Growth Curve Models with Categorical Outcomes

Katherine E. Masyn[1], Hanno Petras[2] and Weiwei Liu[3]
[1]Harvard Graduate School of Education, Cambridge, MA, USA
[2]Research and Development, JBS International, North Bethesda, MD, USA
[3]NORC at the University of Chicago, Bethesda, MD, USA

Overview

Motivated by the limited available literature on the treatment of longitudinal binary and ordinal outcomes in a growth modeling framework, the goal of this entry is to provide an accessible and practical introduction of this topic for a criminological audience. The parameterization of categorical latent growth models is explained by integrating aspects of the more familiar conventional latent growth models and generalized linear models. Emphasis is placed on the process of model building, evaluation, and interpretation. The entry contains an elaboration of how to include predictors of developmental change in the model for covariate-related hypothesis tests along with remarks regarding of the importance of auxiliary information for assessing model validity and utility. Finally, several model extensions including nonlinear change, generalized growth mixture modeling, and longitudinal latent class analysis are discussed.

Introduction

Criminologists typically encounter data on crime and deviance which is skewed and discrete, thus violating the assumptions of ordinary least square (OLS) regression models, which require that the outcome variable is continuous and is (conditionally) normally distributed. Many criminological studies involve binary outcomes, such as arrest versus no arrest, or unordered categorical outcomes, such as judge and jury consensus or disagreement on conviction versus acquittal. Other outcomes consist of categories, which represent a natural ranking or ordering, such as offense severity or self-reported attitudinal items measured on a Likert scale ranging from strongly disagree to strongly agree. Finally, some outcomes in criminological inquiries consist of counts of a particular event, such as the number of police contacts or the number of arrests. Generalized linear models (GLM), originally formulated by Nelder and Wedderburn (1972), represent a flexible generalization of OLS regression to accommodate skewed and discrete outcomes in a regression framework. In addition to classic textbooks (e.g., Agresti 2002; Long 1997), ample guidance exists on direct applications of GLMs for cross-sectional binary, unordered, and ordered categorical outcomes (e.g., Britt and Weisburd 2010) as well as for count outcomes (e.g., MacDonald and Lattimore 2010) in criminological research.

Beyond the explanation and prediction of cross-sectional outcomes, describing and predicting the developmental course of individuals' involvement in criminal and antisocial behavior is a central theme in criminological inquiries. It is well-known that only repeated observations of individuals across time allow for explicit modeling of intra individual change processes and enable the charting of interindividual differences in intra individual age-crime curves including the manifest features of onset, continuation, and cessation in criminal activity (Piquero et al. 2007). In particular, longitudinal data permits proper inferences about stability and change in individual trajectories, differences across individuals with respect to their trajectories, and the predictive effects of time-invariant risk and vulnerability factors as well as time-dependent life events on those trajectories (Piquero 2008). Given the ever-growing interest in not only describing but testing hypotheses related to individual differences in criminal behavior across the life course, more researchers have endeavored to collect longitudinal data on samples of individuals and are making use of statistical models that effectively and appropriately utilize those repeated measures data. In the last two decades, latent growth modeling (LGM; also known as growth curve modeling, latent trajectory analysis, hierarchical linear modeling, linear mixed models, etc.) has emerged as the preferred analytical choice. This preference is in part due to the fact that LGM is more flexible than repeated measures analysis of variance or observed change score analysis in dealing with missing data, unequally spaced time points, complex nonlinear developments, and, importantly, non-normally distributed and discretely scaled repeated measures (Curran et al. 2010).

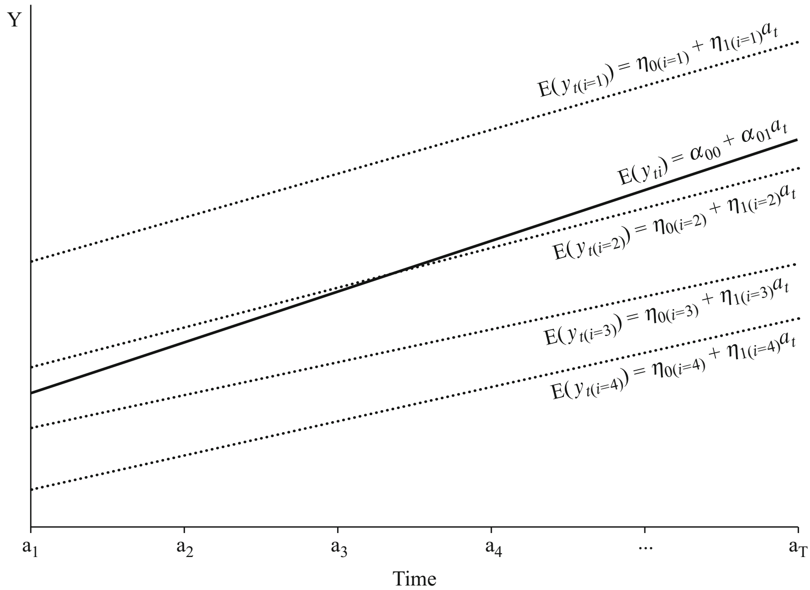
Along with their increasing popularity, vast resources have accumulated to instruct researchers in the application of LGMs with longitudinal continuous (e.g., Bollen and Curran 2006; Muthén 2004; Petras and Masyn 2010) and count outcomes (Kreuter and Muthén 2008; Nagin and Land 1993). With the exception of Feldman et al. (2009); Mehta et al. (2004); Muthén (1996); and Skrondal and Rabe-Hesketh (2004), the discussion regarding the treatment of binary and ordinal outcomes in a growth modeling framework is, by comparison, quite limited, especially for the applied criminology audience. Thus, the goal of this entry is to provide an accessible and practical introduction to the study of change using binary and ordinal repeated measures. The remainder of the entry is organized as follows: First, the conventional growth model for continuous outcomes is briefly introduced followed by a presentation of generalized linear models with binary and ordinal outcomes. Then the process of modeling repeated binary and ordinal measures in a latent growth modeling framework is advanced. This entry is concluded with a summative discussion including an overview of model extensions and alternatives.

