Two-Part Growth Mixture Modeling

Bengt O. Muthén
University of California, Los Angeles

Rough Draft, Do Not Distribute

*

October 22, 2001

*I thank Tihomir Asparouhov for valuable discussions and computational assistance. May Tham, Ching-yun Yu, and Kookhee Kim have provided helpful assistance with displaying the results. The research was supported by grant K02 AA 00230-01 from NIAAA, by grant 40859 from NIMH, and by a grant from NIDA. The work has also benefitted from discussions in Bengt Muthén's Research Apprenticeship Course Hendricks Brown's Prevention Science Methodology group. The email address is bmuthen@ucla.edu.

1 Introduction

This paper considers the analysis of repeated measures data. Conventional random effects growth modeling in the tradition of Laird and Ware (1982) represents unobserved heterogeneity among subjects in the form of random effects, i.e. continuous latent variables. Growth mixture modeling (Muthén & Shedden, 1999; Muthén, 2001a, b; Muthén, Brown, Masyn, Jo, Khoo, Yang, Wang, Kellam, Carlin, & Liao, 2000; Muthén & Muthén, 1998-2001, Appendix 8) offers an important extension of conventional modeling in that more general forms of unobserved heterogeneity can be captured using categorical latent variables (latent classes). Growth mixture modeling as implemented in the Mplus software (Muthén & Muthén, 1998-2001) allows for latent classes that may have different shapes, antecedents, and consequences. A related longitudinal technique, latent class growth analysis (Nagin, 1999), also studies unobserved heterogeneity in the form of categorical latent variables. Growth mixture modeling, however, allows categorical and continuous heterogeneity jointly, capturing potential further continuous heterogeneity among individuals within the latent classes.

1.1 Non-Normal Outcomes

Growth mixture modeling is also useful for describing growth in outcomes that can be seen as continuous but non-normally distributed. For continuous outcomes, Mplus uses the assumption of within-class conditional normality for the set of repeated measures given the set of covariates. This means that the outcomes are not necessarily normal within class due to non-normal covariates. More importantly, the mixture distribution can obtain a very non-normal shape, especially with a majority class having the lowest (or highest) mean and low variance combined with minority classes with higher (or lower) means. This is illustrated by the hypothetical 2-class situation shown in Figure 1. Similar types of distributions were seen in the Muthén, et al. (2000) analysis of teacher-rated aggressive behavior of children in classrooms. Histograms for 200 children observed at 9 time points in grades 1-7 are shown in Figure 2 (the bar to the left of the zero point represents missing data).

FIGURE 1

FIGURE 2

A type of non-normality that cannot be well captured by mixtures of normal distributions arises in studies where a significant number of individuals are at the lowest value of an outcome, for example representing absence of a behavior. Applications include alcohol, drug, and tobacco use among adolescents. Figures 3 and 4 show examples of such situations using alcohol data from 1234 individuals in the Alcohol Misuse Prevention Study (AMPS; Maggs & Schulenberg, 1998) in grades 6-12 (the bar to the left of the zero point represents missing data). Figure 3 shows histograms for the variable TFQ, frequency x quantity of beer, wine, hard liquor measured separately and then summed and collapsed into a 12-point scale for number of drinks, scored as 0 (no drinking), 1 ($<$ 1/year), 2 (1 to 2/year), 3 ($>$ 2 to $<$ 6/year), 4 (6 to $<$ 12/year), 5 (1 to 2/month), 6 ($>$ 2 to $<$ 4/month), 7 (1 to 2/week), 8 ($>$ 2 to 4/week), 9 ($>$ 4 to 7/week), 10 (

> 7 to 10/week), 11 (> 10 to 14/week), 12 (> 14/week). Figure 4 shows histograms for the variable AMOVER, which is the average of three items related to alcohol misuse: During the past 12 months, how many times did you - drink more than you planned to? - feel sick to your stomach after drinking? - get very drunk? Responses were: 0 (never), 1 (once), 2 (two times), 3 (three or more times).

FIGURE 3

FIGURE 4

Figures 3 and 4 show that the lowest value, the absence of the behavior in question, is often the most common outcome. Special modeling is required to properly represent the large percentage of individuals not engaging in the behavior. These distributions cannot be well represented by normal mixtures as in Figure 1. For values above the lowest, however, the distributions are similar to those of figures 1 and 2.

1.2 Related Research

Censored-normal models are often used for outcomes of the kind shown in figures 3 and 4, including classic Tobit regression analysis (Amemiya, 1985; Tobin, 1958) and latent class growth modeling in the TRAJ program (Jones, Nagin, & Roeder, 1998). A recent article by Olsen and Schafer (2001) gives an excellent overview of several related modeling efforts. Censored-normal models have been critized, see e.g. Duan, Manning, Morris, Newhouse (1983), because of the limitation of assuming that the same set of covariates influences both the decision to engage in the behavior and the amount observed. A two-part modeling approach proposed in Olsen and Schafer (2001) avoids this limitation and will be utilized here.

To simplify the discussion, the lowest value will be taken to be zero from here on. It is useful to distinguish between two kinds of zero outcomes. First, individuals may have zero values at a given time point because their behavioral activity is low and is zero during certain periods (”random zeroes”). Second, individuals may not engage in the activity at all and therefore have zeroes throughout all time points of the study (”structural zeroes”). Olsen and Schafer (2001) proposed a two-part model for the case of random zeroes, whereas Carlin, Wolfe, Brown, and Gelman (2001) considered the case of structural zeroes. In both articles, a random-effects logistic regression was used to express the probabilities of non-zeroes versus zeroes.

Olsen and Schafer (2001) studied alcohol use in grades 7-11. To capture the changing zero status across time, the logistic regressions for each time point were expressed as a random-effects growth model. The term two-part model refers to having both a logistic model part to model the probability of non-zero versus zero outcomes (part 1) and a continuous-normal model part for the values of the non-zero outcomes (part 2). In Olsen and Schafer (2000), the two parts have correlated random effects. The two parts are also allowed to have different covariates, avoiding the limitation of censored-normal modeling.

Carlin et al. (2001) studied cigarette smoking among adolescents. A 2-class model was used with a ”zero class” (structural zeroes) representing individuals not susceptible to regular smoking (also referred to as "immunes"). As pointed out in Carlin et al. (2001), an individual with zeroes throughout the study does not necessarily belong to the zero class, but may show zeroes by chance. In their analysis, the estimated proportion of immunes was 69%, while the empirical proportion with all zeroes was 77%. Because of this, an ad hoc analysis based on deleting individuals with all zeroes may lead to distorted results.

Although cross-sectional in nature, related work also includes Deb, Hall, Trivedi and Hall (2000) who compared a two-part model for users and nonusers of health care with a two-class model for infrequent and frequent users (see also Deb, 2001; Deb & Holmes, 2000).

As is clear from figures 3 and 4, missing data is common in longitudinal studies so that whether or not an individual is in the zero category may not be directly observable. Both Olsen and Schafer (2001) and Carlin et al. (2001) allow for missing data, using maximum-likelihood estimation under MAR (missing at random; Little & Rubin, 1987). In this way, individuals with a combination of observed zeroes and missing values can be classified as belonging to the zero class.

1.3 A Generalized Growth Mixture Model

Inspired by Olsen and Schafer (2001) and Carlin et al. (2001), this paper proposes a generalization of growth mixture modeling to handle both random and structural zeroes in a two-part model. Multiple latent classes are used to represent the growth in the probability of non-zero values in part 1 as well as to represent the growth in the non-zero outcomes in part 2. For the part 1 modeling of the probability of non-zero values, this represents a latent class growth alternative to the random effects modeling of Olsen and Schafer (2001) and Carlin et al. (2001), i.e. a model in line with Nagin (1999).

The use of latent classes for the part 1 modeling of the probability of non-zero values may be seen as a semi-parametric alternative to a random effects model in line with Aitkin (1999). Maximum-likelihood estimation for logistic models with random effects typically use Gauss-Hermite quadrature to integrate out the normal random effects. The quadrature uses fixed nodes and weights for a set of quadarture points. As pointed out by Aitkin (1999), a more flexible distributional form is obtained if both the nodes and the weights are estimated and this approach is equivalent to mixture modeling. Aitkin (1999) argues that the mixture approach may be particularly suitable with categorical (binary) outcomes where the usual normality assumption for the random effects has scarce empirical support.

In addition to accounting for random zeroes as in Olsen-Schafer, the proposed part 1 approach incorporates Carlin et al.'s concept of a zero class that has zero probability of non-zero values throughout the study. A further advantage of the proposed approach is that covariates are allowed to have different influence in different classes. This feature has proven important in growth mixture applications (Muthén, 2001a). For the part 2 modeling of the non-zero outcomes, the proposed modeling extends the Olsen-Schafer growth model to a growth mixture model, i.e. letting both categorical and continuous latent variables capture unobserved heterogeneity. This is often necessary to represent qualitatively different types of growth. The classes in part 2 do not necessarily all have to correspond to those of part 1. It is shown that the proposed two-part growth mixture model can be fitted into the general latent variable modeling framework of Mplus (Muthén & Muthén, 1998-2001; Appendix 8), using maximum-likelihood estimation under MAR via the EM algorithm. Mplus input for the analyses presented below can be found at www.statmodel.com.

The remainder of the paper is organized as follows. Section 2 gives a brief overview of the Olsen-Schafer model. Section 3 presents the proposed two-part growth mixture model, including practical aspects of implementing it in Mplus. Section 4 describes in-depth analyses of data from the alcohol misuse study mentioned in the introduction. Section 5 concludes.

2 Two-Part Modeling

In line with Olsen and Schafer (2001), let u_{it} $(i = 1, \ldots, n; t = 1, 2, \ldots, n_i)$ be a binary variable representing the zero or non-zero value of the semicontinuous repeated measures outcome y_{it},

$$u_{it} = \begin{cases} 1 & \text{if } y_{it} > 0 \\ 0 & \text{if } y_{it} = 0, \end{cases} \tag{1}$$

and

$$y_{it} = \begin{cases} m_{it} & \text{if } y_{it} > 0 \\ \mathit{ignored} & \text{if } y_{it} = 0, \end{cases} \tag{2}$$

where m_{it} represents a growth model. For instance, with linear growth,

$$m_{it} = \eta_{0i} + \eta_{1i}\ a_{it} + \epsilon_{it}, \tag{3}$$

where η_{0i} is a random intercept and η_{1i} is a random slope with means α_0, α_1 and covariance matrix $\boldsymbol{\Psi}$, a_{it} represents an age-related variable (such as grade), and ϵ_{it} are time-specific, normally distributed residuals with means zero and covariance matrix $\boldsymbol{\Theta} = \theta \mathbf{I}$.

