

Smoothing Group-Based Trajectory Models Through B-Splines

Brian Francis¹ · Amy Elliott¹ · Mat Weldon¹

Received: 31 August 2015 / Revised: 19 January 2016 / Accepted: 2 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Purpose This paper investigates the use of B-spline smoothers as an alternative to polynomials when estimating trajectory shape in group-based trajectory models. The use of polynomials in these models can cause undesirable curve shapes, such as uplifts at the end of the trajectory, which may not be present in the data. Moreover, polynomial curves are global, meaning that a data point at one end of the trajectory can affect the shape of the curve at the other end.

Methods We use the UK Offenders Index 1963 birth cohort to investigate the use of B-splines. The models are fitted using Latent Gold, and two information criteria (BIC and ICL-BIC) are used to estimate the number of knots of the B-spline, as well as the number of groups. A small simulation study is also presented.

Results A three-group solution was chosen. It is shown that B-splines can provide a better fit to the observed data than cubic polynomials. The offending trajectory groups correspond to the classic groups of adolescent-limited, low-rate chronic and high-rate chronic which were proposed by Moffitt. The shapes of the two chronic trajectory curves from the B-spline fitting are more consistent with the life-course persistent nature of chronic offending than the traditional cubic polynomial curves. The simulation shows improved performance of the B-spline over cubic polynomials.

✉ Brian Francis
B.Francis@Lancaster.ac.uk
Amy Elliott
A.Elliott1@Lancaster.ac.uk
Mat Weldon
M.Weldon@Lancaster.ac.uk

¹ Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

Conclusions The use of B-splines is recommended when fitting group-based trajectory models. Some software products need further development to include such facilities, and we encourage this development.

Keywords Group based trajectory model · Smoothing · Latent class growth analysis · Offenders index

Introduction

In the context of criminology, group-based trajectory models (GBTM) provide a method of examining and understanding the evolving nature of offending over the life course. Trajectory models most commonly examine frequency measures of offending over the life course through arrest or conviction data, although the method has also been used for examining other processes, such as changing offending seriousness over the criminal career [16]. In contrast to other methods, such as the growth curve model [37], such models assume that the underlying population consists of a fixed but unknown number of groups, with distinct trajectories which give the changing estimated mean values for each group over time. The modelling task then is to estimate the number of groups, and the shape of the trajectories, as well as to examine the assignment of individuals to trajectories, and the effect of covariates on trajectory membership.

Since Nagin and Land's seminal work [23], the GBTM method has become increasingly popular. It is now used extensively both in criminology and in other disciplines such as psychology. However, the method has also been subject to controversy. Thus [36], in his encyclopedia entry, identifies that the method has been subject to "unusually robust criticism". Sweeten identifies two controversies, the first concerned with the meaning of "group" and the second suggesting alternative modelling paradigms that may be better than group-based trajectory modelling. The first concern is really focused on the need not to over-interpret the groups formed by the method. Both [24] and [33] have suggested that the method may find groups which may not really exist in the data.

The second concern relating to the form of the model has been where researchers have proposed alternative methods of modelling longitudinal data, either through multilevel or hierarchical models which do not assume a group structure or through growth mixture models. Reviews of these competing approaches have been carried out by [8] and [11].

Bauer [2] gives a thoughtful discussion of these two issues. He divides the applications of GBTM into essentially three groups. *Direct* applications are where the latent trajectories are interpreted as real subgroups in the population. In contrast, in *indirect* applications, there is no interest in the group structure, and the latent classes simply provide a more flexible method of accounting for the individual variation in trajectories—a form of non-parametric random effect. For these, GBTMs may be used even when there is no underlying group structure, and present a way of accounting for non-normality in the heterogeneity. Bauer also identifies a third group which he also labels *indirect*—those applications which recognise

that the latent classes are not real but where the trajectories are still presented and interpreted.

One assumption of the model not usually discussed is that the estimated trajectories are assumed to follow a polynomial curve over age. However, the use of polynomial curves can lead to modeling difficulties. The main issues are:

- (a) Polynomial curves can sometimes produce unexpected changes in direction which tend not to be supported by the data. Various authors have estimated trajectories that show a typical pattern of increase, followed by decrease, followed by an uplift at the end of the time period that cubic polynomials often produce. We illustrate the problem by selecting two illustrative studies from the literature and shown in Fig. 1; [18] estimated juvenile trajectories for indigenous and non-indigenous juveniles, and found that there was an upward turn in their high rate trajectory at age 19 after an earlier increase and decrease before age 19. Bushway et al. [7] also model trajectories of offending behaviour from age 13 to age 22, and estimate an uplift for three of their trajectories. The two papers take different approaches to these trajectory shapes; Marshall comments in the text on the change of shape without suggesting a reason, whereas [7] also comment, but suggest such behaviour to be evidence of intermittency. It is clear that when trajectories are estimated which show a number of changes of direction, then authors are sometimes uncertain how to interpret these shapes and whether such changes in direction are real.
- (b) The fitting of a polynomial curve is “non-local”, meaning that data points in one part of the time axis can influence the shape of the curve in a distant part of the time axis. This is in general not a desirable characteristic. Thus, for example, moving averages (a simple form of smoothing) will average values only in the close neighbourhood of the data point.
- (c) Polynomials are not capable of producing curves that are arbitrarily shaped. It is not possible for a polynomial trajectory curve to be produced that rises steeply in the first part of the time axis, and then exhibits a constant rate thereafter.
- (d) Polynomials are highly correlated and can lead to numerical instability.

Although many research papers do not interpret trajectory shape in detail, some do. For example, [34] estimated pre-employment offending trajectories to check whether there were any non-desisting trajectories before job-entry. The rationale for this was that a job could only be a turning point for those who have not yet started their desistance process. The shape of the underlying trajectories clearly matters in this example. Another example is in child sexual offending, where the aggregated age-crime curve is thought to be bimodal [35, p.5]. There is considerable interest in whether such a curve is produced by offenders stopping offending in early middle age and restarting in later middle age, or whether there are two populations of child sexual offenders with different onset and desistance patterns.

