

A Note on the Use of Mixture Models for Individual Prediction

Veronica T. Cole and Daniel J. Bauer

University of North Carolina at Chapel Hill

Mixture models capture heterogeneity in data by decomposing the population into latent subgroups, each of which is governed by its own subgroup-specific set of parameters. Despite the flexibility and widespread use of these models, most applications have focused solely on making inferences for whole or subpopulations, rather than individual cases. This article presents a general framework for computing marginal and conditional predicted values for individuals using mixture model results. These predicted values can be used to characterize covariate effects, examine the fit of the model for specific individuals, or forecast future observations from previous ones. Two empirical examples are provided to demonstrate the usefulness of individual predicted values in applications of mixture models. The first example examines the relative timing of initiation of substance use using a multiple event process survival mixture model, whereas the second example evaluates changes in depressive symptoms over adolescence using a growth mixture model.

Keywords: growth mixture models, individual prediction, mixture models, person-centered analysis

Recent years have seen a rapid increase in the use of mixture models within the behavioral, health, and social sciences. Examples include latent class analysis (LCA; Lazarsfeld & Henry, 1968), latent profile analysis (LPA; Gibson, 1959), growth mixture models (GMMs; Muthén & Shedden, 1999; Nagin, 1999), and the recently introduced multiple event process survival mixture model (MEPSUM; Dean, Bauer, & Shanahan, 2014). An attractive feature of all of these models is that they decompose the population into a small number of groups, referred to as latent classes, that capture heterogeneity in the processes under study (McLachlan & Peel, 2000).

Applications of mixture models can be distinguished by whether the latent classes are thought to represent natural groups or are simply used as a convenient device with which to model individual differences (Titterton, Smith, & Makov, 1985). In *direct applications* the goal is to

identify the number of truly distinct groups in the population and to characterize these groups relative to one another and in relation to potentially relevant antecedents and consequences (e.g., deRoon-Cassini, Mancini, Rusch, & Bonanno, 2010; Wiesner & Windle, 2004). By contrast, the goal of *indirect applications* is to estimate as many latent classes as necessary to adequately represent the range of individual differences, without concern for the existence or recovery of natural groups. The latent classes are then interpreted to reflect local conditions (Bauer & Shanahan, 2007; Nagin, 2005) or reaggregated to glean insights about the population as a whole (e.g., Gottfredson, Bauer, Baldwin, & Okiishi, 2014; Kelava, Nagengast, & Brandt, 2014; Pek, Chalmers, Kok, & Losardo, 2015). Thus, depending on the nature of the application, inferences could be drawn with respect to the characteristics of the latent classes, the total population, or both.

It is far less common in a mixture analysis for predictions to be made at the level of the individual. This circumstance is at odds with the frequent description of mixture models as being “person oriented” or “person centered” (Bergman & Magnusson, 1997; Laursen & Hoff, 2006; Muthén & Muthén, 2000). Ironically, it is

Correspondence should be addressed to Veronica T. Cole, Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill, NC 27599, USA. E-mail: vcole@email.unc.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hsem.

more routine to compute, plot, and potentially make inferences about the predicted values of individuals when fitting continuous latent variable models (e.g., random effects growth models, factor analysis models, item response theory [IRT] models) despite the fact that these models are generally not regarded as being person centered. Drawing on this parallel literature, we seek to show that similar individual predictions could be made when using mixture models, enhancing both the interpretation and usefulness of the results.

Overall, our goal is thus to demonstrate how the information provided by a mixture model, whether in a direct or indirect application, can be used to make predictions about individuals. Importantly, although there have been a few examples of the use of predicted values in the growth mixture modeling context (Nagin & Tremblay, 2005; Sterba & Bauer, 2014), we believe that this article provides the first general treatment of individual prediction in mixture models up to this point. Importantly, because mixture models could accommodate virtually any parametric distribution of variables within class, we have sought to present individual prediction in a way that is generalizable to any distributional specification. Drawing on a distinction often made for continuous latent variable models (e.g., Skrondal & Rabe-Hesketh, 2009), we explore the computation and use of *marginal predicted values*, which average over the latent variables, and *conditional predicted values*, which take into account an individual's predicted latent variable scores. We note and demonstrate that different predicted values are suited for different purposes. Additionally, we discuss several different ways of approximating uncertainty around predicted values using parametric bootstrapping (Efron & Tibshirani, 1993). Overall, this emphasis on individual prediction brings the application of mixture models into greater concordance with the goals of a person-centered analytic approach (Bauer & Shanahan, 2007; Bergman & Magnusson, 1997; Sterba & Bauer, 2010).

MIXTURE MODEL FORMULATION

Here we provide a general formulation of the finite mixture model. Defining some initial notation, let i index the individual (where $i = 1, 2, \dots, N$) and let k index latent class (where $k = 1, 2, \dots, K$). Let us also define a set of K indicator variables, designated c_{ik} , that have a value of one when case i is a member of class k and a value of zero otherwise. The values of these indicator variables are unobserved, and the vector \mathbf{c}_i of the indicator variables has a multinomial distribution. For a given individual, the values of the endogenous variables (e.g., items, indicators, repeated measures) are contained in the $p \times 1$ vector \mathbf{y}_i , the values of the exogenous variables (e.g., predictors, covariates) are contained in the $q \times 1$ vector \mathbf{x}_i , and the values of any

continuous latent factors that might be present within the model are contained in the $r \times 1$ vector $\boldsymbol{\eta}_i$.

The joint distribution of the (observed and latent) random variables given the fixed and known covariates can be factored as follows:

$$[\mathbf{y}_i, \boldsymbol{\eta}_i, \mathbf{c}_i | \mathbf{x}_i] = [\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\eta}_i, \mathbf{c}_i] [\boldsymbol{\eta}_i | \mathbf{x}_i, \mathbf{c}_i] [\mathbf{c}_i | \mathbf{x}_i] \quad (1)$$

where, following Muthén and Shedden (1999), $[\mathbf{z}]$ indicates a probability density or mass function for the random variable vector \mathbf{z} . Parameter vectors defining the distributions have been suppressed to keep the notation compact (e.g., $[\boldsymbol{\eta}_i | \mathbf{x}_i, \mathbf{c}_i]$ is often specified as a normal distribution for which the parameter vector would consist of conditional factor means, variances, and covariances). Averaging over the latent variables, we obtain the marginal distribution for the observed variables:¹

$$\begin{aligned} [\mathbf{y}_i | \mathbf{x}_i] &= \sum_{k=1}^K P(c_{ik} = 1 | \mathbf{x}_i) \int [\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\eta}_i, c_{ik} = 1] [\boldsymbol{\eta}_i | \mathbf{x}_i, c_{ik} = 1] d\boldsymbol{\eta}_i \\ &= \sum_{k=1}^K P(c_{ik} = 1 | \mathbf{x}_i) [\mathbf{y}_i | \mathbf{x}_i, c_{ik} = 1] \end{aligned} \quad (2)$$

Equation 2 expresses $[\mathbf{y}_i | \mathbf{x}_i]$ as a finite mixture of k component densities $[\mathbf{y}_i | \mathbf{x}_i, c_{ik} = 1]$, integrated over $\boldsymbol{\eta}_i$ and weighted by the mixing probabilities $P(c_{ik} = 1 | \mathbf{x}_i)$. For some specifications the integration can be completed in closed form (e.g., when the endogenous variables are continuous and both $[\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\eta}_i, \mathbf{c}_i]$ and $[\boldsymbol{\eta}_i | \mathbf{x}_i, \mathbf{c}_i]$ are normal), whereas for other specifications numerical approximation methods are required (e.g., when $[\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\eta}_i, \mathbf{c}_i]$ is a multivariate Bernoulli distribution for binary endogenous variables and $[\boldsymbol{\eta}_i | \mathbf{x}_i, \mathbf{c}_i]$ is normal).

The mixing probabilities (which sum to one within person) depend on the covariates through a multinomial regression specification, given as

$$P(c_{ik} = 1 | \mathbf{x}_i) = \frac{\exp(\alpha_{0k} + \boldsymbol{\gamma}'_k \mathbf{x}_i)}{\sum_{j=1}^K \exp(\alpha_{0j} + \boldsymbol{\gamma}'_j \mathbf{x}_i)} \quad (3)$$

where α_{0k} is an intercept for class k and $\boldsymbol{\gamma}_k$ is a $q \times 1$ vector of coefficients conveying the influence of the covariates on the class probabilities. Constraints must be imposed on the values of the parameters in Equation 3 to identify the model; the most common options are to constrain α_{0k} and $\boldsymbol{\gamma}_k$ to zero within a reference class, or to estimate an intercept-free

¹ Note that here and throughout the article, the term *marginal* is applied when marginalizing with respect to the latent variables, although the expression remains conditional on the observed exogenous covariates.

model in which all values of α_{0k} are set to zero (see Huang & Bandeen-Roche, 2004, for a review).

In fitting a mixture model, the primary goal is to estimate and make inferences regarding the model parameters or functions of these parameters. These parameters consist of two types: those that define the within-class distributions of the endogenous observed variables and latent factors and those that capture between-class prediction within the multinomial regression. For instance, in a GMM application, one might estimate parameters that define the class-specific growth trajectories of the repeated measures as well as parameters that capture the effects of predictors on class membership. Typically, these parameters are estimated via maximum likelihood (ML); however, Bayesian methods of estimation are sometimes also implemented (Depaoli, 2013; Tueller & Lubke, 2010).

INDIVIDUAL INFERENCE IN MIXTURE MODELS

Pursuant to the goals of person-centered analysis, it can be particularly interesting to plot the model-predicted values of the endogenous variables for different individuals (either real or hypothetical). Here, we define and distinguish between marginal and conditional predicted values. Both have a number of different but complementary uses; each is explored in turn.

Marginal Prediction

Marginal predicted values summarize what one can predict for the observed endogenous variables based solely on knowledge of the values of the observed exogenous variables. Because both latent class membership and the values of the latent factors are unknown, marginal predicted values average over these latent variables to arrive at an overall prediction for the endogenous variables. That is, the individual values of the latent variables do not inform the prediction.