Olsen and Schafer (2001) specifies a random effects logit model for u. For instance, with linear growth, the logit is expressed as

$$l_{it} = log[P(u_{it} = 1|\eta_{u0i}, \eta_{u1i})/(1 - P(u_{it} = 1|\eta_{u0i}, \eta_{u1i}))] = \eta_{u0i} + \eta_{u1i}\ a_{it}, \tag{4}$$

with right-hand-side quantities defined in line with (3). The random effects of η_{0i}, η_{1i}, η_{u0i}, and η_{u1i} are assumed to follow a joint normal distribution, allowing for correlations between the u and y growth processes.

Continuing the example of linear growth, let $\boldsymbol{\eta}_i = (\boldsymbol{\eta}'_{ui}, \boldsymbol{\eta}'_{yi})'$, where $\boldsymbol{\eta}_{ui} = (\eta_{u0i}, \eta_{u1i})'$ and $\boldsymbol{\eta}_{yi} = (\eta_{0i}, \eta_{1i})'$. Drawing on Olsen and Schafer (2001), the log likelihood for this model can be expressed as $\sum_{i=1}^{n} log\ L_i$ with

$$L_i = \int [\mathbf{u}_i|\boldsymbol{\eta}_i]\ [\mathbf{y}_i|\boldsymbol{\eta}_i]\ [\boldsymbol{\eta}_i]\ d\boldsymbol{\eta}_i \tag{5}$$

$$= \int [\mathbf{u}_i|\boldsymbol{\eta}_{ui}]\ [\boldsymbol{\eta}_{ui}]\ (\int [\mathbf{y}_i|\boldsymbol{\eta}_{ui}, \boldsymbol{\eta}_{yi}]\ [\boldsymbol{\eta}_{yi}|\boldsymbol{\eta}_{ui}]\ d\boldsymbol{\eta}_{yi}\ )\ d\boldsymbol{\eta}_{ui}. \tag{6}$$

Due to conditional independence of the u_t's given $\boldsymbol{\eta}_u$,

$$log\ [\mathbf{u}_i|\boldsymbol{\eta}_{ui}] = \sum_{t=1}^{n_i} [u_{it}\ log\ l_{it} + (1 - u_{it})\ log\ (1 - l_{it})]. \tag{7}$$

Furthermore, $[\boldsymbol{\eta}_{yi}|\boldsymbol{\eta}_{ui}]$ is a conditional normal density and $[\mathbf{y}_i|\boldsymbol{\eta}_{ui}, \boldsymbol{\eta}_{yi}] = [\mathbf{y}_i|\boldsymbol{\eta}_{yi}]$ is a conditional normal density for the $n_i^* \times 1$ vector $\mathbf{y}_i$, where n_i^* is the number of relevant y outcomes for individual i, i.e. $y_{it} > 0$ outcomes. The fact that zero y outcomes are excluded from the y part of the likelihood is a key feature of the two-part approach. Olsen and Schafer (2001) use maximum-likelihood estimation, solving the numerical integrations involved in (6) by Laplace approximation.

3 Mixture Modeling

The two-part growth mixture modeling proposed here uses the same general idea as Olsen-Schafer, but changes the random effects specification of the logit for the u part into a mixture model and combines this with a growth mixture model for the y part. The mixture modeling draws on the formulation in Muthén and Muthén (1998-2001, Appendix 8) as implemented in the Mplus program.

The mixture is represented by $\mathbf{c}_i$, a latent categorical variable with K classes, $\mathbf{c}_i = (c_{i1}, c_{i2}, \ldots, c_{iK})'$, where $c_{ik} = 1$ if individual i belongs to class k and zero otherwise. For simplicity, assume that the age-related variable $a_{it} = a_t$ and let T denote the maximum number of time points, allowing individuals to have missing data at some time points. Define the $T \times 1$ logit vector for the u's, $\mathbf{l}_i = (l_{i1}, l_{i2}, \ldots, l_{in_i})'$. Conditional on an individual being in class k, the growth model for u is expressed as

$$\mathbf{l}_i = \boldsymbol{\Lambda}_{u_k}\ \boldsymbol{\eta}_{ui} + \mathbf{K}_{u_k}\ \mathbf{x}_i, \tag{8}$$

$$\boldsymbol{\eta}_{ui} = \boldsymbol{\alpha}_{u_k} + \boldsymbol{\Gamma}_{u_k}\ \mathbf{x}_i. \tag{9}$$

Here, $\boldsymbol{\Lambda}_{u_k}$ specifies the growth shape, e.g. with linear growth and η_{u0} defined as the initial status,

$$\boldsymbol{\Lambda}_{u_k} = \begin{pmatrix} 1 & 0 \\ 1 & a_2 \\ \vdots & \\ 1 & a_T \end{pmatrix}, \tag{10}$$

i.e. held constant across classes. The vector $\mathbf{x}_i$ contains time-varying and time-invariant covariates. The effects of time-varying covariates are captured in $\mathbf{K}_{u_k}$, while the effects of time-invariant covariates are captured in $\boldsymbol{\Gamma}_{u_k}$. Conditional independence is assumed for the u's given $\mathbf{c}_i$ and $\mathbf{x}_i$,

$$P(u_{i1}, u_{i2}, \ldots, u_{iT}|\mathbf{c}_i, \mathbf{x}_i) = P(u_{i1}|\mathbf{c}_i, \mathbf{x}_i)\ P(u_{iT}|\mathbf{c}_i, \mathbf{x}_i) \ldots P(u_{ir}|\mathbf{c}_i, \mathbf{x}_i). \tag{11}$$

In the Olsen-Schafer model, the logits of (4) vary across individuals, whereas conditional on $\mathbf{c}_i$, $\mathbf{x}_i$, the logits of (8) do not. The non-independence of the u's across time instead arises through the mixture across the K classes similar to latent class analysis (see, e.g. Clogg, 1995). The classes are typically defined as different trends expressed by different α_{u_k} values in (9). This is a latent class growth analysis model in line with Nagin (1999).

The mixture model relates $\mathbf{c}$ to $\mathbf{x}$ by multinomial logistic regression

$$P(c_{ik} = 1|\mathbf{x}_i) = \frac{e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k}\mathbf{x}_i}}{\sum_{j=1}^{K} e^{\alpha_{c_j} + \boldsymbol{\gamma}'_{c_j}\mathbf{x}_i}}, \tag{12}$$

where the last class is a reference class with coefficients standardized to zero, $\alpha_{c_K} = 0$, $\boldsymbol{\gamma}_{c_k} = \mathbf{0}$.

In Olsen and Schafer (2001), the u and the y processes are related through the correlations between their two sets of random effects. The counterpart in the mixture modeling is to let the distribution of random effects for y vary across the latent classes. Following Muthén and Muthén (1998-2001, Appendix 8), multivariate normality is assumed for $\mathbf{y}$ conditional on $\mathbf{x}$ and class k,

$$\mathbf{y}_i = \mathbf{\Lambda}_k \, \boldsymbol{\eta}_{yi} + \mathbf{K}_k \, \mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{13}$$

$$\boldsymbol{\eta}_{yi} = \boldsymbol{\alpha}_k + \mathbf{\Gamma}_k \, \mathbf{x}_i + \boldsymbol{\zeta}_i, \tag{14}$$

where the residual vector $\boldsymbol{\epsilon}_i$ is $N(\mathbf{0}, \mathbf{\Theta}_k)$ and the residual vector $\boldsymbol{\zeta}_i$ is $N(\mathbf{0}, \mathbf{\Psi}_k)$, both assumed to be uncorrelated with other variables. The parameter interpretation is analogous to that of (8), (9). In line with growth mixture modeling, a basic model would only let the means $\boldsymbol{\alpha}_k$ vary across classes, whereas some applications may require the covariance matrices $\mathbf{\Psi}_k$, $\mathbf{\Theta}_k$ to vary as well (see Muthén, 2001 a,b; Muthén, Brown, Masyn, Jo, Khoo, Yang, Wang, Kellam, Carlin, & Liao, 2000).

Mplus carries out maximum-likelihood estimation using the EM algorithm, allowing for missing data under MAR (Little & Rubin, 1987). The observed-data log likelihood is $\sum_{i=1}^{n} log \; L_i$, where

$$L_i = [\mathbf{y}_i, \mathbf{u}_i | \mathbf{x}_i], \tag{15}$$

where $[\mathbf{y}_i, \mathbf{u}_i | \mathbf{x}_i]$ is a mixture distribution defined as

$$\sum_{k=1}^{K} P(c_{ik} = 1 | \mathbf{x}_i) \; N(y_{ij}; \boldsymbol{\mu}_i, \mathbf{\Sigma}_i) \; [\mathbf{u}_i | c_{ik} = 1, \mathbf{x}_i], \tag{16}$$

where $\boldsymbol{\mu}_i$ and $\mathbf{\Sigma}_i$ class-specific mean vectors and covariance matrices derived from (13) and (14).

In the context of the Olsen-Schafer logit model in (4), the α_{u_k} values of (9) can be seen as points along the axes of the joint random effects distribution. In line with this, Aitkin (1999) points out that these mixture parameters can be related to the nodes of a quadrature when integrating out the random effects and that the mixture probabilities correspond to the weights of the quadrature. The mixture can therefore be seen as a non-parametric representation of the normal distribution of $\boldsymbol{\eta}_{ui}$, for example allowing for a skewed distribution that may better represent the data. Essentially, the integrals of (6) are replaced by sums over the latent classes as in (16).

3.1 Two-Part Growth Mixture Modeling In Practice

To carry out two-part growth mixture modeling in the Mplus framework (Muthén & Muthén, 1998-2001) a vector of T u's scored 0 or 1 needs to be created for each person based on the corresponding y being zero or greater than zero. A missing value for y translates into a missing value for u. In addition, a key feature is to change each y that is zero into missing data. Using the MAR ML approach, this has the effect of ignoring such y's in the evaluation of the likelihood, as is done in the Olsen-Schafer approach. In this way, two-part growth mixture modeling is in principle a straightforward application of Mplus mixture modeling of $\mathbf{u}$ and $\mathbf{y}$ jointly. A good modeling strategy is, however, needed given the flexibility of the model. The following five steps are recommended.

