

This article introduces a new SAS procedure written by the authors that analyzes longitudinal data (developmental trajectories) by fitting a mixture model. The TRAJ procedure fits semiparametric (discrete) mixtures of censored normal, Poisson, zero-inflated Poisson, and Bernoulli distributions to longitudinal data. Applications to psychometric scale data, offense counts, and a dichotomous prevalence measure in violence research are illustrated. In addition, the use of the Bayesian information criterion to address the problem of model selection, including the estimation of the number of components in the mixture, is demonstrated.

A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories

BOBBY L. JONES
DANIEL S. NAGIN
KATHRYN ROEDER
Carnegie Mellon University

The study of developmental trajectories is a central theme of developmental and abnormal psychology and of life course studies in sociology and criminology (Fergusson, Lynskey, and Horwood 1996; Loeber and LeBlanc 1990; Moffitt 1993; Patterson 1996; Patterson, DeBaryshe, and Ramsey 1989; Patterson et al. 1998; Patterson and Yoerger 1997; Sampson and Laub 1993). This article demonstrates a new SAS procedure, called TRAJ, developed by the authors for estimating developmental trajectories. The procedure is based on a semiparametric, group-based modeling strategy. Technically, the model is a mixture of probability distributions that are suitably specified to describe the data to be analyzed. The approach is intended to complement two well-established methods for analyzing developmental trajectories—hierarchical modeling (Bryk and Raudenbush 1987, 1992; Goldstein 1995) and latent growth curve modeling (Meredith and Tisak 1990; Muthen 1989; Willett and Sayer 1994). In hierarchical modeling, individual variation in developmental trajectories, which are commonly called growth curves, are captured by a random coefficients modeling strategy. Latent growth curve

modeling uses covariance structure methods. These methods model variation in the parameters of developmental trajectories using continuous multivariate density functions. The group-based approach employs a multinomial modeling strategy. The statistical theory underlying the method has been developed in detail elsewhere (Nagin and Land 1993; Land, McCall, and Nagin 1996; Roeder, Lynch, and Nagin 1999; Nagin and Tremblay 1999; Nagin 1999), so our focus here is on the software itself and its functional capabilities. However, we begin with a brief overview of the underlying statistical theory.

BRIEF OVERVIEW: DERIVATION OF THE LIKELIHOOD

Mixture models are useful for modeling unobserved heterogeneity in a population. An appropriate parametric model $f(\mathbf{y}, \lambda)$ is assumed for the phenomenon to be studied, where $\mathbf{y} = (y_1, y_2, \dots, y_T)$ denotes the longitudinal sequence of an individual's behavioral measurements over the T periods of measurement. However, in contrast to the homogeneous case, it is believed that there are unobserved subpopulations differing in their parameter values. In this case, the marginal density for the data \mathbf{y} can be written,

$$f(\mathbf{y}) = \sum_{k=1}^K \Pr(C = k) \Pr(\mathbf{Y} = \mathbf{y} | C = k) = \sum_{k=1}^K p_k f(\mathbf{y}, \lambda_k). \quad (1)$$

Here p_k is the probability of belonging to class k with corresponding parameter(s) λ_k . The longitudinal nature of the data is modeled by having the parameter(s) λ_k depend on time. Time-stable covariates (risk factors) are incorporated into the model by assuming they influence the probability of belonging to a particular group. Time-dependent covariates can also directly affect the observed behavior, as illustrated in Figure 1.

The risk factors affect the likelihood of a particular data trajectory, but it is assumed that nothing more can be learned about the data (\mathbf{Y}) from risk factors (\mathbf{Z}) given group (C). Thus, we assume the risk factors for subject i , $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iR})$, and the data trajectory for the subject consisting of the repeated measurements over T measurement periods, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT})$, are independent given the group, C_i . Given that there

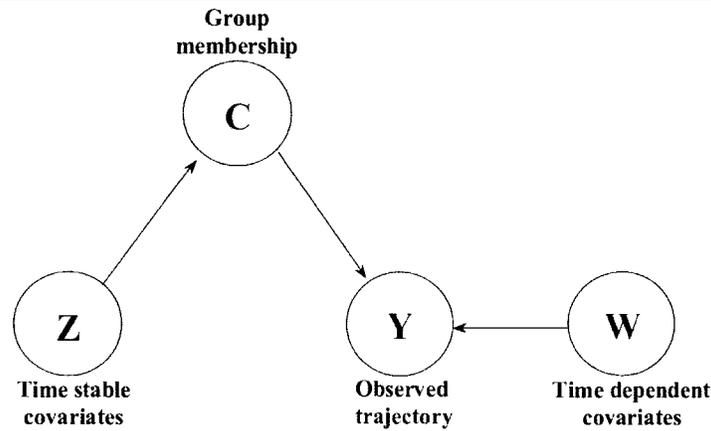


Figure 1: Directed Acyclic Graph Representing the Independence Assumptions

are K groups, we can write the conditional distribution of the observable data for subject i , given risk factors and a time-dependent covariate, $\mathbf{W}_i = (w_{i1}, \dots, w_{it'})$,

$$f(y_i | z_i, \mathbf{w}_i) = \sum_{k=1}^K \Pr(C_i = k | \mathbf{Z}_i = \mathbf{z}_i) \Pr(\mathbf{Y}_i = \mathbf{y}_i | C_i = k, \mathbf{W}_i = \mathbf{w}_i). \quad (2)$$

The time-stable covariate effect on group membership is modeled with a generalized logit function (θ_1 and λ_1 are taken to be zero for identifiability),

$$\Pr(C_i = k | \mathbf{Z}_i = \mathbf{z}_i) = \frac{\exp(\theta_k + \boldsymbol{\lambda}'_k \mathbf{z}_i)}{\sum_{l=1}^K \exp(\theta_l + \boldsymbol{\lambda}'_l \mathbf{z}_i)} \quad (3)$$

