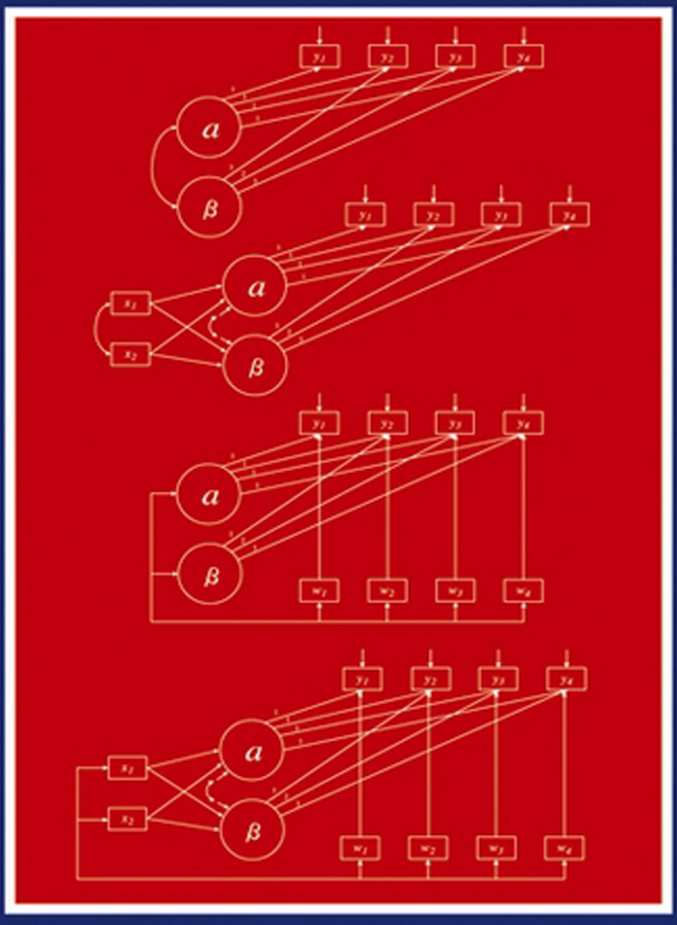
ftp://
SITE AVAILABLE

WILEY

Latent Curve Models

A Structural Equation Perspective



Kenneth A. Bollen

Patrick J. Curran

WILEY SERIES IN PROBABILITY AND STATISTICS

Latent Curve Models

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume.

Latent Curve Models

A Structural Equation Perspective

KENNETH A. BOLLEN

University of North Carolina
Department of Sociology
Chapel Hill, North Carolina

PATRICK J. CURRAN

University of North Carolina
Department of Psychology
Chapel Hill, North Carolina

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Bollen, Kenneth A.
Latent curve models: a structural equation perspective / Kenneth A. Bollen, Patrick J. Curran.
p. cm.—(Wiley series in probability and statistics)
Includes bibliographical references and index.
ISBN-13 978-0-471-45592-9 (cloth)
ISBN-10 0-471-45592-X (cloth)
1. Latent structure analysis. 2. Latent variables. I. Curran, Patrick J., 1965- II. Title.
QA278.6.B65 2005
519.5′35—dc22

2005047028

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

Preface

The last 20 years has witnessed a greater accumulation of longitudinal data in the social sciences than any other period. Accompanying this growth of data has been an interest in methods to analyze such data. Growth curve models or latent curve models have been among some of the most recently popular longitudinal techniques. Although new to some researchers, the historic roots of these techniques are deep. Indeed, these models have their origins in biostatistics, statistics, demography, and other disciplines. Their different disciplinary origins have given the same technique different orientations.

Our book is about growth curve models as they have developed in the latent variable and factor analysis tradition. Borrowing from Meredith and Tisak (1984, 1990), we refer to these as latent curve models (LCMs). They are models in which we have random intercepts and random slopes that permit each case in the sample to have a different trajectory over time. The random coefficients are incorporated into structural equation models (SEMs) by considering them as latent variables. By so doing, we are able to capitalize on all of the strengths of SEMs and to apply them to latent curve models. These include the ability to use maximum likelihood techniques for missing data, to estimate a variety of nonlinear forms of trajectories, to have measures of model fit and diagnostics to determine the source of ill-fit, the inclusion of latent covariates and latent repeated variables, and so on.

One of our goals is to provide a reference source on the LCM approach to longitudinal data. This is a field that is undergoing rapid development, but we do our best to provide many of the most recent developments. The book synthesizes a large amount of work on LCMs as well as providing original results. For instance, we develop new results on the identification of several types of LCMs. We include some of our recently proposed work on the autoregressive latent trajectory (ALT) model and suggest new models for method factors in multiple indicator, repeated latent variable models. Researchers with a background in structural equation models and an interest in longitudinal data are ideally suited for this book.

We have said what the book is about. It also is worthwhile to state what it is not. It is *not* a comprehensive guide to all approaches to growth curve models. There is a valuable literature in biostatistics on growth curve models and another tradition from multilevel models. There is a great deal of overlap in these approaches and

LCMs, and in parts of the book we point out the connections and sometimes the equivalency of models and results. However, a reader looking for comprehensive treatments of growth curve models from these other traditions will not find it here. Those seeking an extensive treatment of these models from a structural equation perspective are our intended audience.

Our book is a multiyear project that would not have happened without the help of numerous people. The most important group has been the Carolina Structural Equation Modeling (CSEM) group that we co-direct. This rotating group of intelligent graduate students and postdocs has been indispensible. They provided feedback on all of the chapters. They provided research assistance in developing empirical examples and a critical audience with whom to discuss results. We consider ourselves incredibly fortunate to have had such a talented group with which to work.

