Methods for Fitting a Markov Latent Class Analysis for the National Crime Victimization Survey

Marcus Berzofsky[1], Paul P. Biemer[2], William Kalsbeek[3]
[1]RTI International, 3040 Cornwallis Rd., RTP, NC 27709
[2]RTI International, 3040 Cornwallis Rd., RTP, NC 27709
[3]University of North Carolina at Chapel Hill, 730 MLK Blvd., Chapel Hill, NC 27599

Abstract
This paper presents the methods for the first assessment of classification error (measurement error for categorical data) in the National Crime Victimization Survey (NCVS). The NCVS is the only national survey that captures information on victimizations that are both reported and unreported to the authorities. To estimate reporting errors we use Markov latent class analysis (MLCA), a modeling technique that can be used to estimate classification error in panel data that does not require a gold-standard (error-free) measurement. This paper proposes a process by which an MLCA can be conducted on complex survey data, ensuring that all key assumptions are met or corrected for such that parameter estimates are valid. To conduct this analysis, we used a special longitudinal file containing all respondent waves from a sample of NCVS households. We determined that a model with fully constrained transition probabilities and partially constrained classification error probabilities fit the NCVS data the best.

Key Words: Markov latent class analysis (MLCA), National Crime Victimization Survey (NCVS), measurement error, classification error, screener questions

1. Introduction

1.1 Background

Classification error, measurement error for categorical data, occurs when a respondent incorrectly identifies their true status (Biemer, 2011). This paper seeks to estimate the levels of classification error in victimization reports for the National Crime Victimization Survey (NCVS). The NCVS is a nationally representative survey of households that estimates the crime victimization rates for nonfatal crimes in the United States. It is also the only national survey that captures information on crimes that are reported to the police as well as crimes that are not (U.S. Department of Justice, 2004). As such, it is a critical survey in our understanding of what crime rates are in the United States and how they are changing over time. Sponsored by the U.S. Bureau of Justice Statistics (BJS) and conducted by the U.S. Census Bureau, the NCVS has been conducted continuously since 1972 (starting as the National Crime Survey [NCS] and becoming the NCVS in 1992).

The NCVS is designed as a rotating panel survey (i.e., a design by which respondents remain in the sample for a set number of interviews and are then replaced by a new sample of respondents) whereby randomly selected households are retained in the sample for seven interviewing waves spaced 6 months apart. The panel design is used to help estimate changes in the victimization rate. The first interviewing wave is called the bounding interview. The bounding interview establishes a time frame for reported victimizations so that victimizations reported in successive interviews are only counted once (Lohr, 1999). The remaining six interviews ask about all crimes that occurred in the preceding 6 months. Prior to 2006, only data from the six interviews that followed the bounding interview were included in the published estimates. Individuals in a selected household that are 12 years old or older are interviewed in each wave. Published estimates are produced on an annual basis. These estimates are based on all interviews whose reference period overlaps with the reporting year. Because of the rotating panel design these interviews include an approximately balanced amount of interviews from each possible time-in-sample wave.

For crime victimization in the NCVS, where the outcome is dichotomous, there are two types of misclassification. A respondent can indicate that they were not a victim of a particular type of crime when they really were (i.e., a false negative response) or a respondent can indicate that they were a victim of a particular type of crime when they really were not (i.e., a false positive response). Understanding the classification error in the NCVS is important for several reasons. For example, if the number of people classified in either classification error type is disproportionately larger (i.e., the absolute number of people misclassified is larger for one type of misclassification) than the other, the published estimates may be biased. If the number of false negatives is higher, then the estimates will be negatively biased (i.e., smaller than they should be), but if the number of false positives is higher, then the estimates will be positively biased (i.e., larger than they should be). These biases can be offsetting (i.e., the false positive rate compensates for the false negative rate) leading to unbiased estimates at the aggregate level. Furthermore, if the classification error rates differ greatly across subpopulations then comparisons among subpopulations may be misleading. For example, if the crime rate for young respondents is larger than the rate for older adults, but younger adults have a much higher false negative rate than older adults, the estimate of the difference between these two groups will be understated. Thus, estimates of the classification error for a key subpopulation can help in the interpretation of differences among subpopulations. Another important purpose of classification error evaluation is quality improvement. Because the NCVS is a continuous survey, a better understanding of which subpopulations have higher rates of misclassification can inform quality improvement efforts to reduce classification errors in the survey. Furthermore, classification error estimates can be used to adjust the estimates for bias, provided the estimates are credible and accurate.

The NCVS interview is split into two sections: the screening section and the incident report. The screening section asks each respondent basic demographic questions as well as seven individual crime victimization screening questions (i.e., a yes/no question screening to determine if a particular event, such as a crime, occurred). Each household has a designated informant that is asked three screening questions on household crimes in addition to the seven individual crime screening questions. All respondents are asked the screening questions at each interviewing wave, while only those that indicate a crime occurred during the screening section are asked to complete an incident report (i.e., a detailed set of questions about the crime). The NCVS interview is a two step process. In the first step a respondent is asked if a crime occurred (screener questions), and in the second step, if the respondent indicated a crime occurred, they are asked a battery of question about the crime (incident report). Because of this two step interviewing process, any potential misclassification regarding the occurrence of a crime occurs at the screening section.

