Published in final edited form as:

Struct Equ Modeling. 2016 March 1; 23(2): 221–233. doi:10.1080/10705511.2014.979932.

Regime Switching Modeling of Substance Use: Time-Varying and Second-Order Markov Models and Individual Probability Plots

Michael C. Neale¹, Shaunna L. Clark¹, Conor V. Dolan², and Michael D. Hunter³

¹Virginia Commonwealth University ²Free University of Amsterdam, The Netherlands ³University of Oklahoma

Abstract

A linear latent growth curve mixture model with regime switching is extended in 2 ways. Previously, the matrix of first-order Markov switching probabilities was specified to be time-invariant, regardless of the pair of occasions being considered. The first extension, time-varying transitions, specifies different Markov transition matrices between each pair of occasions. The second extension is second-order time-invariant Markov transition probabilities, such that the probability of switching depends on the states at the 2 previous occasions. The models are implemented using the R package OpenMx, which facilitates data handling, parallel computation, and further model development. It also enables the extraction and display of relative likelihoods for every individual in the sample. The models are illustrated with previously published data on alcohol use observed on 4 occasions as part of the National Longitudinal Survey of Youth, and demonstrate improved fit to the data.

Keywords

finite mixture; latent growth curve; Markov model; OpenMx; R; regime switching

In recent years, the latent growth curve mixture model (LGCMM) has become a popular tool for analyzing longitudinal data, because it allows for the modeling of subpopulations with distinct developmental trajectories (B. O. Muthén, Khoo, Francis, & Boscardin, 2003; L. K. Muthén & Muthén, 2001; Nagin, 1999). In the LGCMM, there are several subpopulations, referred to as mixture components or classes, that form an overall population. An individual's membership in a subpopulation is unknown, so multiple group approaches cannot be used. Instead, the subpopulations are modeled as a mixture distribution, in which class membership is modeled as the probability of being in a given trajectory class. In the LGCMM, classes are considered distinct when at least one of the growth curve parameters (means, variances or covariances of intercept, slope, or other growth curve factors) is different.

Correspondence should be addressed to Michael C. Neale, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Box 980126, Richmond, VA 23298. mcneale@mac.com.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hsem.

In certain substantive applications, it seems likely that individuals would not remain in the same class over time. For example, an individual might not drink at all during the first two measurement occasions in a study, have a binge drinking episode near the third measurement occasion, and then not drink for the remaining measurement occasions. The LGCMM would fail to capture the switching from abstinence to heavy, binge drinking and the subsequent return to not drinking occurring in this individual, and the presence of such individuals could bias parameter estimates in the LGCMM. Dolan, Schmittmann, Lubke, and Neale (2005) illustrated the concept of switching in the LGCMM. For consistency with that paper and ease of presentation, we use the term *regime* to denote one of the mixture components or trajectory classes in the LGCMM. The terms *regime*, *mixture component*, and *trajectory class* are used synonymously. In the regime switching LGCMM, individuals are allowed to switch between regimes over time. For example, assume that there is an LGCMM with two regimes, one representing individuals who do not smoke and the other a regime of smokers. These regimes would be represented by latent growth curves with flat, zero slopes and distinct intercepts, one high and one low. The high intercept latent growth curve corresponds to the individuals who smoke, and the low intercept is for individuals who do not smoke. Under the regime switching LGCMM, individuals could switch from the nonsmoking to the smoking regime, or the reverse. An individual could make any number of regime switches throughout the course of a study. These regime switches may be modeled using a Markov transition model (Kemeny & Snell, 1976).

Generally speaking, the regime switching models discussed here are a type of dependent mixture model, as contrasted with the more familiar independent mixture model (McLachlan & Peel, 2000). Because the dependency is Markov, these regime switching models could be called hidden Markov models (Zucchini & MacDonald, 2009). The fundamental mathematical background for hidden Markov models (HMMs) is given by Baum and colleagues (Baum & Petrie, 1966; Baum, Petrie, Soules, & Weiss, 1970; Petrie, 1969), but several review articles provide better introduction (e.g., Eddy, 1996, 1998; Rabiner, 1989; Rabiner & Juang, 1986). The primary distinction between what is traditionally called an HMM versus the approach here is sample size. HMMs are a time series method, involving large numbers of temporal observations, whereas here we apply a longitudinal method to only a few temporal observations but with many repeated instances in different individuals. The method here is linked to latent transition analysis but with a more complete measurement model (see Sterba, 2013). The approach is similar to that used by Chow, Grimm, Filteau, Dolan, and McArdle (2013) in that it uses standard structural equation modeling (SEM) with constrained mixture models to represent the dependent structure of the mixture models. It is successful when the number of temporal observations is not large. Several alternative specifications are possible (e.g., Hamaker, Grasman, & Kamphuis, 2010; Visser, 2007; Visser & Speekenbrink, 2010), but they are generally more suited to long time series with few people (Visser, Raijmakers, & Molenaar, 2002). Because the data here involve only a few occasions and many persons, the SEM method with constrained mixtures is advantageous.

The original Dolan et al. (2005) regime switching LGCMM specifies that the transition probabilities are equal across time (i.e., time-invariant), which is consistent with a standard

first-order Markov model. However, such time invariance might not hold for certain traits during periods of growth and development, such as adolescence and young adulthood. For example, compare the probabilities of transitioning from a nondrinking regime to a drinking regime between 9 and 10 years of age to that between ages 20 and 21. The former probability would be low in most populations, but the transition probability at the latter ages could be substantial—especially in the United States where possession and consumption of alcohol changes from illegal to legal at 21 years of age.

A second aspect of the original treatment is that the Markov process modeled was first-order; that is, only the current state influences the probabilities of transition to the next occasion of measurement. The first-order Markov assumption greatly reduces the complexity of a dynamic model, at the potential cost of overlooking some important factors. It seems possible that an individual's substance use history earlier than the immediately prior occasion of measurement could better predict the probability of transition. For example, someone with a prior history of alcohol abuse but who became abstinent—due to, for example, adverse consequences of or treatment for alcohol abuse—might be more likely to switch from abstinence to heavy use than someone who had been a lifelong tee-totaler. Similarly, someone with a prior history of major depression might be at greater risk of a transition to a depressed state than someone with no history (a phenomenon known as *kindling*; Segal, Williams, Teasdale, & Gemar, 1996). The first-order Markov model assigns the same transition probability for two individuals in the same state regardless of their prior history, but this might not accurately model individual differences in the population.

A third, technical, limitation of Dolan et al. (2005) is that it was implemented using a combination of Mx (Neale, Boker, Xie, & Maes, 2003) and R (R Core Team, 2013). The Mx script presented in the appendix of that article is difficult to interpret, generalize, or extend. The approach uses computer memory inefficiently, which restricts its potential utility for applications with more occasions of measurement. These limitations could have resulted in relatively few subsequent applications of the model, despite its intuitive and attractive properties and ability to provide a superior fit to data on substance use.

This article has five main aims. First, we extend the regime switching model to include time-varying transition probabilities, in which each pair of consecutive occasions has a different transition probability matrix. Second, we specify a second-order Markov process. Third, we apply these modified models to the National Longitudinal Survey of Youth (NLSY) data used in Dolan et al. (2005), and demonstrate that the time-varying transition probabilities model fits better. Fourth, we describe the extraction of individual trajectory likelihoods, and plot a number of illustrative cases. Finally, we present an accessible implementation of the regime switching model in the freely available R package OpenMx (Boker et al., 2011). Annotated scripts for the models presented in this article appear in the supplementary materials.

