

Measuring Long Term Individual Trajectories of Offending Using Multiple Methods

Shawn D. Bushway · Gary Sweeten · Paul Nieuwebeerta

Published online: 9 June 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Criminal career researchers and developmental criminologists have identified describing individual trajectories of offending over time as a key research question. In response, recently various statistical methods have been developed and used to describe individual offending patterns over the life-course. Two approaches that are prominent in the current literature are standard growth curve modeling (GCM) and group-based trajectory models (GTM). The goal of this paper is to explore ways in which these different models with different sets of assumptions, do in fact lead to different outcomes on individual trajectories. Using a particularly rich dataset, the criminal career and life-course study, we estimate a unique trajectory for each individual in the sample using the GCM and GTM. We also estimate separate trajectories for each individual directly using the long time series. We then compare these three separate trajectories for each individual. We find that the average trajectories obtained from the different approaches match each other. However, for any given individual, these approaches tell very different stories. For example, each method identifies a rather different set of individuals as desistors. This comparison highlights the strengths and weaknesses of each approach, and more broadly, it reveals the uncertainty involved with measuring long term patterns of change in latent propensity to commit crimes.

S. D. Bushway (✉)

School of Criminal Justice, University at Albany, Albany, NY, USA
e-mail: sbushway@albany.edu

G. Sweeten

School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

P. Nieuwebeerta

Department of Criminology, Leiden University, Leiden, The Netherlands

P. Nieuwebeerta

The Netherlands Institute for the Study of Crime and Law Enforcement (NSCR),
Amsterdam, The Netherlands

P. Nieuwebeerta

Department of Sociology, Utrecht University, Utrecht, The Netherlands

Keywords Trajectories · Growth curves · Life course · Desistance

Introduction

Criminal career researchers have long sought to describe the nature of the criminal career, or path of offending over age. The criminal career paradigm (Blumstein and Cohen 1987; Piquero et al. 2003) focused attention on estimating stochastic parameters that described offending at different stages of individuals' criminal careers. Developmental theorists seek to describe and explain the average path or trajectory of offending over the life-course (Gottfredson and Hirschi 1990; Sampson and Laub 1993; Thornberry 1987; Moffitt 1993). Some of these theorists, most notably Moffitt (1993), advocated typologies, suggesting that there were groups of people with fundamentally different paths. More recently, life course criminology on the other hand started concentrating on *within-individual* developments in crime over time and recognized that the causal factors influencing development may shift as the individual progresses along his or her behavioral pathway. This in turn asks for the temporal ordering of events and of the causal processes leading up to these events to be made explicit. The developmental and life course researchers often differ on the extent to which these trajectories are predetermined. Theories that attempt to more fully address these developmental and life course aspects of criminal behavior have been advanced only in the last 20 years (Thornberry 1997; Farrington 2005).

Until recently, however, two fundamental obstacles existed which prevented further exploitation and testing of the main insight from the criminal career researchers that focus on individual tracks or paths of offending over age.¹ The two obstacles were lack of longitudinal data at the individual level and lack of statistical methods which could provide a more nuanced description of offending over the life-course (Hagan and Palloni 1988).² Since that time, many more individual datasets have been constructed (for some highlights, see Thornberry and Krohn 2003) and a number of statistical approaches have been developed to better describe individual offending over time.

The two most commonly used statistical approaches to the study of offending over the life-course in criminology are standard growth curve models (GCM) and group based trajectory models (GTM). Although much of the discussion in the literature is about the distribution of individual trajectories, these approaches do not estimate individual trajectories. Rather, they make strong (and different) assumptions regarding the distribution of the individual trajectories in the population. The GCM assumes that the population is distributed continuously with individual random effects around each of the key parameters of the growth curve following a known distribution. The key here is that the models assume that the distribution of growth curves can be described by a known parametric distribution, such as a jointly normal distribution. Once this distribution is assumed, the population parameters (mean and variance) can be estimated using standard statistical techniques. GTMs, in contrast, assume that the continuous distribution can be approximated by a discrete number of fixed points. No assumptions about parametric distributions are

¹ But see for example the following empirical work that uses empirical estimates of individual trajectories: Laub et al. (1998), Sampson and Laub (2003, 2005), Nagin et al. (1995), Osgood (2005), Bushway et al. (2003) and Barnett et al. (1989).

² In the 1980s, researchers in the criminal career tradition were forced to assume stable rates of offending for one or two groups of offenders. Sloped trajectories could be achieved through the estimation of a second desistance parameter, which carried with it the unsatisfactory (for some) assumption that desistance was a point in time event, rather than a process (Barnett et al. 1987).

required. The standard growth model is more efficient (i.e. fewer parameters), but requires stronger parametric assumptions than the group-based models.³

There has been a fairly long discussion in criminology about the merits of GTM, and by comparison, GCM. Much of the controversy about the use of GTM involves concerns that the retrospective trajectory groups, which are useful for exploring basic patterns in the data, can then be reified and interpreted as existing prospectively. Critics fear that policymakers will interpret the existence of a chronic group of offenders to mean that there exists a real, identifiably distinct group of offenders who are predetermined to be chronic offenders. This possibility is real, but is not implicit in the method—in fact, the basic GTM statistical model does not identify who is in each group, but rather provides an estimate of the proportion of the population in each group. Bayesian methods are then required to assign a person to a group. Although done less often, GCM can also use Bayesian methods to identify retrospective trajectories or paths for each individual. Neither method says anything about whether these paths are predetermined.

Perhaps part of the problem is that the GTM methods were used almost from the beginning to test typological theories of crime (Nagin et al. 1995), and many of the papers published using GTM focus on between-group variation. For example, the extent to which GTM identifies different groups has been taken as evidence against Moffitt's theory. In reality, typological theories in criminology—like Moffitt's—are not theories about the nature of groups, but about the existence of qualitatively different patterns (and etiologies) of offending over the life course. Therefore, it is vitally important that researchers understand that the GTM's use of groups is nothing more or less than a device by which the underlying distribution of the population can be better explicated. Groups are not a theoretical requirement for this discussion, but rather a statistical device through which parametric assumptions can be avoided.

Of course, since both approaches (GCM and GTM) make different assumptions on the underlying distribution, it is important to understand the relative strengths and weaknesses of each approach. Raudenbush (2001) attempted to lay out some ground rules for when the additional complexity of group based models might be appropriate. His main argument was that if all individuals do follow a similar basic growth pattern, the standard growth curve model will have no difficulty fully capturing the behavior of all individuals. This is especially so, because of the covariance structure of the GCM models (see also Raudenbush 2005). However, he also argues that if individuals do not all follow the same growth pattern, the GCM may not have the flexibility to capture the between individual variation in trajectories. Then the GCM in particular will not capture the behavior of the “non-standard” individuals. In line with this argument, Nagin and his colleagues argue that the GTM, which approximates a continuous distribution of parameters with a discrete number of groups, and thus not bound by a parametric distribution therefore will have more flexibility to accommodate different and also very distinct patterns of individual offending over time (Nagin 2005; Nagin and Tremblay 2005).⁴

³ A third method, growth mixture modeling (GMM) is essentially a combination of the two other methods. It estimates groups, but then allows there to be variation around the parameters that define the groups (Muthén 2001, 2007). We do not apply it in this paper.

⁴ Muthén (2001, 2007) has proposed an alternative, more complex approach, which essentially adds random effects to the group based approach. This random effect approach, based on the standard growth curve model, loosens the restriction that behavior is homogenous within groups. This latter assumption has been a major sticking point for many critics of the group based approach (Raudenbush 2005; Sampson and Laub 2005). Nagin has argued that this added complexity is unnecessary (Nagin 2005). Kreuter and Muthén

The goal of this paper is to use direct estimates of the individual trajectories as a way of evaluating the relative strengths of either GTM or GCM regarding their ability to describe the underlying distribution of offending trajectories in the population. We, however, do not focus on differences in overall model fit (Kreuter and Muthén 2008) or on the underlying distribution of trajectory coefficients. Instead, we use Bayesian techniques to estimate a unique trajectory for each *individual* in the sample using GCM and GTM. In the GCM case, these Bayesian estimates represent a weighted average of the coefficients from the individual trajectory and total sample trajectory. The weights are based on the reliability of the coefficient estimates, which tend to be high for the intercept and slope, but low for the quadratic and cubic terms. In the case of the GTM, the Bayesian estimates are a weighted average of the different group trajectories. We estimate separate trajectories for each individual directly using time-series regression. We do so analyzing data from a particularly rich dataset covering criminal behavior of a large sample over people's full life course, the criminal career and life-course study. To the best of our knowledge, this is the first time anyone has examined the individual trajectories using any of these methods. We compare the individual trajectories from GCM and GTM to that estimated by an individual time series (the individual trajectory model, or ITM) using a number of different comparative metrics. Finally, we use each model to identify individuals as desistors, based on the point estimates of the latent propensities. The models clearly categorize different people as desistors, but we find that there are far more similarities between GCM and GTM relative to ITM. This finding suggests that researchers should focus less on pitting the two models against each other and more on understanding how the models capture the nature of the underlying distribution.

How are Methods for Trajectories Applied?

To understand the background of this paper, it is important to understand why and how methods for trajectories have been applied in life-course criminology. A review of the literature shows four distinct applications.

First, these methods have been used to *describe the average or basic underlying age-crime curve* in the data. Examples with GTM include Broidy et al. (2003), and Nagin and Tremblay (2005). Examples with GCM include Lauritsen (1998), and Raudenbush and Chan (1993), and in a comparison framework, Kreuter and Muthén (2008), which applies both GCM and GTM.

Second, these methods have also been used to *characterize the variety of individual trajectories* underlying the basic pattern. Piquero (2008) reviews the large and growing number of studies (80 at his count) that have applied GTM to longitudinal data sets in criminology. In a sense, these papers focus on average offending over age, but they also begin to look for distinct patterns that can be studied for their own sake, such as chronic offenders. Many of the studies reviewed by Piquero also begin to look for risk factors that can predict "membership" in these different trajectories. Other papers focus specifically on characteristics of the underlying distribution. For example, Hay and Forrest (2006) use GTM to describe the variation in the distribution of self-control trajectories in adolescence. They estimate an eight group model, and conclude that there is a substantial amount of

Footnote 4 continued

(2008) do a side by side comparison in terms of fit using the Cambridge data used by Nagin and Land (1993).

both absolute and relative stability in self control by looking at the paths described by the eight groups. Bushway et al. (2001, 2003) argue for the use of GTM to identify a group of desistors who can then be studied. A number of scholars have also used GTM to look for typologies of offenders as described most prominently by Moffitt (1993). See, for example, work by Nagin et al. (1995), Sampson and Laub (2003) and Blokland et al. (2005).

