Time Varying Grouping Variables in Markov Latent Class Analysis: Some Problems and Solutions

Marcus E. Berzofsky[1] & Paul P. Biemer[1,2]

[1] RTI International, Research Triangle Park, North Carolina, USA

[2] Odum Institute, University of North Carolina, Chapel Hill, North Carolina, USA

Correspondence: Marcus E. Berzofsky, RTI Internatinal, Research Triangle Park, NC, 27709, Canada.
Tel: 1-919-316-3752. E-mail: berzofsky@rti.org

Abstract

Markov latent class analysis (MLCA) is a modeling technique for panel or longitudinal data that can be used to estimate the classification error rates (e.g., false positive and false negative rates for dichotomous items) for discrete outcomes with categorical predictors when gold-standard measurements are not available. Because panel surveys collect data at multiple time points, the grouping variables in the model may either be time varying or time invariant (static). Time varying grouping variables may be more correlated with either the latent construct or the measurement errors because they are measured simultaneously with the construct during the measurement process. However, they generate a large number of model parameters that can cause problems with data sparseness, model diagnostic validity, and model convergence. In this paper we investigate whether more parsimonious grouping variables that either summarize the variation of the time varying grouping variable or assume a structure that lacks memory of previous values of the grouping variables can be used instead, without sacrificing model fit or validity. We propose a simple diagnostic approach for comparing the validity of models that use time-invariant summary variables with their time-varying counterparts. To illustrate the methodology, this approach is applied to data from the National Crime Victimization Survey (NCVS) where greater parsimony and a reduction in data sparseness were achieved with no appreciable loss in model validity for the outcome variables considered. The approach is generalized for application to essentially any MLCA using time varying group variables and its advantages and disadvantages are discussed.

Keywords: Markov Latent Class Analysis, National Crime Victimization Survey, time varying grouping variables, measurement error, classification error, screener questions

1. Introduction

Markov latent class analysis (MLCA) is a modeling technique that can be used to assess the classification error in survey items from a panel or longitudinal survey (Wiggins, 1973; Poulsen, 1982; Van de Pol & de Leeuw, 1986; Van de Pol & Langeheine, 1990). Often in surveys, there is no gold-standard (or error-free) data source available to evaluate the error in survey responses. Therefore, gold-standard techniques cannot be used. Furthermore, even when so-called gold-standard sources—like administrative records or tests using hair or blood samples—do exist, studies have found that these sources are flawed as well, and it is not appropriate to assume they are error free (Visher & McFadden, 1991). MLCA does not require a gold standard to estimate both the true prevalence of the latent variable and the corresponding measurement error. However, one weakness of the methodology is the limitation on the number of covariates that can be included in the model, especially for analyzing more than three time points. A number of authors (see, for example, Biemer & Berzofsky, 2011; Vermunt, Langeheine, & Bockenholt, 1999) have shown that data sparseness can be quite problematic in MLCA and methods for constructing more parsimonious models without sacrificing model validity are needed.

To improve the fit and validity, covariates referred to as *grouping variables* are incorporated into the model. A grouping variable can be any categorical manifest variable that is not intended to be used as an indicator for the latent variables but is correlated to either the outcome of interest or the classification error (Biemer, 2011). In this way, grouping variables are analogous to covariates in a standard regression analysis. Examples of grouping variables include age groups, race/ethnicity, or interview mode. In addition to helping with model identifiability and improving model fit, grouping variables can be used to test hypotheses about the causes and correlates of measurement error and prevalence variations in the population latent variables. Because panel surveys are longitudinal in nature, grouping variables can be either *time invariant (or static)* or *time varying*. Time-invariant grouping variables do not change over time or change as a simple

linear function of time for all respondents (e.g., race/ethnicity or age). Time-invariant grouping variables are denoted by a single letter starting with *G*. Time varying grouping variables change over time in a potentially nonlinear manner. For example, a person's current health status or the mode of the data collection may change over time, in different patterns for each respondent. Time varying grouping variables are denoted by G_t $t=1,\ldots,T$, where t represents the wave number to which that grouping variable corresponds.

The MLC model may be viewed as consisting of 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.

In latent class analysis, a single latent variable is used to represent the true value of the dependent variable. Since MLCA deals with panel or longitudinal data, there is a latent variable to represent the value of the dependent variable at each time point. For our work, both the dependent variable and the latent construct have the same known, fixed number of levels. The dependent variable is defined the same way at each time point, except the reference period shifts at each time point. For example, the Current Population Survey (CPS) is a monthly panel survey that tracks employment status over time and divides the population into one of three categories: employed, unemployed, or not in the labor force. Thus, if the true employment status is being modeled, there is a separate latent variable representing a person's true employment status during each month in the survey.

For the MLC model to be identifiable, a minimum of three time points are required (Biemer, 2011). In a three time–point model, which will be the focus of this paper, the latent variables are represented by X_1, X_2, and X_3, where X_1 is the latent value for the first time period, X_2 is the latent value for the second time period, and X_3 is the latent value for the third time period. Because the latent variables occur chronologically, each latent variable is dependent on the previous values (i.e., X_2 given X_1, hereby notated as $X_2|X_1$ and $X_3|X_2X_1$). In a model with only three time points, four key assumptions must be made for the model to be identifiable. Details of these assumptions are found in Biemer (2011).

The likelihood kernel for the MLC model, with no grouping variables, is

$$\pi_{abc}^{ABC} = \sum_{x_1}\sum_{x_2}\sum_{x_3}\left(\pi_{x_1}^{X_1}\pi_{x_2|x_1}^{X_2|X_1}\pi_{x_3|x_2}^{X_3|X_2}\right)\left(\pi_{a_1|x_1}^{A_1|X_1}\pi_{a_2|x_2}^{A_2|X_2}\pi_{a_3|x_3}^{A_3|X_3}\right) \tag{1}$$

where $\pi_{x_1}^{X_1}\pi_{x_2|x_1}^{X_2|X_1}\pi_{x_3|x_2}^{X_3|X_2}$ is the structural component, and $\pi_{a_1|x_1}^{A_1|X_1}\pi_{a_2|x_2}^{A_2|X_2}\pi_{a_3|x_3}^{A_3|X_3}$ is the measurement component. Figure 1

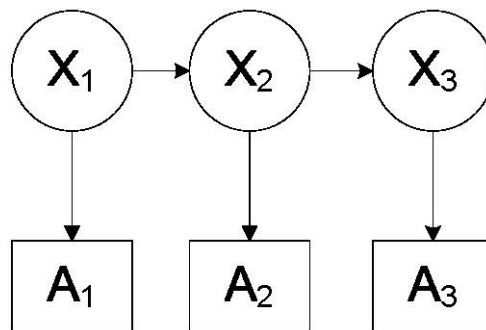presents the path diagram for the basic MLC model.

Figure 1. Path diagram for MLC model with three time points and no grouping variables, with a measurement component that assumes time-homogeneous classification errors (i.e., $A_1|X_1=A_2|X_2=A_3|X_3$).

Changes in a person's actions or behavior that occur over time may predict both the individual's current true and observed state. Therefore, it may be important to fully capture these variations in an MLC model by using time varying grouping variables. However, there are three major costs associated with using time varying grouping variables that are not associated with *time-invariant summary grouping variables*. As a grouping variable, a time-invariant summary grouping variable tries to capture all the information contained in a set of time varying grouping variables in a single static variable. Because the time-invariant summary variable is static across time, like a traditional time-invariant grouping variable, it is denoted by G. For example, mode of data collection (e.g., in person or telephone) is a time varying grouping variable because it can vary across waves. A time-invariant summary grouping variable for mode of data collection may create a category for all respondents who always responded in person, another category for those who always responded via telephone, and a third category for those who used both modes of response.

While time varying grouping variables allow the full modeling of changes over time, they have four costs which make the use of time-invariant summary grouping variables an attractive alternative. The first cost associated with time varying grouping variables is the number of degrees of freedom used. In an MLC model, the number of available degrees of freedom equals the number of cells in the data frequency table minus the number of parameters in the model. Time varying grouping variables maximize the number of parameters in the model thereby minimizing the possible number of degrees of freedom. The second cost associated with using time varying grouping variables is an increased potential for *data sparseness*. Data sparseness is when there is a preponderance of cells in the data frequency table with no or very few observations. Data sparseness can lead to a model with weak identifiability (Biemer, 2011). This was illustrated in Vermunt et al. (1999), where the authors had to collapse a time varying grouping variable representing a student's grade at a particular time to two levels to protect against data sparseness, thus, losing some of the information contained in the time varying grouping variables. Third, a grouping variable based on a respondent's behavior over time may make more theoretical sense than a grouping variable based on the respondent's status at a particular point in time (see Biemer and Bushery (2001) and Biemer (2004) for illustration). Fourth, survey methodologists and others who are particularly interested in the measurement process may consider more elaborate measurement models to test theories regarding the nature and sources of classification error. For this purpose, more parsimonious representations of the model covariates are desirable since these will allow the analyst to maximize the number of possible grouping variables that can be included in the model before the data table becomes too sparse. Time varying grouping variables may better explain the structural and measurement components, but the number of degrees of freedom they require is at odds with the goal of more parsimonious grouping variables. As such, for a particular data set that is being analyzed, it is important to know whether a reduced or time-invariant summary version of a time varying grouping variable will sufficiently capture the information contained in the time varying grouping variable.

