Growth models for categorical response variables: standard, latent-class, and hybrid approaches

Jeroen K. Vermunt

Department of Methodology and Statistics, Tilburg University

1 Introduction

There are three main approaches to the analysis of longitudinal data (Diggle, Liang, and Zeger 1994):

1. conditional, change-score or transitional models
2. marginal or population-average models
3. subject-specific, random-effects, or growth models.

Transitional models such as Markov-type models concentrate on changes between consecutive time points. Marginal models can be used to investigate changes in univariate distributions, and random-effects or growth models study development of individuals over time. These three approaches do not only differ in the questions they address, but also in the way they deal with the dependencies between the repeated measures. Because of their structure, transitional models take the bivariate dependencies between observations at consecutive occasions into account. Growth models capture the dependence by introducing one or more latent variables (random effects). In marginal models, the dependency is not explicitly modeled, but dealt with as found in the data and in general is taken into account in a more ad hoc way in the estimation procedure. Variants of transitional, growth and marginal models have been developed for categorical response variables (Agresti, 2002), as well as hybrids combining these approaches (Vermunt, Rodrigo, and Ato, 2001).

Growth modeling is probably the most popular tool for dealing with longitudinal data in social and behavioral sciences. In econometrics it is usually referred to as panel regression. Basically, growth models are regression models for two-level data – time points nested within individuals – in which time enters as one of the predictors. Here, we deal with growth models for categorical response variables, which implies using mixed-effects variants of the appropriate regression models from the generalized linear modeling (GLM) family, such as random-effects binary, ordinal, and multinomial logistic and Poisson regression.

Whereas in standard growth models unobserved heterogeneity is captured by means of continuous latent variables, it is also possible to work with discrete latent variables, which yields what is referred to as latent-class (LC) growth modeling. The purpose of this tool is to identify subgroups or clusters showing different developmental patterns or trajectories. As will be shown, the LC growth model is a special case of the mixture GLM for two-level data, which makes it is straightforward to deal with categorical response variables, such as binary, nominal, ordinal, and count variables. As in standard growth models, one may use continuous random effects to account for (part of) the unobserved heterogeneity, yielding hybrids between LC and standard mixed models.

Various interesting extensions of the standard LC growth model can be derived using the new multilevel LC methodology proposed by Vermunt (2003, 2004, 2005). For example, one could build a latent classification using multiple indicators and specify a growth model for these latent classifications. Another example is a LC growth model for multilevel data, in which groups may belong to latent classes with different growth patterns or in which groups may differ with respect to the growth class distribution of their members.

Below, I first describe the standard random-effects growth model for categorical response variables. Subsequently, I discuss the LC-based growth modeling approach, as well as hybrid variants combining discrete and continuous random effects. Then, I present an empirical example in which the standard, latent-class, and hybrid methods are applied. Subsequently, I compare the obtained results with the ones that would have been obtained with a Markov model. In the last section, I present extensions of the basic models which involve the use of methods for three-level instead of two-level data.

2 Growth models

Let y_{it} denote the response of case i at occasion t on the response variable of interest, and N and T the number of cases and time points, respectively, where $1 \leq i \leq N$ and $1 \leq t \leq T$. It is not necessary to assume that each case has observation at each time point, which means that the longitudinal data set at hand may be unbalanced and may contain missing values.

Growth modeling involves specifying a random-effects regression model for y_{it}, in which the time-specific responses are assumed to be a function of time (Hedeker, 2004). Depending on the scale type of the dependent variable the regression model will be another member of the GLM family (Agresti et al., 2000). More specifically, after an appropriate transformation $g(\cdot)$, the expected value of the response variable $E(y_{it})$ is assumed to be a linear function of a set of predefined functions of t. In the binary case, the transformation could, for example, be the logit transformation, yielding $g[E(y_{it})] = \log \frac{P(y_{it}=1)}{1-P(y_{it}=1)}$.

The linear model for $g[E(y_{it})]$ has the following form

$$g[E(y_{it})] = \beta_{0i} + \sum_{s=1}^{S} \beta_{si} \cdot f_s(t) = \sum_{s=0}^{S} \beta_{si} \cdot f_s(t), \tag{1}$$

in which $f_s(t)$ is a predefined function of time, and $f_0(t) = 1$ for a compact representation of the constant term. Typical special cases are the linear growth model with $S = 1$ and $f_1(t) = t - 1$, and the quadratic growth model with $S = 2$ and $f_1(t) = t - 1$ and $f_2(t) = (t - 1)^2$. The functions $f_s(t)$ could also take on the form of a set of dummy variables for time points 2 to T, in which case $S = T - 1$, and $f_s(t) = 1$ if $s = t + 1$ and 0 otherwise. For an extended discussion of possible functional forms for the time dependence, see Snijders and Bosker (1999).

The index i appearing in the subscript of each of the regression coefficients in equation (1) indicates that these may be subject specific; that is, each individual may have its own growth curve. Note that with longitudinal data we have to take into account that the observations of the same individual at the various time points are not independent of one another. In fact, we are dealing with a two-level data structure in which time points are nested within cases (Snijders and Bosker, 1999). Equivalent to multilevel analysis, in growth model the dependence between observation is dealt with by assuming that some of the model parameters are "group-specific", where for parsimony, these coefficients are assumed to be random coefficients coming from a multivariate normal distribution; that is $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\beta)$. It is, however, not necessary to assume that all unknown regression coefficients vary across individuals. A simpler variant of the model described in equation (1) is obtained by assuming that only the intercept is a random effect, in which case $\beta_{is} = \beta_s$ for $1 \leq s \leq S$. Inclusion of a random intercept is the minimal requirement for claiming an appropriate treatment of the dependencies between the repeated measures within cases. It amounts to assuming that the association structure between the observations has the form of a compound symmetric covariance matrix. With random change parameters one can pick up autocorrelation structures and changing variances across occasions (Hedeker, 2004; Snijders and Bosker, 1999).

