Published in final edited form as:

Stat Med. 2014 October 15; 33(23): 4116–4134. doi:10.1002/sim.6220.

Latent time-varying factors in longitudinal analysis: a linear mixed hidden Markov model for heart rates

Francesco Lagona^{a,b,*}, Dmitri Jdanov^b, and Maria Shkolnikova^b

^aUniversity of Roma Tre, Rome, Italy

^bMax Planck Institute for Demographic Research, Rostock, Germany

Abstract

Longitudinal data are often segmented by unobserved time-varying factors, which introduce latent heterogeneity at the observation level, in addition to heterogeneity across subjects. We account for this latent structure by a linear mixed hidden Markov model. It integrates subject-specific random effects and Markovian sequences of time-varying effects in the linear predictor. We propose an expectation—maximization algorithm for maximum likelihood estimation, based on data augmentation. It reduces to the iterative maximization of the expected value of a complete likelihood function, derived from an augmented dataset with case weights, alternated with weights updating. In a case study of the Survey on Stress Aging and Health in Russia, the model is exploited to estimate the influence of the observed covariates under unobserved time-varying factors, which affect the cardiovascular activity of each subject during the observation period.

Keywords

heart rate; hidden Markov model; EM algorithm; linear mixed model; longitudinal data

1. Introduction

Longitudinal studies often involve the estimation of the effects of time-constant covariates under time-varying conditions. When, for example, a biomarker is repeatedly observed over time in a sample of subjects, longitudinal profiles are often segmented by the levels of a time-varying factor, which reflect the different states under which each subject is observed. In these cases, there is interest in assessing whether the fixed effect of a time-constant covariate, for example, gender differences, varies with the state of the subject. When the state of a subject is available through a time-varying factor, covariate effects under different conditions can be efficiently estimated by specifying a regression model that includes terms of interaction between the covariates of interest and this factor. When this information is not available, a natural strategy is replacing the time-varying latent factor by a stochastic

Copyright © 2014 John Wiley & Sons, Ltd.

*Correspondence to: Francesco Lagona, DIPES University of Roma Tre, via G. Chiabrera 199, 00165 Rome, Italy.

†lagona@uniroma3.it

Supporting information: Additional supporting information may be found in the online version of this article at the publisher's web site.

process. Hidden Markov models (HMMs) provide a feasible option in this context, by introducing a latent Markov chain that segments longitudinal profiles. Under an HMM, more precisely, repeated outcomes are assumed conditionally independent given the states generated by a latent Markov chain, and the conditional distribution of each outcome depends on the covariates through state-specific fixed effects. As a result, fixed effects are allowed to vary over time, and simultaneously, repeated outcomes are temporally correlated, with an auto-correlation function that depends on the transition probabilities of the Markov chain [1]. The intuitively appealing interpretation of the parameters has motivated HMMs across a variety of case studies. Recent applications of HMMs in a longitudinal context include studies of drinking behavior [2], characterized by alternating latent periods of heavy drinking and abstinence, and studies of brain activity, characterized by different latent sleep stages [3].

Hidden Markov models capture a special source of longitudinal correlation, because of the time-varying heterogeneity of repeated outcomes that are influenced by alternating latent conditions. Recent approaches to longitudinal analysis integrate an HMM with time-constant random effects to account for an additional component of latent heterogeneity, because of unobserved characteristics of the subject. Mixed HMMs are a popular option in this context ([4]; see [5] and [6] for recent reviews). Under a mixed HMM, the conditional distribution of each outcome depends on both subject-specific random effects and time-varying fixed effects that vary according to the states generated by a latent Markov chain. On one side, mixed HMMs extend HMMs by relaxing the conditional independence assumption among repeated outcomes, because observations taken on the same subject share the same random effects. On the other side, mixed HMMs extend mixed effect models, by relaxing the assumption of time-constant fixed effects.

Motivated by a longitudinal study of heart rates, we contribute to this strand of literature by introducing a linear mixed HMM for normal longitudinal data. Independently of the baseline conditions of a subject, the cardiovascular system is characterized by periods of high and low heart rates, associated with specific balances between sympathetic and parasympathetic activities of the autonomic nervous system [7, 8]. Recent empirical evidences [9] suggest that heart variability can be summarized by a small number of states of the heart, associated with these balances. We are interested in estimating the effects of a number of time-constant covariates under different states of the heart, simultaneously accounting for unobserved baseline conditions. This allows to test, for example, whether gender differences are time-constant or they vary across different heart states. Measuring the current balance between activities of the nervous system is extremely difficult and practically impossible in studies that involve hundreds of subjects under normal life conditions. We hence exploit a latent Markov chain to account for latent heart states and propose a mixed linear HMM as a feasible method to estimate time-varying effects on heart activity.

The focus of the literature on mixed HMMs is often on the computational issues that arise in parameter estimation under both Bayesian and likelihood-based settings. Although efficient algorithms are available for the estimation of HMMs [10], the extension of these procedures in the context of mixed HMMs is computationally challenging, because of analytically intractable integrals. Under a Bayesian setting, MCMC methods require ad hoc strategies to

draw samples for the posterior distribution [11,12]. Taking a frequentist approach, quasi-Newton algorithms have been suggested to locate the maximum of the likelihood of mixed HMMs, suitably approximated by Gaussian quadrature methods [4]. These methods are however rarely used, as they require special strategies to avoid numerical instability. More recent proposals are based on EM algorithms, because mixed HMMs can be viewed as missing data models, where random effects and Markovian states are the nonobserved data. These algorithms are based on the iterative maximization of the expected value of a complete-data likelihood function, evaluated with respect to the conditional distribution of the nonobserved data given the observed outcomes. In the context of mixed HMMs, unfortunately, this expected value is often not available in closed form. As a result, proposals are based on stochastic expectation–maximization (EM) algorithms, which replace this expectation by a Monte Carlo average [13].

These difficulties have somehow limited the widespread use of mixed HMMs. If the outcome is normally distributed, however, the likelihood function of a mixed HMM involves integrals that can be computed exactly. Although this is mentioned in the literature [4,5], explicit algorithms for the estimation of normal mixed HMMs have not been illustrated so far. We fill this gap by proposing an EM algorithm that relies on a data augmentation approach to estimation and avoids stochastic approximations. It reduces to the iterative maximization of the expected value of a complete likelihood function, derived from an augmented dataset with case weights (M step), alternated with weights updating (E step). The E step relies on closed-form weights updating equations; the M step involves a closed-form update equation of the fixed effects and a Newton-type iteration that updates the parameters of the random effects distribution.

The rest of the paper is organized as follows. In Section 2, we describe the heart rates data that motivated this paper. The proposed linear mixed HMM is introduced in Section 3. Section 4 is devoted to the EM algorithm exploited for maximum likelihood estimation (additional details are postponed to the Appendix). In Section 5, we describe the results obtained by the proposed approach in the heart rates case study. Finally, Section 6 summarizes relevant points of discussion. The R code used for analysis is available in the Supporting information.

2. Longitudinal heart activity data from the survey on stress aging and health in Russia

Heart activity is traditionally measured by parameters that summarize an electrocardiogram (ECG) in either the frequency or the time domain. In the time domain, an ECG is segmented according to normal and abnormal QRS complexes, which are periodic combinations of ECG deflections. A segmented ECG is exploited to compute heartbeats intervals (also known as normal RR intervals), that is, the times between peaks of successive normal QRS complexes. Time-domain measures of heart activity are typically parameters of the distribution of these intervals. For example, the average heartbeat interval has been widely exploited to establish that elevated heart rate is a risk factor of mortality and morbidity because of cardiovascular and other chronic diseases. More recently, the medical literature has focused on the variability of heartbeat intervals, by exploiting measures that include the

standard deviation of the distribution of the intervals, the root mean square of successive intervals differences, and the proportion of intervals greater than a specific threshold.

Time-domain measures of heart activity have been often regressed on subject-specific covariates to establish population-based norms of heart rate, without which classifying ‘abnormal’ heart activity remains difficult [14]. It has been demonstrated that the variation of heart activity can be partially explained by gender and age differences [15,16] and by a number of biomedical characteristics that reflect the health status of a subject [17]. These studies rely on cross-sectional designs, where heart rate is recorded at a given point in time. Comparability and interpretation of cross-sectional heart activity studies is however complicated by (often unobserved) factors that influence the cardiovascular system during the day and across subjects. It has been therefore stressed the need of long-term ECG data, covering at least 24 h and based on large sample sizes [18]. This has motivated a growing interest in covariate effects on heart rates under time-varying conditions of the subject.

