**Longitudinal Assessment of Measurement Error on the Consumer Expenditure Interview Survey: 1996 – 2010**

Brian Meekins[1]
[1]Bureau of Labor Statistics, 2 Massachusetts Ave, NE Washington, DC, 20210

## Abstract
Previous work by the author used Markov Latent Class Analysis (MLCA) to make aggregate estimates of the underreporting of household expenditure by category (e.g. clothes, furniture, and electricity) by exploiting the four interview, rotating panel design of the Consumer Expenditure Interview Survey (CE). This analysis collapsed a few years of the CE into a pooled single panel. Estimates from this analysis were shown to be consistent with both internal and some external indicators of measurement error. However, there is no "gold standard" for expenditure reports and MLCA models require strong assumptions for identifiability making evaluation of model estimates difficult. The current analysis uses MLCA on data from 15 years of the CE, in order to examine the reliability of measurement error estimates over time. Both sequential and overlapping panels are used in order to assess the sensitivity of MLCA estimates to major survey revisions. Finally, these estimates are compared to survey nonresponse and external benchmarks where available.

**Key Words:** Measurement Error, Markov Latent Class Analysis, Expenditure Reports

## 1. Data

Markov Latent Class Analysis has been used in the past to estimate classification error or misclassification, a type of measurement error, in a number of different surveys, including the Consumer Expenditure Survey (CE) (VandePol and deLeeuw 1986; Tucker 1992; Van de Pol and Langeheine 1997; Bassi et al. 2000; Biemer and Bushery 2000; Tucker, et al. 2002, 2003, 2004, 2005, 2006, and 2008); Meekins et al. (2011)). The advantage of Markov Latent Class Analysis is that these models utilize the repeated measures of the panel design and do not require external data for validation.

The Consumer Expenditure Survey (CE) is one such large panel survey where there is concern about classification error. The interview is long and fairly burdensome asking respondents about their purchases and expenditure in the previous three months. Approximately, 6,000 households are interviewed each year for the CE. Each household is interviewed once every three months, for a total of five times. However, the first interview is not typically used for the analysis of purchases and expenditure, but rather as a "bounding" interview to avoid reverse telescoping - the tendency to include purchases from greater than the prior three months. For this analysis we combined 15 years of the CE, using the four interviews that are used to estimate expenditure for each cohort. A total of 31 commodity categories were analysed. A list of these commodity categories is shown in Table 1.

**Table 1:** CE Commodity Categories

| | | |
|---|---|---|
| Dental | Computer games | Childcare |
| Prescription drugs | Computer equipment | Pets and pet supplies |
| Eye care | Books | Major Vehicle Repairs |
| Clothing | Cable | Minor Vehicle Repairs |
| Infant clothing | Music | License/registration |
| Clothing accessories | Internet (2001+) | HH electricity |
| Clothing services | Sports equipment | HH gas |
| Sewing | Major appliances | HH trash service |

| | | |
|---|---|---|
| Shoes | Minor appliances | Phone |
| Jewelry | Electronics | HH services |
| Events (e.g. sporting/theatre) | | |

In estimating expenditures the CE first asks the respondent if a purchase for a particular commodity or set of commodities was made during the previous three months. If the respondent answers in the affirmative, the respondent is asked the quantity of the expenditure. The CE uses this reported amount (with some editing) to estimate expenditure on reports. While, there is evidence that respondents misreport (either over or under report) expenditure, past research has shown that a large amount of error occurs in the reporting of purchases. This work focuses on the misclassification of reporting purchases.

## 2. Mover-Stayer Model

The models used in this analysis have two components. One part is a straightforward single order Markov Latent Class model, while the other part of the model is called the "mover-stayer" specification, which consists of an additional latent construct that constrains the four latent variables that are estimated in the Markov model. Figure 1. Illustrates the measurement error or classification error part of the model. The four observed variables, A, B, C, D and the four latent construct W, X, Y, Z. Note that there is one indicator for each of the latent constructs.

