A Microlevel Latent Class Model for Measurement Error in the Consumer Expenditure Interview Survey

Clyde Tucker[1] Brian Meekins[1] Paul Biemer[2]
1 Bureau of Labor Statistics, 2 Massachusetts Ave, NE, Washington DC 20212
2 RTI International, PO Box 12194, Research Triangle Park, NC 27709

Abstract
Previous research by Tucker et al. (2005) and Tucker et al. (2006) attempts to identify a latent construct that predicts the amount of measurement error in expenditure reports on the Consumer Expenditure Survey (CEIS). While this work was successful in identifying a construct that predicts measurement error in expenditure reports, it is more sensitive to falsely negative reports of the entire purchase than it is to the underreporting of the amount of expenditure for that purchase. Current research focuses more deeply on the underreporting of expenditure amounts for a number of different commodities. Together, with previously explored indicators such as, the number of contacts, missing on the income question, the length of the interview, and the use of records, we examine new indicators describing the experience of the CU throughout the panel, resulting in a new latent construct. Although we find that our newly developed latent constructs have strong validity, appearing to measure report quality, it does little to explain either the overall level of expenditure or the difference in expenditure reports between interviews.

Key Words: latent class model, measurement error, consumer expenditure

1. Introduction

This work is part of a continuing effort to identify sources of measurement error in the Conusmer Expenditure Interview Survey (CEIS), a household survey of expenditure reports of a variety of different commodity categories (e.g. furniture, clothing, utilities, etc.). Previous efforts have used Markov Latent Class Models to analyze patterns of item missing (where respondents do not report an expenditure in that category), and latent class models to identify characteristics of poor reporting consumer units (CUs). This work extends the work on the latter type of models by adding information pertaining to the experience of the consumer unit throughout all five interviews of the panel, developing a measure of the quality of expenditure reports with strong validity.

2. Consumer Expenditure Survey

The data used in this study consist 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 CU expenditures. We define a 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 total expenditures. A latent variable that adequately accounted for the shared variance among a set of observed response error indicators was created. The observed variables 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 variables taken individually. Each CU then was assigned to a particular class of the latent variable representing its hypothesized level of expenditure underreporting based on the CU's values on the observed variables. See Tucker (1992) for an earlier empirical example.

For this analysis the authors used only second interview data and examined reporters of expenditures while ignoring nonreporters. They wished to develop a model separate from covariates with only indicators of the quality of response. The authors began with the simplest identifiable model composed of three indicators (each with three classes) and a latent variable with three classes. From this point they 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. The 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, LCA software developed by Vermunt (1997). Model selection was based on a number of objective and subjective measures. The authors primarily used the Bayesian Information Criteria (BIC), the L^2 test statistic, and the dissimilarity index. However, for each model the authors 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

In Tucker, Biemer, and Meekins (2005), the authors continued with a more in-depth exploration of micro-level measures of underreporting. In this analysis, only second wave data are used from those respondents actually reporting expenditures in the commodity classes under study (57,184 families interviewed in 1996 through 2001). Thus, we first were interested in the response errors for those respondents reporting expenditures and not those who said they had no expenditures in these categories. Again, the authors assumed response errors come largely in the form of underreports.

In this case, a refined set of manifest indicators of response error were created. These indicators are listed below, with the coding scheme used for each:

1. Number of contacts the interviewer made to complete the interview (1=0-2; 2=3-5; 3=6+)
2. The ratio of respondents to total number of household members (1=<.5; 2>.5)
3. Whether the household was missing a response on the income question (1=present; 2=missing)
4. 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. (1=never; 2=single type or sometimes; 3=multiple types or always)
5. The length of the interview (1<45min; 2=45-90; 3>90)
6. A ratio of expenditures reported for the last month of the 3 month reporting period to the total expenditures for the 3 months (1<.25; 2=.25-.5; 3=>.5)
7. A combination of type of record used and the length of the interview. (1=poor; 2=fair; 3=good) as shown below for the combined variable.
8. Number of expenditure questions within commodity category for which a response was imputed or allocated.

For each of seven expenditure categories: children's clothing, women's clothing, men's clothing, furniture, electricity, minor vehicle expenses, and kitchen accessories, we began with the simplest identifiable model composed of three indicators and a latent variable with three classes. Models were again estimated using lEM. Only three manifest variables were used to maximize cell sizes in the manifest tables. We ran all possible combinations of three indicators for each expenditure class. The analysis involved both "restricted" and "unrestricted" models. Restricted models forced a hypothesized ordering of the manifest indicators to the latent response error (ordering the latent classes in what we believed to be an interpretable manner), while unrestricted models did not. Based on comparisons of the results from restricted and unrestricted models, it was decided to proceed with only restricted models from that point. Combinations of four and five manifest indicators were examined, but all models with more than four variables were of little value. Again, we ran models with several different sets of starting values to avoid reaching only a local solution.
.
The selection of the best model for each expenditure category was based primarily on the BIC and the Dissimilarity Index. The same set of manifest indicators were not used for the best model in each case, but the statistical diagnostics confirm a good fit for all final models chosen.