In this paper, we consider whether, when parsimony is taken into account, a model that substitutes a time-invariant summary grouping variable for a set of time varying grouping variables will fit the data as well and without sacrificing model validity. We propose a process for comparing different summarizations of the time varying grouping variable as a time-invariant summary grouping variable. Then, we apply that process to the National Crime Victimization Survey (NCVS). We examine how different time-invariant summary variables impact model fit and validity.

2. Method

Prior to testing our hypotheses, two initial steps were conducted. First, we reviewed the survey instrument to determine all possible time varying grouping variables and potential latent outcomes. Second, for each latent outcome, we assessed whether an MLCA could be conducted. This assessment was done by determining the condition index for a basic MLC model for that outcome. The condition index denotes whether the MLC model is likely to have 'weak' identifiability. A condition index over 5,000 usually indicates identifiability issues, and an MLCA for that outcome should not be conducted (Biemer, 2011).

To determine if a time-invariant summary grouping variable can be used instead of a set of time varying grouping variables in an MLC model, we propose a nine-step process to compare different summary variables of a time varying grouping variable to a model that uses the time varying grouping variables. To ensure that our process identifies the best fitting model, we propose comparing models generated with a time-invariant summary variable to models with reduced forms of the time varying variables. The steps in the process are as follows:

1. Construct a time-invariant summary grouping variable, G, from the time varying variables.
2. Create a synthetic data set by fitting the fully saturated model (defined below). This model will be referred to as the "population model."
3. Fit a model that replaces the time varying grouping variables (i.e., $G_1G_2G_3$) with the time-invariant summary grouping variable (G) and compare the fit and model estimates.
4. Repeat this process using alternate forms of G.

5. The best form of G is the one that provides the least loss of model fit and whose estimates agree most closely with those of the population model.
6. Fit a model with a reduced form of the time varying grouping variables (i.e., a more parsimonious model compared to the population model) and compare the fit and model estimates.
7. Repeat process for alternative reduced forms of the time varying grouping variables.
8. The best reduced form of the time varying grouping variables is the most parsimonious model that provides the least loss of model fit and whose estimates agree most closely with the population model.
9. Compare the best time-invariant summary model to the best reduced time varying

In the first step of this process, one or more alternate constructions of the time-invariant summary variable are proposed. The goal of this step is to develop variables that are theoretically plausible for the outcome variable being modeled. Furthermore, we want to consider groupings that split the population into homogeneous groups related to their probability of measurement error to satisfy the time homogeneous classification errors assumption.

In the second step, we construct the population model using a synthetic data set based on the fully saturated time varying grouping variable model. The population model best describes the population from which the data come based on the single grouping variable being considered. Although this model may be over parameterized, it serves as the best representation of the population in the absence of knowing the true population model that would include other grouping variables. The likelihood kernel for the population model for cell $(g_1,g_2,g_3,a_1,a_2,a_3)$ of the $G_1 \times G_2 \times G_3 \times A_1 \times A_2 \times A_3$ cross-classification table in terms of X_1, X_2, and X_3 is written as:

$$\pi_{g_1g_2g_3a_1a_2a_3}^{G_1G_2G_3A_1A_2A_3} = \sum_x \pi_{g_1g_2g_3}^{G_1G_2G_3} \pi_{x_1|g_1g_2g_3}^{X_1|G_1G_2G_3} \pi_{x_2|x_1g_1g_2g_3}^{X_2|X_1G_1G_2G_3} \pi_{x_3|x_2g_1g_2g_3}^{X_3|X_2G_1G_2G_3} \pi_{a_1|x_1g_1g_2g_3}^{A_1|X_1G_1G_2G_3} \pi_{a_2|x_2g_1g_2g_3}^{A_2|X_2G_1G_2G_3} \pi_{a_3|x_3g_1g_2g_3}^{A_3|X_3G_1G_2G_3}$$

(2)

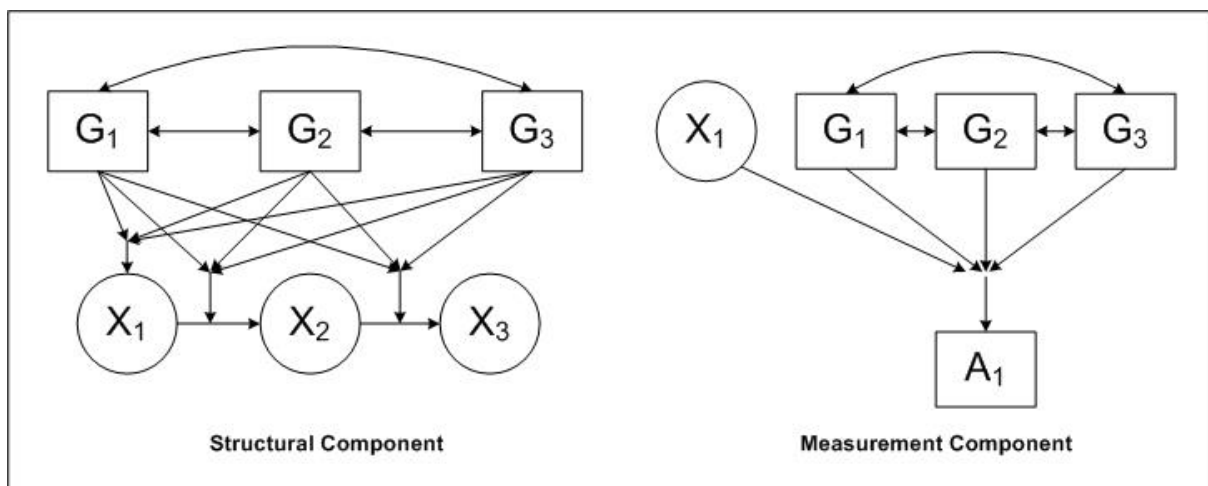The path diagram for the population model is illustrated in Figure 2.

Figure 2. Path diagram for the population model that assumes time-homogeneous classification (i.e., $A_1|X_1G_1G_2G_3 = A_2|X_2G_1G_2G_3 = A_3|X_3G_1G_2G_3$). Double arrows indicate a three-way interaction (i.e., $G_1G_2G_3$). The figure only shows the path diagram for the measurement component of indicator A_1. The diagram for the measurement component is similar for the other two indicators: $A_2|X_2G_1G_2G_3$ and $A_3|X_3G_1G_2G_3$.

The population model is used to generate a synthetic data set. Synthetic data sets are artificial data from a *known* model that simulates the data from the empirical data set being analyzed (Biemer, 2011; Biemer & Berzofsky, 2011). Replacing the actual data with the synthetic population model data has the effect of removing all sources of variation from the data, except variation induced by the variables of interest in the analysis. Thus, the process of determining a more parsimonious representation of the time varying grouping variable can focus on the grouping variable of interest and its effects on the latent and indicator variables. To do otherwise would introduce the potential that the spurious effects of other grouping variables could influence the decision regarding more parsimonious specifications of the grouping variable of interest.

In the third step, using the synthetic data, we fit a model that uses a time-invariant summary variable instead of the time varying grouping variables. All models were computed using ℓEM (Vermunt, 1997). Models were run without survey weights or accounting for the complex design. As described in Vermunt (2002) and Biemer (2011), we used a two-step

estimation process. In the first step, unweighted data are used to estimate the measurement parameters. Because the number of responses that can be obtained is infinite, both Vermunt and Biemer argue that using finite population-based weights is inappropriate for estimating the classification parameters. In the second step, the measurement parameters are fixed based on the results of the first step, and data that are rescaled to the weighted distribution are used to estimate the structural estimates. Since we are mainly interested in the measurement component, we only carried out the first of these two steps for this analysis. However, we do present estimates from the structural component in this paper. Therefore, it is possible that the structural estimates are biased. However, we conducted a preliminary analysis of the weights' impact on the structural component estimates using a simple model, and we found that the weighted and unweighted estimates were not substantively different.

In the fourth and fifth step, we fit a model with each time-invariant summary variable that we developed. For a model to be acceptable, it had to meet all of the following criteria:

1. Index of dissimilarity (D) less than 0.05
2. P-value for the L^2 statistic greater than 0.05
3. Absolute relative bias for the false negative rate compared to the population model 10% or less.

We chose these criteria based on our particular analytic goals. Under different circumstances these criteria can be modified to better fit a different set of goals. For example, if the false positive rate is more important, it can be incorporated into the criteria. If more than one model met all of these criteria, the model with the smallest Bayesian Information Criteria (BIC; Schwarz, 1978) was determined to be the best. To be able to use the BIC as a fit statistic, the same data table was used. For models that used time-invariant summary grouping variables, a data table was created that appears to have $k*$ncells cells, where k is the number of levels for the time-invariant summary grouping variable and ncells is the number of cells the data table has using time varying grouping variables. However, in actuality, many of the new cells are structural zeros (Biemer & Berzofsky, 2011). Structural zeros occur when a value for a time-invariant summary grouping variable cannot occur for a particular set of time varying grouping variable values. For example, for a 2-level time varying grouping variable (1=yes, 2=no) and a time-invariant summary grouping variable, where 1 means the respondent said "yes" in all three time points, a structural zero occurs when the value of the time-invariant summary grouping variable takes the value of 1 and the time varying grouping variables have at least one "no" response in one of the three time periods (e.g., 112). To properly compute the likelihood, a weight statement is required to exclude the structural zeros.