3.1.1 Step 1: Latent Class Analysis of u

A useful starting point is analysis of the u part alone. In step 1, conventional latent class analysis provides an exploratory tool for finding the number and shapes of the different trend classes. Here, the choice of number of classes may be based on the chi-square test against the unrestricted multinomial distribution, the Bayesian Information Criterion (BIC; the lower the value the better the model) and the classification quality based on posterior probabilities as outlined in Muthén and Muthén (1998-2001, Appendix 8); see also Muthén and Muthén (2000).

3.1.2 Step 2: Latent Class Growth Analysis of u

In step 2, a latent class growth analysis is explored (see, e.g. Muthén, 2001, b; Nagin, 1999). Unlike the latent class analysis, latent class growth analysis explicitly utilizes the fact that the different u variables correspond to repeated measures over time, letting the latent classes correspond to different trends. The number of classes and the choice of growth shapes can be based on the latent class analysis of step 1. Latent class growth analysis uses the modeling structure of (8) and (9).

3.1.3 Step 3: Two-Part Growth Mixture Modeling of u and y jointly

Next, the y part is added to the model. The y part of the model uses the model structure in (13), (14). In step 3, a growth mixture model is specified for the y part of the model using the same number of classes as for the u part and letting only the random effect means vary across the classes. The classes found for the u part of the model may, however, not be sufficient to describe the heterogeneity in the y development. The fit of the model to the y data can be studied by comparing for each class estimated moments with moments created by weighting the individual data by the estimated conditional probabilities (Roeder, Lynch & Nagin, 1999). To check how closely the estimated average curve within each class matches the individual data, it is also useful to randomly assign individuals to classes based on individual estimated conditional class probabilities. "Pseudo-class" plots of the observed individual trajectories together with the model-estimated average trajectory can be used to check the model (Bandeen-Roche et al., 1997).

3.1.4 Step 4: Further Exploration of the y Part

The fit of the model may be improved if classes are added that capture further variation in the y part in line with regular growth mixture modeling of y. Step 4 explores the need for additional y classes. As an aid to this, it is useful to perform a separate analysis of the y part of the model. This is different from regular growth mixture modeling given that observations with $y = 0$ have been rescored as missing data. Step 4 also explores the need for allowing class-varying variances for the random effects and/or for the time-specific residuals.

3.1.5 Step 5: Two-Part Growth Mixture Modeling with Covariates

Step 5 adds covariates. Covariates may influence the class membership as in (12), the u outcomes and trend factors as in (8), (9), and the y outcomes and random effects as in (13), (14).

4 An Example: Analysis of Alcohol Misuse

The alcohol misuse outcome AMOVER from the AMPS study, shown in Figure 4 above, serves as an illustrative example of the proposed two-part growth mixture modeling. The AMPS study was administered in forty-nine schools in south-eastern Michigan with a total of $2,666$ students measured seven times from Fall of grade 6 to Spring of grade 12, starting in 1984 (Maggs & Schulenberg, 1998). The current analyses focus on the random half of the sample who were measured in grade 6 and did not have missing data on AMOVER, resulting in $1,234$ students. Each analysis is carried out with and without the covariates of gender (female $= 0$, male $= 1$) and ethnicity (white $= 0$, black $= 1$).

To give a comparison with the results of the two-part growth mixture approach proposed in this paper, results from conventional growth modeling, Olsen-Schafer two-part growth modeling, and growth mixture modeling will first be presented including a comparison of their estimated mean curves and the influence of the two covariates.

4.1 Conventional, Olsen-Schafer Two-Part, and Growth Mixture Modeling

4.1.1 Conventional Modeling

A conventional random effects model was fitted to the development of the alcohol misuse outcome AMOVER over the seven time points. For simplicity, a linear model was chosen. Residual variances were allowed to vary across time. The log likelihood value, # parameters, BIC, and chi-square (df) test against the unrestricted model are: $-4,772.311$, 12, $9,630.038$, and 268.698 (23). The two random effects both have significant variation. Their means are estimated as 0.18 (intercept) and 0.13 (slope). A significant improvement in fit was obtained when allowing correlations between adjacent time-specific residuals, chi-square (df) = 55.647 (17). Here, the log likelihood value, # parameters, and BIC values are: $-4,665.786$, 18, and $9,459.696$. This model modification, however, did not alter the significance of the random effects variances and only slightly changed their estimated means (same to two decimals), so this refinement of the model will not be further considered. It should be noted that the conventional modeling uses maximum-likelihood estimation under the assumption of normally distributed outcomes. The chi-square test of model fit is most likely strongly inflated due to the considerable deviation from normality in these data.

Adding the two covariates Male and Black showed that the intercept had a significant positive relation to Male and an insignificant relation to Black, while the slope had a significant negative relation to Male and a significant negative relation to Black. When allowing for correlated errors, the slope was no longer significantly related to Male although the estimate was still negative.

4.1.2 Growth Mixture Modeling

A linear growth mixture was applied to 2, 3, 4, 5, and 6 classes. The analyses used a class-invariant random effects covariance matrix and class-invariant residual variances allowed to vary across time but having zero correlations. A proper 6-class solution could not be found but resulted in a singular information matrix using several sets of starting values. The 5-class solution resulted in the best BIC value. For this model, the log likelihood value, # parameters, BIC, and entropy values are: $-3,869.041$, 24, $7,908.915$, and 0.881. The within-class variance of the slope is significant, while the variance of the intercept is marginally insignificant. The estimated mean curves and the class percentages are shown in Figure 5.

FIGURE 5

Adding the two covariates Male and Black, showed significant effects on both class membership and the random effects. Being male was found to give a significantly higher odds of being in the high class 1 versus the low normative class 5, as well as being in class 2 versus class 5. Being black was found to give a significantly lower odds of being in the high class 1 versus the low normative class 5, as well as being in class 2 versus class 5. The only significant effect of the covariates on the within-class intercept and slope for y is a negative effect of being black on the slope.

4.1.3 Olsen-Schafer Two-Part Modeling

For the Olsen-Schafer two-part model, linear growth with random intercepts and slopes was used in both parts, letting all four random effects correlate. Integrating out the two random effects for the y part, this reduces to a two-dimensional quadrature for the two random effects for the u part. Here, a 10×10 Gauss-Hermite quadrature was used. The computations were carried out using a not yet available, experimental version of Mplus.

The correlations between the four random effects were estimated as -0.40 (intercept part 1, slope part 1), 0.91 (intercept part 1, intercept part 2), -0.01 (slope part 1, intercept part 2), 0.76 (slope part 1, slope part 2), -0.18 (intercept part 2, slope part 2), where the 0.91 correlation between the two intercepts illustrates the finding in Olsen and Schafer (2001) of high correlation between the two processes. The variances for the random effects were all significantly different from zero, and the random effect means estimated as -2.14 (intercept of part 1), 0.61 (slope of part 1), 0.32 (intercept of part 2), 0.17 (slope of part 2). Here, log likelihood $= -5,203.947$ with 21 parameters, BIC $= 10,654.080$, and $\chi^2_{LR}(122) = 184.985$. The log likelihood and BIC values are not comparable to those of the conventional and growth mixture model analyses because the two-part model brings in the u variables.

Adding the two covariates Male and Black, it was found that they had significant influence on the random effects in both parts of the model. For the u probabilities of part 1, the intercept is not significantly influenced by either covariate, while the slope has a marginally significant negative influence of being male and a significantly negative influence of being black. For the y outcomes in part 2, the only significant effect is a negative influence on the slope of being black.

4.1.4 Comparison of the Three Methods

The differences between the analyses presented above may be summarized as follows. The estimated mean curve for y using the Olsen-Schafer two-part model is considerably higher than for the conventional growth model with an intercept and slope mean of 0.32, 0.17 versus 0.18, 0.13. This is to be expected given that the two-part model estimates refer to individuals who engage in the alcohol misuse at the respective time point, excluding random zeroes, while the conventional model estimates refer to everyone. The estimates from the growth mixture model can be used to compute a mixture estimate over the five classes for the intercept and slope mean, 0.17, 0.07. As expected, these values are closer to those of the conventional model.

Regarding the influence of covariates on the y intercept, the conventional model found a significant positive influence of Male for the intercept and negative influence of Male and Black for the slope, while the growth mixture model and two-part model only found the slope to be significantly negatively influenced by Black. The growth mixture model and the two-part model, however, also have significant influence of the covariates on the latent class and u part of the model, respectively. For the growth mixture model, Male has a positive influence on being in high classes, while Black has a negative influence on being in high classes. For the two-part model, Male and Black have negative effects on the slope of the probability of $y > 0$. In sum, the influence of covariates is assessed differently across the three methods. In particular, the two-part model does not show the same degree of increased alcohol misuse for males as the other two models.

4.2 Two-Part Growth Mixture Modeling

In line with the analysis steps for two-part growth mixture modeling outlined in Section 2, step 1 performs a latent class analysis on the u part of the model using an increasing number of classes. Based on this, step 2 performs a latent class growth analysis focusing more specifically on trends over time. Following this, steps 3 and 4 carry out two-part growth mixture modeling of u and y jointly with classes for y added as needed to those found for u. Finally, step 5 adds the two covariates to the analysis of u and y.

4.2.1 Step 1: Latent Class Analysis of u

To explore the u data, it is of interest to obtain the estimated probabilities for each u variable, pairs of u variables, and the different u response patterns obtained under the unrestricted multinomial distribution, i.e. without imposing a specific model. The study of these u probabilities, however, is made more challenging by the presence of missing data, where for each time point that y is missing for a person, the corresponding u is missing as well. With missing data on u, the EM algorithm described in Little and Rubin (1987; chapter 9.3, pp. 181-185) can be used to compute the ML MAR estimated frequencies in the unrestricted multinomial model. This procedure is incorporated in the Mplus program. The estimated logit (probability) for $u = 1$ at the seven time points are: -1.19 (0.23), -1.01 (0.27), -0.69 (0.33), -0.26 (0.44), 0.09 (0.52), 0.11 (0.53), and 0.61 (0.65). This suggests an approximately linear trend in the logits over time. This trend, however, is seen for the mixture of possibly several underlying latent classes and it is of interest to be able to study trends for each latent class separately. To this aim, latent class analysis is performed for the seven u variables.