The authors also extended the use of substantive diagnostics used in earlier work. For each model they examined both conditional probabilities of the latent variable given each value of each indicator and the conditional probabilities of each indicator given each value of the latent variable. In addition, they also examined the actual probabilities of a case being in a particular latent class given its manifest cell location, as well as the proportion of cases assigned to each manifest cell by the latent class model.

To gain a further understanding of the models, the authors again turned to the expenditure means for the three latent classes. The results, while not completely disconfirming, were not that promising. Across all seven categories of expenditures we analyzed, we found that the three classes of the latent variable failed to distinguish CUs based on their expenditures. However, for kid's clothing, women's clothing, and kitchen accessories, two separate groups could be identified that met our expectations.

By including CUs that reported no expenditures in our analysis, the authors found that, for most commodities, mean expenditure amounts increased monotonically across the latent scale, and the three means were significantly different from one another.

Research by Tucker, Biemer, Meekins, Kesselman (2006) advanced the effort by examining a much larger number of commodity categories (29) and more rigorously examining and validating the results of the latent class models. The "final" model for each of the 29 commodities and overall were selected in a similar manner to past research, using both objective statistics and subjective diagnostic tools. We then assigned CUs certain latent classes in the same way as previous research. Using 30 proportional odds models (each commodity category and overall) we regressed the latent variable on a number of demographics in order to assess the content validity of the latent variable. After finding similar patterns across all commodity categories and verifying the results of the latent class modeling, we examined the mean expenditure amounts using the commodity specific LCA models for expenditure reporters and the overall LCA model for all unit responders. Finally, we examined the expenditure means for each class of the latent variable controlling for key demographics using MANCOVAs and the contribution of the latent variable in predicting expenditure, controlling for demographic variables (such as you would use in weighting or nonresponse adjustment). Consistent with previous research the results of this research provided validation for the latent class approach to modeling measurement error, but a model that could differentiate levels of underreporting (given a report) remained elusive, while models classifying CU's by whether they erroneously omit a report altogether were more successful.

Other research by Meekins, Tucker, and Biemer (2008) used the latent construct developed in Tucker et al. (2006) to examine the relationship of measurement error to subsequent wave nonresponse and bias. It was found that those in the poorest category of reporting were somewhat less likely to respond in subsequent interviews, volunteered expenditure reports in fewer categories, and had more sharply declining overall expenditure amounts in subsequent interviews than their counterparts in the fair and good reporting categories.

4. Current Research

The current research advances this work in a number of ways. Perhaps most importantly the current work includes a number of new indicators that address the experience of the CU within the entire panel. In addition to the indicators used in previous research, the models estimated for this work include:

1. Number of completed interviews.
2. Pattern of attrition combined with the number of completed interviews (those with a pattern of attrition as opposed to a sporadic nonresponse pattern were further penalized).
3. Average number of commodity categories for which CU had at least one expenditure report.
4. The number of interviews in which the third months expenditure to the quarter was between 0.25 and 0.5.
5. Panel averages of some of the interview level indicators.

Model selection was conducted in a manner similar to the previous research where all combinations of three or four indicators were estimated and three and four level latent class variables were considered. Multiple iterations with random start values were used in order to avoid local maxima. Models were initially selected based on fit, reducing the many possible combinations of classes and indicators to a relative few candidates. The remaining candidates were then evaluated based on the relationship of the latent construct to the indicators and other subjective criteria. After a few models were selected the CU was assigned a latent class value based on the probability of being in that class given the indicators in the model. Expenditure means were found for each latent class assignment. The models were further validated by regressing demographic variables on the latent class assignment using proportional odds models. Expenditure means were regressed on the latent class assignment together with demographic variables. Finally, the authors developed a measure of "bias" calculated as follows:

$$Bias' = \begin{cases} 0; X_2 \le X_5 \\ X_2 - X_5; X_2 > X_5 \end{cases}$$

This measure was then regressed in a similar manner to average expenditure on the latent class assignment and demographic variables.