An alternative, mixed modeling, formulation of the growth model described in equation (1) is obtained by using the reparameterization $\beta_{si} = \beta_s + u_{si}$ for $0 \leq s \leq S$, where $\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\beta)$. This yields:

$$g[E(y_{it})] = \sum_{s=0}^{S} \beta_s \cdot f_s(t) + \sum_{s=0}^{S} u_{si} \cdot f_s(t), \tag{2}$$

in which the β_s are called fixed effects and the u_{si} are called random effects.

In the case of categorical response variables, for parameter estimation by means of maximum likelihood (ML), it is useful to parameterize the mixed-effects model as a factor-analytic model with uncorrelated latent variables (random effects) with variances equal to 1. This is necessary to be able to solve the integrals appearing in the log-likelihood function by Gauss-Hermite quadrature. The factor-analytic formulation of the growth model described in equations (1) and (2) equals:

$$g[E(y_{it})] = \sum_{s=0}^{S} \beta_s \cdot f_s(t) + \sum_{s=0}^{S} \sum_{s'=0}^{s} \lambda_{ss'} \cdot F_{s'i} \cdot f_s(t), \tag{3}$$

where $\mathbf{F}_i \sim N(\mathbf{0}, \mathbf{I})$ and the $\lambda_{ss'}$ are elements of a $(S+1)$ by $(S+1)$ lower-diagonal "factor loadings" matrix $\mathbf{\Lambda}$. The connection between the mixed-effects and the factor-analytic parameterization is that $\mathbf{\Sigma}_{\beta}= \mathbf{\Lambda}\mathbf{\Lambda}^{'}$, where $\mathbf{\Lambda}$ is the Cholesky decomposition of $\mathbf{\Sigma}_{\beta}$ (Skrondal and Rabe-Hesketh, 2004; Vermunt and Magidson, 2005).

To see how the Cholesky decomposition works, consider the special case that $S = 1$, yielding a factor-analytic model of the form

$$g[E(y_{it})] = \beta_0 + \beta_1 \cdot f_1(t) + \lambda_{00} \cdot F_{0i} + \lambda_{10} \cdot F_{0i} \cdot f_1(t) + \lambda_{11} \cdot F_{1i} \cdot f_1(t).$$

The connection between the λ parameters and the covariances of the random effects is the following: $\sigma_{00} = (\lambda_{00})^2$, $\sigma_{01} = \lambda_{00} \cdot \lambda_{10}$, and $\sigma_{11} = (\lambda_{10})^2 + (\lambda_{11})^2$.

The factor-analytic parameterization is not only useful for parameter estimation using ML, it offers additional flexibility in terms of constraints that can be imposed on the random effects. For example, assuming "factor loading" λ_{10} to be equal to 0 yields a model in which the random effects u_{0i} and u_{1i} are mutually uncorrelated; and assuming $\lambda_{11} = 0$ yields a model in which u_{0i} and u_{1i} are perfectly correlated. Restricted factor-analytic structures are especially useful in models containing many random coefficients as may, for instance, occur growth models based on time dummies. With $T - 1$ time dummies and an intercept, an unrestricted random part would contain T random coefficients, and as a consequence $(T + 1) \cdot T/2$ unknown parameters in the matrix $\mathbf{\Sigma}_{\beta}$. A much simpler factor-analytic structure with say two factors and some λ elements equal to 0 will most probably do an equally good job in terms of model fit.

The mixed-effects GLM described in equations (1), (2), and (3) can be used with response variables that are continuous, binary, or counts. For dealing with ordinal and nominal response variables, however, which from a GLM perspective are in fact multivariate responses, we need to extend somewhat the above models. Denoting a particular response category as m and the number of response categories as M, an ordinal mixed-effects GLM can be defined as

$$g[E_m(y_{it})] = \beta_{0m} + \sum_{s=1}^{S} \beta_s \cdot f_s(t) + u_{0im} + \sum_{s=1}^{S} u_{si} \cdot f_s(t).$$

Here, $g[E_m(y_{it})]$ may represent one of the $M-1$ logits defining an adjacent-category or cumulative logit model, or the underlying latent variable in an ordinal probit model. The main difference with the dichotomous case is that there are $M-1$ fixed (β_{0m}) and random (u_{0im}) effects associated with the intercept instead of only one. Rather that using a $M-1$ dimensional multivariate normal distribution, Hedeker and Gibbons (1996) proposed restricting these $M-1$ random intercept terms using a single factor, that is, as $u_{0im} = \lambda_{0m} \cdot F_{0i}$ (see also Vermunt and Magidson, 2005).

In the nominal case, we obtain

$$g[E_m(y_{it})] = \sum_{s=0}^{S} \beta_{sm} \cdot f_s(t) + \sum_{s=0}^{S} u_{sim} \cdot f_s(t).$$

where $g[E_m(y_{it})]$ will usually be one of the $M-1$ baseline category logits in a multinomial logistic regression model. Note that there are $M-1$ random effects associated with each term, thus not only with the intercept (Agresti et al., 2000). These can, however, again be restricted using a factor-analytic parametrization with a single factor per term: $u_{sim} = \lambda_{sm} \cdot F_{si}$ (Hedeker, 2003; Vermunt 2005; Vermunt and Magidson, 2005).

