Linear quantile hidden Markov models for longitudinal data with informative drop-out

Maria Francesca Marino∗ Nikos Tzavidis† Marco Alfó‡

Abstract

Linear quantile regression models represent a general class of models that aims at providing a detailed and robust picture of the (conditional) response distribution as function of a set of observed covariates. Longitudinal data represent an interesting potential field of application of such models; due to their peculiar features, they represent a substantial challenge, in that the standard, cross-sectional, model representation needs to be extended for dealing with such kind of data. In fact, repeated observations from the same statistical unit poses a problem of dependence; in a conditional perspective, this dependence could be ascribed to sources of unobserved, individual-specific, heterogeneity. Along these lines, quantile regression models have recently been extended to the analysis of longitudinal, continuous, responses, by modelling dependence via time-constant, see Geraci and Bottai (2007), or time-varying, see Farcomeni (2012), random effects. In this manuscript, we introduce a general quantile regression model for longitudinal, continuous, responses where time-varying and time-constant random parameters with unspecific distribution are jointly taken into account.

∗ Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Italy, mariafrancesca.marino@uniroma1.it
† Department of Social Statistics & Demography, University of Southampton, UK, n.tzavidis@soton.ac.uk
‡ Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Italy, marco.alfo@uniroma1.it
A further feature of longitudinal designs is the presence of partially incomplete sequences, due to some individuals leaving the study before its designed end. The missing data generating process may produce a selection of units which can be informative with respect to the parameters of the longitudinal data model. Therefore, a further extension is needed. To deal with the case of irretrievable drop-out, we introduce a pattern mixture version of the linear quantile hidden Markov model, where we account for time-varying heterogeneity and for changes in the fixed effect vector due to differential propensities to stay in the study. The proposed models are illustrated using a well known benchmark dataset on longitudinal dynamics of CD4 cells and by means of a large scale simulation study, entailing different quantiles and both complete and partially complete (ie subject to drop-out) individual sequences.

1 Introduction

Data from longitudinal studies are often the focus of statical analyses; since measurements recorded on the same individual are likely associated, the adopted statistical modelling tools need to account for such a dependence. A common framework is based on postulating a conditional model which is augmented by some kind of unobserved heterogeneity that is meant to describe the dependence in the data. Since the dependence may be either time-constant and/or time varying, the class of mixed hidden Markov models can provide a general modelling solution. Within this framework, the true and the spurious contagion described by [Heckman (1981)] are captured by time-varying and time-constant, individual-specific, random parameters. The class of mixed HMMs has been introduced by [Altman (2007)] in the case of responses with (conditional) distribution in the exponential family, and generalized by [Maruotti and Rydén (2009)] who have considered time-constant random effects with an unspecified distribution, see [Maruotti (2011)] for a thorough review of the topic. The idea is to describe the (conditional) mean of the response variable in terms of the observed covariates, while ac-
counting for a flexible form of dependence between repeated measurements. We propose to extend this approach to the quantile regression context, joining features from linear quantile mixed effect models, see Geraci and Bottai (2007), and linear quantile hidden Markov models, see Farcomeni (2012), within a single modelling representation. The resulting model, a linear quantile mixed hidden Markov model (lqmHMM), offers robustness against the possible presence of outliers in the observed data and returns reliable and efficient parameter estimates accounting for structured, although unobserved, random variation in the data generating process.

When we consider data from longitudinal studies, a frequent empirical finding is that not all individuals are available at all the intended time occasions; therefore, we are in the presence of partially incomplete data records. The presence of missing information rises a number of challenges because of the potential bias in the parameter estimates that can lead to misleading inferential conclusions. Here, we focus on monotone missingness, that is on those empirical situations where once a unit leaves the study she/he can not re-enter. We describe how a formal model approach can be built to take into account the presence of (potentially non-ignorable) missing data. According to Roy (2003) and Roy and Daniels (2008), we assume that there exists a discrete, ordered, latent variable with a distribution that depends on the propensity to stay in the study; a proxy for this propensity is represented by the number of available measurements for each unit and is assumed to cause changes in fixed parameter estimates. The rationale behind this representation is that the dependence between the observed responses and the missing data process is due to the presence of sources of unobserved heterogeneity shared by individuals with a similar propensity to drop-out.

The paper is structured as follows: in section 2 after discussing the proposal by Farcomeni (2012), we introduce the linear quantile mixed hidden Markov model, where time-varying and time constant random parameters are jointly considered. In section 3 we focus on the proposals by Roy (2003); Roy and Daniels (2008) and apply the latent drop-out class approach to the lqmHMM. Computational details to derive parameter estimates, in a maximum likelihood perspective, for both models we have discussed so far are presented
In section 4, results of a large scale simulation study and the application of the proposed model to the CD4 dataset (Kaslow et al. 1987; Zeger and Diggle 1994) are given in sections 5 and 6 respectively. Last section gives concluding remarks and outlines potential future research agenda.

2 LqmHMM: model specification

We start by introducing some notation. Let $Y_{it}$ be a continuous response variable for unit $i = 1, \ldots, n$ observed at time occasion $t = 1, \ldots, T$; further, let us assume that a set of covariates $x_{it}$ has also been observed, and that the aim is to define a quantile regression model for the conditional distribution of the response variable given the set of observed covariates. Farcomeni (2012) proposes to account for dependence between longitudinal measurements on the same individual by using a hidden Markov representation for the quantile regression model, which we will refer, in the following, with the acronym lqHMM. For a given quantile $\tau \in (0, 1)$, let $\{S_{it}(\tau)\}$ be a homogeneous, first order, hidden Markov chain with state space $S(\tau) = \{1, \ldots, m(\tau)\}$ and initial and transition probabilities given by $\delta_h(\tau) = \Pr(S_{it}(\tau) = h)$ and $q_{kh}(\tau) = \Pr(S_{it}(\tau) = h \mid S_{it-1}(\tau) = k), h, k = 1, \ldots, m(\tau)$. The linear quantile hidden Markov model is specified as follows:

$$Q(y_{it} \mid s_{it}, \tau) = x_{it}' \beta(\tau) + \alpha_{s_{it}(\tau)},$$

where $\alpha_{s_{it}(\tau)}$ represents an individual-specific random intercept that evolves over time according to a first-order, hidden Markov chain, as described before. Since, however, some of the random parameters may be time-varying while others may represent individual-specific features that do not evolve over time, or can be thought of as approximately constant over the analysed time window, the lqHMM representation could be overparameterized. To be more precise, we could increase the number of states of the hidden chain by considering a Cartesian product of time-varying and time-constant supports, but this could result in a quite complex objective function and in an increased number of model parameters. Therefore, for this purpose, we
assume that dependence could be due to different sources of unobserved heterogeneity; some of them are time-varying, while others are time-constant.

Let \( b_i(\tau) = (b_{i1}(\tau), \ldots, b_{iq}(\tau)) \) be an individual-specific random vector with density \( f_b(\cdot; D(\tau)) \), where \( D(\tau) \) is a \((q(\tau) \times q(\tau))\), quantile dependent, covariance matrix and \( E(b_i(\tau)) = 0 \). For a given quantile \( \tau \in (0, 1) \), the \( lqmHMM \) is defined by the following expression

\[
Q(y_{it} | s_{it}, b_i, \tau) = x_{it}' \beta(\tau) + z_{it}' b_i(\tau) + w_{it}' \alpha_{s_{it}(\tau)},
\]

where \( \beta(\tau) \) is a \( p \)-dimensional vector of fixed parameters determining the effect of observed covariates on the response quantiles, while \( b_i(\tau) \) and \( \alpha_{s_{it}(\tau)} \) summarize time-constant and time-varying random deviations from the homogeneous model, respectively. The hidden Markov chain \( \{ S_{it}(\tau) \} \) determines the evolution of the \( \alpha \) parameters over time and is introduced to describe unobserved individual trends. All the model parameters introduced so far are dependent on the specific quantile \( \tau \); however, for ease of notation, such a dependence will be omitted in the following, and model parameters will be denoted free of \( \tau \).

We assume that the random vector \( b_i \) and the hidden process \( \{ S_{it} \} \) are independent as they are meant to capture different sources of unobserved heterogeneity. The distribution of the response variable, at a given time occasion, is defined conditional on the hidden state occupied at the same time and the (time-constant) individual-specific random coefficients. Conditional on \( s_{it} \) and \( b_i \), repeated measures are no longer dependent and the conditional distribution of the individual sequence can be factorized as

\[
f_{y_{i:t} | s_i, b_i, \tau} = \prod_{t=1}^{T} f_{y_{i:t-1} | y_{i1:t-1}, s_{i1:t}, b_i, \tau} = \prod_{t=1}^{T} f_{y_{it} | s_{it}, b_i, \tau},
\]

where \( y_{i1:t-1} \) represents the history of the responses for the \( i \)-th unit up to time \( t - 1 \) and \( s_{i1:t} \) indicates the sequence of hidden states visited up to time \( t \). The proposed model reduces to the model introduced by Geraci and Bottai (2007) for \( m = 1 \) and to the model defined by Farcomeni (2012) for
Parameter estimation is performed in a ML perspective; we follow the proposal of Geraci and Bottai (2007) and assume that responses are (conditionally) independent random variables with an asymmetric Laplace density

\[ Y_{it} \mid s_{it}, b_i, \tau \sim \text{ALD} \left( \mu_{it}[s_{it}, b_i], \sigma, \tau \right), \]

where the location parameter \( \mu_{it} \) is defined by expression (2). In detail, when \( S_{it} = h \), the following regression model holds

\[ \mu_{it}[S_{it} = h, b_i] = x_{it}'\beta + z_{it}'b_i + w_{it}'\alpha_h. \]

By denoting by \( \psi \) the global set of longitudinal model parameters, equation (3) becomes

\[
f_{y \mid sb}(y_i \mid s_i, b_i, \psi, \tau) = \left[ \frac{\tau(1 - \tau)}{\sigma} \right]^{T_i} \exp \left\{ -\sum_{t=1}^{T_i} \rho_{\tau} \left[ \frac{y_{it} - \mu_{it}[s_{it}, b_i]}{\sigma} \right] \right\}.
\]

where \( \rho_{\tau}(\cdot) \) is the standard quantile loss function introduced by Koenker and Bassett (1978). In fact, the assumption that responses have a (conditional) AL distribution is introduced just to build a bridge between ML and the standard minimization of the loss function \( \rho_{\tau}(\cdot) \) originally proposed by Koenker and Bassett (1978). The marginal distribution of the whole sequence of hidden states can be derived by exploiting the Markov property of the hidden chain:

\[
f_s(s_i \mid \delta, Q, \tau) = f_s(s_{i1} \mid \delta, \tau) \prod_{t=2}^{T_i} f_s(s_{it} \mid s_{it-1}, Q, \tau) = \delta_{s_{i1}} \prod_{t=2}^{T_i} q_{s_{it-1}s_{it}}.
\]

Based on the modelling assumptions introduced so far, the observed data likelihood is defined by the following expression

\[
L(\cdot) = \prod_{i=1}^{n} \int \sum_{s_i} \left\{ \prod_{t=1}^{T_i} f_{y \mid sb}(y_{it} \mid s_{it}, b_i, \psi, \tau) \delta_{s_{i1}} \prod_{t=2}^{T_i} q_{s_{it-1}s_{it}} \right\} f_b(b_i \mid D, \tau)db_i.
\]
In order to obtain parameter estimates, we need to solve a (multiple) integral that does not have a closed form expression. If we specify a parametric distribution for the random parameters $b_i$, Monte Carlo EM algorithms (Geraci and Bottai, 2007; Liu and Bottai, 2009) or Gaussian quadrature approximations (Geraci and Bottai, 2014) could be exploited, once they have been extended to deal with the hidden Markov chain.