TRAJ provides the option of modeling three different distributions for $\Pr(\mathbf{Y}_i = \mathbf{y}_i | C_i = k, \mathbf{W}_i = \mathbf{w}_i)$ to analyze count, psychometric scale, and dichotomous data. The zero-inflated Poisson (ZIP) model is useful for modeling the conditional distribution of count data given group membership when there are more zeros than under the Poisson assumption (Lambert 1992). This is common in antisocial and abnormal behavior that is typically concentrated in a small fraction of the

population. For the ZIP model, the probability of observing the data trajectory \mathbf{y}_i given membership in group k is,

$$\begin{aligned} \Pr(\mathbf{Y}_i = \mathbf{y}_i | C_i = k, \mathbf{W}_i = \mathbf{w}_i) \\ = \prod_{y_{ij}=0} [\rho_{ijk} + (1-\rho_{ijk})e^{-\lambda_{ijk}}] \prod_{y_{ij}>0} (1-\rho_{ijk}) \frac{\exp(-\lambda_{ijk})\lambda_{ijk}^{y_{ij}}}{y_{ij}!}. \end{aligned} \quad (4)$$

Note that ρ_{ijk} is the extra-Poisson probability of a zero. Let age_{ij} denote subject i 's age in period j , and w_{ij} subject i 's time-dependent covariate value in period j . The (optional) time-dependent covariate is related linearly to $\log(\lambda_{ijk})$. In addition, a polynomial relationship is used to model the link between age and the model's parameters:

$$\begin{aligned} \log(\lambda_{ijk}) &= \beta_{0k} + age_{ij}\beta_{1k} + age_{ij}^2\beta_{2k} + \dots + w_{ij}\delta_k \text{ and} \\ \log(\rho_{ijk}/(1-\rho_{ijk})) &= \alpha_{0k} + age_{ij}\alpha_{1k} + age_{ij}^2\alpha_{2k} + \dots \end{aligned}$$

The software allows for specification of up to a third-order polynomial in age. It also allows the user to specify different order polynomials across the k trajectory groups. Equations (3) and (4) incorporated into equation (2) give the likelihood of observing the data trajectory of a subject, given his covariate values. The complete likelihood for all subjects is the product of these individual likelihood values.

The censored normal (CNORM) model is useful for modeling the conditional distribution of psychometric scale data, given group membership (Nagin and Tremblay 1999). A distribution allowing for censoring is used because the data tend to cluster at the minimum of the scale (*Min*) and at the scale maximum (*Max*). Hence, the likelihood of observing the data trajectory for subject i , given he belongs to group k , is

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i | C_i = k, \mathbf{W}_i = \mathbf{w}_i) = \prod_{y_{ij}=\min} \Phi\left(\frac{Min-\mu_{ijk}}{\sigma}\right) \prod_{Min < y_{ij} < Max} \frac{1}{\sigma} \phi\left(\frac{y_{ij}-\mu_{ijk}}{\sigma}\right) \prod_{y_{ij}=Max} \left(1-\Phi\left(\frac{Max-\mu_{ijk}}{\sigma}\right)\right),$$

where

$$\mu_{ijk} = \beta_{0k} + age_{ij}\beta_{1k} + age_{ij}^2\beta_{2k} + \dots + w_{ij}\delta_k. \quad (5)$$

The censored normal model is also appropriate for continuous data that are approximately normally distributed, with or without censoring. The uncensored case is handled by specifying a minimum and maximum that lie outside the range of the observed data values.

Finally, the logistic (LOGIT) model is used to model the conditional distribution of dichotomous data, given group membership. The likelihood of observing the trajectory for subject i , given he belongs to group k , is

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i | C_i = k, \mathbf{W}_i = \mathbf{w}_i) = \prod_{y_{ij}} p_{ijk} \prod_{y_{ij}=0} (1 - p_{ijk})$$

with

$$p_{ijk} = \frac{\exp(\beta_{0k} + age_{ij} \beta_{1k} + age_{ij}^2 \beta_{2k} + \dots + w_{ij} \delta_k)}{1 + \exp(\beta_{0k} + age_{ij} \beta_{1k} + age_{ij}^2 \beta_{2k} + \dots + w_{ij} \delta_k)}. \quad (6)$$

Maximum likelihood is used to estimate the model parameters. The maximization is performed using a general quasi-Newton procedure (Dennis, Gay, and Welsch 1981; Dennis and Mei 1979) obtained from Netlib. Standard error estimates are calculated by inverting the observed information matrix. Subjects with some missing longitudinal data values or time-dependent covariate values are included in the analysis. However, subjects with any missing risk factor (time-stable covariate) data are excluded from the analysis.

OVERVIEW OF SOFTWARE

Many researchers are familiar with the SAS preprogrammed statistical procedures to analyze data. In addition, SAS can be programmed through statements in the data step through macros or through the SAS interactive matrix language. A lesser-known fourth option is to develop a customized SAS procedure using a SAS product: SAS/TOOLKIT. Our custom SAS procedure (available for the PC platform only) is a program written in the C programming language that interfaces with the SAS system to perform the model fitting. The executable dynamic link library is distributed to other users who after installation use it just as they would use any preprogrammed SAS procedure. The following

introductory example illustrates the application of the method and the use of the SAS procedure TRAJ.

EXAMPLE 1: MONTREAL LONGITUDINAL STUDY

The data consist of 1,037 boys assessed annually by their teachers at age 6 (spring 1984) and at ages 10 through 15 on scales of physical aggression, opposition, and hyperactivity. The 53 participating schools were located in low socioeconomic areas of Montreal (Canada). Time-stable covariates were recorded, including age of mother and father at the birth of their first child, years of schooling for the mother and father, a home adversity index, and psychometric scale data on inattention, anxiety, and prosocial behavior of each boy at age 6. Consider the opposition score, which ranges from 0 to 10 and measures five items: does not share, irritable, disobedient, blames others, and inconsiderate. Figure 2 shows sample opposition data for nine subjects, illustrating the variability in the trajectory *shapes*. Some never exhibit difficulties; others have difficulties and then seem to learn more adaptive coping strategies, as evidenced by their drop in opposition scores. Also present are subjects who continue to show high levels of oppositional behavior through age 15. Figure 3 shows the distribution of the opposition scores for each year they were recorded. Scores of zero are most frequent. Note also that the opposition scores decrease in frequency as the score increases. Hence, the censored normal distribution seems a sensible choice for modeling these data.

