Chapter 4
Longitudinal Research Using Mixture Models

Jeroen K. Vermunt

Abstract This chapter provides a state-of-the-art overview of the use of mixture and latent class models for the analysis of longitudinal data. It first describes the three basic types mixture models for longitudinal data: the mixture growth, mixture Markov, and latent Markov model. Subsequently, it presents an integrating framework merging various recent developments in software and algorithms, yielding mixture models for longitudinal data that can (1) not only be used with categorical, but also with continuous response variables (as well as combinations of these), (2) be used with very long time series, (3) include covariates (which can be numeric or categorical, as well as time-constant or time-varying), (4) include parameter restrictions yielding interesting measurement models, and (5) deal with missing values (which is very important in longitudinal research). Moreover, it discusses other advanced models, such as latent Markov models with dependent classification errors across time points, mixture growth and latent Markov models with random effects, and latent Markov models for multilevel data and multiple processes. The appendix shows how the presented models can be defined using the Latent GOLD syntax system (Vermunt and Magidson, 2005, 2008).

4.1 Introduction

The aim of this chapter is to provide a state-of-the-art overview of the use of mixture and latent class models for the analysis of longitudinal data. While in the more formal statistical literature the term "latent class model" is typically reserved for a specific type of mixture model (Everitt and Hand, 1981; McLachlan and Peel, 2000), namely for the mixture model for categorical responses described by Lazarsfeld and Henry (1968) and Goodman (1974), in applied fields these terms are used in-

Jeroen K. Vermunt
Department of Methodology and Statistics, Tilburg University, e-mail: j.k.vermunt@uvt.nl

terchangeably. This is also what I will do in this chapter; that is, I will use the terms mixture model and latent class model to denote latent variable models containing one or more discrete latent variables.

In the context of longitudinal research, a mixture model is a latent variable model containing a single or multiple time-constant or time-varying discrete latent variables. The best-known examples are the latent class or mixture growth model (Muthén 2004; Nagin, 1999; Vermunt, 2007), the mixture Markov model (Goodman, 1961; Poulsen, 1990; van de Pol and Langeheine, 1990; Vermunt, 1997a), and the latent or hidden Markov model (Baum, Petrie, Soules, and Weiss, 1970; Bartolucci, Pennoni, and Francis, 2007; Collins and Wugalter, 1992; Mooijaart and van Montfort, 2007; Poulsen, 1990; van de Pol and de Leeuw, 1986; Vermunt, Langeheine, and Böckenholt, 1999; Wiggins, 1973).

Diggle, Liang, and Zeger (1994) distinguished three main approaches for analyzing longitudinal data: (1) marginal or population-average models, (2) random-effects, subject-specific, or growth models, and (3) conditional or transitional models. Marginal models focus on the change in univariate distributions, growth models study individual-level change over time, and transitional models describe changes between consecutive time points. These three approaches do not only differ with regard to the questions they address, but also in how 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 dependencies using latent variables (random effects). In marginal models, dependencies are not explicitly modeled, but dealt with as found in the data and in general are taken into account in a more ad hoc way in the estimation procedure. Variants of transitional, growth, and marginal models have been developed for both continuous and categorical response variables.

Discrete latent variables may be introduced in longitudinal data models for various purposes, the most important of which are dealing with unobserved heterogeneity, dealing with measurement error, and clustering. Or more specific, in context of the three approaches described above, latent classes can be introduced in growth models for clustering and dealing with unobserved heterogeneity (yielding mixture growth models), and in transitional models for dealing with measurement error, static or dynamic clustering, and dealing with unobserved heterogeneity (yielding mixture and latent Markov models). Hagenaars (1990) and Bergsma, Croon, and Hagenaars (2009) used a latent class marginal model for dealing with measurement error in categorical responses.

Starting point of this chapter are the simplest variants of the three basic mixture models for longitudinal data: the mixture growth, mixture Markov, and latent Markov model. Recent developments in software and algorithms have resulted in many extensions of these basic models; that is, mixture models for longitudinal data can nowadays (1) not only be used with categorical, but also continuous response variables (as well as combinations of these), (2) be used with very long time series, (3) include covariates (which can be numeric or categorical, as well as time-constant or time-varying), (4) include parameter restrictions yielding interesting measurement models, and (5) deal with missing values (which is very important in longitudinal research). I will present an integrating framework including all these extended features. Moreover, I will discuss other more advanced features, such as latent Markov models with dependent classification errors across time points, mixture growth and latent Markov models with random effects, and latent Markov models for multilevel data and multiple processes.

There is some overlap between the current chapter and Hagenaars' chapter in this volume, which deals with longitudinal categorical data analysis using the log-linear SEM approach implemented in the LEM software (Vermunt 1997b). On the one hand, this SEM framework is more general than the framework discussed here because it allows defining any type of categorical data model. On the other hand, it is more restricted since it deals with categorical data (responses and covariates) only, and, because it is not tailored for longitudinal data analysis, it can, for example, not be used with long time series.

In the remaining of this chapter, I will first describe the three basic mixture models for longitudinal data analysis, including some of their extensions. Then a general framework is presented containing each of these as special cases, and allowing various interesting combinations. Though several other recent developments could be fit into an even more general framework, these will be discussed as separate extensions in a next section. The last section presents two applications, and the Appendix illustrates how the models concerned can be defined using the Latent GOLD syntax system (Vermunt and Magidson, 2005, 2008).

4.2 The three basic models

Before describing the three basic types of mixture models for longitudinal data, I will introduce the relevant notation. Longitudinal data sets analyzed with the models described in this chapter will typically contain information on multiple response variables from multiple subjects at multiple time points. Let y_{itj} denote the response of subject i on response variable j at occasion t, where $1 \leq i \leq N$, $1 \leq j \leq J$, and $0 \leq t \leq T_i$. Here, N is the number subjects, J the number of response variables, and $T_i + 1$ is the number of measurement occasions for subject i. Note that we use the index i in T_i to be able to deal with the rather common situation in which the number of measurement occasions differ across individuals. The vector collecting the responses of subject i at occasion t is denoted as $\mathbf{y}_{it}$ and the vector collecting all responses of subject i as $\mathbf{y}_i$.

Three remarks have to made about the response variables. First, response variables may also be referred to as output variables, dependent variables, indicators, items, manifest variables, etc. Second, response variables cannot only be categorical variables – in which case $1 \leq y_{itj} = m_j \leq M_j$, with M_j being the number of categories and m_j a particular category of response variable j – but also continuous variables or counts. As we will see below, the scale type of y_{itj} affects its conditional distribution, as well as the type of regression model one may specify to restrict its expected value. Third, often only one response variable is available, in which case the index j can be dropped, yielding the simpler notation y_{it}.

Longitudinal data models may not only contain response variables, but also predictors, also referred to as input variables, independent variables, covariates, concomitant variables, etc. The vectors of time-constant predictors and time-varying predictors at occasion t are denoted by $\mathbf{z}_i$ and $\mathbf{z}_{it}$, respectively. Note that predictors cannot only be numeric but also categorical variables, which will typically be included in the model using a series of dummies or effects. Note also that time and functions of time can be included in the vector of time-varying predictors.

What makes a statistical model a latent class or mixture model is that it contains either a time-constant or a time-varying (or dynamic) discrete latent variable. These two types of latent variables are denoted by w_i and x_{it}, respectively, their number of categories by L and K, and one of their categories by ℓ and k_t. That is, $1 \leq w_i = \ell \leq L$ and $1 \leq x_{it} = k_t \leq K$. To clearly distinguish the two types of latent variables, I will refer to w_i as a latent class and to x_{it} as a latent state.

4.2.1 Mixture growth model

A latent class or mixture growth model is a model for a single response variable y_{it} measured at $T_i + 1$ occasions (Nagin, 1999; Muthén, 2004; Vermunt 2007). In fact, a regression model is specified for y_{it} in which time serves as the only explanatory variable. The aim of growth models is to determine whether individuals differ with respect to the parameters of the growth model, where differences are usually modeled using random effects under the assumption that these come from a multivariate normal distribution.

There are two possible reasons for introducing latent classes in a growth model. First, one may wish to identify (interpretable) clusters of individuals with similar growth parameters. This is similar to the aim of a standard latent class model, with the difference that the observed variables used to find the clusters are repeated measurements of a single response variable rather than multiple items or indicators. A second reason for using a mixture growth model is more technical; that is, one may wish to specify a model with random effects without making strong distributional assumptions about the random effects. This yields what is referred to as a non-parametric maximum likelihood (NPML) approach to random effects modeling, which cannot only be used in the context of longitudinal data analysis but in any type of two-level regression model (Aitkin, 1999; Skrondal and Rabe-Hesketh, 2004; Vermunt 2004; Vermunt and van Dijk, 2001).

A mixture growth model is a statistical model for $f(\mathbf{y}_i|\mathbf{z}_i)$, the probability density of the $T_i + 1$ responses of subject i collected in the vector $\mathbf{y}_i$ conditional on a set of time variables collected in the vector $\mathbf{z}_i$. It can be formulated using the following three equations:

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{\ell=1}^{L} P(w_i = \ell) f(\mathbf{y}_i|w_i = \ell, \mathbf{z}_i) \tag{4.1}$$

$$f(\mathbf{y}_i|w_i = \ell, \mathbf{z}_i) = \prod_{t=0}^{T_i} f(y_{it}|w_i = \ell, \mathbf{z}_{it}), \tag{4.2}$$

$$g[E(y_{it}|w_i = \ell, \mathbf{z}_{it})] = \beta_{0\ell} + \sum_{p=1}^{P} \beta_{p\ell} z_{itp}, \tag{4.3}$$

The first of these three equations indicates that the density $f(\mathbf{y}_i|\mathbf{z}_i)$ is a weighted average of class-specific densities $f(\mathbf{y}_i|w_i = \ell, \mathbf{z}_i)$, where the class proportions $P(w_i = \ell)$ serve as weights. More intuitively, the likelihood of the set of responses $\mathbf{y}_i$ depends on the class membership of person i (on w_i). But because the class membership is unknown, the likelihood is obtained by averaging over the L classes. Note that this kind of reasoning applies to any type of mixture or latent class model.

The second equation states that the joint distribution of $\mathbf{y_i}$ given w_i and $\mathbf{z}_i$ (appearing in the equation 4.1) can be obtained as a product of the $T_i + 1$ univariate marginal distributions $f(y_{it}|w_i = \ell, \mathbf{z}_{it})$. This expresses that the responses are assumed to be independent across time points given a person's class membership, which in the latent class analysis literature is usually referred to as the local independence assumption. The specific form chosen for $f(y_{it}|w_i = \ell, \mathbf{z}_{it})$ depends on the scale type of y_{it}. For example, with binary responses on will often use a binomial distribution, with continuous responses a normal distribution, and with counts a Poisson distribution.