Screener misclassification can occur for one of five reasons

- encoding,
- comprehension error,
- recall error,
- satisficing, and
- social desirability.

Encoding is when an event occurs (in the case of the NCVS a crime), but the respondent (or proxy respondent) is not aware. Encoding may occur when a proxy respondent is used because the proxy does not know if the respondent was victimized. The NCVS tries to minimize the use of proxy respondents to minimize this type of error. Comprehension error is when a respondent either reports a victimization or does not report a victimization because they misunderstand the screener question. Recall error is when a respondent does not recall a crime that may have occurred months ago or may recall it as happening sooner or later than it actually occurred. To minimize classification error due to comprehension and recall error, improvements in the questionnaire were made during the 1992 redesign to clarify the meaning of the questions as well as to improve recall of victimizations (see Biderman & Cantor, 1984; Biderman, Cantor & Reiss, 1982, 1984; Biderman & Lynch, 1981; Bushery, 1981; see also Groves & Couper, 1992, 1993). However, even with the improvements, errors can still frequently occur. For example, comprehension error in the screener questions could still be a source of classification error if the respondent does not know whether what happened to them was a crime and, therefore, reports erroneously
Another possible source of classification error is satisficing (i.e., respondents may not want to report victimizations to make the interview shorter). Since the NCVS is a panel survey, respondents will learn that answering positively to a screener question will necessitate an incident report. To avoid this, respondents may deny crimes that actually occurred—a form of panel conditioning (Yan, 2008). The fifth source for underreporting victimizations is social desirability. A respondent will deny a crime occurred because they are too embarrassed to admit it happened. For this reason, interviewers are instructed to create a private setting for the interview if at all possible.

Although the NCVS has been conducted for almost 40 years, very little is known about the level of classification error in its estimates. Based on our review of the literature, only one other study on misclassification has been conducted. This study was conducted by the Law Enforcement Assistance Administration ([LEAA] 1972) to assess measurement error in the NCS. The study used a reverse-records check technique by which police department records were used to identify known victims during a particular period of time. The victims for which addresses could be obtained were interviewed to determine the underreporting (i.e., false negative) rate of crime. The researchers found that 38.4% of victims of individual crimes did not self-report the crime and 13.8% of property crime victims did not report the crime (LEAA, 1972). More recently, some research has looked into satisficing through time-in-sample bias or panel conditioning (see Lynch, Berbaum, & Planty, 2002; Yan, 2008; Rand & Catalano, 2007) or respondent fatigue (Hart, Rennison, & Gibson, 2005; Hart, 2006). Time-in-sample bias is the term used to describe the response bias that is a function of the number of times a respondent is interviewed in a panel survey. One theory is that this bias is the result of respondents being "conditioned" by prior interviews to misreport in ways that will shorten the interview (Kalton, Kasprzyk, & McMillen, 1989). Denying a victimization will shorten the interview by avoiding further questioning about the victimization, thus reducing respondent burden. Respondent fatigue is the term used to describe a respondent who avoids responding in future waves because they have grown tired of taking the survey (Biderman, 1967; Biderman et al., 1967). Although time-in-sample bias and respondent fatigue may be related to classification error, they do not fully explain why misclassification may occur. Therefore, an analysis of classification error is important to add greater context to the understanding of nonsampling error in the NCVS.

1.2 Purpose

In this paper, we will describe the methods used to estimate the classification error rates in the NCVS and present the model that best fits the NCVS data. The estimates of classification error will be derived from a Markov latent class analysis (MLCA) of the NCVS. MLCA, first proposed by Wiggins (1973) and further developed by Poulsen (1982), Van de Pol & de Leeuw (1986), and Van de Pol & Langeheine (1990), extends the ideas of latent class analysis (LCA) to panel survey data. LCA, developed by Paul Lazarsfeld (1950), treats the true value of a questionnaire item (in this case a screener question or set of screener questions for whether the respondent was a victim of a particular type of crime during the reference period) as an unobservable (latent) variable. It specifies a model for this latent variable, taking into account the interrelationships between observed indicator variables and the grouping (subpopulation) variables. The theoretical development of this model originated with Lazarsfeld and Henry (1968) and Goodman (1974).

In lieu of a gold standard or a test-retest reinterview, MLCA provides information on classification error that can be linked to individual respondents. It requires neither greater respondent burden nor expensive additional data collection. MLCA assumes an error model for the available measurements and estimates the parameters of the error model using maximum likelihood. Thus, the validity of the MLCA estimates hinges on the ability of the model to accurately represent the error-generating mechanism. To obtain an estimable model, MLCA requires at least three panel measurements of the same construct or survey item. The MLCA model then specifies parameters for both the period-to-period changes in the status of the item as well as the measurement errors associated with the observations. Because the NCVS satisfies these estimabilty conditions, MLCA is well suited for estimating classification error in the NCVS.

