Quantile mixed hidden Markov models for multivariate longitudinal data: An application to children's Strengths and Difficulties Questionnaire scores

Luca Merlo[1] | Lea Petrella[2] | Nikos Tzavidis[3]

[1]Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy

[2]MEMOTEF Department, Sapienza University of Rome, Rome, Italy

[3]Department of Social Statistics and Demography, Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, UK

Correspondence
Luca Merlo, Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy.
Email: luca.merlo@uniroma1.it

Abstract

The identification of factors associated with mental and behavioural disorders in early childhood is critical both for psychopathology research and the support of primary health care practices. Motivated by the Millennium Cohort Study, in this paper we study the effect of a comprehensive set of covariates on children's emotional and behavioural trajectories in England. To this end, we develop a quantile mixed hidden Markov model for joint estimation of multiple quantiles in a linear regression setting for multivariate longitudinal data. The novelty of the proposed approach is based on the multivariate asymmetric Laplace distribution which allows to jointly estimate the quantiles of the univariate conditional distributions of a multivariate response, accounting for possible correlation between the outcomes. Sources of unobserved heterogeneity and serial dependency due to repeated measures are modelled through the introduction of individual-specific, time-constant random coefficients and time-varying parameters evolving over time with a Markovian structure respectively. The inferential approach is carried out through the construction of a suitable expectation–maximization algorithm

without parametric assumptions on the random effects distribution.

1 | INTRODUCTION

The occurrence of stressful life events, family environment and poverty is key contributing risk factors to children's emotional and behavioural disorders (Bradley & Corwyn, 2002; Flouri et al., 2010; Goodman et al., 2003; Goodnight et al., 2012; Platt et al., 2016). Mental health problems at an early age can create considerable distress for the child and the family and can have a significant impact on the child's social, emotional and psychological development. Therefore, accurate identification of mental disorders is important for psychologists and clinicians in order to reduce this disruption, avoiding that psychological problems will persist into adulthood, and to better understand the problem for timely treatment recommendations (Becker et al., 2004; Mathai et al., 2004; Van der Meer et al., 2008). This is particularly for children facing major problems where, as risk factors accumulate, emotional and behavioural problems tend to increase considerably (Trentacosta et al., 2008). In this context, one of the most widely and internationally used measure of child mental health is provided by the Strengths and Difficulties Questionnaire (SDQ, see Goodman, 1997; Goodman & Goodman, 2009). It offers a balanced coverage of children and young people's behaviours, emotions and relationships and it has been designed to measure children's emotional and behavioural problems in psychological research. SDQ outcomes are assessed by diagnostic measures of child disorders, and collected by parent- and teacher-reported measures of personal, emotional and social development. On one hand, internalizing behaviours are manifested by inward symptoms such as being withdrawn, fearful or anxious; on the other hand, externalizing behaviours are outward and may be described as aggressive, non-compliant, impulsive or fidgety. The SDQ score is the sum of the main caregiver's responses to a series of items that describe children's internalizing and externalizing problems. This covers five different domains: emotional symptoms, peer problems, conduct problems, hyperactivity and pro-social behaviour. Each domain is measured by five items, for a total of 25 items. For each item, a score equal to 0 is given if the response is not true, 1 if it is somewhat true and 2 if it is certainly true. The internalizing SDQ score is the sum of the scores for responses to the five items in the domains of emotional and peer problems while the externalizing SDQ score is the sum of responses to the five conduct problems and hyperactivity items. Therefore, both SDQ scores range from 0 to 20 (for further details refer to www.sdqinfo.com).

In this paper we focus on emotional and behavioural disorders of children who participated in the Millennium Cohort Study (MCS), which is a longitudinal birth cohort study following children born in the United Kingdom (UK), providing multiple measures of the cohort members' physical, socio-emotional, cognitive and behavioural development over time. An extensive literature has examined and documented the effect of mother's characteristics, neighbourhood context and family risk factors on children's trajectories SDQ scores collected by the MCS (see, e.g. Alfò et al., 2021; Flouri & Sarmadi, 2016; McMunn et al., 2012; Tzavidis et al., 2016; Wickham et al.,

2017). Thanks to the longitudinal structure of the cohort data, these studies can shed light on the evolution of SDQ scores over time and on how they are affected by risk factors and other family and child characteristics.

The analysis of SDQ data poses crucial challenges for statistical modelling. As shown in Figure 1, the distributions of SDQ scores are non-negative, positively skewed and exhibit atypical values. Typically, linear random effect models (Diggle et al., 2002; Goldstein, 2011; Laird & Ware, 1982; Lindsey, 1999) have been implemented with much of the focus being univariate and centred on the conditional mean of the distribution given a set of covariates. Nevertheless, modelling the conditional mean may not offer the best summary as linear models perform badly with non-Gaussian data. Moreover, it is possible that the effect of certain risk factors on the SDQ scores is not the same across the SDQ distribution. Indeed, there is empirical evidence to suggest that socio-economic and parental factors have a more pronounced effects at the top end where children display a high, perhaps abnormal, level of adjustment problems than at the bottom end of the distribution (see e.g. Kiernan & Huerta, 2008; Tzavidis et al., 2016). It is therefore important to identify the predictors of children's disorders, not only at the average but more importantly, in the upper part of the SDQ distribution as this relates to high-risk youths having elevated levels of mental health and conduct problems.

In this context, a quantile regression approach can be more appropriate for getting a complete picture of the entire conditional distribution of children's difficulties than ordinary regression around the mean. Quantile regression provides a way to model the conditional quantiles of a response variable with respect to a set of covariates and may reveal how their effect varies at different parts of the response distribution. Quantile regression methods have become widely used in literature because they are suitable in those situations where skewness, fat-tails, outliers, truncation, censoring and heteroscedasticity arise, and they have been implemented in a wide range of different fields, both in a frequentist paradigm and in a Bayesian setting. For a detailed review and list of references see Koenker (2005) and Koenker et al. (2017).