The third equation shows that the responses are related to the time variables using a regression model from the generalized linear modeling (GLM) family (Agresti, 2002). After applying an appropriate transformation $g(\cdot)$, which in GLM terminology is referred to as a link function, the expected value of y_{it} is modeled as a linear function of a set of P time variables. For example, with $P = 2$, $z_{it1} = t$, and $z_{it2} = t^2$, the expected value of y_{it} would be a quadratic function of time. A key feature is that the regression parameters capturing the time dependence of the responses are assumed to differ across latent classes; that is, each class has its own pattern of change. Note that by defining a regression model for y_{it} one restricts the density $f(y_{it}|w_i = \ell, \mathbf{z}_{it})$ which appears in equation 4.2. In fact, we have a latent class model with restrictions on the class-specific response probabilities/densities which are specified by assuming that the class-specific means are functions of time.

The basic model described in equations (4.1)–(4.3) can be extended in various ways. One important extension is the inclusion of covariates in the model for w_i. Similarly to the model proposed by Dayton and Macready (1988) and van der Heijden, Dessens, and Böckenholt (1996) in the context of standard latent class analysis, this involves replacing $P(w_i = \ell)$ in equation (4.2) by $P(w_i = \ell|\mathbf{z}_i)$ and defining a multinomial logit model for w_i; that is,

$$P(w_i = \ell|\mathbf{z}_i) = \frac{\exp(\gamma_{0\ell} + \sum_{q=1}^{Q} \gamma_{q\ell} z_{iq})}{\sum_{\ell'=1}^{L} \exp(\gamma_{0\ell'} + \sum_{q=1}^{Q} \gamma_{q\ell'} z_{iq})}, \tag{4.4}$$

where for identification we may for example set $\gamma_{0L} = \gamma_{qL} = 0$, yielding what is usually referred to as a baseline category logit model (Agresti, 2002).

Another extension is the inclusion of other predictors than time in the model for y_{it} (in equation 4.3). These could serve as control variables when one is interested determining class-specific change patterns after accounting for the fact that other variables may partially explain the observed change. But other predictors may also be the ones of main interest, in which case the aim of the analysis changes somewhat and the mixture variable will mainly be used to capture unobserved heterogeneity using the NPML approach mentioned above.

4.2.2 Mixture Markov model

As mentioned in the introduction, rather than using a growth model, longitudinal data may also be modeled using a transitional or conditional model. The best-known model from this family is the (first-order) Markov model, which assumes that y_{it} depends on y_{it-1} but not on values at earlier occasions. Similarly to mixture growth models, in mixture Markov models, one will typically have a single response variable. The main reason for using a mixture variant of a Markov model is to deal with unobserved heterogeneity; that is, to account for the fact that transition probabilities/densities are not homogeneous, but instead may differ across (unobserved) subgroups. A more substantive reason may be to find meaningful clusters of individuals with different change patterns. An example of the latter is the application by Dias and Vermunt (2007) in which market segments were identified based on website users' search patterns.

The mixture Markov can be formulated as follows:

$$f(\mathbf{y}_i) = \sum_{\ell=1}^{L} P(w_i = \ell) f(y_{i0}|w_i = \ell) \prod_{t=1}^{T_i} f(y_{it}|y_{it-1}, w_i = \ell). \tag{4.5}$$

As can be seen, the L latent classes are assumed to differ with respect to the initial-state and transition densities. Variants of this model for continuous response variables – referred to as mixture dynamic regression and mixture autoregressive models – were proposed by Kaplan (2005) and Wong and Li (2000). However, most applications of the mixture Markov model concern categorical response variables (Dias and Vermunt, 2007; Poulsen, 1990), in which case the model may also be written as

$$P(\mathbf{y}_i) = \sum_{\ell=1}^{L} P(w_i = \ell) P(y_{i0} = m_0|w_i = \ell) \left[\prod_{t=1}^{T_i} P(y_{it} = m_t|y_{it-1} = m_{t-1}, w_i = \ell) \right]; \tag{4.6}$$

that is, in terms of initial-state and transition probabilities.

Various special cases of the mixture Markov model can be obtained by restricting the transition probabilities. A well-documented special case is the mover-stayer model (Goodman, 1961), which is a two-class model ($L = 2$) where one class (say the second) contains respondents who have a zero probability of making a transition: $P(y_{it} = m_t | y_{it-1} = m_{t-1}, w_i = 2) = 0$ for $m_t = m_{t-1}$. Another special case is a Markov model with a random responder class for which the measurements are independent across time points: $P(y_{it} = m_t | y_{it-1} = m_{t-1}, w_i = 2) = P(y_{it} = m_t | w_i = 2)$.

Various extensions of the simple models described in equations (4.5) and (4.6) are possible, the most important of which is the introduction of predictors affecting the class membership, the initial state, and the transitions. The first extension was discussed above in the context of mixture growth models (see equation 4.4). Covariates can be allowed to affect the initial state and the transitions by defining regression models for y_{i0} and y_{it}, which in the case of a categorical response will be logistic regression models. With Q predictors in the model for y_{i0} and P time-varying predictors in the model for y_{it} conditional on y_{it-1}, we get

$$P(y_{i0} = m | w_i = \ell, \mathbf{z}_{i0}) = \frac{\exp(\beta^0_{\ell m} + \sum_{q=1}^{Q} \beta^0_{q+L,m} z_{i0q})}{\sum_{m'=1}^{M} \exp(\beta^0_{\ell m'} + \sum_{q=1}^{Q} \beta^0_{q+L,m'} z_{i0q})}, \tag{4.7}$$

$$P(y_{it} = m | y_{it-1} = n, w_i = \ell, \mathbf{z}_{it}) = \frac{\exp(\beta_{\ell nm} + \sum_{p=1}^{P} \beta_{p+L,nm} z_{itp})}{\sum_{m'=1}^{M} \exp(\beta_{\ell nm'} + \sum_{p=1}^{P} \beta_{p+L,nm'} z_{itp})}. \tag{4.8}$$

As in a standard multinomial logit model, identifying restrictions on $\beta^0_{\ell m}$ and $\beta^0_{q+L,m}$ are required, for example, they may be fixed to 0 for $m = M$. The same applies to the $\beta_{\ell nm}$ and $\beta_{p+L,nm}$ parameters for which one constraint is needed for each origin state n. A coding referred to as transition coding by Vermunt and Magidson (2008) involves setting $\beta_{\ell nn} = \beta_{p+L,nn} = 0$; that is, the coefficients are fixed to 0 for $m = n$, which implies that the free coefficients can be interpreted as effects on the logit of a transition from n to m.

4.2.3 Latent Markov model

Whereas mixture growth and mixture Markov models contain a static categorical latent variable (w_i), a latent Markov model is a mixture model with a dynamic categorical latent variable – denoted by x_{it}. One of the key elements of this model is that latent-state transitions occurring over time are modeled using a first-order Markov structure. The second key element is that the latent states are connected to one or more observed response variables via a latent class structure with conditional densities $f(y_{itj} | x_{it} = k_t)$. The latent Markov model – which is also referred to as hidden Markov model (Baum et al., 1970; McDonald and Zucchini, 1997), Markov switching or regime switching model (Goldfeld and Quandt, 1973), and latent transition model (Collins and Wugalter, 1992) – can be defined as follows (Poulsen, 1990; van de Pol and de Leeuw, 1986, Wiggins, 1973):

$$f(\mathbf{y}_i) = \sum_{k_0=1}^{K} \sum_{k_1=1}^{K} \cdots \sum_{k_{T_i}=1}^{K} P(x_{i0} = k_0) \left[\prod_{t=1}^{T_i} P(x_{it} = k_t | x_{it-1} = k_{t-1}) \right] \left[\prod_{t=0}^{T_i} \prod_{j=1}^{J} f(y_{itj} | x_{it} = k_t) \right]. \quad (4.9)$$

Besides the Markov assumption for the latent states and the local independence assumption for the responses within occasions, the latent Markov model assumes that responses are independent across occasions conditional on the latent states. The latter implies that the observed associations across time points are assumed to be explained by the autocorrelation structure for the latent states.

The typical applications of this model concern either a single continuous response variable (Schmittmann, Dolan, van der Maas, and Neale, 2005; Dias, Vermunt, and Ramos, 2009), a single categorical response variable (Magidson, Vermunt, and Tran, 2009; Poulsen, 1990; van de Pol and de Leeuw, 1986; Wiggins, 1973), or multiple categorical responses (Bartolucci, Pennoni, and Francis, 2007; Collins and Wugalter, 1992; Paas, Vermunt, and Bijmolt, 2007). With a single continuous response, the model may either be used for clustering or for dealing with unobserved heterogeneity, where contrary to the mixture models described above respondents may switch across clusters or mixture components over time. When applied with a single categorical response variable, one will typically assume that the number of latent states equals the number or categories of the response variable: $K = M$. Moreover, model restrictions are required to obtain an identified model, the most common of which are time-homogeneous transition probabilities or time-homogeneous misclassification probabilities. The aim is to split observed changes in the response into a true change component and a measurement error component. When used with multiple indicators, the model is a longitudinal data extension of the standard latent class model (Hagenaars, 1990). The time-specific latent states can be seen as clusters or types which differ in their responses on the J indicators, and the Markovian transition structure is used to describe and predict changes that may occur across adjacent measurement occasions.

The most straightforward extension of the latent Markov model presented in equation (4.9) involves the inclusion of explanatory variables affecting the initial state and the transition probabilities. Special cases are the multiple-group latent Markov model proposed by van de Pol and Langeheine (1990), the latent Markov model with covariates proposed by Vermunt, Langeheine and Böckenholt (1999), and the input-output model described by Mooijaart and van Montfort (2007). Models with predictors can be defined using similar logistic equations as we used for the mixture Markov model (see equations 4.7 and 4.8), but now for x_{i0} and x_{it} instead of y_{i0} and y_{it} and without conditioning on w_i; that is,

$$P(x_{i0}=k|\mathbf{z}_{i0}) = \frac{\exp(\alpha^0_{0k}+\sum_{q=1}^{Q}\alpha^0_{qk}\,z_{i0q})}{\sum_{k'=1}^{K}\exp(\alpha^0_{0k'}+\sum_{q=1}^{Q}\alpha^0_{qk'}\,z_{i0q})},$$
$$P(x_{it}=k|x_{it-1}=n,\mathbf{z}_{it}) = \frac{\exp(\alpha_{0nk}+\sum_{p=1}^{P}\alpha_{pnk}\,z_{itp})}{\sum_{k'=1}^{K}\exp(\alpha_{0nk'}+\sum_{p=1}^{P}\alpha_{pnk'}\,z_{itp})}.$$

Again, identifying restrictions are needed on the α^0_{0k}, α^0_{qk}, α_{0nk}, and α_{pnk} parameters, where for the latter two one may again use transition coding.