5. Results

Based on the aforementioned procedure three models were chosen to be verified and further analyzed. The fit statistics and overall performance was acceptable for all three of these models. Two models with three indicators were chosen and one model with four indicators was chosen. The three indicator models used the indicators: missing on income, length of interview, average number of commodity categories to differentiate three and four latent classes of reporting quality (lv2b47_3 and lv2b47_4, respectively). The four indicator model (lv2479_3) used all of those used in the three indicator models combined with the number of good interviews in the panel. Fit statistics were very good for all three models. The ordinal constraints did not significantly reduce fit of any of these three models. Table 1 shows the average overall expenditure per quarter by latent class and model. Note that all of the models show significant differences in mean expenditure across the three latent classes.

Table 1: Average Overall Expenditure by Latent Class

	1	**2**	**3**	**4**
lv2479_3	6,946.84	8,920.20	11,985.71	
lv2b47_3	5,790.85	9,045.21	11,535.85	
lv2b47_4	5,361.99	7,802.64	8,975.91	11,535.85

Subsequent analysis regressing the latent construct on a number of demographic variables shows that the relationship between these demographic variables and the latent constructs are consistent and in the expected direction. The demographic variables explain from 17 to 15 percent of the variance in the latent classes, indicating that there is some independence between them.

Table 2: Proportional Odds Model Results for Overall Expenditure

Latent Variable	**2b47_3**		**2b47_4**		**2479_3**	
	Exp(b)	PR(X2)	Exp(b)	PR(X2)	Exp(b)	PR(X2)
Famsize 1	0.763	<.0001	0.805	<.0001	0.756	<.0001
Famsize 2	1.120	<.0001	1.108	<.0001	1.071	<.0001
Age	1.018	<.0001	1.020	<.0001	1.011	<.0001
Educ	0.895	<.0001	0.917	<.0001	0.901	<.0001
ln(incrank/(1-incrank)	1.248	<.0001	1.220	<.0001	1.262	<.0001
Race	1.252	<.0001	1.222	<.0001	1.246	<.0001
Tenure	0.705	<.0001	0.746	<.0001	0.767	<.0001
Urban	1.111	<.0001	1.090	<.0001	1.138	<.0001
Max-rescaled R2		0.17		0.15		0.15

Because the latent variable with four indicators proved to be better at differentiating between the quality of reporting, the remaining analysis is shown with the four indicator model. As is shown in Table 3, the mean expenditure amount for all commodity categories apart from “other vehicle expense” and “drugs“ is significantly different across the latent classes.

Table 3: Mean Commodity Expenditure by Latent Class (lv2479_3)

Latent Class	**1**	**2**	**3**
Electricity	214.42	239.80	256.15
Cable	68.71	80.94	90.53
Trash	12.93	18.24	24.47
Gas	68.20	82.14	101.78
TV	55.46	91.69	152.07
Sports	15.25	27.52	51.26
Furniture	60.18	91.53	148.70
Other Household	17.32	29.74	59.80
Kitchen Accessories	15.86	28.58	59.71
Men’s Apparel	59.63	79.34	109.92
Women’s Apparel	85.48	118.95	180.62
Men’s Clothing Only	49.56	65.62	89.03
Women’s Clothing Only	73.10	100.46	151.16
Men’s Accessories	1.71	2.39	4.00
Women’s Accessories	2.32	3.62	6.44
Men’s Shoes	8.37	11.33	16.89

Latent Class	**1**	**2**	**3**
Women's Shoes	10.07	14.87	23.02
Kid's Apparel	36.49	52.15	75.38
Kid's Clothing Only	30.57	43.30	62.18
Kid's Accessories	0.47	0.79	1.52
Kid's Shoes	5.45	8.06	11.68
Minor Vehicle Repairs	37.43	53.11	81.79
Vehicle Oil	11.87	16.23	22.81
Major Vehicle Repairs	39.75	57.64	89.45
Vehicle Other*	8.98	11.91	22.25
Dental	32.38	59.50	95.10
Eyes	14.37	22.81	39.04
Drugs*	167.55	191.34	190.12
Pets	62.69	81.54	92.33

Table 4 shows the results of the average expenditure per CU per wave regressed on the latent construct controlling for demographics. The marginal gain from introducing the latent construct is small but statistically significant. The effect sizes of the demographics are not significantly diminished and the means of the expenditures by latent class controlling for the demographics variables in the model are significantly different and in the expected direction. These results are similar for each commodity category (not shown), although those commodities with higher probability of measurement error show more differentiation across latent classes.

