Sociological Methods & Research
http://smr.sagepub.com

Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them

Bobby L. Jones and Daniel S. Nagin

Sociological Methods Research 2007; 35; 542
DOI: 10.1177/0049124106292364

Sociological Methods & Research
Volume 35 Number 4
May 2007 542-571
© 2007 Sage Publications
10.1177/0049124106292364
http://smr.sagepub.com
hosted at
http://online.sagepub.com

Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them

Bobby L. Jones
Daniel S. Nagin
Carnegie Mellon, Pittsburgh, PA

This article is a follow-up to Jones, Nagin, and Roeder (2001), which described an SAS procedure for estimating group-based trajectory models. Group-based trajectory is a specialized application of finite mixture modeling and is designed to identify clusters of individuals following similar progressions of some behavior or outcome over age or time. This article has two purposes. One is to summarize extensions of the methodology and of the SAS procedure that have been developed since Jones et al. The other is to illustrate how group-based trajectory modeling lends itself to presentation of findings in the form of easily understood graphical and tabular data summaries.

Keywords: *developmental trajectories; Proc Traj; finite mixture models*

This article is a follow-up to Jones, Nagin, and Roeder (2001), which described an SAS procedure for estimating group-based trajectory models. The Proc Traj software can be downloaded from www.andrew.cmu.edu/user/bjones/index.htm free of charge. Also available for download at the Web site are documentation and a number of supporting SAS macros described in this article and the Jones et al. article.

Group-based trajectory models are designed to identify clusters of individuals following similar progressions of some behavior or outcome over age or time. Psychologists call such progressions *developmental trajectories*. We borrow this term to describe the course of any phenomenon, whether behavioral, biological, or physical. Charting and understanding

---

Authors' Note: The research was supported by generous financial support from the National Science Foundation (SES-9911370) and the National Institute of Mental Health (RO1 MH65611-01A2).

developmental trajectories is among the most fundamental and empirically important research topics in the social and behavioral sciences and medicine.

As developed in Nagin (1999, 2005), the group-based trajectory model is a specialized application of finite mixture modeling. Piquero (2004) reports that more than 60 articles have been published that use group-based trajectory modeling. This article has two purposes. One is to summarize extensions of the methodology and of the SAS procedure that have been developed since Jones et al. (2001). The other is to illustrate how group-based trajectory modeling lends itself to presentation of findings in the form of easily understood graphical and tabular data summaries. Data summaries of this form have the great advantage of being accessible to nontechnical audiences and quickly comprehensible to audiences that are technically sophisticated.

Brief Overview of the Statistical Model

Suppose $Y_i = \{y_{i1} y_{i2}, y_{i3}, \ldots, y_{1T}\}$ represents the longitudinal sequence of measurements on an individual i over T periods, and $P(Y_i)$ denotes the probability of Y_i. The group-based trajectory model assumes that the population is composed of a mixture of J underlying trajectory groups such that

$$P(Y_i) = \sum_j \pi_j P^j(Y_i),$$

where $P^j(Y_i)$ is the probability of Y_i given membership in group j, and π_j is the probability of group j. The basic model also assumes that conditional on membership in group j, the random variables, y_{it}, $t = 1, 2, \ldots, T$, are independent. Thus, $P^j(Y_i) = \prod^{T} p^{jt}(y_{it})$.

The group membership probabilities, π_j, $j = 1, \ldots, J$, are not estimated directly but instead are estimated by a multinomial logit function:

$$\pi_j = e^{\theta_j} / \sum_1^J e^{\theta_j}, \tag{1}$$

where θ_1 is normalized to zero. Estimation of π_j in this fashion ensures that each such probability properly falls between 0 and 1.

The form of $p^{jt}(y_{it})$ is selected to conform to the type of data under analysis. If y_{it} is a count variable, then $p^{jt}(y_{it})$ is assumed to be the Poisson distribution or even more general zero-inflated Poisson (Lambert 1993). For psychometric scale data, $p^{jt}(y_{it})$ is assumed to follow the censored normal distribution to accommodate the possibility of clustering at the scale minimum and maximum. For binary data, $p^{jt}(y_{it})$is assumed to follow the binary logit distribution.

As in most hierarchical and latent curve models, a polynomial relationship is used to model the link between age and behavior. The Proc Traj software allows estimation of up to a fourth-order polynomial. For the Poisson-based model, it is assumed that

$$\ln(\lambda_{it}^{j}) = \beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3 + \beta_4^j Age_{it}^4, \tag{2}$$

where λ_{it}^{j} is the expected number of occurrences of the event of interest (e.g., convictions) of subject i at time t, given membership in group j, and $Age_{\ it}$ is subject i's age at time t. The model's coefficients—$\beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j$, and β_4^j—determine the shape of the trajectory. Because this vector of parameters, which we denote by β^j, is trajectory group specific, its values can vary freely across the j groups.

For the censored normal model, the linkage between age and behavior is established via a latent variable, y_{it}^{*j}, that can be thought of as measuring the potential for engaging in the behavior of interest, say physical aggression. Again, up to a fourth-order polynomial relationship is assumed between y_{it}^{*j} and age:

$$y_{it}^{*j} = \beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3 + \beta_4^j Age_{it}^4 + \varepsilon_{it}, \tag{3}$$

where ε_{it} is a disturbance assumed to be normally distributed with a zero mean and a constant standard deviation σ.

In the case of binary data, it is assumed that conditional on membership in group j, $p^{jt}(y_{it})$ follows the binary logit distribution:

$$p^{jt}(y_{it}) = \frac{e^{\beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3 + \beta_4^j Age_{it}^4}}{1 + e^{\beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3 + \beta_{4t}^j Age_{it}^4}}. \tag{4}$$

Data Sources

Two major longitudinal studies are used to illustrate the modeling extension demonstrated in this article.

Cambridge Study of Delinquent Development (the London study). This study tracked a sample of 411 British males from a working-class area of London. Data collection began in 1961-1962, when most of the boys were age 8. Criminal involvement is measured by the number of convictions for criminal offenses and is available for all individuals in the sample through age 32, with the exception of the eight individuals who died prior to this age. Between ages 10 and 32, a wealth of data was assembled on each individual's psychological makeup; family circumstances, including parental behaviors; and performance in school and work. The modeling demonstrations are based on the 403 individuals who survived to age 32. For a complete discussion of the data set, see Farrington (1990).

Montreal Longitudinal-Experimental Study of Boys (the Montreal study). The 1,037 participants in this study were part of a longitudinal study started in the spring of 1984. All teachers of kindergarten classes in the 53 schools of the lowest socioeconomic areas in Montréal (Canada) were asked to rate the behavior of each boy in their classroom. To control for cultural effects, the boys were included in the longitudinal study only if both their biological parents were born in Canada and their biological parents' mother tongue was French. Assessments were made at age 6 and annually from ages 10 to 17. Wide-ranging measurements of social and psychological function were made based on assessments by parents, teachers, and peers; self-reports by the boy himself; and administrative records from schools and the juvenile court. See Tremblay et al. (1987) for further details on this study.

Modeling and Software Extensions

Maximum likelihood is used for the estimation of the model parameters. The maximization is performed using a general quasi-Newton procedure (Dennis, Gay, and Welsch 1981; Dennis and Mei 1979). Because the procedure may find only a local maximum, multiple starting points are specified so that there is reasonable assurance that the global maximum is located. The variance-covariance matrix for the parameter estimates is obtained from the inverse observed information matrix evaluated at the maximum likelihood parameter estimates. Wald tests for parameters equal to zero are performed for the trajectory parameters and generalized logit parameters. The Bayesian information criterion (BIC) is relied on for model selection as described in Jones et al. (2001).

