Published in final edited form as:

Struct Equ Modeling. 2014 ; 21(2): 263–279. doi:10.1080/10705511.2014.882690.

Determining the Number of Latent Classes in Single- and Multi-Phase Growth Mixture Models

Su-Young Kim

Rutgers, the State University of New Jersey, Center of Alcohol Studies, 607 Allison Road, Piscataway, NJ 08854, Phone: (Cell) 608 - 334 - 6750

Su-Young Kim: suyoun93@gmail.com

Abstract

Stage-sequential (or multiphase) growth mixture models are useful for delineating potentially different growth processes across multiple phases over time and for determining whether latent subgroups exist within a population. These models are increasingly important as social behavioral scientists are interested in better understanding change processes across distinctively different phases, such as before and after an intervention. One of the less understood issues related to the use of growth mixture models is how to decide on the optimal number of latent classes. The performance of several traditionally used information criteria for determining the number of classes is examined through a Monte Carlo simulation study in single- and multi-phase growth mixture models. For thorough examination, the simulation was carried out in two perspectives: the models and the factors. The simulation in terms of the models was carried out to see the overall performance of the information criteria within and across the models, while the simulation in terms of the factors was carried out to see the effect of each simulation factor on the performance of the information criteria holding the other factors constant. The findings not only support that sample size adjusted BIC (ADBIC) would be a good choice under more realistic conditions, such as low class separation, smaller sample size, and/or missing data, but also increase understanding of the performance of information criteria in single- and multi-phase growth mixture models.

Keywords

model selection; class enumeration; growth mixture modeling; multiphase longitudinal data; stage-sequential models; longitudinal data analysis

Growth mixture modeling (GMM; Muthén, 2001a, 2001b; Muthén & Shedden, 1999; Nagin, 1999) is a longitudinal data analysis which is often used to determine whether there exist multiple latent subgroups with different developmental trajectories within a population, and to inference the results to subpopulations. GMM has become an increasingly popular tool for exploring heterogeneity in developmental research in the social sciences (Tofighi & Enders, 2008). Most longitudinal studies, including those that utilize GMM, use single-phase data in which repeated measures represent multiple assessments over a single phase. It is not uncommon, however, for multiple phases to be present in intensive longitudinal data (ILD; Walls & Schafer, 2006), in ecological momentary assessment (EMA)¹; Stone &

Correspondence concerning this article should be addressed to: Su-Young Kim, Center of Alcohol Studies, Rutgers University, 607 Allison Road, Piscataway, New Jersey 08854. suyoun93@gmail.com.

¹Ecological momentary assessment (EMA) is a sampling method to assess subjects' current behaviors and experiences in real time to avoid retrospective recall. In recent years, EMA data collection has been very prosperous due to advances in collecting methods using electrical devices, such as Palm Pilots or cellular phones.

Shiffman, 1994) data, or in prevention/intervention studies. Such multiple-phase (or multiphase) longitudinal data consist of repeated measures within each phase that are repeated across multiple phases, such as yearly math tests during both middle and high school, weekly depression symptom levels before and after an intervention, or the extent of withdrawal severity before and after an attempt to quit smoking. S.-Y. Kim and J.-S. Kim (in press) acknowledged the importance of multiphase longitudinal data in the context of growth mixture modeling, and provided a more thorough discussion of multiphase longitudinal data, with several examples.

For multiphase longitudinal data, state-sequential (or multiphase) growth mixture models may be particularly helpful. For example, Li, Duncan, Duncan, and Hops (2001) examined growth trajectories of adolescent alcohol use during the transition from middle school to high school using a two-phase linear growth mixture model. Among various possible types of stage-sequential GMMs, Kim and Kim (in press) investigated and explicated three distinctive types—traditional piecewise GMM, discontinuous piecewise GMM, and sequential process GMM—that differ in terms of extensions of growth and mixture components of GMM, and gave real data examples of the multiphase GMMs using two-phase EMA data of craving from a smoking cessation study. The first two piecewise GMMs have only one latent class variable regardless of the number of phases present in data, whereas the sequential process GMM has multiple latent class variables (i.e., one in each phase).

Although GMM and its stage-sequential extensions are very useful techniques, it remains unresolved as to how to determine the optimal number of latent classes, known as a class enumeration problem.² Determining the number of latent classes is a critical but difficult problem because there is not a commonly accepted statistical indicator. An appropriate test criterion has to be selected that evaluates competing models with respect to their explanatory power for the data (Entink & Herman, 2009). Many researchers report model comparison indices in their studies, such as Akaike Information Criterion (AIC; Akaike, 1973) and Bayesian Information Criterion (BIC; Schwarz, 1978). Although there is no definitive research on which information criterion (IC) performs best with mixture models or latent class analysis, plenty of books and articles suggest that BIC is a good indicator (Collins, Fidler, Wugalter, & Long, 1993; Hagenaars & McCutcheon, 2002; Magidson & Vermunt, 2004). Much of the purely mathematical or Bayesian literature also recommends BIC (Burnham & Anderson, 2004). AIC has been found to identify a model with one more latent class, compared to the true model, as a correct model in mixture models and latent class analysis (Nylund, Asparouhov, & Muthén, 2007).

In addition to information criteria, some likelihood-based tests—log likelihood difference test, Lo-Mendell-Rubin test (LMR; Lo, Mendell, & Rubin, 2001), and bootstrap likelihood ratio test (BLRT; McLachlan & Peel, 2000)—have also been investigated in mixture models.³ Some other authors suggest using less common techniques. For example, entropy is a summary measure of classification quality based on the posterior probabilities which ranges from 0 to 1 (Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993). Formann (2003) tried to assess overall goodness-of-fit of latent class models using less well-known diagnostic tools,

²There are two important, unresolved issues in utilizing those stage-sequential GMMs: one is determining the required sample size for accurate estimation of parameters and the other is deciding on the optimal number of latent classes. For the former, Kim (in press) carried out a thorough MC study to get tangible sample size requirements for accurate estimation across various simulation conditions, and especially found the close relationship between the sample size required and the number of time points collected.

³To use log likelihood difference test, the degrees of freedom for the difference test should equal the difference in the number of parameters of the two models, i.e., the regularity conditions should be met. Thus, the log likelihood difference test is not applicable for nested LCA or mixture models that differ in the number of classes (Everitt, 1981; McLachlan & Peel, 2000; Nylund et al., 2007; Tofghi & Enders, 2008).

such as the Rudas-Clogg-Lindsay (RCL) index of fit, residual analysis, and methods based on lower order marginals of the contingency table (chi-square tests in the sense of Reiser & Lin, 1999; odds ratios). Muthén (2003) proposed a skewness and kurtosis based test. The procedure relies on testing whether the multivariate skewness and kurtosis tests estimated by the model fit the corresponding sample quantities.