There are four approaches to dealing with this problem. Firstly, higher order polynomials could be used. Sweeten [36] reports that PROC TRAJ, the SAS software addon for group-based trajectory modelling [15], allows polynomials up to order five to be fitted. While using high order polynomials may allow more flexibility in the shape of the trajectory, the method still fails to solve the problem of the non-locality

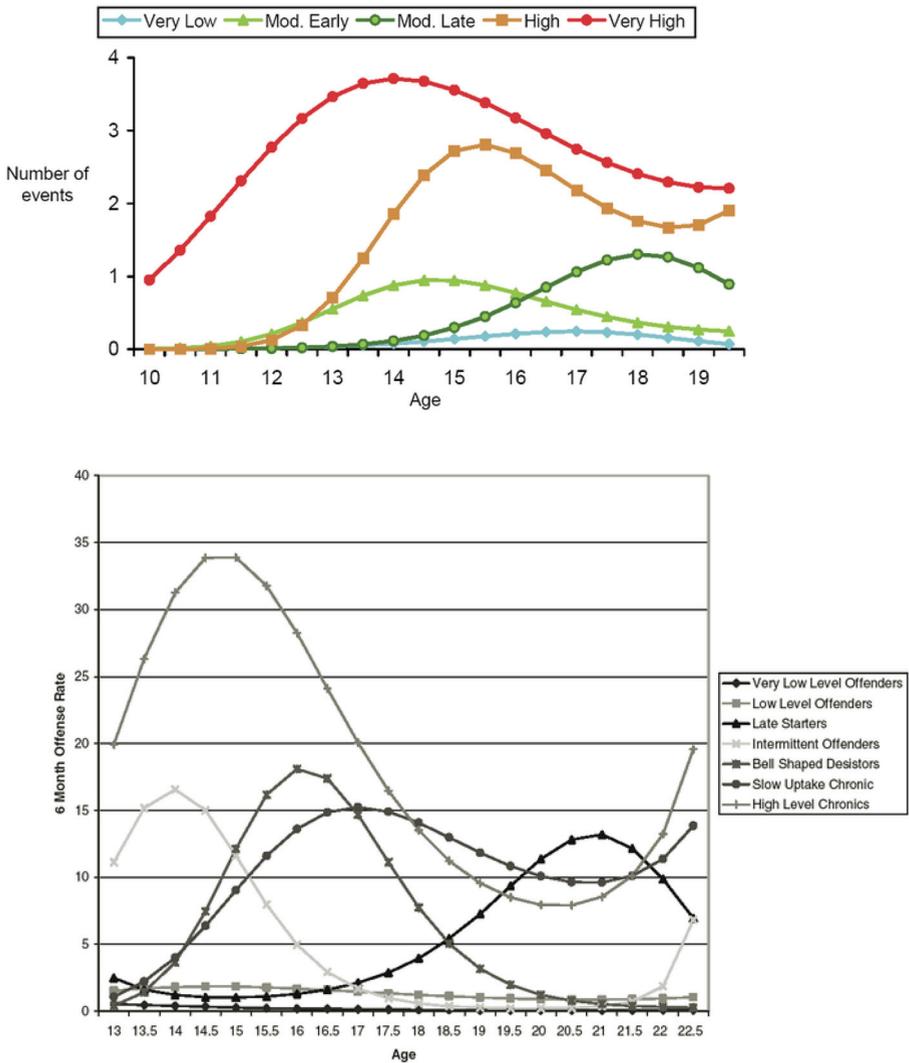


Fig. 1 Two examples of cubic polynomial trajectories from the literature showing uplifts on some trajectories at the high end of the age scale. The *top plot* shows an example from [18] with an uplift for the “high” group, and the *bottom plot* shows three offending trajectories with uplifts from [7] using the Rochester Youth Development Study. Such uplifts may be spurious

of polynomial curves, where a data point at a low age can have a large effect on the fitted curve at a high age. This in our view is an undesirable characteristic. A second solution is to add additional functions of age to the polynomial terms for each trajectory. This is the approach used by [4], who added two additional terms of the form $|((age/10) - 2)^3|$ and $|((age/10) - 3)^3|$ to the model specification of each trajectory. This approach is similar to the fractional polynomial method of [31]. The method is “ad hoc” in the sense that the choice of the constants 2 and 3 are arbitrary and

probably depend on initial examination of the data set under analysis—perhaps an undesirable modelling strategy. A third solution is to fit age as a categorical factor. In effect, this fits age as a stepwise function; the levels of the steps are constant within each time period, and jump between time periods. This method provides local fitting, but the fitting is too local as information from adjacent time periods is not used. In addition, there are typically a large number of parameters to estimate for each trajectory. In this paper, we choose a fourth solution—that of using cubic B-splines. Silverman [32] describes the idea as “a natural and flexible approach to curve estimation, which copes well whether or not the design points are equally spaced”. The flexibility of shape together with its ease of fitting makes it suitable for group-based trajectory models.

Returning to Bauer’s classification above, we recognise that most applications of GBTM want to interpret the resulting trajectories, and this paper is of primary use to the first and the third of these—the direct applications and also the indirect applications which have some interpretation. However, a more flexible form of modelling will also be of considerable utility to the second type of user as the model will account for heterogeneity more precisely. We return to this point in the discussion.

The Group-Based Trajectory Model

The Basic Model

We start by assuming that we observe a set of N offenders who are repeatedly observed over T time periods. For offender i , we observe $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT}, \dots, y_{iT})$.

We also simplify this development by assuming that the observations are counts of some criminal process such as court appearances, offenses or arrests, and that we have no missing data. Furthermore, we assume that all offenders are measured at the same time periods, so that the dataset is “balanced”. We assume that there are K latent groups or trajectories in the data, Then we can write

$$\begin{aligned}
 P(\mathbf{y}_i) &= \sum_{k=1}^K \pi(k) P(\mathbf{y}_i|k) \\
 &= \sum_k \pi(k) \prod_{t=1}^T P(y_{it}|k)
 \end{aligned}
 \tag{1}$$

where the $\pi(k)$ represent the group sizes, with $\sum_k \pi(k) = 1$. The likelihood L can then be written as

$$L = \prod_i P(\mathbf{y}_i) = \prod_i \left(\sum_k \pi(k) \prod_{t=1}^T P(y_{it}|k) \right)
 \tag{2}$$

For count data in the k th trajectory, we assume that the y_{it} are distributed according to some discrete distribution—usually Poisson or negative binomial. Thus, either

$$y_{it} \sim \text{Poisson}(\lambda_{tk}) \text{ with } P(y_{it}|k) = \frac{\lambda_{tk}^{y_{it}} e^{-\lambda_{tk}}}{Y_{it}!}$$

or

$$y_{it} \sim \text{NegBin}(\lambda_{tk}, \theta_k) \text{ with } P(y_{it}|k) = \frac{\Gamma(y_{it} + \theta_k^2)}{y_{it}! \Gamma(\theta_k^2)} \left(\frac{\theta_k}{\theta_k + \lambda_{kt}}\right)^\theta \left(\frac{\lambda_{kt}}{\theta_k + \lambda_{kt}}\right)_{it}^y \tag{3}$$

where λ_{tk} is the mean of the k th trajectory at time period t , and θ_k is a scale parameter for the k th trajectory. We reparameterise so that $\tau_k = \frac{1}{\theta_k}$ represents the overdispersion of the k th trajectory, with the variance of $y_{it}|k$ equal to $\lambda_{tk}(1 + \tau_k)$. If τ_k is zero, then we have Poisson variability for that trajectory.

It is usual to apply polynomial smoothing to the means to each of the K trajectories through a log-linear model. Let the value of the time axis in time period t be x_t . Then, for cubic polynomial fits, we have, for the k th trajectory

$$\ln(\lambda_{tk}) = \beta_{0k} + \beta_{1k}x_t + \beta_{2k}x_t^2 + \beta_{3k}x_t^3 \tag{4}$$

For example, with three latent trajectories with cubic polynomials, we need to estimate two of the three $\pi(k)$, together with either 12 β parameters for the Poisson model, or 12 β parameters and three θ parameters for the negative binomial model.

