Running Head: LONGITUDINAL MODELING OF POPULATION HETEROGENEITY

Longitudinal modeling of population heterogeneity:

Methodological challenges to the analysis of empirically derived criminal trajectory profiles[1]

Frauke Kreuter[2]

Joint Program in Survey Methodology, University of Maryland, College Park, MD

and

Bengt Muthén

University of California, Los Angeles, CA

Longitudinal modeling of population heterogeneity: Methodological challenges to the analysis of empirically derived criminal trajectory profiles

The question of how to model population heterogeneity in longitudinal data has been repeatedly discussed in criminological journals, especially in the context of modeling "criminal careers" as the variable of interest (e.g., Elliot, 1985; Farrington & West, 1993; Nieuwbeerta & Blokland, 2003; Tracy, Wolfgang, & Figlio, 1990). It has been recognized that there is variation in criminal trajectories and any analysis of the age-crime relationship requires a decision on how to model this variation. In Chapter 1 in this volume, Muthén briefly touched on different modeling approaches to capture observed heterogeneity around trajectory profiles. Here we will revisit and extend the examples mentioned. To do so we will use two well-known criminological data sets, the Philadelphia Cohort Study (Tracy et al., 1990) and the Cambridge data (Farrington & West, 1990). Both datasets have been used in the past to describe criminal activity as a function of age.

It is important to note that we are not seeking to provide the final analysis for either one of these datasets. Too many substantive decisions would have to be made that only criminologists could provide. Nonetheless, we will take this opportunity to discuss how the different modeling alternatives are related to each other, what the different modeling results for these applications are, how results from one model can be translated to another, and what further modeling opportunities the different models offer for substantive researchers.

In our analysis we focus on growth mixture models (Muthén, 2001a, 2002, 2004; Muthén & Muthén 1998-2006; Muthén & Shedden, 1999), non-parametric growth models (Aitkin, 1999; Hedeker, 2000; Heinen, 1996), and latent class growth models, also known as group-based trajectory models (e.g., Nagin & Land, 1993; Roeder, Lynch, & Nagin, 1999). The in-depth comparison of these models is important because it can help to answer a key question that substantive researchers have often asked in the past, that is, how to find a balance between mixture classes that are needed according to substantive theory and those needed to model variation around observed growth trajectories (e.g., Bushway, Brame, & Paternoster, 1999; Heckman & Singer, 1984; Nagin & Tremblay, 2005; Sampson & Laub, 2005).

Population-based conviction data

As mentioned above, we decided to use two data sets for our demonstration. The first includes data on convictions of 403 "boys" from Cambridge, England who were followed from ages 10 to 40 (Farrington & West, 1990)[3]. The second dataset includes information on 13,160 males born in Philadelphia in 1958. Annual counts of police contacts are available from age 4 to 26 for this birth cohort. Strengths and limitations of these data have been discussed in various places (Brame, Mulvey, & Piquero, 2001; Moffitt, 1993; Nagin & Land, 1993) and will not be repeated here.

Both datasets share two characteristics. First, the aggregate age-crime curve follows the well-known pattern of increasing yearly convictions throughout the subjects' teenage years and decreasing annual conviction rates thereafter (Farrington, 1986; Hirschi & Gottfredson, 1983). Second, the outcome variable is extremely skewed with a large number of zeros at each point in time. Both characteristics can be seen in Figure 1. The upper panel of Figure 1 displays the average number of police contacts for the Philadelphia data by age-groups. We see the average number of police contacts rise up to the age 16, where it peaks at an average of 0.23 police contacts for the entire cohort of 13,160 males.

Insert Figure 1 about here

The lower panel of Figure 1 displays the number of males with police contacts by age. We see that the change noted in the upper panel of Figure 1 is not just due to an increase in the number of contacts for a particular subset of males, but is also largely due to a changing number of males that experienced police contacts. At age 11, for example, 205 males, which is less than two percent of the entire cohort, had a police contact. At age 16, on the other hand, this number increased to 1800 males (13.7% of the cohort). But there are also a large number of males who never had any contact with the police. In the Philadelphia birth cohort roughly 60% had no recorded police contact in the recorded years between age six and age 26, about 15% had one police contact during this period, and the remaining quarter of the birth cohort shows between 2 and 57 police contacts. In any given year more than 90% of males in the 1958 cohort do not show any policy contact, except for the teenage years (15 through 17) where the number of non-offenders ranged between 86.3 % and 89%.

Insert Figure 2 about here

The distribution of convictions over time looks very similar for the Cambridge study (Figure 2), only a little less smooth due to the smaller number of cases. For the Cambridge males we see also an increase in early adulthood resulting in a peak at age 17 with an average of 0.17 convictions and a decline thereafter. The number of males engaged in crimes increased in a similar way. At age 11 roughly 2% (8 boys) were convicted, while at age 17 there were somewhat more than 10%. Here too most males show no convictions throughout the observational period. Overall, 60% of males in the Cambridge study were never once convicted, and in any given year anywhere between 88.8% and 98.5% of males had zero convictions.

For our modeling example several decisions were made to deal with the large amount of zeros in the outcome variable at each individual time point. Following Roeder et al. (1999) we decided to analyze the data grouped into bi-annual intervals[4], and use zero-inflated Poisson models (Lambert, 1992; Nagin & Land, 1993) in all our analyses of the count outcome variable. These zero-inflated Poisson (ZIP) models were developed for situations in which the count outcome is equal to zero more often than one would expect assuming a Poisson distribution (Hall, 2000). The ZIP model is discussed in detail by Nagin and Land (1993) and Roeder et al. (1999), and will therefore not be explained any further. A brief model description can be found in the Appendix to this chapter.

Comparison of the mixture modeling approaches

The key to the model comparison is a good understanding of the assumptions behind the different growth models, specifically the presence or absence of random effects, as well as a good understanding of non-parametric modeling of the distributions of those random effects. Before we discuss the results we will therefore review the specification of a different growth model using a general latent variable framework.[5] The use of this framework allows a straightforward comparison of the growth models, both in terms of equations and in terms of actual model specification. In our comparison of growth mixture models, we focus on their ability to capture the heterogeneity of trajectories. However, it should be noted that all assumptions underlying regression-based models[6] still apply and need to be considered when applying these even more demanding techniques.

Underlying concepts

In growth curve models, the development of the observed outcome variables over time is characterized as a function of age. Conventional growth modeling can be used to estimate the amount of variation across individuals in the growth parameters (intercepts and slopes) as well as average growth. In other words, in a conventional growth model the individual variation around the estimated average trajectory is expressed in growth parameters that are allowed to vary across individuals (Raudenbush & Bryk, 2002), and that variation of the growth parameters is assumed to take on a normal distribution (Hedeker & Gibbons, 1994). Substantively, it means that one assumes all people in the sample have the same expected criminal trajectory and that the individual variation around this expected trajectory is centered on the estimated intercept and slopes for the whole sample, with symmetric deviation on both sides (e.g., some individuals start their criminal careers earlier and some start later, but on average they start at the estimated intercept).

