Collection of Biostatistics Research Archive
COBRA Preprint Series

Modeling Criminal Careers as Departures from a Unimodal Population Age-Crime Curve: The Case of Marijuana Use

Donatello Telesca[*]   Elena Erosheva[†]

Derek Kreager[‡]   Ross Matsueda[**]

[*]UCLA, dtelesca@ucla.edu

[†]University of Washington

[‡]Pennsylvania State University

[**]University of Washington

Modeling Criminal Careers as Departures from a Unimodal Population Age-Crime Curve: The Case of Marijuana Use

Donatello Telesca, Elena Erosheva, Derek Kreager, and Ross Matsueda

Abstract

A major aim of longitudinal analyses of life course data is to describe the within- and between-individual variability in a behavioral outcome, such as crime. Statistical analyses of such data typically draw on mixture and mixed-effects growth models. In this work, we present a functional analytic point of view and develop an alternative method that models individual crime trajectories as departures from a population age-crime curve. Drawing on empirical and theoretical claims in criminology, we assume a unimodal population age-crime curve and allow individual expected crime trajectories to differ by their levels of offending and patterns of temporal misalignment. We extend Bayesian hierarchical curve registration methods to accommodate count data and to incorporate influence of baseline covariates on individual behavioral trajectories. Analyzing self-reported counts of yearly marijuana use from the Denver Youth Survey, we examine the influence of race and gender categories on differences in levels and timing of marijuana smoking. We find that our approach offers a flexible and realistic model for longitudinal crime trajectories that fits individual observations well and allows for a rich array of inferences of interest to criminologists and drug abuse researchers.

Modeling Criminal Careers as Departures
from a Unimodal Population Age-Crime Curve:
The Case of Marijuana Use [1]

Donatello Telesca
University of California Los Angeles

Elena A. Erosheva
University of Washington

Derek A. Kreager
Pennsylvania State University

Ross L. Matsueda
University of Washington

December 2, 2011

[1]Donatello Telesca is Assistant Professor of Biostatistics, Department of Biostatistics, University of California Los Angeles. Elena Erosheva is Associate Professor of Statistics and Social Work, Department of Statistics and CSSS, University of Washington. Derek Kreager is Associate Professor of Crime, Law and Justice, Department of Sociology, Pennsylvania State University. Ross Matsueda is Blumstein-Jordan Endowed Professor of Sociology, Department of Sociology, University of Washington.

Abstract

A major aim of longitudinal analyses of life course data is to describe the within- and between-individual variability in a behavioral outcome, such as crime. Statistical analyses of such data typically draw on mixture and mixed-effects growth models. In this work, we present a functional analytic point of view and develop an alternative method that models individual crime trajectories as departures from a population age-crime curve. Drawing on empirical and theoretical claims in criminology, we assume a unimodal population age-crime curve and allow individual expected crime trajectories to differ by their levels of offending and patterns of temporal misalignment. We extend Bayesian hierarchical curve registration methods to accommodate count data and to incorporate influence of baseline covariates on individual behavioral trajectories. Analyzing self-reported counts of yearly marijuana use from the Denver Youth Survey, we examine the influence of race and gender categories on differences in levels and timing of marijuana smoking. We find that our approach offers a flexible and realistic model for longitudinal crime trajectories that fits individual observations well and allows for a rich array of inferences of interest to criminologists and drug abuse researchers.

KEY WORDS: Curve Registration, Drug Use, Functional Data, Generalized Linear Models, Individual Trajectories, Longitudinal Data, MCMC, Unimodal Smoothing.

1 INTRODUCTION

An important task in criminology concerns describing individual trajectories of offending across time or age. An adequate description of offending trajectories across age is necessary for describing differences in criminal careers (Blumstein and Cohen 1987), for estimating features of age-crime curves (Hirschi and Gottfredson 1983), such as age-at-onset, and ultimately, for explaining differences in age crime curves using developmental or life course theories (Sampson and Laub 1993).

Most research on criminal careers and age-crime trajectories has been descriptive, following the pioneering work of Wolfgang et al. (1973), who examined age at onset, length of criminal careers, and patterns of desistance. Recent research has turned to model-based approaches, which typically specify individual trajectories as polynomial in age. Such models describe population heterogeneity in individual trajectories either by including random effects for age and age-squared (Raudenbush and Chan 1993), or by specifying a mixture of latent classes of trajectories (Nagin and Land 1993), or by combining latent trajectory classes and random effects (Muthén and Shedden 1999). However, polynomial representations are typically not able to capture nuanced heterogeneity between individuals in their observed patterns of criminal behavior, and research findings are often driven by variability in within-age behavioral amplitude (Gottfredson and Hirschi 1990).

In this article, we propose an alternative approach for analyzing longitudinal crime data. We draw on the criminological work of Hirschi and Gottfredson (1983), who argued that the age-crime curve is invariant across social groups and throughout history. Arguing that a single unimodal age-crime curve underlies all crime, including illicit drug use, they described the curve as rising precipitously from age seven (age of culpability) until the peak years–between ages 13-21, depending on the crime–and then slowly declining thereafter through the life span. Although they claimed invariance in the basic shape of the age-crime curve, they also acknowledged the presence of individual differences in crime trajectories. Specifically, Hirschi and Gottfredson (1983) claimed that individual differences are driven by differences in (time-stable) levels of offending and (time-varying) opportunities to commit crime. We note that these substantive arguments naturally lend themselves to using an appropriately constrained functional data analysis approach for modeling longitudinal crime data.

We develop a Poisson warping regression model that assumes a unimodal mean age-crime curve and defines individual crime trajectories as random functions that deviate from the mean curve according to individual-specific level of offending and time misalignment patterns. We build on curve registration models of Ramsay and Li (1998), who introduced a model for the alignment of a sample of curves via a continuous monotone transformation of a main effect modifier (usually time), and Telesca and Inoue (2008), who formulated a Bayesian hierarchical model for curve registration,

allowing for the borrowing of information across curves. To accommodate discrete observations (counts), we develop a generalized extension of the curve registration models to count data. In addition, we incorporate covariate effects directly on (1) the expected intensity of criminal behavior and on (2) the deviation from the average timing of offenses in a hierarchical fashion.

We are not the first to take a functional data analytic point of view towards longitudinal crime data. Ramsay and Silverman (2002) carried out a functional principal component analysis on a landmark data set originally collected by Glueck and Glueck (1950), and reanalyzed by Sampson and Laub (1993). Our approach to analyzing life course crime trajectories, although functional, is fundamentally different from that of Ramsay and Silverman (2002) as we do not rely on principal components.

Several authors have contributed to the statistical analysis of random curves. Shi et al. (1996) were among the first to introduce flexible semiparametric models for the analysis of a sample of curves based on functional mixed effects modeling. In the analysis of sparsely observed functions, Rice and Silverman (1991), and, more recently, Yao et al. (2005) discuss nonparametric methods based on functional principal component analysis.

Typically, functional data analysis deals with large amounts of data sampled on a fine grid in time or space (Brumback and Lindstrom 2004, Gervini and Gasser 2004). Information on lifetime criminal behavior, however, often comes in the form of many short or sparsely sampled time series (see Elliott et al. 1985 or Harris et al. 2003). High individual heterogeneity in combination with such data structures requires models that capitalize on borrowing information across subjects while maintaining a high level of flexibility in order to provide a reasonable fit to individual observed trajectories.

Our method of hierarchical curve registration with covariates allows us to develop a flexible set of nonparametric representations for individual curves of criminal offending. It deals with data sparsity by combining information across curves in two ways: (1) structurally, by representing individual curves as an affine transformation of a natural crime curve constrained to be unimodal; and (2) stochastically, by assuming conditional dependence (exchangeability) between key parameters contributing to the likelihood function. As we model the crime trajectories in a semi-parametric fashion, we integrate the substantive claims of Hirschi and Gottfredson (1983) with the existing toolkit of functional data analysis methods and accommodate the unimodality constraint for non-Gaussian data. Unlike previous approaches to modeling crime trajectories, our approach explicitly incorporates criminological arguments that the population age crime curve is unimodal and that individual trajectories can be described as departures from the common population curve. We illustrate our approach by analyzing data on marijuana use.
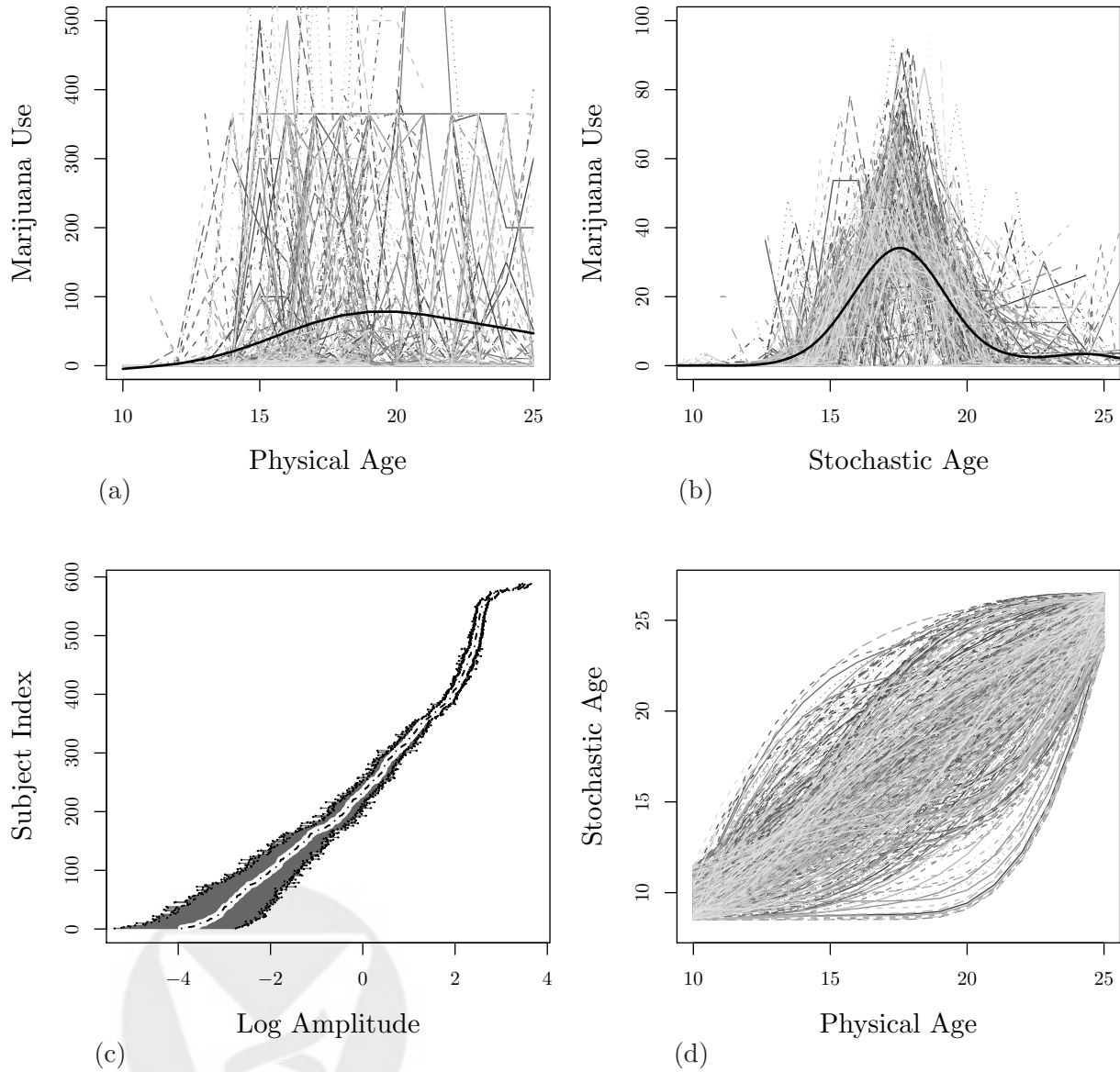
2