This paper is the first application of MLCA to the NCVS. As such, we seek to develop a process by which an MLCA can be conducted on the NCVS that ensures that all model assumptions are met. However, while we apply this process to the NCVS the approach we develop is quite general and can be applied to any panel survey similar to the NCVS.

2. Methods

2.1 Technical Description of MLCA and Its Assumptions

Let X_i denote a latent variable representing a person's true (error-free) crime victimization status for reporting period i for $i = 1,\ldots,T$. For the NCVS X_i can take the values x_i=1, 2, where x_i =1 means the person truly was a victim and x_i =2 means they were not. For each latent variable there is a corresponding indicator variable that represents the survey observation at that time point. Indicators are denoted as A_i for i=1,…,T. In addition to latent and indicator variables, an MLC model may contain one or more grouping variable denoted by G, H, etc. A grouping variable is a manifest variable that does not directly measure the latent variable, but may help explain differences in victimization rates or in the classification error rates. For example, demographic variables, such as age, may help better explain crime rates over time as well as an individual's propensity to be misclassified as a victim. Grouping variables may also be derived from interview paradata, such as the mode of the interview, which may also affect a respondent's misclassification propensity.

Thus, an MLC model has three components: a grouping variable component, a structural component, and a measurement component (Biemer, 2011). The grouping variable component contains the cross product of the grouping variables that are included in either the structural or the measurement component. The structural component describes interrelationships among the latent variables and the grouping variables. The measurement component specifies the relationship between the observed realizations of the latent variables (i.e., dependent variables) and the latent variables. As such, the measurement component contains the classification error parameters.

For an MLC model with four time points (which, as discussed later in the chapter, will be the number of time points used in this analysis), the likelihood kernel may be written as

$$\pi_{ga_1a_2a_3a_4}^{GA_1A_2A_3A_4} = \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4}\pi_g^G \pi_{x_1|g}^{X_1|G}\pi_{x_2|gx_1}^{X_2|GX_1}\pi_{x_3|gx_2}^{X_3|GX_2}\pi_{x_4|gx_3}^{X_4|GX_3}\pi_{a_1|gx_1}^{A_1|GX_1}\pi_{a_2|gx_2}^{A_2|GX_2}\pi_{a_3|gx_3}^{A_3|GX_3}\pi_{a_4|gx_4}^{A_4|GX_4} \quad (1)$$

Where π_g^G, the grouping variable component, represents the probability of being in level g of grouping variable G for g=1,…k. The structural component consists of $\pi_{x_1|g}^{X_1|G}\pi_{x_2|gx_1}^{X_2|GX_1}\pi_{x_3|gx_2}^{X_3|GX_2}\pi_{x_4|gx_3}^{X_4|GX_3}$, where $\pi_{x_1|g}^{X_1|G}$ represents the initial probability of being a victim and $\pi_{x_i|gx_{i-1}}^{X_i|GX_{i-1}}$ for i=2,…T represents the transition probability of being a victim at time i given their victimization status at time i-1. The measurement component consists of $\pi_{a_1|x_1}^{A_1|X_1}\pi_{a_2|x_2}^{A_2|X_2}\pi_{a_3|x_3}^{A_3|X_3}\pi_{a_4|x_4}^{A_4|X_4}$, where $\pi_{a_i|gx_i}^{A_i|GX_i}$ represents the classification probabilities at time i for i=1,…T. A classification error occurs when $a_i \neq x_i$ (i.e., the survey response disagrees with a respondent's true status). For each of these probabilities the capital letter superscript represents the variable being estimated, and the lower case subscripts represent the value of that variable. For example, $\pi_{1|2}^{X_1|G}$ represents the probability of being a victim at the initial time point for a person in group g=2, and $\pi_{2|11}^{A_2|GX_2}$ represents the false negative rate at the second time point in group g=1. MLCA estimates the parameters using maximum likelihood via the EM algorithm.

MLCA makes four critical assumptions that are necessary for model identifiability (Biemer, 2011). The four assumptions are

1. The first-order Markov assumption states that $\pi_{x_i|x_{i-1}}^{X_i|X_{i-1}} = \pi_{x_i|x_{i-1}x_{i-2}}^{X_i|X_{i-1}X_{i-2}}$. Meaning that the transition probabilities only depend on the previous time point (i.e., the probability of being a victim only depends on whether the person was a victim during the previous time period regardless of their status two time periods ago).

2. The independent classification errors (ICE) assumption states that $\pi^{A_1A_2A_3A_4|X_1X_2X_3X_4}_{a_1a_2a_3a_4|x_1x_2x_3x_4} = \pi^{A_1|X_1}_{a_1|x_1}\pi^{A_2|X_2}_{a_2|x_2}\pi^{A_3|X_3}_{a_3|x_3}\pi^{A_4|X_4}_{a_4|x_4}$ for a_t=1, 2, x_t=1, 2, and t=1, 2, 3, 4. Meaning that the classification errors across waves are independent (i.e., the probability that a person accurately reports a crime is independent of the accuracy of reporting a crime at any other time point).
3. The time-homogeneous classification errors assumption states that $\pi^{A_t|X_t}_{a_t|x_t} = \pi^{A|X}_{a|x}$ for $a = a_t$, $x = x_t$, t=1, 2, 3, 4. Meaning that the classification errors for indicator A_t are the same in all waves t=1, 2, 3, 4 (i.e., the accuracy of reporting a crime is the same at all time points).
4. The group-homogeneous error assumption states that individuals in the group have the same probability of being misclassified (i.e., all individuals within a group have the same probability of misclassifying whether they were a victim of a crime).

