Journal of the American Statistical Association

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

Donatello Telesca [a] , Elena A. Erosheva [b] , Derek A. Kreager [c] & Ross L. Matsueda [d]

[a] Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, CA

[b] Departments of Statistics, Department of Social Work, Center for Statistics and the Social Sciences, University of Washington, Seattle, WA

[c] Department of Sociology, Pennsylvania State University, Abington, PA

[d] Department of Sociology, University of Washington, Seattle, WA
Accepted author version posted online: 14 Aug 2012.Published online: 21 Dec 2012.

PLEASE SCROLL DOWN FOR ARTICLE

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

Donatello TELESCA, Elena A. EROSHEVA, Derek A. KREAGER, and Ross L. MATSUEDA

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 model for longitudinal crime trajectories and allows for a rich array of inferences of interest to criminologists and drug abuse researchers. This article has supplementary materials online.

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, Figlio, and Sellin (1973), who examined age at onset, length of criminal careers, and patterns of desistance. Recent research has turned to model-based approaches, such as growth curve or trajectory mixture models, 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 find-

ings are often driven by variability in within-age behavioral amplitude (Gottfredson and Hirschi 1990). More importantly, both growth curve and trajectory mixture models tend to ignore subject-specific variation in the timing of crime. From a statistical perspective, this leads to inconsistent estimates of time-varying population parameters (Brumback and Lindstrom 2004; Gervini and Gasser 2004). From a substantive perspective, ignoring phase variability leads to population estimates of age–crime curves that are not representative of typical individual trajectories as they smear over local features of subject-specific time series. In particular, it has been observed that growth curve and trajectory mixture models are typically unable to capture individual variability in the decline of offending (Bushway, Sweeten, and Nieuwbeerta 2009).

In this article, we propose an alternative approach for analyzing longitudinal crime data. We draw on the growing research literature in criminology on the age–crime curve. Nearly all researchers agree that, overall, the population age–crime curve is unimodal, rising precipitously from age 7 (the age of culpability) until the peak years—between ages 13 and 21 depending on the offense—and then slowly declining throughout the remaining years of the life span. Some researchers have emphasized, on substantive and theoretical grounds, substantial heterogeneity in age–crime curves across groups of individuals. For example, Blumstein, Farrington, and Moitra (1985) posited groups of desisters, persisters, and nonoffenders, and Moffitt (1993) contrasted groups of life-course-persistent chronic offenders versus adolescence-limited normal offenders. By contrast, Hirschi and Gottfredson (1983) injected controversy into criminology by arguing that a single age–crime curve underlies all criminal offenses and is invariant across all social groups and throughout history. Although they claimed invariance in the basic shape

of the age–crime curve, they also acknowledged the presence of modest 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. These substantive arguments naturally lend themselves to using an appropriately constrained functional data analysis approach for modeling longitudinal crime data.

We posit a unimodal population age–crime curve and develop a Poisson warping regression model that 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.

Although the assumption of unimodality is relatively weak, given observed distributions of crime and drug use, we will compare the predictive performance of models with unimodality against a model with an unconstrained shape function. Moreover, our functional data analysis approach is sufficiently flexible to model not only modest individual departures from a population age–crime curve, but also substantial individual departures. Our unimodal constraint assumes a monotonic incline in offending to the peak age followed by a monotonic decline, which is consistent with observed empirical age–crime curves, including the relatively flat individual curves of chronic offending or low offending posited by Blumstein, Farrington, and Moitra (1985). Thus, the model can treat the question of invariance of the population age–crime curve as an empirical question.

We are not the first to take a functional data analytic point of view toward longitudinal crime data. Ramsay and Silverman (2002) carried out a functional principal component analysis on a landmark dataset 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, Weiss, and Taylor (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, Müller, and Wang (2005) discussed 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,

Huizinga, and Ageton 1985; 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 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 semiparametric 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.

