Latent Class Analysis of Consumer Expenditure Reports

Clyde Tucker[1] Brian Meekins[1] Paul Biemer[2] October 2010
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. (2007), working with the Consumer Expenditure Interview Survey (CE), explores the efficacy of measurement error indicators such as: interview length, extent and type of records used, the monthly patterns of reporting, and income question missing in a latent construct. Later research by Tucker, Meekins, and Biemer (2008) extend this latent class model to include indicators of response behavior across multiple interviews in a panel. This research develops a number of plausible models which possess the qualities of reliability and validity, in that they appear to accurately capture measurement error, but prove unable to explain a large amount of the variance associated with expenditure reports. This work extends past research by including the relatively recently recorded indicators from the Contact History Instrument added to the CE since 2005. In addition, this work examines measurement error by mode of collection (also a recently collected item).

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 Consumer 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: 2005 through 2009. 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, CUs 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 CUs are being interviewed for the first time. Only CUs completing and reporting an expense in wave 2 are used in this analysis, for a total of 29,347 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 CUs 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 CUs 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. Based on the results of the models CUs were then assigned to certain classes of reporting in the same way as previous research. The classification variable, corresponding to proo, fair, and good reporting quality was then regressed 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, the authors regressed the expenditure mean for each commodity category and overall on the latent classification controlling for key demographics, examining the contribution of the latent variable in predicting expenditure, controlling for demographic variables (such as one 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. In other research (Tucker, Biemer, Meekins 2009), included indicators that characterized the experience of the CU throughout the course of the entire panel. The new indicators included:

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 month's expenditure to the quarter was between 0.25 and 0.5.
5. Panel averages of some of the interview level indicators.

The three best fitting models selected from this research incorporated indicators found to be important in previous work and new panel indicators. Two models utilized three indicators, while one model used four indicators to examine the quality of expenditure reports. The three indicator models used the indicators: missing on income, length of interview, and average number of commodity categories to differentiate three and four latent classes of reporting quality. The four indicator model used all of those used in the three indicator models combined with the number of good interviews in the panel. Following the same process as prior research, these models were validated with demographic and process variables. The models showed good differentiation of expenditure estimates, even when controlling for demographics and process variables. Of particular note is the contribution of the interaction between income level and the latent class variable to this model. At very high or very low incomes the relationship between level of reporting and reported expenditure is significantly stronger. Indeed, the contribution of the interaction term is much higher than the direct effect of the latent construct. When examining bias (second quarter reported expenditure – fourth quarter reported expenditure), we find similar results. The variable derived from the latent class analysis showed good differentiation in the expected direction. The authors concluded that the latent construct was indeed measuring the quality of reporting, but either lacked the sensitivity needed to adequately predict underreporting, or measurement error is not a strong predictor of the average expenditure reported by the CU or the amount of bias, as measured by the coarse measure

4. Current Research

The current research advances prior work in a two significant ways. Firstly, the data are more current, extending from the second quarter of 2005 to the second quarter of 2009. The more current data also incorporate information collected in the Contact History Instrument (CHI). These data were not collected prior to the second quarter of 2005. These data capture a number of attributes of contact attempts made by field interviewers, including the number of contacts, mode of contact (phone or in-person), and reasons for refusal or noncontact that were recorded by the interviewer. This information is quite detailed. The authors hypothesized that these could add strength to the models constructed in previous work.

The contact history data were refined into indicators for each wave and for the overall panel for each second quarter respondent. In order to combine the detailed reasons for refusal and noncontact (by simple tally), the authors sought guidance from research conducted by Maitland, Casas-Cordero, and Kreuter (2009) and Dixon (2009). Utilizing factor analysis these authors identified the same factors from the CHI in two different surveys. The strongest factors related for reluctance could be grouped by privacy concerns, reluctance, and hostility. Only two factors were identified for reasons for noncontact (gatekeepers or barriers and “other”).

The new indicators are as follows:

1. Indicators of reluctance based on privacy concerns
2. Indicators of reluctance based on time concerns
3. Indicators of especially hostile refusal
4. Any reluctance
5. Indicators of noncontact based on gatekeepers or other barriers
6. Indicators of noncontact based on other problems
7. Proportion of attempts made in-person
8. Proportion of completed interviews that were completed by phone

Model selection was conducted under two separate strategies. The first strategy ran indicators that were examined in previous research in all possible combinations of three and four and ran the “new” CHI indicators in separate latent class analysis also with all possible combinations of three and four indicators. The best performing indicators in these models were then combined. Strategy two grouped all indicators by theoretical concept. Three groups were formed: 1. reluctance indicators; 2. noncontact indicators; 3. all other indictors. After finding the best model using the reluctance and noncontact indicators, other variables were added to the model and tested for fit. 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. Five “best” models were selected. Although three and four category latent class constructs were examined, all of the best fitting models had three latent classes.