Owing to the longitudinal structure of the MCS cohort data (measurements recorded on the same individuals), the potential association between dependent observations should be taken into
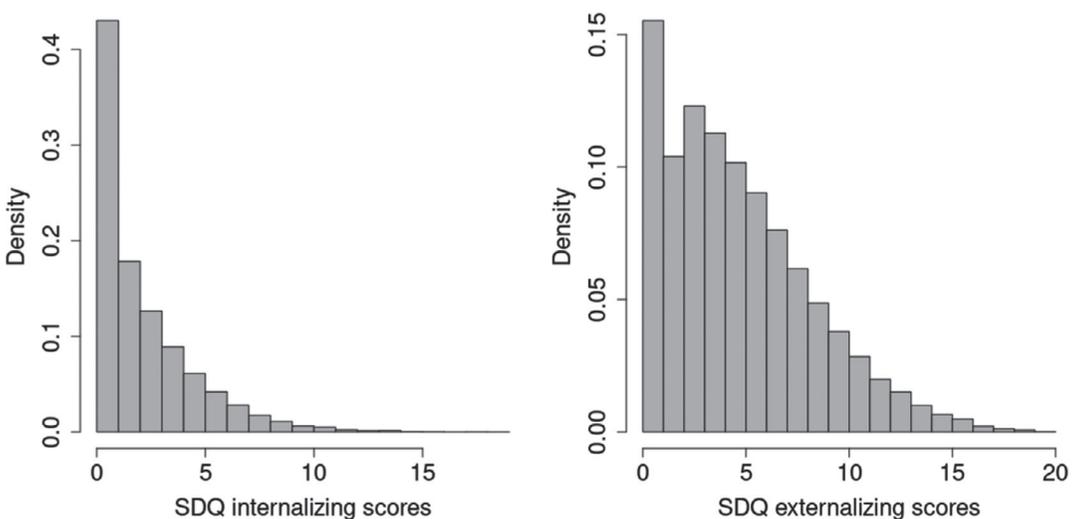
FIGURE 1 Histogram for Strengths and Difficulties Questionnaire internalizing (left) and externalizing (right) problems scores

account in order to provide correct inference. Within the quantile regression literature, random effects models have been presented to accommodate time-constant, within-subject correlation and between subject heterogeneity: Liu and Bottai (2009) and Geraci and Bottai (2014) proposed to add time-constant individual-specific random coefficients in the regression model to capture unobserved heterogeneity. However, when analysing internalizing and externalizing problems, it is reasonable to expect that individual temporal trajectories of SDQ scores vary from child to child. Figure 2 shows the individual trajectories of SDQ scores for a random subset of children, where the overall trend, as estimated by a local polynomial regression, is shown in red along with the 95% confidence bands (highlighted in grey). While the general trend is relatively constant over time, individual trajectories show rapid changes, especially 'U'-shaped curves for externalizing score measurements. For example, trajectories for children exposed to 'high risk' circumstances may become less salient with age, whereas for children at 'low risk', they could become more salient as they make the transition into adolescence. This possibly translates to decreasing and increasing intercepts in the dynamics of SDQ levels that require specific modelling tools. In this case, time-constant random effects can lead to biased estimates (Bartolucci & Farcomeni, 2009; Farcomeni, 2012) meanwhile, their temporal evolution can be better captured by the introduction of time-dependent intercepts.

To account for serial heterogeneity, Farcomeni (2012) suggested the use of hidden Markov models (HMMs). In such a context, a latent homogeneous Markov chain is defined in order to capture the temporal evolution of unobserved heterogeneity and state-dependent parameters are introduced to account for response variability due to time-varying omitted covariates. To handle longitudinal data where both time-constant and time-varying sources of unobserved heterogeneity are present, it is possible to consider the well-known mixed hidden Markov models (MHMMs, see Altman, 2007). The MHMM is obtained by combining the features of HMMs and mixed effects models, encompasses linear mixed models and HMMs as it accommodates time-constant and time-varying sources of variability jointly. In the application of quantile regression to longitudinal data, Marino et al. (2018) introduced a mixed hidden Markov quantile regression model for longitudinal continuous responses, extending the linear quantile mixed model of Geraci and Bottai
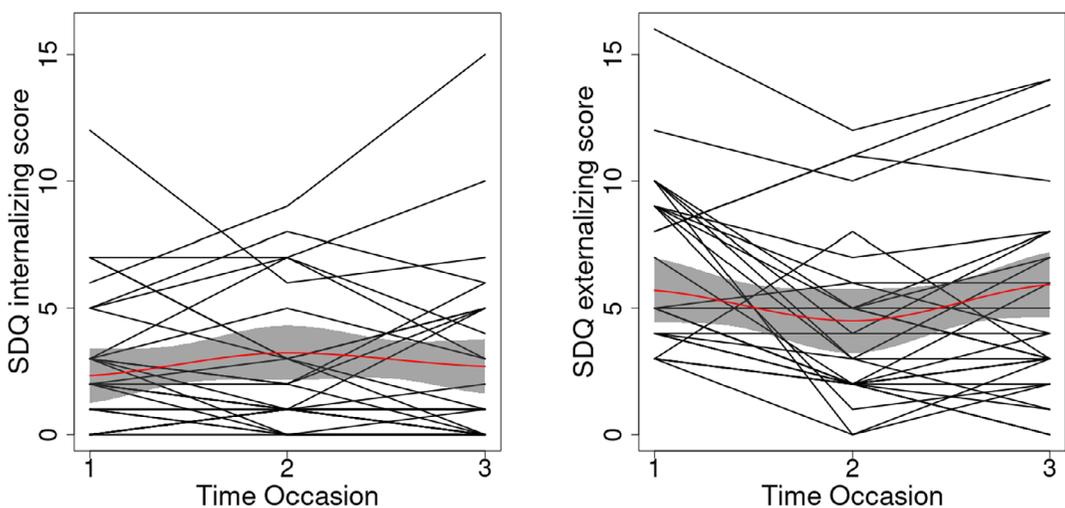
FIGURE 2 Individual trajectories for a random subsample of 30 children for Strengths and Difficulties Questionnaire internalizing (left) and externalizing (right) problems [Colour figure can be viewed at wileyonlinelibrary.com]

(2014) and the linear quantile HMM of Farcomeni (2012). These proposals are, however, designed for univariate dependent variables and consequently, they do not account for the dependence structure between multiple outcomes of interest measured longitudinally.

Another important aspect when analysing SDQ scores concerns the dependence structure between mental and behavioural problems. In the literature, Lilienfeld (2003), Liu (2004), Cicchetti and Toth (2014) and Alfò et al. (2021) have given empirical proof of the existence of correlation between SDQ scores, which places children at a greater risk of developing internalizing–externalizing comorbidity. In this case, univariate approaches completely ignore the correlation structure between children outcomes; by contrast, a multivariate analysis would be able to identify the underlying drivers that affect changes in each of the distribution of the SDQ scores and, at the same time, give valuable insight about the correlation among children disorders.