Although there are many fit index studies with latent class models or mixture models, I am aware of only two studies (Nylund et al., 2007; Tofighi & Enders, 2008) that examined the performance of fit indices in the context of growth mixture modeling. Nylund et al. (2007) examined the performance of the three likelihood-based tests mentioned above as well as several ICs to determine the number of classes for three types of models: latent class analysis (LCA), a factor mixture model (FMA), and a growth mixture model (GMM). For the GMM, consistent AIC (CAIC; Bozdogan, 1987) was overall the best indicator, though BIC was nearly as good as CAIC. However, the main focus of that study was on LCA, so the simulation results for the GMM were limited. Tofighi and Enders (2008) conducted quite a comprehensive Monte Carlo (MC) study to evaluate the performance of nine fit indices that can be used to enumerate the number of latent classes in GMM. Their study examined one manipulated factor at a time while holding the other factors constant, rather than examining fully crossed conditions of the factors. Overall, sample size adjusted BIC (Sclove, 1987) performed best, though LMR also performed quite well. These results are consistent with Yang's simulation study (2006) in which adjusted BIC performed better than BIC in latent class models. As such, likelihood-based tests or other goodness-of-fit tests were not very attractive compared to the performance or the convenience of information-based criteria in the previous GMM studies.

The performance of fit indices in stage-sequential growth mixture models, unfortunately, remains unknown in the literature, to the best of my knowledge. This is a serious knowledge gap for application studies of stage-sequential GMMs because latent classes are used to substantively interpret results and make inferences to the population from which the sample was drawn. The main purpose of the present study is to investigate the performance of fit indices that are used for determining the number of latent classes in stage-sequential GMMs through a Monte Carlo simulation study, so that suggestions can be provided about which fit index performs best in practice. Considering the importance of single-phase GMM in theory and practice, thorough examination is also carried out with single-phase GMM.

It should be noted at this point that likelihood-based tests, such as the LMR test and the BLRT, cannot be used with mixture models that have multiple latent class variables. Put simply, the LMR test and the BLRT compare the improvement in fit between $k-1$ class and k class models, because these tests were designed for the models with a single latent class variable (e.g., LCA or GMM), not for the models with multiple latent class variables (e.g., sequential process GMM). Thus, the focus of the present study is limited to traditionally used ICs, which will be explained in detail later.

In summary, the performance of several information criteria that are traditionally used is investigated in the context of single- and multi-phase GMMs. This examination is aimed at bridging the knowledge gap in the literature and providing a better insight about which of the ICs works best under different conditions of factors (e.g., the number of indicator variables, the separation between classes, sample size, class probabilities, and missing proportions) through a series of Monte Carlo simulations. To make the simulation study more complete and informative, the simulations are carried out in two perspectives: the models and the factors. Specifically, the simulation in terms of the models is performed on the basis of the crossed factors in the four types of GMMs, respectively. The simulation in terms of the factors is performed to examine the influence of each manipulated factor

holding the other factors as constant as possible. While the aim of the first simulation is to see the overall performance of ICs within each model and across the models, the aim of the second simulation is to see the effect of each factor on the performance of information criteria.

The organization of the present study is as follows. The following sections contain overviews of single- and multi-phase GMMs and traditionally used information criteria, for the purpose of the Monte Carlo study. Next, a section addresses method, presenting the procedures for the design and data analysis for the Monte Carlo simulation study. The next section presents the results of the simulations under various conditions, in terms of both the models and the factors. Discussion and conclusions are presented in the final section.

Single- and Multi-Phase Growth Mixture Models

For the purpose of this MC study, an overview of single-phase GMM and three types of stage-sequential GMMs is presented in this section.⁴ These four types of GMMs were also presented in Kim and Kim (in press) with detailed model specifications and empirical examples. For diverse sources of related articles, see also Li et al. (2001), Muthén, Khoo, Francis, and Boscardin (2003), and Muthén and Muthén (2010).

Growth mixture modeling may be viewed as a multi-group extension of latent growth modeling (LGM; Meredith & Tisak, 1990) in that groups are not observed but latent. By integrating continuous latent variables (i.e., growth factors, such as an intercept and a slope) and a categorical latent variable (i.e., a latent class variable) in a model, GMM relaxes the single population assumption of LGM. Although population heterogeneity may also be captured in LGM by estimating variances around fixed-effects growth factors (Bauer & Curran, 2003), all individuals are assumed to follow a single growth trajectory in LGM. GMM estimates mean growth curves for each latent class and captures individual variation around these growth curves by estimating growth factor variances for each class (Muthén & Muthén, 2000).⁵ GMM is basically composed of a growth component, which is the same as LGM, and a finite mixture component. A path diagram is provided in Figure 1.

Piecewise LGM (Bollen & Curran, 2008; Muthén & Muthén, 2010; Raudenbush & Bryk, 2002) provides a better solution than single-phase LGM when the true growth patterns vary across multiphases in longitudinal data. The general procedure of piecewise LGM is to identify some fixed transition point during the time period under study, and to fit a trajectory (typically linear) up to that transition and a trajectory after that transition (Raudenbush & Bryk, 2002). In piecewise LGM, there is only one intercept over the multiple phases, which is usually located in the first phase as a starting point of a growth trajectory, though the location of the intercept is not restricted. Traditional piecewise GMM (Li et al., 2001) in Figure 2a⁶ is a mixture extension of piecewise LGM, or it may also be viewed as a stage-sequential extension of GMM in terms of the growth component. Traditional piecewise GMM has multiple growth components for multiple developmental processes while having only one latent class variable for a mixture component. In sum, traditional piecewise GMM has multiple slope factors, one intercept, and also one latent class variable.

Growth trajectories before and after a transition point are connected at the transition point in traditional piecewise GMM. This may not be a problem at all when there is no radical

⁴A similar overview also appears in Kim (in press) which examined the sample size requirement for accurate estimation of parameters in single- and multi-phase GMMs.

⁵In Nagin's group-based approach (1999), systematic individual differences from the mean trajectory within classes are not allowed, which is also referred to as latent class growth analysis (LCGA). Nagin's LCGA models are not considered in the present study.

⁶The three path diagrams in Figure 2 are reproduced from Kim (in press).

change at or near the transition point. Such growth trajectories, however, may distort true growth shapes in data when a more dynamic change or a discrepancy is expected at the transition point, such as a medication treatment in a depression study or a change of craving level at the quit date in a smoking cessation study (Kim & Kim, in press). By changing the growth component of traditional piecewise GMM, a 'discontinuous' piecewise GMM (see Figure 2b) is introduced. There is an intercept for each phase in this model, and the growth trajectories of the first and the second phases are not necessarily connected at the transition point.