The index of dissimilarity is defined as

$$D = \frac{\sum_{abc} |n_{abc} - \hat{m}_{abc}|}{2n} \tag{3}$$

where n_{abc} denotes the observed frequency for cell (a,b,c), $\hat{m}_{abc}$ denotes the estimated frequency under the hypothesized model, and n is the sample size. The index of dissimilarity represents the smallest proportion of observations that would need to move to other cells for the model to fit the data perfectly (Biemer, 2011). Based on this definition, we expect the most elaborate model to have the smallest index of dissimilarity and the most parsimonious model to have the largest. As such, we are not interested in ranking the models by the index of dissimilarity, but rather using it to verify that all the models have reasonable fit. Biemer (2011) suggests that models with an index of dissimilarity less than 5% have reasonable fit. Thus, if a model has an index of dissimilarity greater than 5%, we will not consider it a good option for achieving our analysis goals, even if it appears to be better based on other measures.

The L^2 statistic is defined as

$$L^2 = 2\sum_i n_i \log\left(\frac{n_i}{\hat{m}_i}\right) \tag{4}$$

where n_i is the observed frequency in cell i, and $\hat{m}_i$ is the estimated expected frequency in cell i. The L^2 has a chi-square distribution with degrees of freedom equal to the number of possible cells in the data table minus the number of model parameters. A p-value over 0.05 indicates a good fitting model.

To assess potential bias caused by using a model with a time-invariant summary grouping variable, we compared the parameter estimates from each time-invariant summary model to the estimates from the population model. In other words, bias is calculated as

$$B\left(p_{POP}\right) = p_j - p_{POP} \tag{5}$$

where p_{pop} is the parameter estimate from the population model, and p_j is the parameter estimate from the jth alternative time-invariant summary grouping variable. The absolute relative bias (ARB) is the absolute bias divided by the parameter estimate from the population model. In other words,

$$\frac{\left|B\left(p_{POP}\right)\right|}{p_{POP}} \tag{6}$$

The BIC is defined as

$$BIC\left(\mathcal{L}\right) = -2\log\mathcal{L} + npar \times \log(n) \tag{7}$$

where $\mathcal{L}$ is the maximum likelihood for the model. The BIC penalizes models with more parameters, and, therefore, the model with the smallest BIC is the most parsimonious model with the best fit. Since the BIC is dependent of the maximum likelihood for that data, the smallest BIC is relative to the particular data set, and, therefore, there is no set range by which a BIC is acceptable.

Steps 6 – 8 follow a similar approach to compare the reduced time varying grouping variable models. In sixth step, we define a set of models that use the time varying grouping variables, but in a more reduced parsimonious fashion than the population model. For example, instead of using the full interaction of time varying grouping variables at each time point (i.e., $G_1G_2G_3$) only the time varying grouping variable from the particular time point may be used. We used a theoretical basis for determining which reduced models we would test. Therefore, a different empirical data set may lead one to create different reduced models. The purpose of these models is to determine if a more parsimonious model that used time varying grouping variables can fit the data and have less bias than a time-invariant summary model. In ninth step, we compared the best time-invariant summary model to the best reduced time varying grouping variable model. For this step, we considered two aspects of the model fit. First, we compared the BIC of the two models. Second, we considered the absolute and relative amount of data sparseness. We defined sparse data as any cell in the data table that had five or fewer observations.

3. Results

3.1 Overview of the NCVS

3.1.1 Sample Design

The NCVS is a household survey that measures crime victimization rates in the United States. The survey uses a multistage probability design to make inferences to all persons 12 years old or older in the United States (U.S. Department of Justice, 2007). A sample of 50,000 households is selected every 6 months and all persons aged 12 years or older in a sampled household are interviewed. The survey uses a rotating panel design in which each household is surveyed every 6 months and remains in the field for 3 and a half years. The reference period for each interview is the past 6 months. The NCVS achieves a 90% response rate.

3.1.2 Conducting an MLCA with the NCVS

For the NCVS the latent variable of interest is whether a particular crime occurred during the past 6 months. These latent constructs are measured through a series of *screener questions*. Screener questions are a short set of questions that help a respondent remember whether a particular event occurred during the reference period (see, for example, Biemer, 2000). The NCVS includes 10 screener questions to probe about crime victimization. However, because these screener questions ask about an overlapping set of crimes, we collapsed them into three latent constructs: victim of a less serious individual crime, victim of a more serious individual crime, and victim of a household crime. The construct of less serious crimes against an individual include crimes such as theft, simple assault, and robbery. The construct of serious crimes against an individual include aggravated assault and rape or sexual assault. The construct of crimes against a household include vandalism, motor vehicle theft, and household burglary.

Prior to conducting our proposed analysis, we determined if any of the models had "weak" identifiability. We did this by running a model with no grouping variables using the synthetic data and calculating the condition number—the ratio of the largest and smallest information matrix eigenvalues (Biemer, 2011). Condition numbers over 5,000 indicate weak identifiability because of large parameter standard errors. Less serious individual crimes and household crimes had condition numbers under 1,000. However, more serious individual crimes had a condition number of 14,650. A review of

the parameter standard errors showed that they were larger than the estimates themselves (e.g., the model estimated that P(X=1)=0.085, but with a standard error of 0.1082). Therefore, no further analysis of grouping variable with the more serious individual crimes was conducted. The remainder of the paper will only discuss less serious individual crimes and household crimes.

To conduct an MLCA, data from all rotation group waves are needed. The most recent public use file containing all waves for a set of respondents is the National Crime Victimization Longitudinal File, 1995–1999 (U.S. Department of Justice, 2007). This file has data from three rotation groups that were released in the third quarter of 1995, the first quarter of 1996, and the third quarter of 1996. These rotation groups contained 26,345 households and 66,706 unique respondents.

In this analysis, we used the first three waves after the bounding interview. Furthermore, we only included respondents that completed the screener questions in all three waves and provided an answer to the time varying grouping variable being tested in each wave. Based on these conditions, the number of respondents used for analysis ranged between 27,845 and 28,000 for the individual-level screeners, and the number of respondents used for analysis ranged between 16,150 and 17,025 for the household-level screeners. Furthermore, in this analysis, we ignored the survey design and assumed the data came from a simple random sample. As noted earlier, we conducted a preliminary analysis of the impact of survey weights and found that it negligibly impacted the parameter estimates.

3.1.3 Time Varying Grouping Variables in the NCVS

Through a review of the NCVS survey instrument we determined that the NCVS survey consists of over 20 different variables that can be used as a grouping variable. These variables consist of respondent characteristics, information from the sampling frame, and interview paradata. Of these variables, six items are time varying: three respondent characteristics and three paradata items.

The survey items used as time varying grouping variables include how often the respondent went out in the evening over the past 6 months, how often the respondent went shopping over the past 6 months, and how often the respondent used public transportation over the past 6 months. For each of these items the respondent could answer "don't know" or one of five responses ranging from "never" to "almost every day (or more frequently)." For purposes of this analysis, the respondent characteristic items were collapsed based on the distribution of the data. The frequency of going out in the evening and frequency of going shopping were collapsed to three levels: 1=almost every day, 2=at least once a week, and 3=once a month or less. The use of public transportation was collapsed into two categories: 1=at least once in the past 6 months and 2=never. For all three of these items, responses of "don't know" were treated as missing.

The paradata items include three 2-level grouping variables. These variables are the interview mode (face-to-face or telephone), the respondent type (self-report or proxy), and whether the respondent was alone while taking the interview.

3.2 Development of Alternative Models

3.2.1 Development of Time-Invariant Summary Models

Three time-invariant summary grouping variables were defined based on a theoretical evaluation of how the grouping variables interact with the outcomes of interest. The model for each of these time-invariant summary grouping variables has the same likelihood kernel for cell (g,a_1,a_2,a_3) of the $GxA_1xA_2xA_3$ cross-classification table in terms of X_1, X_2, and X_3 may be written as

$$\pi_{ga_1a_2a_3}^{GA_1A_2A_3} = \sum_{x_1}\sum_{x_2}\sum_{x_3} \pi_g^G \pi_{x_1|g}^{X_1|G} \pi_{x_2|gx_1}^{X_2|GX_1} \pi_{x_3|gx_2}^{X_3|GX_2} \pi_{a_1|x_1}^{A_1|X_1} \pi_{a_2|x_2}^{A_2|X_2} \pi_{a_3|x_3}^{A_3|X_3} \tag{8}$$

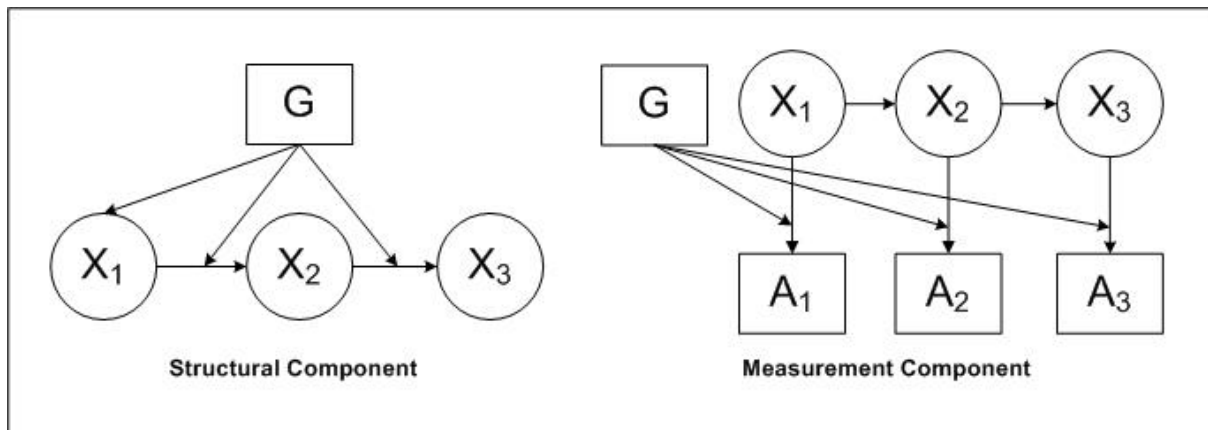Figure 3 illustrates the path diagram for these models.