Given the marginal mixture distribution in Equation 2, the expected value of \mathbf{y}_i given \mathbf{x}_i can be computed as

$$E(\mathbf{y}_i|\mathbf{x}_i) = \sum_{k=1}^K P(c_{ik} = 1|\mathbf{x}_i)E(\mathbf{y}_i|\mathbf{x}_i, c_{ik} = 1) \quad (4)$$

where $E(\mathbf{y}_i|\mathbf{x}_i, c_{ik} = 1)$ is the expected value of the within-class marginal distribution $[y_i|\mathbf{x}_i, c_{ik} = 1]$. Designating the sample estimate for $P(c_{ik} = 1|\mathbf{x}_i)$ as $\hat{\pi}_{ik}$ and the sample estimate for $E(\mathbf{y}_i|\mathbf{x}_i, c_{ik} = 1)$ as $\hat{\mathbf{y}}_{ik}^{(m)}$, the marginal predicted value could then be defined as

$$\hat{\mathbf{y}}_i^{(m)} = \sum_{k=1}^K \hat{\pi}_{ik} \hat{\mathbf{y}}_{ik}^{(m)}. \quad (5)$$

Equation 5 shows the marginal predicted values to be a simple summation of the within-class marginal predicted values, $\hat{\mathbf{y}}_{ik}^{(m)}$, weighted by the mixing probabilities given the covariates, $\hat{\pi}_{ik}$.

One can compute marginal predicted values for each individual in a sample, for a subset of individuals, or for specific configurations of values for the exogenous variables that might be of interest (irrespective of whether they are observed within the sample; i.e., hypothetical individuals). Although they could be put to a variety of purposes, perhaps the most likely potential use of marginal predicted values is to summarize the predictive relationships implied by the model. For instance, when reporting multilevel and latent growth curve models, it is common to generate and plot predicted trajectories to show how change over time in the repeated measures depends on the values of the predictors (Curran, Bauer, & Willoughby, 2004; Preacher, Curran, & Bauer, 2006). In this context, usually the values of one or two exogenous predictors are varied while other exogenous predictors are held constant at their means, permitting the isolation of specific effects.

For instance, in one application of GMM, deRoos-Cassini et al. (2010) examined the development of posttraumatic stress disorder (PTSD) symptoms following traumatic injury over four time points in the 6 months after initial hospitalization, finding four groups: low-symptom (59%), recovering (13%), delayed (6%), and chronic (22%). Class membership was regressed on self-efficacy at Time 1, anger at Time 1, educational level, and whether the traumatic injury was caused by human intention (e.g., an attack). Among other effects, membership in the chronic PTSD symptom class was strongly predicted by human intention, $OR = 7.67$, 95% CI [2.87, 20.49]. The authors, in post-hoc analyses, might be interested in probing this difference by calculating and plotting marginal predicted trajectories for subjects whose injuries were caused by human intention, versus those whose injuries were not, holding all other covariates at their sample averages. Additionally, one could plot marginal predicted trajectories according to multiple covariates (e.g., plotting trajectories according to human intention and self-efficacy); even in the presence of only main effects, it can be very informative to visualize the joint nonlinear effects of multiple predictors. We demonstrate this strategy shortly.

Conditional Prediction

Marginal predicted values incorporate information about only the exogenous variables when generating predictions, averaging over the unknown latent variables. As we show, however, there are some instances in which we might wish to augment our predictions by considering the most likely values of the latent variables for each individual, which requires incorporating not only covariates \mathbf{x}_i but also latent class indicators \mathbf{y}_i . Such inferences can be made through the

computation of conditional predicted values, which are based on the expectation of $[y_i | x_i, \boldsymbol{\eta}_i, \mathbf{c}_i]$ from Equation 1. If the latent variables were observed, this expectation could be computed as

$$E(y_i | x_i, \boldsymbol{\eta}_i, \mathbf{c}_i) = \sum_{k=1}^K c_{ik} E(y_i | x_i, \boldsymbol{\eta}_i, c_{ik} = 1) \quad (6)$$

This expression differs in two important ways from Equation 4. First, we no longer average over latent classes according to the mixing probabilities. Instead, the expected value depends only on the class to which the individual actually belongs (as c_{ik} will equal one only when an individual is a member of class k and will be zero otherwise). Second, we no longer average over latent factors. Within each component the conditional expected value, $E(y_i | x_i, \boldsymbol{\eta}_i, c_{ik} = 1)$, is computed as the expectation of $[y_i | x_i, \boldsymbol{\eta}_i, c_{ik} = 1]$, utilizing knowledge of the specific values of the latent factors. That is, the conditional expected value in Equation 6 incorporates information about the latent factors, whereas the marginal expected value within Equation 4 does not.

To compute conditional predicted values for a given individual requires that we obtain predictions of the latent variable values for the person. There are many potential ways to compute latent variable scores, both for latent factors (Grice, 2001; Skrondal & Laake, 2001; Tucker, 1971) and latent classes (Bolck, Croon, & Hagnaars, 2004; Lanza, Tan, & Bray, 2013; Vermunt, 2010). We do not delve into this extensive literature here. As is common, for the present purposes, we use empirical Bayes's predictors for the latent variables, which take into account both the observed values of y_i as well as x_i . Although it might be somewhat counterintuitive to use y_i to predict c_i and $\boldsymbol{\eta}_i$ and then use the estimated values of c_i and $\boldsymbol{\eta}_i$ to predict y_i , these sorts of predictions (referred to sometimes as *postdictions*; e.g., Skrondal & Rabe-Hesketh, 2009, p. 674) are quite commonly used in multilevel models and IRT to assess model-based predicted values of y_i against observed values; these uses are explored shortly.

For latent class membership, we calculate posterior probabilities of class membership. The posterior probabilities are given by Bayes's Rule as

$$P(c_{ik} = 1 | x_i, y_i) = \frac{P(c_{ik} = 1 | x_i) [y_i | x_i, c_{ik} = 1]}{[y_i | x_i]}. \quad (7)$$

When computed using the model estimates in place of the population parameter values, the posterior probabilities constitute empirical Bayes's predictors and are denoted \hat{c}_{ik} .

Prediction of the factor scores is similarly based on their posterior distribution. The posterior distribution of $\boldsymbol{\eta}_i$ given y_i , x_i , and class membership is given by Bayes's Rule as

$$[\boldsymbol{\eta}_i | y_i, x_i, c_{ik} = 1] = \frac{[y_i | x_i, \boldsymbol{\eta}_i, c_{ik} = 1] [\boldsymbol{\eta}_i | x_i, c_{ik} = 1]}{[y_i | x_i, c_{ik} = 1]}. \quad (8)$$

Taking the expectation of $[\boldsymbol{\eta}_i | y_i, x_i, c_{ik} = 1]$ yields the expected values of the factors for person i assuming he or she is a member of class k . Because class membership is unknown, there are K possible expected values for each individual, one for each latent class to which the individual might be a member. As before, in computing the class-specific expected values, the model estimates are substituted for the population parameters, making these empirical Bayes's predictors of the factor scores. Factor scores computed in this manner are commonly referred to as expected a posteriori scores (EAPs; Bock & Aitkin, 1981). We denote the EAPs as $\hat{\boldsymbol{\eta}}_{ik}$.

Using these empirical Bayes's predictors for the latent variables, the sample analog to Equation 6 for computing conditional predicted values $\hat{y}_i^{(c)}$ is then

$$\hat{y}_i^{(c)} = \sum_{k=1}^K \hat{c}_{ik} \hat{y}_{ik}^{(c)} \quad (9)$$

where $\hat{y}_{ik}^{(c)}$ is the vector of within-class conditional predicted values obtained based on the predicted factor scores, $\hat{\boldsymbol{\eta}}_{ik}$.

Similar to marginal predicted values, one could compute conditional predicted values based on any configuration of values for x_i and y_i , whether these are observed within the sample or simply represent a subset of possible configurations of interest. There are also many potential uses of conditional predicted values. First, they could be used to judge the correspondence between the predicted and observed values for a specific individual, taking into account the latent class structure of the model (rather than averaging over classes). In this way the conditional predicted values could be used to judge the "person fit" of the model, a strategy that has been used extensively in IRT studies (Reise, 2000). In this context, the strength of the association between the individual's estimated ability and their observed score is then used as a measure of person fit, with weak associations indicating potentially aberrant responding (Conijn, Emons, De Jong, & Sijtsma, 2015; Woods, Oltmanns, & Turkheimer, 2008). Similarly, in mixture models, it might be of interest to gauge concordance between predicted and observed values for a random subset of individuals or for selected individuals based on their most likely class. For instance, in the PTSD example discussed earlier (deRoon-Cassini et al., 2010), one could examine whether individuals in some classes more closely follow their predicted trajectories than in other classes, or identify specific individuals, regardless of class, whose trajectories are poorly predicted by the model.

Second, conditional predicted values could be used to visualize the range of individual differences implied by a

model. For instance, in growth modeling applications, plotting conditional predicted values for the repeated measures provides a visual depiction of the full range of individual differences in change over time (rather than just those differences that could be ascribed to the exogenous predictors; Raudenbush, 2001). Finally, one might use incomplete information on the endogenous variables when computing the posterior predicted values to generate predictions about the remaining endogenous variables. Dean, Cole, and Bauer (2015) used this strategy in a survival mixture model of substance abuse initiation in adolescence; given substance use data at age 13, they predicted the pattern of substance use initiation throughout adolescence and young adulthood.

In sum, when fitting a mixture model, we can make individual predictions using either marginal or conditional predicted values. Marginal predicted values take into account only the values of the exogenous variables, whereas conditional predicted values also take into account the predicted values of the latent variables for the individual. Marginal predicted values are well suited to visualizing relationships between exogenous and endogenous variables, averaging over the latent variables. Conditional predicted values are well suited for making individual predictions that are informed by the latent variables, and can be used to evaluate person fit, to visualize individual differences, or for forecasting purposes (among other possibilities). Regardless of the type or use of predicted values, however, an important consideration is that they are computed using sample estimates for the model parameters. Thus, prior to illustrating the use of these predicted values, we consider how best to represent their uncertainty due to sampling error.

QUANTIFYING UNCERTAINTY AROUND INDIVIDUAL PREDICTIONS

There are a number of ways to consider uncertainty around the model-implied predicted values given by Equations 5 and 9. One possibility is to alter the prediction intervals (PIs) developed by Skrondal and Rabe-Hesketh (2009) for random effects models for use with mixture models. Another option is to use parametric bootstrap methods (Efron & Tibshirani, 1993) to generate a number of hypothetical sets of predicted values from the model. This latter approach, which we pursue here, quantifies uncertainty in the individual predicted values by empirically approximating the sampling distribution of the model parameters on which the predicted values are based. We first discuss the computation of confidence intervals (CIs) for the point estimates of the predicted values $\hat{y}_i^{(m)}$ and $\hat{y}_i^{(c)}$ and then consider the computation of PIs for the observations y_i based on these predicted values.