It is worth noting that there are no examples above in which the GCM is used to identify distinctive paths, for either individuals or groups. GCM is almost always used in criminology as a regression model to relate explanatory variables to offending over time for the group on average. In this case, the GCM is basically modeling the residuals to create more efficient estimates, and little attention is paid to the coefficient estimates on the age/time variables. However, as we will show, the GCM model can easily be used for this purpose, and such use can help shed light on the underlying differences between the two models.

A third application of GCM and GTM approaches is as a *control for individual heterogeneity* as described by these trajectories *as they study the impact of time-varying covariates on offending*. These approaches move beyond fixed effect or random effect approaches that only seek to control for stable fixed effects when studying the impact of time-varying covariates.⁵ Prominent examples of the use of GCM for this purpose include the study of marriage and work by Horney et al. (1995) and Laub and Sampson (2003). In reality, however, the GCM is a residual model, meaning that the random effects summarize what is left rather than control for unobserved heterogeneity. Individuals interested in controlling for individual differences have to take the additional step of decomposing the time-varying covariates into between individual and within individual terms as done in Horney et al. (1995). The GCM will provide results that are very similar to what can be achieved through a fixed effect panel model.

Examples that use GTM to control for individual heterogeneity include Lacourse et al. (2003) for gangs, Nagin et al. (2003) for school failure and Laub et al. (1998) and Blokland and Nieuwbeerta (2005) for marriage. In a related approach, researchers have used the groups from the GTM models to match similar individuals in order to assess the impact of various time-varying covariates on crime and delinquency (Haviland and Nagin 2005; Apel et al. 2007; Nieuwbeerta et al. 2009).

Researchers have also used individual trajectories to control for individual differences. Laub et al. (1998), for example, control for posterior probabilities estimated using GTM in a model of marriage. This approach in effect controls for individual trajectories, since each set of posterior probabilities corresponds to a particular individual trajectory.

Fourth, researchers have started to suggest using GCM and GTM to *identify curves/trajectories which they then try to explain using time-varying covariates*. Osgood (2005) presents a good discussion of this issue within the framework of GCM. In this framework, Osgood is primarily worried about explaining the average age-crime curve. Thornberry et al. (2007) and Blokland and Nieuwbeerta (2005) start to look at this issue of explaining patterns of long term change using the GTM. Blokland and Nieuwbeerta (2005) are concerned primarily (but not exclusively) with explaining the average pattern. Thornberry et al. (2007) focus on the behavior of smaller groups of individuals. The logical extension of this work is to explain individual trajectories.

⁵ Standard panel data analysis that only seek to control for a fixed effect include papers by Gordon et al. (2004) on gangs and Paternoster et al. (2003) on the effect of adolescent work on crime. Fixed effect models are essentially individual trajectory models, where the trajectory is a constant, or a flat line (the mean). Researchers attempt to explain variation around the mean with the time-varying covariates.

In a similar vein, Bhati (2007) used individual trajectories to generate estimates of the amount of offending prevented during periods of incarceration. This is similar to what Blokland and Nieuwebeerta (2007) did using GTM. In this case, rather than investigate whether a certain set of time-varying covariates could change the residual trajectory curve, these researchers use the residual trajectory curve to “fill in” missing holes in the offending trajectories. Eggleston et al. (2004) conduct a similar exercise when they estimate GTM with and without controls for time spent in prison. The difference between the two curves is an estimate of the number of crimes “prevented” by prison. In these examples, the purpose is more descriptive rather than causal but nonetheless involves looking at the relationship between time-varying covariates and offending.

This Study

In this paper, we take the focus off of the more conceptual discussions about possible misconceptions that can arise from using average curves from the entire sample or from groups. Instead we ask about the performance of these methods when they are asked to describe the *individual* trajectories thought to exist in the sample. Both the GCM and GTM models were designed to approximate a continuous distribution of individual trajectories (Raudenbush 2005; Nagin 2005). And, although it has rarely been done in this field, both approaches can be used to estimate individual curves.

The central aims of this paper are (1) to examine the quality of the predicted *average* trajectories, (2) to examine the quality of the estimated *individual* trajectories, and (3) to illustrate how the different assumptions of the models lead to different conclusions in criminological research—especially on desistance from crime.

We are aware of no study that has estimated and compared the *individual* trajectories obtainable from each of these (GCM or GTM) models in any context, including offending. Yet the full model in each case describes not just the average path but the full distribution of individual trajectories. Given that criminology as a discipline has been intensely interested in the underlying distribution of the age-crime curve, this failure to describe the actual population of trajectories represents a shortcoming in life-course research.

The ultimate solution to this problem would involve an explicit study of the distribution of intercept, slope and quadratic terms that result when each individual trajectory is estimated separately. This empirical or estimated distribution could then be compared with the distribution of intercept, slope and quadratic terms estimated by GCM. This comparison would be somewhat harder to accomplish in GTM, but some approximate comparison could be made. While relatively straightforward with linear-regression techniques, such an enterprise is daunting in the non-linear world of most offending analyses.⁶

In this paper, we take a less ambitious first step and compare the individual trajectory for each person (ITM) with Bayesian estimates of the individual trajectories from both the GTM and GCM models. These Bayesian estimates from both the GTM and GCM are weighted averages of the individual and group means, and as such are explicitly biased estimates of the individual trajectories. On the other hand, the use of group data allows for much more precision than can be achieved by focusing only on using individual time-series regression (the ITM model). These individual models, which are based on only 40 observations on average, have very wide confidence intervals, and therefore leave much

⁶ This might explain why to our knowledge no one has done this.

doubt as to the actual path of the individual trajectory. The Bayesian estimates represent an attempt to achieve a balance between bias and precision.

This paper's first contribution, therefore, is to show that the two methods (GCM and GTM) appear to describe the average trajectories equally well. Second, the paper shows that there is a great deal of ambiguity about the nature of the individual trajectories. Third, the paper shows that the methods diverge (predictably) in whom they identify as desistors. This divergence can be taken as a measure of the uncertainty surrounding this question, but also highlights the different strengths and weaknesses of the different approaches.

Data

The data set used in this study is compiled from the large-scale criminal career and life-course study (CCLS). The CCLS is a representative sample of 4% of all the cases of offenses tried in The Netherlands in 1977.⁷ The original sample consists of 5,164 individuals (Nieuwbeerta and Blokland 2003).⁸

We place two restrictions on the sample used in the analyses. First, to avoid problems associated with having only a small number of individuals defining offending trajectories at the oldest ages, we limit the offending trajectories to ages for which data was available on at least 200 individuals. This restriction has the effect of ending our observation of the offending trajectories at age 72. Second, here we limit our analysis to the 4,615, individuals on which information on life circumstances is available, which follows the sample selection used in Blokland et al. (2005). For 464 individuals of the 4,615, the decision of the public prosecutor and/or judge led to an acquittal or "innocent" decision. So, these were not convicted in 1977, whereas the remaining 3,971 were.

Police records pertaining to the 1977 offense that led to inclusion in the study offer information on personal characteristics of the sampled individuals. 11% of the sample are female offenders, and 10% are of non-Dutch origin. Approximately 35% are alcohol dependent and 2% drug dependent in 1977. Death records were searched to account for mortality in the data during the follow-up period. In the 25 years following the sampling offense, 17% of the sampled offenders died before 2003—the end of our follow up period (see also Nieuwbeerta and Piquero 2008).

Measures

Extracts from the general documentation files (GDF) of the Criminal Record Office ("rap sheets") were used to construct the entire criminal careers of the sampled individuals. The GDF contain information on every criminal case that has been registered at the Public Prosecutor's Office. These extracts were supplemented with cases that normally would not be mentioned due to expiration periods. In this way, the entire criminal history before 1977 for every individual in the sample was reconstructed. Note that in The Netherlands a person is not given a "blank sheet" upon becoming an adult and the extracts therefore contain information on both adult and juvenile offenses. Next, every new entry between 1977 and 2003 was recorded. Although the GDF contain information on all offenses that have led to any type of judicial action, we choose to use only information on those criminal

⁷ Note that the offenses tried could end as release, acquittals and convictions.

⁸ For more details about the sample, see also Blokland and Nieuwbeerta (2005), and Blokland et al. (2005).

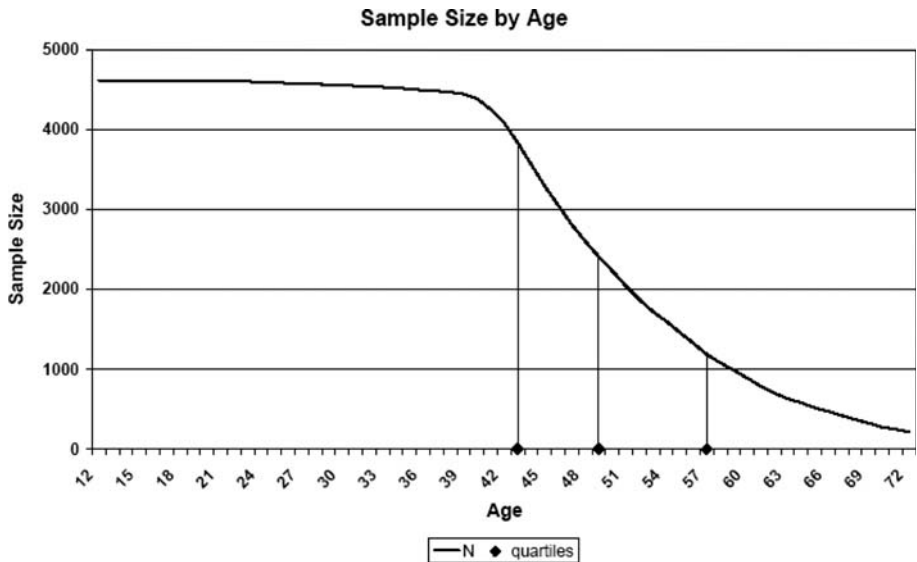


Fig. 1 Sample size by age

law offenses followed by a conviction or a prosecutorial disposition due to policy reasons, thereby excluding non-criminal law offenses (traffic and economic offenses, for example) and cases that resulted in acquittal or a prosecutorial disposition due to technicalities.