Often with complex survey data, like the NCVS data, one or more of these assumptions is not met. For example, the first-order Markov assumption may be violated if the probability that a respondent is victimized during the current time period (e.g., time i) depends on whether that respondent was victimized both in time periods i-1 and i-2. Furthermore, the ICE assumption may be violated if the probability that a respondent accurately reports a particular type of crime depends on whether they accurately reported that type of crime in the previous time period. The time homogeneous classification errors assumption may be violated if, as an individual's time-in-sample increases, the individual becomes less willing to provide accurate answers because of panel conditioning or many other reasons. The group-homogeneous error assumption can be violated if the classification error rates between two groups differ greatly. If one or more of these assumptions are violated, then the estimated model may be invalid and biased. The direction and magnitude of the bias depends on which assumption is violated and the degree to which it is violated (Biemer & Berzofsky, 2011). Therefore, because these assumptions can be easily violated in a survey context, one needs to test each when conducting model selection.

When four or more time points are being modeled some of these assumptions can be relaxed. Although MCLA has been used to analyze complex survey data (see, e.g., Biemer, 2000; Biemer & Bushery, 2001; Biemer, 2004; and Tucker, Biemer, & Meekins, 2011) there is no agreed upon best approach for model selection. As described below, we propose a general approach for fitting MLC models that tests each assumption and, if it fails, relaxes the assumption in ways that produces a well-fitting and well-specified model.

2.2 The Data, Outcomes, and Grouping Variables

To conduct the MLCA, we used data from the *National Crime Victimization Longitudinal File, 1995–1999* (U.S. Department of Justice, 2007). This is the only publicly available data set that contains all seven waves of data for a set of respondents. Because the first interview wave is a bounding interview and is not used by the NCVS for estimation, our analysis was confined to waves 2 through 7. However, we could not simultaneously model all six waves because cross-classifying the data by six waves produced many empty cells—a problem referred to as data sparseness. Sparseness can cause model validity problems and lead to unidentifiable models (Biemer, 2011). Therefore, we split the data into three sets of four waves. We examined waves 2, 3, 4 and 5 together; waves

3, 4, 5 and 6 together; and waves 4, 5, 6, and 7 together. Furthermore, for our analyses, we included only respondents who responded to all four time points being examined. For example, for waves 2, 3, 4, and 5, a respondent must have responded to all four of these waves to be included in the analysis. Based on these criteria, approximately 25,000 observations were included in the crimes against an individual data sets, and 16,700 observations were included in the crimes against a household data sets.

For our analysis, we created three outcome variables based on the 10 screener questions. This was necessary because the NCVS screeners are overlapping (i.e., several screeners are asking about the same type of crime [U.S. Department of Justice, 2007]). Therefore, we wanted to combine the screeners so that we had a set of independent outcomes. These outcomes were less serious crimes against an individual, more serious crimes against an individual, and crimes against a household. Less serious crimes against an individual include crimes such as theft, simple assault, and robbery. More serious crimes against an individual include aggravated assault and rape or sexual assault. Crimes against a household include vandalism, motor vehicle theft, and household burglary. When determining if each of these outcomes was appropriate for MLCA, we determined that the more serious crimes against an individual outcome was too rare and created data that was too sparse, even when only four waves were cross-classified. Therefore, more serious crimes were not included in our analysis.

A total of 14 potential grouping variables were considered for our analysis, including age, owner/renter, gender, education, race, household size, urbanicity, marital status, frequency of going out in the evening, frequency of shopping, and use of public transportation. Several interview paradata variables were also considered, including mode of interview (face-to-face or telephone), respondent type (self or proxy), and presence of others during the interview. Continuous variables like age were discretized. Variables with extremely small cell frequencies were collapsed into fewer categories. Time-varying grouping variables (e.g., going out, frequency of shopping, use of public transportation, and the interview paradata variables) were summarized into a time-invariant summary grouping variable using the approach of Berzofsky, Biemer, and Kalsbeek (2010). Berzofsky et al. (2010) tested these variables and determined that the summarized versions did not introduce bias and helped improve model fit by increasing the number of available degrees of freedom in the model.

2.3 Model Selection

The process for determining the best model consisted of four major steps. These steps essentially test one or more of the four MLC model assumptions and corrects the model in case the assumptions fail. The four steps are

1. Determine what grouping variables should be included in the model to satisfy the group-homogeneity assumption.
2. Test the Markov assumption and adjust the model for failure of this assumption to hold.
3. Test the time-homogeneous classification error rates assumption and correct the model for failure of this assumption to hold.
4. Test the ICE assumption and adjust the model if any dependencies are found.