MODEL SPECIFICATION

The following two subsections briefly recapitulate the description in Dolan et al. (2005), using the same notation. They are prerequisites for the specification of the time-varying and

second-order Markov models, descriptions of which follow. The last subsection describes the computation of individual probabilities and the risk of entering particular states on one or more occasions of measurement. These occasions could be subsequent to those actually measured.

Latent Growth Curve Mixture Model

Following Dolan et al. (2005), we let y_{ij} denote the T -dimensional vector of scores for subject i ($i = 1 \dots N_j$) in regime j ($j = 1 \dots R$). The test scores are univariate measures observed at T occasions: $y_{ij} = [y_{ij1}, y_{ij2}, \dots, y_{ijT}]'$. To ease presentation, we consider only the simple linear LGCMM, but extensions to higher polynomial, exponential or other functional forms (Gompertz, logistic, etc.) would be straightforward (Neale & McArdle, 2000). An individual's vector of scores y_{ij} is a function of latent growth curve and residual error components:

$$y_{ij} = \Lambda \eta_{ij} + \epsilon_{ij} \quad (1)$$

in which η_{ij} is the two-dimensional matrix containing individuals' values on the random level (intercept) factor and random shape (slope) factor, and ϵ_{ij} represents residual terms. We assume $\eta_{ij} \sim N(a_j, \Psi_j)$, where the vector a_j contains the means of the intercepts and slopes, $a_j = [a_{Lj} \ a_{Sj}]$, and Ψ_j contains the variances and covariance of the levels and slopes $(\psi_{Lj}^2, \psi_{LSj}^2, \text{ and } \psi_{Sj}^2)$. Fixed effects could be specified by fixing Ψ_j to zero. Note that the means and variances for the latent variables, η_j , might differ across regimes due to the inclusion of the subscript j . For simplicity, we assume that the residuals are mutually uncorrelated and uncorrected with η . The residuals are distributed $\epsilon_{ij} \sim N(\mathbf{0}, \theta_j)$, where θ_j is the covariance matrix of the residuals, and $\mathbf{0}$ is the vector of zero means. The $T \times 2$ dimensional matrix Λ contains the fixed basis parameters. These could be chosen in many ways, for instance,

$$\Lambda' = \begin{bmatrix} 0 & 1 & 1 & \dots & 1 & 1 \\ 0 & 1 & 2 & \dots & T-1 & T \end{bmatrix}. \quad (2)$$

The basis parameters of the slope factor determine the form of the function, which is linear in this case. The choice of the values in Λ (e.g., if the second row is centered at zero or is polynomial contrast coefficients) affects the correlation between the level and slope factors (Rovine & Molenaar, 1998; Stoel & van den Wittenboer, 2003). As in any other growth curve application, any interpretation of the level-slope correlation should be made with regard to the choice of these basis parameters.

If the residuals ϵ and random effects η are uncorrelated, the within-regime model based on Equation 1 implies:

$$\mu[\vartheta_j] = \Lambda \alpha_j \quad \text{and} \quad \Sigma[\vartheta_j] = \Lambda \Psi_j \Lambda' + \Theta_j. \quad (3)$$

The vector ϑ_j contains the parameters, both fixed and unknown, in α_j , Λ , Ψ_j , and Θ_j . Given R regimes without switching, the unconditional distribution is an R -component multivariate normal mixture. This mixture distribution could be written as:

$$g(y_{ij}; \Sigma[\vartheta_1], \dots, \Sigma[\vartheta_R], \mu[\vartheta_1], \dots, \mu[\vartheta_R], \pi) = \quad (4)$$

$$\sum_{j=1}^R \pi_j f_j(y_{ij}; \Sigma[\vartheta_j], \mu[\vartheta_j]) \quad (5)$$

(McLachlan & Peel, 2000), where $\pi = [\pi_1, \dots, \pi_R]$ is the vector of the mixing proportions, and $f_j(y_{ij}; \Sigma[\vartheta_j], \mu[\vartheta_j])$ is the T -variate normal probability density function (PDF) corresponding to mixture component j . The mixing proportions represent the proportion of cases in each component ($0 < \pi_j < 1$, $\sum \pi_j = 1$). The parameter π_j could be viewed as the prior probability that a given case belongs to component j . Individual probabilities of regime membership given the data could be calculated by Bayes's theorem. Given π_j (the prior probability of belonging to component j) and $f_j(y_{ij}; \Sigma[\vartheta_j], \mu[\vartheta_j])$ (the probability of y_{ij} within, given that one is a member of component j), the probability that individual i belongs to component j , given y_{ij} , is:

$$\text{Prob}(j|y_{ij}) = \frac{[\pi_j f_j(y_{ij}; \Sigma_j, \mu_j)] / g(y_{ij}; \Sigma[\vartheta_1], \dots, \Sigma[\vartheta_R], \mu[\vartheta_1], \dots, \mu[\vartheta_R], \pi)}{\quad} \quad (6)$$

As discussed by McLachlan and Peel (2000), maximum likelihood (ML) estimates of the parameters in $\vartheta_1, \dots, \vartheta_R$ and π may be obtained by maximizing the likelihood function:

$$(\vartheta, \pi; Y) = \sum_{i=1}^N g(y_{ij}; \Sigma[\vartheta_1], \dots, \Sigma[\vartheta_R], \mu[\vartheta_1], \dots, \mu[\vartheta_R], \pi) \quad (7)$$

where N is the total sample size, $\vartheta = [\vartheta_1, \dots, \vartheta_R]$, and Y is the matrix containing the data. Either (quasi-) Newton or the expectation maximization (EM) algorithm can be used to maximize this function (Dolan & van der Maas, 1998; McLachlan & Peel, 2000; B. Muthén & Shedden, 1999; Yung, 1997).

Regime Switching in the Latent Growth Curve Mixture Model

In the LGCM it is assumed that an individual remains in one regime throughout the span of the longitudinal study. We thus have as many components in the mixture as there are regimes. For certain traits (e.g., substance use or depression) it seems likely that individuals would switch between regimes. Suppose that there are $R = 2$ regimes, which we denote as A and B, and $T = 3$ repeated measures. In total we have the following eight (R^T) possible paths through the regimes: A-A-A, A-A-B, A-B-A, A-B-B, B-A-A, B-A-B, B-B-A, and B-B-B. Each possible path represents a component in the mixture. In the standard LGCM, there are R components, but with switching there are R^T components. Following Schmittmann, Dolan, and Neale (2003), we formulate and fit the model with switching as a constrained multivariate normal mixture model (as in Equations 5 and 7). We use a Markov transition model to model switching (Kemeny & Snell, 1976). From this model, we derive mixing

proportions of the mixture model. We introduce the following unconditional initial probabilities: the probability of being in regime A at Occasion 1, q_{1A} , and the probability of being in regime B at Occasion 1, $q_{1B} = 1 - q_{1A}$. In addition, we introduce the conditional transition probabilities, as shown in Table 1.

For instance, in Table 1 $q(B|A)$ is the probability of being in regime B at time t , given that one was in regime A at time $t - 1$. Note that $q(B|A) = 1 - q(A|A)$ and $q(B|B) = 1 - q(A|B)$. By straightforward calculation we derive the probabilities associated with the eight paths; that is, the mixing proportions. These are shown in Table 2 for the time-invariant, time-varying, and second-order Markov models.