1. ***Old model***: This model was based on the best fitting indicators of those examined in previous research. This was done to confirm these results with the more recent data. Indicators: Income missing; Record use combined with interview length; Average number of commodity categories for which CU reported an expenditure; Number of completed interviews
2. ***CHI model***: This model consisted of only the “new” CHI indicators. Indicators: Reluctance due to privacy concerns; Tally of all noncontact problems; Average number attempts.
3. ***Combination model***: This model combined the best fitting indicators from the Old and CHI models. Indicators: All the “Old” model indicators; Average number attempts.
4. ***Reluctance model***: This model combined the best fitting indicators that indicate a degree of reluctance on the part of the CU. Indicators: Income missing; Record use combined with interview length; Number of completed interviews; Reluctance due to time constraints.
5. ***Noncontact model***: This model combined the best fitting indicators that indicate difficulty in making contact with the CU. Indicators: Record use combined with interview length; Tally of noncontact problems; Average number of attempts.

After these five 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. The mean expenditure (across CUs) is significantly lower in the fourth wave interviews compared to that of the second. It is commonly thought that CUs, on average, underreport their expenditures in the fourth wave. Utilizing this as an indicator of poor reporting, the authors regress the latent class assignment on the proportional difference between the total expenditure reported on the second and fourth panel waves of the interview and the standard deviation of the total expenditure across all completed interviews within the panel. The formula for the proportional difference follows:

$$Error = \begin{cases} 0; X_2 \le X_4 \\ (X_2 - X_4)/X_2; X_2 > X_4 \end{cases}$$

5. Results

Although acceptable, the models emerging in the current latent class analysis generally have poorer fit than those selected as the best models in prior research. Table 1, shows the class probabilities associated with each of the latent variables. Note the “previous best” is the best model from previous research, whereas the Old model uses the same indicators but is estimated with the current (2005-2009) data. The difference in the size of the first two classes is somewhat striking considering the only difference in these two are the timeliness of the data.

Table 1: Class Probabilities

	Poor	Fair	Good
Previous Best	.203	.232	.565
Old	.137	.281	.582
CHAI	.245	.434	.321
Combo	.140	.285	.575
Reluctance	.138	.309	.554
Noncontact	.279	.574	.148

Table 2 shows the average overall expenditure per quarter by latent class and model. Corresponding with the change in the relative size of the class mentioned above, unlike the Previous Best model the mean overall expenditure does not differ across the first two classes of the Old model. Other models, including the Combo, Noncontact, and CHI models, suffer from the same lack of differentiation including.

Table 2: Average Overall Expenditure by Latent Class Variable

	Poor	Fair	Good
Previous Best	6,946.84	8,920.20	11,985.71
Old	10,359.92	10,032.08	13,231.43
CHI	12,365.20	12,022.62	10,895.20
Combo	10,725.39	10,492.81	13,014.24
Reluctance*	10,385.60	11,124.69	12,708.89
Noncontact	12,774.90	12,432.04	10,678.05

Looking at the expenditure by some of the commodity types for the Combo and Reluctance models, we see similar inconsistent findings. For some commodities, such as electricity and trash collection, furniture, and dental care, the latent constructs perform as expected. For other categories, such as sports equipment, apparel, and television and other electronics, we see that the latent classes do not perform as expected, with mean expenditure often lowest in the second class. Other models performed as poorly.

Table 3: Mean Commodity Expenditure by Latent Class

Combo	**Poor**	**Fair**	**Good**
Electricity	267.02	301.61	331.37
Trash	7.18	13.08	19.73
Sports	24.86	**20.57**	33.49
Furniture	30.70	42.87	76.90
Kitchen Accessories	5.14	11.48	23.55
Major Vehicle Repairs	17.20	24.67	43.72
Gas	98.03	113.94	135.79
TV	104.71	**75.94**	113.59
Women's Apparel	117.36	**97.86**	127.00
Men's Apparel	**74.04**	**59.72**	**74.99**
Kid's Apparel	47.89	**41.94**	54.59
Dental	15.86	31.40	61.72
Reluctance	**Poor**	**Fair**	**Good**
Electricity	262.42	316.52	325.77
Trash	7.18	14.82	19.04
Sports	24.91	**22.69**	32.41
Furniture	25.76	54.20	73.00
Kitchen Accessories	4.81	14.77	22.41
Major Vehicle Repairs	15.38	29.25	42.47
Gas	98.56	122.61	132.01
TV	96.15	**82.80**	110.52
Women's Apparel	111.43	**108.34**	121.25
Men's Apparel	**74.51**	**66.22**	**71.95**
Kid's Apparel	46.01	47.94	51.74
Dental	15.12	37.94	59.53

The relationship of the demographic variables with any of the latent class variables were consistently in the expected direction. For example, CUs that rent showed a higher propensity of being in the lowest latent class, followed by the second lowest, and finally the best class. Table 4 shows the proportional odds coefficients for the Combo, Reluctance, CHI, and Noncontact models.