Thus, upon completing these four steps, the model will either satisfy all four model assumptions or will have been corrected for any violations of these assumptions.

For all modeling we used LatentGold (Vermunt & Magidson, 2005). LatentGold allows users to account for both the unequal survey weights and the complex survey design, using a pseudo-maximum likelihood method (Pfeffermann, 1993). LatentGold uses both an EM algorithm and a Newton-Raphson algorithm to ensure the global maximum is identified.

In the Step 1, we identified the best grouping variables to help satisfy the homogenous errors assumption. To determine the best grouping variables to satisfy this assumption, we used a combination of a modified forward selection approach and theoretical judgment (i.e., the inclusion of grouping variables based on their theoretical relevance to crime victimization). Under our forward selection approach, we examined each grouping variable individually and selected the variable with the best (smallest) BIC (Schwarz, 1978). The BIC provided the best available measure of model fit because although all models were based on the same data set, some of models were not nested within each other. The forward selection approach was stopped after three grouping variables were identified because it was difficult to achieve model convergence with four or more because of sparseness. Of the remaining variables, we selected two additional variables that may not have been the best empirically, but made theoretical sense for either the structural component or the measurement component. Using a data set with just these five grouping variables and the appropriate indicators, we ran all possible two grouping variable models and chose the model with the smallest BIC (models with three or more grouping variables did not have good fit based on the L^2 statistic). The best two grouping variable model was used in the second step of the model selection process.

In steps 2 and 3 of the model selection process compared different theoretical models to determine which had the best fit. Each model was chosen to test either the first-order Markov assumption or the time-invariant homogenous classification errors assumption. For example, models with a second-order Markov term (i.e., $\pi_{x_3|x_1x_2}^{X_3|X_1X_2}$ and $\pi_{x_4|x_2x_3}^{X_4|X_2X_3}$ parameters were added to the model) were estimated to test the first-order Markov assumption. Also, for the structural component transition probabilities, constrained (i.e., $\pi_{x_2|gx_1}^{X_2|GX_1} = \pi_{x_3|gx_2}^{X_3|GX_2} = \pi_{x_4|gx_3}^{X_4|GX_3}$) and unconstrained (i.e., $\pi_{x_i|gx_{i-1}}^{X_i|GX_{i-1}}$ for i=2, 3, 4 were set free) models were tested to allow for greater flexibility in the measurement component constraints. Furthermore, for the measurement component, constrained (i.e., time-homogeneous error rates), partially constrained (i.e., classification error parameters were assigned to one of two groups with equal classification error probabilities), and unconstrained (i.e., classification errors were set free for all time points) models were estimated to test the time homogeneous classification errors assumption. Moreover, in some models a latent mover-stayer component was estimated in the structural component (Goodman, 1961; Blumen, Kogan, & McCarthy, 1966). In a latent mover-stayer model, the population is partitioned into two latent groups: a mover group for persons who may transition from positive and negative victimization states from time period to time period, and a stayer group for persons who have zero probability of leaving their initial state (e.g., a nonvictim at time 1 who has zero probability of being victimized at times 2, 3, and 4). These models test whether crime victimization behaves in a mover-stayer manner. To test these assumptions, we considered nine different models. These models were

1. Fully constrained transition probabilities and partially constrained classification error probabilities
2. Unconstrained transition probabilities and fully constrained classification error probabilities
3. Unconstrained transition probabilities with a mover-stayer component and fully constrained classification error probabilities

4. Unconstrained transitions probabilities with second-order Markov terms and fully constrained classification error probabilities
5. Unconstrained transition probabilities and partially constrained classification error probabilities
6. Fully constrained transition probabilities and unconstrained classification error probabilities
7. Fully constrained transition probabilities with a mover-stayer component and unconstrained classification error probabilities
8. Unconstrained transition probabilities with a mover-stayer component and partially constrained classification error probabilities
9. Unconstrained transition probabilities with second-order Markov terms and partially constrained classification error probabilities.

Each of the constraints placed on these models were based on a theoretical justification (e.g., relaxing the time-homogeneous errors assumptions make theoretical sense because the error rates increase over time) or a test for one of the MLC model assumptions. Models 1, 6, and 7 assumed constrained transition probabilities. This constraint makes theoretical sense because assuming the probability of a crime occurring from one time period to the next as constant across time points seems plausible. Models 2, 3, 4, 5, 8, and 9 do not constrain the transition probabilities. These models require more parameters to fit, but allow for the transition probabilities to differ. Models 3, 7, and 8 also incorporated a latent mover-stayer variable. This type of model has a good theoretical justification because there is such a large portion of the population that is never victimized over four time periods. Models 4 and 9 included second-order Markov terms to test the first-order Markov assumption. We only included second-order terms without grouping variables to minimize the additional number of parameters added to the model.