Independently of baseline conditions, longitudinal profiles of heartbeat intervals are characterized by periods of high heart activity (HHA) alternated with periods of low heart activity (LHA). It is well known that these periods are associated with distinct balances between sympathetic and parasympathetic activities of the autonomic nervous system [7,8]. The sympathetic and parasympathetic systems are two major divisions of the autonomic nervous system. The sympathetic system is typically responsible of actions requiring quick responses, while the parasympathetic division is related to actions that do not require immediate reaction. Sympathetic and parasympathetic systems work as antagonists in their effect on the heart via chemical mediators. The relative roles of the two systems can be determined by blocking their activity with a pharmacologic antagonist. Measuring this balance during a long-term follow-up is however technically difficult. It is practically impossible in large longitudinal studies that involve hundreds of subjects under normal life conditions. In these cases, the balance between sympathetic and parasympathetic systems takes the form of a latent factor that segments heart activity, and the lack of information about these time-varying conditions of the subject complicates the estimation of covariate effects in longitudinal heart studies. We therefore propose a mixed HMMs to estimate the effect of the available covariates under different balances between sympathetic and parasympathetic activities.

Our case study is drawn from the recent Survey on Stress Aging and Health in Russia [19], which provides one of the few collections of long-term cardiovascular data, observed on a large sample of subjects during a period of 24 h, under normal life conditions. The data are in the form of 1274 subject-specific time series of hourly averages of heartbeat intervals. Details on the Holter device used to collect the data and the procedures of preliminary data preparation are given elsewhere [19]. Figure 1 displays the hourly distribution of the heartbeat intervals in the whole sample. An average interval of x milliseconds (msec) indicates $60 \times 1000/x$ heartbeats per minute (heart rate).

The study includes a number of baseline covariates, measured at the beginning of the follow-up. These covariates include gender, age (in years), and information on cardiovascular events prior to measurement (e.g., experience of myocardial infarction and

stroke). Self-reported covariates include smoking habits as well as the established SF-36 battery that (among other measures) provides a continuous score for the evaluation of physical functioning [20]. Relevant biomedical characteristics include systolic blood pressure, a principal hemodynamic marker [21], and hand grip strength, which indicates individual frailty [22]. To account for atherosclerosis and low-grade inflammation, the study further includes total cholesterol, high sensitivity C-reactive protein, and interleukin-6 as binary factors with clinically established cutoff points [23,24].

About 47% of the subjects are male, with an age ranging between 55 and 90 years. About 55% of the subjects never smoked, 24% was an ex-smoker, and the remaining part of the sample was a current smoker. Percentages of subjects who reported either a myocardial infarction or a stroke were, respectively, 5% and 7%. About 40% was found with a high level (>240 mg/dL) of total cholesterol, 30% with high levels (>3 mg/dL) of C-reactive protein, and about 2% with high levels (>2.0 pg/mL for men and >2.4 pg/mL for women) of interleukin-6. The interquartile ranges of the remaining covariates were 22 – 38 kg (hand grip), 125 – 158 mmHg (systolic blood pressure), and 65–95 (SF-36 physical functioning).

3. A linear mixed hidden Markov model for normal longitudinal data

The data under scrutiny are in the form of n longitudinal profiles $y_i = (y_{it}, t = 0, 1, \dots, T), i = 1, \dots, n$, that include $T + 1$ repeated observations. Linear mixed models (LMMs) provide a general framework to analyze normal longitudinal data. In its simplest form, a LMM extends a linear regression model by introducing random effects at the subject level and can be expressed as follows:

$$\begin{aligned} y_i &= X_i \beta + Z_i b_i + \varepsilon_i \quad i=1 \dots n \\ b_i &\sim N(\mathbf{0}, \mathbf{V}_b) \\ \varepsilon_i &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned} \quad (1)$$

where X_i and Z_i are the covariate and the design matrix, respectively, β is a vector of time-constant fixed effects, b_i is a vector of subject-specific random effects with covariance matrix $\mathbf{V}_b$, and finally, ε_i is a vector of independent errors with variance σ^2 . Under this setting, the outcomes are assumed conditionally independent given the random effects, while their marginal distribution is multivariate normal with mean $X_i \beta$ and variance-covariance matrix $Z_i \mathbf{V}_b Z_i^T + \sigma^2 \mathbf{I}$. This specification assumes one source of longitudinal correlation, because of repeated outcomes that share the same random effects. It can be extended by assuming that the errors are auto-correlated, say $\varepsilon_i \sim N(\mathbf{0}, \mathbf{R})$. In this case, the marginal variance-covariance matrix of the outcomes is given by $Z_i \mathbf{V}_b Z_i^T + \mathbf{R}$ and distinguishes two sources of correlations: a random-effects component, $Z_i \mathbf{V}_b Z_i^T$ and a residual component, $\mathbf{R}$. In real case studies, a few random effects are typically introduced, and $\mathbf{V}_b$ is specified as an unstructured matrix of small size. Matrix $\mathbf{R}$ is instead often structured as the variance-covariance matrix of a Gaussian process, depending on a small number of parameters. Popular is the use of auto-regressive moving-average (ARMA) processes, whose variance-covariance matrix $\mathbf{R}$ depends on auto-regressive (AR) and moving-average parameters. ARMA processes can be often interpreted as AR processes corrupted by measurement error

[25], and in this case, the residual covariance matrix can be structured as the sum $\sigma^2 \mathbf{I} + \mathbf{R}$, where σ^2 is the variance of the measurement error and $\mathbf{R}$ is the variance–covariance matrix of an AR process.

The proposed mixed linear HMM is characterized by an alternative specification of the marginal correlation structure of the data, which distinguishes a random-effects component and a Markov chain component. The random-effects component of heterogeneity is due to repeated outcomes that share the same random effects. The Markov chain component of heterogeneity is due to unobserved time-varying conditions that influence the outcome distribution independently of random effects and are modelled by a latent Markov chain. To illustrate, we assume that such conditions take the form of a time-varying, unobserved factor with K levels. The value taken by this factor for subject i at time t can be conveniently described by a multinomial variable $\xi_{it} = (\xi_{it1} \dots \xi_{itK})$ with one trial and K classes, where $\xi_{itk} = 1$ if the factor takes level k for subject i at time t and 0 otherwise. Under this setting, each subject is associated with a multinomial time series $\xi_i = (\xi_{i0}, \dots, \xi_{iT})$, which indicates the temporal evolution of the latent factor during the follow-up. By interpreting unobserved factor levels as latent classes, ξ_i can be viewed as a segmentation of the longitudinal profile y_i , where each binary component ξ_{itk} indicates the class membership of observation y_{it} . We assume that the n segmentations $\xi_1 \dots \xi_n$ are independent multivariate samples, drawn from the joint distribution of a homogeneous Markov chain with K states. This distribution is assumed known up to a vector π of parameters that includes K initial probabilities, $\pi_k = P(\xi_{i0k} = 1)$, and $K \times K$ transition probabilities, namely, $\pi_{k|h} = P(\xi_{itk} = 1 \mid \xi_{i,t-1,h} = 1)$,

$$\begin{aligned}
 \xi_i &\sim p(\xi; \pi) \\
 p(\xi; \pi) &= \prod_{k=1}^K \pi_k^{\xi_{i0k}} \prod_{t=1}^T \prod_{h=1}^K \prod_{k=1}^K \pi_{k|h}^{\xi_{i,t-1,h} \xi_{itk}} \\
 &\quad \sum_{k=1}^K \pi_k = 1 \\
 &\quad \sum_{k=1}^K \pi_{k|h} = 1, \quad h=1 \dots K
 \end{aligned} \tag{2}$$

We extend the LMM (1) by relaxing the assumption of time-constant covariates effects and assuming that these effects vary over time according to the states visited by a latent Markov chain. Specifically, each latent state is associated with a vector of effects, say $\beta_k, k=1 \dots K$, which indicates the effects of the covariates under the k -th latent condition. Further, the effects of the covariates at time t are defined as

$$\beta(\xi_{it}) = \sum_{k=1}^K \xi_{itk} \beta_k, \quad i=1 \dots n, \quad t=0 \dots T$$

The proposed mixed HMM is obtained by assuming that the outcomes $y_{i0} \dots y_{iT}$ are conditionally independent given the random effects b_i and the segmentation ξ_i and can be expressed as follows:

$$\begin{aligned}
y_{it} &= \mathbf{x}_{it}^T \boldsymbol{\beta}(\boldsymbol{\xi}_{it}) + \mathbf{z}_{it}^T \mathbf{b}_i + \varepsilon_{it} \\
\mathbf{b}_i &\sim N(0, \mathbf{V}_b) \\
\boldsymbol{\xi}_i &\sim p(\boldsymbol{\xi}, \boldsymbol{\pi}) \\
\varepsilon_{it} &\sim N(0, \sigma^2) \text{ i.i.d.}
\end{aligned} \tag{3}$$

where $\mathbf{x}_{it}^T$ includes the observed covariates values for subject i at time t and $\mathbf{z}_{it}^T$ is the row vector of a design matrix $\mathbf{Z}_i$. At the subject level, therefore, the observations share a common random-effects component, $\mathbf{Z}_i \mathbf{b}_i$, such as in a traditional LMM. At the observation level, the data depend on state-dependent regression coefficients, which indicate the influence of the observed covariates under the k -th latent condition of the subject at each time t . The differences between coefficients under two states $h \neq k$, say $\boldsymbol{\beta}_h - \boldsymbol{\beta}_k$, can be therefore interpreted as interaction effects between the observed covariates and an unobserved, time-varying factor with K levels, which evolves according to the transition probabilities matrix of a Markov chain.