Figure 3. Path diagram for models with time-invariant summary grouping variables that assume time-homogeneous classification errors (i.e., $A_1|X_1 G = A_2|X_2 G = A_3|X_3 G$).

The first time-invariant summary grouping variable considered is the *mover-stayer, time-invariant summary variable*. This reduced time-invariant summary variable is based on the concept of a manifest mover-stayer model (see Goodman [1961] or Blumen et al. [1966]), where we split the population based on whether their status of the grouping variable stays constant across time or changes. For those who stay constant, we create a level for each level of the time varying grouping variable. Those who move levels across time are placed in a single category. Therefore, the mover-stayer variable has K+1 levels, where K is the number of levels the time-invariant summary variable has. To illustrate, we will use the "frequency of using public transportation." The move-stayer variable for "using public transportation," which has two levels (1=used at least once in the past 6 months, 2=never), will have three levels as follows:

$$G = \begin{cases} 1 & \text{if } G_1 = 1, \ G_2 = 1, \ G_3 = 1 \\ 2 & \text{if } G_1 = 2, \ G_2 = 2, \ G_3 = 2 \\ 3 & \text{if mixed response pattern} \end{cases}$$

In other words, G=1 is the positive stayer group, G=2 is the negative stayer group, and G=3 is the mover group. A model using this time-invariant summary grouping variable is denoted as M(S1) in this paper, where S1 represents the mover-stayer time-invariant model. This summarization makes theoretical sense because it is plausible that those who always have the same behavior or those who have inconsistent behavior will have similar outcome probabilities and similar classification error probabilities.

The second time-invariant summary variable that we considered is an *extreme behavior variable*. This variable is based on the theoretical concept that those on the extremes act in a similar manner, while those who usually act in a more moderate fashion act similar to each other. This variable has two levels for the extreme response categories (i.e., always or rarely/never) such that those who predominantly have an extreme behavior pattern (i.e., two out of three responses are $G_t=1 t=1,2,3$ or two out of three responses are $G_t=3$ for a 3-level time varying grouping variable with three time points). The remaining respondents represent those whose status is always or usually in the middle of the possible response options. The extreme behavior variable has K+2 levels, where K is the number of levels the time-invariant summary variable has. To illustrate this variable, we will use the frequency of shopping time varying grouping variable, which has three levels (1=almost every day, 2=at least once a week, 3=once a month or less). The extreme behavior time-invariant summary variable for this time varying grouping variable is defined as

$$G = \begin{cases} 1 & \text{if } G_1 = 1 \text{ and } G_2 = 1 \text{ and } G_3 = 1 \\ 2 & \text{if } G_i = 1 \text{ and } G_j = 1 \text{ and } G_k \neq 1 \text{ for } i, j, k = 1,2,3 \ i \neq j \neq k \\ 3 & \text{otherwise} \\ 4 & \text{if } G_i = 3 \text{ and } G_j = 3 \text{ and } G_k \neq 3 \text{ for } i, j, k = 1,2,3 \ i \neq j \neq k \\ 5 & \text{if } G_1 = 3 \text{ and } G_2 = 3 \text{ and } G_3 = 3 \end{cases}$$

In this paper, a model using this time-invariant summary grouping variable is denoted as M(S2) where S2 represents the extreme behavior model.

The third time-invariant grouping variable we considered is a *modified mover-stayer grouping variable*. This variable expands the stayer levels such that there is an additional level for those who predominantly behave in a certain manner (i.e., a respondent indicates two out of three time points as the same behavior). When K>2, the modified mover-stayer variable has K+(K+1) levels, where K is the number of levels the time-invariant summary variable has. When K=2, the modified mover-stayer model is the same as the extreme behavior model. To illustrate this variable, we will use the frequency of shopping time varying grouping variable, which has three levels (1=almost every day, 2=at least once a week, 3=once a month or less). The extreme behavior time-invariant summary variable for this time varying grouping variable is defined as

$$G = \begin{cases} 1 & \text{if } G_1 = 1 \text{ and } G_2 = 1 \text{ and } G_3 = 1 \\ 2 & \text{if } G_i = 1 \text{ and } G_j = 1 \text{ and } G_k \neq 1 \text{ for } i, j, k = 1,2,3 \ i \neq j \neq k \\ 3 & \text{if } G_1 = 2 \text{ and } G_2 = 2 \text{ and } G_3 = 2 \\ 4 & \text{if } G_i = 2 \text{ and } G_j = 2 \text{ and } G_k \neq 2 \text{ for } i, j, k = 1,2,3 \ i \neq j \neq k \\ 5 & \text{if } G_i = 3 \text{ and } G_j = 3 \text{ and } G_k \neq 3 \text{ for } i, j, k = 1,2,3 \ i \neq j \neq k \\ 6 & \text{if } G_1 = 3 \text{ and } G_2 = 3 \text{ and } G_3 = 3 \\ 7 & \text{otherwise} \end{cases}$$

In this paper, a model using this time-invariant summary grouping variable is denoted as M(S3), where S3 represents the modified mover-stayer time-invariant model. Because of the additional parameters added by including this variable, we only considered it when both the mover-stayer and extreme behavior time-invariant summary variables failed our criteria for being an acceptable summary variable.

In addition to the time-invariant summary grouping variable models, we considered two reduced time varying grouping models. These models use the time varying grouping variables, but are more parsimonious than the population model. The first reduced time varying grouping variable model we considered is the *no memory time varying* model. This model assumes that the probability of an outcome occurring only depends on the current status of the grouping variable in which the person is (i.e., $\pi_{x_3|g_1 g_2 g_3}^{X_3|G_1 G_2 G_3} = \pi_{x_3|g_3}^{X_3|G_3}$). The likelihood kernel for the MLC model for cell $(g_1,g_2,g_3,a_1,a_2,a_3)$ of the

$G_1 \times G_2 \times G_3 \times A_1 \times A_2 \times A_3$ cross-classification table in terms of X_1, X_2, and X_3 is written as

$$\pi_{g_1 g_2 g_3 a_1 a_2 a_3}^{G_1 G_2 G_3 A_1 A_2 A_3} = \sum_{x_1} \sum_{x_2} \sum_{x_3} \pi_{g_1 g_2 g_3}^{G_1 G_2 G_3} \pi_{x_1|g_1}^{X_1|G_1} \pi_{x_2|g_2 x_1}^{X_2|G_2 X_1} \pi_{x_3|g_3 x_2}^{X_3|G_3 X_2} \pi_{a_1|x_1}^{A_1|X_1} \pi_{a_2|x_2}^{A_2|X_2} \pi_{a_3|x_3}^{A_3|X_3} \tag{9}$$

Figure 4 illustrates the path diagram for this model. In this paper, we denote this model as M(R1), where R1 represents the no memory model.
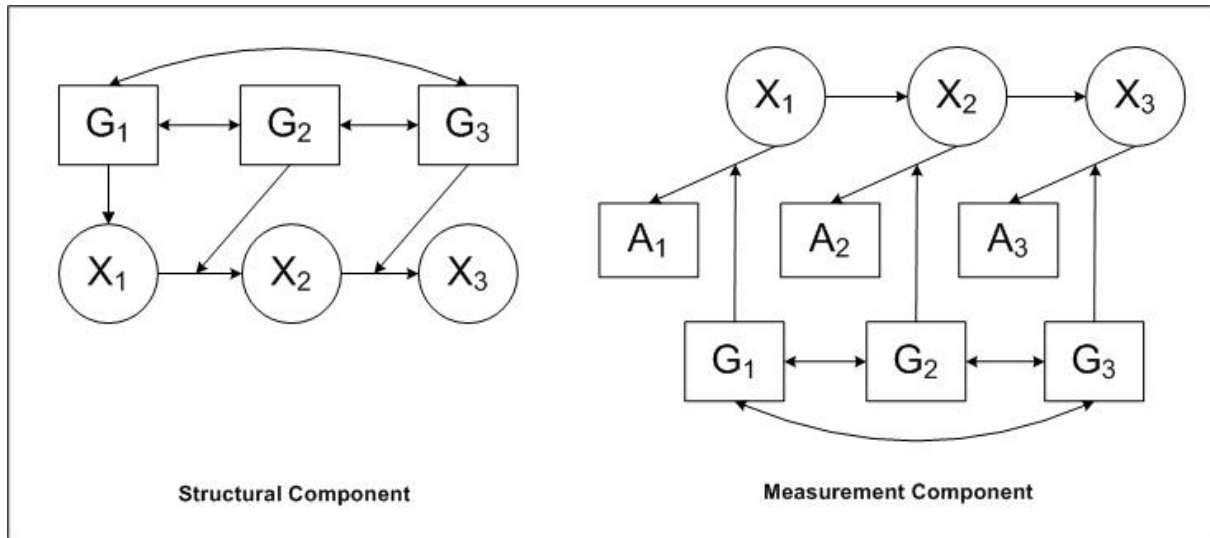
Figure 4. Path diagram for M(R1) that assumes time-homogeneous classification errors (i.e., $A_1|X_1 G_1 = A_2|X_2G_2 = A_3|X_3G_3$.). Double arrows indicate a three-way interaction (i.e., $G_1G_2G_3$).