These equations precisely match the dependent mixture model likelihood given by Zucchini and MacDonald (2009, p. 37). To continue with the $T = 3$ example, the likelihood of a Markov-dependent mixture model is

$$L_M = \sum_{j_3=1}^R \sum_{j_2=1}^R \sum_{j_1=1}^R q_{1j_1} q_{j_2|j_1} q_{j_3|j_2} f_{j_1} f_{j_2} f_{j_3} \quad (8)$$

where f_j is shorthand for the multivariate normal PDF in regime j , $f_j = f_j(y_{ij}; \Sigma[\vartheta_j], \mu[\vartheta_j])$, q_{1j} is the initial probability of regime j , and $q_{j_2|j_1}$ is the probability of transitioning to regime j_2 conditional on being in regime j_1 . The likelihood we use is algebraically equivalent. Making substitutions from Table 2, we then have

$$\sum_{j=1}^{R^3} \pi_j f_j \quad (9)$$

where π_j is defined as in Table 2 and f_j is modified to be the joint multivariate normal PDF that includes the multiple times of measurement. Hence the method transforms a dependent mixture model into a constrained independent mixture model for estimation in a standard SEM format. This is the likelihood for a single row vector of data, as in the standard independent mixture likelihood given in Equation 5. The likelihood of the entire sample is simply the product of row likelihoods as in Equation 7, although equivalently for computational purposes the log-likelihoods are obtained and summed. The modifications for the multivariate normal PDF are described later.

Next, we require the covariance and mean structure model within the components. To accommodate regime switching, we specify the following model within each component:

$$\mu[\theta_k] = \Gamma_k \kappa \quad \text{and} \quad \Sigma[\theta_k] = \Gamma_k \Phi \Gamma_k' + \theta_k. \quad (10)$$

The 4×4 covariance matrix Φ contains the variances and covariances of the levels and slopes in regimes A and B :

$$\Phi = \begin{bmatrix} \psi_{LA} & & & \\ \psi_{LSA} & \psi_{SA} & & \\ \phi_{31} & \phi_{32} & \psi_{LB} & \\ \phi_{41} & \phi_{42} & \psi_{LSB} & \psi_{SB} \end{bmatrix} = \begin{bmatrix} \Psi_A & & \\ \Psi_{AB} & \Psi_B & \end{bmatrix}. \quad (11)$$

The parameters ψ_{LA} , ψ_{LSA} , and ψ_{SA} are the variances and covariance of the levels and slopes in regime A making up the Ψ_A matrix. The ψ_{LB} , ψ_{LSB} , and ψ_{SB} parameters similarly relate to the B regime and Ψ_B . The ϕ_{31} , ϕ_{32} , ϕ_{41} , and ϕ_{42} parameters become the Ψ_{AB} matrix, which can be constrained in various ways or set to zero. We discuss the parameters ϕ_{31} , ϕ_{32} , ϕ_{41} , and ϕ_{42} later.

The four-dimensional vector κ contains the means of the slopes and intercepts in the two regimes: $\kappa = [\alpha_{LA} \quad \alpha_{SA} \quad \alpha_{LB} \quad \alpha_{SB}]' = [\alpha'_A \quad \alpha'_B]'$. The $T \times 4$ matrix of $q_{A|A}$ and $q_{B|B}$. Only the latter two are new relative to the Γ_k contains the fixed factor loadings, in either columns 1 and 2 (regime A), or columns 3 and 4 (regime B), depending on the component. For instance, in Component 1 (A-A-A), the matrix is as follows:

$$\Gamma_1 = \begin{bmatrix} \lambda_1 & z \\ \lambda_2 & z \\ \lambda_3 & z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix} \quad (12)$$

where λ_i denotes the i th row in the matrix of polynomials λ (see earlier), and z denotes the two-dimensional zero row vector. The matrix r_k in a given component k is constructed as follows. If the component specifies that one is in regime A at occasion t , the t th row of Γ_k is $[\lambda_t \ z]$. Conversely, if the component specifies that one is in regime B at occasion t , the t th row of Γ_k is $[z \ \lambda_t]$. For instance, Γ_k is associated with component A-B-A, and given the choice of λ (see Equation 2), we have

$$\Gamma_3 = \begin{bmatrix} \lambda_1 & z \\ z & \lambda_2 \\ \lambda_3 & z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 2 & 0 & 0 \end{bmatrix}. \quad (13)$$

The covariance matrix of the residuals, θ_k , contains the residual variance associated with a given regime, at a given measurement occasion. Assuming that these matrices are diagonal, they equal:

$$\begin{aligned} \theta_1 &= \text{diag} \left(\sigma_{\epsilon 1A}^2 \sigma_{\epsilon 2A}^2 \sigma_{\epsilon 3A}^2 \right) \\ \theta_2 &= \text{diag} \left(\sigma_{\epsilon 1A}^2 \sigma_{\epsilon 2A}^2 \sigma_{\epsilon 3B}^2 \right) \\ &\vdots \\ \theta_7 &= \text{diag} \left(\sigma_{\epsilon 1B}^2 \sigma_{\epsilon 2B}^2 \sigma_{\epsilon 3A}^2 \right) \\ \theta_8 &= \text{diag} \left(\sigma_{\epsilon 1B}^2 \sigma_{\epsilon 2B}^2 \sigma_{\epsilon 3B}^2 \right) \end{aligned} \quad (14)$$

where $\sigma_{\epsilon tB}^2$ ($\sigma_{\epsilon tA}^2$) is the variance of the residual at occasion t in regime B (A).

The number of parameters in the standard LGCM, given $T = 3$ occasions and $R = 2$ regimes, is 17 (in each of the two regimes, A and B, 2 means, 3 (co)variances, 3 residual variances, so $2 * (2 + 3 + 3) = 16$, plus one mixing proportion = 17). Allowing for switching, as specified previously, increases the number of components from two to eight. However, the number of parameters increases only from 17 to 23. Although there are six new mixture components, all these parameters are constrained to be functions of only six new parameters. Only two new mixing proportions must be estimated: the self-transition probabilities for A, $q_{A|A}$, and for B, $q_{B|B}$. The cross-component transition probabilities are functions of these: $q_{B|A} = 1 - q_{A|A}$ and $q_{A|B} = 1 - q_{B|B}$. So, all of the mixing proportions are functions of just three parameters q_{1A} (because $q_{1B} = 1 - q_{1A}$), standard LGCM because the first mixing proportion was already required by that model.

Four additional parameters are required to account for the possible correlations between the random slopes and intercepts of regime A and those of regime B, when a switch takes on prior expectations (e.g., $\phi_{31} = \phi_{42}$, $\phi_{32} = \phi_{41}$). In addition to these three mixing proportions and additional four covariances, we estimate the 16 parameters associated with the means and covariance structure within each regime A and B (2 means, 3 (co)variances, 3 residual variances in each regime, A and B).

Although this model could be fitted using Mx (Neale et al., 2003), it is much simpler to use the OpenMx package in R (Boker et al., 2011). OpenMx is a package for fitting structural equation and other statistical models. It has two primary interfaces for model specification. One is based on SEM paths, which uses the `mxPath()` function, and a second specifies models directly using arbitrary matrix algebras specified with the `mxMatrix()` and `mxAlgebra()` functions. The latter approach is used here, although either approach could in principle be used. Supplementary Materials 1 contains an annotated R script to fit the regime switching model to the NLSY data. The essence of the script is to construct a template model for all R^T regimes (81 in this application where $R = 3$ and $T = 4$, so $R^T = 3^4 = 81$). The regime-specific models are constructed by modifying the predicted means and covariances according to Equations 10, 11, and 14. The 81 mixture components are constructed in a loop, so that specification of this complex mixture distribution model is accomplished in fewer than 100 lines of R code.