Conventional Latent Growth Models

Modern growth models generally treat longitudinal outcomes in one of two ways: (1) as multilevel outcome data, where time or measurement occasions at "Level 1" are nested within persons at "Level 2"; or (2) as multivariate outcome data, where the repeated outcome measures are multiple indicators for latent growth factors values for each individual. Taking a multivariate approach, intra individual change is captured by the measurement model for the growth factors, describing the relationship between individual growth factor values and the observed outcomes over time, and interindividual differences are captured by the structural model, i.e., the mean and variance-covariance structure of the growth factors, describing the distribution of the growth factors in the population of individuals.

Although it is possible to specify analytically equivalent unconditional and conditional growth models across the multilevel and (multivariate) latent growth modeling frameworks, estimated via full-information maximum likelihood (FIML), utilizing the latent variable approach affords access to a variety of modeling extensions not as easily implemented in other frameworks, e.g., models that simultaneously include both antecedents and consequences of the change process, higher-order growth models with multiple indicators of the outcome at each assessment, multiprocess and multilevel growth models, and models that employ both continuous and categorical latent variables for describing population heterogeneity in the change process (for more on growth modeling in a latent variable framework, see, e.g., Bollen and Curran 2006; Muthén 2001, 2004). Given this greater flexibility, it is the multivariate approach to longitudinal data in a latent variable modeling framework that we focus on herein.

The latent growth model specification is a restricted form of a more general structural equation model (SEM; Kline 2010). In the SEM formulation of a latent growth model, there are T repeated measures, $y_t (t = 1, \ldots, T)$, that serve as the indicators or manifest variables, where T is the number of time points or waves during which study participants were assessed. For a *linear* latent growth curve model, there are two latent factors: an intercept growth factor, η_0, and a slope growth factor, η_1. The measurement and



Growth Curve Models with Categorical Outcomes, Fig. 1 Individual trajectories (*dotted lines*) for a hypothetical random sample of four individuals ($i = 1, 2, 3, 4$) drawn from a population with a mean growth trajectory given by the *solid line*

structural portions for an unconditional linear latent growth model are given by

$$\begin{aligned}&\text{Measurement model:}\\&\quad y_{ti} = \eta_{0i} + \eta_{1i}a_t + \varepsilon_{ti},\\&\text{Structural model:}\\&\quad \eta_{0i} = \alpha_{00} + \xi_{0i},\\&\quad \eta_{1i} = \alpha_{10} + \xi_{1i}.\end{aligned} \tag{1}$$

Here, y_{ti} is the observed outcome y for individual $i(i = 1, \ldots, n)$ at time $t(t = 1, \ldots, T)$, a_t is the time score at time t, η_{0i} is the random intercept factor (i.e., the expected outcome on y for individual i at time score $a_t = 0$), and η_{1i} is the random linear slope factor (i.e., the change in the expected outcome on y for individual i for a one unit increase in time, on the scale of a_t). The values for a_t are fixed to define the slope factor as the linear rate of change in y on the observed time metric; for example, in a panel study for which participants were assessed annually for T years, we might use $\mathbf{a} = (0,1,2, \ldots, T - 1)'$ so that one unit on the time metric defined by $\mathbf{a}$ is one year. Typically, the first time score, a_1, is fixed at zero so that the intercept factor can be interpreted as the expected response at the first time of measurement. The ε_{ti}s represent independent and identically distributed measurement and time-specific errors on the y_{ti}s at time t, and the ε_{ti}s are usually assumed to be uncorrelated across time. In the structural model, α_{00} is the population mean of the individual intercept factor values, α_{10} is the population mean of the individual slope factor values, ξ_{0i} is the deviation of η_{0i} from the population mean intercept, α_{00}, and ξ_{1i} is the deviation of η_{1i} from the population mean slope, α_{10}. The distribution of individual intercept factor values and slope factors values is assumed to be multivariate normal, as is the distribution of the ε_{ti}s; the growth factors are assumed to be uncorrelated with the errors. Figure 1 displays the expected individual trajectories (dotted lines) of a hypothetical random sample of four individuals drawn from a population with an overall mean growth trajectory given by the solid line. Intra individual change is represented by each of the individual-specific trajectories (each with person-specific intercept and slope values), and interindividual differences are represented by the variability in individual-specific intercept and slope values relative to the overall mean intercept and slope values.

Extending now to the conditional latent growth model, hypothesized predictors of the interindividual differences can be included in both the measurement model (for time-varying predictors and time-invariant predictors with unrestricted time-varying effects) and the structural model (for time-invariant predictors of the intercept and slope factors) as given by

Measurement model:

$$y_{ti} = \eta_{0i} + \eta_{1i}a_t + \pi_{1t}w_{ti} + \varepsilon_{ti},$$

Structural model: (2)

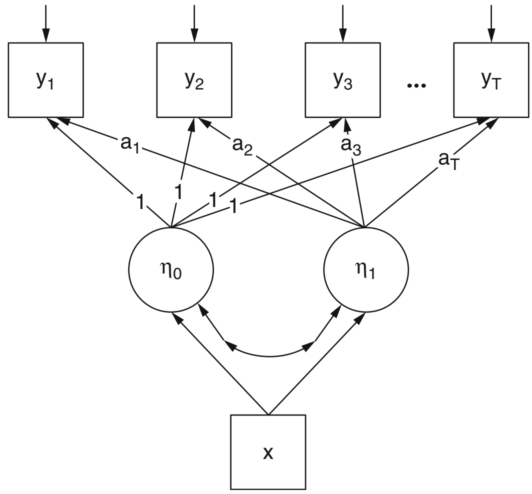
$$\eta_{0i} = \alpha_{00} + \alpha_{01}x_i + \xi_{0i},$$
$$\eta_{1i} = \alpha_{10} + \alpha_{11}x_i + \xi_{1i},$$

where π_{1t} is the difference in the expected outcome, y_t, corresponding to a one unit difference in the time-varying covariate, w, specifically at time t; α_{01} is the difference in the mean of the intercept factor corresponding to the one unit difference in the time-invariant covariate, x; and α_{11} is the difference in the mean of the slope factor corresponding to the one unit difference in the time-invariant covariate, x. The conditional growth model with a predictor for the growth factors is depicted in path diagram form in Fig. 2 using the following diagramming conventions: latent variables are represented by circles, observed variables are represented by rectangles, linear directional relationships are represented by single arrow paths, correlational relationships are represented by double arrow paths, and error terms are represented by unanchored single arrow paths.