The normality assumption of the conventional growth model was challenged by Nagin and Land (1993). These authors adopted a model by Heckman and Singer (1984) that approximates an unspecified continuous distribution of unobserved heterogeneity with a linear combination of discrete distributions (Nagin, Farrington, & Moffitt, 1995). That is, different groups are used to capture the overall variation. Within the latent variable framework one would say the population consists of latent classes each of which has its own growth trajectory; we therefore refer to this approach as latent class growth analysis (LCGA). Over the last few years, LCGA has become attractive for quantitative criminologists and specialists in related fields (e.g., Blokland &

Nieuwbeerta, 2005; Fergusson & Horwood, 2002; Haviland & Nagin, 2005; Piquero, Farrington, & Blumstein, 2003). The group-based modeling approach supports distinctions among subgroups having distinct offending trajectories. It also matches theories like the one proposed by Moffitt (1993) that differentiate between a large group of adolescence-limited offenders and a small subgroup of life-course persistent offenders. However, subgroups found in these latent-class-type models have not always matched theory, and there has been a tension regarding how to interpret the classes found (for a review see Raudenbush, 2005). Recently Nagin and Tremblay (2005) warned about the reification of the trajectory groups when their initial purpose is solely a statistical approximation of a complex continuous distribution. As Raudenbush (2005) pointed out,

> Perhaps we are better off assuming continuously varied growth *a priori* and therefore never tempting our audience to believe in the key misconception that groups of persons actually exist. We would then not have to warn them strongly against "reification" of the model they have been painstakingly convinced to adopt (p. 136).

Looking at the models introduced by Muthén in Chapter 1 of this volume, the decision regarding how to model trajectories might not necessarily be an either/or. A growth mixture model (GMM) would allow for both − a mixed population as predicted by certain criminological theories of different subpopulation trajectories, and a variation in the growth parameters (and likewise the trajectories) within the groups. However, if random effects are allowed within the classes, the growth mixture model also relies on the normality assumption. A nonparametric version of a growth mixture (NP-GMM) can be employed instead, which does not rely on any distributional assumption for the random effects. Instead, the model is specified such that latent classes are used to capture the potentially non-normal distribution within the growth mixture classes. In this sense, the nonparametric growth mixture model resembles the latent class growth model. At the same time it allows for an explicit specification of 'substantively meaningful' trajectory groups and groups that are needed solely to capture the variation in the growth factor(s).

Model specifications

The similarities and differences between these models will become clearer with a closer look at the model specifications. Let us start with a simple model with no covariates other than age, where the age-crime relationship is described by a quadratic growth function. The outcome variable is a count (e.g., number of convictions) and the Poisson parameter λ can be expressed for each individual *i* at time point *t* as a linear combination of the time-related variable *X* with a linear slope parameter β_1 and a quadratic slope parameter β_2, $\ln(\lambda_{it}) = \beta_{0i} + \beta_{1i}x_t + \beta_{2i}x_t^2$. In this specification of a conventional growth model β_{0i} is a random intercept, and β_{1i} and β_{2i} are random slopes. That means that the values for the intercept and slopes are allowed to vary across individuals.

Let us assume for simplicity a conventional growth model with a random intercept factor and no random effects for the slope or quadratic term. The equation above would change to $\ln(\lambda_{it}) = \beta_{0i} + \beta_1 x_t + \beta_2 x_t^2$ with no *i* subscripts on β_1 and β_2. Substantively speaking, every observational unit shows the same development but starting from different values. One can think of this as a shift of the growth trajectories along the vertical axis in pictures like Figure 1 and 2. Using multilevel notation (e.g., Raudenbush & Bryk, 2002), the equation estimated at level two (the individual level) for the growth model with just a random intercept is $\beta_{0i} = \alpha_0 + r_{0i}$. In a latent variable framework, this random effect β_{0i} can be seen as a latent variable with a normal distribution. The latent variable captures the heterogeneity in the intercepts, and again, the normality assumption implies that the individual variation around the expected trajectory is centered on the estimated intercept for the whole sample, with symmetric deviation on both sides. But what if the assumption is violated and the random effects cannot be seen as being normally distributed? This issue will be discussed next in the context of nonparametric growth models.

Nonparametric growth models

In this case the distribution of the random effects can be left unspecified and will be estimated. For estimation the EM algorithm can be employed, and within the E-step numerical integration can be used. In numerical integration the integral is substituted by a finite weighted sum of mass points (nodes) as shown in Chapter 1, Figure 7 (left panel). If one were to approximate a normal distribution with numerical integration, Gauss-Hermite quadrature can be used. In this case nodes and weights of the nodes are known and fixed. However, if an unknown distribution needs to be approximated, the nodes (mass points) and weight of the nodes (masses) can be estimated. Together they provide the necessary parameters to capture the unknown distribution of the random effect (see Chapter 1, Figure 7, right panel; there the mass points determine the location of the bars, and the weight determines their height).

For our example of an age-crime relationship that is described by a quadratic growth function, the random effect of the intercept would now not longer be captured by $\beta_{0i} = \alpha_0 + r_{0i}$, but rather through *N* different nodes that would be called classes in the latent variable framework. The full estimation equation would be expressed by $\ln(\lambda_{it|c_i=n}) = \beta_{0n} + \beta_1 x_t + \beta_2 x_t^2$. The subscript *i* on the intercept growth factor is replaced with *n* indicating a particular class (*node*) in the unknown distribution. With this model specification there is now no longer any within-class variation in the intercept growth factor and there is no error term for the growth factors. Instead, there are *k* different growth factor means. Note that the ‘classes’ do not have a substantive interpretation, but are needed as mass points to approximate the variation in intercepts.

Growth mixture model and latent class growth model

Let us step aside for a moment and consider general mixture-modeling applications. In such a model, be it a growth mixture model or a latent class growth model, the classes are usually perceived as being substantively different in their development. That is, not only do the intercept growth factor means differ across classes, but so too do their slope factor means. Thus, they too should have a subscript *k*, as indicated in the equations in Figure 3. In the growth mixture model, the growth factors, β_{0i}, β_{1i} and β_{2i} may vary randomly among individuals. However, in marked contrast to conventional growth models, these random effect models can be specified for *k* unobserved subpopulations or classes.

Insert Figure 3 about here

The key differences among the classes are typically found in the fixed effects α_0, α_1, and α_2, which may differ for each of the *k* classes of *C* but do not necessarily do so. Likewise, slopes and intercepts may have random effects, but do not necessarily do so. That said, if all variances in the growth factors, β_0, β_1, and β_2, are set to zero, a growth mixture model specification would look like the specification of a latent class growth model (group based trajectory model). If random effects are specified in a growth mixture model they are again latent variables with a normal distribution. Here the normality assumption implies now that for each latent class the individual variation around the expected trajectory is centered on the estimated intercept and slopes for the respective class, with symmetric deviation on both sides. Note that each individual has a probability of membership in each of the classes, and the individual's score on the growth factors can be estimated.

Nonparametric growth mixture model

Having discussed the nonparametric version of a conventional growth model, and the similarities and differences in the latent class growth model and the growth mixture model, we can now combine the different elements and take a closer look at the nonparametric version of a growth mixture model. As the name indicates the nonparametric GMM does not rely on any distributional assumption for the random effects. Instead, the model is now specified such that additional latent classes are estimated to capture the potentially non-normal distribution within the growth mixture classes.