Figure 1: Drug Use (Marijuana). *Panel (a)*: Yearly count for the use of marijuana for 588 subjects from the DYS. A solid black line depicts the structural mean function. *Panel (b)*: Aligned normalized trajectories. In black we report the overall functional convex average $\mathbf{S}(t, \boldsymbol{\beta})$. *Panel (c)*: Subject–specific posterior log–amplitude with associated 95% credible intervals. *Panel (d)*: Subject-specific time scale, characterized by the expected posterior time transformation functions.

3

1.1 Data

We consider marijuana use data from the Denver Youth Survey (DYS) (Esbensen and Huizinga 1990), a longitudinal study of delinquency and drug use in high risk neighborhoods in Denver. Marijuana use is of interest not only to drug researchers and life course scholars but also to criminologists because it is an illegal substance in the United States. The DYS collected data from an accelerated longitudinal design covering the age span from 7 to 25. The peak age of marijuana use is about age 20 (Office of National Drug Control Policy 2008). The survey asked drug use questions starting from age 11.

The DYS identified high risk neighborhoods via a cluster analysis of census variables such as family structure, ethnicity, SES, housing, mobility, marital status, and age composition (e.g., Esbensen and Huizinga 1990). High risk neighborhoods were then defined as the top third in terms of high social disorganization and high official crime rates. These neighborhoods represent the most disadvantaged areas of Denver.

The investigators selected a sample of 20,300 households from high-risk neighborhoods in Denver, and used a screening questionnaire to identify five child and youth cohorts (i.e., 7, 9, 11, 13 or 15 years old in 1988). The overall procedure yielded a sample of 1,528 respondents (for details see Matsueda et al. (2006); Esbensen and Huizinga (1990)). Of these respondents, 1,459 were aged 11 years or older for at least one interviewed year and completed a youth survey that included drug-use counts. Subjects were interviewed in their homes annually from 1988-1992 and 1995-1999 (10 waves).

We consider answers to the survey question "In the past year, how many times have you smoked marijuana?" Our goal is to model individual trajectories of marijuana use over the interval of 10-25 years of age, and to understand differences in these trajectories by race-ethnicity and gender. We selected individuals who had between 4 and 9 longitudinal observations on marijuana use for each individual. The resulting data set had a mean of 7.39 (SD = 1.37) observations per subject. The frequency of marijuana use is highly volatile. Marijuana smokers – those who reported smoking at least once during the observation period – smoked marijuana 42.85 times per year on average with SD = 133.33 and a maximum reported count of yearly marijuana smoking of 999. [1] One commonly reported quantity of marijuana use over the last year is 365 times that corresponds to the once a day frequency of smoking.

The remainder of this article is organized as follows. In Section 2, we introduce a hierarchical model for the semi–parametric analysis of longitudinal count data. We discuss MCMC estimation and inference in Section 3 and analyze lifetime data on marijuana use from the Denver Youth Survey in Section 4. We conclude with a discussion of our contributions and possible model extensions in

---

[1] A few respondents reported using marijuana more than twice a day. For those very few who reported marijuana more then 3 times a day, their answers were truncated to the maximum of 999.

Section 5.

2 HIERACHICAL REGISTRATION

2.1 Poisson Warping Regression Model

In this section, we introduce a general formulation for the functional representation of longitudinal crime data. Let $\mathbf{Y}_i = (Y_{i1}, ..., Y_{ij}, ..., Y_{in})'$ denote an observed vector of offenses for individual i over a discrete sampling design $\mathbf{t} = (t_1, ..., t_j, ..., t_n)$. To simplify notation, we assume that a sampling design $\mathbf{t}$ is common for all individuals but the functional model is flexible to accommodate different sampling times. Technically, observed counts Y_{ij} denote the number of offenses over a reasonable time interval τ, e.g., a month or a year, just before sampling times t_j. The time interval τ is fixed and the same for all observations in the sample. Let $\mathbf{X}_i$ denote a p–dimensional vector of time-stable covariates for individual i.

We assume that individual trajectories of offending are realizations from a functional Poisson process. Thus, the observed count at time t_j for individual i is

$$Y_{ij} \mid \lambda_i(t_j, \mathbf{X}_i) \sim Poisson\{\lambda_i(t_j, \mathbf{X}_i)\}, \tag{1}$$

where $E(Y_{ij} \mid \mathbf{X}_i) = \lambda_i(t_j, \mathbf{X}_i)$. The sampling density of Y_{ij} is given by

$$P(Y_{ij} = y_{ij} \mid \lambda_i(t_j, \mathbf{X}_i)) = \frac{e^{-\lambda_i(t_j, \mathbf{X}_i)} \lambda_i(t_j, \mathbf{X}_i)^{y_{ij}}}{y_{ij}!}.$$

Assume the intensity function $\lambda_i(t_j, \mathbf{X}_i)$ depends on the covariate information $\mathbf{X}_i$ as follows:

$$\lambda_i(t_j, \mathbf{X}_i) = a_i(\mathbf{X}_i)\, \mathbf{S}(t_j, \boldsymbol{\beta}) \circ \mu_i(t_j, \boldsymbol{\phi}_i; \mathbf{X}_i) = a_i(\mathbf{X}_i)\, \mathbf{S}\{\mu_i(t_j, \boldsymbol{\phi}_i; \mathbf{X}_i), \boldsymbol{\beta}\}, \tag{2}$$

where $a_i(\mathbf{X}_i) \geq 0$ is an individual-specific amplitude, $\mathbf{S}(t_j, \boldsymbol{\beta})$ is a mean shape function, and $\mu_i(t_j, \boldsymbol{\phi}_i; \mathbf{X}_i)$ is an individual-specific time transformation function. Consequently, the mean function $\mathbf{S}(t_j, \boldsymbol{\beta})$, evaluated over a subject-specific time scale $\mu_i(t_j, \boldsymbol{\phi}_i; \mathbf{X}_i)$, defines individual-specific mean trajectory of offending. Our notation indicates explicit dependence on $\mathbf{X}_i$ for individual amplitude and time-transformation functions. We define this dependence in Sections 2.3 and 2.4, respectively, by modeling the mean of a_i and $\boldsymbol{\phi}_i$ as a function of covariates $\mathbf{X}_i$.

For the mean shape and time transformation functions, we assume that their functional forms belong to the Sobolev space spanned by linear combinations of cubic B-Spline basis functions (De Boor 1978). Intuitively, this is a vector space containing shapes of virtually arbitrary flexibility, provided it originates from an adequate number of basis functions. See also Peña 1997 for a discussion of B-spline optimality and stability. When modeling the shape function $\mathbf{S}(t, \boldsymbol{\beta})$ we further constrain the functional form to be unimodal.

The modeling framework introduced in equations (1) and (2) is consistent with the substantive arguments of Hirschi and Gottfredson (1983) about the age-crime curve. It starts by assuming

5

a common unimodal shape for the age-crime curve and reflects individual differences in the expected intensity of criminal behavior $a_i(\mathbf{X}_i)$ and deviations from the average timing of offenses $\mu_i(t_j, \boldsymbol{\phi}_i; \mathbf{X}_i)$.