Now that we have provided a brief overview of conventional latent growth modeling in a latent variable framework, we next introduce the foundations of binary and ordinal logistic regression followed by discussion of how repeated measures of a categorical outcome can be analyzed in this same latent variable framework.

Generalized Linear Models for Binary and Ordinal Outcomes

As mentioned in the introduction, categorical and limited dependent variables are quite common in



Growth Curve Models with Categorical Outcomes, Fig. 2 Path diagram for a conditional linear latent growth model with continuous outcomes and a time-invariant predictor of the growth factors

criminology research. These types of outcome variables, be they cross-sectional or longitudinal, violate most if not all the assumptions of the linear models in standard use for continuous outcomes. Generalized linear models (GLMs) are a family of regression models that extend OLS regression to accommodate noncontinuous outcomes while still working with outcome predictors in a linear regression framework. There are different (often equivalent) approaches for parameterizing and estimating GLMs (Long 1997; Skrondal and Rabe-Hesketh 2004). For the purposes of this entry, we will introduce a somewhat less standard specification, known as the *latent response variable* (LRV) formulation for ordinal outcome variables. Our choice of the LRV formulation is based on the ease with which it enables the extension of the latent growth model specification given in the previous section to include categorical longitudinal outcomes.

In the latent response variable formulation, it is assumed that the observed ordinal outcome, y, is a discretized form of an underlying continuous latent response variable, y^*. For example, consider the binary (0/1) outcome of clinical depression, *depress*. One could imagine an underlying continuum of depression, *depress**, such that individuals whose values of *depress**

exceeded a certain level, or threshold, would all be observed with binary outcome *depress* = 1. In general, the relationship between a binary outcome, y, and the latent response variable, y^*, is given by

$$y_i = \begin{cases} 1 \text{ if } y_i^* > \tau_1 \\ 0 \text{ if } y_i^* \leq \tau_1, \end{cases} \quad (3)$$

where τ_1 is the *threshold* for y^* and $\Pr(y = 1) = \Pr(y^* > \tau_1)$. As another example, consider the four-category ordinal outcome, *oppose*, measuring opposition to or disapproval of the death penalty for first degree murder with response categories on a Likert scale: *strongly approve/support* (0), *approve/support* (1), *disapprove/oppose* (2), and *strongly disapprove/oppose* (3). One could imagine an underlying continuum of opposition or disapproval, *oppose**, such that four different ranges of *oppose** defined by three cut points or thresholds map onto the observed values of *oppose*. In general, there are $J - 1$ thresholds that define the relationship between a latent response variable, y^*, and its J-category ordinal form, y, such that

$$\mathrm{y_i} = \begin{cases} 0 \text{ if } -\infty < y_i^* \leq \tau_1 \\ 1 \text{ if } \tau_1 < y_i^* \leq \tau_2 \\ \vdots \\ J-1 \text{ if } \tau_{J-1} < y_i^* \leq \infty \end{cases}, \quad (4)$$

where τ_j is the j^{th} *threshold* for y^*, delineating responses $j - 1$ and j on the scale of y. In this LRV formulation, $\Pr(y \leq j) = \Pr(y^* \leq \tau_{j+1})$ and $\Pr(y = j) = \Pr(y^* \leq \tau_{j+1})\text{-}\Pr(y^* \leq \tau_j)$, where $j \in \{0,1,\ldots,J\text{-}1\}$, $\tau_0 = -\infty$, $\tau_J = \infty$, and $\tau_0 < \tau_1 < \ldots < \tau_{J-1} < \tau_J$. The continuous latent response variable, y^*, is then expressed as the sum of a mean and error term given by

$$y_i^* = \mu_i + \delta_i. \quad (5)$$

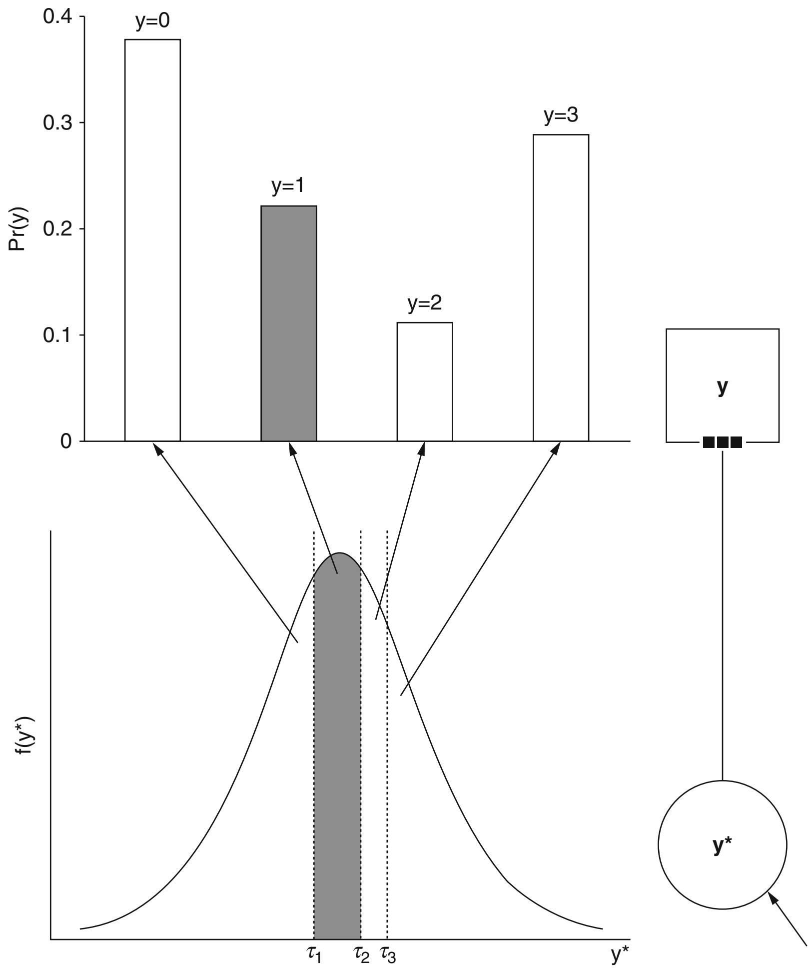
The distribution and scale of the error, δ, must be specified a priori by the analyst; the two most common distributions for δ are the standard normal distribution and the standard logistic distribution. Figure 3 provides a visual representation of the latent response variable formulation for an ordinal outcome with four response categories. The bottom portion of Fig. 3 depicts a standard logistic distribution for y^* with three corresponding thresholds, (τ_1, τ_2, τ_3). All individuals in the population with y^* values between τ_1 and τ_2 will manifest the response value $y = 1$, and the shaded area under the probability density curve for y^* between τ_1 and τ_2 is equal to the probability of the $y = 1$ outcome, as depicted in top portion of Fig. 3. Figure 3 also displays the path diagram representation of the relationship between y^* and y. The three solid squares at the point where the path from y^* meets y indicate the deterministic relationship specified between given values of y^* and resultant observed values on y defined by the three thresholds, (τ_1, τ_2, τ_3).

