Mixed hidden Markov models for quantiles

Maria Francesca Marino and Nikos Tzavidis

Abstract We propose a mixed hidden Markov model for continuous longitudinal data, to the quantile regression perspective. Time-constant and time-varying random parameters are added in the quantile regression model to account for time-invariant and dynamic unobserved factors affecting the variable of interest. A nonparametric maximum likelihood approach is applied to solve the numerical integration problem typically arising in the mixed model framework. Parameter estimates are then obtained by means of an EM algorithm, easily derived by exploiting the forward and backward variables defined in the so called Baum-Welsh recursion.

keywords: Hidden Markov models, mixed effects, nonparametric maximum likelihood

1 Introduction

In the statistical literature, quantile regression [12] has become a quite popular and established technique for the analysis of data. If compared with the standard (mean) regression, it offers a thorough description of the response variable distribution in terms of a given number of observed covariates. When dealing with longitudinal data, dependence between observations coming from the same individual has to be properly taken into account in order to avoid bias in the parameter estimates. Mixed effects models are typically used in that context. By adopting such a model structure, we assume the existence of unobserved heterogeneity (unmeasured covariates) determining the association between longitudinal measurements: continuous, individual-specific, random coefficients are added in the regression model to capture

Maria Francesca Marino
Dipartimento di Scienze Statistiche, Sapienza Universitá di Roma, e-mail: mariafrancesca.marino@uniroma1.it

Nikos Tzavidis
Social Statistics & Demography, Southampton University, e-mail: n.tzavidis@soton.a.uk

this heterogeneity. In the quantile regression framework two main approaches have been developed to deal with longitudinal data: distribution-free and likelihood based methods. The former include, among others, the weighted GEE approach [19] and the penalized quantile regression estimators [11, 15, 8, 7]. The latter are based on the introduction of a parametric distribution for the response variable, in order to derive inference in a maximum likelihood framework. Most of the proposals are based on the asymmetric Laplace distribution [9, 20, 23, 16] which represent a computational convenient trick because of its direct correspondence with the quantile loss function. [9] and [10] consider either Gaussian or asymmetric Laplace distributed random parameters; this assumptions allow to solve the numerical approximation problem, typically observed in the mixed model framework, in a Gaussian quadrature perspective.
In some cases, the assumption that unobserved heterogeneity does not evolve with time can, however, be too restrictive: the response variable can be influenced by unobserved factors having their own dynamics and biased parameter estimates can be obtained by considering a standard mixed effect model. In such a context, [6] suggests the to introduce a hidden Markov chain to describe the temporal evolution of a random intercept, to account for response variability due to time-varying omitted covariates. By postulating an asymmetric Laplace distribution for the response variable, maximum likelihood parameter estimates are obtained by means of an EM algorithm. Here, we aim at extending the idea of [6] by considering a mixed hidden Markov models (see, for a thorough description [21]), where time-constant and time-varying sources of unobserved heterogeneity are jointly modelled. Temporal dynamics are captured by the hidden Markov chain, while time-constant unobserved behaviours are described by semi-parametrically estimating a general random effect distribution (see eg [1, 2]). Firstly introduced in the mixed hidden Markov models framework by [22], the NPML approach allows to relax any parametric assumption about the random effect distribution, offering great flexibility and avoiding possible bias in the parameter estimates caused by misspecification of such a distribution.
The plan of the paper is as follows. In section 2 the mixed quantile regression model is introduced. In sections 3 and 4 we describe the linear quantile mixed hidden Markov model and the proposed EM algorithm to derive maximum likelihood estimates. Last section contains concluding remarks and outlines future developments