Once the parameters of the model have been estimated, then the posterior probability of the membership of trajectory membership for each offender can be estimated. This is given by

$$\hat{p}_{ik} = \hat{P}(k|y_i) = \frac{\hat{\pi}(k) \hat{P}(y_i|k)}{\sum_{k=1}^K \hat{\pi}(k) \hat{P}(y_i|k)} \tag{5}$$

where

$$\hat{P}(y_i|k) = \prod_t \hat{P}(y_{it}|k)$$

for the chosen discrete distribution. Absolute assignment to group can be carried out for any individual i by taking the group with the largest p_{ik} .

Determining the Number of Trajectories K

Various authors have suggested numerous methods for determining the number of groups or classes in the data. Nagin [22], in his book, recommends the use of BIC based on his extensive simulation work. Other authors have suggested alternative methods. For example, Nylund et al. [25] examined the performance of AIC, CAIC, BIC and adjusted BIC, as well as the bootstrap test for a wide range of latent class and growth mixture models, and concluded that the bootstrap test often works well. Aitkin et al. [1] has recently suggested a Bayesian method based on draws from the posterior deviance which is found to perform very well on small and medium samples. We have found from experience that although the BIC works well for small- and medium-sized samples, it is less good for very large samples. An alternative, recommended as the best by [19], in the context of general mixture models, is the ICL-BIC criterion. This is defined to be BIC plus twice the entropy of the model:

$$\text{ICL-BIC} = -2 \log L + \log(n)q + 2 \sum_{ik} \hat{p}_{ik} \log \hat{p}_{ik} \tag{6}$$

where q is the number of parameters in the model, and n is the number of distinct individuals. This criterion was, however, found to be too conservative in Nagin's simulation, finding fewer groups than were actually present. We will examine both BIC and ICL-BIC in this study.

Smoothing Trajectories

We seek to move from the polynomial model by using a more flexible form of smooth function. We replace Eq. 4 with

$$\ln(\lambda_{tk}) = f_k(x_t) \quad (7)$$

We are now estimating k different *smooth functions* or smoothers which change over time, one for each trajectory. Hastie and Tibshirani [13] document a wide range of different methods for estimating the smoothers $f_k(x_t)$ —these include regression splines, local loess smoothing, running line smoothers and kernel smoothers. Here, we choose regression smoothers as a smoothing method. Although many of these other smoothers could be used, regression splines offer one major advantage—they can be used straightforwardly in standard software without the need to construct purpose-written programs. Regression splines are basically a set of piecewise polynomials, usually cubic, that meet at series of knots on the x axis; there is continuity of the piecewise polynomials at these knots and also in the first and second derivatives. The greater the number of knots chosen the more flexible the resulting smoothing function will be.

Regression splines can be fitted by adding a series of terms to the design matrix for each trajectory k . If the original design matrix is defined by $[\mathbf{1}, \mathbf{X}, \mathbf{X}^2, \mathbf{X}^3]$, where $\mathbf{1}$, \mathbf{X} , \mathbf{X}^2 and \mathbf{X}^3 are column vectors containing the values 1, x_{it} , x_{it}^2 and x_{it}^3 , then we augment this matrix with extra columns \mathbf{C}_h —one for each knot h . The extra columns have elements of the form

$$c_{it}^h = \begin{cases} (x_{it} - z_h)^3 & \text{if } (x_{it} > z_h) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where z_h is the location of the h th knot. Thus, a regression spline with three knots will have three extra columns, and have seven degrees of freedom in total for each trajectory. The augmented design matrix is known as the basis of the spline.

B-Splines and Basis Functions

We choose a particular form of regression spline, the B-spline, which is a orthogonal transformation of the regression spline defined above. These have important numerical properties which limit the correlation between adjacent columns in the design matrix. While they can be relatively easily calculated through a recursion formula, the exact mathematical description is beyond this paper, and [5] gives the details. The columns of the basis can be represented graphically: Fig. 2 shows the B-spline basis based on local cubic polynomials for four knots and seven degrees of freedom over the range of x from age 10 to age 45. It can be seen that no basis extends over the entire range of x . This means that the local nature of the fit is ensured. The greater the

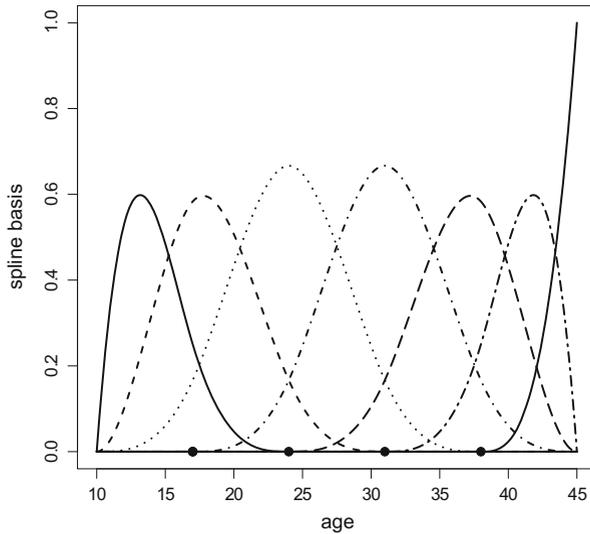


Fig. 2 B-spline basis with four knots and seven degrees of freedom, with the age axis used in the example. Knots are indicated as dots on the x-axis

number of knots specified, the more local these polynomials are and the greater the ability to follow the underlying data. Rodriguez [30] suggests that the most appropriate placement of knots is at the areas where $f(x)$ is changing more rapidly. As with all modelling, a balance needs to be struck between the goodness of fit of the curve to the data, and over fitting by specifying too many knots. It is often appropriate to examine information criterion such as the BIC and ICL-BIC, coupled with the knowledge of the subject and data, to decide the most suitable location and number of knots. A different approach to deal with the number and location of knots is to treat it as a model selection problem [3]. The aim becomes to choose the number of knots that minimises the model fit statistics, by basically treating knot selection as regressor selection. There are some methods that can assist with choosing the appropriate amount of knots. He and Ng [14] use a method called the three-step selection, which compares the change in the Akaike information criterion when adding or removing knots. Alternatively, Osborne et al. [26] use a method called LASSO (Least Absolute Shrinkage and Selection Operator) which is a regression method which involves penalising the absolute size of the regression coefficients. We take a straightforward approach and use the ICL-BIC criterion to choose the number of knots. B-splines are starting to be used in criminology. For example, [17] used B-spline smoothing to describe the general shape of the age-crime curve for males and females and compared the results with group-based trajectory modelling, without, however, incorporating spline fitting into the GBTM modeling. Our focus in this paper is to build on this work and incorporate the B-spline estimation into GBTM modelling.

When fitting B-splines, both the number of knots and the position of the knots need to be specified, as well as the order of the local polynomials. In general, the position of the knots is usually chosen by taking suitable percentiles of the x-axis.