The following statements fit a five-group model to the oppositional behavior data and plot the results (see Figure 4). The justification for the choice of five groups is discussed in the fourth section of this article.

```
PROC TRAJ DATA=MONTREAL OUT=OF OUTPLOT=OP OUTSTAT=OS;
VAR O1-O7;           /* Opposition Variables      */
INDEP T1-T7;        /* Age Variables             */
MODEL CNORM;       /* Censored Normal Model    */
MIN 0;             /* Lower Censoring Point    */
MAX 10;           /* Upper Censoring Point    */
NGROUPS 5;        /* Fit 5 Groups             */
ORDER 3 3 3 3 3; /* Cubic Trajectory for Each Group */
RUN;
%TRAJPLOT (OP, OS, "Opposition Trajectories", "Opposition", "Scaled Age");
```

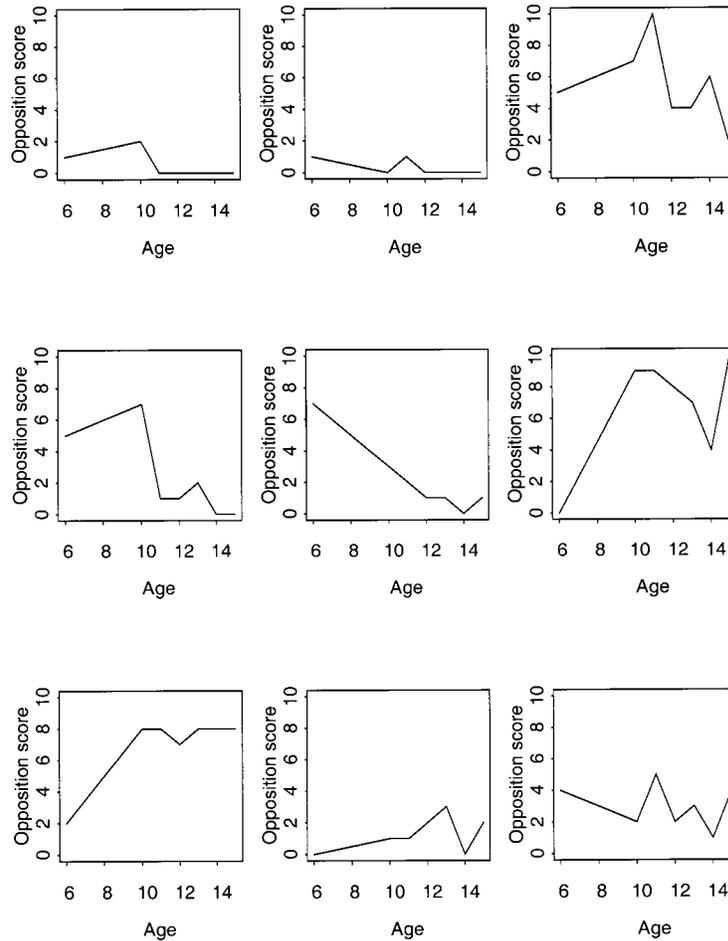


Figure 2: Sample Data (oppositional behavior)

Twenty-two percent of the subjects are classified as exhibiting little or no oppositional behavior (group 1); the largest percentage, 42 percent, exhibit low and somewhat decreasing levels of oppositional behavior (group 2); 18 percent of the subjects show moderate levels of oppositional behavior (group 3); 7 percent of the subjects start out with high levels of oppositional behavior that drops steadily with age (group 4); while the remaining 10 percent exhibit chronic problems with oppositional behavior (group 5).

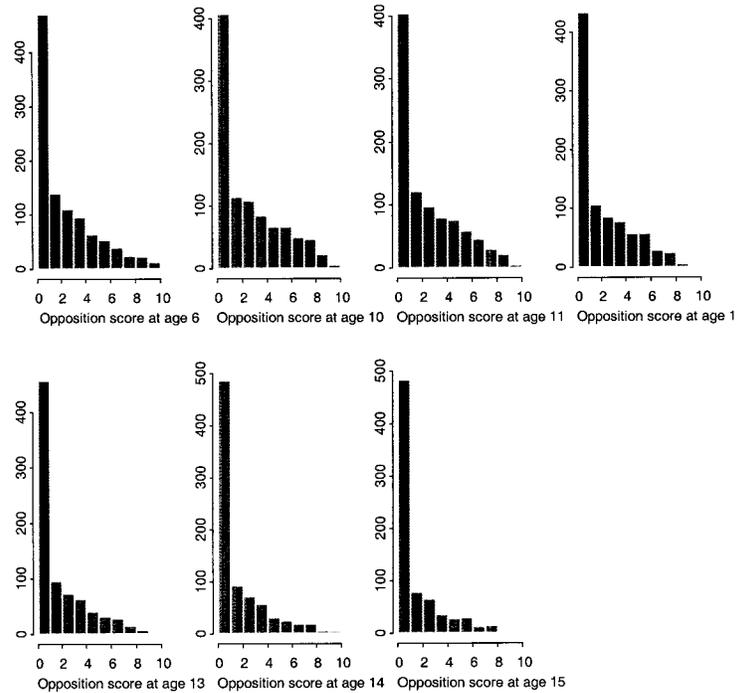


Figure 3: Distribution of Opposition Scores by Age

The next examples illustrate analyses of dichotomous data and Poisson data with extra zeros. It is important to realize that some models are difficult to fit and that there is no guarantee that the procedure will be able to fit the model successfully. In particular, the procedure may find only a local minimum; hence, the process of determining starting values is critical. If the user does not specify starting values (as in the introductory example), the procedure provides default starting values by assuming intercept-only trajectories evenly spaced through the range of the dependent variable. The next example includes the specification of starting values.