Figure 1 shows the number of individuals observed at each age for the final sample. Clearly the follow-up periods for the individuals differ substantially.⁹ The first feature of this figure readers should note is that there is little attrition up until age 40. Second, we have indicated quartiles by length of follow-up in this figure. This shows that we observe one quarter of the sample for 32 years or less (ages 12–43), 50% of the sample for 33–46 years, and a quarter of the sample for over 46 years. 644 persons had no convictions over the entire period they were observed, 917 had convictions in only 1 year, 581 had 2 convictions in two different years, 982 had convictions in 3–5 years, 978 had 6–12 conviction years and 513 had 13 or more years in which they were convicted.

In total our dataset has 184,078 observations, i.e. person-year combinations. A large proportion (87.5%) of these person-year observations are absent any convictions. One conviction is observed in 7.5% of the person-years, 2–5 convictions is observed in 4.6% of the person-years, and just 0.4% of person-years had 6 or more convictions. To simplify the analysis, we therefore focus on prevalence rather than frequency of conviction. This decision allows us to use a simpler logit framework rather than a more complex count model but does not change the fundamental premise of this paper.¹⁰ So, instead of focusing on the rate of conviction, we focus on the probability of conviction in any given year.

⁹ We did a sensitivity analysis in which we tested whether the conclusions differed when reducing the window of observations to the period age 12–50. The results were very similar and did not lead to different conclusions.

¹⁰ We also attempted to model ITM, GCM and GTM curves with Poisson models, retaining the full conviction count information by person-year instead of dichotomizing convictions counts. While this poses little difficulty for GCM and GTM models, it poses difficulties for many of the ITMs. Because we wanted to limit modeling differences between the three methods, we used logistic regression models for each, as they could be consistently applied in each framework.

Methods

Panel Models

We estimate three separate models, the individual trajectory model (ITM), a standard growth curve model (GCM), and a group-based trajectory model (GTM), and use the models to generate individual trajectories from each of the models for each of the 4,615 people in the sample.¹¹

All of the models are cubic polynomial logit models. In other words, we model the probability of offending as a cubic trend over the observed ages, which range from 12 to 72. All models have the same functional form:

$$\text{Ln}\left(\frac{p_t}{1-p_t}\right) = \text{Pr}(Y_t) = \alpha_0 + \alpha_1 \text{Age}_t + \alpha_2 \text{Age}_t^2 + \alpha_3 \text{Age}_t^3 + \varepsilon_t \quad (1)$$

where p_t is the probability of a conviction at age t , and age, age^2 and age^3 the cubic functional form of the relationship between age and the probability of a conviction. This cubic functional form in effect means that the path can go up and down and up again and that the peaks can be asymmetric. Each individual gets his own intercept, slope coefficient, etc. This is assumed to be the same for all three types of models (ITM, GCM and GTM).

We can argue that we need more or less flexibility to truly capture the path. There is, of course, a tension between complicated functional form that can capture variation and simpler functional forms that provide easy to understand descriptions of the basic path. A simple functional form will not capture all meaningful variation, and there well may be interest in explaining variation NOT captured by the retrospective path (Osgood 2005). However, as the models become more flexible, the summary information provided by the paths becomes more complex. Criminologists have generally resolved this debate by using quadratic or cubic functional forms. We follow suit and use the cubic functional form on the conceptual belief that cubic functional form captures the individual age-crime curve path.¹²

The only differences between the ITM, GCM and GTM models are in the parametric assumptions, i.e. in the way the parameters of the models are assumed to be distributed over the population. In the ITM model, there are no assumptions about the way the parameters are distributed over the population. The parameters are unique for every single member of the sample: a separate trajectory is estimated based on that individual's observations. The model estimates are parameters (γ) estimated by a separate analysis for each and every individual i :

$$\begin{aligned} \text{Pr}(Y_{it}) &= \alpha_{0i} + \alpha_{1i} \text{Age}_t + \alpha_{2i} \text{Age}_t^2 + \alpha_{3i} \text{Age}_t^3 + \varepsilon_{it} \\ \text{where } \alpha_{0i} &= \gamma_{0i}; \alpha_{1i} = \gamma_{1i}; \alpha_{2i} = \gamma_{2i}; \alpha_{3i} = \gamma_{3i} \end{aligned} \quad (2)$$

¹¹ For a book length treatment of the GCM and GTM, see respectively Raudenbush and Bryk (2002) and Nagin (2005).

¹² Bushway et al. (2003) and Blokland et al. (2005) have observed that cubic models tend to produce an uptick at the end by definition. Blokland et al. (2005) propose using splines to fix this anomaly. We chose not to use splines in this paper because of the added complexity caused by adding two additional trend terms, particularly for ITM models with limited variation in the dependent variable. As a consequence, all of the pictures will have upticks at the end. While unsightly, we do not believe that the upticks disadvantage any one approach, and therefore do not affect the comparison.

In the GCM model the parameters are assumed to be jointly normally distributed over the population. The model estimates a mean for each of the effect parameters (γ 's) and variation around the means (μ 's):

$$\Pr(Y_{it}) = \alpha_{0i} + \alpha_{1i}\text{Age}_t + \alpha_{2i}\text{Age}_t^2 + \alpha_{3i}\text{Age}_t^3 + \varepsilon_{it} \quad (3)$$

where $\alpha_{0i} = \gamma_{00} + \mu_{0i}$; $\alpha_{1i} = \gamma_{10} + \mu_{1i}$; $\alpha_{2i} = \gamma_{20} + \mu_{2i}$; $\alpha_{3i} = \gamma_{30} + \mu_{3i}$

where μ_{0i} are normally distributed with mean = 0 and variance = σ . These individual random effects on each coefficient are allowed to co-vary.

The GTM assumes that the individual specific parameters differ across groups of individuals but are the same within each group. For each group k , separate trajectories (and thus intercept, slope coefficient, etc.) are estimated:

$$\Pr(Y_{it}) = \alpha_{0i} + \alpha_{1i}\text{Age}_t + \alpha_{2i}\text{Age}_t^2 + \alpha_{3i}\text{Age}_t^3 + \varepsilon_{it} \quad (4)$$

where $\alpha_{0i} = \gamma_{0k}$; $\alpha_{1i} = \gamma_{1k}$; $\alpha_{2i} = \gamma_{2k}$; $\alpha_{3i} = \gamma_{3k}$

Although the GTM assumes that each individual actually belongs to a group, the model itself only estimates the proportion of the population that belongs to each group—it does not identify who belongs to which group. However, Bayesian techniques can be used to estimate posterior probabilities of group membership in each group.

Obtaining Individual Predictions

From the estimated parameters of the models we can obtain predictions for each time point for each individual. For the ITM, the parameters can be used directly to estimate a trajectory for every single individual over the age range in question.¹³ In other words, we get a predicted probability of offending for each year in which we have an observation for the individual.

For the GCM, the full model has random effects on each parameter. Bayesian methods are used to actually estimate an effect for each individual on each parameter. These individual specific parameters can be used to generate a separate trajectory for each individual. These Bayes, or shrinkage, estimates, are a weighted average of the individual estimate from Eq. 2 and the average parameter for the entire sample from Eq. 3.

$$\alpha_{0i}^* = \lambda_i \gamma_{0i} + (1 - \lambda_i) \gamma_{00} \quad (5)$$

where $\lambda_i = \text{reliability} = \text{Var}(\gamma_{00} + \mu_{0i}) / \text{Var}(\gamma_{0i})$

The reliability parameter is the proportion of the total parameter variance that is captured by the random effect around the parameter. Additional variation in the ITM that is not captured by the random effect must be individual specific noise. If the error captured in the two models are essentially the same, then the reliability parameter will be close to 1, and the Bayes estimate will converge to the individual estimate from Eq. 2. However, if the variance around γ_{00} is small relative to the variance in the individual equation, then the individual estimate is unreliable, and the Bayes estimate will converge to the sample average for that parameter.

¹³ The ITM models were inestimable for people with no convictions. As a result, we simply assigned these people a trajectory where the probability was zero throughout the time period.

If the reliability parameter is less than 1, then by definition, the Bayes estimate will be biased away from the individual parameter estimate from the ITM. This deviation is justified on efficiency grounds. In fact, the relative efficiency of the ITM relative to the GCM for any given parameter is the reliability coefficient. This equation makes the tradeoff between bias and efficiency explicit. The estimates of the coefficients for the individual from Eq. 2 are unbiased, but inefficient (large confidence intervals). The estimate of the sample mean from Eq. 3 is a biased, but more efficient, estimate of the individual coefficient. The Bayes estimate represents a weighted average of the two estimates, where the weight is the relative efficiency of the two models.

For the GTM model, the posterior probability of membership in each group for each individual can be generated using Bayesian statistics. The underlying statistical model assumes everyone belongs to a group with probability one—the posterior probability represents our best guess as to which group that person belongs. Imagine a four group model. An individual j might have, for example, a 90% probability of belonging to Group 1, a 10% probability of belonging to Group 2, and no chance of being in Groups 3 or 4. These posterior probabilities can be used to generate an individualized trajectory for each person. In the above example, the predicted value in a year for this person will be 90% of the Group 1 prediction, plus 10% of the Group 2 prediction.¹⁴ In this manner, predicted curves are generated for each individual over the entire observed life-course.

As a result, in each case, we will end up with unique trajectories for each individual from both the ITM, GCM and GTM model. A central aim of this paper is to examine the differences between these obtained individual trajectories.

Comparing Individual Trajectories

To examine how well the GCM and GTM compare with the ITM, we take the simple difference between the estimated probability from the GCM and the ITM, and the GTM and the ITM.¹⁵ We have two basic comparison statistics. The first is to simply subtract predicted probabilities of the ITM from the GTM or GCM for each year covered by the data for that individual to create the *signed difference (SDF)*.¹⁶ This can be thought of as a measure of bias, provided that one remembers that the ITM is also an estimate and not the “true” value. An estimate of zero would imply that the model captures the baseline curve and correctly estimates the probability of being convicted for a person year. A positive SDF indicates that the panel model is overestimating the probability of conviction for that person-year, while a negative SDF indicates that the panel model is underestimating the conviction probability.

The second measure is one of precision. We take the *absolute value of the signed difference (ADF)* which gives an estimate of the average distance of the estimate from the baseline value (ITM) without regard to whether the estimate is too high or low. This is analogous to the standard deviation of the estimate. The SDFs and ADFs can be calculated

¹⁴ Note that creating individual curves from weighted averages of parameters instead of weighted averages of group curves does not take into account covariances and scaling of parameters (e.g. on a logit scale as in this paper), and so yields incorrect individual trajectories (e.g. individual curves that exceed the highest group prediction).