Personal support is at least as important as the professional support in making this book happen. We are lucky to have wives who provide both. We also have the good fortune to each have two daughters who helped us to maintain our perspective on the important things in life. We dedicate the book to our families.

KENNETH A. BOLLEN
PATRICK J. CURRAN

Chapel Hill, North Carolina

CHAPTER ONE

Introduction

Every discipline in the social sciences seeks to understand the sources of stability and change in variables. A psychologist, for instance, maps the development of cognitive abilities in children. An economist traces the economic growth of nations or the sales of companies. A sociologist tracks crime rates across communities. The time interval for change might be daily, weekly, monthly, or annually. The unit of analysis might be children, businesses, or countries. Regardless of the subject area or the time interval, social and behavioral scientists have a keen interest in describing and explaining the time trajectories of their variables.

Although the analysis of longitudinal data has a long and rich history in the social sciences, the past several decades have witnessed a sharp increase in interest in this topic. This resurgence of concern has a variety of sources. First, there is a growing appreciation of the limits of relying solely on cross-sectional data. Many theoretical hypotheses inherently focus on change, and longitudinal data are necessary to fully evaluate such hypotheses. Second, given the wider appreciation of the advantages of longitudinal research designs, there is much greater availability of panel data. Many longitudinal data sets exist that follow children, adults, communities, or countries over many time periods. The increased availability of large and complex longitudinal data sets demands the development of increasingly rigorous statistical analytic methods. Finally, there seems to be greater dissatisfaction with many traditional longitudinal analytic designs. For example, a developmental theory may predict that delinquent behavior develops naturally in children over time and a treatment intervention is designed to deflect this natural trajectory meaningfully. Using a statistical model that evaluates the treatment effect by comparing pre- and posttest group means may not provide an optimal test of the theory that hypothesized change in terms of individual developmental trajectories. Analytical methods are desired that allow for a better correspondence between the theoretical model that gave rise to the research hypotheses and the statistical model that is used to test these hypotheses empirically. For these and many other reasons, the analysis of longitudinal data has become an increasingly important topic in social science research.

Latent Curve Models: A Structural Equation Perspective, by Kenneth A. Bollen and Patrick J. Curran

There is a remarkable array of analytical techniques available for longitudinal empirical data. Examples of these methods include autoregressive models, repeated measures multivariate analysis of variance (MANOVA), raw difference scores, residualized change scores, and random and fixed effects panel data models. Rarely can we conclude that one of these models is inherently correct or incorrect. Rather, these models differ in how well they capture the theoretical orientation guiding the research and how well their assumptions correspond to the characteristics of the empirical data. From this perspective, a particular analytic approach may be well suited for one application but quite poorly suited to another. In this book, the focus is on methods to analyze the temporal trajectories of cases where each case can have a distinct trajectory.

The goal of this chapter is to orient the reader to this type of longitudinal model and to provide background on its origins. In the next section we introduce the conceptualization of these models by contrasting it with that of autoregressive methods and by introducing an empirical data set. In the third section we discuss three basic questions that characterize nearly all trajectory analyses. A brief review of the history of these methods follows, and a discussion of the organization of the rest of the book concludes the chapter.

1.1 CONCEPTUALIZATION AND ANALYSIS OF TRAJECTORIES

The majority of traditional analytical methods focus on change by examining the relation between two adjacent time points and predicting change using parameters that are common across cases. For example, the simple autoregressive model mentioned above is

$$y_{it} = \alpha_t + \rho_{t,t-1} y_{i,t-1} + \epsilon_{it} \tag{1.1}$$

where y_{it} and $y_{i,t-1}$ are the observed values of y for individual i at time t and time $t-1$, respectively; α_t is the time-specific intercept; and $\rho_{t,t-1}$ is the autoregressive coefficient pooled over all cases in the sample. The disturbance (ϵ_{it}) has a mean of zero, is not autocorrelated, and is uncorrelated with $y_{i,t-1}$. The i subscript indexes the cases in the sample, and the t subscript indicates time. Notice that the autoregressive parameter, $\rho_{t,t-1}$, has subscripts that permit it to change over time, but that for a given time point it takes the *same value* for all cases. We can modify Eq. (1.1) to include additional lagged values of y (e.g., the influence of $y_{i,t-2}$) or other explanatory variables with their coefficients (e.g., the influence of an exogenous measure $x_{i,t-1}$), but from this analytical perspective the forces governing change are its past values and the values of covariates.

This book represents a different perspective. The approach posits the existence of *continuous underlying* or *latent trajectories.* The pattern of change in the repeated measures provides information on the trajectories. *Latent* means a process that is not observed directly. The trajectory process is observed only indirectly using the repeated measures. Importantly, this trajectory can differ by individual case. For

example, say that we had assessed reading ability on a sample of children every other year over a period of eight years. A theory of cognitive development might predict that there exists an unobserved underlying ability to read that develops as a continuous function of time. Our interest in the time-specific measures of reading ability is because these repeated measures enable us to estimate the underlying reading trajectory that gives rise to the measures over time. The focus is not on the dependence of current reading level on past reading level across all children, but instead, on the estimation of an underlying developmental trajectory across all time points within *each* child. Our goal, then, will be to estimate a line, or trajectory, that best fits the repeated observations over time for each case. That is, we would like to smooth over our repeated observations to get a more parsimonious estimate of the underlying trajectory that gave rise to the repeated measures. It is this trajectory estimate that is of primary interest in subsequent modeling. An empirical example helps to further illustrate these issues.