For the measurement component, we compared three different types of models: time-homogeneous models, partially constrained models, and unconstrained models. These models relaxed the time-homogeneous classification errors assumption. Models 2, 3, and 4 assumed time-homogeneous classification errors. Models 1, 5, 8, and 9 allowed for partially constrained classification error probabilities. To ensure that these models were identifiable, relaxing constraints in the error component required imposing constraints in the structural component and vice versa. For example, when the transition probabilities in the structural component are fully constrained (model 1), it is possible to relinquish some constraints on the error probabilities. Which constraints were imposed was somewhat driven by theoretical considerations. In the data sets using waves 2, 3, 4, and 5, it seemed plausible to allow wave 2 error probabilities be free and set the remaining time points equal because it was likely that early interviews would have different classification error rates than later interviews. For data sets using waves 3, 4, 5, and 6, we set the first two time points equal as well as the last two time points. Since these data sets spanned the middle of the NCVS panel rotation, it made the most theoretical sense to set the classification error rates equal based on the first and second half of the time points. For the data set using waves 4, 5, 6 and 7, we set the first three time points equal and let the last time point be free. There is some evidence that respondents may provide more accurate responses at their last interview (Meekins, Tucker, & Biemer, 2011); therefore, we allowed wave 7 error probabilities to be unconstrained. However, when the transition probabilities were unconstrained (models 5, 8, and 9), the partial constraint had to be defined as $\pi_{a_1|gx_1}^{A_1|GX_1} = \pi_{a_2|gx_2}^{A_2|GX_2}$ and $\pi_{a_3|gx_3}^{A_3|GX_3} = \pi_{a_4|gx_4}^{A_4|GX_4}$ to ensure an identifiable model (Biemer, 2011). Models 6 and 7 allowed all of the classification error probabilities to be free. As found in Berzofsky et al. (2010) the classification error rates in the NCVS increase over time, and allowing them to all be free allows that trend to be fully realized.

Next the ICE assumption was tested. For each of the nine models the log odds ratio check (LORC, Garret & Zeger, 2000) was used to determine whether the ICE assumption was valid. The LORC was designed for identifying local dependence across multiple indicators in LCA with cross-sectional data (local dependence is the analogous assumption to ICE with cross-sectional data). To ensure that it would perform in a similar manner for longitudinal data, we ran a small simulation study and found that the LORC does properly identify dependencies across time points. If any dependencies were identified, the appropriate direct effect term (i.e., $\pi_{a_i}^{A_i}\pi_{a_j}^{A_j}$ for i=1,…,T, j=1,…,T and $i \neq j$) would be added to the model. However, since we were dealing with longitudinal data, we decided that only direct effects from dependencies involving adjacent time points would be added to the model.

After testing the ICE assumption, the best model was determined by two criteria. First, the model had to have good fit based on the L^2 statistic (i.e., a p-value greater than 0.05). Second, among those models with a good fit, the best model had the smallest BIC. Because non-nested data were being compared, the BIC was the most appropriate statistic for comparing the models.

3. Results

During the first step of model selection, we identified the best grouping variables to help satisfy the homogeneous errors assumption. For less serious crimes against an individual, the three best grouping variables based on our forward selection process were age category, whether one owns their home, and race. For crimes against a household, the best three grouping variables were age category, gender, and whether one owns a home. For both outcomes we chose marital status (single vs. not single) and how often a person goes out in the evening as the two theoretical variables. Models based on just these five grouping variables found that for both outcomes the models with the age category and whether a person owns their home had the best fit based on our criteria.

Table 1 shows the results from the model selection process for less serious crimes against an individual (steps 2 through 4). For each of these models the LORC did not find any dependencies validating the ICE assumption. For each set of waves model 1 (a model with fully constrained transition probabilities and partially constrained classification error probabilities) was considered the best.

Table 1. Fit Statistics for Less Serious Crimes against an Individual Models by Set of Waves Included

Model	Npar	df	L^2	BIC(L^2)	p-value
Waves 2, 3, 4, and 5					
Model 1	42	48	59.1	-426.96	0.130
Model 2	54	36	50.2	-314.36	0.058
Model 3	56	34	28.8	-315.55	0.730
Model 4	56	34	48.6	-295.68	0.050
Model 5	66	24	41.6	-201.43	0.014
Model 6	66	24	33.7	-209.34	0.090
Model 7	67	23	32.3	-200.64	0.095
Model 8	68	22	18.9	-203.91	0.650
Model 9	68	22	34.8	-187.99	0.041
Waves 3, 4, 5, and 6					
Model 1	42	48	57.0	-429.19	0.180
Model 2	54	36	50.8	-313.77	0.052
Model 3	56	34	45.3	-299.03	0.093
Model 4	56	34	45.1	-299.28	0.097
Model 5	66	24	39.2	-203.89	0.026
Model 6	66	24	37.5	-205.53	0.039
Model 7	67	23	35.3	-197.68	0.049
Model 8	68	22	35.8	-187.03	0.032
Model 9	68	22	33.1	-189.72	0.060
Waves 4, 5, 6, and 7					
Model 1	42	48	55.6	-430.54	0.210
Model 2	54	36	45.4	-319.23	0.140
Model 3	56	34	31.0	-313.33	0.620
Model 4	56	34	37.6	-306.79	0.310
Model 5	66	24	36.7	-206.34	0.046
Model 6	66	24	32.4	-210.69	0.120
Model 7	67	23	35.5	-197.44	0.046
Model 8	68	22	28.2	-194.66	0.170
Model 9	68	22	32.1	-190.72	0.076

Table 2 shows the results from the model selection process (steps 2 though 4) for crimes against a household. For each of these models the LORC did not find any dependencies validating the ICE assumption. For each set of waves model 1 (a model with fully constrained transition probabilities and partially constrained classification error probabilities) was considered the best.