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, socioeconomic status, 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 1528 respondents [for details, see Matsueda, Kreager, and Huizinga (2006), and Esbensen and Huizinga (1990)]. Of these respondents, 1459 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 to 1992 and from 1995 to 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
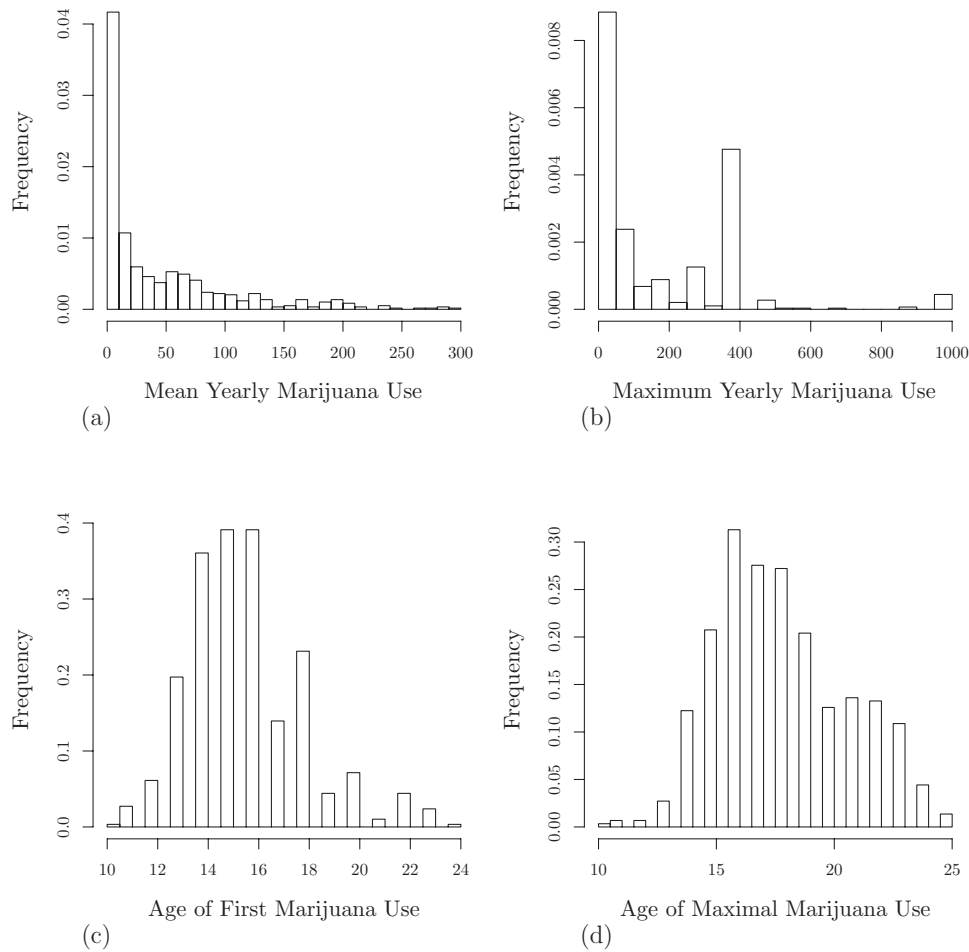
Figure 1. Exploratory summaries. Panels (a, b): Histograms of mean and maximal use of marijuana . Panels (c, d): Histograms of age of first and maximal marijuana use.

who had between 4 and 9 longitudinal observations on marijuana use for each individual. The resulting dataset had a mean of 7.39 (SD = 1.37) observations per subject. Exploratory summaries associated with variability in timing and frequency of marijuana use are reported in Figure 1. In panels (a) and (b), we provide histograms for mean and maximal marijuana use, respectively. 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, with the 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. Ages at first use and at the maximal marijuana use vary substantially as indicated by the histograms in panels (c) and (d) of Figure 1. These summaries illustrate large individual variability in amplitude (level) and phase (timing) among individual trajectories of marijuana smoking.

The remainder of this article is organized as follows. In Section 2, we introduce a hierarchical model for the semiparametric analysis of longitudinal count data. We discuss Markov chain

Monte Carlo (MCMC) estimation and inference in Section 3 and analyze lifetime data on marijuana use from the DYS in Section 4. We conclude with a discussion of our contributions and possible model extensions in 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 $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ij}, \ldots, Y_{in})'$ denote an observed vector of offenses for individual i over a discrete sampling design $\mathbf{t} = (t_1, \ldots, t_j, \ldots, t_n)$. 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, for example, 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

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

---

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

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

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

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) generalizes substantive arguments of Hirschi and Gottfredson (1983) about the age–crime curve. It starts by assuming 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 that 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 $\beta_j \geq 0,\ j = 1, \ldots, K$. 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 exhibit only one possible sign change (Schumaker 1981, theorem 4.76). We combine the unimodality and positivity requirements via the following reparameterization of the shape coefficients $\boldsymbol{\beta}$:

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

where the new coefficients $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_K)'$ are nondecreasing, that is, $0 = \nu_1 \leq \cdots \leq \nu_K < 2\nu^*$, and $\nu^* \geq 0$ 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 kth element of $\boldsymbol{\nu}$ as

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

The variance parameter σ_β^2 can then be interpreted as a smoothing parameter shrinking the shape function toward a piecewise 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:

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

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

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

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

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

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:

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

where

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

---

[2] We discuss unimodal smoothing in more detail in Section 2 of the online supplementary materials. In our analysis, choosing $\nu^* = \sqrt{\bar{Y}}$ 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.

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) imposed 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,

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

The specific form for $\pi(\mathbf{b}_a)$ is described in the next section, 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, opportunities, 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 by 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 $i = 1, \ldots, N$,

$$(\phi_{iq} - \gamma_{iq}) = (\phi_{i(q-1)} - \gamma_{i(q-1)}) + \eta_{iq},$$

with

$$\eta_{iq} \sim \mathcal{N}(0, \sigma_\phi^2) I\{\mathcal{M}\}, \ q = 1, \ldots, Q, \qquad (10)$$

where $\phi_{i0} = \Upsilon_0 = 0$. Here, $\mathcal{M}$ defines a set of random cuts such that $\eta_{iq} - \eta_{i(q-1)} > \Upsilon_{q-1} - \Upsilon_q$, $q = 1, \ldots, Q$, where $|\eta_{i1}| \leq \Delta$ and $|\eta_{iQ}| \leq \Delta$. The variance parameter σ_ϕ^2 is a smoothing parameter that controls the amount of shrinkage of individual time-transformation functions toward 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 2.2 and 2.4, we rewrite the expected number of offenses for subject i at time t_j from Equation (2) as

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

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 $\mathbf{a} = (a_1, \ldots, a_N)'$, and time-transformation coefficients $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1', \ldots, \boldsymbol{\phi}_N')'$ is then

$$\ell(\boldsymbol{\beta}, \mathbf{a}, \boldsymbol{\Phi} \mid \mathbf{Y})$$
$$\propto \sum_{i=1}^{N} \sum_{j=1}^{n} [\, y_{ij} \log\{\lambda_i(t_j, \boldsymbol{\beta}, a_i, \boldsymbol{\phi}_i)\} - \lambda_i(t_j, \boldsymbol{\beta}, a_i, \boldsymbol{\phi}_i]. \quad (13)$$

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 (Eilers and Marx 1996; Lang and Brezger 2004). The shrinkage prior as in Equation (4) automatically shrinks the fixed-effect functions toward 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

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

(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.

