# Latent Markov and growth mixture models for ordinal individual responses with covariates: a comparison

Pennoni, Fulvia and Romeo, Isabella

Dipartimento di Statistica e Metodi Quantitativi, Istituto di Ricerche Farmacologiche Mario Negri

July 2016

# LATENT MARKOV AND GROWTH MIXTURE MODELS FOR ORDINAL INDIVIDUAL RESPONSES WITH COVARIATES: A COMPARISON

Fulvia Pennoni[1] and Isabella Romeo[2]

[1] Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi Milano-Bicocca,
(e-mail: `fulvia.pennoni@unimib.it`)

[2] Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche Mario Negri, Milano.
(e-mail: `isabella.romeo@marionegri.it`)

**ABSTRACT**: We propose a short review between two alternative ways of modeling stability and change of longitudinal data when time-fixed and time-varying covariates referred to the observed individuals are available. They both build on the foundation of the finite mixture models and are commonly applied in many fields. They look at the data by a different perspective and in the literature they have not been compared when the ordinal nature of the response variable is of interest. The latent Markov model is based on time-varying latent variables to explain the observable behavior of the individuals. The model is proposed in a semi-parametric formulation as the latent Markov process has a discrete distribution and it is characterized by a Markov structure. The growth mixture model is based on a latent categorical variable that accounts for the unobserved heterogeneity in the observed trajectories and on a mixture of normally distributed random variable to account for the variability of growth rates. To illustrate the main differences among them we refer to a real data example on the self reported health status.

## 1 Introduction

The analysis of longitudinal or panel data by using latent variable models has a long and rich history mainly in the social sciences. In the past several decades, the increased availability of large and complex data sets, have witnessed a sharp increase in interest in this topic. Nowadays, it demands the development of increasingly rigorous statistical analytic methods which can be proved use-

ful for data reduction as well as for inference. Among different proposals there are two main broad classes of models: one tailored to consider the transition over time and the other focused on growth or trajectory analysis. Both models build on the foundation of finite mixture modeling [1]. Among the former, we recall the latent Markov (LM) model which is mainly used for the analysis of categorical data. Among the second, the growth mixture model (GMM) is originally employed with observed continuous response variables. We propose a comparison between them which accounts for the more recent model improvements since, up to our knowledge, it is not available in the literature except for the reviews made by [2], [3] and some hints provided in the book of [4]. The innovativeness of the proposed comparison is that it is made for measurements on an ordinal scale with the aim to illustrate similarities and differences. We recall the main features of the LM model and the GMM unifying the notation and the terminology for the common parameters and retaining those of each specific literature.

The LM models may be classified as observation-driven models tailored for many types of longitudinal categorical data as showed recently by [5] and [6]. The evolution of individual characteristics of interest over time is represented by a latent process with state occupation probabilities which are time-varying. They are extensions both of the latent class model [7], when there are multiple occasions of measurements and of the Markov chain models for stochastic processes in which an error term is included to the observations. They allow for the unobserved heterogeneity between individuals or within the latent states. Even if the first basic model formulation due to book of [8]* did not include the covariates, at the present days time-constant and time-varying covariates can be added in the measurement or in the latent part of the model. When the model is formulated according to a discrete time dependent latent process it may be classified as a semi-parametric approach. It allows to model with different aims data arising in the applications in fields such as medicine, sociology or biology [see also 9, 10]. Some of its connections with the hidden Markov model[†] which is employed to analyze time series data are illustrated by [12].

---

*At Columbia he introduces the model in a research project in the social sciences with Paul Lazarsfeld (http://www.nasonline.org/publications/biographical-memoirs/memoir-pdfs/lazarsfeld-paul-f.pdf) as principal investigator where he wrote in 1955 his Ph.D. dissertation. The application was devoted to study a single item of human behaviour moving over time in a non-experimental context.

[†]The hidden Markov model was developed also in the social science field to study sudden changes in learning processes by [11].

Conventional growth model or growth curve model (GCM) are viewed either as hierarchical linear models or as structural equation models. Their use in analyzing continuous response variables has been widely discussed in the literature [see, among others, 13, 14]. Their use in modeling and analyzing categorical data has recently received more attention [15, 16].‡

The GCM aims at studying the evolution of a latent individual characteristic in order to estimate the trajectories by accounting for the individual variability about a mean population trend. It imposes an homogeneity assumption, requiring that all individuals follow similar trajectories. The growth mixture model (GMM) proposed by [22] [see also 23, 24] is a generalization of the GCM accounting for the heterogeneity in the observed development trajectories by employing a latent categorical variable. The finite mixture of linear and multinomial regression models allows to disentangle the between-individual differences and the within-individual pattern of changes through time [see also 25, 26]. It is a parametric approach where the population variability in growth is modeled by a mixture of different normally distributed subpopulations.

A specific case of the GMM is the latent-class growth curve model (LCGM) [see, among others, 27, 28, 29], also named as latent class regression model by [30]. Another terminology employed by [31] is latent class growth analysis (LCGA). The multinomial model is used to identify the homogeneous groups of developmental trajectories avoiding the normal distribution assumption on the random effects. The individuals in each class share a common trajectory [32] whithout considering the between-class heterogeneity. Therefore within the LCGM the individual heterogeneity is completely captured by the mean growth trajectories of the latent classes. Whereas GMM allows to model class-specific variance components (intercept and slope variance). For a more complete comparison between GMM and LCGM, see also [33].

We illustrate two recent extensions of the LM model and GMM when the ordinal response variable is made by thresholds imposed on an underlying continuous latent response variable. We show how the discrete support for the latent variable used in the LM model framework can be appropriate in this context. The models are compared on how they allow covariates, how they make inference, on their computational features required to achieve the estimates, and on their ability to classify units and their predictive power. Our proposal to compare them in terms of fitting, parsimony, interpretation and prediction is an attempt to joint the recent literature on these models for panel data. The

---

‡Latent growth modeling was first proposed independently by [17] and [18] in relation to the longitudinal factor analysis and later extended and refined by [19, 20] and [21].

results of the model fitting are illustrating through a dataset on longitudinal study aimed at describing the self perceived health status which appears also in others published scientific articles [see, among others, 34].

The structure of the paper is as follows. In Section 2 we introduce the basic notation for both models and we summarize the main features concerning the estimation issues. In Section 3 we demonstrate the effectiveness of the models explaining their purposes in relation to the applicative example and their results. In the last section we draw some concluding remarks.