The second reduced time varying grouping variable model we considered is a *manifest second-order Markov time varying model*. This model considers all previous time points based on how events transpired over time in the structural component, but uses all time points for each piece of the measurement component. Under this reduced model, the likelihood kernel for the MLC model for cell $(g_1,g_2,g_3,a_1,a_2,a_3)$ of the G_1xG_2xG_3xA_1xA_2xA_3 cross-classification table in terms of X_1, X_2, and X_3 may be written as

$$\pi_{g_1g_2g_3a_1a_2a_3}^{G_1G_2G_3A_1A_2A_3} = \pi_{g_1g_2g_3}^{G_1G_2G_3} \pi_{x_1|g_1}^{X_1|G_1} \pi_{x_2|x_1g_1g_2}^{X_2|X_1G_1G_2} \pi_{x_3|x_2g_1g_2g_3}^{X_3|X_2G_1G_2G_3} \pi_{a_1|x_1g_1g_2g_3}^{A_1|X_1G_1G_2G_3} \pi_{a_2|x_2g_1g_2g_3}^{A_2|X_2G_1G_2G_3} \pi_{a_3|x_3g_1g_2g_3}^{A_3|X_3G_1G_2G_3} \qquad (10)$$

Figure 5 presents the path diagram for this model. We will denote this model as M(R2), where R2 represents the manifest second-order Markov model. This model uses all information available as it occurred.

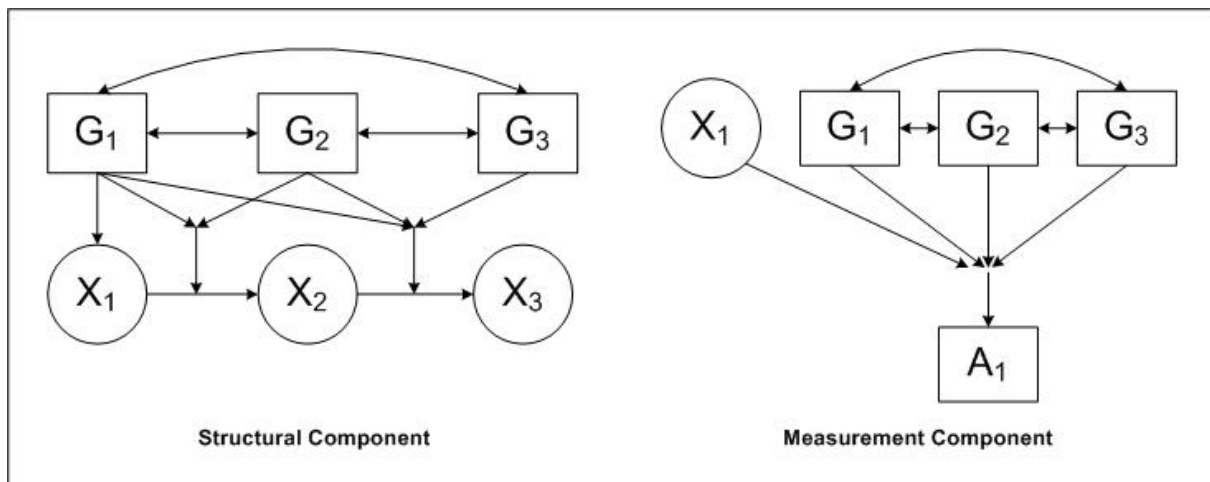Figure 5. Path diagram for M(R2) that assumes time-homogeneous classification (i.e., $A_1|X_1G_1G_2G_3 = A_2|X_2G_1G_2G_3 = A_3|X_3G_1G_2G_3$). Double arrows indicate a three-way interaction (i.e., $G_1G_2G_3$). The figure only shows the path diagram for the measurement component of indicator A_1. The diagram for the measurement component is similar for the other two indicators: $A_2|X_2G_1G_2G_3$ and $A_3|X_3G_1G_2G_3$.

3.2.2 Estimates from the Population Model

Table 1 presents estimates from the population model—the fully saturated model that best represents the classification error and prevalence rates given that only one grouping variable is being included in the model. Although the population model may be over parameterized, its estimates can be considered as a gold standard for comparison because all other potential models are nested within it.

Table 1. Classification Error and Prevalence Estimates1 from the Population Model, by Type of Crime Victimization and Grouping Variable

	Classification Error Rates		Prevalence Rates		
	$\pi_{1\|2}^{A_1\|X_1}$	$\pi_{2\|1}^{A_1\|X_1}$	$\pi_1^{X_1}$	$\pi_1^{X_2}$	$\pi_1^{X_3}$
Less serious crimes					
Going out	0.0284	0.6486	0.1682	0.1268	0.1067
Going shopping	0.0227	0.7007	0.2172	0.1634	0.1478
Public transportation	0.0283	0.6592	0.1746	0.1260	0.1045
Mode of interview	0.0292	0.6656	0.1750	0.1233	0.1030
Respondent type	0.0316	0.6520	0.1611	0.1119	0.0898
Alone during interview	0.0297	0.6565	0.1684	0.1184	0.0982
Household crimes					
Going out	0.0235	0.6450	0.1856	0.1581	0.1399
Going shopping	0.0258	0.6611	0.1845	0.1546	0.1363
Public transportation	0.0232	0.6391	0.1868	0.1561	0.1339
Mode of interview	0.0241	0.6286	0.1682	0.1395	0.1240
Respondent type	0.0285	0.6831	0.1881	0.1331	0.1246
Alone during interview	0.0281	0.6311	0.1597	0.1281	0.1103

[1] Based on unweighted data.

4. Discussion

4.1 Results of Comparisons of Time-Invariant Summary Models

For each of the time-invariant summary models tested, Tables 2 and 3 show the fit statistics and bias, respectively, for each victimization type and grouping variable. Of the 12 grouping variable by victimization types that were tested, 11 had at least one time-invariant summary model that met our criteria for being an acceptable model. For nine of these grouping variables, M(S1) met all of the criteria, and for eight of the grouping variables, M(S2) met all criteria ['respondent type' for household crimes being the one grouping variable where M(S1) met the criteria, but M(S2) did not]. In the eight models where both M(S1) and M(S2) met the criteria, M(S1) had a smaller BIC, making it the best time-invariant summary model based on our criteria. Only the variable "respondent type" for less serious crimes did not have a time-invariant summary model that met the criteria. This grouping variable did not have a model with a p-value greater than 0.05 for the L^2 [p=0.0015 for M(S1) and p=0.0005 for M(S2)]. Furthermore, there were two variables where both M(S1) and M(S2) failed our specified criteria. "Going out" for less serious crimes had a p-value less than 0.05 for the L^2 statistic for both M(S1) (p=0.0011) and M(S2) (p=0.0424). "Going shopping" for household crimes had an ARB greater than 10% for both M(S1) (ARB=11.4%) and M(S2) (ARB=10.3%). In both of these cases, the model M(S3) met the criteria.

Table 2. Fit Statistics for Time-Invariant Summary Models, by Type of Crime Victimization and Grouping Variable

	Model	npar	DF	D	BIC[1]	L^2 p-value
Less serious crimes						
Going out	M(S1)	32	184	0.0215	1.9467	0.0011
	M(S2)	40	176	0.0202	1.9471	0.0424
	M(S3)	56	160	0.0164	1.9485	0.1304
Going shopping	M(S1)	32	184	0.0133	1.7615	0.8161
	M(S2)	40	176	0.0152	1.7622	0.7828
Public transportation	M(S1)	24	40	0.0031	1.1334	0.9100
	M(S2)	32	32	0.0028	1.1342	0.8276
Mode of interview	M(S1)	24	40	0.0035	1.1133	0.9631
	M(S2)	32	32	0.0034	1.1141	0.9227
Respondent type	M(S1)	24	40	0.0025	0.6383	0.0015
	M(S2)	32	32	0.0023	0.6390	0.0005
Alone during interview	M(S1)	24	40	0.0032	0.9114	0.7065
	M(S2)	32	32	0.0027	0.9121	0.6481
Household crimes						
Going out	M(S1)	32	184	0.0219	1.1897	0.5883
	M(S2)	40	176	0.0214	1.1903	0.8061
Going shopping	M(S1)	32	184	0.0174	1.1035	0.5189
	M(S2)	40	176	0.0178	1.1040	0.8625
	M(S3)	56	160	0.0138	1.1054	0.9063
Public transportation	M(S1)	24	40	0.0052	0.6856	0.9783
	M(S2)	32	32	0.0033	0.6863	0.9837
Mode of interview	M(S1)	24	40	0.0039	0.7571	0.9922
	M(S2)	32	32	0.0033	0.7579	0.9912
Respondent type	M(S1)	24	40	0.0012	0.3773	0.0628
	M(S2)	32	32	0.0012	0.3781	0.0104
Alone during interview	M(S1)	24	40	0.0033	0.5756	0.9679
	M(S2)	32	32	0.0028	0.5763	0.9766

[q 1] BIC in units of 100,000.