¹⁵ Comparing parameters, as opposed to predicted probabilities, is challenging because of the different norming that each logit model uses. However, it is possible to generate estimates based on the same norm, and we encourage future work that looks at the parameters themselves, rather than the trajectories.

¹⁶ To calculate this area precisely, we numerically integrate the difference between the curves using 100 points per year. This approach gives us an accurate estimate of the area between the curves.

and reported for each person year. We also report average SDFs and ADFs, averaging over (groups of) persons and/or (groups of) person years.

Identifying Desistors

The study of desistors is the study of people who first reach offending levels at which their propensity to offend is discernably different from zero, and then experience a decline in offending propensity such that this is no longer true. Traditionally, desistors are defined based on their observed behavior, e.g. individuals who do not offend for a certain number of years after an age cutoff are defined as desistors. This period of non-offending ranges from 1 year (Warr 1998) to 11 years (Farrington and Hawkins 1991). In contrast, a number of recent articles have stressed that people should be defined as desistors based on their latent propensity to offend, as revealed by observed behavior, rather than on their observed behavior. Bushway et al. (2003), for example, demonstrated that an emphasis on propensity will lead to the identification of different people than if the focus remains only on observed behavior. The main reason for the difference lies in the fact that not all ‘observed desistors’ have a propensity which is substantively different from zero at the end of the observation period.

Because we are dealing with predicted trajectories of offending probability, we follow the recent developments in the literature and use predicted trajectories obtained from ITM, GCM and GTM models to identify desistors based on their latent propensity. We define desistors as those individuals who have a period where their latent/predicted conviction probabilities are distinguishable from zero followed by at least 5 years when these probabilities are indistinguishable from zero at the end of their career.

Bushway et al. (2001) provide instructions to identify desistors based on the predicted conviction probabilities for each person in all years, estimated from panel models. These instructions are a set of rules in three steps. The first step is to identify a period of stability at the end of each individual’s predicted trajectory. We identify a period of stability as five or more years where the probability of offending differs by no more than .1.¹⁷ Provided a period of stability is observed, the second step is to determine if the probability of offending during this period is discernable from zero. Using the binomial distribution, the probability of observing zero convictions can be calculated using the average level of offending during this period as p , and the length of the stable period as N . People are eligible to be considered desistors if the calculated probability of observing zero convictions during this stable period is above .1.

The third and last step in determining if the person can be identified as a desistor, is to examine a 5–10 year period immediately prior the stable period of non-offending, and to test whether the person was convicted prior to desisting.¹⁸ As in the second step, we again use the binomial distribution with N equal to the number of observed years. However, now we use p equal to average offending during the three peak years during this time period. This provides a less restrictive definition of an offender. If the probability of observing zero

¹⁷ Because the functional form of these models sometimes produces upticks at the end of individual trajectories, we allow this period of stability to end 5 years before the last observed year. Note, in cases where no convictions are observed, this period of stability will extend the length of the observed life-course. These individuals are identified in the steps below.

¹⁸ When possible, this time period is equal in length to the stable non-offending period, but can be shorter if the stable period is longer than 10 years, or if the stable period begins very early in the life-course.

convictions given these parameters is below .1, then the individual is considered an offender during this time period.

To summarize, we operationalize desistors as those individuals who have a period where their latent propensities to offend are distinguishable from zero followed by at least 5 years when their latent propensities are indistinguishable from zero.

Model Selection and Comparison

Model Selection

Figure 2 compares the average raw data for the sample with the average predicted conviction probabilities for the ITM. On average, the ITM fits the raw data reasonably well. The peak in the average ITM is about 2 years later than the peak year for the raw data. This slight misspecification is caused by the limits of the cubic model. A more flexible model would generate an even closer fit. However, most trajectories in the literature are estimated using quadratic or cubic functional form. This reflects an attempt by researchers to balance a desire to fit the data accurately with desire to provide a “clean” summary description of the individual paths of behavior. In this paper, we decide to stay with the cubic logit model for all models.

For the GCM, we estimated parameters with random components on each of the fixed-parameters. The parameter estimates of this model are displayed in the Appendix. Of course, we could have tested whether the coefficients and random effects were significantly different from zero, and only included those random effects in the final model which were statistically significant. However, we are interested in fitting the most general models so as to provide the best possible estimates of the individual specific curves. Or to put it another

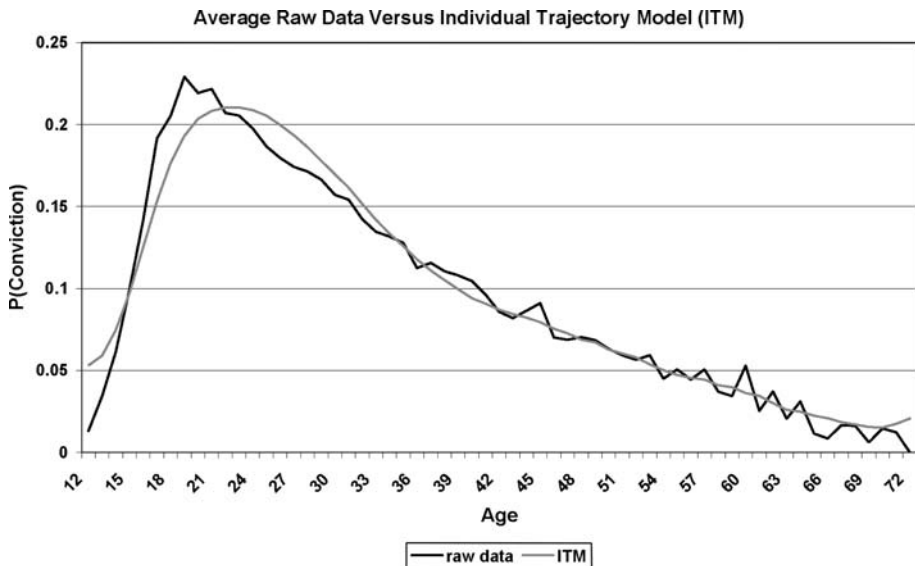


Fig. 2 Average raw data versus individual trajectory model (ITM)

way, we are interested in prediction rather than inference on the individual coefficients. As a result, the significance of the individual coefficients is irrelevant.

Using the parameter estimates, Bayesian methods are then used to actually estimate effect-parameters for each individual on each parameter. These effect-parameters can be used to generate a separate predicted trajectory for each individual. The most important statistic for this model is the reliability of each coefficient, which tells us the relative weight given to the group mean versus the ITM by each parameter. The average reliability of the intercept is .678, of the linear term is .314, of the quadratic term is .143 and of the cubic term is .080. This means that the individual estimate of the intercept from the GCM will be, on average, 32.2% of the group average and 67.8% of the individual estimate. In contrast, the cubic term will be dominated by the group average with only 8% of the value coming from the ITM. These reliabilities are very informative, because they tell us from the beginning that while the description of levels of conviction probabilities will vary across individuals, the story of change in convictions probabilities over time, which is described by the quadratic and cubic terms, will be driven primarily by the average change in the sample, and not the individual changes as described by the ITM.¹⁹

For the GTM, before we could estimate predicted trajectories, we had to decide on the number of groups, such that the model gives a good representation of the data. We used Nagin's advice (Nagin 2005) to choose a model that seemed to capture the most variation in the data with the most parsimonious number of groups. In our case this turns out to be a seven group model. The average maximum group membership probabilities are all above .80 for the 1 through 7 group models, but drop below .80 for the eight group model.²⁰ Adding more groups yields lower BIC scores, but does not add substantively different groups. In addition, it becomes difficult to distinguish between groups as we move beyond seven groups.²¹

Next, using Bayesian methods, posterior probability of membership in each group for each individual are generated for the GTM model. These posterior probabilities then are used to generate an individualized trajectory for each person. The individual curves are estimated as weighted averages of each of these groups. The graph for average predicted curves for the seven groups of the GTM is presented in Fig. 3.²² Note that not all curves are graphed through age 72, as we stopped the graph where the number of estimated cases dropped below 20.

It is important to point out that only one of the seven groups reaches past a .8 probability of conviction at any age. Group 6 peaks at a conviction probability of .818 at age 28. Because individual curves are weighted averages of these seven group average curves, no

¹⁹ It is also important to note that these reliabilities are highly sensitive to the distributional assumptions of the GCM, and therefore should not be taken as "factual" estimates of the overall reliability of the data. To the extent to which the normal distribution does not adequately capture the distribution of growth parameters—for example, if the distribution of the cubic terms is highly skewed—then we will incorrectly conclude that the change is driven primarily by noise. For example, we were able to dramatically increase the reliability of the estimates of the quadratic and cubic terms by adding an over-dispersion parameter to the logit model in GCM (see Skronidal and Rabe-Hesketh, 2007).

²⁰ To be clear, we are referring to the average group membership probabilities for those people who "belonged" to a given group X—meaning that their highest posterior probability was for group X.

²¹ Much like a regression model where R^2 continues to improve as more control variables are included, typically, the accuracy of the individual predicted curves derived from a GTM model will increase as the number of groups increases. Readers should keep this in mind, especially when comparing the GTM to the GCM model.

²² We classified individuals in the group with their highest posterior probability (see also footnote 20).

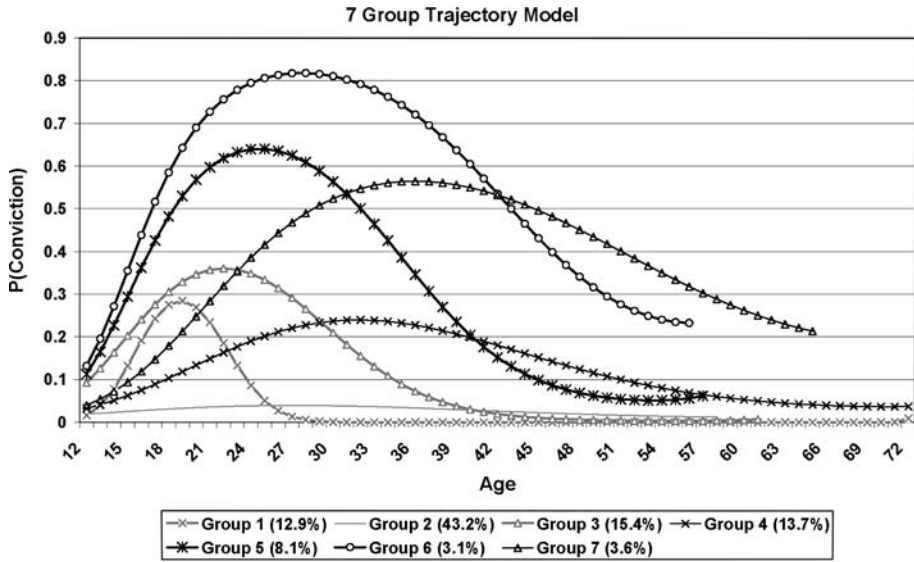


Fig. 3 7 Group trajectory model

individual estimated curve from the GTM will exceed .818. Also, this peak can only come at age 28, the peak for group 6. Likewise, no individual curve from this model will drop below .04 at age 28, which is the estimated value of the group with the lowest predicted conviction probability at this age (Group 1). The basic point is that the maximum and minimum group at each age poses an a priori bound on individual curves derived from this model. Since the individual trajectory is a weighted average of the separate trajectories, no individual curve can have conviction probabilities at a certain age that are higher than the highest curve or lower than the lowest curve.