Confidence Intervals

Let us designate the full vector of model parameters as θ . Under certain assumptions, the ML estimates of the model parameters, $\hat{\theta}$, are asymptotically normally distributed around the true parameter values θ , as follows:

$$\hat{\theta} \sim N\{\theta, V(\hat{\theta})\} \quad (10)$$

Here, $V(\hat{\theta})$ represents the variance covariance matrix of the estimates $\hat{\theta}$.

The parametric bootstrap strategy, described in detail by Pek, Losardo, and Bauer (2011), consists of making some number B of bootstrap draws (e.g., 5,000) from the estimated sampling distribution of the parameter estimates. More specifically, the bootstrap distribution substitutes the ML estimates $\hat{\theta}$ and the estimated covariance matrix of the estimates $\hat{V}(\hat{\theta})$ for their population counterparts, as follows:

$$\tilde{\theta}_b \sim N\{\hat{\theta}, \hat{V}(\hat{\theta})\}. \quad (11)$$

Draws are then taken from this distribution to construct confidence intervals for $\hat{y}_i^{(m)}$ or $\hat{y}_i^{(c)}$.

With each draw from the bootstrap distribution, a new set of marginal or conditional predicted values is calculated based on $\tilde{\theta}_{(b)}$ (where $b = 1, 2, \dots, B$). These values, which we designate $\tilde{y}_{i(b)}^{(m)}$ and $\tilde{y}_{i(b)}^{(c)}$, respectively, are computed as shown in Equations 5 and 9, with the exception that the bootstrapped parameter estimates $\tilde{\theta}_{(b)}$ are used in the computations (rather than original model estimates, $\hat{\theta}$), including when obtaining latent variable scores for conditional predicted values. Using this procedure we can obtain an empirical distribution of marginal or conditional predicted values that reflects the uncertainty of the parameter estimates on which they are based. It might be particularly informative to construct plots of the bootstrapped values. Specifically, to visualize uncertainty due to sampling error, we could plot either the bootstrapped marginal or conditional predicted values $\tilde{y}_{i(b)}^{(m)}$ or $\tilde{y}_{i(b)}^{(c)}$, respectively, around the original (mean) values $\hat{y}_i^{(m)}$ or $\hat{y}_i^{(c)}$.

The set of B bootstrapped values obtained for any given individual could thus be used to quantify or visualize sampling error about the individual predicted values. For instance, for any given endogenous variable, the 2.5th and 97.5th percentiles of the empirical distribution of predicted values demarcate a 95% bootstrapped CI. For marginal predicted values, the procedure for generating bootstrapped CIs could be summarized as follows:

1. Simulate B draws from $\hat{V}(\hat{\theta})$. Denote each vector of parameters as $\tilde{\theta}_{(b)}$.
2. For each set of parameter values, obtain the class-specific expected value $\tilde{y}_{ik(b)}^{(m)}$ for each class. Calculate

prior probabilities of class membership $\tilde{\pi}_{i1(b)} \dots \tilde{\pi}_{iK(b)}$ for case i using Equation 3, and aggregate across classes using Equation 5 to obtain $\tilde{\mathbf{y}}_{i(b)}^{(m)}$.

3. Choose the 97.5th and 2.5th values of $\tilde{\mathbf{y}}_{i(b)}^{(m)}$ from the B draws; these demarcate the bounds of a 95% CI.

For conditional predicted values, the procedure is varied to incorporate the latent variable scores:

1. Simulate B draws from $\hat{V}(\hat{\theta})$. Denote each vector of parameters as $\tilde{\theta}^{(b)}$.
2. For each set of parameter values, calculate subject i 's conditional expected value $\tilde{\eta}_{ik(b)}$ (or EAP) for each class by taking the expectation of the distribution given by Equation 8; use this value to obtain class-specific conditional expected values $\tilde{\mathbf{y}}_{i(b)}^{(c)}$. Calculate posterior probabilities of class membership $\tilde{\tau}_{i1(b)} \dots \tilde{\tau}_{iK(b)}$ for case i using Equation 7, and aggregate across classes using Equation 9 to obtain $\tilde{\mathbf{y}}_{i(b)}^{(c)}$.
3. Choose the 97.5th and 2.5th values of $\tilde{\mathbf{y}}_{i(b)}^{(c)}$ from the B draws; these demarcate the bounds of a 95% CI.

Prediction Intervals

The preceding procedure describes CIs for the predicted values $\hat{\mathbf{y}}_i^{(m)}$ and $\hat{\mathbf{y}}_i^{(c)}$. These intervals indicate sampling error in the estimated expected value of \mathbf{y}_i over repeated sampling, either marginalizing over the latent variables (with $\hat{\mathbf{y}}_i^{(m)}$) or conditioning on their values (with $\hat{\mathbf{y}}_i^{(c)}$). In some instances, however, we might be interested in conveying uncertainty in the potential observed values of \mathbf{y}_i (or any of its individual elements y_{ij}) that correspond to these predicted values; that is, given the predictors \mathbf{x}_i and perhaps also including knowledge of \mathbf{c}_i and $\boldsymbol{\eta}_i$. PIs indicate the range of values of \mathbf{y}_i that might be observed for a real or hypothetical individual within a specified probability (e.g., 95% of the time; Kutner, Nachtsheim, & Neter, 2004, pp. 56–60; Skrondal & Rabe-Hesketh, 2009).

Whereas a CI for $\hat{\mathbf{y}}_i^{(m)}$ or $\hat{\mathbf{y}}_i^{(c)}$ takes into account variability in the parameter estimates $\hat{\theta}$, a PI also must take into account the variance of the random variables in the model, for which the realized values will vary across observations. When computing a marginal PI, the random variables include \mathbf{c}_i , $\boldsymbol{\eta}_i$, and \mathbf{y}_i ; whereas when computing a conditional PI the latent variable values are treated as known and the only random variables are contained within \mathbf{y}_i . The two types of PIs have different uses. For instance, marginal PIs are useful for predicting new observations for new individuals, whereas conditional PIs are useful for predicting new observations for individuals for whom some data have already been collected (i.e., individuals in the sample).

To make PIs for a new observation of \mathbf{y}_i , we augment the resampling procedure outlined earlier by taking draws from the conditional distributions of the random variables. This strategy, borrowed from the multilevel modeling literature (Kovacevic, Huang, & You, 2006; Van Der Leeden, Meijer, & Busing, 2008, pp. 401–419), involves successively drawing from each distribution as follows. For marginal PIs:

1. Simulate B draws from $\hat{V}(\hat{\theta})$. Denote each vector of parameters as $\tilde{\theta}^{(b)}$.
2. For each set of parameter values:
 - a. Draw class membership: Using the bootstrapped parameter values, compute the prior probabilities of class membership using Equation (3); take one random draw of class membership $\tilde{\mathbf{c}}_{i(b)}$ from the multinomial distribution with probabilities $\tilde{\pi}_{i1(b)} \dots \tilde{\pi}_{iK(b)}$.
 - b. Draw $\tilde{\boldsymbol{\eta}}_{i(b)}$: Given membership to class k obtained from step (a), take one random draw from the class-specific marginal distribution of continuous latent variables for subject i 's assigned class, $[\tilde{\boldsymbol{\eta}}_{i(b)} | \mathbf{x}_i, \tilde{\mathbf{c}}_{ik(b)} = 1]$.
 - c. Draw from the predicted distribution of \mathbf{y} : Given the class membership assigned in step (a) and the value of $\tilde{\boldsymbol{\eta}}_{i(b)}$ obtained from step (b), take one random draw, denoted $\tilde{\mathbf{y}}_{i(b)}^{(m)}$, from the implied distribution of \mathbf{y}_i values for the individual, $[\tilde{\mathbf{y}}_{i(b)}^{(m)} | \mathbf{x}_i, \tilde{\boldsymbol{\eta}}_{i(b)}, \tilde{\mathbf{c}}_{ik(b)} = 1]$.
3. Choose the 97.5th and 2.5th values of $\tilde{\mathbf{y}}_{i(b)}^{(m)}$ from the B draws; these demarcate the bounds of a 95% PI around $\hat{\mathbf{y}}_i^{(m)}$. For conditional PIs:

1. Simulate B draws from $\hat{V}(\hat{\theta})$. Denote each vector of parameters as $\tilde{\theta}^{(b)}$.
2. For each set of parameter values:
 - a. Draw class membership: For each class, compute the posterior probability of class membership $\tilde{\tau}_{ik(b)}$ using Equation 7; take one random draw of class membership $\tilde{\mathbf{c}}_{ik(b)}$ from the multinomial distribution with probabilities $\tilde{\tau}_{i1(b)} \dots \tilde{\tau}_{iK(b)}$.
 - b. Draw $\tilde{\boldsymbol{\eta}}_{i(b)}$: Given membership to class k obtained from Step a, take one random draw from the class-specific posterior distribution of continuous latent variables for subject i 's assigned class, $[\tilde{\boldsymbol{\eta}}_{i(b)} | \mathbf{y}_i, \mathbf{x}_i, \tilde{\mathbf{c}}_{ik(b)} = 1]$.
 - c. Draw from the predicted distribution of \mathbf{y} : Given the class membership assigned in Step a and the value of $\tilde{\boldsymbol{\eta}}_{i(b)}$ obtained from Step b, take one random draw, denoted $\tilde{\mathbf{y}}_{i(b)}^{(c)}$, from the implied distribution of \mathbf{y}_i values for the individual, $[\tilde{\mathbf{y}}_{i(b)}^{(c)} | \mathbf{x}_i, \tilde{\boldsymbol{\eta}}_{i(b)}, \tilde{\mathbf{c}}_{ik(b)} = 1]$.

3. Choose the 97.5th and 2.5th values of $\hat{y}_{i(b)}^{(c)}$ from the B draws; these demarcate the bounds of a 95% PI around $\hat{y}_i^{(c)}$.

Importantly, PIs and CIs are both centered around the same expected values (i.e., $\hat{y}_i^{(m)}$ or $\hat{y}_i^{(c)}$), but will differ in width to afford different inferences, with PIs being wider than CIs to encompass the additional range of variability in potential realized values associated with any given expected value.

In summary, by making use of resampling procedures, we can compute CIs around predicted values of y_i to convey the uncertainty of the estimates; further, we can compute PIs by adding another resampling step that accounts for variation in the realized values associated with any given predicted value.