The model depends on a vector $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\eta})$ of parameters that includes the probabilities vector $\boldsymbol{\pi}$ of the latent Markov chain, the vector of state-dependent regression parameters, say $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \dots \boldsymbol{\beta}_K)$, and finally, a vector $\boldsymbol{\eta}$ that includes the residual variance σ^2 and the entries of the variance-covariance matrix $\mathbf{V}_b$. Integrating with respect to the random effects, we obtain the conditional distribution of each profile $\mathbf{y}_i$, given the segmentation $\boldsymbol{\xi}_i$, say

$$\begin{aligned}
f(\mathbf{y}_i | \boldsymbol{\xi}_i; \boldsymbol{\beta}, \boldsymbol{\eta}) &= \int \prod_{t=0}^T \prod_{k=1}^K \phi(y_{it}; \mathbf{x}_{it}^T \boldsymbol{\beta}_k + \mathbf{z}_{it}^T \mathbf{b}_i, \sigma^2)^{\xi_{itk}} \phi(\mathbf{b}_i; \mathbf{0}, \mathbf{V}_b) d\mathbf{b} \\
&= (2\pi)^{-(T+1)/2} |\mathbf{V}_i(\boldsymbol{\eta})|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{e}(\boldsymbol{\xi}_i)^T \mathbf{V}_i(\boldsymbol{\eta})^{-1} \mathbf{e}(\boldsymbol{\xi}_i)\right)
\end{aligned} \tag{4}$$

where $\phi(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates the normal density with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, $\mathbf{V}_i(\boldsymbol{\eta}) = \mathbf{z}_i \mathbf{V}_b \mathbf{z}_i^T + \sigma^2 \mathbf{I}$, and the entries of the residuals vector $\mathbf{e}(\boldsymbol{\xi}_i) = (e(\xi_{i0}), \dots, e(\xi_{iT}))$ depend on the latent states visited by the Markov chain during the follow-up, namely,

$$e(\xi_{it}) = y_{it} - \mathbf{x}_{it}^T \boldsymbol{\beta}_k, \quad i=1 \dots n, t=0, \dots, T$$

As a result, the marginal distribution of each profile $\mathbf{y}_i$ takes the form of a mixture of segmentation-specific components, namely,

$$f(\mathbf{y}_i; \boldsymbol{\theta}) = \sum_{\boldsymbol{\xi} \in \Xi} (p(\boldsymbol{\xi}; \boldsymbol{\pi}) f(\mathbf{y}_i | \boldsymbol{\xi}; \boldsymbol{\beta}, \boldsymbol{\eta})) \tag{5}$$

where Ξ is the set of the K^{T+1} segmentations of $T+1$ observation times. Under this setting, the marginal variance-covariance matrix of each profile is given by

$$\mathbb{E}(\mathbf{y}_i \mathbf{y}_i^T) - \mathbb{E}(\mathbf{y}_i) \mathbb{E}(\mathbf{y}_i^T) = \mathbf{V}_i(\boldsymbol{\eta}) + \mathbf{V}_i(\boldsymbol{\pi}, \boldsymbol{\beta})$$

where $V_i(\boldsymbol{\eta})$ is the random-effects correlation component and $V_i(\boldsymbol{\pi}, \boldsymbol{\beta})$ is the additional contribution of the latent Markov chain to the data correlation structure. This convenient structure of the variance–covariance matrix is the result of the independence assumption between the latent Markov chain and the random effects introduced at the subject level. Specifically, the (t, t') -th generic entry of $V_i(\boldsymbol{\pi}, \boldsymbol{\beta})$ is given by

$$v_{i,t,t'} = \begin{cases} \sum_k^{1\dots K} (\mathbf{x}_{it}^T \boldsymbol{\beta}_k)^2 \pi_{kt} - \left(\sum_{k=1}^K \mathbf{x}_{it}^T \boldsymbol{\beta}_k \pi_{kt} \right)^2 & t=t' \\ \sum_{h,k}^{1\dots K} (\mathbf{x}_{it}^T \boldsymbol{\beta}_h) (\mathbf{x}_{it'}^T \boldsymbol{\beta}_k) \pi_{k|h}^{(t-t')} - \left(\sum_{h=1}^K \mathbf{x}_{it}^T \boldsymbol{\beta}_h \pi_{ht} \right) \left(\sum_{h=1}^K \mathbf{x}_{it'}^T \boldsymbol{\beta}_h \pi_{ht} \right) & t \neq t' \end{cases} \quad (7)$$

where $\pi_{k|h}^{(m)}$ is the m -steps transition probability between states h and k , while π_{kt} is the probability of being in state k at time t .

When $K = 1$, the Markov chain (2) collapses into a deterministic process of ones, that is, $P(\xi_{it1} = 1) = 1, i = 1 \dots n, t = 0 \dots T + 1$, the entries of $V_i(\boldsymbol{\pi}, \boldsymbol{\beta})$ reduce to zero, and the state-dependent regression coefficients $\boldsymbol{\beta}_k$ reduce to the time-constant fixed effects $\boldsymbol{\beta}$ of a LMM with independent regression errors, say

$$\begin{aligned} \mathbf{y}_i | \mathbf{b}_i &\sim \prod_{t=0}^T \varphi(\mathbf{y}_{it}; \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i, \sigma^2) \\ \mathbf{b}_i &\sim \varphi(\mathbf{b}; \mathbf{0}, \mathbf{V}_b) \end{aligned} \quad (8)$$

Under this setting, a LMM can be viewed as a linear mixed HMM with one state.

An interesting limiting case of the model is obtained when the transition probabilities matrix of the Markov chain reduces to an identity $K \times K$ matrix. Under this setting, the state generated by the Markov chain at $t = 0$ is an absorbing state, and each subject is observed under the same state during the whole follow-up. In this case, model (5) collapses into a mixture of LMMs [26]. Recently employed in gene-expression studies, mixtures of LMMs enhance the flexibility of LMMs by allowing for latent classes that cluster subjects according to different sub-populations. In case studies of longitudinal cardiovascular activity, however, time-constant latent classes are too restrictive, as they ignore the temporal evolution of the balance between sympathetic and parasympathetic activities.

A second limiting case is obtained when all the rows of the Markov chain are all equal to the initial probabilities distribution. In this case, the conditional probabilities of the chain do not depend on the previous state, and the Markov process collapses into a sequence of independent multinomial random variables. Under this setting, the model reduces to a multilevel latent class model with random effects at the subject level and time-varying latent classes that operate at the observation level [27,28]. Assuming transition probabilities that do not depend on the previous state can be useful in settings that involve occasional environmental factors, occurring independently during the observation period. It is however unreasonable in longitudinal heart rates studies, where the outcome is influenced by neurophysiological conditions that are correlated in time.

4. Maximum likelihood estimation

Assuming independence among subjects, the parameters of the model can be estimated by maximizing the following likelihood function

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta) = \prod_{i=1}^n \sum_{\xi \in \Xi} p(\xi; \pi) f(y_i | \xi; \beta, \eta) \quad (9)$$

We maximize this likelihood by using an EM algorithm. The R program is available as a Supporting information. The algorithm is based on a complete-data likelihood function, say

$$L_c(\theta) = \prod_{i=1}^n f(y_i | \xi_i; \beta, \eta) p(\xi_i; \pi) \quad (10)$$

obtained by treating the segmentations ξ_i as multivariate missing values. The algorithm alternates an expectation step (E step) and a maximization step (M step), generating a sequence of estimates $\theta^{(s)}$, $s = 1, 2, \dots$ that converge to a local maximum point of the likelihood function. We remark that (10) depends on the latent states of the chain and not on the values of the random intercept. On the contrary, EM algorithms for mixed HMMs typically rely on complete-data likelihood functions that depends on both the latent states and the values of the random effects [5,13]. Conditioning on both latent states and random effects enhances the numerical tractability of the M step of these algorithms. It is however not necessary under the proposed linear mixed HMM, as the complete-data likelihood function (10) allows for a tractable M step.

