ENUMERATING THE CORRECT NUMBER OF CLASSES IN A SEMIPARAMETRIC GROUP-BASED TRAJECTORY MODEL

by

Thomas James Blaze

B.S. in Business Administration, Duquesne University, 2004

Submitted to the Graduate Faculty of

Education in partial fulfillment

of the requirements for the degree of

PhD in Research Methodology

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

Education

This dissertation was presented

by

Thomas James Blaze

It was defended on

December 9, 2013

and approved by

Clement Stone, PhD, Professor, Psychology in Education

Feifei Ye, PhD, Assistant Professor, Psychology in Education

Tammy A. Chung, PhD, Associate Professor, Psychiatry

Dissertation Advisor: Kevin H. Kim, PhD, Associate Professor, Psychology in Education

ENUMERATING THE CORRECT NUMBER OF CLASSES IN A SEMIPARAMETRIC GROUP-BASED TRAJECTORY MODEL

Thomas James Blaze, PhD.

University of Pittsburgh, 2013

The semiparametric group-based trajectory model (GBTM), a special case of the more general growth mixture model, has been and increasingly employed technique for modeling heterogeneous change over time. A benefit of the GBTM is the ability to uncover distinct classes in the population that are characterized by their developmental trajectories. The characteristics of the developmental trajectories are affected by the number of classes extracted during estimation which in turn can affect inference, future investigation, and treatment or intervention. Thus, it is important that the measure(s) being relied upon for class enumeration are as accurate as possible. Only a handful of over 20 measures for class enumeration have been assessed in the context of GBTM using Monte Carlo methods prompting the need for a more thorough investigation. The purpose of this study was to determine if there were differences in the studied enumeration measures (information criteria, likelihood ratio test derivatives, and entropy based statistics and classification indices) abilities to correctly identify a true number of latent classes and to determine the common extraction errors for select enumeration measures in the context of a GBTM. A Monte Carlo study was performed and data were generated for true 4-class censored normal and binary logit models. Manipulated factors were sample size, the number of repeated measures, class mixing proportions, percent missing, and separation among the classes. Data were analyzed using a classification and regression tree approach. The results

demonstrated that there were differences in the enumeration measures abilities to correctly identify the true 4-class solution in both model types. Correct classification rates were highest when the separation among the classes was high, the class mixing proportions were either equal or moderately unequal, and the sample size was 800 or above. The information criteria had the most accurate classification rates while entropy statistics and classification indices had the least accurate classification rates. There were higher rates of under extraction errors overall but in certain conditions some of the enumeration measures showed a tendency to over extract classes. The Bayesian information criterion and the sample size adjusted Bayesian information criterion were the two measures recommended overall.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank the Psychology in Education faculty, Dr. Lane, Dr. Stone, Dr. Kim, and Dr. Ye for their guidance and support during my time in the program. In addition, I would like to thank Drs. Stone, Ye, and Chung for serving on my dissertation committee. I would like to give a special thanks to Dr. Kim for serving as my dissertation committee chair and being a great advisor. I would also like to thank my family for their patients, support and the motivation they provided throughout my entire time in graduate school.

1 INTRODUCTION

A semiparametric group-based trajectory model (GBTM; Land, McCall, & Nagin, 1996; Nagin, 1999; Nagin & Land, 1993) was developed for uncovering distinct but uncategorized classes of individuals who share similar growth trajectories thus allowing for the investigation of class characteristics within a population. For example, while studying child body mass index (BMI) between the ages of 4 and 10, Carter, Dubois, Tremblay, Taljaard, and Jones (2012) found 4 distinct developmental trajectories and examined characteristics of the classes such as socioeconomic status and mothers' smoking habits during pregnancy.

Often cited as a major benefit of the GBTM, classes of distinct developmental trajectories need not be identified prior to analyzing the data; rather classes are extracted as a part of a data analysis. This extraction process is commonly referred to as class enumeration and involves model comparisons. For example, a 3-class solution might be compared to 1-, 2-, 4-, and 5-class solutions in order to arrive at a conclusion that a 3-class solution is optimal. An optimal solution is one that includes the following characteristics: i) it is parsimonious, an overabundance of classes is not extracted, ii) it is encompassing, to the furthest extent possible the classes extracted provide the best explanation of development, and iii) it is interpretable, the developmental classes should make sense relative to the phenomena under investigation.

When employed in order to arrive at an optimal solution, class enumeration via model comparisons has proven to be somewhat problematic in GBTM and related fields. Tofighi and

Enders (2008) go so far as to suggest that extracting an optimal number of classes is one of the most challenging aspects in this type of analysis. The challenge can be attributed to a number of things. Although recommendations have been made (see, for example, Nagin, 2005), there is no common acceptance as to the best method to compare models with different numbers of classes. Even with statistics such as fit indices to guide the process, difficulty in selection between competing models is not uncommon. A theory can and should be used to help the process of class enumeration but this brings with it human subjectivity. Of these challenges to class enumeration, the method of model comparison is the one which can be refined the most relative to the GBTM. This study will utilize Monte Carlo methods to examine an extensive set of statistical measures commonly used for class enumeration in order to select from among competing GBTMs with different numbers of classes. The goal is to narrow down the choice of best enumeration measure(s) both overall and for certain conditions present in data.

1.1 GBTM: AN INCREASINGLY EMPLOYED STATISTICAL TECHNIQUE

The GBTM can be thought of as a combination of finite mixture model and regression. In finite mixture modeling, a finite number of distinct distributions are combined to represent 1 overall distribution of interest. Change over time is examined by regressing an outcome of interest on some measure of time. The resulting combination allows for the identification of distinct classes of developmental trajectories, similar to the distributions in finite mixture modeling, each with its own regression equation to explain how the outcome of interest changes with respect to the chosen measure of time. The unique nature of the GBTM makes it well suited to study particular types of problems. Nagin (2004) suggested use in situations where an outcome of interest has markedly different magnitudes of change for clusters of individuals in a population as opposed to a regular variation around 1 average trajectory. From the time that the GBTM was formally introduced as its own statistical method, an increasing number of applied researchers have been using the method to examine outcomes with markedly different magnitudes of change.

Bauer (2007) demonstrated an increasing use of the GBTM by examining the number of citations per year of Nagin (1999). It was the Nagin (1999) article that first provided applied researchers with a thorough discussion of the GBTM. A similar extended search was conducted and the results are displayed in Figure 1 (Web of Science, Social Science Citation Index, retrieved 2/26/2013). The number of citations increased very steadily from 2000 with 5 citations until 2010 with 88 citations, over a 1600% increase. There has been a slight drop off in the number of citations since 2010 but the numbers in 2011 and 2012 suggest that the GBTM is still being used quite extensively. The method is used in multiple disciplines including psychology, sociology, education, and medicine. Examples of topics studied in psychology are the developmental trajectories of depression, anxiety, and schedule burden (Choi et al., 2012) and

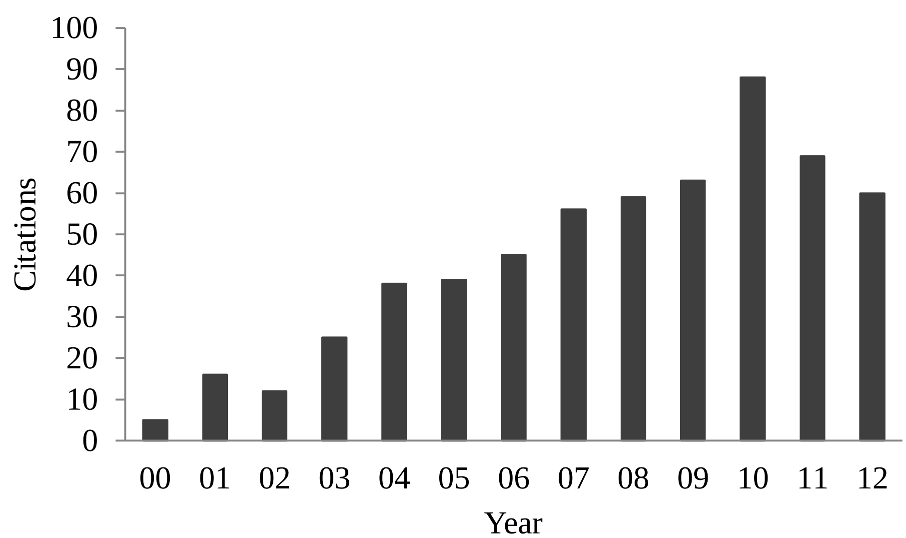

Figure 1. Number of citations by year of Nagin (1999)

the developmental trajectories of posttraumatic stress symptoms (Le Brocque, Hendrikz, & Kenardy, 2010). Examples of topics studied in sociology are the developmental trajectories of risky sexual behavior for individuals in their teenage years into early adulthood (Moilanen, Crockett, Raffaelli, & Jones, 2010) and developmental trajectories of delinquent behavior in adolescence (Reinecke, 2006). In education, the GBTM has been used to study developmental trajectories of mathematics achievement from kindergarten to third grade (Geary et al., 2009) and trajectories of children's educational aspirations and expectations (Lee, Hill, & Hawkins, 2012). Trajectories of overweight classification (overweight, not overweight) for children (Balistreri & Van Hook, 2011) and trajectories of atopic dermatitis in children (Kuss, Gromann, & Diepgen, 2006) have been studied in the medical field. For further discussion on the use of GBTM by discipline see Nagin and Odgers (2010).

1.2 JUSTIFICATION FOR THE STUDY

1.2.1 The use of enumerated classes

The prevalent use of the GBTM warrants further investigation of class enumeration measures for 2 main reasons. First, the importance of class enumeration stems beyond the classes themselves to the ways in which they are interpreted and employed. If a number of classes other than the optimal are enumerated, all subsequent interpretation, inference, and intervention may be adversely impacted. It has been repeatedly emphasized that the classes themselves do not represent literally distinct entities that exist in the population but that they can be used to facilitate a much deeper understanding of a developmental phenomenon being studied (Nagin, 1999; 2005; Nagin & Tremblay, 2005). Even though literally distinct classes do not likely exist, an argument can still be made for the need to select an optimal number of classes. The selection of an incorrect number of classes, that is, a non-optimal number, can have negative consequences. The resulting bias can affect inference and decision making. In discussing growth mixture modeling (GMM), which the GBTM has been shown to be a special case, Bauer (2007) stated, “I am particularly concerned with recovery of the latent class structure, which is often used to develop and evaluate taxonomic theories, identify problematic subgroups, and motivate targeted interventions” (p. 765). These concerns reemphasize the necessity of ensuring an optimal class number due to the manners in which the classes are used.

The running argument suggests that analyses using the GBTM are very rarely complete upon identification of classes. Rather, the identification of classes is typically a starting point for a much more thorough investigation into the problem of interest. The classes themselves are typically interpreted and labeled as part of the process of developing or evaluating taxonomic theories. Problematic classes are identified. Classes are often contrasted with respect to individual-level characteristics. Any individual-level characteristics that help distinguish the classes may then be used to formulate interventions. Predictors can also be added to the GBTM, a process recommended to occur after class enumeration (Nagin, 2005). Time invariant covariates can be added to the GBTM with the purpose of predicting the probability that an individual will belong to a particular class (Roeder, Lynch, & Nagin, 1999). Differential prediction of the covariate by class can also be tested. Another use of the GBTM involving the inclusion of covariates is to assess the alteration of the developmental trajectories for each class. Nagin, Pagani, Tremblay, and Vitaro (2003) discussed a method of including time varying covariates into the GBTM. The inclusion of time varying covariates can facilitate the investigation of 2 primary types of research questions. Within class effects of time varying covariates can be assessed by testing if there is a significant trajectory alteration by a time varying covariate for a particular class. Differences in trajectory alteration by a time varying covariate among the different classes can be assessed to determine the between class effect of the time varying covariate. Because the analytical features of GBTM, such as including covariates, occur after the class enumeration process, it is reasonable to assume that the number of classes extracted will ultimately impact any results.

The uses of the classes outlined above are all susceptible to problems but very much more so when an incorrect number of classes is selected. Example problems are provided for each class use. A common problem that can occur involves the interpretation of biased class trajectories. If a 2-class solution is selected over an optimal 3-class solution, classes in the 2-class solution may be somewhat or even very dissimilar from any of the classes in the 3-class solution. Essentially, the classes' developmental trajectories and other distinguishing characteristics can become qualitatively different between the 2 solutions. In this case, any interpretation or inference of the incorrect 2-class solution will be biased and invalid. This should be argument enough to confirm the importance of accurate and reliable enumeration methods. However, further discussion of more specific problems that can occur will be given to solidify the importance.

Treatments or interventions may target incorrect individuals if individual-level characteristics are used to distinguish problematic classes from non-problematic classes in a solution extracting an incorrect number of classes. If indeed distinguishing, individual-level characteristics will likely distinguish in a particular way given an optimal class solution. When too few classes are extracted, the class mixing that occurs can alter the makeup of the classes relative to individual-level characteristics. The same can occur with over extraction of classes. The makeup of classes will change with respect to individual-level characteristics when individuals are dispersed into the unnecessary classes. Individuals are often targeted for treatment or intervention based on the relationship between class membership and individual-level characteristics. Therefore, targeted individuals of treatment or intervention may change if the makeups of the classes change. As a result, time and money may be wasted on a treatment or an intervention for individuals who truly do not need it. Individuals who are truly in need of treatment or an intervention may be overlooked and miss out.

When including time invariant or time varying covariates into the GBTM, their effect can be mitigated by the number of classes extracted for a couple of reasons. When over extraction occurs, there will be fewer individuals within each class. With fewer individuals per class, the power to detect significant prediction of class membership and alteration of a class trajectory will diminish. When either over or under extraction occurs, the effect of covariates for each class will potentially change simply because the individuals within each class will change. Similar to finding distinguishing individual-level characteristics for each class, the under or over extraction effect on covariates may have an impact on treatments and interventions developed and targeted to specific individuals.

1.2.2 The lack of enumeration measure performance assessment

One of the most common ways to choose between 2 competing models is to use the likelihood ratio difference test (LRT). However, as will be discussed in more detail, the regulatory conditions of the likelihood ratio difference test break down when comparing models with a different number of classes. This led researchers to find or devise other methods of comparing competing models for use with finite mixture models. McLachlan and Peel (2000) discussed many of the commonly used as well as some less commonly used class enumeration measures. The measures can be classified as information criteria, entropy statistics and entropy-penalty based indexes (classification indices), and likelihood ratio test derivatives (Peugh & Fan, 2012). Two of the most commonly used measures for model comparison other than the LRT are the information criteria, Akaike's Information Criterion (AIC; Akaike, 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978). These information criteria have been advocated for the use of comparing models in a variety of statistical techniques such as regression (Kutner, Nachtsheim, Neter, & Li, 2005), factor analysis (Akaike, 1987; Preacher & Merkle, 2012), and structural equation modeling (Kaplan, 2009) in addition to GBTM and other types of finite mixture models. AIC was originally created to correct for biases that occur when attempting to compare multiple models using their respective maximum likelihoods. As a measure of model comparison, the maximum likelihood will always favor a more complicated model. Akaike (1974) suggested that AIC can be used for model comparisons that involve models differing in their parameter space or that share a similar parameter space but place different restrictions on a parameter vector. That is, AIC and other information criteria such as BIC have been recommended for use comparing both nested and non-nested models. The information criteria all take on a similar form of penalized log likelihood. The likelihood measures the fit of the model and the penalty term makes an adjustment for model complexity. Although AIC has been widely used in various statistical techniques, as will be discussed in more detail, AIC was demonstrated to be asymptotically inefficient because its penalty term does not depend on sample size. The development of BIC and additional information criteria was driven by the asymptotic inefficiency of AIC and well as other various reasons. The majority of these information criteria have been used for class enumeration in various types of finite mixture models. The various forms of information criteria, classification indices such as the classification likelihood criterion (CLC; Biernacki & Govaert, 1997), and likelihood ratio test derivatives such as the Vuong-Lo-Mendell-Rubin (VLMR) test (Lo, Mendell, & Rubin, 2001) will all be discussed in more detail.

Extensive sets of class enumeration measures, including those aforementioned, have been thoroughly investigated using Monte Carlo methods for various types of finite mixture models: latent class analysis (Yang, 2006), factor mixture models (Nylund, Asparouhov, & Muthén, 2007), mixture structural equation models (Henson, Reise, & Kim, 2007) and GMM (Peugh & Fan, 2012; Tofighi & Enders, 2008). Classes may not be literally distinct entities existing in the population but there is a solution to this dilemma. A Monte Carlo investigation allows for the optimal number of classes to be known and facilitates the investigation of the best enumeration measure(s). Although there is still no agreement on 1 best enumeration measure for the statistical methods mentioned above, recommendations from the Monte Carlo studies have limited the large set of measures to consider and have helped guide researchers' decisions on which enumeration measure(s) to choose.

Although GBTM is a type of finite mixture model, it is different than those in which the performances of class enumeration measures have been studied. GBTM requires a longitudinal design whereas latent class analysis, factor mixture models, and mixed structural equation models do not. GMM is another type of finite mixture model designed for longitudinal data but GBTM is a less complex model. Many of the enumeration measures to be investigated take into account model complexity based on the number of estimated parameters. As a result, it is likely that the enumeration measures will behave differently in the context of GBTM as opposed to GMM with some being more or less sensitive to the change in the number of classes.

Despite the differences between GBTM and the other types of finite mixture models for which the performances of class enumeration measures have been studied, a thorough Monte Carlo investigation of the performance of enumeration measures has not been conducted for the GBTM. Brame, Nagin, and Wasserman (2006) examined class enumeration measures in a model similar to the GBTM with a Monte Carlo study. The study was limited to 3 enumeration measures and focused on count data at only 1 time point. Had their model contained multiple time points it would have been a GBTM. Thus, it would be useful to conduct a more exhaustive investigation of the enumeration measures in the context of the GBTM because it has not been done and because trajectories other than count data can be investigated using the method.

1.3 PURPOSE AND RESEARCH QUESTIONS

The current Monte Carlo study is intended to fulfill the need for a more exhaustive investigation of the performances of enumeration measures in the context of the GBTM and to refine the class enumeration process so that problems associated with over or under extraction are controlled. Based on the premise that the number of latent classes is known (a true-class solution); multiple enumeration measures will be investigated with the intention of determining the best measure(s) for selecting a model with a true number classes from competing models with different numbers of classes in a GBTM.

This study focused on developmental trajectories of normally distributed and binary outcomes. Educational outcomes are typically continuous variables such as test scores or binary variables such as meeting or not meeting a minimum level of achievement. An example of the latter is a statewide standardized test that measures whether a student has scored at a specific proficiency level or not. Simulated factors were sample size, the number of repeated measures for a fixed time period, class mixing proportions, percent of missing data, and the distance between the trajectories (class separation).

There were 2 research questions in this study. The primary research questions focused on the rate of correct class enumeration. The secondary research question focused on the direction of extraction errors, under extraction or over extraction. Are there differences in the enumeration measures abilities to correctly identify a true number of latent classes? What are the common extraction errors for select enumeration measures?

A description of the GBTM and class enumeration measures will be provided in Chapter 2. In Chapter 3, the study design, outcome measures, factors, and analysis plan will be discussed. Results of the Monte Carlo study will be provided in Chapter 4 followed by a discussion of the study limitations and implications in Chapter 5.

2 LITERATURE REVIEW

2.1 THE GROUP-BASED TRAJECTORY MODEL

2.1.1 Model specifications

Recall that GBTM was developed for the intended purpose of uncovering distinct but uncategorized classes of individuals who share similar growth trajectories. The GBTM differs from other procedures designed to model growth over time (growth trajectories). A latent growth curve model (LGM) (see, for example, McArdle & Epstein, 1987; Meredith & Tisak, 1990; Rogosa & Willett, 1985) and multilevel modeling approach for examining growth over time (see, for example, Bryk & Raudenbush, 1987; Raudenbush & Bryk, 2002) both estimate 1 overall average growth trajectory for all individuals. These methods assume a multivariate normal distribution for the population trajectories (Nagin, 2005). These models, which are considered unconditional models, estimate the mean and covariance structure of the growth trajectory parameters. A drawback of the unconditional model approach is that assuming a continuous distribution of growth trajectories sometimes lead to an incomplete picture of the true growth pattern when there are heterogeneous patterns. GBTM seeks to overcome this obstacle by using a multinomial modeling strategy designed to identify relatively homogenous clusters of developmental trajectories which in turn suggests that the population is composed of a mixture of distinct classes that are defined by their developmental trajectories (Nagin, 1999). Stated differently, the distributions of growth trajectory parameters are conditional on class membership.

As was mentioned, the GBTM can be thought of as a type of finite mixture model. McLachlan and Peel (2000) discuss the general form of a finite mixture model. Let y_i be an observed value of the random vector $\boldsymbol{y}_i$ which when concatenated over individuals, i, produces the entire sample $\boldsymbol{Y}$. An individual probability density function can be written as $f(y_i)$ which in turn can be displayed as a composition of distinct distributions in the form

$$f(y_i) = \sum_{j=1}^{J} \pi_j f^j(y_i) \tag{1}$$

where $f^j(y_i)$ are densities for (j = *1, …, J*) distributions (sometimes distinguished as J classes[1] with different distribution characteristics) and π_j are population proportions making up each distinct distribution. Given that π_j are proportions they cannot be negative and when summed over J will sum to 1. The most common type of mixture involves normal distributions. An example of the probability density function in Equation 1 with 2 classes and normal components is

$$f(y_i) = \pi_1 \phi(y_i|\mu_1, \sigma^2) + \pi_2 \phi(y_i|\mu_2, \sigma^2) \tag{2}$$

where $\phi(\cdot)$ represents the normal density function. An examination of the probabilities making up the GBTM likelihood function will lead to a model form that appears very similar to that of Equation 1.

The GBTMs utilization of latent classes results in the probability of an individual's observed trajectory given a class membership being specified as:

---

[1] j will also be used to index competing models because a different number of classes implies a different model.

$$P^j(Y_i) = \prod^{T} p^j(y_{it}) \tag{3}$$

where $p^j(y_{it})$ is the probability distribution function of the observation of individual *i* at time *t* given membership in class *j* (Nagin, 1999; 2005). The computation of the probability as demonstrated in Equation 3 requires an assumption of conditional independence. The conditional independence assumption primarily serves to reduce the complexity of the model. According to Nagin (2005), the conditional independence assumption implies that the probability of the observed value for individual *i* at time *t* is independent of observed values for the same individual at previously measured time points conditional on membership in class *j*. Although the assumption seems somewhat implausible, it can be justified by examining class level conditionality. Within a class, individual observations will vary randomly around an average trajectory, that is, there will be no predictable pattern to the residuals around the average trajectory. In the population, a class membership is not identified and observed values within an individual will appear to be correlated.

In addition to the probability of an individual's observed trajectory given a class membership, the proportion of individuals making up each distinct trajectory class is required to compute the GBTM likelihood function. This proportion for the *j*-th class will be denoted π_j. The proportions are not estimated directly. The proportions π_j are linked to a set of parameters, θ_j, *j* = 1,…, *J*, using the equation

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{j} e^{\theta_j}}. \tag{4}$$

Unlike π_j, the parameters in the set θ_j are not bound between 0 and 1 and can take on any value, (-∞, ∞). These are optimal conditions for estimation because as Nagin (2005) suggests, it is difficult to force the constraint of π_j being between 0 and 1 during estimation.

The products of the conditional probability of an individual's observed trajectory and the proportion of individuals making up each distinct trajectory class are summed across all *J* classes to form the unconditional probability of the data

$$P(Y_i) = \sum_{j=1}^{J} \pi_j P^j(Y_i)\,. \tag{5}$$

The probability of observing an individual's trajectory given membership in each class is weighted by the proportion of individuals within each of the corresponding classes. This unconditional probability is very similar to the general form of the finite mixture model in Equation 1.

The likelihood function for the entire sample is the product of the unconditional probabilities in Equation 5 taken across all *N* individuals

$$L = \prod_{i=1}^{N} P(Y_i)\,. \tag{6}$$

This is the general form of the likelihood function for the GBTM. However, depending on the scale of the response variable, the specific form of the model and thus the likelihood function can change.

The GBTM has been developed for repeated measures of Poisson, binary, or censored normal data. The different data types require different probability distribution functions and estimation of different parameters but the respective likelihood functions can all be specified in the form given in Equation 6. The models for the censored normal and binary logit distributions are the focus of this study and so further discussion is warranted.

The model for the censored normal distribution with quadratic trajectory[2] is:

$$y_{it}^{*} = \beta_0^j + \beta_1^j X_{it} + \beta_2^j X_{it}^2 + \varepsilon_{it} \qquad (7)$$

where y_{it}^{*} represents the latent outcome for person *i* at time *t*, X_{it} represents a measure of time such as age for individual *i* at time *t*, β_0^j is the value of the outcome for class *j* when X_{it} is zero, and β_1^j and β_2^j represent the linear and quadratic growth parameters for class *j* respectively. Finally, ε_{it} is a disturbance term for person *i* at time *t* and is normally distributed with a mean of zero and variance of σ^2. The latent outcome y_{it}^{*} is not directly observed due to censoring. In order to account for the censoring, Nagin (1999) used a link between the observed outcome y_{it} and the latent variable y_{it}^{*}

$$\begin{aligned} y_{it} &= S_{min} \text{ if } y_{it}^{*} < S_{min} \\ y_{it} &= y_{it}^{*} \text{ if } S_{min} \le y_{it}^{*} \le S_{max}, \text{ and} \\ y_{it} &= S_{max} \text{ if } y_{it}^{*} > S_{max} \end{aligned}$$

where S_{min} and S_{max} are the scale minimum and maximum respectively.

The specification of the likelihood function for censored normal data incorporates the link between the latent response variable and its observed counterpart to obtain $p^j(y_{it})$ using the cumulative distribution function, $\Phi(\cdot)$, and density function, $\phi(\cdot)$, for a normal random variable with a mean of $\beta^j X_{it}$ and standard deviation σ. As demonstrated in Nagin (2005), the probability $p^j(y_{it})$ can be expressed as:

[2] Up to a fourth-order polynomial is allowed using Proc Traj (Jones et al., 2001; Jones & Nagin, 2007).

$$p^j(y_{it} = S_{min}) = \Phi\left(\frac{S_{min} - \beta^j x_{it}}{\sigma}\right), \tag{8.1}$$

$$p^j(y_{it}) = \frac{1}{\sigma}\phi\left(\frac{y_{it} - \beta^j x_{it}}{\sigma}\right) \text{ for } S_{min} \leq y_{it} \leq S_{max}, \text{ and,} \tag{8.2}$$

$$p^j(y_{it} = S_{max}) = 1 - \Phi\left(\frac{S_{max} - \beta^j x_{it}}{\sigma}\right). \tag{8.3}$$

The model for the binary logit distribution with quadratic trajectory can also be expressed using Equation 7 (Nagin, 1999; 2005). For the binary logit model, the latent variable y_{it}^* represents the log of the odds that the observed outcome $y_{it} = 1$ as opposed to $y_{it} = 0$ for individual *i* at time *t*. For the condition $y_{it}^* > 0$, it is assumed that $y_{it} = 1$ and for $y_{it}^* \leq 0$ it is assumed that $y_{it} = 0$ (Nagin, 2005). To arrive at this specification, it is necessary to transform the outcome y_{it} using the logit function,

$$g(y_{it}) = \text{logit} = \log\left(\frac{\alpha_{it}^j}{1 - \alpha_{it}^j}\right), \tag{9}$$

where α_{it}^j represents the probability $y_{it} = 1$ given membership in class *j*. The model is expressed in the form of the binary logit distribution as:

$$\alpha_{it}^j = \frac{\exp\left(\beta_0^j + \beta_1^j X_{it} + \beta_2^j X_{it}^2 + \varepsilon_{it}\right)}{1 + \exp\left(\beta_0^j + \beta_1^j X_{it} + \beta_2^j X_{it}^2 + \varepsilon_{it}\right)} \tag{10}$$

To arrive at the likelihood function in Equation 6, α_{it}^j is substituted in place of $p^j(y_{it})$ in Equation 3.

2.1.2 Estimation and interpretation

The GBTM, in either censored normal or binary logit form, is estimated via maximum likelihood (ML) using a general quasi-Newton procedure in Proc Traj, a SAS procedure specifically designed for the GBTM (Jones, 2012; Jones, Nagin, & Roeder, 2001; Jones & Nagin, 2007).

Standard errors are estimated by inverting the observed information matrix (Jones et al., 2001). Estimation of the GBTM requires the estimation of the class specific trajectory parameters, $\beta^{j\prime}s$, the mixing proportion parameter(s), θ_j, and the error variance, σ^2. For example, assuming a 3-class solution where each class is specified to have a quadratic trajectory, there are 12 parameters to estimate. The 12 parameters are the 3 trajectory coefficients, intercept, linear, and quadratic for each class, 2 class proportions (the third proportion can be determined from the 2 already estimated), and the variance which is constrained to be equal across classes.

The ML estimates of the GBTM have been examined with respect to the asymptotic properties of being unbiased and normally distributed. Loughran and Nagin (2006) studied the 2 desired asymptotic properties of the ML estimates for Poisson data by examining the empirical sampling distributions for a range of sample sizes (500, 1000, 1500). The samples were selected from 1 large sample (13000) whose parameter estimates had very small standard errors and were intended to represent the true parameter values. Laughran and Nagin (2006) concluded that even in relatively small samples (500), the ML estimates are unbiased and their sampling distributions are approximately normally distributed.

In the GBTM, class enumeration occurs during the estimation process. The number of classes must be specified ex ante in order to estimate the model. However, one of the primary goals of the GBTM is to uncover the optimal number of classes for explaining development over time. The solution is to compare multiple models, each estimated with a different number of classes. As mentioned, these model comparisons serve to determine the model with the best fit to the data and ultimately the optimal number of classes. For model comparisons to determine the optimal number of classes, Nagin (1999; 2005) advocated the use of BIC. Nagin (2005) suggested that BIC tends to be conservative, is consistent, and performed well in selecting the correct number of classes in various simulation studies. Furthermore, Nagin (2005) petitioned that BIC should not be the only criterion considered when selecting the correct number of classes; rather BIC and underlying theory should be used in tandem. In spite of Nagin's endorsement of the BIC to determine the optimal number of classes in a GBTM, the consideration of other enumeration measures remains necessary due to the limited investigation of such measures relative to the GBTM. A thorough discussion of the enumeration measures to be considered will be provided after the remaining explanation of the GBTM. The remaining explanation considers the selection of the correct order of the polynomial for each class, the key outputs of the GBTM, and the suggested interpretations for the outputs.

Besides the optimal number of classes, the correct order of the polynomial for each class is another consideration when selecting the final GBTM. Nagin (2005) advocated a 2-stage process for selecting a final model. In the first stage, when selecting the optimal number of classes, a pre-determined order is specified for all classes. For example, all classes could be specified to have a quadratic trajectory. Once the optimal number of classes is determined, the correct order polynomial for each class can be selected based on model comparisons.

The selection of a final model leads to the assessment of the 3 key outputs of GBTM. As stated by Haviland and Nagin (2005), these 3 key outputs are the developmental trajectory of each class, the class mixing proportions, and the posterior probability for an individual belonging to a particular class.

The developmental trajectory of each class is defined by the highest order polynomial and the values of the trajectory coefficients, β^j's. The developmental trajectories are important because they provide insight into the different developmental patterns individuals or other objects of interest follow with respect to some outcome of interest. Stated differently, the developmental trajectories provide a view of the different ways individuals may change over a given time period. The distinct developmental trajectories of each class also serve to explain heterogeneous variability in the population. Nagin and Tremblay (2005) discuss that it is unlikely the population is made up of truly distinct classes. When uncovering developmental trajectories though the selection of the optimal number of classes, the idea is to select as simple of a model as possible that emphasizes all of the distinct features of the population trajectory distribution. Nagin and Tremblay (2005) make clear to emphasize that the idea is not to select the true number of classes as it does not likely exist. Nagin and Tremblay (2005) also point out that the number of classes and the developmental trajectory of each class are functions of the sample size. As the sample size increases, the amount of information available from which to form new and distinct classes also increases. The sample size increase can be of the form of increasing the number of individuals or increasing the number of repeated measures for each individual. Of the 2, the latter has been shown to have a much greater impact.

Another view of the developmental trajectories for each class is that they represent the average behavior trend for a set of individuals (Nagin & Tremblay, 2005). Individuals in a particular developmental trajectory class are not expected to follow the class trajectory exactly just like individuals in a particular social class are not all expected to share the same average income level. There will be random variation of individual trajectories within a given trajectory class.

The distinct developmental trajectories provide a picture of the different ways a particular characteristic of an individual may change over time. The estimated population mixing proportions for each class are intended to provide some insight into the proportion of individuals following one of the distinct developmental trajectories. The mixing proportions are also used in the computation of the posterior probability of class membership.

The posterior probability of a class membership is the probability that an individual belongs to class *j* given their observed trajectory

$$\hat{P}(j|Y_i) = \frac{\hat{P}(Y_i|j)\hat{\pi}_j}{\sum_{j=1}^{J} \hat{P}(Y_i|j)\,\hat{\pi}_j}. \tag{11}$$

As demonstrated through the use of estimates in Equation 11, the posterior probability is computed post estimation with the model parameter estimates. According to Nagin (2005), the posterior probabilities are important because they can be used to find commonalities among members of a particular class, can be used in assessing model fit, and for computing weighted class statistics. Class memberships are also assigned using the posterior probabilities. An individual is assigned to the class for which they have the highest posterior probability. Once class assignments are made, characteristics of individuals within each class can be assessed. Nagin (2005) suggested cross tabulations of class memberships with other characteristics that might be distinguishing factors among the trajectory classes.

Nagin (2005) discusses 3 ways that the posterior probabilities can be used for judging model adequacy, the average posterior probability, the odds of correct classification, and the estimated class proportions versus the assigned proportions. The latter suggests that the estimated class proportions and assigned proportions based on the posterior probabilities should be fairly similar in order to assume good model fit. In addition, the posterior probabilities have been incorporated into various enumeration measures as will be demonstrated in the discussion of such measures.

The average posterior probability ($\overline{PP}_j$) for each class is computed by taking the average posterior probability of belonging to class *j* over all of the individuals classified in class *j* using their maximum posterior probability. Average posterior probabilities closer to 0 are indicative of poor model fit while average posterior probabilities closer to 1 are indicative of good model fit.

The odds of correct classification are computed by:

$$OCC_j = \frac{\overline{PP}_j/1 - \overline{PP}_j}{\hat{\pi}_j/1 - \hat{\pi}_j}. \tag{12}$$

An OCC_j value of 1 suggests that using the posterior probability for assigning individuals to classes is essentially a random process suggesting poor model fit. As the OCC_j increases, the accuracy of assignment increases.

Even though individuals are assigned to a particular class based on their posterior probability, as was previously mentioned, Nagin and Tremblay (2005) are careful to point out that the classes are not reality and individuals cannot be assigned to their true class. The creation of classes and assignment of individuals is simply a method of trying to organize complex information such that it becomes more insightful. Classifying individuals with similar trajectories facilitates questions such as one that was mentioned previously: what characteristics do individuals with similar trajectories have in common. This reemphasizes the point that, assuming an optimal number of classes exists, the answers to these important questions can be biased if under or over extraction of classes occurs.

2.2 CLASS ENUMERATION MEASURES

2.2.1 Introduction

It has already been demonstrated that the GBTM is a type of finite mixture model. Other statistical methods that utilize finite mixture models include but are not limited to latent class analysis (LCA; Muthén, 2001), factor mixture models (FMA; Lubke & Muthén, 2007), and GMM (Muthén, 2001; Muthén & Shedden, 1999). As mentioned by McLachlan and Peel (2000), finite mixture models have also found their way into multiple disciplines including biology, genetics, medicine, psychiatry, and economics to name a few. This widespread use of finite mixture models can be linked back to their intended purpose.

Suppose that the distribution of a variable in the population of interest has an unknown shape and thus an unknown density function. A way of modeling the unknown distributional form is to use a mixture of normal densities. This is one reason for employing a finite mixture model as discussed in Bauer and Curran (2003) and McLachlan and Peel (2000). The other purpose for using finite mixture models is to uncover individuals in the population who can be clustered together in some way to form distinct classes. GBTM employs finite mixture models for the latter purpose. The GBTM attempts to uncover latent classes of distinct developmental trajectories. As previously suggested, the investigation of enumeration measures relative to the GBTM is limited. This study focuses on class enumeration in order to facilitate a broader understanding of available enumeration measures relative to GBTM. The collection of class enumeration measures that will be utilized in this study come from a variety of statistical methods but have all been advocated for use and tested with respect to some type of finite mixture model.

Testing for the optimal number of classes in a finite mixture model is typically accomplished by comparing different models to determine which model is a better fit to the data. In many situations, the statistical comparison of models is done using the LRT which follows a χ^2 distribution with *df* equal to the difference in the number of parameters for the 2 nested models

$$\text{LRT} = 2\left(\log\left[L\{\hat{\Omega}_1\}\right] - \log\left[L\{\hat{\Omega}_0\}\right]\right). \tag{13}$$

Log is the natural logarithm, $L\left(\hat{\Omega}_h\right)$ is the maximum likelihood over the parameters estimates $\hat{\Omega}$ in the null hypothesis model, h = 0, versus the alternative hypothesis model, h = 1. When comparing GBTM for *j* versus *k* classes where $(j \neq k)$, the null hypothesis will contain $\pi_j = 0$ which is on the boundary of the parameters space and thus the LRT cannot be asymptotically χ^2 distributed as discussed in Aitkin and Rubin (1985) and Everitt (1981). Another problem with the LRT in this situation is that the *j* class versus the *k* class solutions may not be nested (Nagin, 2005). For example, consider a 1-class versus a 2-class solution. The individuals from each class in 2-class solution will be combined in the 1-class solution which can cause problems when attempting to determine the correct *df* for the LRT. If the 1-class solution is the optimal solution, the incorrect 2-class solution may produce 2 classes with very similar trajectory coefficients or the estimated proportion of individuals in 1 class may approach zero. Either of these conditions makes it difficult to determine what the true *df* really should be. The problem with the LRT test in the context of finite mixture models has led to a plethora of research for the most efficient class enumeration measure.

2.2.2 Information Criteria

Information criteria were first introduced by Akaike (1974) who created AIC. The criterion was developed to make comparisons among multiple competing models and has since become the foundation for a slew of information based criterion used to select the best model from among competing models. Many of these information criteria can take on a general form as discussed by Sclove (1987),

$$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + a(n)p_j + b(j,n), \tag{14}$$

where $L\left(\hat{\Omega}_j\right)$ is the maximum likelihood over the parameter estimates of the j-th model, $a(n)$ is a penalty term which may or may not vary with sample size, p_j is the number of parameters in model j and $b(j,n)$ is an additional term used for some of the criterion. For the criterion discussed, $b(j,n)$ is equal to 0 and will not be mentioned. This form demonstrates that the information criteria are penalized forms of the log likelihood where the penalty depends on the number of parameters and for some criterion is a function of the sample size. The criteria not only take into account the goodness-of-fit of a model but also the complexity needed to achieve that fit as pointed out by Sclove (1987). Information criteria that take on this form indicate preferred models with smaller values.

Table 1 displays the form of each enumeration measure investigated in this study and how each should be interpreted to determine which model to select. AIC has a constant penalty term of 2, that is, $a(n)$ from Equation 14 is equal to 2 regardless of sample size, AIC in Table 1. In order to choose model $j+1$ over model j with $p_{j+1} > p_j$, the penalty term must not exceed the improvement in model fit as measured by the log likelihood. One of the biggest arguments against the use of AIC for determining the number of classes in a finite mixture model is that the

Table 1. Equations and model selection criteria for the enumeration measures investigated

Measure	Form	Model Selection
AIC	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + 2p_j$	smallest
CAIC	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + ([\log\{n\}] + 1)p_j$	smallest
ssCAIC	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + \left(\left[\log\left\{\frac{n+2}{24}\right\}\right] + 1\right)p_j$	smallest
BIC	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + \log(n)\,p_j$	smallest
B_{10}	$e^{BIC_0} - e^{BIC_1}$	2 or above
ssBIC	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + \log\left(\frac{n+2}{24}\right)p_j$	smallest
HQ	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + \log(\log[n])\,2p_j$	smallest
ssHQ	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + \log\left(\log\left[\frac{n+2}{24}\right]\right)2p_j$	smallest
VLMR	$\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f\left[y_i\middle|\hat{\Omega}_1\right]}{g\left[y_i\middle|\hat{\Omega}_0\right]}\right)^2$	$p \leq .05$, choose alternative model
aVLMR	$\frac{VLMR}{1 + ([p-q]\log[n])^{-1}}$	$p \leq .05$, choose alternative model
sE	$1 - \frac{-\sum_{j=1}^{J}\sum_{i=1}^{n} t_{ij}\log\left(t_{ij}\right)}{n\log(j)}$	largest
NEC	$\frac{E(j)}{L(j) - L(1)}$	smallest
CLC	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + 2E(j)$	smallest
ICL-BIC	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + \log(n)\,p_j + 2E(j)$	smallest
ssICL-BIC	$-2\log\left(L\left[\hat{\Omega}_j\right]\right) + \log\left(\frac{n+2}{24}\right)p_j + 2E(j)$	smallest

penalty term is not a function of sample size and thus it is not asymptotically optimal (Schwarz, 1978). This means that the probability of selecting the correct number of classes does not approach 1 as the sample size increases. Furthermore, the use of a constant penalty term causes AIC to consistently overestimate the number of classes (Celeux & Soromenho, 1996).

A proposed improvement of AIC was given by Bozdogan (1987). The improvement was intended to correct the asymptotic inconsistency of AIC. The improvement was also intended to provide more liberal penalties for increased parameterization of a model. In the finite mixture model context this could mean including more classes. The criterion was called the Consistent Akaike Information Criterion (CAIC) and has the form shown in Table 1 where n denotes the sample size. The penalty term, which would take the place of $a(n)$ from Equation 14, is $([\log\{n\}] + 1)$. Thus, the penalty term for CAIC is larger than that of AIC for all $n > 2$.

Also working to make an improvement over AIC, Schwarz (1978) studied the asymptotic behaviors of Bayes estimators with specific priors and produced BIC which maintained the maximum likelihood estimate as the leading term, BIC in Table 1. The penalty term, which takes the place of $a(n)$ from Equation 14, is $\log(n)$. This penalty term is larger than the AIC penalty term for $n > 8$. Even though the overestimation problem of AIC was overcome by BIC in most contexts, Celeux and Soromenho (1996) suggested that BIC may consistently underestimate the number of classes. Furthermore, Biernacki, Celeux, and Govaert (1998) suggested that BIC may actually overestimate the number of classes given poor fit of the data to the finite mixture model. Specifically, Beirnacki et al. (1998) demonstrated that when a mixture of normal distributions is used to model the true mixture of a uniform and a normal distribution, BIC tends to overestimate the number of classes.

Unlike the likelihood ratio difference test, the information criteria discussed thus far and those yet to be discussed do not have any formal decision rules to suggest how large of a difference determines which of 2 competing models is considered optimal. Subjective judgment is often required to determine how large of a decrease in the criterion value should result in the choice of model $j + 1$ over model j where $p_{j+1} > p_j$. However, with the BIC criterion, an approximation of the Bayes factor (B_{10}) has been suggested when 2 competing models are assumed to be equally likely a priori (Kass & Wasserman, 1995). This form of the Bayes factor is displayed in Table 1. When a Bayes factor is used to select between competing models, it is the ratio of the posterior odds for the 2 models to the prior odds for the 2 models. The important distinction about B_{10} is that recommendations have been made regarding the support it demonstrates for the selection of one model over another. The recommendations for interpreting B_{10} are based on Jeffreys (1961) and were obtained via McLachlan and Peel (2000), Table 2. The potential benefit of using BIC in this fashion is the reduction of subjectivity when selecting between competing models that differ in their number of classes.

Table 2. Scale interpretation of the Bayes factor

B_{10}	2 log B_{10}	Evidence for H_1
< 1	< 0	Negative (supports H_0)
1 to 3	0 to 2	Barely worth mentioning
2 to 12	2 to 5	Positive
12 to 150	5 to 10	Strong
> 150	> 10	Very strong

A sample size adjustment to BIC (ssBIC) was proposed by Sclove (1987). The adjustment reduces the penalty term of BIC and therefore should theoretically reduce the chance of underestimating the number of classes in a finite mixture model. Yang (2006) suggested that the ssBIC will be an improvement over the BIC in 2 particular situations, when the sample size is small or the number of parameters is large. The form of ssBIC is displayed is Table 1. The penalty term which takes the place of $a(n)$ from Equation 14 is $\log\left(\frac{n+2}{24}\right)$. This penalty will always be smaller than that of BIC and CAIC and will only be larger than the penalty term of AIC for sample sizes of 176 and up. Results for ssBIC have generally been favorable. Tofighi and Enders (2008) found ssBIC to perform quite well in the context of GMM with a slight tendency to underestimate the true number of classes whereas Nylund et al. (2007) showed slightly more favorable results for BIC.

Hannan and Quinn (1979) proposed another adjustment of AIC. They suggested that the penalty term should be multiplied by the twice the log of the sample size. The form of HQ is displayed is Table 1. The penalty term which takes the place of $a(n)$ from Equation 14 is $\log(\log[n])\,2$. This penalty term is larger than the penalty term of AIC and ssBIC for sample sizes greater than 15 and smaller than 1207, respectively. The penalty term for HQ is always smaller than the penalty term of BIC and CAIC. There has not been any historical discussion of HQ in the context of class enumeration in finite mixture models. There have been a few recent studies that have investigated its performance in 2 different types of finite mixture models. These studies will be discussed in some detail at a later point.

2.2.3 Likelihood ratio test derivatives

The information criteria discuss thus far are not considered formal statistical tests for determining which of 2 competing models is best. As was mentioned, there is no guideline to suggest how large of a difference there needs to be in a given information criteria value to suggest one model should be selected over another. A derivative of the LRT was developed by Lo et al. (2001) to overcome the LRT's violation of regulatory conditions when used to compare models with different number of classes. Their work was an extension of the work by Vuong (1989). The statistic they developed was specifically designed for determining the number of classes in a mixture model. The VLMR test statistic is displayed in Table 1 where $\boldsymbol{g(y_i|\widehat{\Omega}_0)}$ and $\boldsymbol{f(y_i|\widehat{\Omega}_1)}$ are the probability density functions for the parameter estimates of the null and the alternative model respectively. According to Lo et al. (2001), the asymptotic distribution of the VLMR test statistic is a weighted sum of $p + q$ independent $\boldsymbol{\chi^2_1}$ random variables where p and q represents the number of parameters in the null model and the alternative model respectively. At the time of development, Lo et al. (2001) suggested that their test worked well when used in the context of homoscedastic normal mixtures. They did acknowledge that the test suffered from an inflated type I error rate and suggested an adjusted form of the VLMR test statistic (aVLMR), Table 1. The aVLMR did improve the type I error rate but was found to have poor power especially when sample sizes were small. The VLMR statistic was further scrutinized by Jeffries (2003) who demonstrated that the theorems proposed by Vuong (1989), on which the VLMR is validated, do not hold under certain conditions. If the parameters from a 1-class solution are contained within the parameter space then some parameters are unidentified and difficulties can occur when inverting the necessary information matrix. On the other hand, if parameters from a 1-class solution are excluded from the parameter space, the likelihood ratio test will move toward negative infinity. Jeffries (2003) also conducted a simulation study and found that the proportions for the VLMR statistic do not adhere to a uniform distribution as would be expected.

2.2.4 Entropy statistics and classification indices

Entropy statistics and classification indices are other types of measures used for determining the number of classes in a finite mixture model. The classification indices use a classification likelihood (complete data likelihood) (CL) which has been shown to have a direct link to an entropy measure. Entropy is a value that describes the degree of separation among classes in a finite mixture model. If the classes are well separated entropy will be small but as the degree of overlap among classes increases the value of entropy should increase. Entropy can be computed as

$$E = -\sum_{j=1}^{J}\sum_{i=1}^{n} t_{ij}\log\left(t_{ij}\right) \geq 0 \tag{15}$$

where

$$t_{ij} = \frac{\pi_j f^j\left(y_i|\hat{\Omega}_j\right)}{\sum_{j=1}^{J}\pi_j f^j\left(y_i|\hat{\Omega}_j\right)} \tag{16}$$

is the posterior probability of membership of individual i in class j given the data. This posterior probability is essentially the same as that utilized in the GBTM but provided in a notation consistent with the literature discussing entropy in the context of class enumeration.

Ramaswamy, DeSarbo, Reibstein, and Robinson (1993) proposed a scaled version of entropy (sE) to determine the number of classes in a mixture, Table 1. sE is bounded between 0 and 1 and should approach 1 as the degree of separation among classes increases. Their scaled version is required because as discussed by Celeux and Soromenho (1996), entropy cannot be directly used to determine the number of classes in a finite mixture model.