Other extensions include models with predictors affecting the responses, mixture variants with a time-constant latent variable w_i, models with restrictions on the transition probabilities $P(x_{it}=k_t|x_{it-1}=k_{t-1})$ or the response densities $f(y_{it}|x_{it}=k_t)$, models that relax the assumption that measurement errors are independent across occasions, and models with multiple dynamic latent variables. These and other extensions will be discussed below.

4.3 The mixture latent Markov model

4.3.1 The general model

In the previous section, we described three types of mixture models for longitudinal data analysis. These models contained either a time-constant (w_i) or time-varying (x_{it}) discrete latent variables. In this section, I present the mixture latent Markov with covariates, which can be seen as the encompassing model which contains the three models discussed above as special cases, as well as which allows various interesting extensions and combinations of these. The presented mixture latent Markov model is an expanded version of the mixed Markov latent class model proposed by van de Pol and Langeheine (1990) in the sense that it cannot only be used with categorical but also with continuous responses, it may contain time-constant and time-varying covariates, and it can be used when the number of time points is large. For simplicity of exposition, here, I will restrict myself to models with a single time-constant and a single time-varying latent variable, but in the next section I will present extensions for multiple time-constant and multiple time-varying latent variables.

The general model of interest is the following mixture latent Markov model:

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{\ell=1}^{L}\sum_{k_0=1}^{K}\sum_{k_1=1}^{K}\dots\sum_{k_{T_i}=1}^{K} P(w_i=\ell,\mathbf{x}_i=\mathbf{k}|\mathbf{z}_i)\, f(\mathbf{y}_i|w_i=\ell,\mathbf{x}_i=\mathbf{k},\mathbf{z}_i) \tag{4.10}$$

$$= \sum_{\ell=1}^{L}\sum_{k_0=1}^{K}\sum_{k_1=1}^{K}\dots\sum_{k_{T_i}=1}^{K} P(w_i=\ell|\mathbf{z}_i)P(x_{i0}=k_0|w_i=\ell,\mathbf{z}_{i0}) \left[\prod_{t=1}^{T_i} P(x_{it}=k_t|x_{it-1}=k_{t-1},w_i=\ell,\mathbf{z}_{it})\right] \left[\prod_{t=0}^{T_i}\prod_{j=1}^{J} f(y_{itj}|x_{it}=k_t,w_i=\ell,\mathbf{z}_{it})\right]. \tag{4.11}$$

As many statistical models, the model in equations (4.10) and (4.11) is a model for $f(\mathbf{y}_i|\mathbf{z}_i)$, the (probability) density associated with the responses of subject i conditional on his/her observed covariate values. The right-hand side of equation (4.10) shows that we are dealing with a mixture model containing a time constant latent variable (w_i) and $T+1$ realizations of a time-varying latent variable (collected in the vector $\mathbf{x}_i$). The total number of mixture components (or latent classes) for individual i equals $L \cdot K^{T_i+1}$, which is the product of the number of categories of w_i and x_{it} for $t = 0, 1, 2, ..., T_i$. Equation (4.10) shows that, as in any mixture model, $f(\mathbf{y}_i|\mathbf{z}_i)$ is obtained as a weighted average of class-specific probability densities – here $f(\mathbf{y}_i|w_i = \ell, \mathbf{x}_i = \mathbf{k}, \mathbf{z}_i)$ – where the (prior) class membership probabilities or mixture proportions – here $P(w_i = \ell, \mathbf{x}_i = \mathbf{k}|\mathbf{z}_i)$ – serve as weights (Everitt and Hand, 1981; McLachlan and Peel, 2000).

Equation (4.11) shows the specific structure assumed for the mixture proportions and the class-specific densities. The assumption for $P(w_i = \ell, \mathbf{x}_i = \mathbf{k}|\mathbf{z}_i)$ is that conditional on w_i and $\mathbf{z}_i$, x_{it} is associated only with $x_{i,t-1}$ and $x_{i,t+1}$ and thus not with the states occupied at the other time points – the well-know first-order Markov assumption. For $f(\mathbf{y}_i|w_i = \ell, \mathbf{x}_i = \mathbf{k}, \mathbf{z}_i)$ two assumptions are made: (1) conditionally on w_i, x_{it}, and $\mathbf{z}_{it}$, the J responses at occasion t are independent of the latent states and the responses at other time points, and (2) conditionally on w_i, x_{it}, and $\mathbf{z}_{it}$, the J responses at occasion t are mutually independent, which is referred to as the local independence assumption in latent class analysis (Goodman, 1974).

As can be seen from equation (4.11), the models of interest contain four different kinds of model probabilities/densities:

- $P(w_i = \ell|\mathbf{z}_i)$ is the probability of belonging to a particular latent class conditional on a person's covariate values,
- $P(x_{i0} = k_0|w_i = \ell, \mathbf{z}_{i0})$ is an initial-state probability; that is, the probability of having a particular latent initial state conditional on an individual's class membership and covariate values at $t = 0$,
- $P(x_{it} = k_t|x_{it-1} = k_{t-1}, w_i = \ell, \mathbf{z}_{it})$ is a latent transition probability; that is, the probability of being in a particular latent state at time point t conditional on the latent state state at time point $t-1$, class membership, and time-varying covariate values,

- $f(y_{itj}|x_{it} = k_t, w_i = \ell, \mathbf{z}_{it})$ is a response density, that is, the density corresponding to the observed value for person i of response variable j at time point t conditional on the latent state occupied at time point t, class membership w_i, and time-varying covariate values.

Typically, these four sets of probabilities/densities will be parameterized and restricted by means of regression models from the generalized linear modeling family. As shown in various examples in the previous section, this is especially useful when a model contains covariates, where time itself may be one of the time-varying covariates of main interest.

The three key elements of the mixture latent Markov model described in equation (4.11) are that it can take into account (1) unobserved heterogeneity, (2) autocorrelation, and (3) measurement error. Unobserved heterogeneity is captured by the time-constant latent variable w_i, autocorrelations are captured by the first-order Markov transition process in which the state at time point t may depend on the state at time point $t-1$, and measurement error or misclassification is accounted for by allowing an imperfect relationship between the time-specific latent states x_{it} and the observed responses y_{itj}. Note that these are three of the main elements that should be taken into account in the analysis of longitudinal data; that is, the inter-individual variability in patterns of change, the tendency to stay in the same state between consecutive occasions, and the spurious change resulting from measurement error in observed responses.

4.3.2 Estimation, missing data, and time-unit setting

Parameters of the mixture latent Markov model can be estimated by means of maximum likelihood (ML). For that purpose, it advisable to use a special variant of the expectation maximization (EM) algorithm that is usually referred to as the forward-backward or Baum-Welch algorithm (Baum et al., 1970; McDonald and Zucchini, 1997). This is an EM algorithm in which the E step, which involves computing the relevant set posterior distributions given the current parameter estimates and the observed data, is implemented in a way that is tailored to the models we are dealing with. More specifically, this special algorithm is needed because our model contains a potentially huge number of entries in the joint posterior latent distribution $P(w_i = \ell, \mathbf{x}_i = \mathbf{k}|\mathbf{y}_i, \mathbf{z}_i)$, except for cases where T, L and K are all small. For example, in a fairly moderate sized situation where $T_i = 10$, $L = 2$ and $K = 3$, the number of entries in the joint posterior distribution already equals $2 \cdot 3^{11} = 354294$, a number which is impossible to process and store for all N subjects as has to be done within standard EM. The Baum-Welch algorithm circumvents the computation of this joint posterior distribution making use of the conditional independencies implied by the model; that is, rather than computing the joint distribution and subsequently obtaining the relevant marginals, it computes the relevant marginals directly. For more details, we refer to Vermunt, Tran, and Magidson (2008) who also provided the generalized version of the Baum-Welch algorithm which is required for the estimation of the mixture latent Markov model presented in equation (4.11) and which is implemented in the Latent GOLD 4.5 program (Vermunt and Magidson, 2008). Rather than using ML estimation, it is also possible to estimate these models using Bayesian estimation procedures, an excellent overview of which is provided by Frühwirth-Schnatter (2006).

A common phenomenon in the analysis of longitudinal data is the occurrence of missing data. Subjects may have missing values either because they refused to participate at some occasions or because it is part of the research design. A nice feature of the approach described here is that it can easily accommodate missing data in the ML estimation of the unknown model parameter. Let δ_{it} be an indicator variable taking on the value 1 if subject i provides information for occasion t and 0 if this information is missing. The only required change with missing data is the following modification of the model for the response density $f(\mathbf{y}_i|w_i = \ell, \mathbf{x}_i = \mathbf{k}, \mathbf{z}_i)$:

$$f(\mathbf{y}_i|w_i = \ell, \mathbf{x}_i = \mathbf{k}, \mathbf{z}_i) = \prod_{t=0}^{T_i} [P(\mathbf{y}_{it}|x_{it} = k_t, w_i = \ell, \mathbf{z}_{it})]^{\delta_{it}} .$$

For $\delta_{it} = 1$, nothing changes compared to what we had before. However, for $\delta_{it} = 0$, the time-specific conditional density becomes 1, which means that the responses of a time point with missing values are skipped. Actually, for each pattern of missing data, we have a mixture latent Markov for a different set of occasions. Two limitations of the ML estimation procedure with missing values should be mentioned: (1) it can deal with missing values on response variables, but not with missing values on covariates, and (2) it assumes that the missing data are missing at random (MAR). The first limitation may be problematic when there are time-varying covariates for which the values are also missing. However, in various special cases discussed below – the ones that do not use a transition structure – it is not a problem if time-varying covariates are missing for the time points in which the responses are missing. The second limitation concerns the assumed missing data mechanism: MAR is the least restrictive mechanism under which ML estimation can be used without the need of specifying the exact mechanism causing the missing data; that is, under which the missing data mechanism is ignorable for likelihood-based inference (Little and Rubin, 1987; Schafer, 1997). It is possible to relax the MAR assumption by explicitly defining a not missing at random (NMAR) mechanism as a part of the model to be estimated (Fay, 1986; Vermunt 1997a).