EXAMPLE 2: CAMBRIDGE STUDY OF DELINQUENT DEVELOPMENT

The data consist of 411 subjects from a prospective longitudinal survey conducted in a working-class section of London. Farrington and

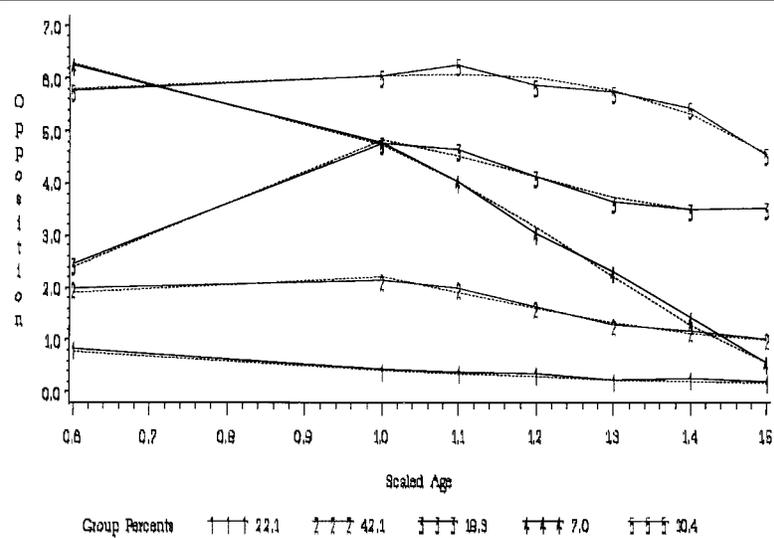


Figure 4: Expected (dashed lines) Versus Observed (solid line) Trajectories

West (1990) provide a detailed discussion of the study. The numbers of criminal offense convictions were recorded annually beginning when the boys were age 10 and continuing through age 32. Because we are dealing with count data, the Poisson model is potentially appropriate here; however, more zeros are present than would be expected in the purely Poisson model, so we use the ZIP model. The following statements fit a four-group model to the offense counts data and plot the results (see Figure 5). The starting values were obtained from an analysis (Roeder et al. 1999) that used cubic trajectories for the four groups.

```
PROC TRAJ DATA=CAMBRDGE OUT=OF OUTPLOT=OP OUTSTAT=OS;
VAR C1-C23;          /* Offense Count Variables */
INDEP T1-T23;       /* Age Variables */
MODEL ZIP;          /* Zero Inflated Poisson Model */
NGROUPS 4;          /* Fit 4 Groups */
ORDER 0 2 0 2;     /* Two Linear and Two Quadratic Groups */
IORDER 1;           /* Linear Zero Inflation */
START -4.8          /* Group 1 - Intercept Only */
-15.5 16.2 -4.5    /* Group 2 - Quadratic Trajectory */
-1.1                /* Group 3 - Intercept Only */
-4.5 5.1 -1.3      /* Group 4 - Quadratic Trajectory */
```

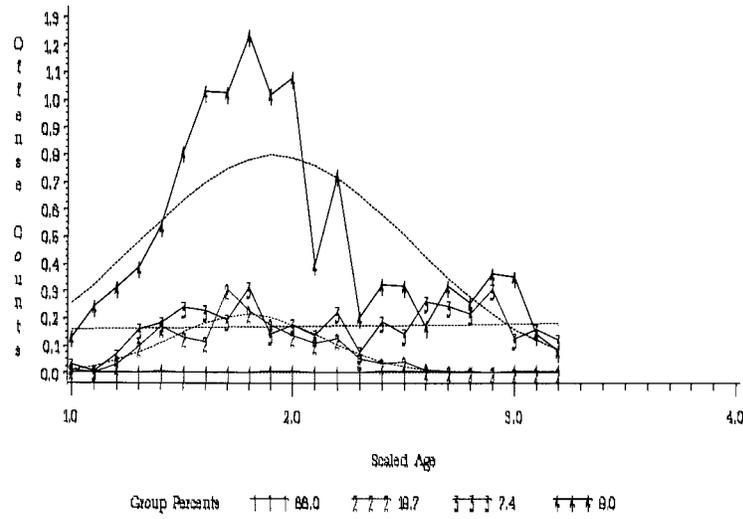


Figure 5: Expected (dashed lines) Versus Observed (solid line) Trajectories

```

-0.2 0.0                /* Linear Zero Inflation          */
-1.2 -2.1 -2.1;        /* Group Proportion Parameters     */
RUN;
%TRAJPLOT (OP, OS, "Offense Counts", "Offense Counts", "Scaled Age");
    
```

Sixty-six percent of the subjects are classified as never convicted (group 1), 19 percent exhibit low conviction rates limited to adolescence (group 2), 7 percent of the subjects show low but persisting conviction rates (group 3), while the remaining 8 percent exhibit the highest conviction rates (group 4).

EXAMPLE 3: CAMBRIDGE DATA PREVALENCE MEASURE

It is common in research on criminal careers to analyze both the frequency of offending measured by offense counts and the absence or presence of offenses (a dichotomous prevalence measure). The analysis on the Cambridge data is repeated, converting the numbers of criminal offense convictions to a dichotomous prevalence measure. The logistic model will be used for the prevalence data. The following statements fit a three-group model to the prevalence measure data and plot the results (see Figure 6).