2 Mixed models for quantile linear regression

Let Y_{it} denote a continuous longitudinal response recorded on $i = 1,...,n$ individuals at times $t = 1,...,T_i$ and let us assume we are interested in analysing how a given set of observed covariates influence its distribution. In a quantile regression framework, a convenient parametric assumption for the response variable is represented by the asymmetric Laplace distribution as it allows to derive inference about model parameters in a maximum likelihood perspective (see [9, 20, 10], for references). For a given quantile τ, the three-parameter asymmetric Laplace distribution is characterized by the following density function:

$$\frac{\tau(1-\tau)}{\sigma}\exp\left\{-\rho_\tau\left(\frac{y_{it}-\mu_{it}(\tau)}{\sigma}\right)\right\}, \tag{1}$$

where $\rho_\tau(u) = u\{\tau - I(u<0)\}$ indicates the quantile loss function while $\mu_{it}(\tau)$ and σ represent the location and the scale parameters, respectively.
In the longitudinal data literature, association between observations coming from the same individual is usually taken into account via the introduction, in the model specification, of individual-specific random parameters; these parameters are meant to capture all the potential sources of unobserved heterogeneity (i.e. unobserved covariates) determining the dependence between subsequent observations. The resulting model is known, in the literature, as random or mixed effect model [14]. Let $\mathbf{b}_i$ be a random parameter vector, distributed with density $f_b(\cdot)$; the basic assumption behind mixed effect models is that, given the random effects, observations coming from the same individual are no longer dependent. This is a form of conditional independence (sometimes referred to as local independence) which leads to the following expression:

$$f_{y|b}(y_{it} \mid y_{it-1},...y_{i1},\mathbf{b}_i) = f_{y|b}(y_{it} \mid \mathbf{b}_i).$$

Based on these considerations, the joint conditional distribution of the longitudinal sequence, for a generic unit i, can be obtained as

$$f_y(\mathbf{y}_i \mid \mathbf{b}_i;\tau) = \prod_{t=1}^{T_i} f_{y|b}(y_{it} \mid \mathbf{b}_i;\tau) = \left[\frac{\tau(1-\tau)}{\sigma}\right]^{T_i}\exp\left\{-\sum_{t=1}^{T_i}\rho_\tau\left(\frac{y_{it}-\mu_{it}(\mathbf{b}_i;\tau)}{\sigma}\right)\right\}$$

where, $\mu_{it}(\mathbf{b}_i;\tau)$ is the τ-th location parameter as a function of fixed and random effects. Following [9] and [10], parameter estimates can be obtained by maximizing the following observed likelihood function

$$L(\cdot) = \prod_{i=1}^{n}\int_{\mathbb{B}}\prod_{t=1}^{T_i} f_y(\mathbf{y}_i \mid \mathbf{b}_i) f_b(\mathbf{b}_i)\mathrm{d}\mathbf{b}_i. \tag{2}$$

In many cases the above integral has not a closed form solution and it has to be numerically approached. Two main proposals are available in order to solve such a problem. [9] and [20] suggest the application of a Monte Carlo EM algorithm to obtain parameter estimates, while [10] suggest the application of a Gaussian quadrature scheme. However, both methods are based on the introduction of parametric assumptions for the random coefficients vector: these can not directly assessed and, thus, a more flexible approach could result more appealing.

3 Mixed hidden Markov models for quantile linear regression

Mixed effect hidden Markov models (mHMMs) extend the idea behind mixed models by assuming the existence of two different sources of unobserved heterogeneity affecting the variable of interest: some unobserved features can have a time-invariant effect on the response, while some others can effect it in a dynamic way. In these contexts, time-constant and time-varying latent variables can be included in the model specification in order to account to such sources of random variation.
Let $\{S_{it}\}, i = 1,...,n, t = 1,...,T_i$ be a homogeneous, first order, hidden Markov chain, taking values in the finite set $\mathscr{S} = \{1,...,m\}$. Let us assume that all the individual share the same initial probability vector $\delta = (\delta_1,...,\delta_m)$ and the same transition probability matrix $\mathbf{Q} = \{q_{hk}\}$. More specifically, for a given unit i, $\delta_h, h = 1,...,m$ represents the prior probability of being in state h at the first time occasion, while $q_{hk}, h, k = 1,...,m$ represents the probability of observing a transition from state h at time $t-1$ to state k at time t.
Moreover, let $\mathbf{b}_i$ be a time-constant, individual-specific, random vector having distribution $f_b(\cdot)$, as defined in section 2. Mixed hidden Markov models are based on the following main assumptions. The random vector $\mathbf{b}_i$ 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 effects $\mathbf{b}_i$. After conditioning on $\mathbf{b}_i$ and s_{it}, longitudinal observations are no longer dependent. Based on these hypothesis the following expression holds:

$$f_{y|sb}(y_{it} \mid y_{i1:t-1}, s_{i1:t}, \mathbf{b}_i) = f_{y|sb}(y_{it} \mid s_{it}, \mathbf{b}_i)$$

where $y_{i1:t-1}$ represents the history of the response, for the i-th subject, up to time $t-1$ and $s_{i1:t}$ is the individual sequence of states up to time t.
In order to derive maximum likelihood estimates, let us assume the response variable is distributed according to an asymmetric Laplace distribution whose location parameter, for a given quantile τ, is defined through the following regression model

$$\mu_{it}\left(S_{it} = h, \mathbf{b}_i; \tau\right) = \mathbf{x}_{it}'\beta_h(\tau) + \mathbf{z}_{it}'\mathbf{b}_i. \tag{3}$$

The basic idea of the above model is the following: the effects of omitted covariates on the response variable could be either time-constant or time-varying. These effects are summarized by the individual-specific random coefficients $\mathbf{b}_i$ and by the state-specific parameters $\beta_h(\tau)$ and the associated (latent) Markov structure, respectively. To simplify the notation, in the following, the dependence on the quantile τ will be omitted whenever clear from the context.
As standard in the hidden Markov model literature, the marginal distribution of the whole sequence of hidden states can be derived by exploiting the Markovian property of the hidden chain, thus leading to

$$f_s(\mathbf{s}_i;\delta,\mathbf{Q}) = f_s(s_{i1})\prod_{t=2}^{T_i} f_s(s_{it} \mid s_{it-1}). \tag{4}$$

Based on such modelling assumptions, the following expression for the likelihood function can be, easily, derived:

$$L(\cdot) = \prod_{i=1}^{n}\int_{\mathbb{B}}\sum_{\mathbf{s}_i}\left\{\prod_{t=1}^{T_i} f_{y|sb}(y_{it} \mid s_{it},\mathbf{b}_i)\delta_{s_{i1}}\prod_{t=2}^{T_i} q_{s_{it-q}s_{it}}\right\} f_b(\mathbf{b}_i)\mathrm{d}\mathbf{b}_i \tag{5}$$

The above integral has to be numerically approached since it can not be solved analytically. As previously highlighted, different approaches can be adopted to this aim; in the present a context we will focus on the nonparametric maximum likelihood approach [13]. Introduced in the mHMM framework by [22], such a method allows to relax any kind of parametric assumption on the random effect distribution, offers a great flexibility and allows overcome problems due to misspecification of the random effect distribution.
The random effect distribution $f_b(\cdot)$ is approximated via a discrete distribution on $G \leq n$ of support points [13, 17, 18]. Let us suppose this discrete distribution puts masses π_g on locations $\mathbf{b}_g$, i.e. $p_g = \Pr(\mathbf{B}_i = \mathbf{b}_g), \sum_g \pi_g = 1, g = 1,...,m$. The likelihood function (5) is then approximated via

$$L(\cdot) = \prod_{i=1}^{n}\sum_{g=1}^{G}\sum_{\mathbf{s}_i}\left\{\prod_{t=1}^{T_i} f_{y|sb}(y_{it} \mid s_{it},\mathbf{b}_g)\delta_{s_{i1}}\prod_{t=2}^{T_i} q_{s_{it-q}s_{it}}\right\}\pi_g \tag{6}$$

where $f_{y|sb}(y_{it} \mid s_{it},\mathbf{b}_g)$ is the conditional distribution of the response variable, for a generic unit i being, at time t, in the hidden state s_{it} and belonging to the g-th component of the finite mixture. The number of mixture components G is treated as fixed and estimated via model selection techniques (see e.g. [4]) while locations and masses are treated as unknown and estimated together with the remaining model parameters, in a mixed HMM framework.