Alternatively, we may leave the distribution of random coefficients unspecified and estimate it by using a discrete distribution on $G(\tau) \leq n$ support points, as proposed by Maruotti and Rydén (2009). In the context of the exponential family, this is known as the nonparametric maximum likelihood (NPML) estimate of the mixing distribution (here $f(y | b_i, D, \tau)$, see eg Aitkin (1996, 1999) and the results by Lindsay (1983a,b). Such a semiparametric approach offers a great flexibility and allows to overcome potential bias due to misspecification of the random coefficient distribution. Locations $b_g$ and masses $\pi_g = \Pr(b_g), \sum_g \pi_g = 1, g = 1, ..., G$ are treated as unknown and estimated through the observed data together with the remaining model parameters. It is important to highlight that also such a discrete distribution is directly related to the level $\tau$ we are considering. The estimated locations and the masses of the mixing distribution, as well as the optimal number of components, may potentially vary when analysing different quantiles of the outcome.

The observed data likelihood in expression (5) can therefore be approximated by

$$L(\cdot) = \prod_{i=1}^{n} \sum_{g=1}^{G} \sum_{s_i} \left\{ \prod_{t=1}^{T_i} f_{y_{it} | s_{it}, b_g, \theta, \tau} \delta_{s_{it}} \prod_{t=2}^{T_i} q_{s_{it-1} s_{it}} \right\} \pi_g,$$

where $f_{y_{it} | s_{it}, b_g; \tau}$ is the conditional distribution of the response variable for a generic unit $i$ belonging to the $g$-th component of the finite mixture and being, at time occasion $t$, in the hidden state $s_{it}$. Maruotti (2011) identifies a number of advantages in the use of the NPML approach to derive parameter estimates in generalized linear mixed HMM models. When compared to Gaussian quadrature techniques, the proposed approach requires
a reduced computational load since its complexity is linear with the integral dimension, while for Gaussian random quadrature, the computational complexity increases exponentially with the dimension of the random coefficient vector. Since locations in the finite mixture are completely free to vary over the corresponding support, extreme departures from the basic (homogeneous) model can be, easily, accommodated.

Even if the maximization of expression (6) can be directly performed by exploiting an appropriate matrices product, it still results quite challenging. As an alternative, a generalization of the EM algorithm for finite mixtures, as described in Aitkin (1996) and Aitkin (1999), could result simpler. Section 4 gives the computational details of the algorithm.

3 Modelling non-ignorable drop-out via latent drop-out classes

In many empirical cases, the measurement process is affected by monotone drop-out (DO); that is, some individuals may leave the study before the planned completion time and thus present incomplete data records. Let \( R_i = (R_{i1}, ..., R_{iT}) \) be the missing data indicator vector for the \( i \)-th individual, with \( R_{it} = 1 \) if \( y_{it} \) has not been observed at time occasion \( t = 1, ..., T \) and zero otherwise. Since we are considering drop-out, \( R_{it} = 1 \implies R_{i't} = 1, t' \geq t = 1, ..., T \).

The individual sequence \( y_i = (y_{i1}, ..., y_{iT}) \) can be partitioned into two different components, \( y^o_i \) and \( y^m_i \), representing the observed and the missing part of \( y_i \), respectively. To keep the notation simple, we will use square brackets to denote a generic distribution; for example, \([y_i, r_i]\) denotes the joint distribution \( f_{yr}(y_i, r_i) \), while \([y_i \mid r_i]\) denotes the conditional distribution \( f_{yr}(y_i \mid r_i) \).

Little (1993) and Little and Rubin (2002) distinguish two classes of models to handle (potentially) non-ignorable missing data: selection and-pattern mixture models. Differences between the two entail the factorization of the joint distribution of the observed and the missing data process. Selection
models are based on the following factorization:

\[
[r_i, y_i \mid \Phi, \gamma] = [r_i \mid y_i, \gamma][y_i \mid \Phi], \quad i = 1, \ldots, n
\]  

(7)

The terminology is due to [Heckman (1976)] and comes from the fact that the conditional density \([r_i \mid y_i, \gamma]\) can be seen as describing a selection mechanism for each unit to either continue or leave the study.

In pattern-mixture models (PMM), see eg [Little (1993)], the joint distribution of \(y_i\) and \(r_i\) is factorized by specifying a marginal model for the missing data process and a model for the complete longitudinal responses, conditional on the drop-out indicator:

\[
[r_i, y_i \mid \Phi, \gamma] = [y_i \mid r_i, \Phi][r_i \mid \gamma], \quad i = 1, \ldots, n.
\]  

(8)

A possible rationale for pattern-mixture models is that each subject has its own propensity to drop-out and the length of the individual sequence is related to such a propensity. Individuals with similar propensities share some common observed and/or unobserved features and the model for the whole population is given by a mixture over these patterns. Further modelling alternatives are discussed by [Little (1995)] and [Rizopoulos and Lesaffre (2014)].

A clear drawback of pattern-mixture models is represented by the weak identifiability of model parameters: the observed data do not provide enough information to completely identify the distribution of the outcome conditional on the patterns of non-response. As noticed by [Roy (2003)], this modelling approach could (potentially) lead to a large number of sparse strata and to biased estimates due to misclassification; a potential solution is to use a model with a reduced number of classes. For this purpose, we will consider the proposal described by [Roy (2003)] and [Roy and Daniels (2008)], and introduce an ordered, latent, categorical variable to represent the effects of the individual-specific propensity to drop-out on model parameter estimates.

Let us remind that \(Y_{it}\) and \(R_{it}\) are the response variable and the missing data indicator for the \(i\)-th individual \(i = 1, \ldots, n\) at occasion \(t = 1, \ldots, T\); \(R_{it} = 1\) if individual \(i\) has not been observed at \(t\). In this case, \(T_i = T - \sum_{t=1}^{T} R_{it}\) in-
indicates the number of measurements available for the $i$-th unit, $i = 1, \ldots, n$. The potential presence of informative drop-out is described by a discrete latent variable with ordered categories, say $\zeta_i = (\zeta_{i1}, \ldots, \zeta_{iG})$, used to represent different propensities to drop-out from the study. Each element $\zeta_{ig}, g = 1, \ldots, G$, is equal to 1 if the $i$-th individual belongs to the $g$-th class and is equal to zero otherwise. Obviously, both $\zeta_i$ and $G$ depend on the chosen quantile $\tau \in (0, 1)$, but to make things easier, we will suppress this dependence.

Sample units belonging to the same latent drop-out (LDO) class are assumed to share some common latent characteristics which explain differences, in terms of the covariates’ effects on the response variable, with respect to individuals from a different class. The probability of having a drop-out propensity lower than a given level is modelled as a monotone function of the length of the individual sequence $T_i$; an increase in the drop-out time progressively increases/decreases the probability of being in one of the first $g$ latent classes, $g = 1, \ldots, G$. The following ordinal regression model holds:

$$
\Pr\left(\sum_{l=1}^{g} \zeta_{il} = 1 \mid T_i\right) = \frac{\exp(\lambda_{0g} + \lambda_i T_i)}{1 + \exp(\lambda_{0g} + \lambda_i T_i)} \quad (9)
$$

under the inequality constraints $\lambda_{01} \leq \cdots \leq \lambda_{0G-1}$. For a given quantile $\tau$, the following $lqHMM$ is specified for units in the $g$-th LDO class:

$$
Q_{\tau}(y_{it} \mid s_{it}, \zeta_{ig}, = 1, \tau) = x_{it}' \beta + z_{it}' b_g + w_{it}' \alpha_{s_{it}}, \quad (10)
$$

where $\beta$ is a $p$-dimensional vector of fixed parameters and $x_{it}$ the corresponding set of covariates. Parameters $b_g$’s and $\alpha_{s_{it}}$’s describe random deviations from the marginal model due to time-constant (drop-out dependent) and time-varying unobserved features, respectively. We assume that, conditional on the hidden state occupied at a given time occasion and the LDO class membership, longitudinal observations from the same individual are independent (local independence assumption); furthermore, conditional on these latent quantities, the longitudinal and the missing data process are no longer dependent. As before we assume that, conditional on $S_{it} = h$ and $\zeta_{ig} = 1$, 

longitudinal measures $y_{it}$ are realizations of an AL distributed random variable $Y_{it}$, with scale and skewness parameters denoted by $\sigma$ and $\tau$, respectively. The location parameter of such a conditional distribution, which also corresponds to the mode and the $\tau$-th quantile, is defined according to expression (10), that is

$$\mu[S_{it} = h, \zeta_{ig} = 1] = x'_it \beta + z'_it b_g + w'_it \alpha_{sit}. \quad (11)$$

Let us denote by $\psi$ the global set of longitudinal model parameters; the joint conditional distribution of the individual sequence $(y_{i1}, ..., y_{iT_i})$ is defined by the following expression

$$f_{y_{i|s}}(y_i | s_i, \zeta_{ig} = 1, \psi, \tau) = \prod_{t=1}^{T_i} f_{y_{it|sb}}(y_{it} | s_{it}, \zeta_{ig} = 1, \psi, \tau) = \left[\frac{\tau(1-\tau)}{\sigma}\right]^T \exp\left\{-\sum_{t=1}^{T_i} \rho_{\tau}\left[\frac{y_{it} - \mu[S_{it}, \zeta_{ig} = 1]}{\sigma}\right]\right\} \quad (12)$$

The (conditional) model for the longitudinal process is then associated to the model describing the distribution for the LDO class variable. Let $\Phi = (\psi, \delta, Q, \lambda)$ be the global set for the parameters of interest and let $\gamma$ be the vector of parameters characterizing the distribution of $T_i$. Based on the modelling assumptions we have introduced so far, and for a generic quantile $\tau \in (0, 1)$, the individual observed data likelihood is given by

$$L_{(i)}(\Phi, \gamma | y_{i}, T_i, \tau) = \int_{\gamma_{m}} \sum_{s_i} \sum_{g=1}^{G} f_{y_i|s, \zeta}(y_i | s_i, \zeta_{ig} = 1, \psi, \tau) f_s(s_i | \delta, Q, \tau) \times \pi_{ig}(T_i | \lambda, \tau) f_T(T_i | \gamma, \tau) dy_{i|m}, \quad (13)$$

where $\pi_{ig}(T_i | \lambda, \tau) = f_{\zeta|T}(\zeta_{ig} = 1 | T_i, \lambda, \tau)$ is the conditional probability for the $i$-th individual to belong to the generic $g$-th LDO class $g = 1, ..., G$. This quantity can be straightforwardly computed as the difference between two adjacent cumulative logits in (9), see Agresti (2010). After conditioning on $\zeta_i$, the longitudinal outcome depends on the dropout process only through the LDO class variable; this allows to directly integrate out the missing data from the above expression. Furthermore, by
assuming that parameters $\Phi$ and $\gamma$ are separate, the marginal distribution of the missing data process $f_T(T_i \mid \gamma, \tau)$ can be left completely unspecified and inference on parameters of interest can be based on the following (individual) conditional observed data likelihood

$$L_{(i)}(\Phi \mid T_i, \tau) = \sum_{s_i} \sum_{g=1}^{G} f_{y_s}(y_i \mid s_i, \zeta_{ig} = 1, \psi, \tau) f_s(s_i \mid \delta, Q, \tau) \pi_{ig}(T_i \mid \lambda, \tau).$$

(14)

By looking at the above expression, it is easy to observe that (14) is a particular version of the likelihood (6) we have derived for the lqmHMM; in fact, the LqHMM+LDO representation reduces to the lqmHMM one when the prior $\pi_{ig}(T_i \mid \lambda, \tau) = \pi_g, \forall i = 1, \ldots, n$ and $g = 1, \ldots, G$, that is if the distribution of the time-constant random parameters is not a function of the number of available measurements for each individual. In that context, mixture components are introduced as a way to approximate intractable integrals on (possibly) continuous, individual-specific, random parameters that represent time-constant unobserved heterogeneity. Here, mixture components are meant to describe time-constant unobserved features that are directly related to the drop-out process and that can be interpreted as grouping individuals with a similar drop-out propensity. This allows us to explain departures from the homogeneous model associated to completers (ie lqHMM) that may be produced by the missingness process. In other words, the lqHMM (Farcomeni, 2012) and the lqmHMM, introduced in Section 2 can be both thought of as MAR versions of the (MNAR) model we obtain by explicitly modelling the dependence between the discrete mixing distribution and the number of measurements available for each unit.