1.1.1 Trajectories of Crime Rates

Our first empirical example consists of crime rates among 359 communities in New York State taken from 1995 *Uniform Crime Reports* (UCR) data.[1] The *crime rate* is the sum of the number of offenses reported to police departments for murder, robbery, aggravated assault, rape, burglary, motor vehicle theft, larceny, and arson converted to rates per 100,000 population. Taking the natural logarithm (ln) of this variable reduces its levels of skew and kurtosis. Crime rates spike upward on weekends, and computing bimonthly figures smoothes out the fluctuations that occur due to differing numbers of weekends within months (e.g., Hipp et al., 2004). Therefore, we use bimonthly crime rates per 100,000 population for the first eight months of 1995 for police units in New York State. Two predictors of these crime trajectories measured in 1990 are *poverty*, as measured by the percentage of the city population whose income is at or below 125% of the poverty level, and *population density*, formed as the community population divided by the square kilometers in the community.[2]

Table 1.1 presents the means, standard deviations, and correlations for the four repeated measures of ln(crime rates) taken on the 359 communities. Several characteristics can be observed. First, the correlations among the repeated measures are moderately high (ranging from 0.71 to 0.83). Lagged crime measures are clearly associated. Second, the means of the repeated measures increase over time, whereas the standard deviations do not have such a consistent trend. This trend reflects a tendency for crime rates to increase from the winter to the summer months. It will be important to select a statistical model that allows for these patterns in the means and the standard deviations.

[1]The UCR consists of reports on crime activity from all police units in the United States to the Federal Bureau of Investigation. These police units include city police, university police, and county sheriffs. In general, the UCR has a very high response rate, as police units covering about 97% of the population respond (U.S. Department of Justice, 1995).

[2]Both of these variables are taken from the 1990 U.S. Census Summary Tape File (STF) 3-A (U.S. Department of Justice, 2000).

Table 1.1 Means, Standard Deviations, and Correlations for Four Repeated Measures of ln(Crime Rate) for 359 Communities in New York State

	Time 1 Crime	Time 2 Crime	Time 3 Crime	Time 4 Crime
Time 1 crime	1.0			
Time 2 crime	0.833	1.0		
Time 3 crime	0.790	0.805	1.0	
Time 4 crime	0.709	0.728	0.819	1.0
Mean	5.319	5.516	5.611	5.763
Standard deviation	0.791	0.754	0.753	0.781

Table 1.2 Poverty, Population Density, and Four Waves of ln(Crime Rates) for Five Communities from the *Uniform Crime Reports*

Community	Poverty	Population Density	Time 1 Crime	Time 2 Crime	Time 3 Crime	Time 4 Crime
1	16.04	0.48	4.01	4.01	5.79	4.69
2	23.05	1.82	6.55	6.68	6.77	7.03
3	18.52	0.77	5.48	5.83	5.82	5.81
4	16.68	0.62	4.43	5.81	5.74	6.62
5	19.34	1.49	6.12	6.35	5.80	4.79

It is very important to consider not only the summary statistics on the repeated measures pooling over all cases, but also to closely examine the individual-level data. Table 1.2 presents the observed data on poverty, population density, and the four repeated measures of crime rates for the first five communities in the sample. These data allow us to consider several interesting questions. First, is there any monotonic increase or decrease in the crime rates over this period? Examining the raw data for the first four communities suggests an upward tendency, but it is far from consistent. For example, for the first community, the four repeated crime measures were 4.01, 4.01, 5.79, and 4.69, indicating no change for the first two times and a larger crime rate followed by one not as large as the third time. The fifth community exhibits a tendency to decrease. Examining just these five communities reveals potentially interesting characteristics of these crime rates across the communities.

Although it is interesting to begin with an examination of the time-specific measures of ln(crime rate) for each community taken at each time point, recall the posited underlying or latent trajectory of crime for each community. That is, it is of less interest that the first community received scores of 4.01, 4.01, 5.79, and 4.69 at each of the four time points than it is to use these repeated observations to estimate an underlying trajectory that *gave rise* to these specific repeated measures over time. A logical starting point is to consider fitting a line

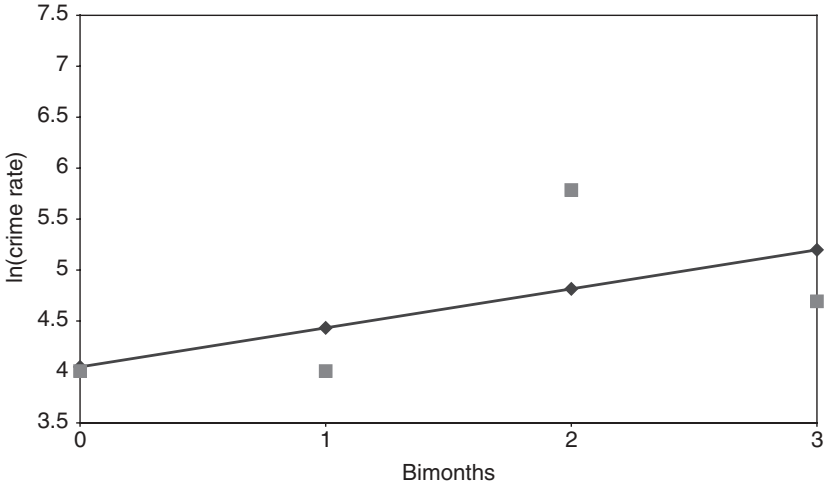


FIGURE 1.1 Observed and predicted crime rates for the first community.

that best fits the repeated measures observed within the first community. Chapter 2 treats this in much greater detail, but we can use ordinary least-squares (OLS) regression to estimate this line (or trajectory) for the first community. The OLS regression line relating ln(crime rate) to time within the first community is presented in Figure 1.1.