## 2 Main notation and illustration of the models

One way to afford the issue of ordinal response variables consists in deriving a conditional probability model from a linear model for a latent response variable. The observed variables are obtained by categorizing the latent continuous response which may be related for example to the amount of understanding, attitude, or wellbeing required to respond in a certain category. Let $Y_{it}$ be the observed ordinal variable for individual $i$, for $i = 1, \ldots, n$ at time $t$, $t = 1, \ldots, T$. We assume an underlying continuous latent variable $Y_{it}^*$, via a threshold model given by

$$Y_{it} = s \quad \text{iff} \quad \tau_{s-1} < Y_{it}^* \leqslant \tau_s$$

where $s = 1, 2, \ldots, S$ and $-\infty = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_{s-1} < \tau_s = +\infty$ are the cut-points by which it is possible to achieve a unique correspondence. With $S$ response categories, there are $S - 1$ threshold parameters, $\tau_s$, $s = 1, 2, \ldots, S - 1$.

### 2.1 Latent Markov models for ordinal data

Under the basic model we assume the existence of a discrete latent process such that

$$Y_{it}^* = \alpha_{it} + \varepsilon_{it},$$

where $\alpha_{i1}, \ldots, \alpha_{iT}$ follows an hidden Markov chain with state space $\xi_1, \ldots, \xi_k$, initial probabilities $\pi_u = p(\alpha_{i1} = \xi_u)$ and transition probabilities $\pi_{u|\bar{u}} = p(\alpha_{it} = \xi_u | \alpha_{i,t-1} = \xi_{\bar{u}})$, $\bar{u}, u = 1, \ldots, k$. Moreover, $\varepsilon_{it}$ is a random error with normal or logistic distribution.

In the case of time-varying or time-fixed covariates collected in the column vectors $\boldsymbol{x}_{it}$, the model is extended as

$$Y_{it}^* = \alpha_{it} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it},$$

including these covariates in the measurement model concerning the conditional distribution of the response variables given the latent process. The covariates are allowed also in the latent part of the model, however the model is better identified when they are stored or in the latent or in the measurement part of the model and the choice is related to the research question and the aims of the analysis.

We assume that the discrete latent process follows a first-order homogeneous Markov chain. The model has a simple structure when the conditional independence of an observed response variable $Y_{it}$ on the other responses given the latent process holds for $i = 1, \ldots, n$, $t = 1, \ldots, T$. This assumption is called the local independence. The conditional distribution of the responses is denoted by $f_t(\mathbf{y}|u, \mathbf{x})$, $u = 1, \ldots, k$, whereas the latent process $U$ is discrete with initial probability function $p(u)$, for $u = 1, \ldots, k$, and transition probability function $p_t(u|\bar{u})$, where $t = 2, \ldots, T$, $u, \bar{u} = 1, \ldots, k$, and $k$ denotes the discrete number of latent states. Therefore a semi-parametric model results. A generalized linear model parameterization [35] allow us to include properly the covariates in the measurement model. In such a way, by using suitable link functions we may allow for specific constraints of interest and we may also reduce the number of parameters.

An effective way to include the covariates in the measurement model is by employing a generalized linear model parameterization

$$\boldsymbol{\eta}_{tux} = \boldsymbol{C} \log[\boldsymbol{M}\boldsymbol{f}_t(u, \boldsymbol{x})],$$

where $\boldsymbol{C}$ is a suitable matrix of contrasts, $\boldsymbol{M}$ is a marginalization matrix with elements 0 and 1, which sums the probabilities of the appropriate cells and the operator log is coordinate wise, $\boldsymbol{f}_t(u, \boldsymbol{x})$ is a $c$-dimensional column vector with elements $f_t(\mathbf{y}|u, \boldsymbol{x})$ for all possible values of $y$ and $\eta_{ty|ux}$, $y = 1, \ldots, s - 1$ denotes each element of $\boldsymbol{\eta}_{tux}$. Within this formulation constraints on the model parameters specific of the application may be conceived.

An interesting formulation for the current context is

$$\eta_{y|ux} = \beta_{1y} + \beta_{2u} + \boldsymbol{x}'\boldsymbol{\beta}_3, \quad y = 1, \ldots, s - 1, u = 1, \ldots, k, \tag{1}$$

where the levels of the $\beta_{1y}$ are cut-points or threshold parameters, $\beta_{2u}$ are intercepts specific of the corresponding latent state, and $\boldsymbol{\beta}_3$ is a vector of parameters for the covariates. The above is possible once we define the global logits [35] on the conditional response mass function

$$\eta_{y|ux} = \log \frac{f(y|u, \boldsymbol{x}) + \cdots + f(s - 1|u, \boldsymbol{x})}{f(0|u, \boldsymbol{x}) + \cdots + f(y - 1|u, \boldsymbol{x})}, \quad y = 1, \ldots, s - 1.$$

We carried out the estimation of the model parameters in two ways: or by using maximum likelihood method through the EM algorithm [36] or by Bayesian methods applying the MCMC algorithm [37]. Within the first choice the log-likelihood is maximized according to the following steps until convergence:

E: step compute the expected value of the complete data log-likelihood given the observed data and the current value of $\boldsymbol{\theta}$, which denotes all the model parameters;

M: step maximize this expected value with respect to $\boldsymbol{\theta}$ and thus update $\boldsymbol{\theta}$.

We use the recursions developed in the hidden Markov literature by [38] and by [39] to compute the quantities of interests. They enable to efficiently compute the expected values of the random variables involved in the complete data log-likelihood

$$
\begin{aligned}
\ell^*(\boldsymbol{\theta}) \;=\; & \sum_t \sum_u \sum_x \sum_y a_{tuxy} \log f_t(\boldsymbol{y}|u,\boldsymbol{x}) + \\
& + \sum_u b_{1ux} \log p(u) + \sum_{t>2} \sum_{\boldsymbol{u}} \sum_u b_{tu\bar{u}} \log p_t(u|\bar{u})
\end{aligned}
$$

where $a_{tuxy}$ is the number of individuals that are in latent state $u$ and provide response $y$, $b_{1u}$ is a frequency of the latent state $u$ at occasion $t$; $b_{tu\bar{u}}$ is the number of transitions from state $\bar{u}$ to state $u$.

As for other mixture models [1] there may be many local optima, therefore the estimation is carried out by considering multiple sets of starting values for the chosen algorithm. A drawback of the EM algorithm is that it can't give a direct quantity to asses the precision of the maximum likelihood estimates. Among other methods, it is possible to consider the missing information principle. In the case of the regular exponential family [40] the observed information is equal to the complete information minus the missing information due to the unobserved components [41, 42]. For an implementation when a directed acyclic Gaussian graphical model is considered with one hidden variable see [43]. The method is suitable for its low computational burden over that required by the maximum likelihood estimation.