The number of knots represents the degree of smoothing (sometimes represented by the degrees of freedom of the basis given by $df = \text{number of knots} + 3$), and can be chosen by criteria such as the BIC or ICL-BIC, or by other statistical methods such as cross-validation. It is common to use local cubic polynomials, and we do so in this paper.

Fitting the Model in Standard Software

We can use standard software to fit the B-spline group-based trajectory model (BGBTM) providing that the software has a flexible enough user interface to allow user-specified design matrices for the trajectories. In general, we recommend that a package such as R can easily be used to calculate the cubic B-spline basis using the `bs()` function in library `splines`, and the B-spline basis can then be provided as time varying covariates in the package of choice. Thus, the R-code

```
library(splines)
xx=bs(age, df=7)
```

can be used to generate the B-spline basis for age with four knots. SAS through the PROC TRAJ procedure [15] is a popular method of fitting group-based trajectory models, and allows time-varying covariates. Thus, specifying the B-splines as time-varying covariates and setting the polynomial order to zero will fit a similar model to ours (it is similar but not identical as there is still an intercept for each trajectory in the model). MPLUS in contrast does not appear to have the flexibility required. We also consider two alternative software packages for trajectory fitting—Latent Gold and R. The package `lcmm` in R [29] can fit group-based trajectory models for continuous and ordinal data, but not for count data.

In this paper, with our focus on count data, we use Latent Gold [38], which is a general package for a wide variety of latent class models. In this software, group-based trajectory models can be fitted using the Latent Class regression option, which required the data to be in long form. We use R to calculate the B-spline basis, and the variables which make up this basis are added to the dataset and become the “predictors” in the regression. The predictor effects are specified as class dependent (that is there is a separate trajectory for each class or group) through the `model` tab. Latent Gold has the facility to fit either the Poisson or the negative binomial by setting the type of the dependent variable either to “count” or to “overdispersed count”. The fitted trajectories can be examined by requesting “estimated values” on the `output` tab.

Dataset and Methods

The England and Wales Offenders Index

We will be illustrating the methodology using official conviction data from the Offenders Index (OI) database for England and Wales. The OI was originally created for researchers to statistically analyse the criminal histories of offenders. The

1963 birth cohort is one of eight birth cohorts (1953, 1958, 1963, 1968, 1973, 1978, 1983 and 1988), each representing a 4-week sample (around a one in thirteen samples) of all offenders born in a specific year. The same 4 weeks were chosen for each cohort year (these are 3–9 March, 19–25 June, 28 September–4 October and 17–23 December). Court reports submitted by the police are built upon to form the database. The data is collected from age 10 (the age of criminal responsibility) with histories followed up till 2008. The 1963 birth cohort was chosen as 1953 and 1958 cohorts have conviction histories in the 1960s, when the criminal justice system had far fewer diversions away from court in the form of cautions and warnings, and thus the changes in the justice system for the two earlier cohorts would be more likely to be confounded with changes in age. The dataset contains the court conviction history from 1973 to the end of 2008. The 1963 cohort has 14,506 offenders who, between them, have accumulated 89,750 offenses over the study period.

The data is structured in a hierarchical form with three levels—the first containing offender details (namely gender and date of birth), the second containing court conviction details for the offender (dates of convictions) and the third level providing details of each separate offense within each court conviction occasion. Not all offenses form part of the OI—with only offenses that are considered standard list offenses are included. Standard list offenses are those which are considered indictable (crown court offenses) or triable either way (either in the magistrates court or the Crown court) and the more serious summary (magistrates court) offenses. Therefore, many minor offenses are excluded.

There are a few disadvantages with the dataset which need to be mentioned. As the dataset is formed from court records, the date of offense is not recorded. As there can be a substantial delay between the offense date and court date, this can cause inaccuracies in the number of offenders committing crimes at specific ages. Data on the length of time from offense to court outcome has only recently been collected in England and Wales; yearly data for the years 2010–2014 suggests the delay is relatively constant, ranging from 151 to 159 days [20]. The dataset also contains no information on cautions, warnings or fixed penalty fines, and no record is kept of unsuccessful prosecutions. Information on death, immigration or emigration is also not recorded.

The way in which the OI database is created is through a matching process using the offenders surname, date of birth and criminal record number (if available). This is subject to matching error and can match records together which do not relate to the same person. This can be a problem with female offenders, who may change surname when they get married. Francis and Crosland [10] also found specific problems with common family names such as Smith or Singh. Due to the lengthy time period of the dataset, the results of analysis are also vulnerable to social changes over time meaning that any changes in the patterns of offending could be influenced by changing attitudes towards certain offenses. Over time, some offenses may be viewed as less serious meaning the offender may receive a lesser disposal such as a caution which may not be recorded as a conviction on the database.

Despite these caveats, the Offenders Index provides an excellent if imperfect resource for examining long-term patterns of offending, with the 1963 cohort providing criminal histories over a 35-year period. In particular, the definition of offenses

over time is extremely consistent with only a few changes to what is considered a standard list offense.

Data Restructuring and Modelling Strategy

We grouped the offense counts into age periods of 2 years (10–11,12–13 up to 44–45). Latent Gold requires the data in long form, and so the data was restructured so that there were multiple lines of data for each offender, one for each 2-year age period. Convicted offenses are naturally overdispersed, as each court appearance potentially generates a number of offenses, and the counts are clustered. One such mechanism is that offenders may seek to have other offenses taken into consideration when they are charged for an offense. This natural overdispersion led us to use the Negative Binomial form of the model. We fitted a sequence of models, from one to eight trajectory groups and with from one knot to five knots (four smoothing degrees of freedom up to eight smoothing degrees of freedom, using 100 different starting values for each model and taking the best solution. For each model, we calculated the BIC and ICL-BIC.

To assess model fit, we compared the mean observed offense counts m_{it} for each trajectory with the fitted mean trajectory counts $\hat{\lambda}_{tk}$, using the posterior probabilities of group assignment as weights:

$$m_{kt} = \sum_i \frac{p_{ik}y_{it}}{n} \tag{9}$$

This gives an indication of how well the fitted trajectories fit the data, on the assumption that the classes or groups are correctly specified.

For the final model, we formed 95 % confidence intervals for each trajectory, which were calculated as follows. If \mathbf{X}_k is the design matrix for trajectory k and $\hat{\eta}_k$ is the fitted linear predictor $\hat{\eta}_k = [\log(\hat{\lambda}_{tk})]$, then the upper (U) and lower (L) limits of the confidence interval are given by

$$[L_{95\%}(\hat{\eta}_k), U_{95\%}(\hat{\eta}_k)] = \hat{\eta}_k \pm Z_{0.975} \text{s.e.}(\hat{\eta}_k) \tag{10}$$

where $Z_{0.975}$ is the 97.5th percentile of the standard normal distribution, and $\text{s.e.}(\hat{\eta})$ is the vector of standard errors of $\hat{\eta}$. This is given by:

$$\text{s.e.}(\hat{\eta}_k) = \sqrt{(\mathbf{X}_k \text{cov}(\beta_k) \mathbf{X}_k^T)} \tag{11}$$

where $\text{cov}(\beta_k)$ is the variance-covariance matrix of β_k . The variance-covariance matrix of the parameter estimates can be extracted from Latent Gold following a model fit, and the above calculations can be carried out in R or other mathematical software.