As pointed out before, parameter estimates can be derived via an indirect maximum likelihood approach, by considering a generalization of the basic EM algorithm (Dempster et al., 1977) that allows to account for the presence of both dynamic and static finite mixtures, such as the one described by Maruotti and Rydén (2009) in the context of standard mixed HMMs. In the following section we will give details of the estimation algorithm.
4 Computational details

Parameter estimates for the \textit{lqmHMM} and the \textit{lqHMM+LDO} can be obtained by considering the Baum-Welch algorithm, see MacDonald and Zucchini (1997), for standard HMMs. Obviously, this algorithm should be adapted and extended to deal with quantiles and with discrete random coefficients. As before, the dependence of model parameters and indicator variables from the specific quantile \(\tau\) will be suppressed for ease of notation. Furthermore, we will refer to LDO classes with the generic term mixture components, to align the terminology of \textit{lqmHMM} and \textit{lqHMM+LDO}; in this perspective, \(\pi_g = \pi_{ig}(T_i \mid \lambda, \tau)\). Let \(u_i(h) = \mathbb{I}[S_{it} = h]\) be the indicator variable for the \(i\)-th subject being in the \(h\)-th state at time occasion \(t\) and let \(u_i(k,h) = \mathbb{I}[S_{it-1} = k, S_{it} = h]\) be equal to 1 if unit \(i\) moves from the \(k\)-th state at time occasion \(t-1\) to the \(h\)-th one at time occasion \(t\). Let \(\zeta_{ig}\) be the indicator variable for the \(i\)-th unit coming from the \(g\)-th component of the finite mixture. Let \(\Phi = (\psi, \delta, Q, \pi)\) be the global set of all model parameters. Parameter estimates can be obtained by means of an EM algorithm (Dempster et al., 1977) by starting from the definition of the complete data log-likelihood:

\[
\ell_c(\Phi \mid \tau) \propto \sum_{i=1}^{n} \left\{ \sum_{h=1}^{m} u_{i1}(h) \log \delta_h + \sum_{t=2}^{T_i} \sum_{h,k=1}^{m} u_{it}(k,h) \log q_{kh} + \sum_{g=1}^{G} \zeta_{ig} \log \pi_g + \right. \\
- T_i \log(\sigma) - \sum_{t=1}^{T_i} \sum_{h=1}^{m} \sum_{g=1}^{G} u_{it}(h) \zeta_{ig} \rho_{\tau} \left[ \frac{y_{it} - \mu_{it}[S_{it} = h, b_g]}{\sigma} \right]. \]

(15)

As it is usually done in the hidden Markov model framework, computation of parameter estimates can be greatly simplified by considering forward and backward variables, as in the Baum-Welch algorithm (Baum et al., 1970; Welch, 2003). In the present framework, forward variables, \(u_{it}(h, g \mid \tau)\), are defined as the joint density of the longitudinal measures up to time \(t\), for a generic individual ending up in the \(h\)-th state, conditional on the \(g\)-th
mixture component:

\[ a_{it}(h, g \mid \tau) = f[y_{1:t}, S_{it} = h \mid g, \tau]. \quad (16) \]

Similarly, the backward variables \( b_{it}(h, g \mid \tau) \) represent the probability of the longitudinal sequence from occasion \( t + 1 \) to the last observation \( T_i \), conditional on being in the \( h \)-th state at time \( t \) and the \( g \)-th mixture component:

\[ b_{it}(h, g \mid \tau) = f[y_{it+1:T_i} \mid S_{it} = h, g, \tau]. \quad (17) \]

Both terms can be computed recursively:

\[
\begin{align*}
    a_{i1}(h, g \mid \tau) &= \delta_{h} f_{y_{i1} \mid \tau} [y_{i1} \mid S_{i1} = h, g, \tau] \\
    a_{it}(h, g \mid \tau) &= \sum_{k=1}^{m} a_{it-1}(k, g \mid \tau) q_{kh} f_{y_{it} \mid \sigma} [y_{it} \mid S_{it} = h, g, \tau] \\
    b_{iT_i}(h, g \mid \tau) &= 1 \\
    b_{it-1}(h, g \mid \tau) &= \sum_{k=1}^{m} b_{it}(k, b_{g}) q_{kh} f_{y_{it} \mid \sigma} [y_{it} \mid S_{it} = k, g, \tau].
\end{align*}
\]

Once these variables have been defined, the E-step can be performed straightforwardly. By considering the indicator variables for both the hidden Markov process and the finite mixture as missing data, conditional expected values, given the observed data and the current parameter estimates, can be easily computed according to the following expressions:

\[
\begin{align*}
    \hat{u}_{it}(h \mid \tau) &= \frac{\sum_{g} a_{it}(h, g \mid \tau) b_{it}(h, g \mid \tau) \pi_{g}}{\sum_{h} \sum_{g} a_{it}(h, g \mid \tau) b_{it}(h, g \mid \tau) \pi_{g}}, \\
    \hat{u}_{it}(k, h \mid \tau) &= \frac{\sum_{g} a_{it-1}(k, g \mid \tau) q_{kh} f_{y_{it} \mid \sigma} [y_{it} \mid S_{it} = h, g, \tau) b_{it}(h, g \mid \tau) \pi_{g}}{\sum_{hk} \sum_{g} a_{it-1}(k, g \mid \tau) q_{kh} f_{y_{it} \mid \sigma} [y_{it} \mid S_{it} = h, g, \tau) b_{it}(h, g \mid \tau) \pi_{g}}, \\
    \hat{\zeta}_{ig}(\tau) &= \frac{\sum_{h=1}^{m} a_{iT_i}(h, g \mid \tau) \pi_{g}}{\sum_{l=1}^{G} \sum_{h=1}^{m} a_{iT_i}(h, l \mid \tau) \pi_{l}}.
\end{align*}
\]
\[ u_{it}(h \mid g, \tau) = \frac{a_{it}(h, g \mid \tau)b_{it}(h, g \mid \tau)\pi_g}{\sum_{k=1}^{m} a_{it}(k, g \mid \tau)b_{it}(k, g \mid \tau)\pi_g}. \]  

(18)

Here, the first three rows correspond to the posterior expected values of state and component membership indicators described so far, while the last one represents the conditional posterior probability of being in state \( h \) at time occasion \( t \), given the individual belongs to the \( g \)-th component of the finite mixture.

In the M-step, model parameter estimates are derived by maximizing the expected complete data log-likelihood with respect to model parameters. Assuming the separability of the parameter spaces in the longitudinal and the missing data process, the maximization can be partitioned into independent sub-problems, thus consistently simplifying the computational complexity.

Standard results for basic HMMs hold for initial and transition probability estimates:

\[ \hat{\delta}_h = \frac{\sum_{i=1}^{n} \hat{u}_{i1}(h \mid \tau)}{n}, \quad \hat{q}_{kh} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \hat{u}_{it}(k, h \mid \tau)}{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \sum_{m}^{m} \hat{u}_{it}(h, k \mid \tau)}. \]  

(19)

Longitudinal model parameters \( \psi \) are estimated by solving a weighted estimating equation, where weights are given by the posterior probabilities of the hidden Markov process and the finite mixture. In detail, such updates are obtained as the solutions of to M-step equations

\[ \arg \min_{\psi} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \sum_{h=1}^{m} \sum_{g=1}^{G} \hat{\zeta}_{ig}(\tau)\hat{u}_{it}(h \mid g, \tau)\rho_r \left[ \frac{y_{it} - \mu_{it}[S_{it} = h, b_g]}{\sigma} \right] \]  

(20)

The computation can be greatly simplified by extending the block algorithm proposed by Farcomeni (2012) for the linear quantile HMM. To describe, three different steps have to be alternated. Fixed parameters \( \beta \) are estimated by solving

\[ \arg \min_{\beta} \sum_{i=1}^{n} \sum_{t=1}^{T_i} \sum_{h=1}^{m} \sum_{g=1}^{G} \hat{\zeta}_{ig}(\tau)\hat{u}_{it}(h \mid g, \tau)\rho_r \left[ \hat{y}_{it} - \hat{x}_{it} \beta \right] \]  

(21)
where $\tilde{y}_{it} = [y_{it} - z'_{it}\hat{b}_g - w'_{it}\hat{\alpha}_h]$.

In the second step, state-dependent parameters $\alpha_h, h = 1, ..., m$ are updated via

$$\arg\min_{\xi_h} \sum_{i=1}^n \sum_{t=1}^{T_i} \sum_{h=1}^m \sum_{g=1}^G \hat{\zeta}_{ig}(\tau) \hat{u}_{it}(h \mid g, \tau) \rho_r [\tilde{y}_{it} - \alpha_h]$$

(22)

where $\tilde{y}_{it} = [y_{it} - x'_{it}\hat{\beta} - z'_{it}\hat{b}_g]$ in the case of $lqmHMM$ and $\tilde{y}_{it} = [y_{it} - x'_{it}\hat{\beta}]$ for the $lqHMM+LDO$.

Finally, the locations $b_g$ for the $lqHMM$ are computed by solving

$$\arg\min_{b_g} \sum_{i=1}^n \sum_{t=1}^{T_i} \sum_{h=1}^m \sum_{g=1}^G \hat{\zeta}_{ig}(\tau) \hat{u}_{it}(h \mid g, \tau) \rho_r [\tilde{y}_{it} - z_{it}b_g]$$

(23)

with $\tilde{y}_{it} = [y_{it} - x'_{it}\hat{\beta} - w'_{it}\hat{\alpha}_h]$.

Prior probabilities are, as usual, updated by

$$\hat{\pi}_g = \frac{1}{n} \sum_{i=1}^n \hat{\zeta}_{ig}(\tau).$$

(24)

in the case of $lqmHMMs$, that is when the prior does not depend on $T_i$. For the $lqHMM+LDO$, drop-out specific parameters are estimated via a constrained optimization of the weighted likelihood for a standard cumulative logit model, with weights given by the posterior probabilities of the LDO class variable:

$$\arg\max_{\lambda} \sum_{g=1}^G \hat{\zeta}_{ig}(\tau) \left\{ \left[ \frac{\exp(\lambda_0 + \lambda_g T_i)}{1 + \exp(\lambda_0 + \lambda_g T_i)} \right] - \left[ \frac{\exp(\lambda_0 + \lambda_{g-1} T_i)}{1 + \exp(\lambda_0 + \lambda_{g-1} T_i)} \right] \right\}.$$  

(25)

The E- and the M-step of the algorithm are iterated until convergence, that is until the (relative) difference between subsequent likelihood values is lower than an arbitrary small amount $\varepsilon$. Penalized likelihood criteria, such as BIC, can be used to jointly identify the best number of mixture components/LDO classes and hidden states. Moreover, as it typically happens in the linear quantile mixed model framework, standard errors for parameter estimates can be obtained by using a non-parametric block bootstrap approach, see eg
5 Simulation study

To investigate the performance of the proposed modelling approach, we have implemented the following simulation study, composed by two different scenarios. First, to analyze the empirical behaviour of the lqmHMM representation, we consider completely observed longitudinal responses (referred to as Scenario 1 in the following). After that, we introduce a non ignorable missing data process, considering two different intensities for the relationship between drop-out and time spent into the study, to analyse the quality of results under the latent drop-out class representation (Scenario 2 in the following).