Extension 1: Confidence Intervals on Trajectories and Group Membership Probabilities

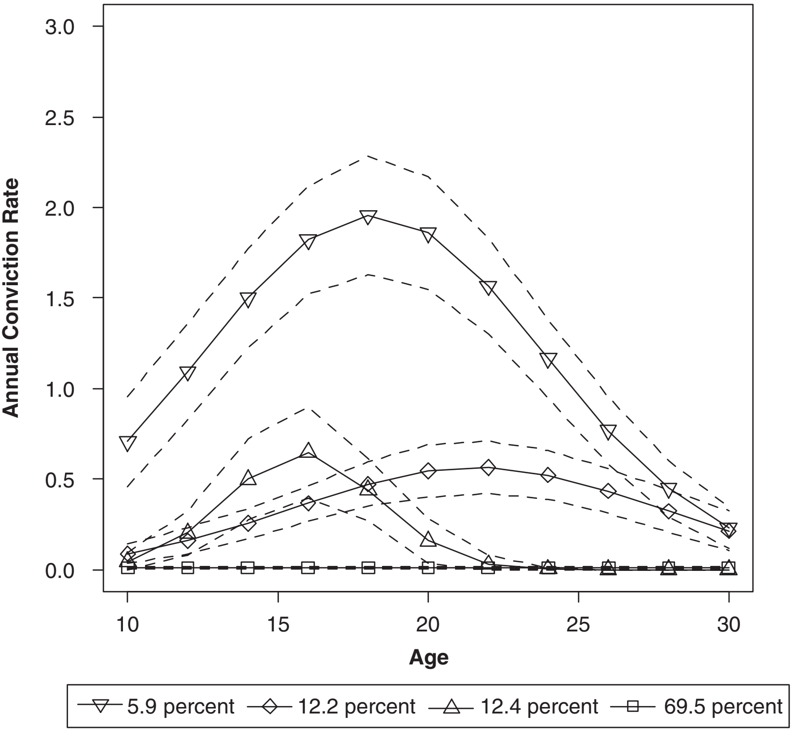
The direct products of the model estimation are the estimates of the vectors that describe each group j's trajectory, β^{j}, and the vector that is composed of the parameters that determine the group membership probabilities, θ. Also estimated is the variance-covariance matrix for the parameter estimates. However, the quantities of actual interest to the user are the group membership probabilities, as specified by equation (1), and the trajectories, as specified by equation (2), (3), or (4). Because these quantities are nonlinear transformations of the parameter estimates, estimation of their standard errors and attendant confidence intervals requires more complex calculations.

One approach to estimating the standard errors of the desired quantities is the bootstrap method (Efron 1979). We did not choose this alternative because the resampling requirements would make model estimation time prohibitive for most users. The other alternative is to approximate the standard error by a first-order Taylor series expansion (Greene 1995). We used this alternative because it is easily implemented and has no material impact on estimation time.

We illustrate this extension with an application to the London data in which we estimate a four-group Poisson-based model. Figure 1 shows the point estimates of the trajectories of λ_t^j and their accompanying 95 percent confidence intervals. The fact that the confidence intervals at most overlap for only relatively short periods indicates that the trajectories are distinct. One group is composed of individuals with a miniscule rate of offending over the entire observation period. This trajectory of negligible offending, which we call the nonoffending group, was modeled by a zero-order (i.e., constant) trajectory in age. Such a zero-order trajectory could also capture the trajectory of a constant high-rate trajectory. Probability of membership in this group is estimated at 69.5 percent, with a 95 percent confidence interval (CI) of 63.3 to 75.7 percent. The three remaining trajectory groups were specified to follow a quadratic function of age. We refer to Group 2 as the "adolescent-limited" group because following a period of active offending during adolescence, the predicted λ_t^j is negligible after age 20. The estimated probability of membership in this group is 12.4 percent (95 percent CI: 7.1 percent, 17.7 percent). Group 4 is referred to as the high-chronic group because of its high offense rate during adolescence and early adulthood and its nonnegligible rate thereafter. This trajectory characterized the behavior of an estimated 5.9 percent of the population (95 percent CI: 3.4 percent, 8.4 percent). The final quadratic trajectory, Group 3, is referred to as the low chronics because members of this group offended at a nonnegligible rate throughout the observation period. The size of this group is estimated at 12.2 percent of the sampled population (95 percent CI: 7.9 percent, 16.5 percent).

The following statements fit this model and produce the graph shown in Figure 1:

Figure 1
Annual Conviction Rate Versus Age: Four-Group Poisson Model



```
proc traj data=london_data out=london_out outstat=crimstat
outplot=crimplot ci95m;
var y1-y11;
indep t1-t11;
model zip;
ngroups 4;
start -5 -.5 0 0 0 0 0 .5 0 0 70 10 10 10;
order 0 2 2 2;
```

*%**trajplotnew*** (crimplot,crimstat, ‘Conviction Rate vs. Age’, ‘Four Group-Poisson Model’,’Lambda’,’Age/10’)

Extension 2: Dual-Trajectory Models

The dual model was designed to analyze the developmental course of two distinct but related outcomes (Nagin and Tremblay 2001). The model can be used to analyze connections between the developmental trajectories of two outcomes that are evolving contemporaneously (e.g., depression and alcohol use) or that evolve over different time periods (e.g., prosocial behavior in childhood and school achievement in adolescence).

The three key outputs of the dual model are as follows: (1) the trajectory groups for both measurement series, (2) the probability of membership in each identified trajectory group, and (3) conditional probabilities linking membership across the trajectory groups of the two respective behaviors.

Loeber (1991) has argued that covert behaviors in childhood, such as opposition, are linked to another form of covert behavior in adolescence, property delinquency. We thus illustrated the dual model with an analysis of the linkage of opposition from ages 6 to 13 with property delinquency from ages 13 to 17. The model is estimated with data from the Montreal-based longitudinal study.

Figure 2 displays the form of the trajectories identified for these two behaviors. Panel A shows the trajectories of opposition from ages 6 to 13, which were a product of the censored normal model. One trajectory starts off low at age 6 and declines steadily thereafter. The second trajectory starts off at a modest level of opposition at age 6, rises slightly until age 10, and then begins a gradual decline. The third trajectory starts off high and remains high over the age period. These trajectories of childhood opposition were estimated to account for 32.0, 45.2, and 22.9 percent of the population, respectively.

Panel B shows the trajectories for property delinquency, which were estimated using the Poisson-based model. The two largest trajectory groups, which were estimated to account for a total of nearly 72 percent of the population, are characterized by low and slightly declining rates of property delinquency. The third trajectory group follows a pattern of rising property delinquency over most of the measurement period, whereas the fourth group remains high over the entire period. Groups 3 and 4 were estimated to account for 19.2 and 8.9 percent of the population, respectively.

Figure 3 reports the key innovation of the joint trajectory model—three alternative representations of the linkage between opposition and property delinquency. One is the probability of membership in each of the property delinquency trajectories, conditional on membership in each of the opposition trajectory groups. These probabilities, which are reported in Panel A,

Figure 2
(A) Trajectories of Opposition From Ages 6 to 13 and
(B) Trajectories of Property Delinquency From Ages 13 to 17