Consider, for example, a growth mixture model with random intercept (and no random effects on the linear and quadratic slope parameters): $\ln(\lambda_{it|c_i=k}) = \beta_{0ki} + \beta_{1k}x_t + \beta_{2k}x_t^2$, and $\beta_{0ki} = \alpha_{0k} + r_{0ki}$. A nonparametric version of this growth mixture model would use classes to capture the variation of the intercept within each of the *k* substantive classes established before. The overall trajectory shape of the *k* substantive classes will not change. That is, each substantive class is still defined by the same slope and quadratic term. However, there will now be additional *n* classes that differ only in the estimated intercept term and are used to capture the distribution of the random effect on the intercept.

It might be helpful here to look at the example model specification in Figure 4 (which resembles to what would be specified in the Mplus software). A quadratic growth model is specified for an observed outcome variable. The slope variance *s* and the variance *q* for the quadratic random effect are set to zero (`s@0 and q@0`). Only a random intercept is estimated in this growth mixture model. In the lower part of the first column of Figure 4 (labeled GMM) a mean intercept β_0 represented by [i], a slope β_1 represented by [s], and a quadratic term β_2 represented by [q] will be estimated for three substantive classes. This results in nine estimated growth parameters for the Poisson part of the growth mixture model, three for each class

----

Insert Figure 4 about here

A nonparametric version of the same growth mixture model is displayed in the middle column of Figure 4, labeled nonparametric GMM. Two latent class variables are specified in this model: *csub*, with three classes that capture the substantive different trajectories found in the GMM model, and *cnp*, to capture the distribution of the random intercept. This second class variable is allowed to have two classes, which vary in their estimated intercept factor means but not in the slope or the quadratic means, as seen in the bottom half of Figure 4. For the first substantive class (csub=1) four parameters are estimated. The estimated value for the two intercepts varies across np-classes that capture the distribution of the random intercept; the value for the estimated slope and quadratic term does not (see the values 2 and 3 for both s and q). Note that by having the slope factor means for the linear and quadratic term [s q] be the same across the np-classes within each substantive class, the assumption is made that the linear and quadratic slope growth factors are uncorrelated with the intercept growth factor. This assumption is not made by the LCGA model, as outlined in the left column of Figure 4, where all growth factor means (intercept, slope and quadratic term) can be different across classes and all growth factor variances are set to zero. That said, a latent class growth model could give a result where the estimated slope factor means ([s] and [q]) vary across substantive classes without being correlated with the estimated mean intercept factor. An LCGA model could therefore lead to the same result as a nonparametric growth mixture model.

The model set-up shows the similarity between these models. But the similarity should not disguise the important theoretical implications of the different model specifications. We will come back to this point when we discuss the different interpretations of the model results for the two data examples in the next section.

Application of the three different modeling strategies

To discuss the different modeling results for the Cambridge data as well as the Philadelphia data, we will build on past results from Kreuter and Muthén (2006), Roeder et al. (1999) as well as D'Unger, Land, McCall, & Nagin (1998). For both data sets we will select a set of growth models that have been shown to represent the data well. We start the comparison of the growth mixture, latent class growth, and nonparametric growth mixture solutions with a discussion of the model statistics followed by a comparison of the resulting mean trajectories, as well as the assignment of most likely class membership.

Discussion of model statistics

A common challenge for all of the latent variable models discussed here is the decision regarding the number of classes needed to best represent the data (see, e.g., McLachlan & Peel, 2000). Objective criteria for doing so have been a matter of some controversy. Kreuter and Muthén (2006) used a different set of statistics to evaluate different models for the Cambridge data. The comparison of the log-likelihood values was used as an indicator for the appropriate number of classes. However, this likelihood ratio test was not used as the sole decision criterion, since for these models it does not have the usual large-sample chi-square distribution due to the class probability parameter being at the border of its admissible space (Muthén, 2004). Therefore, the comparison of log-likelihood values was supplemented by alternative procedure like the Bayesian Information Criterion (BIC; Schwartz, 1978), new mixture tests like the likelihood-ratio test proposed by Lo, Mendell, and Rubin (LMR; 2001), and the bootstrap likelihood ratio-test (McLachlan & Peel, 2000; Nylund, Asparouhov, & Muthén, in press). Results from models fit to the Cambridge data and Philadelphia data are presented in Table 1, and described below.

Insert Table 1 about here

Cambridge data. For the Cambridge data, Kreuter and Muthén (2006) identified a three-class growth mixture model (GMM), a six-class nonparametric growth model (GM np) and a five-class latent class growth model (LCGA) to fit the data best within each of these modeling types. For the three-class GMM, a model with two different developmental trajectories (labeled as GMM (zip) 2+0 in Table 1) showed the best fit. This model has 12 parameters: two parameters for class membership, three parameters (intercept, linear slope and quadratic slope) for both count trajectories, one for the intercept variance, two for the slope and a quadratic slope for the zero-inflation part of the model, and one parameter for the probability of being in the zero class at each point in time. The log-likelihood value for this model is -1,454.7 with a BIC of 2,981.5.

Of the different nonparametric GMM models, a nonparametric model with two support points (nodes) for intercept variation in one non-zero class and three in the other performed best using log-likelihood values, BIC and BLRT as indicators of model fit (labeled as *GMM np zip* in

Table 1). This model, which has 15 parameters, yielded a log likelihood value of -1,444.4 and a BIC of 2,978.8. Among the LCGA models (indicated as *LCGA (zip)* in Table 1), both four- and five-class models performed well. The five-class solution, which has 22 parameters, has a log-likelihood value of -1,441 and a BIC of 3,014.

Philadelphia data. The data from the *Philadelphia Cohort Study* were examined in a similar fashion. Following D'Unger et al. (1998), a random subset of 1000 observations was used for this model comparison. To keep the parallel with the Cambridge data results, we decided to use the results for the five-class latent class growth model (a solution found to fit the data best in the D'Unger et al. study). Also, while the six-class latent class growth model had a slightly better BIC value than the five-class model (6,504 vs. 6,506), the LMR likelihood ratio test failed to reject a five-class model in favor of a six-class model. We matched these results to a three-class GMM model as well as a nonparametric GMM model with three substantive classes and two nonparametric classes.[1] A noticeable departure from the Cambridge data was that, in this case, specifying an explicit zero class was not necessary for any of these models; however, the outcome variable was specified as zero-inflated Poisson, just like it was in the models for the Cambridge data.

For the Philadelphia data the three-class GMM model has a likelihood of -3,173.6 with 15 parameters, resulting in a BIC value of 6,450.8 (see bottom half of Table 1). The nonparametric version of this GMM model, which has 18 parameters, has a slightly higher log-likelihood value of -3,172.4, resulting in a BIC of 6,469.2. The latent class growth model with five classes and 22 parameters has a log-likelihood value of -3,177.1.