Although the repeated observations did not fall on a perfectly straight line, we are estimating the existence of the trajectory for the first community by finding a line that best fits the first community's four bimonthly data points. That is, the line in Figure 1.1 is nothing more than the regression of ln(crime rate) on time within the first community. As with any line, it is characterized completely by an intercept (say, $\widehat{\alpha}$) and a slope (say, $\widehat{\beta}$). Because the initial assessment (January–February) is set equal to zero, the intercept represents the model-implied value of ln(crime rate) at the initial time period.[3] Keep in mind that the $\widehat{\alpha}$ and $\widehat{\beta}$ values characterize the first community. This means that the first community has its *own* estimated intercept term and its *own* estimated slope term. This trajectory (sometimes called a *growth curve* or a *time path*) is an estimate of *intracommunity* (or *within-community*) change for the first community. We have thus *smoothed over* the four repeated measures and have a more parsimonious estimate of the process of change for the first community that is consistent with our theoretical model.

This provides an estimate of intracommunity change in crime rate for the first community, but also of interest are *intercommunity* differences in *intracommunity* change. That is, it is helpful to have a trajectory for *each* community (e.g., intracommunity change) and then to examine the characteristics of these trajectories

[3]We present a detailed discussion of alternative methods for coding time in Chapter 4.

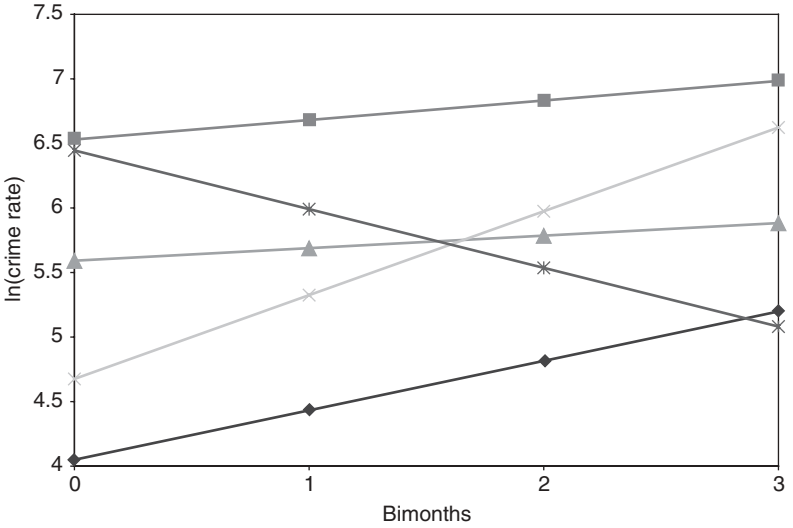


FIGURE 1.2 Model-implied linear trajectories of crime rates for the first five communities.

across all of the communities (e.g., intercommunity differences). To work toward this goal, a useful starting point is simply estimating a unique trajectory for each community using the same OLS regression method as above. So for 359 communities there would be 359 individual regression lines, one for each community. To demonstrate this, we fit individual regression lines to the first five communities in our data set, and these five trajectory estimates are displayed in Figure 1.2.

Figure 1.2 highlights several important concepts. First, the fitted trajectory lines are displayed, but the time-specific crime rates are not. The regression lines represent a *smoothed* version of the repeated observations over time, and a comparison of the observed points to the fitted line would reveal the quality of the fit of the linear trajectory to the observed data. Second, the estimated underlying trajectories are different for each of the five communities. So each line in Figure 1.2 represents the OLS regression of crime rate on time estimated separately for each community, and this necessitates a separate subscript to distinguish the intercept and slope for each community. These parameter estimates now become $\widehat{\alpha}_i$ and $\widehat{\beta}_i$, respectively, where $i = 1, 2, 3, 4, 5$ indicate which $\widehat{\alpha}$ and $\widehat{\beta}$ are associated with which specific community i.

Finally, given the existence of five estimated $\widehat{\alpha}$ values and five estimated $\widehat{\beta}$ values, it is possible to find the mean of the $\widehat{\alpha}$'s and the mean of the $\widehat{\beta}$'s. More specifically, combining these individual trajectory estimates allows estimation of the overall trajectory for the entire set of communities. Further, given the estimation of the mean trajectory, we can examine variability of each community's trajectory around these group means. This variability reflects that some communities are reporting higher levels of initial crime rates, while others are reporting lower initial levels; further, some communities are reporting steeper increases in crime over

time, while others are reporting less steep increases and even decreases. These estimates of variability in community trajectories around the group mean trajectory capture the presence of intercommunity differences in intracommunity change. The larger the magnitude of these variance estimates, the greater evidence there is for individual community variability in trajectories. In Chapter 2 we present these results in detail.

Thus far, the repeated observations allow estimation of a trajectory for each community and for estimating a mean trajectory for the group. The variability of the individual trajectories around this group mean is also available. This model is commonly referred to as an *unconditional trajectory model* because we estimate the means and variances of the trajectories but are not modeling the trajectories as a function of other predictor variables.

However, a key analytic goal is to determine if the individual variability of the trajectories is predictable using other community characteristics. That is, given the presence of significant variability in our set of $\widehat{\alpha}$ estimates, does the magnitude of these $\widehat{\alpha}$'s vary as a function of, say, community poverty? Are higher levels of poverty associated with higher initial levels of crime? Similarly, are higher levels of poverty associated with steeper increases in ln(crime rate) over time? Given that we are interested in the distributions of the $\widehat{\alpha}$ and $\widehat{\beta}$ estimates *conditioned on* one or more predictor variables (e.g., using other community measures to predict our individual trajectory estimates), this model is commonly referred to as a *conditional trajectory model*. We discuss methods by which covariates can predict these trajectory parameters in conditional growth models in Chapter 5.