5.1 Scenario 1: complete individual sequences

Under this completely observed scenario, data have been generated from a two state lqmHMM \((m = 2)\) with initial/transition probabilities given by

\[
\delta = (0.7, 0.3) \quad \text{and} \quad Q = \begin{pmatrix}
0.8 & 0.2 \\
0.2 & 0.8
\end{pmatrix}.
\]

The following longitudinal (conditional) regression model holds for the \(h\)-th state of the Markov chain:

\[
Y_{it} = \alpha_h + (2 + b_i) x_{it1} - 0.8 x_{it2} + \varepsilon_{it}
\]

where \(x_{it1} \sim \text{Unif}[-10, 10]\) and \(x_{it2} = x_{i2} \sim \text{Bin}(1, 0.5)\) is a time-constant binary variable. Different probability distributions have been considered for the random errors, \(\varepsilon_{it}\): a standard Gaussian and a Chi-square distribution with \(\nu = 2\) degrees of freedom, to allow for the presence of skewed data. The time-constant random slopes \(b_i\) capture individual departures from the marginal effect \(\beta_1 = 2\) and represent iid draws from a standard Gaussian and Student’s \(t_3\) distribution, respectively; state-dependent intercepts have been
set to $\alpha_1 = 100$ and $\alpha_2 = 110$. The simulation scheme has been derived from Liu and Bottai (2009).

For each scenario, $B = 250$ samples have been generated. Varying sample size ($n = 100, 200$) and length of the sequence ($T = 5, 10$) have been considered; model parameters have been estimated for three different quantile levels $\tau = (0.25, 0.50, 0.75)$. While we have considered balanced designs only, the algorithm can be readily applied to unbalanced designs as well.

For each simulated dataset, we have estimated a $lqmHMM$ with $m = 2$ hidden states, considered a varying number of mixture components, ($G = 1, \ldots, 15$) and retained the best BIC model. To evaluate the model performance, we have calculated the root mean square error (RMSE) associated to each model parameter. Simulation results for the longitudinal model parameters are reported in Table 1.

Table 1 ABOUT HERE

As we may notice, the accuracy is slightly higher when we consider parameters estimates for the center of the response variable distribution when compared to those for the first and the third quartile; such a difference, however, attenuates with increasing sample size and, more substantially, with increasing number of measurement occasions. As expected, when we look at the tails of the distribution, a smaller amount of information is available and this, somehow, influences the “quality” of parameter estimates. The results obtained for $\tau = 0.25$ are worth of some space for discussion. As we may notice from the linear predictor defined in equation (27), the first quartile of the conditional distribution is mainly associated to units in the first state of the Markov chain and, therefore, the second state intercept $\alpha_2$ is estimated with some additional bias due to a sparse (imprecise) information. However, with increasing size (both in terms of $n$ and $T$), units in the second state are more frequently observed and the quality of results improves.

As expected, parameter estimates for the time-constant, discrete covariate result less accurate than those for the time-varying continuous covariate, even if the difference seems to decrease with increasing $n$ and $T$. When we look
at variance estimate for the individual random parameter, we observe a reduced RMSE when $b_i$ is drawn from a standard Gaussian, while a consistent reduction of the accuracy is found in the $t_3$ case, even if the quality increases when more information, in terms of $n$ and $T$, is available.

It is interesting to notice that in none of the simulation scenarios, the error terms have been drawn from an AL distribution; therefore, the fitted model is always misspecified. Based on these results, we can conclude that such an assumption is rather a working hypothesis to place parameter estimation for standard quantile regression models in a maximum likelihood perspective.

5.2 Scenario 2: partially complete individual sequences

Under this scenario, data have been generated from a linear quantile HMM with two hidden states ($m = 2$) and three latent drop-out classes ($G = 3$). We assume the study has been designed to collect longitudinal measures on a sample of $n = 100, 200$ units at $T = 5, 10$ equally spaced measurement occasions, but some individuals drop-out prematurely. The variable $T_i$ has been drawn from a discrete distribution with $\Pr(T_i = j) = 1/(T - 1), j = 2, ..., T$.

Initial and transition probabilities of the hidden Markov chain have been set to

$$
\delta = (0.7, 0.3) \quad \text{and} \quad Q = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix},
$$

(28)

while two different sets of $\lambda$ parameters have been considered in the analysis. The former, that is $\lambda_{01} = 5, \lambda_{02} = 8.5, \lambda_1 = -1.1$, has been chosen so that class probabilities are strongly related to the drop-out time (“high informative drop-out scenario”). The latter set, that is $\lambda_{01} = 1, \lambda_{02} = 2.75, \lambda_1 = -0.3$, implies that the drop-out time does not strongly influence the LDO class membership (“low informative drop-out scenario”).

For the longitudinal observations, the following regression model holds for
the \( h \)-th state of the Markov chain and the \( g \)-th latent drop-out class:

\[
Y_{it} = \alpha_h + (2 + b_g) x_{it1} - 0.8 x_{it2} + \varepsilon_{it} \tag{29}
\]

where \( x_{it1} \sim \text{N}(1, 3) \) and \( x_{it2} \sim \text{Unif}[0, 10] \). State-dependent intercepts have been set to \( \alpha_1 = 100 \) and \( \alpha_2 = 102.5 \), while latent drop-out dependent parameters have been set to \( b_1 = 0.5, b_2 = 1.5, b_3 = 3 \). The difference between the state-specific intercepts, that is \( (\alpha_1, \alpha_2) \), has been set to a lower value than that one considered in the simulation study described in Chapter ?? to verify whether aliasing between the \( \alpha \)'s and the \( b_g \)'s may occur in such a (not strongly distinguished) scenario. Different probability distributions have been considered for the random errors: a standard Gaussian, a Student’s t distribution with \( \nu = 3 \) degrees of freedom and a Chi-square distribution with \( \nu = 2 \) degrees of freedom, where the latter two distributions allow for heavy tailed and skewed data, respectively.

For each scenario, \( B = 250 \) samples have been generated and model parameters have been estimated for three different quantile levels \( \tau = (0.25, 0.50, 0.75) \). To evaluate the model performance, the root mean square error (RMSE) for each model parameter has been computed. Simulation results are reported in Tables 2-3.

Tables 2-3 about here

As expected, in all the considered simulation scenarios, the quality of parameter estimates increases as the sample size and/or the number of available measures increases. The fixed parameter in the longitudinal data model is always estimated with higher accuracy than the parameters associated to time-constant or time-varying latent variables. Lower RMSEs are observed when modelling the center of the distribution with respect to the corresponding tails, since, in this latter case, a reduced amount of information is available. Based on the linear predictor we have defined, the response variable distribution is almost positive skewed and this can explain why, at \( \tau = 0.75 \), parameters are estimated with higher bias than at \( \tau = 0.25 \). However, the
gap tends to reduce when higher sample sizes and repeated measurements are available. Generally, better results are obtained in the case of Gaussian errors when compared to the heavy tailed and the skewed case, except for the first quartile ($\tau = 0.25$), where $\chi^2$ distributed random errors ensure more information that lead to a higher quality estimates. Based on the parameters we have considered for the simulation, the first state of the hidden Markov chain is the most likely one; as a result $\alpha_1$ is always estimated with some more accuracy than $\alpha_2$.

With respect to the LDO-dependent parameters, it is worth to notice that, for some of the simulation scenarios we have considered, the algorithm faces some difficulties in recovering the true number of LDO classes, especially when a reduced number of repeated measurements is available ($T = 5$), as it is clear by looking at the corresponding RMSEs. In the “low-informative” drop-out scenario, such an issue seems to be overcome when the sample size increases ($n = 200$) for $\tau = 0.25$ and $\tau = 0.50$. For $\tau = 0.75$, a higher number of repeated measurements is, instead, required because of the reduced information that are available in the right region. Similarly, in the “high-informative” drop-out scenario, the problem seems to be solved only when a higher number of measurements is available. In this case, the effect of the missing data process on the observed longitudinal response is higher and this points to the need of longer longitudinal sequences to better approximate the “real” data generating process.

6 Real data application: CD4 data

6.1 Data Description

We illustrate the proposed models by re-analysing the CD4 cell count data, see, among others, Zeger and Diggle (1994). Data come from the Multicenter AIDS Cohort Study (MACS); nearly 5000 gay and bisexual men from Baltimore, Pittsburgh, Chicago and Los Angeles were followed since 1984 to analyse HIV progression. On average, the infection translates into disease after 11 years from contraction; therefore, quite long follow-up periods may
be necessary to observe the progression in a large cohort of individuals. HIV virus destroys T-lymphocytes called CD4 cells, which play a vital role in immune function. The progression of HIV can be assessed by measuring the number of CD4 cells, which, on average, tend to decrease throughout the incubation period.

The analysed data includes 371 men who were seronegative at entry and seroconverted during the follow up. Details of the study design can be found in Kaslow et al. (1987). In our analysis, we consider 2376 measurements from 369 seroconverters, observed in a window ranging from 3 years before to 6 years after seroconversion. Data from two subjects are dropped due to missing values in the covariates. Each individual has been observed from a minimum of 1 to a maximum of 12 measurement occasions.

We start the analysis by assuming a non-informative drop-out scenario; this assumption will be later on relaxed by introducing the LDO representation. We are interested in determining the effect of a given set of covariates on the response variable distribution, while controlling for sources of unobserved heterogeneity. The set of covariates includes: years since seroconversion (negative values indicate the current CD4 measurement was taken before the seroconversion), age at seroconversion (centered around 30), smoking (packs per day), recreational drug use (yes or no), number of sexual partners, depression symptoms as measured by the CESD scale (larger values indicate more severe symptoms). The analysis was conducted on the log transformed CD4 counts, that is log(1 + CD4 count).

6.2 Linear quantile mixed hidden Markov model

As described by Diggle et al. (2002) the cell count is, approximately, constant until seroconversion and then decreases, more quickly at first and less afterwards. To properly model the evolution of individual trajectories over time, we consider both time-varying and time-constant individual-specific random parameters. From a preliminary exploratory analysis and by looking at the fit of a number of regression models, we decide to focus on the \textit{lqmHMM}
below:
\[ Q_\tau(y_{it} \mid s_{it}, b_i) = x_{it}' \beta(\tau) + z_{it}' \delta \beta(\tau) + w_{it}' \alpha_s(\tau) \]

where \( x_{it} \) includes two continuous covariates (years since seroconversion and age), the dummy variable drug (baseline: no) and three discrete variables (packs of cigarette per day, number of sexual partners and CESD score). The vectors associated to (time-constant and time-varying) random parameters, \( z_{it} \) and \( w_{it} \), include the years since seroconversion and a column of ones, respectively. That is, the resulting model allows for a time-varying random intercept, a time-varying random slope associated to years since seroconversion and a set of fixed effects.

We fit the proposed model for a varying number of hidden states and mixture components and for three different quantiles of the distribution, namely \( \tau = (0.25, 0.50, 0.75) \). To reduce the chance of being trapped in local maxima we adopt a multi-start strategy. Following Farcomeni (2012), a first deterministic starting solution is obtained by setting prior and transition probabilities to \( \delta_h = 1/m \) and \( q_{kh} = (1 + sI(h = k))/(m + s) \), for a suitable constant \( s > 0 \). Such a constant acts in a way that \( s \to 0 \Rightarrow q_{kh} \to 1/m \); as \( s \) grows the \( Q \) matrix tends to the identity matrix \( I_m \). Fixed parameters are obtained by fitting a linear QR model for independent observations, while time-varying and time-constant random parameters are set by adding a suitable constant to the corresponding marginal effects; in our implementation, the constant is referred to the locations from the Gaussian quadrature approximation with \( m \) and \( G \) quadrature points, respectively. Random starting points have been obtained by randomly perturbing the deterministic starting solution. For each combination of hidden states and mixture components, we have considered 30 starting points and retain the solution with the best BIC, see Table 4.