Results

Following the above strategy, we fitted a sequence of negative binomial latent trajectory models to the Offenders Index data. Table 1 shows the BIC and ICL-BIC values from these fits.

Table 1 BIC and ICL-BIC values from negative binomial B-spline fits to the Offenders Index data. The number of knots is equal to (df-3)

	4 df	5 df	6 df	7 df	8 df
BIC values					
1 group	310,805.5	310,778.0	310,764.8	310,573.2	310,552.9
2 groups	291,040.2	290,993.6	290,969.2	290,792.1	290,738.0
3 groups	286,719.1	286,659.2	286,564.0	286,478.4	286,483.4
4 groups	283,468.7	283,439.0	283,395.7	283,326.2	283,325.0
5 groups	282,550.6	282,463.2	282,438.0	282,380.2	282,380.7
6 groups	281,886.3	281,815.3	281,782.2	281,713.9	281,707.4
7 groups	281,240.6	281,190.1	281,229.8	281,169.8	281,158.2
8 groups	280,805.5	280,792.5	280,844.2	280,827.8	280,785.8
ICL-BIC values					
1 group	310,805.5	310,778.0	310,764.8	310,573.2	310,552.9
2 groups	292,567.1	292,523.6	292,497.1	292,316.7	292,266.3
3 groups	297,404.6	297,408.2	297,251.7	297,229.5	297,280.7
4 groups	294,674.9	295,230.0	295,144.9	295,103.3	295,117.9
5 groups	295,351.4	295,235.1	295,127.6	295,086.0	295,100.5
6 groups	299,570.4	295,392.7	295,330.3	295,252.1	295,298.4
7 groups	299,086.8	298,423.3	298,358.2	298,423.3	298,576.8
8 groups	301,669.4	301,949.4	302,454.2	301,949.4	301,224.9

We can notice two features. Firstly, holding the degree of smoothing constant, and examining columns of the table, the BIC values continue to decrease from one to eight groups, and no minimum BIC is obtained. This is a common feature of trajectory models with a large number of cases. Looking at the BIC values row-wise, the BIC values suggest that between seven and eight smoothing degrees of freedom is needed depending on the number of groups. Turning now to the ICL-BIC values, and examining columns of the table, we see that the minimum occurs for two groups. We note, however, that the ICL-BIC values are volatile; the values increase for three groups, and then decline again for four groups. Looking at the ICL-BIC values row-wise, we can observe that the minimum value most often occurs for seven degrees of freedom.

It appears that both BIC and ICL-BIC have not really helped us in identifying the number of groups. The BIC values identify too many groups, as the large amount of data allows us to identify many groups. Practically, however, we consider that reporting over eight groups is too complicated. The ICL-BIC, in contrast, identifies too few groups. This chimes with [22], who also reported that ICL-BIC had this characteristic.

We therefore adopt a pragmatic solution, following [22], who commented that “model selection must balance the sometimes competing objectives of model parsimony and capturing the distinctive features of the data” (p75). Our example is

illustrative, and we therefore allow previous studies to influence our choice of the number of groups. D'Unger et al. [9] identified four trajectory groups in the UK London delinquency study, one of which was a non-offending group and three of which were offending groups. As our sample consists solely of offenders, we have no non-offenders, and so we report the three-group trajectory solution. We choose seven smoothing degrees of freedom as this minimises both BIC and ICL-BIC for three groups.

Figure 3 shows the fitted trajectories with 95 % shaded confidence intervals, for the three group model with seven smoothing degrees of freedom, together with the mean observed number of offenses by group using the weighted estimation described earlier. The fitted trajectories appear to follow the observed data well, with no undesirable “polynomial-like” features in the shapes of the curve.

In contrast, Fig. 4 shows the fitted trajectories and observed data for the three group B-spline fit. While the observed mean number of offenses in the bottom part of Fig. 4 looks very similar to the bottom plot of Fig. 3, the fitted trajectory curves are somewhat different. In particular, group 3 fails to follow the trajectory shape of the observed mean counts, dipping too rapidly towards zero.

Simulation Study

One criticism of the above development is that it has been carried out on real data that has an unknown group trajectory structure. To counter this criticism, we therefore carried out a small simulation study to illustrate the methodology on a constructed dataset with a *known* group trajectory structure. We can then examine how well the two methods (B-spline and cubic polynomial) reproduce the known trajectories, and also examine the misclassification between the known group membership and the membership predicted under the competing methods.

Our simulation assumes three trajectories, with known group size proportions of 0.5, 0.4 and 0.1. Table 4 in the [Appendix](#) shows the specified mean counts per year for the three trajectories. A negative binomial distribution was assumed for each trajectory with common shape parameter of 2.00. Furthermore, 14,000 simulated individuals were drawn by first sampling from a multinomial distribution to determine the trajectory membership, and then simulating from the trajectory count distribution for each age from 10 to 45.

Figure 5 shows the fitted three group trajectories to this data, under both the cubic trajectory and B-spline models, and taking the best solution out of 100 different start values. The specified mean counts are shown as data points. It can be seen that the B-spline model reproduces the shape of the trajectories more precisely than the cubic polynomial model. Details of the model fitting are shown in Table 2. It can be seen that, although the group sizes are nearly identical in the two models, the B-spline model has a lower BIC, and the estimates of the scale parameters are better estimated with the B-spline model. Finally, we looked at the assignment of individuals to latent trajectories (using Eq. 5) under the two models using the highest probability method, and compared this to the known membership from the simulation. Table 3 shows these results. The difference here is more dramatic. It can be seen that the