When multivariate response variables are concerned, however, the univariate quantile regression method does not straightforwardly extend to higher dimensions since there is no 'natural' ordering in a p-dimensional space, for $p > 1$. As a consequence, the search for a satisfactory notion of multivariate quantile has led to a flourishing literature on this topic despite its definition is still a debatable issue (see Alfò et al., 2021; Chakraborty, 2003; Charlier et al., 2020; Chavas, 2018; Hallin et al., 2010; Koenker et al., 2017; Kong & Mizera, 2012; Merlo et al., 2021; Stolfi et al., 2018 and the references therein for relevant studies). Recently, Petrella and Raponi (2019) generalized the AL distribution inferential approach of the univariate quantile regression to a multivariate framework by using the multivariate asymmetric Laplace (MAL) distribution defined in Kotz et al. (2012). By using the MAL distribution as a likelihood based inferential tool, the authors sidestep the problem of defining the quantiles of a multivariate distribution, and instead implement joint estimation for the univariate quantiles of the conditional distribution of a multivariate response variable given covariates, accounting for possible correlation among the responses.

In this paper, we generalize the work of Petrella and Raponi (2019) by introducing a MHMM to the longitudinal data setting. From a methodological point of view, we develop a quantile mixed hidden Markov model (QMHMM) to jointly estimate the quantiles of the univariate conditional distributions of a multivariate response, accounting for the dependence structure between the outcomes. This is a flexible approach as time-constant unobserved heterogeneity is described via individual-specific random coefficients while temporal effects are captured through state-specific parameters that evolve over time depending on a hidden Markov chain. In order to prevent inconsistent parameter estimates due to misspecification of the random effects distribution, we adopt the non-parametric maximum likelihood (NPML) approach of Lindsay (1983) where the distribution is left unspecified and approximated by a multivariate discrete finite mixture distribution estimated from the data. Within this scheme, our modelling framework reduces to a multivariate finite mixture of HMM quantile regressions.

As in Petrella and Raponi (2019), we propose to estimate the model parameters through maximum likelihood (ML) by implementing a suitable expectation–maximization (EM) algorithm, which exploits the Gaussian-mixture representation of the MAL distribution where both the hidden Markov chain and the random effects parameters are treated as missing data. Using simulation experiments we assess the validity of the proposed methodology by considering different data generating processes.

From an applied standpoint, the proposed method is then used to model the quantiles of the conditional distributions of SDQ scores of children in England by using the MCS data. Our approach allows to (i) jointly study the effect of demographic and socio-economic risk factors on internalizing and externalizing disorders in childhood, (ii) investigate whether their effect is more pronounced for children experiencing high levels of disorders (i.e., high quantiles) and

(iii) account for the association between SDQ scores and unobserved heterogeneity sources by including into the model time-constant subject-specific random coefficients, without parametric assumptions, and time-dependent random intercepts.

With respect to existing approaches in the literature, the introduced QMHMM can be thought of as a generalization of the univariate QMHMM of Marino et al. (2018) to multivariate longitudinal data. Furthermore, our approach differs substantially from the ones in Kulkarni et al. (2019) and Alfò et al. (2021). The authors consider, respectively, univariate quantile and M-quantile (Breckling & Chambers, 1988) regression models with outcome-specific random effects, where the dependence between responses is captured by time-constant outcome-specific random coefficients. To the best of our knowledge, this is the first attempt to simultaneously estimate multiple conditional quantiles for multivariate longitudinal outcomes, that include both time-constant and time-varying random effects.

The rest of the paper is organized as follows. In Section 3, we introduce the proposed QMHMM regression framework. In Section 2 we describe the MCS data. Section 4 illustrates the EM-based ML approach to estimate model parameters together with M-step updates in closed form. In Section 5 we discuss the empirical application while Section 6 summarizes our conclusions. All proofs are presented in Appendix A, while the simulation study is presented in Appendix B.

2 | THE MCS DATA

The MCS study (http://www.cls.ioe.ac.uk) is a longitudinal, nationally representative study that follows the lives of around 19,000 young people born across UK in 2000–2002. The MCS is designed to provide data about children living and growing up in each of the four countries of the UK namely, England, Wales, Scotland and Northern Ireland, especially for sub-groups of children living in advantaged and disadvantaged circumstances, for children of ethnic minorities and those living in Scotland, Wales and Northern Ireland. In addition to children, it provides data about their families and the broader socio-economic context in which the children live and grow up.

The MCS population is defined as all children born between 1 September 2000 and 31 August 2001 (for England and Wales), and between 24 November 2000 and 11 January 2002 (for Scotland and Northern Ireland), living in the UK at age 9 months, and whose families were eligible to receive child benefits. In order to meet the principle of adequately representing disadvantaged and ethnic minority children, the population was stratified by UK country. For England, the population was stratified, via the stratification of electoral wards extant on 1 April 1998, into three strata. The first stratum (ethnic stratum) is composed of children living in wards where the proportion of ethnic minorities was at least 30% in the 1991 Population Census. The second one (disadvantaged stratum) was based on the Child Poverty Index, recording the proportion of children, in a ward whose families received means-tested benefits. The cut-off value to define a disadvantaged ward was 38.4% receiving such benefits, corresponding to the poorest 25% wards in England and Wales in 1998. The third one (advantaged stratum) consists of all other children. For Wales, Scotland and Northern Ireland, there were just the disadvantaged and advantaged strata. MCS wards were randomly selected within each stratum and country; then a list of all children turning 9 months old during the survey window and living in the selected wards was created. The MCS sample members were first surveyed when they were around 9 months of age, and cohort members continue to remain eligible to be surveyed if they remain living in or returning to the UK. Follow-ups have currently been conducted at 3, 5, 7, 11, 14 and 17 years of age. The main data collection from parents at the first survey was by a face-to-face computer assisted interview and self-completion

(CAPI and CASI). The full set of questions, typically lasting 70–75 min went to a main informant, almost always the mother, and a shorter set of questions, taking around 30 min went to the main informant's partner. For further details, please refer to the technical report of Plewis et al. (2007) and studies of Connelly and Platt (2014) and Joshi and Fitzsimons (2016). This study is widely regarded as the basis of the most reliable estimates of cognitive development problems in young people and it has been thoroughly investigated in the fields of child psychology and pedagogy: see, among others, the works of Griffiths et al. (2011), Goodman and Goodman (2011), Tzavidis et al. (2016), Bell et al. (2019), Ahn et al. (2018) and Alfò et al. (2021).