1.1.2 Data Requirements

As we explicate more fully in Chapter 2, a minimum of three repeated observations are necessary to identify and to estimate underlying linear trajectories and the other parameters in the model. In general, to estimate a polynomial trajectory of degree d, it is necessary to have obtained $d + 2$ repeated observations for identification purposes. So three repeated measures permit a straight line (e.g., a polynomial of degree $d = 1$); four observations fit a quadratic trajectory ($d = 2$); and so on. Conceptually, this is easy to understand. With only two observations, the line of best fit simply connects the two observations and thus reflects a perfect fit of the trajectory to the data. In contrast, the line of best fit for three observations in general reflects some imperfect relation between time and the repeated variables and thus allows a test of the adequacy of model fit. Because of this, our emphasis is on data where there are at least three repeated observations for a large proportion of observed units.[4]

1.1.3 Summary

Conceptualizing stability and change in terms of individual trajectories allows for the articulation and assessment of a wide array of research questions that are not

[4]As we discuss in Chapter 3, we can have missing data for some observations. So it is possible to estimate these models even if fewer than three repeated measures are available for some cases. See Chapter 3 for details.

easily accommodated with other techniques. For example, each case can have a separate trajectory, and a mean group trajectory and variability around the mean are available. We can then incorporate predictor variables and model individual differences in trajectories across cases.

As we explore throughout the book, many interesting and complex questions can be posed about stability and change in a variable. As a starting point, it is helpful to develop a general organizing framework to begin thinking about these many questions of change over time.

1.2 THREE INITIAL QUESTIONS ABOUT TRAJECTORIES

Our focus is on the trajectories of variables across cases. Although many questions about the trajectories are possible, three questions are common in many latent trajectory analyses. These questions relate to (1) the characteristics of the mean trajectory of the entire group, (2) the evaluation of individual differences in trajectories, and (3) the potential incorporation of predictors of individual differences in trajectories. To give substance to these issues, we return to our ln(crime rate) example. Recall that we have four repeated measures of crime rates over eight months for 359 communities in New York State. Further, we have time 1 measures of the percent poverty and the population density for each community. We propose considering three initial questions about the trajectories underlying crime rates.

1.2.1 Question 1: What Is the Trajectory for the Entire Group?

This initial question simply asks: When pooling over all 359 communities in the New York sample, what is the mean trajectory of crime over the four bimonthly measurements? That is, what is the average initial level of the crime rate trajectory? On average, does crime change as a function of time? If so, what is the functional form of change over time? Is this rate of change strictly linear, or is there a curvilinear component in which the rate of change accelerates or decelerates with the passage of time? Evaluation of this first question thus provides an initial insight into the characteristics of the mean trajectory of the crime rates for the *entire sample* of New York communities under study. The mean values of the trajectory parameter estimates are sometimes referred to as the *fixed-effects components* of the trajectory model.

1.2.2 Question 2: Do We Need Distinct Trajectories for Each Case?

Understanding the mean trajectory of crime rates is important, but there is additional information yet to be gained. Of particular interest is whether there is meaningful variability of community trajectories around the mean trajectory. That is, is the mean trajectory reflective of every community in the sample, or are there cases that depart from the group-level trajectory? Depending on our sampling design, we may find that although the mean initial level of crime rate for the group is

a particular value, some communities are reporting initial values well above this, whereas others are well below. Further, although the mean rate of change in crime rates for the group is a specific number, some communities may be increasing at a more rapid rate and some are not increasing at all. The degree to which it is necessary to allow each community to take on a distinct trajectory is a question of *intercommunity* differences in *intracommunity* change. If all communities are within sampling fluctuations of the mean trajectory, there is no evidence of community differences in trajectories; the mean trajectory is thus representative of the trajectory of every community. In contrast, the more the individual community trajectories diverge from the overall group, the greater evidence there is for individual differences in trajectories of change. These variances are sometimes referred to as the *random-effects components* of the trajectory model.

1.2.3 Question 3: If Distinct Trajectories Are Needed, Can We Identify Variables to Predict These Individual Trajectories?

Say that the response to Question 1 indicates that there is a mean trajectory of crime rates of some particular form (e.g., an overall linear trajectory). Further, say that the response to Question 2 indicates that there is meaningful variability of the individual trajectories around the mean intercept and mean slope. The third question addresses whether we can incorporate other explanatory variables to better understand the variability observed in individual trajectories. Do some variables predict which communities start with a higher crime rate relative to those communities that start at lower levels? Similarly, can we use a set of explanatory measures to understand what type of community will increase more steeply in crime relative to those that increase less steeply? This third and final question thus allows us to explore our ability to model individual variability in trajectories and to make an attempt at predicting which cases will start high versus low and which will increase steeply and which will not. The incorporation of predictors leads to the *conditional* trajectory model to which we referred previously.

1.2.4 Summary

There are many questions that we could pose when evaluating trajectories of change. However, three initial questions are a helpful starting point to understanding trajectory analysis. The first question allows us insight into group-level change, the second question explores the presence of individual variability in change, and the third question attempts to model individual differences in change as a function of other explanatory variables of interest. Taken together, this allows us to begin working toward the development of a comprehensive understanding of the individual and group processes of change.

1.3 BRIEF HISTORY OF LATENT CURVE MODELS

The methods that we discuss in this book are not all new. In fact, parts of the methodology that we present have roots more than a century old. In this section

we sketch some of the major developments that have led to the contemporary latent curve models (LCMs) that are the focus of our book. It is not our intention to present an exhaustive literature review or to present the background to all approaches to analyzing growth curve models. Rather, our goal is to focus on LCMs and to show that they derive from over a century of tradition of studying individual and group change. Even the idea of using latent variables or factor analysis to study such change has roots that are nearly a half century old. See Bollen (2005) for a more detailed discussion of the relation between growth curve models and factor analysis. Thus, the LCM is newly popular, but it is not a new idea. Below we identify several time periods that correspond to developments that contributed to the contemporary form of LCM.