---

[1] Note that in the GMM model we allow for a random intercept only. A more detailed analysis of the full Philadelphia cohort study can be found in Muthén and Asparouhov (2006). Note that a model with random effect for the intercept and the linear slope fits the data better and changes the results somewhat. The analysis here is kept simpler for illustrative purposes.

We will look in detail at the results for the trajectories later in the chapter. For now, what is interesting to note at this point is that the log-likelihood values for the three models for each of these datasets are very similar. In fact, they are much more similar to each other than the results within one model type when different numbers of classes are used, or when the models are run without the zero-inflated Poisson specification (see Appendix). The biggest difference between the models is in the number of parameters used to achieve these likelihoods (with the lowest number in GMM and the highest in LCGA) and the values of entropy. Entropy is a measure used for the separation of the latent classes which is based on the posterior class membership probabilities (Muthén & Muthén, 1998-2006). Entropy measures capture how well one can predict class membership given the observed outcomes. Values range from 0 to 1, and high values are preferred. However, entropy measures are by definition a function of the number of classes. If one were to fit a model with as many classes as there are observations, entropy would necessarily be 1. We therefore do not want to overemphasize this value, but will come back to this difference in entropy when we compare class assignment and predictive power of the different models.

Lastly it should be noted that all of these mixture models appear to have a considerably lower BIC values than the conventional growth model with a random intercept term. For the random subset of 1000 observations from the Philadelphia data, the BIC was 6,528.2 for a model with 7 parameters and a log-likelihood value of -3,239.8. The distribution of factor scores presented in the next section will illustrate this effect.

Factor score distribution

The two graphs in Figure 5 show the distribution of intercept factor scores for the conventional growth curve model and the growth mixture model for the Philadelphia data. The picture on the bottom shows for the general growth mixture models a highly skewed distribution of the intercept growth factor. The three-class growth mixture model divides this distribution in three sub-classes. Displayed are here the intercept factors scores for the two biggest classes (roughly 40% and 44 %; according to the most likely class membership 16.1% and 78.5%).

Insert Figure 5 about here

Looking at the bottom picture in Figure 5, we see that there is a high pile of intercept factor scores around zero (-15 on the log-scale) and another pile of factor scores that ranges from -9 to -3 on the log-scale with a mean and median at -6. In the nonparametric version of the growth mixture model these two distributions are represented with two additional sub-classes for each of the mixture classes. The early and higher peaking class has two support points at -13.7 (with a weight of 3%) and -15.5 (with a weight of 17%), the late and very low peaking class has one support point at -5.6 (with a weight of 63%) and -2.6 (with a weight of 12%).

However, looking back at Table 1, the nonparametric version of the GMM has a very similar log-likelihood to the GMM and needs more parameters. It is therefore less parsimonious and does not fit the data better. The normality assumption for the random effect in the GMM model does not seem to be violated so much as to make a nonparametric representation necessary. For the Cambridge data, Kreuter and Muthén (2006) showed the quadrature points for the NP-GMM model and the related intercept factor score distributions. There, too, the nonparametric specification of the random effects in the GMM model appeared to be unnecessary.

Mean trajectories

Growth model results are most often displayed with mean trajectory curves. Figure 6 shows the results for the Philadelphia Cohort study. The three mean trajectories from the GMM are shown in the left-hand panel and the five mean trajectories for the LCGA are shown in the right panel. The results of both models capture three 'themes': a peak in mid-teenage years (age 15) with a steep decline, a peak in later teenage years (age 16/17) with a slower decline and very low peaking in early adulthood (age 19/20). The height of the GMM trajectories is notably lower than those of the LCGA, but one should not forget that additional variance terms are estimated for the GMM that allow for a variation around the means displayed here. One could display the GMM curves not just conditional on the mean but conditional on one standard deviation above and one standard deviation below the mean for any given class. In this case the graphs would look more similar to the NP-GMM result.

Insert Figure 6 about here

Looking back to the model setup in Figure 4, the NP-GMM falls between GMM and LCGA. Like LCGA there is no variance estimated for the trajectory curves, or to phrase it differently, all random effects for intercept, slope, and quadratic terms are set to zero. However, the parameters for slope and intercept are constrained to be equal for some of the classes, which allows for a nonparametric representation of the variation of the intercept in the GMM model. The estimated classes that form the two solid lines in the NP-GMM graph of Figure 6 are variations on the trajectory shape represented with a solid line in the GMM graph of Figure 6. Likewise, the two short-dashed lines are variations of Class 1 in the GMM graph and the dotdashed lines are variations of the dot-dash line in the GMM graph. For all three substantive classes the np-class with the larger estimated class size is marked with a thicker line in Figure 6.

The LCGA results follow similar themes with early and late peaking classes. However there are shifts in form and location of the trajectories compared to the GMM and NP-GMM model results, as well as changes in the estimated class sizes. The two late-peaking classes in LCGA peak at age 19, not 17 as in the GMM and NP-GMM results. Additionally, the peak height of the earlier peaking class trajectories is different from those in the NP-GMM solution. For the Philadelphia data the early high peaking class in the LCGA has an estimated class size of little more than 1% (C 4) and occurs at age 15. In the GMM-NP the early high peaking class has an estimated class size of 3% (Csub1np1) and is not as high peaking as the LCGA class. The even higher peaking class with an estimated class size of almost 1% (Csub2np1) peaks at a later age. And, the LCGA has a late-peaking 4% class 3, peaking at 19, while GMM-NP has 2 later-peaking classes with 1 and 5%, both peaking at age 17.

For the Cambridge data the results for the NP-GMM and the LCGA model were very similar (Kreuter & Muthén, 2006). Figure 7 shows the estimated average number of convictions at each time point plotted in the LCGA estimates against the estimates for each time point in the NP-GMM model for each of the five LCGA classes. The dashed diagonal line is the identity line. Points fall on this line if the estimated means are exactly the same at each time point for both models. While there is some slight deviation from the identity line, this graph nevertheless shows the similarity in the model results for the Cambridge data (further modeling details can be found in Kreuter & Muthén, 2006).

Insert Figure 7 about here

For a substantive researcher attention should be paid to shape and location of the estimated trajectories. If the shape of the trajectories is the same for a set of classes and the only difference between the classes is in how high or low these trajectories start, then there is reason to check carefully if this indicates that the classes are solely variations on the same theme (nonparametric ways to capture a random intercept effect) or if they really have substantive meaning. The latter can be guided by theory, but cannot be decided without extending the models. The relationship of the classes to antecedents and consequences can be information in this regard. For example, if the NP-GMM sub-classes that represent the distribution of the intercept factor scores of the GMM have different predictive power for a distal outcome, or if covariates predict the subclasses differently, then it would be reasonable to relax the restrictions on the growth factor means and move to an LCGA. On the other hand, LCGA might tempt the research to try a substantive interpretation for each of the LCGA classes whereas some of those classes are only needed to capture random variation in the growth factor(s) and GMM would come up with fewer substantively different classes. The model statistics for the Philadelphia cohort study suggest this, with GMM resulting in a higher log-likelihood and lower BIC value than the LCGA. But again, knowledge regarding the interpretation of the classes can only be gained in relating classes and growth factors to variables outside the model (antecedents and consequences).