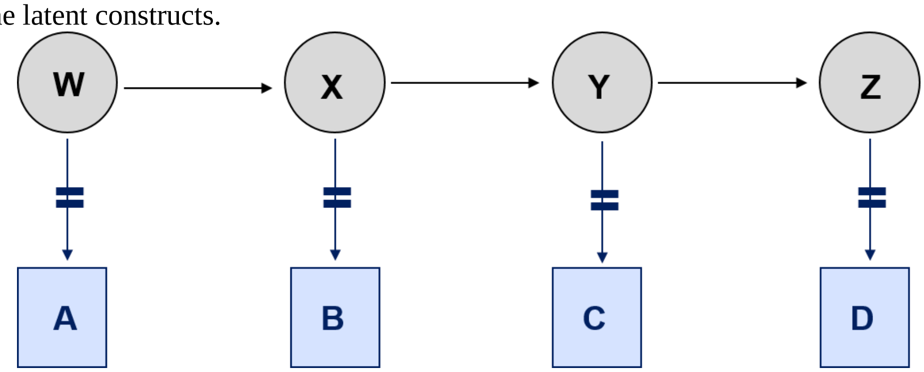


**Figure 1** Measurement Compenent Markov Latent Class Model

The four latent construct W,X, Y, Z are constrained by the latent construct M into three classes: 1.Stayer purchases, where the respondent purchases a given commodity for all four quarters; 2. Stayer non-purchasers, where the respondent does not purchase the commodity in any of the four quarters and: 3. Movers – where the respondent has at least one quarter where they purchase the commodity and one quarter where they do not. The constraints are defined as follows, where M is the mover-stayer latent construct:

$$\boldsymbol{M} = \begin{cases} 1, & P(W{=}1) = P(X{=}1) = P(Y{=}1) = P(Z = 1) = 1 \\ 2, & P(W{=}1) = P(X{=}1) = P(Y{=}1) = P(Z = 1) = 0 \\ 3, & P(W), P(X), P(Y), \text{ and } P(Z) \text{ are unconstrained.} \end{cases}$$

Even with Markov relationships specified this model is underidentified. Therefore, a number of fairly strong assumptions are made in order to estimate. In addition to the Markov assumptions and the specifications of the Mover-Stayer model, we assume that measurement error is equal across all time periods:

$$P(a_i = j \mid w_i = k) = P(b_i = j \mid x_i = k)$$
$$= P(c_i = j \mid y_i = k) = P(d_i = j \mid z_i = K) = q_{jk}$$

We also assume that no respondents report a purchase when no actual purchase was made - there are no false positive reports:

$$P(a_i = 1) \mid (w_i = 2) = 0$$
$$P(a_i = 2) \mid (w_i = 2) = 1$$

Where,
$w_i = 2$, indicates that no purchase was made in the *i*th quarter
$a_i = 1$ a purchase was reported by the respondent in the *i*th quarter
$a_i = 2$ no purchase was reported by the respondent in the *i*th quarter

## 3. Estimation and Measures

Past research has shown that grouping variables are helpful in getting the model to converge and can be helpful in identifying sources of classification error. Variables that were shown in previous research to be helpful in grouping are: family size, refusal to answer income question, record use and interview length (coded in a single variable), and income class (imputed or derived if missing).

Models were estimated using lEM, a free software, created by Jeroen Vermunt. The models were selected using objective diagnostics (L-square, BIC, and dissimilarity index). Multiple iterations were conducted in order to avoid local maxima. The CE is a rotating panel survey with new sample entering each quarter. The models are first run with using these individual cohorts for each quarter. However, in order to improve the consistency of the estimates, the data are pooled to combine cohorts for four consecutive quarters. These we call annual cohorts. They consist of respondents that started their individual cohort in Quarters 1, 2, 3, 4 of each year.

The primary measure of interest is the accuracy rate which is defined as the estimated probability that a purchase was reported given that a purchase was made or $P(a{=}1|w{=}1)$. Ideally, we would wish this probability to be 1. To the extent that this rate is less than 1 it represents misclassification of purchases for that commodity.

## 4. Results

Figure 2 shows the results of the accuracy rate of cable TV purchases (blue line) plotted over the years 1996, quarter 2, to 2009, quarter 2. Also plotted is the $P(A{=}1)$, or the unweighted percentage of respondents that reported a purchase of cable TV in the given commodity in that quarter. Of particular note is the clear drop in both lines, accuracy and reported purchases, in 2005, Quarter 1. This quarter marked significant changes in the design of the CE.