Table 1 gives the latent class analysis results for 2-, 3-, 4-, and 5-class models. The solutions are not contradictory, but represent increasingly more elaborate descriptions. The 2-class solution has a low and a high class with increasing probability of $u = 1$ ($y > 0$). The 3-class solution adds a steeply increasing class, the 4-class solution adds a class that starts high, drops sharply after Spring of 8th grade, and increases again after 10th grade, and the 5-class solution adds a class that starts low and increases sharply after Spring of 8th grade. The 5-class solution is quite good judging from the likelihood-ratio chi-square, and given that BIC increases slightly from 4 to 5 classes a 6th class may not be needed. However, in order to begin the investigation of a zero class ("structural zeroes") as discussed in the introduction, a 6th class is added with parameters fixed to give zero probability of $u = 1$ (i.e. probability one for $y = 0$) at each time point. This is Model 5 in Table 1.

TABLE 1

The zero class of Model 5 is estimated to have 10% or 124 individuals. The classification table given in Table 2 shows that the zero class (class 6) is difficult to distinguish from the low class 5. In contrast, classes 1-4 are reasonably well determined.

TABLE 2

The estimated u logits and probabilities for Model 5 are shown in Figure 6 together with the estimated percentage of individuals in each class. The sharp change between

Spring of 8th grade and the Fall of 10th grade for two of the classes is noteworthy given the transition to high school. This feature will be explored further in the latent class growth analysis presented next.

FIGURE 6

4.2.2 Step 2: Latent Class Growth Analysis of the u Part

The estimated Model 5 shown in Figure 6 above suggests using a linear growth model for the logits of u in combination with a 2-piece linear growth model for the development of classes 2 and 4. The two pieces of the growth model for classes 2 and 4 are for grades $6-8$ and $10-12$, respectively. Given the novelty of the 2-piece latent class growth analysis model, it is shown schematically in Figure 7 using the logit scale. A latent class variable c_t is added to classes 2 and 4 so that individuals in these two classes are either transitioning into a different development after grade 8 ($c_t = 1$), or not ($c_t = 0$). The distinction between the two sub classes $c_t = 1$ and $c_t = 0$ offers interesting developmental information. As drawn, Figure 7 shows that individuals in the $c_t = 1$ sub class starts out with the same alcohol misuse trend as $c_t = 1$ individuals, de-escalates their alcohol misuse upon entry into high school, resume the escalation at the end of high school, although not reaching as high a level at the end of high school as for $c_t = 1$ individuals. Technically, this modeling is a type of confirmatory latent class analysis with two different latent class variables in line with Muthén (2001b).

FIGURE 7

In line with Figure 6, five classes - including a zero class - were chosen for the development of individuals who do not transition ($c_t = 0$). Transitioning was initially allowed for all four non-zero classes, but empty classes for two of these transitions led to a 7-class solution labelled Model 6. As seen in Table 1, Model 6 uses only 16 parameters instead of 34 in the Model 5 latent class growth model. Model 6 has a better BIC value than Model 5 and maintains an acceptable chi-square fit to the multinomial distribution of u (chi-square $=$ 129.373 with 111 df, $p = 0.1122$). In addition, Model 6 provides an interesting explanation of the alcohol misuse development as seen in the estimated probability curves shown in Figure 8. Classes 5 and 6 share the same slow starter development through Spring of grade 8. These two classes make up an estimated 37% of the population, or an estimated 449 individuals. Of the 449 individuals following this early development, 46% are in class 6 which changes trajectory after Spring of grade 8 to an almost zero probability in Fall and Spring of grade 10, followed by an increase by Spring of grade 12. Similarly, of the estimated 203 individuals in classes 2 and 3, sharing the high development through Spring of grade 8, 20% drop down to a zero level in grade 10, followed by an increase by Spring of grade 12. These transition patterns raise interesting questions about why these individuals reduce their alcohol involvement when entering high school and who these individuals are. As mentioned above, transitioning could not be found for individuals in the highest class, class 1, nor for the rapidly escalating class, class 4. Finally, it can be noted that the estimated percentage for the zero class is now 14%.

FIGURE 8

4.2.3 Step 3: Two-Part Growth Mixture Modeling of u and y jointly

The analysis of the u probabilities provides the basis for the full, two-part growth mixture modeling. In line with the methodology described in Section 2, the development in the y part is a growth mixture model with random effect means varying across the classes found for the u part.

The two-part growth mixture model for joint analysis of u and y was based on the 7-class Model 6 shown in Figure 8. The 2-piece development used for the u part is also used for the y part, except that the second piece of the 2-piece development is simplified. Given that the u probability is estimated as zero for the first two time points of the second piece (Fall and Spring of grade 10), the second piece simply consists of a mean parameter for Spring grade 12. The fit statistics for the resulting Model 7 are shown in Table 1. The fit of the u part is somewhat worse, but has not seriously deteriorated. The estimates of the class percentages for Model 6 versus Model 7 (in parenthesis) are: 9 (5), 13 (12), 3 (3), 29 (22), 20 (29) 17 (15), and 10 (14).

4.2.4 Step 4: Further Exploration of the y Part

Exploration of additional classes for the y part included an analysis of the y variables alone, using the rescoring of $y = 0$ to missing data as is used in the two-part model of Step 3. This exploration pointed to four classes. In this case, a zero class is not included given that zero values are missing. A new class type emerged with a high starting point and a negative slope, although this class is relatively small (7%). Adding one class in the joint analysis of u and y, however, the new class was found to have a high starting point but did not show decline. This is Model 8 in Table 1.

As an alternative two-part analysis, two separate and correlated latent class variables can be used, one for the u part and one for the y part in line with Muthén (2001b). In the current application, however, this leads to a very complex model where many of the joint classes are presumably small or empty and this approach will not be pursued here.

As discussed in Section 3.1.3, pseudo-class plots for all classes are useful to check to which extent the model describes the scatter of y observations. Note that to avoid a misleading picture, these plots should use the original observed data on y instead of the analysis data where $y = 0$ is changed to missing. The pseudo-class plots for Model 8 indicate that the observed data show a high degree of variability around the estimated mean curves. This results in lower classification quality as evidenced by the moderate entropy value of 0.578 and makes early prediction of problematic development difficult.

The estimated y mean curves for each of the classes of Model 8 are shown in Figure 9. Figure 9 shows that, except for the additional class, the probability curves for u have not changed much from those of Figure 8 where only the u part was analyzed. It is clear that the additional class has a similar probability curve to the already existing high u class, but is defined by a different y trajectory.

FIGURE 9

Figure 9 shows that the two classes with the highest y means throughout the grades are also the ones with the highest u probabilities. An estimated 10% are found in these two classes and can be regarded as showing early onset of alcohol misuse. The zero class (structural zeroes) still contains an estimated 14% in Model 8. Inspection of estimated posterior probabilities from this model shows that for individuals with their highest posterior probability for this class, the y values are either zero or missing as expected.

Figure 9 also shows the estimated probability of $y = 0$ (i.e. $u = 0$) at each time point, i.e. the probability of observing either a random or a structural zero. This ranges from a high of 0.63 in Fall of grade 6 to a low of 0.20 in Spring of grade 12. The probability estimates are consistently lower than those from the unrestricted multinomial model estimated under MAR reported in section 4.2.1. The section 4.2.1 values, however, were obtained by considering the u part alone. As compared to the observed proportions, the Model 8 probability estimates are lower for the first time points and higher for the remaining time points. The observed proportions, however, ignore that missing values may be distributed unequally across zero and non-zero true values. At the first time point there is no missing data and the unrestricted multinomial probability estimate and observed proportion are the same, 0.77. The lower Model 8 probabilty estimate of 0.63 suggests that Model 8 may need to allow for a deviation from the linear logit shape at this first time point.

It is interesting to consider the improvement in classification when adding the y information. This may be studied via the average posterior probabilities for individuals classified via a highest posterior probability assignment. With the addition of the y information, the class that transitions from a low trajectory to zero misuse in grade 10, class 6 (14%), obtains an improved average posterior probability value of 0.627

compared to only 0.471 when using the latent class growth model for u in step 2. The highest individual posterior probability value for this class is 0.802, obtained for the y outcome vector (0, 0, 0, 0.33, 0, 0, 0). The low progression through the value of 0.33 in Spring of grade 8 indicates that this class includes individuals who do not show a severe form of alcohol misuse.

4.2.5 Step 5: Two-Part Growth Mixture Modeling with Covariates

Taking Model 8 as a starting point, the two covariates are added to study their influence on both parts of the model. In Model 9, the covariates are only allowed to influence the class membership, while in Model 10 they also influence the intercept and slope for both the u and the y part, for simplicity excluding the second piece of the 2-piece development. The covariates are specified to have no influence on the intercept or slope of the zero class.

The fit statistics for Model 9 are given in Table 1, where it may be noted that the likelihood and BIC values can be compared with those for the Olsen-Schafer two-part model with covariates, given as Model 11. Model 9 is nested within Model 10 so that a likelihood-ratio chi-square test is possible. Table 1 shows that the 7 additional parameters of Model 10 relative to Model 9 are worthwhile given that the chi-square value 35.656 with 7 degrees of freedom is significant at the 1% level.

As indicated by the entropy values, Model 10 gives slightly better classification than Model 8 without covariates. The results are on the whole similar for the two models and the Model 10 curves will not be shown. Both models estimate the zero class membership at 14%.

Model 10 finds that being male increases the odds of being in the high class for the $u = 1$ probability development relative to being in the zero class. Being black increases the odds of being in the class that transitions from high to low probability relative to the zero class. Being black decreases the odds of being in the class that shows a sharp increase in the probability relative to the zero class. The intercept of the probability development is significantly lower for Black, using a class-invariant specification of this influence. The slope of the y development is significantly lower for Black, using a class-invariant specification of this influence. These findings are in line with the alcohol literature, indicating a higher probability of problematic alcohol involvement for males and a later onset for blacks.