Computation of the CL was demonstrated by Biernacki and Govaert (1997):

$$CL = \sum_{i=1}^{n} \sum_{j=1}^{J} z_{ij} \log\left(\pi_j f^j \left[y_i | \hat{\Omega}_j\right]\right) \tag{17}$$

where z_{ij} takes on values either 1 or 0 depending on whether or not y_i is determined to come from class j. The CL was also discussed by Biernacki, Celeux, and Govaert (2000) in a form similar to that of Equation 14 but with t_{ij}, the fuzzy classification matrix, replacing z_{ij}, removing the need for "all or nothing" classification which can cause problems in estimation according to Celeux and Soromenho (1996). Another method of computing the CL was elaborated upon by Celeux and Soromenho (1996) and Biernacki and Govaert (1997). This method involves a link that can be made between the log likelihood (L) and the CL through the inclusion of E

$$L(j) = CL(j) + E(j). \tag{18}$$

The partitioning of the log likelihood into the CL and entropy was utilized by Celeux and Soromenho (1996) to produce their normalized entropy criterion (NEC), Table 1, where $\mathrm{L}(1)$ represents the log likelihood of the model with only 1 class. The NEC is intended to select the model that provides the best evidence of a clustering structure by evaluating the separation of classes relative to the change in log likelihood between models. A major drawback of this approach is that an NEC value cannot be computed for the $j = 1$ situation as NEC will become undefined. To overcome this drawback Celeux and Soromenho (1996) proposed a 2-step solution for normal mixtures. The first step requires fitting models in which ($2 \leq j \leq j_{\text{sup}}$), where j_{sup} is a reasonable upper bound for the number of classes to be tested. Once the best of the 2 to j_{sup} models is chosen, j^*, a model is estimated with j^* classes in which the class means and mixing proportions are constrained to be equal. Using the log likelihood and entropy values from this constrained model, $\tilde{L}(1)$ and $\tilde{E}(1)$ respectively, NEC is computed as

$$\frac{\tilde{E}(1)}{\tilde{L}(1) \; - L(1)}. \tag{19}$$

The model with j^* classes is chosen if NEC(j^*) ≤ NEC(1). Biernacki and Govaert (1997) had disappointing results with the rule for selecting between a $j = 1$ and a $j > 1$ model so Biernacki, Celeux, and Govaert (1999) proposed a different method. They suggest that when CL(j) > L(1) for j > 0 then 0 ≤ NEC ≤ 1 and if j^* as defined above does not produce NEC(j^*) ≤ 1 then a 1-class solution should be selected.

NEC, although considered a classification likelihood criterion, does not actually contain the classification likelihood. Biernacki and Govaert (1997) proposed a method of using the CL for selecting the optimal number of classes in a mixture. The method they proposed involved the link between L and CL. They demonstrated that $CL(j) = L(j) - EC(j)$ where

$$EC(j) = -\sum_{j=1}^{J}\sum_{i=1}^{n} z_{ij}\log\left(t_{ij}\right) \geq 0. \tag{20}$$

$E(j)$, the entropy of the fuzzy classification matrix, Equation 15, is the mean of the realization of the random variable $EC(j)$. By replacing $EC(j)$ with $E(j)$ and maximizing the CL, the CLC can be computed as displayed in Table 1. Biernacki and Govaert (1997) proposed using this approach to determine the number of classes in a finite mixture model. This approach is hindered by the fact that the parameter space for j classes Ω_j^* is a subset of Ω_{j+1}^* so CLC(j) ≤ CLC($j+1$) except when imposing a restriction that the mixing proportions are equal. When the mixing proportions are unequal, CLC was found to over extract the number of classes. Results from a simulation suggested that CLC performed well when classes were well separated and mixing proportions were equal (Biernacki & Govaert, 1997).

The classification likelihood was given in a penalized form by Biernacki et al. (1998) as a way to overcome the over extraction problem of CLC as mentioned above. They also wanted a solution to the issue that can occur when BIC is used to determine the optimal number of classes in a mixture. As pointed out by Biernacki et al. (2000), the BIC criterion will be biased towards over extraction when the correct model is not in the family of models considered. The integrated classification likelihood (ICL) that they developed in a form given by McLachlan and Peel (2000) is

$$ICL = -2\log\left(L\left[\hat{\Omega}_j\right]\right) + 2E(j) + 2n\sum_{j=1}^{J}\hat{\pi}_j \log\left(\hat{\pi}_j\right) + p_j \log(n) - 2J\left(n_1, \dots, n_j\right). \qquad (21)$$

In the equation, $J(n_1, \dots, n_j)$ is the log integrated likelihood of the prior Dirichlet distribution adopted by Biernacki et al. (1998) and is defined as

$$\sum_{j=1}^{J}\log\left(\Gamma\left[n_j + \alpha\right]\right) - \log(\Gamma[n + J\alpha]) - J\log(\Gamma[\alpha]) + \log(\Gamma[J\alpha]) \qquad (22)$$

where $\Gamma(\cdot)$ is the gamma function and α was set equal to .5. In this penalized form, the CLC is not supposed to over extract classes particularly in situations when classes are allowed to have unequal mixing proportions. An approximation of the ICL (ICL-BIC) was also given by Biernacki et al. (1998), Table 1. The approximation is BIC with an additional term added. The additional term is 2 times entropy. The ICL-BIC criterion, although an approximate of the ICL, was found to have similar results Biernacki et al. (1998). Only the performance of ICL-BIC will be examined because of its similarity in performance to ICL and relative ease of computation.

2.2.5 Alternative measures

Some alternative enumeration measures will be discussed but not examined in this study due to historically poor performance relative to other enumeration measures, computational difficulty, or computational time. The information complexity criterion (ICOMP) is another information criterion developed in the tradition of AIC but based on the covariance complexity index. The theory for ICOMP came from information criterion (e.g. AIC) and the final estimation criterion of Rissanen (1976). Bozdogan (1990) discussed ICOMP as being a composition in that it takes into account both the covariance matrix of the parameter estimates and residuals. ICOMP can be computed as

$$ICOMP = -2\log\left(L\left[\hat{\Omega}_j\right]\right) + \log\left(\frac{tr\left[I_j^{-1}\right]}{p_j}\right)p_j - \log\left|I_j^{-1}\right| \tag{23}$$

where $tr[\cdot]$ is the trace operator and I_j^{-1} is the inverse Fisher information matrix for model j. McLachlan and Peel (2000) point out that one of the concerns regarding ICOMP is that the form depends on the parameterization of the model. Celeux and Soromenho (1996) cautioned that it can be difficult to compute and demonstrated the form of ICOMP for a normal mixture with p dimensions. Celeux and Soromenho (1996) also investigated the performance of ICOMP and found that it overestimated the number of classes even more so than AIC but only when the classes had different variance matrices. Even so, ICOMP typically performed better than both AIC and BIC in all other conditions. However, in a simulation study performed by Biernacki and Govaert (1997), the performance of ICOMP was generally on par or worse than the performance of BIC across study conditions. The decision to exclude ICOMP from the investigation of enumeration measures in a GBTM stems from its computational difficulties, its limited use, and generally comparable results to BIC.

Bootstrapping has also been employed to facilitate significance tests for the optimal number of classes. The LRT does not necessarily follow a χ^2 distribution when testing for the optimal number of classes in a finite mixture model as previously discussed. The bootstrap procedure does not require a known null distribution of the LRT but is supposed to approximate it through a repeated sampling procedure. McLachlan (1987) discussed the procedure for generating the bootstrap samples for testing a 1-class versus a 2-class solution. The LRT is computed after fitting 1-class and 2-class models to the random sample from the model as specified by the null hypothesis, that is, the random sample is selected based on the maximum likelihood parameter estimates for the model specified in the null hypothesis. This process is repeated *B* times until the bootstrap sample of the LRT is complete. Comparison of the LRT from the original sample to the bootstrap sample of the LRT gives an approximation of the probability of obtaining the original sample LRT value under the null hypothesis. This bootstrapped likelihood ratio difference test will be denoted BLRT. Nylund et al. (2007) have demonstrated comparable to favorable enumeration accuracy of the BLRT compared to AIC, BIC, CAIC, ssBIC, and VLMR for a variety of finite mixture models, LCA, FMA, and GMM. However, the computational time of the BLRT is considerably larger than that for other enumeration measures and therefore will not be considered as part of the current investigation.

Another measure utilizing the bootstrap technique that can be used in determining the number of classes in a finite mixture model is the Efron (bootstrapped) information criterion (EIC) developed by Ishiguro, Sakamoto, and Kitagawa (1997)

$$EIC = -2\log\left(L\left[\hat{\Omega}_j\right]\right) + 2b\left(\hat{F}_n\right) \tag{24}$$

where $b\left(\hat{F}_n\right)$ is approximated by

$$\frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{n}\sum_{i=1}^{n}\log\left[f\{y_{ib}^{*}|\hat{\Omega}_b^{*}\}\right] - \frac{1}{n}\sum_{i=1}^{n}\log\left[f\{y_i|\hat{\Omega}_b^{*}\}\right]\right). \tag{25}$$

The *B* represents the number of bootstrap samples, y_{ib}^{*} is the *i*-th observation vector from the *b*-th bootstrap sample and y_i is the *i*-th observation vector from the original sample. The term $b\left(\hat{F}_n\right)$ is intended to represent the expected value of the bootstrap estimated bias, expectation taken over *Y*. The EIC is similar to AIC except the penalty term is now based on the bootstrap samples rather than dimension of the model, p_j. Ishiguro et al. (1997) suggest that their correction produces an unbiased estimator of -2 log likelihood assuming that the term $b(\hat{F}_n)$ is equal to the true bias when the log likelihood, based on the a maximum likelihood procedure, is used to estimate the expected log likelihood. EIC was excluded from the investigation of enumeration measures in a GBTM because McLachlan and Peel (2000) demonstrated that it did not perform as well as BIC and ICL-BIC in studies that simulated a variety of finite mixture models.

2.2.6 Performances assessment

The performances of the discussed enumeration measures have been briefly summarized in order to provide a general guideline of what to expect in the current Monte Carlo study. In order to provide a more thorough description of the performances of the various enumeration measures, the results of 5 chosen simulation studies will be expanded upon. The studies were chosen for a number of reasons. All of the studies investigated the performances of class enumeration measures in a type of finite mixture model using Monte Carlo methods. One study was chosen because it considered a plethora of enumeration measures. Two studies were chosen because they contained binary items, 1 in an LCA and 1 in both an LCA and FMA. The study that examined enumeration measures in an LCA and FMA also examined them in a GMM. The remaining 2 studies were chosen because they examined enumeration measures in the context of a GMM. The GBTM is a special case of GMM so results from the GMM studies should provide strong evidence for making hypotheses regarding the results of the current study. Evidence for this claim will be given in a comparison of the similarities and differences between GMM and the GBTM. This comparison is also provided to suggest that the investigation of the performances of the enumeration measures in the GBTM is necessary in spite of the current literature already available on the performances of enumeration measures in GMM.

Muthén and Shedden (1999) developed GMM in the structural equation modeling (SEM) framework as a method of studying change over time. Muthén (2001) later identified the GBTM as a special case of GMM which he called latent class growth mixture analysis (LCGA). Muthén (2001) mentioned the class mixing proportions, class trajectories, class assignments based on posterior probabilities, and trajectory variation within each class as key outputs of GMM. The first 3 outputs mentioned are very similar in that they were discussed as the 3 key outputs of the GBTM. The last output, trajectory variation within each class, is one of the differences between the 2 methods. In a GMM, the growth coefficients, $\beta^{j\prime}$s, are not fixed by class but are allowed to vary within each class (i.e., variability among individual trajectories within a class) and can be correlated. Furthermore, the covariance matrix of residuals across time can be different than a homogeneous diagonal matrix (e.g., autocorrelation, Huyuh-Feldt, unstructured). Nagin (2005) discussed some of the advantages and disadvantages of GMM relative to GBTM. The main advantage of GMM is that fewer classes are usually needed to capture variability and specify an adequate model by allowing departure of individual trajectories from their class's average trajectory. The disadvantages of GMM stem from the need for additional parameters to model the within class variability of the growth coefficients and the freely estimated residuals. Nagin (2005) suggests that the additional parameters in GMM make it very difficult to model data types other than normal such as count or binary data. The additional parameters can also blur the classes. Whereas in GBTM, where classes are interpreted as a collection of individuals following a similar trajectory, in GMM the interpretation of classes is not as clear. The within class variability of the growth coefficients can make it more difficult to distinguish between individuals' trajectories even if they are assigned to different classes.

Relative to GMM, GBTM's ability to easily analyze count and binary data and the comparatively larger number of classes commonly extracted are the most notable arguments for a standalone investigation of the performance of enumeration measures in GBTM. However, the performances of enumeration measures in a GMM provide a useful foundation for the study of enumeration measures in a GBTM because of the model similarities between the 2 techniques.

Nylund et al. (2007), Tofighi and Enders (2008) and Peugh and Fan (2012) all studied the performance of class enumeration measures in a GMM. The investigation of enumeration measures in an LCA, Yang (2006) and Nylund et al. (2007), and FMA, Nylund et al. (2007), provide insight into the performance of class enumeration measures with binary variables. Henson et al. (2007) investigated a large number of enumeration measures in the context of a mixture structural equation model, a type of LVMM. In addition, another very recent study was reviewed that examined the performances of 7 enumeration measures (AIC, BIC, ssBIC, VLMR, aVLMR, BLRT, sE) in a latent profile analysis (LPA) with a focus on the effect of the separation among the latent classes (Tein, Coxe, & Cham, 2013). The results of this study were similar to the results of the 5 studies reviewed to assess the historical performances of the enumeration measures investigated and therefore a detailed summary was not provided. However, the recommendations given by Tein et al. (2013) are included in the summary table (see Table 3) of the recommendations (recommended, not recommended, or uncertain) of enumeration measures from each article that was provided to accompany the discussion of their performances.

Henson et al. (2007) evaluated AIC, BIC, ssBIC, CAIC, VLMR, aVLMR, sE, CLC, ICL-BIC, and NEC in the context of a latent variable mixture model (LVMM). They also examined a

Table 3. Performances summary of 15 enumeration measures from 6 simulation studies

Measure	1 (LVMM)	2 (LCA)	3 (LCA)	3 (FMA)	3 (GMM)	4 (GMM)	5 (GMM)	6 (LPA)
AIC	×	×	×	×	×	×	•	×
CAIC	×	×	•	✓	✓	×	×	-
ssCAIC	-	•	-	-	-	•	×	-
BIC	×	×	✓	✓	✓	•	×	✓
B_{10}	-	-	-	-	-	-	-	-
ssBIC	✓	✓	✓	•	•	✓	•	✓
HQ	-	•	-	-	-	-	•	-
ssHQ	-	-	-	-	-	-	•	-
VLMR	•	-	•	•	•	✓	×	✓
aVLMR	•	-	-	-	-	✓	×	✓
sE	×	-	-	-	-	-	•	×
NEC	×	-	-	-	-	-	•	-
CLC	✓	-	-	-	-	-	×	-
ICL-BIC	✓	-	-	-	-	-	×	-
ssICL-BIC	-	-	-	-	-	-	•	-

Note. Statistical model utilized in each article is in parentheses. 1 = Henson, Reise, and Kim (2007); 2 = Yang (2006); 3 = Nylund, Asparouhov, and Muthén (2007); 4 = Tofighi and Enders (2008); 5 = Peugh and Fan (2012); 6 = Tein, Coxe, and Chan (2013); LVMM = latent variable mixture model; LCA = latent class analysis; FMA = factor mixture model; GMM = growth mixture model; ✓ = recommended; • = uncertain; × = not recommended; - = not investigated.

multivariate skew test (MST) and multivariate kurtosis test (MKT) proposed by Muthén (2003) but not discussed here. The test performed poorly in the context of the study and others (see, for example, Tofighi & Enders, 2008). A comparison of the information criteria (AIC, BIC, ssBIC, and CAIC) found that ssBIC recovered the true 2-class solution the most often, AIC performed fairly well but did tend to over extract the number of classes as compared to the other 3, and BIC and CAIC performed very similarly with a strong tendency for under rather than over extraction. VLMR and aVLMR performed similarly well relative to AIC, BIC, and CAIC but did not perform as well as ssBIC. VLMR and aVLMR tended to under extract the number of classes as opposed to over extract but did over extract more so than AIC, BIC, and CAIC.

The entropy statistics and classification indices had a wide range of performance. The sE statistic and the NEC statistic with the rule developed by Biernacki et al. (1999) did not perform as well as the majority of the other enumeration measures examined. They both tended to over extract rather than under extract classes. The rate of correct model recovery by the NEC statistic with the rule developed by Celeux and Soromenho (1996) was slightly less than that of BIC and CAIC. The 2 enumeration measures that performed the best were the CLC and the ICL-BIC with the ICL-BIC holding a slight edge. Furthermore, the CLC and ICL-BIC were the least affected of all the enumeration measures by 2 of the factors manipulated in the study design, sample size and mixing proportions. However, it should be noted that the performances of the enumeration measures were evaluated based on their ability to select a true 2-class solution from competing 1- and 3-class solutions. Entropy cannot be computed for a 1-class solution so sE, CLC, and ICL-BIC could only select between a 2-class and 3-class solution. This situation made it impossible for sE, CLC, and ICL-BIC to make under extraction errors and may have contributed significantly to the superior performance of CLC and ICL-BIC.

Yang (2006) examined the performance of AIC, BIC, CAIC, HQ, ssBIC, and sample size adjusted CAIC (ssCAIC). In addition, the performance of the Draper BIC (DBIC; Draper, 1995) and another variant of AIC proposed by Hurvich and Tsai (1989) (HTAIC) were examined. These latter 2 measures are less commonly encountered in the enumeration literature and they have not been found to outperform all of the other measures being examined and therefore they will not be mentioned further. The ssCAIC was computed by making the same sample size adjustment to CAIC that Sclove (1987) made to BIC such that the n in Table 1 is replaced with $(n+2)/24$, ssCAIC in Table 1. The ssBIC was found to be a top performer across all levels of the 2 conditions examined, sample size and number of classes. Furthermore, it was consistent and performed very well as the number of classes increased. CAIC performed well at larger sample sizes and for a smaller number of latent classes. ssCAIC performed similarly to ssBIC when there were 4 latent classes but under extracted classes when the sample size was 300 or less and there were 5 or 6 latent classes. The performance of HQ was hindered when sample size was small (N = 100) or when the number of classes was 5 or 6 relative to sample sizes of 300 or less. Otherwise, HQ performed very well. The performance of BIC was poor until a sample size of 300 and was greatly affected by the number of latent classes. Not until a sample size of 1000 was BIC's performance on par with the performance of ssBIC and HQ when there were 6 latent classes. AIC had average performance (correct classification rate around 60% to 70%) across all conditions except for when the sample size was 100 and the number of classes was 6. In that condition, performance was poor. Extraction errors varied by enumeration measures and to some extent conditions found in the data. AIC tended to over extract except when the sample size was at the smallest level. BIC, CAIC, ssCAIC, and HQ all had a tendency to under extract rather than over extract classes. ssBIC tended to over extracted rather than under extract classes at the 3 smallest sample sizes. ssBIC only tended to under extract rather than over extract classes when the sample size was N = 200 or N = 300 and there were 6 latent classes.

Nylund et al. (2007) examined the performances of BLRT, VLMR, AIC, CAIC, BIC, and ssBIC. The exclusion of BLRT was already discussed and its performance will not be included in this discussion. The discussion will primarily focus on the 2 best performing enumeration measures in all conditions for both the LCA and FMA, BIC and ssBIC. The LCA was a 4-class solution and the FMA was a 2-class solution. Sample sizes of 200, 500, and 1000 were simulated for both the LCA and FMA. Two of the LCA models were based on an 8-item simple structure where the class sizes were either equal or unequal (5%, 10%, 15%, 70%). The remaining 2 LCA models were based on a 15-item simple structure with equal class proportions or a 10-item complex structure with unequal class proportions (5%, 10%, 15%, 70%). The probability of endorsing specific items was high for multiple classes in the complex structure. For the simple structure, only 1 class had high probabilities for endorsing specific items. The performance of AIC was poor to average at best in both the LCA and FMA. CAIC performed well in the FMA and the 8-item and 15-item simple structures with equal class proportions. When the complex structure was used, CAIC only performed well at the largest sample size. When the class proportions were unequal, CAIC performed poorly at the smaller 2 sample sizes and increased to about average performance at the largest sample size. The rate of correct classification for BIC and ssBIC was within 6% for 7 of the 15 conditions across the 2 models. BIC had better performance than ssBIC when the sample size was 200 for the 8-item simple structure equal class proportions condition and also for the 15-item condition. BIC also outperformed ssBIC for the FMA. ssBIC performed better than BIC at all sample sizes for the 8-item simple structure unequal class proportions condition and at the smaller 2 sample sizes for the 10-item complex structure condition. AIC and VLMR tended to over extract rather than under extract classes. However, VLMR did tend to under extract when class proportions were unequal. CAIC and BIC tended to under extract rather than over extract classes when errors were made. ssBIC tended to over extract classes except for when the class proportions were unequal.

Nylund et al. (2007) also examined the performance of BLRT, VLMR, AIC, CAIC, BIC, and ssBIC in a 2-class GMM with unequal (75%, 25%) mixing proportions. There was quadratic growth in one class and linear growth in the other. The same sample sizes were simulated for the GMM as the LCA and FMA. CAIC had the best performance across all 3 sample sizes. BIC and CAIC selected the correct number of classes 100% of the time at the 2 largest sample sizes. BIC performed fairly well at the smallest sample size but not as well as CAIC. ssBIC was the criterion most affected by sample size. ssBIC ranged from selecting the correct number of classes 66% of the time at the smallest sample size to 100% of the time at the largest sample size. VLMR performed better than and the same as ssBIC at N=200 and N=500, respectively. The performance of VLMR was inconsistent in that the rate of selecting the correct number of classes decreased between sample sizes of N = 500 and N = 1000. The type of extraction error made by each enumeration measure was fairly consistent between the LCA and FMA and GMM models. Of note was the type of error made by VLMR. At the smallest sample size VLMR tended to under extract but as the sample size increased VLMR tended to over extract classes.

Tofighi and Enders (2008) examined BIC, ssBIC, AIC, CAIC, ssCAIC, VLMR, MST, and MKT in a simulated 3-class GMM with quadratic growth in 2 classes and linear growth in the third. They varied the number of repeated measures (4, 7), sample size (400, 700, 1000, 2000), separation of classes (average posterior probability of .8 or .9), mixing proportions (20%, 33%, 47%; 7%, 36%, 57%) and within-class distribution shape (normal, non-normal). The number of repeated measures was found to have little effect on the different measures abilities to select the correct number of classes. Non-normal within-class distributions were found to negatively impact the performances of all enumeration measures compared to when the within-class distributions were normal. As was already discussed, MST and MKT performed very poorly across all conditions. Both BIC and CAIC were very affected by changes in sample size, class separation, and mixing proportions. Not until a sample size of N = 2000 did they perform as well as the best performing enumeration measure, ssBIC. ssCAIC and VLMR were also affected by changes in sample size, class separation, and mixing proportions. The rate at which they selected the correct number of classes was on par with ssBIC at best. Similar to the findings of Nylund et al. (2007), the performance of VLMR was inconsistent in that the rate of selecting the correct number of classes decreased between sample sizes of N=1000 and N=2000. The performance of AIC was relatively unaffected by changes in sample size, class separation, and mixing proportions. The rate at which AIC selected the correct number of classes never exceeded that of ssBIC. Tofighi and Enders (2008) suggested that errors in extraction tended to be under extraction errors with all the enumeration measures except for VLMR. When the sample size reached N=1000, VLMR tended to over extract the number of classes.

Another very recent simulation study that evaluated different enumeration measures in a GMM was performed by Peugh and Fan (2012). They examined a large set of enumeration measures that were also examined in the current study: AIC, CAIC, ssCAIC, BIC, ssBIC, HQ, sE, CLC, NEC, ICL-BIC, VLMR, and aVLMR. Peugh and Fan (2012) also made the same sample size adjustment of Sclove (1987) to HQ and ICL-BIC such that the n in Table 1 is replaced with $(n+2)/24$. These measures will be denoted ssHQ and ssICL-BIC. In their study, they varied sample size (300, 600, 900, 1200, 1500, 3000), magnitude of separation distance among classes measured statistically in multivariate Mahalanobis distance units (.5, .8, 1.2, 2), mixing proportions (33%, 33%, 33%; 50%, 33%, 17%), and different trajectory shapes (different intercepts and same slopes, same intercepts and different slopes, different intercepts and different slopes). A general account of the performance of enumeration measures will be summarized across these conditions.

The type I error rate was considered for each statistic by using a 1-class solution. Peugh and Fan (2012) found that CAIC, BIC, CLC, ICL-BIC, and ssICL-BIC all selected 1-class solution 100% of the time regardless of sample size.

The type II error rate for each measure was considered by using a 3-class solution. Peugh and Fan (2012) reported that they did not experience much change in the performance of the measures for varying degrees of mixing proportion except in one of the trajectory shape conditions and that the performance of all the measures generally decreased as the distance among the classes decreased. When the trajectory shapes were defined as different intercepts and same slopes or same intercepts and different slopes there was poor class recovery; none of the measures selected the correct number of classes even 50% of the time. sE, AIC, and NEC had the best performance in that order except for the same intercepts different slopes condition where at the largest sample size NEC had better performance than AIC.

There was a marked improvement in the majority of the measures for the different intercepts different slopes trajectory shape condition. At the smallest sample size, HQ, sE AIC, ssICL-BIC, and CAIC had the best performance in that order but still only selected the correct number of classes less than 30% of the time. At the largest sample size and with equal class proportions, ssHQ, HQ, ssICL-BIC, and ssBIC had the best performances, all with a recovery rate of 88% or above. Interestingly, the sE statistic, which did so well in other conditions, actually had a decrease in performance when the sample size increased for the different intercepts different slopes condition. At the largest sample size with unequal class proportions, ssHQ, AIC, and HQ had the best recovery rates that ranged from 55% to 62%. Only NEC, VLMR, and aVLMR consistently made over extraction rather than under extraction errors regardless of conditions. AIC tended to under extracted classes rather than over extract classes in all but the largest sample size different intercept different slope conditions. sE tended to over extract rather than under extract in all but the largest sample size, different intercepts different slopes, and unequal class proportions conditions. All of the remaining enumeration measures consistently made under extraction rather than over extraction errors across the various conditions.

Commonalities and differences among the results of the 5 studies were used to make some predictions about the performance of the enumeration measures in the context of GBTM. The information criteria will be considered first. It is hypothesized that the performance of AIC will be average, that it will tend to over extract classes, and that its performance will remain relatively stable across simulated conditions. CAIC and BIC will perform similarly. They will perform well at sample sizes of about 500 or greater and tend to under extract classes at smaller sample sizes. Only Nylund et al. (2007) found BIC to be the best performing of the information criteria overall and to be better performing that ssBIC in a GMM. Nylund et al. (2007) specified a true 2-class solution for their GMM whereas Peugh and Fan (2012) and Tofighi and Enders (2008) both specified a true 3-class solution. It is possible that BIC's tendency to under extract classes is exacerbated when there is a larger true class solution. The performance of B_{10} was not investigated in any of the reviewed studies but it is expected to perform similarly to BIC with the possibility of under extracting classes at a higher rate. The reviewed studies selected the *j* class solution that had the smallest BIC regardless of the difference between the BIC values for the *j* class solution and the *j*-1 and *j*+1 class solutions. B_{10} selects a *j*+1 class solution over a *j* class solution based on the interpretive evidence in Table 2. A larger difference in the BIC values between 2 models will be needed for B_{10} to select the *j*+1 class solution relative to the minimum BIC method of selecting a model. The performance of HQ was only investigated in 2 of the studies but there was some consistency in results. Both Peugh and Fan (2012) and Yang (2006) found HQ to generally outperform AIC, CAIC, and BIC with a slight tendency for under extracting classes. ssBIC was found to outperform BIC and ssCAIC was found to outperform CAIC in the majority of conditions across the various studies that investigated them both. They were not as affected as their non-sample size adjusted counterparts by conditional variations such as changes in sample size or mixing proportions. In the 3 studies that examined both ssBIC and ssCAIC, ssBIC had a better performance overall and was not as affected when mixing proportions became increasingly unequal (Tofighi & Enders, 2008) or when the number of classes increased (Yang, 2004). Tofighi and Enders (2008) found that ssBIC outperformed ssCAIC at smaller sample sizes but Yang (2006) found that ssCAIC held a slight edge at sample sizes of 300 or less when a true 4-class solution was used. The performance of ssHQ was only examined in Peugh and Fan (2012) but it was found to outperform both ssBIC and ssCAIC across the conditions investigated. Based on the collective results, it is hypothesized that ssBIC and ssHQ will be the overall best performing of the information criteria.

The performances of the VLMR and aVLMR tests were found to be almost identical in 3 of the reviewed studies in which they were both investigated. The comparable results enable the following predictions of the performance of the VLMR test to pertain to both likelihood ratio test derivatives. AIC typically outperformed the VLMR test at sample sizes around 500 or less but the VLMR test outperformed AIC at larger sample sizes. Nylund et al. (2007) found that CAIC, BIC, and ssBIC all typically outperformed the VLMR test but Henson et al. (2007), Peugh and Fan (2012), and Tofighi and Enders (2008) all found that the VLMR test outperformed both CAIC and BIC but not ssBIC. The reviewed results suggest that the VLMR test will have an acceptable rate of correct classification and perform similarly to BIC and CAIC but it will not perform as well as ssBIC in the context of the current study. It is also hypothesized that the VLMR test may tend to over extract as opposed to under extract the correct number of classes at sample sizes around 1000 or larger.

The entropy statistics and classification indices to be examined in the current study were only examined in 2 of the reviewed studies. Henson et al. (2007) found that CLC and ICL-BIC recovered the correct number of classes at levels at or above the other enumeration measures examined with ICL-BIC holding a slight edge. As mentioned, Peugh and Fan (2012) found that ICL-BIC correctly identified a true 1-class solution 100% of the time but for some conditions it also selected a 2-class solution 100% of the time when a correct identification should have enumerated 3 classes. The only condition they reported in which ICL-BIC performed slightly well was when the trajectory shapes were specified as different intercepts and slopes and the sample size was at its largest. Based on the aggregate of the results reported in Peugh and Fan (2012), it would be tempting to suggest sE is the best performing entropy statistic or classification index and that NEC outperformed all of the information criteria except ssHQ at smaller sample sizes. However, the performance of sE was at its lowest in the same conditions that ICL-BIC began to have an increase in performance and both ssBIC and ssHQ outperformed NEC in this same condition. Furthermore, Henson et al. (2007) found the performance of sE to be uninteresting in that it was not either very good or very bad. They also found that all of the information criteria they investigated performed better than NEC. As possible evidence against the performance of sE and NEC in Peugh and Fan (2012), consider the performance of AIC. In the majority of conditions reported, Peugh and Fan (2012) found that AIC tended to under extract the correct number of classes rather than over extract the correct number of classes. This result is very inconsistent with the literature which suggested that AIC is known to over identify the number of classes very consistently. The one set of conditions that AIC over extracted more than under extracted the correct number of classes is the exact set that the performance of ICL-BIC improved and sE performed poorly. It is possible that this set of conditions is the only set where the discrepancy among classes was large enough to realize a meaningful and consistent performance of the enumeration measures. In this case, the results of Peugh and Fan (2012) would suggest that HQ, ssHQ, ssICL-BIC, and ssBIC had the best performance which is a bit more consistent with the other results discussed. Peugh and Fan (2012) never found CLC or ICL-BIC to be the best performing enumeration measures as did Henson et al. (2007). However, the performances of CLC and ICL-BIC may have been exaggerated in Henson et al. (2007). As was discussed above, CLC and ICL-BIC could not make under extraction errors. Results of 3 simulation studies in McLachlan and Peel (2000) corroborate the results for ICL-BIC in Henson et al. (2007). McLachlan and Peel (2000) found ICL-BIC to be the best performing of a small set of enumeration measures (AIC, EIC, BIC, CLC, ICL-BIC, ICL). It is somewhat difficult to make an accurate prediction for the best performing of the entropy statistics and classification indices based on 2 studies with somewhat inconsistent results. However, with the added evidence from the results of the McLachlan and Peel (2000) simulation studies, it is assumed that ICL-BIC and possibly CLC and ssICL-BIC will be the best performing of the entropy statistics and classification indices. It is also assumed that they will outperform the non-sample size adjusted information criteria and perform similarly to ssBIC and ssHQ.

As demonstrated, the choice of which enumeration measure or measures to rely on is not an easy one and could possibly be model and condition specific. The 5 reviewed studies were selected because they either examined enumeration measures in GMM, a model similar to the GBTM, had binary logic and normally distributed variables, or examined a large set of the more commonly encountered measures. Although the results from the 5 reviewed studies helped to formulate expectations for the performances of the enumeration measures in the context of the GBTM; it remains to be seen which will actually prove to be the most efficient.

3 METHODS

3.1 INTRODUCTION

A Monte Carlo study was conducted to investigate the effect of sample size, number of repeated measures, class mixing proportions, percent missing, and separation among classes on the enumeration measures' abilities to detect a true[3] number of classes in a censored normal GBTM and binary logit GBTM. A 4 (sample size) by 2 (number of repeated measures) by 3 (class mixing proportions) by 3 (percent missing) by 2 (class separation) (=144 cell) factorial design was performed with 500 replications and 23 enumeration measures computed per cell. Data were generated using SAS Proc IML and then analyzed using Proc TRAJ. Sample size had four levels, 200, 400, 800, and 2000. The number of repeated measures had 2 levels, 5 and 9. Class mixing proportions had 3 levels, equal (25%, 25%, 25%, 25%), moderately unequal (40%, 25%, 20%, 15%), or extremely unequal (60%, 25%, 10%, 5%). Percent missing had 3 levels, 0%, 10%, and 20%. Class separation had 2 levels which varied depending on the model distribution, high separation ($\sigma = 1$) and low separation ($\sigma = 2$) for the binary logit distribution and high separation ($\sigma = 2$) and low separation ($\sigma = 3$) for the censored normal distribution.

[3] The phrase optimal class solution was used for an optimal number of classes in a GBTM. A Monte Carlo study allows for the "true" number of classes to be known so the optimal solution will be the true solution in this context.

3.2 DESIGN

Data were generated for 4 classes measured at either 5 or 9 time points for the censored normal and binary logit models. A 4-class solution was utilized for a couple of reasons. Tofighi and Enders (2008) suggested that the number of classes typically extracted was between 2 and 4 in GMM literature. Nylund et al. (2007), Tofighi and Enders (2008), and Peugh and Fan (2012) specified 2-, 3-, and 3-class solutions, respectively, for examining enumeration measures in the context of GMM. Tofighi and Enders (2008) also suggested that the extraction of more than 4 classes was rare in GMM except when constraining the model to be a GBTM in which case the extraction of more than 4 classes is not uncommon. The reviewed GMM studies all had less than 4 classes and because it is more common to extract a larger number of classes when using GBTM, a 4-class solution was selected. In order to ensure a 4-class solution was acceptable given the numbers of classes historically extracted in the literature, a random sample of 60 studies from the search of articles citing Nagin's 1999 article (Web of Science, Social Science Citation Index, retrieved 2/26/2013) was selected and reviewed. Of the random sample of 60 studies, only 43 of them actually performed a GBTM (Blokland, Nagin, & Nieuwbeerta, 2005; Brook, Zhang, & Brook, 2011; Buckingham-Howes, Oberlander, Hurley, Fitzmaurice, & Black, 2011; Chassin, Pitts, & Prost, 2002; Chen et al., 2012; I.-J. Chung & J. Chun, 2010; Cohen, Piquero, & Jennings, 2010; Côté, Tremblay, Nagin, Zoccolillo, & Vitaro, 2002; de la Sablonniere, Taylor, Perozzo, & Sadykova, 2009; Dolgin et al., 2007; Duchesne, Ratelle, Larose, & Guay, 2007; Dush, Taylor, & Kroeger, 2008; Edwards, Homish, Eiden, Grohman, & Leonard, 2009; Fisher et al., 2010; Gaudreau, Amiot, & Vallerand, 2009; Gildengers et al., 2005; Guadamuz et al., 2012; Guo et al., 2002; Haukka et al., 2011; Higgins, Jennings, Marcum, Ricketts, & Mahoney, 2011; Hynes & Clarkberg, 2005; Kokko, Pulkkinen, Mesiainen, & Lyyra, 2008; Kozyrskyj, Kendall, Jacoby, Sly, & Zubrick, 2010; Lavender et al., 2011; Le Brocque et al., 2010; Letcher, Smart, Sanson, & Toumbourou, 2009; Louvet, Gaudreau, Menaut, Genty, & Deneuve, 2007; Louvet, Gaudreau, Menaut, Genty, & Deneuve, 2009; M'Bailara, Cosnefroy, Vieta, Scott, & Henry, 2013; Murray et al., 2010; Nieuwbeerta, Nagin, & Blokland, 2009; Paciello, Fida, Tramontano, Lupinetti, & Caprara, 2008; Quinsey, Jones, Book, & Barr, 2006; Rodriguez & Tamis-LeMonda, 2011; Rosendahl, Galanti, & Gilljam, 2008; Spano, Rivera, & Bolland, 2010; Sterling, Hendrikz, & Kenardy, 2010; Tan, Dierker, Rose, Li, & Tobacco Etiology Research Network, 2011; van Lier, Vitaro, Barker, Koot, & Tremblay, 2009; Weisburd, Morris, & Groff, 2009; Wiesner, Capaldi, & Kim, 2007; Windle & Wiesner, 2004; Yeh et al., 2011). There were 69 GBTMs run in the 43 reviewed GBTM studies and of those 69, approximately 26% had a 4-class solution. Although it was most common to extract 3 classes, the extraction of 4 classes occurred often enough to provide support for the choice of a 4-class solution.

In this study, the binary logit models for the 4 classes were

$$y_{it}^{(1*)} = -2.14 + .388X_{it} - .091X_{it}^2 + \varepsilon_{it}, \tag{26}$$

$$y_{it}^{(2*)} = 2.912 - .279X_{it} + .02X_{it}^2 + \varepsilon_{it}, \tag{27}$$

$$y_{it}^{(3*)} = -.337 + 1.191X_{it} - .156X_{it}^2 + \varepsilon_{it}, \text{ and} \tag{28}$$

$$y_{it}^{(4*)} = .337 - 1.191X_{it} + .156X_{it}^2 + \varepsilon_{it}, \tag{29}$$

where $y_{it}^{(j*)}$ is the logit for person *i* in group *j* measured at time *t* and X_{it} represents a measure of time such as age for individual *i* at time *t*. The trajectories were based on a combination of the trajectories found in Haukka et al. (2011) and Yeh et al., (2011) and are displayed in Figure 2. Class 1 was intended to represent a group of individuals with a fairly constant low probability of experiencing the outcome of interest, α_{it}^1 = .1 to .15. Class 2 was intended to represent a group

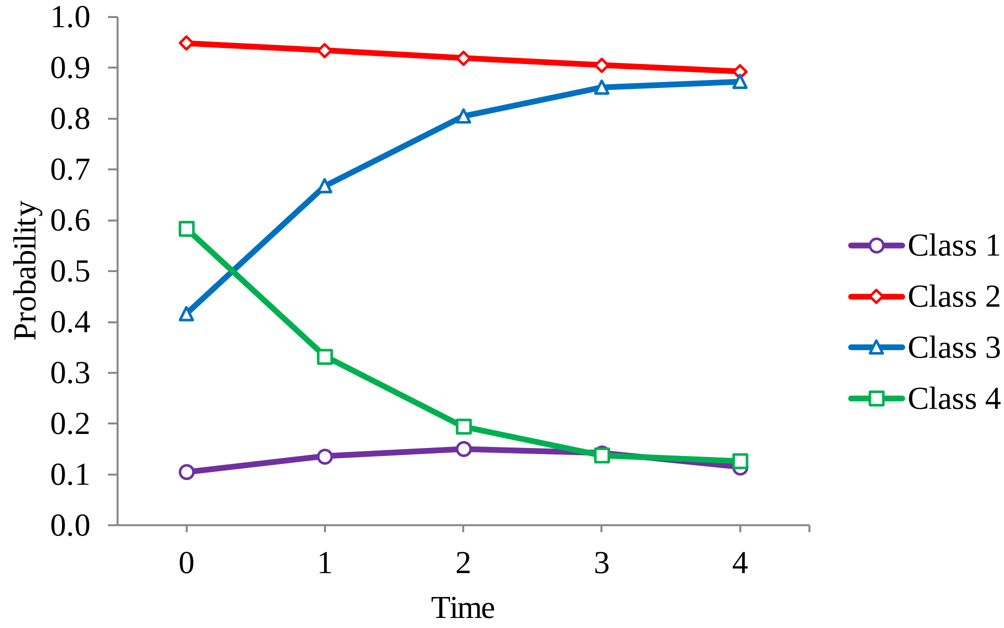

Figure 2. Growth trajectories for the true four class binary logit model

of individuals with a fairly constant high probability of experiencing the outcome of interest, α_{it}^{2} = .89 to .95. Class 3 was intended to represent an at risk group that starts out with a lower probability that changes to a high probability of experiencing the outcome of interest over the measured time period, α_{it}^{3} = .42 to .87. Class 4 was intended to represent a moderate declining group that starts out with a higher probability that changes to a low probability of experiencing the outcome of interest, α_{it}^{4} = .58 to .13. The 4-class solution of a high class, a low class and 2 intersecting intermediate classes was selected because it appears to be somewhat common. The trajectories in Haukka et al. (2011) and Yeh et al., (2011) were similar to trajectories in 3 of the 18 reviewed studies that had a 4-class solution (Kozyrskyj et al., 2010; Lavender et al., 2011; M’Bailara et al., 2013) and one study that had a 5-class solution (Windle & Wiesner, 2004). The class trajectories in these studies can all be characterized as having 2 classes intersecting between a high and a low class but with differences in trajectory shape. The class trajectories in Windle and Wiesner (2004) can be characterized the same way except there were 2 low classes.

The trajectories for the censored normal model are displayed in Figure 3. These trajectories were computed by adding a constant value of 20 to the intercept in Equations 26-29. The scale maximum was set at 30 so that $p(y_{it} \geq 30) < .01$ for Class 2, $t = 2, \ldots, 5$ or 9. The interpretation of the classes for the censored normal model is similar to the interpretation of the classes for the binary logit model but with scores replacing probability. Class 1 was intended to represent a group of individuals with a fairly constant low score, $\bar{y}_t^1 = 17.86$ to 18.27. Class 2 was intended to represent a group of individuals with a fairly constant high score, $\bar{y}_t^2 = 22.12$ to 22.91. Class 3 was intended to represent an at risk group that starts out with a lower score that changes to a higher score over the measured time period, $\bar{y}_t^3 = 19.66$ to 21.93. Class 4 was intended to represent a moderate declining group that starts out with a higher score that changes

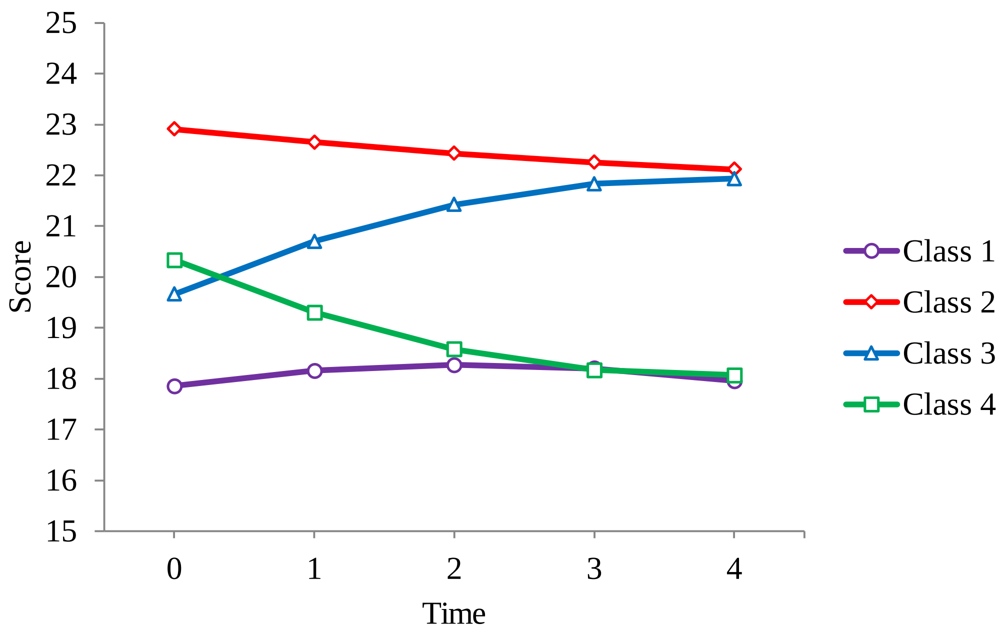

Figure 3. Growth trajectories for the true four class censored normal model

to a lower score, $\bar{y}_t^4 = 20.34$ to 18.07. BMI is an example of an individual characteristic that might result in similar trajectories over a particular time frame given a different score scale.

3.2.1 Factor: Class separation

The error variance was manipulated in order to control the separation among classes. The error term was specified as $N(0,1)$ or $N(0,2)$ in the binary logit model or $N(0,2)$ or $N(0,3)$ in the censored normal model. The error variance was chosen so that the classes' average posterior probabilities[4] in the binary logit model varied. In the high separation condition ($\sigma = 1$), the highest average posterior probability for a particular class was .94 with a median of .95 and a semi-interquartile range (SIQR) of .03. Also in the high separation condition ($\sigma = 1$), the lowest average posterior probability for a particular class was .76 with a median of .76 and a SIQR of .06. In the low separation condition ($\sigma = 2$), the highest average posterior probability for a particular class was .83 with a median of .84 and a SIQR of .055. Also in the low separation condition ($\sigma = 2$), the lowest average posterior probability for a particular class was .60 with a median of .60 and a SIQR of .05. The McFadden and Cox and Snell R^2 were computed to provide an estimate of the amount of variability explained by the linear and quadratic trend in the true 4-class solution[5]: McFadden $R^2 = .118$ and Cox and Snell $R^2 = .430$ when $\sigma = 1$; McFadden $R^2 = .011$ and Cox and Snell $R^2 = .068$ when $\sigma = 2$. These values suggest that approximately (high separation - $\sigma = 1$, 12% to 43%; low separation - $\sigma = 2$, 1% to 7%) of the variability in the outcome is being explained by time. The classes' average posterior

[4] Computation averaged over sample size, number of repeated measures, class mixing proportions, and percent missing and was based on the 22675 replications where the 4-class solution converged.
[5] Computation of R^2 based on 50 replications of the $N = 2000$, equal class proportions, 0% missing, and 5 repeated measures condition.

probabilities[6] in the censored normal model varied. In the high separation condition ($\sigma = 1$), the highest average posterior probability for a particular class was .86 with a median of .86 and a semi-interquartile range (SIQR) of .04. Also in the high separation condition ($\sigma = 1$), the lowest average posterior probability for a particular class was .67 with a median of .67 and a SIQR of .05. In the low separation condition ($\sigma = 2$), the highest average posterior probability for a particular class was .80 with a median of .79 and a SIQR of .055. Also in the low separation condition ($\sigma = 2$), the lowest average posterior probability for a particular class was .60 with a median of .60 and a SIQR of .05. McFadden $R^2 = .007$ and Cox and Snell $R^2 = .157$ when $\sigma = 2$ and McFadden $R^2 = .002$ and Cox and Snell $R^2 = .039$ when $\sigma = 3$[7]. These values suggest that approximately (high separation - $\sigma = 2$, 1% to 16%; low separation - $\sigma = 3$, .2% to 4%) of the variability in the outcome is being explained by time. Tofighi and Enders (2008) used a similar method of varying the error variance to alter the separation among trajectories while maintaining an approximate average posterior probability for each class. Nagin (2005) suggested that an average posterior probability of around .7 or above for all classes is indicative of good model fit. Based on reviewed literature, practitioners of the GBTM technique seem to adhere to this suggestion. Therefore, it appeared acceptable to select error variances that maintained reasonable average posterior probabilities for all classes. Allowing the posterior probabilities of some classes to go below .7 enabled the performances of the enumeration measures to be assessed when classes were not so clearly distinguishable.

[6] Computation averaged over sample size, number of repeated measures, class mixing proportions, and percent missing and was based on the 35468 replication where the 4-class solution converged.
[7] Computation of R^2 based on 50 replications of the N = 2000, equal class proportions, 0% missing, and 5 repeated measures condition.

3.2.2 Factor: Sample size

There were 4 sample sizes, 200, 400, 800, and 2000. The sample sizes are similar to those used in Nylund et al. (1997), Tofighi and Enders (2008), and Peugh and Fan (2012). The sample sizes are also reflective of sample sizes commonly found in the GBTM literature. The sample sizes were considered to be small, small to medium, medium, and large sample sizes.

3.2.3 Factor: Number of repeated measures

The number of repeated measures was considered because of the potential impact it could have on the classes' posterior probabilities. As mentioned, Tofighi and Enders (2008) examined the effect of having different numbers of repeated measures and concluded that it had a minor impact on the class enumeration measures they examined. However, they did not examine the performances of entropy statistics or classification indices which take into account the separation of classes by incorporating posterior probabilities. The performances of entropy statistics and classification indices were examined in this study so it was desirable to determine if the number of repeated measures had an impact.