Table 3. Bias in Estimates Between Time-Invariant Summary Variable Recode Models and the Population Model, by Type of Crime Victimization and Grouping Variable Based on Synthetic Data

		Classification Error Rates		Prevalence Rates		
	Model	$\pi_{1\|2}^{A_1\|X_1}$	$\pi_{2\|1}^{A_1\|X_1}$	$\pi_1^{X_1}$	$\pi_1^{X_2}$	$\pi_1^{X_3}$
Less serious crimes						
Going out	M(S1)	0.0016	-0.0398	-0.0221	-0.0201	-0.0198
	M(S2)	0.0020	-0.0405	-0.0235	-0.0214	-0.0208
	M(S3)	0.0019	-0.0410	-0.0233	-0.0214	-0.0206
Going shopping	M(S1)	0.0042	-0.0421	-0.0395	-0.0356	-0.0359
	M(S2)	0.0056	-0.0699	-0.0574	-0.0508	-0.0516
Public transportation	M(S1)	0.0002	-0.0005	-0.0010	-0.0005	-0.0025
	M(S2)	0.0003	0.0000	-0.0007	-0.0005	-0.0021
Mode of interview	M(S1)	0.0000	-0.0169	-0.0093	-0.0071	-0.0070
	M(S2)	-0.0001	-0.0168	-0.0089	-0.0067	-0.0066
Respondent type	M(S1)	-0.0005	-0.0023	0.0002	0.0002	0.0004
	M(S2)	-0.0002	-0.0010	0.0001	0.0005	0.0006
Alone during interview	M(S1)	-0.0005	-0.0020	0.0001	0.0001	-0.0010
	M(S2)	-0.0005	-0.0019	0.0003	0.0003	-0.0008
Household crimes						
Going out	M(S1)	0.0036	-0.0530	-0.0335	-0.0336	-0.0290
	M(S2)	0.0042	-0.0498	-0.0336	-0.0339	-0.0291
Going shopping	M(S1)	0.0045	-0.0754[1]	-0.0457	-0.0444	-0.0418
	M(S2)	0.0044	-0.0679[1]	-0.0427	-0.0430	-0.0411
	M(S3)	0.0031	-0.0497	-0.0326	-0.0333	-0.0336
Public transportation	M(S1)	0.0014	-0.0091	-0.0081	-0.0079	-0.0057
	M(S2)	0.0015	-0.0078	-0.0078	-0.0077	-0.0055
Mode of interview	M(S1)	0.0019	-0.0258	-0.0159	-0.0178	-0.0167
	M(S2)	0.0007	-0.0166	-0.0092	-0.0105	-0.0092
Respondent type	M(S1)	-0.0001	0.0044	0.0033	0.0038	0.0041
	M(S2)	-0.0002	-0.0099	-0.0057	-0.0020	-0.0042
Alone during interview	M(S1)	0.0014	-0.0274	-0.0150	-0.0125	-0.0105
	M(S2)	0.0011	-0.0314	-0.0161	-0.0144	-0.0130

[1] Absolute relative bias greater than 10%.

4.2 Results of Comparisons of Reduced Time Varying Grouping Variable Models

For each of the time varying grouping variable models tested, Tables 4 and 5 show the fit statistics and bias, respectively, for each victimization type and grouping variable. For all 12 grouping variables and victimization types, at least one of the two reduced models tested met the criteria. Nine of the 12 grouping variable and victimization types M(R1) met the criteria. M(R1) failed twice because of a L^2 p-value being less than 0.05 ["going out" for less serious crimes (p=0.0003) and "mode of interview" for less serious crimes (p=0.0000)] and failed once because of an absolute relative bias for the false negative rate being greater than 10% ["going shopping" for less serious crimes (ARB=11.7%)]. M(R2) met the criteria in 11 of the 12 grouping variables and victimization types tested. M(R2) for "going shopping" for household crimes had an absolute relative bias greater than 10% (ARB=10.3%). In the eight cases where both M(R1) and M(R2) met the criteria, M(R1) always had a smaller BIC making it the better model.

Table 4. Fit Statistics for Reduced Time Varying Grouping Variable Models, by Type of Crime Victimization and Grouping Variable

	Model	npar	DF	D	BIC[1]	L^2 p-value
Less serious crimes						
Going out	M(R1)	48	168	0.0211	1.9481	0.0003
	M(R2)	156	60	0.0040	1.9571	0.9995
Going shopping	M(R1)	48	168	0.0177	1.7630	0.5346
	M(R2)	156	60	0.0040	1.7728	0.9381
Public transportation	M(R1)	22	42	0.0100	1.1334	0.0511
	M(R2)	50	14	0.0024	1.1358	0.6797
Mode of interview	M(R1)	22	42	0.0124	1.1137	0.0000
	M(R2)	50	14	0.0009	1.1156	0.9998
Respondent type	M(R1)	22	42	0.0031	0.6377	0.4015
	M(R2)	50	14	0.0011	0.6402	0.8083
Alone during interview	M(R1)	22	42	0.0082	0.9115	0.0026
	M(R2)	50	14	0.0016	0.9137	0.8388
Household crimes						
Going out	M(R1)	48	168	0.0243	1.1912	0.1559
	M(R2)	156	60	0.0056	1.2001	0.9999
Going shopping	M(R1)	48	168	0.0196	1.1050	0.2043
	M(R2)	156	60	0.0045	1.1140	0.9979
Public transportation	M(R1)	22	42	0.0106	0.6855	0.4714
	M(R2)	50	14	0.0020	0.6879	0.9798
Mode of interview	M(R1)	22	42	0.0063	0.7569	0.9818
	M(R2)	50	14	0.0025	0.7595	0.7924
Respondent type	M(R1)	22	42	0.0016	0.3767	0.9806
	M(R2)	50	14	0.0003	0.3793	0.9120
Alone during interview	M(R1)	22	42	0.0065	0.5754	0.8758
	M(R2)	50	14	0.0015	0.5779	0.9672

[1] BIC in units of 100,000.

Table 5. Bias in Estimates Between Reduced Time Varying Grouping Variable Models and the Population Model, by Type of Crime Victimization and Grouping Variable Based on Synthetic Data

		Classification Error Rates		Prevalence Rates		
	Model	$\pi_{1|2}^{A_1|X_1}$	$\pi_{2|1}^{A_1|X_1}$	$\pi_1^{X_1}$	$\pi_1^{X_2}$	$\pi_1^{X_3}$
Less serious crimes						
Going out	M(R1)	-0.0002	-0.0236	-0.0114	-0.0129	-0.0128
	M(R2)	0.0004	-0.0367	-0.0182	-0.0173	-0.0155
Going shopping	M(R1)	0.0063	-0.0819[1]	-0.0648	-0.0561	-0.0573
	M(R2)	0.0009	-0.0371	-0.0290	-0.0227	-0.0250
Public transportation	M(R1)	0.0000	-0.0087	-0.0045	-0.0059	-0.0050
	M(R2)	0.0005	-0.0095	-0.0068	-0.0061	-0.0065
Mode of interview	M(R1)	-0.0004	-0.0186	-0.0090	-0.0062	-0.0062
	M(R2)	0.0005	-0.0151	-0.0096	-0.0079	-0.0075
Respondent type	M(R1)	0.0000	-0.0026	-0.0014	-0.0002	-0.0006
	M(R2)	-0.0018	0.0043	0.0072	0.0062	0.0056
Alone during interview	M(R1)	-0.0083	0.0269	0.0383	0.0347	0.0316
	M(R2)	-0.0015	0.0096	0.0091	0.0073	0.0052
Household crimes						
Going out	M(R1)	-0.0017	-0.0082	-0.0007	-0.0030	-0.0048
	M(R2)	0.0004	-0.0564	-0.0271	-0.0300	-0.0200
Going shopping	M(R1)	0.0011	-0.0380	-0.0223	-0.0242	-0.0232
	M(R2)	0.0011	-0.0679[1]	-0.0355	-0.0333	-0.0314
Public transportation	M(R1)	0.0036	-0.0292	-0.0231	-0.0221	-0.0191
	M(R2)	0.0028	-0.0211	-0.0178	-0.0161	-0.0139
Mode of interview	M(R1)	0.0025	-0.0356	-0.0213	-0.0212	-0.0199
	M(R2)	0.0017	-0.0423	-0.0218	-0.0222	-0.0192
Respondent type	M(R1)	-0.0015	-0.0356	-0.0171	0.0079	0.0041
	M(R2)	-0.0008	-0.0433	-0.0228	-0.0016	0.0067
Alone during interview	M(R1)	0.0002	-0.0250	-0.0113	-0.0089	-0.0072
	M(R2)	0.0017	-0.0373	-0.0199	-0.0165	-0.0149

[1] Absolute relative bias greater than 10%.

4.3 Comparison of the Best Time-Invariant Summary Models and the Best Time Varying Models

After considering the best time-invariant summary model and the best time varying grouping variable model for each grouping variable and victimization type, we compared the two to see if the time-invariant summary grouping variable could be used as an alternative to the time varying grouping variable model. To assess this, we considered the model fit and the sparseness of the data.