Thus far, we assumed that there were no other predictors than the time variable itself. Equivalent to standard random-effects model, growth models can easily be extended to include both time-constant (between-subject) and time-varying (within-subject) predictors, denoted as z_{pi} and z_{qit}, respectively. Suppose we have a single between-subject predictor z_{1i} indicating whether case i belongs to the treatment group ($z_{1i} = 1$) or the control group ($z_{1i} = 0$). If we assume that the treatment affects both the initial value y_{i1} and the single change term $f_1(t)$, in the mixed-effects model formulation we obtain

$$g[E(y_{it})] = \beta_0 + \beta_1 \cdot f_1(t) + \beta_2 \cdot z_{1i} + \beta_3 \cdot z_{1i} \cdot f_1(t) + u_{0i} + u_{1i} \cdot f_1(t).$$

In the alternative but equivalent two-level model formulation, one would specify that $\beta_{si} = \beta_s + \beta_{S+s} \cdot z_{1i} + u_{si}$.

The random-effects growth models presented in this section are extremely valuable tools for dealing with longitudinal data. There are, however, also several problematic aspects associated with the use of these methods:

- with categorical response variables, they may become computationally very intensive when there are more than two or three random effects. The reason for this is that the integrals appearing in the likelihood function must be solve using approximation methods, such as linearization methods, numerical integration methods, or (Bayesian) simulation techniques. A possible way out to this problem is to reduce the dimensionality of the problem by using the more restricted factor-analytic specification introduced above. Although usually approximation methods perform very well, this not always the case (Lesaffre and Spiessens, 2001; Rabe-Hesketh, Skrondal and Pickles, 2002; Rodriguez and Goldman, 2001).

- they rely on the untestable assumption that random coefficients come from a multivariate normal distribution (Aitkin, 1999; Vermunt and Van Dijk, 2001). Results obtained with these methods may be biased when this rather strong distributional assumption is violated.

- it is not at all straightforward to interpret the parameters associated with the random effects (the variances and covariances of the u_{si} terms). A common solution to this problem is to depict selected estimated individual-specific curves, for example, the curves at 0, 1 and 2 standard deviations from the mean.

3 Latent-class-based and hybrid growth models

Some of the problems associated with the parametric random-coefficients approach discussed in the previous section may be circumvented by adopting a latent-class-based nonparametric random coefficients approach. This group-based approach is usually referred to as latent class (LC) or mixture regression analysis (Vermunt and Van Dijk, 2001; Wedel and DeSarbo, 2002). Using the LC regression framework, we obtain the following LC growth model:

$$g[E(y_{it}|k)] = \sum_{s=0}^{S} \beta_{sk} \cdot f_s(t), \tag{4}$$

where k denotes a particular latent class, and again $f_0(t) = 1$. The transformed expected value of y_{it} given that case i belongs to latent class k is a function of time, where parameters may differ across latent classes. The output obtained from growth modeling using LC regression analysis is thus K class-specific growth curves, where K is the number of classes or mixture components. Nagin (1999) proposed using the mixture growth model for the analysis of developmental trajectories with the purpose to identify distinctive groups of individual trajectories. Vermunt and Van Dijk (2001) and Vermunt and Hagenaars applied the method in the context of panel studies with categorical response variables.

Although this may not directly be clear from equation (4), also in this model we are assuming that the regression coefficients are random – that they differ across individuals according to some distribution. In fact, we assume that the individual-specific parameters come from a K-mass discrete distribution with unknown locations $\boldsymbol{\beta}_k$ and unknown weights π_k. By increasing the number of classes K till the log-likelihood function does no longer increase, we obtain what is referred to as the nonparametric ML estimator for the random-effects model of interest (Aitkin, 1999; Skrondal and Rabe-Hesketh, 2004; Vermunt 2004). Nagin (1999) referred to the situation in which a smaller number of classes is retained than the maximum that can be identified as a semi-parametric ML estimator.

The similarity with the parametric approach described in the previous section becomes even clearer if one sees that the means and covariances of the random effects can easily be obtained from the β_{sk} and π_k parameters; that is

$$\beta_s = \sum_{k=1}^{K} \beta_{sk} \cdot \pi_k, \; \sigma_{ss'} = \sum_{k=1}^{K} (\beta_{sk} - \beta_s) \cdot (\beta_{s'k} - \beta_{s'}) \cdot \pi_k.$$

In most applications, these numbers will be similar in terms of size to their parametric counterparts (fixed effects β_s and entries of Σ_β). It should also be noted that we could write β_{sk} in terms of a fixed and a random effect: $\beta_{sk} = \beta_s + u_{sk}$.

Similar to the parametric approach, it is possible to assume that certain growth parameters vary across (classes of) individuals, while others do not. Such a constraint amounts to equating a particular growth parameter across classes: $\beta_{sk} = \beta_{s.}$. This is, however, not the only type of constraint that can be imposed. It is also possible to equate parameters across selected classes and to fix growth parameters to 0 or some other value in selected classes. The latter options make possible to specify LC models in which the functional form of the time dependency differs across classes. For example, in a 4-class model, in class 1 the time dependence may be quadratic, in class 2 unrestricted ($S = T - 1$ dummies), and in classes 3 and 4 there may be no time dependence ($S = 0$; no change), where in class 4 the intercept is fixed to a very high negative value, yielding a so-called mover-stayer or zero-inflated model. This flexibility of having class-specific models makes it possible to test very specific hypotheses about developmental trajectories.