An issue strongly related to missing data is the one of unequally spaced measurement occasions. As long as the model parameters defining the transition probabilities are assumed to be occasion specific, no special arrangements are needed. If this is not the case, unequally spaced measurements can be handled by defining a grid of equally spaced time points containing all measurement occasions. Using this technique, the information on the extraneous occasions can be treated as missing data for all subjects. An alternative is to use a continuous-time rather than a discrete-time framework (Böckenholt, 2005), which can be seen as the limiting case in which the elapsed time between consecutive time points in the grid approaches zero.

	Model name	Transition structure	Unobserved heterogeneity	Measurement error
I.	Mixture latent Markov	yes	yes	yes
II.	Mixture Markov	yes	yes	no
III.	Latent Markov	yes	no	yes
IV.	Standard Markov*	yes	no	no
V.	Mixture latent growth	no	yes	yes
VI.	Mixture growth	no	yes	no
VII.	Standard latent class	no	no	yes
VIII.	Independence*	no	no	no

*: This model is not a latent class model.

Another issue related to missing data is the choice of the time variable and the corresponding starting point of the process. The most common approach is to use calender time as the time variable and to define the first measurement occasion to be $t = 0$. However, one may, for example, also use age as the relevant time variable, as I do in the second empirical example. Although children's ages at the first measurement vary between 11 and 17, I use age 11 as $t = 0$. This implies that for a child that is 12 years of age information at $t = 0$ is treated as missing, for a child that is 13 years of age information a $t = 0$ and $t = 1$ is treated as missing, etc.

4.3.3 The most important special cases

Table 4.3.3 lists the various special cases that can be derived from the mixture latent Markov model defined in equation in (4.11) by assuming that one or more of its three elements – transition structure, measurement error, and unobserved heterogeneity – is not present or needs to be ignored because the data is not informative enough to deal with it. Models I-III and V-VII are latent class models, but IV and VIII are not. Model VII differs from models I-VI in that it is a model for repeated cross-sectional data rather than a model for panel data. Below we describe the various special cases in more detail.

4.3.3.1 Mixture latent Markov

First of all, it is possible to define simpler versions of the mixture latent Markov model itself. Actually, the mixed Markov latent class model proposed by van de Pol and Langeheine (1990) which served as an inspiration for our model is the special case in which responses are categorical and in which no covariates are present. van de Pol and Langeheine (1990) proposed a variant in which the four types of model probabilities could differ across categories of a grouping variable (see also Langeheine and van de Pol, 2002). A similar model is obtained by replacing the $\mathbf{z}_i$ and $\mathbf{z}_{it}$ by a single categorical covariate coded using a set of dummy predictors.

4.3.3.2 Mixture Markov

The mixture Markov model for a categorical response variable (Poulsen, 1990; Dias and Vermunt, 2007) is the special case of the model presented in equation (4.11) when there is a single response variable ($J = 1$) that is assumed to be measured without error, which is specified by $K = M$ and $P(y_{it} = m_t|x_{it} = k_t) = 1$ if $m_t = k_t$ and 0 otherwise. Note that y_{it} is assumed not to depend on w and $\mathbf{z}_{it}$ but only on x_t. The mover-stayer model (Goodman, 1961) can be obtained by setting $L = 2$ and fixing the transition probabilities to 0 for the second class: $P(x_{it} = k_t|x_{tt-1} = k_{t-1}, w_i = 2, \mathbf{z}_{it}) = 0$ if $k_t = k_{t-1}$ and 0 otherwise. Note that the mover-stayer constraint cannot only be imposed in the mixture Markov but also in the mixture latent Markov, in which case transitions across imperfectly measured stated are assumed not to occur in the stayer class.

4.3.3.3 Latent Markov model

The latent Markov, latent transition, or hidden Markov model (Baum et al., 1970; Collins and Wugalter, 1992; Mooijaart and van Montfort, 2007; van de Pol and de Leeuw, 1996; Vermunt, Langeheine, and Böckenholt, 1999, Wiggins, 1973) is the special case of the mixture latent Markov that is obtained by eliminating the time-constant latent variable w_i from the model; that is, by assuming that there is no unobserved heterogeneity or that it can be ignored. The latent Markov model can be obtained without modifying the formulae, but by simply assuming that $L = 1$; that is, that all subject belong to the same latent class.

4.3.3.4 Markov model

By assuming both perfect measurement as in the mixture Markov model and absence of unobserved heterogeneity as in the latent Markov model, one obtains a standard Markov model, which is no longer a mixture model. This model can further serve as a simple starting point for longitudinal applications with a single response variable, where one wishes to assume a Markov structure. It provides a baseline for comparison to the three more extended models discussed above. Use of these more extended models makes sense only if they provide a significantly better description of the data than the simple Markov model.

4.3.3.5 Mixture latent growth model

Now we turn to latent class models for longitudinal research that are not transition or Markov models. These mixture growth models assume that dependencies between measurement occasions can be captured by the time-constant latent variable w_i. The most extended variant is the mixture latent growth model, which is obtained from the mixture latent Markov model by imposing the constraint $P(x_{it} = k_t | x_{i,t-1} = k_{t-1}, w_i = \ell, \mathbf{z}_{it}) = P(x_{it} = k_t | w_i = \ell, \mathbf{z}_{it})$. This model is a variant for longitudinal data of the multilevel latent class model proposed by Vermunt (2003): subjects are the higher-level units and time points the lower-level units. It should be noted that application of this very interesting model with categorical responses requires that there be at least two response variables ($J \geq 2$).

In mixture growth models one will typically pay a lot of attention to the modeling of the time dependence of the state occupied at the different time points. The latent class or mixture approach allows identifying subgroups (categories of the time-constant latent variable w_i) with different change patterns (Nagin, 1999). The extension provided by the mixture latent growth model is that the dynamic dependent variable is itself a (discrete) latent variable which is measured by multiple indicators.

4.3.3.6 Mixture growth model

The mixture or latent class growth model (Nagin, 1999, Muthén, 2004; Vermunt, 2007) for a categorical response variable can be seen as a restricted variant of the mixture latent growth model; i.e., as a model for a single indicator measured without error. The extra constraint is the same as the one used in the mixture Markov model: $K = M$ and $P(y_{it} = m_t | x_{it} = k_t) = 1$ if $m_t = k_t$ and 0 otherwise. A more natural way to define the mixture growth model is by omitting the time-varying latent variable x_{it} from the model specification as was done in equations (4.1) and (4.2).

4.3.3.7 Standard latent class model

When we eliminate both w_i and the transition structure, we obtain a latent class model that assumes observations are independent across occasions. This is a realistic model only for the analysis of data from repeated cross-sections; that is, to deal with the situation in which observations from different occasions are independent because each subject provides information for only one time point.

4.4 Other extensions

The previous section presented a general mixture model for longitudinal analysis, which contained the three basic models and various of their extensions as special cases. This section describes several other interesting extensions, which could be fit into an even more general mixture model for the longitudinal analysis.

4.4.1 Ordered states

The first extension concerns a latent Markov model for dichotomous or ordered polytomous responses in which the latent states can be interpreted as ordered categories. Examples are developmental stages of children, disease stages of patients, and states representing degrees of agreement in attitude measurement. It may of course turn out that the estimation of an unrestricted latent Markov model yields the hypothesized ordering of the latent states. However, it is also possible to force the latent states to be ordered by imposing constraints on the model parameters.

One class of restrictions concerns the relationship between latent states and responses. Bartolucci, Pennoni, and Francis (2007) and Vermunt and Georg (2002) presented various of such models, which can be seen as longitudinal data variants of the discretized item response theory models described by Heinen (1996) and Vermunt (2001). Two possible restrictions for multicategory items are

$$\log \frac{P(y_{itj} = m | x_{it} = k)}{P(y_{itj} = m-1 | x_{it} = k)} = \beta_{0jm} + \beta_{1j}\, \nu_k,$$

and

$$\log \frac{P(y_{itj} \geq m | x_{it} = k)}{P(y_{itj} < m | x_{it} = k)} = \beta_{0jm} + \beta_{1j}\, \nu_k,$$

where the former defines an adjacent category ordinal logit model for y_{itj} and the latter a cumulative logit model (Agresti, 2002). Note that ν_k represents the location of latent state k, which can either be fixed a priori or treated as a free parameter to be estimated. Vermunt and Hagenaars (2004) gave an extended overview of longitudinal models for ordinal responses, which also included various types of mixture models.

Another way to obtain latent states that can be interpreted as ordered categories is via restrictions on the transition probabilities. An example in which latent states represent (ordered) developmental stages was provided by Collins and Wugalter (1992). According to the underlying developmental psychology theory, children may make a transition to a next stage but will never return to a previous stage. In terms of the latent Markov model parameters, this means that $P(x_{it} = k_t | x_{it-1} = k_{t-1}) = 0$ for $k_t < k_{t-1}$.

4.4.2 Continuous latent variables

The mixture models discussed so far contained only discrete latent variables. However, in many applications, it may be useful to include also continuous latent variables in the model, which can play the role of latent factors in a measurement model or the role of random effects in a regression model. Below, I describe several situations in which time-constant or time-varying continuous latent variables may be used in the model for the transitions or in the model for the responses. I will denote continuous latent variables by F.

4.4.2.1 Time-constant affecting transitions

As a way to account for unobserved heterogeneity, it may be useful to expand the latent Markov model with random effects in the regression models for the initial state and transition probabilities. An example was provided by Pavlopoulos, Muffels, and Vermunt (2009) in an application of wage mobility. Their model contains two continuous latent variables, one affecting the initial state and the other the transitions.

Note that not only continuous random effects can be used to model unobserved heterogeneity, but also the mixture variable w_i can be used for this purpose. The choice between the two approaches depends on the assumptions one wishes to make about the nature of the unobserved heterogeneity; that is, whether it can be assumed to be continuous and normally distributed or whether a discrete specification – for example, using a mover-stayer structure – is more appropriate.

4.4.2.2 Time-constant affecting responses

Not only the transitions, but also the responses can be affected by time-constant continuous latent variables. In latent Markov models this would be a way to model dependencies between responses across occasions using an approach which is similar to the random-effects latent class models proposed in the biomedical field (Hadgu and Qu, 1998). Such a model is obtained by replacing $f(y_{itj}|x_{it}=k_t)$ with $f(y_{itj}|x_{it}=k_t,F_i)$ and defining a regression model for y_{itj} where F_i enters as one of the predictors.