1.3.1 Early Developments: The Nineteenth Century

Although philosophical discussions of change have been traced back as far as Aristotle (Zeger and Harlow, 1987), the earliest development of statistical analysis of data over time appears to be early in the nineteenth century. The studies were oriented toward the change in aggregates or groups rather than in individuals. Gompertz (1820) presents one of the first discussions of modeling stability and change from a trajectory perspective. The key development of Gompertz (1820, 1825) was the method of estimating a single trajectory that would characterize continuous mortality rates over time for a group of individuals. The general goal of this work was to use mathematical modeling to estimate the "nature of the function of mortality" (p. 230), a topic that today we would classify as a topic of demography.

In the early nineteenth century, predictions of mortality were based on extrapolating values from actuarial tables which were prone to significant error. Gompertz (1820) proposed using polynomial mortality curves to provide better estimates of predicted mortality rates. He stated that ". . . besides the usual mode of . . . working from calculated tables . . . subject to the errors of numerous interpolations . . . it may be easier to work them directly without such reduction, or by reducing the terms to geometrical progressions for short periods . . ." (p. 235). Further, he considered his proposed methods to be rather general, such that his object was to ". . .propose a plan of analysis and notation, which I conceive may be applied with utility to most problems likely to occur, and capable of suggesting a variety of new speculations in the pursuit of this science" (p. 216). This is the earliest application that we have identified that addresses what we might consider to be "trajectory analysis" with the focus on finding "laws" of development for a group of individuals.

During this same time period, Verhulst (1845, 1847) was working on a similar problem. Specifically, Verhulst was attempting to develop mathematical models of population growth over time, and he proposed that population increases may best be described using a logarithmic function of time. As with Gompertz, Verhulst was primarily interested in discovering "laws" of population growth over time, and he, too, focused on a demographic topic. That is, his focus was on the trajectory of

population growth for the entire group, and he did not consider variability between individual trajectories.

Finally, Quetelet (1835) focused on mathematical models of growth of the human body. He opens his 1835 paper by stating, "Man is born, grows up, and dies, according to certain laws which have never been properly investigated" (p. 317). A key difference between Quetelet and other early theorists was that in addition to seeking lawful developmental processes, he also expressed an interest in difference in human development as a function of other mitigating influences. For example, he describes the relation between physical development and factors such as socioeconomic status and gender. However, he still did not explicitly move from the desire to find lawful relations over time into the estimation of interindividual differences in intraindividual change.

In sum, there are clear lines of development of trajectory analysis that can be traced back to the early nineteenth century. The work of Gompertz, Verhulst, and Quetelet combine to provide the theoretical foundation upon which later developments in trajectory modeling rest.

1.3.2 Fitting Group Trajectories: 1900–1937

Studies from the early twentieth century continued to focus on aggregate change. The majority of these early applications utilized rather complex functional forms of growth (e.g., nonlinear polynomials and logistic curves) to examine change for an entire group of individuals. That is, these early works attempted to address the *group-level trajectory* described in Question 1. For example, Robertson (1908) used logistic curves to model the growth of white rats and then attempted to apply these same models to the growth of humans. He concluded that human development required multiple logistic curves superimposed upon one another to adequately characterize the patterns of change. As with earlier theorists, Robertson was working under the assumption that there is an underlying law of growth that is to be discovered and modeled.

Reed, Pearl, and colleagues drew upon Robertson's early work as they evaluated developmental trajectories in problems as diverse as the weight of the human brain, national food consumption, influenza, and the population problem (e.g., Reed, 1921; Pearl, 1924, 1925; Reed and Pearl, 1927; Reed and Love, 1932). Their goal was to develop mathematical equations to describe growth processes with greater accuracy than was possible with other methods available at that time. Although they believed that the fitting of an equation to summarize a growth process increased the accuracy of prediction, they also believed that a given mathematical model did not necessarily aid in an understanding of the causes and regulatory factors of growth.

In sum, the first third of the twentieth century gave rise to many alternative formulations of mathematical forms of trajectories applied to a wide array of biological and social science problems. Despite the advances made in the variety of functional forms of trajectories, the goal was still to estimate a single underlying trajectory that best characterized the process for the entire group under consideration.

1.3.3 Fitting Individual and Group Trajectories: 1938–1950s

Whereas the analytical developments and empirical applications of the early twentieth century focused primarily on the trajectory of change for the entire group, the 1930s gave rise to the significant development of estimating unique trajectories for each individual within the group and modeling these trajectories as a function of other predictor variables. Of key importance is the work of Wishart (1938), who applied growth curves to weight gain in bacon pigs. His focus was on an experimental design with pigs randomly assigned to diets of low, medium, or high protein content, and the trajectory of the weight of the pigs was the outcome of interest. Wishart fit by hand a parabolic equation to the weight of each pig over a 16-week period. He then used these pig-specific growth trajectory estimates as the dependent variables in an ANOVA model to predict the rate of growth from the experimental conditions and gender. To our knowledge, this is the first attempt to consider not only the developmental trajectory for the entire group of individual cases, but to treat the individual trajectory estimates as varying randomly over individual cases.

Some years later, Griliches (1957) used a similar approach to modeling the rate of growth of hybrid corn in various regions of the United States. Griliches fit S-curves to the percentage of hybrid corn acreage of each of the regions and computed the origin, slope, and ceiling of these trajectories. These trajectory estimates for each region were then regressed on various measures to predict growth and yield over time.