Sequential process GMM (Kim & Kim, in press; Muthén & Muthén, 2010) is a stage-sequential adaptation of single-phase GMM by applying multiple, single-phase GMMs for multiphase longitudinal data. Sequential process GMM has multiple mixture components as opposed to the single mixture component in the previous two types of piecewise GMM. In other words, the model has multiple latent class variables, which makes it possible for individuals to change their latent class membership between phases. A sequential process GMM estimates not only parameters corresponding to the proportion of individuals in each latent class, but also a transition probability matrix, which consists of the probability estimates of latent class membership at the next observed time, conditional on latent class membership at the previous time. A path diagram for sequential process GMM is provided in Figure 2c, in which $c1$ and $c2$ are the first and the second latent class variables.

Information Criteria

Information criteria for model selection were originally derived by Akaike (1973) who used the Kullback-Leibler information measure (Kullback & Leibler, 1951) to discriminate between competing models. Schwarz (1978) derived another major class of information criteria (BIC) by using Bayesian statistics. AIC and BIC are generally the most widely used statistical criteria for model selection.

$$AIC = -2\loglikelihood + 2p, \quad (1)$$

$$BIC = -2\loglikelihood + p\log(N), \quad (2)$$

where p is the number of free (estimated) parameters and N is the number of subjects. AIC uses only the number of parameters as a penalty term, whereas BIC uses the number of both parameters and subjects as a penalty. Model selection literatures have a long running debate about using AIC or BIC, which is beyond the purpose of this study.

Since the introduction of AIC and BIC, many modified information criteria have been derived or proposed. Sclove (1987) suggested a sample size adjusted BIC (ADBIC).

$$ADBIC = -2\loglikelihood + p\log\left(\frac{N+2}{24}\right). \quad (3)$$

Yang (2006) performed a simulation study to explore the performance of information criteria in a set of LCA models, and the results indicated that the ADBIC was the best indicator of the information criteria considered, which included AIC and BIC.

While BIC is consistent in the sense that it identifies the correct class model more frequently as sample size increases, AIC is not consistent. Bozdogan (1987) derived a consistent version of AIC (CAIC).

$$CAIC = -2\loglikelihood + p(\log(N) + 1). \quad (4)$$

CAIC is quite similar to BIC, but penalizes more severely for model complexity than BIC, and the difference is the number of free parameters, p . Tofighi and Enders (2008) examined a sample size adjusted CAIC where N is replaced by $(N + 2)/24$ (the same sample size adjustment applied to the ADBIC). This index is referred to as ADCAIC in the present study.

$$ADCAIC = -2\loglikelihood + p \left(\log \left(\frac{N+2}{24} \right) + 1 \right). \quad (5)$$

The five ICs considered in the present study to assess single- and multi-phase GMMs are BIC, ADBIC, AIC, CAIC, and ADCAIC. In their general form, information criteria are based on the log likelihood of a fitted model, where each of the information criteria applies a different penalty for the number of model parameters, sample size, or both. Because of the different penalties used in the ICs, it is possible that each of the information criteria points toward a different class solution as the best model (Nylund et al., 2007). These criteria are either available in statistical software packages such as Mplus, or can be easily computed using the provided output from software packages (e.g., CAIC or ADCAIC).

Method – The Monte Carlo study

Unlike the common use of Monte Carlo studies that investigates the performance of statistical estimators under various conditions, the use of the present MC study is to decide on the optimal number of latent classes in some types of growth mixture models with the aim of investigating the performance of traditionally used information criteria.

Study Design and Data Generation

For the purposes of the present MC study, models to be studied were decided first, which were the single-phase GMM and the three types of multiphase GMMs. Once the models are chosen, the two types of specifications that are allowed to vary in a simulation study should be decided: Monte Carlo variables and population variables. The Monte Carlo variables include the sample size and the number of samples to be generated (replications). For the present study, 500 and 2,000 were chosen for the sample sizes, and 100 replications were generated in each model. Then, population values for each parameter of the model must be decided. Empirical parameter estimates from previous research may provide a good clue for determining appropriate population values in a Monte Carlo study (Muthén & Muthén, 2002). The population variables, discussed in detail later, for the present simulations were 1) class probabilities, 2) the degree to which classes were separated, 3) the number of indicator variables (time points), and 4) missing data proportions. The single- and multi-phase growth mixture models investigated were generated on the basis of the four crossed design factors (population variables) and the Monte Carlo variables. All of the data generations and simulations were carried out using Mplus 6 (Muthén & Muthén, 2010) on a personal computer under Windows 7.

The specification of the manipulated factors determined the generation of the simulated data, which was also closely related to the specific aims in the simulation study. Each manipulated factor, except the number of indicator variables, was characterized by two conditions, one theoretical and the other realistic, because one of the goals in this study was to provide practical and realistic information to researchers who are interested in using growth mixture models with real longitudinal data. The conditions generated in this model

selection study were similar to those shown in the previous Monte Carlo study of determining sample sizes in single- and multi-phase GMMs (Kim, in press).⁷ Throughout the conditions, the single-phase GMM and the two piecewise GMMs had a true 4 class model,⁸ and the sequential process GMM had a true 2×2 (2 in the first phase and 2 in the second phase, 4 class patterns in total) class model. All the models had linear growth parameters with continuous outcomes and a covariate.⁹ Simulation conditions are described in greater detail in terms of both the models and the factors, as discussed previously.

Model Perspective—The aim of the simulation through the models was to see the performance of the traditionally used information criteria within and across the models. The first factor was class probability in two cases: even and uneven. Even class probabilities meant each latent class comprised the same amount of proportion of the generated population: 25%, 25%, 25%, and 25% for the 4 class GMM and piecewise GMMs, and 50% and 50% in the first phase and 50% and 50% in the second phase for the 2×2 class sequential process GMM. Uneven class probabilities were 50%, 20%, 20% and 10% for GMM and piecewise GMMs, and 75% and 25% in the first phase, and 70% and 30% in the second phase for sequential process GMM. Even class probability designated the theoretical condition, while uneven class probability designated the realistic condition.

The second population factor was the degree to which classes were separated (class separation) in two conditions: high class separation and low class separation. These conditions differed in the distributions of their means and variances of growth parameters. The condition of the high class separation was defined by having parameter means that were particularly high or low for a given class so that these parameters discriminated among the classes (Kim, in press; Nylund et al., 2007; Tofighi & Enders, 2008). The condition of the low class separation, on the other hand, was defined by the absence of any single parameter that was particularly high or low for a specific class. During a preliminary simulation phase of this study, the intercept differences between adjacent classes were more sensitive in estimation than were the slope differences, so class separation was defined only on the basis of the intercepts. In the present study, it was defined as high class separation when intercept mean differences between adjacent classes were 3 standard deviations apart, while it was defined as low class separation when the differences were 1.5 standard deviations apart. This choice was made because GMM results for real data showed differences typically of 1 to 2 standard deviations between intercepts across adjacent latent classes in the author's experience.