In mixture growth modeling, it is very common to use a combination of discrete and continuous latent variables, where the continuous latent variables capture the unobserved heterogeneity within latent classes (Muthén, 2004; Vermunt, 2007). This involves replacing $f(y_{it}|w_i=\ell,\mathbf{z}_{it})$ by $f(y_{it}|w_i=\ell,\mathbf{z}_{it},\mathbf{F}_i)$ or, equivalently, by allowing $\beta_{0\ell}$ and $\beta_{p\ell}$ (see equation 4.3) to be random effects.

4.4.2.3 Time-varying affecting responses

Rather than using time-constant continuous latent variables, it is also possible to work with time-varying continuous latent variables. One possible application is in a latent Markov model for multiple responses which cannot be assumed to be locally independent within time points. The time-varying continuous latent variables would capture unobserved time-specific factors which vary across individuals and which are independent across occasions. Such a model can be obtained by replacing $f(y_{itj}|x_{it}=k_t)$ by $f(y_{itj}|x_{it}=k_t, F_{it})$ and defining a regression model for y_{itj} where F_{it} enters as a predictor.

Another, very different, type of use of time-varying continuous variables in latent Markov models is as common factors in a factor analytic model for the response variables. In other words, the continuous latent variables define a factor analytic measurement model for the responses. Changes in the factor mean(s) could be modeled using either a mixture growth or a latent Markov model, which defines two longitudinal data variants of the mixture factor analysis model proposed by Yung (1997).

For the situation that there is one common factor, the variant using a latent Markov structure to model the change in the factor means may have the following form:

$$f(\mathbf{y}_i) = \sum_{k_0=1}^{K}\sum_{k_1=1}^{K}\cdots\sum_{k_{T_i}=1}^{K} P(x_{i0}=k_0)\left[\prod_{t=1}^{T_i} P(x_{it}=k_t|x_{it-1}=k_{t-1})\right]$$
$$\left\{\prod_{t=0}^{T_i}\int\left[f(F_{it}|x_{it}=k_t)\prod_{j=1}^{J} f(y_{itj}|F_{it})\right]dF_{it}\right\},$$

where the last part shows that the distribution of the latent factor F_{it} depends on x_{it} and that F_{it} affects the responses. Regression models for F_{it} and y_{itj} complete the model specification.

4.4.3 Multiple, multilevel, and higher-order processes

This subsection presents extensions of latent Markov models for multiple, multilevel, and higher-order processes. These have in common that they require including an additional time-constant or time-varying discrete latent variable in the model. I will use a number as a subscript to denote the latent variable number (e.g., x_{it}^1 and x_{it}^2), and an asterisk to refer to a latent variable at a higher level of a nested structure (e.g., x_{it}^*).

4.4.3.1 Parallel processes

The latent Markov models described so far assume that there is a single Markov process of interest, which is possibly affected by time-constant and time-varying predictors. Suppose one has a categorical time-varying predictor which cannot be assumed to be measured without error. As suggested by Vermunt, Langeheine, and Böckenholt (1999) as a possible extension of their model, a latent Markov structure could also be defined for such a time-varying predictor. This yields a latent Markov model with two latent variables x_{it}^1 and x_{it}^2, where x_{it}^1 is related to the first set of J^1 response variables and x_{it}^2 to the other set of J^2 responses. Assuming that there are no (other) covariates, such a model has the following form:

$$f(\mathbf{y}_i) = \sum_{k_0^1=1}^{K^1}\sum_{k_1^1=1}^{K^1}\cdots\sum_{k_{T_i}^1=1}^{K^1}\sum_{k_0^2=1}^{K^2}\sum_{k_1^2=1}^{K^2}\cdots\sum_{k_{T_i}^2=1}^{K^2} P(x_{i0}^1 = k_0^1, x_{i0}^2 = k_0^2)$$
$$\left[\prod_{t=1}^{T_i} P(x_{it}^1 = k_t^1, x_{it}^2 = k_t^2 | x_{it-1}^1 = k_{t-1}^1, x_{it-1}^2 = k_{t-1}^2)\right]$$
$$\left\{\prod_{t=0}^{T_i}\left[\prod_{j=1}^{J^1} f(y_{itj}|x_{it}^1 = k_t^1)\right]\left[\prod_{j=J^1+1}^{J^1+J^2} f(y_{itj}|x_{it}^2 = k_t^2)\right]\right\}.$$

Additional attention is required with respect to the joint probability of x_{it}^1 and x_{it}^2 given x_{it-1}^1 and x_{it-1}^2, which may be decomposed in a specific way and/or modeled using a logistic regression equation. A meaningful specification is, for example, a model in which x_{it}^1 and x_{it}^2 are both affected by x_{it-1}^1 and x_{it-1}^2 but are not associated with one another, yielding what is sometimes referred to as a cross-lagged panel model. This involve decomposing the joint transition probability of x_{it}^1 and x_{it}^2 by

$$P(x_{it}^1 = k_t^1 | x_{it-1}^1 = k_{t-1}^1, x_{it-1}^2 = k_{t-1}^2)\, P(x_{it}^2 = k_t^2 | x_{it-1}^1 = k_{t-1}^1, x_{it-1}^2 = k_{t-1}^2).$$

Another possibility is that the causal effect goes in one direction; that is, x_{it}^2 affects x_{it}^1 but x_{it}^1 is not affected by x_{it}^2 or x_{it-1}^2. This can be specified as follows:

$$P(x_{it}^1 = k_t^1 | x_{it-1}^1 = k_{t-1}^1, x_{it}^2 = k_t^2) P(x_{it}^2 = k_t^2 | x_{it-1}^2 = k_{t-1}^2). \qquad (4.12)$$

A specification for correlated processes that are not causally related is obtained by allowing x_{it1} and x_{it2} to be associated and omitting the cross-lagged direct effects from the logistic model for x_{it1} and x_{it2}.

4.4.3.2 State-trait models

Eid and Langeheine (1999) proposed a discrete latent variable variant of the state-trait model. This model is obtained by expanding the latent Markov model with a J time-constant discrete latent variables, each of which affects one of the J responses.

The time-varying latent variable (representing the state) is assumed to be independent of the J time-constant latent variables (representing the traits). A state-trait model can be defined as follows:

$$f(\mathbf{y}_i) = \sum_{\ell^1=1}^{L^1} \sum_{\ell^2=1}^{L^2} \cdots \sum_{\ell^J=1}^{L^J} \sum_{k_0=1}^{K} \sum_{k_1=1}^{K} \cdots \sum_{k_{T_i}=1}^{K} P(w_i^1 = \ell^1, w_i^2 = \ell^2, ..., w_i^J = \ell^J)$$
$$P(x_{i0} = k_0) \left[\prod_{t=1}^{T_i} P(x_{it} = k_t | x_{it-1} = k_{t-1}) \right]$$
$$\left[\prod_{t=0}^{T_i} \prod_{j=1}^{J} f(y_{itj} | x_{it} = k_t, w_i^j = \ell^j) \right].$$

A more restricted variant of this model is obtained by assuming that the states are independent across occasions: $P(x_{it} = k_t | x_{it-1} = k_{t-1}) = P(x_{it} = k_t)$.

Eid and Langeheine (1999) worked with categorical y_{itj} variables for which they defined logistic models. These contained main effects of the state at time point t and the trait for response j but no interaction term; that is,

$$\log \frac{P(y_{itj} = m | x_{it} = k, w_i^j = \ell)}{P(y_{itj} = M | x_{it} = k, w_i^j = \ell)} = \beta_{0jm} + \beta_{1jkm} + \beta_{2j\ell m}.$$

4.4.3.3 Second-order model

As indicated earlier, one of the key assumptions of the latent Markov model is that the latent state transitions can be described with a first-order Markov structure. This assumption can be relaxed, for example, by allowing x_{it} to be affected not only by x_{it-1}, but also by x_{it-2}, which involves replacing $P(x_{it} = k_t | x_{it-1} = k_{t-1})$ by $P(x_{it} = k_t | x_{it-1} = k_{t-1}, x_{it-2} = k_{t-2})$ for $t \geq 2$. Though most software for latent Markov modeling does not allow defining such a second-order process, it can be defined with a trick which involves using a second time-varying latent variable x_{it}^2. The cross-lagged effect of x_{it}^1 (the variable of interest) on x_{it}^2 is restricted in such a way that $P(x_{it}^2 = k_t | x_{it-1}^1 = k_{t-1}) = 0$ for $k_t \neq k_{t-1}$, which implies that the lag one of the second latent variable (x_{it-1}^2) is in fact the lag two of the first latent variable (x_{it-2}^1). The second-order latent Markov model can now be obtained by allowing the transition probability for x_{it}^1 to depend on the lag of the second latent variable, which yields a model of the form

$$f(\mathbf{y}_i) = \sum_{k_0^1=1}^{K^1} \sum_{k_1^1=1}^{K^1} \cdots \sum_{k_{T_i}^1=1}^{K^1} \sum_{k_0^2=1}^{K^2} \sum_{k_1^2=1}^{K^2} \cdots \sum_{k_{T_i}^2=1}^{K^2} P(x_{i0}^1 = k_0^1) P(x_{i0}^2 = k_0^2)$$
$$P(x_{i1}^1 = k_1^1 | x_{i0}^1 = k_0^1) \left[\prod_{t=2}^{T_i} P(x_{it}^1 = k_t^1 | x_{it-1}^1 = k_{t-1}^1, x_{it-1}^2 = k_{t-1}^2) \right]$$
$$\left[\prod_{t=1}^{T_i} P(x_{it}^2 = k_t^2 | x_{it-1}^1 = k_{t-1}^1) \right]$$
$$\left[\prod_{t=0}^{T_i} \prod_{j=1}^{J} f(y_{itj} | x_{it}^1 = k_t^1) \right].$$

4.4.3.4 Processes for nested time units

Another interesting extension of the simple latent Markov model was recently presented by Rijmen et. al (2008). In their application there were two nested time units: the higher-level concerned changes occurring between days and the lower-level changes occurring between (non-sleeping) hours within days. The proposed model consists of two nested latent Markov models, one for between-day transitions and one for within-day transitions. A slight expansion of our notation is needed to write down the relevant model formulae. Let h, i, and t be the indices for a person, a day, and an hour, respectively. For the rest, notation is kept as much as possible as above, with the exception that quantities referring to the higher-level process get an asterisk as a superscript. The higher-level (between-day) model for person h can now be defined as

$$f(\mathbf{y}_h) = \sum_{k_0^*=1}^{K^*} \sum_{k_1^*=1}^{K^*} \cdots \sum_{k_{T_h^*}^*=1}^{K^*} P(x_{h0}^* = k_0^*) \left[\prod_{i=1}^{T_h^*} P(x_{hi}^* = k_i^* | x_{hi-1}^* = k_{i-1}^*) \right]$$
$$\left[\prod_{i=0}^{T_h^*} f(\mathbf{y}_{hi} | x_{hi}^* = k_i^*) \right],$$

which has the structure of a standard latent Markov model. The lower-level (within-day) model describing the hourly changes specifies a latent Markov model for $f(\mathbf{y}_{hi} | x_{hi}^* = k_i^*)$,

$$f(\mathbf{y}_{hi}|x^*_{hi}=k^*_i) = \sum_{k_0=1}^{K}\sum_{k_1=1}^{K}\dots\sum_{k_{T_i}=1}^{K} P(x_{hi0}=k_0|x^*_{hi}=k^*_i)$$
$$\left[\prod_{t=1}^{T_{hi}} P(x_{hit}=k_t|x_{hit-1}=k_{t-1},x^*_{hi}=k^*_i)\right]$$
$$\left[\prod_{t=0}^{T_{hi}}\prod_{j=1}^{J} f(y_{hitj}|x_{hit}=k_t,x^*_{hi}=k^*_i)\right].$$

Note that this is, in fact, a mixture latent Markov model in which the initial-state and transition probabilities and possibly also the response densities depend on the higher-level latent state occupied by person h at day i (x^*_{hi}).