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 equally space the knots for the time-transformation functions.

In the following section, we describe an MCMC algorithm for posterior simulation based on fixed numbers of spline basis functions for the 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 $\mathbf{a} = (a_1,\ldots,a_N)'$, an $(N \times Q)$ matrix of individual-specific time-transformation coefficients $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1',\ldots,\boldsymbol{\phi}_N')'$, 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 a 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 online supplement.

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 an MH sampler with appropriately scaled transition kernels $q(\boldsymbol{\phi}^{\text{old}}, \boldsymbol{\phi}^{\text{new}})$. Given that $\phi_{iq} < \phi_{i(q+1)}$ ($\forall\ i = 1,\ldots,N$, $q=1,\ldots,Q$), we consider proposal densities of the form $q(\phi_{iq}^{\text{old}}, \phi_{iq}^{\text{new}}) = \mathcal{N}(\phi_{iq}^{\text{old}}, s_{iq}^2)I\{\mathcal{M}\}$, 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 $\mathbf{a} = (a_1,\ldots,a_N)'$ 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 the Newton–Raphson method. Defining $\Sigma_{b_a}^* = (H_{b_a} \mid_{\hat{\mathbf{b}}_a})^{-1}$, we obtain the transition kernel on the basis of the overrelaxed proposal $q(\mathbf{b}_a^{\text{old}}, \mathbf{b}_a^{\text{new}}) = 2\hat{\mathbf{b}}_a - \mathbf{b}_a^{\text{old}} + MVt(0, \tau_{b_a}\Sigma_{b_a}^*, \eta)$. 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 an 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 $\boldsymbol{v}$ 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 an MH scan with transition kernels based on conditional proposals $q(\nu_j^{\text{old}}, \nu_j^{\text{new}}|\boldsymbol{\nu}_{\setminus j})$, $j = 1, \ldots, K$. Defining $\Omega_\nu = \text{cov}(\boldsymbol{\nu})^{-1}$, the logarithm of the target conditional posterior density is

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

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 an acceptance rate between 35% and 65%.

3.3 Model Interpretation and Inference

Given baseline covariate information $\mathbf{x} = (x_1, \ldots, x_p)'$, 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

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

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:

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

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) discussed and compared ways to identify "early-starters" and "desisters" 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) and (16) and pointwise $100(1 - \alpha)\%$ highest posterior density (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:

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

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

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

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:

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

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

$$E\{\mathbf{Y}(t) \mid a = 1; \mathbf{x}\} = \mathbf{S}(t, \boldsymbol{\beta}) \circ \left(t + b_{\phi1} + \sum_{\kappa=2}^{p} x_k b_{\phi k}\right).$$

The mean age–crime trajectory for the baseline combination of covariates is $\mathbf{S}(t + b_{\phi1}, \boldsymbol{\beta})$, 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.

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 nonusers 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 2(a). A few observations indicating marijuana smoking more than 500 times per 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 nine basis functions (five interior knots) 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

---

[3]The inclusion of observations with a shorter time series would not affect the population estimates; however, posterior inference on subjects with fewer then four records can be misleading due to weak identifiability of the subject-level parameters.
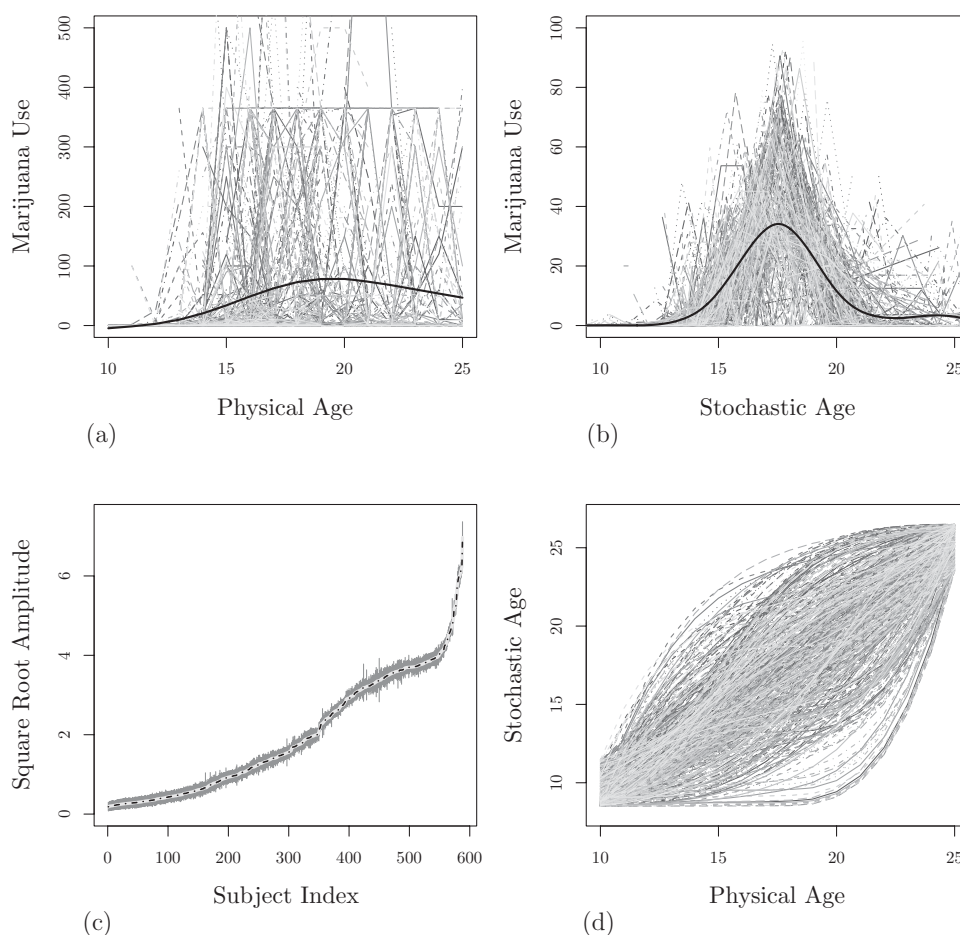
Figure 2. Drug use (marijuana). Panel (a): Yearly count for the use of marijuana for 588 subjects from the DYS. A solid black line depicts the structural mean function. Panel (b): Aligned normalized trajectories. In black we report the overall functional convex average $\mathbf{S}(t, \boldsymbol{\beta})$. Panel (c): Subject-specific posterior square root amplitude with associated 95% credible intervals. Panel (d): Subject-specific time scale, characterized by the expected posterior time-transformation functions.