4.1. Expectation step

Given the estimate $\theta^{(s-1)}$, obtained at the end of step $s-1$, the E step evaluates the expected value of the complete log-likelihood

$$\begin{aligned} l_c(\theta) = \log L_c(\theta) &= \sum_{i=1}^n \log p(\xi_i; \pi) + \sum_{i=1}^n \log f(y_i | \xi_i; \beta, \eta) \\ &= l_c(\pi) + l_c(\beta, \eta) \end{aligned} \quad (11)$$

with respect to the conditional distribution of the latent segmentations $\xi = (\xi_1 \dots \xi_n)$, given the observed values $\mathbf{y} = (\mathbf{y}_1 \dots \mathbf{y}_n)$, computed at $\theta = \theta^{(s-1)}$, say

$$Q\left(\theta; \hat{\theta}^{(s-1)}\right) = \mathbb{E}(l_c(\theta) | \mathbf{y}; \hat{\theta}^{(s-1)}) = \mathbb{E}(l_c(\pi | \mathbf{y}); \hat{\theta}^{(s-1)}) + \mathbb{E}\left(l_c(\beta, \theta) | \mathbf{y}; \hat{\theta}^{(s-1)}\right) \quad (12)$$

The first expectation in the right-hand side of (12) is given by

$$\mathbb{E}\left(l_c(\pi) | \mathbf{y}; \hat{\theta}^{(s-1)}\right) = \sum_{i=1}^n \sum_{k=1}^K \hat{\pi}_{i0k}^{(s-1)} \log \pi_k + \sum_{i=1}^n \sum_{t=1}^T \sum_{h=1}^K \sum_{k=1}^K \hat{\pi}_{i,t-1,t,hk}^{(s-1)} \log \pi_{k|h} \quad (13)$$

where

$$\hat{\pi}_{i0k}^{(s-1)} = \mathbb{E} \left(\xi_{i0k} | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) = P \left(\xi_{i0k} = 1 | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) \quad (14)$$

$$\hat{\pi}_{i,t-1,t,hk}^{(s-1)} = \mathbb{E} \left(\xi_{i,t-1,h} \xi_{itk} | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) = P \left(\xi_{itk} = 1, \xi_{i,t-1,h} = 1 | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) \quad (15)$$

are, respectively, the posterior probabilities of the states at time $t = 0$ and the bivariate posterior probabilities of the states at two successive times.

The second expectation in the right-hand side of (12) is given by

$$\begin{aligned} \mathbb{E} \left(l_c \left(\boldsymbol{\beta}, \boldsymbol{\eta} | \mathbf{y}; \hat{\boldsymbol{\theta}}^{(s-1)} \right) \right) &= -\frac{n(T+1)}{2} \log(2\pi) - \frac{n}{2} \log |V(\boldsymbol{\eta})| - \frac{1}{2} \sum_{i=1}^n \mathbb{E} \left(\left(\mathbf{e}(\boldsymbol{\xi}_i)^T V(\boldsymbol{\eta})^{-1} \mathbf{e}(\boldsymbol{\xi}_i) \right) | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) \\ &= -\frac{n(T+1)}{2} \log(2\pi) - \frac{n}{2} \log |V(\boldsymbol{\eta})| - \frac{1}{2} \sum_{i=1}^n \sum_{h=1}^K \sum_{k=1}^K \sum_{t=0}^T \sum_{\tau=0}^T e_{ith} \hat{\pi}_{it\tau hk}^{(s-1)} v^{(t\tau)}(\boldsymbol{\eta}) e_{i\tau k} \end{aligned} \quad (16)$$

where

$$e_{itk} = y_{it} - \mathbf{x}_{it}^T \boldsymbol{\beta}_k, \quad i=1 \dots n, t=0 \dots T, k=1 \dots K$$

indicates the residual of subject i at time t in state k , while $v^{(t\tau)}(\boldsymbol{\eta})$ indicates the (t, τ) th entry of the inverse covariance matrix $V^{-1}(\boldsymbol{\eta})$ and, finally,

$$\hat{\pi}_{it\tau hk}^{(s-1)} = \mathbb{E} \left(\xi_{i\tau h} \xi_{itk} | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) = P \left(\xi_{itk} = 1, \xi_{i\tau h} = 1 | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) \quad (17)$$

indicates the subject-specific posterior bivariate probabilities of two states h and k at two different times t and τ .

In summary, the expectation step reduces to the computation of the posterior probabilities (14), (15), and (17). The Appendix describes a forward-backward (FB) recursion to compute the univariate posterior probabilities

$$\hat{\pi}_{itk}^{(s-1)} = \mathbb{E} \left(\xi_{itk} | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) = P \left(\xi_{itk} = 1 | \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)} \right) \quad (18)$$

and the bivariate probabilities (15). At $t = 0$, (18) is the required initial posterior probabilities (14). Furthermore, the univariate and bivariate probabilities (18) and (15) can be exploited to compute the posterior transition probabilities

$$\hat{\pi}_{itk|t-1,h}^{(s-1)} = P\left(\xi_{itk}=1|\xi_{i,t-1,h}=1, \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(s-1)}\right) = \frac{\hat{\pi}_{i,t-1,t,hk}^{(s-1)}}{\hat{\pi}_{i,t-1,h}^{(s-1)}} \quad (19)$$

The univariate posterior probabilities (18) and the posterior transition probabilities (19) specify a (non-homogeneous) Markov chain for each subject i . The bivariate probabilities (17) can be hence obtained by exploiting the Chapman–Kolmogorov equations for nonhomogeneous Markov chains.

4.2. Maximization step

The M step involves the maximization of the expected complete log-likelihood (12). This is the sum of two components, say

$$Q\left(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}^{(s-1)}\right) = Q\left(\boldsymbol{\pi}; \hat{\boldsymbol{\theta}}^{(s-1)}\right) + Q\left(\boldsymbol{\beta}, \boldsymbol{\eta}; \hat{\boldsymbol{\theta}}^{(s-1)}\right)$$

which depend on independent parameters and can therefore be maximized separately. Maximization of the first component with respect to $\boldsymbol{\pi}$ provides the closed-form solutions

$$\begin{aligned} \hat{\pi}_k^{(s)} &= \frac{\sum_{i=1}^n \hat{\pi}_{i0k}^{(s-1)}}{\sum_{k=1}^K \sum_{i=1}^n \hat{\pi}_{i0k}^{(s-1)}} \\ \hat{\pi}_{k|h}^{(s)} &= \frac{\sum_{i=1}^n \sum_{t=1}^T \hat{\pi}_{i,t-1,t,hk}^{(s-1)}}{\sum_{k=1}^K \sum_{i=1}^n \sum_{t=1}^T \hat{\pi}_{i,t-1,t,hk}^{(s-1)}} \end{aligned}$$

Maximization of the second component of (12) is facilitated by data augmentation. Precisely, we append K copies of $\mathbf{y}_i$ in a $(T+1)K \times 1$ vector $\mathbf{y}_i^{\text{aug}}$ and define an augmented residual vector as

$$\mathbf{e}_i^{\text{aug}} = \mathbf{y}_i^{\text{aug}} - (\mathbf{I} \otimes \mathbf{X}_i) \boldsymbol{\beta}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \dots \boldsymbol{\beta}_K)$ and $\otimes$ indicates the Kronecker product. We additionally define K^2 weights matrices

$$\mathbf{W}_{hk}^{(s-1)} = \left(\hat{\pi}_{it\tau hk}^{(s-1)}, t, \tau = 0, \dots, T \right)$$

which include the $(T+1)$ posterior bivariate probabilities (17), computed at the previous E step. We finally introduce an augmented, blocks matrix, where each block is a copy of $\mathbf{V}^{-1}(\boldsymbol{\eta})$, whose entries are weighted by the posterior bivariate probabilities (17), namely,

$$\sum_i^{\text{aug}}(\boldsymbol{\eta}) = \left(W_{hk}^{(s-1)} \odot V(\boldsymbol{\eta})^{-1}, h, k=1 \dots K \right) \quad (20)$$

where $\odot$ is the element-wise product operator (Hadamard product) between two matrices of equal dimension. Under this setting, the second component in the right-hand side of (12) takes the following matrix form

$$Q\left(\boldsymbol{\beta}, \boldsymbol{\eta}; \hat{\boldsymbol{\theta}}^{(s-1)}\right) = -\frac{n(T+1)}{2} \log(2\pi) - \frac{n}{2} \log|V(\boldsymbol{\eta})| - \frac{1}{2} \sum_{i=1}^n (\mathbf{e}_i^{\text{aug}})^T \sum_i^{\text{aug}}(\boldsymbol{\eta}) \mathbf{e}_i^{\text{aug}}. \quad (21)$$

We first maximize (21), evaluated at $\boldsymbol{\eta} = \boldsymbol{\eta}^{(s-1)}$, obtaining the closed-form solution

$$\hat{\boldsymbol{\beta}}^{(s)} = \left(\sum_{i=1}^n \mathbf{X}_i^T \left(\sum_i^{\text{aug}}(\hat{\boldsymbol{\eta}}^{(s-1)}) \right)^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i^T \left(\sum_i^{\text{aug}}(\hat{\boldsymbol{\eta}}^{(s-1)}) \right)^{-1} \mathbf{y}_i^{\text{aug}}.$$

Secondly, we maximize (21), evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(s)}$, with respect to $\boldsymbol{\eta}$, and obtain the updated estimate $\boldsymbol{\eta}^{(s)}$ by exploiting a traditional Newton-type algorithm.