As in the parametric case, small modifications are needed for dealing with ordinal and nominal response variables. For ordinal responses, we again have a model with category-specific intercept terms:

$$g[E_m(y_{it}|k)] = \beta_{0mk} + \sum_{s=1}^{S} \beta_{sk} \cdot f_s(t);$$

and for nominal responses a model with category-specific intercepts and slopes:

$$g[E_m(y_{it}|k)] = \sum_{s=0}^{S} \beta_{smk} \cdot f_s(t).$$

As in standard growth models, in LC-based growth models, one can include time-varying and time-constant predictors in the regression equation for the dependent variable. Whereas in the standard growth models, regressing the latent variable on time-constant predictors is equivalent to including those predictors in the regression model for the outcome variable of interest, in the discrete random-effects models discussed in this section this is not the case. As a result, there are two alternative ways to deal with time-constant covariates: they can be used in the regression model for the response variable or in the regression model for the latent classes. The latter can be accomplished by means of a multinomial logistic regression model of the form

$$\log \frac{\pi_{k|\mathbf{z}_i}}{\pi_{1|\mathbf{z}_i}} = \gamma_{0k} + \sum_{p=1}^{P} \gamma_{pk} \cdot z_{pi}.$$

One could, for example, be interested in knowing whether belonging to the treatment group (say $z_{1i} = 1$) increases the probability of belonging to a latent class with a favorable trajectory. This is an alternative to a model in which treatment is assumed to have a direct effect on (changes in) the response variable.

The LC-based approach has various advantages compared to the parametric models discussed in the previous section. That is,

- it is much less computationally intensive when applied with categorical response variables. The log-likelihood function contains a sum over K latent classes rather than a set of integrals that should be approximated by, for example, numerical integration methods. Since the log-likelihood can be computed exactly, no approximations are needed when applied with categorical responses. In fact, no special provisions are needed for dealing with such variables.
- it does not rely on nontestable assumptions about the distributions of the random effects.
- it yields much easier to interpret results. Rather than a set means, variances and covariances summarizing the N observed trajectories, one obtains K basic trajectories.

The LC-based approach does not only have advantages compared to the standard approach, but also certain weak points. These are that

- researchers are confronted with additional model selection issues. One not only needs to decide about the functional form of the time dependence and about which parameters should be assumed to vary across individuals, but also about the necessary number of latent classes, as well as about the form for each of the class-specific time functions. A way to simplify the model selection is to 1) work with the same type of time dependence for all classes, 2) determine the number of classes using BIC (Bayesian information criterion), and 3) investigate whether the selected model can be simplified by equating parameters across classes or fixing certain coefficients to 0.

- local maxima are more common in nonparametric random-effects models than in parametric models. However, current software has provision for dealing with the local maxima problem by using multiple sets of random starting values for the model parameters (Vermunt and Magidson, 2005).

- the LC approach has what is sometimes referred to the problem of overextraction (Bauer and Curran, 2003). It may occur that a large number of latent classes is needed to fully capture the variation in the growth parameters, and that these classes differ mainly with respect to the intercept. This can be seen as an artifact of the LC method since it is usually not what the analyst is looking for. Most likely, he is interested in finding groups with different change parameters rather than different intercepts. A way out to this problem is the hybrid methodology described below, which involves combining discrete with continuous unobserved heterogeneity (Muthén, 2004).

A simple hybrid growth model that in most application will resolve the overextraction problem is obtained by expanding the LC regression model described in equation (4) with a random intercept (Lenk and DeSarbo, 2000; Vermunt and Magidson, 2005). That is,

$$\begin{aligned} g[E(y_{it}|k)] &= \sum_{s=0}^{S} \beta_{sk} \cdot f_s(t) + u_{0i} \\ &= \sum_{s=0}^{S} \beta_{sk} \cdot f_s(t) + \lambda_{00} \cdot F_{0i}, \end{aligned} \tag{5}$$

Note that this model relaxes the basic assumption of the LC model that latent classes are homogenous with respect to all model parameters by allowing for within-class heterogeneity with respect to the intercept. Another way to see the random intercept is as a manner to relax the assumption that the (time-specific) responses are independent within latent classes.

As far as the random part of the model described in equation (5) is concerned, there is two possible extensions. One is to included random change parameters in addition to a random intercept. Another interesting extension is to allow the variance of u_{0i} to vary across latent classes or, equivalently, to replace λ_{00} by λ_{00k}, where $\sigma_{00k} = (\lambda_{00k})^2$. Such a model not only assumes that there is within-class heterogeneity, but also that the amount of within-class heterogeneity may differ across classes, an assumption that may be more realistic in certain applications.

Any of the extensions of the LC regression model discussed above – equality, zero and other fixed-value restrictions, class-specific time functions, ordinal and nominal responses, and explanatory variables affecting the responses and/or the classes – can also be used in hybrid models.

4 An empirical example

The empirical example I will use to illustrate the various types of growth models discussed above is taken from Hedeker and Gibbon's (1996) MIXOR program. It concerns a dichotomous outcome variable "severity of schizophrenia" measured at 7 occasions (consecutive weeks). This binary outcome was obtained by collapsing a severity score ranging from 1 to 7 into two categories, where a 1 indicates that the severity score was at least 3.5 (severe), and 0 that is was smaller than 3.5 (non severe).