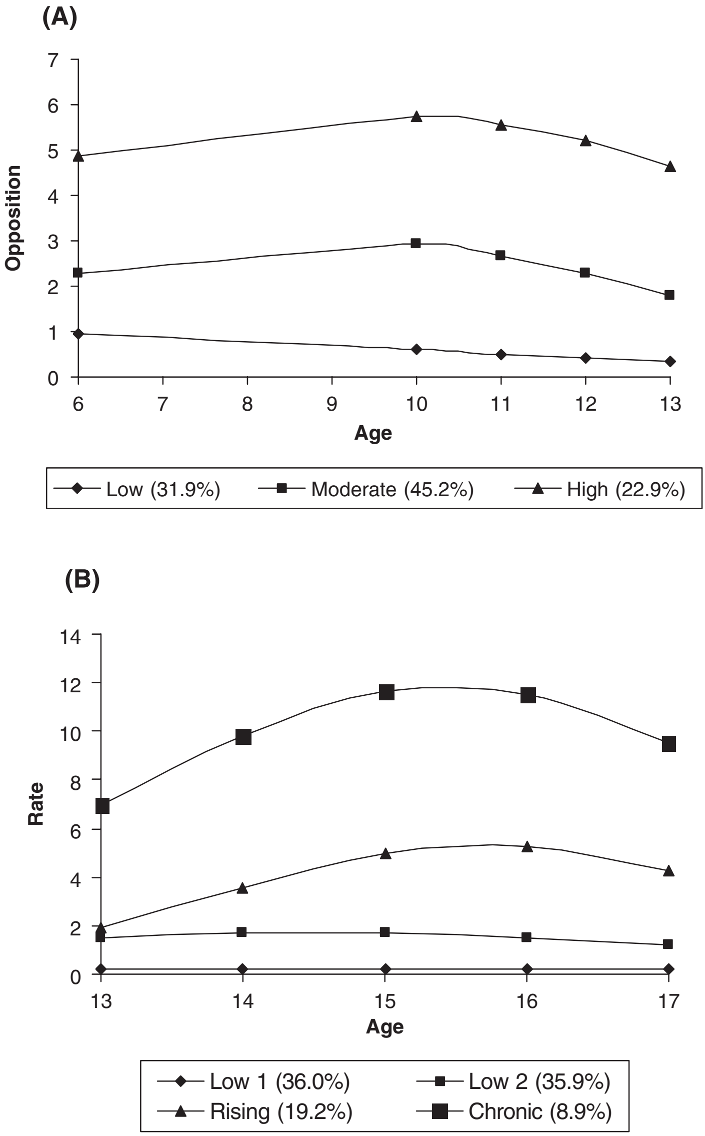
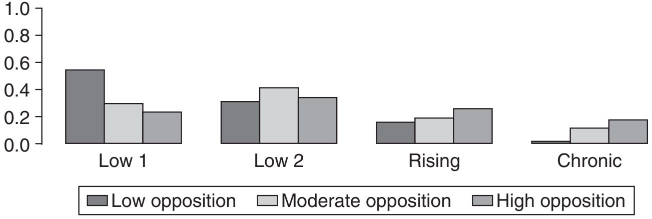
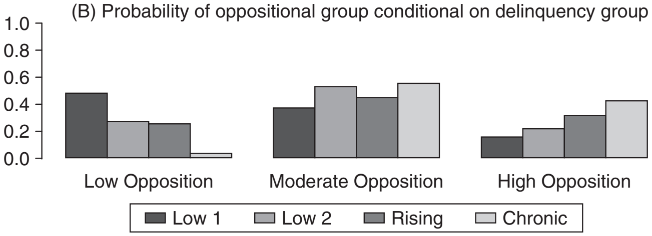


Figure 3
The Relationship of Opposition (Ages 6-13)
With Property Delinquency (Ages 13-17)

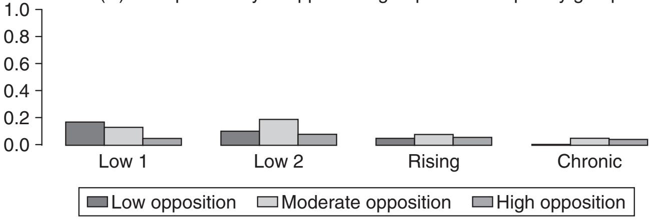
(A) Probability of delinquency group conditional on opposition group



(B) Probability of oppositional group conditional on delinquency group



(C) Joint probability of opposition group and delinquency group



can be interpreted as the probability of transitioning from each childhood opposition trajectory to each of the property delinquency trajectories. Because the probabilities are conditional on membership in a given opposition trajectory group, each column of probabilities in Panel A sums to 1. Panel B reports the reverse set of conditional probabilities: the probability of membership in each of the opposition trajectories conditional on membership in each of the property delinquency trajectory groups. In this panel, each row of probabilities sums to 1. These probabilities measure the composition of each adolescent trajectory group in terms of the childhood trajectory of origin. The third form of representation, reported in Panel C, is the joint probability of membership in a specific property delinquency trajectory and a specific trajectory of the opposition trajectory group. This panel enumerates the probabilities of all the possible combinations of opposition and property delinquency trajectory groups. Thus, the 12 joint probabilities sum to 1.

However represented, the results show a strong relationship between the developmental trajectories for these two behaviors. Panel A shows that the boys who were least oppositional from ages 6 to 13 were least likely to be members of the two higher trajectories of property delinquency. Indeed, their probability of membership in the chronic trajectory was nearly zero. By contrast, the probability of transition to the chronic trajectory from the high-opposition trajectory was 17 percent. Notwithstanding these overall tendencies, the transition probabilities also make clear that the childhood opposition trajectory is not even close to being a certain predictor of the subsequent trajectory of property delinquency. The entries in Panel B make this clear. With the exception of the chronic trajectory, all of the property trajectories were composed of large contingents of boys from each of the opposition trajectories. Compared to simply correlating the number of acts of property delinquency at each year from ages 13 to 17 with opposition at each age from 6 to 13, the dual model provides a far richer, yet still comprehensible, summary of the relationships in the data.

The following statements were used to fit the dual model described in this example:

```
proc traj data=combine out=b outstat=crimstat outplot=crimplot
outstat2=crimstat2 outplot2=crimplot2;
id id;
var opp1-opp5;
indep t1-t5;
model cnorm;
max 10;
ngroups 3;
order 2 2 2;
var2 del13-del17;
indep2 t5-t9;
model2 zip;
ngroups2 4;
order2 2 2 2 2;
run;
```

Extension 3: The Dual-Trajectory Model Generalized to Include Predictors of Conditional Probabilities

We next demonstrate an extension that allows the conditional probabilities linking trajectories across behaviors to vary as a function of individual-level variables. The objective is to provide the statistical basis for analyzing whether the conditional probabilities linking trajectories vary as a function of individual-level characteristics and experiences. For example, in the context of the childhood opposition/adolescent delinquency model, experiencing a major life event at age 11 or 12 might alter the probability of transition to the various delinquency trajectories. To illustrate this extension, we extend the prior example with an analysis of how the probabilities of transition from each of the childhood opposition trajectories to each of adolescent property delinquency trajectories vary as a function of the individual having experienced two potential turning-point events at the time of the transition—not living with both parents at age 11 or 12, hereafter referred to as "broken home 11-12," and using drugs at age 12.

A full description of the technical details of how the basic dual-trajectory model is extended to analyze such effects is described in Nagin (2005). Here we provide only a brief outline. Let Y_1 and Y_2 denote the two longitudinal series to be modeled in a dual-trajectory format, and let j and k index the J and K trajectory groups for Y_1 and Y_2, respectively. In the basic dual model, the direct products of model estimation are the parameters that specify the J and K trajectories of Y_1 and Y_2; the probability of membership in each of Y_1's J trajectories, π_j; and the conditional probability of membership in Y_2's kth trajectory given membership in Y_1's jth trajectory, $\pi_{k|j}$. To ensure that $\pi_{k|j}$ is always between 0 and 1, this quantity is not directly estimated. Instead, a set of parameters $\gamma^0_{k|j}$ is estimated in which

$$\pi_{k|j} = e^{\gamma^0_{k|j}} \Big/ \sum_k e^{\gamma^0_{k|j}} \qquad j = 1, \dots, J.$$

Let w_i denote a vector of variables that are thought to be associated with $\pi_{k|j}$. The impact of these variables on $\pi_{k|j}$ is measured by a set of parameter vectors γ'_k, whereby

$$\pi_{k|j}(w_i) = e^{\gamma^0_{k|j} + \gamma'_k w_i} \Big/ \sum_k e^{\gamma^0_{k|jk} + \gamma'_k w_i} \qquad j = 1, \dots, J.$$

To avoid an unmanageable proliferation of parameters, the model assumes that the effects of the variables included in w_i do not depend on the trajectory group membership for Y_1. Under this assumption, the

Table 1
The Impact of Living in a Broken Home and Drug Use on Trajectory Transition Probabilities (Low 1 is the Comparison Group)

Delinquency Group and Variable	Coefficient	*Z* Score
Low 2		
Broken home 11-12	0.114	0.48
Drug use 12	0.784	6.00
Rising		
Broken home 11-12	0.123	0.47
Drug use 12	0.959	7.06
Chronic broken home 11-12	0.627	1.81
Drug use 12	1.37	9.23

various values of $\gamma'_{k|j}$ are equal across the J trajectory groups in Y_1; therefore, they can be denoted by γ'_k. Conceptually, this modification amounts to assuming that the influence of a particular variable on the probability of transition to a specific trajectory group k of Y_2, as measured by γ'_k, does not interact with trajectory membership for Y_1.