4 Computational details

Even if the likelihood (6) can be directly computed, it is cumbersome to maximize because of the presence of a multiple summation over all possible sequence of hidden states $\mathbf{s}_i$. To overcome the problem, parameter estimation can be performed by means of the EM algorithm. Introduced by [5], it is frequently used in the presence of latent variables, because of its robustness and ease of application.
Let $u_i(h) = \mathbb{I}[S_{it} = h]$ be the indicator variable for the i-th subject in the h-th state at time t and let $u_{it}(h,k) = \mathbb{I}[S_{it-1} = h, S_{it} = k]$ be equal to 1 if the i-th subject moves from the h-th state to the k-th at time t. Moreover, let $\eta_i(g) = \mathbb{I}[\mathbf{B}_i = \mathbf{b}_g]$ be an indicator variable equal to one if the i-th unit comes from the g-th mixture component. Parameter estimation starts from the definition of the following complete data

log-likelihood

$$\ell_c(\cdot) \propto \sum_{i=1}^{n} \left\{ \sum_{h=1}^{m} u_{i1}(h) \log \delta_h + \sum_{t=2}^{T_i} \sum_{h,k=1}^{m} u_{it}(h,k) \log q_{hk} + \sum_{g=1}^{G} \eta_i(g) \log \pi_g \right.$$
$$\left. - T_i \log(\sigma) - \sum_{t=1}^{T_i} \sum_{h=1}^{m} \sum_{g=1}^{G} \left[u_{it}(h) \eta_i(g) \rho_\tau \left(\frac{y_{it} - \mu_{it}(S_{it} = h, \mathbf{b}_g; \tau)}{\sigma} \right) \right\}. \quad (7)$$

At each step of the EM algorithm, the E step require the computation the posterior probabilities of the indicator variables $u_{it}(h), u_{it}(h,k)$ and $\eta_i(g)$, given the observed data and the current parameter estimates. As usually done in the hidden Markov model framework, such a computation can be greatly simplified by considering the well know forward and backward variables, as in the Baum-Welsh algorithm [3]. Adapted to the present framework, forward variables, $a_{it}(h, \mathbf{b}_g)$, are defined as being 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:

$$a_{it}(h, \mathbf{b}_g) = f\left[y_{i1:t}, S_{it} = h \mid \mathbf{b}_g\right]. \quad (8)$$

As observed by [3], these terms can be recursively computed through

$$a_{i1}(h, \mathbf{b}_g) = \delta_h f_{y|sb}\left[y_{i1} \mid S_{i1} = h, \mathbf{b}_g\right] \quad (9)$$
$$a_{it}(h, \mathbf{b}_g) = \sum_{k=1}^{m} a_{it-1}(k, \mathbf{b}_g) q_{kh} f_{y|sb}\left[y_{it} \mid S_{it} = h, \mathbf{b}_g\right].$$

The backward variables, $b_{it}(h, \mathbf{b}_g)$, are similarly defined and represent the probability of the longitudinal sequence from occasion $t+1$ to the last observation, conditional on being in the h-th state at time t and the g-th mixture component:

$$b_{it}(h, \mathbf{b}_g) = f\left[y_{it+1:T_i} \mid S_{it} = h, \mathbf{b}_g\right]. \quad (10)$$

As for the forward variables, also backward variables can be derived recursively:

$$b_{iT_i}(h, \mathbf{b}_g) = 1 \quad (11)$$
$$b_{it-1}(h, \mathbf{b}_g) = \sum_{k=1}^{m} b_{it}(k, \mathbf{b}_g) q_{hk} f_{y|sb}\left[y_{it} \mid S_{it} = k, \mathbf{b}_g\right],$$

where T_i represent the last measurement occasion for the i-th unit. For a detailed description of the general form and the properties of the Baum-Welsh algorithm, see the seminal paper by [3] and the reference monograph by [24].
Computation of the expected complete data log-likelihood, given the observed data and the current parameter estimates, leads to the following expression