The model selection may be based on a Likelihood Ratio (LR) test between the model with $k$ classes and that with $k+1$ classes for increasing values of $k$, until the test is not rejected. However, we need to employ the bootstrap to obtain a $p$-value for *LR* test. It is based on a suitable number of samples simulated from the estimated model with $k$ classes [44]. In [45] the best parsimonious model is the result of a statistic based on the parametric bootstrap which is

consistent. Their results relay on the assumption that the best model is one among those with the proposed number of classes.

We select the number of latent states according to the most common information criteria: the Akaike Information Criterion (AIC) [46] and the Bayesian Information Criterion (BIC) [47]. We recall that when we select the states according to the model with the smallest value of BIC, we decrease the maximum of the log-likelihood value considering also the number individuals. Their performance has been deeply studied in the literature on mixture models [see, among others, 1, Chapter 6]. They are also employed in the hidden Markov literature for time-series, where they are penalized by the number of time occasions [see, among others, 48]. The BIC is usually preferred to AIC, as the latter tends to overestimate the number of latent states but it may be too strict in certain cases [see, among others, 49]. The theoretical properties of BIC in the LM models framework are still not well established. However, BIC is a commonly accepted model choice criterion even for these models as well as to choose the number of latent classes for the latent class model [see, among others, 50]. In [6] this criterion is also used together with other diagnostic statistics measuring the goodness-of-classification. In a more recent study [51] compare the performance of some likelihood and classification based criteria, such as an entropy measure, for selecting the number of latent states when a multivariate LM model is fitted to the data.

An interesting feature of the LM model concerns prediction. As showed in [6] the local decoding allows prediction of the latent state for each individual at each time occasion by maximizing the estimated posterior function of the latent process. The global decoding employing the [52] algorithm, [see also 53] allows to get the most a posteriori likely predicted sequence of states for each individual. The joint conditional probability of the latent states given the responses and the covariates $\hat{f}_{U|X,Y}(\boldsymbol{u}|\boldsymbol{x},\boldsymbol{y})$ is computed by using a forward recursion according to the maximum likelihood estimates of the model parameters, where $\boldsymbol{u}$ denotes a configuration of the latent states. The optimal state

$$\hat{u}_t^* = \arg\max_u \hat{r}_t(u)\hat{p}_{t+1}(u|\hat{u}_{(t-1)}^*).$$

is found by considering $\hat{r}_1(u) = \hat{p}(u|\boldsymbol{x})\prod_t \hat{f}_1(y_1|u_1,x)$, for $u = 1,\ldots,k$; and computing in a similar way $\hat{r}_t(v)$, for $t = 2,\ldots,T$ and $v = 2,\ldots,k$; then maximizing such that $\hat{u}_T^* = \arg\max_u \hat{r}_T(u)$.

## 2.2 Growth mixture models

The GCMs provide the estimated shapes of the individual trajectories accounting for within and between individual differences. The measurement model concerning the observed responses deals with individual growth factors. The latent model is related to the means, to the variances and covariances of the growth factors to explain between individuals differences. First we recall the LGCM and then the GMM. The LGCM without covariates is defined by the following equations

$$
\begin{aligned}
Y_{it}^* &= \alpha_i + \lambda_t \beta_i + \lambda_t^2 q_i + \varepsilon_{it}, \\
\alpha_i &= \mu_\alpha + \zeta_{\alpha i}, \\
\beta_i &= \mu_\beta + \zeta_{\beta i},
\end{aligned}
\tag{2}
$$

where $\alpha_i$ and $\beta_i$ are the intercept and the slope of the growth factor for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, respectively. To allow identifiability the intercept growth factor is fixed to 1. Therefore, it equally influences all the repeated measures across the waves and it remains constant across time for each individual. Different values can be assigned to the time indicator $\lambda_t$, to produce growth curves of different shapes that are linearly or not linearly dependent on time. In order to define a linear growth model with equidistant time points the time scores for the slope growth factor are fixed at $0, 1, 2, \ldots, T-1$, and the quadratic growth factor $q_i$ is fixed at 0 [see, among others, 14]. The first time score is fixed at zero and the intercept growth factor can be interpreted as the expected response at the first time point. The time scores for the quadratic growth factor are fixed at $0, 1, 4, \ldots, (T-1)^2$ to allow for a quadratic shape of the trajectory.

The measurement errors $\varepsilon_{it}$ in Equation (2) are not correlated across time, they are i.i.d. disturbances. Because there is no intercept term in the measurement model, the mean structure of the repeated measures is determined entirely by the means of the latent trajectory factors. In the structural model, the parameters $\mu_\alpha$ and $\mu_\beta$ are the population means of the intercept and of the slope, respectively; $\zeta_{\alpha i}$ is the deviation of $\alpha_i$ from the population mean intercept and $\zeta_{\beta i}$ is the deviation of $\beta_i$ from the population mean slope. The are assumed with a multivariate normal distribution with zero means and variances denoted by $\psi_{\alpha\alpha}$ and $\psi_{\beta\beta}$, respectively and they are uncorrelated with $\varepsilon_{it}$. The covariance of the intercept and of the slope is denoted by $\psi_{\alpha\beta}$. When the response is ordinal or categorical the thresholds are assumed to be equal for each measurement occasion by imposing the constraint $\tau_{st} = \tau_s$ for $t = 1, \ldots, T$ and also the constraint $\mu_\alpha = 0$ is required.

In the conditional growth model the time-fixed covariates are included as predictors of the growth factors or as direct predictors of the response variable. Time-varying covariates can only be included as predictors in the measurement model according to the following equations

$$
\begin{aligned}
Y_{it}^* &= \alpha_i + \lambda_t \beta_i + \lambda_t^2 q_i + \boldsymbol{\omega}_{it} \boldsymbol{\gamma}_t + \varepsilon_{it}, \\
\alpha_i &= \mu_\alpha + \boldsymbol{x}_i' \boldsymbol{\gamma}_\alpha + \zeta_{\alpha i}, \\
\beta_i &= \mu_\beta + \boldsymbol{x}_i' \boldsymbol{\gamma}_\beta + \zeta_{\beta i},
\end{aligned} \tag{3}
$$

for $i = 1, \ldots, T$ and $t = 1, \ldots, T$, where $\boldsymbol{\gamma}_\alpha$ and $\boldsymbol{\gamma}_\beta$ are vectors of parameters for the time-fixed covariates $\boldsymbol{x}_i$ on $\alpha_i$ and $\beta_i$, respectively, and $\boldsymbol{\gamma}_t$ is the vector of parameters for the time-varying covariates $\boldsymbol{\omega}_{it}$ on the measurement model.