In total, there is information on 437 cases. However, for none of the cases there is complete information. For 42 cases, we have observations at 2, for 66 at 3, for 324 at 4, and for 5 at 5 time points. There are 434, 426, 14, 374, 11, 9, and 335 observations at the 7 time points.

Besides the repeated measures for the response variable, there is one time-constant predictor, treatment (0=control group; 1=treatment group). The treatment is a new drug that is expected to decrease the symptoms related to schizophrenia. The main research question to be answered with this data set is whether the treatment reduces the symptoms related to schizophrenia.

[INSERT FIGURE 1 ABOUT HERE]

Figure 1 depicts the observed probability of being in the severe state at each of the seven occasions for the treatment and control group. As can be seen, at the start of the study, almost all cases belong to the severe category in both the treatment and control group. At each of the next time points, the treatment group has a lower probability of having severe schizophrenia symptoms than the control group, showing that there is evidence for a treatment effect.

[INSERT TABLE 1 ABOUT HERE]

In the analysis of this data set, I followed Hedeker and Gibbon's (1996) suggestion to set $S = 1$, with $f_1(t) = \sqrt{t-1}$, and to use a binary logit model. This yields a model in which the logit of severity is a function of the square root of time. Though there is no strong theoretical motivation for using this functional form for the time dependence, there is a good empirical motivation: in a simple model without random effects, this model fits the time-specific response probabilities much better than a linear or a quadratic model, and almost as good as model with an unrestricted time dependence. Table 1 reports the test results obtained by applying various of the models described in the previous two sections to the schizophrenia data set. Besides the log-likelihood value (LL), the number of parameters (# par), and the BIC value, the table provides information on the Wald test for the treatment effect (Wald value, degrees of freedom, and p value). The latter is a test for the time-treatment interaction in Models 1-5 and for the treatment effect on class membership in Models 6-15.

As can be seen from Table 1, the estimated models differ in

- the way they capture unobserved heterogeneity. Model 1 contains no random effects, Model 2 is a standard growth model, Models 3-8 and 12-13 are LC-based models, and Models 9-11 and 14-15 are hybrid models

- how treatment enters in the equation. In Models 1-5 treatment affect the response, whereas in the other models it affects class membership

- the assumed number of latent classes, ranging from 1 to 4.

- whether there is one latent class (the last one) with a different (quadratic) time dependence. This is specified by defining $f_2(t) = t - 1$ and $f_3(t) = (t - 1)^2$, and setting the parameters corresponding to these two terms to 0 in all but class K and the parameter corresponding to $f_1(t)$ to 0 in class K. This specification is used in Models 12-15.

For parameter estimation, I used version 4.0 of the Latent GOLD program (Vermunt and Magidson, 2005), which uses ML estimation. For numerical integration with Gauss-Hermite quadrature, I used the program's default setting of 10 quadrature points per dimension.

Model 1 without random effects serves as a baseline model. Comparison of the log-likelihood and BIC values for Models 2-5 with the ones of Model 1 indicates clearly that it is necessary to take into account individual variation in the growth parameters. When using a LC-based approach, three latent classes seem to be needed. Whereas the log-likelihood value of the three-class model (Model 4) is clearly higher than of the standard growth model (Model 2), according to the BIC criterion the latter is somewhat better. What is most striking is that the treatment effect is not significant if we do not take unobserved heterogeneity into account (Model 1), but it turn out to be if we do so. It should be noted that Models 2 and 4 give similar answers concerning the significance and size of the treatment effect.

Models 6-8 differ from 3-5 in that treatment is assumed to affect class membership instead of having a direct effect on the outcome variable. The latter models fit somewhat worse (have lower log-likelihood and higher BIC values). Models 9-11 differ from 6-8 in that they contain a random intercept allowing for within-class heterogeneity. When using such a specification, no more than two latent classes are needed, and the Wald test for the treatment effect shows an even more significant result (see Model 9).

Models 12-15 are variants of Models 6-7 and 9-10 in which the time dependence in class K (the last class) is assumed to be quadratic instead of a function $\sqrt{t-1}$. It turns out that especially Model 14, which is a modified version of Model 9, does a very good job. It has a much lower log-likelihood value than Model 9, and despite the extra parameter a much lower BIC value. Moreover, it indicates that there is even more evidence for a treatment effect.

[INSERT TABLE 2 ABOUT HERE]

[INSERT FIGURE 2 ABOUT HERE]

Table 2 reports the parameter estimates obtained with Model 14. For each latent class, we have a set of parameters describing the time dependence of the logit of the probability of being in the severely schizophrenic state – Intercept and SQRT-TIME in class 1 and Intercept, TIME, and SQ-TIME in class 2 – as well the standard deviation of the random effect indicating how much the intercept varies within classes. The size of latter parameter, which is assumed to be equal across latent classes, indicates that there is quite some variation within classes. To get a better idea on what all these coefficients mean, Figure 2 depicts the estimated growth curves for the two latent classes. The depicted time- and class-specific probabilities of being severely depressed that are obtained by marginalizing over (integration out) the continuous random effects. The obtained figure is similar to Figure 1, with the difference that the two development patterns are much smoother and more different from one another. It can also be seen why the quadratic curve was needed for class 2: after a small drop in weeks 1 and 2, the probability of a severe form of schizophrenia increased again, a pattern that cannot be described by a monotonic function.