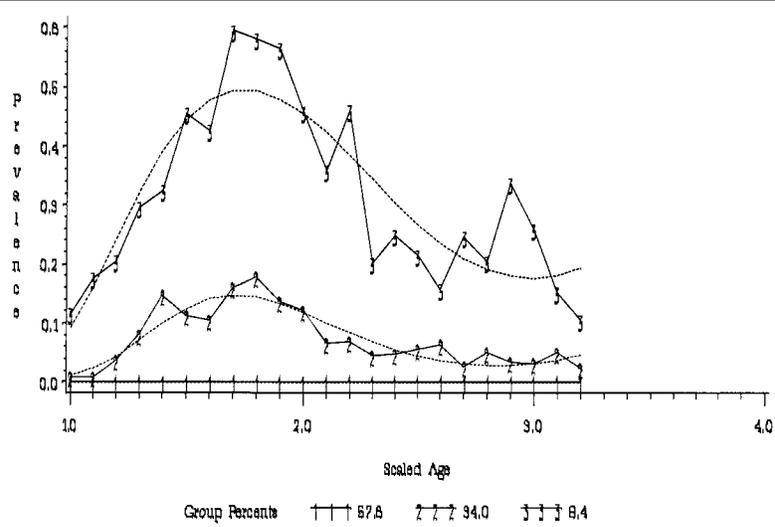


Figure 6: Expected (dashed lines) Versus Observed (solid line) Trajectories

```

PROC TRAJ DATA=CAMBRIDGE OUT=OF OUTPLOT=OP OUTSTAT=OS;
VAR C1-C23;           /* Prevalence Variables      */
INDEP T1-T23;        /* Age Variables             */
MODEL LOGIT;         /* Logistic Model           */
NGROUPS 3;           /* Fit 3 Groups             */
ORDER 3 3 3;        /* Cubic Trajectories       */
RUN;

%TRAJPLOT (OP, OS, "Prevalence Measure", "Prevalence", "Scaled Age");

```

Fifty-eight percent of the subjects are classified as never convicted, 34 percent have a low prevalence rate that peaks during adolescence, and the remaining 8 percent exhibit the highest prevalence rate.

EXAMPLE 4: INTRODUCING TIME-STABLE COVARIATES INTO THE MODEL

A common objective of social science research is to establish whether a trait (e.g., being prone to oppositional behavior) is linked to measured covariates (e.g., risk factors). Previous applications of the semiparametric approach categorized subjects by latent trait from observable behavior (Nagin, Farrington, and Moffitt 1995; Laub, Nagin, and Sampson 1998). The group assignments were then fit to the co-

variates with standard linear models. However, this classify-analyze procedure does not account for the uncertainty in group assignment and can lead to bias (Clogg 1995; Roeder et al. 1999). This final example illustrates the inclusion of risk factors directly into the model. In so doing, this approach accounts for assignment uncertainty automatically.

Suppose we were interested in investigating whether and to what degree inattention, verbal IQ, and an adverse home life are risk factors for elevated levels of opposition. Figure 7 shows the distribution of measures of each of these factors for the subjects in the Montreal study. The procedure automatically drops observations with missing data in the risk factor variables. Of the subjects, 174 have missing values in the risk factors and are omitted from the analysis. The following statements perform the risk analysis on the remaining 863 subjects.

```
PROC TRAJ DATA=MONTREAL OUT=OF OUTPLOT=OP OUTSTAT=OS;
VAR O1-O7;           /* Opposition Variables          */
INDEP T1-T7;        /* Age Variables                 */
MODEL CNORM;        /* Censored Normal Model        */
MIN 0;              /* Lower Censoring Point        */
MAX 6;              /* Upper Censoring Point        */
NGROUPS 5;          /* Fit 5 Groups                  */
ORDER 3 3 3 3 3;    /* Cubic Trajectory for Each Group */
RISK VERBALIQ,      /* Risk Factors                   */
INATTENT,ADVERSTY;
RUN;
```

In Table 1, we present the risk factor parameter estimates, standard errors, tests for the hypothesis that the parameter equals zero, and *p* values for the tests. Figure 8 illustrates the marginal relationships of the risk factors—inattention, adversity, and verbal IQ—to the likelihood of belonging to the highest opposition category versus the lowest opposition category. Included in the plots are the sample values (a small amount of noise has been added to the plot points to separate them): low opposition group on the bottom and high opposition group on the top of each graph. As adversity in the home and inattention scores increase, so does the likelihood of problems with high oppositional behavior. However, as verbal IQ increases, the likelihood of belonging to the high opposition group decreases.

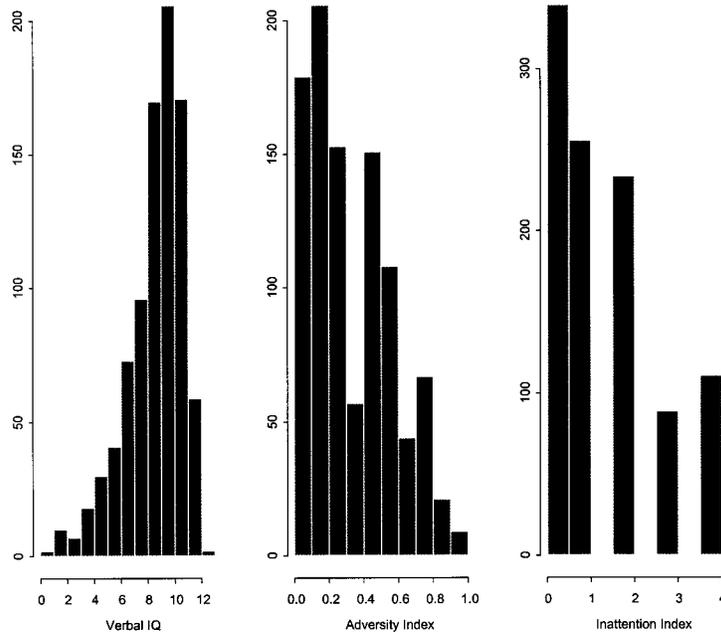


Figure 7: Distribution of Verbal IQ, Adversity, and Inattention Index

*EXAMPLE 5: MONTREAL LONGITUDINAL STUDY
WITH A TIME-VARYING COVARIATE*

A trajectory defines the developmental course of a behavior over age (or time). Trajectories, however, are not deterministic functions of age. External events may deflect a trajectory. For example, Laub et al. (1998) examine the impact of marriage on deflecting trajectories of offending from high levels of criminality toward desistance. Life events may also have transitory affects on enduring trajectories of behavior. For example, spells of mental illness may temporarily alter trajectories of high-level productivity.