The conditional model for ordinal outcomes is specified so that the observed predictors of the categorical outcome are related to cumulative response probabilities via the latent response variables as follows:

$$\begin{aligned} y_i^* &= \mu_{i|x_i} + \delta_i, \\ \mu_{i|x_i} &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}. \end{aligned} \quad (6)$$

Figure 4 displays a plot of linear regression model of latent response variable, y^*, underlying a four-category ordinal outcome, versus a single predictor, x. As with a standard linear regression model, the expected values of y^* given x falls along the line, $\beta_0 + \beta_1 x$, and the distributions of y^* values in the population at each given value of x are the same in shape and variance (as shown in Fig. 4 by the probability density curves at the x-values, $x = x_1$, $x = x_2$, and $x = x_3$). For model identification, the intercept, β_0, is suppressed, i.e., fixed at zero. The linear model implies that the difference in the expected value of y^* corresponding to a one unit difference in x is equal to β_1 across the entire range of x. The conditional model given in Eq. 6 also makes the assumption of *threshold invariance*, that is, assuming that the thresholds defining the

Growth Curve Models with Categorical Outcomes, Fig. 3 Graphical and path diagram representations of the latent response variable formulation for an observed ordinal outcome with four response categories



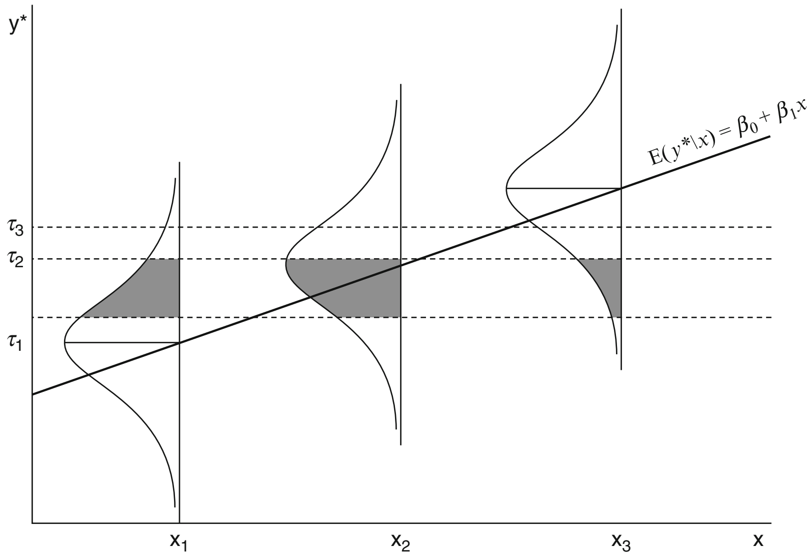
relationship between y^* and y do *not* depend on (i.e., are invariant) relative to x. To assist in visualizing how the linear functional relationship between y^* and x and the assumption of threshold invariance translate to the relationship between observed response probabilities of y and x, the three thresholds for y^*, which do *not* depend on x, are drawn as horizontal lines in Fig. 4. The area under the probability density function for y^* between τ_1 and τ_2, corresponding to $\Pr(y = 2|x)$, at $x = x_1$, $x = x_2$, and $x = x_3$, is shaded making it easy to see that even though the mean of y^* shifts linearly across the range of x, the response category probabilities do *not* change linearly or even monotonically.

Using a standard logistic distribution for δ, the linear regression for y^* on a single predictor (easily replaced with a linear combination of multiple predictors as given in Eq. 6) and the assumption of threshold invariance translates to the following relationship between the response category probabilities and the predictor:

$$\Pr(y_i \le j|x_i) = \frac{1}{1 + \exp\left(-\tau_{j+1} + \beta_1 x_i\right)}, \quad (7)$$

or, equivalently,

$$\log\left(\frac{\Pr(y_i \le j|x_i)}{\Pr(y_i > j|x_i)}\right) = \tau_{j+1} - \beta_1 x_i. \quad (8)$$



Growth Curve Models with Categorical Outcomes, Fig. 4 Plot of the linear regression line for a latent response variable, y^*, underlying an observed four-category ordinal outcome variable, y, versus a single predictor, x. Probability density curves depict the distribution of y^* values at three different values of x. Dashed horizontal lines show the threshold values for y^* delineating the four value ranges mapping on to the four response categories of y. Shaded areas of the probability density show the changing Pr ($y = 2|x$) across the three values of x