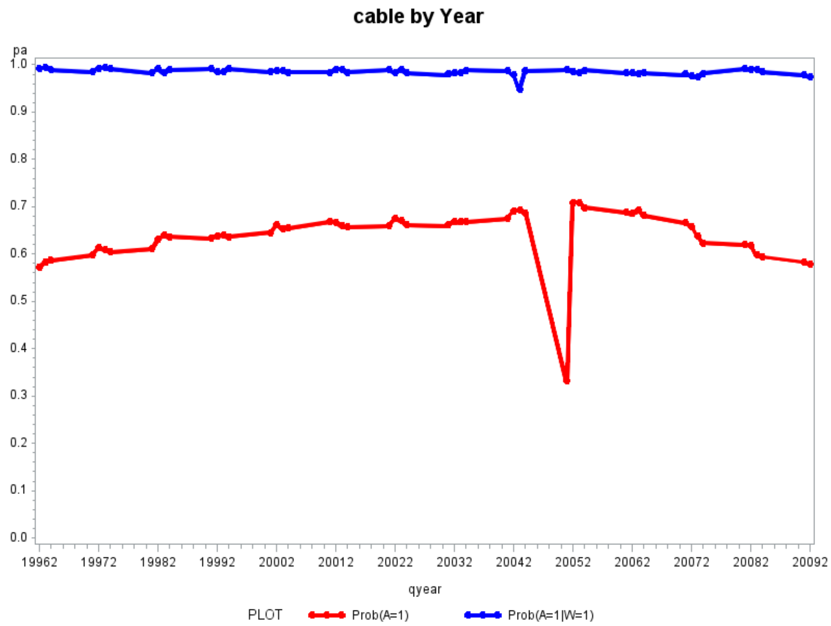


**Figure 2:** Cable TV by Year (Individual Cohort)

Figure 3 shows cable TV by family size by year. Controlling for family size, we see only a dip in accuracy for single family households *(G=1)*. This would indicate that these households were primarily responsible for the misreporting of cable TV purchases in that quarter.

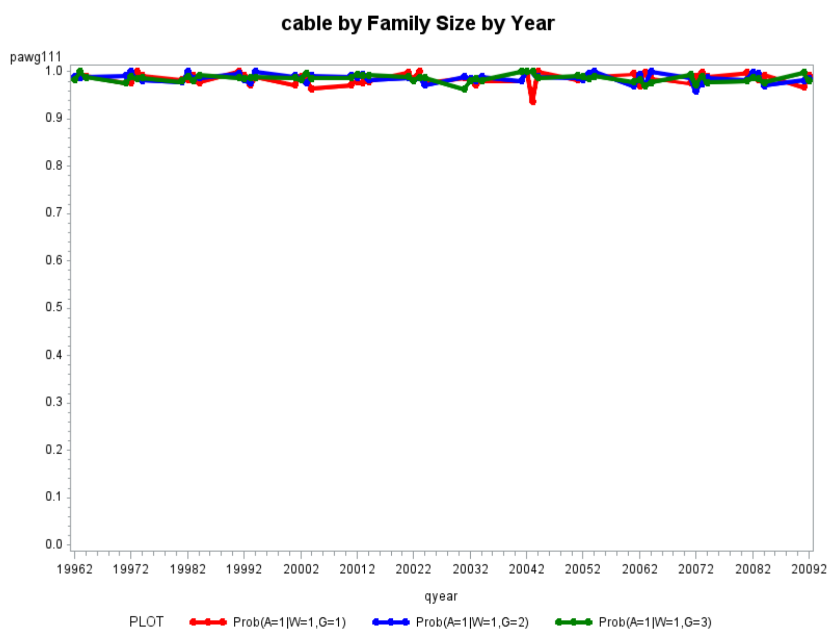


**Figure 3:** Cable TV by Family Size and Year (Individual Cohort)

Figure 4, shows a plot for the accuracy and reported purchases of electricity. Here we see that the change in the incidence of reporting, *P(A=1)* in the 2005, Quarter 1 is not reflected in the accuracy rate.