Obtained results suggest to select a model with \( m = 4 \) hidden states at all the considered quantiles, highlighting the presence of a quite strong time-varying unobserved heterogeneity. Time-constant random slopes, capturing
individual-specific effects of the time since seroconversion on the CD4 cell counts, are approximated by a number of mixture components that decreases as we move through the right tail of the response variable distribution. Based on results reported in Table 4, we set $G = 5, 4, 3$ for $\tau = 0.25, 0.50$ and $\tau = 0.75$, respectively.

In Table 5, we show the estimated parameters for the longitudinal data model, with the corresponding 95% confidence intervals (within brackets), computed by considering $B = 1000$ block bootstrap samples.

We may notice that, for all the analysed quantiles, age seems to play a minor role, while the packs of cigarettes and the number of sexual partners have a positive and significant effect on the log count of CD4 cells. The recreational use of drugs acts positively on the tails of the response variable distribution, while such an effect is negligible at the median level, as it is clear by looking at the corresponding confidence intervals. As suggested by Zeger and Diggle (1994), the positive effect associated to these “risk” factors may indicate immune response stimulation or, simply, the effect of selection bias with healthier men that stay longer in the study and choose to continue their usual practices. More severe depression, summarized by higher values for the CEDS score, implies a slight decrease in the number of T-lymphocytes. Also for the time since seroconversion the effect is negative; that is, with increasing lag from the date of seroconversion, the number of CD4 cells decreases. Such an effect attenuates when moving towards higher quantiles of the response variable distribution, thus suggesting that the progression of the human immunodeficiency virus is slower for healthier men. The estimated variance of the random parameter results significantly different from zero, pointing out the presence of substantial individual-specific departures from the homogeneous effect of $\text{Time}_{\text{sero}}$ on the T-lymphocytes count. Also, this variability reduces when moving towards the right tail of the response distribution.

State-dependent intercepts identify increasing baseline CD4 levels (higher
locations correspond to a higher number of cells) and, as expected, the corresponding estimates increase when moving towards the right tail of the response variable distribution.

Table 6 reports the estimated initial and transition probabilities for the hidden Markov chain. The combination of these results with the intercept values reported in Table 6 helps understand the evolution of CD4 cell counts over time, holding fixed the covariate effects.

For all the analysed quantiles, the estimated initial probabilities suggest that most of the sample units start the study with an intermediate level of CD4 cell counts ($\delta_2 + \delta_3 > 0.75$), while few observations start with more extreme (lower or higher) levels. For $\tau = 0.50$ and $\tau = 0.75$, transitions between hidden states over time are quite unlikely ($q_{hh} > 0.8$) and, if any transition is observed, units tend to move towards hidden states characterized by lower levels of T-lymphocytes. For $\tau = 0.25$, we may observe a slightly different evolution of the response over time. Estimated initial and transition probabilities highlight that for sicker men the number of CD4 cells in the blood tends to repeatedly increase and decrease over the follow-up time, and this is particularly evident for “lower” hidden states. It is interesting to notice that transitions towards the first hidden state (with the lower CD4 log-count) are unlikely ($\sum_{k=1}^{m} q_{k1} < 0.15$) and, if any transition is observed, with high probability individuals in the next occasion move towards states characterized by higher CD4 cell count levels ($q_{11} = 0.284$), that is, the observed decrease is just temporary.

6.3 Linear quantile hidden Markov model with latent drop-out classes

As already discussed, each individual has been observed from a minimum of 3 years before to 12 years after seroconversion: some of them leave the study before the planned completion time, thus, presenting incomplete information. Table 7 shows the number of individuals remaining in the study at each visit.
As it can be noticed, only a small proportion has been observed until the end of the follow-up time.

Table 7

In this section, we propose to account for the presence of a possible informative missing data process. By looking at figure 1, which represents the distribution of the response variable at each visit as a whole and stratified by whether or not units drop-out from the study between the current and the subsequent time occasion, a general reduction of the CD4 level over time can be noticed, which is coherent with the results obtained under the MAR assumption (i.e. the lqmHMM specification). However, it is also clear that units dropping-out from the study between occasions \( t \) and \( t + 1 \) present CD4 cell count levels that are quite lower than those registered for units that remain in the study beyond time \( t + 1 \), especially during the first few measurement occasions.

Figure 1

This suggests the presence of some dependence between the longitudinal and the missing data process which should be taken into consideration to avoid potential bias in the parameter estimates or, at least, to have a clearer interpretation of the results. For this purpose, we estimate a \( lqHMM+LDO \) and compare the results with those we obtain with the corresponding MAR model, i.e. the \( lqmHMM \) presented in section 6.2.

More in detail, the following \( lqHMM+LDO \) classes is considered

\[
Q_\tau(y_{it} | s_{it}, \zeta_{itg} = 1) = x'_{it}\beta(\tau) + z'_{it}b(\tau) + w'_{it}\alpha_{s_{it}(\tau)}
\]

where the vectors \( x_{it}, z_{it} \) and \( w_{it} \) have the same meanings than before. Therefore, the resulting model considers a set of fixed effect covariates, a time-varying random intercept and a time-constant random slope for years since seroconversion, which varies with the LDO class.

We fit the proposed model for a varying number of states and of mixture components, for three quantile levels, \( \tau = (0.25, 0.50, 0.75) \). To avoid local
maxima, hidden Markov and longitudinal model parameters are initialized by following the same multi-start strategy described for the \textit{lqmHMM} in section 6.2. Initial values of parameters related to the missing data process are set by estimating an ordered logit model on the response variable obtained by discretizing the observed time to drop-out and randomly perturbing such categories to avoid singularities. For each combination of hidden states and LDO classes, we consider 30 starting points and retain the solution with the best BIC value. Results are reported in Table 8.

We select the model with $m = 5$ hidden states and $G = 5$ LDO classes when modelling the first quartile of the distribution ($\tau = 0.25$); for the median and the third quartile, solutions with $m = 4$ and $G = 4$ return the lowest BIC values. However, by looking at the parameter estimates for the LDO class model at $\tau = 0.75$, two $\lambda_{0g}$ parameters do not significantly differ from zero and the corresponding confidence intervals overlap. Therefore, we decide to perform the research of the optimal number of classes within the set of models with $G = 1, \ldots, 3$ to avoid spurious solutions; as a result, for $\tau = 0.75$, we retain the solution with $m = 4$ hidden states and $G = 3$ latent drop-out classes.

We report in Table 9 the estimated parameters for the longitudinal data model, with the corresponding 95% confidence intervals (within brackets) computed by means of $B = 1000$ block bootstrap samples.

Regarding the fixed covariate effects, we may notice that, for all the analysed quantiles, the age and the recreational use of drugs seems to play no role in determining the evolution of the response over time, while the higher the number of smoked cigarettes and of sexual partners the higher is the level of T-lymphocytes in the blood. As we have discussed before, these positive effects may be due to immune stimulation or to selection bias with healthier men that stay longer into the study and decide to continue their usual practices. On the other hand, the CD4 cell counts reduces when more severe depression symptoms are registered and when the time since seroconversion increases. This latter effect slightly reduces when moving towards the right
tail of the response variable distribution, highlighting the presence of healthier men in the sample having a slower progression of the illness over time. As before, state-dependent intercepts identify increasing CD4 baseline levels (“higher” hidden states correspond to a higher number of T-lymphocytes in the blood) and, as expected, higher values are observed when moving from the left to the right tail of the response distribution.

By matching these results with the estimated initial and transition probabilities reported in Table 10, deeper information on the evolution of individual trajectories can be obtained. Focusing the attention on \(\tau = 0.50\) and \(\tau = 0.75\), it is clear that most of the individuals in the sample starts the study with an intermediate level of CD4 cell counts \((\delta_2 + \delta_3 > 0.7)\); transitions between states are quite unlikely \((q_{hh} > 0.8)\) and, if any transition is observed, individuals tend to move towards hidden states characterized by lower CD4 baseline levels. For the left tail of the response distribution \((\tau = 0.25)\), a different evolution of individual trajectories over time is observed and this is worth of same space for discussion. As for the other quartiles, individuals mainly start the study with an intermediate level of T-lymphocytes in the blood \((\delta_3 + \delta_4 > 0.7)\). However, in this case, estimated transition probabilities suggest that sicker men can experience repeated increases and decreases in the CD4 count over the follow-up time. More in detail, the first hidden state is related to low response values; transitions towards this state are unlikely \((\sum_{k=1}^{m} q_{k1} < 0.1)\) and, if any transition is observed, for most of the individuals an increment of the T-lymphocyte levels is subsequently registered \((q_{11} = 0.264)\). On the other hand, transitions from and to the remaining hidden states highlight the presence of units that move towards higher baseline CD4 levels at first; after they reach a given level, the number of T-lymphocytes in the blood tends to remain constant or to decrease once more during the observational time period.

Results obtained under the \(lqHMM+LDO\) formulation do not substantially differ from those presented in section 6.2 for \(lqmHMM\). The only remarkable difference is related to the additional hidden state identified for the
\textit{lqHMM+LDO} at \( \tau = 0.25 \). It is worth to notice that such a solution has been obtained only once within the set of starting points we have considered for the analysis and represents quite an isolated point in the parameter space. This explains somehow differences between the two modelling approaches and also the wider confidence intervals that are associated to \( \alpha_1 \) and \( q_{tk}, k = 1, \ldots, m, \) with respect to those related to the remaining estimates. More in detail, the highly parametrized structure of \textit{lqHMM+LDO} may lead to flat likelihood regions and, therefore, to multiple optimal solutions that make difficult to obtain reliable parameter estimates that are necessary to build the block bootstrap CIs.

In Table 11, we show the estimated LDO-dependent locations in the longitudinal data model together with the locations of the finite mixture estimated under the \textit{lqmHMM} formulation. As it may be noticed, for both models, the negative effect of Time\textsubscript{sero} on the HIV progression reduces when moving from \( \tau = 0.25 \) to \( \tau = 0.75 \). Estimated parameters suggest that individuals belonging to “lower” mixture components/LDO classes register a steep reduction in the (log) number of T-lymphocytes when a unit increase in the time since seroconversion is observed; such an effect progressively reduces for units belonging to “higher” categories. As it may be easily noticed, similar results are obtained for \textit{lqmHMM} and \textit{lqHMM+LDO} at \( \tau = 0.50 \) and \( \tau = 0.75 \). As regards the left tail of the response variable distribution, some differences may be observed, thus suggesting that the missing data process mainly influences the distribution of individuals with lower CD4 cell count levels (sicker men).

A deeper insight into results obtained from \textit{lqHMM+LDO} can be obtained by analysing the estimated LDO class model; see estimates reported in Table 12. For all the analysed quantiles, the negative effect of the time to drop-out (\( \hat{\lambda}_1 < 0 \)) suggests that “lower” LDO classes identify groups of individuals with shorter longitudinal sequences: the probability of belonging to one of the first \( g \) classes reduces when the number of available measures increases.
This result is much clearer when looking at Figure 2, where we represent the estimated LDO class probabilities as a function of the number of available measurements. We may notice that the probability of extreme categories is lower than that of central classes and, as expected, when the length of the observation window increases, “lower” LDO class probabilities decrease in favour of the “higher” ones, regardless of the considered quantile.

By matching these results with those presented in Table 11, we can conclude that the time since seroconversion strongly influences the T-lymphocytes levels for units that leave the study earlier in time, while this effect tends to decrease for units remaining under observation for a longer period.

To conclude, we report in Figures 3-4 the longitudinal trajectories of individuals classified in the different mixture components/LDO classes by fitting the $lqmHMM$ presented in Section 6.2 and the $lqHMM+LDO$ discussed so far, respectively. Local polynomial regression curves (blue lines), 95% confidence intervals (gray bands) and mean values (blue dots) are reported to highlight the general trend. Due to the missing data process, wider confidence intervals are observed for the last measurement occasions.

The two modelling approaches we are comparing seem to return similar classifications of units; however, when assuming that the mixture component probabilities depend on $T_i$, we may be able to correct some anomalies (units incorrectly classified) obtained under the $lqmHMM$ formulation.