Model Comparisons

The basic model fit statistics are provided in Table 1. We have 4,615 people with on average 40 years per person. Our final dataset has 184,078 observations. Table 1 lists the number of parameters, and the log likelihood from each model. The ITM model uses 18,460 parameters, and therefore it is not surprising that the log likelihood for the ITM is substantially better (almost 20,000 log likelihood points) than the log likelihood for the other models.²³ Or to put it another way, the ITM log likelihood is 34% bigger than the next biggest log likelihood. Given this flexibility, it is reasonable to assume that the ITM model will generate the best unbiased estimate of the individual trajectories under the assumption that latent trajectories exist, follow a cubic polynomial, and map to behavior in a manner captured by the logistic model. That the log likelihood is not zero means that the ITM does not capture all of the variation in the data—the cubic trajectories are abstractions

²³ This log likelihood comes from adding up the log likelihood from each of the 4,615 models. Each model has 4 parameters.

Table 1 Model fit statistics and trajectory comparisons

Model $N = 4,615$, NT = 184,078	Model fit statistics				Trajectory comparisons	
	# Parameters	Log likelihood	BIC	AIC	Mean SDF (SD)	Mean ADF (SD)
Individual trajectory model (ITM)	18,460	-36,554.7	228,857.7	110,029.5	Ref.	Ref.
Growth curve model (GCM)	14	-55,520.0	111,158.2	111,068.1	-0.0008 (.105)	0.065 (.084)
7 Group trajectory model (GTM)	34	-55,604.6	111,621.4	111,277.2	-0.0008 (.115)	0.071 (.093)

Bold indicates the model winner by that statistic

that summarize the behavior over the life-course. Given the far fewer parameters used by the other models (only 14 in the case of the cubic GCM and 34 in the cubic GTM), they will, however, be much more efficient. The goal of this paper is to understand the tradeoffs we make when we use the GCM and GTM models, and explore some of the implications of our reliance on these models.

While generic model fit is not an explicit focus of our paper, it is nonetheless interesting to compare the models. The ITM model loses in dramatic fashion to the GCM and the GTM models when we calculate the BIC scores, which penalizes models on the basis of the number of parameters. However, using the AIC, which penalizes models less severely for the number of parameters, the ITM model is the preferred model.²⁴ The added flexibility of the model captures real variation in the underlying data.

Between the GCM and GTM, the GCM wins according to all three measures of model fit. This suggests that the GCM is able to parsimoniously and accurately capture the individual variation across time better than the GTM. This result differs from that found by Kreuter and Muthén (2008).²⁵

Results

Average Predicted Trajectories and Overall Error

The first aim of this paper is to examine the quality of the predicted *average* trajectories. Therefore, the average trajectories for the ITM, the GCM and the GTM models are all presented in Fig. 4. These trajectories represent the average of all of the individual curves from each of the models. In other words, we first predicted the conviction probabilities for each individual for each year, and then averaged the predicted values to create the picture

²⁴ We use the standard formulas for BIC and AIC:

$$\text{BIC} = -2 * \text{LL} + k * \ln(N)$$

$$\text{AIC} = -2 * \text{LL} + 2 * k$$

where LL is the log likelihood, k is the number of parameters used, and N is the sample size.

²⁵ As a comparison, we also estimated the ten-group GTM, which does in fact have better model fit statistics than the GCM. The GTM10 uses 49 parameters to achieve this and is not the model which we chose on the basis of criteria discussed by Nagin (2005). It is, however, the preferred model when using BIC scores and, may therefore be the right model for use when using BIC scores to make model comparisons.

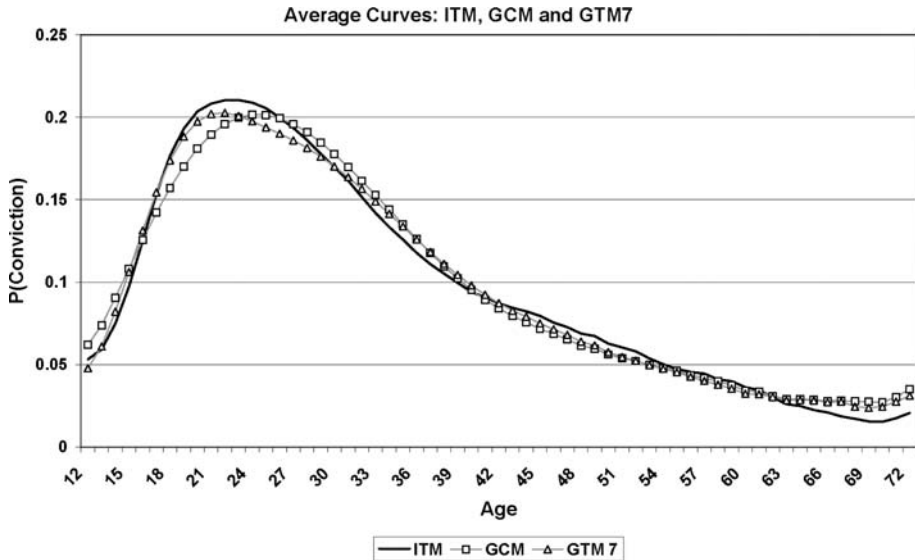


Fig. 4 Average curves: ITM, GCM and GTM7

in Fig. 4. Perhaps the most important takeaway from Fig. 4 is that the models all give very similar average curves, indicating that they do reasonably well at describing the average change over the life-course.

In order to perform this comparison more closely, we calculate the signed difference or SDF by subtracting the ITM’s predicted probability from the panel model’s predicted probability for each and every person-year. A positive SDF indicates that the panel model is overestimating the curve for that person-year (relative to the ITM), while a negative SDF indicated that the panel model is underestimating the curve. An SDF of zero occurs when the two models generate an identical estimate. The average SDF of a model is thus a measure of bias.

Table 1 presents the average results for the full sample of all 184,078 person-years. When examining the mean SDF, the first finding is striking: the average distance between the predicted probabilities from the panel models and the ITM are identical and very small—less than .08 of a percentage point. In other words, on average both the GCM and GTM panel models precisely estimate the ITM trajectories.

Predicted Individual Trajectories

The second aim of this paper is to examine the quality of the estimated *individual* trajectories. Although, on average both panel models precisely estimate the ITM trajectories, the question can be raised whether the models produce adequate predictions for each individual and for each year.

A first answer to this question comes from the standard deviation of the SDF (see Table 1). The standard deviation is large, ranging from 10.5 percentage points for the GCM to 11.5 percentage points for the GTM. This suggests that while on average the estimates are unbiased, there is substantial variation in the degree to which the individual probability in any given person-year is accurately captured by the panel models.

A second answer to this question—that is in line with the first answer—comes from the absolute difference or ADF, which gives us a sense of how far off on *average* an estimate from the GCM or GTM is from the ITM estimate that we are using as the baseline. This average ADF is thus a measure of precision. The mean ADFs (also presented in Table 1), are the average difference between the predicted numbers from the ITM and the GCM/GTMs without regard to the sign of the difference. The GCM is off on average by 6.5 percentage points, while the GTM is off by 7.1 percentage points. Moreover, these also have wide standard deviations, suggesting that mispredictions larger than 10 percentage points on either side of the ITM estimate are not unusual.

This variation in the estimated conviction probabilities is captured by Fig. 5a and b, which provide SDF histograms for the GCM and GTM, respectively. The distributions of all 184,078 person-years in our data are presented. For the GCM (Fig. 5a), the good news is that for over 60% of the person-year observations, the bias is less than 5 percentage points, meaning that the GCM either over or underpredicts the predicted probability of offending (from the ITM) by no more than 5 percentage points. Somewhat more troubling is the fact that the GCM overestimates the predicted probability of offending by 5–15 percentage points for another 20% of the person-year values. Most of these are years where the probability of offending is either zero or very close to zero but the model predicts a non-zero probability of offending. It is also somewhat troubling that 17% of the time the GCM underestimates the probability of offending by at least 5 percentage points, and 7% of the time, the probability of offending is underestimated by at least 15 percentage points. These underestimates are most likely to occur for periods where the ITM predicted probability of offending is relatively high.

As the average ADFs already indicated (see Table 1), the story is quite similar for the GTM (see Fig. 5b). Although for about 63% of the person-year observations, the bias is less than 5 percentage points, somewhat troubling is the fact that the GTM overestimates the predicted probability of offending by 5–15 percentage points for almost 16% of the person year values. So, both GTM and GCM flatten out the curves, deemphasizing change.

In order to provide some intuition about the nature of these predictions, we present data from two selected individuals. These individuals were selected because they highlight the strengths and weaknesses of the different approaches. Figure 6a shows the predicted curves of conviction probabilities created using each of the three methods for the first person selected. This first individual has a simple pattern of involvement, with convictions at age 18, 20, 26, 35, and 42. The ITM predicts a curve that starts at a probability of offending of .07 at age 12, peaks at .20 at age 26, and slowly declines to a probability of .02 by age 52 when the person exits our sample. The GCM curve closely matches the ITM curve, underestimating the ITM prediction by less than 1 percentage point in each of the 40 person-years in this example. The GTM curve is also fairly close to the ITM curve, although it overestimates it by about 1 percentage point over the course of this person's life. The ADF's are of similar magnitude, indicating as in the picture that the predictions are fairly close on average. All in all, this is an example where both GCM and GTM panel models capture the individual trajectory described by the ITM quite well.