In order to analyse the impact of child and mother personal characteristics, neighbourhood context and family risk factors on children disorders, the following set of predictors is considered. ALE 11 measures the number, out of 11 events, of potentially stressful life events experienced by the family between two consecutive sweeps. The events, classified on the basis of the scale proposed by Tiet et al. (1998), are family member died, negative change in financial situation, new step-parent, sibling left home, child got seriously sick or injured, divorce or separation, family moved, parent lost job, new natural sibling, new stepsibling and mother diagnosed with or treated for depression. SED 4 measures the household's socio-economic disadvantage condition by combining information on overcrowding (more than 1.5 people per room excluding the bathroom and kitchen), not owning a home, receipt of means-tested income support and income poverty. Mother's personal characteristics and distress psychological indicators are included such as maternal education (no qualification (baseline), university degree or General Certificate of Secondary Education (GCSE)) and maternal depression, Kessm, measured by the Kessler score and diagnosed by a doctor. Furthermore, child's age in years, centred around the mean, age year scal, the quadratic effect of child's age, age2 year scal, ethnicity (non-white (baseline) or white) and gender (female (baseline) or male) were included in the model. Finally, three explanatory variables evaluate the area characteristics. IMD is a time varying variable which measures neighbourhood deprivation by the index of multiple-deprivation score. A design variable which allows for the stratification of the MCS sampling design: the stratification variable of the MCS consists of three categories, namely the advantaged stratum (baseline category), the ethnic stratum, eth stratum and the disadvantaged stratum, dis stratum. To allow for comparison, the considered predictors are the same as those in Tzavidis et al. (2016).

The data that we use in this paper are SDQ internalizing, SDQ_{Int}, and SDQ externalizing, SDQ_{Ext}, scores recorded on children who were observed at all measurement occasions, that is, the considered sample consists of $N = 5342$ units and $T_i = T = 3$, for all $i = 1, \ldots, N$. As is customary in child psychology, SDQ scores are treated as though they are continuous variables. Table 1 presents the main descriptive statistics of continuous and categorical variables considered in the sample. The average values of ALE 11, SED 4 and Kessm are 1.405, 0.555 and 2.597 but, there are cases with much higher scores as demonstrated by their maximum values. 43% of children have mothers who hold a degree and 48% have mothers with GCSE or other qualification. Around 50% of children are males and, in relation to ethnicity, 89% of members are white. The sample also includes 8.2% and 38% of families from the ethnic and disadvantaged strata respectively. Furthermore, the empirical correlation between SDQ_{Int} and SDQ_{Ext} equals to 0.369. As expected, internalizing and externalizing problems are positively correlated, justifying the joint modelling approach we adopt in this paper.

This exploratory analysis and all the preliminary considerations raised in the introduction suggest that a QMHMM that accounts for time-constant and time-dependent unobserved heterogeneity sources offers an approach to modelling the data. In order to account for the data features and provide correct statistical inferences, the modelling approach we propose defines a

TABLE 1 Summary statistics for the Millennium Cohort Study (MCS) data

Variable	Minimum	First quartile	Median	Mean[a]	Third quartile	Maximum
SDQ$_{Int}$	0	1	2	2.493	4	19
SDQ$_{Ext}$	0	2	5	5.144	7	20
ALE 11	0	1	1	1.405	2	7
SED 4	0	0	0	0.555	1	4
Kessm	0	0	2	2.597	4	24
Degree				43.429		
GCSE				47.997		
White				89.012		
Male				50.337		
IMD	1	3	6	5.619	8	10
Eth stratum				8.199		
Dis stratum				38.132		

[a]Means for dummy variables are reported in %.

joint quantile regression for multivariate longitudinal data to study the effects of environmental, parental and child factors across different quantiles of the SDQ distributions. This can offer a more complete picture of the determinants of children's problems useful for clinicians and educationalists, in order to design programs preventing the onset of psychopathologies and addressing the incidence of disorders in early childhood.

3 | METHODOLOGY

Let $\mathbf{Y}_{it} = (Y_{it}^{(1)}, \ldots, Y_{it}^{(p)})$ be a continuous p-variate response variable vector and $\mathbf{X}_{it} = (X_{it}^{(1)}, \ldots, X_{it}^{(k)})$ be a k-dimensional vector of explanatory variables for subject $i = 1, \ldots, N$ and time occasion $t = 1, \ldots, T_i$. Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$ denote p quantile indexes with $\tau_j \in (0, 1)$, for $j = 1, \ldots, p$ and let $S_{it}(\boldsymbol{\tau})$, $i = 1, \ldots, N$, $t = 1, \ldots, T_i$ be a homogeneous, first-order, aperiodic and irreducible hidden Markov chain defined over a discrete states space $S = \{1, \ldots, M\}$ with initial and transition probabilities denoted by $\mathbf{q} = (q_1, \ldots, q_M)$ and $\mathbf{Q} = \{q_{jk}\}$ over $S \times S$ common to all subjects respectively. Finally, let $\mathbf{b}_i(\boldsymbol{\tau})$ be a time-constant, subject-specific, random effects matrix having distribution $f_{\mathbf{b}}(\cdot \mid \mathbf{X}_{it}, \boldsymbol{\tau})$ with support $\mathcal{B}$, where $\mathbb{E}(\mathbf{b}_i(\boldsymbol{\tau})) = 0$ is used for parameter identifiability. We assume that the τ_jth quantile of each of the jth components of $\mathbf{Y}_{it}$ can be modelled as a function of explanatory variables. Let $\boldsymbol{\beta}(\boldsymbol{\tau}) = (\boldsymbol{\beta}_1(\boldsymbol{\tau}), \ldots, \boldsymbol{\beta}_p(\boldsymbol{\tau}))$ be the $k \times p$ matrix of unknown quantile regression coefficients. Then, the QMHMM is defined as follows:

$$\mathbf{Y}_{it} = \mathbf{X}_{it}\boldsymbol{\beta}(\boldsymbol{\tau}) + \mathbf{Z}_{it}\mathbf{b}_i(\boldsymbol{\tau}) + \mathbf{W}_{it}\boldsymbol{\alpha}_{S_{it}}(\boldsymbol{\tau}) + \varepsilon_{it}(\boldsymbol{\tau}) \tag{1}$$

where $\mathbf{Z}_{it}$ is a subset of $\mathbf{X}_{it}$, $\mathbf{W}_{it}$ is a further subset of $\mathbf{X}_{it}$ whose effects are assumed to vary over time, $\varepsilon_{it}(\boldsymbol{\tau})$ denotes a p-dimensional vector of error terms with univariate component-wise quantiles (at fixed levels $\tau_1, \ldots, \tau_p$ respectively) equal to zero and where the coefficients matrix $\boldsymbol{\alpha}_{S_{it}}(\boldsymbol{\tau})$ evolves over time according to the hidden Markov chain, $S_{it}(\boldsymbol{\tau})$, and takes one of the values in the set $\{\boldsymbol{\alpha}_1(\boldsymbol{\tau}), \ldots, \boldsymbol{\alpha}_M(\boldsymbol{\tau})\}$. In particular, the parameters $\mathbf{b}_i(\boldsymbol{\tau})$, $i = 1, \ldots, N$, and $\{\boldsymbol{\alpha}_1(\boldsymbol{\tau}), \ldots, \boldsymbol{\alpha}_M(\boldsymbol{\tau})\}$ are

designed to account for within-individual dependence by considering unobserved time-constant and time-varying sources of unobserved heterogeneity respectively.