A final check of Model 10 is made possible due to the fact that the analyses used the random half of the sample for which data collection started already in Fall of grade 6. An ordinary cross validation is not possible because the other half of the sample does not have the outcome for Fall of grade 6, but a check on the stability of the findings is obtained by simply analyzing the full sample, yielding a sample size of $2,580$. The findings are very similar in the full and partial samples, including the class percentages. The zero class is now estimated at 13%. The significant covariate findings of Model 10 are still significant using the full sample. In addition, two new relationships are found significant due to the larger sample: the odds of being in the class that has the highest y development (and second highest probability development) is significantly decreased for blacks; and the odds of being in the low, transitioning class is significantly increased for males. The latter finding sheds new light on the development of alcohol involvement for adolescents. Here, a group of individuals, estimated as 38% of the population (estimated as 37% in the earlier Model 6 using u only), follow the same relatively low, increasing trajectory for both u and y through Spring of grade 8, but an estimated 35% of this group (estimated as 46% in the earlier Model 6) transition to almost zero involvement in grade 10, with a subsequent increase by Spring of grade 12. The male to female odds is estimated at 2.90. The gender relationship for this low, transitioning class may warrant further research.

4.3 Final Comparison of Results Across Models

Regular growth mixture modeling, ignoring the preponderance of zeroes, gives results that are qualitatively similar to those of the final 8-class two-part growth mixture model, Model 10. However, the chosen 5-class solution in section 4.1.2 is not as informative and more than 5 classes could not be fitted. Model 10 benefitted from the exploratory modeling of the u part, which led to the suggestion for a 2-piece model.

Comparing Model 10 with Model 11, the Olsen-Schafer two-part model, it is seen that the results agree for the slope of the y development, but Model 11 did not find a significant effect of the covariates on the u probability intercept as did Model 10. Also, Model 11 found negative influence on the slope of the u probabilities from both covariates, but this was not found in Model 10. The differences may be due to the greater detail of Model 10 due to the use of several latent classes, allowing the covariates to influence the class membership probabilities, and allowing for a 2-piece growth model in two of the classes.

As an interesting aside, in the fitting of the models, a relatively large residual was found for the bivariate distribution of u at Fall and Spring of grade 10. To better capture the relationship of those two u variables in the Olsen-Schafer models, a model was fitted with an extra random effect for those two occasions, leading to a considerable improvement in model fit. This model modification did not, however, affect the influence of the covariates. The 2-piece modeling approach of Model 10, arrived at from the exploratory latent class analyses, also serves to fit this part of the model better given that grade 10 is allowed to deviate from the regular development for portions of the sample. The 2-piece modeling, however, accomplishes this in a way that provides a more useful interpretation in terms of transitions.

5 Conclusions

The ability to carry out growth mixture modeling for data where a large percentage is at the lowest value is important for many types of applications, such as when analyzing early development. The proposed two-part growth mixture model therefore meets a great need for better modeling. The alcohol misuse example shows the strength of the proposed approach. Both the development of the probability of misuse and the amount of misuse can be modeled in a very flexible way. The class-specific 2-piece growth model illustrates the detail that can be read out of the data. An interesting aspect of the modeling is the possibility to estimate the proportion of individuals who are in the "zero class", i.e. having structural zeroes. These modeling features are useful for substantive researchers interested in onset and transitions of behavior.

The proposed two-part growth model is quite general. This has the disadvantage of requiring an extensive, stepwise modeling process. It has the strong advantage, however, of each step adding to the knowledge about the development. The two-part growth mixture model also benefits from being incorporated in a general latent variable modeling framework, so that many modeling variations and extensions are possible. For example, it is possible to analyze two processes such as alcohol and tobacco use simultaneously, exploring their relationships. Other extensions are also available, such as analyzing sequential processes (Muthén, Khoo, Francis, & Kim Boscardin) and distal outcomes predicted by the latent classes (Muthén & Shedden, 1999). Furthermore, the general framework allows for outcomes that are ordered categorical instead of continuous, leading to a two-part latent class growth analysis.

References

Aitkin, M. A general maximum likelihood analysis of variance components in generalized linear models. Biometrics, 1999, V55, 117-128.

Amemiya, T. (1985). Advanced Econometrics. Cambridge, Mass.: Harvard University Press.

Carlin, J.B., Wolfe R., Brown, C.H., Gelman, A. (2001). A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. Forthcoming in Biostatistics.

Clogg, C.C. (1995). Latent class models. In G. Arminger, C.C. Clogg & M.E. Sobel (eds.), Handbook of statistical modeling for the social and behavioral sciences (pp. 311-359). New York: Plenum Press.

Deb, P. (2001). A Discrete Random Effects Probit Model with Application to the Demand for Preventive Care. Health Economics, 10, 371-383.

Deb, P. & Holmes. A.M. (2000). Estimates of use and costs of behavioral health care: a comparison of standard and finite mixture models. Health Economics, 9, 475-489.

Deb, P., Hall, C., Trivedi, P.K., & Hall, W. (2000). The structure of demand for health care: Latent class versus two-part models. Under review.

Duan, N., Manning, W.G., Morris, C.N. & Newhouse, J.P.(1983). A comparison of alternative models for the demand for medical care. Journal of Business and Economic Statistics, 1, 115-126.

Jones, B.L., Nagin, D.S. & Roeder, K. (1998). A SAS procedure based on mixture models for estimating developmental trajectories.

Little, R.J., & Rubin, D.B. (1987). Statistical analysis with missing data. New York: John Wiley & Sons.

Maggs, J.L & Schulenberg, J. (1998). Reasons to drink and not to drink: altering trajectories of drinking through an alcohol misuse prevention program. Applied Developmental Science, 2, 48-60.

Muthén, B. (2001a). 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.M. & Sayer, A. (Eds.), New Methods for the Analysis of Change (pp. 291-322). Washington, D.C.: APA.

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

Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S., Yang, C., Wang, C.P., Kellam, S. J.B. Carlin, & J. Liao (2000). General growth mixture modeling in randomized preventive interventions. Accepted for publication in Biostatistics.

Muthén, B., Khoo, S.T., Francis, D. & Kim Boscardin, C. (2000). Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. Multilevel Modeling: Methodological Advances, Issues, and Applications (in press). S.R. Reise & N. Duan (Eds). Mahaw, NJ: Lawrence Erlbaum Associates.

Muthén, L.K. & Muthén, B. (1998-2001). Mplus User's Guide. Los Angeles, CA: Muthén & Muthén.

Muthén, B. & Muthén, L. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. Alcoholism: Clinical and Experimental Research, 24, 882-891.

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

Nagin, D.S. (1999). Analyzing developmental trajectories: a semi-parametric, group-based approach. Psychological Methods, 4, 139-157.

Olsen, M. K, & Schafer, J., L. (2001). A two-part random effects model for semicontinuous longitudinal data. Journal of the American Statistical Association, 96, 730-745.

Tobin, J (1958). Estimation of relationships for limited dependent variables. Econometrica, 26, 24-36.

Table 1

Analysis Results

Model	# classes	log L	# par's	BIC	Entropy	Chi-Square	D.f.	p
Latent Class Analysis								
1	2	-3,205.580	15	6,517.931	0.626	398.390	112	0.0000
2	3	-3,089.375	23	6,342.465	0.612	165.988	104	0.0001
3	4	-3,065.454	29[1]	6,337.331	0.648	118.150	98	0.0081
4	5	-3,057.035	33[2]	6,348.965	0.622	101.311	94	0.2849
5	6	-3,054.913	34[2]	6,351.838	0.574	97.066	93	0.3659
Latent Class Growth Analysis								
6	7	-3,071.066	16	6,256.021	0.489	129.373	111	0.1122
Two-Part Growth Mixture Modeling								
7	7	-5,133.579	39	10,544.761	0.573	149.551	111	0.0086
8	8	-5,104.211	44	10,521.614	0.578	146.244	108	0.0084
Two-Part Growth Mixture Modeling with Covariates								
9	8	-5,008.650	58	10,428.911	0.610	142.838	108	0.0139
10	8	-4,990.822	65[3]	10,442.931	0.610	_____[4]	_____[4]	_____[4]
Olsen-Schafer Two-Part Model with Covariates								
11	1	-5,108.766	29	10,423.337	_____	_____[4]	_____[4]	_____[4]

[1] Two item logits fixed at ±15.
[2] Four item logits fixed at ±15.
[3] Parameter fixed at –10 for class 5 regressed on Black.
[4] Not available with covariates influencing the probabilities.

Table 2

Average Class Probabilities by Class for Model 5

	1	**2**	**3**	**4**	**5**	**6**
Class 1	0.701	0.088	0.136	0.023	0.052	0.000
Class 2	0.013	0.861	0.044	0.000	0.082	0.000
Class 3	0.105	0.005	0.756	0.049	0.086	0.000
Class 4	0.020	0.000	0.188	0.765	0.026	0.000
Class 5	0.013	0.006	0.119	0.129	0.566	0.168
Class 6	0.000	0.000	0.000	0.012	0.460	0.527