2.2 Mean Shape $\mathbf{S}(t, \boldsymbol{\beta})$ and Unimodal Smoothing

Let the mean shape function $\mathbf{S}(t, \boldsymbol{\beta})$ be a mapping $\mathbf{S}(t, \boldsymbol{\beta}) : \mathcal{T} \longrightarrow \mathbb{R}^+$, where $\mathcal{T} = [t_1 - \Delta, \, t_n + \Delta]$ is the observed sampling interval $[t_1, \, t_n]$ that is extended by a temporal misalignment window $\Delta < \infty$ (Telesca and Inoue (2008)). Assume the functional form of the average shape $\mathbf{S}(t, \boldsymbol{\beta})$ is a linear B-spline combination $\mathbf{S}(t, \boldsymbol{\beta}) = \mathcal{S}_B(t)' \boldsymbol{\beta}$, where $\mathcal{S}_B(t)$ is a set of K basis functions of order 4 evaluated at time t and $\boldsymbol{\beta}$ is a p-dimensional vector of spline coefficients. To ensure positivity of $\mathbf{S}(t, \boldsymbol{\beta})$, it is sufficient to require positivity of the shape coefficients $\beta_j \geq 0$, $j = 1, ..., K$. To ensure unimodality of $\mathbf{S}(t, \boldsymbol{\beta})$, it is sufficient to require the first derivative $\partial \mathbf{S}(t, \boldsymbol{\beta}) / \partial t$ to exibit only one possible sign change (Schumaker 1981, Theorem 4.76). We combine the unimodality and positivity requirements via the following reparametrization of the shape coefficients $\boldsymbol{\beta}$:

$$\beta_k = \nu^{*2} - (\nu_k - \nu^*)^2, \quad k = 1, ..., K, \tag{3}$$

where the new coefficients $\boldsymbol{\nu} = (\nu_1, ..., \nu_K)'$ are nondecreasing, i.e., $0 = \nu_1 \leq \cdots \leq \nu_K$, and ν^* is a fixed modal pivot [2]. We place a second order shrinkage prior distribution on $\boldsymbol{\nu}$. In particular, assuming $\nu_0 = \nu_1 = 0$, we model the generic k^{th} element of $\boldsymbol{\nu}$ as

$$\nu_k = 2\nu_{k-1} - \nu_{k-2} + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_\beta^2). \tag{4}$$

The variance parameter σ_β^2 can then be interpreted as a smoothing parameter shrinking the shape function towards a piece-wise linear trajectory.

2.3 Amplitude Parameters $a_i(\mathbf{X}_i)$ and Amplitude Regression

The notion that individual criminal propensity is constant across the life span but varies among individuals is common in the criminology literature. Gottfredson and Hirschi (1990) introduced the hypothetical concept of self-control that could explain this variation. The amplitude regression part of our model allows us to test the relationship between individual criminal propensity and observed covariates.

We model the dependence of individual-specific amplitude a_i on covariates $\mathbf{X}_i$ in a generalized linear fashion:

$$E(a_i \mid \mathbf{X}_i, \mathbf{b}_a) = \exp\{\mathbf{X}_i' \mathbf{b}_a\}, \quad i = 1, \ldots, N, \tag{5}$$

---

[2]In our analysis choosing $\nu^* = \sqrt{\overline{\mathbf{Y}}}$ provided a reasonable reference scale on the magnitude of $\mathbf{S}(t, \boldsymbol{\beta})$, for the prior on $\mathbf{b}_a$ to be reasonably centered around 0.

6

where $\mathbf{b}_a$ is a p-dimensional vector of amplitude regression coefficients. To specify a prior distribution for a_i with the mean given by equation (5), we exploit the Gamma-Poisson conjugacy and assume

$$(a_i \mid \mathbf{b}_a, b_0; \mathbf{X}_i) \sim \mathcal{G}\left(b_0, b_0 \exp\{-\mathbf{X}_i \mathbf{b}_a\}\right).$$ (6)

In this formulation, $1/\sqrt{b_0}$ represents the coefficient of variation.

The prior distribution of a_i in equation (6) has two appealing properties. First, due to conjugacy, the conditional posterior density of a_i is

$$P(a_i \mid \mathbf{Y}_i, \boldsymbol{\beta}, \boldsymbol{\phi}_i, \mathbf{b}_a, b_0; \mathbf{X}_i) \propto a_i^{\{b_0 + \sum_j Y_i(t_j) - 1\}} \, \exp\left\{-\left(\frac{b_0}{\exp\{\mathbf{X}_i' \mathbf{b}_a\}} + \sum_j \mathbf{S}\{\mu_i(t_j, \boldsymbol{\phi}_i), \boldsymbol{\beta}\}\right) a_i\right\},$$ (7)

which corresponds to Gamma distribution with shape parameter $(b_0 + \sum_j Y_i(t_j))$ and rate $(b_0 \exp\{-\mathbf{X}_i' \mathbf{b}_a\} + \sum_j \mathbf{S}\{\mu_i(t_j, \boldsymbol{\phi}_i), \boldsymbol{\beta}\})$. In addition, the marginal distribution of observed offense counts Y_{ij}, integrating over a_i, is Negative Binomial:

$$
\begin{aligned}
P(Y_{ij} \mid \boldsymbol{\beta}, \boldsymbol{\phi}_i, \mathbf{b}_a, b_0; \mathbf{X}_i) &= \int_0^\infty P(Y_{ij} \mid \boldsymbol{\beta}, \boldsymbol{\phi}_i, a_i, \mathbf{b}_a, b_0) P(a_i \mid \mathbf{b}_a, b_0) da_i \\
&= \frac{\Gamma(Y_{ij} + b_0)}{b_0 \ Y_{ij}!}(1 - W_{ij})^{b_0} \ W_{ij}^{Y_{ij}},
\end{aligned}
$$ (8)

where

$$W_{ij} = \left(\frac{\exp\{\mathbf{X}_i' \mathbf{b}_a\} \mathbf{S}\{\mu_i(t_j, \boldsymbol{\phi}_i), \boldsymbol{\beta}\}}{\exp\{\mathbf{X}_i' \mathbf{b}_a\} \mathbf{S}\{\mu_i(t_j, \boldsymbol{\phi}_i), \boldsymbol{\beta}\} + b_0}\right)^{Y_{ij}}.$$

This form allows for natural modeling of overdispersion in the marginal distribution of counts. Here, small values of b_0 indicate extra variability beyond that explained by the Poisson.

In the presence of amplitude parameters a_i, scale identifiability is often an issue. In Gaussian models, for example, Gervini and Gasser (2004) and Brumback and Lindstrom (2004) impose summation constraints of the kind $\sum_i^N a_i = N$. From a Bayesian perspective, scale identification can be achieved by modeling dependence between the a_i at the population level (see Telesca and Inoue 2008).

The amplitude part of the model is completed with priors for the coefficient of variation and for the regression coefficients respectively

$$b_0 \sim \mathcal{G}(\lambda_a, \lambda_b), \qquad \mathbf{b}_a \sim \pi(\mathbf{b}_a).$$ (9)

The specific form for $\pi(\mathbf{b}_a)$ is described in the next section, in order to relate amplitude and phase effects.

2.4 Time Transformation Functions $\mu_i(t, \boldsymbol{\phi}_i)$ and Phase Regression

Criminologists specify multiple dynamic influences on trajectories of offending. For example, changes in crime and drug use over time are attributed to changes in peer groups, opportu-

nities, school experiences, and neighborhood contexts. In our model, we use individual time-transformation functions and a phase shift to account for such dynamic individual-specific influences. In addition, we model the phase shift as a linear function of time-stable covariates, gender and race-ethnicity. This allows us to test whether certain groups of individuals start their criminal careers on average earlier than other groups, controlling for differences in amplitudes.

We allow time transformation functions to map the original time scale onto random image sets enclosed in an extended sampling interval $\mathcal{T} = [t_1 - \Delta, \, t_n + \Delta]$, that is $\mu_i(t, \boldsymbol{\phi_i}) : [t_1, \, t_n] \longrightarrow \mathcal{T}$ (Telesca and Inoue (2008)). As before, $[t_1, t_n]$ is the observed time interval and $\Delta < \infty$ is a temporal misalignment window. We require subject-specific time transformation functions $\mu_i(t, \boldsymbol{\phi_i})$ to be strictly monotone, $\partial \mu_i(t, \boldsymbol{\phi}_i)/\partial t > 0$ (Ramsay and Li 1998), to prevent time reversibility and to define a bijection between the original time scale t and the transformed time scale $\mu_i(t, \boldsymbol{\phi}_i)$.

Let $\mathcal{S}_\mu(t)$ denote a set of Q B-spline basis functions of order 4, evaluated at time t. We define the subject-specific time transformation functions as linear combinations $\mu_i(t, \boldsymbol{\phi}_i) = \mathcal{S}_\mu(t)' \boldsymbol{\phi}_i$ for a given Q–dimensional vector of basis coefficients, $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{iQ})'$. Imposing the ordering $\phi_{i1} < \cdots < \phi_{iq} < \cdots < \phi_{iQ}$ provides us with a sufficient condition for time transformation functions $\mu_i(t, \boldsymbol{\phi}_i)$ to be monotone (Brumback and Lindstrom 2004). Additionally, imposing boundary conditions $(t_1 - \Delta \leq \phi_{i1} \leq t_1 + \Delta)$ and $(t_n - \Delta \leq \phi_{iQ} \leq t_n + \Delta)$ allows for the time transformations $\mu_i(t, \boldsymbol{\phi}_i)$ to map the original time scale t onto random intervals not bigger than $[t_1 - \Delta, \, t_n + \Delta]$ and not smaller than $[t_1 + \Delta, \, t_n - \Delta]$. This last requirement rules out possible degeneracies, provided that the temporal misalignment window is such that $\Delta << (t_n - t_1)/2$.

Let Υ be a Q–dimensional vector of identity coefficients, so that $\mathcal{S}_\mu(t)\Upsilon' = t$. Following the penalization approach introduced in Lang and Brezger (2004), we assume that individual time transformation coefficients $\boldsymbol{\phi}_i$ arise from a first-order random walk shrinkage prior. Thus, for all $i = 1, \ldots, N$,

$$(\phi_{iq} - \gamma_{iq}) = (\phi_{i(q-1)} - \gamma_{iq-1}) + \eta_q, \ \text{with} \ \eta_q \sim \mathcal{N}(0, \sigma_\phi^2) I\{\mathcal{M}\}, \ q = 1, \ldots, Q, \quad (10)$$

where $\phi_{i0} = \Upsilon_0 = 0$. Here, $\mathcal{M}$ defines a set of random cuts such that $\eta_q - \eta_{q-1} > \Upsilon_{q-1} - \Upsilon_q$, $q = 1, \ldots, Q$, where $|\eta_1| \leq \Delta$ and $|\eta_Q| \leq \Delta$. The variance parameter σ_ϕ^2 is a smoothing parameter that controls the amount of shrinkage of individual time transformation functions towards the identity transformation $\mu_i(t, \Upsilon) = t$.