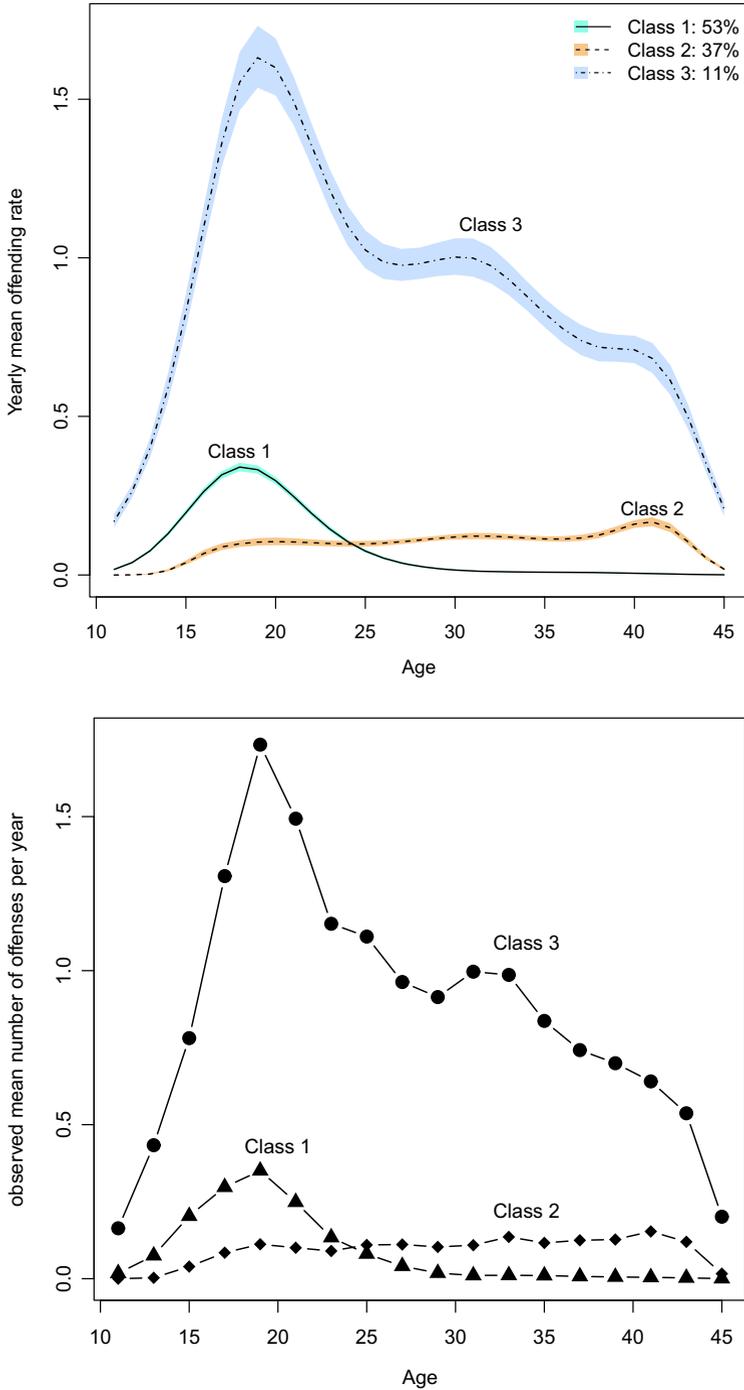


Fig. 3 The top graph shows the fitted B-spline trajectories for the three group model with seven degrees of freedom smooth fitted trajectories identified from the B-spline analysis

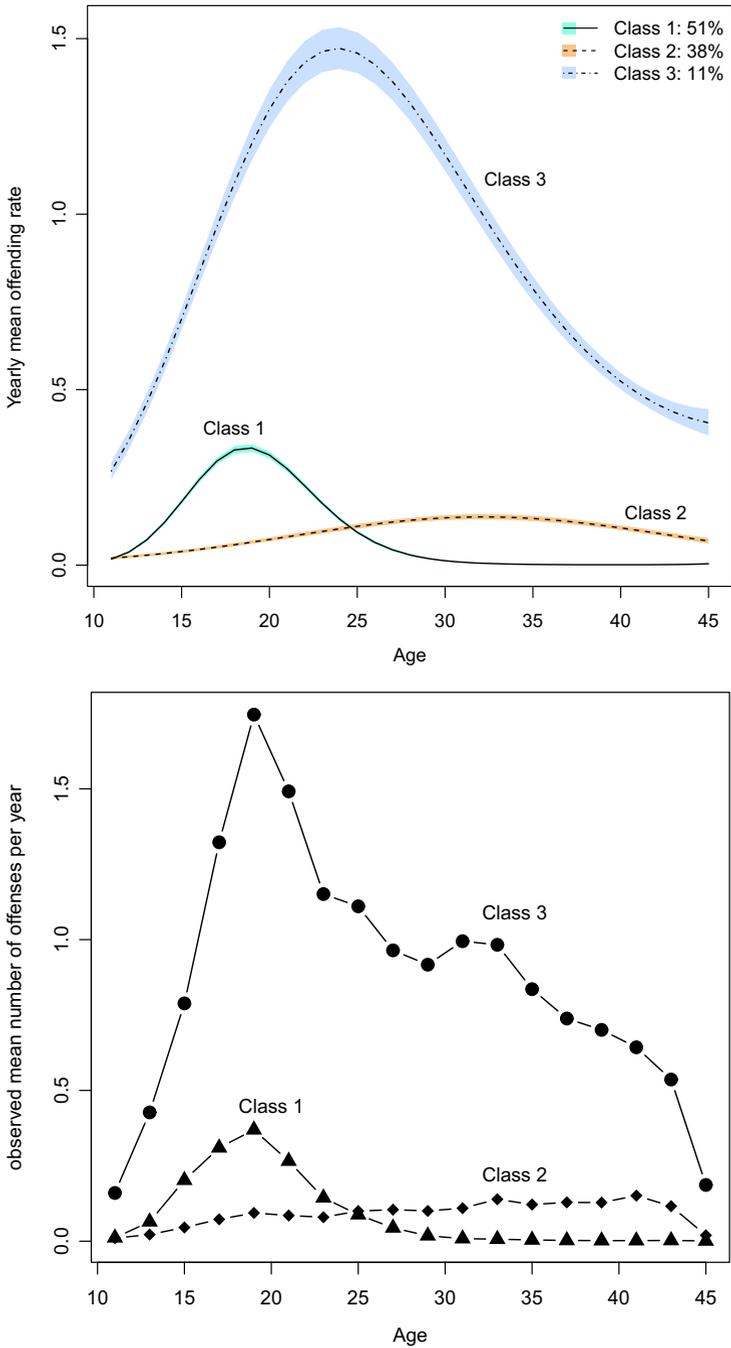


Fig. 4 The *top graph* shows the fitted *cubic* polynomial trajectories for the three group model. 95 % Confidence intervals are shown as lighter shaded areas surrounding each curve. The *bottom graph* shows the observed mean offense counts for the three fitted trajectories identified from the cubic polynomial analysis

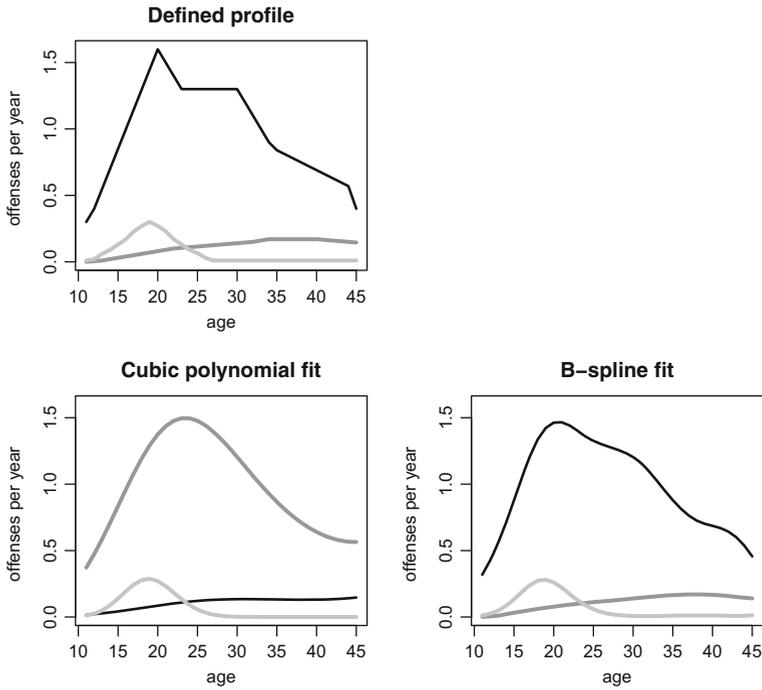


Fig. 5 Simulation results. The *top graph* shows the defined mean profiles listed in the [Appendix](#) and used to construct the simulated data set, and the *bottom two graphs* show the cubic polynomial trajectories and the B-spline trajectories fitted to the simulated data