in (14) (Gelfand and Ghosh 1998). Furthermore, we consider five basis functions (one interior knot) for the individual random time transformations (see Figure 7). 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 justi-

fied 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
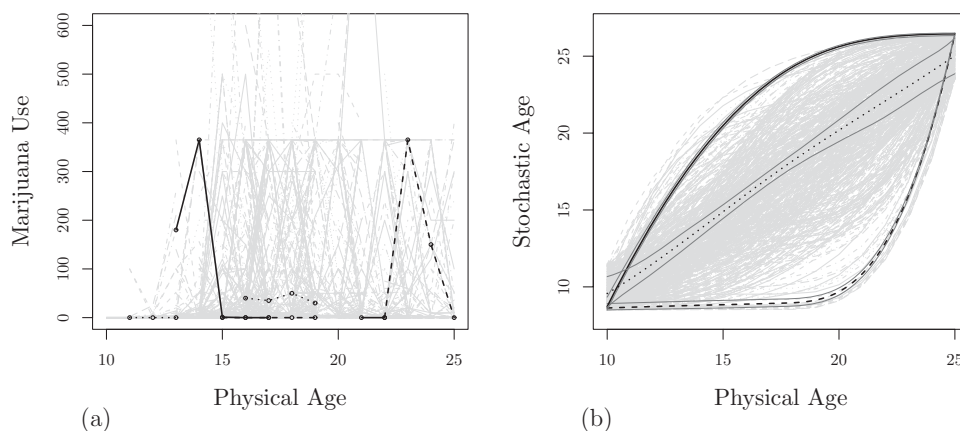

Figure 3. Drug use (marijuana). Panel (a): Yearly count for the use of marijuana for three subjects exhibiting different timing of marijuana use. Panel (b): Subject-specific time scale and associated 95% credible bands for the same subjects highlighted in panel (a).

Table 1. Amplitude and phase regression parameters [stars denote 95% credible interval (CI) that do not cover the value zero]

Effect	Amplitude		Phase (years)	
	$E(b_a\|Y)$	95% CI	$E(b_\phi\|Y)$	95% CI
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]

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 2(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 20. 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 2(c), shows posterior median estimates of the individual amplitude parameters on the log scale, with corresponding 95% 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 Figure 2(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) ≈ 4.5 times higher than the structural mean, with the estimated peak at about 33 × 4.5 = 148 times per year.

Figure 2(d), 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 3 illustrates how large differences in the timing of marijuana use are reflected in individual-specific estimates of time-transformation functions. Figure 3(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 Figure 3(b). We note that for the average user, the time transformation is close to identity, that is, 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) ≈ 0.72 to 1, when compared with their male counterparts. African American males from disadvantaged areas of Denver, on the other hand, seem to be using marijuana more frequently (1.55 to 1), when compared with 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 with 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 4, 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 4 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 5, we report expected frequencies of marijuana use for a subsample of six subjects in the DYS chosen from the race and gender categories to illustrate a representative range of individuals' amplitude and timing, obtained with Equation (16). Black dots and solid lines indicate the observed and expected

---

[4]National survey data covering the years of our survey show that observed race and gender differences in age at onset of marijuana use are small and change signs across survey years (Gfroerer, Wu, and Penne 2002).
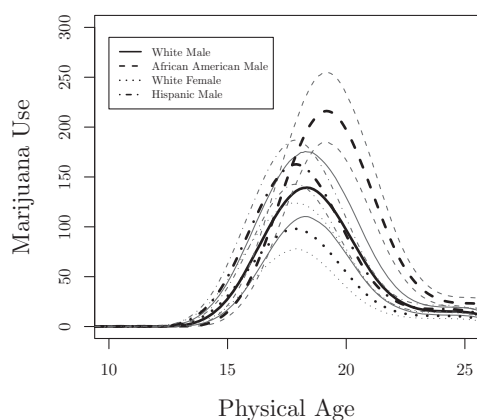
Figure 4. Drug use (marijuana) mean population curves for some race and gender categories.