Table 2. Fit Statistics for Crimes against a Household Models by Set of Waves Included

Model	Npar	df	L^2	BIC(L^2)	p-value
Waves 2,3,4, and 5					
Model 1	42	48	52.8	-411.37	0.290
Model 2	54	36	47.5	-300.63	0.095
Model 3	56	34	41.4	-287.33	0.180
Model 4	56	34	43.9	-284.88	0.120
Model 5	66	24	40.3	-191.77	0.020
Model 6	66	24	36.3	-195.77	0.051
Model 7	67	23	35.6	-186.76	0.045
Model 8	68	22	34.6	-178.14	0.043
Model 9	68	22	36.3	-176.41	0.028
Waves 3, 4, 5, and 6					
Model 1	42	48	63.9	-400.88	0.062
Model 2	54	36	46.9	-301.63	0.100
Model 3	56	34	46.9	-282.30	0.070
Model 4	56	34	42.3	-286.84	0.150
Model 5	66	24	33.8	-198.54	0.088
Model 6	66	24	36.1	-196.25	0.053
Model 7	67	23	39.9	-182.83	0.016
Model 8	68	22	36.0	-176.95	0.030
Model 9	68	22	29.9	-183.13	0.120
Waves 4, 5, 6, and 7					
Model 1	42	48	53.3	-411.65	0.280
Model 2	54	36	39.5	-309.19	0.310
Model 3	56	34	36.3	-293.10	0.360
Model 4	56	34	31.4	-298.00	0.600
Model 5	66	24	33.4	-199.07	0.095
Model 6	66	24	31.1	-201.34	0.150
Model 7	67	23	34.9	-187.94	0.054
Model 8	68	22	25.1	-188.01	0.290
Model 9	68	22	26.3	-186.84	0.240

4. Conclusions

This paper presented the methods we used model classification error in the NCVS. We use MLCA to model the classification error rates. While our approach is tailored specifically to the NCVS it can easily be generalized to other panel surveys. The approach tests for and corrects any model assumption violation through the use of grouping variables, comparing different models, and using the LORC. For the NCVS, we found that a model that constrains the transition probabilities and partially constrains the classification error probabilities best fits the data and produces the best estimates of classification error. A future paper will present the estimated classification error rates from our MLCA.

References

Berzofsky, M. E., Biemer, P. P., & Kalsbeek, W. D. (2010). Time varying covariates in Markov latent class analysis: Some problems and solutions. In *Proceedings of the ASA Section on Survey Methodology, Joint Statistical Meetings*, 3785–3799. Vancouver, BC.

Biderman, A. D. (1967). Surveys of population samples for estimating crime incidence. Ann. *American Academy of Political and Social Science, 374*, 16–33.

Biderman, A. D., Johnson, L. A., McIntyre, J., & Weir, A. W. (1967). Field survey I: Report on a pilot study in the District of Columbia on victimization and attitudes toward law enforcement. President's Commission on Law Enforcement and Administration of Justice Field Survey I. Washington, DC: US Government Printing Office.

Biderman, A. D. & Lynch, J. P. (1981). *Recency bias in data on self-reported victimization*. Paper presented at the proceedings of the Social Statistics Section, American Statistical Association, Alexandria, VA.

Biderman, A. D. & Cantor, D. (1984). *A longitudinal analysis of bounding, respondent conditioning and mobility as sources of panel bias in the National Crime Survey*. Paper presented at the proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria, VA.

Biderman, A. D., Cantor, D. & Reiss, A. J., Jr. (1982). *A quasi-experimental analysis of personal victimization reporting by household respondents in the National Crime Survey*. Paper presented at the proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria, VA.

Biderman, A. D., Cantor, D. & Reiss, A. J., Jr. (1984). *Procedural biases and conceptual incongruities in operationalizations of the distinction between household and personal victimization*. Unpublished memorandum, Bureau of Social Science Research, Washington, DC.

Biemer, P. (2004). An analysis of classification error for the revised Current Population Survey employment questions. *Survey Methodology*, *30*(2), 127–140.

Biemer, P. P. (2000). *An application of Markov latent class analysis for evaluating reporting error in Consumer Expenditure Survey Screening Questions* (prepared for the U.S. Bureau of Labor Statistics). Research Triangle Park, NC: RTI International.

Biemer, P.P. (2011). *Latent Class Analysis of Survey Error*. Hoboken, NJ: John Wiley & Sons.

Biemer, P., & Bushery, J. (2001). On the validity of Markov latent class analysis for estimating classification error in labor force data. *Survey Methodology, 26*(2), 136–152.