misclassification rate is far lower for the B-spline method than for the traditional cubic fit, with a substantially greater amount of misclassification between groups 1 and 2 for the cubic model. In summary, the results support the real data analysis above, with the B-spline method giving a more flexible curve shape, and a better classification rate.

Table 2 Results of simulation: estimated group sizes and estimated dispersion

	Cubic model			B-spline model		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Estimated group sizes (%)	49.61	40.11	10.29	49.67	40.04	10.28
True group sizes (%)	50.05	39.64	10.31	50.05	39.64	10.31
Known scale parameter θ_k	2.000	2.000	2.000	2.000	2.000	2.000
Estimated scale parameter $\hat{\theta}_k$	2.382	1.975	2.045	2.041	2.009	2.033
BIC	393,557			391,978		
ICL-BIC	398,011			397,641		

Table 3 Results of simulation: misclassification rates and assignment tables for the two models

Cubic model

Estimated group membership

Known group membership	1	2	3	Total
1	5819	408	0	6227
2	1186	5141	11	6338
3	1	2	1432	1435
Total	7006	5551	1443	14,000
Misclassification rate:	0.1149			

B-spline model

Estimated group membership

Known group membership	1	2	3	Total
1	6560	660	1	7221
2	446	4889	6	5341
3	1	1	1436	1438
Total	7007	5550	1443	14,000
Misclassification rate:	0.0796			

Discussion and Conclusions

The results and small-scale simulation study presented in this paper show that the use of B-splines improve both the estimation of trajectory shape and the classification rate. Thus, B-splines in group-based trajectory modelling gives improved insight into the shapes of trajectories. Like many other writers (for example [28]), we can relate our trajectories in the real data example to Moffitt's taxonomy [21]. For the three offending groups shown in Fig. 3, the largest group is group 1 (51 % of offenders), which can be labelled as adolescent-limited. The shape of this trajectory appears to be quadratic. However, the remaining trajectories are not polynomial shaped. Group 2 appears to coincide with Moffitt's low rate chronic group. This has a nearly constant trajectory shape from around age 20 to age 40. Group 3 (the high-rate chronic group) also has a non-polynomial shape. From peaking at age 20 at around one and a half offenses per year, the trajectory flattens off at age 25 at about one offense a year, then declines to around half an offense a year by age 40. The trajectory has a far more complex offending shape than the cubic trajectory presented in Fig. 4, which has its peak at age 25 followed by a strong decline thereafter. The shape of the spline trajectory in our view fits better to the idea of a chronic or life-course persistent offender. However, as has been pointed out by [6], these substantive results would also support general life course criminological theories such the cumulative disadvantage theory of [27] and indeed Gottfredson and Hirshi's general theory of crime [12].

The disadvantage of cubic polynomial trajectories in generating uplifts not supported by the data has been highlighted in the introduction by an example from sexual offending. In that example, the aim was to examine whether there is evidence that some sexual offenders desist in early middle age, only to restart at a later age.

A method that generates false uplifts must then be suspect for these types of investigation, and the B-spline method offers strong advantages for these problems. However, no matter which method is used, researchers must also take care that their trajectories do not extend beyond the range of the data. Additionally, if the B-spline method detects uplifts, examination of the number of offenses in the region of the uplift will help to determine the criminological importance of such a result.

Another disadvantage of cubic polynomials that we identified in the introduction was that data points in the early part of the time axis can influence the shape of the curve in the later part of the time axis. While this can be seen mathematically, no empirical studies have been undertaken on GBTMs to investigate this problem, and we highlight this for future research.

One referee suggested that observed mean counts may be an easier way of understanding the underlying shape of trajectories. Certainly in this paper, the bottom graphs in Figs. 3 and 4 appear to be very similar. The procedure would then be to fit cubic polynomial trajectories and then to plot the observed mean counts at each age point rather than the fitted curves. We think this idea has promise for those without access to B-spline software, but it would be an approximate method as the assignment weights for the means are determined by the cubic trajectory model, rather than the better B-spline model. It would give a similar result to that obtained by fitting “age” as a factor.

The model is valuable to all three of the forms of application identified by [2] described in the introduction section. Clearly, it is of use to those applications who wish to interpret trajectories either as real subgroups or as approximations to reality with individual variability around each trajectory. But it is also of use to those who simply want to use trajectories to account for heterogeneity in the population through a non-parametric discrete mixing distribution. For these applications, mixtures of splines will provide more flexibility than mixtures of polynomials.

The model can be extended in various ways. We can easily relax the earlier assumption of count data, so that trajectory models can be fitted to binary or continuous observations—this will simply change the distributional assumption and the link function in Eq. 4. For continuous data, the `lcm` package in R can be used to fit the models. Missing observations can be incorporated easily as the likelihood can be constructed for only those time points that are observed. This would essentially use a full information maximum likelihood analysis and would assume a missing at random process.

Other extensions to the model presented here are possible but would require programming work. For example, the Zero-inflated Poisson (ZIP) model, which allows for intermittency, is popular in criminological applications as it is included in PROC TRAJ. However, the full form of the ZIP model is not available in Latent Gold. Where the Nagin and Land ZIP model specifies a different immune proportion for each time point and for each trajectory, the Latent Gold model ZIP model assumes a constant immune probability over all trajectories and time points. Other forms of smoothing technology could also be used in place of B-splines, and this would also require programming. We would expect results with other smoothers to be similar.

In conclusion, we can recommend the use of B-spline trajectories. They provide a more flexible way of fitting trajectories, and the results of such analyses are not constrained to the often unrealistic shapes of cubic curves. Some software can already

be used to fit these models. We also suggest that the assumption of negative binomial counts may be more realistic in many examples than the more common assumption of Poisson counts. We encourage software writers to add both facilities to packages that do not currently support them.