Out of the total sample, 65% is estimated to belong to latent class 1 and 35% to latent class 2. These numbers are 76% and 24% for the treatment group and 34% and 66% for the control group. The treatment effect on class member is given in terms of a logistic regression coefficient and its asymptotic standard error in the lower part of Table 2 – the odds of begin in class 1 instead of 2 is exp(1.78) higher for the treatment than for the control group. The encountered treatment effect shows, on the one hand, that there is a rather strong relation between treatment and class membership, but, on the other hand, that this relationship is far from perfect.

5 Markov models

Whereas this paper dealt with growth models for repeatedly observed categorical response variables, as was already mentioned in the introduction, there are other alternatives for analyzing this type of data. The most important alternative is the standard Markov model or one of its variants (Collins and Wugalter, 1992; Langeheine and Van de Pol, 1994; Vermunt, Langeheine, and Böckenholt, 1999).

To understand the fundamental difference between a growth model and a Markov model, recall that in the former we try to describe and explain individual-level differences in the probability of being in a certain state at a particular time point. In the empirical example, we investigated whether this individual-level probability is time and treatment dependent. In Markov models on the other hand, we describe and explain aggregate transition probabilities; that is, the overall probability of being in a certain state given the state at the previous time point, possibly after controlling for covariates. In the empirical application, the question of interest could, for example, be whether the probability of experiencing a transition between schizophrenic and healthy is larger for the treatment than for the control group.

Mixed Markov and latent class Markov models are two important LC-based variants of the standard Markov model. In a mixed Markov model, individual are assumed to belong to one of K classes that differ with respect to their transition probabilities. In the LC Markov model, the state variable is assumed to be measured with error, yielding a model in which transitions occur across time-specific *latent* states which are connected to observed states by means of a probabilistic mechanism.

I estimated a standard stationary first-order Markov model, a mixed Markov model with 2 latent classes, and a LC Markov model with 2 latent states per time point to the schizophrenia data set. The obtained log-likelihood values – -846, -844, and -844, respectively – show that these models fit as good as the best fitting growth models described above. These log-likelihood values also show that the standard Markov model would suffice, indicating that there is no evidence for unobserved heterogeneity in the transition probabilities as assumed in the mixed Markov model, nor for measurement error in the time-specific schizophrenia state as assumed in the LC Markov model.

The estimated transition probabilities obtained with the stationary first-order Markov model indicate that the treatment group has a higher probability of a transition out of the schizophrenic state (0.23 versus 0.05 for control group) and a lower probability of a transition into the schizophrenic state (0.08 versus 0.16). This is in agreement with the evidence for a positive treatment effect as obtained with the growth modeling approach.

6 Extensions

Various interesting extensions of the growth models described and illustrated in the previous sections can be obtained using the multilevel LC framework that was recently proposed by Vermunt (2003, 2004, 2005). Two of these are growth models for multiple response variables and growth models for individuals nested within groups. Both applications yield data structures with three instead of two levels of nesting; that is, multiple responses nested within time points and time points nested within cases, and time points nested within cases and cases nested within groups.

Suppose that rather than with single indicator we would measure the severity of schizophrenia using multiple indicators. Let $y_{it\ell}$ be the response of case i at occasion t on item ℓ and L the total number of items. A possible manner to deal with these multiple indicators is to construct a time-specific latent typology using a standard LC model (Goodman, 1974; Magidson and Vermunt, 2004). The time-specific LC model would be of the well-known form:

$$P(\mathbf{y}_{it} = \mathbf{m}) = \sum_{x=1}^{X} P(x_{it} = x) \prod_{\ell=1}^{L} P(y_{it\ell} = m_\ell | x_{it} = x),$$

where x_{it} denotes the class membership of case i at occasion t, x a particular latent class and X the number of latent classes. The growth model is no longer specified for the observed responses, but instead for the time-specific class memberships x_{it}. Depending on whether one uses a standard or a LC-based growth model, that part of the model would take the form a mixed-effect multinomial logistic regression model

$$\log \frac{P(x_{it} = x)}{P(x_{it} = 1)} = \sum_{s=0}^{S} \beta_{sx} \cdot f_s(t) + \sum_{s=0}^{S} u_{sxi} \cdot f_s(t).$$

or a LC multinomial logistic regression model

$$\log \frac{P(x_{it} = x|k)}{P(x_{it} = 1|k)} = \sum_{s=0}^{S} \beta_{sxk} \cdot f_s(t).$$

This multilevel LC approach to multiple response growth modeling makes sense if the multiple responses are indicators of a single underlying latent variable. The proposed method is, in fact, similar to LC Markov modeling (Vermunt, Langeheine, and Böckenholt, 1999), but rather than modeling dependencies between the time-specific latent states using a transition structure, here these are modeled using random effects.

Nagin (2001) proposed an alternative growth modeling approach for multiple responses that makes more sense if these are *not* indicators of the same underlying variable. His approach amounts to specifying a separate LC growth structure for each response variable, where the multiple class memberships are allowed to be correlated with one another. Nagin's model can, in fact, be specified as standard LC regression model with specific equality constraints across the latent classes so that these can be interpreted as the categories of a joint discrete latent variable.

The three-level modeling approach cannot only be used to deal with multiple response variables, but also with situations in which individuals are nested within groups, such as children nested within schools, employees nested within firms, patients nested within therapists, or citizens nested within regions. In such situations, one may wish to investigate how the parameters of the specified growth model differ across groups. There are at least four types of hierarchical data extensions of the growth models discussed in the previous section that may be of interest:

1. inclusion of group-level random effects in a standard growth model;
2. inclusion of a group-level discrete latent variable affecting the response variable in a LC-based growth model;
3. inclusion of group-level random effects in the model for the individual-level growth classes;
4. inclusion of a group-level discrete latent variable affecting the individual-level growth classes.