We incorporate covariate effects by modeling the average phase shift as a linear function of covariates $\mathbf{X}_i$:

$$\gamma_{iq} = E[\phi_{iq} | \mathbf{X}_i, \mathbf{b}_\phi] = \Upsilon_q + \mathbf{X}_i' \mathbf{b}_\phi, \ \ i = 1, \ldots, N. \quad (11)$$

Finally, we complete the model with priors over phase and amplitude regression coefficients $(\mathbf{b}_a', \mathbf{b}_\phi')' \sim N(\mathbf{b_0}, \Sigma_b)$ with conditionally conjugate hyperprior $\Sigma_b \sim IW(\nu_b, c_b \, I_{2p})$.

3 ESTIMATION AND INFERENCE

Our modeling approach can be essentially summarized as follows. Marijuana use in time is assumed to arise as the realization of a functional Poisson process with mean structure (2). Dependence on covariate information is included through amplitude effects (5) and phase shifts (11).

3.1 Likelihood Function

Using the B-Spline representations for the mean shape and time transformation functions described in Sections (2.4) and (2.2), we rewrite the expected number of offenses for subject i at time t_j from equation (2) as:

$$\lambda_i(t_j, \boldsymbol{\beta}, a_i, \boldsymbol{\phi}_i) = a_i \, \mathcal{S}_B(t_j)' \boldsymbol{\beta} \circ \mathcal{S}_\mu(t_j)' \boldsymbol{\phi}_i = a_i \, \mathcal{S}_B\{\mathcal{S}_\mu(t_j)' \boldsymbol{\phi}_i\}' \boldsymbol{\beta}. \tag{12}$$

Here we omit the explicit dependence of a_i and $\boldsymbol{\phi}_i$ on covariates $\mathbf{X}$ to simplify notation. The log-likelihood function of shape coefficients $\boldsymbol{\beta}$, amplitude parameters $\mathbf{a} = (a_1, ..., a_N)'$ and time transformation coefficients $\boldsymbol{\Phi} = (\boldsymbol{\phi_1}', ..., \boldsymbol{\phi}_N')'$ is then

$$\ell(\boldsymbol{\beta}, \mathbf{a}, \boldsymbol{\Phi} | \mathbf{Y}) \propto \sum_{i=1}^{N} \sum_{j=1}^{n} \left[Y_i(t_j) \, log\{\lambda_i(t_j, \boldsymbol{\beta}, a_i, \boldsymbol{\phi}_i)\} - \lambda_i(t_j, \boldsymbol{\beta}, a_i, \boldsymbol{\phi}_i) \right]. \tag{13}$$

The above formulation of the likelihood depends on the choice of the number and locations of the spline knots for the mean shape $\mathbf{S}(t, \boldsymbol{\beta})$ and time transformation functions $\mu_i(t, \boldsymbol{\phi}_i)$. Because the mean shape $\mathbf{S}(t, \boldsymbol{\beta})$ is estimated from multiple individual trajectories, several authors in functional data analysis recommend selecting a large number of knots. For example, placing knots at every sampling time point can allow for a high level of shape flexibility. The level of smoothness is then selected automatically or ad hoc via likelihood or prior penalization schemes (Lang and Brezger 2004, Eilers and Marx 1996). The shrinkage prior as in equation (4) automatically shrinks the fixed effect functions towards a linear regression. In our case of highly sparse longitudinal offense data, however, we observed some sensitivity to the choice of the number of basis functions. To select the number of basis functions, we therefore recommend applying a model selection criterion based on the minimization of a posterior predictive loss (Gelfand and Ghosh 1998).

Let Y^o denote the observed counts and Y^p denote the predicted counts. Following Gelfand and Ghosh (1998), we obtain the deviance version of the posterior predictive loss criteria for the Poisson case as

$$D_k(m) = \sum_{ij} \{h_{ij} - h(\mu_{ij}^p)\} + \sum_{ij} \left\{ \frac{h(\mu_{ij}^p) + k\, h(Y_{ij}^o)}{k+1} - h\left(\frac{\mu_{ij}^p + k\, Y_{ij}^o}{k+1} \right) \right\}, \quad i = 1, ..., N, \; j = 1, ..., n, \tag{14}$$

where $h(x) = (x + 1/2)log(x + 1/2) - x$, $h_{ij} = E\{h(Y_{ij}^p \mid \mathbf{Y})\}$, and $\mu_{ij}^p = E\{Y_{ij}^p \mid \mathbf{Y}\}$. Here, m denotes the number of basis functions in the model to be evaluated.

9

Different considerations apply for the subject–specific time transformation functions. These maps carry structural smoothness as they are constrained to be monotone. The strict monotonicity requirement counterbalances the small number of observations associated with each individual trajectory and suggests parsimony in the choice of the number of knots. Because the time scale is stochastic, the exact placement of knots is less important in this case; thus we place knots for time transformation functions as equally spaced.

In the following section, we describe a Markov chain Monte Carlo (MCMC) algorithm for posterior simulation based on fixed numbers of spline basis for mean shape $\mathcal{S}_B(\cdot)$ and time transformation $\mathcal{S}_\mu(\cdot)$ functions.

3.2 Posterior Simulation via MCMC

For the Poisson warping regression model described in Section 2, the full parameter set $\boldsymbol{\theta}$ includes an N-dimensional vector of individual-specific random amplitude coefficients $\mathbf{a} = (a_1, ..., a_N)'$, an $(N \times Q)$ matrix of individual-specific time transformation coefficients $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1', ..., \boldsymbol{\phi}_N')'$, a p-dimensional vector of population-level shape coefficients $\boldsymbol{\beta}$, and population level regression and smoothing parameters $\mathbf{b}_a$, $\mathbf{b}_\phi$, b_0, σ_ϕ^2 and σ_β^2.

We seek inference about $\boldsymbol{\theta}$ and functionals of $\boldsymbol{\theta}$ through the posterior probability $P(\boldsymbol{\theta} \mid \mathbf{Y}; \mathbf{X}) \propto p(\mathbf{Y} \mid \boldsymbol{\theta}; \mathbf{X}) P(\boldsymbol{\theta}; \mathbf{X})$, where $p(\mathbf{Y} \mid \boldsymbol{\theta}; \mathbf{X})$ is described by the log-likelihood in equation (13) and $P(\boldsymbol{\theta}; \mathbf{X})$ represents the joint prior distribution. Recall the dependence on covariates for amplitude and time-transformation parameters through their respective prior distributions (equations (6) and (11)). Because the posterior distribution is not available in closed form, we base our inferences on an MCMC simulation from the joint posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{Y}; \mathbf{X})$ (for a recent review, see Gamerman 1997). We use a Gibbs sampler (Gelfand and Smith 1990) whenever conditional posterior quantities are available in standard distributional form. Otherwise, we derive an efficient sampling scheme, combining Gibbs steps with Metropolis-Hastings (MH) steps (Hastings 1970) in a hybrid sampler (Tierney 1994).

Sampling phase regression coefficients $\mathbf{b}_\phi$ *and smoothing parameters* σ_ϕ^2 *and* σ_β^2: The prior model induces likelihood conjugacy in the conditional posterior distribution of the phase regression coefficients $\mathbf{b}_\phi$ and the smoothing parameters σ_ϕ^2 and σ_β^2. For these quantities, it is therefore straightforward to devise an efficient Gibbs sampler based on direct simulation from their complete conditional distributions that we include in the Appendix.

Sampling time transformation coefficients $\boldsymbol{\phi}$: Taking advantage of the fact that the time transformation coefficients $\boldsymbol{\phi}$ have compact support $\mathcal{T} = [t_1 - \Delta, t_n + \Delta]$, we implement a MH sampler with appropriately scaled transition kernels $q(\boldsymbol{\phi}^{old}, \boldsymbol{\phi}^{new})$. Given that $\phi_{iq} < \phi_{i(q+1)}$, $(\forall \ i = 1, ..., N, \ q = 1, ..., Q)$, we consider proposal densities of the form $q(\phi_{iq}^{old}, \phi_{iq}^{new}) = \mathcal{N}(\phi_{iq}^{old}, s_{iq}^2) I\{\mathcal{M}\}$, where $\mathcal{M}$

is the compact support defined in equation (10), Section 2.4. During the MCMC simulation, for each set of individual-specific time transformation coefficients, we start from some value s^2 for the variance of the proposal density and recalibrate the individual proposal variances s_{iq}^2 at burn-in to achieve an acceptance rate between 35% and 65% (Roberts and Rosenthal 2001).

Sampling amplitude parameters $\mathbf{a}$, $\mathbf{b_a}$ *and* b_0:   We use a Gibbs sampler and simulate directly from the conditional posterior distribution given in equation (7) to update individual amplitude parameters $\mathbf{a} = (a_1, ..., a_N)'$ one at a time.

The conditional posterior distributions for regression coefficients $\mathbf{b}_a$ and for the coefficient of variation b_0 are not available in closed form. We implement MH scans with proposal distributions informed by the respective target densities. If Ω_{b_a} defines the inverse of the prior covariance matrix (i.e., the concentration matrix) on amplitude regression coefficients $\mathbf{b}_a$, the conditional posterior density of $\mathbf{b}_a$ can be written as:

$$log\{P(\mathbf{b}_a \mid \mathbf{a}, b_0, \Omega_{b_a})\} = -\frac{1}{2} \mathbf{b}_a' \Omega_{b_a} \mathbf{b}_a - b_0 \sum_i \{\mathbf{X}_i' \mathbf{b}_a + a_i \exp(-\mathbf{X}_i' \mathbf{b}_a)\},$$

with gradient vector

$$g_{b_a} = -\frac{1}{2} \mathbf{b}_a' \Omega_{b_a} - b_0 \sum_i \{\mathbf{X}_i' - a_i \mathbf{X}_i' \exp(-\mathbf{X_i}' \mathbf{b}_a)\},$$

and Hessian matrix

$$H_{b_a} = -\frac{1}{2} \Omega_{b_a} - b_0 \sum_i \{a_i \mathbf{X}_i \mathbf{X}_i' \exp(-\mathbf{X}_i' \mathbf{b}_a)\}.$$

Given g_{b_a} and H_{b_a}, we approximate the conditional posterior mode $\hat{\mathbf{b}}_a$ numerically via Newton-Raphson. Defining $\Sigma_{b_a}^* = (H_{b_a} \mid_{\hat{\mathbf{b}}_a})^{-1}$, we obtain the transition kernel on the basis of the over-relaxed proposal $q(\mathbf{b}_a^{old}, \mathbf{b}_a^{new}) = 2\hat{\mathbf{b}}_a - \mathbf{b}_a^{old} + MVt(0, \tau_{b_a} \Sigma_{b_a}^*, \eta)$. The parameters τ_{b_a} and η can be used to tune the MH acceptance ratio.