Note that the no-interaction model continues to allow the intercepts, $\gamma'_{k|j}$, to vary freely across Y_1's J trajectories. Consequently, the transition probabilities, $\pi_{k|j}$, will be different for two individuals with identical characteristics w_i but who had followed different trajectories for behavior Y_1. Thus, the no-interaction model still allows the trajectory that is followed for Y_1 to influence the probability of trajectory group membership for Y_2, even controlling for the variables in w_i.

We turn now to our illustrative application of this model extension, in which the conditional probabilities linking trajectories of opposition from ages 6 to 13 with trajectories of property delinquency from ages 13 to 17 vary as a function of living in a broken home at age 11 or 12 and level of drug use at age 12. Table 1 shows the γ'_k coefficient estimates with 95 percent confidence intervals. For each trajectory group, the estimates should be interpreted as the effect of its associated variable on the probability of transition to that delinquency trajectory relative to the low-delinquency trajectory. The results indicate that *controlling* for the childhood opposition trajectory, living in a broken home at age 11 or 12, has no effect on the transition probabilities, with the possible exception of those related to the chronic trajectory. By contrast, drug use at age 12 has a positive and highly significant effect on the probability of transition to the Low 2, rising, and chronic trajectories relative to the Low 1 trajectory.

Figure 4
Probability of Transition From Each Childhood Opposition Trajectory to the Rising Delinquency Trajectory

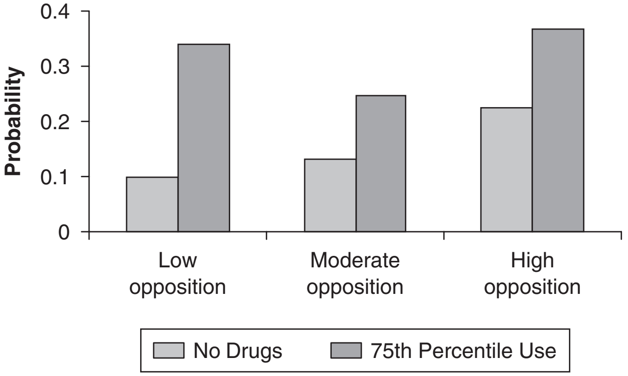


Figure 4 provides a different perspective on these results. It reports the probability of transition to the rising property delinquency trajectory from each of the childhood opposition trajectories for two prototypical individuals—one who does not use drugs at age 12 and one who is at the 75th percentile of the age 12 drug use distribution. Observe that regardless of the childhood opposition trajectory group, drug use at age 12 greatly increases the risk of transition to the rising delinquency trajectory group.

Extension 4: Multitrajectory Modeling

The dual-trajectory model is designed to measure the linkages between the trajectories of two distinct but related outcomes. Conceptually, it is straightforward to extend the dual model to more than two outcomes, but as a practical matter, the addition of more outcomes results in an unmanageable proliferation of probability matrices linking the trajectories for the various outcomes. For example, an extension to three behaviors requires estimation of three matrices of joint probabilities—one for Outcomes 1 to 2, another for Outcomes 1 to 3, and still another for Outcomes 2 to 3. An extension to four outcomes requires the estimation of six matrices of joint probabilities. Still, there are many circumstances where it would be valuable to link trajectories of three or more outcomes of interest. The multitrajectory model is designed to provide this capacity in a model of manageable size.

Figure 5 provides an illustrative application of the multitrajectory model to three distinct but related forms of delinquency—property delinquency, violent delinquency, and vandalism. As can be seen from the Figure 5, each trajectory group is now defined not by one trajectory but by three trajectories: one for property delinquency, a second for violent delinquency, and a third for vandalism. As in the basic model, the size of the group is measured by the probability of group membership. As in the basic model, this probability can be linked to characteristics of the individual. Also, as in the base model, Proc Traj computes posterior probabilities of group membership and related group membership assignments. Time-varying covariates can also be added to the trajectories for each of behaviors, and the probability of group membership can be related to characteristics of the individual.

This multitrajectory model is a form of the so-called constrained dual-trajectory model described in Chapter 8 of Nagin (2005). The Poisson-based form of the model is also laid out in Brame, Mulvey, and Piquero (2001) and also applied in Piquero et al. (2002). The basic form of the likelihood is

$$P(Y_1, Y_2, \ldots, Y_K) = \sum_j \pi_j \prod_k f_k^j(Y_k),$$

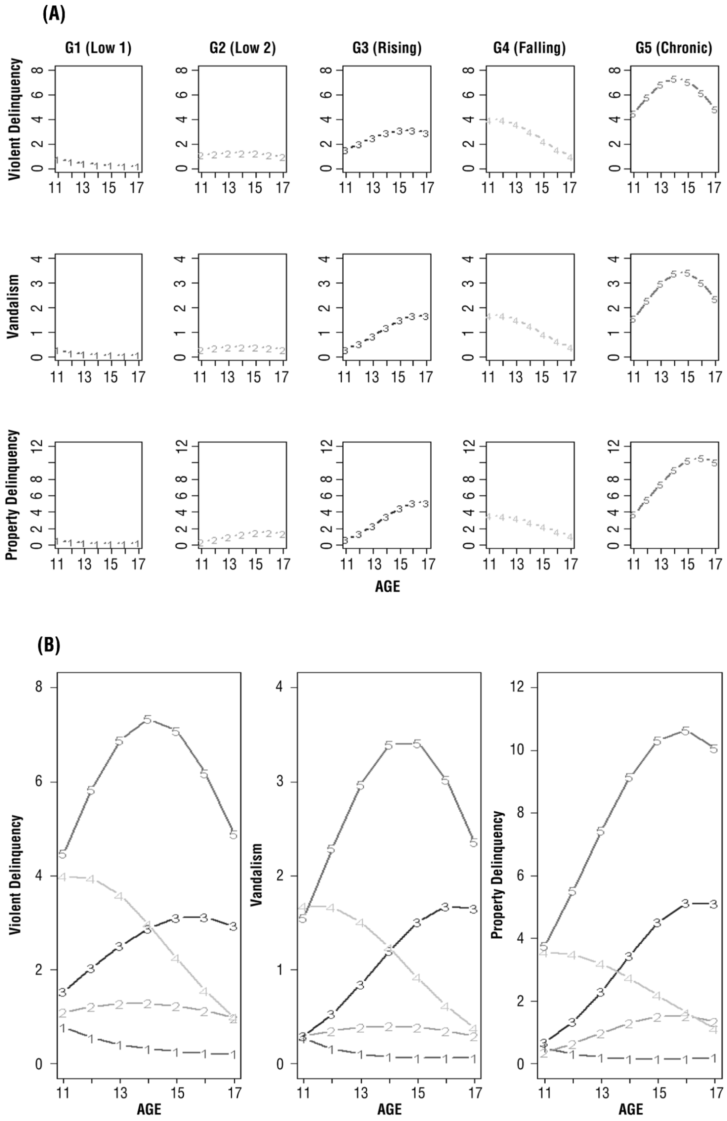
where k is an index of the number of different outcome trajectories in each trajectory group j. Note that $f(*)$, which defines the distribution for each such outcome by trajectory group, can be different across the outcomes. In the example given above, the trajectories for each of the forms of delinquency were modeled using the Proc Traj Poisson model option. However, the use of the same distribution across outcomes is not required. For example, one outcome might be modeled by the censored normal, another by the logit option, and still another by the Poisson option.