4.3. Missing values

We have illustrated an EM algorithm under the assumption of complete-data information, as the study that motivated this paper included complete longitudinal profiles. The algorithm can be however easily extended to a setting with missing (at random) values, by slightly modifying the E step. With missing data, the E step reduces to evaluate the expected values of $l_c(\boldsymbol{\pi})$ and $l_c(\boldsymbol{\beta}, \boldsymbol{\eta})$, with respect to the conditional distribution of the missing data (latent states and missing outcomes) given the observed data, computed at the estimate $\boldsymbol{\theta}^{(s-1)}$ available at the previous step of the algorithm. On the one side, the expectation of $l_c(\boldsymbol{\pi})$ reduces to the evaluation of the posterior probabilities (18) and (15), which can be computed by using a backward–forward procedure that uses the marginal distributions of the observed data [1,29]. On the other side, we observe that $l_c(\boldsymbol{\beta}, \boldsymbol{\eta})$ depends on the outcomes through a linear combination of the sufficient statistics $\xi_{itk} y_{it}$, $\xi_{itk} y_{it}^2$ and $\xi_{ith} \xi_{it'k} y_{it} y_{it'}$. If these statistics include unobserved outcomes, we can compute the expectation of $l_c(\boldsymbol{\beta}, \boldsymbol{\eta})$ by replacing these statistics with their expectations. These expectations are easy to evaluate, by recalling the normal distribution (4) of each $\mathbf{y}_i$ given the segmentation $\boldsymbol{\xi}_i$, and take the following form:

$$\begin{aligned} \mathbb{E}(\xi_{itk} y_{it}) &= \mathbb{E}(\xi_{itk}; \boldsymbol{\theta}^{(s-1)}) \mathbb{E}(y_{it} | \xi_{itk}=1; \boldsymbol{\theta}^{(s-1)}) = \hat{\pi}_{itk}^{(s-1)} \mathbf{x}_{it}^T \boldsymbol{\beta}_k^{(s-1)} \\ \mathbb{E}(\xi_{itk} y_{it}^2) &= \mathbb{E}(\xi_{itk}; \boldsymbol{\theta}^{(s-1)}) \mathbb{E}(y_{it}^2 | \xi_{itk}=1; \boldsymbol{\theta}^{(s-1)}) = \hat{\pi}_{itk}^{(s-1)} \left(v^{(tt)}(\boldsymbol{\eta}^{(s-1)}) - (\mathbf{x}_{it}^T \boldsymbol{\beta}_k^{(s-1)})^2 \right) \\ \mathbb{E}(\xi_{itk} \xi_{it'k} y_{it} y_{it'}) &= \mathbb{E}(\xi_{ith} \xi_{it'k}; \boldsymbol{\theta}^{(s-1)}) \mathbb{E}(y_{it} y_{it'} | \xi_{ith}=1, \xi_{it'k}=1; \boldsymbol{\theta}^{(s-1)}) \\ &= \hat{\pi}_{itt'hk}^{(s-1)} \left(v^{(tt')}(\boldsymbol{\eta}^{(s-1)}) - (\mathbf{x}_{it}^T \boldsymbol{\beta}_h^{(s-1)}) (\mathbf{x}_{it'}^T \boldsymbol{\beta}_k^{(s-1)}) \right) \end{aligned}$$

4.4 Computational details

The EM algorithm may converge to local maxima of the log-likelihood function or to singularities at the edge of the parameter space, where the log-likelihood is unbounded [30]. It is hence necessary to run the algorithm several times from different starting points to improve the chances to locate the global maximum. We followed a short-runs strategy (known as the emEM algorithm [31]) to initialize the EM algorithm, by running the EM algorithm from 10 random initializations, and stopping the algorithm when the increase of two successive log-likelihoods fell below $10^{-2}\%$, that is, without waiting for full convergence. In the case of mixtures of Gaussian distributions, divergence to singularities is fast and can be detected within these short EM runs [32]. We selected the outputs of the EM short runs and checked for spurious solutions, where this effect did not occur. Then, the short-run output maximizing the likelihood was used to initialize a longer run of the EM algorithm, stopped when the increase of two successive log-likelihoods fell below $10^{-5}\%$. We repeated this procedure 10 times and selected the maximum likelihood output. The final steps of the long-run EM algorithm were replaced by a direct numerical maximization of the log-likelihood function, exploiting a quasi-Newton method [33]. This combination of the EM algorithm and numerical optimization is often called a hybrid algorithm [34] and provides a compromise between the large circle of convergence provided by the EM algorithm and the high speed of direct numerical maximization [1]. In our case study, this hybrid approach was especially useful in the estimation of linear mixed HMMs with more than three states. Moreover, the quasi-Newton routine has the added benefit of producing, as a by-product, the estimated variance–covariance matrix of the parameter estimators (as given by the inverse of the observed Fisher information matrix).

5. Results

5.1. An analysis with time-constant fixed effects

The data described in Section 2 were preliminary examined by fitting a number of linear mixed models, which included functions of time as a time-varying covariate, in addition to the available time-constant covariates. The best results were obtained by exploiting a cubic trajectory, that is, by augmenting the columns of the covariate matrix X with the terms $t - \bar{t}$, $(t - \bar{t})^2$, and $(t - \bar{t})^3$, $t = 1, \dots, 24$; and $\bar{t} = 12$. Table I displays the BIC values obtained by including different components of correlation in the model. As a reference value, the top-left entry of the table (BIC=375,342.7) indicates the BIC of the linear regression model

$$y_{it} = \beta_0 + \beta_1(t - \bar{t}) + \beta_2(t - \bar{t})^2 + \beta_3(t - \bar{t})^3 + \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_{it} \quad (22)$$

where $\mathbf{x}_i^T$ is the row profile of the time-constant covariates described in Section 2, and the errors are independent.

The first row of Table I displays the results obtained by progressively including a number of random-effects components in model (22), namely, a random intercept b_{i1} , a linear component $(t - \bar{t})b_{i2}$, and a quadratic component $(t - \bar{t})^2 b_{i3}$, assuming an unstructured variance–covariance matrix of the included random effects and independent errors. We used the lme function of the nlme R-library to estimate these models, by centering the continuous

covariates. The BIC favors the least parsimonious model, where longitudinal correlation is explained by means of subject-specific quadratic trajectories (BIC=354,373.3).

One may avoid incorporating random effects in the model to account for dependence among repeated observations, choosing to use auto-correlated errors. The first column of Table I displays the BIC values under a number of ARMA processes, estimated by the `gls` function of the R-library `nlme`. Interestingly, most of the BIC values in this column are smaller than the BIC value obtained by considering quadratic trajectories and independent errors. By comparing the first row and the first column of Table I, a linear model with ARMA(2,2) errors model gives the best results.

The rest of the table displays the results obtained by combining different random-effects structures and error processes. According to the BIC, the best compromise between goodness of fit and parsimony is given by a LMM with quadratic random trajectories and an ARMA(2,1) structure of the errors. The estimates under this model are displayed in Table II. Some results are in keeping with previous studies of cardiac activity. Heart rate decreases with age, blood pressure, and good physical strength. Bradycardia is also typical in hearts whose electrical system was damaged by heart attacks or myocarditis. Finally, it is well known that heart rate increases with smoking, atherosclerosis, and low-grade inflammation. Somehow expectedly, the difference between nonsmokers and ex-smokers is not significant at a 95% confidence level. Unexpectedly, instead, not significant are also the effects of gender and physical functioning. This is in sharp contrast with well-established facts about heart rate. It is well known that men have a lower heart rate than women and that heart rates tend to decrease under good physical conditions [16, 35]. This puzzle can be solved by recalling that heart rates are usually measured under resting conditions at a given point in time, while the subjects in our study are repeatedly observed during the day, under normal life conditions. It is possible that some effects are confounded by latent time-varying conditions that evolve during the observation period and interact with some observed covariates. Moreover, the complex ARMA(2,1) correlation structure of the errors seems indicating that an important source of correlation has not been adequately captured by subject-specific, polynomial trajectories of time. This is somehow expected, as subjects under normal life conditions are characterized by different activities schedules during the day, which do not necessarily evolve according to a polynomial function of time. More meaningful trajectories could be in principle be included in the analysis, by using alternative time-varying covariates, which are however not available in the study under scrutiny. In general, little is known on time-varying predictors of heart rates, and on the way, they are linked to cardiac activity under normal life conditions. Self-reported variables such as sleep-wake cycles and working/resting periods are rarely used because of their inaccuracy.

5.2. A linear mixed hidden Markov model with two states

Possible time-varying effects can be captured by a linear mixed HMM that clusters the data according to a number of latent states of the subject and simultaneously estimates the effects of the covariates under each state. We discuss here the results obtained by fitting a model with two states and a subject-specific random intercept, say M_2 . Compared with the LMM discussed in the previous section, say M_1 (BIC=344,214.9), the BIC under model M_2 is

equal to 343,683.3, indicating that combining a random intercept and a hidden Markov chain is a better compromise between goodness of fit and model parsimony than combining subject-specific trajectories and auto-correlated errors.