In this study, the number of repeated measures was varied as either 5 or 9 while maintaining a constant time period. For the 5 observation condition, the values of X_{it} in Equations 26-29 were specified as 0, 1, 2, 3, and 4. For the 9 observation condition, the values of X_{it} in Equations 26-29 were specified as 0, .5, 1, 1.5, 2, 2.5, 3, 3.5, and 4. In a GBTM it is assumed that everyone is measured at the same time relative to the value of X_{it} (i.e., fixed factor). For example, if X were intended to represent age, each individual could be measured at 11, 12, 13, 14, and 15 years old but they may have been measured on different occasions or dates. This is the reason that the same values of X_{it} can be specified for all individuals. The numbers of observations were specified as either 5 or 9 to reflect those found in the GBTM literature. Over 75% of the reviewed GBTM studies had a number of repeated measures between 3 and 9.

3.2.4 Factor: Class mixing proportions

The class mixing proportions condition specified the number of subjects out of the total sample size that were in each class. As illustrated in the review of studies assessing the performances of enumeration measures in a GMM, different class mixing proportions where shown to have an effect. The class proportions were set as either equal (25%, 25%, 25%, 25%), moderately unequal (40%, 25%, 20%, 15%), or extremely unequal (60%, 25%, 10%, 5%). The classes' sample sizes for different levels of sample size and class mixing proportions are displayed in Table 4. In the reviewed GBTM studies, the largest class proportion was 91% and the smallest was 2%. The largest class proportion in the majority of the studies was between 40% and 70%

Table 4. Classes' sample sizes by overall sample size and class mixing proportions

Sample Size	Equal (25%, 25%, 25%, 25%)	Moderately Unequal (40%, 25%, 20%, 15%)	Extremely Unequal (60%, 25%, 10%, 5%)
200	50, 50, 50, 50	80, 50, 40, 30	120, 50, 20, 10
400	100, 100, 100, 100	160, 100, 80, 60	240, 100, 40, 20
800	200, 200, 200, 200	320, 200, 160, 120	480, 200, 80, 40
2000	500, 500, 500, 500	800, 500, 400, 300	1200, 500, 200, 100

Note: Sample sizes are for Class 1, Class 2, Class 3, and Class 4, respectively.

and the smallest was typically between 5% and 15%. The selected class proportions conditions provide an accurate representation of the range of class proportions typically found in the GBTM literature while providing an ideal equal proportion condition in which each class has enough individuals to be accurately estimated.

3.2.5 Factor: Percent missing

Missed observations are an inevitable part of longitudinal data collection and analysis (Feldman & Rabe-Hesketh, 2012). Longer study durations give participants more time to drop out or become unreachable and thus un-measureable for follow up observations. Missing data can reduce the amount of information available from which a model is estimated. Furthermore, mishandling of missing data can bias results (Feldman & Rabe-Hesketh, 2012). There has been a historical emphasis on finding optimal methods for handling such data (Duncan & Duncan, 1994; Enders & Bandalos, 2001; Ferrer, Hamagami, & McArdel, 2004; Little & Rubin, 1989; Newman, 2003; Raykov, 2005; Shin, Davison, & Long, 2009). In the trajectory analysis context, missing data are handled by estimating the model using all available information. For example, if there are 5 measured time points but an individual was only measured at the first 3 time points, the product of $p^j(y_{it})$ in Equation 3 is only taken over the first 3 time points. Nagin (2005) suggests that handling missing data in this fashion requires the assumption that data are missing completely at random (MCAR; Rubin, 1976). However, even if the MCAR assumption is met, it is still likely that missing data will have an effect on the ability to distinguish among classes and also the average posterior probability for each class. It is of interest to determine the extent of missing data which will not impact the enumeration measures' abilities to identify the true class solution.

The percent missing was specified as 0%, 10%, or 20% monotonic missing and met the MCAR assumption. Specifying the missing data to be monotonic missing emulates a situation where a participant drops out of the study or is lost to follow up and never returns. The monotonic missing situation was created by implementing the rule

$$\begin{cases} P\left(y_{it}^{j} = missing \middle| y_{i(t-1)}^{j} \neq missing\right) = P(missing) \\ P\left(y_{it}^{j} = missing \middle| y_{i(t-1)}^{j} = missing\right) = 1. \end{cases} \qquad t = 2, \ldots, 5 \text{ or } 9$$

There were 5 data patterns for the condition with 5 observations, the complete case, the case with data missing at the fifth time point, the case with data missing at the fourth and fifth time points, the case with missing data at the third, fourth and fifth time points, and the case with missing data all time points except the first. Similarly, there were 9 data patterns for the condition with 9 observations. Twenty percent was chosen for the largest missing percent because it is viewed as being large enough while still being in an acceptable range. The amount of missing data was larger than 20% in some of the reviewed GBTM studies but 50% missing, for example, could be considered unacceptable in many studies.

3.3 PROCEDURE

The sample size, number of repeated measures, class mixing proportions, percent missing and class separation conditions were looped and 250 replications were performed for each condition. As was mentioned, data was generated with SAS Proc IML. A seed number for random number generation was set based on the conditions and replication numbers. A 3 × 4 parameter matrix and a design matrix were created. The design max was 5 × 3 or 9 × 3 when the number of observations was specified to be 5 or 9, respectively. Four class sample size variables were created by multiplying the total sample size by each class's mixing proportion. Four class specific design matrixes were created by repeating the design matrix by each classes sample size. Four $(n^j \times t) \times 1$ residual vectors were randomly generated, one for each class. An outcome vector for each class was generated by multiplying the class specific design matrix with the class specific column of the parameter matrix and then adding the class specific residual matrix. The outcome vectors were then merged and restructured to produce the $n \times t$ matrix of observations. For the binary logit model, $y_{it}^{(j)}$ was recoded as 1 if $y_{it}^{(j*)} > 0$ and recoded as 0 if $y_{it}^{(j*)} \leq 0$. For the censored normal model, the constant 20 points were added to each observation and $y_{it}^{(j)}$ was recoded as 30 if $y_{it}^{(j*)} > 30$ at this point in the data generation process. The observations were combined into a dataset with a group number variable and one variable representing the time of measurement for each of the repeated measures.

The data was analyzed for 1-, 3-, 4-, and 5-class solutions. The 1-class solution was run to obtain the log likelihood needed for the computation of NEC and the 3-, 4- and 5-class solutions were run to assess the performance of the enumeration measures. The analysis of the 4-class solution and data collection will be discussed but it should be noted that the same process was used to analyze and collect data from the 3- and 5-class solutions. A 4-class solution with quadratic trajectories was run in Proc TRAJ with a maximum score boundary of 30. The SAS log was written to a text file in order to screen for convergence issues. Any convergence issues such as singular convergence, inability to compute standard errors, or extreme standard errors were recorded. Next, entropy was computed using the posterior probabilities provided in a Proc TRAJ output dataset. The enumeration measures were then computed. BIC, CAIC, HQ, ssBIC, ssCAIC, ssHQ, ICL-BIC, and ssICL-BIC were each computed using n and the number of non-missing observations as the sample size. BIC, CAIC, HQ, ssBIC, ssCAIC, ssHQ, ICL-BIC, and ssICL-BIC computed with the number of non-missing observations as the sample size were denoted BIC2, CAIC2, HQ2, ssBIC2, ssCAIC2, ssHQ2, ICL-BIC2, and ssICL-BIC2. This was done for exploratory purposes because Proc TRAJ computes BIC using both the number of subjects and the number of non-missing observations as the sample size.

The VLMR and aVLMR tests were not performed during the computation of the enumeration measures in the procedure discussed above. The computation of the null distribution to obtain the *p*-value for the VLMR and aVLMR tests is complicated. Furthermore, validated VLMR and aVLMR test results could easily be obtained using MPlus. Therefore, MPlus was used to conduct the two tests by running a GMM constrained to be a GBTM. The parameter estimates from the 4- and 5-class solutions were read into MPlus as starting values to maintain as similar solutions as possible between Proc TRAJ and MPlus. The resulting VLMR and aVLMR output was then read back into SAS and appended to the dataset containing the other enumeration measures.

The convergence issues and enumeration measures from the 3-, 4-, and 5-class solutions were combined into one dataset with three rows. This dataset was then appended to a .csv file. Similarly, the parameter estimates, standard error estimates, and average posterior probabilities from the 3-, 4-, and 5-class solutions were combined into one dataset with three rows. This dataset was also appended to a .csv file. Appending the results to a .csv file marked the completion of one replication for one condition of the simulation. The process was repeated for the full number of replications for each condition.

Each of the 15 enumeration measures in Table 1 and the 8 enumeration measures computed with the number of non-missing observations as the sample size were used to select from among 3-, 4-, and 5-class solutions for each generated data set in order to assess their performance. The comparison of enumeration measures for estimated 3- and 5-class solutions to a true 4-class solution was a common method employed in simulation studies examining enumeration measures. It is assumed that if an enumeration measure does not select a 3-class solution over a true 4-class solution then it would not select a 2-class solution over a true 4-class solution. The same statement can be made for 5- and 6-class solutions.

3.4 MEASURES

The performance of an enumeration measure was assessed by determining the rate at which it selected the true 4-class solution. The 3-, 4-, or 5-class solution with the smallest AIC, CAIC, ssCAIC, BIC, ssBIC, HQ, ssHQ, NEC, CLC, ICL-BIC, ssICL-BIC, BIC2, CAIC2, HQ2, ssBIC2, ssCAIC2, ssHQ2, ICL-BIC2, and ssICL-BIC2 was considered the best fitting for each respective measure. The largest value for sE indicated the best fitting of the 3 models. If the VLMR and aVLMR tests were significant ($p \leq .05$) when comparing a 3-class solution to a 4-class solution then the 4-class solution was selected as the better fitting of the 2 solutions. Similarly, if the VLMR and aVLMR tests were significant ($p \leq .05$) when comparing a 4-class solution to a 5-class solution then the 5-class solution was selected as the better fitting of the 2 solutions. A B_{10} value of 2 or greater when comparing 3- and 4-class solutions or the 4- and 5-class solutions indicated that the solution with a larger number of classes was a better fit.

The proportion of times each enumeration measures selected the 3-, 4- and 5-class solution was recorded. The proportions served to indicate the direction of errors made by the enumeration measures. In addition, a dichotomous variable was created for each enumeration measure and coded as 1 if an enumeration measure selected the true 4-class solution or 0 if it selected the 3- or 5-class solution.

3.5 ANALYSIS PLAN

Solutions with convergence issues were identified. Replications where at least 1 solution failed to converge and at least 1 converged were dealt with in the same manner as was done in Tofighi and Enders (2008). If the 3- and 5-class solutions failed to converge but the 4-class solution converged, all 15 enumeration measures were given credit for identifying the true 4-class solution. If the 5-class solution failed to converge and an enumeration measure favored the 4-class solution over the 3-class solution, the enumeration measure was given credit for identifying the true 4-class solution. Similarly, if the 3-class solution failed to converge and an enumeration measure favored the 4-class solution over the 5-class solution the enumeration measure was given credit for identifying the true 4-class solution. Replications with convergence issues for the 3-, 4-, and 5-class solutions were flagged.

Initially, a repeated measures logistic regression was to be performed on the 23 dichotomized enumeration measures predicted by sample size, number of repeated measures, class mixing proportions, percent missing, class separation, all 10 possible two way interaction effects, and all 10 possible three way interaction effects. The model was to be estimated using a generalized estimating equation (GEE) method. This method was going to be used because the repeated observations (23 dichotomous enumeration measures) required that the correlation among scores be taken into account. Liang and Zeger (1986) developed GEE as an extension of the general linear model (GLM) to handle repeated measures with a focal point on regression. GEE handles the correlation among repeated measures through the specification of a working correlation matrix. However, there were 2 factors that caused convergence issues when attempting to run the GEE model. There was multicollinearity among some of the dichotomized enumeration measure performance variables. Also, some of the enumeration measures never selected the correct 4-class solution in certain cells of the design. As an alternative, a classification and regression trees (CART) analysis was performed.

A CART analysis can be performed on both continuous and categorical outcome variables and with both continuous and categorical predictor variables. The basic idea behind a CART analysis is to classify a case based on a set of observed predictors taking into account the possibility that different relationships may be present between variables in different parts of the parameter space (Breiman, Friedman, Olshen, & Stone, 1984). An advantage of this type of analysis is that it is a non-parametric approach and does not make any assumptions about an underlying statistical model. A CART analysis was utilized because it allowed for the exploratory investigation of complex interactions among the predictor variables without being affected by the high correlations among some of the enumeration measures.

Figure 4 demonstrates a simple CART for the purpose of explaining the procedure. A CART analysis starts with all observations in what is called the root node and splits the root node into two child nodes (Node 1 and Node 2) based on some predictor variable, in this case the categorical predictor variable X, such that the resulting child nodes are as homogeneous as possible with respect to the dichotomous outcome variable, Y. At this point, Node 1 and Node 2 reside at level 1 in the CART. It is important to note that a node can only ever be split into 2 child nodes in a CART analysis. After the root node has been split into the child nodes at level 1 (Node 1 and Node 2), if possible, these nodes are further split into even more homogeneous child nodes based on some predictor variable. As demonstrated in Figure 4, Node 1 was further split into child nodes, Node 3 and Node 4, based on the categorical predictor variable Z but Node 2 was not split. At this point, Node 3 and Node 4 reside at level 2. Node 1 and Node 2 were child nodes of the root node but Node 1 is the parent node of Node 3 and Node 4. This process is

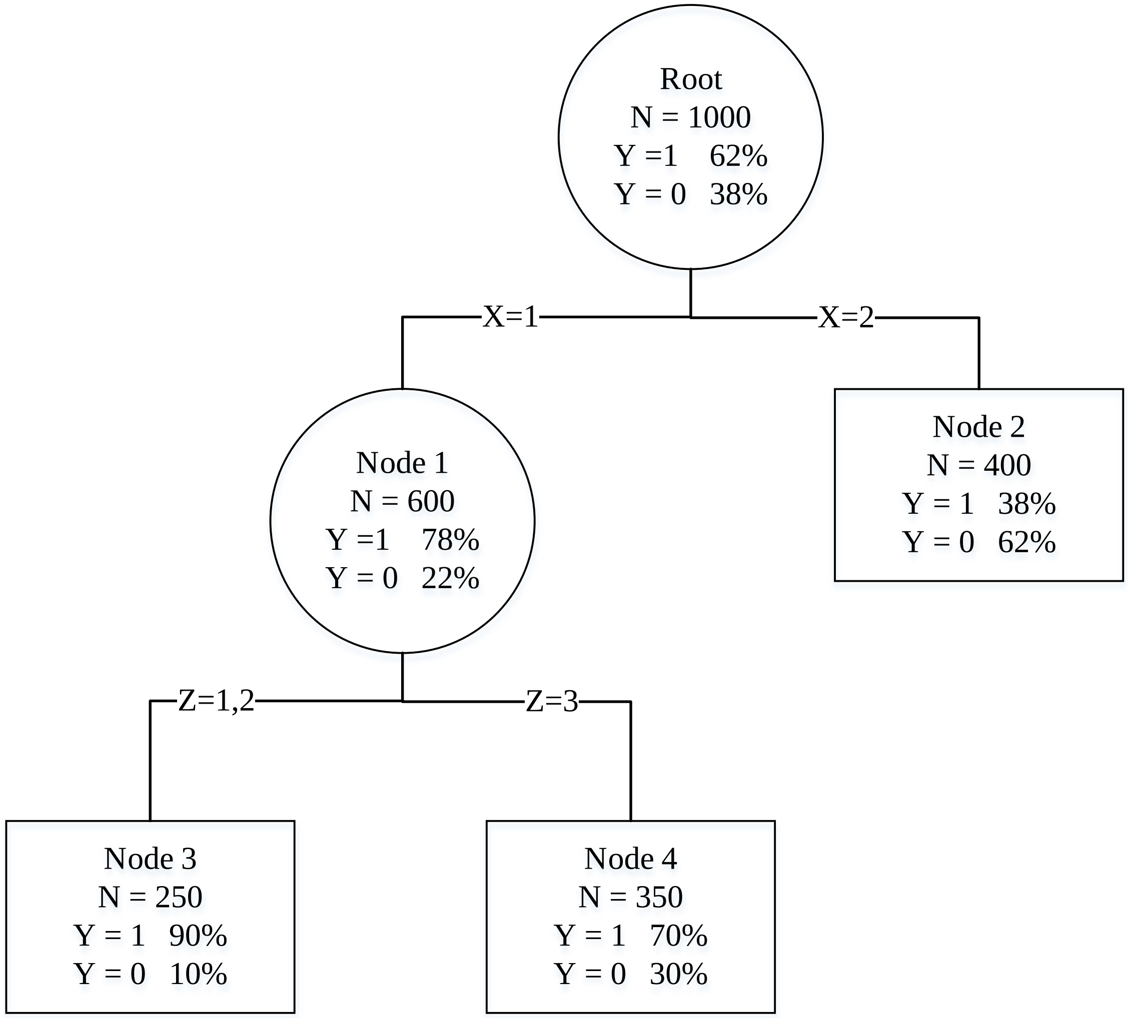

Figure 4. Example classification and regression tree

repeated until all nodes can no longer be split, either because the node contains only one observation or because the improvement in prediction of the outcome variable does not meet some predetermined value. A node that does not have any child nodes is called a terminal node, Node 2, Node 3, and Node4 in the example Figure 4. Along with the information provided by the CART nodes, the sensitivity, specificity, and an overall prediction rate for the final model are provided.

4 RESULTS

4.1 CENSORED NORMAL MODEL

4.1.1 Model convergence

Across the conditions, the 3-, 4-, and 5-class solutions all failed to converge in 16 of the total 36,000 replications. These non-converging replications were most frequent when the class mixing proportions were extremely unequal and there was low separation among the classes.

Across the conditions, the 3-, 4-, and 5-class solutions all converged in approximately 92% of the replications. The percent of specific convergence patterns across the class number solutions are displayed in Table 5. For example, pattern 1 demonstrates that the 3-, 4-, and 5-class solution all converged in approximately 92% of all replications while pattern 7

Table 5. Percent of class specific convergence patterns for the censored normal model

Pattern	3-Class	4-Class	5-Class	Frequency	Percent
1	✓	✓	✓	33044	91.8
2	×	×	×	16	< .1
3	×	×	✓	23	< .1
4	×	✓	×	45	.1
5	×	✓	✓	181	.5
6	✓	×	×	136	.4
7	✓	✓	×	2198	6.1
8	✓	×	✓	357	1.0

✓ = Converged; × = Did not converge.

demonstrates that the 3- and 4-class solutions converged but the 5-class solution failed to converge in approximately 6% of all replications. When the 3-, 4-, and 5-class solutions all failed to converged, pattern 2, the replications (16) were excluded from the analysis. Failure to converge was most common in the 5-class solution at 6.6% of all replications. The 4-class solution failed to converge in 1.5% of the replications and the 3-class solution failed to converge in .7% of the replications. Increases in sample size, percent missing, the difference in class proportions, and a decrease in separation among the classes all resulted in a larger number of replications that failed to converge for the 3-, 4-, and 5-class solutions. When the number of repeated measures increased from 5 to 9, there was a decrease in the number of replications that failed to converge for the 3-class solution but an increase in the number of replications that failed to converge for the 4- and 5-class solutions.

Non-convergence for the true 4-class solution was examined in more depth. There were 4 out of the 144 cells where 10% to 12% of the 4-class solutions failed to converge. The 4-class solution failed to converge at a rate less than 10% for the remaining 140 cells. The 4-class solution failed to converge most often when the sample size was large (N = 800 or N = 2000), the class mixing proportions were extremely unequal, and there was low separation among the classes. Similar patterns were observed for the non-convergence of the 3- and 5-class solutions. It was assumed that when the class mixing proportions were extremely unequal and the sample size is large, the larger classes are more dominant than when the sample size was small.

4.1.2 CART analysis results

CART analysis was performed on correct classification rate (correct: selected the 4-class solution, incorrect: did not select the 4-class solution) by sample size, the number of repeated measures, class mixing proportions, percent missing, separation among the classes, and enumeration measures. The overall correct classification of the CART analysis model was 82%. The model was able to correctly classify 92% of the incorrect outcomes while only correctly classifying 57% of the correct outcomes.

At the level of the root node, which contained 827,632 observations, the rate of correctly selecting the 4-class solution was 28% averaged over all conditions and all enumeration measures. At level 1, the root node was split by the performance of the enumeration measures into two child nodes which will be identified as the "good" and "bad" performing groups for ease of identification. The enumeration measures by group are,

"good" {AIC, ssCAIC, ssBIC, ssBIC2, HQ, HQ2, ssHQ, ssHQ2, VLMR, aVLMR}

and

"bad" {BIC, BIC2, B_{10}, CAIC, CAIC2, ssCAIC2, sE, NEC, CLC, ICL-BIC, ICL-BIC2, ssICL-BIC, ssICL-BIC2}.

Averaged over all of the enumeration measures in each group, the correct classification rate was 45% in the good performing group and 14% in the bad performing group. The measures' rates of selecting the correct 4-class solution ranged from 32% to 56% in the good performing group and 3% to 28% in the bad performing group. At level 2, separation among the classes and enumeration measures were the most important predictors. At level 3, class mixing proportions, separation among the classes, and enumeration measures were the most important predictors. At level 4, sample size, class mixing proportions, separation among the classes, percent missing, and enumeration measures were the most important predictors. At level 5, the lowest level of the CART, sample size, the number of repeated measures, class mixing proportions, and enumeration measures were the most important predictors. Figure 5 displays the CART diagram for predicting the correct selection of the true 4-class solution for the censored normal model.

There were 16 terminal nodes in the good performing group which will each be discussed in turn. The factors and factor levels that separated these 16 terminal nodes are summarized in Table 6. In Node 31, ssHQ, VLMR, aVLMR, and ssBIC2 performed similarly with an average correct classification rate of 42% when there was high separation among the classes, the class mixing proportions were either equal or moderately unequal, and the sample size was low (200 or 400). Given the same conditions in Node 32, AIC, HQ, ssBIC, ssCAIC, HQ2, and ssHQ2 performed noticeably better with an average correct classification rate of 67%.

In Node 33, AIC, HQ, ssBIC, ssCAIC, HQ2, and ssBIC2 had an average correct classification rate of 88% when the separation among the classes was high, the class mixing proportions were either equal or moderately unequal, and the sample size was medium to high (800 or 2000). Given the same conditions in Node 34, ssHQ, VLMR, aVLMR, and ssHQ2 achieved an average correct classification rate of 70%.

In Node 35, ssBIC, ssCAIC, VLMR, aVLMR and ssBIC2 had an average correct classification rate of 25% when there was high separation among the classes, the mixing proportions were extremely unequal, and the number of repeated measures was 5. In Node 36, where the number of repeated measures was increased to 9 holding the separation and mixing proportion conditions constant, the average correct classification rate for the 5 enumeration measures increased to 41%.

In Node 37, AIC, HQ, ssHQ, HQ2, and ssHQ2 had an average correct classification rate of 53% when there was high separation among the classes, the mixing proportions were extremely unequal, and the sample size was small to medium (200, 400, or 800). In Node 38,

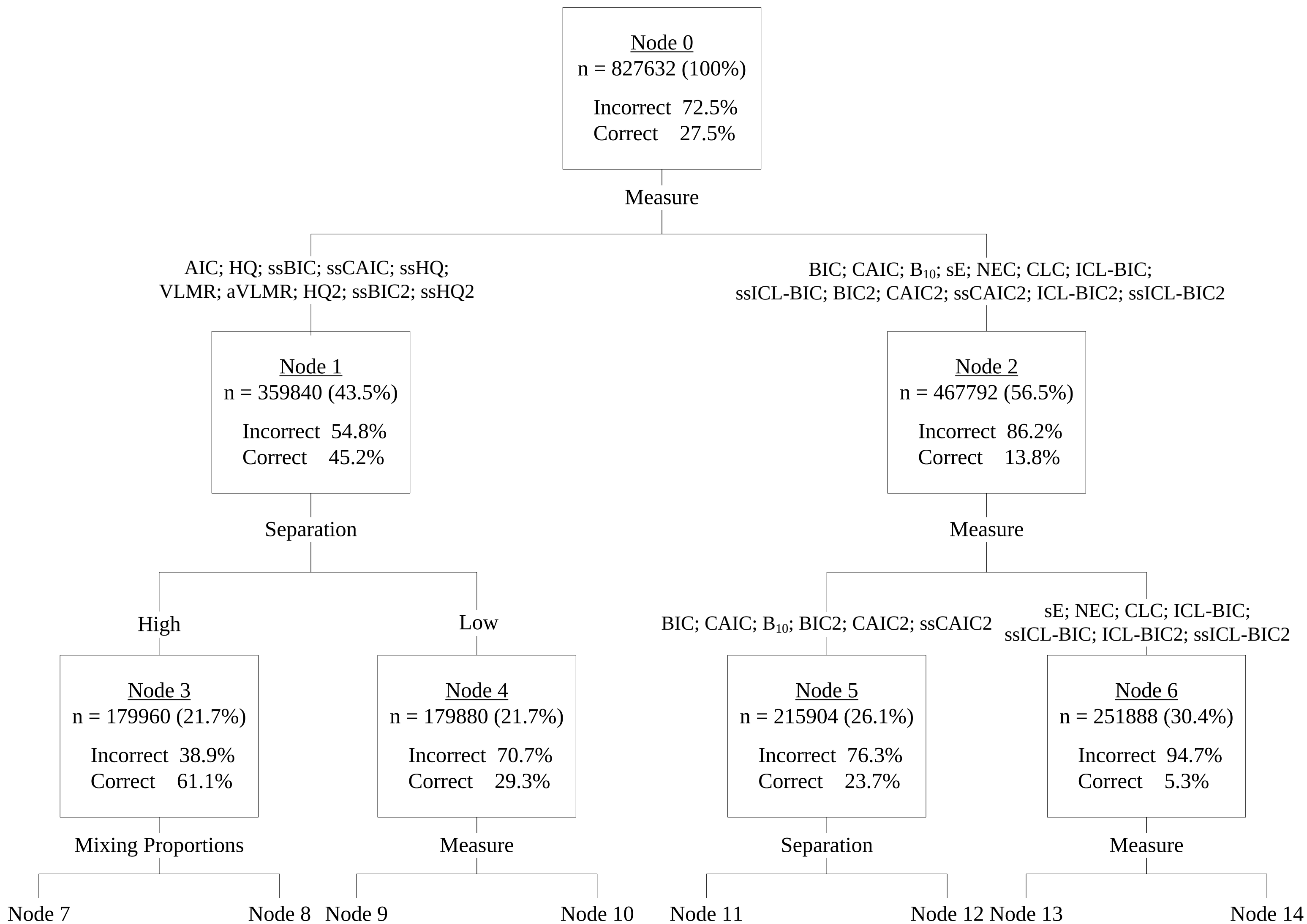

Figure 5. CART predicting correct selection of the true 4-class solution for the censored normal model

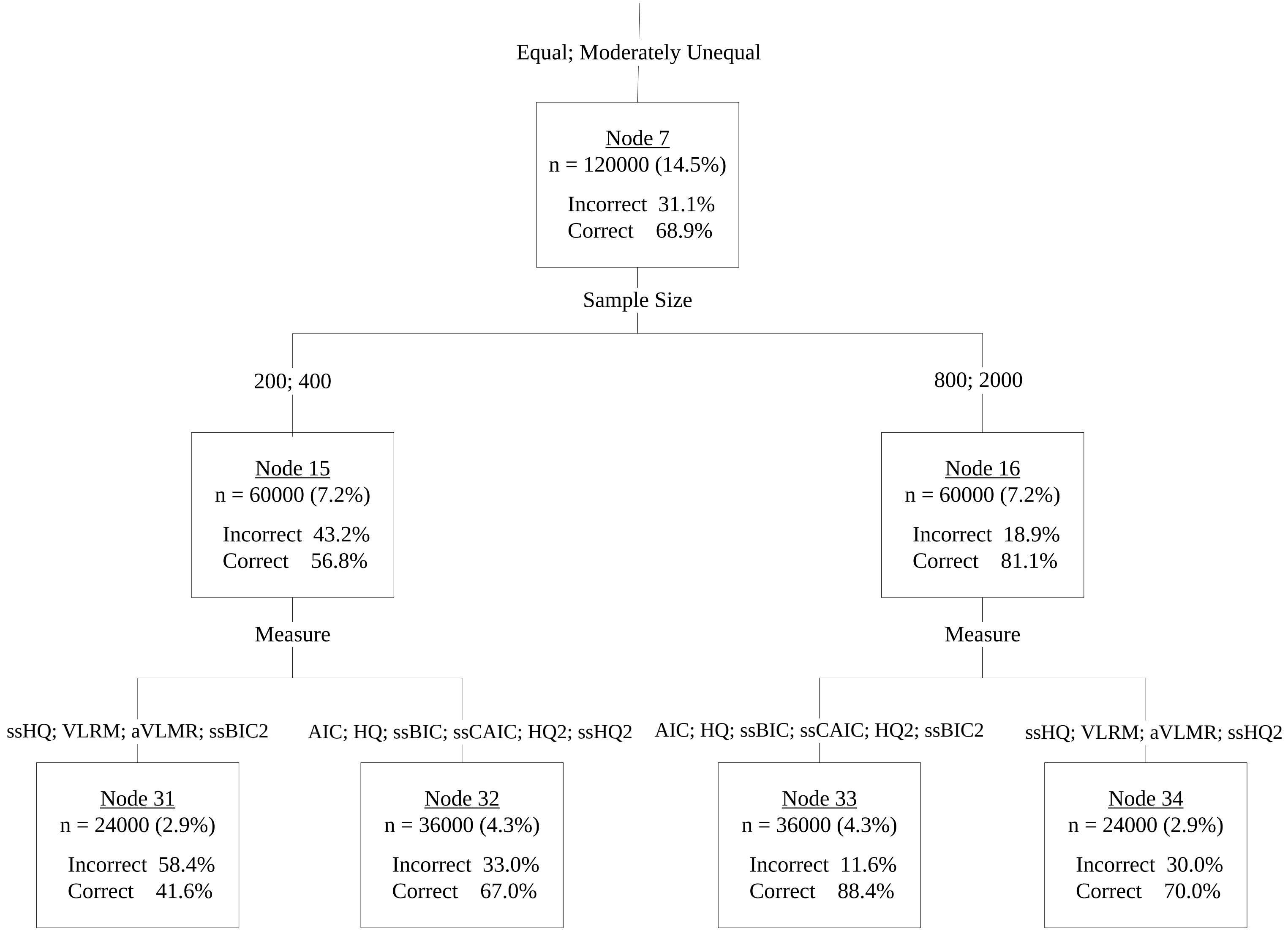

Figure 5 (continued).

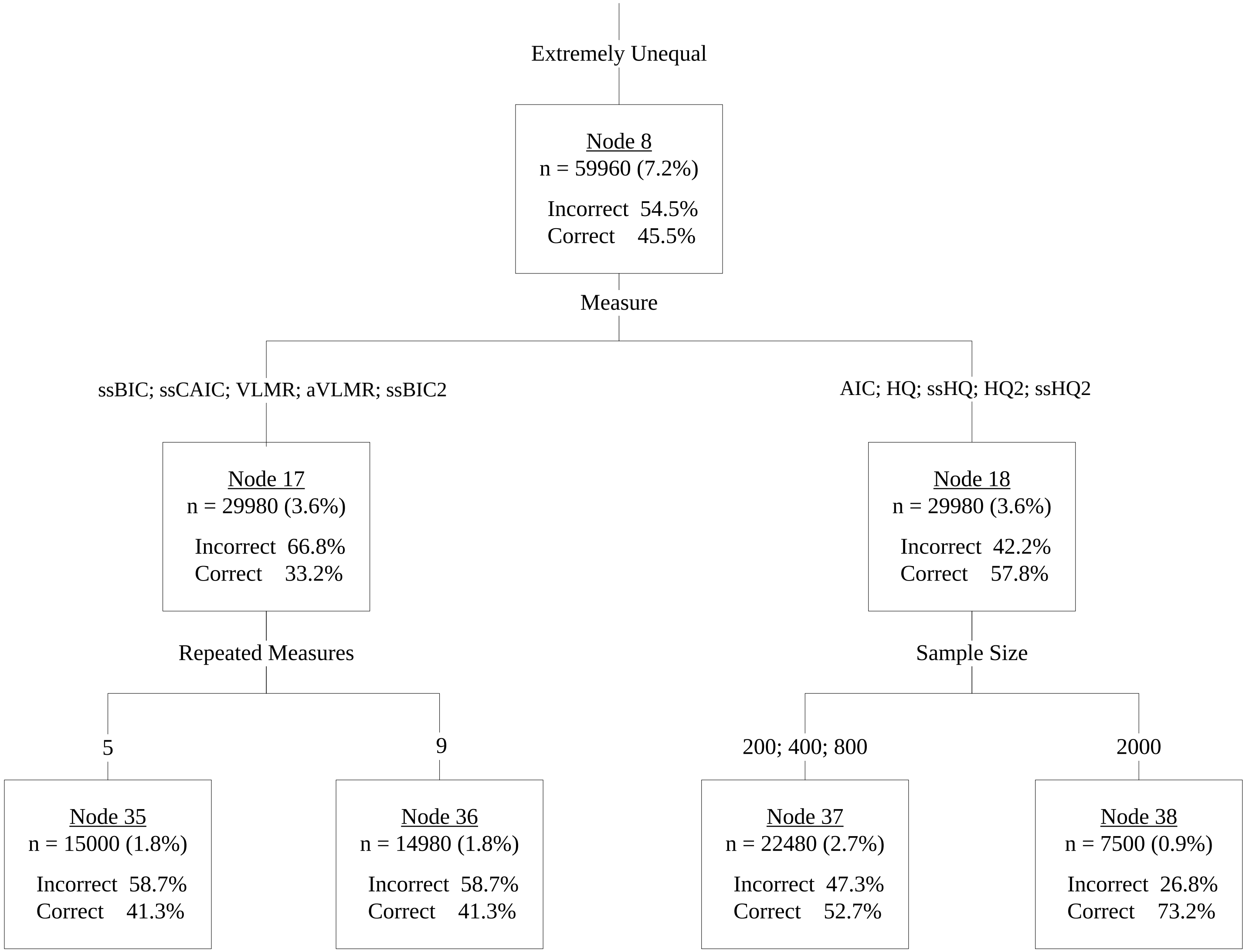

Figure 5 (continued).

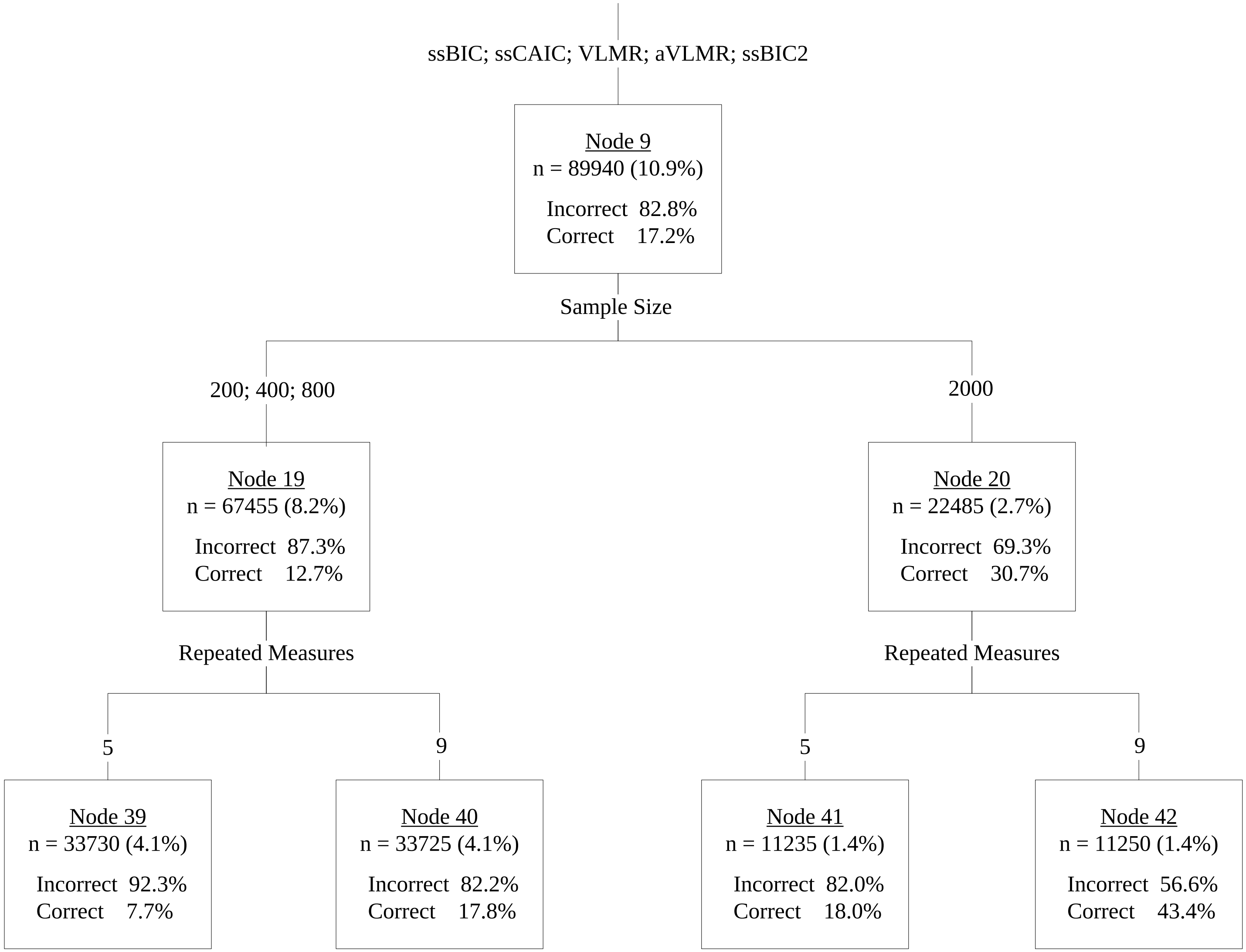

Figure 5 (continued).

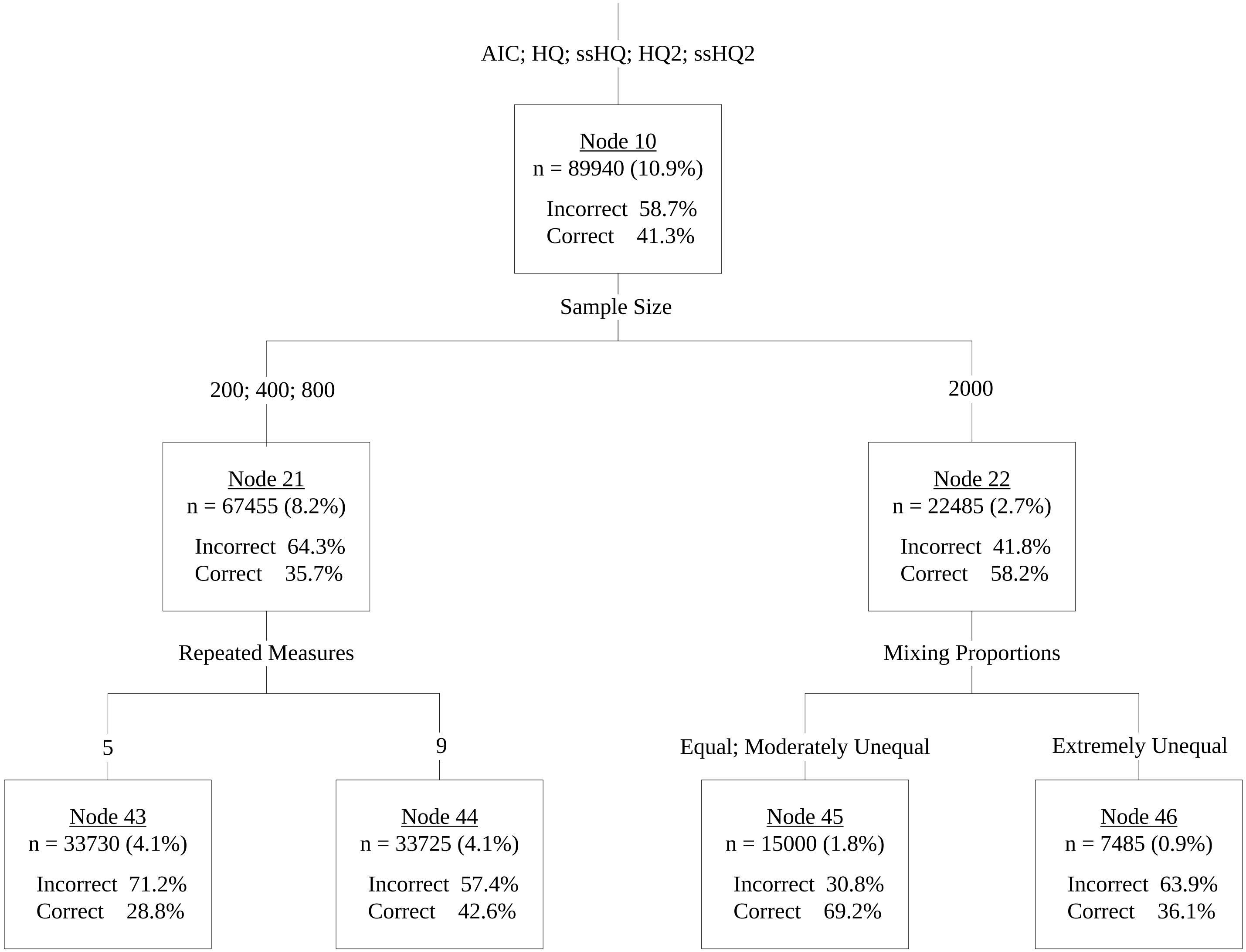

Figure 5 (continued).

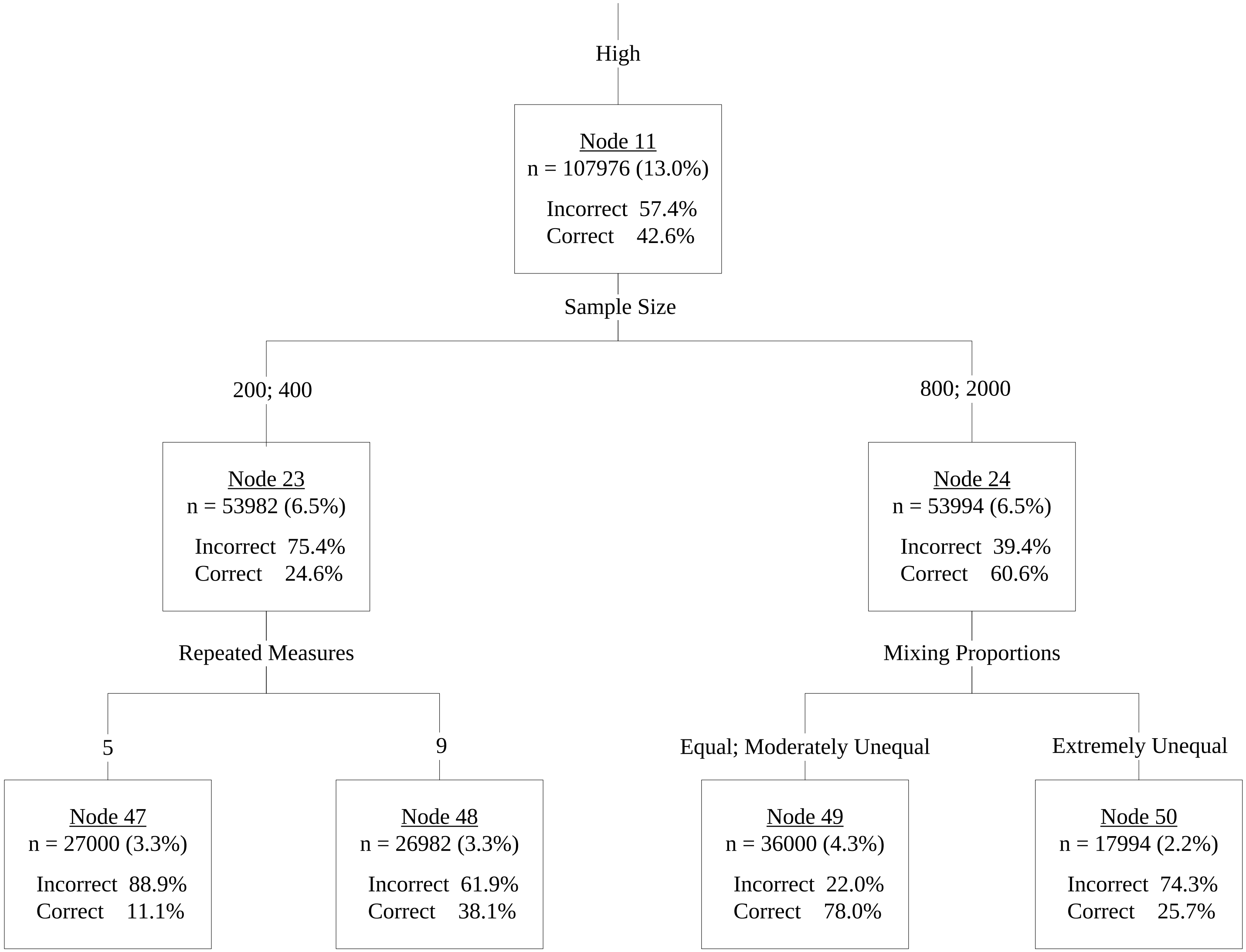

Figure 5 (continued).

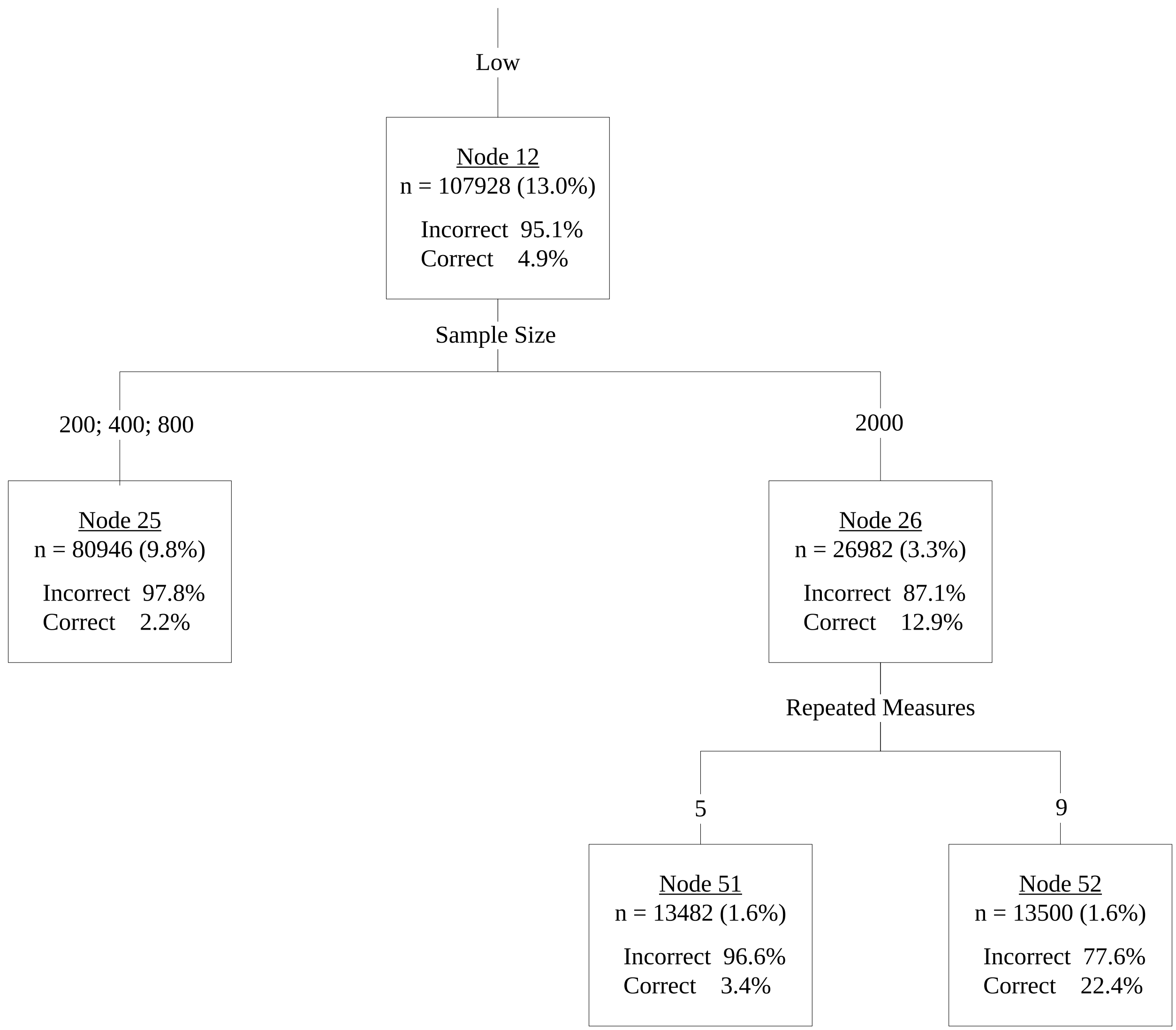

Figure 5 (continued).