In this example, we extend the basic model presented in example 1 by introducing a time-varying covariate into the trajectory model. Specifically, we add to the base model relating opposition to age a binary variable equal to 2 if by the age t the individual had been held back in school, 1 if the individual has not been held back. The objec-

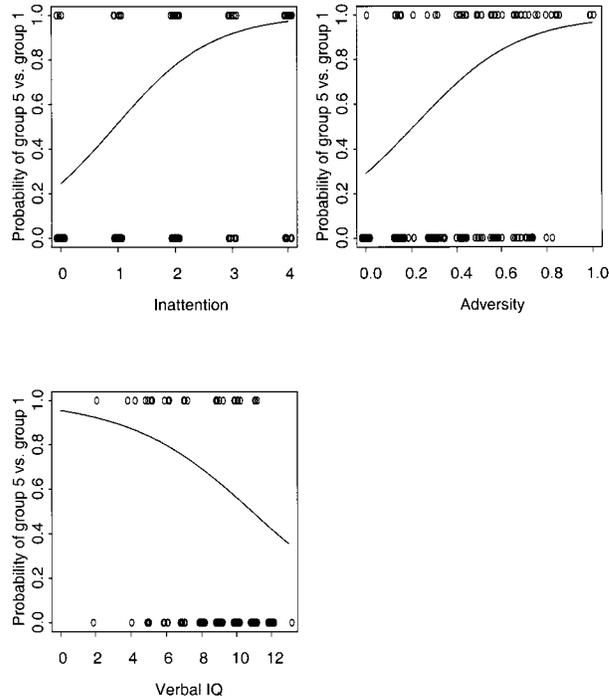


Figure 8: Probability of Belonging to Group 5 (high opposition) Versus Group 1 (low opposition) as a Function of Risk Factor

tive is to test whether for some trajectory groups school failure is associated with an increase in opposition. Note that the structure of the model allows for the possibility that the impact may vary by trajectory group. The number of students held back ranges from 51 at age 6 to 516 at age 15.

The following statements fit a five-group model to the oppositional behavior data.

```

PROC TRAJ DATA=MONTREAL OUT=OF OUTPLOT=OP OUTSTAT=OS;
VAR O1-O7;                /* Opposition Variables          */
INDEP T1-T7;              /* Age Variables                  */
MODEL CNORM;              /* Censored Normal Model         */
MIN 0;                    /* Lower Censoring Point         */
MAX 10;                   /* Upper Censoring Point         */
NGROUPS 5;               /* Fit 5 Groups                   */

```

TABLE 1: Risk Factor Parameter Estimates, Errors, Tests, and *p* Values

| <i>Group</i> | <i>Parameter</i> | <i>Estimate</i> | <i>Error</i> | <i>Test</i> | <i>p Value</i> |
|--------------|------------------|-----------------|--------------|-------------|----------------|
| 2 | Constant | 1.96 | 0.79 | 2.49 | .013 |
| | Inattention | 0.26 | 0.15 | 1.82 | .069 |
| | Adversity | 2.83 | 0.63 | 4.46 | .000 |
| | Verbal IQ | -0.41 | 0.08 | 5.19 | .000 |
| 3 | Constant | 0.80 | 0.72 | 1.11 | .268 |
| | Inattention | 0.48 | 0.12 | 3.98 | .000 |
| | Adversity | 0.98 | 0.49 | 1.99 | .046 |
| | Verbal IQ | -0.10 | 0.07 | -1.48 | .140 |
| 4 | Constant | -4.61 | 1.37 | -3.36 | .001 |
| | Inattention | 1.21 | 0.16 | 7.72 | .000 |
| | Adversity | 2.92 | 0.79 | 3.71 | .000 |
| | Verbal IQ | 0.11 | 0.12 | 0.90 | .366 |
| 5 | Constant | -2.46 | 1.30 | -1.90 | .058 |
| | Inattention | 1.18 | 0.20 | 5.91 | .000 |
| | Adversity | 4.27 | 0.99 | 4.33 | .000 |
| | Verbal IQ | -0.28 | 0.11 | -2.60 | .009 |

ORDER 3 3 3 3 3;
 TCOV C1-C7;
 RUN;

/* Cubic Trajectory for Each Group */
 /* Time Varying Covariate (Held Back) */

Expected opposition trajectories for subjects never held back and always behind are given in Figure 9. Note that this was done as one way to illustrate the effect of the time-varying covariate. Other plots are possible by changing when subjects begin to be behind grade. We see that there is an increase in opposition for those behind grade in groups 2, 3, and 5. There is little effect in the lowest opposition group (group 1) and in the steadily decreasing group (group 4). Those behind grade in group 4 showed lower opposition in the first period. This is explained because of the 55 subjects classified to group 4 (the smallest group), 4 were behind grade in the first period and all had low opposition scores relative to the rest of the group.

*USING THE BAYESIAN INFORMATION CRITERION (BIC)
 FOR MODEL SELECTION*

One possible choice for testing the hypothesis of the number of components in a mixture is the likelihood ratio test. However, the null

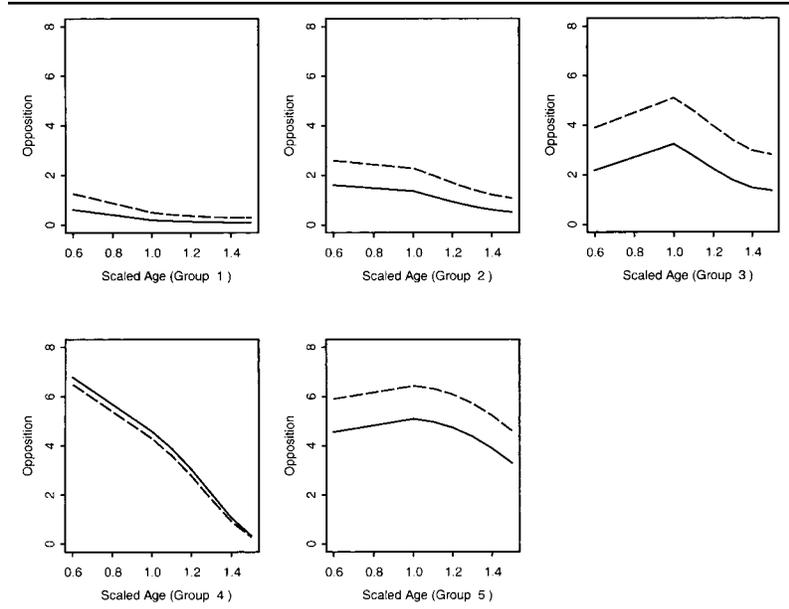


Figure 9: Expected Opposition Trajectories for Subjects Who Have Never Been Held Back (solid lines) Versus Subjects Who Have Always Been Behind Grade (dashed line)