Our objective is to provide joint estimation of the p quantiles of the univariate conditional distributions of $\mathbf{Y}_{it}$ taking into account for potential correlation among the dependent variables. The QMHMM framework is based on the following central assumptions, which are standard in mixed effects models. The random effects $\mathbf{b}_i(\boldsymbol{\tau})$ are independent of the hidden Markov chain, $S_{it}(\boldsymbol{\tau})$, as they are meant to capture different unobserved characteristics, and furthermore, it is assumed that the covariates $\mathbf{X}_{it}$ are uncorrelated with $\mathbf{b}_i(\boldsymbol{\tau})$, that is $f_{\mathbf{b}}(\cdot \mid \mathbf{X}_{it}, \boldsymbol{\tau}) = f_{\mathbf{b}}(\cdot \mid \boldsymbol{\tau})$. Regarding the longitudinal responses, they must satisfy the contemporary dependence and conditional independence conditions. The former states that for the ith subject at time t, the distribution of $\mathbf{Y}_{it}$, given the state variables $(S_{i1}(\boldsymbol{\tau}), \dots, S_{iT_i}(\boldsymbol{\tau}))$ and the time-constant individual-specific random effects $\mathbf{b}_i(\boldsymbol{\tau})$, depends only on the current state $S_{it}(\boldsymbol{\tau})$; the latter entails that the responses $(\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iT_i})$ are conditionally independent, given the hidden state occupied at time t by $S_{it}(\boldsymbol{\tau})$ and the individual-specific random coefficients $\mathbf{b}_i(\boldsymbol{\tau})$. These assumptions imply that the following equality holds:

$$f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{y}_{i1:t-1}, \mathbf{x}_{i1:t-1}, s_{i1:t}, \mathbf{b}_i, \boldsymbol{\tau}) = f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{x}_{it}, s_{it}, \mathbf{b}_i, \boldsymbol{\tau}) \tag{2}$$

where $\mathbf{y}_{i1:t-1}$ and $\mathbf{x}_{i1:t-1}$ represent the history of the responses and the observed covariates for the ith subject up to time $t-1$, respectively, and $s_{i1:t}$ is the individual sequence of states up to time t.

Generalizing the approach of Petrella and Raponi (2019), for the model in (1) we consider the MAL distribution, $\mathcal{MAL}(\boldsymbol{\mu}, \mathbf{D}\tilde{\boldsymbol{\xi}}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D})$, (see Kotz et al., 2012) having density function:

$$f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{x}_{it}, s_{it}, \mathbf{b}_i, \boldsymbol{\tau}) = \frac{2 \exp\{(\mathbf{y}_{it} - \boldsymbol{\mu}_{it})' \mathbf{D}^{-1}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\xi}}\}}{(2\pi)^{p/2}|\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}|^{1/2}} \left(\frac{\tilde{m}_{it}}{2+\tilde{d}}\right)^{\nu/2} K_\nu\left(\sqrt{(2+\tilde{d})\tilde{m}_{it}}\right), \tag{3}$$

where the location parameter $\boldsymbol{\mu}_{it}$ is defined by the MHMM:

$$\boldsymbol{\mu}_{it} = \boldsymbol{\mu}(s_{it}, \mathbf{b}_i, \boldsymbol{\tau}) = \mathbf{X}_{it}\boldsymbol{\beta}(\boldsymbol{\tau}) + \mathbf{Z}_{it}\mathbf{b}_i(\boldsymbol{\tau}) + \mathbf{W}_{it}\boldsymbol{\alpha}_{s_{it}}(\boldsymbol{\tau}), \tag{4}$$

$\mathbf{D}\tilde{\boldsymbol{\xi}}$ is the skew parameter with $\mathbf{D} = \mathrm{diag}[d_1, \dots, d_p]$, $d_j > 0$ and $\tilde{\boldsymbol{\xi}} = [\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_p]'$ having generic element $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}, j = 1, \dots, p$. $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite matrix such that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}$, with $\boldsymbol{\Psi}$ being an unstructured correlation matrix of dimension p and $\boldsymbol{\Lambda} = \mathrm{diag}[\sigma_1, \dots, \sigma_p]$, with $\sigma_j^2 = \frac{2}{\tau_j(1-\tau_j)}, j = 1, \dots, p$. Moreover, $\tilde{m}_{it} = (\mathbf{y}_{it} - \boldsymbol{\mu}_{it})'(\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})^{-1}(\mathbf{y}_{it} - \boldsymbol{\mu}_{it})$, $\tilde{d} = \tilde{\boldsymbol{\xi}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\xi}}$, and $K_\nu(\cdot)$ denotes the modified Bessel function of the third kind with index parameter $\nu = (2-p)/2$.