Time-Varying Transition Probabilities

Three models for time-varying transition probabilities are considered. First, it is possible to specify a saturated switching model, in which $R^T - 1$ independent probabilities are estimated for the R^T component mixture of regime switching patterns. This model is attractive as a benchmark against which the fit of simpler models involving fewer parameters can be compared. In practice, however, this model can be difficult to fit to real data, because the number of parameters is large ($R^T - 1$ independent proportions) and because mixture distributions generally are subject to local minima.

Second, time-varying transition probabilities are readily specified as $R - 1$ different parameter matrices for each pair of occasions. This specification employs $R - 1$ parameters for the initial probabilities, and $R(R - 1)(T - 1)$ parameters for the transition matrices; that is, $(R - 1)(R(T - 1) + 1)$ in total. In most applications this is considerably fewer than the R^T

– 1 parameters of the saturated model, but substantially more than for the time-invariant first-order Markov process that uses only $R^2 - 1$. The time-invariant first-order Markov model has $\mathcal{O}(R^2)$ parameters¹, whereas the time-variant first-order Markov model has $\mathcal{O}(R^2T)$ parameters, and the saturated switching model has $\mathcal{O}(R^T)$.

Generally, $\mathcal{O}(R^2) < \mathcal{O}(R^2T) < \mathcal{O}(R^T)$. Thus, when the number of occasions of measurement is high, the saturated switching model becomes unattractive due to the very large number of parameters needed to specify the probabilities, which in turn requires both substantial sample size and significant computational effort for the estimation.

Third, a second-order Markov model can be implemented, in which transition probabilities are a function of both the current state and the one prior to that (e.g., Shamshad, Bawadi, Hussin, Majid, & Sanusi, 2005). With R regimes, there are effectively R^2 prior states (A-A, A-B, B-A, B-B in the preceding two-regime example), and R possible current states. The probabilities of transitioning into the R current states must sum to unity for each of the prior states, so there are $R^2(R - 1)$ independent free parameters for the transitions. The transition probabilities between the first two occasions (for which only one prior state is known) are obtained as the average of the R transition probabilities. For example, the probability of an A-A transition is the average of the A-A-A and the B-A-A transition probabilities. This model (like the first-order Markov model) has the advantage that the number of parameters required does not increase with the number of occasions of measurement. It has fewer parameters than time-varying transitions model when $R < T - 1$.

Computation and Graphical Display of Posterior Probabilities

In certain applications, the probability that an individual enters a certain state or regime might be of particular interest. For example, the probability that an individual will switch to a regime with dangerous levels of substance use, abuse, or dependence is especially relevant for clinical practice, or for research studies aiming to identify factors that affect this risk. Similarly, there might exist depressive states in which suicide risk is high, and the ability to predict an individual's chance of entering that state would be very valuable. Here we consider some possibilities for these calculations.

The saturated switching and the time-varying models have the disadvantage of limited utility for predicting future outcomes. They begin with the assumption that the process itself is changing, so extrapolation to future states is purposefully difficult with these models. Possibly, the transition probabilities between the final two occasions of measurement could be assumed to apply to future occasions, under the assumption that the transitions have reached asymptotic values. Some empirical support for this could be obtained by comparing a model in which the transition probability estimates between occasions $t - 1$ and t are set equal to the corresponding estimates between occasions $t - 2$ and $t - 1$. Thus there could be a best estimate of the $R \times R$ transition probability matrix, which is also the case for the first-order Markov process. To obtain regime membership probabilities at occasion $t + 1$, the current regime membership probabilities π_t can be postmultiplied by the transition

¹The symbol $\mathcal{O}$ is used for Big $\mathcal{O}$ notation of the order of magnitude, rather than the exact numeric formula. Thus, $\mathcal{O}(R^2)$ is read “on the order of R^2 ”.

probability matrix $\mathbf{P}$, and this process can be repeated to obtain π_{t+x} for any value of x . If the transition matrix $\mathbf{R}$ is regular (i.e., there are no zero entries of $\mathbf{P}^z$ for at least one value of $z \geq 1$) then the regime membership probability vector π will approach a steady state (Bertsekas & Tsitsiklis, 2008). Thus regardless of their initial probabilities, in the limit as $z \rightarrow \infty$ all individuals have the same expected steady state regime membership probabilities. However, over a finite interval, differences between individuals' posterior probability vectors generate differences in expected regime membership probabilities.

The vector of individual posterior probabilities that individual i is in regime k are readily computed as the likelihood of the individual's data given that they are in regime k , divided by the sum of these likelihoods over all regimes, per Equation 6. This method of posterior probabilities is analogous to the Viterbi optimal decoding algorithm for hidden Markov models (Viterbi, 1967). The posterior probabilities can be aggregated in various ways, for example, over all trajectories that include at least one heavy use state. Also valuable are graphical displays of individual posterior probabilities. One approach is to plot the occasions of measurement on the x axis, the regimes on the y axis, and draw a separate line for each of the trajectories. The trajectories can be distinguished using a color scheme, such that sequences A-A-A to B-B-B would be ordered with the last occasion of measurement changing most rapidly. These trajectories can be drawn with line thickness proportional to the individual posterior probability computed using Equation 6. To anticipate the following application below, Figure 1a shows a four occasion and three regime plot, with equiprobable regime trajectories. In the color version, trajectories range from Light-Light-Light-Light in red, through the rainbow to violet for Heavy-Heavy-Heavy-Heavy, with the later occasions changing most rapidly (so that, e.g., Light-Light-Light-Heavy would be colored red and Heavy-Heavy-Heavy-Light would be violet). Individual data points would be superimposed on this plot, using the right-hand y axis. Note that this equiprobable plot is not equivalent to the individual plot for someone whose data are all missing. In this case, the prior probabilities of the regimes, computed from the initial probability and transition probability matrices, would be correct. Figure 1b shows this prior probability plot for the application, which we now describe.

APPLICATION

Dolan et al. (2005) found that the regime switching model fitted data on alcohol use considerably better ($-2\log L = 352$) than did an LGCM without switching, but there might be room for further improvement. In particular, it seems possible that the probability of transition from one regime to another could differ according to the age at which the transition occurs. Therefore, we fit an extended regime switching model to the same data as Dolan et al. (2005). These data consisted of four repeated measures ($T = 4$) relating to alcohol consumption among adolescents from the NLSY (NLSY97; downloaded from <http://www.bls.gov/nls/nlsy97.htm>) for years 1998, 1999, 2000, and 2001. At each occasion, participants responded to the question "In the past 30 days, on the days you drank alcohol, about how many drinks did you usually have?" The original analysis was restricted to 737 White males and females who were aged 13 or 14 in the year 1998, and who indicated that they regularly drank alcohol. Here we focus on extending the three-regime switching model LGCM3-SW2, which provided the best fit (lowest $-2\log L$ and lowest Bayesian Information

Criterion [BIC]) to the data. The growth curve parameter estimates correspond to Light (L), Moderate (M), and Heavy (H) drinking regimes.

Latent Growth Curve Mixture Model With Time-Invariant Regime Switching

The time-invariant switching model (LGCM3-SW2 in Dolan et al. 2005) includes regime switching and certain restrictions on the level and slope parameters. These empirically justified restrictions were zero mean slope for the heavy group, zero variance for initial level of any group, and zero variance level for the heavy group. The OpenMx input file for this model is shown in Supplementary Materials 1 (Script A). Fit statistics for all models are shown in Table 3. Estimates of the initial probabilities, the level, slope, and residual parameters are shown in Table 4, and the transition probability matrix parameters and likelihood-based 95% confidence intervals (Neale & Miller, 1995) are shown in Table 5. These results agree with those found using Mx (Neale et al, 2003) as reported by Dolan et al. (2005).