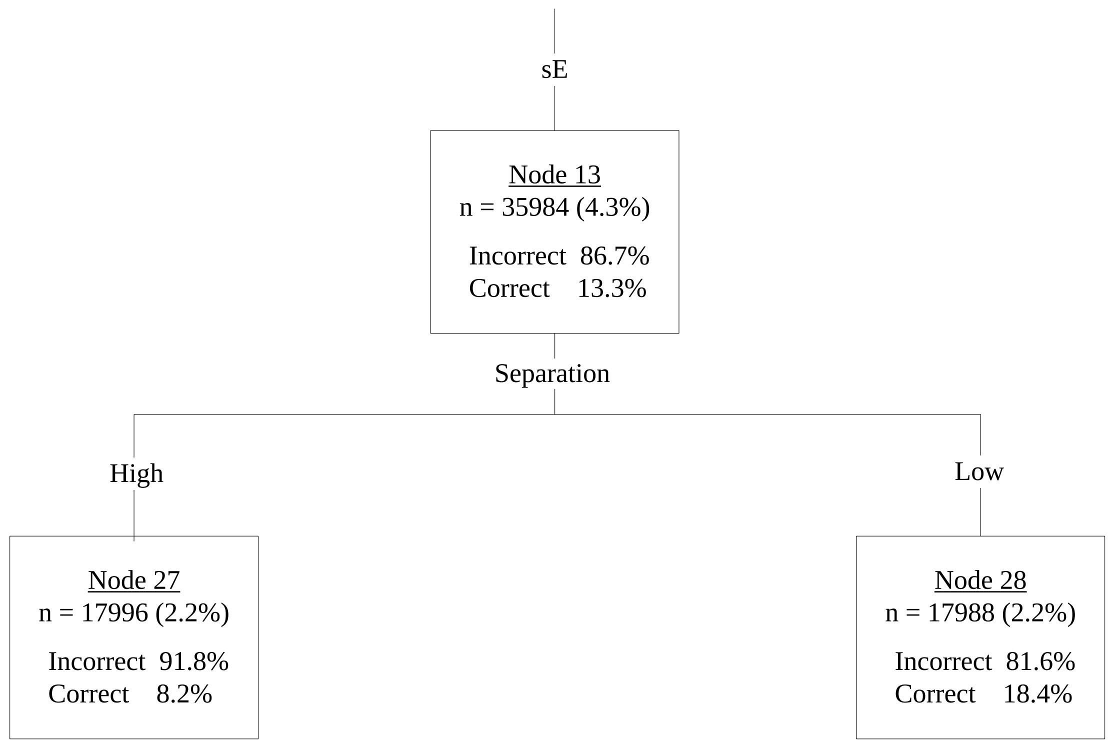

Figure 5 (continued).

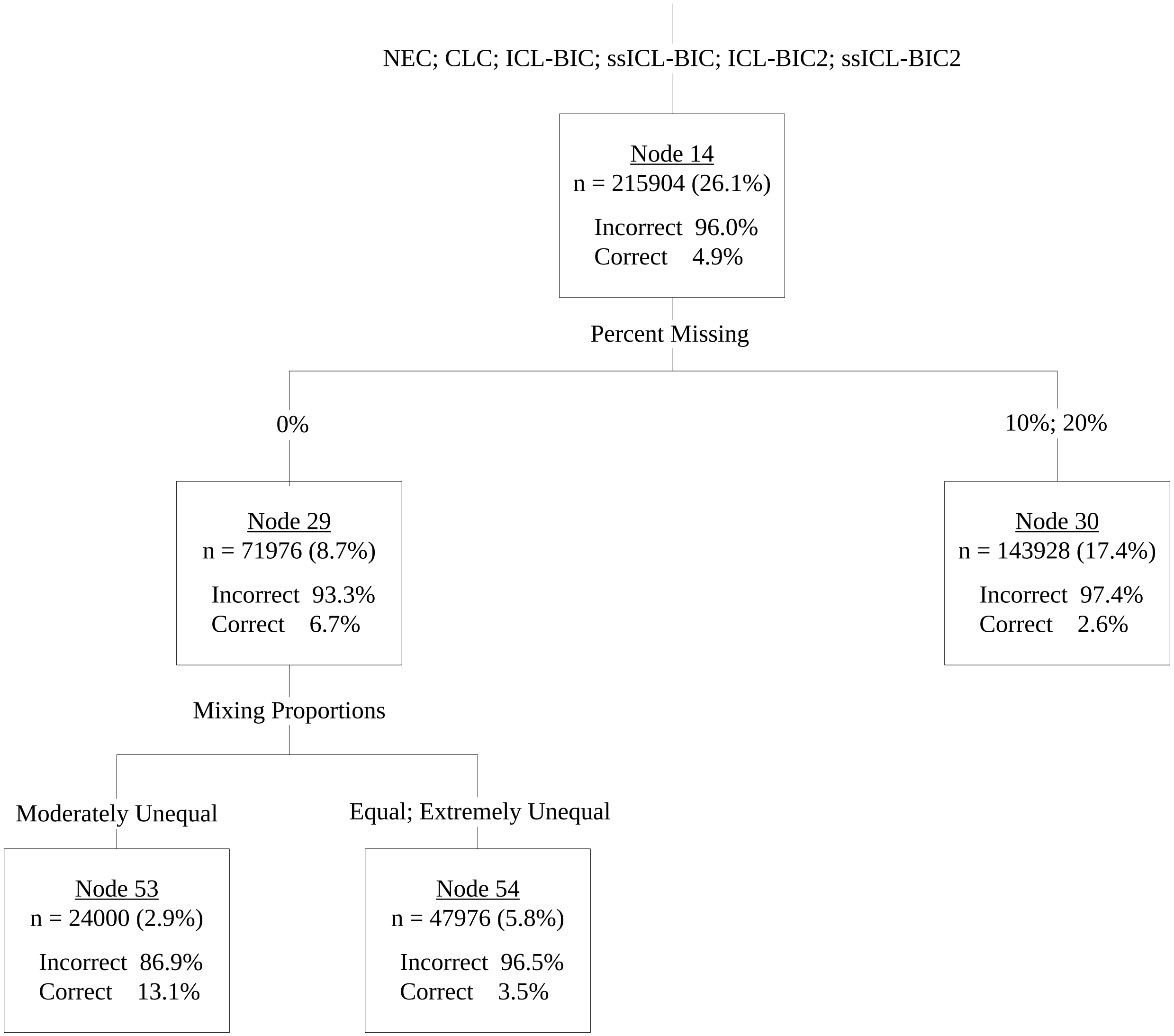

Figure 5 (continued).

Table 6. Terminal node summary of the good performing enumeration measures for the censored normal model

Node	Sample Size	No. Obs.	Mixing Proportions	Class Separation	Measure	Percent Correct	Percent of Total
31	200; 400		Equal; Moderately Unequal	High	ssHQ; VLMR; aVLMR; ssBIC2	41.6	2.9
32	200; 400		Equal; Moderately Unequal	High	AIC; HQ; ssBIC; ssCAIC; HQ2; ssHQ2	67.0	4.3
33	800; 2000		Equal; Moderately Unequal	High	AIC; HQ; ssBIC; ssCAIC; HQ2; ssBIC2	88.4	4.3
34	800; 2000		Equal; Moderately Unequal	High	ssHQ; VLMR; aVLMR; ssHQ2	70.0	2.9
35		5	Extremely Unequal	High	ssBIC; ssCAIC; VLMR; aVLMR; ssBIC2	25.1	1.8
36		9	Extremely Unequal	High	ssBIC; ssCAIC; VLMR; aVLMR; ssBIC2	41.3	1.8
37	200; 400; 800		Extremely Unequal	High	AIC; HQ; ssHQ; HQ2; ssHQ2	52.7	2.7
38	2000		Extremely Unequal	High	AIC; HQ; ssHQ; HQ2; ssHQ2	73.2	0.9
39	200; 400; 800	5		Low	ssBIC; ssCAIC; VLMR; aVLMR; ssBIC2	7.7	4.1
40	200; 400; 800	9		Low	ssBIC; ssCAIC; VLMR; aVLMR; ssBIC2	17.8	4.1
41	2000	5		Low	ssBIC; ssCAIC; VLMR; aVLMR; ssBIC2	18.0	1.4
42	2000	9		Low	ssBIC; ssCAIC; VLMR; aVLMR; ssBIC2	43.4	1.4
43	200; 400; 800	5		Low	AIC; HQ; ssHQ; HQ2; ssHQ2	28.8	4.1
44	200; 400; 800	9		Low	AIC; HQ; ssHQ; HQ2; ssHQ2	42.6	4.1
45	2000		Equal; Moderately Unequal	Low	AIC; HQ; ssHQ; HQ2; ssHQ2	69.2	1.8
46	2000		Extremely Unequal	Low	AIC; HQ; ssHQ; HQ2; ssHQ2	36.1	0.9

when the sample size was increased to large (2000) while holding the separation and mixing proportion conditions constant, the average correct classification rate for the 5 enumeration measures increased to 73%.

In Node 39, ssBIC, ssCAIC, VLMR, aVLMR, and ssBIC2 performed similarly with an average correct classification rate of 8% when there was low separation among the classes, the sample size was small to medium (200, 400, or 800), and there were 5 repeated measures. In Node 40, the rate of correct classification increased to 18% for the same 5 measures in the same low separation and sample size conditions when the number of repeated measures increased to 9. The average rate of correct classification for ssBIC, ssCAIC, VLMR, aVLMR, and ssBIC2 was also higher in Node 41 (18%) relative to Node 39 (8%). Both Nodes 41 and 39 had low separation among the classes and 5 repeated measures but differed on sample size. The sample size in Node 41 was large (2000) while the sample size in Node 39 was small to medium (200, 400, or 800). In Node 42, the same 5 measures has an average correct classification rate of 43% when there was low separation among the classes, the sample size was large (2000), and the number of repeated measures was 9.

In Node 43, AIC, HQ, ssHQ, HQ2, and ssHQ2 had an average correct classification rate of 29% when there was low separation among the classes, the sample size was small to medium (200, 400, or 800), and there were 5 repeated measures. In Node, 44, the rate of correct classification increased to 43% for the same 5 measures in the same low separation and sample size conditions when the number of repeated measures increased to 9.

In Node 45, AIC, HQ, ssHQ, HQ2, ssHQ2 had an averaged correct classification rate of 69% when the separation among the classes was low, the sample size was large (2000), and the mixing proportions were either equal or moderately unequal. Holding the levels of separation and sample size constant, the same 5 enumeration measures had a decrease in average correct classification rate to 36% when the class mixing proportions became extremely unequal, Node 46.

There were 12 terminal nodes in the bad performing group. The factors and factor levels that separated these 12 terminal nodes are summarized in Table 7. Before discussing each of the terminal nodes it is of interest to note that the enumeration measures in the bad performing group were further split by type. The information criteria (24% correct classification rate) and the enumeration measures (5% correct classification rate) were split into separate nodes.

In Node 47, BIC, CAIC, B_{10}, BIC2, CAIC2, and ssCAIC2 had an average correct classification rate of 11% when there was high separation among the classes, the sample size was small (200 or 400), and the number of repeated measures was 5. The same 6 enumeration measures had an increase in correct classification rate to an average of 38% when the number of repeated measures increased to 9 and the levels of separation and sample size conditions were held constant, Node 48.

In Node 49, BIC, CAIC, B_{10}, BIC2, CAIC2, and ssCAIC2 had an average correct classification rate of 78% when there was high separation among the classes, the sample size was medium to large (800 or 2000) and the class mixing proportions were either equal or moderately unequal. However, Node 50 shows that even with high separation and a medium to large sample size (800 or 2000), the average correct classification rate for the 6 enumeration measures was only 26% given that class proportions were extremely unequal.

In Node 25, BIC, CAIC, B_{10}, BIC2, CAIC2, and ssCAIC2 had an average correct classification rate of 2% when there was low separation among the classes and the sample size was small to medium (200, 400, or 800). The same 6 enumeration measures were further

Table 7. Terminal node summary of the bad performing enumeration measures for the censored normal model

Node	Sample Size	No. Obs.	Mixing Proportions	Percent Missing	Class Separation	Measure	Percent Correct	Percent of Total
47	200; 400	5			High	BIC; CAIC; B_{10}; BIC2; CAIC2; ssCAIC2	11.1	3.3
48	200; 400	9			High	BIC; CAIC; B_{10}; BIC2; CAIC2; ssCAIC2	38.1	3.3
49	800; 2000		Equal; Moderately Unequal		High	BIC; CAIC; B_{10}; BIC2; CAIC2; ssCAIC2	78.0	4.3
50	800; 2000		Extremely Unequal		High	BIC; CAIC; B_{10}; BIC2; CAIC2; ssCAIC2	25.7	2.2
25	200; 400; 800				Low	BIC; CAIC; B_{10}; BIC2; CAIC2; ssCAIC2	2.2	9.8
51	2000	5			Low	BIC; CAIC; B_{10}; BIC2; CAIC2; ssCAIC2	3.4	1.6
52	2000	9			Low	BIC; CAIC; B_{10}; BIC2; CAIC2; ssCAIC2	22.4	1.6
27					High	sE	8.2	2.2
28					Low	sE	18.4	2.2
53			Moderately Unequal	0%		NEC; CLC; ICLBIC; ssICLBIC; ICLBIC2; ssICLBIC2	13.1	2.9
54			Equal; Extremely Unequal	0%		NEC; CLC; ICLBIC; ssICLBIC; ICLBIC2; ssICLBIC2	3.5	5.8
30				10%; 20%		NEC; CLC; ICLBIC; ssICLBIC; ICLBIC2; ssICLBIC2	2.6	17.4

affected by the number of repeated measures when the sample size increased to large (2000). In Node 51, when the number of repeated measures was 5, the average rate of correct classification was 3%. In Node 52, when the number of repeated measures increased to 9, the average rate of correct classification rose to 22%.

In Node 27, sE had a correct classification rate of 8% when there was high separation among the classes. Somewhat contradictory to expectation, sE's rate of correct classification rose to 18% when the separation among the classes was low, Node 28.

In Node 53, NEC, CLC, ICL-BIC, ssICL-BIC, ICL-BIC2, and ssICL-BIC2 had an average correct classification rate of 13% when there was no missing data and the class mixing proportions were moderately unequal. In Node 54, the same 6 enumeration measures had a 4% average correct classification rate when the class mixing proportions were equal or extremely unequal and there was no missing data. In Node 30, where the percent missing was at 10% or 20%, the class mixing proportions did not have a substantial effect and the average rate of correct classification was 3%.

4.1.3 Results summary

The 16 terminal nodes in the good performing group and the 12 terminal nodes in the bad performing group, ordered by correct classification rates, are displayed in Figure 6 and Figure 7, respectively. To better address specific themes that emerged in the CART, the nodes were first split into quartiles by performance. The 7 nodes in the first quartile all had a correct classification rate of 11% or below. The 7 nodes in the second quartile had correct classification rates between 13% and 26%. The 7 nodes in the third quartile had correct classification rates

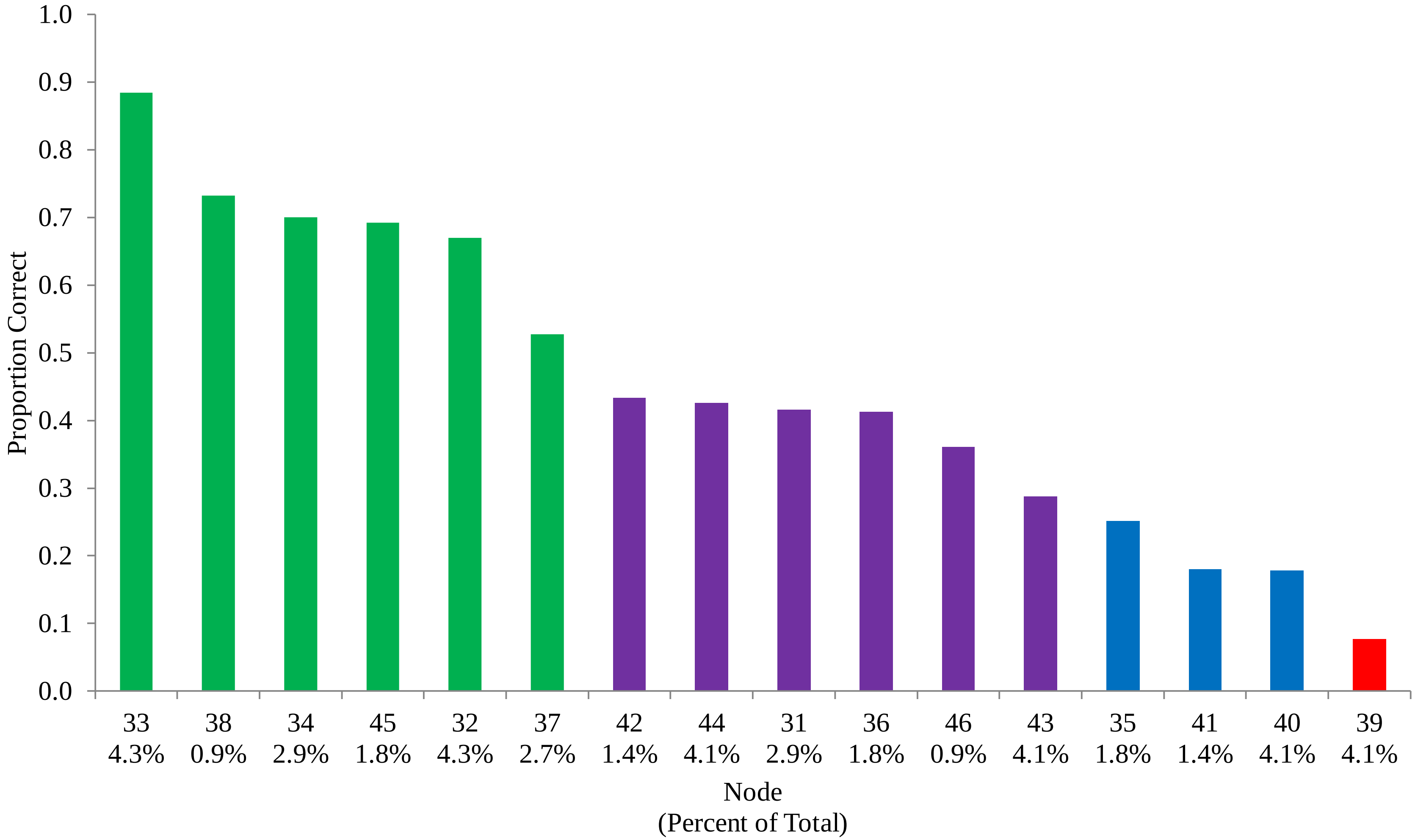

Figure 6. Proportion of times the 4-class solution was identified as correct by good performing enumeration measure nodes for the censored normal model

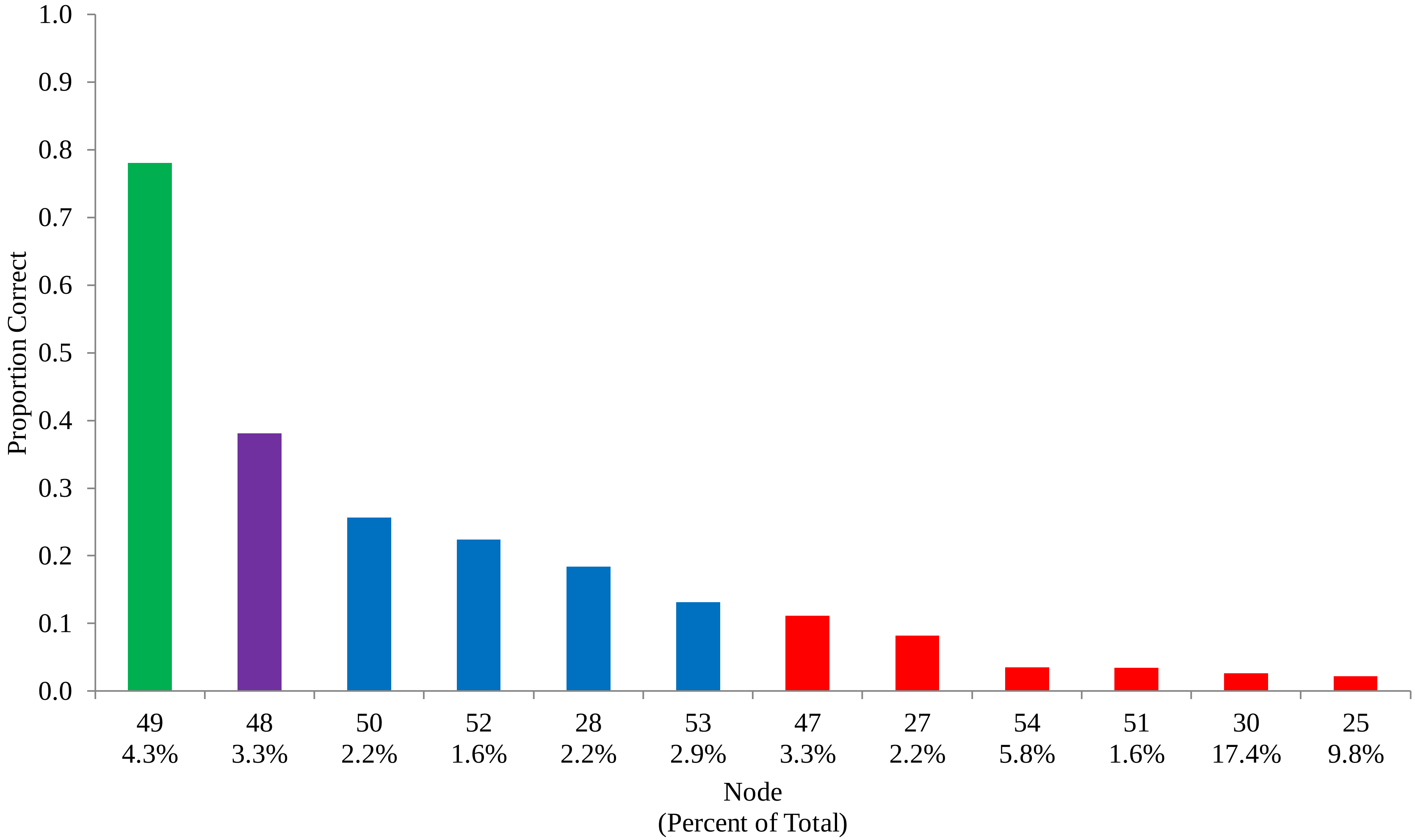

Figure 7. Proportion of times the 4-class solution was identified as correct by bad performing enumeration measure nodes for the censored normal model

between 29% and 43%. The 7 nodes in the fourth quartile had correct classification rates between 53% and 88%.

The 7 nodes in the fourth quartile were all affected by different combinations and levels of enumeration measures, sample size, class mixing proportions, and the separation among the classes. Three of these 7 best performing nodes (33, 34, 49) had sample sizes of medium to large (800 or 2000), equal and moderately unequal class mixing proportions, and high separation among the classes and all contained information criteria or the likelihood ratio test derivatives. AIC, HQ, HQ2, and ssHQ2 were present in the remaining 4 nodes that typically had ideal conditions in at least two of the three factors – sample size (2000), class mixing proportions (equal, moderately unequal), and separation among the classes (high).

Overall, given the ideal conditions of a medium to large sample size (800 or 2000), equal or moderately unequal class mixing proportions, and high separation among the classes, if an enumeration measure was an information criteria or likelihood ratio test derivative, a node was always in the top quartile. Also, as long as there was high separation among the classes, regardless of sample size or class mixing proportions, when the enumeration measure was AIC, HQ, HQ2, ssHQ, or ssHQ2, the node was always in the top quartile. ssBIC and ssCAIC were the only other enumeration measures in a node in the top quartile given non ideal conditions. These 2 information criteria performed relatively well given small sample sizes (200 or 400) as long as the class mixing proportions were equal or moderately unequal and there was high separation among the classes.

The 7 nodes in the third quartile were affected by different combinations and levels of sample size, class mixing proportions, separation among the classes, and the number of repeated measures and all were affected by enumeration measures. As with the top quartile, only information criteria and the likelihood ratio test derivatives were present in nodes in the third quartile. AIC, HQ, ssHQ, HQ2, and ssHQ2 were grouped together in 3 nodes in the third quartile (44, 46, 43) all of which had low separation among the classes. The other 4 nodes in the third quartile all had what can be considered ideal conditions in two of the three non-enumeration measure factors affecting them.

The 7 nodes in the second quartile were affected by different combinations and levels of sample size, class mixing proportions, separation among the classes, the number of repeated measures and percent missing and all were affected by enumeration measures. The nodes in the second quartile contained the information criteria in the bad performing group (BIC, CAIC, B_{10}, BIC2, CAIC2, ssCAIC2), ssBIC, ssCAIC, VLMR and aVLMR, and the entropy based statistics. Four of the 6 nodes in the second quartile affected by the separation among the classes had low separation and all 3 nodes affected by class mixing proportions had moderately to extremely unequal class mixing proportions.

The 7 nodes in the first quartile were affected by different combinations and levels of sample size, class mixing proportions, separation among the classes, the number of repeated measures and percent missing and all were affected by enumeration measures. The enumeration measures in the nodes in the first quartile were primarily entropy based statistics and the information criterion in the bad performing group (BIC, CAIC, logB10, BIC2, CAIC2, ssCAIC2) but ssBIC, ssCAIC, VLMR, aVLMR, and ssBIC2 were also present in 1 node.

A closer examination of the performances of certain enumeration measures was warranted in order to facilitate a final decision on which measure or measures to use. AIC, HQ, HQ2, ssHQ, and ssHQ2 had the highest correct classification rates in the majority of conditions. However, these measures performed similarly. For further examination, AIC was selected because of its prevalence in the literature. HQ was selected because it held a very slight edge over HQ2 and ssHQ and is slightly easier to compute than ssHQ2. BIC and ssBIC were also selected for further examination because they performed well in certain conditions and they are prevalent in the literature. The performances of these 4 select enumeration measures was investigated overall and for two specific conditions, when conditions were poor or unfavorable for correct classification and when conditions were good or favorable for correct classification.

The proportion of 3-, 4-, and 5-class solutions selected as correct by AIC, BIC, HQ, and ssBIC across all conditions are displayed in Figure 8. AIC and HQ performed very similarly with HQ holding a very slight edge in correctly selecting the true 4-class solution. Overall, AIC and HQ both tended to under extract rather than over extract classes with HQ having a higher

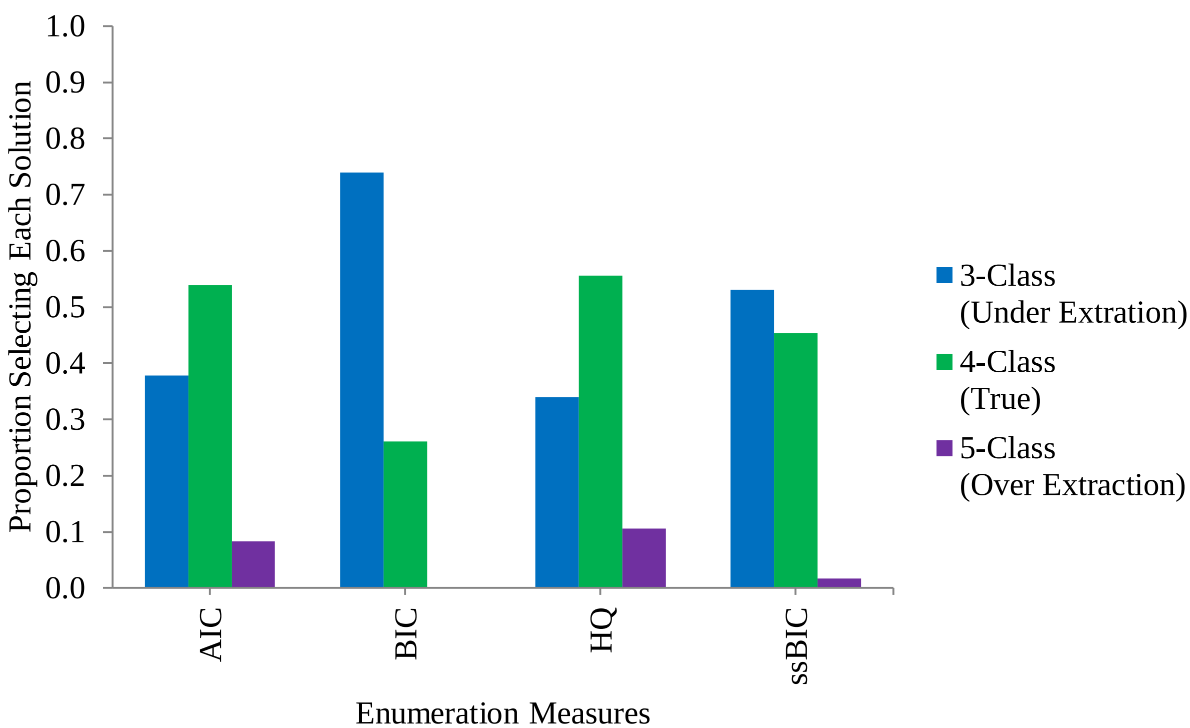

Figure 8. Proportion of 3-, 4-, and 5-class solutions identified as correct by select enumeration measures for the censored normal model

over extraction rate by a very slim margin. BIC had an overall correct classification rate of 26% which was approximately 30% below that of HQ and approximately 20% below that of ssBIC. Although ssBIC did not have quite as high of a correct classification rate as AIC and HQ, it almost exclusively under extracted rather than over extracted classes. Cohen's Kappa was computed to assess the overall rate of agreement between AIC, BIC, HQ, and ssBIC. Cohen's Kappa is a measure of agreement between 2 observers, in this case enumeration measures, beyond that which is expected just based on chance (Cohen, 1960). Landis and Koch (1977) provided interpretative guidelines for Cohen's Kappa, Table 8. Cohen's Kappas between AIC,

Table 8. Interpretation guidelines for Cohen's Kappa

Kappa	Strength of Agreement
< 0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

BIC, HQ and ssBIC over all conditions is displayed in Table 9. AIC and HQ had a very high level of agreement. BIC had a fair level of agreement with both AIC and HQ but a moderate level of agreement with ssBIC. ssBIC had a substantial level of agreement with AIC and a moderate level of agreement with HQ.

The proportion of 3-, 4-, and 5-class solutions selected as correct by AIC, BIC, HQ, and ssBIC in unfavorable conditions – small sample sizes (200 or 400), 5 repeated measures, extremely unequal mixing proportions, and low separation among the classes - are displayed in Figure 9. Given the unfavorable conditions, HQ had the highest correct classification rate at

Table 9. Cohen's Kappa between 4 select measures over all conditions for the censored normal model

	1	2	3	4
1. AIC	-			
2. BIC	.333	-		
3. HQ	.900	.294	-	
4. ssBIC	.652	.576	.571	-

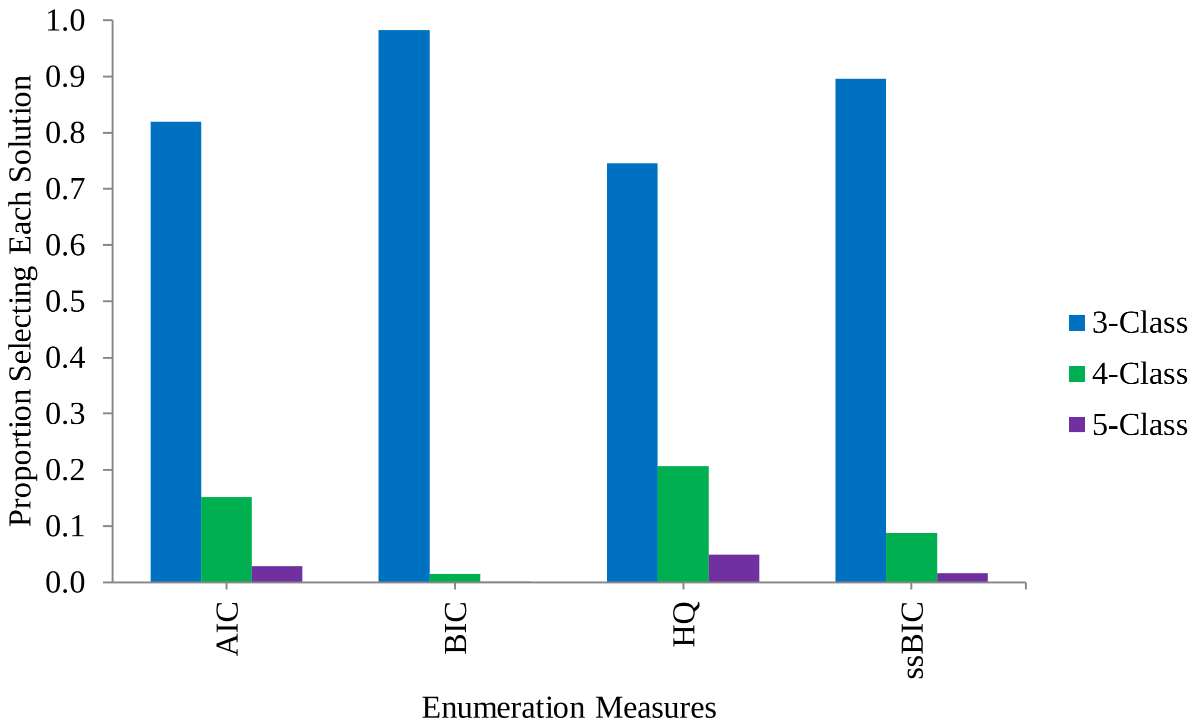

Figure 9. Proportion of 3-, 4-, and 5-class solutions identified as correct by select enumeration measures for sample sizes of 200 and 400, 5 repeated measures, extremely unequal class mixing proportions, and low separation among the classes for the censored normal model

21% followed by AIC at 15%, ssBIC at 9% and BIC at 2%. As with the overall miss classification proportions, AIC, BIC, HQ and ssBIC all tended to under extract rather than over extract classes but HQ and AIC did have the highest over extraction rates at 11% and 8%, respectively. Cohen's Kappas between AIC, BIC, HQ, and ssBIC for the unfavorable conditions are displayed in Table 10. There was a substantial level of agreement between AIC and HQ.

Table 10. Cohen's Kappa between 4 select measures for sample sizes of 200 and 400, 5 repeated measures, extremely unequal class mixing proportions, and low separation among the classes for the censored normal model

	1	2	3	4
1. AIC	-			
2. BIC	.144	-		
3. HQ	.774	.101	-	
4. ssBIC	.671	.265	.470	-

There was a slight level of agreement between BIC and both AIC and HQ and a fair level of agreement between BIC and ssBIC. There was a moderate level of agreement between ssBIC and HQ and a substantial level of agreement between ssBIC and AIC.

The proportion of 3-, 4-, and 5-class solutions selected as correct by AIC, BIC, HQ, and ssBIC in favorable conditions – medium to large sample sizes (800 or 2000), 9 repeated measures, equal and moderately unequal mixing proportions, and high separation among the classes - are displayed in Figure 10. Given the favorable conditions, ssBIC had the highest correct classification rate at 98% followed by BIC at 92%, AIC at 86% and HQ at 86%. AIC and HQ both tended to over extract as opposed to under extract classes given the favorable conditions. BIC never over extracted classes and ssBIC, which almost never made a mistake, had very small rates of both under and over extraction. Cohen's Kappas between AIC, BIC, HQ,

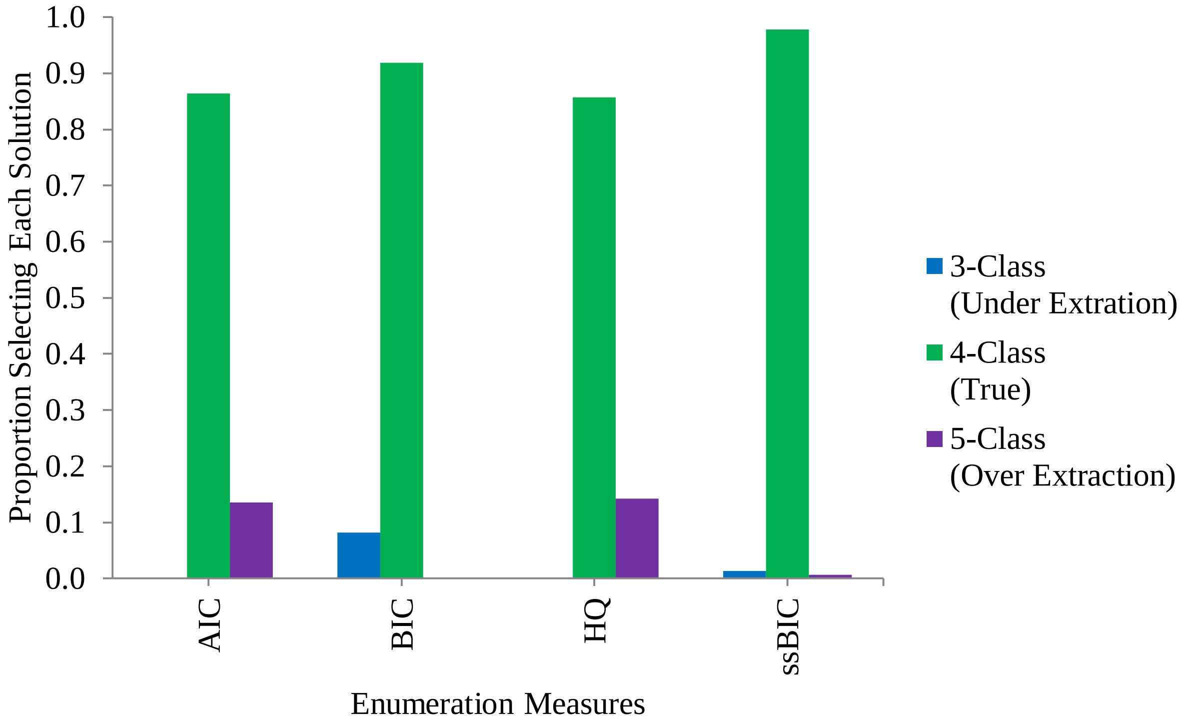

Figure 10. Proportion of 3-, 4-, and 5-class solutions identified as correct by select enumeration measures for sample sizes of 800 and 2000, 9 repeated measures, equal and moderately unequal class mixing proportions, and high separation among the classes for the censored normal model

and ssBIC for the favorable conditions are displayed in Table 11. There was a very high level of agreement between AIC and HQ. There was a very slight level of agreement between BIC and both AIC and HQ and a fair level of agreement between BIC and ssBIC. There was a slight level of agreement between ssBIC and both HQ and AIC.

A boxplot of the correct classification rates of AIC, BIC, HQ, and ssBIC over all conditions is displayed in Figure 11 in order to demonstrate their variability in performance. The boxplot reiterates the findings from examining the 4 enumeration measures in an unfavorable and favorable condition. AIC and HQ performed very similarly and had better average correct classification rates than BIC and ssBIC. They also had the least variability in performance. BIC

Table 11. Cohen's Kappa between 4 select measures for sample sizes of 800 and 2000, 9 repeated measures, equal and moderately unequal class mixing proportions, and high separation among the classes for the censored normal model

	1	2	3	4
1. AIC	-			
2. BIC	.015	-		
3. HQ	.939	.013	-	
4. ssBIC	.101	.264	.087	-

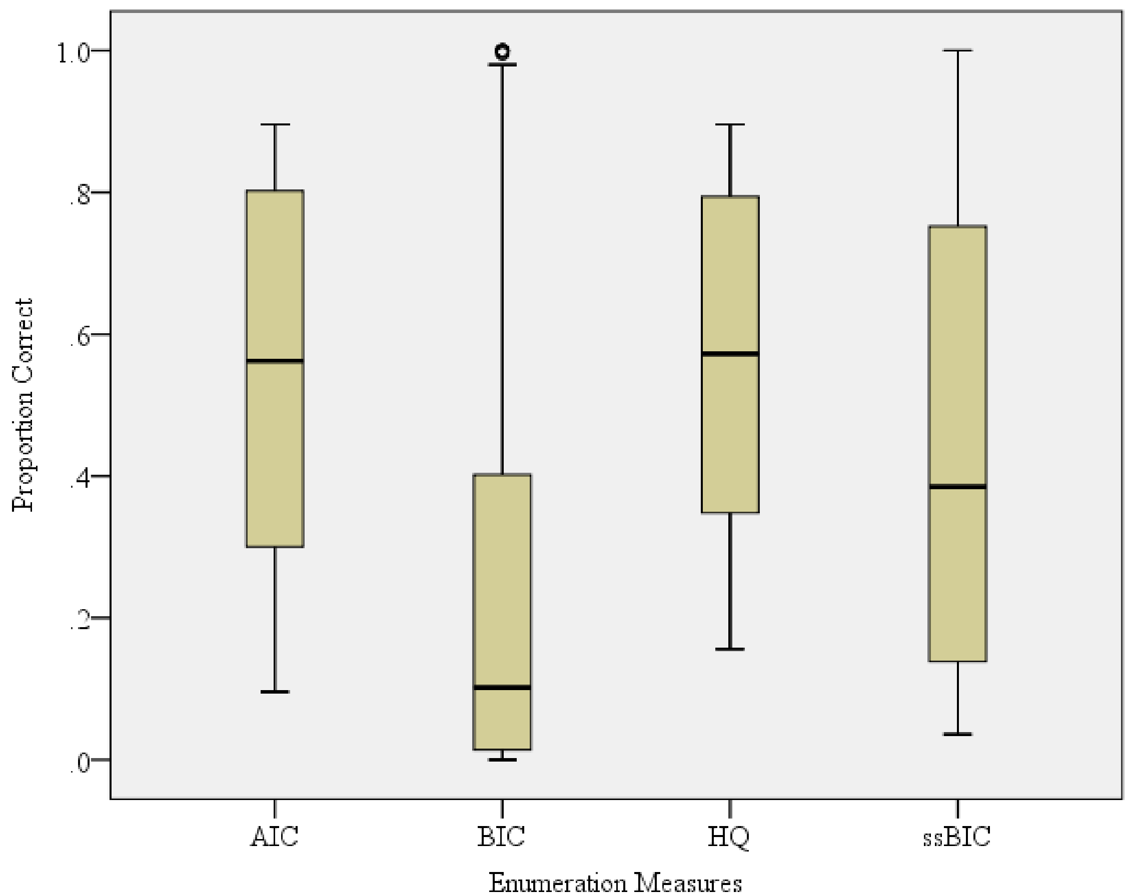

Figure 11. Rate of correct classification across conditions for select enumeration measures for the censored normal model

had the lowest average correct classification rate and was positively skewed with a high correct classification rate in fewer conditions. However, there were certain data conditions where BIC had a higher correct classification rate than AIC and HQ which was demonstrated in Figure 10. ssBIC had a higher average correct classification rate than BIC but not as high as AIC and HQ. ssBIC, like BIC, had more variability in correct classification rate than AIC and HQ. Also, as demonstrated in Figure 10, given certain data condition, ssBIC had a higher correct classification rate than AIC and HQ. The larger variability in the correct classification rates of BIC and ssBIC and the smaller variability in the correct classification rates of AIC and HQ suggested that BIC and ssBIC were more affected by the manipulated factors than AIC and HQ.

4.2 BINARY LOGIT MODEL

4.2.1 Model convergence

Analysis of the binary logit data resulted in a floating point zero divide error in 1,840 (5%) of the replications. When this error occurred, the simulation crashed and no results for a replication were recorded. For these replications, it is unknown which class number solution failed to converge. In order to determine which conditions were more likely to result in a floating point zero divide error, the 25 conditions with an error rate of 10% or more were further examined. The floating point zero divide error occurred more often when the mixing proportions were equal and when there was low separation among the classes. The remaining convergence statistics pertain to the 34,160 replications for which there was no floating point zero divide error and results were recorded.

Across the conditions, the 3-, 4-, and 5-class solutions all failed to converge in 3,480 (10%) of the total 34,160 replications. The maximum number of replications that completely failed to converge in any one cell was 119. The percent of specific convergence patterns across the class number solutions are displayed in Table 12. Failure to converge was most common in

Table 12. Percent of class specific convergence patterns for the binary logit model

Pattern	3-Class	4-Class	5-Class	Frequency	Percent*
1	✓	✓	✓	10882	31.9
2	×	×	×	3480	10.2
3	×	×	✓	376	1.1
4	×	✓	×	1388	4.1
5	×	✓	✓	574	1.7
6	✓	×	×	6221	18.2
7	✓	✓	×	9831	28.8
8	✓	×	✓	1408	4.1

*Percent based on the 34160 replications for which results were recorded.
✓ = Converged; × = Did not converge

the 5-class solution at 61% of the recorded replications. For example, pattern 1 demonstrates that the 3-, 4-, and 5-class solution all converged in approximately 32% of all replications while pattern 7 demonstrates that the 3- and 4-class solutions converged but the 5-class solution failed to converge in approximately 29% of all replications. When the 3-, 4-, and 5-class solutions all failed to converged, pattern 2, the replications (3480) were excluded from the analysis. The 4-class solution failed to converge in 34% of the recorded replications and the 3-class solution failed to converge in 17% of the recorded replications. An increase in sample size, increase in repeated measures, a decrease in percent missing, and a decrease in separation among the classes all resulted in a smaller number of replications that failed to converge for the 3-, 4-, and 5-class solutions. When the mixing proportions went from equal to moderately unequal, there was an increase in the number of replications that failed to converge for the 3-class solution but a decrease in the number of replications that failed to converge for the 4- and 5-class solutions. However, when the mixing proportions went from moderately unequal to extremely unequal, there was a decrease in the number of replications that failed to converge for the 3-class solution but an increase in the number of replications that failed to converge for the 4- and 5-class solutions. Overall, non-convergence was most common when the number of repeated measures and sample size were low, there was 20% missing data, and the separation among the classes was high.

4.2.2 CART analysis results

CART analysis was performed on correct classification rate (correct: selected the 4-class solution, incorrect: did not select the 4-class solution) by sample size, the number of repeated measures, class mixing proportions, percent missing, separation among the classes, and enumeration measures. The overall correct classification of the CART analysis model was 81%. The model was able to correctly classify 91% of the incorrect outcomes while only correctly classifying 55% of the correct outcomes.

At the level of the root node, which contained 705,640 observations, the rate of correctly selecting the 4-class solution was 29% averaged across condition and enumeration measures. As with the CART for the censored normal model, at level 1, the root node was split by the performance of the enumeration measures into two child nodes which will be identified as the "good" performing and "bad" performing groups for ease of identification. The enumeration measures by group are,

"good" { AIC, BIC, CAIC, HQ, ssBIC, ssCAIC, ssHQ, B_{10}, BIC2, CAIC2, HQ2, ssBIC2, ssCAIC2, and ssHQ2} and

“bad” {VLMR, aVLMR, sE, NEC, CLC, ICL-BIC, ssICL-BIC, ICL-BIC2, and ssICL-BIC2}.

The enumeration measures in the good performing group consisted of all the information criteria. The enumeration measures in the bad performing group consisted of the likelihood ratio test derivatives and the entropy based statistics. Averaged over all of the enumeration measures in each group, the correct classification rate was 39% in the good performing group and 11% in the bad performing group. The measures’ rates of selecting the correct 4-class solution ranged from 27% to 52% in the good performing group and 8% to 19% in the bad performing group. At level 2, separation among the classes and enumeration measures where the most important predictors. At level 3, sample size, percent missing, the number of repeated measures, and enumeration measures were the most important predictors. At level 4, the number of repeated measures, class mixing proportions, separation among the classes, and enumeration measures were the most important predictors. At level 5, the lowest level of the CART, sample size, the number of repeated measures, class mixing proportions, separation among the classes, and enumeration measures were the most important predictors. Figure 12 displays the CART diagram for predicting the correct selection of the true 4-class solution for the censored normal model.

There were 16 terminal nodes in the good performing group which will each be discussed in turn. The factors and factor levels that separated these 16 terminal nodes are summarized in Table 13. In Node 31, BIC, CAIC, ssCAIC,B_{10}, BIC2, CAIC2, ssBIC2, and ssCAIC2 performed similarly with an average correct classification rate of 16% when there was high separation among the classes, the sample size was small (200 or 400), and the number of repeated measures was 5. Given the same conditions in Node 32, AIC, HQ, ssBIC, ssHQ, HQ2, and ssHQ2 had an average correct classification rate of 32%.

