Issues surrounding Markov Latent Class Analysis for Assessing Measurement Error in Complex Surveys: A Case Study Using the National Crime Victimization Survey

Marcus Berzofsky[1] and Paul P. Biemer[1,2]
[1]RTI International, Research Triangle Park, NC, USA
[2]University of North Carolina, Chapel Hill, NC, USA
Corresponding author: Marcus Berzofsky, e-mail: berzofsky@rti.org

Key words: Markov Latent Class Analysis (MLCA), National Crime Victimization Survey (NCVS), measurement error, classification error

Abstract

Classification error, measurement error for categorical data, occurs when a respondent incorrectly identifies their true status. 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. MLCA is a useful technique because it does not require a gold standard (or error-free) data source, which is often not available in a survey setting, in order to estimate the classification error. However, there are many issues that need to be considered and addressed in order to ensure unbiased and identifiable estimates. This paper first discusses what issues should be considered. Then, the paper discusses how the authors addressed these issues in an analysis of the U.S. National Crime Victimization Survey (NCVS).

1. Introduction

1.1 Usefulness of Markov Latent Class Analysis to Assess Measurement Error

Classification error, measurement error for categorical data, occurs when a respondent incorrectly identifies their true status (Biemer, 2011). 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, therefore, it is not appropriate to assume that 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.

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.

1.2 Understanding the 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.

2. Issues Associated with Complex Survey Data

When conducting an MCLA with complex survey data there are usually two types of issues one should address: general issues and survey specific issues. General issues are common situations that arise when analyzing most complex surveys. Survey specific issues are situations that only occur because of the design or nature of the specific survey being analyzed.

2.1 General Issues

General issues one should address when conduction an MLCA include:

- Assumption violations. The first issue that needs to be addressed is determining if any model assumptions have been violated. Violations in one or more model assumptions can lead to biased estimates and invalid conclusions (Berzofsky, 2011). In addition, model violations can lead to weak identifiability (Berzofsky & Biemer, 2012). MLCA has 5 key assumptions are usually made
    - Simple random sample. This assumption is almost always violated in a complex survey setting
    - First-order Markov assumption. Presumes that the response probability at a given time point is only dependent on the most recent previous time point.
    - Independent classification error (ICE). Assumes that classification errors across waves are independent.
    - Time-homogeneous classification errors. The classification errors for an indicator are the same across all time points
    - Homogeneous error probabilities. All individuals in the same latent class have the same probability of being misclassified
- Sparse data. Sparse data occurs when too many of the cells in the data table are zero. Because the frequency of the data table is the cross product of all grouping variables and indicators in the model, this can occur when too many time points are being analyzed simultaneously or if the grouping variables have too many levels. Sparse data can cause issues with model convergence.
- Missing data. Most software used to conduct MLCA use list-wise deletion to handle cases with missing values. This applies to both grouping variables and indicators.

- Time-varying covariates. When dealing with longitudinal data covariates (survey items used for grouping variables) can either be static (i.e., not change over time) or time-varying (i.e., change over time in a non-linear fashion). Examples of time-varying covariates in the NCVS include marital status and how often a person goes out in the evening. See Berzofsky, Biemer, and Kalsbeek (2010) for more information on handling time-varying covariates in an MLCA.

2.2 Issues Specific to the NCVS

In addition to the general issues, there are often survey specific issues. These issues may not be unique to a particular survey, but may not be applicable to all surveys. For the NCVS two additional issues that need to be considered are:

- Rare events. Crime victimization is a relatively rare event (annual prevalence usually less than 10% for most crime types). This can lead to sparse data and unidentifiable models.

- Sensitive events. A sensitive event is an event where it is likely for a respondent to falsely indicate an event did not occur when it did (i.e., a false negative response), but unlikely to indicate an event did occur when it did not (i.e., a false positive response). Crime victimization is a sensitive event (especially for very serious types of crime like rape or sexual assault). This may allow one to make certain restrictions on the false positive rate to increase the available degrees of freedom.

3. Issues Investigated

When analyzing the NCVS we used the following techniques to address these issues (see Berzofsky, Biemer, & Kalsbeek (2011) for more details on the model fitting process and the measurement error estimates).