One of the key benefits of the MAL distribution is that, using Equations (1) and (3), and following Kotz et al. (2012), $\mathbf{Y} \sim \mathcal{MAL}(\boldsymbol{\mu}, \mathbf{D}\tilde{\boldsymbol{\xi}}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D})$ can be written as a location-scale mixture, having the following representation:

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{D}\tilde{\boldsymbol{\xi}}\tilde{C} + \sqrt{\tilde{C}}\mathbf{D}\boldsymbol{\Sigma}^{1/2}\mathbf{Z} \tag{5}$$

where $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ denotes a p-variate standard normal distribution and $\tilde{C} \sim \exp(1)$ has a standard exponential distribution, with $\mathbf{Z}$ being independent of $\tilde{C}$. In particular, the constraints imposed on $\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{\Lambda}$ represent necessary conditions for model identifiability for any fixed quantile level $\tau_1, \dots, \tau_p$ and guarantee that $\boldsymbol{\mu}_{it}^{(j)}$ is the τ_jth conditional quantile function of $Y_{it}^{(j)}$ given $S_{it}(\boldsymbol{\tau})$ and $\mathbf{b}_i$, for $j = 1, \dots, p$. As shown in Petrella and Raponi (2019), using this approach we

are able to conduct inference on the quantiles of the univariate conditional distributions of $\mathbf{Y}_{it}$ simultaneously, taking into account the possible correlation between the outcomes. For a given quantile level τ, following Kotz et al. (2012) and by simple calculations it is possible to show that the covariance matrix of $\mathbf{Y}$ can be written as:

$$\mathbf{S} = \mathbf{D}(\tilde{\xi}\tilde{\xi}' + \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda})\mathbf{D}, \tag{6}$$

where the off-diagonal elements of $\mathbf{S}$ provide an indirect measure of association between the outcomes.

Two remarks are also noteworthy regarding the methodology introduced above. First, our model can be thought of as an extension to multivariate longitudinal data of: (i) the linear quantile hidden Markov model by Farcomeni (2012) when $\mathbf{W}_{it} = \mathbf{1}$ and $\mathbf{b}_i(\tau) = \mathbf{0}$ for all $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$; (ii) the linear quantile mixed model (LQMM) proposed in Geraci and Bottai (2014) when there is only one state of the hidden Markov chain, that is, $M = 1$. Second, the proposed approach differs substantially from the ones by Kulkarni et al. (2019) and Alfò et al. (2021). In the former, the authors consider univariate quantile regression models where the dependence across time and responses is captured by time-constant outcome-specific normally distributed random coefficients. In the latter, the proposed method targets a different set of location parameters, that is, the M-quantiles (Breckling & Chambers, 1988) of the distribution of the dependent variables, which are more difficult to interpret than quantiles. The authors then define univariate M-quantile (Breckling & Chambers, 1988) regression models with outcome-specific random effects, where dependence between outcomes for each unit is introduced by assuming correlated, subject-specific random effects in the univariate models.

Estimation of model parameters can be pursued using a ML approach. To ease the notation, unless specified otherwise, hereinafter we omit the quantile levels vector τ, yet all model parameters are allowed to depend on the p quantile indexes. Thus, let us denote by $\mathbf{\Phi}_\tau = (\boldsymbol{\beta}, \mathbf{D}, \mathbf{\Psi}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M, \mathbf{q}, \mathbf{Q})$ the set of model parameters. Given the modelling assumptions introduced so far, the observed data likelihood is defined by:

$$L(\mathbf{\Phi}_\tau) = \prod_{i=1}^{N} \int_{\mathcal{B}} \left\{ \sum_{S^{T_i}} \left[\prod_{t=1}^{T_i} f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{x}_{it}, s_{it}, \mathbf{b}_i) \right] q_{s_{i1}} \prod_{t=2}^{T_i} q_{s_{it-1}s_{it}} \right\} f_{\mathbf{b}}(\mathbf{b}_i) \, d\mathbf{b}_i. \tag{7}$$

The maximization of the likelihood in Equation (7) generally may prove to be excessively cumbersome because it involves a multidimensional integral over the random coefficient distribution $f_{\mathbf{b}}(\cdot)$ and a summation over M^{T_i} terms for each unit. In addition, the choice of an appropriate distribution for the random effects is not straightforward. Ideally $f_{\mathbf{b}}(\cdot)$ should be data driven and resistant to misspecification (Marino & Farcomeni, 2015), otherwise an incorrect distributional assumption for the random effects has unfavourable influence on statistical inferences (see Agresti et al., 2004; Maruotti, 2011; Neuhaus et al., 2013). In the next section, we discuss how we specify the random effects distribution and how we may avoid evaluating the integral in Equation (7) for ML estimation.

3.1 | Specification of the random coefficient distribution

In the literature, typically the Gaussian distribution is a convenient choice for $f_{\mathbf{b}}(\cdot)$ from a computational point of view. In this case, we may approximate the integral in Equation (7) using

Gaussian quadrature or adaptive Gaussian quadrature schemes (see Crowther et al., 2014; Pinheiro & Chao, 2006; Rabe-Hesketh et al., 2005). A disadvantage of such approaches lies in the required computational effort, which is exponentially increasing with the number of the random parameters. For this reason, potential alternatives proposed the use of simulation methods such as Monte Carlo and simulated ML approaches (McCulloch, 1997). However, for samples of finite size and short individual sequences, these methods may not provide a good approximation of the true mixing distribution (Alfò et al., 2017). As a robust alternative to the Gaussian choice, the multivariate symmetric Laplace or multivariate Student t distributions have been considered by Geraci and Bottai (2014) and Farcomeni and Viviani (2015). However, a parametric assumption on the distribution of the random coefficients could be rather restrictive and misspecification of the mixing distribution can lead to biased parameter estimates (see Alfò & Maruotti, 2010). Following Marino et al. (2018), in this work we exploit the approach based on the non-parametric maximum likelihood (NPML) estimation of Laird (1978) and extend it to the multivariate context. In particular, we do not parametrically specify $f_\mathbf{b}(\cdot)$ but we approximate it by using a discrete distribution defined on $G < N$ multivariate locations, $\mathbf{b}_g(\boldsymbol{\tau})$, with associated probabilities defined by:

$$\pi_g(\boldsymbol{\tau}) = \Pr(\mathbf{b}_i(\boldsymbol{\tau}) = \mathbf{b}_g(\boldsymbol{\tau})), \tag{8}$$

with $\pi_g \geq 0$, $\forall g = 1, \ldots, G$ and $\sum_{g=1}^{G} \pi_g = 1$. More concisely, we can write:

$$\mathbf{b}_i(\boldsymbol{\tau}) \sim \sum_{g=1}^{G} \pi_g(\boldsymbol{\tau}) \delta_{\mathbf{b}_g}(\boldsymbol{\tau}), \tag{9}$$

where δ_θ is a one-point distribution putting a unit mass at θ. With this approach, the parametric problem is thus converted to a semiparametric one, where $\mathbf{b}_g(\boldsymbol{\tau})$ and $\pi_g(\boldsymbol{\tau})$ define the discrete probability distribution of the random effects defined on G distinct support points. In this context, time-constant unobserved heterogeneity in the data is represented by a finite mixture with unknown proportions $\pi_g(\boldsymbol{\tau})$ and locations $\mathbf{b}_g(\boldsymbol{\tau})$ common to all subjects in the gth group. Since locations and masses are completely free to vary over the corresponding support, this is a flexible method that can readily accommodate a wide range of shapes, including fat-tailed and asymmetric distributions, and it is more robust against deviations from model assumptions. For the interested reader, a detailed survey about this method can be found in Aitkin and Alfó (1998), Alfò and Aitkin (2000), Aitkin and Alfò (2003), Alfò et al. (2017, 2021) and Merlo et al. (2021), for example.