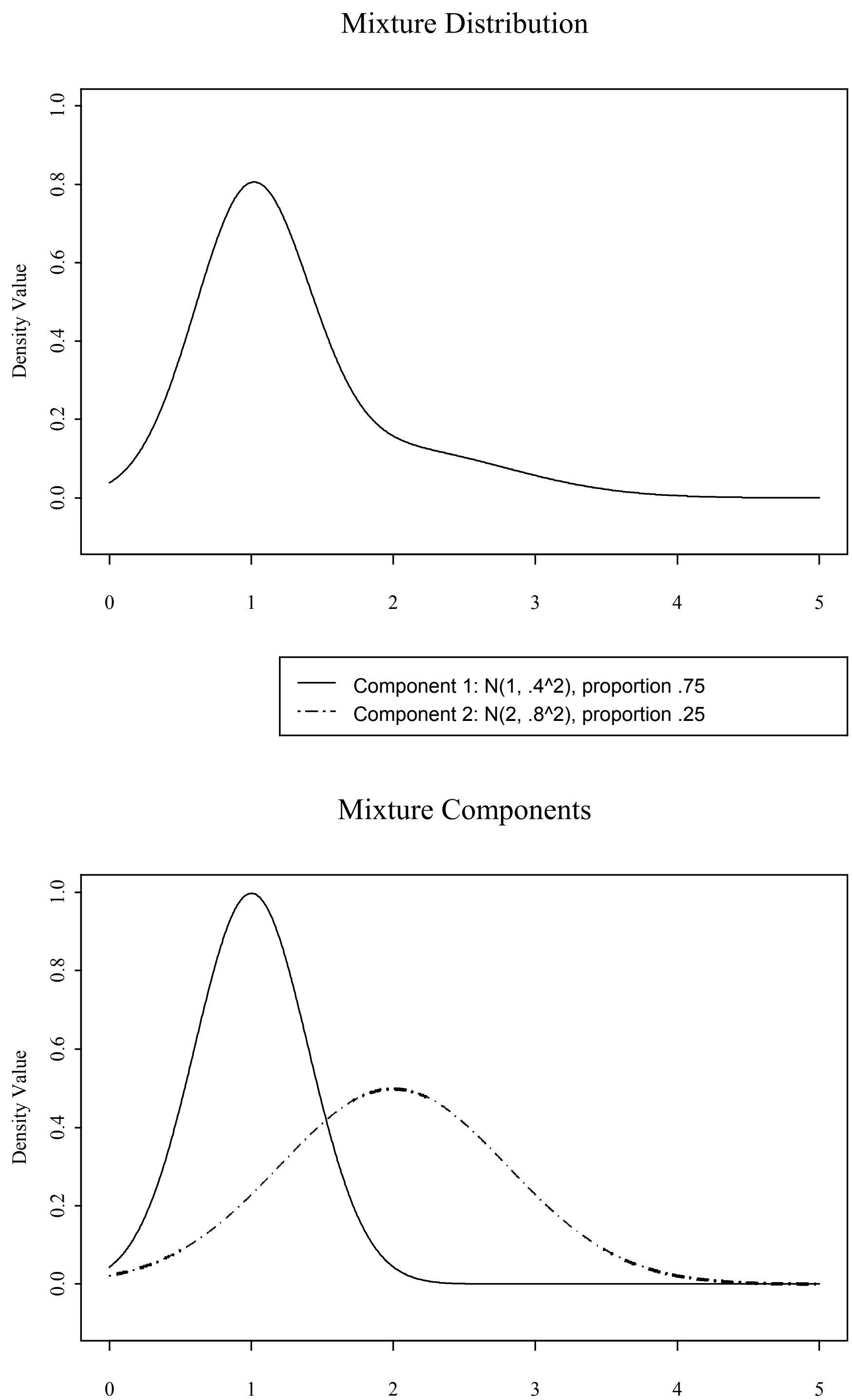

Figure 1. Mixture distributions

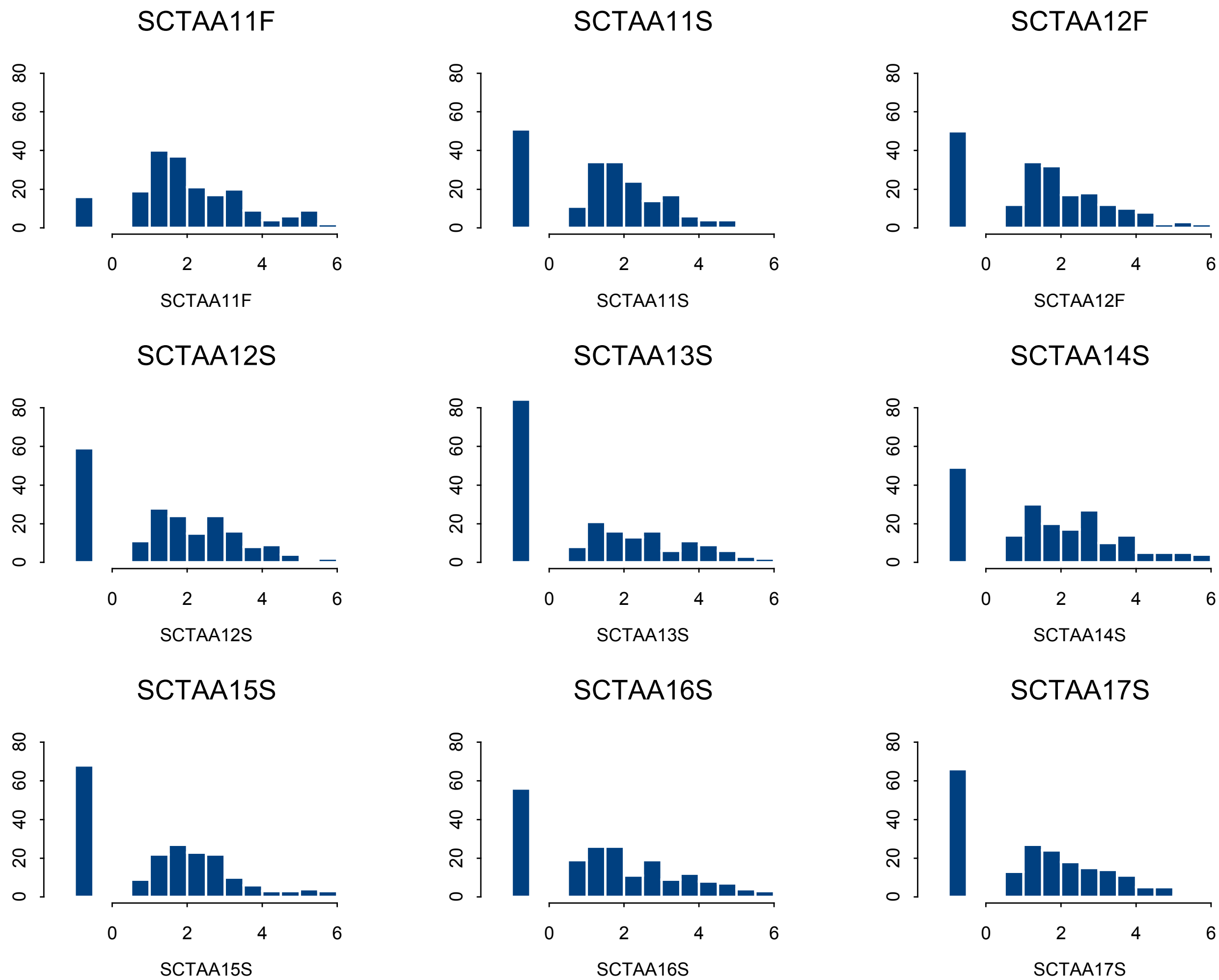

Figure 2. Aggressive behavior distributions at nine time points

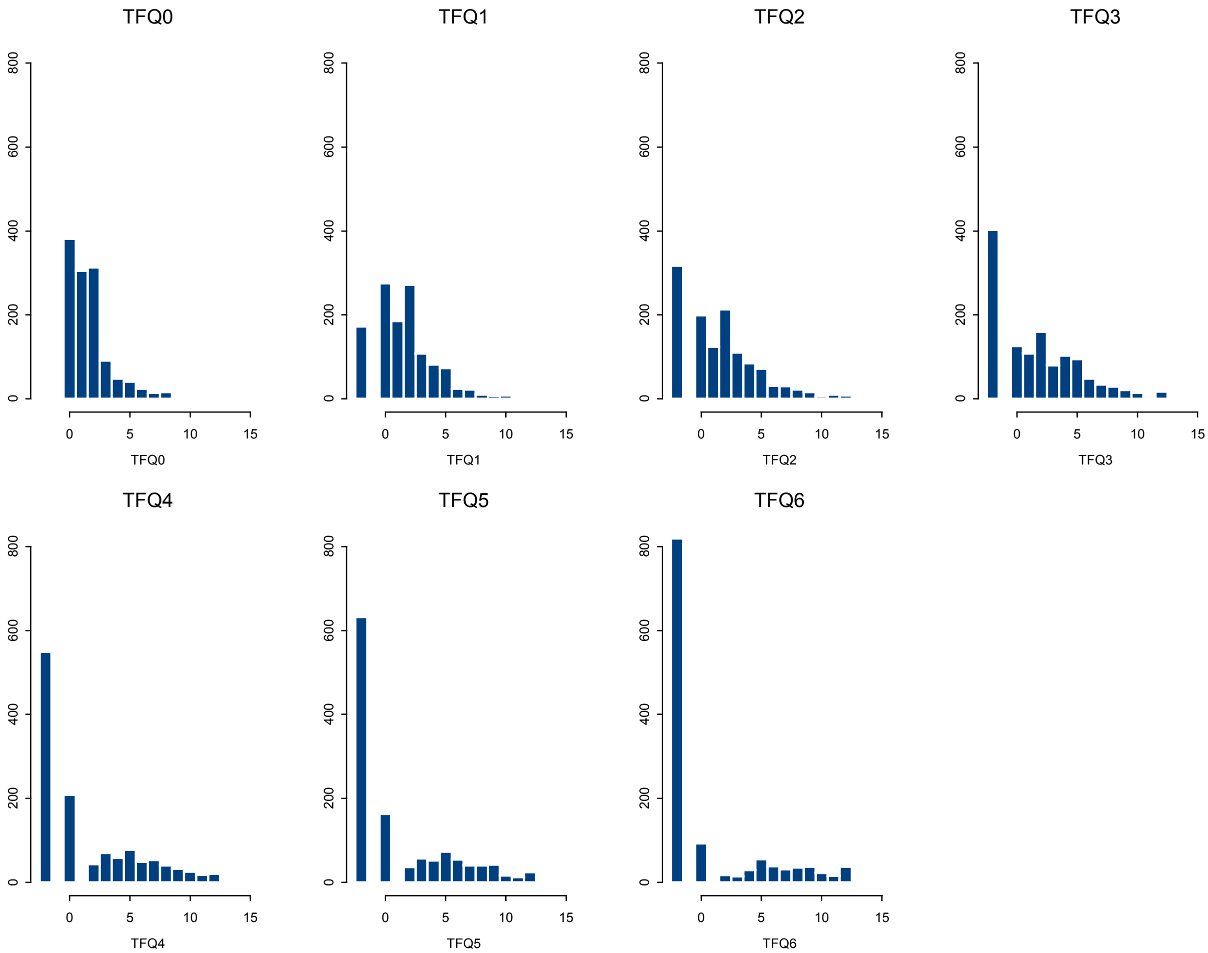

Figure 3. TFQ distributions at seven time points