In terms of model fit, the results differed depending on how many levels the time varying grouping variable had. For variables with three levels ("going out" and "going shopping"), the best time-invariant summary model had a smaller BIC than the best time varying grouping variable model. The one instance where the time varying grouping model performed better was the "going shopping" model for household crimes. In this case, M(S1) and M(S2) failed to meet our required criteria for being an acceptable model. M(S3) met the criteria, but the additional parameters increased its BIC (BIC=110538) to be larger than the BIC of M(R1) (BIC=110496). The best time-invariant summary model for "going out"

for less serious individual crimes was also M(S3), but because M(R1) did not meet the criteria for being an acceptable model it was not compared to M(S3). When M(S3) for "going out" for less serious crimes was compared to M(R2) it had a smaller BIC [BIC=194847 for M(S3) and BIC=195710 M(R2)]. However, for variables with two levels, the best time varying grouping model had a smaller BIC in seven out of the eight compared models. In one case where the best time-invariant summary variable model had better fit than the best time varying grouping variable model—"mode of interview" for less serious individual crimes—M(R1) failed to meet the required criteria for being an acceptable model. In this case, the BIC for M(S1) (BIC=111331) was smaller than the BIC for M(R2) (BIC=111564). In other words, in all cases where M(R1) was an acceptable model, its BIC was always smaller than the best time-invariant summary model when the time varying grouping variable had two levels.

However, when the sparseness of the data was taken into account the time-invariant summary models always performed better regardless of the how many levels the time varying grouping variable had. Table 6 shows the number of cells (ncells) in a data table that each model had [note that M(R1) and M(R2) share the same data table and, therefore, have equivalent patterns of sparseness, thus, only the sparseness from M(R1) is presented], the number of cells in the data table that had five or fewer observations, and the percentage of the cells that were small. For all grouping variables and victimization types the time-invariant summary models significantly reduced the amount of sparseness in both absolute and relative terms. This difference was most dramatic in models where the time varying grouping variable had three levels. This is because the difference in the number of cells between the time varying grouping variable models and the time-invariant summary models increases as the number of levels in the time varying grouping variable increases. For example, the difference in the number of cells between M(R1) and M(S1) when the time varying grouping variable has three levels is 184, but it is 40 when the time varying grouping variable has two levels. Interestingly, the one case where no time-invariant summary models performed well—"respondent type" for less serious crimes—had over 30% of the cells in its data table with fewer than five observations for all models. Namely, M(S1) had 33.3% of cells with five or less observations, and M(S2) had 37.5% of cells with five or fewer observations.

Thus, although reducing the number of parameters in a time varying grouping variable model improved the model fit, it did not improve the amount of sparseness in the data. Because our analytic goals were to both find a parsimonious model with good fit and a low level of data sparseness so that we could incorporate additional grouping variables into our model, we determined that in all cases, except "respondent type" for less serious individual crimes, the best time-invariant summary variable could be used to replace the time varying grouping variables.

As a final check on our findings, we ran the best time-invariant summary model and best time varying model using the actual NCVS data (i.e., the data used to create the population model). This check allowed us to confirm that the synthetic data were acting as a proper proxy for the true data set. The parameter estimates for these models are presented in Table 7. In all cases the absolute relative bias of the false negative rate compared to the population model was less than 10%. The largest was 5.1% [M(R1) for the respondent type variable for household crimes]. In fact, 10 of the 23 models had an absolute relative bias less than 1% for the false negative rate. Therefore, we concluded that the synthetic data were a successful proxy for the actual data, while providing the benefit of removing all sources of variation from the data, except those induced by the variables of interest in the analysis.

Table 6. Percentage of Table Cells with Five or Fewer Observations, by Time Varying Grouping Variable and Model Data Table

	Model	ncells	No. of small cells	Percentage
Less serious crimes				
Going out	M(R1)	216	59	27.3
	M(S1)	32	1	3.1
	M(S2)	40	1	2.5
	M(S3)	56	1	1.8
Going shopping	M(R1)	216	82	38.0
	M(S1)	32	2	6.3
	M(S2)	40	2	5.0
Public transportation	M(R1)	64	5	7.8
	M(S1)	24	0	0.0
	M(S2)	32	0	0.0
Mode of interview	M(R1)	64	5	7.8
	M(S1)	24	0	0.0
	M(S2)	32	0	0.0
Respondent type	M(R1)	64	37	57.8
	M(S1)	24	8	33.3
	M(S2)	32	12	37.5
Alone during interview	M(R1)	64	13	20.3
	M(S1)	24	3	12.5
	M(S2)	32	4	12.5
Household crimes				
Going out	M(R1)	216	75	34.7
	M(S1)	32	2	6.3
	M(S2)	40	2	5.0
Going shopping	M(R1)	216	98	45.4
	M(S1)	32	5	15.6
	M(S2)	40	7	17.5
	M(S3)	56	9	16.1
Public transportation	M(R1)	64	7	10.9
	M(S1)	24	0	0.0
	M(S2)	32	0	0.0
Mode of interview	M(R1)	64	11	17.2
	M(S1)	24	1	4.2
	M(S2)	32	2	6.3
Respondent type	M(R1)	64	47	73.4
	M(S1)	24	11	45.8
	M(S2)	32	17	53.1

Table 7. Classification Error and Prevalence Estimates from Best Time-Invariant Summary and Time Varying Grouping Variable Models Based on Actual NCVS Data

		Classification Error Rates		Prevalence Rates		
Model		$\pi_{1\|2}^{A_1\|X_1}$	$\pi_{2\|1}^{A_1\|X_1}$	$\pi_1^{X_1}$	$\pi_1^{X_2}$	$\pi_1^{X_3}$
Less serious crimes						
Going out	M(S3)	0.0290	0.6538	0.1699	0.1258	0.1021
	M(R1)	0.0261	0.6502	0.1753	0.1308	0.1108
Going shopping	M(S1)	0.0283	0.6731	0.1831	0.1312	0.1114
	M(R2)	0.0231	0.6779	0.1990	0.1492	0.1300
Public transportation	M(S1)	0.0292	0.6639	0.1749	0.1259	0.1013
	M(R1)	0.0303	0.6539	0.1666	0.1164	0.0942
Mode of interview	M(S1)	0.0322	0.6561	0.1617	0.1121	0.0894
	M(R2)	0.0285	0.6521	0.1693	0.1188	0.0967
Respondent type[1]	M(R1)	0.0313	0.6583	0.1651	0.1159	0.0927
Alone during interview	M(S1)	0.0305	0.6550	0.1658	0.1154	0.0939
	M(R2)	0.0296	0.6667	0.1743	0.1228	0.0997
Household crimes						
Going out	M(S1)	0.0314	0.6234	0.1553	0.1259	0.1080
	M(R1)	0.0271	0.6552	0.1820	0.1506	0.1263
Going shopping	M(S3)	0.0272	0.6349	0.1666	0.1356	0.1160
	M(R1)	0.0243	0.6493	0.1814	0.1493	0.1286
Public transportation	M(S1)	0.0268	0.6280	0.1724	0.1425	0.1217
	M(R1)	0.0307	0.6238	0.1615	0.1297	0.1093
Mode of interview	M(S1)	0.0257	0.6340	0.1667	0.1337	0.1164
	M(R1)	0.0278	0.6357	0.1632	0.1299	0.1137
Respondent type	M(S1)	0.0284	0.6877	0.1914	0.1370	0.1286
	M(R1)	0.0269	0.6483	0.1717	0.1417	0.1294
Alone during interview	M(S1)	0.0286	0.6458	0.1659	0.1319	0.1169
	M(R1)	0.0278	0.6354	0.1627	0.1300	0.1135

[1] The respondent type grouping for less serious crimes did not have any time-invariant summary variable models that meet the evaluation criteria.

5. Discussion

5.1 A General Process for Assessing Time-Invariant Summary Variables

The paper presents a process that can be generalized to any data set being analyzed. In this paper, the focus was determining if a time-invariant summary variable could replace time varying grouping variables to reduce data sparseness and increase the number of available degrees of freedom to allow additional grouping variables to be included in the model. To that end, our evaluation criteria put an emphasis on models that had good fit, while taking parsimony into account and reducing the sparseness of the data. However, the basic process of using the population model to create synthetic data and test alternative models could just as easily be applied to a different set of evaluation criteria. For example, if a particular data set being analyzed has very few potential grouping variables (time varying or time invariant)

then less emphasis can be placed on sparseness as long as the model is stable. Moreover, our analysis focused on the bias in only the false negative rate. This was done for two reasons. First, our analytic goals focused on estimating the classification error rates in the NCVS. Therefore, although it is important to have a structural component with good fit, the accuracy of those estimates is not as critical. Second, crime victimization has a relatively small false positive rate because, regardless of the type crime, it is less likely for a person who was not truly a victim to indicate that they were than for a person who was truly a victim to say they were not. Therefore, even a small change can result in a large relative bias even if it is not substantively different. This could lead to the rejection of certain models that may be perfectly good. Because of this, we chose not to include the false positive rate in our evaluation criteria. However, if an analyst's main objectives were to estimate the structural component or if the expected false positive rates were larger, the bias in these estimates could easily be incorporated into the evaluation criteria.