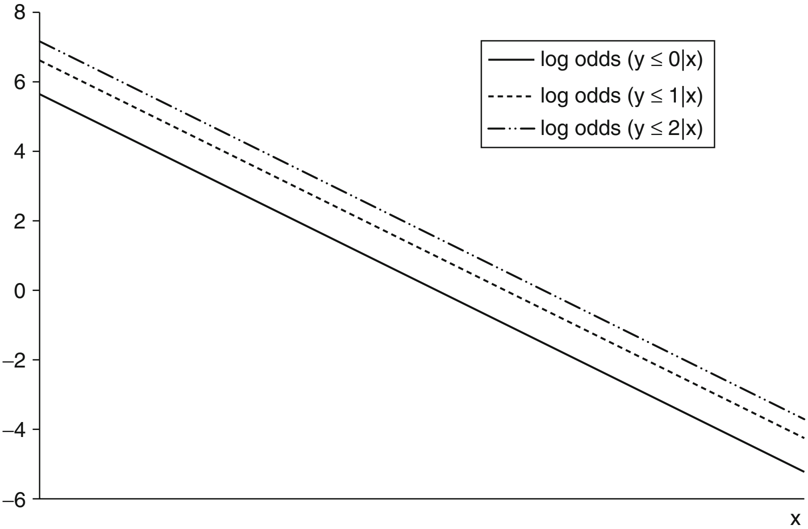
For a binary variable (with $J = 2$ categories), Eq. 8 reduces to the familiar logistic regression equation:

$$\log\left(\frac{\Pr(y_i = 1|x_i)}{\Pr(y_i = 0|x_i)}\right) = -\tau_1 + \beta_1 x_i. \qquad (9)$$

In Fig. 5 we can see the representation of the relationships between the cumulative response category log odds and the predictor as described in Eq. 8 – that cumulative log odds for the different response categories all vary linearly as a function of x and that the lines for each of the three cumulative log odds are *parallel* with the distance between them determined by the differences in the threshold values.

The functional form of each of the three lines in Fig. 5 and the fact that they are all linear is a direct consequence of linearity assumption for the relationship between y^* and x. Assuming a linear relationship between y^* and x constrains the relationship between y and x such that cumulative response category log odds differ identically and linearly for every one unit difference in x. Consequently, β_1 is interpreted not just as the difference in the expected value of y^* corresponding to a positive difference of one unit on x but also as the log odds ratio for responding at or below a given response category (rather than above) corresponding to a *negative* difference of one unit on x, i.e., the odds for a response less than or equal to category j differ by a factor of exp $(-\beta_1)$ for every one additional unit on x, for all $j = 0$, 1, or 2. Thus, the cumulative odds for each response category differ identically and proportionally for every one unit difference in x. The linearity assumption is also known as the *proportional odds* assumption.

The equidistance between each pair of lines in Fig. 5 across the full range of x is a direct consequence of assuming that the thresholds defining the relationship between y^* and y do not depend on x. Thus, the differences in the cumulative log odds across the response categories (i.e., the vertical distance between the lines) at any given value of x are constant across all values of x. The assumption of *threshold invariance*, meaning that the thresholds do not depend on x, is also known as the *parallel regression* assumption. Both the proportional odds assumption and the



Growth Curve Models with Categorical Outcomes, Fig. 5 Plot of the cumulative response log odds for *y* versus *x* based on a linear regression model of y^* on *x* with threshold invariance

parallel regression assumption can be relaxed (and tested), and this will be mentioned again in the growth modeling context in the next section.

Building upon the overviews of both conventional latent growth models and cross-sectional latent response variable models for binary and ordinal dependent variables, the intersection of these two modeling approaches that enables the study of change in binary and ordinal longitudinal outcomes is introduced next.

Latent Growth Models for Binary and Ordinal Outcomes

The specification for the latent growth model for binary or order categorical outcomes is built on the same latent response variable formulation used with cross-sectional data. It is assumed that there is a continuous latent response variable, y_{ti}^*, underlying the observed response on the J-category ordinal outcome, y_{ti}, for individual $i(i = 1, \ldots, n)$ at time $t(t = 1, \ldots, T)$ with the relationship between y_{ti}^* and y_{ti} given by

$$y_{ti} = \begin{cases} 0 \text{ if } -\infty < y_{ti}^* \le \tau_{1t} \\ 1 \text{ if } \tau_{1t} < y_{ti}^* \le \tau_{2t} \\ \vdots \\ J-1 \text{ if } \tau_{(j-1)t} < y_{ti}^* \le \infty \end{cases}, \quad (10)$$

where τ_{jt} is the j^{th} *threshold* for y_t^*, delineating responses $j - 1$ and j on the scale of y_t. Usually, in the model building taxonomy, it is common to begin with models that make the assumption of *longitudinal threshold invariance*, meaning that the set of $J - 1$ thresholds, $(\tau_{1t}, \ldots, \tau_{(J-1)t})$, are the same at each wave; i.e., $\tau_{jt} = \tau_j, \forall t$. The longitudinal threshold invariance assumption can then be evaluated by testing the improvement in model fit when threshold invariance is relaxed.

The continuous latent response variable, y_{ti}^*, at each wave, t, is expressed as the sum of a mean and error term given by

$$y_{ti}^* = \mu_{ti} + \delta_{ti}. \quad (11)$$

As before, the distribution and scale for the error at each wave, δ_{ti}, must be specified a priori by the analyst – for the purposes of this entry, the standard logistic distribution for the error terms will be used. A latent growth model is then specified for the individual μ_{ti} values in a similar way as for observed continuous repeated measures. A linear latent growth model expresses the expected value on the latent response variable for individual i at time t as a function of the intercept and growth factors values; that is,

$$\mu_{ti} = \eta_{0i} + \eta_{1i} a_t, \quad (12)$$

or, equivalently, combining Eqs. 11 and 12,

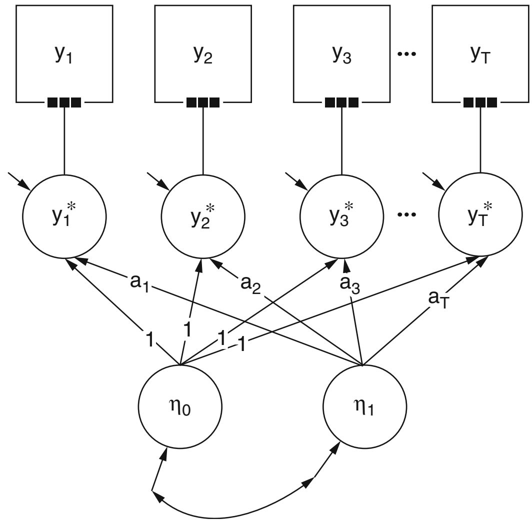
$$y_{ti}^{*} = \eta_{0i} + \eta_{1i}a_t + \delta_{ti}. \quad (13)$$