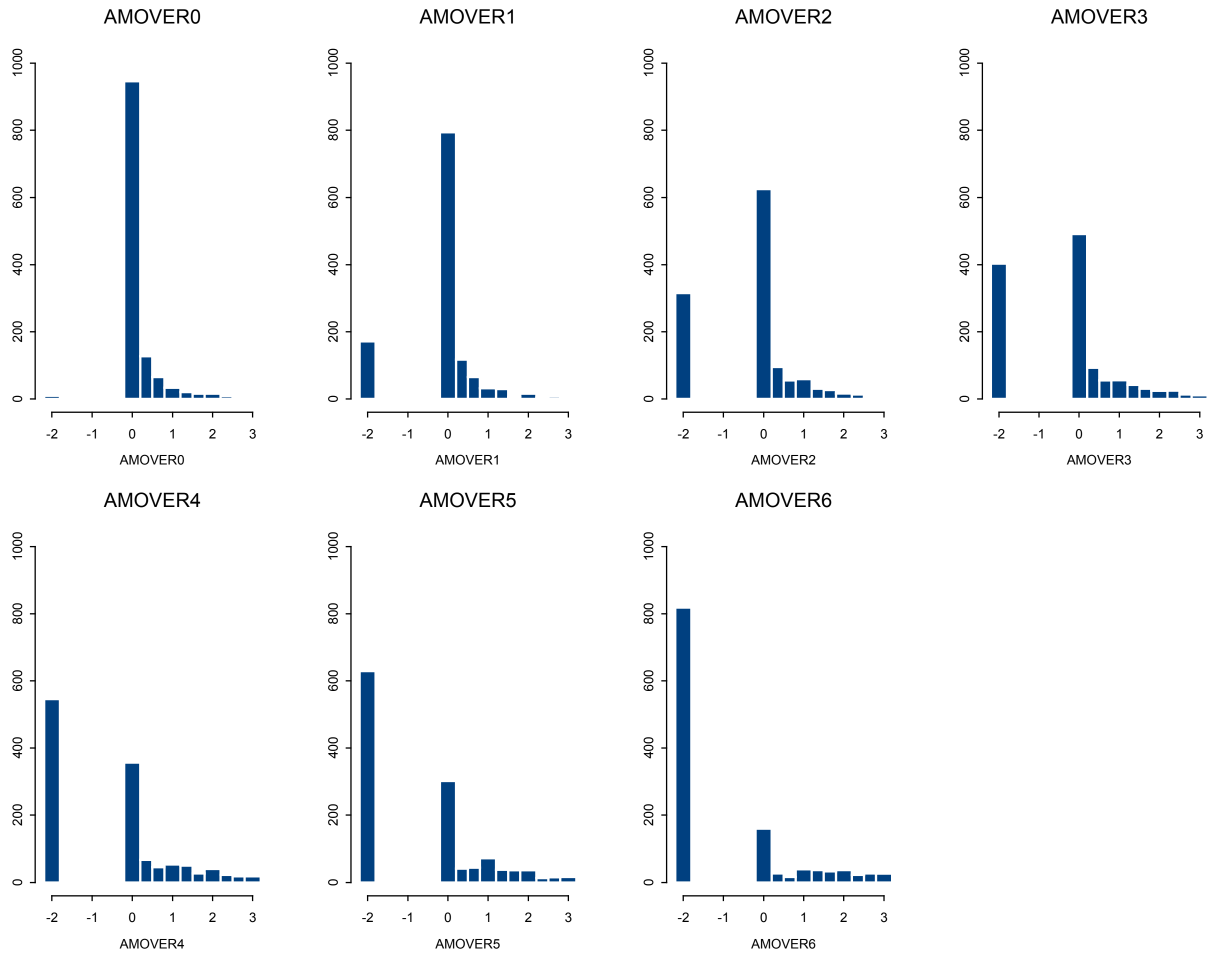

Figure 4. AMOVER distributions at seven time points

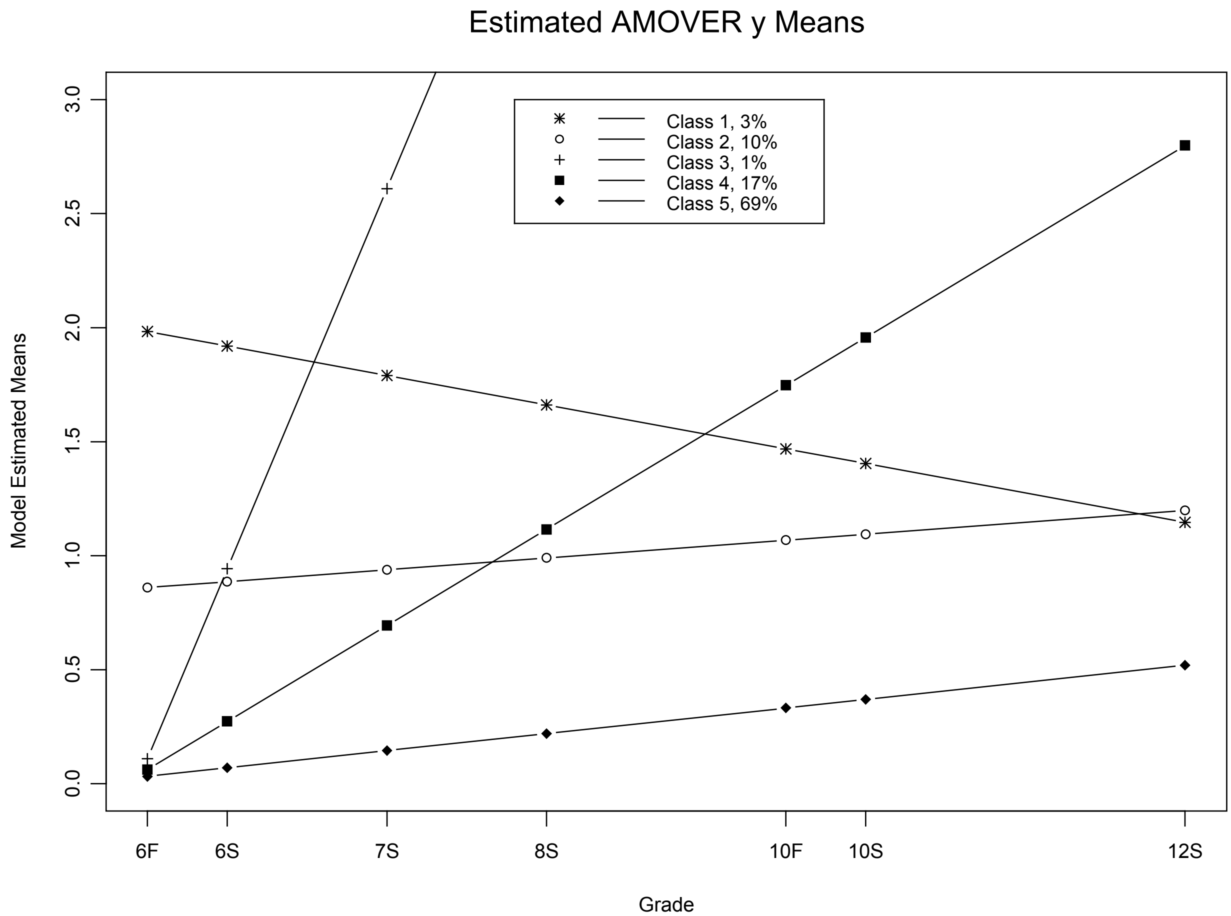

Figure 5. Estimated means from growth mixture modeling

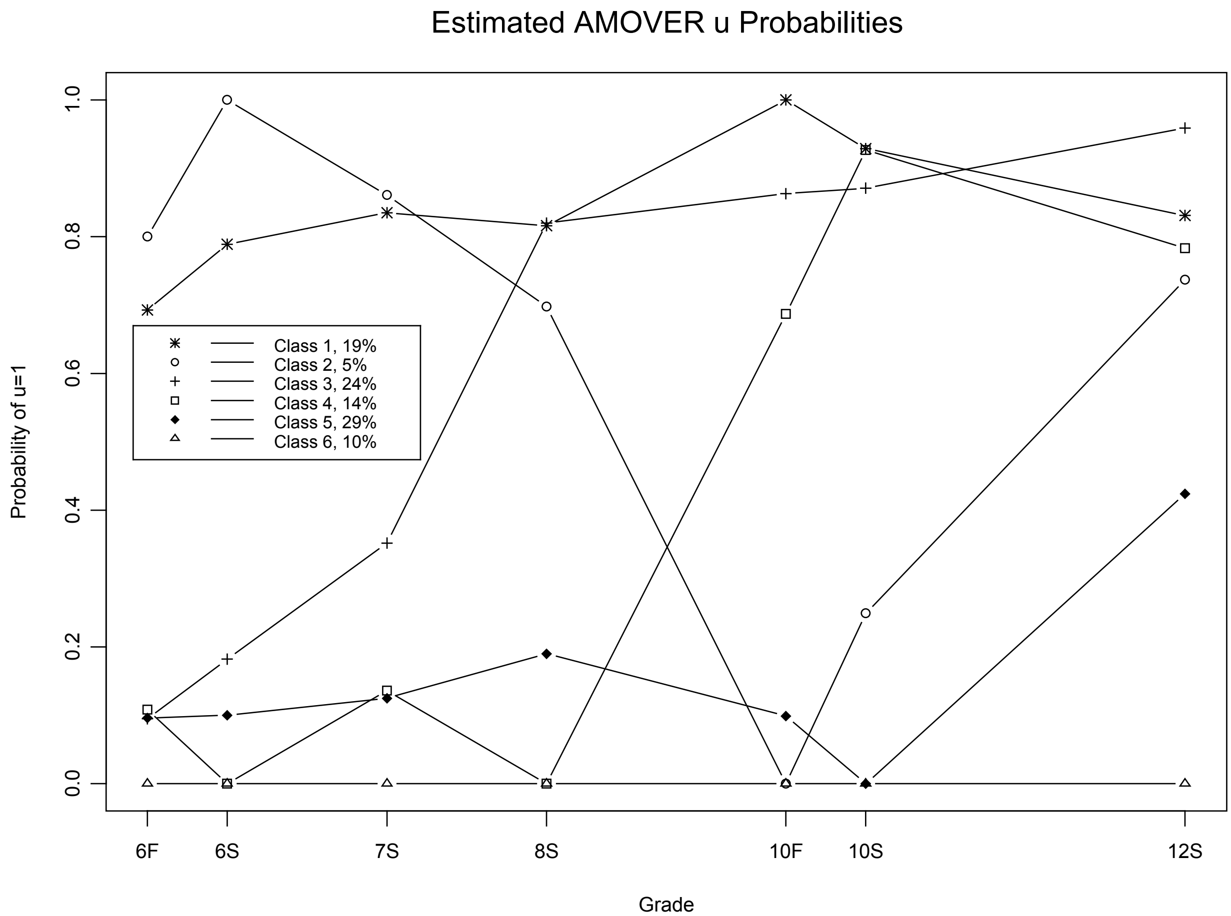

Figure 6. Estimated probabilities from latent class analysis (Model 5)