counts of yearly marijuana use, respectively, and the dashed lines represent pointwise 90% HPD 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 the 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 are available. We carried out a formal assessment of the goodness of fit by comparing posterior predictive distributions with corresponding summary statistics obtained empirically from the data. In Figure 6(a)–(b), we report 90% posterior predictive intervals for the individual yearly average and maximal marijuana smoking levels, as well as the corresponding summary statistics. The model provides us with an excellent coverage
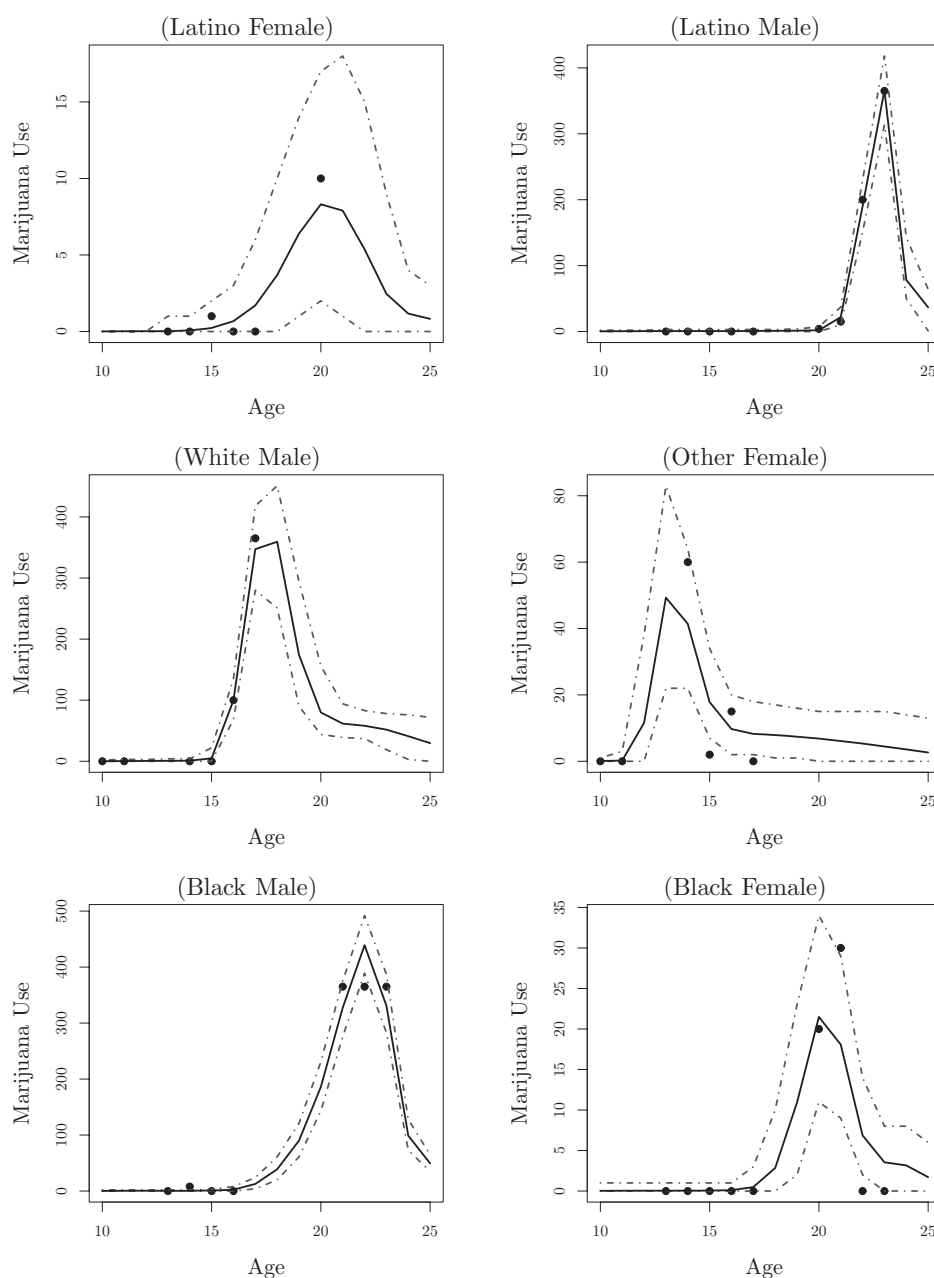