ASSUMPTIONS OF THE RESAMPLING APPROACH

The parametric bootstrap offers a relatively computationally inexpensive way of obtaining intervals when analytical derivations require approximations or are otherwise intractable, as would often be the case for the models considered here. A fully nonparametric solution, which would involve re-estimation of the model on B redrawn samples from the raw data, would be considerably more expensive computationally and potentially infeasible (Davison & Hinkley, 1997, pp. 15–22; Yung & Bentler, 1996).

It is important to recognize, however, that the parametric bootstrap approach invokes a number of assumptions that must be met to yield valid coverage rates. Above all, it assumes that the parametric specification of $\hat{\theta}$ in Equation 10 is correct—that is, that $\hat{\theta}$ is a consistent estimator of θ , that $\hat{\theta}$ is asymptotically normally distributed, and that the sample size is sufficiently large to approach asymptotics. Whether these properties hold will depend on the estimator of $\hat{\theta}$, with each method of estimation invoking its own set of assumptions. For normal theory ML, Satorra (1990) divided these into structural assumptions, such as the inclusion of all relevant variables and correct specification of the relationships between them; and distributional assumptions, including that errors are normally distributed and homogeneous across all levels of predictors and outcomes (i.e., homoscedasticity), and observations are independent and identically distributed. Also included in this latter category is that the distributions of the latent and observed variables are specified correctly (e.g., normal, binomial, Poisson). Within mixture models, both structural and distributional misspecification can compromise class enumeration procedures and lead to inconsistent within-class estimates even when the correct number of classes is selected (Bauer & Curran, 2003, 2004; Hoeksma & Kelderman, 2006; Morin et al., 2011; Van Horn et al., 2012).

The parametric bootstrap also requires that $\hat{V}(\hat{\theta})$ is a consistent estimator of $V(\hat{\theta})$ in Equation 11. Under normal theory ML, $V(\hat{\theta})$ is given by the inverse of the expected Fisher information matrix (Eliason, 1993); $\hat{V}(\hat{\theta})$ can be estimated as the second derivative of the log-likelihood evaluated at the ML estimator (i.e., the observed Fisher information matrix; Efron & Hinkley, 1978). Because there is typically no closed-form analytic solution for $\hat{V}(\hat{\theta})$ when the expectation maximization (EM; Dempster, Laird, & Rubin, 1977) algorithm is used, $\hat{V}(\hat{\theta})$ is evaluated numerically by most statistical packages using the method of Louis (1982); this is the method used here. The naive estimate of $\hat{V}(\hat{\theta})$ will be consistent under the same structural and distributional assumptions described by Satorra (1990). Note, however, that there are alternative ways to compute $\hat{V}(\hat{\theta})$ with varying degrees of robustness to model misspecification (Arminger & Schoenberg, 1989; Browne & Arminger, 1995), nonnormality (when distributions are specified as normal; Satorra & Bentler, 1994; Yuan & Bentler, 1997), and heteroscedasticity (Huber, 1967; White, 1982).

Thus, if the assumptions of the normal theory ML estimator are met, then parametric bootstrap estimates of variability should also show good asymptotic performance and nominal coverage rates. By implication, considerable attention should be paid to the specification of the model when implementing parametric bootstrapping (Efron & Tibshirani, 1993; Preacher & Selig, 2012). Without intending to discount the importance of avoiding specification errors, it is also worth recognizing that research on indirect applications of mixture models have documented good performance for estimates and inferences made by aggregating information over classes, despite the fact that the fitted model is not literally correct in the population (Bauer, Baldasaro, & Gottfredson, 2012; Nagin, 2005, pp. 48–54; Pek et al., 2011; Sterba & Bauer, 2014). Thus, although additional research is needed on this point, the predicted values $\hat{y}_i^{(m)}$ and $\hat{y}_i^{(c)}$ and associated CIs and PIs might be relatively robust to the violation of structural assumptions, distributional assumptions, or both, provided the model is specified in such a way that the estimated parameters are still able to capture the primary features of the data generating process.

We now turn to two empirical demonstrations of the utility of individual predictions in mixture models.

EMPIRICAL EXAMPLE 1: PATTERNS OF SUBSTANCE USE INITIATION

For our first example, we revisit an application of a multivariate survival mixture model to trajectories of substance use initiation in adolescence. The goal of this analysis was to characterize patterns of onset of substance use across multiple categories of licit and illicit drugs (e.g., alcohol,

tobacco, marijuana, cocaine). Here we build on this analysis by focusing specifically on predicting the onset of harder drug use based on the time of initiation of alcohol and marijuana use.

Sample and Measures

Data come from the National Survey on Drug Use and Health (NSDUH), and consist of substance use initiation questionnaires taken from respondents on a yearly basis between ages 10 and 30. Subjects ($N = 55,772$; 52% female) were asked whether and at what age they initiated use of the following substances: alcohol, tobacco, marijuana, nonmedical use of prescription drugs (NMUP), hallucinogens, cocaine, inhalants, stimulants, or heroin. The sample was ethnically diverse, with 62% of respondents categorized as White, 13% as African American, 16% Hispanic, and 9% as other. Sex and ethnicity were included in the model as covariates. A more detailed description of the data and sample are available in Dean et al. (2015).

Model Fitting

The model applied to the data is the MEPSUM model introduced by Dean et al. (2014). Briefly, the MEPSUM model characterizes the relative timing of multiple events. For each of the nine drug classes, a binary indicator variable was created at each age that was scored zero if the event had not yet occurred, one if the event occurred at that specific age, and missing otherwise (i.e., if the event had already occurred or the individual was not observed at that age, indicating censoring). The vector of indicator variables constitutes the endogenous variables for the model, or \mathbf{y}_i .

The MEPSUM model conforms to Equation 2 with $[\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\eta}_i, c_{ik} = 1]$ defined by assuming that the indicator variables within \mathbf{y}_i are conditionally independent and Bernoulli distributed. The probability parameters, which in this case correspond to hazard rates, vary across classes and are designated \mathbf{h}_{ik} . In this application, we implemented the logistic link function for the hazards and modeled change in the logit for each substance as a quadratic function of age. Similar to a multivariate growth model, this entailed the definition of a latent intercept, linear change, and quadratic change factor for each substance. Thus, the model was defined as

$$\mathbf{v}_{ik} = \boldsymbol{\Lambda} \boldsymbol{\eta}_i \quad (12)$$

where \mathbf{v}_{ik} is the vector of logits corresponding to \mathbf{h}_{ik} . The factor loading matrix $\boldsymbol{\Lambda}$ was defined to be equal across classes (hence the absence of a k subscript) and to consist of nine blocks reflecting quadratic change for each substance. Specifically, using decade as the metric of time, the block of factor loadings corresponding to any given substance s was specified as

$$\boldsymbol{\Lambda}_s = \begin{pmatrix} 1 & 0 & 0 \\ 1 & .1 & .01 \\ 1 & .2 & .04 \\ \vdots & \vdots & \vdots \\ 1 & 2 & 4 \end{pmatrix} \quad (13)$$

Each block thus defined three factors, for a total of 27 factors across the nine substances. Contrasting with a standard multivariate growth analysis, in the MEPSUM model the variances and covariances of the factors are all set to zero, that is, $\boldsymbol{\Psi}_k = \mathbf{0}$; only the factor means $\boldsymbol{\alpha}_k$ are estimated (hence the within-class distribution for the continuous latent variables drops out of Equation 2 and the integral resolves). Class membership was regressed on coding variables representing sex and ethnicity using the multinomial specification given in Equation 3. Finally, for conveying results, the model-implied hazards are cumulated to produce lifetime distribution functions (see Dean et al., 2014, for computational details).

Full details of model fitting and estimates are available in the original report (Dean et al., 2015). In brief, fit indexes favored a six-class solution, shown in Figure 1. A plurality of subjects fell into the abstainer class (35.7%), characterized by a low risk of any substance use over time. Users of soft drugs (alcohol, tobacco, and marijuana) fell into three remaining classes, characterized by late onset of soft drug use (26.7%), early onset of soft drug use (12.1%), and a progressive, steady hazard of initiating soft drug use (8.2%). Users of hard drugs (predominantly cocaine, hallucinogens, and NMUP) fell into the remaining two classes: late hard drug onset (10.9%), and early hard drug onset (6.3%).

Individual Predictions

For this model, we focus on conditional predicted values; however, for completeness we also provide the formula for marginal predicted values. Specifically, for this model the marginal predicted values in Equation 5 correspond to

$$\begin{aligned} \hat{\mathbf{y}}_i^{(m)} &= \sum_{k=1}^K \hat{\pi}_{ik} \hat{\mathbf{y}}_{ik}^{(m)} \\ &= \sum_{k=1}^K \hat{\pi}_{ik} \hat{\mathbf{h}}_{ik}^{(m)} \end{aligned} \quad (14)$$

where $\hat{\mathbf{y}}_i^{(m)}$ can be interpreted as the predicted hazard rate given the covariate scores, averaging over latent classes.

Here, $\hat{\mathbf{h}}_{ik}^{(m)}$ is the model-implied marginal hazard rate within class k , calculated by first computing the predicted logit as $\hat{\mathbf{v}}_{ik}^{(m)} = \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\alpha}}_k$ and then inverting the logistic link function to obtain the predicted probability (expected value) corresponding to each element in $\hat{\mathbf{v}}_{ik}^{(m)}$.