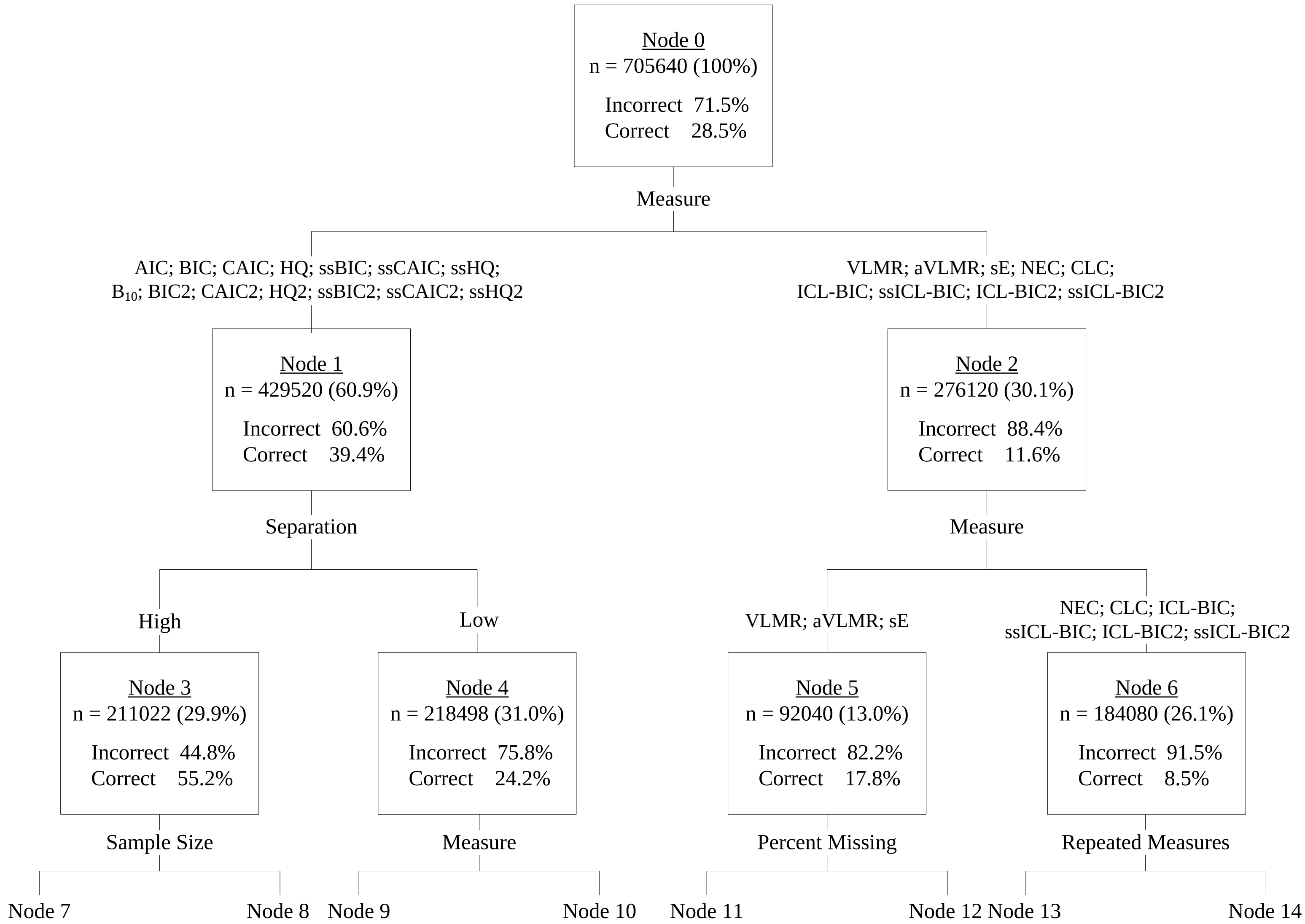

Figure 12. CART predicting correct selection of the true 4-class solution for the censored normal model

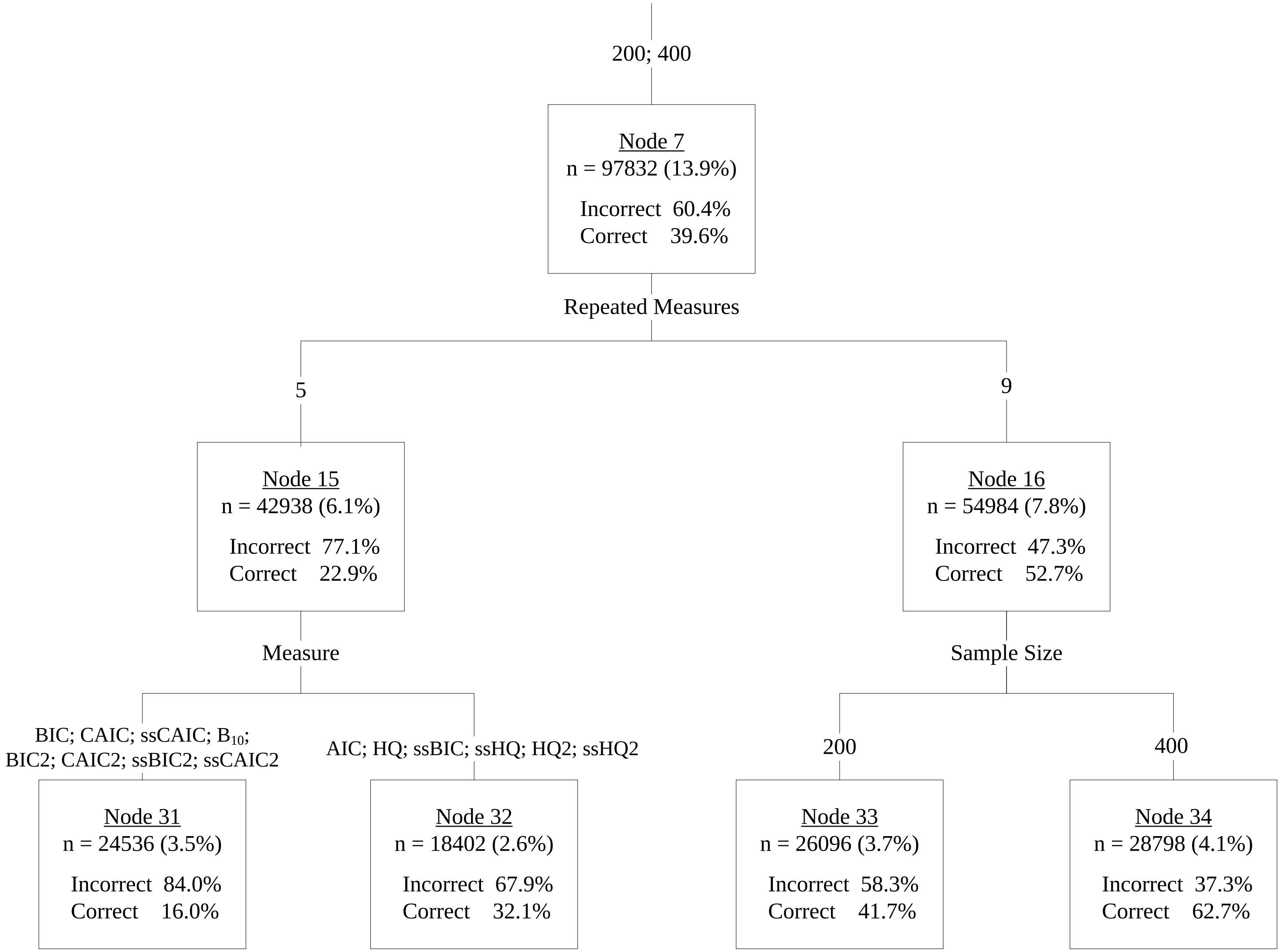

Figure 12 (continued).

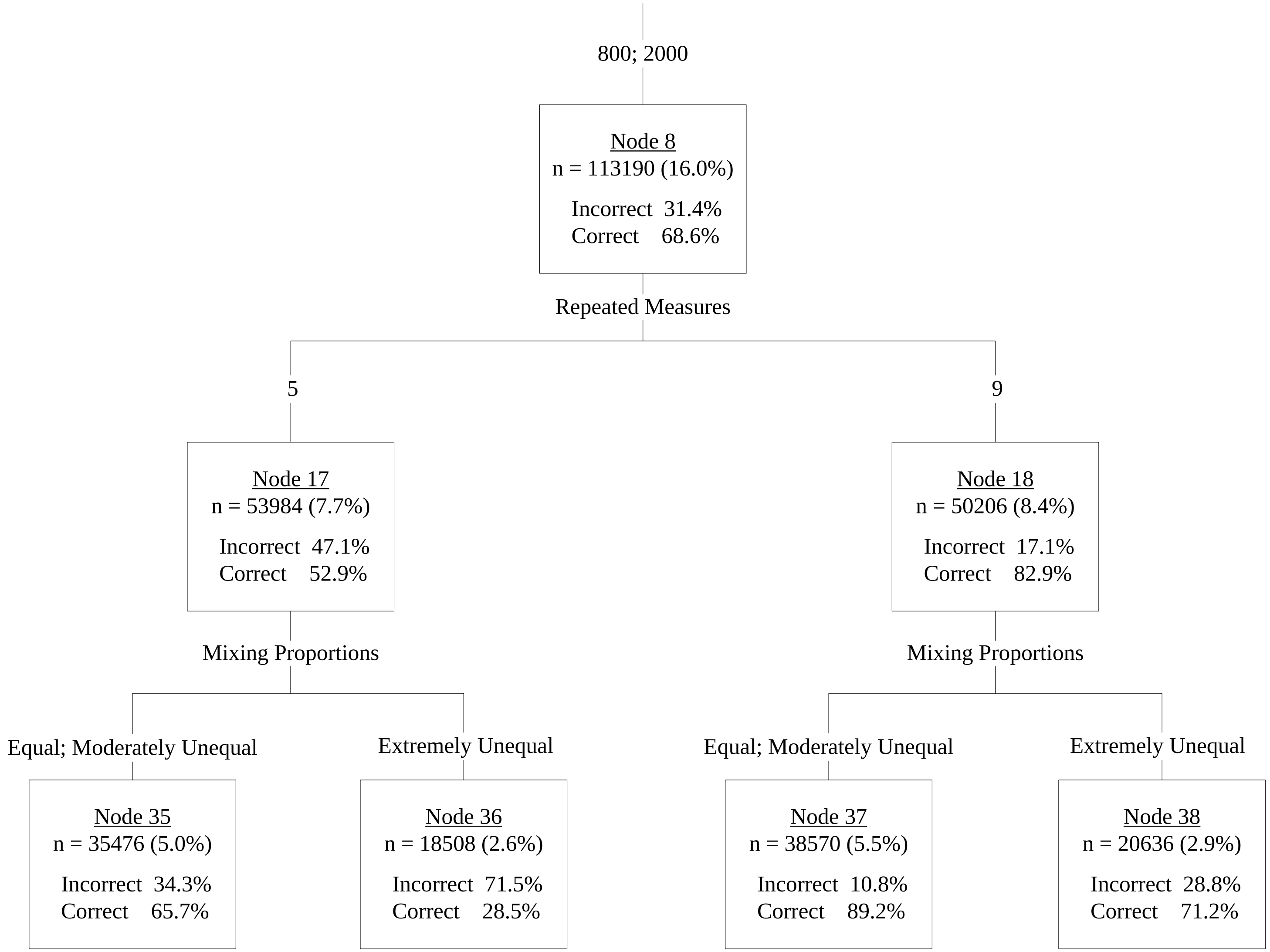

Figure 12 (continued).

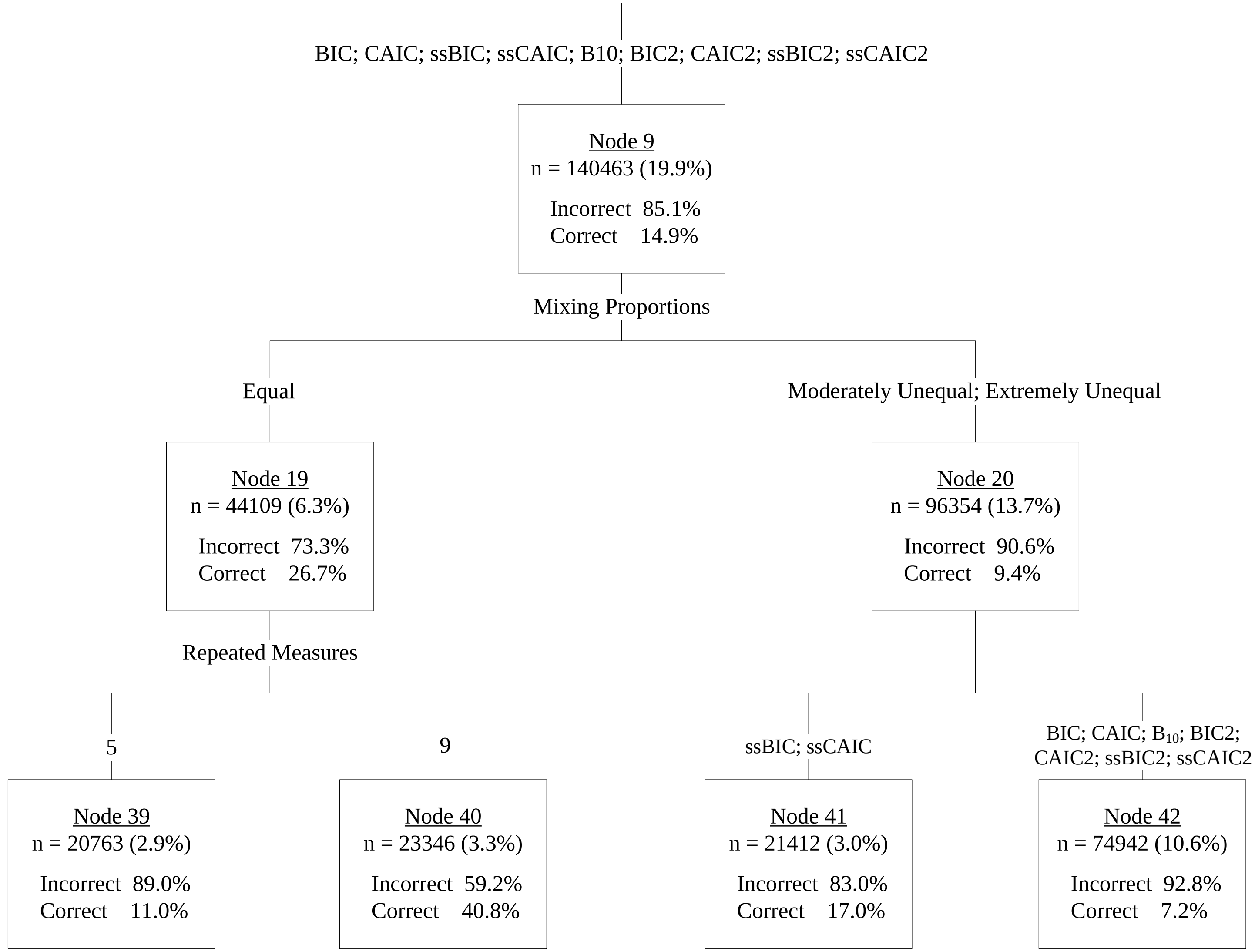

Figure 12 (continued).

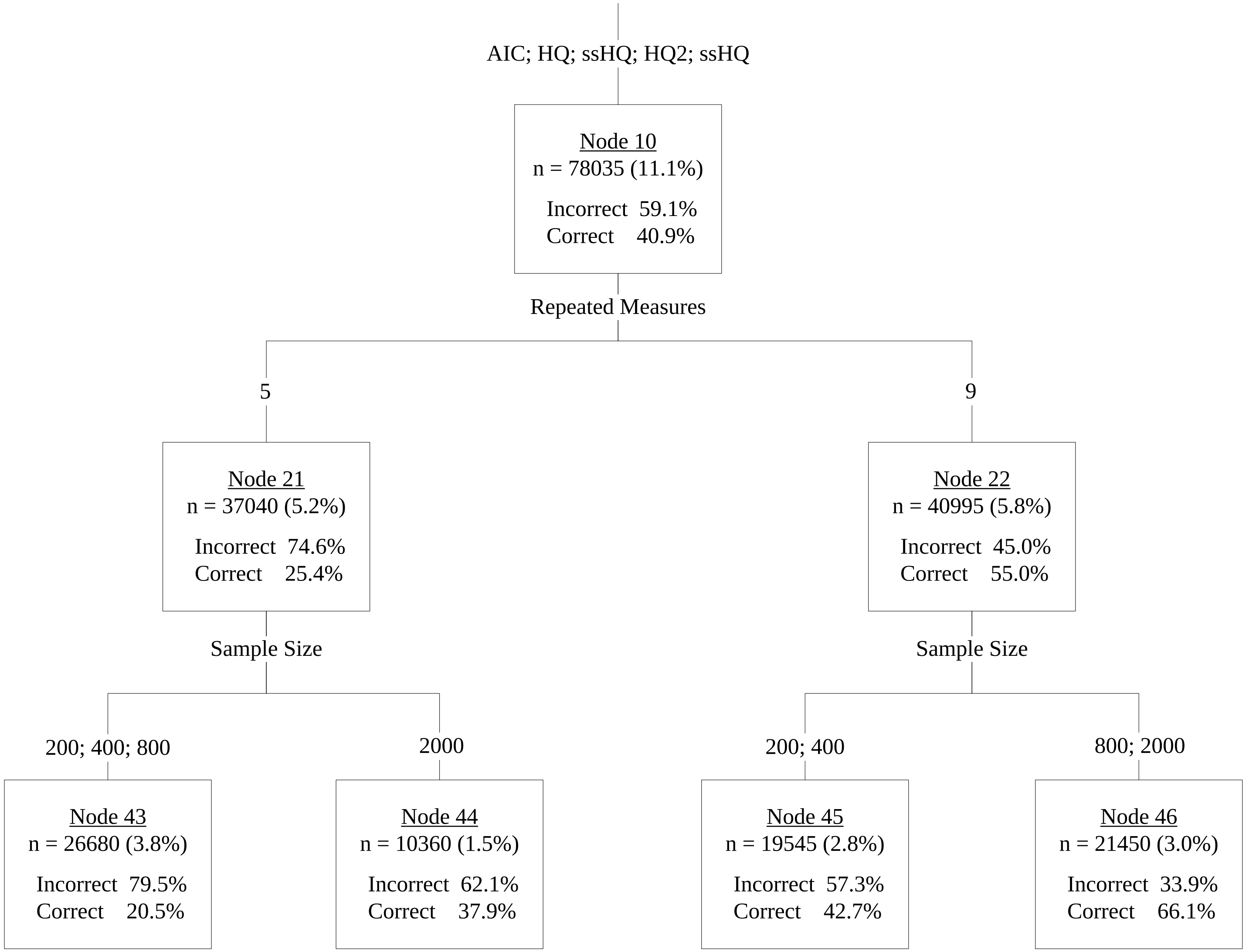

Figure 12 (continued).

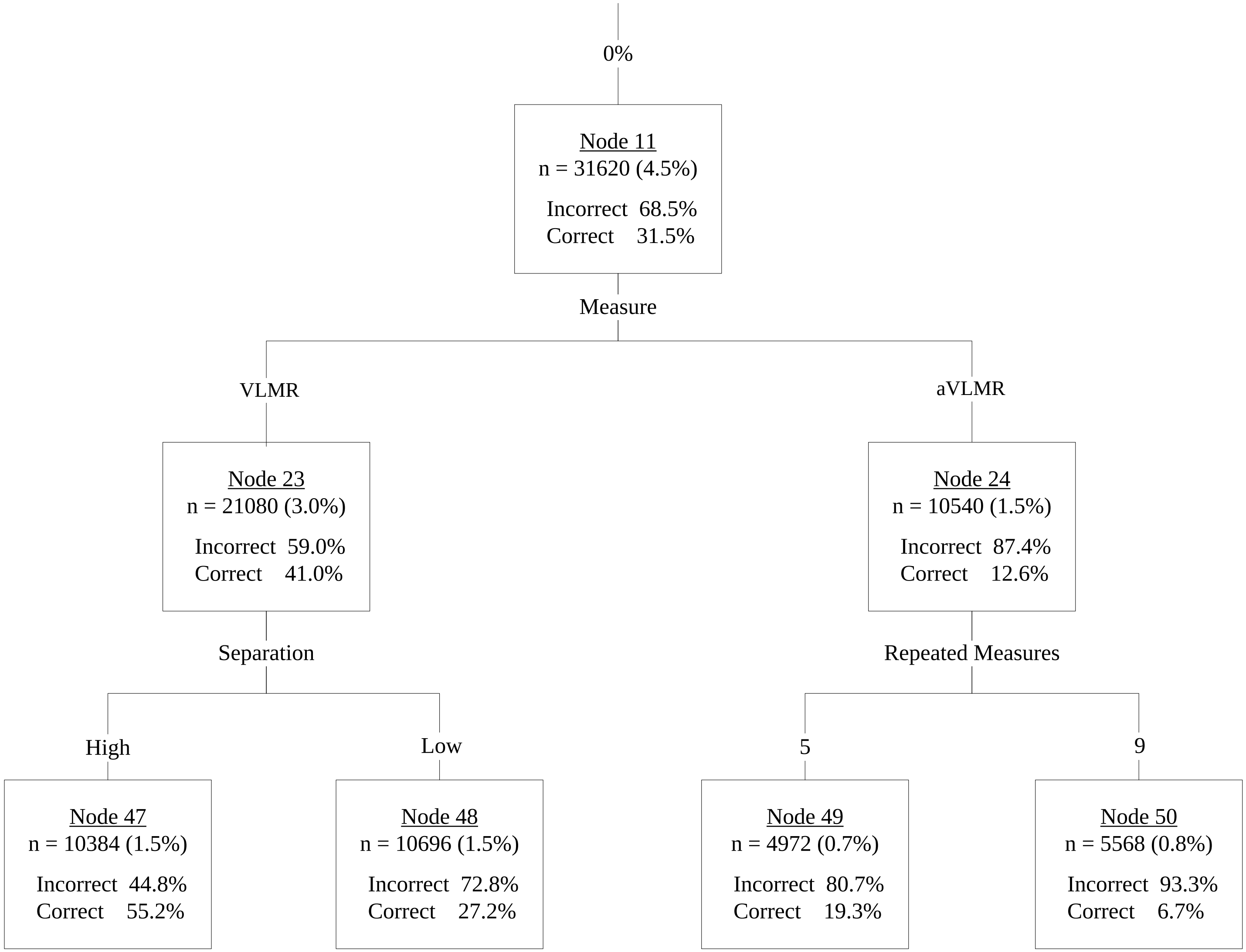

Figure 12 (continued).

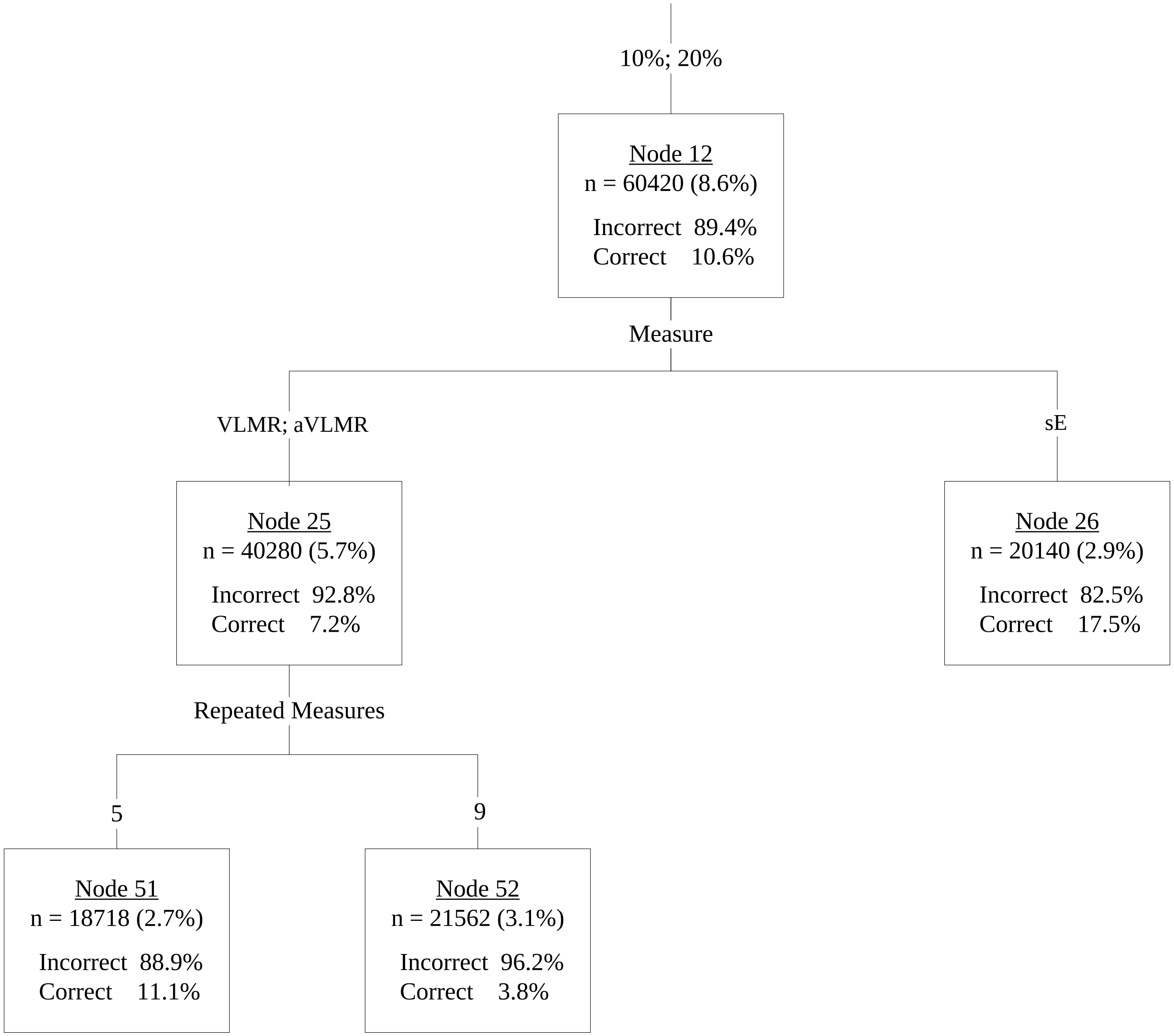

Figure 12 (continued).

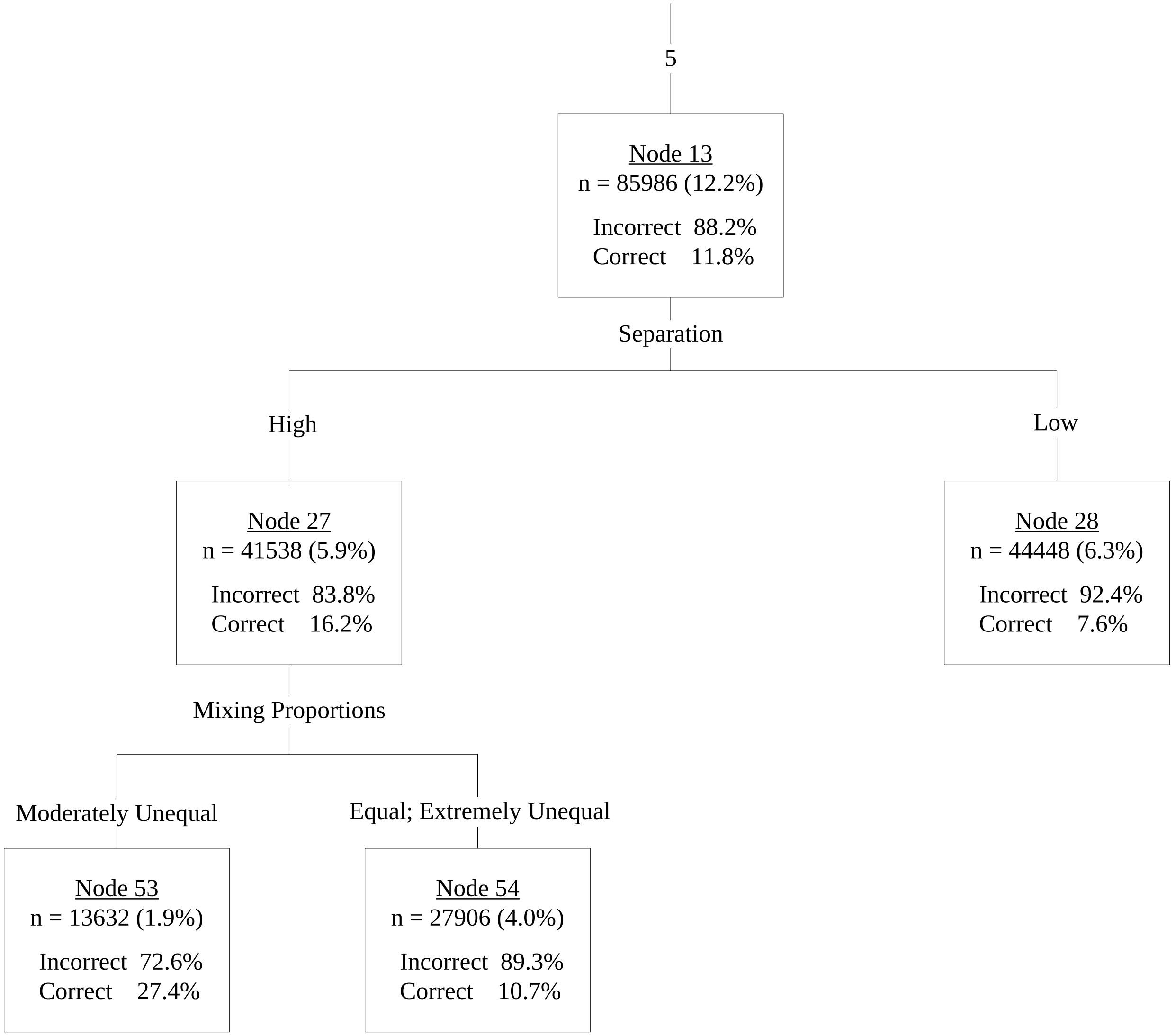

Figure 12 (continued).

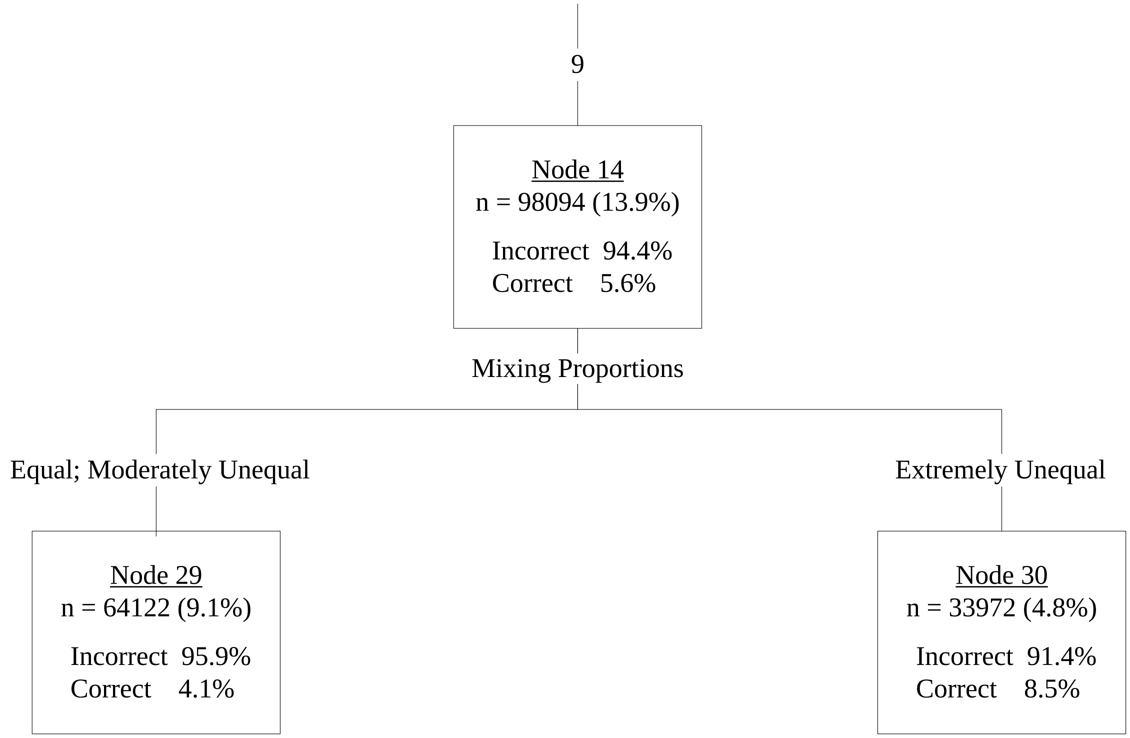

Figure 12 (continued).

Table 13. Terminal node summary of the good performing enumeration measures for the binary logit model

Node	Sample Size	No. Obs.	Mixing Proportions	Class Separation	Measure	Percent Correct	Percent of Total
31	200; 400	5		High	BIC; CAIC; ssCAIC; B_{10}; BIC2; CAIC2; ssBIC2; ssCAIC2	16.0	3.5
32	200; 400	5		High	AIC; HQ; ssBIC; ssHQ; HQ2; ssHQ2	32.1	2.6
33	200	9		High	AIC; BIC; CAIC; HQ; ssBIC; ssCAIC; ssHQ; B_{10}; BIC2; CAIC2; HQ2; ssBIC2; ssCAIC2; ssHQ2	41.7	3.7
34	400	9		High	AIC; BIC; CAIC; HQ; ssBIC; ssCAIC; ssHQ; B_{10}; BIC2; CAIC2; HQ2; ssBIC2; ssCAIC2; ssHQ2	62.7	4.1
35	800; 2000	5	Equal; Moderately Unequal	High	AIC; BIC; CAIC; HQ; ssBIC; ssCAIC; ssHQ; B_{10}; BIC2; CAIC2; HQ2; ssBIC2; ssCAIC2; ssHQ2	65.7	5.0
36	800; 2000	5	Extremely Unequal	High	AIC; BIC; CAIC; HQ; ssBIC; ssCAIC; ssHQ; B_{10}; BIC2; CAIC2; HQ2; ssBIC2; ssCAIC2; ssHQ2	28.5	2.6
37	800; 2000	9	Equal; Moderately Unequal	High	AIC; BIC; CAIC; HQ; ssBIC; ssCAIC; ssHQ; B_{10}; BIC2; CAIC2; HQ2; ssBIC2; ssCAIC2; ssHQ2	89.2	5.5
38	800; 2000	9	Extremely Unequal	High	AIC; BIC; CAIC; HQ; ssBIC; ssCAIC; ssHQ; B_{10}; BIC2; CAIC2; HQ2; ssBIC2; ssCAIC2; ssHQ2	71.2	2.9
39		5	Equal	Low	BIC; CAIC; ssBIC; ssCAIC; B_{10}; BIC2; CAIC2; ssBIC2; ssCAIC2	11.0	2.9
40		9	Equal	Low	BIC; CAIC; ssBIC; ssCAIC; B_{10}; BIC2; CAIC2; ssBIC2; ssCAIC2	40.8	3.3
41			Moderately Unequal; Extremely Unequal	Low	ssBIC; ssCAIC	17.0	3.0
42			Moderately Unequal; Extremely Unequal	Low	BIC; CAIC; B_{10}; BIC2; CAIC2; ssBIC2; ssCAIC2	7.2	10.6
43	200; 400; 800	5		Low	AIC; HQ; ssHQ; HQ2; ssHQ2	20.5	3.8
44	2000	5		Low	AIC; HQ; ssHQ; HQ2; ssHQ2	37.9	1.5
45	200; 400	9		Low	AIC; HQ; ssHQ; HQ2; ssHQ2	42.7	2.8
46	800; 2000	9		Low	AIC; HQ; ssHQ; HQ2; ssHQ2	66.1	3.0

In Node 33, all of the enumeration measures in the good performing group had an average correct classification rate of 42% when the separation among classes was high, the number of repeated measures was 9 and the sample size was small (200). In Node 34, all of the enumeration measures in the good performing group had an average correct classification rate of 63% given a small sample size (400) and the same high separation and number of repeated measures conditions as Node 33.

In Node 35, all of the enumeration measures in the good performing group had an average correct classification rate of 66% when there was high separation among the classes, the sample size was medium to large (800 or 2000), the number of repeated measures was 5, and the class mixing proportions were either equal or moderately unequal. In Node 36, where the class mixing proportions were extremely unequal but the remaining conditions were the same as Node 35, the average correct classification rate was 29%.

In Node 37, all of the enumeration measures in the good performing group had an average correct classification rate of 89% when there was high separation among the classes, the number of repeated measures was 9, and the class mixing proportions were either equal or moderately unequal. There was an increase in correct classification rate of approximately 23% between Node 35 and 37 when the number of repeated measures increased from 5 to 9 given the same levels of enumeration measures, separation among the classes, and sample size. In Node 38, where the class mixing proportions were extremely unequal but the remaining conditions were the same as Node 37, the average correct classification rate was 71%.

In Node 39, BIC, CAIC, ssBIC, ssCAIC, B_{10}, BIC2, CAIC2, ssBIC2, and ssCAIC2 performed similarly with an average correct classification rate of 11% when there was low separation among the classes, the class mixing proportions were equal, and there were 5 repeated measures. In Node 40, the rate of correct classification increased to 41% for the same 9 measures in the same low separation and class mixing proportion conditions when the number of repeated measures increased to 9.

In Node 41, ssBIC and ssCAIC performed similarly with an average correct classification rate of 17% when there was low separation among the classes and the class mixing proportions were either moderately unequal or extremely unequal. Given the same class separation and class mixing proportions conditions in Node 42 as in Node 41, the average correct classification rate of BIC, CIAC, B_{10}, BIC2, CAIC2, ssBIC2, and ssCAIC2 was 7%.

In Node 43, AIC, HQ, ssHQ, HQ2, and ssHQ2 performed similarly with an average correct classification rate of 21% when there was low separation among the classes, the number of repeated measures was 5, and the sample size was small to medium (200, 400, or 800). The same 5 measures had an average correct classification rate of 38% in Node 44 when the sample size was large (2000) and given the same separation among the classes and number of repeated measures conditions as Node 43. In Node 45, AIC, HQ, ssHQ, HQ2, and ssHQ2 had an average correct classification rate of 43% when there was low separation among the classes, the number of repeated measures was 9 and the sample size was small (200 or 400). Node 46 had medium to large sample sizes (800 or 2000), which were larger than Node 45, but the two nodes had the same number of repeated measures (9) and enumeration measures (AIC, HQ, ssHQ, HQ2, ssHQ2). In Node 46, the average correct classification rate was 66%, an increase of approximately 23% over Node 45.

There were 12 terminal nodes in the bad performing group. The factors and factor levels that separated these 12 terminal nodes are summarized in Table 14. In Node 47, VLMR and aVLMR performed similarly with a correct classification rate of 55% when there was no missing

Table 14. Terminal node summary of the bad performing enumeration measures for the binary logit model

Node	No. Obs.	Mixing Proportions	Percent Missing	Class Separation	Measure	Percent Correct	Percent of Total
47			0%	High	VLMR; AVLMR	55.2	1.5
48			0%	Low	VLMR; AVLMR	27.2	1.5
49	5		0%		sE	19.3	0.7
50	9		0%		sE	6.7	0.8
51	5		10%; 20%		VLMR; AVLMR	11.1	2.7
52	9		10%; 20%		VLMR; AVLMR	3.8	3.1
26			10%; 20%		sE	17.5	2.9
53	5	Moderately Unequal		High	NEC; CLC; ICLBIC; ssICLBIC; ICLBIC2; ssICLBIC2	27.4	1.9
54	5	Equal; Extremely Unequal		High	NEC; CLC; ICLBIC; ssICLBIC; ICLBIC2; ssICLBIC2	10.7	4.0
28	5			Low	NEC; CLC; ICLBIC; ssICLBIC; ICLBIC2; ssICLBIC2	7.6	6.3
29	9	Equal; Moderately Unequal			NEC; CLC; ICLBIC; ssICLBIC; ICLBIC2; ssICLBIC2	4.1	9.1
30	9	Extremely Unequal			NEC; CLC; ICLBIC; ssICLBIC; ICLBIC2; ssICLBIC2	8.5	4.8

and there was high separation among the classes. When there was no missing and low separation among the classes, VLMR and aVLMR had an average correct classification rate of 27%, Node 48.

In Node 49, sE had a correct classification rate of 19% when there was no missing and the number of repeated measures was 5. The correct classification rate of sE decreased to 7% when there was no missing and number of repeated measures increased to 9, Node 50.

In Node 51, VLMR and aVLMR performed similarly with an average correct classification rate of 11% when there was 10% or 20% missing and the number of repeated measures was 5. In Node 52, where there was 10% or 20% missing and the number of repeated measures increased to 9, the average correct classification rate of VLMR and aVLMR was 4%, a decrease of 7% relative to Node 51. In Node 26, the correct classification rate of sE was 18% when there was 10% or 20% missing.

In Node 53, NEC, CLC, ICL-BIC, ssICL-BIC, ICL-BIC2, and ssICL-BIC2 performed similarly with an average correct classification rate of 27% when the number of repeated measures was 5, there was high separation among the classes, and the class mixing proportions were moderately unequal. In Node, 54, the same 6 enumeration measures had an average correct classification rate of 11% when the number of repeated measures was 5, there was high separation among the classes and the class mixing proportions were equal or extremely unequal. In Node 28, NEC, CLC, ICL-BIC, ssICL-BIC, ICL-BIC2, and ssICL-BIC2 had an average correct classification rate of 8% when the number of repeated measures was 5 and there was low separation among the classes.

In Node 29, NEC, CLC, ICL-BIC, ssICL-BIC, ICL-BIC2, and ssICL-BIC2 performed similarly with an average correct classification rate of 4% when there were 9 repeated measures and the class mixing proportions were equal or moderately unequal. The average correct classification rate for the same 6 enumeration measures was 9% when there were 9 repeated measures and the class mixing proportions were extremely unequal, Node 30.

4.2.3 Results summary

The 16 terminal nodes in the good performing group and the 12 terminal nodes in the bad performing group, ordered by correct classification rates, are displayed in Figure 13 and Figure 14, respectively. To better address specific themes that emerged in the CART, the nodes were first split into quartiles by performance. The 7 nodes in the first quartile had a correct classification rates of slightly less than 11% or below. The 7 nodes in the second quartile had correct classification rates between 11% and 21%. The 7 nodes in the third quartile had correct classification rates between 27% and 42%. The 7 nodes in the fourth quartile had correct classification rates between 43% and 89%.

The 7 nodes in the fourth quartile were affected by different combinations and levels of all the factors, enumeration measures, sample sizes, class mixing proportions, the number of repeated measures, percent missing, and separation among the classes. AIC, BIC, CAIC, HQ, ssBIC, ssCAIC, ssHQ, B_{10}, BIC2, CAIC2, HQ2, ssBIC2, ssCAIC2, and ssHQ2 performed similarly and were in 4 nodes in the top quartile (37, 38, 35, 34). AIC, HQ, ssHQ, HQ2, and ssHQ2 performed similarly in the only 2 nodes (46, 45) in the fourth quartile to have low separation among the classes. Overall, a node was most likely to be in the fourth quartile when it contained information criteria, especially AIC, HQ, ssHQ, HQ2, and ssHQ2, there was high separation among the classes, it had a medium to large sample size (800 or 2000), and there were 9 repeated measures.

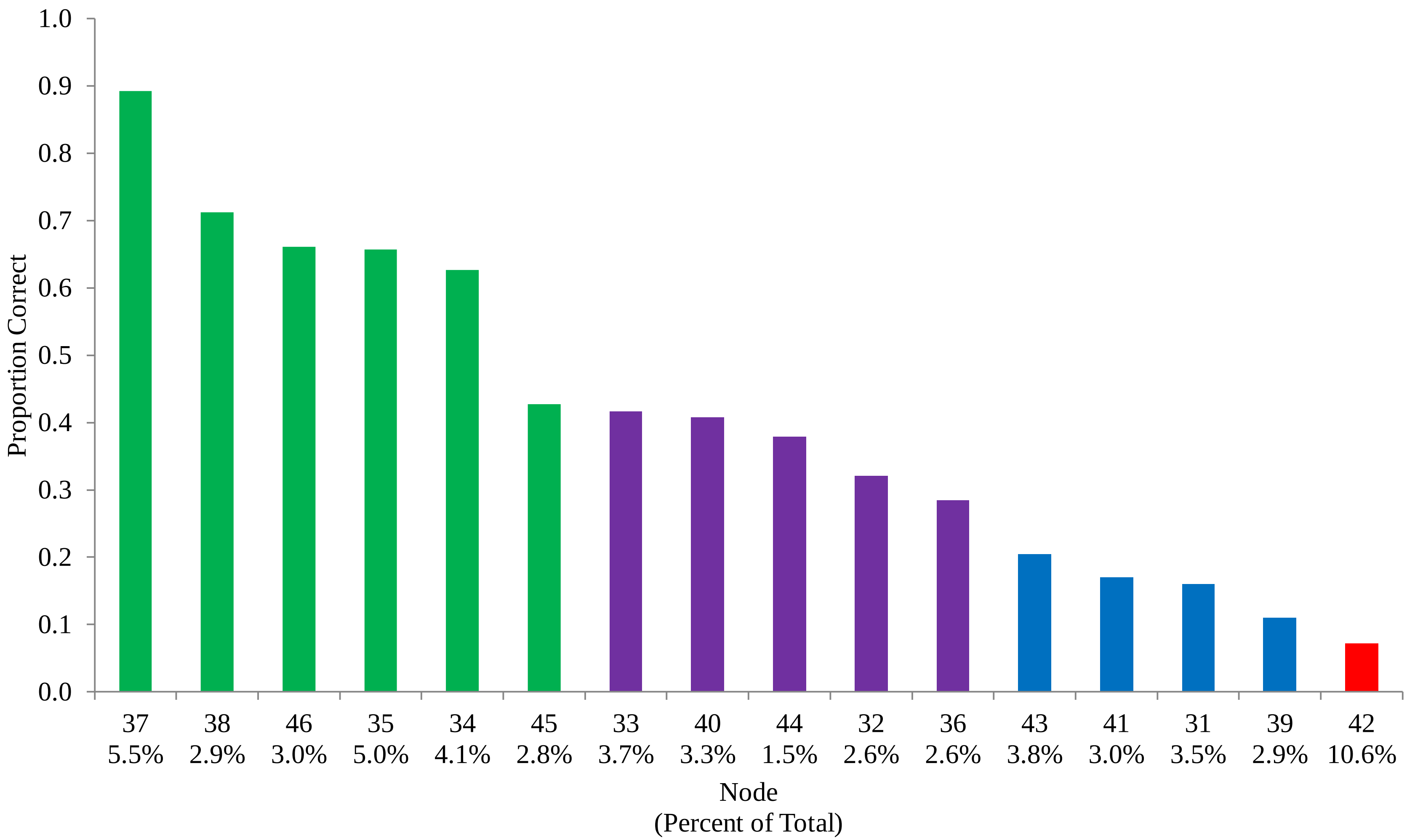

Figure 13. Proportion of times the 4-class solution was identified as correct by good performing enumeration measure nodes for the binary logit model

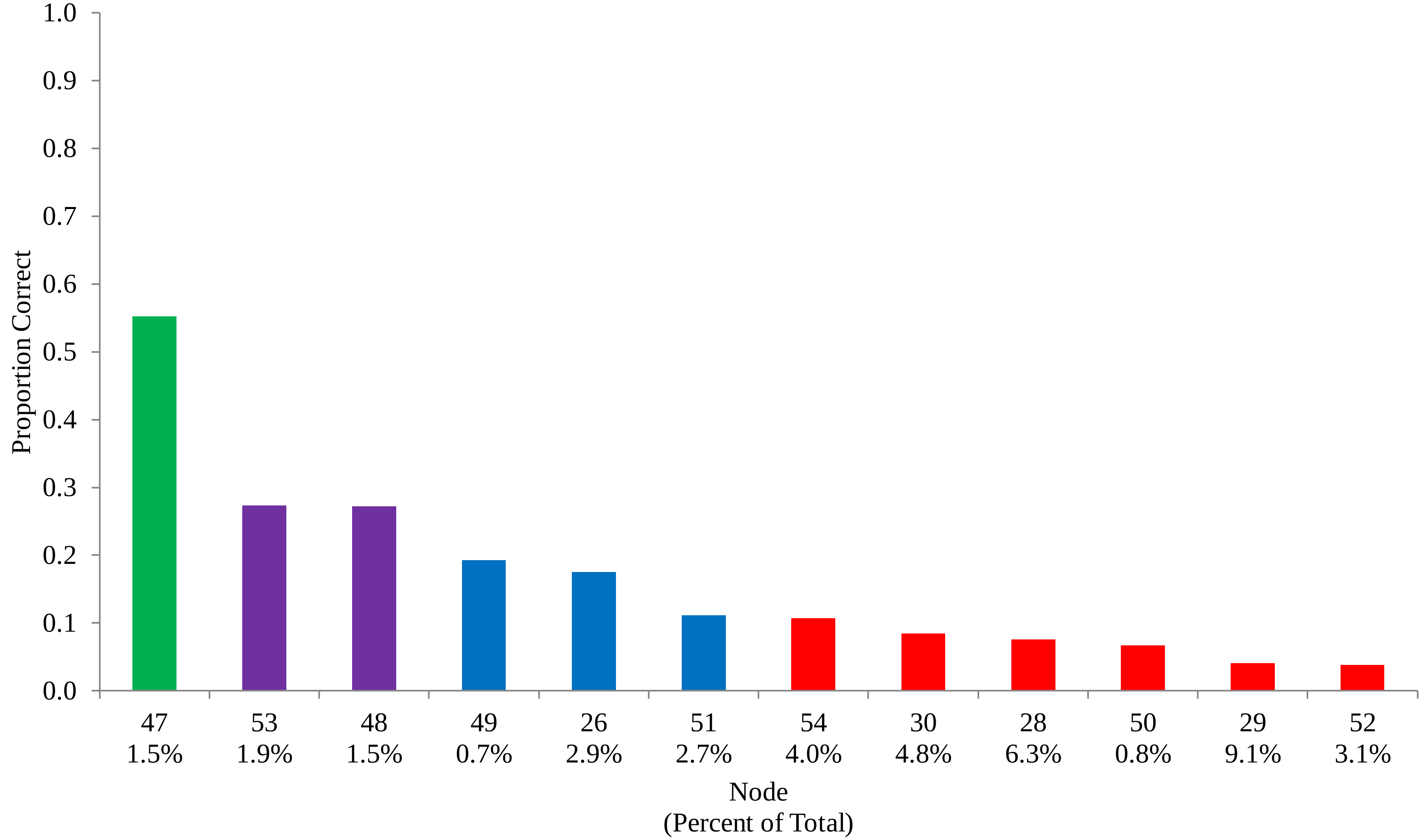

Figure 14. Proportion of times the 4-class solution was identified as correct by bad performing enumeration measure nodes for the binary logit model

The 7 nodes in the third quartile were affected by different combinations and levels of all the factors, enumeration measures, sample sizes, class mixing proportions, the number of repeated measures, percent missing, and separation among the classes. There were not any real clear patterns or combinations of factors and factor levels that emerged in the third quartile. However, the nodes in the third quartile tended to have 5 repeated measures as opposed to 9 repeated measures and 5 of the 7 nodes contained information criteria.