4.4.3.5 Multilevel data

Vermunt (2003, 2004) proposed multilevel extensions of various types of mixture models that may also be useful in longitudinal data analysis. That is, when the observations for which we have longitudinal data are nested within higher-level units. Examples are longitudinal data on children which are nested within school, repeated measures data on patients nested within hospitals, and panel data from respondents nested within regions.

Palardy and Vermunt (in press) presented a multilevel mixture growth model for such data sets and illustrated the model with an application in which higher-level units (schools) are clustered based on the learning rates of children. Vermunt (2004) presented an application using a similar, but slightly simpler, multilevel mixture growth model. Denoting a higher-level unit by h, the higher-level part of this model is

$$f(\mathbf{y}_h|\mathbf{z}_h) = \sum_{\ell^*=1}^{L^*} P(w^*_h=\ell^*)\prod_{h=0}^{I^*_h} f(\mathbf{y}_{hi}|w^*_h=\ell^*,\mathbf{z}_{hi}),$$

where I_h is the number of persons belonging to higher-level unit or group h. The lower-level part is

$$f(\mathbf{y}_{hi}|w^*_h=\ell^*,\mathbf{z}_{hi}) = \sum_{\ell=1}^{L} P(w_{hi}=\ell)\prod_{t=0}^{T_{hi}} f(y_{hit}|w_{hi}=\ell,w^*_h=\ell^*,\mathbf{z}_{hit}).$$

As in the mixture growth model described in equations (4.1) and (4.3), the regression model for y_{hit} specifies how the higher- and lower-level latent classes differ in term of the growth parameters.

Yu and Vermunt (in progress) developed a multilevel extension of the latent Markov model. The structure of this model is similar to that of a mixture latent Markov model, with the important difference that the mixture is at the group level and thus not at the individual level. The model can be formulated as follows:

$$f(\mathbf{y}_i) = \sum_{\ell^*=1}^{L^*} P(w_h^* = \ell^*) \prod_{i=1}^{I_h^*} f(\mathbf{y}_{hi}|w_h^* = \ell^*).$$

The lower-level part defines the structure for $f(\mathbf{y}_{hi}|w_h^* = \ell^*)$ which is the same as the lower-level part of the multilevel process model described above, except for that the conditioning is on $w_h^* = \ell^*$ instead of $x_{hi}^* = k_i^*$; that is,

$$f(\mathbf{y}_{hi}|w_h^* = \ell^*) = \sum_{k_0=1}^{K} \sum_{k_1=1}^{K} \dots \sum_{k_{T_i}=1}^{K} P(x_{hi0} = k_0|w_h^* = \ell^*)$$
$$\left[\prod_{t=1}^{T_{hi}} P(x_{hit} = k_t|x_{hit-1} = k_{t-1}, w_h^* = \ell^*)\right]$$
$$\left[\prod_{t=0}^{T_{hi}} \prod_{j=1}^{J} f(y_{hitj}|x_{hit} = k_t, w_h^* = \ell^*)\right].$$

4.4.3.6 Dependent classification errors

One of the assumptions of the latent Markov model is that responses are independent across time points conditional on the latent states, an assumption that may be unrealistic in certain applications. However, it is sometimes possible to relax this assumption, which is sometimes referred to as ICE (independent classification errors).

Above, we already discussed a non-ICE model; that is, a latent Markov model with a time-constant continuous latent variable affecting the responses at the different time points. In this model, it is assumed that an unobserved individual factor is causing correlations between measurement errors. This is a good non-ICE model when these correlations are (almost) equally strong between each pair of occasions.

However, typically, correlations between errors are much stronger between adjacent time points. Possible mechanisms leading to such correlated errors are that making an error at one occasion increases the likelihood of making an error at the next occasion (Manzoni et al., in progress), or that experiencing a transition increases the likelihood of making an error (see also Hagenaars, 1988). Bassi et al. (2000) proposed a non-ICE latent Markov model for employment status measurements obtained using a very specific retrospective data collection design (see also Hagenaars' chapter in this volume).

Here, I would like to discuss the non-ICE specification proposed by Manzoni et al. (in progress). Their application concerned a latent Markov model with two measures of a person's monthly employment status (employed, self employed, unemployed, and not employed) for a period of about a year. The first measure is a retrospective report of the last year and the second is a retrospective report on the same period collected ten years later. The aim of the analysis was to determine the quality latter report. Because respondents are likely to misplace or forget unemployment spells when these occurred a long time ago, it is clearly incorrect to assume that errors in the second measure are uncorrelated across occasions. Manzoni et al. proposed a correlated measurement error model which involves replacing the response probability $P(y_{it2} = m_{t2}|x_{it} = k_t)$ by $P(y_{it2} = m_{t2}|x_{it} = k_t, y_{it-1,2} = m_{t-1,2}, x_{it-1} = k_{t-1})$; that is, a model in which $y_{it,2}$ is not only affected by x_{it}, but also by $y_{it-1,2}$ and x_{it-1}. Moreover, restrictions were imposed on the way the lagged observed and latent states affect the measurement error. One restriction yielded a specification in which respondents making an error at $t-1$ have a different (higher) probability of making an error at t. So, in fact, two sets of error probabilities were estimated, one for respondents reporting correctly at $t-1$ ($m_{t-1,2} = k_{t-1}$) and another for respondents reporting incorrectly ($m_{t-1,2} \neq k_{t-1}$). Various alternative specifications were also investigated.

4.5 Applications

This section presents two applications of the mixture models for longitudinal data described in this chapter. The first application concerns a repeated measures experimental study and is used to illustrate the mixture growth model, including the more advanced model with continuous random effects. The second application concerns a longitudinal survey and is used to illustrate the latent Markov and mixture latent Markov model, as well as the latent Markov model for parallel processes. For parameter estimation, I used version 4.5 of the Latent GOLD program (Vermunt and Magidson, 2005, 2008). Examples of syntax files can be found in the Appendix.

4.5.1 A mixture growth model

The empirical example I will use to illustrate mixture growth modeling 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.

Table 4.1 Test results for the mixture growth models estimated with the schizophrenia data

Model	Log-likelihood	BIC	# Parameters
A1: 1-class growth	-704	1421	2
A2: 2-class growth	-625	1286	6
A3: 3-class growth	-608	1277	10
A4: 4-class growth	-601	1287	14
B2: 2-class growth with squared time for class 2	-620	1282	7
B3: 3-class growth with squared time for class 3	-597	1261	11
C2: B2 with random intercept	-601	1250	8
C3: B3 with random intercept	-595	1263	12

Whereas Vermunt (2007) used the same data set for a more extended comparison of various types of growth models, here the focus will be on the mixture growth models described in this chapter. More specifically, it will be shown that two growth classes can be identified – one class with decreasing severity and one class without – and that patients receiving the treatment are much more likely to be belong to the decreasing severity class than the control group. Moreover, it will be shown that using random effects may yield a simpler solution with a smaller number of latent classes.

In the analysis of this data set, I followed Hedeker and Gibbon's (1996) suggestion to set $P = 1$, with $z_{it1} = \sqrt{t}$, 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 latent classes nor random effects, this model fits the time-specific response probabilities much better than a linear or a quadratic model, and almost as well as a model with an unrestricted time dependence.

Table 4.1 reports the log-likelihood value, the number of parameters, and the BIC value obtained by applying various of the models described in the previous two sections to the schizophrenia data set. Models A1-A4 are 1 to 4-class mixture growth models using the $\sqrt{t}$ time dependence and containing treatment as a covariate affecting the class membership. Based on the BIC value, one would select the 3-class model as the best one. Models B2 and B3 modify models A2 and A3 in the sense that one latent class (the last one) has a different (quadratic) time dependence. This is specified by defining $z_{it2} = t$ and $z_{it3} = t^2$, and setting the parameters corresponding to these two terms to 0 in all but class K and the parameter corresponding to z_{it1} to 0 in class K. As can be seen from the BIC values, Models B2 and B3 fit better than Models A2 and A3, which indicates that it makes sense to assume another type of time dependence for one of the classes. It can also be seen that the 3-class model is still preferred to the 2-class model. Models C2 and C3 are variants of Models B2 and B3 containing a random intercept to allow for within class heterogeneity. As can be seen, these models have lower BIC values than Models B2 and B3. More-

Table 4.2 Parameter estimates obtained with Model C2

Model for Responses	Class 1			Class 2		
	β or λ	s.e.	z-value	β or λ	s.e.	z-value
Intercept	9.16	1.22	7.49	6.95	0.92	7.56
StdDev Random Intercept	3.50	0.55	6.42	3.50	0.55	6.42
TIME				- 3.82	1.01	-3.77
SQ-TIME				1.13	0.30	3.77
SQRT-TIME	-4.98	0.65	-7.66			
Model for Latent Classes	Class 1					
	γ	s.e.	z-value			
Intercept	-0.64	0.31	-2.08			
Treatment	1.80	0.36	4.99			

over, under this specification, the simpler 2-class model (C2) performs better than the 3-class model (C3).