By looking at Figure 4, it is clear that when the dependence of the longitudinal data process on the missing data one is explicitly formalized, more homogeneous clusters can be identified. This result is particularly evident for the first quartile of the response variable distribution ($\tau = 0.25$), that is for responses associated to sicker men which often represent the main target of inference. These results, together with the lower BIC values obtained
under $lqHMM+LDO$ with respect to $lqmHMM$ (see Tables 4-8), suggest an improvement in the quality of results and, thus, render $lqHMM+LDO$ an interesting modelling solution for the analysis of these data. However, the gain in the log-likelihood values we obtain when moving from $lqmHMM$ to $lqHMM+LDO$ could be primarily ascribed to the prior probabilities that, in the former model, are constant across individuals while, in the latter, are defined as a function of individual-specific features. In our formulation, we assume that such features are connected to a differential propensity to stay in the study, which, in turn, is summarized by $T_i$. Nevertheless, this propensity is unobservable and, therefore, we may not conclude that the missing data process is truly informative, since $T_i$ could represent other, unobserved, individual characteristics not linked to the propensity to drop-out. A sensitivity analysis to check the non-ignorability of the missing data generating process represents a further step to validate the model.

7 Concluding remarks

In this manuscript we have discussed a class of mixed hidden Markov quantile regression models for longitudinal continuous responses; a general dependence structure is considered by allow the measurement from each statistical units share both time-varying and time constant random parameters, thus providing an extension to the models proposed by Geraci and Bottai (2007) and Farcomeni (2012). Both unobserved heterogeneity sources are modelled by using discrete distribution, in a nonparametric fashion, in order to produce a conditional model which should be robust under a series of empirical situations, as shown in the simulation study. Since unobserved heterogeneity in this context may arise due to omitted covariates or be influenced by patterns of drop-out, we allow the time-constant random parameters to have a distribution which is dependent on the observed pattern of drop-out (i.e. the number of measurements for each individual) through a drop-out related ordered latent class, as suggested by Roy (2003) and Roy and Daniels (2008). The simulation study and the re-analysis of a well known benchmark dataset, the CD4 cells data of Kaslow et al. (1987), give quite encouraging results,
showing how the proposed models can be easily applied to complex data structures.

References

A. Agresti. *Analysis of ordinal categorical data*. John Wiley & Sons, 2010.

M. Aitkin. A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and computing*, 6(3):251–262, 1996.

M. Aitkin. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55(1):117–128, 1999.

R.J. Altman. Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, 102(477):201–210, 2007.

L. E Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

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

P. J. Diggle, P. J. Heagerty, K.Y. Liang, and S. L. Zeger. *Analysis of longitudinal data*, volume 25 of *Oxford Statistical Science Series*. Oxford University Press, second edition, 2002.

A. Farcomeni. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Statistics and Computing*, 22(1):141–152, 2012.

M. Geraci and M. Bottai. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–54, 2007.
M. Geraci and M. Bottai. Linear quantile mixed models. *Statistics and Computing*, 24(3):461–479, 2014.

J. J. Heckman. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement*, pages 475–492. NBER, 1976.

J.J. Heckman. The incidental parameters problem and the problem of initial conditions in estimating discrete time-discrete data stochastic processes and some Monte Carlo evidence. In C.F. Manski and D. McFadden, editors, *Structural analysis of discrete data*. MIT Press, 1981.

R. A. Kaslow, D.G. Ostrow, R. Detels, J. P. Phair, B. F. Polk, C.R. Rinaldo, et al. The multicenter aids cohort study: rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology*, 126(2):310–318, 1987.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

S. Lahiri. Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, pages 386–404, 1999.

B. G. Lindsay. The geometry of mixture likelihoods: a general theory. *The Annals of Statistics*, 11(1):86–94, 1983a.

B. G. Lindsay. The geometry of mixture likelihoods, Part II: the exponential family. *The Annals of Statistics*, 11(3):783–792, 1983b.

R. J. A. Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993.

R. J. A. Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431):1112–1121, 1995.
R. JA Little and D. B Rubin. *Statistical analysis with missing data*. Wiley, 2002.

Y. Liu and M. Bottai. Mixed-Effects Models for Conditional Quantiles with Longitudinal Data. *The International Journal of Biostatistics*, 5(1):1–24, 2009.

I.L. MacDonald and W. Zucchini. *Hidden Markov and other models for discrete-valued time series*, volume 110. CRC Press, 1997.

A. Maruotti. Mixed hidden markov models for longitudinal data: An overview. *International Statistical Review*, 79(3), 2011.

A. Maruotti and T. Rydén. A semiparametric approach to hidden markov models under longitudinal observations. *Statistics and Computing*, 19(4): 381–393, 2009.

D. Rizopoulos and E. Lesaffre. Introduction to the special issue on joint modelling techniques. *Statistical methods in medical research*, 23(1):3–10, 2014.

J. Roy. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, 59(4):829–836, 2003.

J. Roy and M. Daniels. A general class of pattern mixture models for non-ignorable dropout with many possible dropout times. *Biometrics*, 64(2): 538–545, 2008.

L. R. Welch. Hidden markov models and the baum-welch algorithm. *IEEE Information Theory Society Newsletter*, 53(4):10–13, 2003.

S. L. Zeger and P. J. Diggle. Semiparametric models for longitudinal data with application to CD4 cell numbers in hiv seroconverters. *Biometrics*, 50(3):689–699, 1994.
Table 1: Simulation study. lqmHMM at different quantiles. RMSE for longitudinal parameter estimates. Number of mixture components chosen through BIC.

|       | $\tau = 0.25$ |       |       |       | $\tau = 0.50$ |       |       |       |       | $\tau = 0.75$ |       |       |       |
|-------|---------------|-------|-------|-------|---------------|-------|-------|-------|-------|---------------|-------|-------|-------|
|       | $b_i$         | $\varepsilon_{it}$ | $\alpha_1$ | 0.415 | 0.363 | 0.708 | 0.122 | 0.256 | 0.317 | 0.231 | 0.386 | 0.613 | 0.122 | 0.272 | 0.281 | 0.279 | 0.526 | 0.593 | 0.099 | 0.100 | 0.156 | 0.099 | 0.100 | 0.156 | 0.201 | 0.460 | 0.832 | 0.135 | 0.132 | 0.392 | 0.146 | 0.138 | 0.425 | 0.131 | 0.139 | 0.506 | 0.135 | 0.132 | 0.399 | 0.108 | 0.104 | 0.419 | 0.113 | 0.109 | 0.452 | 0.109 | 0.103 | 0.399 | 0.122 | 0.089 | 0.420 | 0.092 | 0.096 | 0.426 | 0.099 | 0.094 | 0.390 | 0.070 | 0.076 | 0.408 | 0.076 | 0.076 | 0.379 | 0.078 | 0.071 | 0.342 | 0.070 | 0.076 | 0.408 | 0.076 | 0.076 | 0.379 | 0.078 | 0.071 | 0.342 |
|       | $\alpha_2$    | 4.344 | 0.471 | 5.361 | 0.126 | 0.272 | 0.281 | 0.279 | 0.526 | 0.593 | 0.099 | 0.100 | 0.156 | 0.287 | 0.415 | 0.684 | 0.150 | 0.257 | 0.287 | 0.201 | 0.460 | 0.832 | 0.135 | 0.132 | 0.392 | 0.146 | 0.138 | 0.425 | 0.131 | 0.139 | 0.506 | 0.135 | 0.132 | 0.399 | 0.108 | 0.104 | 0.419 | 0.113 | 0.109 | 0.452 | 0.109 | 0.103 | 0.399 | 0.122 | 0.089 | 0.420 | 0.092 | 0.096 | 0.426 | 0.099 | 0.094 | 0.390 | 0.070 | 0.076 | 0.408 | 0.076 | 0.076 | 0.379 | 0.078 | 0.071 | 0.342 | 0.070 | 0.076 | 0.408 | 0.076 | 0.076 | 0.379 | 0.078 | 0.071 | 0.342 |

$n = 100, T = 5$

$n = 100, T = 10$

$n = 200, T = 5$

$n = 200, T = 10$
Table 2: Simulation study. lqHMM+LDO at different quantiles. RMSE for longitudinal parameter estimates for $\lambda = (1, 2.75, -0.3)$. Parameters $m$ and $G$ chosen through BIC.

| $\varepsilon_{it}$ | $N$ | $t_3$ | $\chi^2$ | $N$ | $t_3$ | $\chi^2$ | $N$ | $t_3$ | $\chi^2$ |
|-------------------|-----|-------|----------|-----|-------|----------|-----|-------|----------|
| $n = 100, T = 5$  |     |       |          |     |       |          |     |       |          |
| $\alpha_1$       | 0.502 | 0.838 | 0.631    | 0.529 | 0.489 | 0.938    | 1.151 | 1.083 | 0.977    |
| $\alpha_2$       | 1.912 | 2.370 | 1.850    | 0.661 | 0.543 | 1.316    | 1.765 | 1.601 | 2.186    |
| $\beta$          | 0.047 | 0.057 | 0.049    | 0.039 | 0.044 | 0.062    | 0.067 | 0.075 | 0.082    |
| $bg_1$           | 0.119 | 0.422 | 0.073    | 0.050 | 0.063 | 0.281    | 0.614 | 0.618 | 0.562    |
| $bg_2$           | 0.917 | 0.906 | 0.891    | 0.172 | 0.142 | 0.671    | 1.407 | 1.375 | 1.338    |
| $bg_3$           | 1.298 | 1.272 | 1.191    | 0.225 | 0.131 | 0.122    | 0.194 | 0.573 | 0.401    |
| $n = 100, T = 10$|     |       |          |     |       |          |     |       |          |
| $\alpha_1$       | 0.230 | 0.683 | 0.381    | 0.176 | 0.206 | 0.654    | 0.241 | 0.326 | 0.579    |
| $\alpha_2$       | 0.688 | 1.108 | 0.281    | 0.230 | 0.231 | 0.457    | 0.221 | 0.276 | 2.009    |
| $\beta$          | 0.029 | 0.035 | 0.018    | 0.027 | 0.026 | 0.033    | 0.028 | 0.034 | 0.044    |
| $bg_1$           | 0.046 | 0.071 | 0.032    | 0.038 | 0.040 | 0.169    | 0.048 | 0.051 | 0.065    |
| $bg_2$           | 0.097 | 0.098 | 0.056    | 0.034 | 0.036 | 0.301    | 0.041 | 0.046 | 0.060    |
| $bg_3$           | 0.103 | 0.044 | 0.078    | 0.034 | 0.034 | 0.048    | 0.039 | 0.041 | 0.062    |
| $n = 200, T = 5$ |     |       |          |     |       |          |     |       |          |
| $\alpha_1$       | 0.190 | 0.308 | 0.142    | 0.170 | 0.213 | 0.436    | 1.006 | 1.029 | 0.888    |
| $\alpha_2$       | 0.722 | 1.115 | 0.175    | 0.271 | 0.275 | 0.466    | 1.625 | 1.944 | 2.186    |
| $\beta$          | 0.029 | 0.035 | 0.019    | 0.024 | 0.027 | 0.033    | 0.052 | 0.051 | 0.060    |
| $bg_1$           | 0.034 | 0.041 | 0.023    | 0.032 | 0.032 | 0.038    | 0.571 | 0.556 | 0.489    |
| $bg_2$           | 0.038 | 0.041 | 0.025    | 0.032 | 0.035 | 0.038    | 1.385 | 1.352 | 1.283    |
| $bg_3$           | 0.045 | 0.050 | 0.029    | 0.041 | 0.043 | 0.052    | 0.196 | 0.194 | 0.360    |
| $n = 200, T = 10$|     |       |          |     |       |          |     |       |          |
| $\alpha_1$       | 0.151 | 0.311 | 0.133    | 0.134 | 0.141 | 0.484    | 0.154 | 0.216 | 0.639    |
| $\alpha_2$       | 0.318 | 0.805 | 0.161    | 0.147 | 0.158 | 0.433    | 0.186 | 0.208 | 2.183    |
| $\beta$          | 0.021 | 0.023 | 0.012    | 0.019 | 0.018 | 0.023    | 0.019 | 0.023 | 0.032    |
| $bg_1$           | 0.032 | 0.034 | 0.018    | 0.028 | 0.026 | 0.032    | 0.032 | 0.034 | 0.043    |
| $bg_2$           | 0.028 | 0.033 | 0.015    | 0.024 | 0.024 | 0.030    | 0.026 | 0.029 | 0.042    |
| $bg_3$           | 0.028 | 0.032 | 0.017    | 0.023 | 0.022 | 0.028    | 0.026 | 0.027 | 0.045    |
Table 3: Simulation study. lqHMM+LDO at different quantiles. RMSE for longitudinal parameter estimates for $\lambda = (5, 8.5, -1.1)$. Parameters $m$ and $G$ chosen through BIC.