The 7 nodes in the second quartile were affected by different combinations and levels of all the factors, enumeration measures, sample sizes, class mixing proportions, the number of repeated measures, percent missing, and separation among the classes. There were not any real clear patterns or combinations of factors and factor levels that emerged in the second quartile. There were different mixtures of information criteria in 4 of the nodes, sE was by itself in 2 nodes, and VLMR and aVLMR were grouped together in 1 node. The nodes affected by the number of repeated measures in the second quartile all had 5 repeated measure. Also, 3 of the 4 nodes affected by the separation among the classes had low separation.

The 7 nodes in the first quartile were affected by different combinations and levels of enumeration measures, class mixing proportions, the number of repeated measures, percent missing, and separation among the classes. Four of the 7 nodes contained the entropy based statistics NEC, CLC, ICL-BIC, ssICL-BIC, ICL-BIC2, and ssICL-BIC2 and one of the other nodes contained sE. The number of repeated measures affected 6 nodes in the first quartile. Of the 6 effected nodes, 4 had 9 repeated measures.

As with the censored normal model, AIC, BIC, HQ, and ssBIC were further examined for the binary logit model. The proportion of 3-, 4-, and 5-class solutions selected as correct by AIC, BIC, HQ, and ssBIC across all conditions are displayed in Figure 15. Overall, the

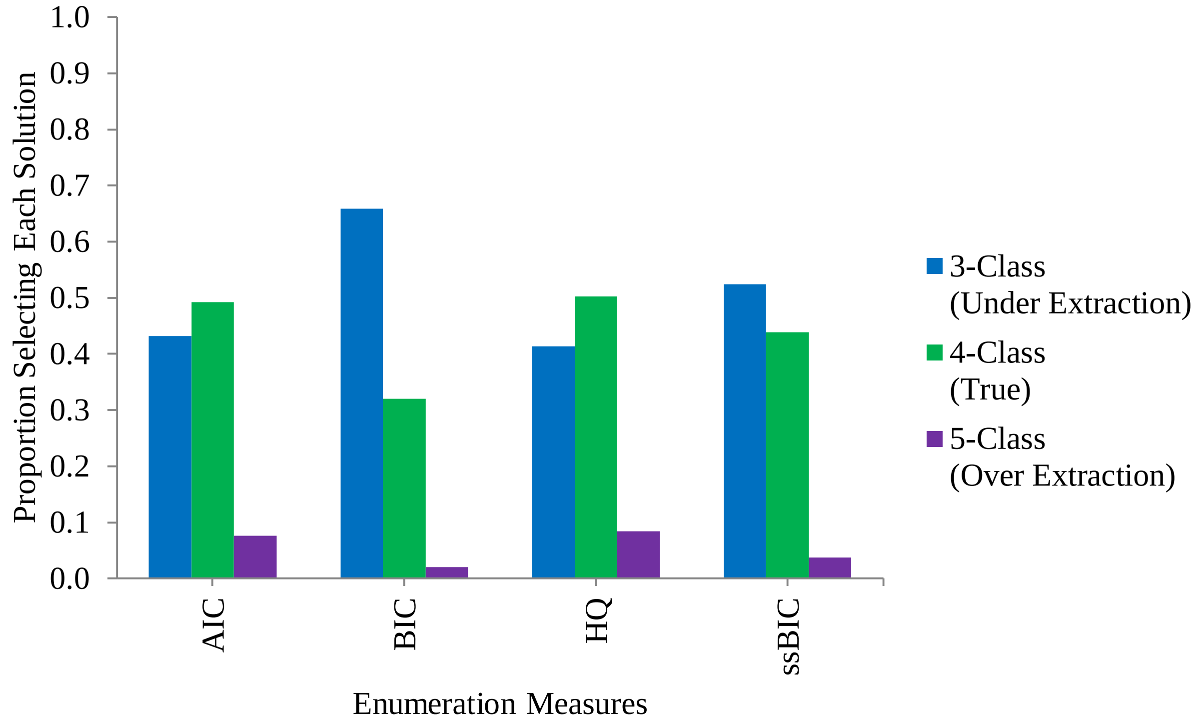

Figure 15. Proportion of 3-, 4-, and 5-class solutions identified as correct by select enumeration measures for the binary logit model

performances of AIC and HQ were virtually identical with HQ having a 50% correct classification rate and AIC having a 49% correct classification rate. The correct classification rate of ssBIC was slightly less than AIC and HQ at 44% while BIC only had a correct classification rate of 32%. All 4 enumeration measures tended to under extract rather than over extract classes but AIC and HQ had the highest rates of over extraction at 7.6% and 8.4%. Cohen's Kappas between AIC, BIC, HQ, and ssBIC over all conditions are displayed in Table 15. There was a very high level of agreement between AIC and HQ. There was a moderate level of agreement between BIC and both AIC and HQ and a substantial level of agreement between BIC and ssBIC. There was a substantial level of agreement between ssBIC and both AIC and HQ.

Table 15. Cohen's Kappa between 4 select measures over all conditions for the binary logit model

	1	2	3	4
1. AIC	-			
2. BIC	.562	-		
3. HQ	.957	.535	-	
4. ssBIC	.791	.735	.755	-

The proportion of 3-, 4-, and 5-class solutions selected as correct by AIC, BIC, HQ, and ssBIC in unfavorable conditions – small sample sizes (200 or 400), 5 repeated measures, extremely unequal mixing proportions, and low separation among the classes - are displayed in Figure 16. All of the enumeration measures had very poor correct classification rates given the unfavorable conditions. HQ had the highest correct classification rate at 10% followed by AIC at 8%, ssBIC at 6% and BIC at 5%. Under extraction was practically always the error made by all 4 measures. Cohen's Kappas between AIC, BIC, HQ, and ssBIC for the unfavorable conditions are displayed in Table 16. There was a very high level of agreement between AIC and both HQ and ssBIC. There was also a very high level of agreement between BIC and ssBIC. There was a substantial level of agreement between BIC and both AIC and HQ and also HQ and ssBIC. The proportions of 3-, 4-, and 5-class solutions selected as correct by AIC, BIC, HQ, and ssBIC in favorable conditions – medium to large sample sizes (800 or 2000), 9 repeated measures, equal and moderately unequal mixing proportions, and high separation among the classes - are displayed in Figure 17. The correct classification rates and extraction errors of AIC and HQ were almost indistinguishable. Similarly, the correct classification rates and extraction errors of BIC and ssBIC were practically the same. BIC and ssBIC had correct classification

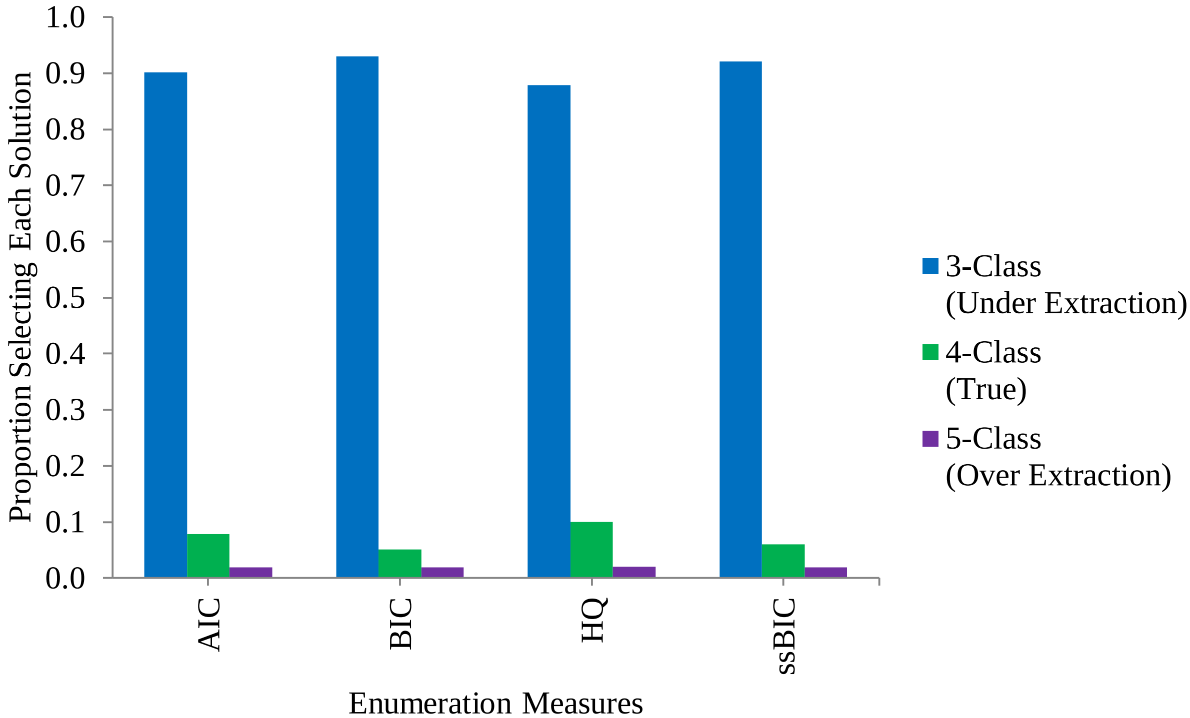

Figure 16. Proportion of 3-, 4-, and 5-class solutions identified as correct by select enumeration measures for sample sizes of 200 and 400, 5 repeated measures, extremely unequal class mixing proportions, and low separation among the classes for the binary logit model

Table 16. Cohen's Kappa between 4 select measures for sample sizes of 200 and 400, 5 repeated measures, extremely unequal class mixing proportions, and low separation among the classes for the binary logit model

	1	2	3	4
1. AIC	-			
2. BIC	.545	-		
3. HQ	.980	.534	-	
4. ssBIC	.598	.903	.586	-

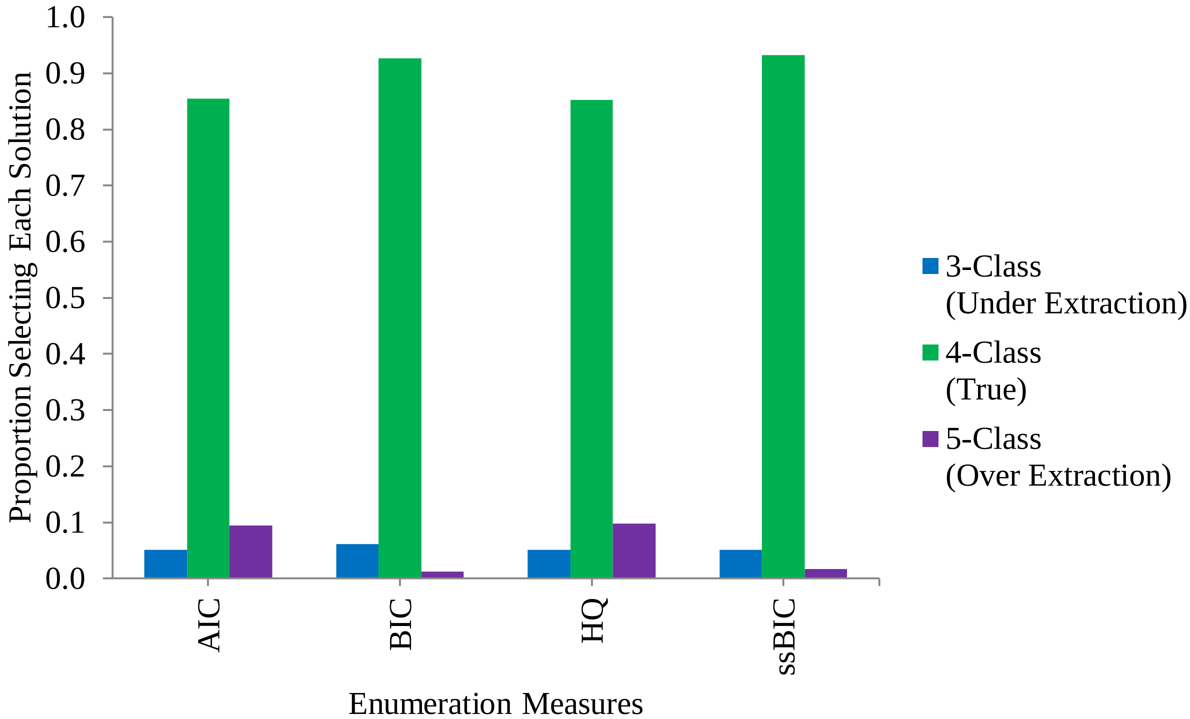

Figure 17. Proportion of 3-, 4-, and 5-class solutions identified as correct by select enumeration measures for sample sizes of 800 and 2000, 9 repeated measures, equal and moderately unequal class mixing proportions, and high separation among the classes for the binary logit model

rates of approximately 93% while AIC and HQ had correct classification rates of approximately 85%. BIC and ssBIC both made over extraction errors less than 2% of the time while AIC and HQ made over extraction errors about 9% of the time. AIC, HQ and ssBIC made under extraction errors at a rate of 5% while BIC made under extraction errors at a rate of 6%. Cohen's Kappas between AIC, BIC, HQ, and ssBIC for the favorable conditions are displayed in Table 17. There was a very high level of agreement between AIC and HQ and between BIC and ssBIC. There were moderate levels of agreement between the remaining pairs of measures.

A boxplot of the correct classification rate of AIC, BIC, HQ, and ssBIC over all conditions is displayed in Figure 18 in order to demonstrate their variability in performance. The boxplot for the binary logit model was very similar to that of the boxplot for the censored normal

Table 17. Cohen's Kappa between 4 select measures for sample sizes of 800 and 2000, 9 repeated measures, equal and moderately unequal class mixing proportions, and high separation among the classes for the binary logit model

	1	2	3	4
1. AIC	-			
2. BIC	.767	-		
3. HQ	.863	.638	-	
4. ssBIC	.854	.910	.721	-

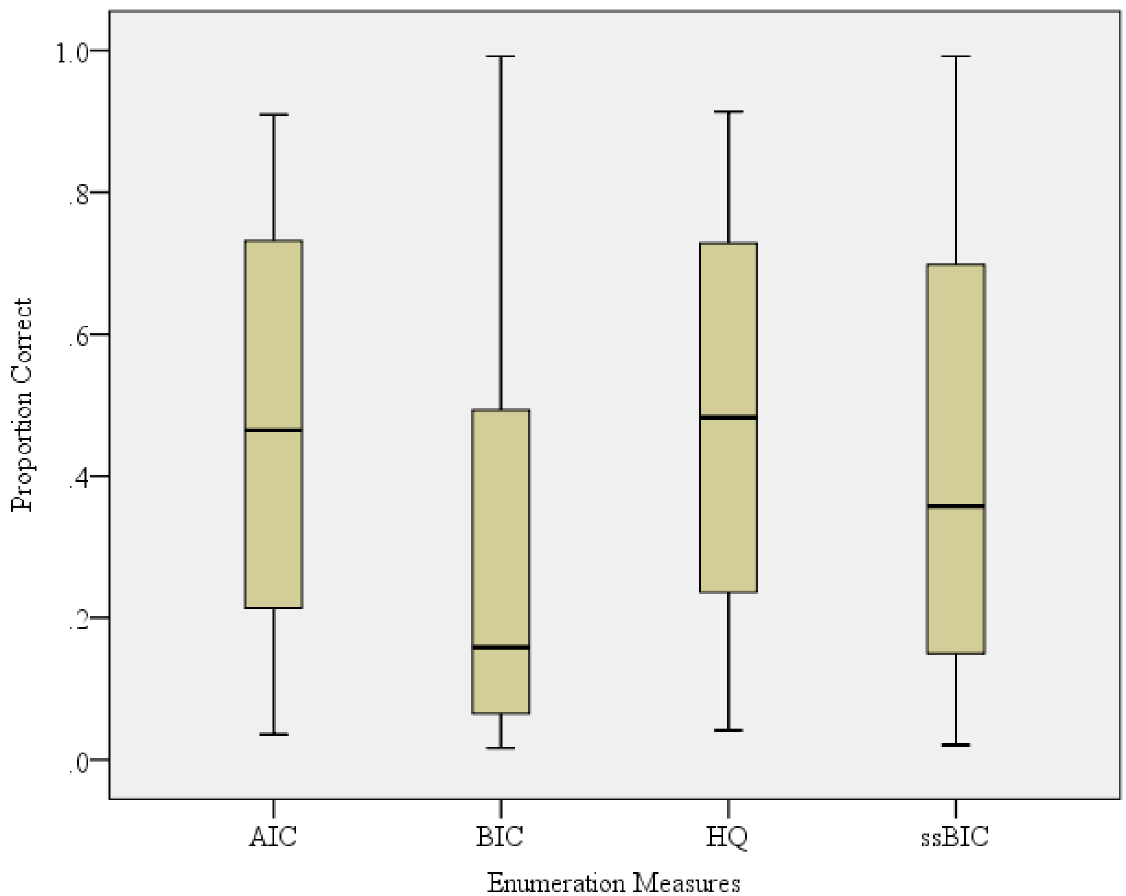

Figure 18. Rate of correct classification across conditions for select enumeration measures for the binary logit model

model. AIC and HQ had higher average correct classification rates than BIC and ssBIC and less variability across the conditions. The performance of BIC was positively skewed across the conditions for the binary logit model which was similar to censored normal model. ssBIC had a higher average correct classification rate than BIC but they both covered about the same range of correct classification rates across all of the conditions. Both BIC and ssBIC had higher correct classification rates than AIC and HQ for certain conditions.

5 DISCUSSION

The purpose of the current study was to investigate the performance of a large set of enumeration measures in the context of a GBTM. The investigated enumeration measures were classified into three categories, information criteria, likelihood ratio test derivatives, and entropy based statistics or classification indices. Data were simulated for a true 4-class GBTM with censored normal outcomes and true 4-class GBTM with binary logit outcomes. Both models were analyzed as a 3-, 4-, and 5-class GBTM with the intention of identifying the effect of different data conditions on the correct classification rates of the enumeration measures and to select an enumeration measure or a set of measures that provided the highest rate(s) of correctly selecting the true 4-class solution. In maintaining consistency with the previous literature, the sample size, number of repeated measures, class mixing proportions, percent of missing data, and the separation among the classes were manipulated.

5.1 CENSORED NORMAL MODEL AND BINARY LOGIT MODEL: RESULT COMPARISONS

Results from the CART analysis for the censored normal and binary logit models were very similar. In both CART, the most important predictor at level 1 was enumeration measures. There was a clear "good" performing group and "bad" performing group of enumeration measures in the results for both models. Whereas some of the information criteria were in the bad performing group for the censored normal model, all of the information criteria were in the good performing group for the binary logit model. This suggests that there was a smaller difference in correct classification rate among the information criteria for the binary logit model. The enumeration measures in the good performing group tended to perform best when the sample size was medium to large (800 or 2000), the mixing proportions were equal to moderately unequal, and there was high separation among the classes for both models. The number of repeated measures was also important in the binary logit model. Most of the enumeration measures, especially the good performing, had a higher correct classification rates when the number of repeated measures was 9 as opposed to 5. The percent of missing data did not have a major effect on the good performing enumeration measures in either model.

5.2 RECOMMENDATIONS

5.2.1 Recommended measures

The 2 measures recommended based on the results of this study are ssBIC and BIC. AIC and HQ are also recommended but they are recommended with caution and primarily for a particular data condition. ssBIC and BIC were two of the highest performing measures given ideal data conditions although their correct classification rates became very low given small sample sizes, extremely unequal class mixing proportions, and low separation among the classes. AIC and HQ had two of the highest overall average correct classification rates and were two of the best performing measures given non-ideal data conditions. AIC, HQ, ssBIC, and BIC performed similarly across model types, censored normal and binary logit. The performances of all 4 measures were also reasonably consistent with the 5 performance assessment studies summarized in Chapter 2. However, the recommendations were not always the same. Table 18 summarizes the recommendations (recommended, not recommended, or uncertain) of the 15 enumeration measures from the 5 reviewed studies, Tein et al. (2013), and the current study. The recommendations of BIC and ssBIC are consistent with some of the previous studies but the recommendations of AIC and HQ are not. AIC was not recommended in any of the reviewed studies but this was typically due to its asymptotic inefficiency. HQ was only studied in 2 of the reviewed studies. However, in those 2 studies, it performed relatively well but it was neither recommended nor not recommended. The authors of the 5 performance assessment studies summarized in Chapter 2 all tended to favor enumeration measures that demonstrated asymptotic efficiency and rightly so. This is a very desirable property for any enumeration measure. However, the results of this study demonstrated that given certain conditions present in the data, such as small sample size, or evidence gathered based on analysis results, such as extremely unequal class mixing proportions, it may be advantageous to also consider enumeration measures such as AIC and HQ. This is especially true for the censored normal model when class mixing proportions are extremely unequal. Figure 19 displays the classification rates of AIC, BIC, HQ and ssBIC for the censored normal model when the sample size was large (n=2000) the class mixing proportions were extremely unequal and the separation among the classes was high. Even with a large sample size and high separation among the classes, when the class mixing proportions were extremely unequal, the correct classification rates of BIC and ssBIC were only 30% and 49%, respectively, for the censored normal model. Given the same conditions, AIC and HQ both had a correct classification rate of 74%. It is quite possible that a theoretically

Table 18. Performance summary of 15 enumeration measures from 6 reviewed simulation studies and the current study

Measure	1 (LVMM)	2 (LCA)	3 (LCA)	3 (FMA)	3 (GMM)	4 (GMM)	5 (GMM)	6 (LPA)	7 (GBTM)
AIC	✗	✗	✗	✗	✗	✗	•	✗	✓
CAIC	✗	✗	•	✓	✓	✗	✗	-	✗
ssCAIC	-	•	-	-	-	•	✗	-	✗
BIC	✗	✗	✓	✓	✓	•	✗	✓	✓
B_{10}	-	-	-	-	-	-	-	-	✗
ssBIC	✓	✓	✓	•	•	✓	•	✓	✓
HQ	-	•	-	-	-	-	•	-	✓
ssHQ	-	-	-	-	-	-	•	-	✗
VLMR	•	-	•	•	•	✓	✗	✓	✗
aVLMR	•	-	-	-	-	✓	✗	✓	✗
sE	✗	-	-	-	-	-	•	✗	✗
NEC	✗	-	-	-	-	-	•	-	✗
CLC	✓	-	-	-	-	-	✗	-	✗
ICL-BIC	✓	-	-	-	-	-	✗	-	✗
ssICL-BIC	-	-	-	-	-	-	•	-	✗

Note. Statistical model utilized in each article is in parentheses. 1 = Henson, Reise, and Kim (2007); 2 = Yang (2006); 3 = Nylund, Asparouhov, and Muthén (2007); 4 = Tofighi and Enders (2008); 5 = Peugh and Fan (2012); 6 = Tein, Coxe, and Cham (2013); 6 = current study; LVMM = latent variable mixture model; LCA = latent class analysis; FMA = factor mixture model; GMM = growth mixture model; LPA = latent profile analysis; GBTM = semi-parametric group based trajectory model; ✓ = recommended; • = uncertain; ✗ = not recommended; - = not investigated.

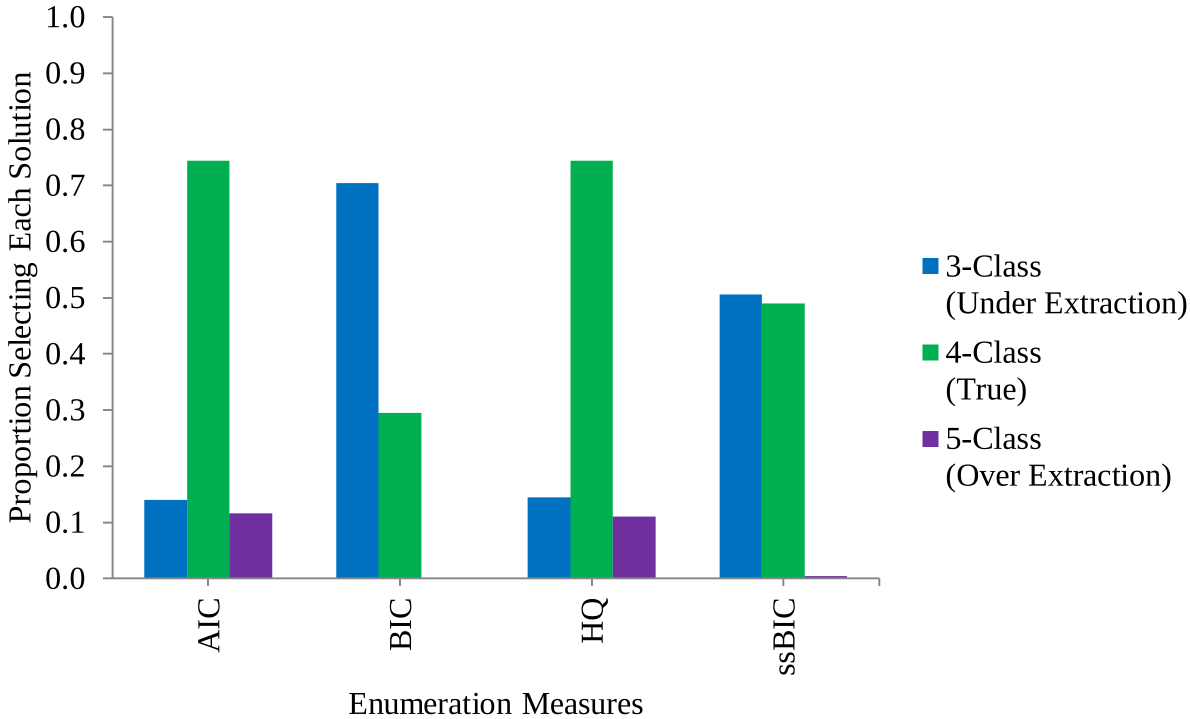

Figure 19. Proportion of 3-, 4-, and 5-class solutions identified as correct by select enumeration measures for a sample size of 2000, extremely unequal class mixing proportions, and high separation among the classes for the censored normal model

important but proportionately smaller class exists and AIC and HQ were 2 of the best enumeration measures at correctly classifying the model in this condition for the censored normal model. AIC, HQ, BIC, and ssBIC performed similarly in the same condition for the binary logit model. Thus, although AIC and HQ are recommended, they are recommended with caution given the recommendations in the reviewed studies and the limited scope of their recommended use here. They were both recommended despite their striking similarity in performance because each has a unique advantage. AIC is a default output of Proc Traj and Mplus so it is more widely available whereas HQ incorporates sample size into its penalty of the log likelihood.

In previous studies, the effect of sample size on the performance of AIC was typically less than that of other enumeration measures. Furthermore, the variability in the performance ofAIC was smaller overall and it was routinely one of the best performing measures given non-ideal data conditions such as small sample size or extremely unequal mixing proportions. In this study, the performance of AIC was very similar to previous studies with slightly more variability by data conditions. There was a higher tendency for AIC to make under extraction errors in this study. This may be due to the differences in the levels of separation factor between current and previous studies. The results from Tein et al. (2013) demonstrated that AIC tended to make under extraction errors when the separation among the classes was low but made over extraction errors when the separation among the classes was high. Class separation in this study could have been lower overall contributing the tendency of AIC to make under extraction errors. The relatively superior performance of AIC in non-ideal data conditions could be an artifact of its tendency to over parameterize a model, that is, over extract the number of classes (Bozdogan, 1987). Regardless, AIC's ease of computation and wide spread use still warrant its consideration, especially when the data conditions create a situation where there is low power to detect a higher albeit correct number of classes.

The performance of HQ was very similar to that of AIC. Not only did the two measures have comparable correct classification rates by manipulated factors but they also had substantial to high levels of agreement as measured by Cohen's Kappa for both the censored normal and binary logit models, .9 and .957 respectively . The comparable performances of AIC and HQ is consistent with the literature where it was noted by McQuarrie and Tsai (1998) in a regression context. HQ has not been thoroughly investigated in conjunction with class enumeration nor has it been historically employed for class enumeration. However, HQ was shown to perform well for model selection in regression at large sample sizes (McQuarrie & Tsai, 1998). The performance of HQ was investigated in 2 of the 5 previous studies reviewed in Chapter 2 and it was found to be comparable to the performance in this study. Yang (2006) found the performance of HQ to be somewhat consistent for various sample sizes although it tended to under extract the number of classes when the true solution consisted of 5 or 6 classes and the sample size was small (100 or 200). Peugh and Fan (2012) reported good performance of HQ relative to the other enumeration measures investigated although sample size had a larger effect on its performance than in this study. Despite the similar performances of HQ and AIC when used for class enumeration in a GBTM, as Miller (2002) suggests, HQ has better asymptotic properties which warrants its consideration.

As noted, the performance of ssBIC was more variable than the performances of both AIC and HQ. ssBIC was the best performing of the enumeration measures when the sample size was medium to large (800 or 2000), there were equal or moderately unequal class mixing proportions, and there was high separation among the classes. However, the performance of ssBIC was only on par with AIC and HQ when the sample size was small (200 or 400), the class mixing proportions were equal or moderately unequal, and there was high separation among the classes. When the class mixing proportions were extremely unequal or there was low separation among the classes, ssBIC did not perform as well as AIC and HQ regardless of the levels of the other factors. The performance of ssBIC was very consistent with its performances reported in the 5 previous studies. In those studies, ssBIC was typically one of the few measures advocated for use or at least recognized as one of the better performing measures. The performances of ssBIC in this study and in previous studies suggest that it should be examined for the purpose of class enumeration in a GBTM.

BIC, like ssBIC, had one of the highest correct classification rates when the sample size was medium or large (800 or 200), the class mixing proportions were equal or moderately unequal, and there was high separation among the classes. However, also like ssBIC, the correct classification rate of BIC dropped significantly when data conditions were not ideal. Although ssBIC always had a higher correct classification rate than BIC, BIC is still recommended for use for a couple of reasons. BIC is the currently recommended measure to use for class enumeration in a GBTM by Nagin (see, for example, Nagin, 2005). BIC consistently under extracts rather than over extracts the number of classes. This property was demonstrated in this study and in the 5 previous studies and is considered important based on results in 2 of the 5 previous studies and in Tein et al. (2013). Nylund et al. (2007), Yang (2006), and Tein et al. (2013) reported that BIC consistently under extracted classes when in error. They also reported that ssBIC sometimes over extracted classes when in error. This characteristic of ssBIC's performance is inconsistent with the results of the current study but there is a reasonable explanation. It is quite possible that there was higher separation among the classes in the studies done by Nylund et al. (2007) and Yang (2006). In fact, Tein et al. (2013) demonstrated that ssBIC begins to over extract classes as the separation among the classes becomes very high. In the high separation condition, the tendency of BIC to under extract classes when in error was to its benefit. Very high separation among the classes caused ssBIC to make over extraction errors at the expense of the correct classification rate. However, very high separation only caused the correct classification rate of BIC to increase as the rate of under extraction decreased without an increase in the rate of over extraction. Therefore BIC is recommended for class enumeration in a GBMT because of the similar performances in this study and previous studies - where it consistently tended to make under extraction errors or extract the correct number of classes - and its support for use in the literature.

5.2.2 Non-recommended measures

Several measures were not recommended because they performed poorly or they were similar but did not outperform the recommended measures. For example, the performance of ssCAIC was comparable to that of ssBIC but ssBIC always held a slight edge regardless of condition. Similarly, the performances of the enumeration measures that used the number of non-missing observations as the sample size (CAIC2, ssCAIC2, BIC2, ssBIC2, HQ2, ssHQ2, ICL-BIC2, ssICL-BIC2) had lower or comparative classification rates to their standard computation counterparts and were not recommended.

Contrary to expectation, none of the entropy based statistics or classification indices are recommended. sE, NEC, CLC, ICL-BIC and ssICL-BIC were not recommended because they all performed poorly. The best performing of them only correctly classified 13% and 16% of the solutions for the censored normal model and binary logit model respectively. The performances of CAIC, ssCAIC, B_{10}, ssHQ, VLMR, and aVLMR were not terrible and had overall correct classification rates ranging from 19% to 52% for both model type. However, CAIC and B_{10} did not perform as well as BIC, ssCAIC did not perform as well as ssBIC, ssHQ did not perform as well as HQ, and VLMR and aVLMR did not perform as well as any of the recommended measures. The performances of CAIC, ssCAIC, ssHQ, ICL-BIC, VLMR, aVLMR and ssICL-BIC were all considered consistent with their performances in the literature. Any slight discrepancies could most likely be attributed to differences in models, number of trajectories and trajectory patterns, and factors or the levels of factors manipulated. sE, NEC, and CLC were examined in both Henson et al. (2007) and Peugh and Fan (2012) and their performances were somewhat inconsistent with this study. sE, NEC, and CLC all made over extraction errors at a higher rate than under extraction errors in both reviewed articles. In Peugh and Fan (2012), sE and NEC outperformed the information criteria in certain conditions and CLC almost always over extracted the number of classes. These disparities in performances may require additional research to fully understand but one plausible explanation is the difference in models between the studies. This study used a GBTM and individual trajectories were not allowed to vary within each class. Peugh and Fan (2012) used a GMM and individual trajectories were allowed to vary within each class. The added variability of the GMM may necessitate additional classes to produce better overall class separation by effectively reducing the amount of within-group variability.

5.3 LIMITATIONS

The results and subsequent conclusions of this study are somewhat constrained by the data generation model. A true 4-class solution was utilized. The extraction of a 4-class solution was found to occur fairly often in the literature but it is not uncommon to see as small as 2 to as large as 7 or more classes extracted. This is meaningful because Yang (2006) reported that the number of latent classes had an impact on the classification rates of enumeration measures.

In addition to having a constant number of classes, the class trajectories were not varied for either the censored normal or the binary logit model. There is an infinite number of trajectories that a class can follow in a GBTM and only one set was utilized in this study for both the censored normal and binary logit models. The class trajectories can vary with respect to their highest order polynomial and magnitude of each growth coefficient. This point is relevant because it has been demonstrated that differences in growth trajectories can impact the performance of enumeration measures even while holding the separation among classes constant (Peugh & Fan, 2012).

The enumeration measures were evaluated based on their ability to select the true 4-class solution from competing 3- and 5-class solutions. An assumption was made for this study regarding how many classes below and above the true 4-class solution should be estimated for comparison. The assumption stated that if an enumeration measure does not select a 3-class solution over a true 4-class solution then it would not select a 2-class solution over a true 4-class solution. The same statement can be made for 5- and 6-class solutions. This logic is likely to hold in most circumstances but there could be instances, although rare, when there is an exception. It is assumed that further comparing the 4-class solution to a 1-, 2-, 6-, and 7-class solution or larger would not have drastically impacted the results. Most likely the correct classification rates of all measures would have remained stable but the extraction errors would have had more variability over the incorrect class number solutions. However, this brings up an important topic, determining if the data truly came from a mixture. Bauer and Curran (2003) demonstrated that multiple classes can be improperly extracted even when only 1 class exists if data is nonnormal. It remains to be determined how the enumeration measures evaluated in this study would perform given a true 1-class solution and whether or not the most powerful measures would also maintain a reasonable Type I error rate.

Viewing each generated sample as a "true" 4-class solution also places some restriction on the interpretability of the results. It has repeatedly been stated that classes are not intended to represent literally distinct entities that exist in the population (Nagin, 1999; 2005; Nagin &

Tremblay, 2005). Even though data was generated for 4 separate classes and combined into one sample, it is possible that in some circumstances – for example, when there was low separation among the classes – the data would appear to have been generated from 3-classes. Although the data was always considered to have come from a true 4-class solution, there may not have been enough separation for this to be completely accurate.

Another assumption was made in this study regarding non-convergence. Any time at least one but not all of the 3-, 4-, or 5-class solutions failed to converge; the enumeration measures had less than 3 solutions to select from. For example, if a 5-class solution failed to converge, the enumeration measures selected between the remaining 3-class and 4-class solutions. A sensitivity analysis was conducted to determine the extent that this assumption affected the results. A CART analysis was run for both the censored normal model and the binary logit model using only those replications where all 3 class number solutions converged. It was determined that treating non-converging solutions in this manner had little effect on the results of the CART analysis for the censored normal model. In fact, the results were almost identical. The number of non-converging solutions was larger for the binary logit model and there were more discrepancies in the CART. However, the important predictors persisted in both sets of results and the enumeration measures recommended for use remained the same.

Some additional limitations were due to the way the missing data was specified and the way VLMR and aVLMR were computed. Missing data was specified as MCAR. Missing data had little effect on the performances of the majority of the enumeration measures regardless of model type. Nagin (2005) discussed that the GBTM can handle data that is MCAR but it is not suited for the more complex types of missing data, MAR and MNAR. Had the type of missing data been specified as MAR or MNAR, to test the GBTM outside of its recommended bounds, missing data may have had a larger impact on the performance of the enumeration measures.

It was previously mentioned that VLMR and aVLMR were computed using MPlus due to the computational difficulty of the null distribution and because the tests have been validated in MPlus. Steps were taken to ensure that the 4- and 5- class solutions were as similar as possible between Proc TRAJ and MPlus. A potential problem is that Proc TRAJ and MPlus use different optimization algorithms and there could have been differences in a final estimated model. However, any differences were expected to be minimal because Blaze and Kim (2013) have demonstrated that Proc TRAJ and MPlus produce very similar model estimates.

5.4 FUTURE RESEARCH

The current study helped to build a case for the use of specific enumeration measures in a GBTM. The progression of research on this topic can focus on the limitations in the data generation model. Data for this study was generated based on 4-class censored normal and binary logit population models. The enumeration measures performed similarly for these different model types and it is expected that they would have comparable performances for a GBTM with a Poisson distribution. However, it is recommended that the performances of the enumeration measures be examined with respect to a GBTM with a Poisson distribution.

The performances of the enumeration measures should also be investigated in the context of a 2-, 3-, 5-, and 6-class population model. The 43 studies discussed in Chapter 3 that employed a GBTM in practice were the basis for this suggested range of classes. Of the 69 GBTMs run in the 43 reviewed GBTM studies, more than 6 classes were extracted in only 1.

Along similar lines, data could be generated with different class trajectories while maintaining the degree of separation among the classes, which was done by Peugh and Fan (2012) in the context of GMM.

Investigating the performances of the enumeration measures when model assumptions break down would also be an important contribution. This study focused on an ideal situation where all of the assumptions of the GBTM were met. The data for each class at each time point was specified to be normally distributed, the covariance matrix was a homogenous diagonal matrix, and missing data were generated to meet the MCAR assumption. Practitioners would benefit from knowing the degree of non-normality or heterogeneity of variance among the classes that could cause significant drops in the correct classification rates of enumeration measures. Also, because data often do not meet the MCAR assumption, it would be advantageous to study the behavior of the enumeration measures when missing data is MAR or MNAR.

The entropy based statistics and classification indices performed poorly in this study compared to the information criteria and likelihood ratio test derivatives. This is not entirely inconsistent with the literature although they were not expected to perform as poorly as they did. Peugh and Fan (2012) reported the superior performance of sE and NEC to some of the information criteria in certain data conditions. Also, McLachlan and Peel (2000) reported on various simulation studies involving finite mixture models in which ICL-BIC performed very well. It could be worthwhile to assess entropy and how it behaves depending on the number of classes extracted. For example, the distribution of sE could be assessed relative to the number of classes extracted. Entropy is a measure of the degree of separation among classes in a finite mixture model. Examining the behavior of entropy more closely should help to determine if it can be a useful measure of the degree of separation among classes and ultimately a tool that can be used for class enumeration.