Table III displays the estimates of the covariate effects and their standard errors under the two states (columns 1–4). This table furthermore includes the estimated initial distribution and the transition probabilities of the chain. According to model M_2 , heart rates are characterized by the alternation of two persistent states.

All the covariates that were significant (at a 95% confidence level) under M_1 are still significant under the two states inferred by M_2 , with concordant signs and smaller standard errors. Differences between nonsmokers and ex-smokers remain not significant under the two states. Remarkably, gender and physical functioning (not significant under M_1) are significant under M_2 . Compared with women, men have lower heart rates under both states. Better physical conditions are associated with lower heart rates.

In addition, Table III includes the estimated interaction effects $\beta_2 - \beta_1$ and their standard errors (columns 5 and 6). Some of the covariates are associated with a significant interaction effect, indicating that the influence of these covariates on heart activity varies according to the state under which the subject is observed. This finding explains, at least in part, why cross-sectional studies, based on heart rates observed at a given point in time, may give different results when they are used to estimate covariate effects on heart activity. It also contributes to the literature that describes heart activity by means of specific balances of the nervous system [9], by providing evidence that some covariates have a different effects under these balances. Not all the significant covariates are however associated with a significant interaction effect. Blood pressure, cholesterol levels, smoking, and previous myocardial infarction seem to have a time-constant effect on cardiovascular activity during the day, regardless of the state of the subject.

The results of Table III depend on the segmentation of longitudinal profiles inferred by M_2 . Figure 2 summarizes the main features of this segmentation. The top-left panel of the figure shows the conditional distribution of heartbeat intervals given the most probable state, evaluated by comparing the estimated posterior probabilities π_{ik}^* (modal allocation). These two distributions are similar to the two main components of cardiac activity found by Costa *et al.* [9] and can be, respectively, interpreted as conditions of HHA and LHA, associated with two typical balances of the autonomous nervous system. At the subject level, the temporal evolution of these states largely depends on daily schedules, for example, sleep–wake cycles or working–resting periods. By averaging these evolutions across subjects, we obtain the background cardiac rhythm in the study population. The top-right panel of Figure 2 shows the hourly proportion $\hat{p}_{t1}$ of the subjects who have been allocated by M_2 to a HHA state (solid line) and the proportion $\hat{p}_{t2} = 1 - \hat{p}_{t1}$ of those who have been allocated to a LHA state (dashed line). According to this allocation, most subjects are in HHA state at the beginning of the study: this proportion decreases with time, reaching a peak during night, it then increases in the morning. It thus reflects the association between typical balances of the nervous system and diurnal/nocturnal hours, known as the circadian heart rhythm.

The bottom panels of Figure 2 depict the temporal trajectories of the average effects, obtained from the components of the vector

$$\hat{\beta}_{kt} = \hat{p}_{1t}\hat{\beta}_1 + \hat{p}_{2t}\hat{\beta}_2$$

The bottom-left panel includes the covariates with a significant interaction effect (Table III), that is, with a significant time-varying influence on heart rates. The remaining covariates are included in the bottom-right panel of the figure. Table III shows that most of the effects under a LHA state are larger (in absolute value) than their counterparts estimated under a HHA state. Nocturnal hours are associated with larger proportions of subjects in a LHA state and, as a result, with larger average effects. Average effects decrease (in absolute value) during the diurnal hours, when larger proportions of subjects are in a HHA state. It is possible that nocturnal heart rate is simply a better measure than diurnal heart rate, and as a result, the association between heart activity and the included covariates is stronger during night than during diurnal hours. This hypothesis is biologically plausible but still debated in the medical literature [17].

5.3. Linear mixed hidden Markov models with more than two states

We have estimated a number of alternative linear mixed HMMs, by increasing the number of states up to $K = 5$. For each model, a BIC statistic has been computed to compare the models in terms of goodness of fit and parsimony. We have also computed the quality of the segmentation under each model through the classification entropy index

$$H = - \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^K \hat{\pi}_{itk} \log \hat{\pi}_{itk}$$

which measures the degree of class separation. This index has been exploited to compute the BIC Integrated Classification Likelihood (ICL) statistic, obtained by the sum $\text{BIC} + 2H$. It penalizes the BIC statistic for poorly segmented profiles [36, 37], and it is often exploited to select the number of latent states of HMMs [29,38].

Table IV summarizes the results. The BIC statistic suggests a model with three states, namely, M_3 . However, M_2 relies on a better segmentation than M_3 , and it is supported by the BIC-ICL statistic. BIC and ICL-BIC disagree because BIC selects a mixture of overlapping components, which improve the fit at the price of classes that are poorly separated. This behavior of the BIC statistic has been extensively studied in the context of mixtures models [39]. Figure 3 summarizes the output of M_3 . This model introduces a state of intermediate heart activity that reaches up to half of the sample in the late evening (Figure 3; top-left and top-right panels). M_3 takes therefore the form of a mixture of three components, which are in this case overlapping, thus not adding value for the physiological interpretation but just improving the fitted density within each latent class. The bottom panels of Figure 3 show the time-varying average effects of the covariates under M_3 . Compared with the patterns found under M_2 , the effects trajectories keep a similar shape, thus confirming the findings obtained

by M_2 . However, under M_3 , some covariates have a reduced influence compared with that estimated by M_2 . It is possible that the richer heterogeneity structure, introduced in M_3 by an additional third state, captures part of the heterogeneity explained by the observed covariates.

6. Discussion

Combining multiple sources of heterogeneity is a natural strategy to model complex correlation structures in longitudinal studies. We propose a linear mixed HMM that integrates subject-specific random effects and a latent Markov chain that segments longitudinal profiles. Random effects are exploited to account for unobserved baseline conditions. The latent Markov chain accounts for the additional source of correlation because of the interaction between the available covariates and a time-varying latent factor, which reflects different conditions under which a subject is observed.

In a case study of longitudinal heart rates, the model is useful to estimate the effects of the observed covariates under different latent conditions of the heart, associated with specific balances between sympathetic and parasympathetic activities of the autonomic nervous system. Our findings indicate that the effect of selected covariates, including gender differences and physical functioning, significantly varies according to two specific balances of the autonomic nervous system, which reflect two states of low and high cardiac activity, respectively. In this case study, a linear mixed HMM with two states performs better than traditional linear mixed models that combine subject-specific polynomial trajectories and auto-correlated errors. As expected, subjects under normal life conditions are characterized by different activities schedules during the day, which do not necessarily evolve according to a polynomial function of time. More meaningful parametric trajectories are difficult to specify as little is known on time-varying predictors of heart rates, and on the way, they are linked to cardiac activity. Under this setting, a HMM segmentation of trajectories avoids possibly misspecified parametric assumptions and, in summary, provides a method to describe subject-specific trajectories that are otherwise difficult to model by a parsimonious combination of parametric trajectories with subject-specific random coefficients and auto-correlated errors.

The proposed model relies on a homogeneous Markov chain. In the analysis of heartbeat intervals, an homogeneous transition probability matrix reflects the assumption of a background circadian rhythm that is shared by all the subjects of the population. In principle, this assumption could be relaxed by assuming that the transition probabilities depend on suitable covariates. Little is however known on the factors that modulate the transitions between different balances of the nervous system.

We describe heartbeat intervals in terms of two states of high and low cardiac activity. In our case study, a larger number of states lead to a mixture of overlapping components that improves the fit, without adding value for physiological interpretation. We remark that this parsimonious two-state model aims at detecting time-varying effects on hourly averages of heartbeat intervals, which may hide short peaks of activity. At a finer temporal resolution, the model could likely involve a large number of states. At the hour level, where the interest

is on large cycles of heart activity, our findings indicate that a model with two states of high and low cardiac activity is a good compromise between goodness of fit and interpretable inference.

An EM algorithm is proposed for maximum likelihood estimation. EM algorithms for mixed HMMs are traditionally based on a complete likelihood function, which treats both the latent states and the random effects as missing values [13]. The E step thus involves the computation of expectations with respect to the predictive distribution of both the latent states and the random effects, given the observed data. An alternative approach to maximum likelihood estimation relies on the direct numerical maximization of the log-likelihood function, integrating the conditional observation process with respect to both the distribution of the latent states and the random effects [4]. Both approaches have advantages and disadvantages, because, on one side, EM algorithms are numerically more stable than procedures of numerical maximization, but, on the other side, numerical procedures are computationally much faster than EM algorithms. Taking advantage of the normality of both the outcome and the random effects, the proposed EM is an intermediate proposal between these two approaches. We integrate the conditional observation process only with respect to the distribution of the random effects and obtain a complete log-likelihood function that depends on the latent states. This allows for an E step with closed-form updating equations and a M step that essentially reduces to the estimation of a linear mixed model on an augmented dataset with case weights.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute on Aging (grant numbers R01AG026786, R01AG16790, R01AG16661, and P01-AG020166), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant number R24HD047879), and the Dynasty Foundation (Russia). The funding agencies had no role in the study design; in the collection, analysis, or interpretation of the data; in writing the manuscript; or in the decision to submit it for publication. We thank James W. Vaupel at MPIDR (Rostock, Germany) for his leadership in designing the SAHR. Svetlana Shalnova from the National Research Centre for Preventive Medicine (NRCPM) in Moscow (Russia) co-ordinated the SAHR field works. Viktoria Metelskaya from NRCPM for provided expertise in biochemical measurements. Evgeny Andreev at the New Economic School (Moscow) and Alexander Deev at NRCPM were responsible for data cleaning and pre-processing work.