For the coefficient of variation b_0, we use a MH step to sample from the conditional posterior density. We consider a Gamma proposal with shape v_0 and rate $v_0/\hat{b}_0$; the parameter v_0 can be used to tune the MH acceptance ratio, while the conditional moment estimator of b_0 is defined as $\hat{b}_0 = \sum_i \frac{\exp\{2\mathbf{X}_i' \mathbf{b}_a\}}{\sum(a_i - \exp\{\mathbf{X}_i' \mathbf{b}_a\})^2}$.

Sampling parameters $\boldsymbol{\nu}$ *of common shape function*:   Recall that we reparameterized the common shape function $\mathbf{S}(t, \boldsymbol{\beta})$ with new nondecreasing coefficients $\boldsymbol{\nu}$ from equation (3). We update $\boldsymbol{\nu}$ one parameter at a time using a MH scan with transition kernels based on conditional proposals $q(\nu_j^{old}, \nu_j^{new} | \boldsymbol{\nu}_{\setminus j})$, $j = 1, ..., K$. Defining $\Omega_\nu = Cov(\boldsymbol{\nu})^{-1}$, the logarithm of the target conditional posterior density is

$$log\{P(\nu_j \mid \mathbf{Y}, \boldsymbol{\nu}_{\setminus j}, \mathbf{a}, \boldsymbol{\phi})\} \propto -\frac{1}{2} \boldsymbol{\nu}' \Omega_\nu \boldsymbol{\nu} + \sum_{ij} [Y_i(t_j) log\{\mathcal{S}_B(\mu_i(t, \phi_i))' \boldsymbol{\beta}(\boldsymbol{\nu})\} - \mathcal{S}_B(\mu_i(t, \phi_i))' \boldsymbol{\beta}(\boldsymbol{\nu})\}].$$

11

Given a fixed modal pivot ν^*, the conditional posterior support of ν_j is $[l(\nu_j), u(\nu_j)]$, with $l(\nu_j) = \max(0; \nu_{j-1})$ and $u(\nu_j) = \min(\nu_{j+1}; 2\nu^*)$, $j = 2, \ldots, K-1$. Furthermore, for $j = K$, $l(\nu_K) = \min(\nu_{K-1}, \nu^*/2)$ and $u(\nu_K) = \infty$. We fix $\nu_1 = 0$, corresponding to the assumption of no marijuana use at time t_1. We consider independent Gaussian proposals with support defined by $(l(\nu_j), u(\nu_j))$ and recalibrate the scale of the transition kernel at burn-in to achieve acceptance rate between 35% and 65%.

3.3 Model Interpretation and Inference

Given baseline covariate information $\mathbf{x} = (x_1, ..., x_p)'$, the mean function $\mathbf{S}\{\mu(t), \boldsymbol{\beta}\}$ is well defined for any $t \in [t_1, t_n]$. If we focus on the expected intensity of criminal behavior, the marginal expected count at time t can be written as:

$$E\{\mathbf{Y}(t) \mid \mathbf{x}\} = \exp\{\mathbf{x}'\mathbf{b}_a\}\, \mathbf{S}\{(t + \mathbf{x}'\mathbf{b}_\phi), \boldsymbol{\beta}\}, \tag{15}$$

allowing us to describe average trajectories of offending conditional on covariate values.

Posterior predictions of individual trajectories can be obtained conditioning on individual-specific amplitudes and time transformation functions:

$$E\{\mathbf{Y}_i(t) \mid \mu(t) = \mu_i(t); \mathbf{x}_i\} = a(\mathbf{x}_i)\mathbf{S}\{\mu_i(t, \phi_i; \mathbf{x}_i), \boldsymbol{\beta}\}, \quad t \in [t_1, t_n]. \tag{16}$$

Such model-based individual predicted trajectories are of considerable interest to criminologists for describing and explaining the development of crime and deviance over the life-course. For example, Bushway, Sweeten and Nieuwbeerta (2009) discuss and compare ways to identify "early-starters" and "desistors" by examining individual predictions from other longitudinal data analysis approaches. In contrast to other methods, however, our model allows naturally for examining marginal covariate effects on two key features of the age-crime distribution – criminal propensity and the timing of criminal careers – across all individuals in the sample.

Given MCMC draws from the posterior distribution of model parameters $\boldsymbol{\theta}$ and a fine grid of time points in $T = [t_1, t_n]$, we obtain pointwise summaries of curves given by equations (15, 16) and pointwise $100(1 - \alpha)\%$ HPD intervals using the method described by Chen and Shao (1999).

We find it convenient to include an intercept term in the model by letting the first column of the design matrix $\mathbf{X}$ to be the column of 1s. Thus, if marginal effects of covariates on the expected intensity of criminal behavior are of interest, one can examine the marginal expected count at time t, conditional on the identity time transformation for the natural age-crime curve:

$$E\{\mathbf{Y}(t) \mid \mu(t) = t; \mathbf{x}\} = \exp\{\mathbf{x}'\mathbf{b}_a\}\, \mathbf{S}(t, \boldsymbol{\beta}), \quad t \in [t_1, t_n]. \tag{17}$$

Given that $x_1 = 1$ by convention, we rewrite the expectation in (15) as

$$E\{\mathbf{Y}(t) \mid \mu(t) = t; \mathbf{x}\} = \mathbf{S}(t, \boldsymbol{\beta}) \exp\{b_{a1}\} \prod_{\kappa=2}^{p} \exp\{x_\kappa b_{a\kappa}\}.$$
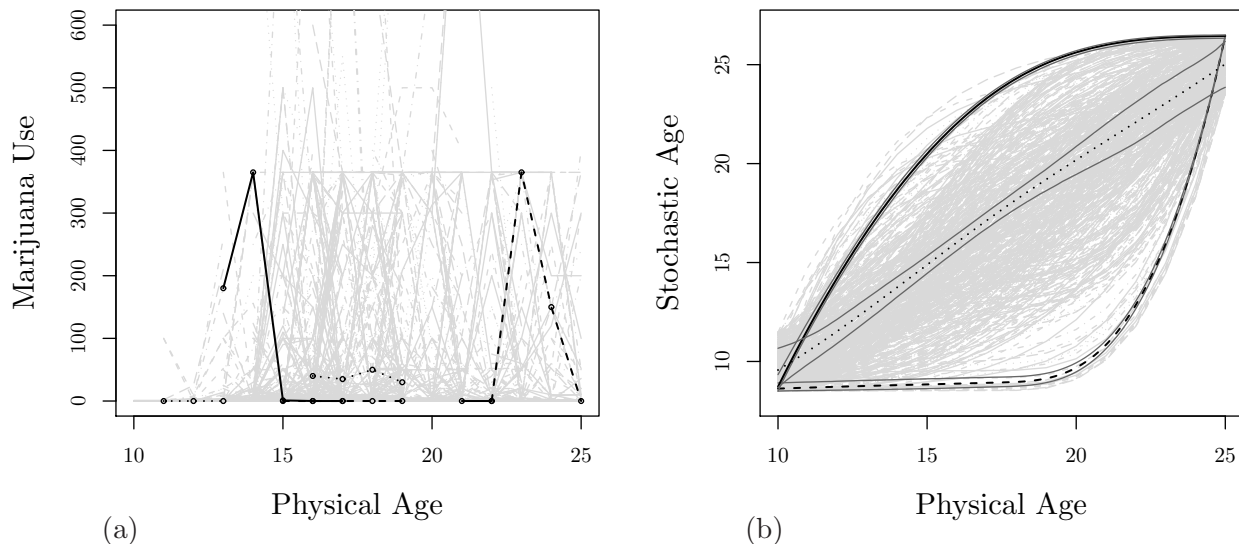
12

Figure 2: Drug Use (Marijuana). *Panel (a)*: Yearly count for the use of marijuana for 3 subjects exhibiting different timing of marijuana use. *Panel (b)*: Subject-specific time scale and associated 95% credible bands for the same subjects highlighted in panel (a).

The intensity of offending for the baseline combination of covariates is then $\exp\{b_{a1}\}$, and a multiplicative effect on intensity associated with a unit increase in covariate x_κ is $\exp\{b_{a\kappa}\}$, $\kappa = 2, \ldots, p$, all else being equal.