Table 4.2 reports the parameter estimates obtained with Model C2. 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. Figure 4.5.1 depicts the estimated growth curves for the two latent classes, which are obtained by marginalizing over (integration out) the continuous random effects. Class 1 contains the patients for which the probability of severe symptoms of schizophrenia decreases during the study. It can now 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, 66% is estimated to belong to latent class 1 and 34% to latent class 2. These numbers are 76% and 24% for the treatment group and 35% and 65% 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 4.2 – the odds of begin in class 1 instead of 2 is $\exp(1.80)$ 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.

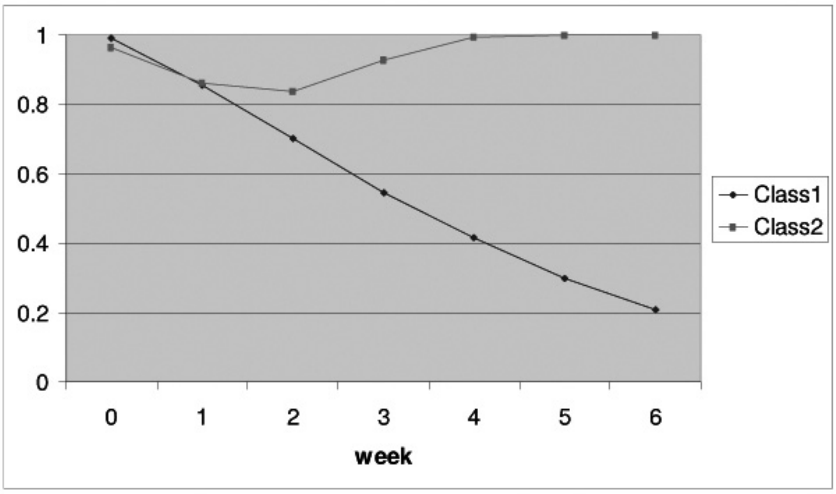

Fig. 4.1 Class-specific trajectories obtained with Model C2.

4.5.2 A mixture Latent Markov model

The latent Markov models described above will be illustrated with the nine-wave National Youth Survey (Elliott, Huizinga, and Menard, 1989) for which data were collected annually from 1976 to 1980 and at three year intervals after 1980. At the first measurement occasion, the ages of the 1725 children varied between 11 and 17. To account for the unequal spacing across panel waves and to use age as the time scale, we define a model for 23 time points ($T+1=23$), where $t=0$ corresponds to age 11 and the last time point to age 33. For each subject, we have observed data for at most 9 time points (the average is 7.93) which means that the other time points are treated as missing values.

We study the change in a dichotomous response variable "drugs" indicating whether young persons used hard drugs during the past year (1=no; 2=yes). It should be noted that among the 11 years of age nobody in the sample reported to have used hard drugs, which is something that will be taken into account in our model specification. Time-varying predictors are age and age squared, and time-constant predictors are gender and ethnicity. In the second step of the analysis, I will introduce alcohol use during the past year as a time-varying covariate containing measurement error.

A preliminary analysis showed that there is a clear age-dependence in the reported hard-drugs use which can well be described by a quadratic function: usage first increases with age and subsequently decreases. That is why we used this type of time dependence in all reported models. Age and age-squared are used as time-

Table 4.3 Test results for the Markov models estimated with the drugs use data

Model	Log-likelihood	BIC	# Parameters
A1. Markov	-4143	8330	6
A2. Latent Markov with K=2	-4009	8078	8
A3. Mover-stayer latent Markov with K=2	-4000	8068	9
A4. Mixture latent Markov with L=2 and K=2	-3992	8066	11
A5. A4 with Gender & Ethnicity effects on W_i	-3975	8061	15
B1. A5 with Markov model for Alcohol	-9328	18789	18
B2. B1 with Alcohol affecting X_{it}	-9185	18520	20
B3. B2 with Alcohol measured with error	-8912	17989	22

dependent covariates in the regression model for the latent transition probabilities (see also equation 4.8); that is,

$$\log \frac{P(x_{it}=k'|x_{it-1}=k, w_i=\ell, \text{age}_{it})}{P(x_{it}=k|x_{it-1}=k, w_i=\ell, \text{age}_{it})} = \alpha_{0kk'} + \alpha_{\ell kk'} + \alpha_{L+1,kk'} \cdot \text{age}_{it} + \alpha_{L+2,kk'} \cdot (\text{age}_{it})^2,$$

where the α coefficients are fixed to 0 for $k'=k$ and for $\ell=1$. For the initial-state, we do not have a model with free parameters but we simply assume that all children start in the no-drugs state at age 11.

Table 4.3 reports the fit measures for the estimated models, where Models A1 to A4 do not contain covariates gender and ethnicity. Among these models, the most general model – the mixture latent Markov model – performs best. By removing measurement error, simplifying the mixture into a mover-stayer structure, or eliminating the mixture structure, the fit deteriorates significantly. Model A5 is a mixture latent markov model in which we introduced covariates in the model for the mixture proportions: sex and/or ethnicity seem to be significantly related to the mixture component someone belongs to.

As a final step, we investigated whether alcohol use affects hard drugs use. We specified three additional models: Model B1 in which alcohol does not affect drugs use, Model B2 in which alcohol use at age t affects the transitions in the model for drugs, and Model B3 in which alcohol use is treated as a time-varying covariate measured with error. The latter model is a latent Markov model for two parallel processes. We used a specification in which alcohol use affects the drugs-use transitions but in which the reversed effect is absent (see equation 4.12). In Models B1 and B2, we specified a Markov model without measurement error for alcohol use in order to be able to compare the BIC values across these three models. Note that as far as the modeling of drugs use is concerned, Model B1 is, in fact, equivalent to Model A5, but their log-likelihood values cannot be compared because alcohol is introduced as an additional response variable in Model B1. Comparison of the fits measures for Models B1 and B2 shows that alcohol use has a significant effect on the drugs use transitions, and comparison of the fit measures for Models B2 and B3 shows that there is evidence that alcohol use is measured with error.

One set of parameters of the final model (B3) are the probabilities of the measurement models for drugs and alcohol. These show that the latent states are rather strongly connected to the two observed states: $P(y_{it1} = 1|x^1_{it} = 1) = 0.99$ and $P(y_{it1} = 2|x^1_{it} = 2) = 0.83$ for drugs use; $P(y_{it2} = 1|x^2_{it} = 1) = 0.87$ and $P(y_{it1} = 2|x^2_{it} = 2) = 0.99$ for alcohol use.

The most relevant coefficients in the model for the drugs use transitions are the effects of alcohol (x^2_{it}) and of w_i. The former show that being in the latent alcohol use state increases the probability of moving into the drugs use state ($\alpha = 4.61; S.E = 1.33$) and decreases the probability of exciting the drugs use state ($\alpha = -1.86; S.E = 0.52$). The parameters for w_i show that class 1 is the low-risk class having a lower probability than class 2 of entering into the use state ($\alpha = -1.19; S.E = 0.36$) and a much higher probability of leaving the non-use state ($\alpha = 4.16; S.E. = 0.63$). This means that class 1 contains young people that quit the drug-use state quickly when they get into this state.

The parameters in the logistic regression model for w_i shows that males are less likely to be in the low-risk class than females ($\gamma = -0.67; S.E. = 0.20$). Moreover, blacks are more likely ($\gamma = 0.41; S.E = 0.26$), hispanics less likely ($\gamma = -0.75; S.E. = 0.52$), and other ethnic groups less likely ($\gamma = -0.09; S.E = 0.70$) to be in the low-risk class than whites, but these ethnicity effects are non significant.

Appendix: Examples of Latent GOLD syntax files

The Latent GOLD 4.5 software package (Vermunt and Magidson, 2008) implements the mixture models described in this article. In this appendix, I provide examples of syntax files used for the empirical applications.

The data should be in the format of a person-period file, where for Markov type models it is important to include also periods with missing values in the file since each next record for the same subject is assumed to be the next time point. The definition of a model contains three main sections: "`options`", "`variables`" and "`equations`".

The mixture growth models A1 to A4 from Table 4.1 can be defined as follows:

```
options
   output parameters standarderrors estimatedvalues;
 variables
   caseid id;
   dependent severity binomial;
   independent sqrttime, treatment;
   latent W nominal 2;
 equations
   W        <- 1 + treatment;
   severity <- 1 | W + sqrttime | W;
```

In the above `options` section, only the commands related to the output options are shown. It is indicated that we wish to output parameters and standard errors of the parameters, as well as the estimates for the model probabilities.

In the `variables` section we define the `caseid` variable connecting the multiple records of a person, the `latent`, `dependent`, and `independent` variables to be used in the analysis, as well as various attributes of these variables, such as their scale types and, for categorical latent variables, also their number of categories. Note that the model above is a two-class mixture model since we specified "`latent W nominal 2;`".

The `equation` section contains 2 equations: one for the mixture variable (`W`) and another for the response variable. The logit model for `W` contains an intercept (the term "`1`") and the effect of treatment. The model for the response variable `severity` contains an intercept and an effect of square root time. Both parameters are assumed to vary across latent classes, which is achieved by the conditioning "`| W`".

The more complex final two-class model C2 – containing a continuous random effect and a different time dependence for classes 1 and 2 – is defined as follows:

```
options
  output parameters standarderrors estimatedvalues;
variables
  caseid id;
  dependent severity binomial;
  independent sqrttime, time, sqtime, treatment;
  latent W nominal 2, F continuous;
equations
  W        <- 1 + treatment;
  severity <- 1 | W + (b1) sqrttime | W + (b2) time | W
              + (b3) sqtime | W + F;
  b1[2]=0; b2[1]=0; b3[1]=0;
```

As can be seen, the model contains two additional predictors (`time` and `sqtime`) and a continuous latent variable (`F`). These are all used as predictors in the regression model for the response variable. It can also be seen that three of the regression coefficient get labels, which is needed to be able to define the three constraints at the bottom. These restrictions indicate that `sqrttime` has no effect in class 2, and that `time` and `sqtime` have no effect in class 1.

The syntax for Markov models is somewhat more complicated than for growth models. As an example, this is the setup for model A5 appearing in Table 4.3, a mixture latent Markov model with two covariates affecting the mixture distribution and with a quadratic time dependence of the transition logits:

```
options
  missing includeall;
  output parameters=first standarderrors estimatedvalues;
variables
  caseid id;
  dependent drugs nominal;
  independent gender nominal, ethnicity nominal, age, age2;
```

```
    latent W nominal 2, X nominal dynamic 2;
  equations
    W     <- 1 + gender + ethnicity;
    X[=0] <- (-100) 1;
    X     <- (~tra) 1 | X[-1] + (a~tra) W | X[-1]
             + (~tra) age | X[-1] + (~tra) age2 | X[-1];
    drugs <- (b~err) 1 | X;
```

Compared to the specification above, the `options` section contains the statement "`missing=includeall`" indicating that records with missing values should be retained in the analysis and the output option "`parameters=first`" requesting dummy coding with the first category as the reference category for nominal variables. A new element in the `variables` section is the keyword "`dynamic`" which indicates that the nominal latent variable `X` may change its value over time (in this case, it is a two-state time-varying latent variable).