The unconditional GMM is defined by a latent categorical variable $U$ accounting for the unobserved heterogeneity in the development among individuals. It represents a mixture of subpopulations whose membership is inferred by the data [for a review, see, among others, 14, 54]. It is characterized by the following equations

$$
\begin{aligned}
Y_t^* &= \sum_{u=1}^k p_u (\alpha_u + \lambda_{tu} \beta_u + \varepsilon_{tu}), \\
\alpha_u &= \mu_{\alpha u} + \boldsymbol{x}' \boldsymbol{\gamma}_{\alpha u} + \zeta_{\alpha u}, \\
\beta_u &= \mu_{\beta u} + \boldsymbol{x}' \boldsymbol{\gamma}_{\beta u} + \zeta_{\beta u},
\end{aligned}
$$

for $i = 1, \ldots, n$ and $t = 1, \ldots T$, where $p_u$ is the probability of belonging to latent class $u$, for $u = 1, \ldots, k$ which defines the latent trajectory, with the constraints $p_u \geqslant 0$ and $\sum_{u=1}^k p_u = 1$, where $k$ is equal to the number of mixture components. The thresholds $\tau_s$ are unknown and they are estimated and constrained to be equal across time and latent classes. The intercepts of the growth factors may vary across latent classes. With categorical response variables the growth factor referred to the last class is constrained to zero for identifiability issues and the others are estimated from the model. The variances and covariance of the growth factors can be allowed to be class-specific or constrained to be equal. Residuals of the growth factors and of the measurement model are assumed normally distributed within each latent class. As in Equation (3) only time-fixed covariates may be included to infer the latent class through a multinomial logistic regression model since the latent variable is typically viewed as time invariant. Therefore, the GMM reduces to the GCM when $k = 1$ and to the LCGM when the within-class growth factor variance and covariances $\psi_{\alpha u}, \psi_{\beta u}, \psi_{\alpha \beta u}$ are set to zero for all $u$, $u = 1, \ldots, k$. In the latter case, the between-individual variability is captured only by the latent class

membership. The thresholds are estimated with the mean cumulative response probabilities for a specific response category at each measurement occasion by the estimated distribution of the latent growth factors.

The maximum likelihood estimation of the model parameters when there are categorical response variables and continuous latent variables requires numerical methods. The computation is carried out by using Monte Carlo integration [14, 55]. As in the standard normal mixture models, imposing constraints on the covariance matrices of the latent classes ensures the absence of singularities and potentially reduces the number of local solutions [22, 26]. The model selection concerns the choice of the number of the latent classes and the order of the polynomial of the group's trajectories. The most common applied empirical procedure is the following: first the order of the polynomial is assessed by estimating both linear and not linear unconditional GCM, or GMM with $k = 1$, GMM(k) in the following. Then, the number of groups is determined according to the unconditional model in order to avoid an over-extraction of the latent classes [see also 56]. Finally, the covariates are added in the model as predictors of the latent classes.

The LR test is employed for the model selection also by considering the bootstrap [see, among others, 57] as illustrated in the previous section. The number of latent classes is selected according to the AIC or BIC illustrated in Section 2.1. The relative entropy measure [58] is commonly employed to state the goodness of classification

$$E_k = 1 - \frac{\sum_{i=1}^{n} \sum_{u=1}^{k} -\hat{p}_{iu} \log(\hat{p}_{iu})}{n \log(k)}, \tag{4}$$

where $\hat{p}_{iu}$ is the estimated posterior probability of belonging to the $u$-th latent class, $k$ is the number of latent classes and $n$ is the sample size. The values approach one when the latent classes are well separated. However, we notice that it differs from the normalized entropy criterion (NEC) defined by [59] which instead divides the entropy index $EN(\hat{p}_{iu})$ by the difference between the log-likelihood of the model with $k$ classes and the one with just one class. The above criteria may lead to a model lacking of interpretability in terms of latent classes or in which only few individuals are allocated in a class. As suggested by many authors such a choice needs also to be guided by the research question as well as by theoretical justification and interpretability [60, 61, 62]. The optimal number of classes derived from the LCGM is always bigger than the optimal number of classes derived from GMM. Within the LCGM, individuals with slightly different growth parameters are sooner defined to a different class compared with the GMM [see, among others, 63].

# 3 Real data example: the health and retirement study

In order to show the main differences among the models illustrated in the previous section we consider a longitudinal study aimed at describing the self-perceived health status. The latter is a frequently used way to establish health policy and care as the repeated subjective health assessment reflects the own perception of health and how it is going to evolve over time. It is recorded by one item with response categories defined so that an ordinal variable results. The data is taken from version I of the RAND HRS data, collected by the University of Michigan[§]. The 30,406 respondents, were asked to express opinions on their health status at $T = 8$ approximately equally spaced occasions, from 1992 to 2006. After considering only individuals with no missing data on the variables, we ended up with a sample of $n = 7,074$ individuals. The response variable SRHS is measured on a scale based on five ordered categories: "poor", "fair", "good", "very good" and "excellent". For every individual some covariates are also available: gender, race, education and age (at each time occasion). The study relies on the investigation of the population heterogeneity in the health status perception as well as on prediction of features needs especially tailored for those elders which are identified to share the most difficult health conditions.

First of all, we summarize the estimation process for both models presented in Section 2 and then we make some comparisons on the estimated quantities. The estimation of the LM models is undertaken on the R [64] package LMest V2.2 that is available on Comprehensive R Archive Repository (CRAN). This version accounts also for covariates on the latent part of the model and it handles also missing values on the responses. The estimation of the growth models is undertaken under the commercial software MPLUS V7.2.

For the LM model parameterized as in Equation (1) we applied the model search procedure as illustrated in Section 2.1 to find the best model among those with a number of latent states from 1 to 11. The search strategy which is implemented to account for the multimodality of the likelihood function is based on estimating the same model with the same number of states by using deterministic and random starting values for the EM algorithm. The number of different random starting values sis proportional to the number of latent states. The relative log-likelihood difference is evaluated by considering a tolerance level equal to 1e-10. The model is estimated for an increasing number of latent

---

[§]See also `http://www.cpc.unc.edu/projects/rlms-hse` and `http://www.hse.ru/org/hse/rlms`

states while checking for the replication of likelihood values. The best model is that with 9 latent states. It is selected according to the BIC values showed in Table 1 that also show the number of free parameters.