Figure 5. Drug use (marijuana). Lifetime marijuana use profiles for six random subjects from different race and gender categories. For each profile, the solid line represents the median posterior expected count and the dot-dashed lines represent the associated 90% pointwise prediction intervals.
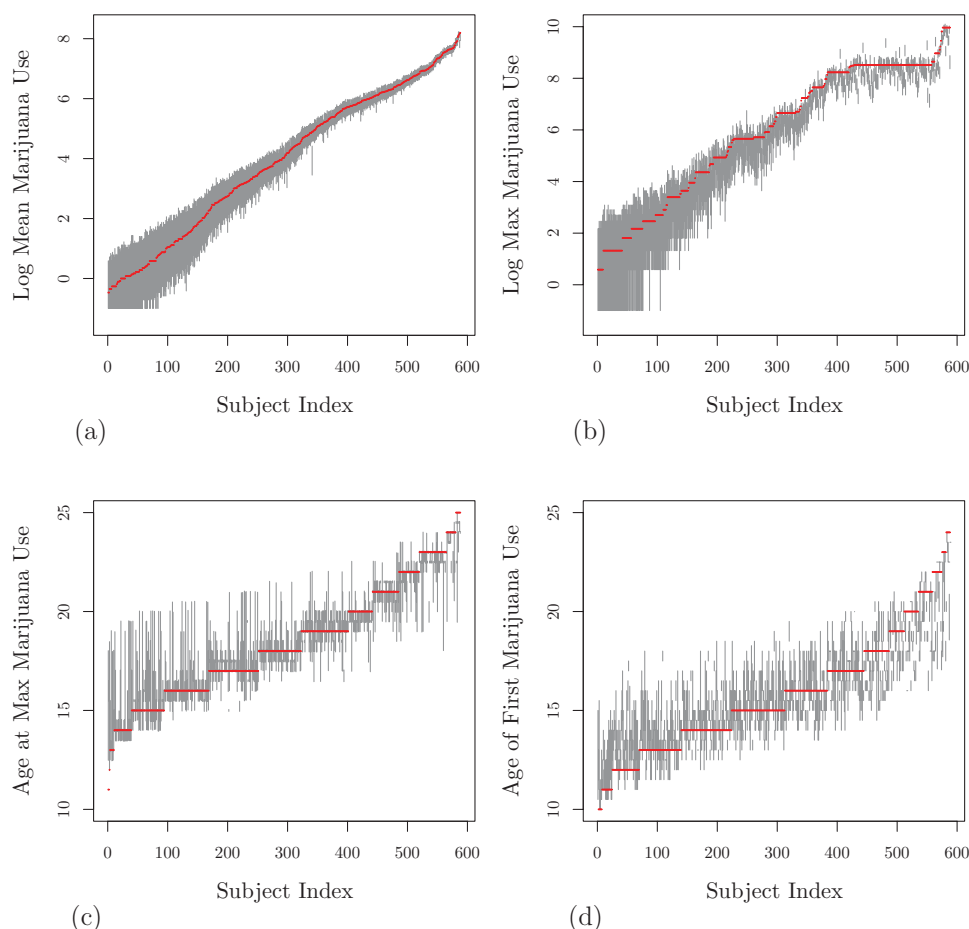
Figure 6. Posterior predictive checks. Panels (a, b): (90%) Posterior predictive intervals versus summary statistics associated with average marijuana use and maximal level of marijuana use. Panels (c, d): (90%) Posterior predictive intervals and data points corresponding to the age at maximal marijuana use and the age of first use. The online version of this figure is in color.

for the observed average levels of marijuana use (94%) and maximal frequencies of marijuana use (86%). Figure 6(c)–(d), report 90% posterior predictive intervals for the individual age at maximal marijuana use and for the age at first use, as well as the corresponding observed ages. The posterior predictive intervals follow the observed ages fairly well across the time interval. Note that because the available data do not allow us to obtain exact empirical estimates of the timing summaries, that is, we could only say that the first marijuana smoking event happened during a particular year but are unable to distinguish when it happened exactly, examining numerical coverage values for age-related summaries would not be appropriate.

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 modeling trajectories of 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 individual variability in phase and amplitude. We find little support for the

claims that differences in the shape of the age–crime curve are merely due to differences in rates of offending; for example, Hirschi and Gottfredson (1983) argued that racial differences in age at onset are merely due to racial differences in rates of offending. In 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 individual 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 age–crime curve, evaluated over a random individual-specific time-transformation scale, and with a random individual-specific amplitude.

The analysis we present in this article 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 sampling model would take into account tendencies to report smoking marijuana with a rounded frequency (most
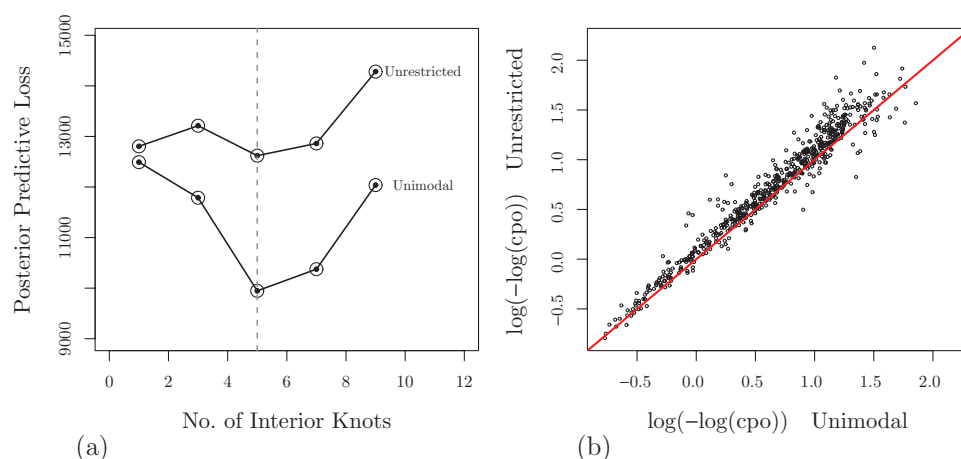
Figure 7. Unimodal versus unrestricted age–crime curve. Panel (a): Posterior predictive loss analysis comparing unimodal and unrestricted models fitted specifying a varying number of interior knots. Panel (b): Conditional predictive ordinate analysis comparing unrestricted and unimodal models with five interior knots. The online version of this figure is in color.

commonly, in our example, 365). Second, observations in the dataset are right censored at age 25 and may be left censored at the age when individuals enter the survey. Third, the flexible specification of the mean function 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 over-parameterization concerns via structural modeling constraints, shrinkage priors, and model selection via posterior predictive checks.