$$Q(\cdot) = \sum_{i=1}^{n} \left\{ \sum_{h=1}^{m} \hat{u}_{i1}(h) \log \delta_h + \sum_{t=2}^{T_i} \sum_{h,k=1}^{m} \hat{u}_{it}(h,k) \log q_{hk} + + \sum_{g=1}^{G} \hat{\eta}_i(g) \log \pi_g \right.$$
$$\left. - T_i \log(\sigma) - \sum_{t=1}^{T_i} \sum_{h=1}^{m} \sum_{g=1}^{G} \left[\hat{\eta}_i(g) \hat{u}_{it}(h \mid g) \rho_\tau \left(\frac{y_{it} - \mu_{it}(S_{it} = h, \mathbf{b}_g; \tau)}{\sigma} \right) \right\}, \quad (12)$$

where $\hat{u}_{it}(h)$, $\hat{u}_{it}(h,k)$ and $\hat{\eta}_i(g)$ represent the expected values of the corresponding binary variables previously introduced; these are computed with respect to the posterior marginal distribution of the hidden Markov process and the finite mixture, respectively. All of these quantities can be easily obtained by exploiting the forward and backward variables (9) and (11), according to

$$\hat{u}_{it}(h) = \frac{\sum_g a_{it}(h, \mathbf{b}_g) b_{it}(h, \mathbf{b}_g) \pi_g}{\sum_h \sum_g a_{it}(h, \mathbf{b}_g) b_{it}(h, \mathbf{b}_g) \pi_g}$$
$$\hat{u}_{it}(h,k) = \frac{\sum_g a_{it-1}(h, \mathbf{b}_g) q_{hk} f_{y|sb}(y_{it} \mid S_{it} = h, \mathbf{b}_g) b_{it}(k, \mathbf{b}_g) \pi_g}{\sum_{hk} \sum_g a_{it-1}(h, \mathbf{b}_g) q_{hk} f_{y|sb}(y_{it} \mid S_{it} = h, \mathbf{b}_g) b_{it}(k, \mathbf{b}_g) \pi_g}.$$
$$\hat{\eta}_i(g) = \frac{\sum_{h=1}^{m} a_{iT_i}(h,g) \pi_g}{\sum_{g=1}^{G} \sum_{h=1}^{m} a_{iT_i}(h,g) \pi_g}$$

Furthermore, $\hat{u}_{it}(h \mid g)$ indicates the conditional posterior probability of being in state h at time occasion t, for a generic unit i belonging to the g-th mixture components and is computed as

$$\hat{u}_{it}(h \mid g) = \frac{a_{it}(h,g) b_{it}(h,g) \pi_g}{\sum_{h=1}^{m} a_{it}(h,g) b_{it}(h,g) \pi_g}.$$

The M-step of the EM algorithm require the maximization with respect to model parameters. Closed form solutions can be found for those parameters related to the hidden Markov process and for the prior probabilities of the finite mixture. The former are obtained by means of the following expressions

$$\hat{\delta}_h = \frac{\sum_{i=1}^{n} \hat{u}_{i1}(h)}{n}, \quad \hat{q}_{hk} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \hat{u}_{it}(h,k)}{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \sum_{k=1}^{m} \hat{u}_{it}(h,k)} \quad j,k = 1, \dots, m \qquad (13)$$

while the latter are estimated as

$$\hat{\pi}_g = \frac{1}{n} \sum_{i=1}^{n} \hat{\eta}_i(g) \quad g = 1, \dots, G. \qquad (14)$$

Longitudinal parameters, β's and $\mathbf{b}$'s, are simultaneously estimated; defining $\Psi = (\phi, \beta_1, \dots, \beta_m, \mathbf{b}_1, \dots, \mathbf{b}_g)$ as the set of longitudinal model parameters, the corresponding update is obtained as the solutions of the M-step equations

$$\frac{\partial Q(\cdot)}{\partial \Psi} = \sum_{i=1}^{n} \sum_{t=1}^{T_i} \sum_{h=1}^{m} \sum_{g=1}^{G} \hat{\eta}_i(g) \hat{u}_{it}(h \mid g) \frac{\partial}{\partial \Phi} \left\{ \rho_\tau \left(\frac{y_{it} - \mu_{it}(s_{it}, \mathbf{b}_g; \tau)}{\sigma} \right) \right\} = 0 \quad (15)$$