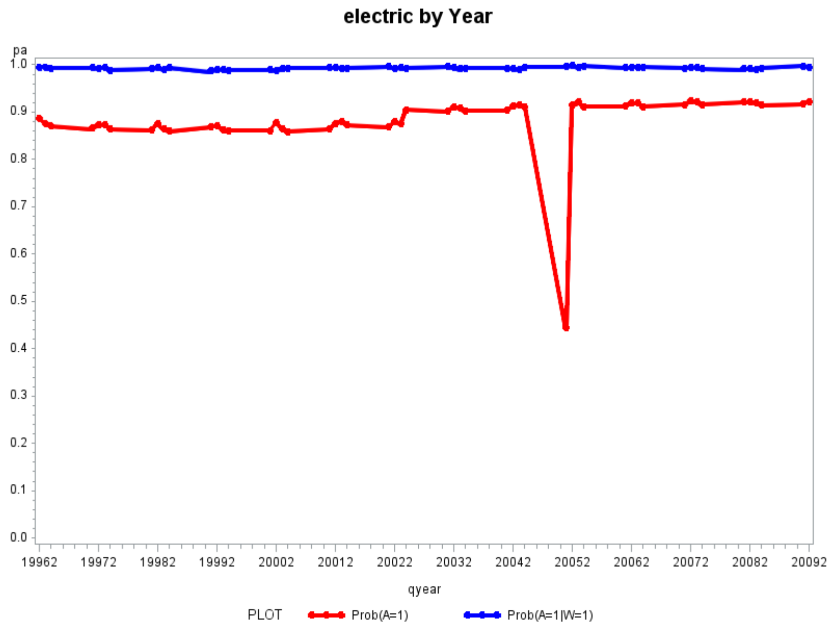


**Figure 4:** Electricity by Year (Individual Cohort)

While the plots for cable TV and electricity show a fairly flat trend from 1996 to 2009, the plot for music (Figure 5) appears to show a marked decrease in the accuracy of the reporting of music purchases. To a lesser extent the proportion of respondents reporting a music purchase is declining. This seems to make sense due to the changing nature of how music has been purchased over those years (an increase in internet sales and easy to miss online purchase of single tracks). Also of note in Figure 5 is the noisy nature of the line. Indeed, music is purchased and reported as a purchase much less frequently than cable TV and electricity. These creates cells with a small number of cases, and creates difficulty for estimation of Markov Latent Class Models.

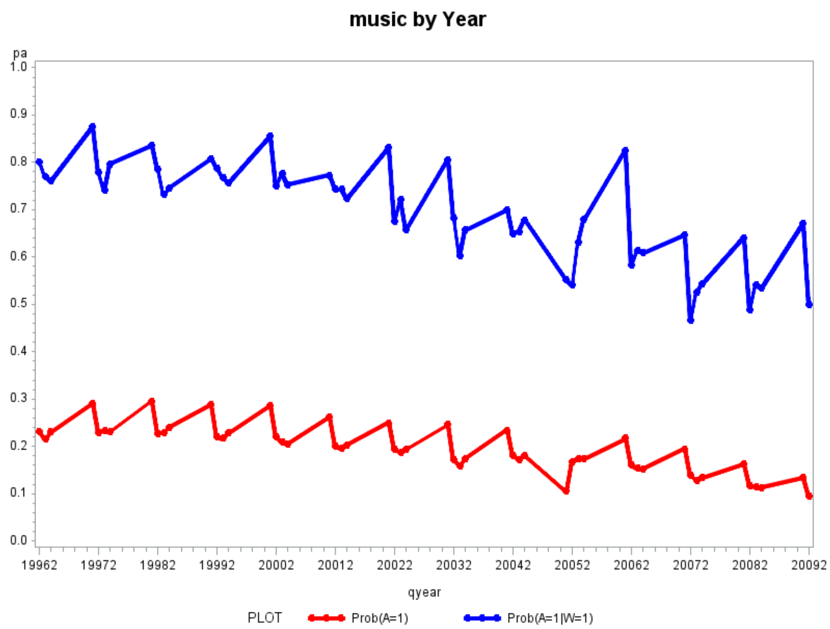


**Figure 5:** Music by Year (Individual Cohort)

In response to noisy estimates for a number of commodities, the authors chose to pool data from a grouping of four consecutive cohorts, creating an "annual" estimate. As shown in Figure 6, this greatly reduces the amount of noise in the data for music purchases.