APPENDIX A

CENSORED NORMAL MODEL DATA GENERATION SYNTAX

```
%MACRO diss_cnorm;
%LET dir=c:\tempsas\Dissertation\work_cnorm;
%LET df3 = 12;
%LET df4 = 16;
%LET df5 = 20;
%LET s1 = 3;
%LET s2 = 4;
%LET s3 = 5;
%LET mrep = 50;

DATA condition;
INFILE "&dir\last_rep.txt" DELIMITER=',' end=last;
INPUT n obs mix miss sigma rep;
CALL SYMPUT ('i',n);
CALL SYMPUT ('j',obs);
CALL SYMPUT ('k',mix);
CALL SYMPUT ('l',miss);
CALL SYMPUT ('m',sigma);
CALL SYMPUT ('rep',rep);
IF last;
RUN;

%IF &rep = &mrep %THEN %DO;
  %IF &m < 2 %THEN %DO;
    %LET m = &m+1;
      %LET rep = 1;
  %END;
  %ELSE %DO;
    %IF &l < 3 %THEN %DO;
        %LET l = &l+1;
        %LET m = 1;
        %LET rep = 1;
      %END;
    %ELSE %DO;
        %IF &k < 3 %THEN %DO;
          %LET k = &k+1;
            %LET l = 1;
```

```
            %LET m = 1;
            %LET rep = 1;
        %END;
      %ELSE %DO;
          %IF &j < 2 %THEN %DO;
              %LET j = &j+1;
              %LET k = 1;
              %LET l = 1;
              %LET m = 1;
              %LET rep = 1;
            %END;
            %ELSE %DO;
              %IF &i < 4 %THEN %DO;
                %LET i = &i + 1;
                %LET j = 1;
                %LET k = 1;
                %LET l = 1;
                %LET m = 1;
                  %LET rep = 1;
              %END;
              %ELSE %DO;
                %LET rep = &rep+1;
              %END;
            %END;
      %END;
    %END;
  %END;
%END;
%ELSE %DO;
  %LET rep = &rep+1;
%END;

%DO %WHILE (&i <= 4);
  %IF &i = 1 %THEN %LET n = 200;
  %IF &i = 2 %THEN %LET n = 400;
  %IF &i = 3 %THEN %LET n = 800;
  %IF &i = 4 %THEN %LET n = 2000;

  %DO %WHILE (&j <= 2);
    %IF &j = 1 %THEN %LET obs = 5;
      %IF &j = 2 %THEN %LET obs = 9;

      %DO %WHILE (&k <= 3);
        %IF &k = 1 %THEN %LET mix = {.25 .25 .25 .25};
        %IF &k = 2 %THEN %LET mix = {.4 .25 .2 .15};
        %IF &k = 3 %THEN %LET mix = {.6 .25 .1 .05};

      %DO %WHILE (&l <= 3);
          %IF &l = 1 %THEN %LET miss = 0;
        %IF &l = 2 %THEN %LET miss = .1;
        %IF &l = 3 %THEN %LET miss = .2;

            %DO %WHILE (&m <= 2);
              %IF &m = 1 %THEN %LET sigma = 2;
              %IF &m = 2 %THEN %LET sigma = 3;

          %DO %WHILE (&rep <= &mrep);
```

```
    DATA _NULL_;
SET condition;
FILE "&dir\reps.txt" DLM = ',' MOD;
n = &i;
obs = &j;
mix = &k;
miss = &l;
sigma = &m;
rep = &rep;
PUT n obs mix miss sigma rep;
RUN;

DATA _NULL_;
SET condition;
FILE "&dir\last_rep.txt" DLM = ',';
n = &i;
obs = &j;
mix = &k;
miss = &l;
sigma = &m;
rep = &rep;
PUT n obs mix miss sigma rep;
RUN;

PROC IML;
      seed = &i*100000000 + &j*10000000 + &k*1000000 + &l*100000
+ &m*10000 + &rep;
      CALL RANDSEED(seed);
      nobs = 0;
      obs = &obs;
      sigma = &sigma;
      n = &n;
      miss = &miss;
      mix = &mix;

      beta = {-2.143 2.912 -.337 .337, .388 -.279 1.191 -1.191, -
.091 .02 -.156 .156};
      IF obs = 5 THEN t = {1 0 0, 1 1 1, 1 2 4, 1 3 9, 1 4 16};
      ELSE t = {1 0 0, 1 .5 .25, 1 1 1, 1 1.5 2.25, 1 2 4, 1 2.5
6.25, 1 3 9, 1 3.5 12.25, 1 4 16};
      pmx = J(n,obs,.);
      CALL RANDGEN(pmx,'UNIFORM');
      n1 = round(n*mix[1]);
      n2 = round(n*mix[2]);
      n3 = round(n*mix[3]);
      n4 = round(n*mix[4]);
      t1 = repeat(t,n1);
      t2 = repeat(t,n2);
      t3 = repeat(t,n3);
      t4 = repeat(t,n4);
      e1 = J(n1*obs,1,.);
      e2 = J(n2*obs,1,.);
      e3 = J(n3*obs,1,.);
      e4 = J(n4*obs,1,.);
      CALL RANDGEN(e1,'NORMAL',0,sigma);
      CALL RANDGEN(e2,'NORMAL',0,sigma);
      CALL RANDGEN(e3,'NORMAL',0,sigma);
```

```
                  CALL RANDGEN(e4,'NORMAL',0,sigma);
                  y1 = t1*beta[,1] + e1;
                  y2 = t2*beta[,2] + e2;
                  y3 = t3*beta[,3] + e3;
                  y4 = t4*beta[,4] + e4;

                  y = y1//y2//y3//y4;

                  gn1 = repeat(1,n1);
                  gn2 = repeat(2,n2);
                  gn3 = repeat(3,n3);
                  gn4 = repeat(4,n4);

                  gn = gn1//gn2//gn3//gn4;
                  id = shape(1:n,n,1);

                  yw = shape(y,n,obs);
                  ynw = yw + 20;

                  DO o=1 TO n;
                    DO p=1 TO obs;
                      IF ynw[o,p] > 30 THEN ynw[o,p] = 30;
                        IF ynw[o,p] < 0 THEN ynw[o,p] = 0;
                      IF (p ^= 1 & miss ^= 0) THEN DO;
                        IF(pmx[o,p] <= miss | missing(ynw[o,p-1])=1) THEN
ynw[o,p] = .; ELSE nobs = nobs+1;
                        END;
                  ELSE DO; nobs = nobs+1;
                  END;
                    END;
                  END;
                  CALL SYMPUT ('nobs',LEFT(CHAR(nobs)));
                  IF obs = 5 THEN DO;
                    ynw =
id||gn||J(n,1,0)||J(n,1,1)||J(n,1,2)||J(n,1,3)||J(n,1,4)||ynw;
                    CREATE trajdata_n FROM ynw
[colname={id,gn,t1,t2,t3,t4,t5,y1,y2,y3,y4,y5}];
                    APPEND FROM ynw;
                  END;
                  ELSE IF obs=9 THEN DO;
                    ynw =
id||gn||J(n,1,0)||J(n,1,0.5)||J(n,1,1)||J(n,1,1.5)||J(n,1,2)||J(n,1,2.5)||J(n
,1,3)||J(n,1,3.5)||J(n,1,4)||ynw;
                    CREATE trajdata_n FROM ynw
[colname={id,gn,t1,t2,t3,t4,t5,t6,t7,t8,t9,y1,y2,y3,y4,y5,y6,y7,y8,y9}];
                    APPEND FROM ynw;
                  END;
                  QUIT;

                  FILENAME el
'C:\tempsas\Dissertation\work_cnorm\error_log.txt';

                  /*****1-CLASS MODEL TO OBTAIN LOG LIKELIHOOD FOR NEC******/
                  PROC PRINTTO LOG = el NEW;
                  RUN;
```

```
                  PROC TRAJ data=trajdata_n outplot = op1 outstat = os1 out =
of1 outest = oe1;
                  ID id gn;
                  VAR y1-y&obs;
                  INDEP t1-t&obs;
                  MODEL cnorm;
                  MIN 0;
                  MAX 30;
                  NGROUPS 1;
                  ORDER 2;
                  RUN;

                  PROC PRINTTO;
                  RUN;

                  *check for errors and/or warnings;
                  DATA e_log1;
                  INFILE el dlm='10'X dsd;
                  LENGTH message $256;
                  INPUT message;
                  IF FIND(message, "ERROR: Singular convergence")~=0 THEN
issue = 1;
                  ELSE IF FIND(message,"WARNING: Unable to calculate standard
errors")~=0 THEN issue = 2;
                  ELSE IF FIND (message,"ERROR: Floating Point Overflow")~=0
THEN issue = 3;
                  ELSE IF FIND (message,"ERROR: Floating Point Zero
Divide")~=0 THEN issue = 4;
                  ELSE IF FIND (message,"WORK.SSCPTTZ has 0 observations and
0 variables")~=0 THEN issue = 5;
                  ELSE issue = 0;
                  IF (issue = 0 & (FIND(message,
"ERROR")~=0|FIND(message,"WARNING")~=0)) THEN issue = 6;
                  DROP message;
                  RUN;

                  PROC SORT DATA=e_log1 nodupkey;
                  BY issue;
                  RUN;

                  DATA e_log1;
                  SET e_log1;
                  flag = 1;
                  RUN;

                  PROC TRANSPOSE DATA=e_log1 OUT=e_log1 PREFIX=c1_issue;
                  ID issue;
                  VAR flag;
                  RUN;

                  DATA _null_;
                  dset=open('e_log1');
                  CALL SYMPUT ('chk1',varnum(dset,'c1_issue1'));
                  CALL SYMPUT ('chk2',varnum(dset,'c1_issue2'));
                  CALL SYMPUT ('chk3',varnum(dset,'c1_issue3'));
                  CALL SYMPUT ('chk4',varnum(dset,'c1_issue4'));
                  CALL SYMPUT ('chk5',varnum(dset,'c1_issue5'));
```

```
                    CALL SYMPUT ('chk6',varnum(dset,'c1_issue6'));
                    RUN;

                    DATA e_log1;
                    SET e_log1;
                    IF &chk1 = 0 THEN c1_issue1 = 0;
                    IF &chk2 = 0 THEN c1_issue2 = 0;
                    IF &chk3 = 0 THEN c1_issue3 = 0;
                    IF &chk4 = 0 THEN c1_issue4 = 0;
                    IF &chk5 = 0 THEN c1_issue5 = 0;
                    IF &chk6 = 0 THEN c1_issue6 = 0;
                    DROP _NAME_ c1_issue0;
                    RUN;
                    *end check for errors and/or warnings;

                    DATA ll_oneclass;
                    RETAIN c1_issue1 c1_issue2 c1_issue3 c1_issue4 c1_issue5
c1_issue6;
                    SET oe1;
                    IF _N_ = 1 THEN SET e_log1;
                    IF _TYPE_ ~= 'PARMS' THEN DELETE;
                    DROP _MODEL_ _MODEL2_ _TYPE_ _NAME_ INTERC1 LINEAR1 QUADRA1
                         SIGMA1 _AIC_ _BIC1_ _BIC2_ _CONVERGE_;
                    RENAME _loglik_ = c1loglik;
                    RUN;
                    /*****END 1-CLASS MODEL TO OBTAIN LOG LIKELIHOOD FOR
NEC*****/


                    /*****3-CLASS MODEL AND OUTPUT*****/
                    PROC PRINTTO LOG = el NEW;
                    RUN;

                    PROC TRAJ data = trajdata_n outplot = op3 outstat = os3 out
= of3 outest = oe3;
                    ID id gn;
                    VAR y1-y&obs;
                    indep t1-t&obs;
                    MODEL cnorm;
                    MIN 0;
                    MAX 30;
                    NGROUPS 3;
                    ORDER 2 2 2;
                    RUN;

                    PROC PRINTTO;
                    RUN;

                    *check for errors and/or warnings;
                    DATA e_log3;
                    INFILE el dlm='10'X dsd;
                    LENGTH message $256;
                    INPUT message;
                    IF FIND(message, "ERROR: Singular convergence")~=0 THEN
issue = 1;
                    ELSE IF FIND(message,"WARNING: Unable to calculate standard
errors")~=0 THEN issue = 2;
```

```
                  ELSE IF FIND (message,"ERROR: Floating Point Overflow")~=0
THEN issue = 3;
                  ELSE IF FIND (message,"ERROR: Floating Point Zero
Divide")~=0 THEN issue = 4;
                  ELSE IF FIND (message,"WORK.SSCPTTZ has 0 observations and
0 variables")~=0 THEN issue = 5;
                  ELSE issue = 0;
                  IF (issue = 0 & (FIND(message,
"ERROR")~=0|FIND(message,"WARNING")~=0)) THEN issue = 6;
                  DROP message;
                  RUN;

                PROC SORT DATA=e_log3 nodupkey;
                  BY issue;
                  RUN;

                  DATA e_log3;
                  SET e_log3;
                  flag = 1;
                  RUN;

                  PROC TRANSPOSE DATA=e_log3 OUT=e_log3 PREFIX=issue;
                  ID issue;
                  VAR flag;
                  RUN;

                  DATA _null_;
                  dset=open('e_log3');
                  CALL SYMPUT ('chk1',varnum(dset,'issue1'));
                  CALL SYMPUT ('chk2',varnum(dset,'issue2'));
                  CALL SYMPUT ('chk3',varnum(dset,'issue3'));
                  CALL SYMPUT ('chk4',varnum(dset,'issue4'));
                  CALL SYMPUT ('chk5',varnum(dset,'issue5'));
                  CALL SYMPUT ('chk5',varnum(dset,'issue6'));
                  RUN;

                  DATA e_log3;
                  RETAIN issue1 issue2 issue3 issue4 issue5 issue6;
                  SET e_log3;
                  IF &chk1 = 0 THEN issue1 = 0;
                  IF &chk2 = 0 THEN issue2 = 0;
                  IF &chk3 = 0 THEN issue3 = 0;
                  IF &chk4 = 0 THEN issue4 = 0;
                  IF &chk5 = 0 THEN issue5 = 0;
                  IF &chk6 = 0 THEN issue6 = 0;
                  DROP _NAME_ issue0;
                  RUN;
                  *end check for errors and/or warnings;

                  *Compute entropy and add to oe dataset;
                  DATA of3;
            SET of3;
            IF grp1prb = 0 THEN pp1 = 0; ELSE pp1 = grp1prb*log(grp1prb);
                  IF grp2prb = 0 THEN pp2 = 0; ELSE pp2 =
grp2prb*log(grp2prb);
            IF grp3prb = 0 THEN pp3 = 0; ELSE pp3 = grp3prb*log(grp3prb);
            pp_j = pp1+pp2+pp3;
```

```
        RUN;

              PROC MEANS DATA=of3 NOPRINT;
              VAR pp_j;
              OUTPUT SUM=pp_sum OUT=fuzzy3;
              RUN;

              DATA fuzzy3;
              SET fuzzy3;
              DROP _TYPE_ _FREQ_;
              RUN;

              DATA oe3;
              SET oe3;
              IF _N_ = 1 THEN SET fuzzy3;
              IF _N_ = 1 THEN SET ll_oneclass;
              entropy = -1*pp_sum;
              DROP pp_sum;
              RUN;
              *End compute entropy;

              *Create dataset with enumeration measures;
              DATA em3;
              SET oe3;
              IF _TYPE_ ~= 'PARMS' THEN DELETE;
              DROP _MODEL_ _MODEL2_ _TYPE_ _NAME_ INTERC1-INTERC3
LINEAR1-LINEAR3 QUADRA1-QUADRA3 SIGMA1 THETA2-THETA3 _CONVERGE_;
              AIC = -2*_LOGLIK_ + 2*&df3;
              BIC = -2*_LOGLIK_ + log(&n)*&df3;
              CAIC = -2*_LOGLIK_ + (log(&n)+1)*&df3;
              HQ = -2*_LOGLIK_ + log(log(&n))*&df3;
              ssBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df3;
              ssCAIC = -2*_LOGLIK_ + (log((&n+2)/24)+1)*&df3;
              ssHQ = -2*_LOGLIK_ + log(log((&n+2)/24))*&df3;
              sE = 1 - (entropy/(&n*log(&s1)));
              NEC = entropy/(_LOGLIK_ - c1loglik);
              CLC = -2*_LOGLIK_ + 2*entropy;
              ICLBIC = -2*_LOGLIK_ + log(&n)*&df3 + 2*entropy;
              ssICLBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df3 + 2*entropy;
              BIC2 = -2*_LOGLIK_ + log(&nobs)*&df3;
              CAIC2 = -2*_LOGLIK_ + (log(&nobs)+1)*&df3;
              HQ2 = -2*_LOGLIK_ + log(log(&nobs))*&df3;
              ssBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df3;
              ssCAIC2 = -2*_LOGLIK_ + (log((&nobs+2)/24)+1)*&df3;
              ssHQ2 = -2*_LOGLIK_ + log(log((&nobs+2)/24))*&df3;
              ICLBIC2 = -2*_LOGLIK_ + log(&nobs)*&df3 + 2*entropy;
              ssICLBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df3 +
2*entropy;
              CALL SYMPUT ('bic3',BIC);
              RUN;
              *End create dataset with enumeration measures;

              *Compute average posterior probabilities;
              PROC SORT DATA=of3;
              BY group;
              RUN;
```

```
PROC MEANS DATA=of3 NOPRINT;
VAR GRP1PRB GRP2PRB GRP3PRB;
BY group;
OUTPUT OUT=desc3;
RUN;

DATA desc3;
SET desc3;
IF _stat_ ~= 'MEAN' THEN DELETE;
IF group = 1 THEN avgpp = GRP1PRB;
ELSE IF group = 2 THEN avgpp = GRP2PRB;
ELSE IF group = 3 THEN avgpp = GRP3PRB;
DROP _TYPE_ _FREQ_ _STAT_ GRP1PRB GRP2PRB GRP3PRB group;
RUN;

PROC SORT DATA=desc3;
BY DESCENDING avgpp;
RUN;

DATA desc3;
SET desc3;
id = _N_;
RUN;

PROC TRANSPOSE DATA=desc3 OUT=pp_means3 PREFIX=PP;
ID id;
VAR avgpp;
RUN;

DATA pp_means3;
SET pp_means3;
PP4 = .;
PP5 = .;
DROP _NAME_;
RUN;
*End compute average posterior probabilities;
/*****END 3-CLASS MODEL AND OUTPUT*****/

/*****4-CLASS MODEL AND OUTPUT*****/
PROC PRINTTO LOG = el NEW;
RUN;

PROC TRAJ data = trajdata_n outplot = op4 outstat = os4 out
= of4 outest = oe4;
ID id gn;
VAR y1-y&obs;
indep t1-t&obs;
MODEL cnorm;
MIN 0;
MAX 30;
NGROUPS 4;
ORDER 2 2 2 2;
RUN;

PROC PRINTTO;
RUN;
```

```
                    *check for errors and/or warnings;
                    DATA e_log4;
                    INFILE el dlm='10'X dsd;
                    LENGTH message $256;
                    INPUT message;
                    IF FIND(message, "ERROR: Singular convergence")~=0 THEN
issue = 1;
                    ELSE IF FIND(message,"WARNING: Unable to calculate standard
errors")~=0 THEN issue = 2;
                    ELSE IF FIND (message,"ERROR: Floating Point Overflow")~=0
THEN issue = 3;
                    ELSE IF FIND (message,"ERROR: Floating Point Zero
Divide")~=0 THEN issue = 4;
                    ELSE IF FIND (message,"WORK.SSCPTTZ has 0 observations and
0 variables")~=0 THEN issue = 5;
                    ELSE issue = 0;
                    IF (issue = 0 & (FIND(message,
"ERROR")~=0|FIND(message,"WARNING")~=0)) THEN issue = 6;
                    DROP message;
                    RUN;

                    PROC SORT DATA=e_log4 nodupkey;
                    BY issue;
                    RUN;

                    DATA e_log4;
                    SET e_log4;
                    flag = 1;
                    RUN;

                    PROC TRANSPOSE DATA=e_log4 OUT=e_log4 PREFIX=issue;
                    ID issue;
                    VAR flag;
                    RUN;

                    DATA _null_;
                    dset=open('e_log4');
                    CALL SYMPUT ('chk1',varnum(dset,'issue1'));
                    CALL SYMPUT ('chk2',varnum(dset,'issue2'));
                    CALL SYMPUT ('chk3',varnum(dset,'issue3'));
                    CALL SYMPUT ('chk4',varnum(dset,'issue4'));
                    CALL SYMPUT ('chk5',varnum(dset,'issue5'));
                    CALL SYMPUT ('chk6',varnum(dset,'issue6'));
                    RUN;

                    DATA e_log4;
                    SET e_log4;
                    RETAIN issue1 issue2 issue3 issue4 issue5 issue6;
                    IF &chk1 = 0 THEN issue1 = 0;
                    IF &chk2 = 0 THEN issue2 = 0;
                    IF &chk3 = 0 THEN issue3 = 0;
                    IF &chk4 = 0 THEN issue4 = 0;
                    IF &chk5 = 0 THEN issue5 = 0;
                    IF &chk6 = 0 THEN issue6 = 0;
                    DROP _NAME_ issue0;
                    RUN;
                    *end check for errors and/or warnings;
```

```
*Compute entropy and add to oe dataset;
DATA of4;
SET of4;
IF grp1prb = 0 THEN pp1 = 0; ELSE pp1 =
grp1prb*log(grp1prb);
IF grp2prb = 0 THEN pp2 = 0; ELSE pp2 =
grp2prb*log(grp2prb);
IF grp3prb = 0 THEN pp3 = 0; ELSE pp3 =
grp3prb*log(grp3prb);
IF grp4prb = 0 THEN pp4 = 0; ELSE pp4 =
grp4prb*log(grp4prb);
pp_j = pp1+pp2+pp3+pp4;
RUN;

PROC MEANS DATA=of4 NOPRINT;
VAR pp_j;
OUTPUT SUM=pp_sum OUT=fuzzy4;
RUN;

DATA fuzzy4;
SET fuzzy4;
DROP _TYPE_ _FREQ_;
RUN;

DATA oe4;
SET oe4;
IF _N_ = 1 THEN SET fuzzy4;
IF _N_ = 1 THEN SET ll_oneclass;
entropy = -1*pp_sum;
DROP pp_sum;
RUN;
*End compute entropy;

*Create dataset with enumeration measures;
DATA em4;
SET oe4;
IF _TYPE_ ~= 'PARMS' THEN DELETE;
DROP _MODEL_ _MODEL2_ _TYPE_ _NAME_ INTERC1-INTERC4
LINEAR1-LINEAR4 QUADRA1-QUADRA4 SIGMA1 THETA2-THETA4 _CONVERGE_;
AIC = -2*_LOGLIK_ + 2*&df4;
BIC = -2*_LOGLIK_ + log(&n)*&df4;
CAIC = -2*_LOGLIK_ + (log(&n)+1)*&df4;
HQ = -2*_LOGLIK_ + log(log(&n))*&df4;
ssBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df4;
ssCAIC = -2*_LOGLIK_ + (log((&n+2)/24)+1)*&df4;
ssHQ = -2*_LOGLIK_ + log(log((&n+2)/24))*&df4;
sE = 1 - (entropy/(&n*log(&s2)));
NEC = entropy/(_LOGLIK_ - c1loglik);
CLC = -2*_LOGLIK_ + 2*entropy;
ICLBIC = -2*_LOGLIK_ + log(&n)*&df4 + 2*entropy;
ssICLBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df4 + 2*entropy;
BIC2 = -2*_LOGLIK_ + log(&nobs)*&df4;
CAIC2 = -2*_LOGLIK_ + (log(&nobs)+1)*&df4;
HQ2 = -2*_LOGLIK_ + log(log(&nobs))*&df4;
ssBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df4;
ssCAIC2 = -2*_LOGLIK_ + (log((&nobs+2)/24)+1)*&df4;
```

```
                    ssHQ2 = -2*_LOGLIK_ + log(log((&nobs+2)/24))*&df4;
                    ICLBIC2 = -2*_LOGLIK_ + log(&nobs)*&df4 + 2*entropy;
                    ssICLBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df4 +
2*entropy;
                    logB10 = &bic3-BIC;
                    B10 = exp((&bic3-BIC)/2);
                    CALL SYMPUT ('bic4',BIC);
                    RUN;
                    *End create dataset with enumeration measures;

                    *Compute average posterior probabilities;
                    PROC SORT DATA=of4;
                    BY group;
                    RUN;

                    PROC MEANS DATA=of4 NOPRINT;
                    VAR GRP1PRB GRP2PRB GRP3PRB GRP4PRB;
                    BY group;
                    OUTPUT OUT=desc4;
                    RUN;

                    DATA desc4;
                    SET desc4;
                    IF _stat_ ~= 'MEAN' THEN DELETE;
                    IF group = 1 THEN avgpp = GRP1PRB;
                    ELSE IF group = 2 THEN avgpp = GRP2PRB;
                    ELSE IF group = 3 THEN avgpp = GRP3PRB;
                    ELSE IF group = 4 THEN avgpp = GRP4PRB;
                    DROP _TYPE_ _FREQ_ _STAT_ GRP1PRB GRP2PRB GRP3PRB GRP4PRB
group;
                    RUN;

                    PROC SORT DATA=desc4;
                    BY DESCENDING avgpp;
                    RUN;

                    DATA desc4;
                    SET desc4;
                    id = _N_;
                    RUN;

                    PROC TRANSPOSE DATA=desc4 OUT=pp_means4 PREFIX=PP;
                    ID id;
                    VAR avgpp;
                    RUN;

                    DATA pp_means4;
                    SET pp_means4;
                    pp5 = .;
                    DROP _NAME_;
                    RUN;
                    *End compute average posterior probabilities;
                    /*****END 4-CLASS MODEL AND OUTPUT*****/

                    /*5-CLASS MODEL AND OUTPUT*/
                    PROC PRINTTO LOG = el NEW;
                    RUN;
```

```
                  PROC TRAJ data = trajdata_n outplot = op5 outstat = os5 out
= of5 outest = oe5;
                  ID id gn;
                  VAR y1-y&obs;
                  indep t1-t&obs;
                  MODEL cnorm;
                  MIN 0;
                  MAX 30;
                  NGROUPS 5;
                  ORDER 2 2 2 2 2;
                  RUN;

                  PROC PRINTTO;
                  RUN;

                  *check for errors and/or warnings;
                  DATA e_log5;
                  INFILE el dlm='10'X dsd;
                  LENGTH message $256;
                  INPUT message;
                  IF FIND(message, "ERROR: Singular convergence")~=0 THEN
issue = 1;
                  ELSE IF FIND(message,"WARNING: Unable to calculate standard
errors")~=0 THEN issue = 2;
                  ELSE IF FIND (message,"ERROR: Floating Point Overflow")~=0
THEN issue = 3;
                  ELSE IF FIND (message,"ERROR: Floating Point Zero
Divide")~=0 THEN issue = 4;
                  ELSE IF FIND (message,"WORK.SSCPTTZ has 0 observations and
0 variables")~=0 THEN issue = 5;
                  ELSE issue = 0;
                  IF (issue = 0 & (FIND(message,
"ERROR")~=0|FIND(message,"WARNING")~=0)) THEN issue = 6;
                  DROP message;
                  RUN;

                  PROC SORT DATA=e_log5 nodupkey;
                  BY issue;
                  RUN;

                  DATA e_log5;
                  SET e_log5;
                  flag = 1;
                  RUN;

                  PROC TRANSPOSE DATA=e_log5 OUT=e_log5 PREFIX=issue;
                  ID issue;
                  VAR flag;
                  RUN;

                  DATA _null_;
                  dset=open('e_log5');
                  CALL SYMPUT ('chk1',varnum(dset,'issue1'));
                  CALL SYMPUT ('chk2',varnum(dset,'issue2'));
                  CALL SYMPUT ('chk3',varnum(dset,'issue3'));
                  CALL SYMPUT ('chk4',varnum(dset,'issue4'));
```

```
CALL SYMPUT ('chk5',varnum(dset,'issue5'));
CALL SYMPUT ('chk6',varnum(dset,'issue6'));
RUN;

DATA e_log5;
RETAIN issue1 issue2 issue3 issue4 issue5 issue6;
SET e_log5;
IF &chk1 = 0 THEN issue1 = 0;
IF &chk2 = 0 THEN issue2 = 0;
IF &chk3 = 0 THEN issue3 = 0;
IF &chk4 = 0 THEN issue4 = 0;
IF &chk5 = 0 THEN issue5 = 0;
IF &chk6 = 0 THEN issue6 = 0;
DROP _NAME_ issue0;
RUN;
*end check for errors and/or warnings;

*Compute entropy and add to oe dataset;
DATA of5;
SET of5;
IF grp1prb = 0 THEN pp1 = 0; ELSE pp1 =
grp1prb*log(grp1prb);
IF grp2prb = 0 THEN pp2 = 0; ELSE pp2 =
grp2prb*log(grp2prb);
IF grp3prb = 0 THEN pp3 = 0; ELSE pp3 =
grp3prb*log(grp3prb);
IF grp4prb = 0 THEN pp4 = 0; ELSE pp4 =
grp4prb*log(grp4prb);
IF grp5prb = 0 THEN pp5 = 0; ELSE pp5 =
grp5prb*log(grp5prb);
pp_j = pp1+pp2+pp3+pp4+pp5;
RUN;

PROC MEANS DATA=of5 NOPRINT;
VAR pp_j;
OUTPUT SUM=pp_sum OUT=fuzzy5;
RUN;

DATA fuzzy5;
SET fuzzy5;
DROP _TYPE_ _FREQ_;
RUN;

DATA oe5;
SET oe5;
IF _N_ = 1 THEN SET fuzzy5;
IF _N_ = 1 THEN SET ll_oneclass;
entropy = -1*pp_sum;
DROP pp_sum;
RUN;
*End compute entropy;

*Create dataset with enumeration measures;
DATA em5;
SET oe5;
IF _TYPE_ ~= 'PARMS' THEN DELETE;
```

```
                  DROP _MODEL_ _MODEL2_ _TYPE_ _NAME_ INTERC1-INTERC5
LINEAR1-LINEAR5 QUADRA1-QUADRA5 SIGMA1 THETA2-THETA5 _CONVERGE_;
                  AIC = -2*_LOGLIK_ + 2*&df5;
                  BIC = -2*_LOGLIK_ + log(&n)*&df5;
                  CAIC = -2*_LOGLIK_ + (log(&n)+1)*&df5;
                  HQ = -2*_LOGLIK_ + log(log(&n))*&df5;
                  ssBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df5;
                  ssCAIC = -2*_LOGLIK_ + (log((&n+2)/24)+1)*&df5;
                  ssHQ = -2*_LOGLIK_ + log(log((&n+2)/24))*&df5;
                  sE = 1 - (entropy/(&n*log(&s3)));
                  NEC = entropy/(_LOGLIK_ - c1loglik);
                  CLC = -2*_LOGLIK_ + 2*entropy;
                  ICLBIC = -2*_LOGLIK_ + log(&n)*&df5 + 2*entropy;
                  ssICLBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df5 + 2*entropy;
                  BIC2 = -2*_LOGLIK_ + log(&nobs)*&df5;
                  CAIC2 = -2*_LOGLIK_ + (log(&nobs)+1)*&df5;
                  HQ2 = -2*_LOGLIK_ + log(log(&nobs))*&df5;
                  ssBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df5;
                  ssCAIC2 = -2*_LOGLIK_ + (log((&nobs+2)/24)+1)*&df5;
                  ssHQ2 = -2*_LOGLIK_ + log(log((&nobs+2)/24))*&df5;
                  ICLBIC2 = -2*_LOGLIK_ + log(&nobs)*&df5 + 2*entropy;
                  ssICLBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df5 +
2*entropy;
                  logB10 = &bic4-BIC;
                  B10 = exp((&bic4-BIC)/2);
                  RUN;
                  *End create dataset with enumeration measures;

                  *Compute average posterior probabilities;
                  PROC SORT DATA=of5;
                  BY group;
                  RUN;

                  PROC MEANS DATA=of5 NOPRINT;
                  VAR GRP1PRB GRP2PRB GRP3PRB GRP4PRB GRP5PRB;
                  BY group;
                  OUTPUT OUT=desc5;
                  RUN;

                  DATA desc5;
                  SET desc5;
                  IF _stat_ ~= 'MEAN' THEN DELETE;
                  IF group = 1 THEN avgpp = GRP1PRB;
                  ELSE IF group = 2 THEN avgpp = GRP2PRB;
                  ELSE IF group = 3 THEN avgpp = GRP3PRB;
                  ELSE IF group = 4 THEN avgpp = GRP4PRB;
                  ELSE IF group = 5 THEN avgpp = GRP5PRB;
                  DROP _TYPE_ _FREQ_ _STAT_ GRP1PRB GRP2PRB GRP3PRB GRP4PRB
GRP5PRB group;

                  RUN;

                  PROC SORT DATA=desc5;
                  BY DESCENDING avgpp;
                  RUN;

                  DATA desc5;
                  SET desc5;
```

```
    id = _N_;
    RUN;

    PROC TRANSPOSE DATA=desc5 OUT=pp_means5 PREFIX=PP;
    ID id;
    VAR avgpp;
    RUN;

    DATA pp_means5;
    SET pp_means5;
    DROP _NAME_;
    RUN;
    *End compute average posterior probabilities;
    /*****END 5-CLASS MODEL AND OUTPUT*****/

    /*****COMPUTE VLMR AND AVLMR P-VALUES USING MPLUS*****/
    DATA _null_;
    SET of4;
    FILE "&dir\trajdata.dat";
    PUT y1-y&obs;
    RUN;

    %DO c = 4 %TO 5;
  DATA _NULL_;
SET oe&c;
IF _TYPE_ ~= 'PARMS' THEN DELETE;
FILE "&dir\getvlmr.inp" LRECL=1500;
PUT "TITLE:";
PUT "LMR for GBTM";
PUT "DATA:";
PUT " FILE IS &dir\trajdata.dat;";
PUT "VARIABLE:";
PUT " NAMES ARE t1-t&obs;";
PUT " USEVARIABLES ARE t1-t&obs;";
PUT " CLASSES = c(&c);";
PUT " MISSING = .;";
PUT "ANALYSIS:";
PUT " TYPE = mixture;";
PUT " STARTS = 100 10;";
PUT "MODEL:";
PUT ' %OVERALL%';
IF &obs = 5 THEN PUT " i s q | t1@0 t2@1 t3@2 t4@3 t5@4;";
ELSE PUT " i s q | t1@0 t2@0.5 t3@1 t4@1.5 t5@2 t6@2.5 t7@3
t8@3.5 t9@4;";
PUT " i with s@0;";
PUT " i with q@0;";
PUT " s with q@0;";
PUT " i@0;";
PUT " s@0;";
PUT " q@0;";
PUT " t1 (1);";
PUT " t2 (1);";
PUT " t3 (1);";
PUT " t4 (1);";
PUT " t5 (1);";
IF &obs = 9 THEN DO;
  PUT " t6 (1);";
```

```
        PUT " t7 (1);";
        PUT " t8 (1);";
        PUT " t9 (1);";
      END;
      PUT '%c#1%';
      PUT "[i*" INTERC1 " s*" LINEAR1 " q*" QUADRA1"];";
      PUT '%c#2%';
      PUT "[i*" INTERC2 " s*" LINEAR2 " q*" QUADRA2"];";
      PUT '%c#3%';
      PUT "[i*" INTERC3 " s*" LINEAR3 " q*" QUADRA3"];";
      PUT '%c#4%';
      PUT "[i*" INTERC4 " s*" LINEAR4 " q*" QUADRA4"];";
      IF &c = 5 THEN DO;
        PUT '%c#5%';
        PUT "[i*" INTERC5 " s*" LINEAR5 " q*" QUADRA5"];";
      END;
      PUT "OUTPUT:";
      PUT " TECH11;";
      PUT "SAVEDATA:";
      PUT " RESULTS ARE &dir\trajresult.dat;";
      RUN;

      OPTIONS NOXWAIT;
      X CALL "c:\Program Files (x86)\Mplus\mplus.exe"
        "&dir\getvlmr.inp"
        "&dir\getvlmr.out";

      DATA vlmr&c;
      INFILE "&dir\trajresult.dat";
      INPUT #5 vlmr 130-143 vlmr_df 146-159 #6 vlmr_m 2-15 vlmr_sd
18-31 vlmr_p 34-47 avlmr 50-63 avlmr_p 66-79;
      RUN;
    %END;
    /*****END COMPUTE VLMR AND AVLMR P-VALUE USING MPLUS*****/

    /*BUILD OUTPUT DATASETS*/
    DATA em3;
    MERGE e_log3 em3;
    n = &i;
    obs = &j;
    mix = &k;
    miss = &l;
    sigma = &m;
    rep = &rep;
    class = &s1;
    RUN;

    DATA oe3se;
    SET oe3;
    IF _TYPE_ ~= 'STDERR' THEN DELETE;
    DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ SIGMA1 THETA2-THETA3;
    RENAME interc1-interc3 = ise1-ise3;
    RENAME linear1-linear3 = lse1-lse3;
    RENAME quadra1-quadra3 = qse1-qse3;
    RUN;
```

```
            DATA oe3;
            RETAIN n obs miss mix sigma rep class INTERC1 LINEAR1 QUADRA1
INTERC2 LINEAR2 QUADRA2 INTERC3 LINEAR3 QUADRA3
                   INTERC4 LINEAR4 QUADRA4 INTERC5 LINEAR5 QUADRA5 SIGMA1
THETA2-THETA5 ISE1 LSE1 QSE1 ISE2 LSE2 QSE2
                   ISE3 LSE3 QSE3 ISE4 LSE4 QSE4 ISE5 LSE5 QSE5 PP1-PP5;
            SET oe3;
            IF _N_ = 1 THEN SET pp_means3;
            IF _N_ = 1 THEN SET oe3se;
            IF _TYPE_ ~= 'PARMS' THEN DELETE;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s1;
            INTERC4 = .;
            LINEAR4 = .;
            QUADRA4 = .;
            INTERC5 = .;
            LINEAR5 = .;
            QUADRA5 = .;
            THETA4 = .;
            THETA5 = .;
            ISE4 = .;
            LSE4 = .;
            QSE4 = .;
            ISE5 = .;
            LSE5 = .;
            QSE5 = .;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ c1loglik entropy c1_issue1-c1_issue5;
            RUN;

            DATA em4;
            MERGE e_log4 em4 vlmr4;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s2;
            RUN;

            DATA oe4se;
            SET oe4;
            IF _TYPE_ ~= 'STDERR' THEN DELETE;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ SIGMA1 THETA2-THETA4;
            RENAME interc1-interc4 = ise1-ise4;
            RENAME linear1-linear4 = lse1-lse4;
            RENAME quadra1-quadra4 = qse1-qse4;
            RUN;

            DATA oe4;
```

```
            RETAIN n obs miss mix sigma rep class INTERC1 LINEAR1 QUADRA1
INTERC2 LINEAR2 QUADRA2 INTERC3 LINEAR3 QUADRA3
                   INTERC4 LINEAR4 QUADRA4 INTERC5 LINEAR5 QUADRA5 SIGMA1
THETA2-THETA5 ISE1 LSE1 QSE1 ISE2 LSE2 QSE2
                   ISE3 LSE3 QSE3 ISE4 LSE4 QSE4 ISE5 LSE5 QSE5 PP1-PP5;
            SET oe4;
            IF _N_ = 1 THEN SET pp_means4;
            IF _TYPE_ ~= 'PARMS' THEN DELETE;
            IF _N_ = 1 THEN SET oe4se;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s2;
            INTERC5 = .;
            LINEAR5 = .;
            QUADRA5 = .;
            THETA5 = .;
            ISE5 = .;
            LSE5 = .;
            QSE5 = .;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ c1loglik entropy c1_issue1-c1_issue5;
            RUN;

            DATA em5;
            MERGE e_log5 em5 vlmr5;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s3;
            RUN;

            DATA oe5se;
            SET oe5;
            IF _TYPE_ ~= 'STDERR' THEN DELETE;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ SIGMA1 THETA2-THETA5;
            RENAME interc1-interc5 = ise1-ise5;
            RENAME linear1-linear5 = lse1-lse5;
            RENAME quadra1-quadra5 = qse1-qse5;
            RUN;

            DATA oe5;
            RETAIN n obs miss mix sigma rep class INTERC1 LINEAR1 QUADRA1
INTERC2 LINEAR2 QUADRA2 INTERC3 LINEAR3 QUADRA3
                   INTERC4 LINEAR4 QUADRA4 INTERC5 LINEAR5 QUADRA5 SIGMA1
THETA2-THETA5 ISE1 LSE1 QSE1 ISE2 LSE2 QSE2
                   ISE3 LSE3 QSE3 ISE4 LSE4 QSE4 ISE5 LSE5 QSE5 PP1-PP5;
            SET oe5;
            IF _N_ = 1 THEN SET pp_means5;
            IF _TYPE_ ~= 'PARMS' THEN DELETE;
```

```
            IF _N_ = 1 THEN SET oe5se;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s3;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ c1loglik entropy c1_issue1-c1_issue5;
            RUN;

            DATA outdata_measures;
            SET em3 em4 em5;
            RUN;

            DATA outdata_params;
            SET oe3 oe4 oe5;
            RUN;
            /*****END BUILD OUTPUT DATASETS*****/

                  /*****WRITE OUTPUT DATASETS TO A TEXT FILE*****/
                  DATA _null_;
            SET outdata_measures;
            FILE "&dir\enumeration_measures.txt" DLM=',' MOD LRECL=1500;
            PUT n obs mix miss sigma rep class issue1-issue6 c1_issue1-
c1_issue6 _LOGLIK_ _BIC1_ _BIC2_ _AIC_
                      c1loglik aic bic caic hq ssbic sscaic sshq entropy se
nec clc iclbic ssiclbic bic2 caic2 hq2
                        ssbic2 sscaic2 sshq2 iclbic2 ssiclbic2 logb10 b10
vlmr vlmr_df vlmr_m vlmr_sd vlmr_p avlmr avlmr_p;
            RUN;

                  DATA _null_;
                  SET outdata_params;
                  FILE "&dir\estimates.txt" DLM=',' MOD LRECL=1500;
                  PUT n obs mix miss sigma rep class interc1 linear1 quadra1
interc2 linear2 quadra2 interc3 linear3 quadra3
                      interc4 linear4 quadra4 interc5 linear5 quadra5 ise1
lse1 qse1 ise2 lse2 qse2 ise3 lse3 qse3
                        ise4 lse4 qse4 ise5 lse5 qse5 sigma1 theta2-theta5
pp1-pp5;
                  RUN;
            /*****END WRITE OUTPUT DATASETS TO A TEXT FILE*****/

                  /*****CLEANUP AFTER YOURSELF*****/
            PROC DATASETS LIBRARY=work NOLIST;
            DELETE desc3-desc5 em3-em5 e_log1 e_log3-e_log5 fuzzy3-fuzzy5
ll_oneclass
                   oe1 oe3-oe5 oe3se oe4se oe5se of1 of3-of5 op1 op3-op5 os1
os3-os5
                   pp_means3-pp_means5 trajdata_n vlmr4 vlmr5 sscpttz
outdata_measures outdata_params;
            RUN;

                DM "out;clear;log;clear;";
                  /*****END CLEANUP AFTER YOURSELF*****/
```

```
                  %LET rep = &rep+1;
              %END;
              %LET m = &m+1;
              %LET rep = 1;
        %END;
            %LET l = &l+1;
            %LET m = 1;
            %LET rep = 1;
      %END;
        %LET k = &k+1;
        %LET l = 1;
        %LET m = 1;
        %LET rep = 1;
    %END;
      %LET j = &j+1;
      %LET k = 1;
      %LET l = 1;
      %LET m = 1;
      %LET rep = 1;
  %END;
  %LET i = &i+1;
  %LET j = 1;
  %LET k = 1;
  %LET l = 1;
  %LET m = 1;
  %LET rep = 1;
%END;

%MEND;

%diss_cnorm;
```