In fact, the class probability and the class separation factors are closely related (Kim, in press). Nylund et al. (2007) characterized these factors as “simple structure vs. complex structure.”¹⁰ Growth parameter specifications for the single- and multi-phase growth mixture models are summarized in Table 1, consistent with the definitions above. Within-class variance estimates of growth parameters are also provided parenthetically.

⁷Kim (in press) investigated the issue of the sample size requirement for accurate estimation of parameters in single- and multi-phase GMMs, and the current article tried to examine the issue of determining the latent classes in the same models. Therefore, factor specifications for the two MC studies were quite similar, though the purposes, the procedures, and the results were very different.

⁸Through a review of several single- and multi-phase GMMs, the number of extracted latent classes was between 2 and 7 (Duncan, Duncan, Strycker, Okut, & Li, 2002; Hix-Small, Duncan, Duncan, & Okut, 2004; Kim & Kim, in press; Li et al., 2001; Muthén, 2001b, 2001c; Muthén & Muthén, 2000), and 4 was the average.

⁹Covariates are important in correctly specifying the model, in finding the proper number of classes, and in correctly estimating class proportions and class membership (Muthén, 2004). The performance of all information criteria was worse with covariates than without covariates in Tofighi and Enders (2008).

¹⁰High class separation with even class probabilities was defined as a simple structure, which is similar to a factor analysis model in which there are unique items that identify each of the factors (i.e., no or low cross-loadings). On the other hand, low class separation with possibly uneven class probabilities was defined as a complex structure (Kim, in press; Nylund et al., 2007).

The third factor was the number of indicator variables (or the number of time points). For single-phase GMM, the numbers of indicators specified were 4 and 7. For the three types of stage-sequential GMMs, the numbers of indicators were '4 and 4' (4 indicators in the first phase and 4 in the second phase), and '7 and 7.' The time variable (i.e., slope factor loadings of indicators) took on values of 0, 1, 2, and 3 for the 4 indicators, and values of 0, 1, 2, 3, 4, 5, and 6 for the 7 indicators.

The last factor to consider was what proportions of the data were missing. Two conditions were considered: one with complete data and the other with 20% of all possible responses missing. A condition characterized by missing data is more plausible than one with complete data for many longitudinal data collections in psychological or social science studies. Thus, missing data represents the realistic condition, and complete data represents the theoretical condition. It is very common in multiphase longitudinal data that they are sparser in the second phase as the experimental time moves far from the critical or intervention point. The generated data were dense in the first phase, relatively sparse in the beginning of the second phase, and even sparser over time in the second phase. Since the estimation with 20% missing data took an extremely long time compared to the estimation without any missing data,¹¹ the missing data factor was not fully crossed with the other design factors. Instead, simulations with missing data were carried out under several select conditions.

Factor Perspective—The aim of the simulation through the factors was to see the effect of each factor on the performance of the information criteria, while holding the other factors constant. In this simulation study, there were three types of factors: the first was four population variables (e.g., the number of indicators, class separation, class probability, and missingness), the second was Monte Carlo variables (e.g., sample size and the number of replications), and the last was the number of starting values sets. Among these seven individual factors, three factors (the number of indicators, sample size, and the number of starting values sets) were investigated, because they affected the performance of information criteria, and also because they could be, in practice, controlled by researchers. It is very difficult to control missing proportion. It is almost impossible to control class probability in real data analysis. Thus, it would be practically more meaningful to examine the effects of these controllable factors.¹²

Similar to Beauchaine and Beauchaine (2002) or Tofighi and Enders (2008), a normative condition for the factors was decided: uneven class probability (50%, 20%, 20% and 10% for GMM and piecewise GMMs, and 75% and 25% in the first phase, and 70% and 30% in the second phase for sequential process GMM), low class separation (1.5 standard deviations apart), 4 (or 4 and 4) indicator variables, complete data, a sample size of 2,000, 100 replications, and 200 starting value sets. Then, each factor varied while the other factors were held constant at their normative values.

First, the effect of the number of indicators was examined in each model. For the single-phase GMM, 4, 5, 6, and 7 indicators were considered, and for the three stage-sequential GMMs, 4 and 4, 5 and 5, 6 and 6, and 7 and 7 were considered. Second, the effect of the sample size was examined. It took on values of 500, 1,000, 2,000, and 4,000. Lastly, the effect of the number of starting values sets was examined. In the estimation process for mixture models, multiple starting value sets are typically utilized to avoid the local

¹¹Estimation time with missing data were about 100 times longer than without missing data especially when the number of indicators were '7 and 7' in sequential process GMM. The actual time was more than 10 days per model with a modern personal computer, and a single condition required five models to be estimated with sequential process GMM.

¹²The effects of all the factors were examined at the initial stage. To save space, and to be more realistic, only the results of the three controllable factors are provided in this study.

maximum problem. In Mplus, the number of initial stage random sets of starting values as well as the final stage optimizations can be specified by users. A variety of starting value specifications was considered to avoid local solutions, and sufficient random value sets (200 initial and 20 final) were chosen for the whole simulations in the present study. Even though it was rationally expected that greater numbers of starting values sets produce fewer local solutions, resulting in improved performance of the information criteria, the number of starting value sets was more influential than expected, especially in terms of the convergence rate and local maximum problems. The number of random starting value sets took on values of 10, 50, 100, 200, and 300 for the simulation.

Data Analysis Strategy

There are two ways to carry out a Monte Carlo simulation study in Mplus: internal or external. An external Monte Carlo study was performed in which multiple data sets were generated and saved outside Mplus in the first step, and these data were analyzed and the results were summarized in the second step using Mplus.

Data sets were generated according to the population model attributes, and single-phase GMM and stage-sequential GMMs were analyzed using a series of models that differ in the number of latent classes—a true class model and models that were one class different from the true class model in each phase. The true class model (4 classes) and misspecified models (3 and 5 classes) were estimated for GMM and two piecewise GMMs using maximum likelihood estimation with robust standard error under the generated conditions. Also, the true class model (2×2 class pattern) and misspecified models (1×2, 2×1, 2×3, and 3×2 class patterns) were estimated for sequential process GMM. The residual variances were constrained to be equal across classes during the estimation. The covariances between intercepts and slopes were also invariant across classes.