In contrast, Fig. 6b provides the case of a person where the panel models do not do a very good job of capturing the predicted curves from the ITM. In this case, the person is convicted at ages 13, 40, 42, 44 and 49. As described by the ITM, the probability of conviction starts out at nearly 50% with a quick decline to zero by age 18, followed by a late peak of 40% probability of conviction at age 44. Neither the GCM nor the GTM captures the two-peaked nature of the ITM. The curves, which borrow heavily from the average trends, do not have the flexibility to capture the kind of intermittent offending

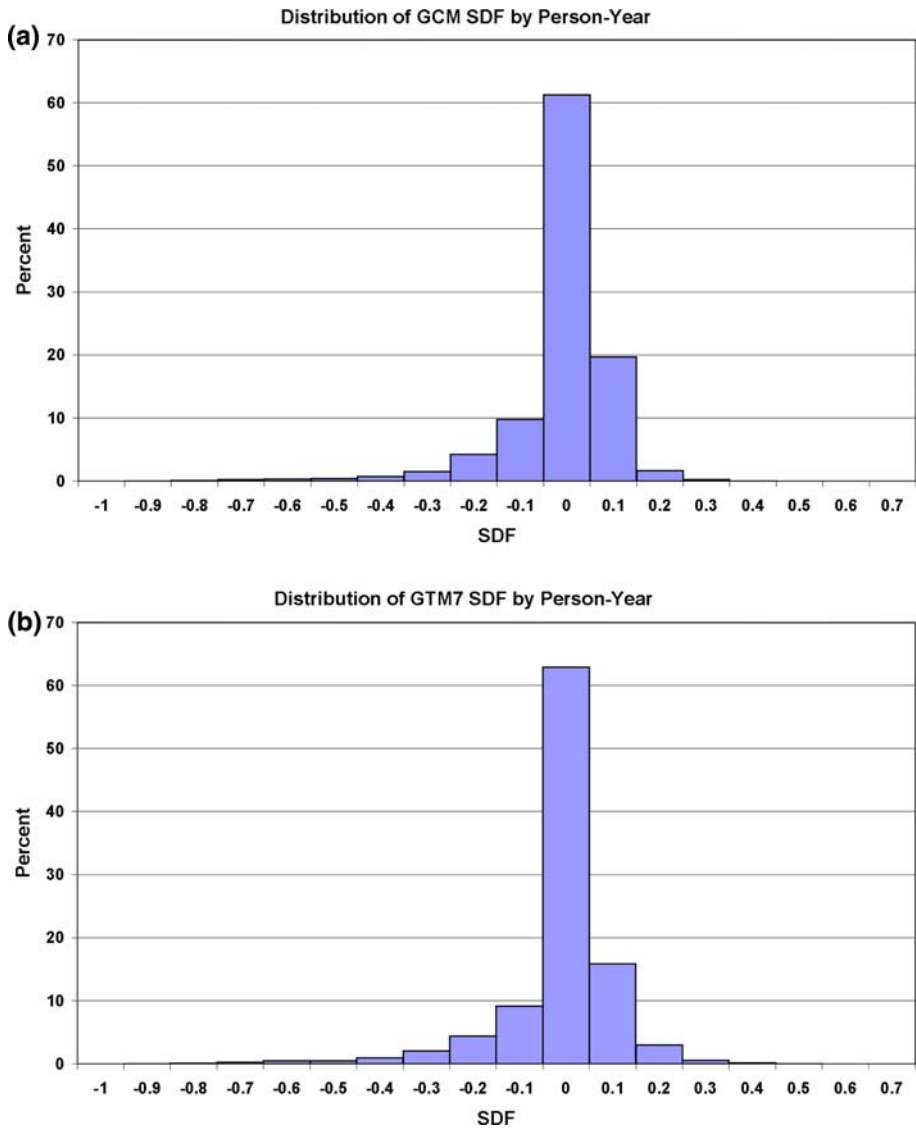


Fig. 5 (a) Distribution of GCM SDF by person-year. (b) Distribution of GTM7 SDF by person-year

described by the ITM model. However, the SDF measures of bias suggest that the models do quite well on average. The GCM, on average, underestimates the ITM by 2.7 percentage points every year and the GTM overestimates the ITM by less than a percentage point every year. This result masks the fact that the models both over and underestimate the ITM rather grossly at different points of time. This is captured more accurately by the average ADF, which is 11 percentage points for the GCM model and 14 percentage points for the GTM. In other words, on average the GCM and GTM predictions miss the ITM by 11 and 14 percentage points in any given year, respectively.

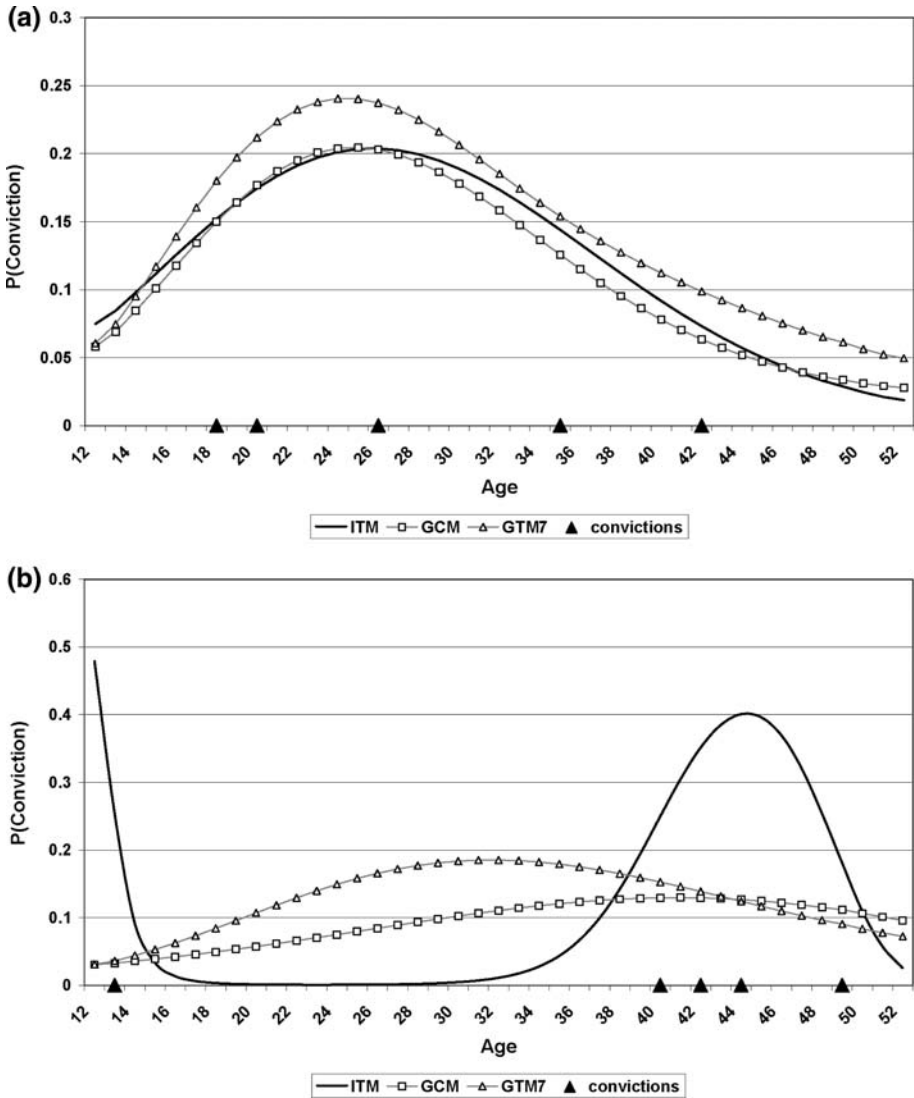


Fig. 6 (a) Example of a well-approximated individual curve. (b) Example of a poorly approximated individual curve

It is important to acknowledge that the ITM itself is also an estimate, and therefore, it is possible that the GCM and GTM are closer to the truth than the ITM. However, by definition, the individual trajectories of the GCM and GTM are biased towards the group-wide average. As a result, we find the divergence of these models from the point estimate of the ITM to be informative about the extent of the possible errors, particularly for atypical cases such as this one, which obviously diverge from the group or average pattern in the data.

Identifying Desistors

The third aim of this paper is to examine how the different assumptions of the models lead to different conclusions in criminological research—especially on desistance from crime. The question becomes how the different models identify people desisting from crime (desistors), and whether there are substantial differences in who gets identified as a desistor. This difference will be informative about the strengths and weaknesses of the different models, and shed some light on the substantive question of how to study long term patterns of behavior such as desistance.

We define desistors as those individuals who have a period where their latent/predicted conviction probabilities are distinguishable from zero followed by at least 5 years when these probabilities are indistinguishable from zero at the end of their career. In the “[Method](#)” section we discussed at length the three steps we used to identify desistors: (1) identify a period of stability at the end of the career; (2) determine whether the probability of offending during that period is not discernable from zero; and (3) determine whether the probability of offending is discernable from zero in the 5–10 years prior the non-active period.

We applied these three steps to our data of the individual predicted trajectories derived from ITM, GCM, and GTM models. Results are reported in [Table 2](#). The models provide quite different pictures of the proportion of the sample that are considered desistors. Using the ITM, we identify 60.8% of the sample as desistors, but the GTM only identifies 36.4% and the GCM only 27.5% as desistors. So, the ITM, which is the most flexible of the trajectories, has the highest percentage of people who are first high enough to be considered an offender before descending to a level low enough to be considered a desistor. The GTM is the next most flexible method a priori, given that it provides a method that is a weighted average of different group trajectories, identifies 40% fewer people as desistors than the ITM. The biggest difference is not in the number of people who are identified as having a period of stable non-offending in the latter part of their career, but in the number of people who are identified as having a period of stable offending in the 10 years prior to that period. Clearly, the differences arise primarily from Step 3 in our procedure to define individuals as persistors.

The GCM, which found that the reliability of the quadratic and cubic terms was very low, does not present a very “curvy” picture for the individual trajectories—but rather presents a picture of change at the individual level that largely mirrors that for the entire sample. This means that the GCM finds both a higher percentage of life course persistors—people who never descend to a level close to zero, and a higher percentage of never offenders, people whose predicted offending propensity never moves appreciably away from zero. The net result is that the GCM identifies 55% fewer desistors than the ITM, and 24% fewer than the GTM.