Table 4: MANCOVA Results: Total Expenditures

	Baseline model		With LV	
	Estimate	p-value	Estimate	p-value
F	2740.45	<.0001	2096.07	<.0001
R^2	.39		.41	
F[Contribution of Latent variable]:			21.21	<.0001
F[Contribution of Interaction Term]:			519.11	<.0001

Total Expenditure: Least Squared Means Controlling for All Other Variables in the Model*

		p-values for differences in LSMean		
Class	Mean	Poor	Fair	Good
Poor	6,905.96		<.01	<.01
Fair	7,293.90	<.01		<.01
Good	7,556.87	<.01	<.01	

*Tukey-Kramer adjustment for multiple comparisons

Of particular note is the contribution of the interaction between income level and the latent class variable. As shown in Figure 1, at very high or very low incomes the relationship between latent class and reported income is much stronger. Indeed, the contribution of the interaction term is much higher than the direct effect of the latent construct.

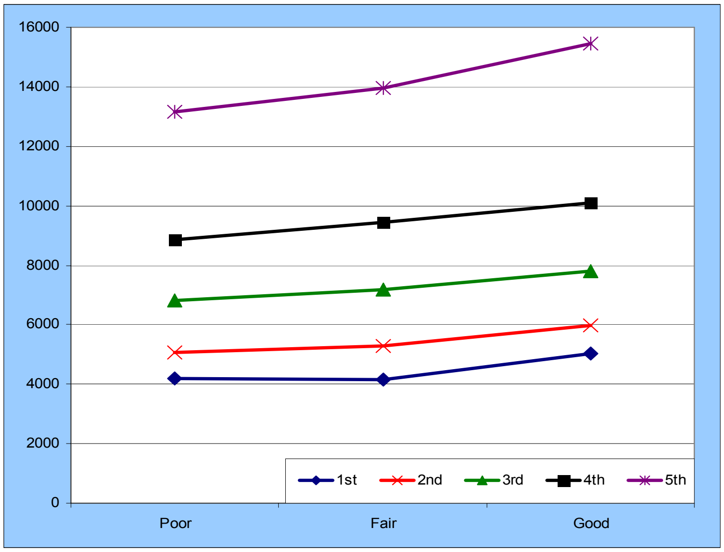

Figure 1: Latent Class by Average Expenditure by Income Rank Percentile

Turning to our measure of "bias" we see similar results. Table 5 shows summarizes the outcome of regressing the "bias" measure on the latent construct controlling for demographic variables. Note again the strong interaction term, but also how little of the variation in bias our model explains. However, the latent construct seems to differentiate between the poorest reporters and the other two classes.

Table 5: MANCOVA Results: Bias

	Baseline model		With LV	
	Estimate	p-value	Estimate	p-value
F	89.2	<.0001	67.8	<.0001
R^2	.02		.03	
F[Contribution of Latent variable]:			12.51	<.0001
F[Contribution of Interaction Term]:			13.38	<.0001

Total Expenditure: Least Squared Means Controlling for All Other Variables in the Model*

		p-values for differences in LSMean		
Class	Mean	Poor	Fair	Good
Poor	1,300.67		<.01	<.01
Fair	1,567.46	<.01		.07
Good	1,730.65	<.01	.07	

*Tukey-Kramer adjustment for multiple comparisons

6. Discussion

Based on the analysis we are confident that the latent construct is indeed measuring the quality of reporting, but either this latent construct lacks the sensitivity needed to adequately predict underreporting, or measurement error is not a strong predictor of the average expenditure reported by the CU and the amount of bias, as measured by our admittedly coarse measure. We show differences in expenditure amounts in the expected direction. For commodity categories where previous research has indicated that there is a great deal of measurement error, the model seems to explain more of the variation in expenditure amount and bias. Finally, incorporating the latent construct in our models appears to enhance and clarify the effect of income on reported expenditure. All of these serve to validate the construct.

However, although laying the groundwork for future measures, the current latent construct is probably not useful for adjustment as it does not explain much of the variation in expenditure or "bias."

Future research will attempt to develop indicators that may be able to further differentiate reporting quality among expenditure reporters. In addition, the authors intend to add a Markov component to this micro analysis, using both time varying and time invariant covariates to assist in predicting levels of measurement error. The goal is to eventually produce overall estimates of underreporting by commodity category and respondent characteristics.

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 Meekins, B. (2005). "Estimating the Level of Underreporting of Expenditures among Expenditure Reporters: A Further Micro-Level Latent Class Analysis." Proceedings of the ASA, Survey Research Methods Section, Toronto, CA.

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.

Meekins, B. and Tucker, C, and Biemer, P. (2008), "Measurement Error and Subsequent Panel Nonresponse in the Consumer Expenditure Survey." Paper presented at the 2008 International Total Survey Error Workshop, Research Triangle Park, NC.

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.

Tucker, C., Biemer, P., Meekins, B., Kesselman V. (2006), "A Micro-Level Latent Class Analysis of Underreporting on the Consumer Expenditure Survey,"*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