For the process of determining the number of latent classes, the models were fitted to generated data sets, and correctly specified numbers of models were counted in terms of each information criterion. In the case of GMM, for example, 3, 4, and 5 class growth mixture models were fitted to the generated data sets, and each fit criterion was compared across those three models that differed in the number of classes. When the model with the lowest criterion value was the true 4 class model, it was counted as a correct one. The same strategy was applied to the case of stage-sequential GMMs. Our main interest was to determine how often each information criterion correctly identified the true class model and to compare the patterns of the performance of the five information criteria.

Results

Results include the performance of the five information criteria in identifying the correct model, that is, correctly specified numbers of replications in terms of each information criterion. The estimation results showed a very minimal non-convergence rate, approximately 1% of the total. Hence, non-converged replications were ignored and simply discarded when counting the numbers of correctly specified replications.

General Results through the Models

Table 2 provides the proportion of replications in which the true 4 class model was correctly identified in GMM. There were several noteworthy findings. First, AIC performed consistently poorly across all the conditions. The AIC correctly identified the true 4 class model 41% of the time on average, and that proportion was close to the probability of choosing the 4 class model randomly out of 3, 4, or 5 class models. Second, when classes were highly separated (i.e., theoretical condition), all the ICs except AIC performed well

across all conditions. On average, each of BIC, CAIC, and ADCAIC identified the true class model 96% of the time, and ADBIC did 92% of the time. When class separation was low (i.e., realistic condition), however, ADBIC performed best across overall, identifying the true class model 62% of the time. BIC, CAIC, and ADCAIC identified 34%, 28%, and 52% of the time, respectively. BIC, which was one of the best ICs under high class separation, identified almost none of the time under low class separation when the number of indicators were small (4), and sample size was also small ($n = 500$). Third, the simulations with 20% missing data were carried out under eight select conditions because of interminable estimation time as previously indicated. Overall performance of the ICs was almost the same as the case of complete data when class separation was high. However, when the class separation was low, the performance of the ICs in the case of missing data was much worse than in the case of complete data. The patterns of the IC performance were similar to the case of complete data though, meaning ADBIC was overall the best indicator with missing data.

When the conditions were theoretical (i.e., high class separation), BIC, ADBIC, CAIC, and ADCAIC performed quite well overall. These results were consistent with Nylund et al. (2007) in which CAIC performed best, BIC was almost as good as CAIC, and ADBIC was a bit worse than CAIC and BIC. In the present study, determining the best performing IC under the condition of high class separation did not seem to be very meaningful because all ICs performed very well. On the other hand, when the conditions were realistic or complex (e.g., low class separation, missing data, and/or smaller sample size), ADBIC performed best among the five information criteria. Tofighi and Enders (2008) also found that ADBIC performed best under low class separation conditions, although in their study, ADBIC also performed best under high class separation, which is not consistent with the results of the present study as well as those of Nylund et al. (2007).

The results of the traditional piecewise GMM analyses are provided in Table 3. The patterns of the IC performance were quite similar to those found in the GMM results. For example, AIC performed poorly across all conditions, BIC performed very well in high class separation, and ADBIC generally performed better than the others in low class separation. The overall performance of the ICs in the traditional piecewise GMM was a little better than the overall performance of the ICs in the GMM. Table 4 presents the results of the discontinuous piecewise GMM analyses. Again, the performance patterns were similar to the GMM or the traditional piecewise GMM results, and the actual numbers were even a bit higher than the previous results. The results of the previous three models were quite similar to each other probably because there was only one latent class variable across all phases in all three types of GMMs: the single-phase GMM, the traditional piecewise GMM, and the discontinuous piecewise GMM.

The results of the sequential process GMM analyses are provided in Table 5. The performance of the ICs with the sequential process GMM was similar to the results of the previous three types of GMMs in some parts. First, AIC performed poorly again, which was irrespective of any condition. AIC correctly identified the true 2×2 class model 25% of the time across all generated conditions, which is a little bit higher than the probability of choosing the 2×2 class model randomly out of 1×2 , 2×1 , 2×2 , 2×3 , and 3×2 class models. Second, when classes were highly separated, each of BIC, CAIC, and ADCAIC identified the true class model 91% of the time, and ADBIC identified 84% of the time on average. When class separation was low, ADBIC performed best overall, 40% of the time. BIC, CAIC, and ADCAIC identified the true model 24%, 20%, and 34% of the time, respectively. Third, when the data were 20% missing, overall performance of the ICs was similar to the case of complete data, even under the low class separation conditions.

One peculiar finding was quite a big difference in the performance of the ICs between even and uneven class probabilities in the sequential process GMM. While there was not much difference between even and uneven class probabilities in the previous model results, the performance of the ICs in the condition of uneven class probability performed approximately 50% better than in the condition of even class probability. That is, the ICs performed arguably better when there was a dominant latent class. This may require further research in the future. Lastly, it should be noted that it is not relevant to compare actual frequencies across the previous three GMMs and the sequential process GMM, because the numbers of misspecified models are different (2 vs. 4).

Effects of the Specific Factors

The influences of the number of indicator variables, sample size, and the number of starting value sets on the performance of the information criteria were investigated in this section. Each factor was examined while holding the other factors constant: uneven class probability, low class separation, 4 (or 4 and 4) indicator variables, complete data, sample size of 2,000, and 200 starting values sets. The results of each investigation are illustratively summarized in Figures 3 to 5.

First, the influence of the number of indicators is provided in Figure 3. For all of the four types of GMMs, all ICs except AIC identified the true class model nearly 100% of the time when the number of indicators reached 5 (or 5 and 5). At the smallest number of indicators in each model, however, the performance of the ICs could be discriminated across the models. For the two piecewise GMMs, all ICs but AIC performed very well (almost 100%). For the GMM, BIC and CAIC did not perform very well (around 40%), while ADBIC and ADCAIC performed quite well (more than 80%). For the sequential process GMM, BIC, ADBIC, CAIC, and ADCAIC identified the true class model around 60–70% of the time.

Figure 4 shows the influence of sample size in performance of the ICs. First, as the sample size increased, the performance of the ICs clearly improved too, except for AIC. Surely, it was because all but AIC were consistent indicators. Second, the two piecewise GMMs required relatively smaller numbers of subjects to ensure the use of the ICs compared to the GMM or the sequential process GMM. Lastly, across the four types of GMMs, ADBIC performed best overall among the five ICs. ADBIC identified the true class model, except when it was for the sequential process GMM, more than 90% of the time when the sample size reached 2,000. In addition, across the simulations in this study, BIC and CAIC were quite sensitive to sample size and performed poorly in the smallest sample size ($N = 500$) when compared to ADBIC and ADCAIC.