One note of caution should be added here. So far we only discussed random variation around an estimated mean intercept trajectory. That is, we discussed that shifts along the vertical axis are a random effect that might call for a nonparametric representation. A similar thought can be brought up for shifts along the horizontal axis. For these example data sets this would mean that the number of convictions or police contacts in certain life periods is more predictive of later consequences than at which age those convictions or police contacts happen. This shows once again that successful modeling needs to combine substantive and technical knowledge.

Summary

This chapter illustrates and compares three different modeling approaches for longitudinal data. The approaches considered include latent class growth analysis as well as parametric and nonparametric versions of growth mixture models. The analyses show that researchers might not want to make *a priori* decisions regarding whether to assume continuously varied growth or to rely entirely on substantive classes to capture the variation in growth. A growth mixture model where random effects are allowed within classes can be an alternative. Or, if the normality assumption is questioned, a nonparametric version may be considered.

For both the Philadelphia and Cambridge data we saw that mixture models are needed to summarize the developmental trajectories. We also saw that the variation in the intercept random effect of a GMM can be represented in a nonparametric way. The resulting estimates for the nonparametric model seem to fit the data slightly better than the other two mixture specifications for the Cambridge data. Note that in the Cambridge data the resulting trajectories for an LCGA model are very similar to the NP-GMM results. This is not the case for the the Philadelphia data. The parametric and non-parametric version of the GMM model show very similar growth trajectories. However, the resulting trajectories of the LCGA differ in class sizes and slope compared to the parametric and non-parametric GMM.

In both cases, but in particular for the Philadelphia data, the next analysis step would be to try models that relate the growth factor variation in the GMM model, as well as the different classes in NP-GMM and LCGA, to distal outcomes or antecedents. If it is the case that the NP-

GMM sub-classes have different predictive power regarding a distal outcome, for example, then this could be sign that these classes represent more than a mere variation on a random intercept. In addition the full Philadelphia data could be explored with models that allow for random linear and quadratic slopes (see Muthén & Asparouhov, 2006).

References

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, *55*, 117–128.

Berk, R. (2004): *Regression analysis: A constructive critique.* Thousand Oaks, CA: Sage.

Blokland A. A. J., & Nieuwbeerta, P. (2005). The effects of life circumstances on longitudinal trajectories of offending. *Criminology*, *43*, 1203-1240.

Brame, R., Mulvey, E. P., & Piquero A.R. (2001). On the development of different kinds of criminal activity. *Sociological Methods & Research*, *29*, 319-341.

Bushway, S., Brame, R., & Paternoster, R. (1999). Assessing stability and change in criminal offending: A comparison of random effects, semiparametric, and fixed effects modeling strategies. *Journal of Quantitative Criminology*, *15*, 23-61.

D'Unger, A. V., Land, K. C., McCall, P. L., & Nagin, D. S. (1998). How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses. *American Journal of Sociology*, *103*, 1593-1630.

Elliot, D. (1985). *National Youth Survey 1976-1980*: *Wave I-V*. Ann Arbor, MI: Behavioral Research Institute, Inter-University Consortium for Political and Social Research.

Farrington, D. (1986). *The Cambridge study on delinquency: Long term follow-up*. Cambridge, MA: Cambridge University Press.

Farrington, D. P., & West, D. J. (1990). *The Cambridge study in delinquent development: A long-term follow-up of 411 London males*. In H. J. Kerner & G. Kaiser (Eds.), *Kriminalitaet*. Berlin: Springer.

Farrington, D. P., & West, D. J. (1993). Criminal, penal and life histories of chronic offenders: Risk and protective factors and early identification. *Criminal Behaviour and Mental Health*, *3*, 492-523.

Fergusson, D. M., & Horwood L. J. (2002). Male and female offending trajectories. *Development and Psychopathology*, *14*, 159-177.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, *56*, 1030-1039.

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, *52*, 271-320.

Hedeker, D. (2000). *A fully semi-parametric mixed-effects regression model for categorical outcomes*. Paper presented at the Joint Statistical Meetings, Indianapolis, IN.

Hedeker, D., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, *50*, 933-944.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.

Hirschi, T., & Gottfredson, M. (1983). Age and the explanation of crime. *American Journal of Sociology, 89*, 552-584.

Kreuter, F., & Muthén. B. (2006). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. Under review. http://www.statmodel.com/download/kreutermuthen2006_34.pdf

Lambert, D. (1992). Zero-inflated Poisson regression with application to defects in manufacturing. *Technometrics*, *34*, 1-14.

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

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

Moffitt, T. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674-701.

Muthén, B. (2001a). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1-33). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B. (2001b). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 289-322). Washington, D.C.: American Psychological Association.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*, 81-117.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage.

Muthén, B. & Asparouhov, T. (2006). Growth mixture analysis: Models with non-Gaussian random effects. Forthcoming. In G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs (Eds.), *Advances in Longitudinal Data Analysis*. Chapman & Hall/CRC Press.

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

Muthén, L., & Muthén, B. (1998–2006). Mplus user's guide. Los Angeles, CA.

Nagin, D. S., Farrington, D. P., & Moffitt, T. E. (1995). Life-course trajectories of different types of offenders. *Criminology*, *33*, 111-139.

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

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

Nieuwbeerta, P., & Blokland, A. (2003). *Criminal careers in adult Dutch offenders* (codebook and documentation). Leiden: NCSR.

Nylund, K. L., Asparouhov, T., & Muthén, B. (in press). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal.*

Piquero, A. R., Farrington, D. P., & Blumstein, A. (2003). The criminal career paradigm: Background and recent developments. *Crime and Justice: A Review of Research*, *30*, 359-506.

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

Raudenbush, S. W. (2005): How do we study "What Happens Next"? *The Annals of the American Academy of Political and Social Science*, *602*, 131-144.

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

Sampson, R. J., & Laub, J. H. (2005). Seductions of methods: Rejoinder To Nagin and Tremblay's "Developmental Trajectory Groups: Fact or Fiction?" *Criminology*, *43*, 905-913.

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

Tracy, P., Wolfgang, M. E., & Figlio, R. M. (1990). *Delinquency careers in two birth cohorts*. New York: Plenum Press.

Appendix

Note that the ZIP model is already a special case of a finite mixture model with two classes. For each observation the probability is estimated to be in either the "zero" class, those with structural zeros, or the "non-zero" class, those with random zeros or at least one conviction. For the zero class a zero count occurs with probability one. For the non-zero class, the probability of a zero count is expressed with a Poisson process. The interesting feature for the ZIP, or its expression as a two-class model, is that the probability of being in the zero class can be modeled by covariates that are different from those that predict the counts for the non-zero class. The same is true when allowing for a zero class in the growth trajectory modeling.