Table 2 Identification of desistors (% of 4,615)

Step	ITM (%)	GCM (%)	GTM (%)
(1) Stable period identified	94.5	96.9	97.5
(2) Non-offender during stable period	87.7	82.0	85.6
(3a) At least 5 years observed prior to stable period	70.5	49.6	45.7
(3b) Desistor	60.8	27.5	36.4

Clearly, if one wants to study desistors, the method matters. Each model will identify different (number of) people as persistors and desistors, and therefore, it is reasonable to conclude that we will reach different conclusions about the causes of desistance.

We like to stress though that this exercise does not necessarily tell us which method is “right”. Rather, the exercise highlights the problems that arise when working with point estimates of offending propensity, which can only be estimated with substantial error. The GCM chooses to be very conservative when dealing with what it perceives to be error around the coefficients of change, and the result is estimates of individual trajectories which rely heavily on the average path traveled by the members of the sample. This will necessarily affect who will be identified as desistors. In contrast, the GTM is less conservative, because it uses a weighted average of the trajectories across the groups, rather than the overall sample mean. However, the GTM also does not have as much flexibility at the ITM, again because it discounts some of the variation at the individual level as noise. As we add more groups, we add flexibility, and identify more desistors, but also lose efficiency. The ITM identifies the most desistors, but the confidence errors around the point estimates are huge. In fact, of the 3,971 individuals for whom individual trajectories could be estimated (1 or more conviction year), 84% of the individual trajectories from the GCM and 71% of the individual trajectories from the GTM fall within the 95 percent confidence intervals of the ITM.

Conclusion

In this paper, we presented and compared individual trajectories estimated from two different panel models, i.e. growth curve models and group trajectory models. The exercise was noteworthy on its face because it represents a novel approach to understanding the differences between these two models. Of course, for some life course research, the individual trajectories are largely uninteresting. For example, if we care about the overall age-crime curve, the description of the individual curves is not useful or important. And indeed, our analysis shows that each method does a very similar job in describing the average curve.

Our results, however, are informative with respect to recent attempts to use these methods to describe individual distributions of trajectories. For example, our comparison shows that both GCM and GTM will do a poor job of capturing the trajectories of individuals who offend late in the life-course. Moreover, our analyses show that the different methods result in different numbers that will be identified as desistors of crime. The biggest difference between the approaches is not in the number of people who are identified as having a period of stable non-offending in the latter part of their career, but in the number of people who are identified as having a period of stable offending in the 10 years prior to that period.

Our results suggest that some reflection on earlier research on criminal trajectories, including our own, might be in order. First, our comparison shows that both GCM and GTM will do a poor job of capturing the trajectories of individuals who offend late in the life-course. Yet, these patterns are exactly the trajectories which Sampson and Laub (2003) and Blokland et al. (2005) were trying to identify using GTM in their papers on life course persistors. More generally GTM (and GCM if used this way) will generally miss the types of cases that do not follow the general trend, a fact which might account for the very few papers that find late starters (Piquero 2008; Eggleston and Laub 2002; Bushway et al. 2003; van der Geest et al. 2009). Both the GTM and GCM methods simply do a poor job of

identifying paths that do not follow the basic trend, since the methods use the average trend to compensate for “instability” in the individual trajectories.

Thus, the use of GCM and GTM may lead to the conclusion that latent traits develop more smoothly than they actually do in practice, a finding which might cast doubt about the conclusions that are sometimes reached about the nature of the distributions from standard panel models. For example, consider the important discussion of the time-varying nature of self-control discussed by Hay and Forrest (2006). They estimate a GTM for measures of self control, and find 8 groups with largely (but not completely) parallel paths. On the basis of this finding, they conclude that there is a “strong element of truth to Gottfredson and Hirschi’s claim of stability, but that stability in self-control is not the rule for everyone (Hay and Forrest 2006, p. 766).” Our results raise questions about the strength of that conclusion. We have shown that the groups themselves do not capture the full nature of the individual distribution of trajectories captured by the ITM. A weighted average of the curves presented by Hay and Forrest which takes the posterior probabilities of assignment into account might show considerably more crossing than anticipated in their analysis, especially if the small changing groups have non-trivial weight for many people in the sample.

But even if Hay and Forrest had focused on the individual trajectories estimated by the GTM, we have shown that the GTM is predisposed to flatten the individual curves because of its reliance on the average path to generate estimates for the individual. This means that the GTM could overestimate the stability in the sample, and could potentially predispose the researcher to conclude there is more stability in self-control than there truly is. In fact, researchers apply the GTM and GCM models to their observational data in particular since they are interested in long term and enduring changes and because they realize their observational data is measured with error. So, researchers are explicitly trying to “smooth out” the large share of within individual change. We feel that the use of Bayesian techniques to identify the individual trajectories makes this “smoothing out” function both more obvious and transparent. Future research that attempts to compare the distribution of parameters of the ITM models with the parameters estimated from the GCM models would shed light on both the reasonableness of the assumptions of the GCM, and the tradeoffs that are made between the group and the individual in the GCM data.

Second, our example with the desistors identifies a problem that arises when researchers want to talk about point estimates of individual propensities over time, as advocated by Bushway et al. (2001, 2003). Since not everyone belongs to the group with equal probability, the curve of the desistor group does not fully describe the path of each and every individual. And, either panel model will tend to ascribe more stability to the individual curve than a more individualized estimate. This may lead to the inclusion of people in the desistor group whose activity seems to defy a common sense definition of a desistor. However, the long term variation described by the path of the so-called desistor group is still valid variation in the data that we as researchers might want to explain.

A related issue is that the ITM model identified nearly twice as many desistors as the GTM and GCM models using the same stochastic definition of desistors. This difference is driven by the need in the definition to be able to observe a change from active offending to a period of stable non-offending. Using GCM and GTM, we find roughly 30% fewer of the convicted offenders have ever had a stable period of non-zero offending propensity than when we use ITM.

From a policy perspective, these models suggest that a substantial proportion of the sample of convicted offenders in The Netherlands in 1977 are offenders in name

only—they are stochastically not different from zero, and therefore any treatment or effort to get them to change is not really necessary. Looking at the correlates of desistance for this subsample would be a waste of time, since there is no real change to explain. Therefore, the GCM and GTM focus the policymakers' attention on the smaller 30% of the whole sample who appear to experience real change.

In contrast, the ITM includes much of the variation which the GTM and GCM essentially eliminate as unreliable noise. Therefore, the ITM will identify the GTM and GCM “non-offenders” as people who do indeed experience real change from non-zero offending probability to something approaching zero. Therefore, the ITM says that studying these people whom the GCM and GTM categorize as non-changers will be worthwhile. Moreover, from a supervision perspective, the ITM approach says that more of these people are indeed “risky” at the time of their conviction, and therefore are susceptible to or in need of intervention to become desistors. The GCM and GTM perspective says that a fairly large number of people are largely innocuous, with their “true” offending propensity consistent with very low levels of offending propensity.

While the ITM is undoubtedly the best unbiased model of the individual trajectories, the confidence intervals around these individual models are very large. The GCM and GTM models attempt to deal with the error by producing more efficient estimates, in effect trading off bias for efficiency. Our results show clearly that both approaches do about equally well on average, although both models deviate quite a bit from the predicted paths from the ITM. That analysis also brought real meaning to the low reliability estimates for the quadratic and cubic terms on the GCM. Osgood (2005) is right—there is a lot of variation in the data that is not captured in either the GCM or GTM models.²⁶ This is both a strength and a weakness, depending on the question being asked.

This message is both important and sobering for researchers interested in studying the type of long term change captured by the logic of trajectories—the reality is that we can only estimate our dependent variable with a great deal of error. Attempts to estimate this variable will necessarily be highly dependent on the modeling assumptions, which should be made explicit by the researcher. The GCM model's reliability measures provide a useful warning light about the potential for noise to affect conclusions. Unfortunately, GTM does not provide such a warning signal. On the other hand, the warning signal in GCM is highly dependent on the correct modeling of the functional form of the variation. While we are aware of the possibility of many different functional forms for this variation (Raudenbush 2005), we are aware of little empirical work that explores the correlation between different distributional forms of the random effect and the reliability measures for the change coefficients.

Epilogue

A reasonable person could argue that our own models were not flexible enough. For example, our imposition of the cubic logistic model artificially constrained the paths to

²⁶ Consider for example the log likelihoods reported in Table 2. The total log likelihood to be explained in our dataset was $-69,450.9$. The ITM LL was $-36,554.7$, which means that the ITM captures more variation which could then be explained by covariates. In contrast the GCM LL was $-55,520$, which means that the individual trajectories measured by the GCM model have captured 42% of the variation captured by the ITM models. The GTM model, with a log likelihood of $-55,604.6$, also captures 42% of this variation.

look smooth when they are not, thereby eliminating interesting variation that still may be explainable. While we have no fundamental objection to such an argument in principle, we do think that such an argument taken to the extreme can lead to problems. At some point, we believe the researcher needs to assert the existence of the individual trajectory with the specified functional form as a summary statement of an individual's underlying propensity to offend. A researcher who is unwilling to do this is left to study offending events (not rates). Such a “rely on the data” approach is the extreme on the continuum of individual trajectories. The ITM, which attempts to smooth out the radical shifts suggested by the data, is the first move away from a model that assumes that an offender is either fully an offender or fully a non-offender.

The biggest drawback to the “no propensities” approach is that this approach homogenizes all offenders and non-offenders in a manner that is antithetical to much criminological research. We know not all offenders are equally “criminal”. Although prediction is not easy, information about past offending, for example, is a well-accepted predictor of future offending (Gendreau et al. 1996). The reverse is also true—not all “non-offenders” are equal, and in fact individuals with longer spells of non-offending much more closely resemble non-offenders than people with only a short spell of non-offending (Kurlychek et al. 2006). As such, we find the “solution” of ignoring the stochastic nature of the data to be unsatisfactory. We think it both more responsible and interesting to acknowledge the uncertainty, and use the models of latent propensity with a strong awareness of their limitations.

These limitations should lead empirical life-course researchers to be a bit more humble about the ability of any one model to explain the data. In this context, we think the debate about the “correct” trajectory model, pitting GTM against GCM, is largely pointless. Both have relative strengths when trying to pursue life-course research and we think this paper demonstrated that both methods could provide useful insights into the life-course, particularly about the average change over time. In fact, in this paper we have shown how GCM can be used in ways for which it has yet to be used in the literature. However, we have concerns about the ability of either approach to approximate the full complexity of the individual trajectories.