Figure 5 shows the influence of the number of starting value sets. The values from 10 to 300 were the numbers of random starting sets in the initial stage in Mplus. The number of final optimizations was 10% of each of the initial stage sets, except for the first level (10 initial stage sets and 2 final optimizations). Since too many replications showed non-convergence in GMM with 10 starting value sets, counting the correctly specified number of replications was omitted. For the GMM and the traditional piecewise GMM, the performance of the ICs was overall satisfactory and gradually went up as the number of starting value sets increased. For the discontinuous piecewise GMM, the ICs performed very well across all the levels of starting value sets. For the sequential process GMM, the performance of the ICs was poor at 10, 50, and 100 starting value sets, but the performance of the ICs, with an exception of AIC, improved markedly from 100 to 200, and continued to improve up to 300 starting value sets. From these results, the larger numbers of starting value sets appear crucial to ensure the performance of the ICs especially in the sequential process GMM.

Direction of Extraction Errors

The values in tables and figures provide only the correctly specified numbers of replications in terms of each information criterion, not any information about the direction of errors that were committed (i.e., extracting fewer or more classes than the true number of classes). When errors were made, they were closely related with the class separation factor. The ICs tended to extract more classes when classes were clearly separated (high class separation), but fewer classes when classes were not clearly separated (low class separation). It makes sense because low class separation means there is no clear border between classes and may lead to extracting a fewer number of classes.

Discussion and Conclusion

The purpose of the present study was to comprehensively examine the performance of five traditionally used information criteria that could be used to decide on the optimal number of latent classes in GMM and in stage-sequential GMMs. To achieve this aim, a series of Monte Carlo simulations were conducted using Mplus. The findings from this research not only reinforce previous findings concerning the performance of information criteria in mixture models (Nylund et al., 2007; Tofighi & Enders, 2008; Yang, 2006), but also provide entirely new information concerning stage-sequential GMMs. The overall results from the present study support a conclusion that the performance of information criteria in growth mixture models depends on several factors, such as class probability, class separation, number of indicators, proportion of missing data, sample size, and number of starting value sets.

The simulations produced several important specific findings. First, information criteria performed better as: 1) the separation between classes increased, 2) the number of indicators increased, 3) sample size increased, and 4) the number of starting value sets increased. Second, the performance of AIC was pointless, meaning the performance was rather like tossing a coin, needless to say it was disappointing overall. These results are consistent with Nylund et al. (2007) and Tofighi and Enders (2008). Third, ADBIC performed best overall, although BIC, CAIC, and ADCAIC performed somewhat better than ADBIC under the condition of high class separation (theoretical condition). In more realistic conditions (e.g., low class separation, smaller sample size, and/or missing data), however, ADBIC was clearly better than the other information criteria overall. These results are consistent with Nylund et al. (2007) in that BIC and CAIC performed better than ADBIC for simple structure GMM, and consistent with Tofighi and Enders (2008) in that ADBIC was the best indicator for the low class separation condition in GMM. Tofighi and Enders (2008), however, found ADBIC also performed better than other ICs in the condition of high class separation, probably because latent classes in their high class separation condition were not as well separated as those in the present study or in Nylund et al. (2007).

There were also somewhat unexpected results in the present simulation study. First, for the sequential process GMM, the performance of the ICs was worse with even class probabilities than with uneven class probabilities, though the reason was unclear. Across the single-phase GMM and the two piecewise GMMs, any differences in the performance of the ICs between even vs. uneven class probability simulation conditions were minor. Second, a very large number of starting value sets was required to ensure the use of traditionally used information criteria. From a different perspective, local or improper solutions may be avoided via a very large number of starting value sets. For the single-phase GMM, the correctly specified number of replications rapidly increased up to 200 starting value sets, and continued to improve up to 300 starting value sets, as seen in Figure 5. For the sequential process GMM, the correctly specified number of replications rapidly increased up to 300 starting value sets, leaving room to improve.¹³ For the two piecewise GMMs

(traditional and discontinuous), ICs performed well overall across the different starting value set condition.

There are some limitations because this is a simulation study. First, the conditions chosen for the factors were somewhat arbitrary. For example, the high and low class separation conditions were rather subjective, based on the author's experience with growth mixture models. The conditions in each factor, which were characterized as "theoretical vs. realistic," were also based on the author's experience. Second, only several select conditions were examined with the presence of missing data across all the four types of GMMs, because of the interminable estimation. Therefore, interpretation of and generalization from the results are limited to the specific conditions examined in this study.

Nevertheless, the findings support several recommendations for practice to ensure adequate performance of the traditionally used ICs. First, researchers need to collect as many subjects as possible. Second, if they have already collected a limited sample of subjects, they should collect as many time points (indicators) as possible. Third, if the number of time points are minimal as well as the sample size is not large enough, increasing starting value sets seems to be a good option to ensure the use of the ICs, as long as estimation time is tolerable. Finally, the sample size adjusted BIC, ADBIC, is recommended for typical GMM studies. As Tofighi and Enders (2008) cautioned, however, determining the number of latent classes needs to be made after considering a variety of evidence, not by just a single index, until we know more about the conditions under which information criteria performs well. The present study contributes towards expanding our knowledge base with regard to class enumeration (i.e., determining the optimal number of latent classes) using information criteria in single- and multi-phase growth mixture models.

Acknowledgments

The present study was supported, in part, by grants from the National Institute on Alcohol Abuse and Alcoholism (R01 AA019511 and R01 AA019511-02S1).

References

- Akaike, H. Information theory and an extension of the maximum likelihood principle. In: Petrov, BN.; Csaki, BF., editors. *Second International Symposium on Information Theory*. Academiai Kiado; Budapest: 1973. p. 267-281.
- Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*. 2003; 8:338–363. [PubMed: 14596495]
- Beauchaine TP, Beauchaine RJ III. A comparison of maximum covariance and *k*-means cluster analysis in classifying cases into known taxon groups. *Psychological Methods*. 2002; 7:245–261. [PubMed: 12090413]
- Bollen, KA.; Curran, PJ. *Latent curve models: A structural equation modeling perspective*. New Jersey: John Wiley; 2008.
- Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytic extensions. *Psychometrika*. 1987; 52:345–370.
- Burnham KP, Anderson DR. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*. 2004; 33:261–304.
- Collins LM, Fidler PL, Wugalter SE, Long JD. Goodness-of-fit testing for latent class models. *Multivariate Behavioral Research*. 1993; 28(3):375–389.

¹³For reference, the default value of random start option in Mplus is 10 random sets of initial starting values and 2 final stage optimizations.