Appendix A

The task of computing the posterior probabilities (18) and (15) from an estimate $\theta^{(s-1)}$ is generally referred to as the HMM-smoothing numerical issue, and it is typically solved by specifying the posterior probabilities in terms of suitably normalized functions, which can be computed recursively, avoiding unpractical summations over the state space of latent Markov chains and numerical underflow and overflow. In the literature, this approach is known as the FB recursion, and it can be implemented in a number of different ways ([10]; ch. 3). We describe in the succeeding text the FB recursion we exploited in this paper. Let

$$L_{itk}(\theta) = \int \phi(y_{it}; \mathbf{x}_{it}^T \beta_k + Z_i \mathbf{b}, \sigma^2) \phi(\mathbf{b}; \mathbf{0}, \mathbf{V}_b) d\mathbf{b}$$

be the conditional contribution to the likelihood of observation y_{it} . Furthermore, let $\mathbf{y}_{i,0:t}$ and $\xi_{i,0:t}$ be, respectively, the outcome profile and the segmentation within subject i , truncated at t . Finally, let

$$L_{i,0:t}(\theta) = \sum_{\xi_{0:t}} p(\xi_{0:t}; \pi) f(\mathbf{y}_{i,0:t} | \xi_{0:t}; \beta, \eta)$$

be the likelihood contribution of the first t observations of subject i . We run a forward and a backward iteration. During the forward iteration, we exploit the output of the s -th step of the EM algorithm to compute the probabilities $\psi_i^{(t)}(k) = P(\xi_{itk} = 1 | \mathbf{y}_{i,0:t-1})$, the likelihood ratios

$$c_{it} = \frac{L_{i,0:t}(\hat{\theta}_s)}{L_{i,0:t-1}(\hat{\theta}_s)}, \text{ and the forward probabilities } \bar{a}_{it}(k) = P(\xi_{itk} = 1 | \mathbf{y}_{i,0:t}), \text{ as follows.}$$

Forward recursion (for each $i = 1 \dots n$):

- initialization:

$$\begin{aligned} \psi_i^{(0)}(k) &= \hat{\pi}_k^{(s)} \\ c_{i0} &= \sum_{k=1}^K \psi_i^{(0)}(k) L_{i0}(\hat{\theta}_s) \\ \bar{a}_{i0}(k) &= \frac{\psi_i^{(0)}(k) L_{i0}(\hat{\theta}_s)}{c_{i0}} \end{aligned}$$

- for $t = 1, \dots, T$

$$\begin{aligned} \psi_i^{(t)}(k) &= \sum_{h=1}^K \bar{a}_{i,t-1}(h) \hat{\pi}_k^{(s)} \\ c_{it} &= \sum_{k=1}^K \psi_i^{(t)}(k) L_{it}(\hat{\theta}_{ks}) \\ \bar{a}_{it}(k) &= \frac{\psi_i^{(t)}(k) L_{it}(\hat{\theta}_{ks})}{c_{it}} \end{aligned}$$

At the end the forward recursion, we store the values $c_{i0} \dots c_{iT}$ and $\bar{a}_{i0}(k) \dots \bar{a}_{iT}(k)$, $i = 1, \dots, n$. The sequence $c_{i0} \dots c_{iT}$ can be exploited to compute the value taken by the log-likelihood at the s -th step of the EM algorithm, as follows:

$$\log L(\hat{\theta}_s) = \sum_{i=1}^n \sum_{t=0}^T \log c_{it}$$

We then run a backward recursion, by computing the ratios $\bar{\varphi}_{it}(k) = \frac{f(\mathbf{y}_{i,t+1:T} | \xi_{itk}=1)}{\prod_{l=t}^T c_{il}}$, as follows.

Backward recursion (for $i = 1 \dots n$):

- initialization: $\bar{\varphi}_{iT}(k) = \frac{1}{c_{iT}}$
- for $t = T-1, T-2, \dots, 0$

$$\bar{\varphi}_{it}(k) = \frac{\sum_{h=1}^K \hat{\pi}_{h|k}^{(s)} L_{i,t+1}(\hat{\theta}_{ks}) \bar{\varphi}_{i,t+1}(h)}{c_{it}}$$

At the end of the backward recursion, we store the values of $\bar{\varphi}_{i0}(k) \dots \bar{\varphi}_{iT}(k)$ and compute the univariate posterior probabilities (18) as

$$\hat{\pi}_{itk}^{(s)} = \frac{\bar{\alpha}_{it}(k) \bar{\varphi}_{it}(k)}{\sum_{k=1}^K \bar{\alpha}_{it}(k) \bar{\varphi}_{it}(k)}$$

and the posterior bivariate probabilities (15) as

$$\hat{\pi}_{i,t-1,t,hk}^{(s)} = \bar{\alpha}_{it}(k) \hat{\pi}_{k|h}^{(s)} L_{i,t+1}(\hat{\theta}) \bar{\varphi}_{i,t+1}(k)$$

References

1. Zucchini, W.; MacDonald, IL. Hidden Markov Models For Time Series: An Introduction Using R. Chapman & Hall; Boca Raton: 2009.
2. Shirley KE, Small DS, Lynch KG, Maisto SA, Oslin DW. Hidden Markov models for alcoholism treatment trial data, the annals of applied statistics. Annals of Applied Statistics. 2010; 4:366–395. volume 4, number 1 (2010), 366–395.
3. Langrock R, Swihart BJ, Caffo BS, Punjabi NM, Crainiceanu CM. Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms. Statistics in Medicine. 2013; 32(19):3342–3356. [PubMed: 23348835]
4. Altman, RaM. Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. Journal of the American Statistical Association. 2007; 102:201–210.
5. Maruotti A. Mixed hidden Markov models for longitudinal data: an overview. International Statistical Review. 2011; 79(3):427–454.
6. Bartolucci, F.; Farcomeni, A.; Pennoni, F. Latent Markov Models for Longitudinal Data. Chapman and Hall/CRC; Boca Raton: 2012.
7. Stein PK, Kleiger RE. Insights from the study of heart variability. Annual Review of Medicine. 1999; 50:249–261.
8. Tsuji H, Venditti FJ, Manders ES. Determinants of heart rate variability. Journal of the American College of Cardiology. 1996; 28:1539–46. [PubMed: 8917269]
9. Costa T, Boccignone G, Ferraro M. Gaussian mixture model of heart rate variability. PLoS ONE. 2012; 7(5):1–9. e37731.

10. Cappé, O.; Moulines, E.; Rydén, T. *Inference in Hidden Markov Models*. Springer; New York: 2005.
11. Seltman, HJ. *Case Studies in Bayesian Statistics*. Gatsonis, C., editor. Springer-Verlag; New York: 2002. p. 397-405.
12. Detilleux JC. The analysis of disease biomarker data using a mixed hidden Markov model. *Genetics Selection Evolution*. 2008; 40(5):491–509.
13. Delattre M. Inference in mixed hidden Markov models and applications to medical studies. *Journal de la Société Française de Statistique*. 2010; 151:90–105.
14. Nunan D, Sandercock GR, Brodie DA. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *Pacing Clin Electrophysiol*. 2010; 33:1407–1417. [PubMed: 20663071]
15. Plichart M, Thomas F, Empana JP, Bean K, Périer MC, Celermajer DS, Hanon O, Danchin N, Pannier B, Jouven X. Gender-specific trends in heart rate in the general population from 1992–2007: a study of 226,288 French adults. *European Journal of Preventive Cardiology*. 2013; 20:61–72. [PubMed: 22345675]
16. Smetana P, Malik M. Sex differences in cardiac autonomic regulation and in repolarisation electrocardiography. *Pflügers Archiv - European Journal of Physiology*. 2013; 465(5):699–717. [PubMed: 23404618]
17. Johansen CD, Olsen RH, Pedersen LR, Kumarathurai P, Mouridsen MR, Binici Z, Intzilakis T, Køber L, Sajadieh A. Resting, night-time, and 24 h heart rate as markers of cardiovascular risk in middle-aged and elderly men and women with no apparent heart disease. *European Heart Journal*. 2013; 34(23):1732–1739. [PubMed: 23306958]
18. Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*. 1996; 17:354–381. [PubMed: 8737210]
19. Shkolnikova M, Shalnova S, Shkolnikov VM, Metelskaya V, Deev A, Andreev E, Jdanov D, Vaupel JW. Biological mechanisms of disease and death in Moscow: rationale and design of the survey on stress aging and health in Russia (Sahr). *BMC Public Health*. 2009; 9:293–308. [PubMed: 19678931]
20. RAND. [Accessed on August 4, 2011] Medical outcomes study: 36-item short form survey. 2011. Retrieved August 4, 2011 from http://www.rand.org/health/surveys_tools/mos/mos_core_36item.html
21. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetière P, Jousilahti P, Keil U, Njølstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the score project. *European Heart Journal*. 2003; 24:987–1003. [PubMed: 12788299]
22. Syddall H, Cooper C, Martin F, Briggs R, Sayer AA. Is grip strength a useful single marker of frailty? *Age and Ageing*. 2003; 32:650–656. [PubMed: 14600007]
23. NCEP (National Cholesterol Education Program) L, National Heart BI. Technical Report. National Institutes of Health; 2001. Third report of the NCEP expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel iii): executive summary NIH publication no. 01-3670.
24. Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, Criqui M 3rd, Fadl YY, Fortmann SP, Hong Y, Myers GL, Rifai N, Smith SC Jr, Taubert K, Tracy RP, Vinicor F. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: a statement for healthcare professionals from the centers for disease control and prevention and the american heart association. *Circulation*. 2003; 107:499–511. [PubMed: 12551878]
25. Granger CWJ, Morris MJ. Time series modelling and interpretation. *Journal of the Royal Statistical Society Series A*. 1976; 139:246–257.
26. Celeux G, Martin O, Lavergne C. Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modelling*. 2005; 5:243–267.
27. Vermunt JK. Latent class and finite mixture models for multilevel datasets. *Statistical Methods in Medical Research*. 2008; 17:33–51. [PubMed: 17855746]