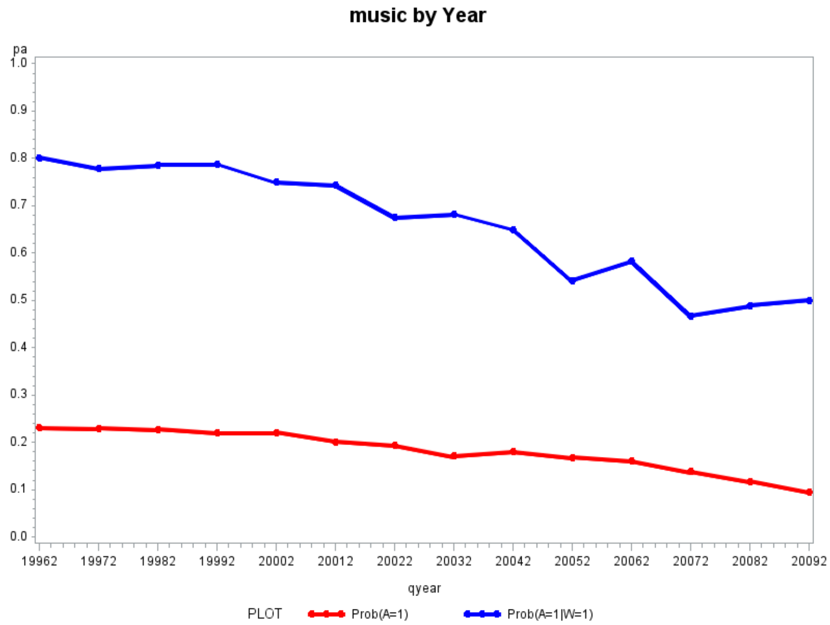


**Figure 6:** Music by Year (Annual Cohort)

The declining trend in accuracy of reporting is quite clear in this plot. For the remainder of the analysis, the combined annual cohorts will be used. Figure 7, shows the annual cohort for electricity. Note the much higher accuracy rate, as we would expect for a purchase with a high degree of regularity, where supporting records (energy bills) can usually be produced.

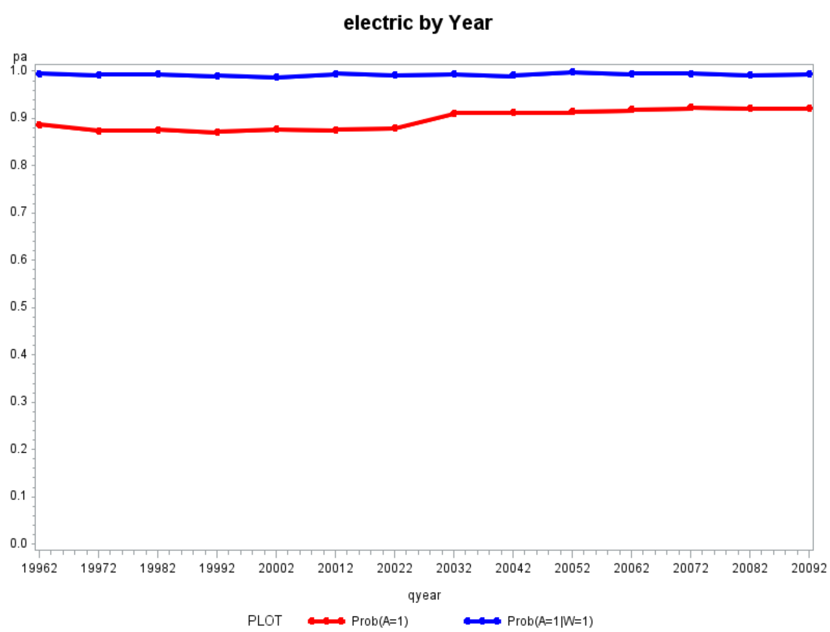


**Figure 7:** Electricity by Year (Annual Cohort)

Figure 8, shows the annual cohort for minor appliances. These items, such as blenders and hair dryers, are purchased at much lower regularity and are unlikely to be supported by records or documentation. Therefore, the accuracy rate is quite low. Figure 9, shows the accuracy rate for music again, but broken out by income class. As you can see the data again become very noisy, even when using the annual cohorts, with little to no ability to diagnose a trend.