| $\varepsilon_{it}$ | $\tau = 0.25$ | $\tau = 0.50$ | $\tau = 0.75$ |
|--------------------|----------------|----------------|----------------|
|                    | $N$ $t_3$ $\chi^2_2$ | $N$ $t_3$ $\chi^2_2$ | $N$ $t_3$ $\chi^2_2$ |
| $n = 100, T = 5$  |                |                |                |
| $\alpha_1$        | 0.320          | 0.262          | 0.648          |
|                    | 0.470          | 0.273          | 0.796          |
|                    | 0.207          | 0.571          | 0.684          |
| $\alpha_2$        | 1.061          | 1.352          | 1.587          |
|                    | 1.292          | 0.404          | 2.806          |
|                    | 0.396          | 0.526          | 2.498          |
| $\beta$           | 0.041          | 0.036          | 0.047          |
|                    | 0.048          | 0.043          | 0.057          |
|                    | 0.027          | 0.051          | 0.064          |
| $b_{g_1}$         | 0.049          | 0.045          | 0.112          |
|                    | 0.051          | 0.048          | 0.140          |
|                    | 0.033          | 0.061          | 0.159          |
| $b_{g_2}$         | 0.144          | 0.058          | 0.190          |
|                    | 0.100          | 0.056          | 0.228          |
|                    | 0.035          | 0.090          | 0.248          |
| $b_{g_3}$         | 0.598          | 0.638          | 0.601          |
|                    | 0.574          | 0.504          | 0.584          |
|                    | 0.416          | 0.545          | 0.574          |
| $n = 100, T = 10$ |                |                |                |
| $\alpha_1$        | 0.221          | 0.211          | 0.211          |
|                    | 0.310          | 0.191          | 0.306          |
|                    | 0.145          | 0.504          | 0.567          |
| $\alpha_2$        | 0.588          | 0.331          | 0.210          |
|                    | 1.004          | 0.263          | 0.260          |
|                    | 0.184          | 0.528          | 1.965          |
| $\beta$           | 0.029          | 0.024          | 0.030          |
|                    | 0.036          | 0.026          | 0.034          |
|                    | 0.017          | 0.030          | 0.047          |
| $b_{g_1}$         | 0.046          | 0.047          | 0.050          |
|                    | 0.065          | 0.047          | 0.065          |
|                    | 0.032          | 0.051          | 0.080          |
| $b_{g_2}$         | 0.040          | 0.038          | 0.036          |
|                    | 0.046          | 0.035          | 0.043          |
|                    | 0.023          | 0.040          | 0.056          |
| $b_{g_3}$         | 0.037          | 0.038          | 0.036          |
|                    | 0.042          | 0.032          | 0.046          |
|                    | 0.021          | 0.039          | 0.061          |
| $n = 200, T = 5$  |                |                |                |
| $\alpha_1$        | 0.187          | 0.165          | 0.380          |
|                    | 0.285          | 0.182          | 0.538          |
|                    | 0.138          | 0.451          | 0.548          |
| $\alpha_2$        | 0.746          | 0.211          | 0.824          |
|                    | 1.149          | 0.234          | 1.367          |
|                    | 0.193          | 0.415          | 2.207          |
| $\beta$           | 0.030          | 0.026          | 0.028          |
|                    | 0.033          | 0.025          | 0.037          |
|                    | 0.018          | 0.032          | 0.046          |
| $b_{g_1}$         | 0.032          | 0.025          | 0.085          |
|                    | 0.035          | 0.030          | 0.048          |
|                    | 0.020          | 0.037          | 0.066          |
| $b_{g_2}$         | 0.039          | 0.032          | 0.141          |
|                    | 0.063          | 0.034          | 0.113          |
|                    | 0.024          | 0.043          | 0.114          |
| $b_{g_3}$         | 0.268          | 0.269          | 0.237          |
|                    | 0.257          | 0.316          | 0.296          |
|                    | 0.236          | 0.259          | 0.324          |
| $n = 200, T = 10$ |                |                |                |
| $\alpha_1$        | 0.132          | 0.120          | 0.146          |
|                    | 0.271          | 0.174          | 0.190          |
|                    | 0.127          | 0.440          | 0.515          |
| $\alpha_2$        | 0.269          | 0.170          | 0.161          |
|                    | 0.903          | 0.194          | 0.213          |
|                    | 0.161          | 0.448          | 1.971          |
| $\beta$           | 0.020          | 0.018          | 0.021          |
|                    | 0.039          | 0.019          | 0.023          |
|                    | 0.012          | 0.022          | 0.034          |
| $b_{g_1}$         | 0.030          | 0.031          | 0.034          |
|                    | 0.058          | 0.030          | 0.041          |
|                    | 0.021          | 0.038          | 0.054          |
| $b_{g_2}$         | 0.028          | 0.023          | 0.029          |
|                    | 0.042          | 0.024          | 0.033          |
|                    | 0.017          | 0.029          | 0.038          |
| $b_{g_3}$         | 0.025          | 0.021          | 0.025          |
|                    | 0.040          | 0.022          | 0.026          |
|                    | 0.016          | 0.025          | 0.048          |
Table 4: CD4 data. BIC values for lqmHMM for different choices of \( m \) and \( G \) at different quantiles.

| Mixture Components | Hidden States | 1    | 2    | 3    | 4    | 5    | 6    |
|--------------------|---------------|------|------|------|------|------|------|
| \( \tau = 0.25 \)  | 1             | 3940.50 | 3530.48 | 3408.42 | 3312.98 | 3286.67 | 3298.81 |
|                    | 2             | 3202.25 | 2923.33 | 2780.49 | 2772.64 | 2760.39 | 2772.33 |
|                    | 3             | 2963.50 | 2747.21 | 2676.26 | 2637.39 | 2619.05 | 2627.48 |
|                    | 4             | 2757.55 | 2660.63 | 2541.63 | 2510.71 | 2478.88 | 2487.22 |
|                    | 5             | 2688.98 | 2527.39 | 2504.99 | 2494.63 | 2486.83 | 2507.90 |
| \( \tau = 0.50 \)  | 1             | 3434.26 | 3014.58 | 2898.11 | 2850.34 | 2833.41 | 2874.04 |
|                    | 2             | 2733.04 | 2522.56 | 2412.12 | 2381.26 | 2374.18 | 2381.82 |
|                    | 3             | 2523.15 | 2345.76 | 2280.27 | 2266.98 | 2256.28 | 2268.15 |
|                    | 4             | 2410.07 | 2268.98 | 2236.89 | 2236.29 | 2236.84 |
|                    | 5             | 2377.80 | 2298.48 | 2272.29 | 2266.93 | 2267.97 | 2278.48 |
| \( \tau = 0.75 \)  | 1             | 3491.69 | 3134.31 | 2986.87 | 2951.28 | 2951.44 | 2940.33 |
|                    | 2             | 2823.44 | 2549.81 | 2495.15 | 2455.24 | 2441.21 | 2444.51 |
|                    | 3             | 2470.11 | 2337.64 | 2290.17 | 2242.39 | 2244.89 | 2256.24 |
|                    | 4             | 2370.11 | 2307.19 | 2289.67 | 2281.06 | 2240.10 | 2230.01 |
|                    | 5             | 2356.16 | 2295.80 | 2251.37 | 2252.42 | 2252.06 | 2274.55 |
Table 5: CD4 data. Estimated longitudinal model parameters for lqmHMM at different quantiles. 95% bootstrap confidence intervals are reported within brackets.

|        | $\tau = 0.25$       | $\tau = 0.50$       | $\tau = 0.75$       |
|--------|---------------------|---------------------|---------------------|
|        | $[m = 4, G = 5]$    | $[m = 4, G = 4]$    | $[m = 4, G = 3]$    |
| $\alpha_1$ | 5.593 (5.403; 5.677) | 6.054 (5.994; 6.133) | 6.203 (6.071; 6.273) |
| $\alpha_2$ | 6.124 (6.066; 6.166) | 6.432 (6.368; 6.530) | 6.580 (6.517; 6.628) |
| $\alpha_3$ | 6.540 (6.489; 6.587) | 6.750 (6.689; 6.837) | 6.876 (6.804; 6.934) |
| $\alpha_4$ | 6.915 (6.847; 6.995) | 7.055 (7.023; 7.231) | 7.256 (7.168; 7.373) |
| Age     | 0.000 (−0.004; 0.002) | 0.004 (−0.001; 0.008) | 0.000 (−0.005; 0.005) |
| Drugs   | 0.044 (0.000; 0.092)  | 0.057 (−0.014; 0.110) | 0.061 (0.003; 0.113)  |
| Packs   | 0.056 (0.041; 0.071)  | 0.043 (0.015; 0.054)  | 0.044 (0.015; 0.062)  |
| Partners| 0.006 (0.001; 0.012)  | 0.005 (0.001; 0.012)  | 0.011 (0.003; 0.016)  |
| CESD    | −0.004 (−0.005; −0.001) | −0.004 (−0.006; −0.002) | −0.004 (−0.006; −0.002) |
| TimeSero| −0.175 (−0.206; −0.150) | −0.140 (−0.164; −0.114) | −0.123 (−0.145; −0.102) |
| $\sigma_b$ | 0.219 (0.200; 0.360) | 0.134 (0.105; 0.165) | 0.102 (0.088; 0.133) |
Table 6: CD4 data. Estimated initial and transition probabilities for lqmHMM, at different quantiles. 95% bootstrap confidence intervals are reported within brackets.