**Table 1.** *Fit statistics for an increasing number of latent states from 1 to 11 of the LM model with covariates.*

|        | log-likelihood | AIC       | BIC       | #*par* |
|--------|----------------|-----------|-----------|--------|
| LM(1)  | -80,623.52     | 161,267.0 | 161,335.7 | 10     |
| LM(2)  | -69,789.21     | 139,604.4 | 139,693.6 | 13     |
| LM(3)  | -65,707.82     | 131,451.6 | 131,575.2 | 18     |
| LM(4)  | -63,968.06     | 127,986.1 | 128,157.7 | 25     |
| LM(5)  | -63,293.98     | 126,656.0 | 126,889.3 | 34     |
| LM(6)  | -63,062.23     | 126,214.5 | 126,523.4 | 45     |
| LM(7)  | -62,894.29     | 125,904.6 | 126,302.7 | 58     |
| LM(8)  | -62,739.12     | 125,624.2 | 126,125.3 | 73     |
| LM(9)  | -62,645.69     | 125,471.4 | 126,089.1 | 90     |
| LM(10) | -62,615.99     | 125,450.0 | 126,198.2 | 109    |
| LM(11) | -62,650.58     | 125,561.2 | 126,453.5 | 130    |

**Table 2.** *Estimated support points and parameters referred to the initial probabilities of the chain of the LM(9) model.*

| Latent state | Support points | Initial probabilities |
|--------------|----------------|-----------------------|
| 1            | -8.657         | 0.047                 |
| 2            | -4.941         | 0.117                 |
| 3            | -2.456         | 0.192                 |
| 4            | -1.147         | 0.028                 |
| 5            | -0.224         | 0.213                 |
| 6            | 2.062          | 0.189                 |
| 7            | 4.303          | 0.121                 |
| 8            | 5.159          | 0.213                 |
| 9            | 7.357          | 0.067                 |

**Table 3.** *Estimates of the vector of the regression parameters of the LM(9) model.*

| Coefficient | *Female* | *Non-white* | *Some college* | *College and above* | *Age* | *$Age^2$* |
|---|---|---|---|---|---|---|
| β | -0.185 | -1.341 | 1.37 | 2.461 | -0.125 | -0.001 |
| s.e. | 0.075 | 0.109 | 0.092 | 0.104 | 0.007 | 0.026 |

**Table 4.** *Estimates of the transition probabilities under the LM(9) model, (probabilities out of the diagonal greater than 0.1 are in bold).*

| | | | | $\hat{\pi}_{u|\bar{u}}$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 0.796 | **0.182** | 0.000 | 0.001 | 0.006 | 0.001 | 0.002 | 0.012 | 0.000 |
| 2 | 0.053 | 0.822 | **0.106** | 0.001 | 0.000 | 0.000 | 0.000 | 0.017 | 0.000 |
| 3 | 0.008 | 0.020 | 0.868 | 0.004 | 0.061 | 0.001 | 0.000 | 0.038 | 0.000 |
| 4 | 0.026 | 0.013 | 0.001 | 0.336 | 0.006 | 0.039 | **0.155** | **0.292** | **0.132** |
| 5 | 0.002 | 0.024 | 0.015 | 0.000 | 0.887 | 0.066 | 0.005 | 0.000 | 0.000 |
| 6 | 0.000 | 0.004 | 0.024 | 0.003 | 0.024 | 0.896 | 0.044 | 0.001 | 0.003 |
| 7 | 0.001 | 0.004 | 0.001 | 0.052 | 0.025 | 0.009 | 0.845 | 0.001 | 0.062 |
| 8 | 0.018 | 0.061 | **0.189** | **0.301** | **0.153** | 0.000 | 0.000 | 0.278 | 0.000 |
| 9 | 0.000 | 0.000 | 0.000 | 0.050 | 0.006 | 0.051 | 0.072 | 0.000 | 0.821 |

The estimated cut-points of the LM(9) are $\hat{\tau}_1 = 8.261$ $\hat{\tau}_2 = 4.559$, $\hat{\tau}_3 = 0.800$, $\hat{\tau}_4 = -3.470$. The estimated initial probabilities are reported in Table 2 together with the support points. The categories of response variables have been reversed from excellent to poor and the estimated support points are arranged in increasing order to interpret the resulting latent states from the best (latent state 1) to the worst (latent state 9). We notice that the 11% and 19% of individuals are in second and in the third latent state which are better states with respect to latent states 6 and 8. The matrix of the estimated transition probabilities is reported in Table 4. We notice that the only greater probabilities then 0.10 in the elements adjacent to the diagonal are those of the transition from the first to the second latent state and from the second to the third. For the latent states 4 the probability to moving to latent state 7 to 9 are higher then 0.10. For the latent state 8 the probability of moving to latent state 3 to 5 are
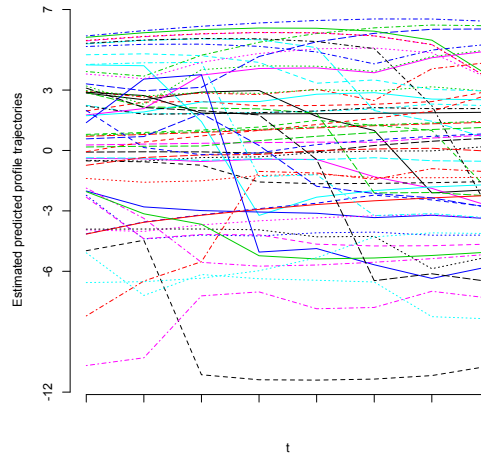
**Figure 1.** *Individual profiles for a selected group of individuals for the LM(9) model.*

higher than 0.10, therefore we conclude that the individuals feeling not well at the beginning of the survey have high probability to improve their health conditions over time.

Table 3 shows the effect of the covariates on the probability of reporting a certain level of the health status. In particular, women tend to report worse health status than men (the odds ratio for females versus males is equal to $(\exp(-0.185) = 0.831)$, whereas white individuals have higher probability of reporting a good health status with respect to non-white (the odds ratio for non-white versus white is equal to $(\exp(-1.341) = 0.261)$. We also observe that better educated individuals tend to have a better opinion about their health status. Finally, the effect of age is decreasing over time and its trend is linear as the quadratic term of age is not significant. In Figure 1 we compare the individual response profiles of the LM(9) model obtained by using the estimated posterior probabilities according to the rules illustrated in Section 2.1. They refer only to white females over 65 years old at the third wave of interview which are highly educated. Some profiles are less regular then others: they detect those females whose health status may strongly decline due to events that are not observed through the covariates.