The assumption of a unimodal population age–crime curve plays a key role in the proposed analysis framework and provides us with robustness to alternative model specifications involving varying flexibility of the population age–crime curve. Furthermore, this structural restriction is associated with improved predictive performance in this case study, when compared with the models without the unimodality restriction. A formal comparison of models fitted with and without the uni-modality restriction is summarized in Figure 7. Figure 7(a) compares the posterior predictive loss associated with the two models, evaluated for a varying number of spline basis knots. In our implementation, the posterior predictive loss is based on deviance calculations as defined in (14). Our derivation follows the argument by Gelfand and Ghosh (1998), which describe the posterior predictive loss as a penalized goodness-of-fit measure, where the penalty term is associated with the magnitude of the predictive variance. For all levels of complexity, the posterior predictive loss associated with the unimodal curve model is lower than that associated with the unrestricted mean model. We obtain a detailed comparison of the two best models by calculating the conditional predictive ordinate (CPO) $p(y_i \mid y_{(i)})$ for each individual in the sample (Pettit 1990; Geisser 1993). This summary is comparable with the classical cross-validation procedure. Here, we evaluate the predictive density of y_i given observations $y_{(i)}$, excluding counts from subject i. Figure 7(b), shows that for a clear majority of subjects, predictions based on the unimodal curve model outperform those obtained without imposing structural constraints (i.e., have higher CPO

and smaller values of the transformed CPO plotted in Figure 7(b)). These results indicate that the unimodality assumption is appropriate for modeling marijuana smoking trajectories. Further empirical work is needed to confirm that the unimodality assumption is appropriate for modeling behavioral trajectories of drug use and crime in general. This assumption would indeed be unnecessary if longer, less sparse time series of crime behavior would be available. In which case, one may also question the assumption of a unique population age–crime curve. We however maintain that further modeling extensions may not be warranted in the analysis of the DYS data explored in this article; certainly not without substantial methodological developments.

Our method of Bayesian hierarchical unimodal curve registration presents a novel approach to modeling data on behavioral trajectories that relies on a common underlying unimodal mean curve and models individual trajectories by specifying individual-specific deviations in phase and amplitude from the mean curve. Popular current approaches to the analysis of longitudinal behavioral trajectories most often rely on growth curve models (Raudenbush and Chan 1993) and mixture trajectory models (e.g., Nagin and Land 1993). Similar to our approach, growth curve models also assume that all individuals share the same mean trajectory of crime. However, the mean trajectory is most commonly specified by a polynomial, and individual deviations from the mean are characterized by individual random intercept, slope, and quadratic terms. Group-based trajectory models (Nagin and Land 1993; Roeder, Lynch, and Nagin 1999) 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. Growth curve mixture models of Muthén and Shedden (1999) also assume latent classes of trajectories but, in addition, incorporate within-group polynomial random effects, similar to the growth curve models. In contrast to our curve registration approach, the existing approaches for longitudinal analysis of behavioral outcomes can only *implicitly* account for individual differences in timing, whether at the group or individual level. Because time-varying population estimates can only be consistent if individual trajectories have been properly

aligned to a standardized time scale (Brumback and Lindstrom 2004; Gervini and Gasser 2004; Telesca and Inoue 2008), substantive implications of failing to account for phase variability in the data can be enormous. For example, without adjusting for phase variability, estimates of individual quantities related to timing, such as age at peak use, may not be reliable and findings of multiple latent groups of individual trajectories may be spurious. We believe that our approach has great potential to provide researchers with new insights for modeling drug and alcohol use, as well as other behavioral outcomes.

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 the 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, and to different shapes of the natural crime curve. Moreover, it would relax the assumption of a single population age–crime curve and attempt to fit models of multiple group trajectories, such as Blumstein, Farrington, and Moitra (1985) and Moffitt (1993).

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.

SUPPLEMENTARY MATERIALS

Supplementary materials to this article include a Monte Carlo study assessing coverage and estimation performance of the proposed method as well as a more detailed discussion of unimodal smoothing. Also included is a web appendix summarizing full conditional distributions used in the posterior simulation.

[Received August 2008. Revised May 2012.]

REFERENCES

Blumstein, A., and Cohen, J. (1987), "Characterizing Criminal Careers," *Science*, 237, 985–991. [1427]

Blumstein, A., Farrington, D. F., and Moitra, S. (1985), "Delinquency Careers: Innocents, Desisters, and Persisters," *Crime and Justice*, 6, 187–219. [1427,1428,1439]

Brumback, L. C., and Lindstrom, M. J. (2004), "Self Modeling With Flexible, Random Time Transformations," *Biometrics*, 60, 461–470. [1427,1428,1431,1439]

Bushway, S., Sweeten, G., and Nieuwbeerta, P. (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, 259–286. [1427,1433]

Chen, M. H., and Shao, Q. M. (1999), "Monte Carlo Estimation of Bayesian Credible and HPD Intervals," *Journal of Computational and Graphical Statistics*, 8, 69–92. [1433]

De Boor, C. (1978), *A Practical Guide to Splines*, Berlin: Springer-Verlag. [1430]

Eilers, P. H. C., and Marx, B. D. (1996), "Flexible Smoothing With *B*-Splines and Penalties," *Statistical Science*, 11, 89–102. [1431]

Elliott, D. S., Huizinga, D., and Ageton, S. S. (1985), *Explaining Delinquency and Drug Use*, Beverley Hills, CA: Sage Publications. [1428]

Esbensen, F.-A., and Huizinga, D. (1990), "Community Structure and Drug Use: From a Social Disorganization Perspective," *Justice Quarterly*, 7, 691–709. [1428]