The structural portion of the latent growth model is identical to the specification for observed continuous repeated measures; that is,

$$\begin{aligned} \eta_{0i} &= \alpha_{00} + \xi_{0i}, \\ \eta_{1i} &= \alpha_{10} + \xi_{1i}, \end{aligned} \quad (14)$$

where $\xi_{\bullet i}$ s are specified to have a multivariate normal distribution. Since the location and scale of the latent response variable is indeterminate, an additional restriction must be placed, fixing $\alpha_{00} = 0$, similar to fixing $\beta_0 = 0$ for y^* in the cross-sectional model. The path diagram representation of an unconditional linear latent growth model for observed ordinal outcomes is depicted in Fig. 6. Although the above specification expresses the change in the latent response variable over time as a linear function of the time metric, the same approaches can be used as with observed continuous outcomes to investigate interindividual differences in curvilinear and other forms of nonlinear trajectories of change.

To gain a better understanding of what specifying a linear latent growth model for the latent response variable with longitudinal threshold invariance implies with respect to interindividual difference in intra individual change in the observed ordinal outcome, consider Fig. 7 which displays expected individual growth trajectories (dotted lines) on the latent response variable underlying an observed four-category ordinal variable for three hypothetical individuals, $i = 1$, 2, and 3, and the population overall mean growth trajectory (solid line). The time-invariant thresholds are shown by the horizontal dashed lines. Each individual in the population has their own expected y^* trajectory as given by Eq. 12. Each individual's cumulative response probabilities at each point in time are determined by the thresholds and the distribution of y_{ti}^{*} centered at μ_{ti}, as depicted by the separated density curves for each of the three hypothetical individuals at time points a_2, a_3, and a_4 shown in Fig. 7. Assuming that δ_{ti} are independently and



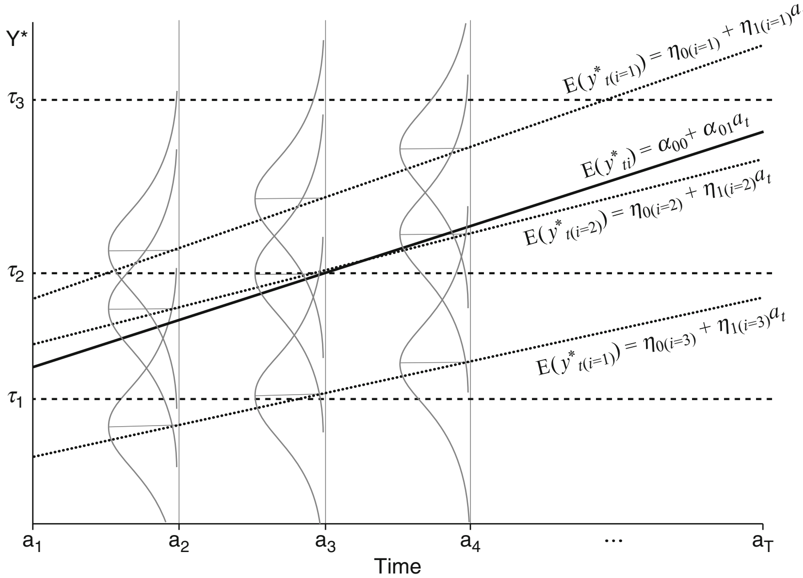
Growth Curve Models with Categorical Outcomes, Fig. 6 Path diagram for an unconditional linear latent growth model for the latent response variables underlying a set of observed longitudinal ordinal outcomes

identically distributed standard logistic, the relationship between the observed ordinal response and the growth factors is given by

$$\log\left(\frac{\Pr(y_{ti} \le j)}{\Pr(y_{ti} > j)}\right) = \tau_j - (\eta_{0i} + \eta_{1i}a_t). \quad (15)$$

Thus, $\tau_j - \eta_{0i}$ is the cumulative log odds for response category j for individual i when $a_t = 0$, and $-\eta_{1i}$ is the change in the cumulative log odds (or cumulative log odds ratio) for response category j for individual i corresponding to a one unit increase in the time metric of a_t.

It can be seen from Eq. 15 that the longitudinal threshold invariance assumption (i.e., $\tau_{jt} = \tau_j, \forall t$) implies that the difference in cumulative log odds for any two response categories will be the same at any given fixed point in time, a_t, across the entire time span for a given individual i. Assuming a linear function for the intra individual change process in the latent response variable under the threshold invariance assumption implies that the change in the cumulative log odds for a given individual and for a given response category corresponding to a one unit difference in the time metric is the same across the entire time span and across all response categories. As mentioned



Growth Curve Models with Categorical Outcomes, Fig. 7 Individual trajectories (*dotted lines*) of latent responses to repeated measure of an observed four-category ordinal outcome variable for a hypothetical random sample of three individuals ($i = 1, 2, 3$) drawn from a population with a mean growth trajectory for y^* given by the *solid line* with *dashed horizontal lines* show the time-invariant threshold values for y^* delineating the four value ranges mapping on to the four response categories of y

before, both the linearity and threshold invariance assumptions can be relaxed and tested. However, in order for the structural portion of the latent growth model to be identified, *at least one threshold* must be held invariant across time. For a binary longitudinal outcome with only one threshold at each time point, complete threshold invariance must be imposed.