For the growth models the best model within the class of GMMs is detected according the model strategy illustrated at the end of Section 2.2. As first step,

we estimate two GMMs without covariates with just one latent class in which the respondents's opinions about their health are specified as a function of linear and nonlinear growth patterns. The GMM with a quadratic effect shows a log-likelihood equal to -63,996.8 and a BIC value equal to 128,100 with 12 parameters. This model is to be preferred according to the BIC value as the GMM without the quadratic effect results in a log-likelihood equal to -63,116.3 and a BIC value equal to 128,303.5 with 8 parameters. The $\chi^2 = 1,761$ with 4 degrees of freedom is significant. However, we notice that the numerical integration becomes more computationally demanding as it reaches three dimension of integration and we could not reach the convergence when we increased the number of latent classes. Therefore, we choose to retain the linear form. As second step, the hypothesis of homogeneity within groups has to be rejected since the log-likelihood of the linear model under this assumption decreases to -83,152.7. However, we should note that the numerical integration becomes increasingly more computationally demanding as the number of latent factors increases. When we consider only intercept and slope in the model, we have two dimensions of integration and the computer burden is then moderate. When also the quadratic term is considered we reach three dimensions of integration and the computer burden increases exponentially. Therefore, we decide to use a linear GMM as a comparison with the LM model. The estimated model parameters denotes that the perception of a good health status decreases over time. The variances of the intercept and of the slope factor are significant, indicating the existence of individual differences in growth trajectories. As third step, we fit the selected GMM model without covariates by considering the existence of mixture distributions from to 2 to 5 with varying patterns of the growth trajectory. Table 5 shows the results for an increasing number of latent classes 1 up to 5. According to the BIC values we select the

**Table 5.** *Selection of the number of latent classes of the GMM without covariates.*

| Latent class | Log-likelihood | BIC | #par | Entropy |
|---|---|---|---|---|
| 1 | -64,116.3 | 128,303.5 | 8 | 1.000 |
| 2 | -64,092.3 | 128,282.2 | 11 | 0.599 |
| 3 | -63,982.3 | 128,088.7 | 14 | 0.719 |
| 4 | -63,982.2 | 128,115.1 | 17 | 0.428 |
| 5 | -63,977.2 | 128,131.7 | 20 | 0.746 |

model with three latent classes also because we realized that the models with a higher number of components did not reach the convergence criteria. The

model with 4 classes have the same log-likelihood of the one with 3 classes. While, the best log-likelihood value of the model with 5 classes is not replicated with different starting values as that of the model with 3 classes. As last step, only time-fixed covariates are included in the model of Equation (3), considered as constants across the latent classes. Their coefficients are significant with the exception of the quadratic effect of age, which is then excluded and the resulted model has a log-likelihood equal to -63,421.0 and a BIC value equal to $127,143.3$ with 34 parameters. The entropy value as in Equation (4) is equal to 0.763. The estimated probabilities of GMM(3) and the average conditional probability of belonging to each latent class are displayed in Table 6. This a common employed way to asses the tenability of the selected model as the average posterior probability of group membership for each trajectory is considered as an approximation of the reliability of the trajectory. The posterior probabilities are used to assign each individual membership to the trajectory that best matches, where the values of 0.70 or 0.80 are reference in the literature to group individuals in the same latent class with a similar pattern of change. Table 6 shows the classification probabilities for the selected GMM(3) by considering the most likely latent class membership (row) by the average conditional probabilities (column). We notice that contrary to our expectation the diagonal values referred to the first and third latent class are low meaning that the classes are not properly identified. In fact the percentage of units belonging to those classes is 10.8% and 3.2%, respectively. From Table 7 the estimated coefficients of the covariates on the growth factor are not high and differently from the LM model the sign of the female coefficient is reversed, therefore females tend to report better health status then man. Probably due to the poor convergence of the selected model. The high education shows the highest positive estimated coefficient on the intercept factor.

As shown in Table 8 the estimated covariance is negative meaning that the individuals with the highest values of the intercepts at the first occasion (e.g. with better perceived health) change more rapidly into a worse perception. Figure 2 illustrates the estimated trajectories where the first latent class identifies the individuals with an initial poor health status and a slow decline in their health, the second latent class those with a better initial health status and a slightly faster decline compared to the first class and the third latent class individuals perceiving a strong worsening of their health status over time.

**Table 6.** *Classification probabilities for the GMM(3) with covariates according to the most likely latent class membership (row) by the average conditional probabilities (column).*

|         | 1     | 2     | 3     |
|---------|-------|-------|-------|
| Class 1 | 0.436 | 0.556 | 0.008 |
| Class 2 | 0.022 | 0.973 | 0.005 |
| Class 3 | 0.028 | 0.436 | 0.537 |

**Table 7.** *Estimates of the vector of the regression parameters of the intercept and slope growth factor of the GMM(3) with covariates.*

| Coefficient | Female | Non-white | Some college | College and above | Age |
|-------------|--------|-----------|--------------|-------------------|-----|
| $\gamma_\alpha$ | 0.265 | -1.506 | 1.037 | 1.876 | -0.044 |
| s.e. | (0.103) | (0.170) | (0.136) | (0.148) | (0.009) |
| $\gamma_\beta$ | 0.005 | 0.032 | -0.040 | -0.071 | 0.000 |
| s.e. | (0.012) | (0.015) | (0.016) | (0.018) | (0.001) |

**Table 8.** *Estimates of the structural parameters of GMM(3) with covariates.*

| Coefficient | Estimates | s.e. | Coefficient | Estimates | s.e. |
|-------------|-----------|------|-------------|-----------|------|
| $\mu_{\alpha(1)}$ | -6.734 | 0.498 | $\mu_{\beta(1)}$ | -0.105 | 0.090 |
| $\mu_{\alpha(2)}$ | -2.302 | 0.443 | $\mu_{\beta(2)}$ | -0.193 | 0.069 |
| $\mu_{\alpha(3)}$ | 0.000 | 0.000 | $\mu_{\beta(3)}$ | -1.292 | 0.118 |
| $\psi_\alpha$ | 6.501 | 0.422 | $\psi_\beta$ | 0.065 | 0.005 |
| $\psi_{\alpha\beta}$ | -0.272 | 0.039 | | | |