The Proc Traj syntax for estimating the model shown in Figure 5 is shown as follows:

```
PROC TRAJ DATA=MONTREAL OUTPLOT=OP OUTSTAT=OS
OUTPLOT2=OP2 OUTSTAT2=OS2 OUTPLOT3=OP3 OUTSTAT3=OS3;
VAR bat89-bat95; INDEP T1-T7; MODEL ZIP; ORDER 2 2 2 2 2;
VAR2 det89-det95; INDEP2 T1-T7; MODEL2 ZIP; ORDER2 2 2 2 2 2;
VAR3 vol89-vol95; INDEP3 T1-T7; MODEL3 ZIP; ORDER3 2 2 2 2 2;
MULTGROUPS 5;

RUN;
```

Figure 5
(A) Five-Group Multiple-Trajectory Model and
(B) Trajectory of Plots Organized by Behavior



Extension 5: Relating Trajectory Groups to a Subsequent Outcome Variable

The dual model provides the capability to analyze the linkages between the developmental trajectories of two related outcomes. Here we demonstrate an option that links trajectory groups measured up to t periods for a specified outcome with another outcome variable, O_i, measured only once on or after the termination of the trajectory. An example of the type an analysis where this model would apply is the relationship of trajectories of hyperactivity from ages 6 to 15 to performance on an academic achievement test at age 18. To accomplish this, the trajectory mixture likelihood is modified to incorporate the probability of the observed outcome, O_i:

$$P(Y_i, O_i) = P(O_i) \sum_j \pi_j P^j(Y_i).$$

We assume that the observed outcome depends only on the trajectory groups. This dependency is modeled through a linear predictor (explained below).

The form of $P(O_i)$ is selected to conform to the outcome data type. For a count outcome, $P(O_i)$ can be specified according to the Poisson, zero-inflated Poisson, or negative binomial distribution. For continuous outcomes, $P(O_i)$ can be accommodated using the normal distribution. Psychometric scale outcomes are accommodated by taking $P(O_i)$ to follow a censored normal distribution. For binary data, $P(O_i)$ follows a binary logit distribution. The linear predictor, $\phi_1\pi_{i1} + \phi_2\pi_{i2} + \cdots + \phi_J\pi_{iJ}$, is used to model the relationship between posterior group membership and outcome, where $\pi_{i1}, \ldots, \pi_{iJ}$ are the posterior probabilities of group membership. For example, if $P(O_i)$ was specified to follow the Poisson distribution, then $\ln(\lambda) = \phi_1\pi_{i1} + \phi_2\pi_{i2} + \cdots + \phi_J\pi_{iJ}$, where λ is the expected number of occurrences for the outcome of interest.

In the case of binary outcome data, it is assumed that $P(O_i)$ follows the binary logit distribution:

$$P(O_i = 1) = \frac{e^{\phi_1\pi_{i1} + \phi_2\pi_{i2} + \cdots + \phi_J\pi_{iJ}}}{1 + e^{\phi_1\pi_{i1} + \phi_2\pi_{i2} + \cdots + \phi_J\pi_{iJ}}}.$$

To illustrate, we investigate how the number of sexual partners at age 14 might differ by opposition trajectory groups in the childhood opposition model. The following statements fit the model with the outcome variable described in this example:

Table 2
Average Annual Number of Sexual Partners at Age 14 and 95 Percent Confidence Intervals by Opposition Trajectory Group (Ages 6-13)

Opposition Trajectory Group	Average Number of Sexual Partners	95 Percent Confidence Interval
Low	0.36	(0.31, 0.41)
Moderate	0.67	(0.46, 0.98)
High	1.62	(1.49, 1.75)

```
proc traj data=a out=b outstat=crimstat outplot=crimplot
outcomeresults=crimresults;
id id;
var opp1-opp5;
indep t1-t5;
model cnorm;
max 10;
ngroups 3; order 0 2 2;

outcome nbp14;
omodel poisson;

run;
```

Table 2 gives the average number of sexual partners at age 14 with 95 percent confidence intervals by opposition trajectory group (ages 6-13). These quantities were calculated from the ϕ_k coefficient estimates and standard errors. The results show that the number of sexual partners differs significantly by childhood opposition trajectory group, with greater levels of oppositional behavior associated with higher numbers of sex partners. The estimated average number of sexual partners in the low-opposition group is 0.36 (95 percent CI: 0.31, 0.41). In contrast, the average number of sexual partners in the moderate-opposition group is 0.67 (95 percent CI: 0.46, 0.98), while the number of partners in the high-opposition group is 1.62 (95 percent CI: 1.49, 1.75).

Extension 6: Calculating Group Membership Probabilities as a Function of Time-Stable Covariates

Roeder, Lynch, and Nagin (1999) described an extension of the basic model that allows group membership probabilities to vary as a function of time-stable characteristics of the individual. Because such characteristics act as predictors of trajectory group membership, they should be established no later than the initial period of the trajectory. Denoting such predictors by x_i, the Roeder et al. model extension generalized equation (1) to follow a multinomial logit function that is a function of x_i:

$$\pi_j(x_i) = e^{x_i \theta_j} / \sum_1^J e^{x_i \theta_j}. \tag{5}$$

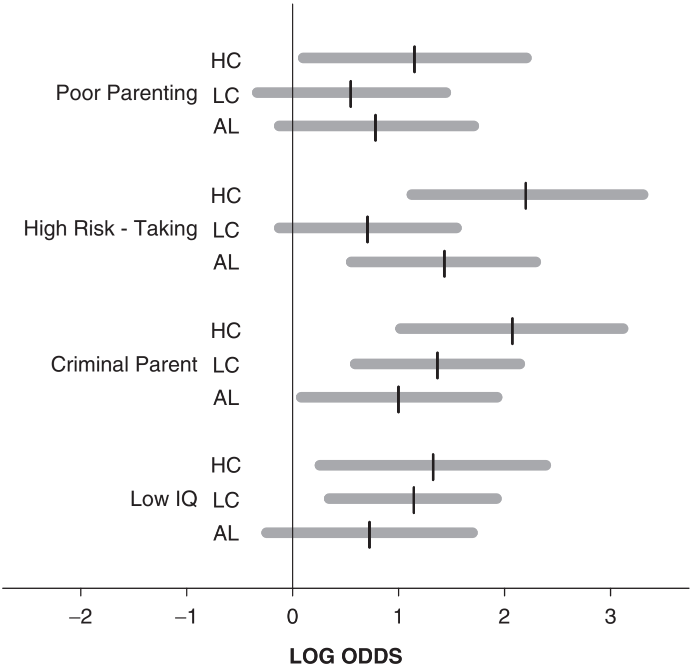
Figure 6 reports the results of an analysis in which the basic four-group model described in the prior example is extended to allow trajectory group probabilities to vary as a function of four classic childhood risk factors for criminal and delinquent behavior in adolescence and beyond: low IQ, having a parent with a criminal record, high risk-taking behavior, and poor parenting. More specifically, the low-IQ variable distinguishes individuals in the lower quartile of the IQ distribution. The high risk-taking variable identifies individuals in the upper quartile of an index of risk taking. The poor parenting variable distinguishes individuals whose parents were identified as having a poor or neglectful parenting style, and the criminal parents' variable identifies individuals with either a father or mother with a criminal record.