Option 1 yields a standard three-level mixed-effects model, while option 2 yields its LC-based nonparametric counterpart (Vermunt, 2004). These models make it possible to assume that growth curves differ both across groups and individuals. Suppose that the patients participating in the schizophrenia study are nested within therapists, and that it makes sense to assume that there is a therapist effect on the outcome variable. In the standard random effects approach, this might give rise to a mixed-effects model of the form:

$$g[E(y_{jit})] = \sum_{s=0}^{S} \beta_s \cdot f_s(t) + \sum_{s=0}^{S} u_{sij} \cdot f_s(t) + \sum_{s=0}^{S} v_{sj} \cdot f_s(t),$$

where the index j refers to the therapist and v_{sj} are random therapist effects. Similarly, it is possible to expand the LC-based model with therapist effects by assuming that each therapist belongs to one of G classes of therapist. Denoting a particular group-level class by g, the three-level LC regression model can be defined as

$$g[E(y_{jit}|k, g)] = \sum_{s=0}^{S} \beta_{sk} \cdot f_s(t) + \sum_{s=0}^{S} v_{sg} \cdot f_s(t),$$

where v_{sg} are the group-level growth parameters.

Options 3 and 4 listed above are variants of the multilevel LC models described by Vermunt (2003, 2005). The basic idea is that groups may differ with respect to the sizes of the growth classes. In the context of our schizophrenia application, this might mean that the probability of belonging to class 1 – the class that showed a significant improvement over time – varies between therapists. Depending on whether one uses a standard or LC-based random-effects approach for this part of the model, one obtains

$$\log \frac{\pi_{k|\mathbf{z}_{ij},\mathbf{v}_j}}{\pi_{1|\mathbf{z}_{ij},\mathbf{v}_j}} = \gamma_{0k} + \sum_{p=1}^{P} \gamma_{pk} \cdot z_{pij} + v_{0kj},$$

where v_{0kj} is a group-level random intercept, or

$$\log \frac{\pi_{k|\mathbf{z}_{ij},g}}{\pi_{1|\mathbf{z}_{ij},g}} = \gamma_{0kg} + \sum_{p=1}^{P} \gamma_{pk} \cdot z_{pij},$$

where the index g in γ_{0kg} indicates that the intercept in the model for the individual-level classes differs across group-level classes. Whereas in these specifications only the intercept is assumed to be group specific, also the covariate effects – for example, the treatment effect – can be assumed to differ across groups (therapists).

References

Agresti, A. (2002). *Categorical Data Analysis.* New York: Wiley.

Agresti A., Booth, J.G., Hobert, J.P., and Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, 30, 27-80.

Aitkin (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55, 218-234.

Bauer, D.J., and Curran, P.J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectories. *Psychological Methods*, 8, 338-364.

Collins, L. M., and Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, 27, 131-157.

Diggle, P.J., Liang, K.Y., and S.L. Zeger (1994). *Analysis of Longitudinal Data.* Oxford: Clarendon Press.

Goodman, L.A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I - A modified latent structure approach. *American Journal of Sociology*, 79, 1179-1259.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, 22, 1433-1446.

Hedeker, D. (2004). Introduction to growth modeling. D. Kaplan (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Chapter 12, 215-234. Thousand Oakes: Sage Publications.

Hedeker, D., and Gibbons, R.D. (1996). MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine,* 49, 157-176.

Langeheine, R., and Van de Pol (1994). Discrete-time mixed Markov latent class models. In A. Dale and R. B. Davies, (eds.) *Analyzing social and political change: A casebook of methods.* (pp. 170-197). London: Sage.

Lenk, P.J., and DeSarbo, W.S. (2000). Bayesian inference for finite mixture models of generalized linear models with random effects, *Psychometrika*, 65, 93-119.

Lesaffre, E. and Spiessens, B. (2001), On the effect of the number of quadrature points in a logistic random-effects model: an example, *Applied Statistics,* 50, 325-335.

Magidson, J., and Vermunt, J.K. (2004). Latent class models. D. Kaplan (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Chapter 10, 175-198. Thousand Oakes: Sage Publications.

Muthén, B. (2004). Latent variable analysis. Growth mixture modeling and related techniques for longitudinal data. D. Kaplan (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Chapter 19, 345-368. Thousand Oakes: Sage Publications.

Nagin, D.S. (1999). Analyzing developmental trajectories: a semiparametric group-based approach. *Psychological Methods*, 4, 139-157.

Nagin, D.S., and Tremblay, R.E. (2001). Analyzing developmental trajectories of distinct but related behaviors: a group-based method. *Psychological Methods*, 6, 18-34.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2002), Reliable estimation of generalised linear mixed models using adaptive quadrature, *The Stata Journal,* 2, 1-21.

Rodriguez, G. and Goldman, N. (2001). Improved estimation procedures for multilevel models for binary response: a case study. *Journal of the Royal Statistical Society*, *Series A*, 164, 339-355.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models.* London: Chapman & Hall/CRC.

Snijders, T.A.B., and Bosker, R.J. (1999). *Multilevel Analysis* . London: Sage Publications.

Vermunt, J.K. (2003). Multilevel latent class models. *Sociological Methodology*, 33, 213-239.

Vermunt, J.K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, 58, 220- 233.