## 4   Concluding remarks

We propose a comparison between the latent Markov (LM) models and the Growth Mixture Models (GMMs) when the interest lies in modelling longitudinal ordinal responses and time-fixed and time-varying individual covariates. The interest on this topic is relevant since in many different contexts the ordinal data are a way to account for the importance given by an item or measure something do not directly observable. The LM model is a data-driven model which
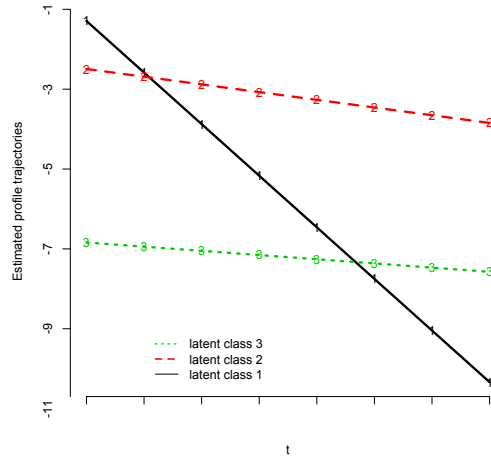
**Figure 2.** *Response profile plot for the GMM(3) with covariates.*

relays on a latent stochastic process following a first-order Markov chain with the fundamental principle to estimate transitions between latent states and to capture the influence of time-varying and time-fixed covariates on the observed transitions. GMM exploits a latent categorical variable to allows the unobserved heterogeneity in observed development trajectories. The latent variable is time invariant and it describes the trend through a polynomial function allowing for time-fixed covariates. We illustrate the main features of the models and their performance by referring to a specific application based on real data in which the ordinal response variable of interest describes the self-perceived health status. The aim is also to estimate a life expectancy for longevity. We can summarize the main differences between the LM and the GMM according to the following characteristics: *i*) the model estimation and selection procedure leading to the choice of the number of the latent states or classes; *ii*) the way they relate the conditional probabilities of the responses to the available individual covariates. *iii*) the model capability to use the posterior probabilities for creating profiles of group membership.

We show that the LM model outperforms the GMM mainly because it is more rigorous on each of the above points. With reference to *i*) the model choice is more complex for the GMM and it starts with the model without covariates. We found that the Monte Carlo integration for the GMM with a

number of latent classes up to three leads to improper solutions. The selection of the best model is more straight for the LM model, however it requires a search strategy to properly initialize the EM algorithm and therefore it is computational demanding when the number of latent states of the model is high. With reference to *ii*) the covariates are better handled by the LM model since they are allowed according to suitable parametrization for categorical data such as global logits. While in the LM model the covariates may affect the measurement part of the model or may influence the latent process, in the GMM they can affect both but in the measurement model only time-fixed covariates are allowed. Then, when the interest is on detecting subpopulations in which individuals may be arranged according to their perceived health status the LM is more appropriate. The GMM can be useful when just a mean trend is of interest and the expected subpopulations are not many. With reference to *iii*) the predictions of the LM model are based on local and global decoding. The first is based on the maximization of the estimated posterior probability of the latent process and the second on a well known algorithm developed in the hidden Markov model literature to get the most aposteriori likely predictive sequence. In the GMM the prediction is based on the maximum posterior probability and as showed in the example it may be not precise when the internal reliability of the model is poor.

We conclude that, due to the asymptotic properties of the algorithm used to estimate the posterior probabilities, the LM model should be recommended especially when the prediction of the latent states is one of the main interest in the data analysis. The GMM leads to select a lower number of subpopulations compared to the LM. However, this is not always a desirable property since when the data are rich, as in the applicative example, it may not be of interest to compress extremely their information. Within the LM model it is possible to detect also a reversible transition between the latent states. The consideration of the time dimension in the structural form made by the GMM is inadequate to explain the latter feature of the data. The results proposed by the applicative example may be useful when the interest is to evaluate the needs of group of elders to prevent the fast decrease of their health, or to investigate better the reasons why the health conditions are improved when the age is increasing and therefore plan specific interventions for the group of interest.

## Acknowledgements

## References

[1] G. J. McLachlan and D. Peel. *Finite mixture models*. Wiley, 2000.

[2] S. W. Raudenbush. Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual review of psychology*, 52:501–525, 2001.

[3] J K Vermunt. Longitudinal research using mixture models. In V. K. Montfort, J. Oud, and A. Satorra, editors, *Longitudinal research with latent variables*, pages 119–152. Springer, 2010.

[4] S. Menard. *Handbook of longitudinal research: Design, measurement, and analysis*. San Diego, 2008.

[5] F. Bartolucci, A. Farcomeni, and F. Pennoni. Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates (with discussion). *Test*, 23:433–486, 2014.

[6] F. Bartolucci, A. Farcomeni, and F. Pennoni. *Latent Markov models for longitudinal data*. Chapman and Hall/CRC press, Boca Raton, 2013.

[7] P. F. Lazarsfeld and N. W Henry. *Latent structure analysis*. Houghton Mifflin, Boston, 1968.

[8] L. M. Wiggins. *Panel Analysis: Latent probability models for attitude and behaviour processes*. Elsevier, Amsterdam, 1973.

[9] F. Pennoni and G. Vittadini. Two competing models for ordinal longitudinal data with time-varying latent effects: an application to evaluate hospital efficiency. *QdS, Journal of methodological and applied statistics*, 15:53–68, 2013.

[10] F. Pennoni and G. Vittadini. Hidden markov and mixture panel data models for ordinal variables derived from original continuous responses. In *Proceedings of the 3rd International Conference on Mathematical, Computational and Statistical Sciences*, pages 98–106, Dubai, 2015.

[11] G. A. Miller. Finite Markov processes in psychology. *Psychometrika*, 17:149–167, 1952.

[12] I. Visser and M. Speekenbrink. Comment on: Latent markov models: a review of a general framework for the analysis longitudinal data with. *Test*, 23:478–483, 2014.

[13] T. E. Duncan, S. C. Duncan, L. A. Strycker, F. Li, and A. Alpert. *An introduction to latent variable growth curve modeling: Concepts, issues, and application*. Lawrence Erlbaum Associates, London, 1999.

[14] K. A. Bollen and P. J. Curran. *Latent curve models: a structural equation perspective*, volume 467. Wiley-Interscience, 2006.

[15] B. O. Muthén. Beyond sem: general latent variable modeling. *Behaviormetrika*, 29:81–118, 2002.

[16] T. Lu, W. Poon, and Y. Tsang. Latent growth curve modeling for longitudinal ordinal responses with applications. *Computational Statistics & Data Analysis*, 55:1488–1497, 2011.