Each coefficient estimate is associated with a particular risk factor for a particular trajectory group. Each such estimate can be interpreted as the log of the odds ratio of the impact of that risk factor on the probability of membership in the specified group relative to the nonoffending group. In general, each of the risk factors significantly distinguishes each of the criminal trajectory groups from the nonoffender group.

An alternative way of representing the impact of one or more of these risk factors is through calculations that directly represent their impact on the probability of group membership. Figure 7 illustrates estimates of group membership probabilities for the London model for six scenarios about the level of the four predictor variables—low IQ, having a parent with a criminal record, high risk taking, and poor parenting. Scenario 1 assumes that all are 0. This is equivalent to calculating group membership probabilities for individuals with none of the above risk factors for delinquency. Scenarios 2 through 5 report these same probabilities for individuals with only one of the four risk factors included in the model. In Scenario 6, the group membership probabilities are computed for individuals with all four of the delinquency risks.

The calculations illustrate the concept of cumulative risk (Rutter et al. 1975). Rutter et al. (1975) argued that in most circumstances, no single

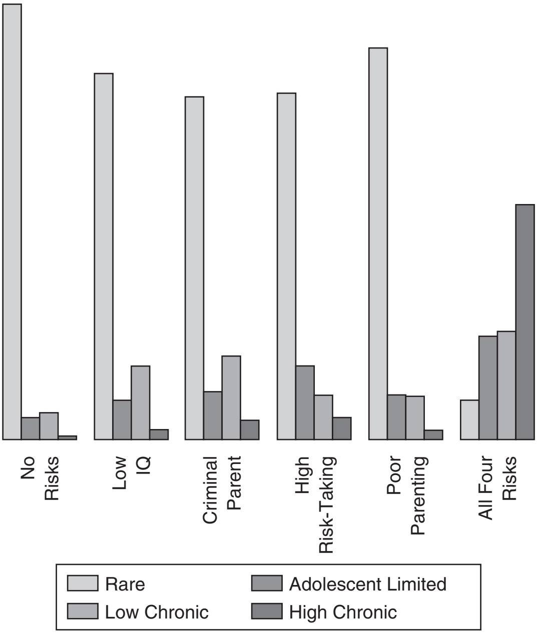
Figure 6
Log Odds With 95 Percent Confidence Intervals for Risk Factors by Group



Note: HC = high chronic; LC = low chronic; AL = adolescent limited.

risk factor is decisive in determining an individual's vulnerability to psychopathology but that the accumulation of such risks is decisive. The calculations show that each risk factor increases the probability of membership in one or more of the delinquent trajectory groups, but no single factor dramatically shifts the probabilities away from those in the no-risk scenario. For example, consider Scenario 3. The model predicts that the probability of membership in the rare group is .70 for individuals who have at least one parent with a criminal record but who have none of the other risk factors. The counterpart prediction for the high-chronic group for these individuals is .039. In contrast, the predicted probabilities of

Figure 7
Predicted Group Membership Probabilities: London Data



membership in the rare and high-chronic groups for individuals with no risks are .89 and .006, respectively. Thus, the criminal parent risk factor materially reduces the rare group probability and increases the high-chronic group probability. Still, the basic ordering of the probabilities remains the same—the rare group is much more likely than the high-chronic group. However, the presence of all four risks results in a dramatic shift. The probability of membership in the high-chronic group increases from nearly 0 in the no-risk scenario to .48 in the all four risk scenarios.

The calculations that form the basis for entries in Figure 7 are based on the logit model coefficient estimates illustrated in Figure 6. Proc Traj, in combination with standard SAS data-handling features, can be used to perform these calculations with the following set of commands:

```
data predict;
input lowiq pbeh daring crimpar;
cards;
0 0 0 0
1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1
1 1 1 1
run;
data combine;
set predict london_data;
run;

proc traj data=combine out=outdata outstat=crimstat outplot=crimplot;
var y1-y11;
indep t1-t11;
model zip;
ngroups 4;
start -4.57 -7 5.96 -1.38 -20.5 25.4 -8.0 -4.37 5.55 -1.52
-1.7 0 0 0 0 -1.7 0 0 0 0 -2.5 0 0 0 0;
order 0 2 2 2;
iorder -1 -1 -1 -1;
risk lowiq crimpar daring pbeh;
run;

proc print data=b (obs=6);
run;
```

The first two data steps preceding the call to Proc Traj append six dummy records to the top of the estimation data set. In the six dummy records, the values of the four risk variables are set to correspond to the six scenarios for which we want to calculate group membership variables. However, because the number of convictions in each of the periods from ages 10 to 32 is left unspecified in the dummy records, SAS specifies their values as “missing.” As a consequence, these records have no impact on the estimated model coefficients, their standard errors, or the calculation of various model fit statistics, such as the Bayesian information criterion. Thus, they are referred to as dummy records.

The final processing step involves printing out the first six records of a data set called “outdata.” This data set is produced by Proc Traj. It includes the posterior probability of membership in each trajectory group for each case in the estimation data set. The posterior probability of membership in group *j* for case *i* is based on application of Bayes's theorem. For each case *i*, group *j*'s posterior probability is calculated using the estimated model coefficients and the data for that case. For real cases with actual conviction data, this calculation takes into account the history of conviction. For example, an individual whose convictions are limited to his or her adolescence will have a high probability of membership in the adolescent-limited group. However, for the dummy cases, in which the conviction history is deliberately specified as missing, the calculation reduces to the desired quantity—that specified by equation (5).

Extension 7: Wald Tests of the Equality of Coefficient Estimates

The two primary outputs of model estimation are the model's parameter estimates and their variances and covariances. The latter form the basis for the test statistics reported in the output of Proc Traj of whether the coefficient estimates are significantly different from zero. For example, the confidence intervals showing whether the various risk factors were significant predictors of membership in one of the criminal trajectories compared to the nonoffender trajectory, given in Figure 6, were easily derived directly from the Proc Traj output.

However, many other interesting and important statistical tests can be conducted based on the variance-covariance estimates of the parameter estimates. One is whether the various risk factor coefficient estimates differ across the three criminal trajectory groups. Differences in their magnitude would imply that they differentially predict membership in the various criminal trajectory groups. Alternatively, if they are individually significantly different than zero but the null hypothesis of their equality cannot be rejected, this would imply that while the risk factor is a significant predictor of criminality, it does not predict among its alternative developmental courses.

As elaborated in Nagin (2005), such hypotheses can be tested with the Wald test. The Wald test is a chi-square-based test with degrees of freedom (df) equal to the number of equality constraints being tested. A newly developed SAS macro called trajtest can be used to conduct Wald tests of the equality of one or more coefficients. For example, a test of the equality of the low IQ coefficients across the three criminal trajectory groups yielded a Wald-based chi-squared statistic of .98 with 2 df, which is far short of significance. This result implies that low IQ is not a significant predictor of which of the criminal trajectories an individual might follow. Thus, while the results in Figure 6 showed that low IQ was a risk factor for criminality, the chi-square-based statistic implies that low IQ does not distinguish among the specific developmental courses of criminality. On the other hand, the companion test for the risk-taking variable was significant at the .05 level. This implies that risk taking was not only a significant predictor of criminality but also of its specific trajectory.

The Wald test can also be used to test whether trajectories are distinctive in the sense that they are not parallel. Two or more trajectories would be considered parallel if the constant terms of the trajectories (i.e., β_0^j)were significantly different but the coefficients of higher order terms of the polynomials describing the trajectories of the respective groups (i.e. $\beta_1^j, \beta_2^j, \ldots$)were not. We applied this testing procedure to compare the low- and high-chronic trajectories. We found that both their intercepts and their linear and quadratic terms were significantly different according to the Wald test. The chi-square statistic for the intercept contrast is 15.71 ($df = 1$), and the companion chi-square test statistic for the equality of the linear and quadratic terms was 16.44 ($df = 2$).