|       | 1               | 2               | 3               | 4               |
|-------|-----------------|-----------------|-----------------|-----------------|
| **τ = 0.25** |                 |                 |                 |                 |
| δ     | 0.083 (0.036; 0.134) | 0.408 (0.316; 0.512) | 0.411 (0.304; 0.501) | 0.098 (0.055; 0.157) |
|       | 0.284 (0.116; 0.466) | 0.700 (0.510; 0.860) | 0.000 (0.000; 0.025) | 0.016 (0.000; 0.069) |
| 1     | 0.083 (0.034; 0.125) | 0.675 (0.574; 0.754) | 0.232 (0.153; 0.334) | 0.010 (0.000; 0.032) |
| 2     | 0.031 (0.008; 0.062) | 0.126 (0.077; 0.179) | 0.787 (0.705; 0.844) | 0.055 (0.019; 0.119) |
| 3     | 0.011 (0.000; 0.033) | 0.044 (0.000; 0.090) | 0.015 (0.000; 0.095) | 0.930 (0.854; 0.977) |
| **τ = 0.50** |                 |                 |                 |                 |
| δ     | 0.120 (0.059; 0.242) | 0.434 (0.306; 0.543) | 0.326 (0.213; 0.449) | 0.120 (0.059; 0.162) |
|       | 0.927 (0.775; 1.000) | 0.073 (0.000; 0.216) | 0.000 (0.000; 0.034) | 0.000 (0.000; 0.016) |
| 1     | 0.107 (0.052; 0.174) | 0.830 (0.734; 0.925) | 0.063 (0.000; 0.144) | 0.000 (0.000; 0.016) |
| 2     | 0.013 (0.000; 0.062) | 0.055 (0.000; 0.108) | 0.898 (0.838; 0.961) | 0.034 (0.000; 0.071) |
| 3     | 0.043 (0.000; 0.079) | 0.000 (0.000; 0.041) | 0.019 (0.000; 0.092) | 0.938 (0.866; 0.994) |
| **τ = 0.75** |                 |                 |                 |                 |
| δ     | 0.119 (0.041; 0.192) | 0.359 (0.224; 0.497) | 0.367 (0.232; 0.488) | 0.155 (0.098; 0.217) |
|       | 0.861 (0.757; 0.958) | 0.139 (0.041; 0.243) | 0.000 (0.000; 0.000) | 0.000 (0.000; 0.000) |
| 1     | 0.125 (0.069; 0.194) | 0.810 (0.701; 0.885) | 0.065 (0.000; 0.171) | 0.000 (0.000; 0.022) |
| 2     | 0.021 (0.000; 0.062) | 0.094 (0.031; 0.184) | 0.858 (0.770; 0.921) | 0.026 (0.000; 0.063) |
| 3     | 0.019 (0.000; 0.048) | 0.000 (0.000; 0.044) | 0.088 (0.007; 0.192) | 0.894 (0.782; 0.965) |
Table 7: CD4 data. Number of individuals in the study at each time occasion.

| Visit | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|       | 369 | 364 | 340 | 315 | 268 | 225 | 173 | 133 | 92  | 54  | 33  | 10  |

Table 8: CD4 data. BIC values for lqHMM+LDO for different choices of $m$ and $G$ at different quantiles.

| LDO classes | Hidden States | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    |
|-------------|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|
|             | $\tau = 0.25$|       |       |       |       |       |       |       |       |       |        |        |        |
|             | 1             | 3940.50 | 3525.57 | 3400.96 | 3305.67 | 3279.57 |       |       |       |       |        |        |        |
|             | 2             | 3292.25 | 2919.41 | 2773.66 | 2761.84 | 2755.28 |       |       |       |       |        |        |        |
|             | 3             | 2963.50 | 2741.63 | 2669.37 | 2636.84 | 2612.29 |       |       |       |       |        |        |        |
|             | 4             | 2757.55 | 2660.38 | 2537.89 | 2509.15 | 2522.63 |       |       |       |       |        |        |        |
|             | 5             | 2688.98 | 2551.40 | 2522.53 | 2474.39 | 2460.52 |       |       |       |       |        |        |        |
|             | $\tau = 0.50$|       |       |       |       |       |       |       |       |       |        |        |        |
|             | 1             | 3434.26 | 3010.15 | 2895.63 | 2847.35 | 2829.65 |       |       |       |       |        |        |        |
|             | 2             | 2733.04 | 2517.26 | 2406.67 | 2377.40 | 2369.86 |       |       |       |       |        |        |        |
|             | 3             | 2523.15 | 2343.65 | 2280.14 | 2266.03 | 2259.80 |       |       |       |       |        |        |        |
|             | 4             | 2410.07 | 2265.06 | 2233.48 | 2231.14 | 2233.14 |       |       |       |       |        |        |        |
|             | 5             | 2377.80 | 2291.68 | 2265.61 | 2259.65 | 2244.27 |       |       |       |       |        |        |        |
|             | $\tau = 0.75$|       |       |       |       |       |       |       |       |       |        |        |        |
|             | 1             | 3491.69 | 3137.21 | 2987.71 | 2953.09 | 2953.71 |       |       |       |       |        |        |        |
|             | 2             | 2823.44 | 2551.39 | 2491.32 | 2453.73 | 2448.10 |       |       |       |       |        |        |        |
|             | 3             | 2470.11 | 2335.57 | 2287.38 | 2240.26 | 2242.14 |       |       |       |       |        |        |        |
|             | 4             | 2370.11 | 2308.85 | 2225.93 | 2203.98 | 2223.70 |       |       |       |       |        |        |        |
|             | 5             | 2356.17 | 2290.69 | 2248.35 | 2240.17 | 2242.98 |       |       |       |       |        |        |        |
Table 9: CD4 data. Estimated longitudinal model parameters for lqHMM+LDO at different quantiles. 95% bootstrap confidence intervals are reported within brackets.

|       | $\tau = 0.25$ | $\tau = 0.50$ | $\tau = 0.75$ |
|-------|---------------|---------------|---------------|
|       | $[m = 5, G = 5]$ | $[m = 4, G = 4]$ | $[m = 4, G = 3]$ |
| $\alpha_1$ | 5.046 (3.937; 5.286) | 6.043 (5.931; 6.114) | 6.198 (6.069; 6.282) |
| $\alpha_2$ | 5.880 (5.730; 5.918) | 6.416 (6.323; 6.502) | 6.579 (6.512; 6.628) |
| $\alpha_3$ | 6.193 (6.126; 6.256) | 6.719 (6.647; 6.825) | 6.872 (6.801; 6.934) |
| $\alpha_4$ | 6.582 (6.508; 6.634) | 7.040 (6.973; 7.215) | 7.243 (7.167; 7.370) |
| $\alpha_5$ | 6.936 (6.846; 7.026) |  |  |
| Age    | $-0.004 (-0.007; 0.000)$ | $0.004 (-0.001; 0.007)$ | $0.000 (-0.004; 0.005)$ |
| Drugs  | 0.048 (-0.013; 0.124) | 0.072 (-0.006; 0.145) | 0.064 (0.007; 0.115) |
| Packs  | 0.032 (0.024; 0.051) | 0.042 (0.014; 0.054) | 0.044 (0.018; 0.064) |
| Partners | 0.011 (0.005; 0.016) | 0.005 (0.000; 0.012) | 0.011 (0.002; 0.016) |
| CESD   | $-0.003 (-0.006; -0.001)$ | $-0.004 (-0.006; -0.002)$ | $-0.004 (-0.006; -0.002)$ |
| $T_{time}$ | $-0.157 (-0.187; -0.127)$ | $-0.146 (-0.175; -0.119)$ | $-0.131 (-0.155; -0.108)$ |
Table 10: CD4 data. Estimated initial and transition probabilities for lqHMM+LDO, at different quantiles. 95% bootstrap confidence intervals are reported within brackets.

|    | 1          | 2          | 3          | 4          | 5          |
|----|------------|------------|------------|------------|------------|
| \( \tau = 0.25 \) |            |            |            |            |            |
| \delta | 0.010 (0.000; 0.026) | 0.125 (0.054; 0.187) | 0.365 (0.254; 0.477) | 0.358 (0.263; 0.470) | 0.142 (0.073; 0.212) |
| 1   | 0.264 (0.000; 0.610) | 0.588 (0.000; 0.855) | 0.092 (0.000; 0.616) | 0.000 (0.000; 0.427) | 0.056 (0.000; 0.285) |
| 2   | 0.049 (0.000; 0.115) | 0.471 (0.167; 0.791) | 0.480 (0.143; 0.797) | 0.000 (0.000; 0.000) | 0.000 (0.000; 0.000) |
| 3   | 0.027 (0.000; 0.050) | 0.139 (0.052; 0.212) | 0.592 (0.430; 0.706) | 0.225 (0.130; 0.376) | 0.017 (0.000; 0.049) |
| 4   | 0.006 (0.000; 0.022) | 0.030 (0.000; 0.077) | 0.165 (0.103; 0.299) | 0.737 (0.640; 0.815) | 0.062 (0.013; 0.128) |
| 5   | 0.004 (0.000; 0.015) | 0.008 (0.000; 0.030) | 0.047 (0.000; 0.096) | 0.095 (0.000; 0.180) | 0.845 (0.764; 0.947) |
| \( \tau = 0.50 \) |            |            |            |            |            |
| \delta | 0.118 (0.048; 0.215) | 0.411 (0.295; 0.544) | 0.341 (0.221; 0.454) | 0.130 (0.066; 0.180) |           |
| 1   | 0.939 (0.753; 1.000) | 0.061 (0.000; 0.246) | 0.000 (0.000; 0.000) | 0.000 (0.000; 0.010) |           |
| 2   | 0.104 (0.060; 0.193) | 0.846 (0.711; 0.912) | 0.050 (0.000; 0.155) | 0.000 (0.000; 0.020) |           |
| 3   | 0.015 (0.000; 0.053) | 0.049 (0.004; 0.101) | 0.901 (0.836; 0.955) | 0.036 (0.000; 0.084) |           |
| 4   | 0.041 (0.000; 0.082) | 0.000 (0.000; 0.036) | 0.045 (0.000; 0.126) | 0.914 (0.835; 0.981) |           |
| \( \tau = 0.75 \) |            |            |            |            |            |
| \delta | 0.122 (0.040; 0.205) | 0.361 (0.219; 0.510) | 0.360 (0.221; 0.488) | 0.157 (0.094; 0.225) |           |
| 1   | 0.866 (0.749; 0.973) | 0.134 (0.027; 0.251) | 0.000 (0.000; 0.000) | 0.000 (0.000; 0.000) |           |
| 2   | 0.122 (0.065; 0.199) | 0.813 (0.695; 0.894) | 0.065 (0.000; 0.168) | 0.000 (0.000; 0.023) |           |
| 3   | 0.020 (0.000; 0.055) | 0.099 (0.039; 0.194) | 0.854 (0.758; 0.916) | 0.027 (0.000; 0.064) |           |
| 4   | 0.018 (0.000; 0.045) | 0.000 (0.000; 0.045) | 0.091 (0.012; 0.207) | 0.891 (0.765; 0.960) |           |
Table 11: CD4 data. Estimated LDO-dependent parameters in the longitudinal data model for lqHMM+LDO at different quantiles. 95% bootstrap confidence intervals are reported within brackets.

|         | τ = 0.25 |         | τ = 0.50 |         | τ = 0.75 |
|---------|----------|---------|----------|---------|----------|
| lqHMM+LDO |         |         |          | lqmHMM  |         |
| b_1     | -0.849   | -0.740  | (−1.079; −0.662) |         |
| b_2     | -0.434   | -0.300  | (−0.324; −0.256) |         |
| b_3     | -0.220   | -0.164  | (−0.181; −0.133) |         |
| b_4     | -0.123   | -0.053  | (−0.095; −0.035) |         |
| b_5     | -0.020   | 0.026   | (−0.011; 0.045)  |         |

Table 12: CD4 data. Estimated LDO class model parameters for lqHMM+LDO at different quantiles. 95% bootstrap confidence intervals are reported within brackets.

|         | τ = 0.25 |         | τ = 0.50 |         | τ = 0.75 |
|---------|----------|---------|----------|---------|----------|
|         | [m = 5, G = 5] | [m = 4, G = 4] | [m = 4, G = 3] |         |
| λ_{01}  | -3.583   | -2.112  | -1.388   | -1.388  | 0.615    |
| λ_{02}  | -1.159   | 0.013   | 2.102    | 1.295   | 4.379    |
| λ_{03}  | 0.514    | 2.002   | 5.299    |         |         |
| λ_{04}  | 1.926    | 3.116   |          |         |         |
| λ_1     | -0.290   | -0.318  | -0.324   | -0.324  | -0.066  |
Figures

Figure 1: CD4 data. Response variable distribution at each time occasion.
Figure 2: CD4 data. Estimated LDO probabilities as a function of the number of available measurements for lqHMM+LDO at different quantiles.
Figure 3: CD4 data. Longitudinal trajectories within mixture components for lqmHMM at different quantiles.
Figure 4: CD4 data. Longitudinal trajectories within mixture components for lqHMM+LDO at different quantiles.

\[ \tau = 0.25 \]

\[ \tau = 0.50 \]

\[ \tau = 0.75 \]