Latent Growth Curve Mixture Model With Time-Varying Switching

We extended the three-regime switching LGCM with time-varying transition probabilities as described previously. This model adds 12 free parameters. There are nine parameters for the conditional transition probabilities for the transitions between occasions 2 and 3, and nine for the 3 to 4 transition, but in each case only six are independent, as the transition probabilities from any regime must sum to unity. The fit indexes for this model (Table 3) show that the regime switching LGCM with time-varying transition probabilities fits the data better, according to the likelihood ratio test, $\chi^2(12) = 23.8, p < .001$, and the Bayesian information criterion (Schwarz, 1978).

Several features of the time-varying transition probability estimates are notable. First, at the second and third transitions the estimate of the probability of remaining in the moderate or heavy use regimes is greater than it is at the first transition. Commensurate with that difference, the probability of remaining in the light use regime at the third transition is lower than it was at younger ages. These changes are consistent with increases in alcohol consumption with age. More interestingly, once individuals are in the moderate or heavy use regimes they are less likely to revert to lower use levels. This pattern is consistent with the habitual use of alcohol and the development of dependence. Second, it is possible to predict the longer term class membership probabilities, assuming that the transition probabilities have stabilized at the values for the transitions between the third and fourth occasions. This assumption might be incorrect, but the potential utility of the method can be illustrated.

Latent Growth Curve Mixture Model With Second-Order Markov Switching

The second-order Markov model has the same number of parameters as the time-varying switching model, but it fits the data slightly less well ($\chi^2 = 3.2$; see Table 3). Parameter estimates of the growth curve components and the initial probabilities (Table 4) are very similar to those of the other two models. The top right panel of Table 5 shows the transition probabilities between the first and second occasions, which are the averages of the estimates in the three lower right panels. These estimates are very similar to those of the first-order Markov model (top left panel). The probabilities in the lower three right panels are not

strictly comparable to those on the left, because they are based on both the prior occasion and the one before it. Individuals in the light use group at both prior occasions (LL) are most likely ($p = .55$) to be light users on the third occasion, but have a substantial probability of switching to the moderate regime ($p = .42$). There is only a small chance ($p = .03$) of such users switching to the heavy regime. Moderate–light (ML) users are also likely to stay light users ($p = .85$), but those with a prior history of heavy then light use (HL) have an appreciable chance of switching to heavy use ($p = .20$). In general, those with a decreasing pattern (ML, HM, and HL) have a relatively low risk of heavy use, $p = .04$, $.07$, and $.20$, respectively. Somewhat counterintuitively the HL group has a higher point estimate of heavy use probability (.20) than the HM, the 95% confidence intervals overlap considerably. Possibly, this heavy–light–heavy pattern indicates that an appreciable proportion of the population follows a pattern of remission and relapse. Across the estimates, the width of the confidence intervals varies substantially (e.g., LL to light, .44 to .64, vs. HL to light, .19 to 1.00), due to the relative density of the data relevant to estimating these statistics. Consistently light users are more common than those with a heavy then light use pattern. Reducing regimes, such as MM, MH, HM, or HH to light are relatively improbable, with estimates close to zero. Others—particularly consistent ones such as MMM and HHH—are relatively common, as are certain dramatic transitions such as LH to heavy.

Individual Probability Plots

We selected the best fitting, time-varying switching model to compute individual posterior probability plots, although the plots based on the second-order Markov plots are very similar, so this choice was not critical. Posterior probability plots (using the template described earlier) were drawn for all 737 subjects in the sample, and are available online as supplementary material, as are those for the second-order Markov. Figure 2 presents several illustrative examples. The top left panel shows data from an individual with alcohol consumption data vector [1, 2, 10, 15], which is close to the means for the light regime at on the first two occasions (1.2 and 1.4), and near the stable mean of 10.1 for the second two occasions. Accordingly, the most probable switching pattern for this individual follows the LLHH pattern, the thick orange line on the plot.

The top right panel of Figure 2 shows a relatively uncommon data pattern [10, 1, 1, 3] of heavy use at a young age, followed by subsequent light and then moderate use. Possibly, the first occasion data were an outlier due to misunderstanding the question. The bottom two panels illustrate cases with missing data. On the right the last occasion data are missing, and the likelihood of observing each of the outcomes is equal (1/3). The weighted likelihoods are not equal and result in a strong prediction that this individual will be in the moderate class at the (unmeasured) 2001 occasion. We can make use of the normalized weighted likelihoods at the third occasion and the model parameters (bottom left panel of Table 5) to predict the outcomes for the fourth occasion. Multiplying the posterior probabilities at occasion 3 by the transition probability matrix between occasions 3 and 4 yields:

$$\begin{bmatrix} 0 & 0.93 & 0.07 \end{bmatrix} \begin{bmatrix} .52 & .45 & 0.03 \\ .06 & .86 & 0.09 \\ .03 & .25 & 0.71 \end{bmatrix} = \begin{bmatrix} 0.05 & 0.82 & 0.13 \end{bmatrix}.$$

These probabilities differ markedly from both the initial population probabilities ([0.47 0.37 0.16]) and the estimated probabilities at Time 4 [0.36 0.33 0.30]. Consistent with the graph at Time 3, moderate is the most likely class for this individual at Time 4.

DISCUSSION

In this section we consider first the substantive findings, and second the methodological developments presented here and additional possible extensions.

Application to Alcohol Use

Three models were fitted to NLSY data on self-report daily alcohol consumption collected on four occasions when the study participants were between 13 and 17 years of age. Dolan et al. (2005) had previously found that a regime switching model with time-invariant switching probabilities fitted the data much better than a model with no regime switching. Here, we found that a model with time-varying transition probabilities fits the data better still. This is unsurprising given the age range of the sample, when both the availability of alcohol and the social acceptability of its use are likely to have changed. Estimates of the growth curve parameters were largely unchanged by the modification. The largest differences in the switching parameters were from the light drinking class. With increasing age, there was less chance that someone in the light drinking regime would stay in it, and more chance that they would switch to the moderate class. This is consistent with greater availability of alcohol at later ages.

We also found that a second-order Markov model fit the data better than the time-invariant model. However, it fit worse than the time-varying model, and yielded parameter estimates that were less easy to interpret than those of the time-varying model. Thus although this model is conceptually appealing because it can be used for prediction of regime switching probabilities beyond the range of the study, in this case it appears less suitable for the data.

Finally, it is worth noting that the data suggest surprisingly high levels of alcohol consumption by some of the teenage participants in this study. They were asked “In the past 30 days, on the days you drank alcohol, about how many drinks did you usually have?” Failure to attend to the central clause “on the days you drank alcohol” would yield rather different responses than the per diem intent of the question. Indeed, more than 20 participants responded 15 to 25 when they were as young as 13, which seems more consistent with a monthly total than a daily quantity. However, the great majority of the sample responded in the range of 1 to 6, which is a reasonable level of daily consumption. With that caution in mind, removing those reporting more than 12 per day actually had relatively little effect on the parameter estimates indicating that these findings were not solely due to a misinterpretation of the question. Future studies might consider wording such questions as “In the past 30 days, on the days you drank alcohol, about how many drinks *per day* did you usually have?”