The syntax of the calls to the trajtest macro for conducting the above reported tests was as follows:

```
%TRAJTEST ('lowiq2=lowiq3=lowiq4')
   /* low iq equality test */
%TRAJTEST ('daring2=daring3=daring4')
   /* risk taking equality test */
%TRAJTEST ('interc3=interc4')
   /* intercept equality test */
%TRAJTEST ('linear3=linear4,quadra3=quadra4')
   /* linear & quadratic equality test */
```

Extension 8: Plotting the Impact of Time-Varying Covariates

Jones et al. (2001) demonstrated the capability of adding time-varying covariates beyond age or time into the specification of the trajectory. As developed at length in Nagin et al. (2003) and Nagin (2005), the objective of this modeling extension is to analyze the effect of turning-point events, such as divorce, or therapeutic interventions, such as counseling, on the behavioral trajectories. (We use the term *effect* advisedly in full recognition of the difficulties of causal interpretation in observational data. See

Haviland and Nagin [2005] and Nagin [2005] for a discussion of causal inference in the context of group-based trajectory modeling.)

Here we demonstrate a new feature that aids in the graphical presentation of the estimated effect of the covariate of interest on each trajectory. Specifically, the function called plottcov directs Proc Traj to calculate the trajectory for each group under a user-specified set of values of the covariate of interest. These calculations are done *postmodel estimation* based on the estimated values of the coefficients that define the trajectory over time. These include the coefficients measuring the effects of all the covariates included in the trajectory.

We demonstrate this capability with a group-based trajectory analysis of the effect of gang membership on violent delinquency. This analysis is based on self-reports from the Montreal data on violent delinquency from ages 11 to 17 and companion self-reports of whether or not the boy was involved in a delinquent group at that age. With the addition of gang membership as a time-varying covariate, the form of the trajectory for each group was as follows:

$$\ln(\lambda_t^j) = \beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \alpha_1^j gang_{it},$$

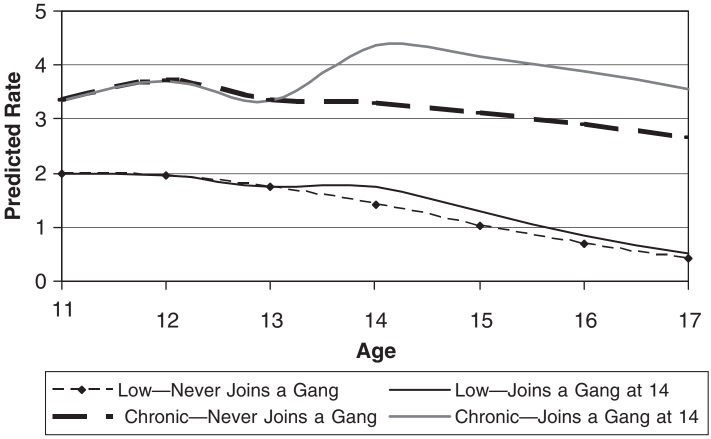
where $gang_{it}$ is a dichotomous indicator variable that equals 1 if subject i reports being in a gang in period t. We model self-reported frequency of violent delinquency as a Poisson random variable, and thus each trajectory is defined by the Poisson rate parameter, λ_t^j, over time.

A five-group model was estimated. Note that the model allows the effect of the gang membership status variable to vary freely across trajectory groups, which allows for the possibility of differential effects. In this analysis, the estimate of α_1^j was positive and highly significant for each group, which implies that gang membership is associated with increased violence.

Figure 8 is a graphical depiction of the plottcov-generated calculations for two of the five trajectories. For each trajectory, the graph compares the predicted trajectory for an individual who stays out of a gang from ages 11 to 17 with that of an individual who stays out of gangs from ages 11 to 13 but at age 14 joins a gang and remains a gang member thereafter. As indicated, the gang effect was positive and statistically significant for all groups, but the figure shows that the effect size was larger for what we call the moderate-chronic group than for what we call the low group.

The following commands were used to create the predicted trajectories shown in Figure 8:

Figure 8
Illustration of Estimated Effect of Gang Membership on Violent Delinquency for the Low 2 and Moderate Chronic Trajectories



```
proc traj data=a out=b outstat=viostat outplot=vioplot1;
    var vio11-vio17;
    indep age11-age17;
        model zip;
            tcov gang11-gang17;
            plottcov 0 0 0 0 0 0 0;
        ngroups 5;
        order 2 2 2 2 2;
run;

proc print data=vioplot1;
run;

proc traj data=a out=b outstat=viostat outplot=vioplot2;
    var vio11-vio17;
    indep age11-age17;
    model zip;
    tcov gang11-gang17;
    plottcov 0 0 0 1 1 1 1;
```

```
    ngroups 5;
    order 2 2 2 2 2;
    run;

proc print data=vioplot2;
run;
```

The key commands are in bold. The command “tcov gang11-gang17” adds the gang membership status variable to the trajectory. Note that additional covariates beyond gang membership status at each age could have been included in the tcov command. The command “plottcov 0 0 0 0 0 0 0” directs Proc Traj to calculate the predicted trajectories postestimation under the assumption that the gang membership status variable equals 0 in all periods. This command produced the predicted trajectories under the non-gang-joining condition shown in Figure 8. The predicted trajectories were stored in the file named “vioplot1” and were printed out using SAS’s standard proc print function. To obtain the predicted trajectories for the condition where the individual joins a gang at age 14 (i.e., plottcov 0 0 0 1 1 1 1), the entire process has to be repeated. Even though the plottcov calculations were performed postestimation, technical obstacles make it impossible to include more than one use of plottcov in a call to Proc Traj. Thus, the entire model must be reestimated to do the second set of calculations that were required to produce Figure 8.

The figure itself was created by entering the predicted trajectories in vioplot1 and vioplot2 into a stand-alone graphing software package. Also, note that the two vioplot files also included the actual trajectories of violent delinquency for each trajectory group. For each trajectory group, the calculation of the actual trajectory is based on the actual values of violent delinquency for each individual weighted by their posterior probability of membership in that trajectory. Because actual gang membership experience has a pronounced positive association with violent delinquency, the actual trajectories are no longer comparable to the predicted trajectories under the hypothetical set of conditions specified in the plottcov command.

Extension 9: Sample Weights and Exposure Time

The final example illustrates the capacity to incorporate sample weights and adjust for exposure time in the ZIP model. These options will be shown using data from the Rochester Youth Development Study.

The Rochester Youth Development Study tracked a sample of students who were in the seventh and eighth grades in Rochester, New York, public schools for the spring semester of the 1988 school year. Males and students from high-crime areas were more heavily sampled since it was assumed that they were at greater risk for offending. The project was a 12-wave prospective study where members of the sample and one of their parents were interviewed semiannually from 1988 to 1992 and annually from 1994 to 1996.

Due to the oversampling of individuals from high-crime areas, the data for study participants must be weighted in model estimation to obtain a model that provides unbiased estimates of the representation of each trajectory group in the population and also to obtain valid standard errors. A robust estimator of the variance-covariance matrix for the parameters from the weighted likelihood function is obtained using a sandwich estimator.

In addition, because the time between assessment periods was not constant over the course of the study and across study participants, it is imperative that we account for differences across individuals and over time in the participant's "exposure time" to commit crimes between assessments. For example, in the early stages of the study when assessments were made semiannually, counts of self-reported delinquent acts referred to approximately 6 months of exposure time, whereas later in the study, the counts referred to approximately 12 months of exposure time. More generally, it is critical that time between assessments be taken into account in any analysis of event frequency.