Acknowledgments We are grateful to three anonymous referees who provided useful and insightful comments which improved this paper. This work was supported by the UK Economic and Social Research Council (ESRC) (award numbers ES/K006460/1 and ES/J500094/1). This study was a re-analysis of existing data that are held by the Ministry of Justice and available on request.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Table 4 Trajectory means used in simulation study (“Simulation Study” section)

Age	Group trajectory means used in simulation offenses per year		
	Group 1	Group 2	Group 3
11	0.01	0	0.3
12	0.02	0.005	0.4
13	0.06	0.01	0.55
14	0.09	0.02	0.7
15	0.13	0.03	0.85
16	0.17	0.04	1
17	0.23	0.05	1.15
18	0.27	0.06	1.3
19	0.3	0.07	1.45
20	0.27	0.08	1.6
21	0.23	0.09	1.5
22	0.17	0.1	1.4
23	0.13	0.105	1.3
24	0.09	0.11	1.3
25	0.065	0.115	1.3
26	0.03	0.12	1.3
27	0.01	0.125	1.3
28	0.01	0.13	1.3
29	0.01	0.135	1.3
30	0.01	0.14	1.3
31	0.01	0.145	1.2
32	0.01	0.15	1.1
33	0.01	0.16	1

Table 4 (continued)

Age	Group trajectory means used in simulation offenses per year		
	Group 1	Group 2	Group 3
34	0.01	0.17	0.9
35	0.01	0.17	0.84
36	0.01	0.17	0.81
37	0.01	0.17	0.78
38	0.01	0.17	0.75
39	0.01	0.17	0.72
40	0.01	0.17	0.69
41	0.01	0.165	0.66
42	0.01	0.16	0.63
43	0.01	0.155	0.6
44	0.01	0.15	0.57
45	0.01	0.145	0.4
Group sizes	50 %	40 %	10 %

References

- Aitkin, M., Vu, D., & Francis, B. (2015). A new Bayesian approach for determining the number of components in a finite mixture. *Metron*, 73(2), 155–175.
- Bauer, D. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42, 757–786.
- Berk, R. (2008). *Statistical learning from a regression perspective*. New York: Springer.
- Blokland, A., Nagin, D., & Nieuwebeerta, P. (2005). Life span offending trajectories of a Dutch conviction cohort. *Criminology*, 43, 919–954.
- de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer-Verlag.
- Brame, R., Mulvey, E., & Piquero, A. (2001). On the development of different kinds of criminal activity. *Sociological Methods and Research*, 29, 319–341.
- Bushway, S.D., Thornberry, T., & Krohn, M.D. (2003). Desistance as a developmental process: a comparison of static and dynamic approaches. *Journal of Quantitative Criminology*, 19, 129–153.
- Bushway, S.D., Sweeten, G., & Nieuwebeerta, P. (2009). Measuring long term individual trajectories of offending using multiple methods. *Journal of Quantitative Criminology*, 25, 259–286.
- D'Unger, A., Land, K., McCall, P., & Nagin, D. (1998). How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses of the London, Philadelphia, and Racine cohorts studies. *American Journal of Sociology*, 103, 1593–1630.
- Francis, B., & Crosland, P. (2002). The police national computer and the offenders index: can they be combined for research purposes? Tech rep., Home Office, London. <http://eprints.lancs.ac.uk/50042/1/pncandoir170.pdf>.
- Francis, B., & Liu, J. (2015). Modelling escalation in crime seriousness: a latent variable approach. *Metron*, 73(2), 277–297.
- Gottfredson, M., & Hirshi, T. (1990). *A general theory of crime*. Stanford: Stanford University Press.
- Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman and Hall.
- He, X., & Ng, P. (1999). COBS: qualitatively constrained smoothing via linear programming. *Computational Statistics*, 14, 315–337.
- Jones, B., & Nagin, D. (2007). Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological Methods and Research*, 35, 542–571.

16. Liu, J., Francis, B., & Soothill, K. (2011). A longitudinal study of escalation in crime seriousness. *Journal of Quantitative Criminology*, 27, 175–196.
17. Liu, S. (2015). Is the shape of the age-crime curve invariant by sex? evidence from a national sample with flexible non-parametric modeling. *Journal of Quantitative Criminology*, 31, 93–123.
18. Marshall, J. (2006). *Juvenile offending trajectories: a South Australian study*. Tech. rep. Office of Crime Statistics and Research: Department of Justice, Government of South Australia. http://www.ocsar.sa.gov.au/docs/research_reports/JOT.pdf.
19. McLachlan, G., & Peel, D. (2000). *Finite mixture models* harvard university press. New York: Wiley.
20. Ministry of Justice (2015). Criminal Court Statistics Quarterly: July to September 2015: main tables (Table T5). Available at <https://www.gov.uk/government/statistics/criminal-court-statistics-quarterly-july-to-september-2015>.
21. Moffitt, T. (1993). 'Life-course-persistent' and 'adolescent limited' antisocial behavior—a developmental taxonomy. *Psychological Review*, 100, 674–701.
22. Nagin, D. (2005). *Group-based modeling of development*. Cambridge: Harvard Univ Press.
23. Nagin, D., & Land, K.C. (1993). Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, 31, 327–362.
24. Nagin, D., & Tremblay, R.E. (2005). Developmental trajectory groups: fact or a useful statistical fiction? *Criminology*, 43, 873–904.
25. Nylund, K., Asparouhov, T., & Muthen, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling*, 16, 535–569.
26. Osborne, M.R., Presnell, B., & Turlach, B.A. (1998). Knot selection for regression splines via the LASSO. *Proceedings of the 30th Symposium on the Interface: Computing Science and Statistics*, 30, 44–49.
27. Patterson, G.R. (1993). Orderly change in a stable world: the antisocial trait as a Chimera. *Journal of Consulting and Clinical Psychology*, 61, 911–919.
28. Piquero, A., Farrington, D., & Blumstein, A. (2007). *Key issues in criminal careers research*. Cambridge: Cambridge University Press.
29. Proust-Lima, C., Philipps, V., Diakite, A., & Liqueur, B. (2015). lcmm: Extended Mixed Models Using Latent Classes and Latent Processes. <http://CRAN.R-project.org/package=lcmm,rpackageversion1.7.2>.
30. Rodriguez, G. (2001). Smoothing and non-parametric regression. Tech rep., Princeton University, NJ, <http://data.princeton.edu/eco572/smoothing.pdf>.
31. Royston, P., & Altman, D.G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society Series C*, 43, 429–467.
32. Silverman, B. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society Series B*, 47, 1–52.
33. Skarðhamar, T. (2010). Distinguishing facts and artefacts in group-based modelling. *Criminology*, 48, 295–320.
34. Tr, S.karðhamar., & Savolainen, J. (2014). Changes in criminal offending around the time of job entry: a study of employment and desistance. *Criminology*, 52(2).
35. Smallbone, S., Marshall, W.L., & Wortley, R. (2008). *Preventing child sexual abuse: Evidence, Policy and practice Willan*. UK: Cullompton.
36. Sweeten, G. (2014). Group-based trajectory models. In Bruinsma, G., & Weisburd, D. (Eds.) *Encyclopedia of criminology and criminal justice*, (pp. 1991-2003). New York: Springer.
37. Verbyla, A. (1986). Conditioning in the growth curve model. *Biometrika*, 73, 475–483.
38. Vermunt, J.K., & Magidson, J. (2005). *Latent GOLD 4.0 user's guide statistical innovations Inc*. Belmont: MA.