If the marginal effects of covariates on the expected timing of criminal behavior are of interest, we can examine the mean trajectory of offending over time t, substituting the identity transformation for the amplitude:

$$E\{\mathbf{Y}(t) \mid a = 1; \mathbf{x}\} = \mathbf{S}(t, \boldsymbol{\beta}) \circ (t + \mathbf{x}'\mathbf{b}_\phi). \tag{18}$$

As before, given that $x_1 = 1$, we rewrite (18) as

$$E\{\mathbf{Y}(t) \mid a = 1; \mathbf{x}\} = \mathbf{S}(t, \boldsymbol{\beta}) \circ \left(t + b_{\phi 1} + \sum_{\kappa=2}^{p} x_k b_{\phi k} \right).$$

The mean age-crime trajectory for the baseline combination of covariates is $\mathbf{S}(t + b_{\phi 1}, \boldsymbol{\beta})$, and an additive phase effect associated with a unit increase in covariate x_κ is $b_{\phi \kappa}$, $\kappa = 2, \ldots, p$, all else being equal. These phase effects can be interpreted as shifts in the timing of criminal careers. Thus, positive coefficients $b_{\phi \kappa}$ indicate an earlier participation in crime on average.
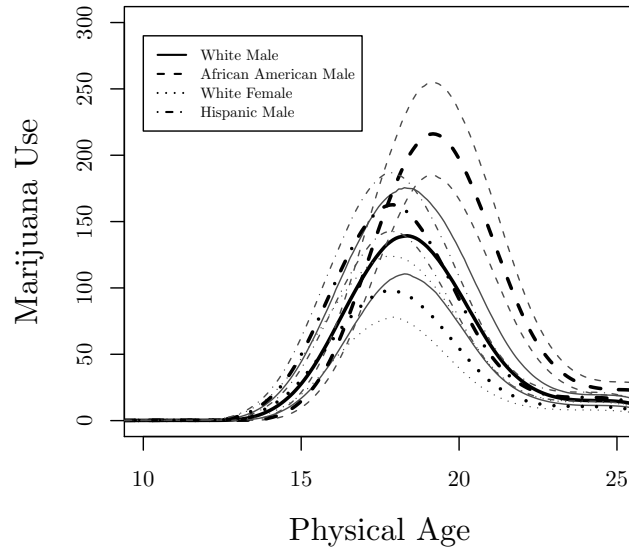
13

Figure 3: Drug use (marijuana) mean population curves for some race and gender categories.

4 CASE STUDY

We restrict our analysis to marijuana users who had at least four, not necessarily consecutive, observations during the course of the study.[3] We define marijuana users as those who reported smoking marijuana at least once. After removing 867 non-users and 22 marijuana users who had fewer than four observed time-points, we are left with a subset of 588 marijuana offenders for analysis. Our inferences are based on 15,000 (thinned by 20) samples from the posterior distribution, after discarding a conservative 50,000 iterations for burn–in. We assessed convergence using the R package BOA (Bayesian Output Analysis; Smith and Brian 2005).

Longitudinal observations of marijuana use are reported in Figure 1, panel (a). A few observations indicating marijuana smoking more than 500 times/year have been cut off for ease of visualization. In this figure, the solid black superimposed curve is the overall smoothed mean. This summary does not resemble any of the individual trajectories as it smears over the variability in both timing and frequency of drug use.

We fit the model introduced in Section 2 using 11 basis functions for the shape function $\mathbf{S}\{t, \boldsymbol{\beta}\}$, defined on the extended time interval $[t_1 - \Delta, t_n + \Delta]$. This choice was made to minimize the posterior predictive loss introduced in (14) (Gelfand and Ghosh 1998). Furthermore, we consider 5

---

[3]The inclusion of observations with a shorter time series would not affect the population estimates; however, posterior inference on subjects with fewer then 4 records can be misleading due to weak identifiability of the subject level parameters.

basis functions for the individual random time transformations. The misalignment window Δ can be interpreted as the maximal size of a linear shift. A natural constraint for the size of Δ is given by the half width of the time domain $(t_n - t_1)/2$, but more stringent values may be justified in order to avoid degeneracies in the time transformation functions. In our application, we choose a more conservative $\Delta = 1.5$.

We place relatively diffuse $\mathcal{G}(0.1, \text{rate} = 10)$ priors on the shape precision $1/\sigma_\beta^2$ and $\mathcal{G}(0.1, \text{rate} = 1)$ on the time transformation precision $1/\sigma_\phi^2$. The amplitude-phase prior covariance Σ_b is assigned a proper Inverse Wishart prior $IW(12, 100\, I_{10})$. Finally we complete the model specifying a prior distribution for the coefficient of variation $1/\sqrt{b_0}$. As we are considering a sample of users we require $b_0 > 1$ and define a shifted gamma prior $\mathcal{G}(\lambda_a = 1.1, \lambda_b = 0.1)$ on $(b_0 - 1)$. A proper scale-informative prior on $\mathbf{b}_a$ is used to define 'soft' identifiability constraints. The constraint $b_0 > 1$ assures that the prior mode of a_i is greater than 0.

Figure 1, panel (b), shows observed frequencies of drug use that have been normalized by removing individual differences in timing and in amplitude for all individuals. We obtained these quantities by evaluating observed frequencies on the inverse transformed time scale $E(\mu_i^{-1}(t, \boldsymbol{\phi_i}) \mid \mathbf{Y})$, including the phase shift, and dividing by the expected amplitude of offense $E(a_i \mid \mathbf{Y})$. We superimpose normalized observed counts with a smoothed functional convex average $\mathbf{S}(t, \boldsymbol{\beta})$. This figure shows a typical pattern of marijuana use for an average individual in our sample of marijuana smokers from the most disadvantaged areas of Denver. The average individual starts smoking marijuana during adolescence, continues with higher intensity through college age, and then drops off marijuana smoking after reaching twenty. We observe a thin left tail of occasional use before the peak years and a thicker right tail of occasional use after the peak years. This pattern is generally consistent with the claims of Hirschi and Gottfredson (1983) and previous empirical research on the age-crime curve.

Figure 1, panel (c), shows posterior median estimates of the individual amplitude parameters on the log scale, with corresponding 95% highest posterior density (HPD) credible intervals. We observe that variability in amplitude is an important source of variation in marijuana smoking trajectories. Estimated log amplitude parameters are from about -4 to 4; the range of these estimates is much wider than the width of a typical 95% credible interval. A log amplitude equal to 0 corresponds approximately to a marijuana smoking trajectory at the level of the overall functional convex mean (solid black line in panel (b)), with the estimated peak smoking at about 33 times per year. In comparison, the average log amplitude of 1.5 for white males corresponds to marijuana smoking frequency that is $\exp(1.5) \approx 4.5$ times higher than the structural mean, with the estimated peak at about $33 * 4.5 = 148$ times per year.

Panel (d), Figure 1, shows the posterior expected estimates of individual time transformation functions, indicating that phase variability is another important source of variation in self-reported

marijuana use. Figure 2 illustrates how large differences in the timing of marijuana use are reflected in individual-specific estimates of time-transformation functions. Panel (a) highlights individual trajectories for an early user (solid line), an average user (dotted line) and a late user (dashed line). The corresponding estimated time transformation functions and associated 95% credible bands are reported in panel (b). We note that for the average user the time transformation is close to identity, i.e., the stochastic age of this person is similar to his or her physical age. The late user's stochastic age is kept frozen in time until his or her physical age of about 18; this individual then goes through the marijuana use period much faster than an average marijuana smoker in our sample. The early user exhibits a similarly quick period of marijuana smoking but at much earlier ages.

We use two approaches to investigate how overall intensity and timing of drug use depends on race and gender. First, in Table 1, we report posterior estimates of the amplitude and phase regression parameters for the covariates in the model (indicators for 'female', as well as 'African American', 'Latino' and 'other' case categories; we use 'white male' as the baseline category). We find that, overall in our sample, females use marijuana less frequently. For example, white females use marijuana with an overall frequency that is about $\exp(-0.334) \approx 0.72$ to 1, when compared to their male counterparts. African Americans males from disadvantaged areas of Denver, on the other hand, seem to be using marijuana more frequently (1.55 to 1), when compared to their Caucasian counterparts. We find the same significant differences in terms of timing of drug use, with females starting to use marijuana on average about 5.5 months earlier when compared to white males and African Americans starting to use marijuana on average about 10 months later. [4] We did not estimate gender by race interactions in our model as some subgroups only included a small (< 25) number of subjects. Second, in Figure 3, we report predicted mean population curves of marijuana use for some race and gender subgroups, obtained with equation (15). The predicted mean curves in Figure 3 complement our findings from Table 1, illustrating differences in marijuana use by race and gender.

Examining individual data, we find that the estimated expected crime trajectories fit the observed data well. In Figure 4, we report expected frequencies of marijuana use for an illustrative subsample of six subjects in the DYS, obtained with equation (16). Black dots and solid lines indicate the observed and expected counts of yearly marijuana use, respectively, and the dashed lines represent pointwise 90% highest posterior density prediction intervals. This figure shows how our model formulation appears to provide a remarkably close fit to individual profiles. Based on information that is shared across subjects, this modeling framework allows for individual-specific predictions for all time points within time interval T, including those points where the individual did not have observations. Wider prediction bands illustrate higher uncertainty in model predictions where no subject-specific data is available.