- Duncan, TE.; Duncan, SC.; Strycker, LA.; Okut, H.; Li, F. Growth mixture modeling of adolescent alcohol use data. 2002. Retrieved from <http://people.oregonstate.edu/~acock/growth-curves/mixture1-30-02.pdf>
- Entink, K.; Herman, R. Statistical Models for Responses and Response Times. 2009. Retrieved from <http://doc.utwente.nl/60452/1/thesisRKleinEntink.pdf>
- Everitt BS. A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distribution. *Multivariate Behavioral Research*. 1981; 16:171–180.
- Formann AK. Latent class model diagnostics-A review and some proposals. *Computational Statistics & Data Analysis*. 2003; 41:549–559.
- Hagenaars, JA.; McCutcheon, AL. Applied latent class analysis. Cambridge University Press; Cambridge: 2002.
- Hix-Small H, Duncan TE, Duncan SC, Okut H. A multivariate associative finite growth mixture modeling approach examining adolescent alcohol and marijuana use. *Journal of Psychopathology and Behavioral Assessment*. 2004; 26(4):255–270.
- Kim S-Y. Sample size requirements in single- and multi-phase growth mixture models: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*. in press.
- Kim S-Y, Kim J-S. Investigating stage-sequential growth mixture models with multiphase longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*. in press.
- Kullback S, Leibler RA. On information and sufficiency. *Annals of Mathematical Statistics*. 1951; 22:79–86.
- Li F, Duncan TE, Duncan SC, Hops H. Piecewise growth mixture modeling of adolescent alcohol use data. *Structural Equation Modeling: A Multidisciplinary Journal*. 2001; 8(2):175–204.
- Lo Y, Mendell N, Rubin D. Testing the number of components in a normal mixture. *Biometrika*. 2001; 88:767–778.
- Magidson, J.; Vermunt, JK. Latent class models. In: Kaplan, D., editor. *The Sage handbook of quantitative methodology for the social sciences*. Newbury Park, CA: Sage; 2004. p. 175-198.
- McLachlan, G.; Peel, D. Finite mixture models. New York: Wiley; 2000.
- Meredith W, Tisak J. Latent curve analysis. *Psychometrika*. 1990; 55(1):107–122.
- Muthén, B. Latent variable mixture modeling. In: Marcoulides; Schumacker, editors. *New developments and techniques in structural equation modeling*. Mahwah, NJ: Erlbaum; 2001a. p. 1-33.
- Muthén, B. Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In: Collins, L.; Sayer, A., editors. *New methods for the analysis of change*. Washington, DC: American Psychological Association; 2001b. p. 291-322.
- Muthén, B. Two-part growth mixture modeling. 2001c. Retrieved from <http://gseis.ucla.edu/faculty/muthen/articles/Article094.pdf>
- Muthén B. Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods*. 2003; 8:369–377. [PubMed: 14596497]
- Muthén, B. Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan, D., editor. *Handbook of quantitative methodology for the social sciences*. Newbury Park, CA: Sage; 2004. p. 345-368.
- Muthén, B.; Khoo, S-T.; Francis, DJ.; Boscardin, CK. Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. In: Reise, SR.; Duan, N., editors. *Multilevel modeling: Methodological advances, issues, and applications*. Mahaw, NJ: Lawrence Erlbaum Associates; 2003. p. 71-89.
- Muthén B, Muthén L. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*. 2000; 24(6):882.
- Muthén B, Muthén L. How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*. 2002; 4:599–620.
- Muthén, L.; Muthén, B. *Mplus: Statistical analysis with latent variables user's guide 6.0*. Los Angeles: Muthén & Muthén; 2010.

- Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*. 1999;463–469. [PubMed: 11318201]
- Nagin DS. Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*. 1999; 4:139–157.
- Nylund KL, Asparouhov T, Muthén B. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*. 2007; 14(4):535–569.
- Ramaswamy V, DeSarbo WS, Reibstein DJ, Robinson WT. An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*. 1993:103–124.
- Raudenbush, SW.; Bryk, AS. *Hierarchical linear models: Applications and data analysis methods*. SAGE Publications Inc; 2002.
- Reiser, M.; Lin, Y. A goodness-of-fit test for the latent class model when expected frequencies are small. In: Sobel, ME.; Becker, MP., editors. *Sociological Methodology*. Blackwell; Oxford: 1999. p. 81-111.
- Schwarz G. Estimating the dimension of a model. *The annals of statistics*. 1978; 6(2):461–464.
- Sclove L. Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*. 1987; 52:333–343.
- Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*. 1994; 16(3):199–202.
- Tofighi, D.; Enders, CK. Identifying the correct number of classes in growth mixture models. In: Hancock, GR.; Samuelsen, KM., editors. *Advances in latent variable mixture models*. Information Age; Greenwich, CT: 2008. p. 317-341.
- Walls, TA.; Schafer, JL. *Models for intensive longitudinal data*. Oxford University Press; USA: 2006.
- Yang CC. Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis*. 2006; 50:1090–1104.

Figure 1. A growth mixture model. *y1* to *y4* are the four indicator variables. *int* is an intercept, and *slp* is a slope. *c* is a latent class variable. *X* represents covariates, and *U* represents outcome variables.

Figure 2.

Three stage-sequential growth mixture models. 'a' is a traditional piecewise GMM, 'b' is a discontinuous piecewise GMM, and 'c' is a sequential process GMM. $y1$ to $y4$ are the first phase measures while $y5$ to $y8$ are the second phase measures. int is an intercept, and slp is a slope. c is a latent class variable. X represents covariates, and U represents outcome variables.

Figure 3.

The influence of the number of indicators. The other factors were held constant as follows: uneven class probability, low class separation, complete data, sample size of 2,000, and 200 starting values sets.

Figure 4.

The influence of sample size. The other factors were held constant as follows: uneven class probability, low class separation, 4 (or 4 and 4) indicator variables, complete data, and 200 starting values sets.

Figure 5.

The influence of the number of starting value sets. The other factors were held constant as follows: uneven class probability, low class separation, 4 (or 4 and 4) indicator variables, complete data, and sample size of 2,000.

Table 1

Growth Parameter Specifications in Data Generation

Model	Class separation	Class patterns			
		class 1	class 2	class 3	class 4
GMM	High	int	3 (1.0)	6 (1.0)	9 (1.0)
		slp	.01 (.05)	.5 (.05)	-.3 (.05)
	Low	int	3 (1.0)	4.5 (1.0)	6 (1.0)
		slp	.01 (.05)	.5 (.05)	-.3 (.05)
TPGMM	High	int	3 (1.0)	6 (1.0)	9 (1.0)
		slp1	.01 (.05)	.5 (.05)	-.3 (.05)
		slp2	-.2 (.04)	.1 (.04)	.01 (.04)
	Low	int	3 (1.0)	4.5 (1.0)	6 (1.0)
DPGMM	High	slp1	.01 (.05)	.5 (.05)	-.3 (.05)
		slp2	-.2 (.04)	.1 (.04)	.01 (.04)
		int1	1 (1.0)	4 (1.0)	7 (1.0)
		slp1	.01 (.05)	.5 (.05)	-.3 (.05)
SPGMM	High	int2	1 (1.0)	6 (1.0)	5 (1.0)
		slp2	-.2 (.04)	.1 (.04)	.01 (.04)
		int1	3 (1.0)	4.5 (1.0)	6 (1.0)
		slp1	.01 (.05)	.4 (.05)	-.3 (.05)
TPGMM	Low	int2	1 (1.0)	6 (1.0)	5 (1.0)
		slp2	-.2 (.04)	.1 (.04)	.01 (.04)
		int1	3 (1.0)	4.5 (1.0)	6 (1.0)
		slp1	.01 (.05)	.4 (.05)	-.3 (.05)