[17] C. R. Rao. Some statistical methods for comparison of growth curves. *Biometrics*, 14:1–17, 1958.

[18] L. R. Tucker. Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23:19–23, 1958.

[19] W. Meredith and J. Tisak. "tuckerizing" curves. In *annual meeting of the Psychometric Society, Santa Barbara, CA*, 1984.

[20] W. Meredith and J. Tisak. Latent curve analysis. *Psychometrika*, 55: 107–122, 1990.

[21] J. J. McArdle and D. Epstein. Latent growth curves within developmental structural equation models. *Child development*, 58:110–133, 1987.

[22] B. O. Muthén and K. Shedden. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55:463–469, 1999.

[23] B. O. Muthén and L. K. Muthén. Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research*, 24:882–891, 2000.

[24] B. O. Méthen. Latent variable mixture modeling. In R. E. Schumacker G. A. Marcoulides, editor, *New developments and techniques in structural equation modeling*, pages 1–33. Lawrence Erlbaum Associates, 2001.

[25] R. H. Hoyle. *Handbook of structural equation modeling*. Guilford Publication, 2012.

[26] J. R. Hipp and D. J. Bauer. Local solutions in the estimation of growth mixture models. *Psychological Methods*, 11:36–53, 2006.

[27] D. S. Nagin. Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological methods*, 4:139–157, 1999.

[28] D. S. Nagin and R. E. Tremblay. Developmental trajectory groups: fact or a useful statistical fiction? *Criminology*, 43:873–904, 2005.

[29] D. S. Nagin and R. E. Tremblay. Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological methods*, 6:18–34, 2001.

[30] J. K. Vermunt and L. Van Dijk. A nonparametric random-coefficients

approach: the latent class regression model. *Multilevel Modelling Newsletter*, 13:6–13, 2001.

[31] B.O. Muthén. Second–generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class–latent growth modeling. In A. Sayer L. M. Collins, editor, *New methods for the analysis of change*, pages 291–322. Washington, D.C., APA, 2001.

[32] D. J. Bauer and P. J. Curran. The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological methods*, 9:3–29, 2004.

[33] F. Kreuter and B. O. Muthén. Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology*, 24:1–31, 2008.

[34] F. Bartolucci, S. Bacci, and F. Pennoni. Longitudinal analysis of self-reported health status by mixture latent auto-regressive models. *Journal of the Royal Statistical Society C*, 63:267–288, 2014.

[35] P. McCullagh. Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B*, 42:109–142, 1980.

[36] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

[37] P. J. Green. Reversible jump Markov chain Monte Carlo computation and bayesian model determination. *Biometrika*, 82:711–732, 1995.

[38] L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41:164–171, 1970.

[39] L. R. Welch. Hidden markov models and the baum-welch algorithm. *IEEE Information Theory Society Newsletter*, 53:10–13, 2003.

[40] M. A. Tanner. *Tools for statistical inference*. Springer, New York, 1996.

[41] T. Orchard and M. A. Woodbury. A missing information principle: theory and applications. In *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, pages 697–715, Berkeley, 1972.

[42] T. A. Louis. Finding the observed information matrix when using the em-algorithm. *Journal of the Royal Statistical Society, Series B*, 44:226–233, 1982.

[43] F. Pennoni. *Issues on the estimation of latent variable and latent class models*. Scholars' Press, Saarbücken, 2014.

[44] Z. Feng and C. E. McCulloch. Using bootstrap likelihood ratios in finite mixture models. *J. R. Statist. Soc., series B*, 58:609–617, 1996.

[45] R. C. H. Cheng and W. B. Liu. The consistency of estimators in finite mixture models. *Scandinavian Journal of Statistics*, 28:603–616, 2001.

[46] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and Csaki F., editors, *Second International symposium on information theory*, pages 267–281, Budapest, 1973. Akademiai Kiado.

[47] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[48] S. Boucheron and E. Gassiat. An information-theoretic perspective on order estimation. In T. Rydén O. Cappé, E. Moulines, editor, *Inference in Hidden Markov Models*, pages 565–602. Springer, 2007.

[49] D. Rusakov and D. Geiger. Asymptotic model selection for naive Bayesian networks. In *Proceedings of the eighteenth conference on uncertainty in artificial intelligence*, pages 438–455. Morgan Kaufmann Publishers Inc., 2002.

[50] J. Magidson and J. K. Vermunt. Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, 31: 223–264, 2001.

[51] S. Bacci, S. Pandolfi, and F. Pennoni. A comparison of some criteria for states selection in the latent Markov model for longitudinal data. *Advances in Data Analysis and Classification*, 8:125–145, 2014.

[52] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13:260–269, 1967.

[53] B. H. Juang and L. R. Rabiner. Hidden Markov models for speech recognition. *Technometrics*, 33:251–272, 1991.

[54] K. E. Masyn, H. Petras, and W. Liu. Growth curve models with categorical outcomes. In *Encyclopedia of Criminology and Criminal Justice*, pages 2013–2025. Springer, 2014.

[55] M. Wang and T. E. Bodner. Growth mixture modeling identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods*, 10:635–656, 2007.

[56] K. L. Nylund and K. E. Masyn. Covariates and latent class analysis: Results of a simulation study. In *society for prevention research annual meeting*, 2008.

[57] K. L. Nylund, T. Asparouhov, and B. O. Muthén. Deciding on the number of classes in latent class analysis and growth mixture modeling: a monte carlo simulation study. *Structural Equation Modeling*, 14:535–569, 2007.

[58] V. Ramaswamy, W. S. DeSarbo, D. J. Reibstein, and W T Robinson. An empirical pooling approach for estimating marketing mix elasticities with pims data. *Marketing Science*, 12:103–124, 1993.

[59] G. Celeux and G. Soromenho. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of classification*, 13:195–212, 1996.

[60] D. J. Bauer and P. J. Curran. Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological methods*, 8:338, 2003.

[61] D. J. Bauer and P. J. Curran. Overextraction of latent trajectory classes: much ado about nothing? reply to rindskopf (2003), muthén (2003), and cudeck and henly (2003). *Psychological Methods*, 8:384–393, 2003.

[62] B. O. Muthén. Statistical and substantive checking in growth mixture modeling: comment on bauer and curran (2003). *Psychological Methods*, 8:369–377, 2003.

[63] J. Twisk and T. Hoekstra. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. *Journal of clinical epidemiology*, 65:1078–1087, 2012.

[64] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `http://www.R-project.org/`.