Similarly, the scale parameter is updated via the following expression

$$\hat{\sigma} = \frac{1}{\sum_{i=1}^{n} T_i} \sum_{t=1}^{T_i} \sum_{h=1}^{m} \sum_{g=1}^{G} \hat{\eta}_i(g) \hat{u}_{it}(h \mid g) \rho_\tau (y_{it} - \mu_{it}(h, \mathbf{b}_g; \tau)) \quad (16)$$

The E and the M steps are alternated repeatedly until the difference between two subsequent likelihoods is lower than a fixed constant ε, that is

$$\ell^{(r+1)} - \ell^{(r)} < \varepsilon.$$

Once the algorithm has reached the convergence for a given number of mixture components, parameter estimates are computed for different values of G: a formal comparison between penalized likelihood criteria associated to estimated models allows to identify the best number of mixture components.

5 Conclusions and further developments

We propose a mixed hidden Markov model for conditional quantiles in the longitudinal data framework. Such a model has, at least, a twofold aim. On one hand, time-constant and time-varying unobserved heterogeneity, determining the dependence between longitudinal observations, are jointly taken into account. On the other hand, it offers a complete overview on the effect that observed covariates have on the response variable distribution, correcting the potential bias on parameter estimates coming from possible outlier observations. We adopt a non-parametric maximum likelihood approach to obtain parameter estimates in an expectation-maximization perspective. Parametric distributional assumptions on the random effect are relaxed in order to avoid misspecification of such component and obtain a higher flexibility: since locations are completely free to vary over the corresponding support, extreme departures from the basic (homogeneous) model can be, easily, accommodated. Moreover, both the Markovian and the finite mixture process allow to classify subjects in clusters with common value of the random parameters: in health sciences, this results particularly useful as sample units can be divided in groups characterized by similar propensity to the event of interest. A comparison of parameter estimates obtained via the NPML approach and the Gaussian quadrature rule proposed by [9, 10], which should be properly extended to deal with mixed HMM models, can be explored. Furthermore, in our development we have considered continuous response variables. An extension to categorical data could be of great interest as well.

References

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

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

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

[4] D. Boehning. The em algorithm with gradient function update for discrete mixtures with known (fixed) number of components. *Statistics and Computing*, 13(3):257–265, 2003.

[5] A.P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38, 1977.

[6] A. Farcomeni. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Stat. Comput.*, 22, 2012.

[7] A. F. Galvao. Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164(1):142–157, 2011.

[8] A. F. Galvao and G. V. Montes-Rojas. Penalized quantile regression for dynamic panel data. *J. Statist. Plann. Inference*, 140(11):3476–3497, 2010.

[9] M. Geraci and M. Bottai. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–54, 2007.

[10] M. Geraci and M. Bottai. Linear quantile mixed models. *Statistics and Computing*, pages 1–19, 2013.

[11] R. Koenker. Quantile regression for longitudinal data. *J. Multivariate Anal.*, 91(1):74–89, 2004.

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

[13] N. Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811, 1978.

[14] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

[15] C Lamarche. Robust penalized quantile regression estimation for panel data. *J. Econometrics*, 157(2):396–408, 2010.

[16] D. Lee and T. Neocleous. Bayesian quantile regression for count data with application to environmental epidemiology. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, 59(5):905–920, 2010.

[17] B. G. Lindsay. The geometry of mixture likelihoods: a general theory. *Ann. Statist.*, 11(1):86–94, 1983.

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

[19] S. R. Lipsitz, G. M. Fitzmaurice, G. Molenberghs, and L. P. Zhao. Quantile regression methods for longitudinal data with drop-outs: Application to cd4 cell counts of patients infected with the human immunodeficiency virus. *Journal*

of the Royal Statistical Society: Series C (Applied Statistics), 46(4):463–476, 1997.

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

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

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

[23] Y. Yuan and G. Yin. Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, 66(1):105–114, 2010.

[24] W. Zucchini and I.L. MacDonald. *Hidden Markov models for time series*, volume 110 of *Monographs on Statistics and Applied Probability*. CRC Press, 2009.