Gamerman, D. (1997), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, London: Chapman & Hall. [1432]

Geisser, S. (1993), *Predictive Inference: An Introduction*, London: Chapman & Hall. [1438]

Gelfand, A. E., and Ghosh, S. K. (1998), "Model Choice: A Minimum Posterior Predictive Loss Approach," *Biometrika*, 85, 1–11. [1432,1434,1438]

Gelfand, A. E., and Smith, A. F. (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409. [1432]

Gervini, D., and Gasser, T. (2004), "Self-Modelling Warping Functions," *Journal of the Royal Statistical Society,* Series B, 66, 959–971. [1427,1428,1431,1439]

Gfroerer, J., Wu, L., and Penne, M. (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. [9]

Glueck, S., and Glueck, E. (1950), *Unraveling Juvenile Delinquency*, New York: The Commonwealth Fund. [1428]

Gottfredson, M., and Hirschi, T. (1990), *A General Theory of Crime*, Stanford, CA: Stanford University Press. [1427,1430]

Harris, K. M., Florey, F., Tabor, J., Bearman, P. S., Jones, J., and Udry, J. R. (2003), "The National Longitudinal Study of Adolescent Health: Research Design," WWW document [online]. Available at *http://www.cpc.unc.edu/projects/addhealth/design*. [1428]

Hastings, W. K. (1970), "Monte Carlo Sampling Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109. [1432]

Hirschi, T., and Gottfredson, M. R. (1983), "Age and the Explanation of Crime," *American Journal of Sociology*, 89, 552–548. [1427,1428,1430,1435,1437]

Lang, S., and Brezger, A. (2004), "Bayesian *p*–Splines," *Journal of Computational and Graphical Statistics*, 13, 183–212. [1431]

Malfait, N., and Ramsay, J. O. (2003), "The Historical Functional Linear Model," *The Canadian Journal of Statistics*, 31, 115–128. [1439]

Matsueda, R. L., Kreager, D. A., and Huizinga, D. (2006), "Deterring Delinquents: A Rational Choice Model of Theft and Violence," *American Sociological Association*, 71, 95–122. [1428]

Moffitt, T. E. (1993), "Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy," *Psychological Review*, 100, 674–701. [1427,1438,1439]

Muthén, S., and Shedden, K. (1999), "Finite Mixture Modeling With Mixture Outcomes Using the EM Algorithm," *Biometrics*, 55, 463–469. [1427,1438]

Nagin, D. S., and Land, K. C. (1993), "Age, Criminal Careers, and Population Heterogeneity: Specification and Estimation of a Nonparametric, Mixed Poisson Model," *Criminology*, 31, 327–362. [1427,1438]

Office of National Drug Control Policy (2008), *Marijuana Sourcebook 2008—Marijuana: The Greatest Cause of Illegal Drug Abuse*, Washington, DC: Author. [1428]

Peña, J. (1997), "*B*-Splines and Optimal Stability," *Mathematics of Computation*, 66, 1555–1560. [1430]

Pettit, L. (1990), "The Conditional Predictive Ordinate for the Normal Distribution," *Journal of the Royal Statistical Society,* Series B, 52, 175–184. [1438]

Ramsay, J. O., and Li, X. (1998), "Curve Registration," *Journal of the Royal Statistical Society,* Series B, 60, 351–363. [1428,1431]

Ramsay, J. O., and Silverman, B. W. (2002), *Applied Functional Analysis: Methods and Case Studies*, New York: Springer-Verlag. [1428,1439]

Raudenbush, S., and Chan, W. (1993), "Application of a Hierarchical Linear Model to the Study of Adolescent Deviance in an Overlapping Cohort Design," *Journal of Clinical Consulting Psychology*, 61, 941–951. [1427,1438]

Rice, J. A., and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data Are Curves," *Journal of the Royal Statistical Society,* Series B, 53, 233–243. [1428]

Roberts, O. R., and Rosenthal, J. S. (2001), "Optimal Scaling of Various Metropolis–Hastings Algorithms," *Statistical Science*, 16, 351–367. [1432]

Roeder, K., Lynch, K. G., and Nagin, D. S. (1999), "Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology," *Journal of the American Statistical Association*, 94, 766–776. [1438]

Sampson, R. J., and Laub, J. H. (1993), *Crime in the Making: Pathways and Turning Points Through the Life Course*, Cambridge, MA: Harvard University Press. [1427,1428]

Schumaker, L. L. (1981), *Spline Functions. Basic Theory*, New York: Wiley. [1430]

Shi, M., Weiss, R. E., and Taylor, J. M. G. (1996), "An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome Using Flexible Random Curves," *Applied Statistics*, 45, 151–163. [1428]

Smith, A., and Brian, J. (2005), *Bayesian Output Analysis Program*. Available at *http://www.public-health.uiowa.edu/boa/BOA.pdf*. [1433]

Telesca, D., and Inoue, L. Y. T. (2008), "Bayesian Hierarchical Curve Registration," *Journal of the American Statistical Association*, 103, 328–339. [1428,1430,1431,1439]

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1728. [1432]

Wang, H., and Heitjan, D. (2008), "Modeling Heaping in Self-Reported Cigarette Counts," *Statistics in Medicine*, 27, 3789–3804. [1437]

Wolfgang, M., Figlio, R., and Sellin, T. (1973), *Delinquency in a Birth Cohort*, Chicago, IL: University of Chicago Press. [1427]

Yao, F., Müller, H. J., and Wang, J. L. (2005), "Functional Data Analysis of Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100 (470), 577–590. [1428]