28. Lagona F, Picone M. Model-based clustering of multivariate skew data with circular components and missing values. *Journal of Applied Statistics*. 2012; 39(5):927–945.
29. Bulla J, Lagona F, Maruotti A, Picone M. A multivariate hidden markov model for the identification of sea regimes from incomplete skewed and circular time series. *Journal of Agricultural, Biological, and Environmental Statistics*. 2012; 17(4):544–567.
30. Wu CFJ. On the convergence properties of the EM algorithm. *The Annals of Statistics*. 1983; 11:95–103.
31. Biernacki C, Celeux G, Govaert G. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*. 2003; 41(3-4):561–575.
32. Ingrassia S, Rocci R. Degeneracy of the EM algorithm for the MLE of multivariate gaussian mixtures and dynamic constraints. *Computational Statistics and Data Analysis*. 2011; 55:1715–1725.
33. Nash, JC. *Compact Numerical Methods For Computers: Linear Algebra and Function Minimisation*. Wiley; New York: 1979.
34. Bulla J, Berzel A. Computational issues in parameter estimation for stationary hidden Markov models. *Computational Statistics*. 2008; 23(1):1–18.
35. Stor EJ, McDonnell B, Thompson J, Stone K, Bull T, Houston R, Cockcroft J, Shave R. Left ventricular mechanics in humans with high aerobic fitness: adaptation independent of structural remodelling, arterial haemodynamics and heart rate. *The Journal of Physiology*. 2012; 590(9): 2107–2119. [PubMed: 22431336]
36. Biernacki C, Celeux G, Govaert G. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22:719–725.
37. McLachlan, G.; Peel, D. *Finite Mixture Models*. Wiley; New York: 2000.
38. Lagona F, Picone M. Maximum likelihood estimation of bivariate circular hidden Markov models from incomplete data. *Journal of Statistical Computation and Simulation*. 2013; 83:1223–1237.
39. Baudry JP, Raftery AE, Celeux K, Lo G, Gottardo R. Combining mixture components for clustering. *Journal of Computational and Graphical Statistics*. 2010; 19:332–353. [PubMed: 20953302]

Figure 1.

Distribution of hourly heartbeat intervals of 1274 subjects, observed during 24 h under normal life conditions.

Figure 2.

Results of a linear mixed hidden Markov model with two states. Top left: the distribution of heartbeat intervals given the most probable state (HHA: high heart activity; LHA: low heart activity). Top right: the estimated proportions of subjects under HHA (solid line) and LHA (dashed line) conditions. Bottom: time-varying effects of the covariates on the length of heartbeat intervals.

Figure 3.

Results of a linear mixed hidden Markov model with three states. Top left: the distribution of heartbeat intervals given the most probable state (HHA: high heart activity; IHA: intermediate heart activity; LHA: low heart activity). Top right: the estimated proportions of subjects under HHA (solid line), LHA (dashed line), and IHA (dotted line) conditions. Bottom: time-varying effects of the covariates on the length of heartbeat intervals.

Table I

Linear mixed models with a cubic trend: BIC values.

Errors component	Random-effects component			
	None	Constant	Linear	Quadratic
White noise	375,342.7	369,486.7	357,110.0	354,373.3
AR(1)	345,404.9	344,747.2	344,767.8	344,663.2
MA(1)	357,058.1	347,295.7	347,091.7	346,009.4
ARMA(1,1)	345,366.8	344,359.9	344,380.5	344,225.9
AR(2)	345,377.9	344,358.6	344,379.1	344,208.9
MA(2)	350,623.0	344,921.4	344,915.9	344,456.5
ARMA(1,2)	345,323.7	344,365.0	344,385.6	344,218.9
ARMA(2,1)	345,068.6	344,362.8	344,383.4	344,214.9
ARMA(2,2)	344,954.6	344,368.9	344,389.4	344,224.0

AR, auto-regressive; ARMA, auto-regressive moving-average; MA, moving-average.

Table II

Fixed effects and variance components under the best linear mixed model with a population cubic trajectory.

Fixed effect	Estimate	Standard error	
Intercept	933.235	6.059	
Time	21.860	0.306	
Time ²	-1.689	0.026	
Time ³	-0.189	0.003	
Age (years)	3.074	0.428	
Female (reference: male)	-7.450	9.176	
Ex-smoker (reference: never smoked)	-5.893	6.716	
Current smoker (reference: never smoked)	-19.931	7.593	
Myocardial infarction reported (reference: not reported)	31.030	9.852	
Stroke reported (reference: not reported)	30.348	10.760	
Systolic blood pressure (mmHg)	0.222	0.110	
High total cholesterol (>240 mg/dL)	-16.456	5.255	
High C-reactive protein (>3 mg/dL)	-16.260	5.705	
High Interleukin-6 (m: >2 pg/mL; f: >2.4 pg/mL)	-31.588	6.973	
Hand grip strength (kg)	1.200	0.421	
Physical functioning (SF36)	0.204	0.133	
Random effects	Standard deviation	Correlation	
		Time	Time ²
Intercept	98.395		
Time	0.004	-0.018	
Time ²	0.593	-0.561	0.033
Residual standard deviation	89.670		
Errors auto-correlation			
AR(1)	0.437		
MA(1)	0.092		
AR(2)	0.727		

AR, auto-regressive; MA, moving-average.

Table III

Fixed effects and variance components under a linear mixed hidden Markov model with two states.

Fixed effect	High activity state		Low activity state		Interaction	
	Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
Intercept	489.45	21.51	624.65	33.17	135.20	18.43
Age (years)	2.63	0.23	3.65	0.36	1.02	0.20
Female (reference: male)	-12.13	4.96	-24.07	7.65	-11.94	4.28
Ex-smoker (reference: never smoked)	-5.49	3.64	10.76	5.58	-5.27	3.07
Current smoker (reference: never smoked)	-25.39	4.13	-30.34	6.37	-5.15	3.51
Myocardial infarction reported (reference: not reported)	25.77	5.44	18.61	8.51	-7.16	4.67
Stroke reported (reference: not reported)	33.15	5.88	53.97	9.50	20.82	5.52
Systolic blood pressure (mmHg)	0.21	0.06	0.19	0.09	-0.02	0.05
High total cholesterol (>240 mg/dL)	-18.98	2.85	-19.96	4.37	-0.98	2.41
High C-reactive protein (>3 mg/dL)	-17.02	3.10	-32.04	4.78	-15.02	2.63
High Interleukin-6						
(m: >2 pg/mL; f: >2.4 pg/mL)	-30.83	3.83	-38.72	5.92	-7.89	3.23
Hand grip strength (kg)	1.25	0.23	2.45	0.35	1.20	0.19
Physical functioning (SF36)	0.30	0.07	0.49	0.11	0.19	0.06
Random intercept standard deviation			41.49			
Residual standard deviation			85.88			
Destination state						
Origin state	High activity	Low activity				
High activity	0.92	0.08				
Low activity	0.08	0.92				
$t = 0$	0.95	0.05				

Table IV

Linear mixed hidden Markov models: BIC values.

Model	Number of parameters	BIC	Classification entropy	ICL-BIC
M_2	31	343,683.3	4025.3	351,734.0
M_3	49	339,634.9	6140.2	351,915.3
M_4	69	340,721.0	8341.1	357,403.2
M_5	91	341,362.4	9971.0	361,304.4

ICL, Integrated Classification Likelihood.