Methodological Development

This study extends the Dolan et al. (2005) methodological development of the regime switching LGCM in three important ways. First, implementation of the model in the R

package OpenMx makes it much more accessible, because only knowledge of the very popular R language is required. The previous treatment involved a combination of Mx and R, an unusual combination of skills. The R environment makes it simple to prepare the data, fit the models, and produce customized plots, such as those in Figures 1 and 2. A further problem with the implementation in Mx was that its memory usage was inefficient due to limitations of the Mx program language. Larger models involving more than three classes or time series with more than four occasions could not have been fitted due to memory limitations. These restrictions have been lifted by the OpenMx implementation, although long time series or large numbers of mixture components are still potentially demanding in terms of memory and CPU time for optimization. Speed has, however, improved, and not only due to hardware developments. The time-invariant model took 1 minute on an early 2011 Macbook Pro (2.3 GHz Intel Core i7 single-threaded) compared to over an hour using Mx on a Windows 1.8 GHz machine used in the earlier article. The time-varying and second-order Markov models took 2 and 3 minutes, respectively.

Two new models were presented. First was the time-varying transitions model, which specifies a different transition probability matrix between each pair of occasions. The model suffers the disadvantage that it requires more parameters as the number of occasions of measurement increase. It is also more difficult to extrapolate from the observed occasions to others either earlier or (more typically of interest) beyond the end of the series. Nevertheless, whether the transition probabilities are equal across different measurement intervals is an empirical question. Arguably, only when the transition matrix appears to have stabilized should the transition matrix between the last two occasions be used to extrapolate into future occasions. An intermediate model, with time-varying transitions early in the series, followed by time-invariant transitions in the later part of the series, could be used to test for stabilization of the transition probabilities.

A second-order Markov model was also implemented. To make use of information from occasions earlier than the immediately preceding one is a key advance. It sets the stage for implementing not only other higher order Markov processes, but for processes that depend on, for example, having entered a particular regime on any previous occasion of measurement. In certain substantive applications, the influence of more distal states seems likely to affect transition probabilities. For example, relapse (reuptake of a substance use habit after a period of abstinence or low use) is a very common problem with treatments for many substances of abuse and dependence (Hunt, Barnett, & Branch, 1971; Spear, Ciesla, & Skala, 1999; Xie, McHugo, Fox, & Drake, 2005). A period of reduced use or abstinence is a prerequisite for relapse, and is consistent with (at least) a second-order rather than simply first-order Markov process. One extension to the regime switching model would be to add an n -order Markov process so that substance use at the current time (T) can be dependent on all previous measurement occasions. Note that hybrid models are also possible. For example, the transition probabilities between the first and second occasion might be specified to have separate sets of free parameters, with a second-order model between subsequent occasions. Such specifications might fit the data better, while not greatly increasing the number of parameters.

An important extension for many psychological and behavioral traits is to extend the model to other scales of measurement. As presented here, the regime switching model is only appropriate for continuous outcomes with normally distributed residuals (i.e., conditionally normal). However, many substance abuse measures, for example, are neither continuous nor normally distributed. They might be ordered categorical, e.g., the frequency of use being daily, weekly, monthly, or yearly, or count data, such as a tally of the number of drinks in the last week. The modifications needed to adapt the existing R script to analyze ordinal data by normal theory full-information maximum likelihood are relatively straightforward. However, with current hardware they would become computationally intensive for time series with more than, for example, 10 occasions (and this would require $2^{10} = 1,024$ switching components even with only two regimes).

The treatment here focuses on the trajectories of a single variable over a small number of occasions. A relatively simple extension would be to include one or more measures of different variables. These could be antecedent, concurrent, or subsequent to the measures in the time series. These measures might be outcomes of substantive interest, or risk factors thought to influence either the growth curve components or the probability of switching between them. For example, the probability of switching from a moderate to a heavy smoking regime might be moderated by the polymorphisms in the nicotinic receptor *CHRNA5*, associated with smoking (Saccone et al., 2007). Similarly, the mean of the heavy smoking growth curve component might be directly influenced by this polymorphism. The addition of known risk factors of substance abuse, such as family history of drug use, diagnosis of conduct disorder, or low socioeconomic status, into the regime switching LGCM can help predict who is likely to be in each regime and aid in identifying those individuals who are at risk for transitioning into the heavy use regime. Through early identification of at-risk individuals, limited prevention resources can be spent on those who need it most. Outcomes of substance use trajectories, such as lung cancer or cardiovascular disease, might be better predicted from either the growth curve components or the probabilities of switching. A time series with a sufficient number of assessments can provide a within-person measure of lability (to switch between classes), which in turn might predict probabilities such as drug initiation, cessation, and relapse.

Two further types of multivariate extension are worthy of mention. First is the simultaneous analysis of two variables, both measured repeatedly. This is an active area of research, for example, the assessment of coupling (also known as parallel process growth modeling) in dynamical systems analysis (Boker & Laurenceau, 2007; Hu & Boker, 2013). The possibilities for regime switching models of coupled variables are quite extensive. One would be to allow covariance between the growth curve components across the two variables, which is commonly used for bivariate growth curves. Of possibly greater substantive interest would be to model switching probabilities of one variable conditional on the status of the other. For example, a switch to a heavy drinking regime might increase the probability of subsequently using other drugs of abuse. The second type of multivariate extension would be for the analysis of data from relatives, such as twins. This would permit estimation of heritabilities of growth curve variance components, and familial resemblance for trajectories. The latter could be obtained in a two-step analysis from estimated individual trajectory likelihoods, but a single-step analysis is generally preferable, as it retains

information about the statistical precision and therefore enables valid statistical inference in both steps. These two types of bivariate analysis, different variables and pairs of relatives, both suffer from the curse of dimensionality. The number of possible paired trajectories increases as the square of the number of components in the corresponding univariate models. The relatively simple three-regime, four-occasion model used in this article would require $81^2 = 6,561$ mixture components. There seems to be no obvious impediment to the analysis of mixtures with very large numbers of components, but parallel computation could prove especially useful to decrease CPU time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING

This research was supported by National Institute on Drug Abuse grants DA-18673 and DA-26119.

REFERENCES

- Baum LE, Petrie T. Statistical inference from probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*. 1966; 37:1554–1563.
- Baum LE, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*. 1970; 41:164–171.
- Bertsekas, DP.; Tsitsiklis, JN. Introduction to probability. 2nd. Athena Scientific; Nashua, NH: 2008.
- Boker, S.; Laurenceau, J. Coupled dynamics and mutually adaptive context. In: Little, TD.; Bovaird, JA.; Card, NA., editors. Modeling contextual effects in longitudinal studies. Taylor & Francis; New York, NY: 2007. p. 299–324.
- Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Fox J. OpenMx: An open source extended structural equation modeling framework. *Psychometrika*. 2011; 76:306–311. [PubMed: 23258944]
- Chow S-M, Grimm KJ, Filteau G, Dolan CV, McArdle JJ. Regime-switching bivariate dual change score model. *Multivariate Behavioral Research*. 2013; 48:463–502. doi: 10.1080/00273171.2013.787870. [PubMed: 26742002]
- Dolan CV, Schmittmann VD, Lubke GH, Neale MC. Regime switching in the latent growth curve mixture model. *Structural Equation Modeling*. 2005; 12:94–119.
- Dolan CV, van der Maas HLJ. Fitting multivariate normal mixtures subject to structural equation modeling. *Psychometrika*. 1998; 63:227–253.
- Eddy SR. Hidden Markov models. *Current Opinion in Structural Biology*. 1996; 6:361–365. [PubMed: 8804822]
- Eddy SR. Profile hidden Markov models. *Bioinformatics Review*. 1998; 14:755–763.
- Hamaker, EL.; Grasman, RP.; Kamphuis, JH. Regime-switching models to study psychological processes. In: Molenaar, PCM.; Newell, KM., editors. Individual pathways of change: Statistical models for analyzing learning and development. American Psychological Association; Washington, DC: 2010. p. 155–168.
- Hu Y, Boker SM. Permutation tests of coupled latent differential equations. *Multivariate Behavioral Research*. 2013; 48:160. [PubMed: 26789222]
- Hunt WA, Barnett LW, Branch LG. Relapse rates in addiction programs. *Journal of Clinical Psychology*. 1971; 27:455–456. [PubMed: 5115648]
- Kemeny, JG.; Snell, JL. Finite markov chains. Springer Verlag; New York, NY: 1976.
- McLachlan, GJ.; Peel, D. Finite mixture models. Wiley; New York, NY: 2000.