In this setting, the observed data likelihood in Equation (7) reduces to:

$$L(\boldsymbol{\Phi}_\tau) = \prod_{i=1}^{N} \sum_{g=1}^{G} \left\{ \sum_{S^{T_i}} \left[\prod_{t=1}^{T_i} f_\mathbf{Y}(\mathbf{y}_{it} \,|\, \mathbf{x}_{it}, s_{it}, \mathbf{b}_g) \right] q_{s_{i1}} \prod_{t=2}^{T_i} q_{s_{it-1}s_{it}} \right\} \pi_g, \tag{10}$$

where $\boldsymbol{\Phi}_\tau = (\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\Psi}, \mathbf{b}_1, \ldots, \mathbf{b}_G, \pi_1, \ldots, \pi_G, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M, \mathbf{q}, \mathbf{Q})$ denotes the vector of model parameters and $f_\mathbf{Y}(\mathbf{y}_{it} \,|\, \mathbf{x}_{it}, s_{it}, \mathbf{b}_g)$ represents the response distribution of unit i being in the state s_{it} at time t and belonging to the gth component of the finite mixture, which is assumed to follow the MAL black as in Equation (3) with location parameter given by:

$$\boldsymbol{\mu}_{it} = \boldsymbol{\mu}(s_{it}, \mathbf{b}_g, \boldsymbol{\tau}) = \mathbf{X}_{it}\boldsymbol{\beta}(\boldsymbol{\tau}) + \mathbf{Z}_{it}\mathbf{b}_g(\boldsymbol{\tau}) + \mathbf{W}_{it}\boldsymbol{\alpha}_{s_{it}}(\boldsymbol{\tau}). \tag{11}$$

By looking at the likelihood in Equation (10), one can see that it resembles the likelihood of a finite mixture of HMM models with G homogeneous sub-populations where the presence of latent time-constant heterogeneity is described by discrete multivariate random effects. From the estimation perspective, locations $\mathbf{b}_g$ and corresponding probabilities π_g are unknown parameters which need to be estimated along with other model parameters. The number of mixture components G is unknown, and it is usually treated as fixed and estimated via penalized likelihood criteria (see e.g. Böhning, 1999). Furthermore, as an important by-product, the computational complexity of the likelihood evaluation in Equation (10) is of linear order with respect to G, which greatly facilitates the implementation of EM-type algorithm, as is described in the following section.

4 | MAXIMUM LIKELIHOOD ESTIMATION AND INFERENCE

As mentioned in the previous sections, the MAL density represents a convenient tool to jointly model the univariate quantiles of the conditional distribution of a multivariate response variable in a quantile regression framework. In this section we introduce a ML approach to estimate and make inference on model parameters and build a suitable EM algorithm (Dempster et al., 1977). We will show that the M-step update of all model parameters can be easily obtained in closed form, hence reducing the computational burden of the algorithm compared to direct maximization of the likelihood in Equation (10). Specifically, we derive the EM algorithm by exploiting the Gaussian location-scale mixture representation in Equation (5) of the MAL distribution under the constraints on $\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{\Lambda}$.

4.1 | The EM algorithm

The EM algorithm alternates between an expectation (E) step, which defines the expectation of the complete log-likelihood evaluated of the current parameters estimates, and a maximization (M) step, which computes parameter estimates by maximizing the expected complete log-likelihood obtained in the E-step. The complete log-likelihood, the expected complete log-likelihood function and the optimal parameter estimators are given below in the following propositions.

Given the representation in Equation (10), let us denote by w_{ig} the indicator variable that is equal to 1 if the ith unit belongs to the gth component of the finite mixture, and 0 otherwise. Similarly, let u_{itj} be equal to 1 if unit i is in state j at time t and 0 otherwise; let v_{itjk} be equal to 1 if unit i is in state j at time $t-1$ and in state k at time t, and 0 otherwise. Finally, we denote by z_{itjg} the indicator of the ith individual being in state j at time t and coming from the gth component of the mixture. The expected complete data log-likelihood is presented in the following proposition.

Proposition 1 *For any fixed $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$, G mixture components and M hidden states, the complete data log-likelihood function is proportional to*:

$$\ell_c(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) \propto \sum_{i=1}^{N} \left\{ \sum_{g=1}^{G} w_{ig} \log \pi_g + \sum_{j=1}^{M} u_{i1j} \log q_j + \sum_{t=2}^{T_i} \sum_{j=1}^{M} \sum_{k=1}^{M} v_{itjk} \log q_{jk} \right.$$

$$-\frac{1}{2}T_i \log |\mathbf{D\Sigma D}| + \sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg}(\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})'\mathbf{D}^{-1}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\xi}}$$

$$-\frac{1}{2}\sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg}\frac{1}{\tilde{C}_{itjg}}(\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})'(\mathbf{D\Sigma D})^{-1}(\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})$$

$$-\frac{1}{2}\tilde{\boldsymbol{\xi}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\xi}}\sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg}\tilde{C}_{itjg}\Bigg\}, \tag{12}$$

where $\tilde{C}_{itjg}$ *is a latent variable that follows an exponential distribution with parameter* 1.

In the E-step of the algorithm, the presence of the unobserved indicator variables w_{ig}, u_{itj}, v_{itjk} and z_{itjg} is handled by taking their conditional expectation given the observed data and the current parameter estimates. Calculation of such quantities may be addressed via an adaptation of the forward and backward variables; see Welch (2003). Similarly, the conditional expectations of $\frac{1}{\tilde{C}_{itjg}}$ and $\tilde{C}_{itjg}$ are considered.