The `equations` section contains 4 equations: one for the mixture variable (`W`), one for the initial state (`X[=0]`), one for the state at time point t (`X`) conditional on the state at $t-1$ (`X[-1]`), and one for the response variable at time point t (`drugs`). The logit model for `W` contains an intercept as well as effects of gender and ethnicity. The model for `X[=0]` contains an intercept that is fixed to -100, which indicates that everyone starts in latent state 1. The model for `X` is parameterized in such a way that the intercept and the effects of `W`, `age`, and `age2` can be interpreted as effects on the logit of a transition (as in the equation provided in the text). This is achieved by the conditioning "`| X[-1]`" combined with "˜tra" in the parameter label, which yields a coding for the logit coefficients in which the no transition category serves as the reference category. The model for the response variable `drugs` contains an intercept which varies across latent states, with the same type of coding as used for the transition (for the dependent variable called error coding). Note that removing "˜tra" and "˜err" does not change the model but only the identifying constraints that are imposed in the parameter set concerned. As can be seen, two parameter sets get labels (`a` and `b`), which will be used below to define models with parameter restrictions.

The 2-class mixture can be changed into a mover-stayer structure with the additional line "`a = -100;`" which fixes the transition probabilities to 0 for the second class. A latent Markov model is obtained either by removing `W` from the `variables` and `equations` sections or by setting its number of categories to 1. A standard Markov is obtained with the extra line "`b = -100;`". This fixes the logit parameters in the model for the response variable to -100, which because of the special error coding (induced with "˜err") yields a perfect relationship between `X` and `drugs`.

The model in which alcohol is used as a time-varying covariate measured with error (Model B3 of Table Table 4.3 is obtained by including `alcohol` as a second dependent variable and defining a second dynamic latent variable `X2`. The `equations` section of this more advanced model contains also equations for the initial state and the transitions of `X2`, includes `X2` in the equation for `X`, and defines a measurement equation for `alcohol`; that is,

```
equations
  W       <- 1 + gender + ethnicity;
  X[=0]   <- (-100) 1;
  X2[=0]  <- 1;
  X       <- (~tra) 1 | X[-1] + (~tra) W | X[-1]
             + (~tra) age | X[-1] + (~tra) age2 | X[-1]
             + (~tra) X2 | X[-1];
  X2      <- (~tra) 1 | X2[-1];
  drugs   <- 1 | X;
  alcohol <- 1 | X2;
```

References

Agresti, A. (2002). *Categorical data analysis*. New York: Wiley.
Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics, 55,* 218-234.
Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics, 41,* 164-171.
Bartolucci, F., Pennoni, F., & Francis, B. (2007). A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society, Ser. A, 170,* 115-132.
Bassi, F., Hagenaars, J. A., Croon, M. A., & Vermunt, J. K. (2000). Estimating true changes when categorical panel data are affected by uncorrelated and correlated errors. *Sociological Methods and Research, 29,* 230-268.
Bergma, W., Croon, M. A., & Hagenaars, J. A. (2009). *Marginal models for dependent, clustered and longitudinal categorical data*. Dordrecht, NL: Springer.
Böckenholt, U. (2005). A latent Markov model for the analysis of longitudinal data collected in continuous time: States, durations, and transitions. *Psychological Methods, 10,* 65-82.
Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27,* 131-157.
Dayton C.M., & Macready, G. B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association, 83,* 173-178.
Dias, J. G., & Vermunt, J. K. (2007). Latent class modeling of website users' search patterns: Implications for online market segmentation. *Journal of Retailing and Consumer Services, 14,* 359-368.
Dias, J. G., Vermunt, J. K., & Ramos, S. (2009). Mixture hidden Markov models in finance research. In A. Fink, L. Berthold, W. Seidel, & A. Ultsch (Eds.), *Advances in data analysis, data handling and business intelligence*. Berlin-Heidelberg: Springer
Diggle, P. J., Liang, K. Y., & Zeger (1994), S. L. *Analysis of longitudinal data*. Oxford: Clarendon Press.
Eid, M., & Langeheine, R. (1999). Measuring consistency and occasion specificity with latent class models: A new model and its application to the measurement of affect. *Psychological Methods, 4,* 100-116.
Elliott, D. S., Huizinga, D., & Menard, S. (1989). *Multiple problem youth: Delinquency, substance use, and mental health problems*. New York: Springer-Verlag.
Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.
Fay, R. E. (1986). Causal models for patterns of nonresponse. *Journal of the American Statistical Association, 81,* 354-365.
Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models.* New York, NY: Springer.

Goldfeld, S., & Quandt, R. (1973). A Markov model for switching regressions. *Journal of Econometrics, 1,* 3-16.
Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association, 56,* 841-868.
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.
Hagenaars, J. A. (1988). Latent structure models with direct effects between indicators: Local dependence models. *Sociological Methods and Research, 16,* 379-405.
Hagenaars, J. A. (1990). *Categorical longitudinal data: Loglinear analysis of panel, trend and cohort data.* Newbury Park: Sage.
Hadgu, A., & Qu, Y. (1998). A biomedical application of latent class models with random effects. *Applied Statistics, 47,* 603-616.
Hedeker, D., & Gibbons, R. D. (1996). MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine, 49,* 157-176.
Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oakes, CA: Sage Publications.
Kaplan, D. (2005). Finite mixture dynamic regression modeling of panel data with implications for dynamic response analysis. *Journal of Educational and Behavioral Statistics*, 30, 169-187.
Langeheine, R., & van de Pol, F. (2002). Latent Markov chains. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis.* (pp. 304-341). Cambridge, UK: Cambridge University Press.
Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.
Little, R.J., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.
Magidson, J., Vermunt, J. K., & Tran, B. (2009). Using a mixture latent Markov model to analyze longitudinal U.S. employment data involving measurement error. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New Trends in Psychometrics* (pp. 235-242). Tokyo: Universal Academy Press.
Manzoni, A., Vermunt, J.K., Luijkx, R., & Muffels, R. (in preparation). Memory bias in retrospectively collected employment careers: A model based approach to correct for measurement error.
McDonald, I. L., & Zucchini, W. (1997), *Hidden Markov and other models for discrete valued time series*. London: Chapman and Hall.
McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
Mooijaart, A., & van Montfort, K. (2007). Latent Markov models for catagorical variables and time-dependent covariates. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences.* (pp. 1-18). Mahwah, NJ: Lawrence Erlbaum.
Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345-368). Thousand Oakes, CA: Sage.
Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric group-based approach. *Psychological Methods, 4,* 139-157.
Paas, L. J., Vermunt, J. K., & Bijmolt, T. H, (2007). Discrete-time discrete-state latent Markov modelling for assessing and predicting household acquisitions of financial products. *Journal of the Royal Statistical Society, Ser. A (Statistics in Society), 170,* 955-974.
Palardy, G., & Vermunt, J. K. (in press). Multilevel growth mixture models for classifying group-level observations. *Journal of Educational and Behavioral Statistics.*
Pavlopoulos, D., Muffels, R., & Vermunt, J. K. (2009). Training and low-pay mobility: The case of the UK, the Netherlands and Germany. *Labour, 21,* 37-59.
Poulsen, C. S. (1990). Mixed Markov and latent Markov modelling applied to brand choice behaviour. *International Journal of Research in Marketing, 7,* 519.
Rijmen, F., Vansteelandt, K., & de Boeck, P. (2008). Latent class models for diary method data: Parameter estimation by local computations. *Psychometrika, 73,* 167-182.
Schafer, J. L. (1997). *Statistical analysis with incomplete data*. London: Chapman & Hall.

Schmittmann, V. D., Dolan, C.V., van der Maas, H. L. J., & Neale, M. C. (2005). Discrete latent Markov models for normally distributed response data. *Multivariate Behavioral Research, 40,* 461-488.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. London: Chapman & Hall/CRC.

van de Pol, F., & de Leeuw, J. (1986). A latent Markov model to correct for measurement error. *Sociological Methods and Research, 15,* 118-141.

van de Pol, F., & Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, 213-247.

van der Heijden, P. G. M., Dessens, J., & Böckenholt, U. (1996). Estimating the concomitant variable latent class model with the EM algorithm. *Journal of Educational and Behavioral Statistics, 21,* 215-229.

Vermunt, J. K. (1997a). *Log-linear models for event histories.* Thousand Oakes, CA: Sage.

Vermunt, J. K. (1997b). *LEM: A general program for the analysis of categorical data: User's manual*. Tilburg, The Netherlands: Tilburg University.

Vermunt, J. K. (2001). The use of latent class models for defining and testing non-parametric and parametric item response theory models. *Applied Psychological Measurement, 25,* 283-294.

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. (2007). Growth models for categorical response variables: Standard, latent-class, and hybrid approaches. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 139-158). Mahwah, NJ: Lawrence Erlbaum.

Vermunt, J. K., & Georg, W. (2002). Longitudinal data analysis using log-linear path models with latent variables. *Metodología de las Ciencias del Comportamiento, 4,* 37-53.

Vermunt, J. K., & Hagenaars, J. A. (2004). Ordinal longitudinal data analysis. In R. C. Hauspie, N. Cameron, & L. Molinari (Eds.), *Methods in human growth research* (pp. 374-393). Cambridge, UK: Cambridge University Press.

Vermunt, J. K., Langeheine, R., & 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., & Magidson, J. (2005) *Latent GOLD 4.0 user's guide*. Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2008). *LG-syntax user's guide: Manual for Latent GOLD 4.5 syntax module*. Belmont, MA: Statistical Innovations.

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 373-385). Burlington, MA: Elsevier.

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

Wiggins, L. M. (1973). *Panel analysis.* Amsterdam: Elsevier.

Wong, C. S, & Li, W.K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society, Ser. B, 62,* 95-115.

Yu, H. T., & Vermunt, J. K. (in preparation). Multilevel latent Markov model for nested longitudinal discrete data.

Yung, Y. F. (1997). Finite mixtures in confirmatory factor analysis models. *Psychometrika, 62,* 297-330.