- Muthén, BO.; Khoo, S.; Francis, DJ.; Boscardin, CK. Analysis of reading skills development from kindergarden through first grade: An application of growth mixture modeling to sequential processes. In: Reise, SP.; Dua, N., editors. *Multilevel modeling: Methodological advances, issues, and applications*. Erlbaum; Mahwah, NJ: 2003. p. 71-89.
- Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*. 1999; 55:463–469. [PubMed: 11318201]
- Muthén, LK.; Muthén, BO. *Mplus* [Computer program]. Muthén & Muthén; Los Angeles, CA: 2001.
- Nagin D. Analyzing developmental trajectories: Semi-parametric, group-based approach. *Psychological Methods*. 1999; 4:139–177.
- Neale, MC.; Boker, SM.; Xie, G.; Maes, HHM. Mx: Statistical modeling. 6th. Department of Psychiatry, Virginia Commonwealth University; Richmond, VA: 2003.
- Neale MC, McArdle JJ. Structured latent growth curves for twin data. *Twin Research*. 2000; 3:165–177. [PubMed: 11035490]
- Neale MC, Miller MB. The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*. 1995; 27:113–120. [PubMed: 9145549]
- Petrie T. Probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*. 1969; 40:97–115.
- R Core Team. R: A language and environment for statistical computing [Computer software manual]. Author; Vienna, Austria: 2013. Retrieved from <http://www.R-project.org/>
- Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*. 1989; 77:257–286.
- Rabiner LR, Juang BH. An introduction to hidden Markov models. *IEEE ASSP Magazine*. 1986; 3:4–16.
- Rovine MJ, Molenaar PCM. The covariance between level and shape in the latent growth curve model with estimated basis vector coefficients. *Methods of Psychological Research Online*. 1998; 3:95–107.
- Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, Bierut LJ. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human Molecular Genetics*. 2007; 16:36–49. doi: 10.1093/hmg/ddl438. [PubMed: 17135278]
- Schmittmann VD, Dolan CV, Neale MC. Discrete latent Markov models for normally distributed response data. *Multivariate Behavior Research*. 2003; 40:461–488.
- Schwarz G. Estimating the dimension of a model. *Annals of Statistics*. 1978; 6:461–464.
- Segal ZV, Williams JM, Teasdale JD, Gemar M. A cognitive science perspective on kindling and episode sensitization in recurrent affective disorder. *Psychological Medicine*. 1996; 26:371–380. [PubMed: 8685293]
- Shamshad A, Bawadi MA, Hussin WMAW, Majid TA, Sanusi SAM. First and second order Markov chain models for synthetic generation of wind speed time series. *Energy*. 2005; 30:693–708.
- Spear SF, Ciesla JR, Skala SY. Relapse patterns among adolescents treated for chemical dependency. *Substance Use and Misuse*. 1999; 34:1795–1815. [PubMed: 10540973]
- Sterba SK. Understanding linkages among mixture models. *Multivariate Behavioral Research*. 2013; 48:775–815. doi:1080/00273171.2013.827564. [PubMed: 26745595]
- Stoel RD, van den Wittenboer G. Time dependence of growth parameters in latent growth curve models with time invariant covariates. *Methods of Psychological Research Online*. 2003; 19:21–41.
- Visser, I. depmix: An R-package for fitting mixture models on mixed multivariate data with Markov dependencies, [R-package manual]. 2007. Retrieved from <http://cran.r-project.org>
- Visser L, Raijmakers MEJ, Molenaar PCM. Fitting hidden Markov models to psychological data. *Science Progress*. 2002; 10:185–199.
- Visser I, Speekenbrink M. depmixS4: An R package for hidden Markov models. *Journal of Statistical Software*. 2010; 36(7):1–21. Retrieved from <http://www.jstatsoft.org/v36/i07/>.
- Viterbi AJ. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*. 1967; 13:260–269.

- Xie H, McHugo GJ, Fox MB, Drake RE. Substance abuse relapse in a ten-year prospective follow-up of clients with mental and substance use disorders. *Psychiatric Services*. 2005; 56:1282–1287. doi: 10.1176/appi.ps.56.10.1282. [PubMed: 16215196]
- Yung Y. Finite mixtures in confirmatory factor-analysis models. *Psychometrika*. 1997; 62:297–330.
- Zucchini, W.; MacDonald, IL. Hidden Markov models for time series: An introduction using R. Chapman & Hall/CRC; Boca Raton, FL: 2009.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 1.

Template reference plots for individual posterior probability plots. Each line traces a possible trajectory between three regimes (low, moderate, and heavy shown on the left y axis) across four occasions of measurement (1998–2001 shown on the x axis). (a) The equiprobable trajectories. (b) Prior probability plots based on the time-varying transition probability matrices estimated using the National Longitudinal Study of Youth data on alcohol consumption. The right axis is to enable concurrent plotting of individual data points superimposed on the posterior trajectory likelihoods.

FIGURE 2.

Four illustrative individual posterior probability plots from National Longitudinal Study of Youth data. The lines trace all possible trajectories among three regimes (low, moderate, and heavy shown on the left y axis) across four occasions of measurement (1998–2001 shown on the x axis). Observed data are plotted against the right axis. Clockwise from the top left: A common pattern of increasing alcohol use with age, an unusual pattern of decreasing alcohol use, an individual with missing data at the fourth occasion, and an individual with missing data on the first occasion.

TABLE 1

Example Conditional Transition Probabilities

t – I			
		A	B
<i>t</i>	A	$q_{A/A}$	$q_{A/B}$
	B	$q_{B/A}$	$q_{B/B}$

Note. Constraints on probabilities: $q_{A/A} + q_{B/B} = 1$, $q_{A/B} + q_{B/A} = 1$ (conditional transition probabilities).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2

Mixing Proportions in Three Models Given 2 Regimes (A & B) and $T = 3$ Occasions

Pattern	Prop.	Switching Model			second-order Markov
		Equal	Time-varying		
AAA	π_1	$q_{1A} * q_{A A} * q_{A A}$	$q_{1A} * q_{A A1} * q_{A A2}$		$q_{1A} * q_{A (A1,A2)}$
AAB	π_2	$q_{1A} * q_{A A} * q_{B A}$	$q_{1A} * q_{A A1} * q_{B A2}$		$q_{1A} * q_{B (A1,A2)}$
ABA	π_3	$q_{1A} * q_{B A} * q_{A B}$	$q_{1A} * q_{B A1} * q_{A B2}$		$q_{1A} * q_{A (A1,B2)}$
ABB	π_4	$q_{1A} * q_{B A} * q_{B B}$	$q_{1A} * q_{B A1} * q_{B B2}$		$q_{1A} * q_{B (A1,B2)}$
BAA	π_5	$q_{1B} * q_{A B} * q_{A A}$	$q_{1B} * q_{A B1} * q_{A A2}$		$q_{1B} * q_{A (B1,A2)}$
BAB	π_6	$q_{1B} * q_{A B} * q_{B A}$	$q_{1B} * q_{A B1} * q_{B A2}$		$q_{1B} * q_{B (B1,A2)}$
BBA	π_7	$q_{1B} * q_{B A} * q_{B B}$	$q_{1B} * q_{B B1} * q_{A B2}$		$q_{1B} * q_{A (B1,B2)}$
BBB	π_8	$q_{1B} * q_{B B} * q_{B B}$	$q_{1B} * q_{B B1} * q_{B B2}$		$q_{1B} * q_{B (B1,B2)}$