It is important at this point to make the reader aware of one critical difference between the latent growth model for observed continuous outcomes and the LGM for observed ordinal outcomes. For continuous outcomes, the mean growth trajectory for the population is just a linear function of the population mean intercept factor and population mean slope factor. This is only the case at the latent response variable level of the categorical LGM – it is *not* the case for the mean population responses for the observed ordinal variable across time. In other words, it is incorrect to just plug the means of η_0 and η_1 into Eq. 15 to get the population mean cumulative response probabilities because the average of the individual cumulative response probabilities is not the same as the cumulative response probability corresponding to the mean growth trajectory for the latent response variable. That is,

$$E[\Pr(y_{ti} \le j)] \ne \frac{1}{1 + \exp\left(-\tau_j + (\alpha_{00} + \alpha_{10}a_t)\right)}. \quad (16)$$

To compute the mean cumulative response probability for the population at a given point in time requires calculating and then averaging the individual cumulative response probabilities for all members of the population, that is,

$$E[\Pr(y_{ti} \le j)] = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{1+\exp\left(-\tau_j+(\eta_{0i}+\eta_{1i}a_t)\right)}\right)\right]. \quad (17)$$

The above expression can also be written as a double integral equation, integrating over both η_0 and η_1. Obtaining model-estimated mean cumulative response probabilities for a specific response category at a specific measurement occasion requires numeric integration based on the model-estimated distribution of the latent growth factors which is a post-estimation option in some modeling software, such as Mplus V6.12 (Muthén and Muthén 1998–2011a).

It is straightforward to extend the unconditional ordinal LGM to include time-invariant and time-varying predictors, just as is done with observed continuous repeated measures. Time-invariant predictors can be included as predictors of the growth factors, or direct predictors of the expected latent response variable across time and time-varying predictors can be included as direct predictors of the expected latent response variable across time, as given by

$$\begin{aligned} y_{ti}^* &= \mu_{ti} + \delta_{ti}, \\ \mu_{ti} &= \eta_{0i} + \eta_{1i} a_t + \pi_{1t} w_{ti}, \\ \eta_{0i} &= \alpha_{00} + \alpha_{01} x_i + \xi_{0i}, \\ \eta_{1i} &= \alpha_{10} + \alpha_{11} x_i + \xi_{1i}, \end{aligned} \quad (18)$$

where δ_{ti} are assumed to be *i.i.d.* standard logistic and α_{00} is fixed at zero for identification.

The unconditional and conditional ordinal latent growth models presented in this section can all be estimated using maximum likelihood estimation. These growth models can be more computationally intensive than the cross-sectional models for observed ordinal outcomes as maximizing the likelihood function with continuous latent predictors (i.e., the latent growth factors) for the ordinal responses requires numeric integration with one dimension of integration for each growth factor. Full-information maximum likelihood (FIML) is available for these models which allows the inclusion of cases with incomplete data on the longitudinal outcomes under the missing-at-random (MAR) assumption (for more about missing data, see, e.g., Enders 2010).

Discussion

This entry has discussed the application of latent growth curve modeling to categorical outcomes by integrating aspects of the conventional latent growth model with the generalized linear model. The unconditional growth model with respect to parameterization and estimation was presented, and concise guidelines were provided for how to conduct the model building and evaluation process. It was then elaborated on how to include covariates to evaluate hypotheses about the relationship between predictors and the developmental change process.

There are several interesting augmentations of and alternatives to these models that are currently available to applied researchers. Although in Eq. 18 the change in the expected latent response outcome was expressed as a linear function of the time metric, it is possible (with an adequate number of repeated observations on each subject) to investigate interindividual differences in curvilinear and other forms of nonlinear trajectories of change. The two most common approaches are (1) to freely estimate $T - 2$ of the time scores loadings for η_1 (fixing one loading at zero to define the intercept location and fixing one loading at unity for identification and to set the slope factor metric) or (2) to use additional growth factors (beyond the intercept and slope factors) to accommodate curvilinear polynomial functions of times (e.g., adding a third, quadratic growth factor with loadings fixed at the values a_t^2). Alternative specifications of time can also be easily accommodated, including piece-wise linear growth models as well as exponential and sinusoidal models of change (see, e.g., Blozis et al. 2007; Bollen and Curran 2006).

In this entry, the application of a latent growth model to repeatedly measured categorical outcomes was discussed. In this model intra-individual change over time is estimated by person-specific growth parameters (e.g., intercept and slope), and interindividual differences are modeled by allowing for individual variation around the estimated growth factor means. Notably, this model is somewhat restrictive in that it assumes population homogeneity in the growth trajectories, i.e., that the growth factors for all persons in the sample are identically distributed. However, etiological as well as prevention and intervention evidence exists to suggest that criminal and antisocial behavior in the overall population may be better represented by a mixing of unobserved heterogeneous subgroups of individuals characterized by differently distributed developmental trajectories, differential risk factors, and differential responses to behavioral and policy interventions. Fortunately, it is relatively straightforward to extend the categorical latent growth model to a generalized linear growth *mixture* model (Feldman et al. 2009). As is the case for continuous outcomes, different latent trajectory classes can be characterized by class-varying mean and variance structures for the growth factors. For example, the developmental change process for one latent subgroup may be perfectly captured by an intercept and a linear slope, while for another subgroup, an additional nonlinear slope is needed.

In addition to modeling nonlinear change and developmental heterogeneity, it is possible that developmental differences cannot be captured by continuous or discrete individual variability around a structured function of time (i.e., intercept, linear, and nonlinear slope). Longitudinal latent class analysis (LLCA: Vermunt et al. 2008; also known as repeated measure latent class analysis, RMLCA: Collins and Lanza 2010) can be used in such situations. LLCA is the application of latent class analysis to repeated outcomes, and unobserved heterogeneity in the developmental response profiles is captured exclusively by a categorical latent variable. In comparison to a growth model which models time-scaled change, LLCA models the longitudinal patterns of discrete states (see, e.g., Feldman et al. 2009; Liu et al. 2010; Liu et al., in press).