Table 4: Proportional Odds Model Results

Latent Variable	**Combo**		**Reluctance**		**CHI**		**Noncontact**	
	Exp(b)	PR(X^2)	Exp(b)	PR(X^2)	Exp(b)	PR(X^2)	Exp(b)	PR(X^2)
Famsize 1	.887	.0145	1.129	.0102	.907	.0030	1.099	.0838
Famsize 2	.948	.8415	1.060	.9456	.999	.0869	1.024	.1812
Age	1.007	<.0001	1.007	<.0001	1.025	<.0001	1.021	<.0001
Educ	.962	.2868	.990	.7562	.970	.3119	0.988	.6784
Inc rank1	.782	.0009	.920	.9954	1.211	<.0001	1.407	<.0001
Inc rank2	.809	.0039	.846	<.0001	1.071	.2961	1.138	.0877
Race	1.420	<.0001	1.439	<.0001	1.263	<.0001	1.135	<.0001
Tenure	.805	<.0001	.880	.0003	.931	.0310	.962	.2140
Urban	.878	.0477	.947	.3516	.833	.0006	.855	.0019
Cell complete	.929	.2783	1.016	.8026	.918	.1496	.865	.0099
Max-rescaled R^2		.018		.016		.055		.057

Table 5 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 by the introduction of the latent construct. The means of the expenditure by latent class controlling for the demographics variables in the model are significantly different but are not in the expected direction, where the Fair category of the latent construct has the lowest means expenditure. These results are similar for a number of commodity categories (not shown), although for some commodities the latent construct does perform as expected controlling for demographics.

Table 5: MANCOVA Results: Total Expenditures Combo Model

	Baseline model		With LV	
	Estimate	p-value	Estimate	p-value
F	1569.60	<.0001	971.82	<.0001
R^2	.41		.41	
F[Contribution of Latent variable]:			46.93	<.0001
F[Contribution of Interaction Term]:			8.32	<.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	10,263.94		<.01	.89
Fair	9,165.52	<.01		<.01
Good	10,383.46	.89	<.01	

*Scheffe adjustment for multiple comparisons

Table 6 shows summarizes the results of regressing the proportional difference of the Wave 2 expenditure and Wave 4 expenditure measure on the latent construct controlling for demographic variables. Again the contribution of the latent construct to the model is statistically significant, however the adjusted R^2 is extremely small overall. In this model the latent construct performs better than in the model for total expenditure. where the proportional difference (a sign of bad reporting) is smaller for the Good reporting group and highest for the Poor reporting group. The differences are statistically significant.

Table 6: MANCOVA Results: Proportional Difference Combo Model

	Baseline model		With LV	
	Estimate	p-value	Estimate	p-value
F	1.48	.1296	8.41	<.0001
R^2	.00		.00	
F[Contribution of Latent variable]:			27.52	<.0001
F[Contribution of Interaction Term]:			7.15	<.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	.207		.04	<.01
Fair	.148	.04		<.01
Good	.117	<.01	<.01	

*Tukey-Kramer adjustment for multiple comparisons

6. Discussion

Overall, the latent constructs did not perform as well as they did in previous research. While the model fit was still good, the CHI variables grouped more closely with themselves than with any other indicators and seemed to contribute little to the efficacy of the latent construct in predicting reporting error. Overall the strongest indicators across models were income missing, record use combined with interview length, number of completed interviews, reluctance due to time constraints, and average number of attempts. These indicators were also strong in previous research. As in previous research the latent constructs lack 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 total expenditure, or the proportional difference in Wave 2 and Wave 4 expenditure reports.

Unlike previous research we do not consistently show differences in expenditure amounts in the expected direction across the three classes. For many commodity categories and for overall expenditure we can only differentiate between two classes of reporting quality. The current latent constructs appear to be relatively blunt instruments (although the only instruments we have), and are probably not useful for adjustment as they do not explain much of the variation in expenditure or change in expenditure.

Future research will attempt to develop indicators that may be able to further differentiate reporting quality among expenditure reporters. One direction is indicated from the results of Tables 1 and 2. The models emphasizing noncontact (CHI and Noncontact) have very different class probabilities than models emphasizing reluctance (Combo and Reluctance). In addition, the expenditure means across class are quite different for these models. It is possible that utilizing two different, two-class latent constructs, corresponding to the CU's reason for bad reporting, will result in a more effective model.

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.

Dixon, J. (2009). "Modeling the Difference in Interview Characteristics for Different Respondents." *Proceedings of the Section on Government Statistics*. Washington, DC: American Statistical Association.

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.

Tucker, C., Biemer, P., and Meekins, B. 2008. "A Microlevel Latent Class Model for Measurement Error in the Consumer Expenditure Interview Survey." *Proceedings of the Section on Survey Research Methods*, Washington, DC: American Statistical Association.

Maitland, A., Casas-Cordero, C., and Kreuter, F. (2009), “An Evaluation of Nonresponse Bias Using Paradata from a Health Survey.” *Proceedings of the Section on Government Statistics*. Washington, DC: American Statistical Association.

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