TABLE 2: Interpretation of $2\log_e(B_{10})$

| $2\log_e(B_{10})$ | (B_{10}) | Evidence Against H_0 |
|-------------------|------------|------------------------|
| 0 to 2 | 1 to 3 | Not worth mentioning |
| 2 to 6 | 3 to 20 | Positive |
| 6 to 10 | 20 to 150 | Strong |
| > 10 | > 150 | Very strong |

hypothesis (i.e., three components versus more than three components) is on the boundary of the parameter space, and hence the classical asymptotic results do not hold (Ghosh and Sen 1985). To circumvent this problem, we follow the lead of D’Unger et al. (1998) and use the change in the BIC between models as an approximation to the log of the Bayes factor (Kass and Wasserman 1995). Keribin (1997) demonstrated that, under certain conditions, this approximation is valid for testing the number of components in a mixture. Raftery (1995) and Kass and Raftery (1995) are good references for Bayes factors. Also,

TABLE 3: Tabulated Bayesian Information Criterion (BIC) and $2\log_e(B_{10})$ (opposition data)

| <i>Number of Groups</i> | <i>BIC</i> | <i>Null Model</i> | $2\log_e(B_{10})$ |
|-------------------------|------------|-------------------|-------------------|
| 1 | -12,524.06 | | |
| 2 | -11,818.92 | 1 | 1,410.28 |
| 3 | -11,685.81 | 2 | 266.22 |
| 4 | -11,683.27 | 3 | 5.08 |
| 5 | -11,669.70 | 4 | 27.14 |
| 6 | -11,678.51 | 5 | -17.62 |

Fraley and Raftery (1998) address the use of Bayes factors in model-based clustering. The Bayes factor (B_{10}) gives the posterior odds that the alternative hypothesis is correct when the prior probability that the alternative hypothesis is correct equals one-half.

The BIC (Schwarz 1978), the log-likelihood evaluated at the maximum likelihood estimate less one-half the number of parameters in the model times the log of the sample size, tends to favor more parsimonious models than likelihood ratio tests when used for model selection. To maintain consistent usage with that of Jeffreys (1961) and Kass and Raftery (1995), we use the BIC log Bayes factor approximation,

$$2\log_e(B_{10}) \approx 2(\Delta\text{BIC}), \quad (7)$$

where ΔBIC is the BIC of the alternative (more complex) model less the BIC of the null (simpler) model. The log form of the Bayes factor is interpreted as the degree of evidence favoring the alternative model (see Table 2).

Table 3 tabulates the BIC for model fits to the oppositional behavior data. Based on the results, the five-group model is favored.

CONCLUSION

We demonstrated the use of a new SAS procedure that we wrote to analyze longitudinal data by fitting a mixture model. We illustrated the use of the TRAJ procedure through applications to psychometric scale data (oppositional behavior) using the censored normal mixture, offense counts using the ZIP mixture, and an offense prevalence

measure using the logistic mixture. Time-stable covariates (risk factors) were incorporated into the model by assuming that the risk factors are independent of the developmental trajectories, given group membership. A time-dependent covariate can also directly affect the observed behavior trajectory. In addition, the use of the BIC to address the problem of model selection, including the estimation of the number of components in the mixture, was demonstrated. While we focused on applications from research on antisocial behavior, any application that proposes to differentiate observations by type or category can be analyzed by our method. The procedure, with online documentation, is available from the authors free of charge at <http://lib.stat.cmu.edu/~bjones/traj.html>.

REFERENCES

- Bryk, Anthony S. and Stephen W. Raudenbush. 1987. "Application of Hierarchical Linear Models to Assessing Change." *Psychology Bulletin* 101:147-58.
- . 1992. *Hierarchical Linear Models for Social and Behavioral Research: Application and Data Analysis Methods*. Newbury Park, CA: Sage.
- Clogg, Clifford C. 1995. "Latent Class Models." In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. New York: Plenum.
- Dennis, John E., David M. Gay, and Roy E. Welsch. 1981. "An Adaptive Nonlinear Least-Squares Algorithm." *ACM Transactions on Mathematical Software* 7:348-83.
- Dennis, John E. and Howell W. Mei. 1979. "Two New Unconstrained Optimization Algorithms Which Use Function and Gradient Values." *Journal of Optimization Theory and Applications* 28:453-83.
- D'Unger, Amy V., Kenneth C. Land, Patricia L. McCall, and Daniel S. Nagin. 1998. "How Many Latent Classes of Delinquent/Criminal Careers? Results From Mixed Poisson Regression Analyses of the London, Philadelphia, and Racine Cohorts Studies." *American Journal of Sociology* 103:1593-630.
- Farrington, David P. and Donald J. West. 1990. "The Cambridge Study in Delinquent Development: A Prospective Longitudinal Study of 411 Males." In *Criminality: Personality, Behavior, and Life History*, edited by Hans-Jürgen Kerner and G. Kaiser. New York: Springer-Verlag.
- Fergusson, David M., Michael T. Lynskey, and L. John Horwood. 1996. "Factors Associated With Continuity and Change in Disruptive Behavior Patterns During Childhood and Adolescence." *Journal of Abnormal Child Psychology* 24:533-53.
- Fraley, Chris and Adrian E. Raftery. 1998. "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis." *Computer Journal* 41:578-88.
- Ghosh, Jayanta K. and Pranab K. Sen. 1985. "On the Asymptotic Performance of the Log Likelihood Ratio Statistic for the Mixture Model and Related Results." In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, vol. 3, edited by Lucien M. LeCam and Richard A. Olshen. Monterey, CA: Wadsworth.