---

[4] National survey data covering the years of our survey shows that observed race and gender differences in age at
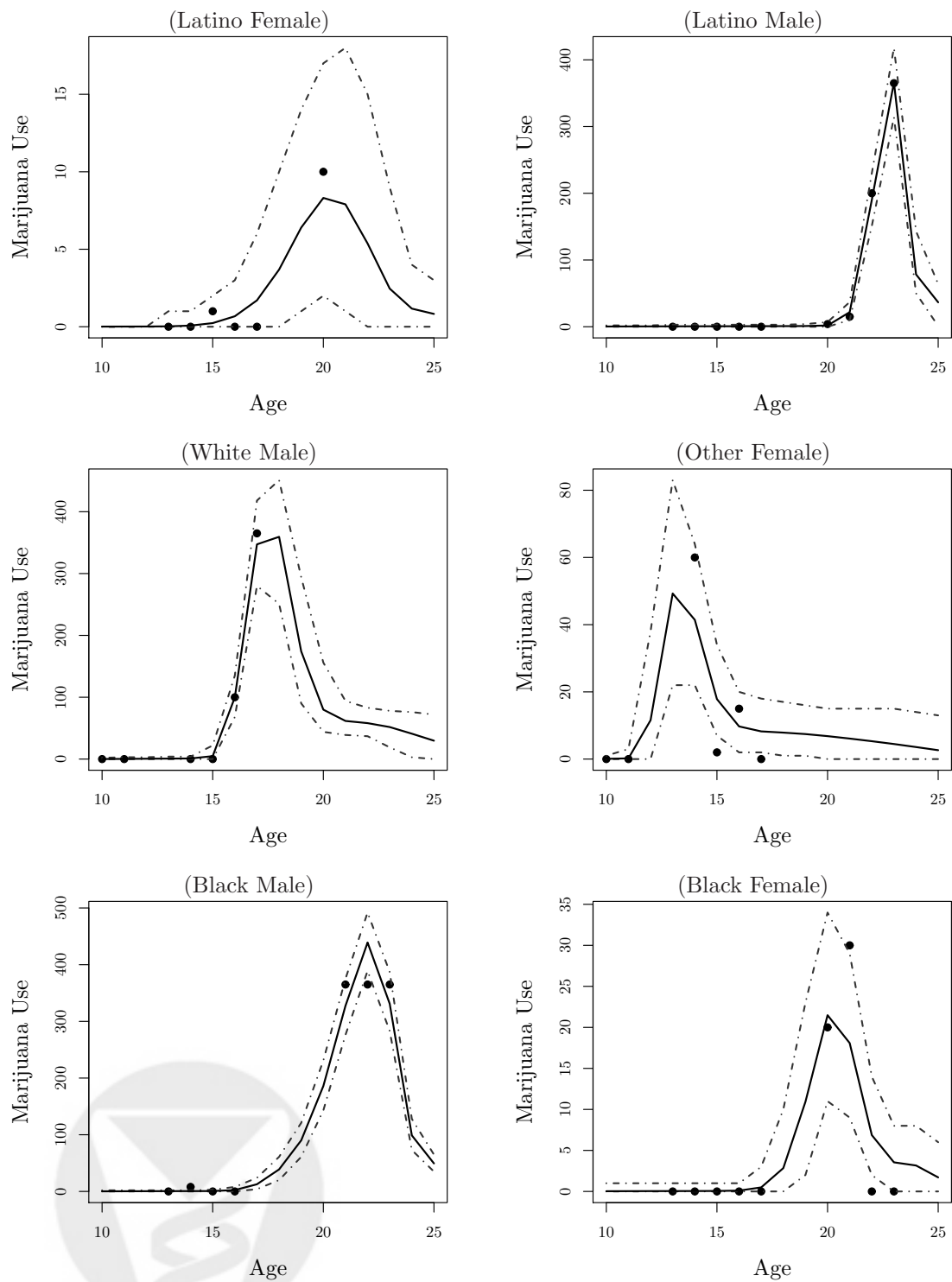
16

Figure 4: Drug Use (Marijuana). Lifetime marijuana use profiles for six random subjects from different race and gender categories. For each profile, the solid line represents the median posterior expected count and the dot–dashed lines represent the associated 95% pointwise prediction intervals.

17

	Amplitude		Phase (years)	
Effect	$E(b_a\|Y)$	95% C.I.	$E(b_\phi\|Y)$	95% C.I.
Baseline				
White Males	1.502	[1.240, 1.842]	-0.741	[−1.169, −0.312]
Main Effects				
Female	-0.334	[−0.501, −0.167]*	0.461	[0.213, 0.713]*
Latino	0.154	[−0.107, 0.396]	0.420	[−0.017, 0.864]
African American	0.436	[0.175, 0.678]*	-0.864	[−1.325, −0.385]*
Other	0.394	[0.068, 0.715]*	0.172	[−0.383, 0.718]

Table 1: Amplitude and Phase Regression Parameters. (*Stars denote 95% C.I. which do not cover the value zero.*)

In sum, our model provides two important features in modeling individual crime trajectories. First, it allows us to estimate a common age-crime curve and fit individual trajectories as departures from that mean curve. Second, it allows us to disentangle variation in individual crime trajectories due to differences in level (amplitude) and timing (phase shift) of offenses. This approach gives us new insights into offending trajectories for marijuana use. We find the shape of the estimated common age-crime curve to be consistent with prior empirical research. Like previous research, we find race and gender differences in levels of marijuana smoking; unlike previous research, our finding controls for the common age-crime curve and phase variability. We find little support for the claim that group differences in the shape of the age-crime curve are merely due to group differences in rates of offending. For example, Hirschi and Gottfredson (1983) argue that racial differences in age-at-onset are merely due to racial differences in rates of offending. By contrast, we find significant race and gender differences in timing as indicated by the shift of marijuana use trajectories while controlling for race and gender differences in amplitude. Moreover, we do not detect any correlation, a posteriori, between amplitude and shift parameters, which would be expected if differences in timing were merely due to differences in amplitude.

5 DISCUSSION

In this article we propose a generalized warping regression method for the analysis of longitudinal crime data. We model subject-specific expected patterns of offenses as arising from a natural unimodal crime curve, evaluated over a random individual-specific time transformation scale, and with a random individual-specific amplitude.

The analysis we present in this paper has several limitations. First, we chose to ignore the issue of heaping in the distribution of self-reported counts (Wang and Heitjan 2008). A more realistic

onset of marijuana use are small and change signs across survey years (Gfroerer et al. 2002).

sampling model would take into account tendencies to report smoking marijuana with a rounded frequency (most commonly, in our example, 365). Second, observations in the data set are right-censored at age 25 and may be left-censored at the age when individuals enter the survey. Third, the model may be subject to criticism of overparameterization. Although these issues are important and deserve further attention, we believe that the hierarchical Bayesian formulation together with a flexible mean function in our model may already provide reasonable adjustments to heaping biases and issues of censored data. We alleviate overparametrization concerns via structural modeling constraints, shrinkage priors and model selection via posterior predictive checks.

Individual trajectories predicted by our model appear to capture the full nature of the individual longitudinal observations of marijuana use, given that the observations conform to the unimodality assumption. This result contrasts with the fit of individual predicted trajectories based on polynomial trajectory models with random effects (Raudenbush and Chan 1993) and group-based models (Nagin and Land 1993), which have been criticized for failing to fit some atypical individual profiles such as those of "late starters" (Bushway et al. 2009).

Group-based trajectory models of Nagin and Land (1993) aim to identify latent classes of crime trajectories, following from a theoretical taxonomy developed by Moffitt (1993) who described a population of offenders as a mixture of two different groups, adolescence-limited and life-course persistent offenders. An extension by Roeder et al. 1999 included time-stable covariates as risk factors influencing individual membership in latent group trajectories. Muthén and Shedden (1999) developed a model similar to that of Nagin and Land (1993) that incorporates within-group random effects. In the past decade, there has been an explosion in the number of publications on group-based data analyses of longitudinal crime data (Sampson and Laub 2005b). However, the merits of group-based modeling and the interpretation of the meaning of groups is still a point of contention and controversy in the criminology literature (Sampson and Laub 2003, Raudenbush 2005; Nagin and Tremblay 2005d; Nagin and Tremblay 2005a; Sampson and Laub 2005a; Nagin and Tremblay 2005b; Nagin and Tremblay 2005c). In addition, research findings using these methods can be overwhelmingly driven by tremendous variability in within-age behavioral amplitude common in crime data. For example, group trajectories of offending reported in studies often identify several groups that exhibit similar shapes, differing only in level (e.g., Nagin and Tremblay 1999, Sampson and Laub 2003, Bushway et al. 2009). Finally, we note that growth curve mixture models are only able to *implicitly* define other quantities of particular interest to criminologists, such as age at onset and level of offending, whether at the group or individual level. Typically, researchers simply eyeball the curves for quantities of interest. This makes it difficult to carry out model-based inference about covariate effects on these important quantities.

Our model formulation *explicitly* accounts for variability in individual amplitude (level of offending) and timing, and includes covariate effects on these quantities. While the interpretation

of covariate effects on amplitude is straightforward, covariate effects on timing are incorporated in the shift parameter and should be interpreted as such. Technically, the shift parameter in our model reveals differences in timing that are conditional on the individual curves being transformed to a common shape. The observed summary statistics may not correspond to the differences in timing/shift revealed by the model because observed differences in timing may be confounded with observed differences in shape. In the marijuana example, however, race and gender differences in timing found with our model were similar in magnitude and significance to observed race and gender differences in age at maximal use. Thus, in a regression of age at maximal use on race and gender, females were estimated to reach the age at maximal use significantly earlier than males and blacks significantly later than whites. If we were interested in examining race and gender differences in other specific timing features of the age-crime curve, such as age at onset and age at desistance, the model would have to be extended to incorporate time-stable covariates as time-varying effects. While this approach would induce higher flexibility in the modeling of longitudinal counts, the interpretation can be more challenging. On the other hand, the appealing features of direct inference about functional changes in the mean structure suggest that extending our model to incorporate time-varying effects of covariates may be worthwhile.

Another direction for a possible extension of our model relates to the foregoing discussion about identifying distinct groups of offending trajectories. Ramsay and Silverman (2002) carried out a functional principal component analysis on arrest data in an attempt to confirm or disprove the existence of distinct groups of criminal offenders and found "no real evidence of strong groups." A group mixture reformulation of our model may allow for the classification of different features of the age-crime curve, from intensity of offense, to typical offending ages, to different shapes of the natural crime curve.

Another important question in criminology is to understand how criminal behavior changes in association with time-dependent covariate information. For example, do individual departures from a natural crime curve correspond to changes in life course transitions, such as high school dropout, entrance into college, parenthood, and entrance into the labor force? To address these questions, one needs to incorporate time-dependent covariates. This could be achieved, for example, by integrating our warping regression model with the historical functional linear model of Malfait and Ramsay (2003).

The above potential extensions to our model would capitalize on the strengths of our general approach to modeling crime trajectories. These strengths include (1) an individual-level model that is both flexible and realistic, and allows for differences in amplitude and timing of offenses; (2) a model that incorporates criminologists' specifications of an invariant age-crime curve with individual departures based on individual differences in crime propensity and life situations; and (3) an estimation procedure that borrows information between the population average age-crime

curve and individual departures from that curve.

References

Blumstein, A. and J. Cohen (1987). Characterizing criminal careers. *Science 237*(4818), 985–991.

Brumback, L. C. and M. J. Lindstrom (2004). Self modeling with flexible, random time transformations. *Biometrics 60*(2), 461–470.

Bushway, S., G. Sweeten, and P. Nieuwbeerta (2009). Measuring long term individual trajectories of offending using multiple methods measuring long term individual trajectories of offending using multiple methods. *Journal of Quantitative Criminology 25*(3), 259–286.

Chen, M. H. and Q. M. Shao (1999). Monte carlo estimation of bayesian credible and hpd intervals. *Journal of Computational and Graphical Statistics 8*(9), 69–92.