More formally, for the present application this model can be represented as follows: At each individual time point a count outcome variable U_{ti} (the number of conviction at each time point t for individual i) is distributed as ZIP (Roeder et al., 1999).

$$U_{ti} \sim \begin{cases} 0 \quad \text{with probability} \quad p_{ij} \\ \text{Poisson}(\lambda_{ti}) \text{ with probability} \quad 1 - p_{ij} \end{cases}.$$

The parameters p_{it} and λ_{it} can be represented with $\text{logit}(\rho_{ti}) = \log[\rho_{\text{ti}}/1 - \rho_{\text{ti}}] = \text{X}_{\text{t}}\gamma_{ti}$

and $\ln(\lambda_{ti}) = \text{X}_{ti}\beta_i$.

Footnotes

[1] We thank Tihomir Asparouhov, Shawn Bushway, John Laub, Katherine Masyn, Daniel Nagin, Paul Nieuwbeerta and the participants of the 2006 CILVR conference at the University of Maryland for stimulating discussions that shaped our perspective for this paper. Michael Lemay was of great help in data preparation and analysis.

[2] The authors contributed equally to this chapter.

[3] Eight boys died during the observational period and are not included in the analysis.

[4] The Cambridge data were grouped into bi-annual intervals starting with age 10. The Philadelphia data were also grouped into bi-annual intervals starting with age 5. In addition, we reduced the age range for the Cambridge data to age 10-27 (see Kreuter & Muthén, 2006).

[5] For an overview of this modeling framework, see Muthén (2002). For a step-by-step introduction to applying latent variable models to longitudinal data, see Muthén (2004).

[6] See Berk (2004) for a comprehensive summary.

Table 1

Growth Mixture Model, Nonparametric Representation of Growth Mixture Model, and Latent Class Growth Analysis (All With Zero-Inflation)

Model	**Classes**	**LogL**	**No. of parameters**	**BIC**	**Entropy**
		Cambridge Data			
GMM (zip)	2 + 0	-1,454.7	12	2,981.5	0.493
GMM np zip	2 + 3 + 0	-1,444.4	15	2,978.8	0.660
LCGA (zip)	5	-1,441.0	22	3,014.0	0.814
		Philadelphia Cohort Study ’58 (subset n=1000)			
GMM (zip)	3	-3,173.6	15	6,450.8	0.238
GMM np zip	2+2+2	-3,172.4	18	6,469.2	0.704
LCGA (zip)	5	-3,177.1	22	6,506.1	0.817

Figure 1. Distribution of police contacts over the ages 5 to 26 for the Philadelphia Cohort Study

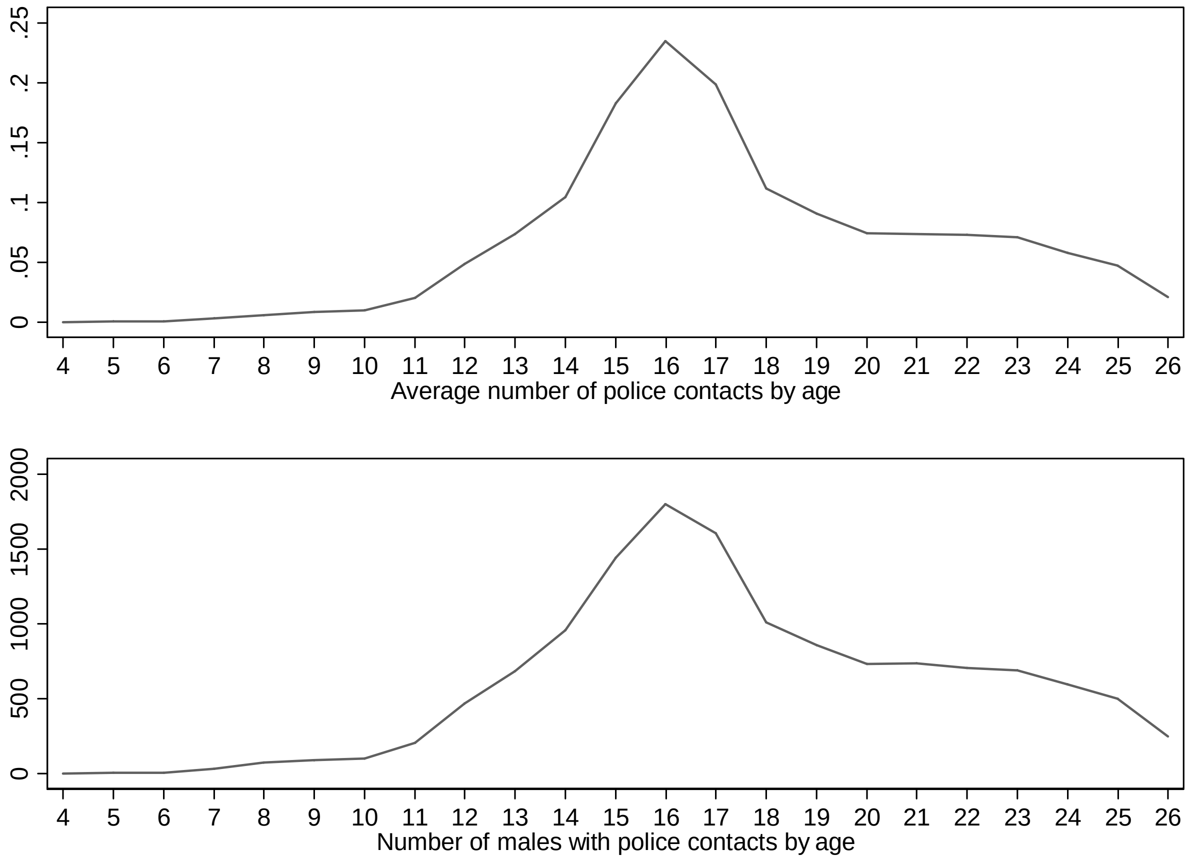


Figure 2. Distribution of convictions over the ages 10 to 40 for the Cambridge Study

