeling and Estimation Errors in Consumer Expenditure Reports,” *Proceedings of the Section on Survey Research Methods*, Washington, DC: American Statistical Association.
Tucker, C., Biemer, P., Meekins, B., and Shields, J. (2004), “Estimating the Level of Underreporting of Expenditures among Expenditure Reporters: A Micro-Level Latent Class Analysis,”*Proceedings of the Section on Survey Research Methods*, Washington, DC: American Statistical Association.
Van de Pol, F. and de Leeuw, J. (1986). “A Latent Markov Model to Correct for Measurement Error,” Sociological Methods & Research, Vol. 15, Nos. 1-2, pp 118-141.
Van de Pol, F. and Langeheine, R. (1997). “Separating Change and Measurement Error in Panel Surveys with an Application to Labor Market Data,” in L. Lyberg, et al (eds.) Survey Measurement and Process Quality, John Wiley & Sons, NY.
Vermunt, J. (1997). lEM: A General Program for the Analysis of Categorical Data, Tilburg, University

Appendix

Indicator Coding

- #contacts (1=0-2; 2=3-5; 3=6+)
- Resp/hh size (1= <.5; 2= .5+)
- Income missing (1=present; 2=missing)
- Records use (1=never; 2=single type or sometimes; 3=multiple types and always)
- Interview length (1= <45; 2=45-90; 3= 90+)
- Month3 expn/all (1= <.25; 2= .25-.5; 3= +.5)
- Combined records and length (1= poor; 2= fair; 3=good)

Demographic Coding

- CU size (1=1; 2=2; 3=3+)
- Age (1= 30<; 2= 30-49; 3=50+)
- Education (1=< H.S.; 2= H.S.+)
- Income rank (1= <=.25; 2=.25-.75 and missing; 3=+.75)
- Race (1= White; 2= Other)
- Tenure (1= renter; 2= owner)
- Urban (1= urban; 2= rural)

LCM Fit by Commodity

	BIC	Diss Index
Kid’s Clothing	-7.4029	0.0040
Women’s Clothing	-23.1575	0.0221
Men’s Clothing	-244.0034	0.0258
Furniture	-239.6923	0.0327
Electricity	-8.6450	0.0021
Minor Vehicle Repairs	-221.5239	0.0306
Kitchen Accessories	-119.7008	0.0288

Expenditure Means by Latent Class

		Value of Latent Class		
		1 = Poor	2	3 = Good
Kid's Clothing	Mean	44.90(a)	59.62(b)	71.09(c)
	n	24,666	12,001	6,331
Women's Clothing	Mean	99.00(b)	148.08(a)	152.94(a)
	n	21,316	11,281	10,401
Men's Clothing	Mean	78.98(b)	107.04(a)	105.46(a)
	n	36,080	842	6,076
Furniture	Mean	117.25(a)	66.22(b)	266.63(c)
	n	23,437	16,315	3,246
Electricity	Mean	230.47(a)	198.87(b)	223.30(c)
	n	32,905	4,377	5,716
Minor Vehicle Expenditures	Mean	39.47(a)	57.03(b)	82.28(c)
	n	9,864	26,288	6,846
Kitchen Accessories	Mean	23.27(b)	52.51(a)	47.58(a)
	n	26,589	2,934	13,.475

Proportional Odds Model Results for Minor Vehicle

	Exp(b)	PR(X^2)
Famsize 1	.801	<.0001
Famsize 2	1.063	<.0001
Age 1	.862	<.0001
Age 2	1.033	.0293
Educ	.984	.1492
Inclass 1	.776	<.0001
Inclass 2	.762	<.0001
Race	1.056	<.0001
Tenure	.909	<.0001
urban	1.104	<.0001
Max-rescaled R^2		.0566