De Boor, C. (1978). *A Practical Guide to Splines*. Berlin: Springer-Verlag.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with *B*-splines and penalties. *Statistical Science 11*, 89–102.

Elliott, D. S., D. Huizinga, and S. S. Ageton (1985). *Explaining Delinquency and Drug Use*. Beverly Hills: Sage Publications.

Esbensen, F.-A. and D. Huizinga (1990). Community structure and drug use: From a social disorganization perspective. *Justice Quarterly 7*, 691–709.

Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall Ltd.

Gelfand, A. E. and S. K. Ghosh (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika 85*, 1–11.

Gelfand, A. E. and A. F. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Gervini, D. and T. Gasser (2004). Self-modelling warping functions. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 66*(4), 959–971.

Gfroerer, J., L. Wu, and M. Penne (2002). Initiation of marijuana use: Trends, patterns, and implications. Technical Report Analytic Series: A-17, DHHS Publication No. SMA 02-3711, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, Rockville, MD.

Glueck, S. and E. Glueck (1950). *Unraveling Juvenile Delinquency*. New York: The Commonwealth Fund.

Gottfredson, M. and T. Hirschi (1990). *A General Theory of Crime.* Stanford, CA : Stanford University Press.

Harris, K. M., F. Florey, J. Tabor, P. S. Bearman, J. Jones, and J. R. Udry (2003). *The National Longitudinal Study of Adolescent Health: Research Design.* WWW document. URL: http://www.cpc.unc.edu/projects/addhealth/design.

Hastings, W. K. (1970). Monte carlo sampling using markov chains and their applications. *Biometrika 57*, 97–109.

Hirschi, T. and M. R. Gottfredson (1983). Age and the explanation of crime. *American Journal of Sociology 89*, 552–548.

Lang, S. and A. Brezger (2004). Bayesian p–splines. *Journal of Computational and Graphical Statistics 13*(1), 183–212.

Malfait, N. and J. O. Ramsay (2003). The historical functional linear model. *The Canadian Journal of Statistics 31*(2), 115–128.

Matsueda, R. L., D. A. Kreager, and D. Huizinga (2006). Deterring delinquents: A rational choice model of theft and violence. *American Sociological Association 71*, 95–122.

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

Muthén, S. and K. Shedden (1999). Finite mixture modeling with mixture outcomes using the em algorithm. *Biometrics 55*, 463–469.

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

Nagin, D. S. and R. E. Tremblay (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development 70*, 1181–1196.

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

Nagin, D. S. and R. E. Tremblay (2005b). From seduction to passion: A response to sampson and laub. *Criminology 43*(4), 873–904.

Nagin, D. S. and R. E. Tremblay (2005c). What has been learned from group-based trajectory modeling? examples from physical aggression and other problem behavior. *Annals of the American Academy of Political and Social Science 602*(82-117).

Nagin, D. S. and R. E. Tremblay (2005d). What has been learned from group-based trajectory modeling? examples from physical aggression and other problem behaviors. *The Annals of the American Academy of Political Social Science 602*, 82–117.

Office of National Drug Control Policy (2008). *Marijuana Sourcebook 2008–Marijuana: The Greatest Cause of Illegal Drug Abuse.* Office of National Drug Control Policy.

Peña, J. (1997). *B−splines and optimal stability. Mathematics of Computation 66*, 1555–1560.

Ramsay, J. O. and X. Li (1998). Curve registration. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 60*, 351–363.

Ramsay, J. O. and B. W. Silverman (2002). *Applied Functional Analysis: Methods and Case Studies.* New York: Springer-Verlag.

Raudenbush, S. (2005). How do we study 'what happens next? *Annals of the American Academy of Political and Social Science 602*, 131–144.

Raudenbush, S. and W. Chan (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Clinical Consulting Psychology 61*(6), 941–951.

Rice, J. A. and B. W. Silverman (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B: Methodological 53*, 233–243.

Roberts, O. R. and J. S. Rosenthal (2001). Optimal scaling of various metropolis–hastings algorithms. *Statistical Science 16*(4), 351–367.

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

Sampson, R. and J. Laub (2003). Life-course desisters? trajectories of crime among delinquent boys followed to age 70. *Criminology 43*(4), 905–914.

Sampson, R. and J. H. Laub (2005a). Deductions of method: rejoinder to nagin and tremblay's "developmental trajectory groups: fact or fiction?". *Criminology 43*(4), 905–914.

Sampson, R. J. and J. H. Laub (1993). *Crime in the Making: Pathways and Turning Points Through the Life Course.* Cambridge, MA: Harvard University Press.

Sampson, R. J. and J. H. Laub (2005b). A life-course view of the development of crime. *Annals of the American Academy of Political and Social Science 602*, 12–45.

Schumaker, L. L. (1981). *Spline Functions. Basic Theory.* New York : John Wiley & Sons.

Shi, M., R. E. Weiss, and J. M. G. Taylor (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics 45*, 151–163.

Smith, A. and J. Brian (2005). http://www.public-health.uiowa.edu/boa/BOA.pdf.

Telesca, D. and L. Y. T. Inoue (2008). Bayesian hierarchical curve registration. *Journal of the American Statistical Association 103 (481)*, 328–339.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics 1994* (22), 1701–1728.

Wang, H. and D. Heitjan (2008). Modeling heaping in self-reported cigarette counts. *Statistics in Medicine 27*(19), 3789–3804.

Wolfgang, M., R. Figlio, and T. Sellin (1973). *Delinquency in a Birth Cohort.* Chicago: University of Chicago Press.

Yao, F., H. J. Müller, and J. L. Wang (2005). Functional data analysis of sparse longitudinal data. *JASA 100*(470), 577–590.

Appendix: Full Conditional Distributions

Shape smoothing parameters σ_β^2: For ease of notation we define the precision of the jth fixed effect $h_\beta = 1/\sigma_\beta^2$. From Section 2.2 we have that $P(h_\beta \mid a_\beta, b_\beta) \propto h_\beta^{(a_\beta-1)} exp\{-b_\beta h_\beta\}$. Given the fixed effect coefficients $\boldsymbol{\nu}$, the conditional posterior density of h_β can be written as $P(h_\beta \mid Y, \boldsymbol{\theta}_{\backslash h_\beta}) \propto P(\boldsymbol{\nu} \mid h_\beta)P(h_\beta \mid a_\beta, b_\beta)$, where $P(\boldsymbol{\nu} \mid h_\beta) \propto h_\beta^{K/2} exp\{-h_\beta/2\,\boldsymbol{\nu}'\Omega\boldsymbol{\nu}\}$ and Ω is the banded concentration structure arising from a second order random walk, described in Section 2.2. It is trivial to show that: $P(h_\beta \mid \mathbf{Y}, \boldsymbol{\theta}_{\backslash h_\beta}) \propto h_\beta^{(a_\beta^*-1)} exp\{-b_\beta^* h_\beta\}$, corresponding to the density function of a Gamma random variable with shape $a_\beta^* = a_\beta + K/2$, and rate $b_\beta^* = b_\beta + \boldsymbol{\nu}'\Omega\boldsymbol{\nu}/2$.

Time transformation smoothing parameter σ_ϕ^2: For ease of notation we define the time transformation precision $h_\phi = 1/\sigma_\phi^2$. From Section 2.4 we have $P(h_\phi \mid a_\phi, b_\phi) \propto h_\phi^{a_\phi-1} exp\{-b_\phi\,h_\phi\}$. Given the matrix of time transformation coefficients $\boldsymbol{\phi}$, the conditional posterior density of h_ϕ can be written as $P(h_\phi \mid \mathbf{Y}, \boldsymbol{\theta}_{\backslash h_\phi}) \propto P(\boldsymbol{\phi} \mid \boldsymbol{\gamma}, h_\phi)P(h_\phi \mid a_\phi, b_\phi)$, where $P(\boldsymbol{\phi} \mid \boldsymbol{\gamma}, h_\phi) \propto \prod_{i=1}^N h_\phi^{Q/2} exp\{-h_\phi/2\,(\boldsymbol{\phi}_i - \boldsymbol{\gamma}_i)'\Xi(\boldsymbol{\phi}_i - \boldsymbol{\gamma}_i)\}$ and Ξ is a banded concentration structure arising from a first order random walk, given $(\phi_{i0} - \gamma_{i0}) = 0$ for all $i = 1, ..., N$. It easily follows that $P(h_\phi \mid \mathbf{Y}, \boldsymbol{\theta}_{\backslash h_\phi}) \propto h_\phi^{a_\phi^*-1} exp\{-b_\phi^*\,h_\phi\}$, a Gamma random variable with shape $a_\phi^* = a_\phi + N \times Q/2$, and rate $b_\phi^* = b_\phi + 1/2 \times \sum_{i=1}^N (\boldsymbol{\phi}_i - \boldsymbol{\gamma}_i)'\Xi(\boldsymbol{\phi}_i - \boldsymbol{\gamma}_i)$.

Phase Regression parameters $\mathbf{b}_\phi$: Using standard conjugate analysis it is easy to show that $P(\mathbf{b}_\phi \mid \boldsymbol{\phi}) =_d N_p(m^*, V^*)$, where $V^* = (\sum_{i=1}^N \sum_{q=1}^Q \mathbf{X}_i'\mathbf{X}_i + \Omega)^{-1}$ and $m^* = V^*(\sum_i \sum_q X_i\tilde{\phi}_{iq})$. Where $\tilde{\phi}_{iq}$ is the q-th element of $(\phi_i - \Upsilon)(\Xi)^{1/2}$ and Ξ is defined as before.

Amplitude-Phase Covariance Σ_b: using standard conjugate analysis it is easy to show that the conditional posterior distribution $P(\Sigma_b \mid \mathbf{b}_a, \mathbf{b}_\phi, v_0, \Phi_0) =_d IW(p + v_0, (S_b + \Phi_0))$, where $S_b = \sum_{k=1}^p (b_{ak}, b_{\phi k})(b_{ak}, b_{\phi k})'$.