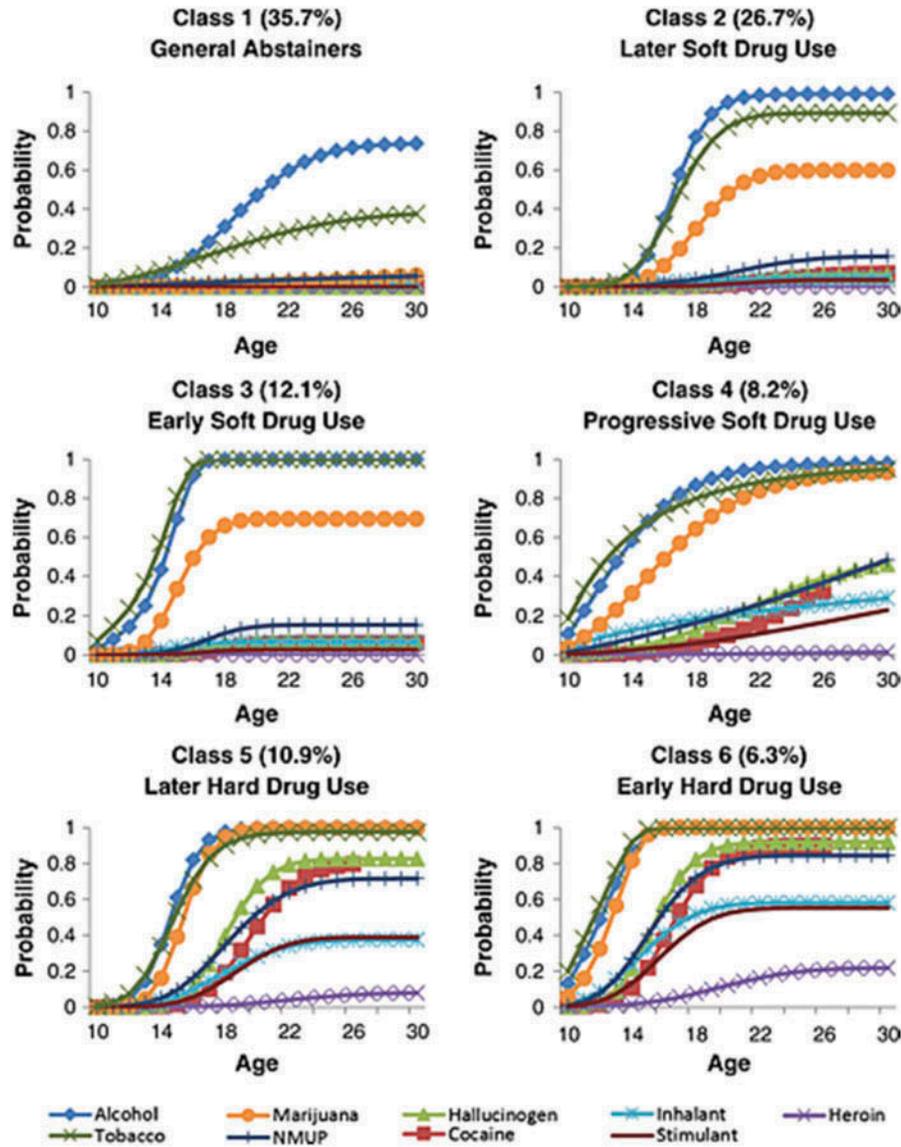


FIGURE 1 Empirical Example 1: Cumulative probability of substance use initiation over time. *Note.* Figure originally appears as Figure 4 in Dean et al. (2015).

Similarly, the conditional predicted values in Equation 9 are computed as

$$\begin{aligned} \hat{\mathbf{y}}_i^{(c)} &= \sum_{k=1}^K \hat{\tau}_{ik} \hat{\mathbf{y}}_{ik}^{(c)} \\ &= \sum_{k=1}^K \hat{\tau}_{ik} \hat{\mathbf{h}}_{ik}^{(c)} \end{aligned} \quad (15)$$

and can be interpreted as the predicted hazard for person i given her values for the covariates as well as her posterior predicted class membership. Here, $\hat{\mathbf{h}}_{ik}^{(c)}$ is the predicted hazard conditional on the individual's latent factor scores

and is calculated by computing the predicted logit within class k as $\hat{\mathbf{v}}_{ik}^{(c)} = \hat{\Lambda} \hat{\boldsymbol{\eta}}_{ik}$ and applying the inverse link function. In this case, because the variances of the latent factors are null, the predicted factor scores collapse to the class means, and $\hat{\mathbf{v}}_{ik}^{(c)} = \hat{\Lambda} \hat{\boldsymbol{\eta}}_{ik} = \hat{\Lambda} \hat{\boldsymbol{\alpha}}_k$. Thus, because in this model there is no within-class variation in the factor scores, $\hat{\mathbf{h}}_{ik}^{(c)} = \hat{\mathbf{h}}_{ik}^{(m)}$ and the only difference between the marginal predicted values in Equation 14 and the conditional predicted values in Equation 15 is whether the within-class hazards are weighted by the predicted or posterior class probabilities. Weighting by the predicted probabilities averages over classes, whereas weighting by the posterior probabilities conditions on the individual's class membership.

In the current application we used conditional predicted values to evaluate the likelihood of engaging in hard drug use given early involvement in softer drug use. A number of studies have linked the early use of marijuana (Kandel, 2002; Robins, Darvish, & Murphy, 1970) to later hard drug use. Given growing concerns about the abuse of prescription medications (Boyd, McCabe, Cranford, & Young, 2006; Compton & Volkow, 2006), we centered our examination on NMUP onset and how it could vary depending on the timing of onset of marijuana use, holding alcohol use constant. Conditional predicted values are best suited to evaluating this question because the observed timing of marijuana use can be used to provide information on class membership. We computed conditional predicted values for NMUP for two hypothetical individuals: (a) a White male subject with alcohol use onset at age 12 and marijuana use onset at age 13; and (b) a White male subject with alcohol use onset at age 12 and marijuana use onset at age 16. In computing the posterior probabilities of class membership from Equation 7, we coded all of the remaining indicator variables in \mathbf{y}_i as missing (unobserved). For ease of interpretation, we used the predicted hazards for these individuals to compute lifetime distribution functions.

The predicted lifetime distribution functions are shown as the solid bold lines in Figure 2. To convey the sampling error in these predicted values, we also generated and

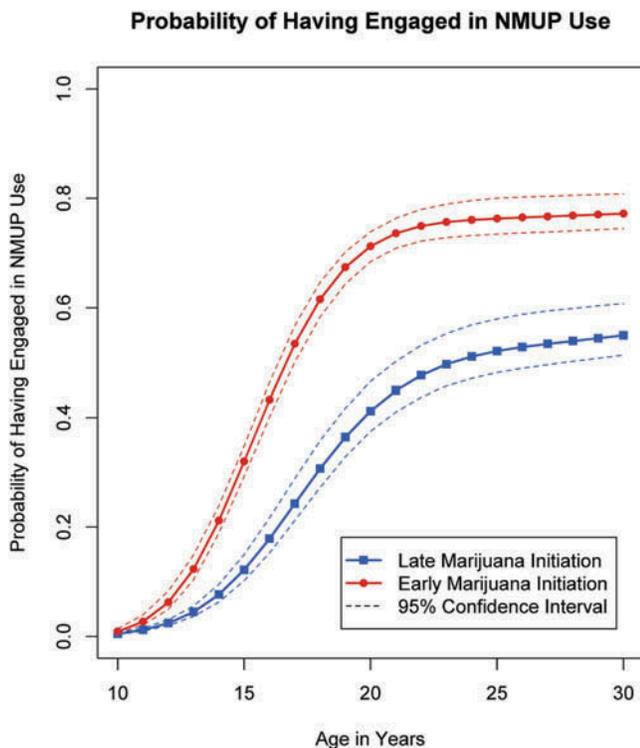


FIGURE 2 Empirical Example 1: Predicted probability of nonmedical prescription drug use (NMUP) for two subjects differing in age of marijuana use initiation.

plotted $B = 2,500$ bootstrapped estimates of the predicted values. The plotted intervals are point-wise confidence intervals as opposed to PIs; thus, they convey uncertainty in the expected survival function. Examination of these functions indicates that for both subjects the lifetime probability of NMUP use increases rapidly over adolescence and then begins to asymptote in the early 20s. Earlier initiation of marijuana use, however, results in an earlier, more rapid, and more pronounced increase in the likelihood of NMUP use. By age 20, an early-onset marijuana user has a roughly 70% probability of having engaged in NMUP, compared to approximately 40% for the late-onset marijuana user.

In sum, this application illustrated how conditional predicted values can be used for forecasting purposes. A subset of the endogenous variables (referencing marijuana and alcohol use) was used to infer posterior probabilities of class membership, which in turn were used to generate conditional predicted values for other endogenous variables at later points in time (and for other substances). In this manner we were able to enhance our understanding of the interdependence of substance use onset times implied by the MEPSUM model, in particular, the relation between early marijuana use and NMUP use.

EMPIRICAL EXAMPLE 2: THE DEVELOPMENT OF DEPRESSIVE SYMPTOMATOLOGY

For our second example, we demonstrate how to plot predicted trajectories from a GMM, with a specific focus on changes in depressive symptoms during the transition from adolescence to adulthood.

Sample and Measures

Data were drawn from the 1997 National Longitudinal Survey of Youth (NLSY97). For the purposes of this demonstration, we included data for individuals who were 14 years old in 1997 from assessments made in 2000, 2002, 2004, 2006, 2008, and 2010 (i.e., during the transition to adulthood), and for whom no more than half of the selected repeated measures were missing ($N = 1,460$). The sample was 51% male, and relatively ethnically diverse, with 27%, 19%, 1%, and 53% of respondents identifying as Black, Hispanic or Latino, mixed race, and neither Black nor Hispanic or Latino, respectively.

The main outcome of interest, depression, was measured as the sum of five 4-point items from the Center for Epidemiological Studies Depression scale (CES-D) (Radloff, 1977). Potential values range from 5 to 20, with higher values indicating higher levels of depression. Predictors include race (coded 0 for non-Black/non-Hispanic/Latino, 1 otherwise), gender (coded 1 for males, 0 for females), parent-rated physical health (coded from 1–5, with lower scores representing better overall health),

and college attendance (coded 1 if the subject attended college by age 23, 0 otherwise). The latter predictor was included based on research indicating that clinical patterns among college students might differ from their non-college-attending counterparts (Blanco et al., 2008; Gfroerer, Greenblatt, & Wright, 1997).

Model Fitting

The GMM fit to the data allowed for maximum flexibility in the shape of change over time observed within classes and was initially described in detail by Ram and Grimm (2009). In brief, whereas a typical GMM might constrain growth within each class to follow a linear or lower order polynomial function, here we estimate the functional form of growth freely within each class via a freed-loading latent basis model (Meredith & Tisak, 1990).

With reference to Equation 2, in this application we assumed that, conditional on the latent growth factors, the repeated measures are normally distributed within classes; that is $[y_i | x_i, \eta_i, c_{ik} = 1]$ references a normal distribution with conditional mean vector μ_{ik} and covariance matrix Σ_k . These conditional moments are structured according to a latent growth model such that

$$\begin{aligned} \mu_{ik} &= \Lambda_k \eta_i \\ \Sigma_k &= \text{DIAG}(\sigma_{1k}^2, \sigma_{2k}^2, \dots, \sigma_{Jk}^2) \end{aligned} \tag{16}$$

where J is the number of repeated measures, σ_{jk}^2 is the residual variance of each repeated measure at time j in class k , and **DIAG**() is an operator that places the enclosed elements in a diagonal matrix, implying that the repeated measures are conditionally independent.