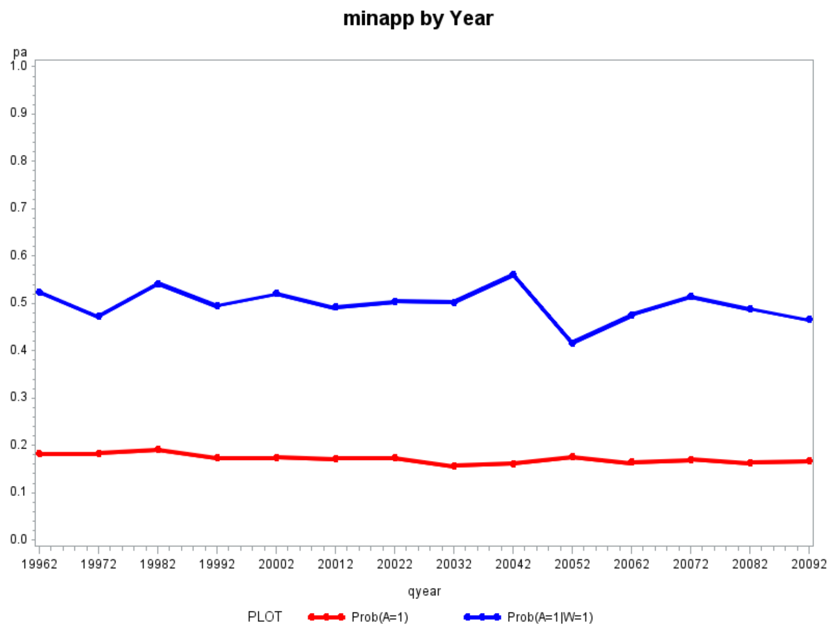


**Figure 8:** Minor Appliances by Year (Annual Cohort)

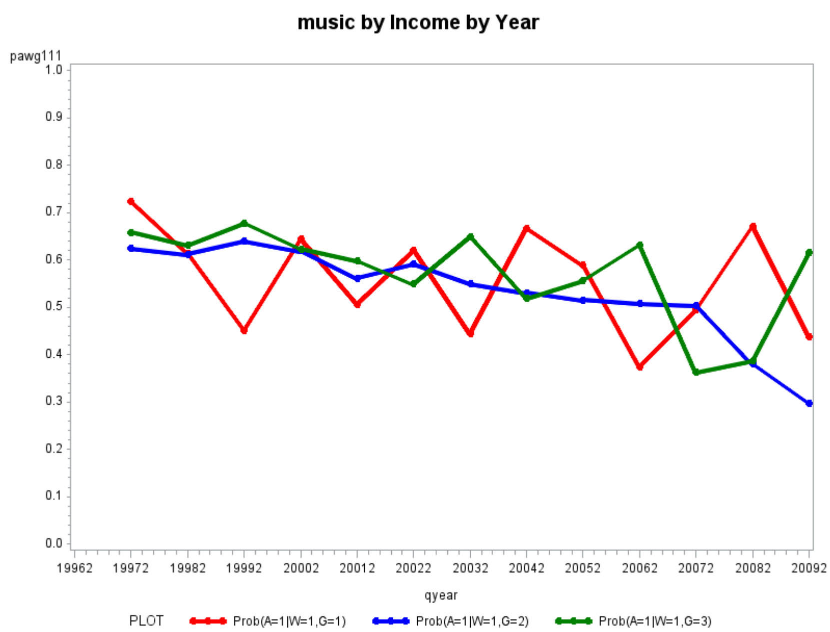


**Figure 9:** Music by Income by Year (Annual Cohort)

## 5. Conclusion

Markov Latent Class Analysis, in general, and the Mover-Stayer model, in particular, show great promise for the detection of change in classification error *over time*. Results show that reasonable pooling of data using four consecutive cohorts shows consistency in estimates of accuracy rates. Results appear to make sense with what we know about purchases on the CE – specifically, that irregular purchases without supporting records are reported with less accuracy. This finding was consistent over all commodity categories that were examined. Commodities like electricity, cable TV, and even household internet service (purchases by many fewer households), had very high accuracy of reporting, while music, minor appliances, and minor vehicle repairs had far less accuracy.

This technique, however, does require a significant sample size. This is the case, especially, if one wants to examine the covariates of classification error. Introducing even a single grouping variable in the model greatly increases the inconsistency of the estimates. Larger grouping of cohorts is necessary if one is to examine these variables. It then becomes a balance between precision from estimation and precision by time. One solution to this issue may be to use smaller groups to identify departures from a trend, which these models seem capable of detecting (see comments regarding 2005 quarter 1), or identify a point in time based on interest, such as changes in the survey administration or instrument. Having identified these time points, one can then group cohorts on either side of that time point to the extent needed to truly examine the possible sources of the change in measurement error.

## References

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., 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, Brian, Paul Biemer, and Clyde Tucker. (2011). "Latent Class Analysis of Measurement Error in the Consumer Expenditure Survey." *Proceedings of the Joint Statistical Meetings*. Miami, FL.

Tucker, Clyde, Brian Meekins, and Paul Biemer. (2011). "Estimating Underreporting of Consumer Expenditures Using Markov Latent Class Analysis." *Survey Research Methods*. July.

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.