The work of Wishart (1938) and Griliches (1957) highlights the major shift in moving from focusing on the estimation of the trajectory for the entire group to estimating the specific trajectory for each individual in the group. This allowed both the examination of individual differences in trajectories and the prediction of trajectories by other explanatory variables.

1.3.4 Trajectory Modeling with Latent Variables: 1950s–1984

Until the early 1950s, all trajectory models were estimated using ANOVA, ANCOVA, and MANOVA types of analytic approaches. That is, the individual trajectory estimates were typically computed by hand, and differences in the means of the trajectory estimates as a function of group membership were then examined. A significant shift in this general analytic strategy occurred in the middle of the twentieth century, and several key developments moved the estimation of trajectory models into the latent variable framework. In this framework, the growth process is governed by processes that are not observed directly, only indirectly through the repeated measures.

To our knowledge, the first person to propose using latent variables within the factor analytic framework to do trajectory modeling was Baker (1954). Baker reasoned that a factor analytic model could “indicate separate influences operating on the relative growth of individuals. These separate influences may be identified as environmental, physiological, or genetical effects. In any case the structure of the growth phenomenon is clearly exhibited with respect to time” (p. 137).

JOHNSON and KOTZ · Distributions in Statistics
JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 1, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 2, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Discrete Multivariate Distributions
JOHNSON, KOTZ, and KEMP · Univariate Discrete Distributions, *Third Edition*
JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
JUREČKOVÁ and SEN · Robust Statistical Procedures: Aymptotics and Interrelations
JUREK and MASON · Operator-Limit Distributions in Probability Theory
KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
KADANE AND SCHUM · A Probabilistic Analysis of the Sacco and Vanzetti Evidence
KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data, *Second Edition*
KASS and VOS · Geometrical Foundations of Asymptotic Inference
† KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster Analysis
KEDEM and FOKIANOS · Regression Models for Time Series Analysis
KENDALL, BARDEN, CARNE, and LE · Shape and Shape Theory
KHURI · Advanced Calculus with Applications in Statistics, *Second Edition*
KHURI, MATHEW, and SINHA · Statistical Tests for Mixed Linear Models
* KISH · Statistical Design for Research
KLEIBER and KOTZ · Statistical Size Distributions in Economics and Actuarial Sciences
KLUGMAN, PANJER, and WILLMOT · Loss Models: From Data to Decisions, *Second Edition*
KLUGMAN, PANJER, and WILLMOT · Solutions Manual to Accompany Loss Models: From Data to Decisions, *Second Edition*
KOTZ, BALAKRISHNAN, and JOHNSON · Continuous Multivariate Distributions, Volume 1, *Second Edition*
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Volumes 1 to 9 with Index
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Supplement Volume
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 1
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 2
KOVALENKO, KUZNETZOV, and PEGG · Mathematical Theory of Reliability of Time-Dependent Systems with Practical Applications
LACHIN · Biostatistical Methods: The Assessment of Relative Risks
LAD · Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction
LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*
LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE · Case Studies in Biometry
LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*
LAWLESS · Statistical Models and Methods for Lifetime Data, *Second Edition*
LAWSON · Statistical Methods in Spatial Epidemiology
LE · Applied Categorical Data Analysis
LE · Applied Survival Analysis

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

LEE and WANG · Statistical Methods for Survival Data Analysis, *Third Edition*
LEPAGE and BILLARD · Exploring the Limits of Bootstrap
LEYLAND and GOLDSTEIN (editors) · Multilevel Modelling of Health Statistics
LIAO · Statistical Group Comparison
LINDVALL · Lectures on the Coupling Method
LINHART and ZUCCHINI · Model Selection
LITTLE and RUBIN · Statistical Analysis with Missing Data, *Second Edition*
LLOYD · The Statistical Analysis of Categorical Data
LOWEN and TEICH · Fractal-Based Point Processes
MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in Statistics and Econometrics, *Revised Edition*
MALLER and ZHOU · Survival Analysis with Long Term Survivors
MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel
MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of Reliability and Life Data
MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets
MARCHETTE · Random Graphs for Statistical Pattern Recognition
MARDIA and JUPP · Directional Statistics
MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with Applications to Engineering and Science, *Second Edition*
McCULLOCH and SEARLE · Generalized, Linear, and Mixed Models
McFADDEN · Management of Data in Clinical Trials
* McLACHLAN · Discriminant Analysis and Statistical Pattern Recognition
McLACHLAN, DO, and AMBROISE · Analyzing Microarray Gene Expression Data
McLACHLAN and KRISHNAN · The EM Algorithm and Extensions
McLACHLAN and PEEL · Finite Mixture Models
McNEIL · Epidemiological Research Methods
MEEKER and ESCOBAR · Statistical Methods for Reliability Data
MEERSCHAERT and SCHEFFLER · Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice
MICKEY, DUNN, and CLARK · Applied Statistics: Analysis of Variance and Regression, *Third Edition*
* MILLER · Survival Analysis, *Second Edition*
MONTGOMERY, PECK, and VINING · Introduction to Linear Regression Analysis, *Third Edition*
MORGENTHALER and TUKEY · Configural Polysampling: A Route to Practical Robustness
MUIRHEAD · Aspects of Multivariate Statistical Theory
MULLER and STOYAN · Comparison Methods for Stochastic Models and Risks
MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and Nonlinear Optimization
MURTHY, XIE, and JIANG · Weibull Models
MYERS and MONTGOMERY · Response Surface Methodology: Process and Product Optimization Using Designed Experiments, *Second Edition*
MYERS, MONTGOMERY, and VINING · Generalized Linear Models. With Applications in Engineering and the Sciences
† NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
† NELSON · Applied Life Data Analysis
NEWMAN · Biostatistical Methods in Epidemiology
OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
OKABE, BOOTS, SUGIHARA, and CHIU · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams, *Second Edition*
OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