Each class-specific matrix of factor loadings, Λ_k , was minimally constrained to define a latent intercept and shape factor, as follows:

$$\Lambda_k = \begin{pmatrix} 1 & 0 \\ 1 & \lambda_{22k} \\ 1 & \lambda_{32k} \\ 1 & \lambda_{42k} \\ 1 & \lambda_{52k} \\ 1 & 1 \end{pmatrix}. \tag{17}$$

Similarly, we assumed the latent growth factors to be normally distributed within classes such that $[\eta_i | x_i, c_{ik} = 1]$ references a normal distribution with mean vector α_k and variance-covariance matrix Ψ_k . Each of the elements of the mean vector α_k was freely estimated; however, to decrease model complexity and facilitate model convergence, only the variance of the intercept was estimated and it was constrained to equality across classes; thus

$$\alpha_k = \begin{pmatrix} \alpha_{1k} \\ \alpha_{2k} \end{pmatrix}; \Psi_k = \begin{pmatrix} \psi_{11} & \\ 0 & 0 \end{pmatrix}. \tag{18}$$

Because both $[y_i | x_i, \eta_i, c_{ik} = 1]$ and $[\eta_i | x_i, c_{ik} = 1]$ are normal, the within-class marginal distribution of the repeated measures $[y_i | x_i, c_{ik} = 1]$ in Equation 2 resolves to a normal probability distribution with an implied mean vector of $\Lambda_k \alpha_k$ and covariance matrix of $\Lambda_k \Psi_k \Lambda_k' + \Sigma_k$. Last, covariate effects on class membership were modeled via the multinomial regression specification given in Equation 3.

As in any application, one could consider alternative model specifications to the one used here. For instance, if the within-class distributions of depression are thought to be skewed, then one might specify $[y_i | x_i, \eta_i, c_{ik} = 1]$ as a skew-normal distribution (Asparouhov & Muthén, 2016). Likewise, one could consider alternative specifications of the within-class covariance matrix for the latent factors. Without a strong basis for advocating for one specification over another, we sought here to implement the simplest model specification that we believed would adequately capture individual differences in change over time.

Models with successively larger numbers of classes were tested, and a three-class model was chosen on the basis of the Bayesian information criterion (BIC; Schwarz, 1978), which balances fit and parsimony (BIC = 34343.96, 34334.82, and 34349.38 for models with two, three, and four classes, respectively). To determine whether free estimation of the functional form of growth was necessary, we also examined the fit of a three-class linear GMM; we found the freed-loading GMM to fit significantly better than the corresponding linear model, $\chi^2(12) = 97.30, p < .001$. The residual variance of the indicators σ_{jk}^2 was also allowed to differ over classes k and time points j ; this parameterization showed significantly better fit than a model in which residual variance was constrained to be equal across time, $\chi^2(15) = 85.55, p < .001$, or across classes, $\chi^2(12) = 786.33, p < .001$.

Figure 3 shows the predicted trajectories for these three classes. A plurality of cases (40.2%) fall into Class 1, which is characterized by generally low levels of depressive symptoms that decrease very slightly over time. The other two classes are characterized by either a pattern of symptoms that start out relatively low but increase linearly (Class 2, 23.9%), or high overall levels of symptoms that remain relatively stable and decrease very slightly (Class 3, 35.9%). Of the covariate effects examined, gender and college attendance were the only significant predictors of class membership. Membership to Class 3, which was characterized by high overall levels of depressive symptoms, was positively associated with being female and negatively associated with college attendance. Specifically, women were significantly more likely to be in Class 3 than Class 1, $\gamma = 0.362, z = 2.156, p = .031$; or Class 2, $\gamma = 0.869, z = 3.541, p < .001$. College nonattendees were significantly more likely to be in Class 3 than Class 1,

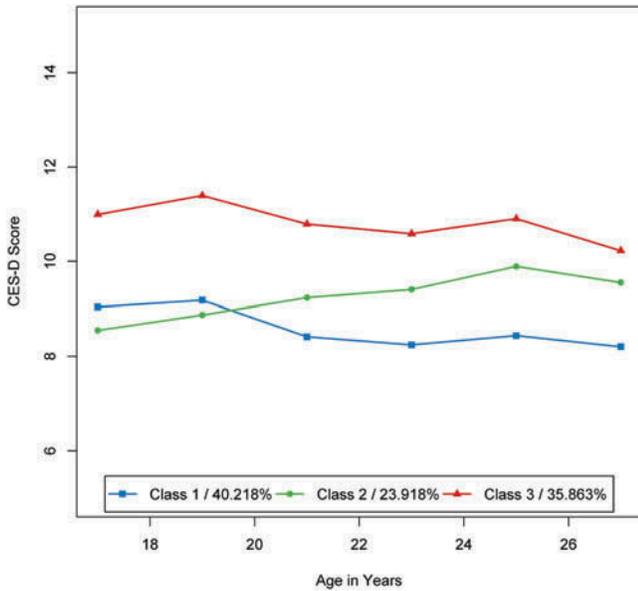


FIGURE 3 Empirical Example 2: Mean trajectories and class membership proportions for each class.
 Note. CES-D = Center for Epidemiological Studies Depression scale.

$\gamma = 1.018, z = 5.661, p < .001$; or Class 2, $\gamma = 0.805, z = 2.970, p = .003$. To better understand the results of the model we now compute predicted values. Marginal predicted values are used to visualize the effects of the covariates, whereas conditional predicted values are used to evaluate the person fit of the model.

Individual Predictions

The analysis at the whole-sample level suggested that both gender and college attendance are linked to depressive symptoms. To better understand this relationship, we computed marginal predicted values for four hypothetical individuals, varying the values of gender and college attendance, but holding all nonfocal covariates at their sample averages. For this model, the marginal predicted values in Equation 5 can be expressed as

$$\hat{\mathbf{y}}_i^{(m)} = \sum_{k=1}^K \hat{\pi}_{ik} \hat{\mathbf{y}}_{ik}^{(m)} = \sum_{k=1}^K \hat{\pi}_{ik} (\hat{\Lambda}_k \hat{\mathbf{a}}_k) \tag{19}$$

Thus, based on the covariate values, we first obtained the predicted probabilities of class membership $\hat{\pi}_{i1}, \hat{\pi}_{i2}$, and $\hat{\pi}_{i3}$. Then, weighting the model-implied estimated class mean trajectories ($\hat{\mathbf{y}}_{ik}^{(m)} = \hat{\Lambda}_k \hat{\mathbf{a}}_k$) by these probabilities, we obtained the marginal predicted trajectories shown in the top panel of Figure 4. These predicted trajectories help to clarify the nature of the predictor effects, showing how, over

this span of development, depressive symptoms are more severe in women and in those who did not attend college. The shape of change is consistent, generally declining with age with the exception of episodic increases at 19 and 25 years of age. We generated marginal 95% CIs using $B = 2,500$ for these trajectories. As shown in the top panel of Figure 4, the set of plausible values overlaps almost completely for male and female participants; however, as shown in the bottom panel, the trajectories diverge somewhat more between college attendees and nonattendees. This is an informative finding, as it contextualizes the results from the covariate logistic regression: Although both college attendance and gender had significant effects on membership to a more highly symptomatic class, the overall predicted difference between college attendees and nonattendees in the development of depressive symptoms is more robust than that between male and female participants.

Marginal trajectories predict the course of depressive symptoms based solely on the covariates of interest, and do not incorporate information about the individual values for the latent variables. Because the latent variables are unobserved, marginal predicted values truly represent what one would expect based only on the known information. However, for some purposes it is useful to incorporate inferred information about the latent variables when computing the predicted values. For the fitted GMM, the conditional predicted values in Equation 9 are obtained as

$$\hat{\mathbf{y}}_i^{(c)} = \sum_{k=1}^K \hat{\tau}_{ik} \hat{\mathbf{y}}_{ik}^{(c)} = \sum_{k=1}^K \hat{\tau}_{ik} (\hat{\Lambda}_k \hat{\boldsymbol{\eta}}_{ik}) \tag{20}$$

Here we display the person fit of the model for two cases from each class. For each class k , two cases were selected at random among those with $\hat{\tau}_{ik} > .95$. Conditional predicted values were plotted against observed values, shown in Figure 5. Additionally, given the bootstrapping procedure described earlier, PIs were generated using $B = 2,500$ predicted values to approximate the uncertainty of the prediction. Across the six individuals, a few subjects (e.g., Subjects 321, 379, and 372) have at least one data point that lies far from the line of prediction; however, all individual data points are enclosed within the corresponding PIs.

Visual examination of Figure 5 suggests varying levels of closeness between observed data points and the model-predicted values. The observed scores of some individuals appear to increase more or less rapidly than predicted, or change more erratically. Nevertheless, the observed values are within the range that the model predicts should be observed, suggesting that the model provides reasonably good fit for these individuals. For other individuals, examination of plots such as these might suggest poor person fit,