Model	Class separation	Class patterns			
		class 1	class 2	class 3	class 4
	slp2			-.1 (.04)	.3 (.04)
	int1	3 (1.0)	4.5 (1.0)		
Low	slp1	.01 (.05)	.4 (.05)		
	int2			4 (1.0)	5.5 (1.0)
	slp2			-.1 (.04)	.3 (.04)

Note. Values in the table indicate the parameter mean specifications for intercepts and slopes with variances for those parameters in parentheses. *int* is the intercept parameter and *slp* is the slope parameter. TPGMM stands for traditional piecewise growth mixture modeling; DPGMM stands for discontinuous piecewise growth mixture modeling; SPGMM stands for sequential process growth mixture modeling.

Table 2

Frequency of Correctly Identified Models in GMM

Missing	Class probability	Class separation	Indicators	Sample size	BIC	ADBIC	AIC	CAIC	ADCAIC
Complete	Even	High	4	500	93	81	34	92	93
				2,000	100	96	36	100	100
			7	500	92	87	51	92	92
				2,000	100	100	42	100	100
	Low		4	500	3	29	37	0	16
				2,000	55	79	31	43	77
			7	500	21	42	36	13	40
				2,000	90	90	49	90	90
	Uneven	High	4	500	89	76	41	85	87
				2,000	100	98	40	100	100
			7	500	100	85	49	100	98
				2,000	100	99	41	100	100
20% Missing	Even	Low	4	500	1	33	32	0	13
				2,000	49	93	45	37	81
			7	500	40	76	51	26	69
				2,000	100	100	40	100	100
	Uneven	High	4	500	92	84	40	85	94
				2,000	100	98	42	100	100
			4	500	2	28	34	2	13
				2,000	26	70	38	15	60
	Low		4	500	90	85	52	82	92
				2,000	100	99	46	100	100
			4	500	0	23	30	0	9
				2,000	22	71	36	12	55

Note. Values in the table are the numbers of correctly identified models by each information criterion out of 100 replications.

Table 3

Frequency of Correctly Identified Models in Traditional Piecewise GMM

Missing	Class probability	Class separation	Indicators	Sample size	BIC	ADBIC	AIC	CAIC	ADCAIC
Complete	Even	High	4 and 4	500	93	84	51	93	92
				2,000	99	96	42	99	98
			7 and 7	500	99	87	56	99	96
				2,000	100	99	31	100	100
		Low	4 and 4	500	10	32	39	4	26
				2,000	86	86	35	86	86
			7 and 7	500	49	69	39	67	71
				2,000	99	98	33	99	99
	Uneven	High	4 and 4	500	96	83	38	96	96
				2,000	100	98	31	100	100
			7 and 7	500	98	86	46	98	96
				2,000	100	99	40	100	100
		Low	4 and 4	500	10	45	33	6	36
				2,000	97	96	38	97	97
20% Missing	Even	High	4 and 4	500	54	66	45	39	77
				2,000	98	98	47	98	98
			7 and 7	500	92	82	44	92	93
				2,000	100	98	37	100	100
		Low	4 and 4	500	9	29	30	3	20
				2,000	77	76	34	74	76
	Uneven	High	4 and 4	500	97	81	32	97	96
				2,000	100	100	39	100	100
			7 and 7	500	10	42	25	3	31
				2,000	99	98	34	99	99

Note. Values in the table are the numbers of correctly identified models by each information criterion out of 100 replications.

Table 4

Frequency of Correctly Identified Models in Discontinuous Piecewise GMM

Missing	Class probability	Class separation	Indicators	Sample size	BIC	ADBIC	AIC	CAIC	ADCAIC
Complete	Even	High	4 and 4	500	98	89	39	97	97
				2,000	100	99	29	100	99
			7 and 7	500	100	87	30	100	99
				2,000	100	100	29	100	100
		Low	4 and 4	500	26	85	38	12	75
				2,000	100	99	38	100	99
			7 and 7	500	100	91	31	100	99
	Uneven	High	4 and 4	500	99	90	42	99	99
				2,000	100	98	32	100	100
			7 and 7	500	98	87	37	98	97
				2,000	100	100	35	100	100
		Low	4 and 4	500	25	85	39	15	75
				2,000	99	97	43	99	99
20% Missing	Even	High	4 and 4	500	95	86	42	93	95
				2,000	100	99	40	100	99
			7 and 7	500	100	89	33	97	98
				2,000	100	98	26	100	99
		Low	4 and 4	500	12	73	18	8	47
				2,000	98	99	32	94	100
	Uneven	High	4 and 4	500	99	90	38	94	97
				2,000	100	100	40	100	100
			7 and 7	500	93	93	32	93	92
				2,000	93	93	32	90	93

Note. Values in the table are the numbers of correctly identified models by each information criterion out of 100 replications.

Table 5

Frequency of Correctly Identified Models in Sequential Process GMM

Missing	Class probability	Class separation	Indicators	Sample size	BIC	ADBIC	AIC	CAIC	ADCAIC
Complete	Even	High	4 and 4	500	73	66	35	72	70
				2,000	88	88	43	88	88
			7 and 7	500	72	61	35	72	72
				2,000	98	97	37	98	98
		Low	4 and 4	500	0	16	23	0	4
				2,000	4	15	15	1	12
			7 and 7	500	1	26	15	1	18
	Uneven	High	4 and 4	500	99	81	22	99	99
				2,000	100	100	23	100	100
			7 and 7	500	100	81	23	100	99
				2,000	100	98	33	100	100
		Low	4 and 4	500	1	29	18	1	15
				2,000	66	68	21	58	70
20% Missing	Even	High	4 and 4	500	28	58	31	13	56
				2,000	100	97	39	100	100
			7 and 7	500	83	78	31	83	84
				2,000	86	85	36	86	86
		Low	4 and 4	500	0	14	15	0	4
				2,000	3	21	20	0	14
	Uneven	High	4 and 4	500	100	74	26	100	97
				2,000	100	99	18	100	100
			7 and 7	500	5	32	13	2	13
				2,000	56	77	21	43	73

Note. Values in the table are the numbers of correctly identified models by each information criterion out of 100 replications.