5.2 Analysis of the Measurement Error Estimates

Although the focus of this paper was to determine if a model that uses a time-invariant summary variable instead of time varying grouping variables can be used to reduce the number of parameters included in a model, the analyses also offer a preliminary look at the classification error rates and true prevalence rates of crime victimization in the NCVS. Based on our review of the literature, there are no existing papers examining the classification error rates in the NCVS screener questions. Therefore, this analysis offers a preliminary glimpse at how well the NCVS screener items identify crime victimization. However, it is important to point out that these estimates are not comparable to published victimization rates. First, the victimization rates published from the NCVS are based on an incident report that is administered after one or more of the screener items are endorsed. Second, the victimization rates published are based on all interviews that took place during the reporting year. This includes a mixture of respondents from all six possible waves. Our analysis groups' respondents are based on which wave a respondent is in regardless of when in time that interview took place. Therefore, the prevalence rates estimated in our analysis represent how the reporting of crime victimization changes over time in study rather than how it changes over a specific period of time.

Table 1 shows the parameter estimates from the population models for each time varying grouping variable in the NCVS. Two characteristics of these estimates are worth noting. First, the false negative rates seem implausibly high at first glance. They are around 65% for less serious individual crimes and 61% for household crimes. Given these rates, it is important to note that steps were taken to ensure that these estimates were not caused by latent variable flippage[1] (Biemer, 2011). Second, for both victimization types, the prevalence of crime decreases as time in sample increases. It is possible that these two phenomena are related. There is evidence that respondents in the NCVS suffer from *time in sample bias* or *panel conditioning,* which has led to a downward bias in victimization rates as the number of interviews completed increases (see, for example, Lynch, Berbaum, & Planty, 2002; Yan, 2008). However, respondents of the NCVS have not been found to endure "fatigue bias," whereby after reporting a victimization and enduring a longer interview they do not participate in the subsequent panel (Hart, Rennison, & Gibson, 2005). Time-in-sample bias is the term used to describe how respondents can be "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. The Bureau of Justice Statistics has found that time-in-sample bias occurs in the NCVS, but because annual estimates for the NCVS are essentially averages of the panel estimates across all six interview frequencies, the effects of time-in-sample bias change little over time and, therefore, have no effects on the year-to-year estimates that are produced (Rand & Catalano, 2007).

To illustrate the effects of time-in-sample bias on the false negative rate, we compared the estimates from when a respondent's first interview occurred to the estimates from a respondent's first three waves. For each victimization type, Figure 6 shows the prevalence rate over time based on when the respondent's first bounded interview occurred (for the data we used, a respondent's first bounded interview occurred during a 6-month period starting at one of three times: first quarter 1996 [96Q], third quarter 1996 [96Q3], or first quarter 1997 [97Q1]). By using the respondent's first bounded interview, we mitigate the potential bias caused by panel conditioning. The figure shows that for both victimization types, crime victimization rates are relatively flat during that 18-month period. These findings are consistent with published NCVS estimates, which show only a slight decrease in victimization rates during the period from 1995 to 1997 (Klaus, 2002). Furthermore, Figure 6 shows that based on time in sample, the prevalence rate decreases as the wave number increases. This corroborates the findings by Yan (2008) and Rand and Catalano (2007), but also implies that if the true victimization rates are flat over time, then there needs to be a sizable false negative rate to compensate for the decreasing number of reported victimizations as respondents participate longer in the panel.

---

[1]Latent variable flippage is the misalignment of the categories of one or more of the latent variables X_1, X_2, and X_3 and the categories of the indicators A_1, A_2, and A_3.
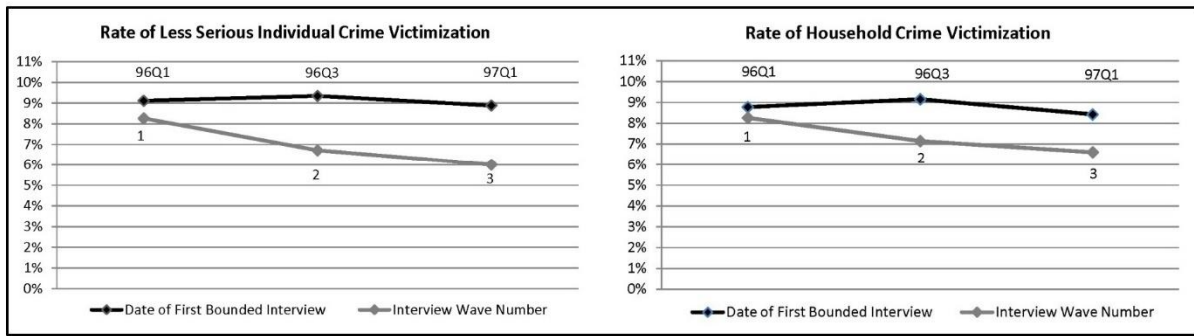
Figure 6. Crime victimization rates based on when first bounded interview occurred and interview wave number by victimization type

Although this initial analysis indicates that there is some level of false negatives, we wanted to assess whether the estimated rates from the MLCA were reasonable. Based on Klaus (2002) and ignoring the panel conditioning, one would assume that, to obtain a flat victimization rate, the same number of respondents would indicate being a victim at each wave (albeit possibly a different set of respondents). Therefore, if we assume the prevalence rate is really flat, we can calculate the average false negative rate using the formula

$$\pi_{2|1}^{A_t|X_t} = \frac{\left(n_{v_{t-1}} - n_{v_t}\right) + \pi_{1|2}^{A_t|X_t}\left(n - n_{v_{t-1}}\right)}{n_{v_{t-1}}} \tag{11}$$

where n_{v_t} is the number of victimizations reported at time t for $t=2, 3$, and n is the number of people who responded to all three waves analyzed. If we assume a constant false positive rate of 2.75% (approximately the rate we estimated for both victimization types), we obtain a false negative rate of 61.3% and 54.9% for less serious individual crimes and 51.4% and 47.2% for household crimes between time points 1 and 2 and time points 2 and 3, respectively. These rates are comparable to our model-based estimates indicating that the model estimates are plausible.

Figure 6 shows a monotonically decreasing prevalence rate over time in sample. When assuming that the true prevalence rate is flat over time, this indicates that the false negative rate is increasing over time. However, the results of equation (11) indicate that the false negative rate is not constant over time, but rather it increased at a greater rate between the first and second interview than between the second and third interview. This makes sense given the decreasing slope in the victimization rates as the interview wave number increases (Figure 6), and it indicates a potential violation in the time homogeneous classification error rates assumption. When the time homogeneous classification error rate assumption is made, but does not hold the resulting estimated classification, error rates are roughly averages of the classification error rates at each time point (Biemer, 2011). This only exacerbates the resulting prevalence estimates over time. To relax this assumption and still have model identifiability, more waves need to be analyzed.

6. Conclusions

Our paper develops a process by which analysts can determine if a time-invariant summary variable can replace a time varying grouping variable in an MLC model. To accurately measure the change in model fit that one variable has compared to the model that generated the underlying data, we used simulated data rather than empirical data from a panel survey. However, because we wanted to apply our results immediately to a specific data set, we propose to simulate the data by creating synthetic data that have similar characteristics to the data set of interest. Thus, as long as the assumption that the actual survey data behave like the synthetic data hold, the conclusions regarding the best variable can be applied directly to the data set being analyzed. However, because the synthetic data are tailored to a particular survey, the conclusions drawn cannot be generalized to other data sets. In other words, the process is general, but the results are specific.

For synthetic data similar to the NCVS, the results of the process indicate that a time varying summary variable can be used as an alternative specification to a set of time varying grouping variables. Therefore, an MLCA of the NCVS can incorporate time-invariant summary grouping variables.

For an MLCA of the NCVS data, this analysis indicates that further analysis of the time homogeneous classification errors assumption needs to be conducted. It is likely that a model with more than three time points will be necessary to accurately

estimate the classification error rates. With the need to expand the number of time points, the ability to use time-invariant summary variables will be beneficial because of the additional sparseness created by increasing the number of time points analyzed. Moreover, this paper does not suggest which grouping variable or set of grouping variables best reduces heterogeneity in the measurement error, but rather indicates that for most time varying grouping variables in the NCVS, a time-invariant summary variable can be used instead. Further analysis, which incorporates time-invariant grouping variables, needs to be conducted to determine the best set of grouping variables for the NCVS.

Acknowledgments

References

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

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

Biemer, P., & Berzofsky, M. (2011). Some issues in the application of latent class models for questionnaire design. In J. Madans, K. Miller, G. Willis, & A. Maitland (Eds.). *Questionnaire Evaluation Methods, 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

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

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.

Klaus, P. A. (2002). Crime and the nation's households, 2000 (with trends, 1994 – 2000). U.S. Department of Justice, Office of Justice Programs. Bureau of Justice Statistics Bulletin, NCJ 194107. Retrieved from http://bjs.ojp.usdoj.gov/content/pub/pdf/cnh00.pdf.

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.

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.

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. (1997). LEM: A general program for the analysis of categorical data. Tilburg Netherlands: Department of Methodology and Statistics, Tilburg University. Retrieved from http://spitswww.uvt.nl/web/fsw/mto/lem/manual.pdf.

Vermunt, J. K. (2002). Comments on "Latent class analysis of complex sample survey data." *Journal of the American Statistical Association, 97*, 736-737.

Vermunt, J. K., Langeheine, R., & Bockenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics, 24*(2), 179-207.

Visher, C. A., & McFadden, K. (1991). A comparison of urinalysis technologies for drug testing in criminal justice. National Institute of Justice Research in Action. Washington, DC: U.S. Department of Justice.

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.

Copyrights