Vermunt J.K. (2005). Mixed-effects logistic regression models for indirectly observed outcome variables. *Multivariate Behavioral Research*, in press.

Vermunt, J.K. Langeheine, R., and Böckenholt, U. (1999). Latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 178-205.

Vermunt, J.K., and Hagenaars, J.A. (2004). Ordinal longitudinal data analysis. In: R. Hauspie, N. Cameron and L. Molinari (eds.). *Methods in Human Growth Research*, Chapter 15. Cambridge University Press.

Vermunt, J.K. and Magidson, J. (2005). *Technical Guide to Latent GOLD 4.0: Basic and Advanced.* Belmont Massachusetts: Statistical Innovations Inc.

Vermunt, J.K., Rodrigo, M.F., Ato, M. (2001) Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods and Research*, 30, 170- 196.

Vermunt, J.K. and Van Dijk. L. (2001). A nonparametric random-coefficients approach: the latent class regression model. *Multilevel Modelling Newsletter*, 13, 6-13.

Wedel, M. and DeSarbo, W. (2002), Mixture regression models, in: J. Hagenaars and A. McCutcheon (eds.), *Applied Latent Class Analysis*, Cambridge University Press, Cambridge, 366-382.

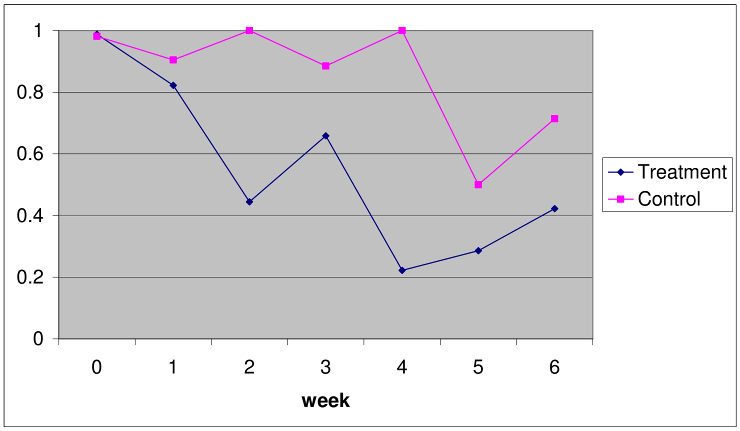

Figure 1: Observed trajectories for the treatment and control group

Model	# Classes	Random	Treatment	Wald	DF	P value	LL	# par	BIC
1	1		response	2.7	1	1.0E-01	-681	4	1386
2	1	u_{0i}, u_{1i}	response	11.3	1	7.6E-04	-614	7	1270
3	2		response	9.5	1	2.1E-03	-619	7	1280
4	3		response	10.8	1	1.0E-03	-607	10	1275
5	4		response	17.5	1	2.9E-05	-602	13	1283
6	2		classes	9.5	1	2.1E-03	-625	6	1286
7	3		classes	12.4	2	2.0E-03	-608	10	1277
8	4		classes	11.2	3	1.1E-02	-601	14	1287
9	2	u_{0i}	classes	21.5	1	3.6E-06	-613	7	1268
10	3	u_{0i}	classes	12.0	2	2.5E-03	-604	11	1274
11	4	u_{0i}	classes	11.2	4	1.1E-02	-601	15	1293
12	2 (1 sqr)		classes	15.9	1	6.5E-05	-620	7	1282
13	3 (1 sqr)		classes	19.3	2	6.5E-05	-597	11	1261
14	2 (1 sqr)	u_{0i}	classes	24.8	1	6.2E-07	-601	8	1250
15	3 (1 sqr)	u_{0i}	classes	19.1	2	7.0E-05	-595	12	1263

Table 1: Test results for the growth models estimated with the schizophrenia data

Model for Responses	Class 1			Class 2		
	β or λ	s.e.	z-value	β or λ	s.e.	z-value
Intercept	8.79	1.17	7.54	6.76	0.88	7.64
StdDev Random Intercept	3.31	0.53	6.27	3.31	0.53	6.27
TIME			- 3.78	1.01	-3.75	
SQ-TIME			1.12	0.29	3.79	
SQRT-TIME	-4.81	0.62	-7.80			

Model for Latent Classes	Class 1		
	γ	s.e.	z-value
Intercept	-0.65	0.31	-2.13
Treatment	1.78	0.36	4.98

Table 2: Parameter estimates obtained with Model 14

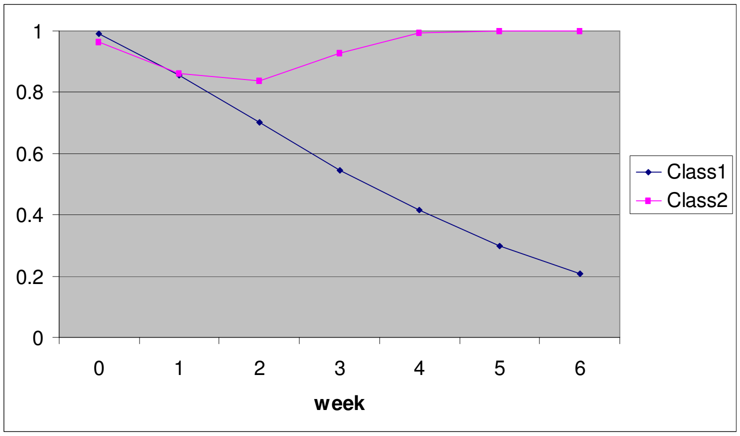

Figure 2: Class-specific trajectories obtained with Model 14