Blumen, I., Kogan, M., & McCarthy, P. J. (1966). Probability models for mobility. In P. F. Lazarsfeld, & N. W. Henry (Eds.), *Readings in mathematical social science,* 318-334. Cambridge, MA: MIT Press

Bushery, J. M. (1981). *Results of the NCS reference period research experiment*. Unpublished memorandum, U.S. Department of Commerce, Bureau of the Census, Washington, DC.

Garrett, E. S., & Zeger, S. L. (2000). Latent class model diagnosis. *Biometrics, 56*(4), 1055-1067.

Goodman, L. A. (1961). Statistical methods for the Mover-Stayer model. *Journal of the American Statistical Association. 56*(296), 841-868.

Goodman, L. A. (1974). Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. *Biometrika*, *61*, 215–231.

Groves, R. M. & Couper, M. P. (1992). *Correlates of nonresponse in personal visit surveys*. Paper presented at the proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria, VA.

Groves, R. M. & Couper, M. P. (1993). *Multivariate analysis of nonresponse in personal visit surveys*. Paper presented at the proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria, VA.

Hart, T.C., (2006). *Respondent fatigue in self-report victim surveys: Examining a source of nonsampling error from three perspectives* (Doctoral dissertation). Retrieved from http://scholarcommons.usf.edu/etd/2551.

Hart, T. C., Rennison, C. M., & Gibson, C. (2005). Revisiting respondent 'fatigue bias' in the National Crime Victimization Survey. *Journal of Quantitative Criminology, 21*(3), 345-363.

Kalton, G., Kasprzyk, D., & McMillen, D. (1989). Nonsampling errors in Panel Surveys. In G. Kalton, D. Kasprzyk & D. McMillen (Eds.). *Panel Surveys*. New York: Wiley & Sons.

Law Enforcement Assistance Administration (LEAA) (1972). *San Jose methods test o known crime victims*. Statistics Technical Report No. 1. Washington, DC: Author.

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, & J. Clausen (Eds.), *Studies on Social Psychology in World War II, Vol. 4, Measurement and Prediction*. Princeton, NJ: Princeton University Press.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Lohr, S. L. (1999). Sampling: Design and Analysis. Pacific Groves, CA: Brooks/Cole Publishing Company.

Lynch, J. P., Berbaum, M. L., & Planty, M. (2002). Investigating repeated victimization in the NCVS (report prepared for the U.S. Department of Justice, National Institute of Justice. NCJ 193415). Retrieved from http://www.ncjrs.gov/pdffiles1/nij/grants/193415.pdf.

Meekins, B., Tucker, C., & Biemer, P. (2011). Incorporating nonresponse propensity in a Markov latent class measurement error model of Consumer Expenditure. Paper presented at the 2011 International Total Survey Error Workshop, Quebec City, Quebec.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Survey Review*, *61*, 317–337.

Poulsen, C. A. (1982), Latent structure analysis with choice modeling applications. Aarhus School of Business Administration and Economics, Aarhus, Denmark.

Rand, M., & Catalano, S. (2007). Criminal victimization, 2006. U.S. Department of Justice, Office of Justice Programs. *Bureau of Justice Statistics Bulletin*, NCJ 219413. Retrieved from http://www.rainn.org/pdf-files-and-other-documents/News-Room/press-releases/2006-ncvs-results/NCVS%202006-1.pdf.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2): 461–464.

Tucker, C., Biemer, P. P., & Meekins, B. (2011). Estimating underreporting of consumer expenditures using Markov latent class analysis. *Survey Research Methods*, *5* (2), 39 – 51.

U.S. Department of Justice, Bureau of Justice Statistics (2004). The nation's two crime measures. NCJ 122705. http://bjs.ojp.usdoj.gov/content/pub/pdf/ntcm.pdf

U.S. Department of Justice, Bureau of Justice Statistics. (2007). *National Crime Victimization Survey longitudinal file, 1995-1999* [Computer Bibliographic Citation: file]. Conducted by U.S. Department of Commerce, Bureau of the Census. ICPSR04414-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor]. doi:10.3886/ICPSR04414. Retrieved from http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/04414/detail.

Van de Pol, R., & De Leeuw, J. (1986). A latent Markov model to correct for measurement error. *Sociological Methods and Research, 15*, 118–141.

Van de Pol, F., & Langeheine, R. (1990), Mixed Markov latent class models. In C. C. Clogg, (Ed.), *Sociological Methodology* (pp. 213–247). Blackwell: Oxford.

Vermunt, J. K., & Magidson, J. (2005). *Technical guide to Latent Gold 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations.

Wiggins, L. M. (1973). *Panel analysis, latent probability models for attitude and behavior processing*. Elsevier SPC: Amsterdam.

Yan, T. (2008). Panel conditioning: A cross-cultural perspective. *Proceedings of the 3mc 2008 conference*. Retrieved from http://www.csdiworkshop.org/pdf/3mc2008_proceedings/session_37/yan_oct08.pdf on March 21, 2011.