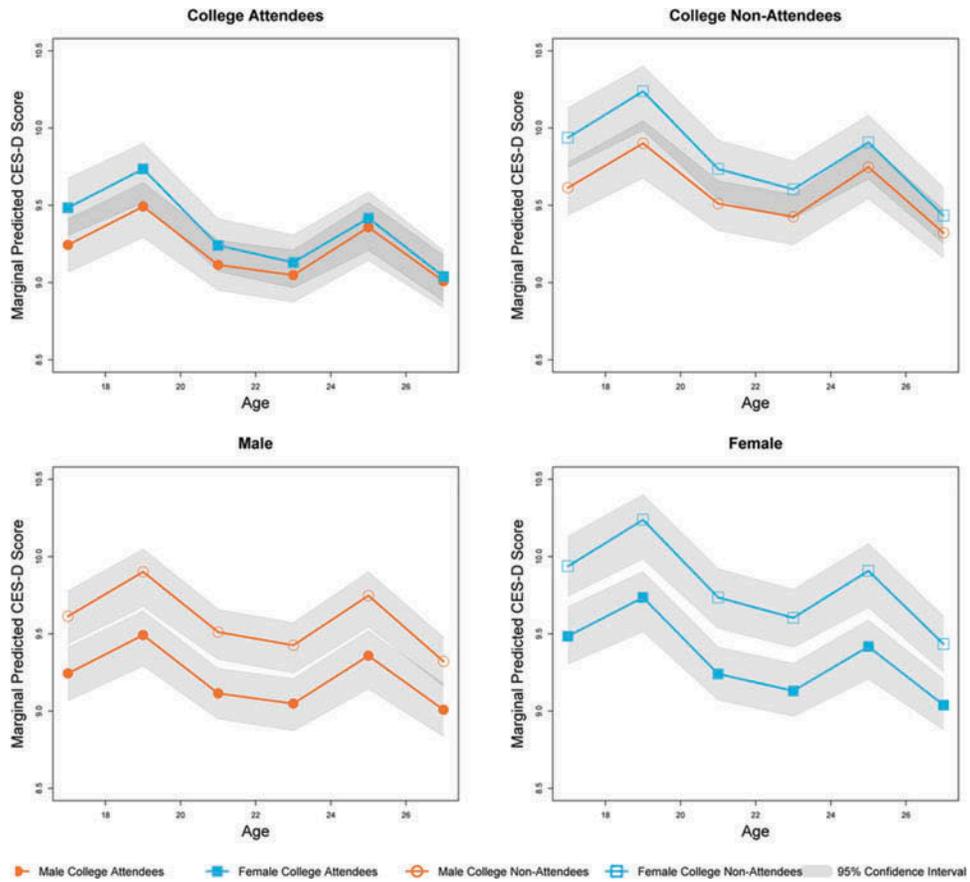


FIGURE 4 Empirical Example 2: Marginal predicted values of the Center for Epidemiological Studies Depression scale (CES-D) for four subjects differing in gender and college attendance.

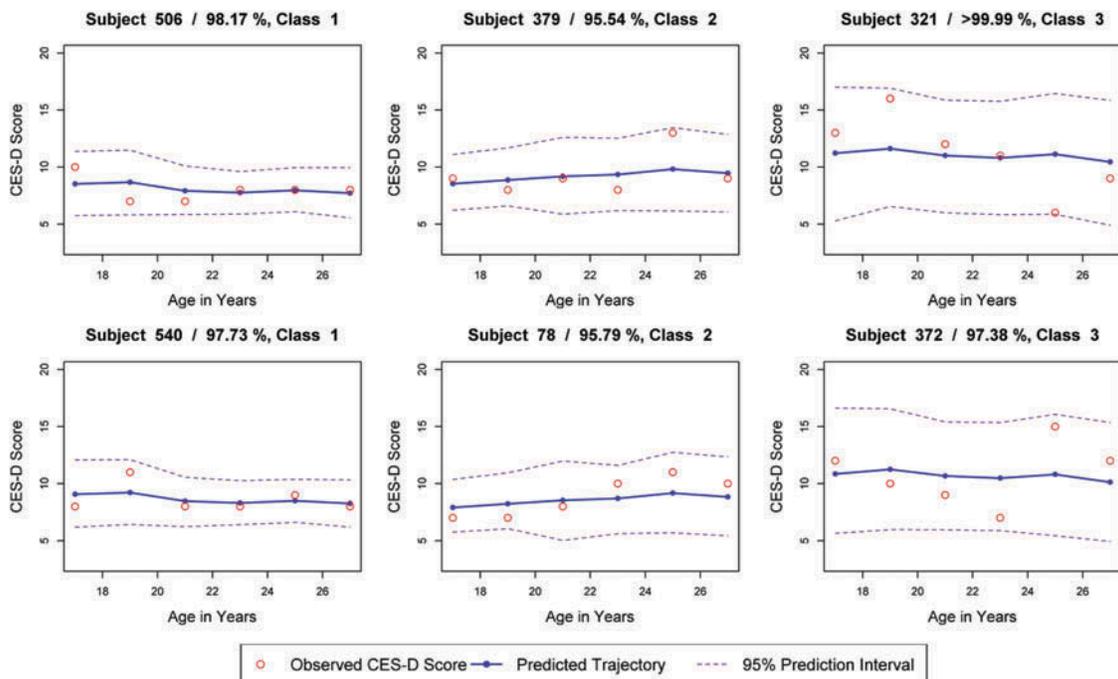


FIGURE 5 Empirical Example 2: Conditional predicted values and observed values of the Center for Epidemiological Studies Depression scale (CES-D) for six subjects.

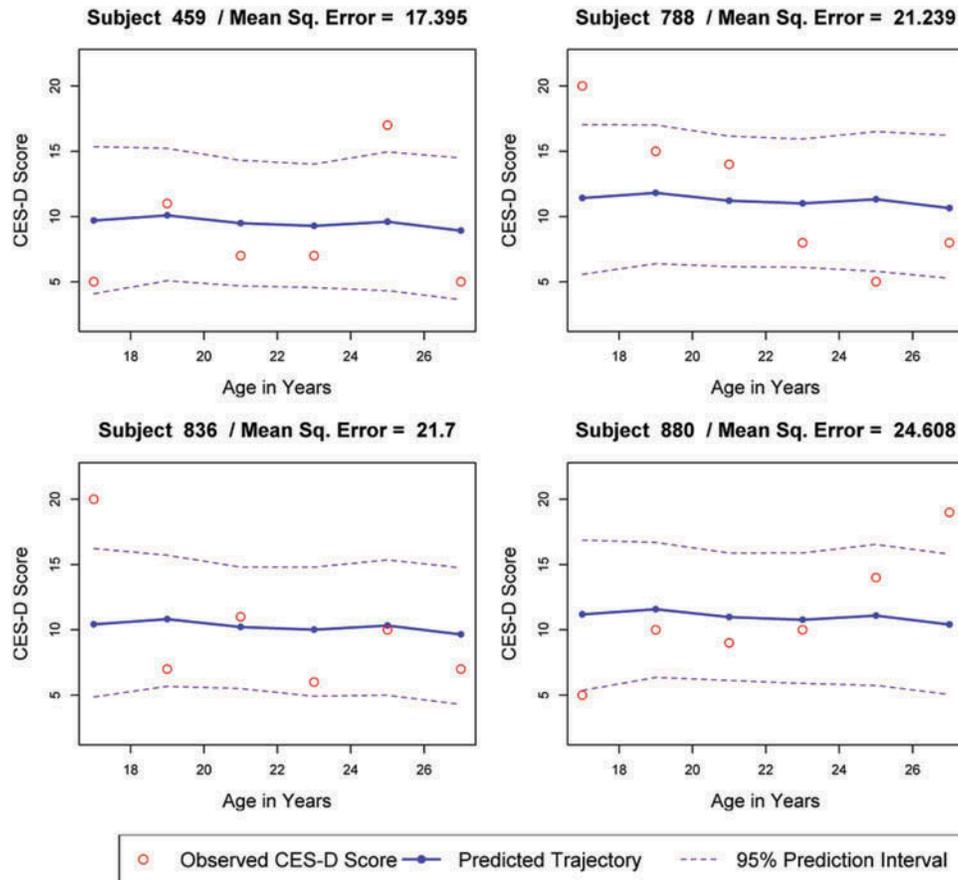


FIGURE 6 Empirical Example 2: Conditional predicted values and observed values of the Center for Epidemiological Studies Depression scale (CES-D) for four subjects with poor model fit.

prompting the analyst to consider refinements to the model (e.g., the addition or subtraction of a trajectory class). The predicted and observed values for four such cases are shown in Figure 6; among cases with complete data, these cases are the four with the highest mean squared distance between the observed and predicted data points. Visual examination suggests that the depressive symptoms of some individuals (particularly Subjects 788 and 880) are characterized by a more systematic trend than the model is capturing.

In sum, marginal predicted values helped us to illustrate how we would expect individuals differing in their covariate values to differ in their trajectories of depression, averaging over the distributions of the unknown latent variables. In contrast, conditional predicted values allowed us to incorporate information about the latent variables to examine the predictions and fit of the model at the individual level.

DISCUSSION

This report explored a method for using mixture model results to make inferences about individuals, whether hypothetical or observed. We described and illustrated the

use of both marginal and conditional predicted values. Additionally, a method for quantifying uncertainty around these predicted values was introduced. In the first empirical example, we explored the link between early marijuana use and subsequent nonmedical use of prescription medications. In the second example, we used predicted values to examine the roles of gender and college attendance as potential risk factors for the maintenance of depressive symptoms in early adulthood, and we examined how well the model fit the observed data at the individual level.

Making individual predictions based on mixture models represents a break from the more common practice of using these models for the sole purpose of making inferences about latent subpopulations or the population as a whole. Computing and using predicted values for individuals is, however, a natural extension of methods used in the multilevel modeling (Afshartous & De Leeuw, 2005; Skrondal & Rabe-Hesketh, 2009), latent curve modeling (Preacher et al., 2006), and IRT (Reise, 2000) frameworks. This work builds on these methods by showing how the results obtained from any mixture model—longitudinal or cross-sectional, with discrete indicators or continuous—can be used to obtain marginal and conditional individual-level predicted values. We have also discussed a

method for quantifying the uncertainty around predicted values that is tailored to the unique challenges presented by mixture models. In particular, the algorithms we present for forming CIs and PIs explicitly model uncertainty in class membership, which is critical in making inferences from mixture model results (Asparouhov & Muthén, 2016; Vermunt, 2010; Wang, Brown, & Bandeen-Roche, 2005). Our use of parametric bootstrapping also avoids the potential intractability of analytical solutions for some models, for instance, as might arise when specifying different distribution functions across classes (Fang, Li, & Sun, 2005; Mallick & Gelfand, 1994). As we previously discussed, this parametric bootstrapping approach makes a number of assumptions, but there are reasons to believe that the intervals that result might be relatively robust so long as the fitted model provides a sufficient approximation to the data generating process. More research is needed to clarify this possibility.

In sum, mixture models allow for the modeling of increasingly complex relationships between variables, and our understanding of the substantive implications of the results can often be aided by computing and plotting the individual-level predicted values (whether for hypothetical or sampled individuals). Moreover, focusing on the predictions these models afford for specific individuals brings their application into greater alignment with a person-centered data analytic approach.

ACKNOWLEDGMENTS

The authors would like to thank Danielle Dean and Jolynn Pek for helpful comments on this article.

FUNDING

This work was supported by National Institutes of Health Grants F31 DA040334 (Fellow: Veronica Cole) and R01 DA034636 (PI: Daniel Bauer). The content is solely the responsibility of the authors and does not represent the official views of the National Institute on Drug Abuse or the National Institutes of Health.