- Goldstein, Harvey. 1995. *Multilevel Statistical Models*. 2d ed. London: Arnold.
- Jeffreys, Harold. 1961. *Theory of Probability*. 3d ed. London: Oxford University Press.
- Kass, Robert E. and Adrian E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90:773-95.
- Kass, Robert E. and Larry Wasserman. 1995. "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion." *Journal of the American Statistical Association* 90:928-34.
- Keribin, Christine. 1997. "Consistent Estimation of the Order of Mixture Models." Working Paper No. 61. Laboratoire Analyse et Probabilité, Université d'Évry-Val d'Essonne, Évry, France.
- Lambert, Diane. 1992. "Zero-Inflated Poisson Regressions, With an Application in Manufacturing." *Technometrics* 34:1-13.
- Land, Kenneth C., Patricia McCall, and Daniel S. Nagin. 1996. "A Comparison of Poisson, Negative Binomial, and Semiparametric Mixed Poisson Regression Models With Empirical Applications to Criminal Careers Data." *Sociological Methods & Research* 24:387-440.
- Laub, John H., Daniel S. Nagin, and Robert J. Sampson. 1998. "Good Marriages and Trajectories of Change in Criminal Offending." *American Sociological Review* 63:225-38.
- Loeber, Rolf and Marc LeBlanc. 1990. "Toward a Developmental Criminology." In *Crime and Justice: An Annual Review of Research*, vol. 12, edited by Michael Tonry and Norval Morris. Chicago: University of Chicago Press.
- Meredith, William and John Tisak. 1990. "Latent Curve Analysis." *Psychometrika* 55(1):107-22.
- Moffitt, Terrie E. 1993. "Adolescence-Limited and Life-Course Persistent Antisocial Behavior: A Developmental Taxonomy." *Psychological Review* 100:674-701.
- Muthen, Bengt O. 1989. "Latent Variable Modeling in Heterogeneous Populations." *Psychometrika* 54(4):557-85.
- Nagin, Daniel S. 1999. "Analyzing Developmental Trajectories: A Semi-Parametric, Group-Based Approach." *Psychological Methods* 4:139-77.
- Nagin, Daniel S., David P. Farrington, and Terrie E. Moffitt. 1995. "Life-Course Trajectories of Different Types of Offenders." *Criminology* 33:111-39.
- Nagin, Daniel S. and Kenneth C. Land. 1993. "Age, Criminal Careers, and Population Heterogeneity: Specific Estimation of a Nonparametric, Mixed Poisson Model." *Criminology* 31:327-62.
- Nagin, Daniel S. and Richard E. Tremblay. 1999. "Trajectories of Boys' Physical Aggression, Opposition, and Hyperactivity on the Path to Physically Violent and Non Violent Juvenile Delinquency." *Child Development* 70:1181-96.
- Patterson, Gerald R. 1996. "Some Characteristics of a Developmental Theory for Early-Onset Delinquency." In *Frontiers of Developmental Psychopathology*, edited by Mark F. Lenzenweger and Jeffrey J. Haugaard. Oxford, UK: Oxford University Press.
- Patterson, Gerald R., Barbara D. DeBaryshe, and E. Ramsey. 1989. "A Developmental Perspective on Antisocial Behavior." *American Psychologist* 44:329-35.
- Patterson, Gerald R., Marion S. Forgatch, Karen L. Yoerger, and Mike Stoolmiller. 1998. "Variables That Initiate and Maintain an Early-Onset Trajectory for Juvenile Offending." *Development and Psychopathology* 10:531-47.
- Patterson, Gerald R. and Karen L. Yoerger. 1997. "A Developmental Model for Late-Onset Delinquency." In *Motivation and Delinquency*, edited by D. Wayne Osgood. Lincoln: University of Nebraska Press.
- Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research (With Discussion)." In *Sociological Methodology*, edited by Peter V. Marsden. Cambridge, MA: Blackwell.

- Roeder, Kathryn, Kevin G. Lynch, and Daniel S. Nagin. 1999. "Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology." *Journal of the American Statistical Association* 94:766-76.
- Sampson, Robert J. and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.
- Schwarz, Gideon 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6:461-64.
- Willett, John B. and Aline G. Sayer. 1994. "Using Covariance Structure Analysis to Detect Correlates and Predictors of Individual Change Over Time." *Psychological Bulletin* 116(2):363-81.

Bobby L. Jones is a Ph.D. candidate in the Department of Statistics at Carnegie Mellon University. He is currently working on his dissertation, "Analyzing Longitudinal Data With Latent Class Models." He is the coauthor (with Shohini Ghose, James P. Clemens, Perry R. Rice, and Leno M. Pedrotti) of "Photon Statistics of a Single Atom Laser," which appeared in Physics Review A (1999).

Daniel S. Nagin is the Teresa and H. John Heinz III Professor of Public Policy at the H. John Heinz III School of Public Policy and Management, Carnegie Mellon University. He has written widely on deterrence, developmental trajectories and criminal careers, tax compliance, and statistical methodology. His recent publications include "Analyzing Developmental Trajectories: A Semi-Parametric, Group-Based Approach" in Psychological Methods (1999) and "Trajectories of Boys' Physical Aggression, Opposition, and Hyperactivity on the Path to Physically Violent and Nonviolent Juvenile Delinquency" (with Richard E. Tremblay) in Child Development (1999).

Kathryn Roeder is professor of statistics at the Carnegie Mellon University. Her research has focused on the development of statistical methodology for the analysis of heterogeneous data using mixture models and semiparametric methods. She is interested in criminology and the genetic basis of psychiatric disorders. Recent publications include "Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology" (with Kevin G. Lynch and Daniel S. Nagin) in the Journal of the American Statistical Association (1999) and "Genomic Control for Association Studies" (with Bernie Devlin) in Biometrics (1999).