Note. For $T > 3$ occasions, the second-order Markov model has fewer parameters than the time-varying model. Constraints on probabilities: $\sum \pi_i = 1$ (mixing proportions); $q_{1A} + q_{1B} = 1$ (initial probabilities); $q_{A|A} + q_{B|B} = 1$ & $q_{A|A} + q_{B|A} = 1$ (conditional transition probabilities for constant switching); and $q_{A|A_v} + q_{B|A_v} = 1$, $q_{A|B_v} + q_{B|B_v} = 1$ for $v = 1$ and 2 (conditional transition probabilities for time-dependent switching). Between the first two occasions, the second-order Markov model is aggregated over the first two states in the left column; that is, $q_{A|A} = \pi_1 + \pi_5$, $q_{B|A} = \pi_2 + \pi_6$, and so on.

TABLE 3

Model Comparison Statistics for Three Models: Regime Switching With Time-Invariant Transition Probabilities, Regime Switching Model With Time-Varying Transition Probabilities; and the Second-Order Markov

Model	NC	NPAR	-2*LogL	BIC
Time-invariant transitions (LGCM3-SW2)	81	18	9918	10037
Time-varying transitions	81	30	9893	9996
Second-order Markov	81	30	9896	9999

Note. NC = number of components; NPAR = number of parameters; -2LogL = $-2 \times \log$ -likelihood; BIC = Bayesian information Criterion.

TABLE 4

A Maximum Likelihood Parameter Estimates of Three Models of Regime Switching

Model	Time-Invariant	Time-Varying	Second-Order Markov
Light regime			
Initial probability	0.48	0.47	0.47
Level <i>M</i>	1.17	1.19	1.17
Level <i>SD</i>	0.0* ^a	0.0* ^a	0.0 ^a
Slope <i>M</i>	0.22	0.18	0.20
Slope <i>SD</i>	0.19	0.17	0.18
Residual <i>SD</i>	0.56	0.55	0.53
<i>M</i> T1	1.17	1.19	1.17
<i>M</i> T2	1.39	1.37	1.37
<i>M</i> T3	1.62	1.55	1.56
<i>M</i> T4	1.84	1.73	1.76
Moderate regime			
Initial probability	0.36	0.37	0.37
Level <i>M</i>	3.36	3.43	3.40
Level <i>SD</i>	0.0* ^a	0.0* ^a	0.0* ^a
Slope <i>M</i>	0.46	0.38	0.39
Slope <i>SD</i>	0.31	0.35	0.29
Residual <i>SD</i>	1.61	1.57	1.62
<i>M</i> T1	3.37	3.43	3.31
<i>M</i> T2	3.82	3.81	3.74
<i>M</i> T3	4.28	4.19	4.17
<i>M</i> T4	4.74	4.57	4.60
Heavy regime			
Initial probability	0.16	0.16	0.16
Level <i>M</i>	10.15	10.13	10.11
Level <i>SD</i>	0.0* ^a	0.0* ^a	0.0* ^a
Slope <i>M</i>	0.0* ^a	0.0* ^a	0.0* ^a
Slope <i>SD</i>	0.0* ^a	0.0* ^a	0.0* ^a
Residual <i>SD</i>	4.83	4.83	4.83
<i>M</i> T1	10.15	10.13	10.11
<i>M</i> T2	10.15	10.13	10.11
<i>M</i> T3	10.15	10.13	10.11
<i>M</i> T4	10.15	10.13	10.11

Note. Time-invariant has the same transition matrix between consecutive occasions, Time-varying has three separate transition matrices and Second-order Markov has a 3×3 matrix for time 1 to time 2, and a 9×3 matrix subsequently (e.g., time 3 depends on status at both time 1 and time 2)

^aFixed parameter.

Table 5

Estimated Transition Probabilities (and 95% Maximum-Likelihood Confidence Intervals) for Three Regime Switching Models

First-Order Markov				Second-Order Markov			
Time-Invariant Transition Probabilities				Time 1–Time2			
	Light	Moderate	Heavy		Light	Moderate	Heavy
Light	.64 [.58, .69]	.29 [.23, .36]	.07 [.03, .11]	L	.66 [.56, .75]	.25 [.15, .35]	.09 [.04, .16]
Moderate	.12 [.07, .18]	.71 [.63, .78]	.17 [.12, .23]	M	.11 [.06, .21]	.72 [.61, .79]	.17 [.10, .25]
Heavy	.05 [.00, .11]	.26 [.18, .36]	.69 [.60, .77]	H	.07 [.01, .18]	.24 [.16, .36]	.69 [.57, .78]
Time-Varying Transition Probabilities				Time 1 & 2 (2 & 3) to Time 3 (4)			
	Time 2				Light at Time 2		
Time 1	Light	Moderate	Heavy		Light	Moderate	Heavy
Light	0.67 [.57, .77]	0.24 [.14, .35]	0.09 [.03, .16]	LL	.55 [.44, .64]	.42 [.31, .54]	.03 [.00, .10]
Moderate	0.13 [.02, .26]	0.63 [.48, .79]	0.23 [.14, .35]	LM	.33 [.17, .50]	.40 [.18, .61]	.27 [.09, .46]
Heavy	0.14 [.03, .30]	0.33 [.15, .54]	0.53 [.34, .72]	LH	.17 [.00, .51]	.02 [.00, .48]	.81 [.33, .100]
	Time 3				Moderate at Time 2		
Time 2	Light	Moderate	Heavy		Light	Moderate	Heavy
Light	.60 [.50, .71]	.31 [.19, .43]	.09 [.02, .17]	ML	.85 [.43, 1.00]	.11 [.00, .53]	.04 [.00, .37]
Moderate	.14 [.06, .24]	.66 [.53, .78]	.20 [.11, .30]	MM	.01 [.00, .11]	.83 [.69, .93]	.16 [.07, .28]
Heavy	.00 [.00, .07]	.25 [.10, .39]	.75 [.60, .90]	MH	.01 [.00, .29]	.52 [.19, .80]	.47 [.15, .78]
	Time 4				Moderate at Time 2		
Time 3	Light	Moderate	Heavy		Light	Moderate	Heavy
Light	.52 [.38, .66]	.45 [.31, .59]	.03 [.00, .10]	HL	.58 [.19, 1.00]	.22 [.00, .60]	.20 [.00, .43]
Moderate	.06 [.00, .16]	.86 [.73, .97]	.09 [.01, .18]	HM	.00 [.00, .29]	.93 [.58, 1.00]	.07 [.00, .34]
Heavy	.03 [.00, .13]	.25 [.09, .42]	.71 [.56, .86]	HH	.03 [.00, .13]	.19 [.05, .37]	.77 [.61, .92]

Note. Each 3×3 panel contains transition probability matrices for switching from time in the row to the time in the column. The top left shows time-invariant regime switching, with time-invariant transition probabilities. The three left panels below it are the estimates for the time-varying transition probabilities model. Estimates of the second-order Markov model are shown in the right four panels. The first three are estimates for Time 1 to Time 2, and the lower three comprise the 9×3 second-order transition, matrix probabilities.