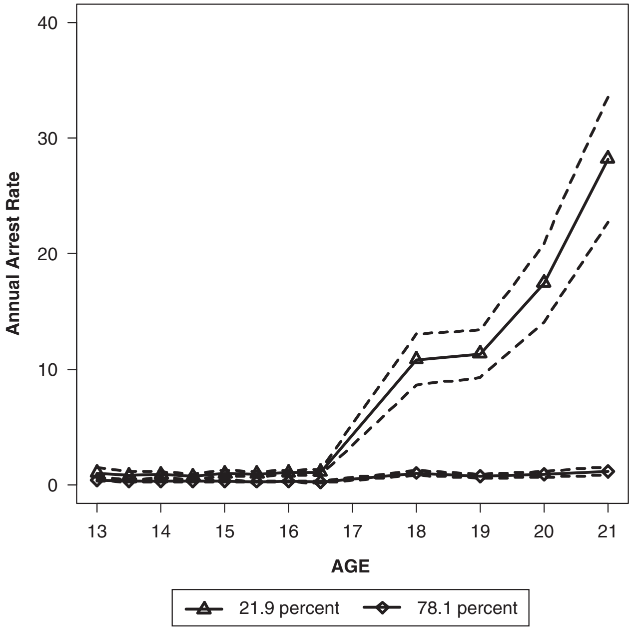
An adjustment for exposure time is also required if individuals are somehow placed in a situation where they are restricted from engaging in the activity of interest. For example, the exposure time adjustment was first demonstrated in the Piquero et al. (2001) analysis of the arrest histories of individuals who had been under the supervision of the California Youth Authority. The purpose of the exposure time adjustment was to account for spells of imprisonment, during which times individuals cannot be arrested for crimes that would appear on their arrest record.

The use of exposure variables modifies the Poisson rate equation as follows:

$$\ln(\lambda_t^j) = \beta_0^j + \beta_1^j(Age_{it} * Expos_{it}) + \beta_2^j(Age_{it} * Expos_{it})^2.$$

The following statements fit a two-group model to the arrest counts:

Figure 9
Annual Arrest Rate Versus Age



```
proc traj data=a out=b outstat=crimstat outplot=crimplot;
    id id;
    weight wt;
    var ac1-ac12;
    indep t1-t12;
    expos e1-e12;
    model zip;
    ngroups 2; order 2 2; iorder 0 2;
run;
```

Figure 9 shows the estimates for the two trajectories of λ_t^j. The first group is composed of individuals with a very low rate of offending over the 12-wave period. The size of this group is estimated at 78.1 percent. The other trajectory group (21.9 percent) shows very high and increasing levels of offending during the 9th through 12th waves.

Conclusion

The purpose of this article was to describe extensions of a SAS-based procedure for estimating group-based trajectory models that have been developed since the appearance of Jones et al. (2001). These include the following:

1. reporting confidence intervals on trajectories and group membership probabilities,
2. calculating group membership probabilities as a function of time-stable covariates,
3. conducting Wald tests of the equality of coefficient estimates,
4. plotting the impact of time-varying covariates,
5. estimating dual-trajectory models with the ability to include predictors of the conditional probabilities that link trajectories of two different outcomes,
6. relating trajectory groups to a subsequent outcome variable,
7. including sample weights and adjusting for exposure time in the ZIP model, and
8. linking trajectories of three or more outcomes of interest with the multitrajectory model.

It our hope that these expanded capabilities will increase the use of group-based trajectory modeling in appropriate problem domains involving the analysis of longitudinal data.

References

Brame, R., E. P. Mulvey, and A. R. Piquero. 2001. "On the Development of Different Kinds of Criminal Activity." *Sociological Methods & Research* 29:319-41.

Dennis, J. E., D. M. Gay, and R. E. Welsch. 1981. "An Adaptive Nonlinear Least-Squares Algorithm." *ACM Transactions on Mathematical Software* 7:348-83.

Dennis, J. E. and H. H. W. Mei. 1979. "Two New Unconstrained Optimization Algorithms Which Use Function and Gradient Values." *Journal of Optimization Theory and Applications* 28:453-83.

Efron, B. 1979. "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7:1-26.

Farrington, D. P. 1990. "Childhood Aggression and Adult Violence: Early Precursors and Laterlife Outcomes." In *The Development of Childhood Aggression*, edited by D. J. Pepler and K. H. Rubin. Hillsdale, NJ: Lawrence Erlbaum.

Greene, W. H. 1995. *Econometrics*. New York: Macmillan.

Haviland, A. and D. S. Nagin. 2005. "Causal Inference With Group-Based Trajectory Models." Working paper, Carnegie Mellon University.

Jones, B. L., D. Nagin, and K. Roeder. 2001. “A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories.” *Sociological Methods & Research* 29:374-93.

Lambert, D. 1993. “Zero Inflated Poisson Regression With an Application to Defects in Manufacturing.” *Technometrics* 34:1-13.

Loeber, R. 1991. “Questions and Advances in the Study of Developmental Pathways.” In *Models and Integrations*, edited by D. Cicchetti and S. Toth. New York: University of Rochester Press.

Nagin, D. 1999. “Analyzing Developmental Trajectories: Semi-Parametric, Group-Based Approach.” *Psychological Methods* 4:39-177.

———. 2005. *Group-Based Modeling of Development*. Cambridge, MA: Harvard University Press.

Nagin, D., L. Pagani, R. Tremblay, and F. Vitaro. 2003. “Life Course Turning Points: A Case Study of the Effect of School Failure on Interpersonal Violence.” *Development and Psychopathology* 15:343-61.

Nagin, D. S. and R. E. Tremblay. 2001. “Analyzing Developmental Trajectories of Distinct but Related Behaviors: A Group-Based Method.” *Psychological Methods* 6 (1): 18-34.

Piquero, A. 2004. “What Have We Learned About the Natural History of Criminal Offending From Longitudinal Studies?” Paper presented at the National Institute of Justice, Washington, DC.

Piquero, A., A. Blumstein, R. Brame, R. Haapanen, E. Mulvey, and D. S. Nagin. 2001. “Assessing the Impact of Exposure Time and Incapacitation on Longitudinal Trajectories of Offending.” *Journal of Adolescent Research* 16:54-76.

Piquero, A. R., R. Brame, P. Mazzerole, and R. Haaplanen. 2002. “Crime in Emerging Adulthood.” *Criminology* 40:137-70.

Roeder, K., K. Lynch, and D. S. Nagin. 1999. “Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology.” *Journal of the American Statistical Association* 94:766-76.

Rutter, M., B. Yule, D. Quintin, O. Rowlands, W. Yule, and W. Berger. 1975. “Attainment and Adjustment in Two Geographical Areas: The Prevalence of Psychiatric Disorder.” *British Journal of Psychiatry* 126:493-509.

Tremblay, R. E., L. Desmarais-Gervais, C. Gagnon, and P. Charlebois. 1987. “The Preschool Behavior Questionnaire: Stability of Its Factor Structure Between Culture, Sexes, Ages, and Socioeconomic Classes.” *International Journal of Behavioral Development* 10:467-84.

Bobby L. Jones is a principal research analyst and statistician, Heinz School of Public Policy and Management, Carnegie Mellon University. Recent publications include “Mixture Models for Linkage Analysis of Affected Sibling Pairs and Covariates” (with Bernie Devlin, Silviu Bacanu, and Kathryn Roeder) in *Genetic Epidemiology* (2002) and “Underdevelopment of the Postural Control System in Autism” (with Nancy Minshew, Kim Sung, and Joseph Furman) in *Neurology* (2004).

Daniel S. Nagin is the Teresa and H. John Heinz III Professor of Public Policy and Statistics, Heinz School of Public Policy and Management, Carnegie Mellon University. His research focuses on the evolution of criminal and antisocial behaviors during the life course, the deterrent effect of criminal and noncriminal penalties on illegal behaviors, and the development of statistical methods for analyzing longitudinal data. He is the author of *Group-Based Modeling of Development* (2005).