- Rare events. Due to the rare nature of crimes, in order to have our models converge, we needed to create composite crime measures. We developed two composite measures that we were able to analyze: less serious person crimes and household crimes. A third composite crime, more serious person crimes, was still too rare to be modeled.

- Sensitive data. While preliminary models did find the false positive rate to be small (between 1% and 3%), we did set any restrictions on it.

- Sparse data. We initially attempted to model all six bounded waves simultaneously. However, this proved to create a data table that was too sparse. Therefore, we created three sets of data using each set of four consecutive waves (i.e., waves 2 – 5, waves 3 – 6, and waves 4 – 7).

- Model assumptions. Our analysis developed nine models which we could compare to assess each of the model assumptions. For each assumption, we made the following model adjustments to test the assumptions:
    - Simple random sample. We use LatentGold (Vermunt & Magidson, 2005) to analyze the NCVS data. LatentGold allows one to incorporate the sample design parameters and utilizes pseudo-maximum likelihood in its estimation process.

- First-order Markov assumption. In order to test this assumption we looked at two other models for the structural component: second-order Markov and Mover-stayer. Second-order Markov allows the outcome to be dependent on the last two periods rather than only the most recent. Mover-stayer divides the respondents into two categories: movers, those who change status over time, and stayers, those that do not change across any period. Our analysis found that the first-order Markov assumption held.

- ICE. In order to test the ICE assumption we used the log-odds ratio test (LORC; Garret & Zeger, 2000). The LORC determines if a direct effect (e.g., the interaction of two indicators) will address the dependency between two time points. Our analysis found that there were no dependencies between any indicators.

- Time-homogeneous classification errors. For this assumption we tested models where the classification error rates between the first two time points and the second two time points were allowed to vary. We found that this was significant in the NCVS. Thus, our models allowed for this flexibility. In order for this model to be identifiable, we constrained the transition probabilities to be equal.

- Homogeneous error probabilities. For this assumption we added grouping variables. We tested both static and time-varying covariates. We found that for less serious person crimes age category, whether a person owns a home, race were the best grouping variables. For household crimes, age category, gender, and whether one owns a home.

- Missing data. In order to address missing data we restricted our models to those that responded to all four of the waves being analyzed. This increased the number of respondents that would have been included if we restricted the file to those that responded to all six waves.

4. Conclusions

MLCA is an effective tool for evaluating measurement error in complex surveys. However, there are several issues that need to be considered during the modeling process to ensure valid estimates. This paper discusses several of those issues and how we addressed them in an analysis of measurement error in the NCVS. Because our analysis was more interested in estimating the classification error in the NCVS rather than the true estimates of crime victimization, addressing these issues allowed us to create a model that was more flexible in estimating the measurement component by incorporating additional restrictions on the structural component. This tradeoff was acceptable for our purposes, but may not be applicable for a model with different goals.

References

Berzofsky, M.E. (2011). Methods and approaches for evaluating the validity of latent class models with applications. An unpublished dissertation.

Berzofsky, M. E., & Biemer, P. P. (2012). *Weak identifiablility in Latent Class Analysis*. In *Proceedings of the ASA Section on Survey Methodology, Joint Statistical Meetings*, San Diego, CA.

Berzofsky, M. E., Biemer, P. P., and 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, Vancouver, BC, 3785–3799.

Berzofsky, M.E., Biemer, P.P., Kalsebeek, W.D. (2011). Methods for Fitting a Markov Latent Class Analysis for the National Crime Victimization Survey. . In Proceedings of the ASA Section on Survey Methodology, Joint Statistical Meetings, San Diego, CA, 5351–5365.

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

Garrett, E. S., and Zeger, S. L. (2000). Latent class model diagnosis. Biometrics. 56(4): 1055–1067.

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

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

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

Van de Pol, R., and De Leeuw, J. (1986). A latent Markov model to correct for measurement error. Sociol Meth Res. 15: 118–141.

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

Vermunt, J. K., and Magidson, J. (2005). Technical Guide to Latent Gold 4.0: Basic and Advanced. Statistical Innovations, Belmont, MA.

Visher, C. A., & K. McFadden. (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.