As a result, we believe that researchers need to spend more time considering the individual trajectories of offending in future studies. These studies will then also shed more light on the discussion of whether the identification of groups (regardless of methodology and models) helps us understand the longitudinal patterning of criminal behavior. Alternatively, researchers can focus on identifying a “latent trait” which is common to many offenders (including groups), but we would encourage more substantive discussion about the vast amount of “residual” variation that is left over after this trait is identified. This would include, in our opinion, more attention to the underlying nature of the model that identifies much of the individual variation as “unreliable”.

Acknowledgments The authors would like to thank Frauke Kreuter, Bobby Brame, Arjan Blokland, Raymond Paternoster, three anonymous reviewers, the editors and seminar participants at the American Society for Criminology 2007 meetings in Atlanta and the 2008 Analysis of Criminal Career Data Conference in London, UK for helpful comments. We are particularly indebted to Daniel Nagin and Wayne Osgood who both spent considerable effort to help us understand the specific characteristics and (dis)advantages of the models discussed in the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

See Table 3.

Table 3 Parameter estimates

	Coefficient (SE)	Random effects (<i>p</i> value)	Reliability estimates				
<i>GCM model^a</i>							
Intercept	−1.974 (.025)	1.454 (.000)	.678				
Slope	−.659 (.019)	.846 (.000)	.314				
Quadratic	−.416 (.010)	.306 (>.50)	.143				
Cubic	.142 (.004)	.125 (.436)	.080				
Log likelihood = −55,520.0							
	Group1	Group2	Group3	Group4	Group5	Group6	Group7
<i>GTM model^b</i>							
Intercept	−29.68	−5.92	−10.49	−7.80	−11.62	−11.19	−8.68
Slope	34.42	2.33	10.41	4.91	11.71	10.97	5.97
Quadratic	−12.25	−.60	−3.32	−1.11	−3.48	−2.96	−1.25
Cubic	1.10	.04	.29	.07	.30	.24	.08
Proportion	.129	.432	.154	.137	.081	.031	.036
Log likelihood = −55,604.6							

^a In order to aid convergence of GCM models, age was converted from 12–72 to −1.8 to 4.2

^b Age in the GTM models ranges from 1.2 to 7.2 instead of 12–72. As such, without transformation, coefficients are not directly comparable between the GCM and GTM models

References

- Apel R, Bushway S, Brame R, Haviland A, Nagin D, Paternoster R (2007) Unpacking the relationship between adolescent employment and antisocial behavior: a matched samples comparison. *Criminology* 45:67–97
- Barnett A, Blumstein A, Farrington DP (1987) Probabilistic models of youthful criminal careers. *Criminology* 25:83–108
- Barnett A, Blumstein A, Farrington DP (1989) A prospective test of a criminal career model. *Criminology* 27:373–385
- Bhati AS (2007) Estimating the number of crimes averted by incapacitation: an information theoretic approach. *J Quant Criminol* 23:355–375
- Blokland AAJ, Nieuwbeerta P (2005) The effects of life circumstances on longitudinal trajectories of offending. *Criminology* 43:1203–1240
- Blokland AAJ, Nieuwbeerta P (2007) Selectively incapacitating frequent offenders: costs and benefits of various penal scenarios. *J Quant Criminol* 23:327–353
- Blokland AAJ, Nagin D, Nieuwbeerta P (2005) Life span offending trajectories of a Dutch conviction cohort. *Criminology* 43:919–954
- Blumstein A, Cohen J (1987) Characterizing criminal careers. *Science* 238:985–991
- Broidy LM, Nagin DS, Tremblay RE, Brame B, Dodge KA, Fergusson D, Horwood JL, Loeber R, Laird R, Lynam DR, Moffitt TE, Pettit GS, Vitaro F (2003) Developmental trajectories of childhood disruptive behavior disorders and adolescent delinquency: a six-site, cross-national study. *Dev Psychol* 39: 222–245

- Bushway SD, Piquero AR, Broidy LM, Cauffman E, Mazerolle P (2001) An empirical framework for studying desistance as a process. *Criminology* 39:491–516
- Bushway SD, Thornberry TP, Krohn MD (2003) Desistance as a developmental process: a comparison of static and dynamic approaches. *J Quant Criminol* 19:129–153
- Eggleston EP, Laub J (2002) The onset of adult offending: a neglected dimension of the criminal career. *J Crim Justice* 30:603–622
- Eggleston EP, Laub JH, Sampson RJ (2004) Methodological sensitivities to latent class analysis of long-term criminal trajectories. *J Quant Criminol* 20:1–26
- Farrington DP (2005) Childhood origins of antisocial behavior. *Clin Psychol Psychother* 12:177–190
- Farrington DP, Hawkins JD (1991) Predicting participation, early onset and later persistence in officially recorded offending. *Crim Behav Ment Health* 1:1–33
- Gendreau P, Little T, Goggin C (1996) A meta-analysis of the predictors of adult offender recidivism: what works! *Criminology* 34:575–607
- Gordon RA, Lahey BB, Kawai E, Loeber R, Stouthamer-Loeber M, Farrington DP (2004) Antisocial behavior and youth gang membership: selection and socialization. *Criminology* 42:55–88
- Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford
- Hagan J, Palloni A (1988) Crimes as social events in the life course: reconceiving a criminological controversy. *Criminology* 26:87–100
- Haviland AM, Nagin DS (2005) Causal inferences with group based trajectory models. *Psychometrika* 70:1–22
- Hay C, Forrest W (2006) The development of self-control: examining self-control theory's stability thesis. *Criminology* 44:739–774
- Horney J, Osgood DW, Marshall IH (1995) Criminal careers in the short-term: intra-individual variability in crime and its relation to local life circumstances. *Am Sociol Rev* 60:655–673
- Kreuter F, Muthén B (2008) Analyzing criminal trajectory profiles: bridging multilevel and group-based approaches using growth mixture modeling. *J Quant Criminol* 24:1–31
- Kurlychek M, Brame R, Bushway SD (2006) Scarlet letters and recidivism: does an old criminal record predict future offending? *Criminol Public Policy* 5(3):483–504
- Lacourse E, Nagin D, Tremblay RE, Vitaro F, Claes M (2003) Developmental trajectories of boys' delinquent group membership and facilitation of violent behaviors during adolescence. *Dev Psychopathol* 15:183–197
- Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge
- Laub JH, Nagin DS, Sampson RJ (1998) Trajectories of change in criminal offending: good marriages and the desistance process. *Am Sociol Rev* 63:225–238
- Lauritsen JL (1998) The age-crime debate: assessing the limits of longitudinal self-report data. *Soc Forces* 77:127–154
- Moffitt TE (1993) Adolescence-limited and life-course persistent antisocial behavior: a developmental taxonomy. *Psychol Rev* 100:674–701
- Muthén B (2001) Latent variable mixture modeling. In: Marcoulides GA, Schumacker RE (eds) *New developments and techniques in structural equation modeling*. Lawrence Erlbaum Associates, New York, pp 1–33
- Muthén B (2007) Latent variable hybrids: overview of old and new models. In: Hancock GR, Samuelsen KM (eds) *Advances in latent variable mixture models*. Information Age Publishing, Charlotte, pp 1–24
- Nagin DS (2005) *Group-based modeling of development*. Harvard University Press, Cambridge
- Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed poisson model. *Criminology* 31:327–362
- Nagin DS, Tremblay RE (2005) What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. *Ann Am Acad Polit Soc Sci* 602:82–117
- Nagin DS, Farrington DP, Moffitt T (1995) Life course trajectories of different types of offenders. *Criminology* 33:111–139
- Nagin DS, Pagan L, Tremblay RE, Vitaro F (2003) Life course turning points: the effect of grade retention on physical aggression. *Dev Psychopathol* 15:343–361
- Nieuwebeerta P, Blokland AAJ (2003) Criminal careers of adult Dutch offenders. Codebook and documentation. NCSR, Leiden
- Nieuwebeerta P, Piquero AR (2008) Mortality rates and causes of death of convicted Dutch criminals 25 years later. *J Res Crime Delinq* 45:256–286
- Nieuwebeerta P, Blokland AAJ, Nagin DS (2009) The relationship between first imprisonment and criminal career development: a matched samples comparison. *J Quant Criminol* (forthcoming)
- Osgood DW (2005) Making sense of crime and the life course. *Ann Am Acad Polit Soc Sci* 602:196–211

- Paternoster R, Bushway S, Brame R, Apel R (2003) The effect of teenage employment on delinquency and problem behaviors. *Soc Forces* 82:297–335
- Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the life course. In: Liberman A (ed) *The long view of crime*. Springer, New York, pp 23–78
- Piquero AR, Farrington DP, Blumstein A (2003) The criminal career paradigm. *Crime Justice* 30:359–506
- Raudenbush SW (2001) Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annu Rev Psychol* 52:501–525
- Raudenbush SW (2005) How do we study “What happens next”? *Ann Am Acad Polit Soc Sci* 602:131–144
- Raudenbush SW, Bryk AS (2002) *Hierarchical linear models: applications and data analysis methods*, 2nd edn. Sage, Newbury Park
- Raudenbush SW, Chan WS (1993) Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *J Consult Clin Psychol* 61:941–951
- Sampson RJ, Laub JH (1993) *Crime in the making: pathways and turning points through life*. Harvard University Press, Cambridge
- Sampson RJ, Laub JH (2003) Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology* 41:319–339
- Sampson RJ, Laub JH (2005) A life-course view of the development of crime. *Ann Am Acad Polit Soc Sci* 602:12–45
- Skrondal A, Rabe-Hesketh S (2007) Redundant overdispersion parameters in multilevel models for categorical responses. *J Educ Behav Stat* 32:419–430
- Thornberry TP (1987) Toward an interactional theory of delinquency. *Criminology* 25:863–892
- Thornberry TP (1997) *Developmental theories of crime and delinquency*. Transaction, New Brunswick
- Thornberry TP, Krohn MD (eds) (2003) *Taking stock of delinquency: an overview of findings from contemporary longitudinal studies*. Kluwer Academic/Plenum Publishers, New York
- Thornberry TP, Bushway SD, Krohn MD, Lizotte A (2007) Accounting for behavioral change. Working paper
- van der Geest V, Blokland A, Bijleveld C (2009) Delinquent development in a sample of high-risk youth: shape, content, and predictors of delinquent trajectories from age 12 to 32. *J Res Crime Delinq* 46:111–143
- Warr M (1998) Life-course transitions and desistance from crime. *Criminology* 36:183–215