APPENDIX B

BINARY LOGIT MODEL DATA GENERATION SYNTAX

```
%MACRO diss_binary;
%let dir=c:\tempsas\Dissertation\work_binary;
%let df3 = 11;
%let df4 = 15;
%let df5 = 19;
%let s1 = 3;
%let s2 = 4;
%let s3 = 5;
%LET mrep = 50;

DATA condition;
INFILE "&dir\last_rep.txt" DELIMITER=',' end=last;
INPUT n obs mix miss sigma rep;
CALL SYMPUT ('i',n);
CALL SYMPUT ('j',obs);
CALL SYMPUT ('k',mix);
CALL SYMPUT ('l',miss);
CALL SYMPUT ('m',sigma);
CALL SYMPUT ('rep',rep);
IF last;
RUN;

%IF &rep = &mrep %THEN %DO;
  %IF &m < 2 %THEN %DO;
    %LET m = &m+1;
      %LET rep = 1;
  %END;
  %ELSE %DO;
    %IF &l < 3 %THEN %DO;
        %LET l = &l+1;
        %LET m = 1;
        %LET rep = 1;
      %END;
    %ELSE %DO;
        %IF &k < 3 %THEN %DO;
          %LET k = &k+1;
            %LET l = 1;
```

```
            %LET m = 1;
            %LET rep = 1;
        %END;
      %ELSE %DO;
          %IF &j < 2 %THEN %DO;
            %LET j = &j+1;
            %LET k = 1;
            %LET l = 1;
            %LET m = 1;
            %LET rep = 1;
          %END;
          %ELSE %DO;
            %IF &i < 2 %THEN %DO;
              %LET i = &i + 1;
              %LET j = 1;
              %LET k = 1;
              %LET l = 1;
              %LET m = 1;
                %LET rep = 1;
            %END;
            %ELSE %DO;
              %LET rep = &rep+1;
            %END;
          %END;
      %END;
    %END;
  %END;
%END;
%ELSE %DO;
  %LET rep = &rep+1;
%END;

%DO %WHILE (&i <= 4);
  %IF &i = 1 %THEN %LET n = 200;
  %IF &i = 2 %THEN %LET n = 400;
  %IF &i = 3 %THEN %LET n = 800;
  %IF &i = 4 %THEN %LET n = 2000;

  %DO %WHILE (&j <= 2);
    %IF &j = 1 %THEN %LET obs = 5;
      %IF &j = 2 %THEN %LET obs = 9;

      %DO %WHILE (&k <= 3);
        %IF &k = 1 %THEN %LET mix = {.25 .25 .25 .25};
        %IF &k = 2 %THEN %LET mix = {.4 .25 .2 .15};
        %IF &k = 3 %THEN %LET mix = {.6 .25 .1 .05};

      %DO %WHILE (&l <= 3);
          %IF &l = 1 %THEN %LET miss = 0;
        %IF &l = 2 %THEN %LET miss = .1;
        %IF &l = 3 %THEN %LET miss = .2;

            %DO %WHILE (&m <= 2);
              %IF &m = 1 %THEN %LET sigma = 1;
              %IF &m = 2 %THEN %LET sigma = 2;

          %DO %WHILE (&rep <= &mrep);
```

```
DATA _NULL_;
SET condition;
FILE "&dir\reps.txt" DLM = ',' MOD;
n = &i;
obs = &j;
mix = &k;
miss = &l;
sigma = &m;
rep = &rep;
PUT n obs mix miss sigma rep;
RUN;

DATA _NULL_;
SET condition;
FILE "&dir\last_rep.txt" DLM = ',';
n = &i;
obs = &j;
mix = &k;
miss = &l;
sigma = &m;
rep = &rep;
PUT n obs mix miss sigma rep;
RUN;


PROC IML;
      seed = &i*100000000 + &j*10000000 + &k*1000000 + &l*100000
+ &m*10000 + &rep;
      CALL RANDSEED(seed);
      nobs = 0;
      obs = &obs;
      sigma = &sigma;
      n = &n;
      miss = &miss;
      mix = &mix;

      beta = {-2.143 2.912 -.337 .337, .388 -.279 1.191 -1.191, -
.091 .02 -.156 .156};
      IF obs = 5 THEN t = {1 0 0, 1 1 1, 1 2 4, 1 3 9, 1 4 16};
      ELSE t = {1 0 0, 1 .5 .25, 1 1 1, 1 1.5 2.25, 1 2 4, 1 2.5
6.25, 1 3 9, 1 3.5 12.25, 1 4 16};
      pmx = J(n,obs,.);
      CALL RANDGEN(pmx,'UNIFORM');
      n1 = round(n*mix[1]);
      n2 = round(n*mix[2]);
      n3 = round(n*mix[3]);
      n4 = round(n*mix[4]);
      t1 = repeat(t,n1);
      t2 = repeat(t,n2);
      t3 = repeat(t,n3);
      t4 = repeat(t,n4);
      e1 = J(n1*obs,1,.);
      e2 = J(n2*obs,1,.);
      e3 = J(n3*obs,1,.);
      e4 = J(n4*obs,1,.);
      CALL RANDGEN(e1,'NORMAL',0,sigma);
      CALL RANDGEN(e2,'NORMAL',0,sigma);
```

```
                    CALL RANDGEN(e3,'NORMAL',0,sigma);
                    CALL RANDGEN(e4,'NORMAL',0,sigma);
                    y1 = t1*beta[,1] + e1;
                    y2 = t2*beta[,2] + e2;
                    y3 = t3*beta[,3] + e3;
                    y4 = t4*beta[,4] + e4;

                    y = y1//y2//y3//y4;

                    gn1 = repeat(1,n1);
                    gn2 = repeat(2,n2);
                    gn3 = repeat(3,n3);
                    gn4 = repeat(4,n4);

                    gn = gn1//gn2//gn3//gn4;
                    id = shape(1:n,n,1);

                    yw = shape(y,n,obs);
                    ybw = J(n,obs,.);

                    DO o=1 TO n;
                      DO p=1 TO obs;
                        IF yw[o,p] > 0 THEN ybw[o,p] = 1; ELSE ybw[o,p] = 0;
                          IF (p ^= 1 & miss ^= 0) THEN DO;
                          IF(pmx[o,p] <= miss | missing(ybw[o,p-1])=1) THEN
ybw[o,p] = .; ELSE nobs = nobs+1;
                          END;
                          ELSE DO; nobs = nobs+1;
                    END;
                      END;
                    END;
                    CALL SYMPUT ('nobs',LEFT(CHAR(nobs)));
                    IF obs = 5 THEN DO;
                      ybw =
id||gn||J(n,1,0)||J(n,1,1)||J(n,1,2)||J(n,1,3)||J(n,1,4)||ybw;
                      CREATE trajdata_b FROM ybw
[colname={id,gn,t1,t2,t3,t4,t5,y1,y2,y3,y4,y5}];
                      APPEND FROM ybw;
                    END;
                    ELSE IF obs=9 THEN DO;
                      ybw =
id||gn||J(n,1,0)||J(n,1,0.5)||J(n,1,1)||J(n,1,1.5)||J(n,1,2)||J(n,1,2.5)||J(n
,1,3)||J(n,1,3.5)||J(n,1,4)||ybw;
                      CREATE trajdata_b FROM ybw
[colname={id,gn,t1,t2,t3,t4,t5,t6,t7,t8,t9,y1,y2,y3,y4,y5,y6,y7,y8,y9}];
                      APPEND FROM ybw;
                    END;
                    QUIT;

                    FILENAME el
'C:\tempsas\Dissertation\work_binary\error_log.txt';

                    /*****1-CLASS MODEL TO OBTAIN LOG LIKELIHOOD FOR NEC******/
                    PROC PRINTTO LOG = el NEW;
                    RUN;
```

```
PROC TRAJ data=trajdata_b outplot = op1 outstat = os1 out =
of1 outest = oe1;
ID id gn;
VAR y1-y&obs;
INDEP t1-t&obs;
MODEL logit;
NGROUPS 1;
ORDER 2;
RUN;

PROC PRINTTO;
RUN;

*check for errors and/or warnings;
DATA e_log1;
INFILE el dlm='10'X dsd;
LENGTH message $256;
INPUT message;
IF FIND(message, "ERROR: Singular convergence")~=0 THEN
issue = 1;
ELSE IF FIND(message,"WARNING: Unable to calculate standard
errors")~=0 THEN issue = 2;
ELSE IF FIND (message,"ERROR: Floating Point Overflow")~=0
THEN issue = 3;
ELSE IF FIND (message,"ERROR: Floating Point Zero
Divide")~=0 THEN issue = 4;
ELSE IF FIND (message,"WORK.SSCPTTZ has 0 observations and
0 variables")~=0 THEN issue = 5;
ELSE issue = 0;
IF (issue = 0 & (FIND(message,
"ERROR")~=0|FIND(message,"WARNING")~=0)) THEN issue = 6;
DROP message;
RUN;

PROC SORT DATA=e_log1 nodupkey;
BY issue;
RUN;

DATA e_log1;
SET e_log1;
flag = 1;
RUN;

PROC TRANSPOSE DATA=e_log1 OUT=e_log1 PREFIX=c1_issue;
ID issue;
VAR flag;
RUN;

DATA _null_;
dset=open('e_log1');
CALL SYMPUT ('chk1',varnum(dset,'c1_issue1'));
CALL SYMPUT ('chk2',varnum(dset,'c1_issue2'));
CALL SYMPUT ('chk3',varnum(dset,'c1_issue3'));
CALL SYMPUT ('chk4',varnum(dset,'c1_issue4'));
CALL SYMPUT ('chk5',varnum(dset,'c1_issue5'));
CALL SYMPUT ('chk6',varnum(dset,'c1_issue6'));
RUN;
```

```
                    DATA e_log1;
                    SET e_log1;
                    IF &chk1 = 0 THEN c1_issue1 = 0;
                    IF &chk2 = 0 THEN c1_issue2 = 0;
                    IF &chk3 = 0 THEN c1_issue3 = 0;
                    IF &chk4 = 0 THEN c1_issue4 = 0;
                    IF &chk5 = 0 THEN c1_issue5 = 0;
                    IF &chk6 = 0 THEN c1_issue6 = 0;
                    DROP _NAME_ c1_issue0;
                    RUN;
                    *end check for errors and/or warnings;

                    DATA ll_oneclass;
                    RETAIN c1_issue1 c1_issue2 c1_issue3 c1_issue4 c1_issue5
c1_issue6;
                    SET oe1;
                    IF _N_ = 1 THEN SET e_log1;
                    IF _TYPE_ ~= 'PARMS' THEN DELETE;
                    DROP _MODEL_ _MODEL2_ _TYPE_ _NAME_ INTERC1 LINEAR1 QUADRA1
                         SIGMA1 _AIC_ _BIC1_ _BIC2_ _CONVERGE_;
                    RENAME _loglik_ = c1loglik;
                    RUN;
                    /*****END 1-CLASS MODEL TO OBTAIN LOG LIKELIHOOD FOR
NEC*****/


                    /*****3-CLASS MODEL AND OUTPUT*****/
                    PROC PRINTTO LOG = el NEW;
                    RUN;

                    PROC TRAJ data = trajdata_b outplot = op3 outstat = os3 out
= of3 outest = oe3;
                    ID id gn;
                    VAR y1-y&obs;
                    indep t1-t&obs;
                    MODEL logit;
                    NGROUPS 3;
                    ORDER 2 2 2;
                    RUN;

                    PROC PRINTTO;
                    RUN;

                    *check for errors and/or warnings;
                    DATA e_log3;
                    INFILE el dlm='10'X dsd;
                    LENGTH message $256;
                    INPUT message;
                    IF FIND(message, "ERROR: Singular convergence")~=0 THEN
issue = 1;
                    ELSE IF FIND(message,"WARNING: Unable to calculate standard
errors")~=0 THEN issue = 2;
                    ELSE IF FIND (message,"ERROR: Floating Point Overflow")~=0
THEN issue = 3;
                    ELSE IF FIND (message,"ERROR: Floating Point Zero
Divide")~=0 THEN issue = 4;
```

```
                 ELSE IF FIND (message,"WORK.SSCPTTZ has 0 observations and
0 variables")~=0 THEN issue = 5;
                 ELSE issue = 0;
                 IF (issue = 0 & (FIND(message,
"ERROR")~=0|FIND(message,"WARNING")~=0)) THEN issue = 6;
                 DROP message;
                 RUN;

               PROC SORT DATA=e_log3 nodupkey;
                 BY issue;
                 RUN;

                 DATA e_log3;
                 SET e_log3;
                 flag = 1;
                 RUN;

                 PROC TRANSPOSE DATA=e_log3 OUT=e_log3 PREFIX=issue;
                 ID issue;
                 VAR flag;
                 RUN;

                 DATA _null_;
                 dset=open('e_log3');
                 CALL SYMPUT ('chk1',varnum(dset,'issue1'));
                 CALL SYMPUT ('chk2',varnum(dset,'issue2'));
                 CALL SYMPUT ('chk3',varnum(dset,'issue3'));
                 CALL SYMPUT ('chk4',varnum(dset,'issue4'));
                 CALL SYMPUT ('chk5',varnum(dset,'issue5'));
                 CALL SYMPUT ('chk6',varnum(dset,'issue6'));
                 RUN;

                 DATA e_log3;
                 RETAIN issue1 issue2 issue3 issue4 issue5 issue6;
                 SET e_log3;
                 IF &chk1 = 0 THEN issue1 = 0;
                 IF &chk2 = 0 THEN issue2 = 0;
                 IF &chk3 = 0 THEN issue3 = 0;
                 IF &chk4 = 0 THEN issue4 = 0;
                 IF &chk5 = 0 THEN issue5 = 0;
                 IF &chk6 = 0 THEN issue6 = 0;
                 DROP _NAME_ issue0;
                 RUN;
                 *end check for errors and/or warnings;

                 *Compute entropy and add to oe dataset;
                 DATA of3;
            SET of3;
            IF grp1prb = 0 THEN pp1 = 0; ELSE pp1 = grp1prb*log(grp1prb);
                 IF grp2prb = 0 THEN pp2 = 0; ELSE pp2 =
grp2prb*log(grp2prb);
            IF grp3prb = 0 THEN pp3 = 0; ELSE pp3 = grp3prb*log(grp3prb);
            pp_j = pp1+pp2+pp3;
            RUN;

                 PROC MEANS DATA=of3 NOPRINT;
                 VAR pp_j;
```

```
OUTPUT SUM=pp_sum OUT=fuzzy3;
RUN;

DATA fuzzy3;
SET fuzzy3;
DROP _TYPE_ _FREQ_;
RUN;

DATA oe3;
SET oe3;
IF _N_ = 1 THEN SET fuzzy3;
IF _N_ = 1 THEN SET ll_oneclass;
entropy = -1*pp_sum;
DROP pp_sum;
RUN;
*End compute entropy;

*Create dataset with enumeration measures;
DATA em3;
SET oe3;
IF _TYPE_ ~= 'PARMS' THEN DELETE;
DROP _MODEL_ _MODEL2_ _TYPE_ _NAME_ INTERC1-INTERC3
LINEAR1-LINEAR3 QUADRA1-QUADRA3 SIGMA1 THETA2-THETA3 _CONVERGE_;
AIC = -2*_LOGLIK_ + 2*&df3;
BIC = -2*_LOGLIK_ + log(&n)*&df3;
CAIC = -2*_LOGLIK_ + (log(&n)+1)*&df3;
HQ = -2*_LOGLIK_ + log(log(&n))*&df3;
ssBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df3;
ssCAIC = -2*_LOGLIK_ + (log((&n+2)/24)+1)*&df3;
ssHQ = -2*_LOGLIK_ + log(log((&n+2)/24))*&df3;
sE = 1 - (entropy/(&n*log(&s1)));
NEC = entropy/(_LOGLIK_ - c1loglik);
CLC = -2*_LOGLIK_ + 2*entropy;
ICLBIC = -2*_LOGLIK_ + log(&n)*&df3 + 2*entropy;
ssICLBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df3 + 2*entropy;
BIC2 = -2*_LOGLIK_ + log(&nobs)*&df3;
CAIC2 = -2*_LOGLIK_ + (log(&nobs)+1)*&df3;
HQ2 = -2*_LOGLIK_ + log(log(&nobs))*&df3;
ssBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df3;
ssCAIC2 = -2*_LOGLIK_ + (log((&nobs+2)/24)+1)*&df3;
ssHQ2 = -2*_LOGLIK_ + log(log((&nobs+2)/24))*&df3;
ICLBIC2 = -2*_LOGLIK_ + log(&nobs)*&df3 + 2*entropy;
ssICLBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df3 +
2*entropy;
CALL SYMPUT ('bic3',BIC);
RUN;
*End create dataset with enumeration measures;

*Compute average posterior probabilities;
PROC SORT DATA=of3;
BY group;
RUN;

PROC MEANS DATA=of3 NOPRINT;
VAR GRP1PRB GRP2PRB GRP3PRB;
BY group;
OUTPUT OUT=desc3;
```

```
RUN;

DATA desc3;
SET desc3;
IF _stat_ ~= 'MEAN' THEN DELETE;
IF group = 1 THEN avgpp = GRP1PRB;
ELSE IF group = 2 THEN avgpp = GRP2PRB;
ELSE IF group = 3 THEN avgpp = GRP3PRB;
DROP _TYPE_ _FREQ_ _STAT_ GRP1PRB GRP2PRB GRP3PRB group;
RUN;

PROC SORT DATA=desc3;
BY DESCENDING avgpp;
RUN;

DATA desc3;
SET desc3;
id = _N_;
RUN;

PROC TRANSPOSE DATA=desc3 OUT=pp_means3 PREFIX=PP;
ID id;
VAR avgpp;
RUN;

DATA pp_means3;
SET pp_means3;
PP4 = .;
PP5 = .;
DROP _NAME_;
RUN;
*End compute average posterior probabilities;
/*****END 3-CLASS MODEL AND OUTPUT*****/

/*****4-CLASS MODEL AND OUTPUT*****/
PROC PRINTTO LOG = el NEW;
RUN;

PROC TRAJ data = trajdata_b outplot = op4 outstat = os4 out
= of4 outest = oe4;
ID id gn;
VAR y1-y&obs;
indep t1-t&obs;
MODEL logit;
NGROUPS 4;
ORDER 2 2 2 2;
RUN;

PROC PRINTTO;
RUN;

*check for errors and/or warnings;
DATA e_log4;
INFILE el dlm='10'X dsd;
LENGTH message $256;
INPUT message;
```

```
                  IF FIND(message, "ERROR: Singular convergence")~=0 THEN
issue = 1;
                  ELSE IF FIND(message,"WARNING: Unable to calculate standard
errors")~=0 THEN issue = 2;
                  ELSE IF FIND (message,"ERROR: Floating Point Overflow")~=0
THEN issue = 3;
                  ELSE IF FIND (message,"ERROR: Floating Point Zero
Divide")~=0 THEN issue = 4;
                  ELSE IF FIND (message,"WORK.SSCPTTZ has 0 observations and
0 variables")~=0 THEN issue = 5;
                  ELSE issue = 0;
                  IF (issue = 0 & (FIND(message,
"ERROR")~=0|FIND(message,"WARNING")~=0)) THEN issue = 6;
                  DROP message;
                  RUN;

                  PROC SORT DATA=e_log4 nodupkey;
                  BY issue;
                  RUN;

                  DATA e_log4;
                  SET e_log4;
                  flag = 1;
                  RUN;

                  PROC TRANSPOSE DATA=e_log4 OUT=e_log4 PREFIX=issue;
                  ID issue;
                  VAR flag;
                  RUN;

                  DATA _null_;
                  dset=open('e_log4');
                  CALL SYMPUT ('chk1',varnum(dset,'issue1'));
                  CALL SYMPUT ('chk2',varnum(dset,'issue2'));
                  CALL SYMPUT ('chk3',varnum(dset,'issue3'));
                  CALL SYMPUT ('chk4',varnum(dset,'issue4'));
                  CALL SYMPUT ('chk5',varnum(dset,'issue5'));
                  CALL SYMPUT ('chk6',varnum(dset,'issue6'));
                  RUN;

                  DATA e_log4;
                  SET e_log4;
                  RETAIN issue1 issue2 issue3 issue4 issue5 issue6;
                  IF &chk1 = 0 THEN issue1 = 0;
                  IF &chk2 = 0 THEN issue2 = 0;
                  IF &chk3 = 0 THEN issue3 = 0;
                  IF &chk4 = 0 THEN issue4 = 0;
                  IF &chk5 = 0 THEN issue5 = 0;
                  IF &chk6 = 0 THEN issue6 = 0;
                  DROP _NAME_ issue0;
                  RUN;
                  *end check for errors and/or warnings;

                  *Compute entropy and add to oe dataset;
                  DATA of4;
                  SET of4;
```

```
                  IF grp1prb = 0 THEN pp1 = 0; ELSE pp1 =
grp1prb*log(grp1prb);
                  IF grp2prb = 0 THEN pp2 = 0; ELSE pp2 =
grp2prb*log(grp2prb);
                  IF grp3prb = 0 THEN pp3 = 0; ELSE pp3 =
grp3prb*log(grp3prb);
                  IF grp4prb = 0 THEN pp4 = 0; ELSE pp4 =
grp4prb*log(grp4prb);
                  pp_j = pp1+pp2+pp3+pp4;
                  RUN;

                  PROC MEANS DATA=of4 NOPRINT;
                  VAR pp_j;
                  OUTPUT SUM=pp_sum OUT=fuzzy4;
                  RUN;

                  DATA fuzzy4;
                  SET fuzzy4;
                  DROP _TYPE_ _FREQ_;
                  RUN;

                  DATA oe4;
                  SET oe4;
                  IF _N_ = 1 THEN SET fuzzy4;
                  IF _N_ = 1 THEN SET ll_oneclass;
                  entropy = -1*pp_sum;
                  DROP pp_sum;
                  RUN;
                  *End compute entropy;

                  *Create dataset with enumeration measures;
                  DATA em4;
                  SET oe4;
                  IF _TYPE_ ~= 'PARMS' THEN DELETE;
                  DROP _MODEL_ _MODEL2_ _TYPE_ _NAME_ INTERC1-INTERC4
LINEAR1-LINEAR4 QUADRA1-QUADRA4 SIGMA1 THETA2-THETA4 _CONVERGE_;
                  AIC = -2*_LOGLIK_ + 2*&df4;
                  BIC = -2*_LOGLIK_ + log(&n)*&df4;
                  CAIC = -2*_LOGLIK_ + (log(&n)+1)*&df4;
                  HQ = -2*_LOGLIK_ + log(log(&n))*&df4;
                  ssBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df4;
                  ssCAIC = -2*_LOGLIK_ + (log((&n+2)/24)+1)*&df4;
                  ssHQ = -2*_LOGLIK_ + log(log((&n+2)/24))*&df4;
                  sE = 1 - (entropy/(&n*log(&s2)));
                  NEC = entropy/(_LOGLIK_ - c1loglik);
                  CLC = -2*_LOGLIK_ + 2*entropy;
                  ICLBIC = -2*_LOGLIK_ + log(&n)*&df4 + 2*entropy;
                  ssICLBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df4 + 2*entropy;
                  BIC2 = -2*_LOGLIK_ + log(&nobs)*&df4;
                  CAIC2 = -2*_LOGLIK_ + (log(&nobs)+1)*&df4;
                  HQ2 = -2*_LOGLIK_ + log(log(&nobs))*&df4;
                  ssBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df4;
                  ssCAIC2 = -2*_LOGLIK_ + (log((&nobs+2)/24)+1)*&df4;
                  ssHQ2 = -2*_LOGLIK_ + log(log((&nobs+2)/24))*&df4;
                  ICLBIC2 = -2*_LOGLIK_ + log(&nobs)*&df4 + 2*entropy;
                  ssICLBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df4 +
2*entropy;
```

```
                    logB10 = &bic3-BIC;
                    B10 = exp((&bic3-BIC)/2);
                    CALL SYMPUT ('bic4',BIC);
                    RUN;
                    *End create dataset with enumeration measures;

                    *Compute average posterior probabilities;
                    PROC SORT DATA=of4;
                    BY group;
                    RUN;

                    PROC MEANS DATA=of4 NOPRINT;
                    VAR GRP1PRB GRP2PRB GRP3PRB GRP4PRB;
                    BY group;
                    OUTPUT OUT=desc4;
                    RUN;

                    DATA desc4;
                    SET desc4;
                    IF _stat_ ~= 'MEAN' THEN DELETE;
                    IF group = 1 THEN avgpp = GRP1PRB;
                    ELSE IF group = 2 THEN avgpp = GRP2PRB;
                    ELSE IF group = 3 THEN avgpp = GRP3PRB;
                    ELSE IF group = 4 THEN avgpp = GRP4PRB;
                    DROP _TYPE_ _FREQ_ _STAT_ GRP1PRB GRP2PRB GRP3PRB GRP4PRB
group;
                    RUN;

                    PROC SORT DATA=desc4;
                    BY DESCENDING avgpp;
                    RUN;

                    DATA desc4;
                    SET desc4;
                    id = _N_;
                    RUN;

                    PROC TRANSPOSE DATA=desc4 OUT=pp_means4 PREFIX=PP;
                    ID id;
                    VAR avgpp;
                    RUN;

                    DATA pp_means4;
                    SET pp_means4;
                    pp5 = .;
                    DROP _NAME_;
                    RUN;
                    *End compute average posterior probabilities;
                    /*****END 4-CLASS MODEL AND OUTPUT*****/

                    /*5-CLASS MODEL AND OUTPUT*/
                    PROC PRINTTO LOG = el NEW;
                    RUN;

                    PROC TRAJ data = trajdata_b outplot = op5 outstat = os5 out
= of5 outest = oe5;
                    ID id gn;
```

```
                    VAR y1-y&obs;
                    indep t1-t&obs;
                    MODEL logit;
                    NGROUPS 5;
                    ORDER 2 2 2 2 2;
                    RUN;

                    PROC PRINTTO;
                    RUN;

                    *check for errors and/or warnings;
                    DATA e_log5;
                    INFILE el dlm='10'X dsd;
                    LENGTH message $256;
                    INPUT message;
                    IF FIND(message, "ERROR: Singular convergence")~=0 THEN
issue = 1;
                    ELSE IF FIND(message,"WARNING: Unable to calculate standard
errors")~=0 THEN issue = 2;
                    ELSE IF FIND (message,"ERROR: Floating Point Overflow")~=0
THEN issue = 3;
                    ELSE IF FIND (message,"ERROR: Floating Point Zero
Divide")~=0 THEN issue = 4;
                    ELSE IF FIND (message,"WORK.SSCPTTZ has 0 observations and
0 variables")~=0 THEN issue = 5;
                    ELSE issue = 0;
                    IF (issue = 0 & (FIND(message,
"ERROR")~=0|FIND(message,"WARNING")~=0)) THEN issue = 6;
                    DROP message;
                    RUN;

                    PROC SORT DATA=e_log5 nodupkey;
                    BY issue;
                    RUN;

                    DATA e_log5;
                    SET e_log5;
                    flag = 1;
                    RUN;

                    PROC TRANSPOSE DATA=e_log5 OUT=e_log5 PREFIX=issue;
                    ID issue;
                    VAR flag;
                    RUN;

                    DATA _null_;
                    dset=open('e_log5');
                    CALL SYMPUT ('chk1',varnum(dset,'issue1'));
                    CALL SYMPUT ('chk2',varnum(dset,'issue2'));
                    CALL SYMPUT ('chk3',varnum(dset,'issue3'));
                    CALL SYMPUT ('chk4',varnum(dset,'issue4'));
                    CALL SYMPUT ('chk5',varnum(dset,'issue5'));
                    CALL SYMPUT ('chk6',varnum(dset,'issue6'));
                    RUN;

                    DATA e_log5;
                    RETAIN issue1 issue2 issue3 issue4 issue5 issue6;
```

```
                  SET e_log5;
                  IF &chk1 = 0 THEN issue1 = 0;
                  IF &chk2 = 0 THEN issue2 = 0;
                  IF &chk3 = 0 THEN issue3 = 0;
                  IF &chk4 = 0 THEN issue4 = 0;
                  IF &chk5 = 0 THEN issue5 = 0;
                  IF &chk6 = 0 THEN issue6 = 0;
                  DROP _NAME_ issue0;
                  RUN;
                  *end check for errors and/or warnings;

                  *Compute entropy and add to oe dataset;
                  DATA of5;
                  SET of5;
                  IF grp1prb = 0 THEN pp1 = 0; ELSE pp1 =
grp1prb*log(grp1prb);
                  IF grp2prb = 0 THEN pp2 = 0; ELSE pp2 =
grp2prb*log(grp2prb);
                  IF grp3prb = 0 THEN pp3 = 0; ELSE pp3 =
grp3prb*log(grp3prb);
                  IF grp4prb = 0 THEN pp4 = 0; ELSE pp4 =
grp4prb*log(grp4prb);
                  IF grp5prb = 0 THEN pp5 = 0; ELSE pp5 =
grp5prb*log(grp5prb);
                  pp_j = pp1+pp2+pp3+pp4+pp5;
                  RUN;

                  PROC MEANS DATA=of5 NOPRINT;
                  VAR pp_j;
                  OUTPUT SUM=pp_sum OUT=fuzzy5;
                  RUN;

                  DATA fuzzy5;
                  SET fuzzy5;
                  DROP _TYPE_ _FREQ_;
                  RUN;

                  DATA oe5;
                  SET oe5;
                  IF _N_ = 1 THEN SET fuzzy5;
                  IF _N_ = 1 THEN SET ll_oneclass;
                  entropy = -1*pp_sum;
                  DROP pp_sum;
                  RUN;
                  *End compute entropy;

                  *Create dataset with enumeration measures;
                  DATA em5;
                  SET oe5;
                  IF _TYPE_ ~= 'PARMS' THEN DELETE;
                  DROP _MODEL_ _MODEL2_ _TYPE_ _NAME_ INTERC1-INTERC5
LINEAR1-LINEAR5 QUADRA1-QUADRA5 SIGMA1 THETA2-THETA5 _CONVERGE_;
                  AIC = -2*_LOGLIK_ + 2*&df5;
                  BIC = -2*_LOGLIK_ + log(&n)*&df5;
                  CAIC = -2*_LOGLIK_ + (log(&n)+1)*&df5;
                  HQ = -2*_LOGLIK_ + log(log(&n))*&df5;
                  ssBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df5;
```

```
            ssCAIC = -2*_LOGLIK_ + (log((&n+2)/24)+1)*&df5;
            ssHQ = -2*_LOGLIK_ + log(log((&n+2)/24))*&df5;
            sE = 1 - (entropy/(&n*log(&s3)));
            NEC = entropy/(_LOGLIK_ - c1loglik);
            CLC = -2*_LOGLIK_ + 2*entropy;
            ICLBIC = -2*_LOGLIK_ + log(&n)*&df5 + 2*entropy;
            ssICLBIC = -2*_LOGLIK_ + log((&n+2)/24)*&df5 + 2*entropy;
            BIC2 = -2*_LOGLIK_ + log(&nobs)*&df5;
            CAIC2 = -2*_LOGLIK_ + (log(&nobs)+1)*&df5;
            HQ2 = -2*_LOGLIK_ + log(log(&nobs))*&df5;
            ssBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df5;
            ssCAIC2 = -2*_LOGLIK_ + (log((&nobs+2)/24)+1)*&df5;
            ssHQ2 = -2*_LOGLIK_ + log(log((&nobs+2)/24))*&df5;
            ICLBIC2 = -2*_LOGLIK_ + log(&nobs)*&df5 + 2*entropy;
            ssICLBIC2 = -2*_LOGLIK_ + log((&nobs+2)/24)*&df5 +
2*entropy;
            logB10 = &bic4-BIC;
            B10 = exp((&bic4-BIC)/2);
            RUN;
            *End create dataset with enumeration measures;

            *Compute average posterior probabilities;
            PROC SORT DATA=of5;
            BY group;
            RUN;

            PROC MEANS DATA=of5 NOPRINT;
            VAR GRP1PRB GRP2PRB GRP3PRB GRP4PRB GRP5PRB;
            BY group;
            OUTPUT OUT=desc5;
            RUN;

            DATA desc5;
            SET desc5;
            IF _stat_ ~= 'MEAN' THEN DELETE;
            IF group = 1 THEN avgpp = GRP1PRB;
            ELSE IF group = 2 THEN avgpp = GRP2PRB;
            ELSE IF group = 3 THEN avgpp = GRP3PRB;
            ELSE IF group = 4 THEN avgpp = GRP4PRB;
            ELSE IF group = 5 THEN avgpp = GRP5PRB;
            DROP _TYPE_ _FREQ_ _STAT_ GRP1PRB GRP2PRB GRP3PRB GRP4PRB
GRP5PRB group;

            RUN;

            PROC SORT DATA=desc5;
            BY DESCENDING avgpp;
            RUN;

            DATA desc5;
            SET desc5;
            id = _N_;
            RUN;

            PROC TRANSPOSE DATA=desc5 OUT=pp_means5 PREFIX=PP;
            ID id;
            VAR avgpp;
            RUN;
```

```
    DATA pp_means5;
    SET pp_means5;
    DROP _NAME_;
    RUN;
    *End compute average posterior probabilities;
    /*****END 5-CLASS MODEL AND OUTPUT*****/

    /*****COMPUTE VLMR AND AVLMR P-VALUES USING MPLUS*****/
    DATA _null_;
    SET of4;
    FILE "&dir\trajdata.dat";
    PUT y1-y&obs;
    RUN;

    %DO c = 4 %TO 5;
  DATA _NULL_;
SET oe&c;
IF _TYPE_ ~= 'PARMS' THEN DELETE;
FILE "&dir\getvlmr.inp" LRECL=1500;
PUT "TITLE:";
PUT "LMR for GBTM";
PUT "DATA:";
PUT " FILE IS &dir\trajdata.dat;";
PUT "VARIABLE:";
PUT " NAMES ARE t1-t&obs;";
PUT " USEVARIABLES ARE t1-t&obs;";
      PUT " CATEGORICAL ARE t1-t&obs;";
PUT " CLASSES = c(&c);";
PUT " MISSING = .;";
PUT "ANALYSIS:";
PUT " TYPE = mixture;";
PUT " STARTS = 100 10;";
      PUT " ALGORITHM = INTEGRATION;";
      PUT "MODEL:";
PUT ' %OVERALL%';
IF &obs = 5 THEN PUT " i s q | t1@0 t2@1 t3@2 t4@3 t5@4;";
ELSE PUT " i s q | t1@0 t2@0.5 t3@1 t4@1.5 t5@2 t6@2.5 t7@3
t8@3.5 t9@4;";
PUT " i with s@0;";
PUT " i with q@0;";
PUT " s with q@0;";
PUT " i@0;";
PUT " s@0;";
PUT " q@0;";
PUT '%c#1%';
PUT "[i*" INTERC1 " s*" LINEAR1 " q*" QUADRA1"];";
PUT '%c#2%';
PUT "[i*" INTERC2 " s*" LINEAR2 " q*" QUADRA2"];";
PUT '%c#3%';
PUT "[i*" INTERC3 " s*" LINEAR3 " q*" QUADRA3"];";
PUT '%c#4%';
PUT "[i*" INTERC4 " s*" LINEAR4 " q*" QUADRA4"];";
IF &c = 5 THEN DO;
  PUT '%c#5%';
  PUT "[i*" INTERC5 " s*" LINEAR5 " q*" QUADRA5"];";
END;
```

```
               PUT "OUTPUT:";
               PUT " TECH11;";
               PUT "SAVEDATA:";
               PUT " RESULTS ARE &dir\trajresult.dat;";
               RUN;

               OPTIONS NOXWAIT;
               X CALL "c:\Program Files (x86)\Mplus\mplus.exe"
                 "&dir\getvlmr.inp"
                 "&dir\getvlmr.out";

               DATA vlmr&c;
               INFILE "&dir\trajresult.dat";
               INPUT #6 vlmr 66-79 vlmr_df 82-95 vlmr_m 98-111 vlmr_sd 114-127
vlmr_p 130-143 avlmr 146-159 #7 avlmr_p 2-15;
               RUN;
             %END;
             /*****END COMPUTE VLMR AND AVLMR P-VALUE USING MPLUS*****/

             /*BUILD OUTPUT DATASETS*/
             DATA em3;
             MERGE e_log3 em3;
             n = &i;
             obs = &j;
             mix = &k;
             miss = &l;
             sigma = &m;
             rep = &rep;
             class = &s1;
             RUN;

             DATA oe3se;
             SET oe3;
             IF _TYPE_ ~= 'STDERR' THEN DELETE;
             DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ SIGMA1 THETA2-THETA3;
             RENAME interc1-interc3 = ise1-ise3;
             RENAME linear1-linear3 = lse1-lse3;
             RENAME quadra1-quadra3 = qse1-qse3;
             RUN;

             DATA oe3;
             RETAIN n obs miss mix sigma rep class INTERC1 LINEAR1 QUADRA1
INTERC2 LINEAR2 QUADRA2 INTERC3 LINEAR3 QUADRA3
                    INTERC4 LINEAR4 QUADRA4 INTERC5 LINEAR5 QUADRA5 SIGMA1
THETA2-THETA5 ISE1 LSE1 QSE1 ISE2 LSE2 QSE2
                    ISE3 LSE3 QSE3 ISE4 LSE4 QSE4 ISE5 LSE5 QSE5 PP1-PP5;
             SET oe3;
             IF _N_ = 1 THEN SET pp_means3;
             IF _N_ = 1 THEN SET oe3se;
             IF _TYPE_ ~= 'PARMS' THEN DELETE;
             n = &i;
             obs = &j;
             mix = &k;
             miss = &l;
             sigma = &m;
             rep = &rep;
```

```
            class = &s1;
            INTERC4 = .;
            LINEAR4 = .;
            QUADRA4 = .;
            INTERC5 = .;
            LINEAR5 = .;
            QUADRA5 = .;
            THETA4 = .;
            THETA5 = .;
            ISE4 = .;
            LSE4 = .;
            QSE4 = .;
            ISE5 = .;
            LSE5 = .;
            QSE5 = .;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ c1loglik entropy c1_issue1-c1_issue5;
            RUN;

            DATA em4;
            MERGE e_log4 em4 vlmr4;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s2;
            RUN;

            DATA oe4se;
            SET oe4;
            IF _TYPE_ ~= 'STDERR' THEN DELETE;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ SIGMA1 THETA2-THETA4;
            RENAME interc1-interc4 = ise1-ise4;
            RENAME linear1-linear4 = lse1-lse4;
            RENAME quadra1-quadra4 = qse1-qse4;
            RUN;

            DATA oe4;
            RETAIN n obs miss mix sigma rep class INTERC1 LINEAR1 QUADRA1
INTERC2 LINEAR2 QUADRA2 INTERC3 LINEAR3 QUADRA3
                   INTERC4 LINEAR4 QUADRA4 INTERC5 LINEAR5 QUADRA5 SIGMA1
THETA2-THETA5 ISE1 LSE1 QSE1 ISE2 LSE2 QSE2
                   ISE3 LSE3 QSE3 ISE4 LSE4 QSE4 ISE5 LSE5 QSE5 PP1-PP5;
            SET oe4;
            IF _N_ = 1 THEN SET pp_means4;
            IF _TYPE_ ~= 'PARMS' THEN DELETE;
            IF _N_ = 1 THEN SET oe4se;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s2;
```

```
            INTERC5 = .;
            LINEAR5 = .;
            QUADRA5 = .;
            THETA5 = .;
            ISE5 = .;
            LSE5 = .;
            QSE5 = .;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ c1loglik entropy c1_issue1-c1_issue5;
            RUN;

            DATA em5;
            MERGE e_log5 em5 vlmr5;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s3;
            RUN;

            DATA oe5se;
            SET oe5;
            IF _TYPE_ ~= 'STDERR' THEN DELETE;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ SIGMA1 THETA2-THETA5;
            RENAME interc1-interc5 = ise1-ise5;
            RENAME linear1-linear5 = lse1-lse5;
            RENAME quadra1-quadra5 = qse1-qse5;
            RUN;

            DATA oe5;
            RETAIN n obs miss mix sigma rep class INTERC1 LINEAR1 QUADRA1
INTERC2 LINEAR2 QUADRA2 INTERC3 LINEAR3 QUADRA3
                   INTERC4 LINEAR4 QUADRA4 INTERC5 LINEAR5 QUADRA5 SIGMA1
THETA2-THETA5 ISE1 LSE1 QSE1 ISE2 LSE2 QSE2
                   ISE3 LSE3 QSE3 ISE4 LSE4 QSE4 ISE5 LSE5 QSE5 PP1-PP5;
            SET oe5;
            IF _N_ = 1 THEN SET pp_means5;
            IF _TYPE_ ~= 'PARMS' THEN DELETE;
            IF _N_ = 1 THEN SET oe5se;
            n = &i;
            obs = &j;
            mix = &k;
            miss = &l;
            sigma = &m;
            rep = &rep;
            class = &s3;
            DROP _AIC_ _BIC1_ _BIC2_ _CONVERGE_ _LOGLIK_ _MODEL_ _MODEL2_
_NAME_ _TYPE_ c1loglik entropy c1_issue1-c1_issue5;
            RUN;

            DATA outdata_measures;
            SET em3 em4 em5;
            RUN;
```

```
            DATA outdata_params;
            SET oe3 oe4 oe5;
            RUN;
            /*****END BUILD OUTPUT DATASETS*****/

                  /*****WRITE OUTPUT DATASETS TO A TEXT FILE*****/
                  DATA _null_;
            SET outdata_measures;
            FILE "&dir\enumeration_measures.txt" DLM=',' MOD LRECL=1500;
            PUT n obs mix miss sigma rep class issue1-issue6 c1_issue1-
c1_issue6 _LOGLIK_ _BIC1_ _BIC2_ _AIC_
                      c1loglik aic bic caic hq ssbic sscaic sshq entropy se
nec clc iclbic ssiclbic bic2 caic2 hq2
                        ssbic2 sscaic2 sshq2 iclbic2 ssiclbic2 logb10 b10
vlmr vlmr_df vlmr_m vlmr_sd vlmr_p avlmr avlmr_p;
            RUN;

                  DATA _null_;
                  SET outdata_params;
                  FILE "&dir\estimates.txt" DLM=',' MOD LRECL=1500;
                  PUT n obs mix miss sigma rep class interc1 linear1 quadra1
interc2 linear2 quadra2 interc3 linear3 quadra3
                      interc4 linear4 quadra4 interc5 linear5 quadra5 ise1
lse1 qse1 ise2 lse2 qse2 ise3 lse3 qse3
                        ise4 lse4 qse4 ise5 lse5 qse5 sigma1 theta2-theta5
pp1-pp5;
                  RUN;
                  /*****END WRITE OUTPUT DATASETS TO A TEXT FILE*****/

                  /*****CLEANUP AFTER YOURSELF*****/
            PROC DATASETS LIBRARY=work NOLIST;
            DELETE desc3-desc5 em3-em5 e_log1 e_log3-e_log5 fuzzy3-fuzzy5
ll_oneclass
                   oe1 oe3-oe5 oe3se oe4se oe5se of1 of3-of5 op1 op3-op5 os1
os3-os5
                   pp_means3-pp_means5 trajdata_b vlmr4 vlmr5 sscpttz
outdata_measures outdata_params;
            RUN;

                DM "out;clear;log;clear;";
                  /*****END CLEANUP AFTER YOURSELF*****/
                  %LET rep = &rep+1;
            %END;
              %LET m = &m+1;
              %LET rep = 1;
            %END;
            %LET l = &l+1;
            %LET m = 1;
            %LET rep = 1;
        %END;
        %LET k = &k+1;
        %LET l = 1;
        %LET m = 1;
        %LET rep = 1;
    %END;
      %LET j = &j+1;
      %LET k = 1;
```

```
      %LET l = 1;
      %LET m = 1;
      %LET rep = 1;
  %END;
  %LET i = &i+1;
  %LET j = 1;
  %LET k = 1;
  %LET l = 1;
  %LET m = 1;
  %LET rep = 1;
%END;

%MEND;

%diss_binary;
```

BIBLIOGRAPHY

Aitkin, M., & Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, 47*(1), 67-75.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317-332.

Balistreri, K. S., & Van Hook, J. (2011). Trajectories of overweight among US school children: A focus on social and economic characteristics. *Maternal and Child Health Journal*, *15*, 610-619.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*(4), 757-786.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338-363.

Biernacki, C., Celeux, G., & Govaert, G. (1998). *Assessing a mixture model for clustering with the integrated classification likelihood* (Tech. Rep. No. 3521). Rhône-Alpes, France: INRIA.

Biernacki, C., Celeux, G., & Govaert, G. (1999). An improvement of the NEC criterion for assessing the number of clusters in a mixture model. *Pattern Recognition Letters*, *20*, 267-272.

Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(7), 719-725.

Biernacki, C., & Govaert, G. (1997). Using the classification likelihood to choose the number of clusters. *Computing Science and Statistics, 29*, 451–457.

Blaze, T. J., & Kim, K. H. (2013, April). *Examining prediction accuracy, posterior probability, and parameter recovery in a semi-parametric trajectory analysis.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Blokland, A. A. J., Nagin, D. S., & Nieuwbeerta, P. (2005). Life span offending trajectories of a Dutch conviction cohort. *Criminology*, *43*(4), 919-954.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3), 345-370.

Bozdogan, H. (1990). On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics-Theory and Methods*, *19*, 221-278.

Brame, R., Nagin, D. S., & Wasserman, L. (2006). Exploring some analytical characteristics of finite mixture models. *Journal of Quantitative Criminology*, *22*(1), 31-59.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* New York, NY: Chapman & Hall/CRC.

Brook, J. S., Zhang, C., & Brook, D. W. (2011). Developmental Trajectories of Marijuana Use From Adolescence to Adulthood Personal Predictors. *Archives of Pediatrics & Adolescent Medicine*, *165*(1), 55-60.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*(1), 147-158.

Buckingham-Howes, S., Oberlander, S. E., Hurley, K. M., Fitzmaurice, S., & Black, M. M. (2011). Trajectories of adolescent mother-grandmother psychological conflict during early parenting and children's problem behaviors at age 7. *Journal of Clinical Child and Adolescent Psychology, 40*(3), 445-455.

Carter, M. A., Dubois, L., Tremblay, M. S., Taljaard, M., & Jones, B. L. (2012). Trajectories of childhood weight gain: The relative importance of local environment versus individual social and early life factors. *Plos One* 7(10): e47065. doi:10.1371/journal.pone.0047065.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*, 195-212.

Chassin, L., Pitts, S. C., & Prost, J. (2002). Binge drinking trajectories from adolescence to emerging adulthood in a high-risk sample: Predictors and substance abuse outcomes. *Journal of Consulting and Clinical Psychology*, *70*(1), 67-78.

Chen, Q., Just, A. C., Miller, R. L., Perzanowski, M. S., Goldstein, I. F., Perera, F. P., & Whyatt, R. M. (2012). Using latent class growth analysis to identify childhood wheeze phenotypes in an urban birth cohort. *Annals of Allergy Asthma & Immunology, 108*(5), 311-U122.

Choi, C-W. J., Stone, R. A., Kim, K. H., Ren, D., Schulz, R., Given, C. W.,…Sherwood, P. R. (2012). Group-based trajectory modeling of caregiver psychological distress over time. *Annals of Behavioral Medicine, 44,* 73-84.

Chung, I.-J., & Chun, J. (2010). Co-occurring patterns of smoking and alcohol consumption among Korean adolescents. *American Journal on Addictions, 19*(3), 252-256.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Cohen, M. A., Piquero, A. R., & Jennings, W. G. (2010). Studying the costs of crime across offender trajectories. *Criminology & Public Policy, 9*(2), 279-305.

Côté, S., Tremblay, R. E., Nagin, D., Zoccolillo, M., & Vitaro, F. (2002). The development of impulsivity, fearfulness, and helpfulness during childhood: patterns of consistency and change in the trajectories of boys and girls. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43*(5), 609-618.

de la Sablonniere, R., Taylor, D. M., Perozzo, C., & Sadykova, N. (2009). Reconceptualizing relative deprivation in the context of dramatic social change: the challenge confronting the people of Kyrgyzstan. *European Journal of Social Psychology, 39*(3), 325-345.

Dolgin, M. J., Phipps, S., Fairclough, D. L., Sahler, O. J. Z., Askins, M., Noll, R. B., . . . Katz, E. R. (2007). Trajectories of adjustment in mothers of children with newly diagnosed cancer: A natural history investigation. *Journal of Pediatric Psychology, 32*(7), 771-782.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society (Ser. B), 57,* 45-97.

Duchesne, S., Ratelle, C. F., Larose, S., & Guay, F. (2007). Adjustment trajectories in college science programs: Perceptions of qualities of parents' and college teachers' relationships. *Journal of Counseling Psychology, 54*(1), 62-71.

Dush, C. M. K., Taylor, M. G., & Kroeger, R. A. (2008). Marital happiness and psychological well-being across the life course. *Family Relations, 57,* 211-226.

Duncan, S.C., & Duncan, T.E. (1994). Modeling incomplete longitudinal substance use data using latent variable growth curve methodology. *Multivariate Behavioral Research, 29,* 313-338.

Edwards, E. P., Homish, G. G., Eiden, R. D., Grohman, K. K., & Leonard, K. E. (2009). Longitudinal prediction of early childhood discipline styles among heavy drinking parents. *Addictive Behaviors*, *34*(1), 100-106.

Enders, C.K., & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430-457.

Everitt, B. S. (1981). A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions. *Multivariate Behavioral Research*, *16*, 171-180.

Feldman, B. J., & Rabe-Hesketh, S. (2012). Modeling achievement trajectories when attrition is informative. *Journal of Educational and Behavioral Statistics*, *37*(6), 703-736.

Ferrer, E., Hamagami, F., & McArdel, J.J. (2004). Modeling latent growth curves with incomplete data using different types of structural equation modeling multilevel software. *Structural Equation Modeling*, *11*(3), 452-483.

Fisher, W. H., Banks, S. M., Roy-Bujnowski, K., Grudzinskas, A. J., Jr., Simon, L. J., & Wolff, N. (2010). Categorizing temporal patterns of arrest in a cohort of adults with serious mental illness. *Journal of Behavioral Health Services & Research*, *37*(4), 477-490.

Gaudreau, P., Amiot, C. E., & Vallerand, R. J. (2009). Trajectories of affective states in adolescent hockey players: Turning point and motivational antecedents. *Developmental Psychology*, *45*(2), 307-319.

Geary, D.C., Bailey, D.H., Littlefield, A., Wood, P., Hoard, M.K., & Nugent, L. (2009). First-grade predictors of mathematical learning disability: A latent class trajectory analysis. *Cognitive Development*, *24*, 411-429.

Gildengers, A. G., Houck, P. R., Mulsant, B. H., Dew, M. A., Aizenstein, H. J., Jones, B. L., . . . Reynolds, C. F. (2005). Trajectories of treatment response in late-life depression - Psychosocial and clinical correlates. *Journal of Clinical Psychopharmacology*, *25*(4), S8-S13.

Guadamuz, T. E., Lim, S. H., Marshal, M. P., Friedman, M. S., Stall, R. D., & Silvestre, A. J. (2012). Sexual, behavioral, and quality of life characteristics of healthy weight, overweight, and obese gay and bisexual men: Findings from a prospective cohort study. *Archives of Sexual Behavior*, *41*(2), 385-389.

Guo, J., Chung, I.-J., Hill, K. G., Hawkins, J. D., Catalano, R. F., & Abbott, R. D. (2002). Developmental relationships between adolescent substance use and risky sexual behavior in young adulthood. *Journal of Adolescent Health*, *31*(4), 354-362.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society (Ser. B)*, *41*(2), 190-195.

Haukka, E., Leino-Arjas, P., Ojajärvi, A., Takala, E.-P., Viikari-Juntura, E., & Riihimaki, H. (2011). Mental stress and psychosocial factors at work in relation to multiple-site musculoskeletal pain: A longitudinal study of kitchen workers. *European Journal of Pain, 15*(4), 432-438.

Haviland, A. M., & Nagin, D. S. (2005). Causal inferences with group based trajectory models. *Psychometrika*, *70*(3), 557-578.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling*, *14*(2), 202-226.

Higgins, G. E., Jennings, W. G., Marcum, C. D., Ricketts, M. L., & Mahoney, M. (2011). Developmental trajectories of nonsocial reinforcement and offending in adolescence and young adulthood: An exploratory study of an understudied part of social learning theory. *Journal of Criminal Justice, 39*(1), 60-66.

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297-307.

Hynes, K., & Clarkberg, M. (2005). Women's employment patterns during early parenthood: A group-based trajectory analysis. *Journal of Marriage and Family, 67*(1), 222-239.

Ishiguro, M., Sakamoto, Y., & Kitagawa, G. (1997). Bootstrapping log-likelihood and EIC, an extension of AIC. *Annals of the Institute of Statistical Mathematics, 49*(3), 411-434.

Jeffreys, S. H. (1961). *Theory of Probability, (3rd ed)*. Oxford: Claredon Press.

Jeffries, N. O. (2003). A note on ‘Testing the number of components in a normal mixture’. *Biometrika*, *90*(4), 991-994.

Jones, B. L. (2012). Traj (Version 9) [Software]. Available from http://www.contrib.andrew.cmu.edu/~bjones/

Jones, B. L., & Nagin, D. S. (2007). Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological Methods & Research, 35*(4), 542-571.

Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research, 29*(3), 374-393.

Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, *90*, 928-934.

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions (2nd ed)*. Thousand Oaks, CA: Sage.

Kokko, K., Pulkkinen, L., Mesiainen, P., & Lyyra, A.-L. (2008). Trajectories based on postcomprehensive and higher education: Their correlates and antecedents. *Journal of Social Issues, 64*(1), 59-76.

Kozyrskyj, A. L., Kendall, G. E., Jacoby, P., Sly, P. D., & Zubrick, S. R. (2010). Association between socioeconomic status and the development of asthma: Analyses of income trajectories. *American Journal of Public Health*, *100*(3), 540-546.

Kuss, O., Gromann, C., & Diepgen, T. L. (2006). Model-based clustering of binary longitudinal atopic dermatitis disease histories by latent class mixture model. *Biometrical Journal*, *48*(1), 105-116.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models (5th ed)*. New York, NY: McGraw-Hill.

Land, K. C., McCall, P. L., & Nagin, D. S. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models with empirical applications to criminal careers data. *Sociological Methods & Research*, *24*(4), 387-442.

Landis, J. R., & Koch, G. G. (1977). The measurement of observed agreement for categorical data. *Biometrics*, *33*, 159-174.

Lavender, J. M., De Young, K. P., Franko, D. L., Eddy, K. T., Kass, A. E., Sears, M. S., & Herzog, D. B. (2011). An investigation of the joint longitudinal trajectories of low body weight, binge eating, and purging in women with anorexia nervosa and bulimia nervosa. *International Journal of Eating Disorders, 44*(8), 679-686.

Le Brocque, R. M., Hendrikz, J., & Kenardy, J. A. (2010). The course of posttraumatic stress in children: Examination of recovery trajectories following traumatic injury. *Journal of Pediatric Psychology, 35*(6), 637-645.

Lee, J. O., Hill, K. G., & Hawkins, J. D. (2012). The role of educational aspirations and expectations in the discontinuity of intergenerational low-income status. *Social Work Research*, *36*(2), 141-151.

Letcher, P., Smart, D., Sanson, A., & Toumbourou, J. W. (2009). Psychosocial precursors and correlates of differing internalizing trajectories from 3 to 15 Years. *Social Development*, *18*(3), 618-646.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13-22.

Little, R. A., & Rubin, D.B. (1989). The analysis of social science data with missing values. *Sociological Methods and Research, 18,* 292-326.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767-778.

Loughran, T., & Nagin, D. S. (2006). Finite sample effects in group-based trajectory models. *Sociological Methods & Research, 35*(2), 250-278.

Louvet, B., Gaudreau, P., Menaut, A., Genty, J., & Deneuve, P. (2007). Longitudinal patterns of stability and change in coping across three competitions: A latent class growth analysis. *Journal of Sport & Exercise Psychology, 29*(1), 100-117.

Louvet, B., Gaudreau, P., Menaut, A., Genty, J., & Deneuve, P. (2009). Revisiting the changing and stable properties of coping utilization using latent class growth analysis: A longitudinal investigation with soccer referees. *Psychology of Sport and Exercise, 10*(1), 124-135.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling*, *14*, 26-47.

M'Bailara, K., Cosnefroy, O., Vieta, E., Scott, J., & Henry, C. (2013). Group-based trajectory modeling: A novel approach to examining symptom trajectories in acute bipolar episodes. *Journal of Affective Disorders, 145*(1), 36-41.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110-133.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics, 36*(3), 318-324.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York, NY: Wiley.

McQuarrie, A. D. R., & Tsai, C.-L. (1998). *Regression and Time Series Model Selection.* River Edge, NJ: World Scientific.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107-122.

Miller, A. (2002). *Subset selection in regression (2nd ed).* New York, NY: Chapman & Hall/CRC.

Moilanen, K. L., Crockett, L. J., Raffaelli, M., & Jones, B. L. (2010). Trajectories of sexual risk from middle adolescence to early adulthood. *Journal of Research on Adolescence, 20*(1), 114-139.

Murray, E. T., Roux, A. V. D., Carnethon, M., Lutsey, P. L., Ni, H., & O'Meara, E. S. (2010). Trajectories of neighborhood poverty and associations with subclinical atherosclerosis and associated risk factors. *American Journal of Epidemiology, 171*(10), 1099-1108.

Muthén, B. O. (2001). Latent variable mixture modeling. In G.A. Marcoulides & R.E. Schumacher (Eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1-33). Lawrence Erlbaum Associates.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*(3), 369-377.

Muthén B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463-469.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*(2), 139-157.

Nagin, D. S. (2004). Response to "Methodological Sensitivities to Latent Class Analysis of Long-Term Criminal Trajectories". *Journal of Quantitative Criminology, 20*(1), 27-35.

Nagin, D.S. (2005). *Group-based modeling of development.* Cambridge, MA: Harvard University Press.

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology, 31*(3), 327-362.

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology, 6*, 109-138.

Nagin, D. S., Pagani, L., Tremblay, R. E., & Vitaro, F. (2003). Life course turning points: The effect of grade retention on physical aggression. *Development and Psychopathology, 15*, 343-361.

Nagin, D. S. & Tremblay, R. E. (2005). Developmental trajectory groups: Fact or a useful statistical fiction? *Criminology, 43*(4), 873-904.

Newman, D.A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6*(3), 328-362.

Nieuwbeerta, P., Nagin, D. S., & Blokland, A. A. J. (2009). Assessing the impact of first-time imprisonment on offenders' subsequent criminal career development: A matched samples comparison. *Journal of Quantitative Criminology, 25*(3), 227-257.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 533-569.

Paciello, M., Fida, R., Tramontano, C., Lupinetti, C., & Caprara, G. V. (2008). Stability and change of moral disengagement and its impact on aggression and violence in late adolescence. *Child Development*, *79*(5), 1288-1309.

Peugh, J., & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? A simulation study examining GMM's performance characteristics. *Structural Equation Modeling, 19*, 204-226.

Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, *17*(1), 1-14.

Quinsey, V. L., Jones, G. B., Book, A. S., & Barr, K. N. (2006). The dynamic prediction of antisocial behavior among forensic psychiatric patients - A prospective field study. *Journal of Interpersonal Violence*, *21*(12), 1539-1565.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, *12*(1), 103-124.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed)*. Thousand Oaks, CA: Sage.

Raykov, T. (2005). Analysis of longitudinal studies with missing data using covariance structure modeling with full-information maximum likelihood. *Structural Equation Modeling, 12*, 493-505.

Reinecke, J. (2006). Longitudinal analysis of adolescents' deviant and delinquent behavior: Applications of latent class growth curves and growth mixture models. *Methodology*, *2*(3), 100-112.

Rissanen, J. (1976). Minmax entropy estimation of models for vector processes. In R. K. Mehra & D. G. Lainiotis (Eds.), *System identification* (pp. 97-119). New York, NY: Academic Press.

Rodriguez, E. T., & Tamis-LeMonda, C. S. (2011). Trajectories of the home learning environment across the first 5 years: Associations with children's vocabulary and literacy skills at prekindergarten. *Child Development, 82*(4), 1058-1075.

Roeder, K., Lynch, K. G., & Nagin, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association, 94*, 766-776.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, *50*(2), 203-228.

Rosendahl, K. I., Galanti, M. R., & Gilljam, H. (2008). Trajectories of smokeless tobacco use and of cigarette smoking in a cohort of Swedish adolescents: Differences and implications. *Nicotine & Tobacco Research, 10*(6), 1021-1027.

Rubin, D.B. (1976). Inference and missing data. *Biometrika, 61,* 581-592.

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

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333-343.

Shin, T., Davison, M.L., & Long, J.D. (2009). Effects of missing data methods in structural equation modeling with nonnormal longitudinal data. *Structural Equation Modeling, 16*(1), 70-98.

Spano, R., Rivera, C., & Bolland, J. M. (2010). Are chronic exposure to violence and chronic violent behavior closely related developmental processes during adolescence? *Criminal Justice and Behavior, 37*(10), 1160-1179.

Sterling, M., Hendrikz, J., & Kenardy, J. (2010). Compensation claim lodgement and health outcome developmental trajectories following whiplash injury: A prospective study. *Pain, 150*(1), 22-28.

Tan, X., Dierker, L., Rose, J., Li, R., & Tobacco Etiology Res, Network. (2011). How spacing of data collection may impact estimates of substance use trajectories. *Substance Use & Misuse, 46*(6), 758-768.

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling, 20*, 640-657.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Charlotte, NC: Information Age.

van Lier, P. A. C., Vitaro, F., Barker, E. D., Koot, H. M., & Tremblay, R. E. (2009). Developmental links between trajectories of physical violence, vandalism, theft, and alcohol-drug use from childhood to adolescence. *Journal of Abnormal Child Psychology, 37*(4), 481-492.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nexted hypotheses. *Econometrica, 57*, 307-333.

Weisburd, D., Morris, N. A., & Groff, E. R. (2009). Hot spots of juvenile crime: A longitudinal study of arrest incidents at street segments in Seattle, Washington. *Journal of Quantitative Criminology, 25*(4), 443-467.

Wiesner, M., Capaldi, D. M., & Kim, H. K. (2007). Arrest trajectories across a 17-year span for young men: Relation to dual taxonomies and self-reported offense trajectories. *Criminology, 45*(4), 835-863.

Windle, M., & Wiesner, M. (2004). Trajectories of marijuana use from adolescence to young adulthood: Predictors and outcomes. *Development and Psychopathology, 16*(4), 1007-1027.

Yang, C. C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistic & Data Analysis, 50*, 1090-1104.

Yeh, H.-H., Chen, C.-Y., Fang, S.-Y., Chang, I.-S., Wu, E. C.-H., & Lin, K.-M. (2011). Five-year trajectories of long-term benzodiazepine use by adolescents: Patient, provider, and medication factors. *Psychiatric Services, 62*(8), 900-907.