In the above-discussed three modeling extensions (nonlinear change, generalized linear growth mixture model, and longitudinal latent class analysis), the inclusion of antecedents and distal outcomes plays an important role. The use of such auxiliary information, potentially derived from substantive theory, is highly relevant to determine the concurrent and prognostic validity of specific growth factors and developmental trajectory profiles derived from a particular data set (Kreuter and Muthén 2008; Petras and Masyn 2010). That is to say, the inclusion of auxiliary information in these models is a necessary step in understanding as well as evaluating the fidelity and utility of the resultant trajectory profiles from a given study. In the simplest case, auxiliary information can consist of observed univariate or multivariate variables measuring predictors or distal outcomes. However, the auxiliary information could itself be a latent variable with its own measurement model and can consist of repeated measures which are observed sequentially or concurrently and modeled simultaneously with the change process of the outcome.

Clearly, these models hold great potential for aiding empirical investigations of developmental theories of normative and non-normative behaviors and risky outcomes across the lifespan. In no way is this more evident than in the marked increase in their use among applied researchers in criminology and other behavioral sciences. However, there is still much opportunity in the realm of methods development to capitalize on the potential of these models and extensions to better accommodate the complexities of developmental theories. And, as with any statistical tool, the research question, along with previous theoretical and empirical work, should guide these models' application in a particular study, with thoughtful and purposeful choices for model specification, selection, and interpretation.

Acknowledgment Dr. Weiwei Liu received support through a training grant from the National Institute of Mental Health while working on this entry (T32 MH18834). Correspondence concerning this article should be addressed to Hanno Petras, Ph.D., at JBS International, Inc., 5515 Security Lane, Suite 800, North Bethesda, MD 20852-5007, USA. Phone: (240) 645-4921. Email: hpetras@jbsinternational.com.

Related Entries

- ▶ Crime Specialization, Progression, and Sequencing
- ▶ Desistance From Crime
- ▶ Pathways to Delinquency
- ▶ General Theory of Crime
- ▶ Group-Based Trajectory Models and Developmental Change
- ▶ Group-Based Trajectory Models
- ▶ Interrupted Time Series Models
- ▶ Longitudinal Crime Trends at Places
- ▶ Longitudinal Studies in Criminology

Recommended Reading and References

Agresti A (2002) Categorical data analysis, 2nd edn. Wiley, New York
Blozis SA, Conger KJ, Harring JR (2007) Nonlinear latent curve models for longitudinal data. Int J Behav Dev 31:340–346
Bollen KA, Curran PJ (2006) Latent curve models: a structural equation perspective. Wiley, Hoboken
Britt CL, Weisburd D (2010) Logistic regression models for categorical outcome variables. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 649–682
Collins LM, Lanza ST (2010) Latent class and latent transition analysis with applications in the social, behavioral, and health sciences. Wiley, Hoboken
Curran PJ, Obeidat K, Losardo D (2010) Twelve frequently asked questions about growth curve modeling. J Cogn Dev 2:121–136
Enders CK (2010) Applied missing data analysis. Guildford Press, New York
Feldman BJ, Masyn KE, Conger RD (2009) New approaches to studying problem behaviors: a comparison of methods for modeling longitudinal, categorical adolescent drinking data. Dev Psychol 3:652–676
Kline RB (2010) Principles and practice of structural equation modeling, 3rd edn. Guilford Press, New York
Kreuter F, Muthén B (2008) Analyzing criminal trajectory profiles: bridging multilevel and group-based approaches using growth mixture modeling. J Quant Criminol 24:1–31
Liu LC, Hedeker D, Segawa E, Flay BR (2010) Evaluation of longitudinal intervention effects: an example of latent growth mixture models for ordinal drug-use outcomes. J Drug Issues 40:27–43
Liu W, Lynne-Landsman S, Petras H, Masyn K, Ialongo N The evaluation of two first grade preventive interventions on childhood aggression and adolescent marijuana use: a latent transition longitudinal mixture model. *Prev Sci* (in press)
Long JS (1997) Regression models for categorical and limited dependent variables. Sage, Thousand Oaks
MacDonald JM, Lattimore PK (2010) Count models in criminology. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 683–698
Mehta PD, Neale MC, Flay BR (2004) Squeezing interval change from ordinal panel data: latent growth curves with ordinal outcomes. Psychol Methods 9(3):301–333
Muthén BO (1996) Growth modeling with binary responses. In: Eye AV, Clogg C (eds) Categorical variables in developmental research: methods of analysis. Academic, San Diego, pp 37–54
Muthén B (2001) Second-generation structural equation modeling with a combination of categorical and continuous latent variables: new opportunities for latent class/latent growth modeling. In: Collins LM, Sayer A (eds) New methods for the analysis of change. APA, Washington, DC, pp 291–322
Muthén B (2004) Latent variable analysis: growth mixture modeling and related techniques for longitudinal data. In: Kaplan D (ed) Handbook of quantitative methodology for the social sciences. Sage, Newbury Park, pp 345–368
Muthén B, Muthén LK (1998–2011). Mplus (Version 6.12) [Computer software]. Muthén & Muthén, Los Angeles
Muthén LK, Muthén B (1998–2011) Mplus user's guide, 6th edn. Muthén & Muthén, Los Angeles
Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. Criminology 31:327–362
Nelder J, Wedderburn R (1972) Generalized linear models. J R Stat Soc 135(3):370–384
Petras H, Masyn K (2010) General growth mixture analysis with antecedents and consequences of change. In: Piquero A, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 69–100
Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the life course. In: Liberman AM (ed) The long view of crime – a synthesis of longitudinal research. Springer, New York, pp 23–78
Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analyses of the Cambridge study in delinquent development. Cambridge University Press, Cambridge
Skrondal A, Rabe-Hesketh S (2004) Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models. Chapman Hall, London
Vermunt JK, Tran B, Magidson J (2008) Latent class models in longitudinal research. In: Menard S (ed) Handbook of longitudinal research: design, measurement, and analysis. Elsevier, Burlington, pp 373–385

GST

▶ Individual Characteristics and General Strain Theory

Guilds

▶ History of Police Unions