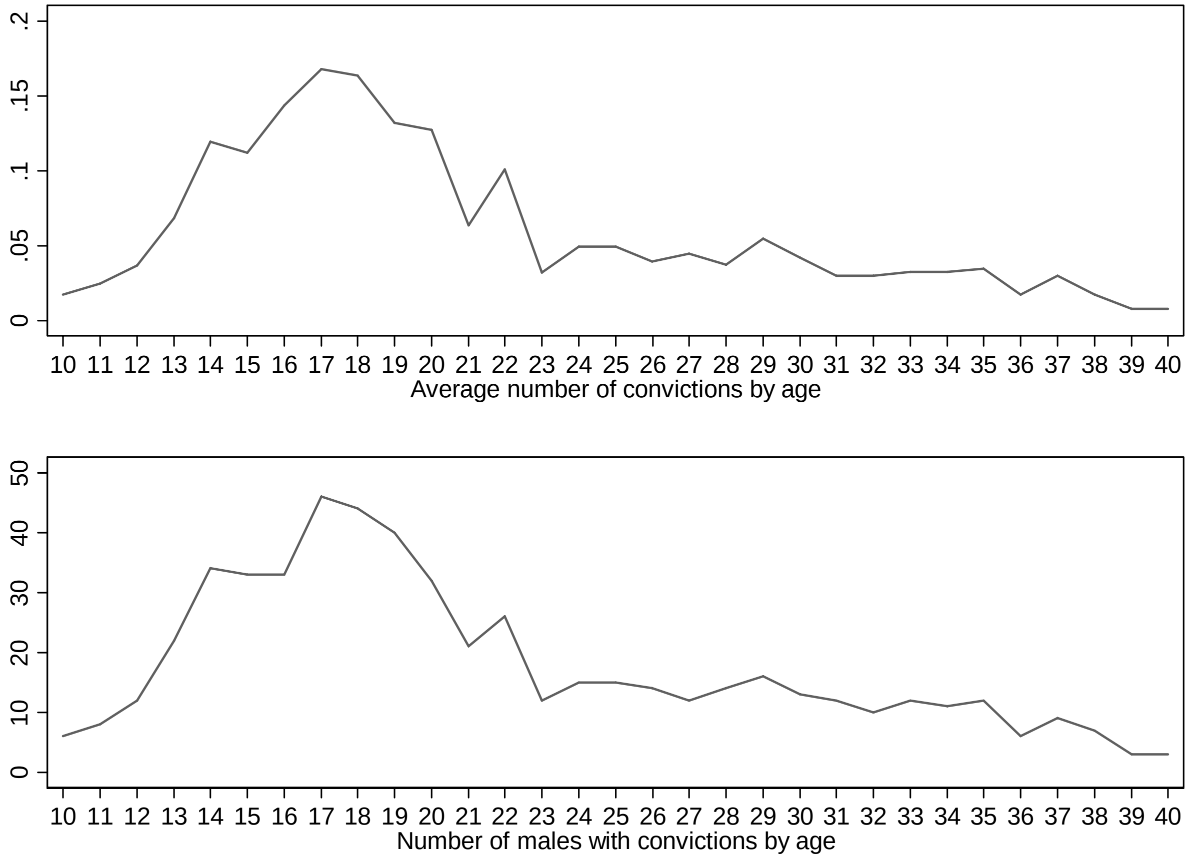


Figure 3. Specification of general growth mixture model and latent class growth model

Growth Mixture Model	Latent Class Growth Model
$\ln(\lambda_{it\|c_i=k}) = \beta_{0ki} + \beta_{1ki}x_t + \beta_{2ki}x_t^2$	$\ln(\lambda_{it\|c_i=k}) = \beta_{0k} + \beta_{1k}x_t + \beta_{2k}x_t^2$
$\beta_{0ki} = \alpha_{0k} + r_{0ki}$	$\beta_{0k} = \alpha_{0k}$
$\beta_{1ki} = \alpha_{1k} + r_{1ki}$	$\beta_{1k} = \alpha_{1k}$
$\beta_{2ki} = \alpha_{2k} + r_{2ki}$	$\beta_{2k} = \alpha_{2k}$

Figure 4. Example model specification for GMM, NP-GMM and LCGA

GMM CLASSES = csub(3);	*Non-parametric GMM* CLASSES = csub(3) cnp(2);	*LCGA* CLASSES = csub(5);
s@0; q@0;	model csub: %csub#1% [s-q]; %csub#2% [s-q]; %csub#3% [s-q];	i@0; s@0; q@0;

Estimated Growth Parameters

GMM

csub	[i	s	q]
1	1	2	3
2	4	5	6
3	7	8	9

Non-parametric GMM

csub	cnp	[i	s	q]
1	1	1	2	3
1	2	4	2	3
2	1	5	6	7
2	2	8	6	7
3	1	9	10	11
3	2	12	10	11

LCGA – 5 classes

csub	[i	s	q]
1	1	2	3
2	4	5	6
3	7	8	9
4	10	11	12
5	13	14	15

Figure 5. Distribution of intercept growth factor scores for the conventional growth model and the two biggest classes of the GMM model (class assignment according to most likely class membership) for the Philadelphia data.