REFERENCES

- Afshartous, D., & De Leeuw, J. (2005). Prediction in multilevel models. *Journal of Educational and Behavioral Statistics, 30*, 109–139. doi:10.3102/10769986030002109
- Arminger, G., & Schoenberg, R. J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika, 54*, 409–425. doi:10.1007/BF02294626
- Asparouhov, T., & Muthén, B. (2016). Structural equation models and mixture models with continuous nonnormal skewed distributions. *Structural Equation Modeling, 23*, 1–19. doi:10.1080/10705511.2014.947375
- Bauer, D. J., Baldasaro, R. E., & Gottfredson, N. C. (2012). Diagnostic procedures for detecting nonlinear relationships between latent variables. *Structural Equation Modeling, 19*, 157–177. doi:10.1080/10705511.2012.659612
- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8*, 338. doi:10.1037/1082-989X.8.3.338
- Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*, 3. doi:10.1037/1082-989X.9.1.3
- Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 255–284). Mahwah, NJ: Erlbaum.
- Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology, 9*, 291–319. doi:10.1017/S095457949700206X
- Blanco, C., Okuda, M., Wright, C., Hasin, D. S., Grant, B. F., Liu, S.-M., & Olfson, M. (2008). Mental health of college students and their non-college-attending peers: Results from the national epidemiologic study on alcohol and related conditions. *Archives of General Psychiatry, 65*, 1429–1437. doi:10.1001/archpsyc.65.12.1429
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–459. doi:10.1007/BF02293801
- Bolck, A., Croon, M., & Hagenars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*(1), 3–27. doi:10.1093/pan/mp001
- Boyd, C. J., McCabe, S. E., Cranford, J. A., & Young, A. (2006). Adolescents' motivations to abuse prescription medications. *Pediatrics, 118*, 2472–2480. doi:10.1542/peds.2006-1644
- Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean-and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York, NY: Plenum.
- Compton, W. M., & Volkow, N. D. (2006). Abuse of prescription drugs and the risk of addiction. *Drug and Alcohol Dependence, 83*, S4–S7. doi:10.1016/j.drugalcdep.2005.10.020
- Conijn, J. M., Emons, W. H., De Jong, K., & Sijtsma, K. (2015). Detecting and explaining aberrant responding to the Outcome Questionnaire–45. *Assessment, 22*, 513–524. doi:10.1177/1073191114560882
- Curran, P. J., Bauer, D. J., & Willoughby, M. T. (2004). Testing main effects and interactions in latent curve analysis. *Psychological Methods, 9*, 220. doi:10.1037/1082-989X.9.2.220
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application* (Vol. 1). Cambridge, UK: Cambridge University Press.
- Dean, D. O., Bauer, D. J., & Shanahan, M. J. (2014). A discrete-time multiple event process survival mixture (MEPSUM) model. *Psychological Methods, 19*, 251. doi:10.1037/a0034281
- Dean, D. O., Cole, V., & Bauer, D. J. (2015). Delineating prototypical patterns of substance use initiations over time. *Addiction, 110*, 585–594. doi:10.1111/add.12816
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 34*, 1–38.
- Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods, 18*, 186. doi:10.1037/a0031609
- deRoon-Cassini, T. A., Mancini, A. D., Rusch, M. D., & Bonanno, G. A. (2010). Psychopathology and resilience following traumatic injury: A latent growth mixture model analysis. *Rehabilitation Psychology, 55*(1), 1. doi:10.1037/a0018601
- Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika, 65*, 457–483. doi:10.1093/biomet/65.3.457

- Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, 57 53–56.
- Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice* (Vol. 96). Thousand Oaks, CA: Sage.
- Fang, H. B., Li, G., & Sun, J. (2005). Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scandinavian Journal of Statistics*, 32(1), 59–75. doi:10.1111/j.1467-9469.2005.00415.x
- Gfroerer, J. C., Greenblatt, J. C., & Wright, D. A. (1997). Substance use in the US college-age population: Differences according to educational status and living arrangement. *American Journal of Public Health*, 87(1), 62–65. doi:10.2105/AJPH.87.1.62
- Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, 24, 229–252. doi:10.1007/BF02289845
- Gottfredson, N. C., Bauer, D. J., Baldwin, S. A., & Okiishi, J. (2014). Using a shared parameter mixture model to estimate change during treatment when treatment termination is related to recovery speed. *Journal of Consulting and Clinical Psychology*, 82, 813–827. doi:10.1037/a0034831
- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6, 430–450. doi:10.1037/1082-989X.6.4.430
- Hoeksma, J. B., & Kelderman, H. (2006). On growth curves and mixture models. *Infant and Child Development*, 15, 627–634. doi:10.1002/(ISSN)1522-7219
- Huang, G. H., & Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, 69(1), 5–32. doi:10.1007/BF02295837
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam & J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (vol. 1, pp. 221–233). Berkeley, CA: University of California Press.
- Kandel, D. B. (2002). *Stages and pathways of drug involvement: Examining the gateway hypothesis*. Cambridge, UK: Cambridge University Press.
- Kelava, A., Nagengast, B., & Brandt, H. (2014). A nonlinear structural equation mixture modeling approach for nonnormally distributed latent predictor variables. *Structural Equation Modeling*, 21, 468–481. doi:10.1080/10705511.2014.915379
- Kovacevic, M., Huang, R., & You, Y. (2006). Bootstrapping for variance estimation in multi-level models fitted to survey data. In *ASA Proceedings of the Survey Research Methods Section* (pp. 3260–3269). Alexandria, VA.
- Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models*. Chicago, IL: McGraw-Hill/Irwin.
- Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, 20, 1–26. doi:10.1080/10705511.2013.742377
- Laursen, B. P., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, 52, 377–389. doi:10.1353/mpq.2006.0029
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44, 226–233.
- Mallick, B. K., & Gelfand, A. E. (1994). Generalized linear models with unknown link functions. *Biometrika*, 81, 237–245. doi:10.1093/biomet/81.2.237
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122. doi:10.1007/BF02294746
- Morin, A. J., Maiano, C., Nagengast, B., Marsh, H. W., Morizot, J., & Janosz, M. (2011). General growth mixture analysis of adolescents' developmental trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling*, 18, 613–648. doi:10.1080/10705511.2011.607714
- Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24, 882–891. doi:10.1111/j.1530-0277.2000.tb02070.x
- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463–469. doi:10.1111/j.0006-341X.1999.00463.x
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4, 139. doi:10.1037/1082-989X.4.2.139
- Nagin, D. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.
- Nagin, D. S., & Tremblay, R. E. (2005). Developmental trajectory groups: Fact or a useful statistical fiction?. *Criminology*, 43, 837–904. doi:10.1111/j.1745-9125.2005.00026.x
- Pek, J., Chalmers, R. P., Kok, B. E., & Losardo, D. (2015). Visualizing confidence bands for semiparametrically estimated nonlinear relations among latent variables. *Journal of Educational and Behavioral Statistics*, 40, 402–423. doi:10.3102/1076998615589129
- Pek, J., Losardo, D., & Bauer, D. J. (2011). Confidence intervals for a semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, 18, 537–553. doi:10.1080/10705511.2011.607072
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448. doi:10.3102/10769986031004437
- Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 6(2), 77–98. doi:10.1080/19312458.2012.679848
- Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3), 385–401.
- Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, 33, 565–576. doi:10.1177/0165025409343765
- Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, 52(1), 501–525. doi:10.1146/annurev.psych.52.1.501
- Reise, S. P. (2000). Using multilevel logistic regression to evaluate person-fit in IRT models. *Multivariate Behavioral Research*, 35, 543–568. doi:10.1207/S15327906MBR3504_06
- Robins, L. N., Darvish, H. S., & Murphy, G. E. (1970). The long-term outcome for adolescent drug users: A follow-up study of 76 users and 146 nonusers. In J. Zubin & A. M. Freedman (Eds.), *Proceedings of the annual meeting of the American Psychopathological Association* (Vol. 59, pp. 159–178). New York, NY: Grune & Stratton.
- Satorra, A. (1990). Robustness issues in structural equation modeling: A review of recent developments. *Quality and Quantity*, 24, 367–386. doi:10.1007/BF00152011
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464. doi:10.1214/aos/1176344136
- Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, 66, 563–575. doi:10.1007/BF02296196

- Skrondal, A., & Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*, 659–687. doi:10.1111/rssa.2009.172.issue-3
- Sterba, S. K., & Bauer, D. J. (2010). Matching method with theory in person-oriented developmental psychopathology research. *Development and Psychopathology*, *22*, 239–254. doi:10.1017/S0954579410000015
- Sterba, S. K., & Bauer, D. J. (2014). Predictions of individual change recovered with latent class or random coefficient growth models. *Structural Equation Modeling*, *21*, 342–360. doi:10.1080/10705511.2014.915189
- Titterton, D. M., Smith, A. F., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions* (Vol. 7). New York, NY: Wiley.
- Tucker, L. R. (1971). Relations of factor score estimates to their use. *Psychometrika*, *36*, 427–436. doi:10.1007/BF02291367
- Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling*, *17*, 165–192. doi:10.1080/10705511003659318
- Van Der Leeden, R., Meijer, E., & Busing, F. M. (2008). Resampling multilevel models. In J. deLeeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 401–433). New York, NY: Springer.
- Van Horn, M. L., Smith, J., Fagan, A. A., Jaki, T., Feaster, D. J., Masyn, K., ... Howe, G. (2012). Not quite normal: Consequences of violating the assumption of normality in regression mixture models. *Structural Equation Modeling*, *19*, 227–249. doi:10.1080/10705511.2012.659622
- Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*, 450–469. doi:10.1093/pan/mpq025
- Wang, C. P., Brown, C. H., & Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: Examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association*, *100*, 1054–1076. doi:10.1198/016214505000000501
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–25. doi:10.2307/1912526
- Wiesner, M., & Windle, M. (2004). Assessing covariates of adolescent delinquency trajectories: A latent growth mixture modeling approach. *Journal of Youth and Adolescence*, *33*, 431–442. doi:10.1023/B:JOYO.0000037635.06937.13
- Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2008). Detection of aberrant responding on a personality scale in a military sample: An application of evaluating person fit with two-level logistic regression. *Psychological Assessment*, *20*, 159–168. doi:10.1037/1040-3590.20.2.159
- Yuan, K. H., & Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association*, *92*, 767–774. doi:10.1080/01621459.1997.10474029
- Yung, Y. F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 195–226). Mahwah, NJ: Erlbaum.