PALTA · Quantitative Methods in Population Health: Extensions of Ordinary Regressions
PANKRATZ · Forecasting with Dynamic Regression Models
PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
* PARZEN · Modern Probability Theory and Its Applications
PEÑA, TIAO, and TSAY · A Course in Time Series Analysis
PIANTADOSI · Clinical Trials: A Methodologic Perspective
PORT · Theoretical Probability for Applications
POURAHMADI · Foundations of Time Series Analysis and Prediction Theory
PRESS · Bayesian Statistics: Principles, Models, and Applications
PRESS · Subjective and Objective Bayesian Statistics, *Second Edition*
PRESS and TANUR · The Subjectivity of Scientists and the Bayesian Approach
PUKELSHEIM · Optimal Experimental Design
PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics
† PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
QIU · Image Processing and Jump Regression Analysis
* RAO · Linear Statistical Inference and Its Applications, *Second Edition*
RAUSAND and HØYLAND · System Reliability Theory: Models, Statistical Methods, and Applications, *Second Edition*
RENCHER · Linear Models in Statistics
RENCHER · Methods of Multivariate Analysis, *Second Edition*
RENCHER · Multivariate Statistical Inference with Applications
* RIPLEY · Spatial Statistics
RIPLEY · Stochastic Simulation
ROBINSON · Practical Strategies for Experimenting
ROHATGI and SALEH · An Introduction to Probability and Statistics, *Second Edition*
ROLSKI, SCHMIDLI, SCHMIDT, and TEUGELS · Stochastic Processes for Insurance and Finance
ROSENBERGER and LACHIN · Randomization in Clinical Trials: Theory and Practice
ROSS · Introduction to Probability and Statistics for Engineers and Scientists
† ROUSSEEUW and LEROY · Robust Regression and Outlier Detection
* RUBIN · Multiple Imputation for Nonresponse in Surveys
RUBINSTEIN · Simulation and the Monte Carlo Method
RUBINSTEIN and MELAMED · Modern Simulation and Modeling
RYAN · Modern Regression Methods
RYAN · Statistical Methods for Quality Improvement, *Second Edition*
SALTELLI, CHAN, and SCOTT (editors) · Sensitivity Analysis
* SCHEFFE · The Analysis of Variance
SCHIMEK · Smoothing and Regression: Approaches, Computation, and Application
SCHOTT · Matrix Analysis for Statistics, *Second Edition*
SCHOUTENS · Levy Processes in Finance: Pricing Financial Derivatives
SCHUSS · Theory and Applications of Stochastic Differential Equations
SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
* SEARLE · Linear Models
SEARLE · Linear Models for Unbalanced Data
SEARLE · Matrix Algebra Useful for Statistics
SEARLE, CASELLA, and McCULLOCH · Variance Components
SEARLE and WILLETT · Matrix Algebra for Applied Economics
SEBER and LEE · Linear Regression Analysis, *Second Edition*
† SEBER · Multivariate Observations
† SEBER and WILD · Nonlinear Regression
SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
* SERFLING · Approximation Theorems of Mathematical Statistics

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

SHAFER and VOVK · Probability and Finance: It's Only a Game!
SILVAPULLE and SEN · Constrained Statistical Inference: Inequality, Order, and Shape Restrictions
SMALL and McLEISH · Hilbert Space Methods in Probability and Statistical Inference
SRIVASTAVA · Methods of Multivariate Statistics
STAPLETON · Linear Statistical Models
STAUDTE and SHEATHER · Robust Estimation and Testing
STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
STYAN · The Collected Papers of T. W. Anderson: 1943–1985
SUTTON, ABRAMS, JONES, SHELDON, and SONG · Methods for Meta-Analysis in Medical Research
TANAKA · Time Series Analysis: Nonstationary and Noninvertible Distribution Theory
THOMPSON · Empirical Model Building
THOMPSON · Sampling, *Second Edition*
THOMPSON · Simulation: A Modeler's Approach
THOMPSON and SEBER · Adaptive Sampling
THOMPSON, WILLIAMS, and FINDLAY · Models for Investors in Real World Markets
TIAO, BISGAARD, HILL, PEÑA, and STIGLER (editors) · Box on Quality and Discovery: with Design, Control, and Robustness
TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics
TSAY · Analysis of Financial Time Series, *Second Edition*
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
VAN BELLE · Statistical Rules of Thumb
VAN BELLE, FISHER, HEAGERTY, and LUMLEY · Biostatistics: A Methodology for the Health Sciences, *Second Edition*
VESTRUP · The Theory of Measures and Integration
VIDAKOVIC · Statistical Modeling by Wavelets
VINOD and REAGLE · Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments
WALLER and GOTWAY · Applied Spatial Statistics for Public Health Data
WEERAHANDI · Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models
WEISBERG · Applied Linear Regression, *Third Edition*
WELSH · Aspects of Statistical Inference
WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment
WHITTAKER · Graphical Models in Applied Multivariate Statistics
WINKER · Optimization Heuristics in Economics: Applications of Threshold Accepting
WONNACOTT and WONNACOTT · Econometrics, *Second Edition*
WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
WOODWORTH · Biostatistics: A Bayesian Introduction
WOOLSON and CLARKE · Statistical Methods for the Analysis of Biomedical Data, *Second Edition*
WU and HAMADA · Experiments: Planning, Analysis, and Parameter Design Optimization
YANG · The Construction Theory of Denumerable Markov Processes
* ZELLNER · An Introduction to Bayesian Inference in Econometrics
ZHOU, OBUCHOWSKI, and McCLISH · Statistical Methods in Diagnostic Medicine

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.