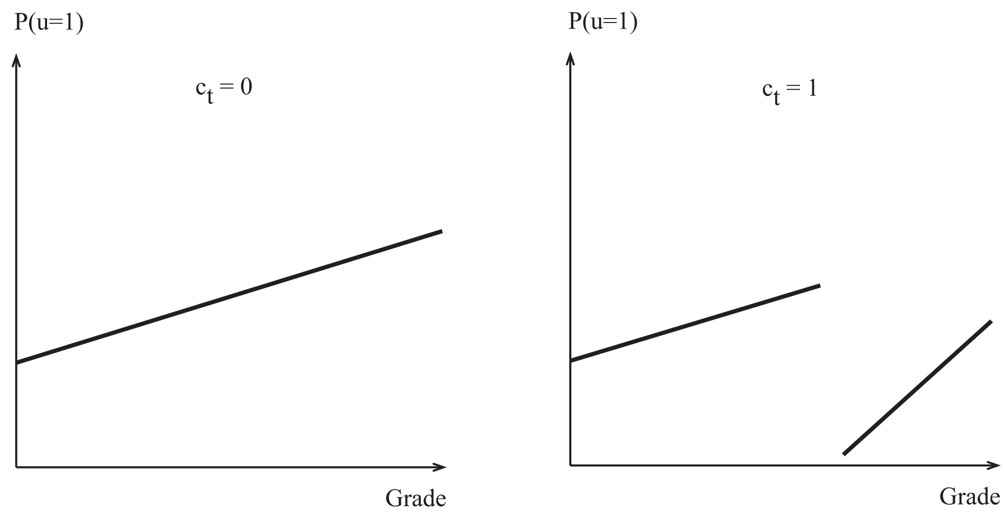

Figure 7. Hypothetical two-piece growth model for u probabilities in logit scale

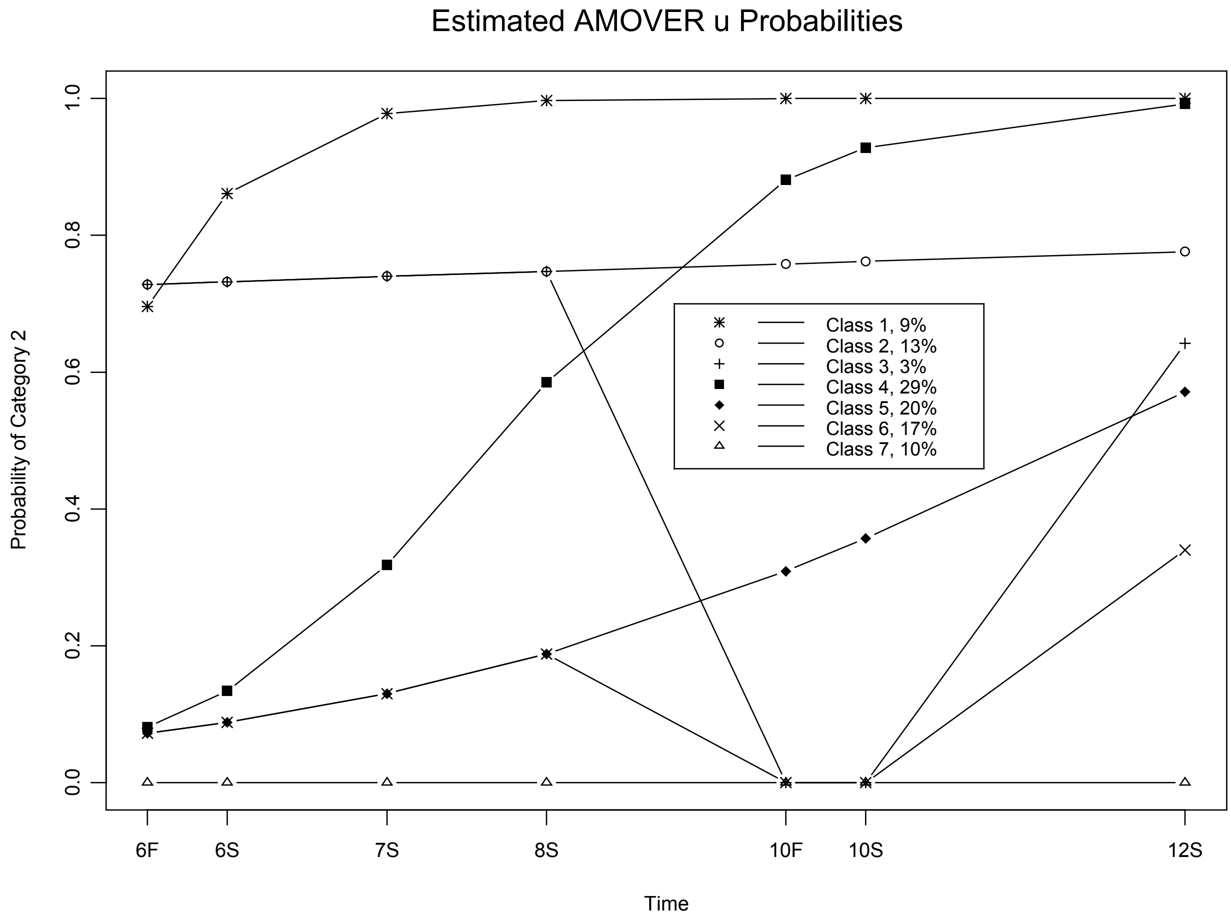

Figure 8. Estimated probabilities from latent class growth analysis (Model 6)

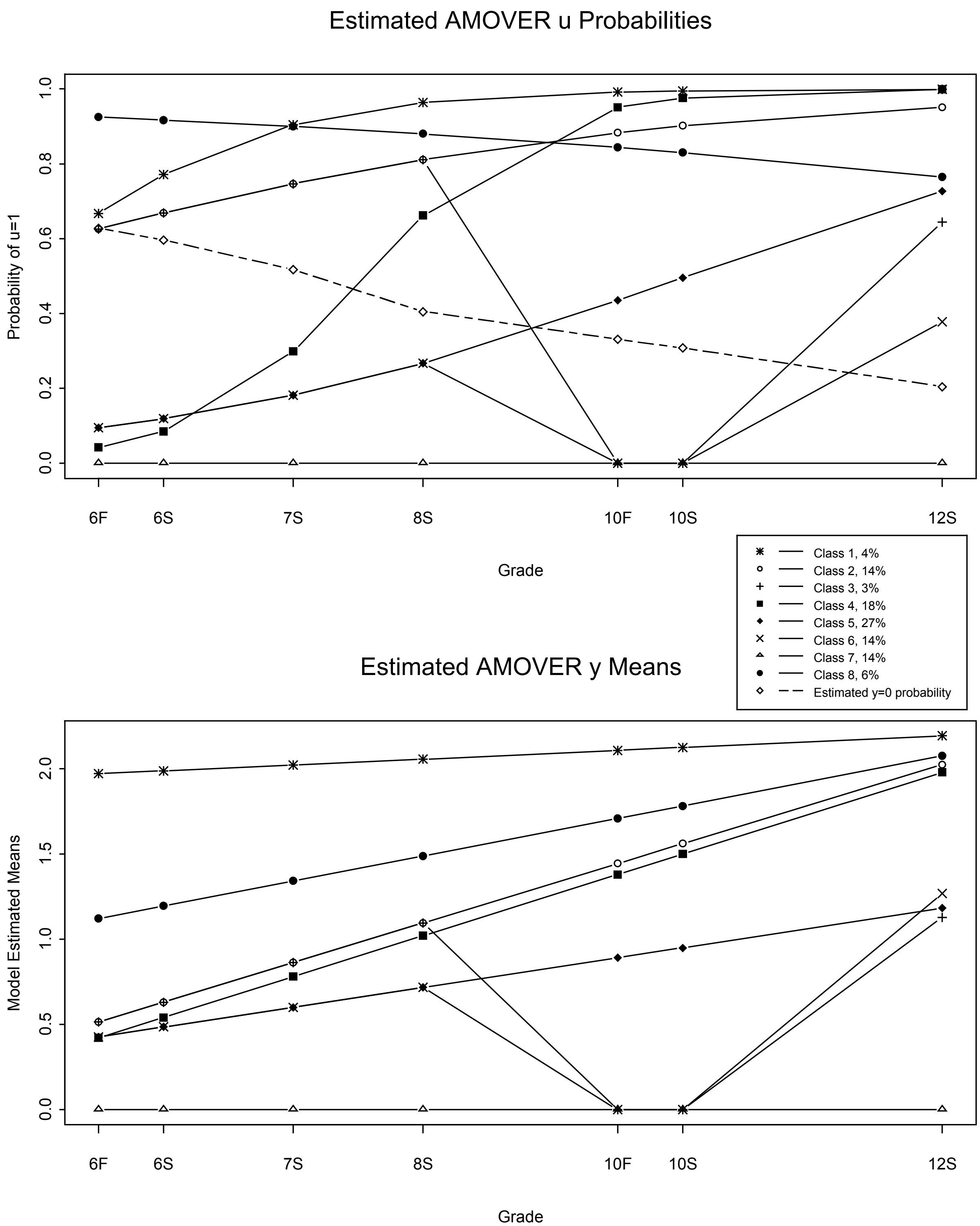

Figure 9. Estimated probabilities and means from two-part growth mixture modeling (Model 8)