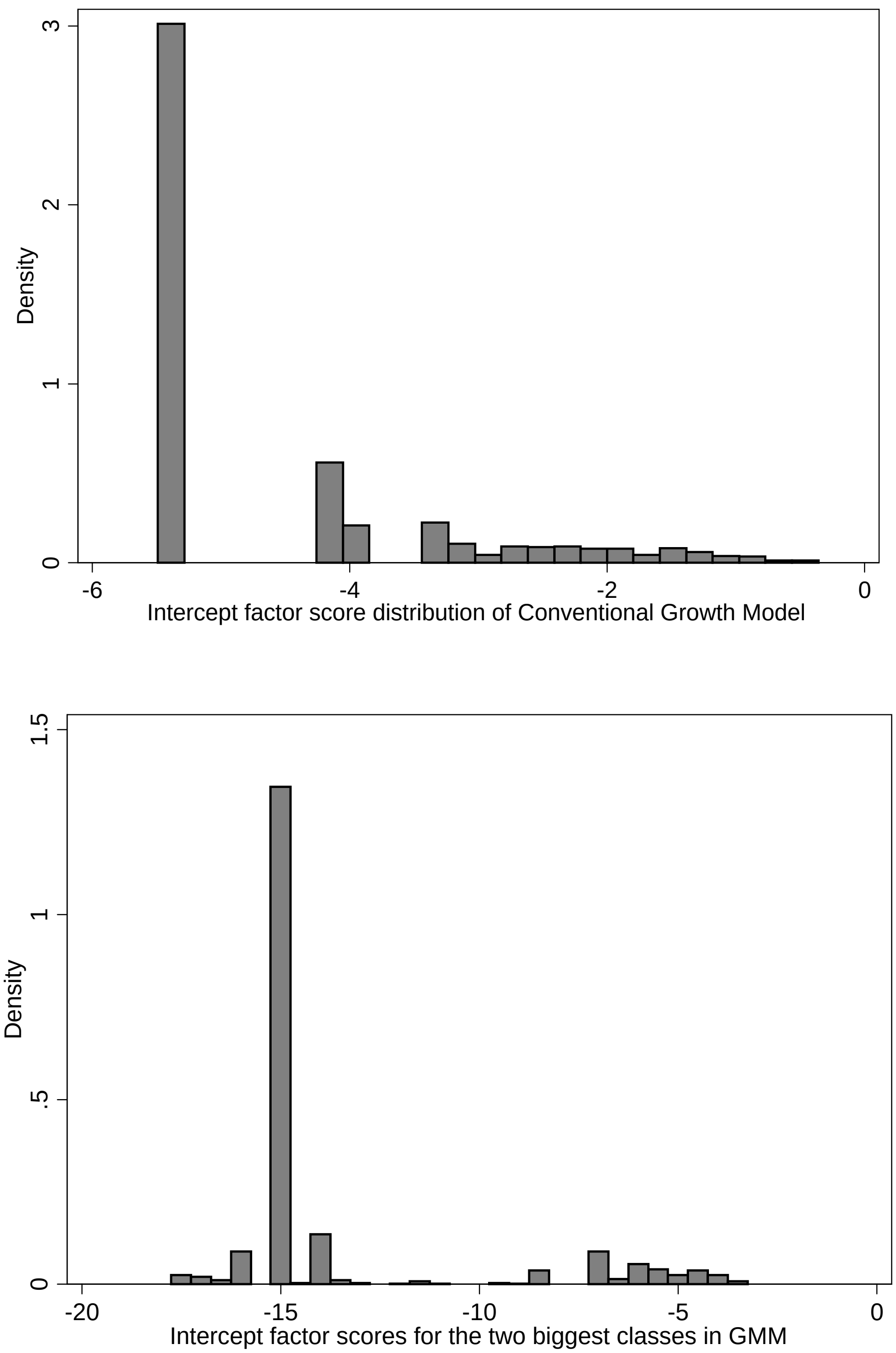
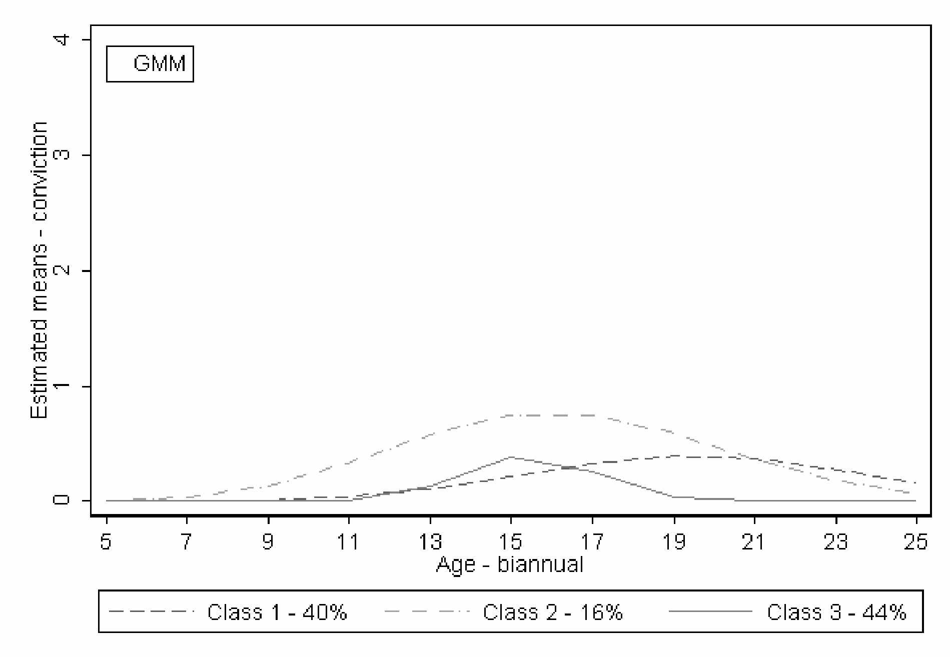
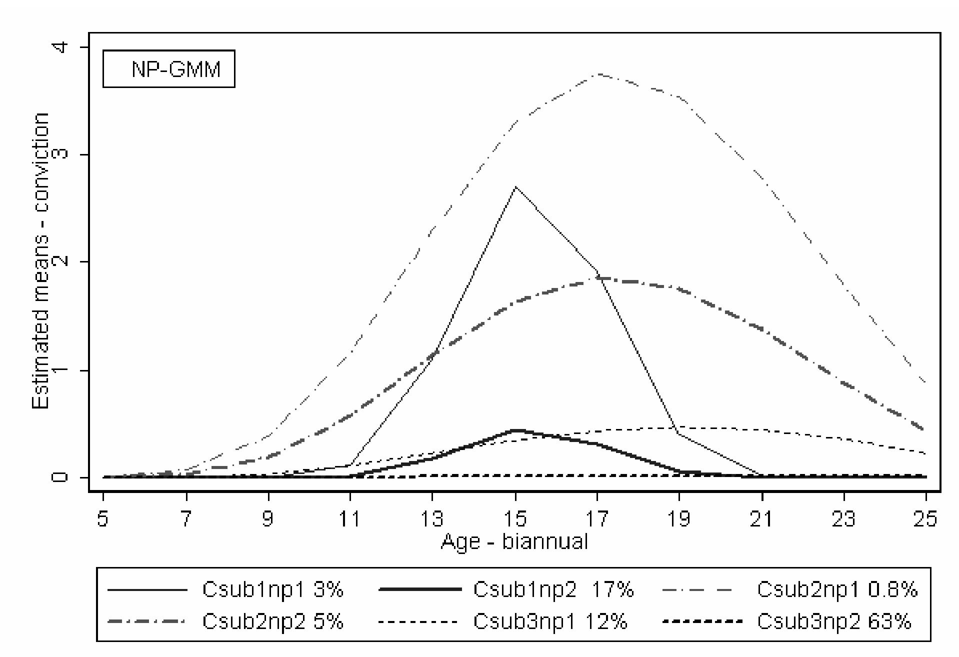


Figure 6. Mean trajectories for the three-class, its nonparametric GMM, and latent class growth model





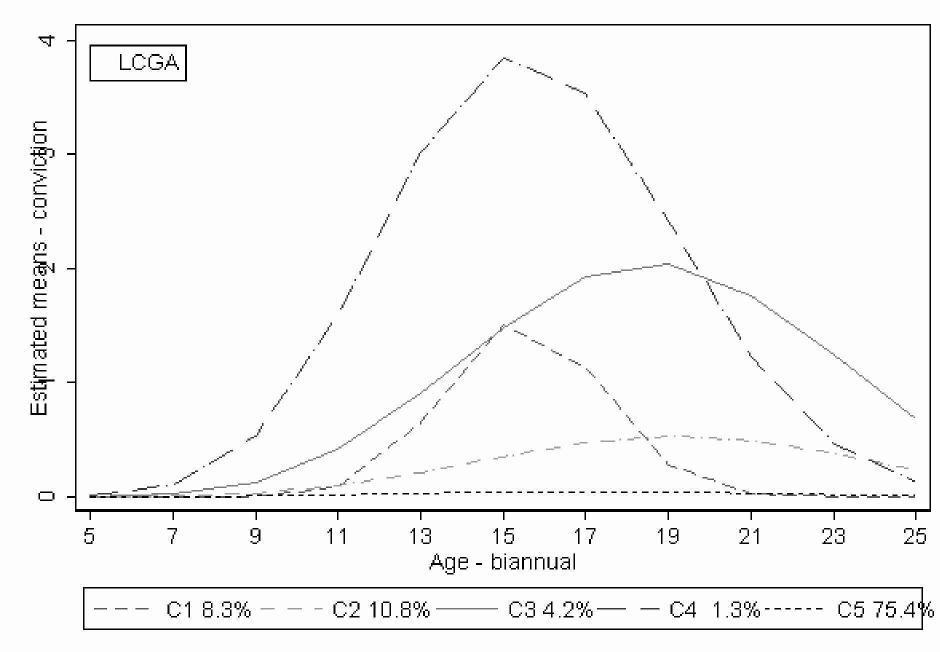

LCGA
Estimated means - conviction
4
3
2
1
0
5
7
9
11
13
15
17
19
21
23
25
Age - biannual
C1 8.3%
C2 10.8%
C3 4.2%
C4 1.3%
C5 75.4%

Figure 7. Cambridge data: LCGA estimated average number of convictions at each time point plotted against NP-GMM estimates for each time point for each of the five LCGA classes