For the implementation of the algorithm, forward and backward variables are defined for the longitudinal measures. We define the probability of observing the partial sequence ending up in state j at time t, given the gth component, as:

$$a_{it}(j,g) = f(\mathbf{y}_{i1:t}, S_{it} = j|\mathbf{b}_g) \quad \text{and} \quad a_{i1}(j,g) = q_j f(\mathbf{y}_{i1}|S_{i1} = j, \mathbf{b}_g). \tag{13}$$

The quantity $a_{it}(j,g)$ can be rewritten using the following recurrence relationship:

$$a_{it}(j,g) = \sum_{h=1}^{M} a_{it-1}(h,g)q_{hj}f(\mathbf{y}_{it} \mid S_{it} = j, \mathbf{b}_g). \tag{14}$$

Backward variables are defined as the probability of the longitudinal sequence from time $t + 1$ to the last available observation T_i, conditional on being in state j at time t, given the gth component:

$$b_{it}(j,g) = f(\mathbf{y}_{it+1:T_i} \mid S_{it} = j, \mathbf{b}_g) \quad \text{and} \quad b_{iT_i}(j,g) = 1. \tag{15}$$

Accordingly, the backward variable $b_{it}(j,g)$ can be rewritten as:

$$b_{it}(j,g) = \sum_{k=1}^{M} b_{it+1}(k,g)q_{jk}f(\mathbf{y}_{it+1} \mid S_{it+1} = k, \mathbf{b}_g). \tag{16}$$

Finally, the expected values of w_{ig}, u_{itj}, v_{itjk} and z_{itjg} can be computed as:

$$\hat{w}_{ig} = \frac{\pi_g \sum_{j=1}^{M} a_{iT_i}(j,g)}{\sum_{g=1}^{G} \pi_g \sum_{j=1}^{M} a_{iT_i}(j,g)},$$

$$\hat{z}_{itjg} = \frac{a_{it}(j,g)b_{it}(j,g)\pi_g}{\sum_{g=1}^{G}\sum_{j=1}^{M} a_{it}(j,g)b_{it}(j,g)\pi_g},$$

$$\hat{u}_{itj} = \sum_{g=1}^{G} \hat{z}_{itjg},$$

$$\hat{v}_{itjk} = \frac{\sum_{g=1}^{G} a_{it-1}(j,g) q_{jk} f(\mathbf{y}_{it} \mid S_{it} = k, \mathbf{b}_g) b_{it}(k,g) \pi_g}{\sum_{g=1}^{G} \sum_{j=1}^{M} \sum_{k=1}^{M} a_{it-1}(j,g) q_{jk} f(\mathbf{y}_{it} \mid S_{it} = k, \mathbf{b}_g) b_{it}(k,g) \pi_g}. \tag{17}$$

By substituting the corresponding posterior expectations in Equation (17) into the complete data likelihood in Equation (12), the expected complete data log-likelihood function is provided in the following proposition.

Proposition 2 *For any fixed* $\boldsymbol{\tau} = (\tau_1, \dots, \tau_p)$, G *mixture components and* M *hidden states, the expected complete data log-likelihood function is proportional to*:

$$\mathcal{O}(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) \propto \sum_{i=1}^{N} \left\{ \sum_{g=1}^{G} \hat{w}_{ig} \log \pi_g + \sum_{j=1}^{M} \hat{u}_{i1j} \log q_j + \sum_{t=2}^{T_i} \sum_{j=1}^{M} \sum_{k=1}^{M} \hat{v}_{itjk} \log q_{jk} \right.$$

$$- \frac{1}{2} T_i \log |\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}| + \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} \hat{z}_{itjg} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})' \mathbf{D}^{-1} \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\xi}}$$

$$- \frac{1}{2} \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} \hat{z}_{itjg} \hat{\tilde{z}}_{itjg} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})' (\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})^{-1} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})$$

$$\left. - \frac{1}{2} \tilde{\boldsymbol{\xi}}' \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\xi}} \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} \hat{z}_{itjg} \hat{\tilde{c}}_{itjg} \right\}, \tag{18}$$

where

$$\hat{\tilde{c}}_{itjg} = \left(\frac{\tilde{m}_{itjg}}{2+\tilde{d}} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\tilde{d})\tilde{m}_{itjg}} \right)}{K_{\nu}\left(\sqrt{(2+\tilde{d})\tilde{m}_{itjg}} \right)}, \quad \hat{\tilde{z}}_{itjg} = \left(\frac{2+\tilde{d}}{\tilde{m}_{itjg}} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\tilde{d})\tilde{m}_{itjg}} \right)}{K_{\nu}\left(\sqrt{(2+\tilde{d})\tilde{m}_{itjg}} \right)} - \frac{2\nu}{\tilde{m}_{itjg}}, \tag{19}$$

with

$$\tilde{m}_{itjg} = (\mathbf{y}_{it} - \boldsymbol{\mu}_{it})' (\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})^{-1} (\mathbf{y}_{it} - \boldsymbol{\mu}_{it}), \quad \tilde{d} = \tilde{\boldsymbol{\xi}}' \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\xi}}. \tag{20}$$

Therefore, the EM algorithm can be implemented as follows:

E-step: At rth iteration of the algorithm, let $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)}$ denote the current parameter estimates. Then, conditionally on the observed data and $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)}$, calculate the conditional expectations in Equations (17) and (19). We denote such quantities $\hat{w}_{ig}^{(r)}, \hat{z}_{itjg}^{(r)}, \hat{u}_{itj}^{(r)}, \hat{v}_{itjk}^{(r)}$, and $\hat{\tilde{c}}_{itjg}^{(r)}, \hat{\tilde{z}}_{itjg}^{(r)}$.

M-step: Use $\hat{w}_{ig}^{(r)}, \hat{z}_{itjg}^{(r)}, \hat{u}_{itj}^{(r)}, \hat{v}_{itjk}^{(r)}$, and $\hat{\tilde{c}}_{itjg}^{(r)}, \hat{\tilde{z}}_{itjg}^{(r)}$ to maximize $\mathcal{O}(\boldsymbol{\Phi}_{\boldsymbol{\tau}} \mid \hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)})$ with respect to $\boldsymbol{\Phi}_{\boldsymbol{\tau}}$, and obtain the update parameter estimates. Based on the introduced modelling assumptions, the maximization can be partitioned into orthogonal subproblems, that is, the maximization with respect to the fixed, hidden Markov chain and discrete mixing distribution parameters can be performed separately. The initial probabilities q_j, transition probabilities q_{jk} and mixing proportions π_g are estimated by:

$$\hat{q}_j^{(r)} = \frac{\sum_{i=1}^{N} \hat{u}_{i1j}^{(r)}}{N}, \quad \hat{q}_{jk}^{(r)} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{v}_{itjk}^{(r)}}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{k=1}^{M} \hat{v}_{itjk}^{(r)}}, \quad \hat{\pi}_g^{(r)} = \frac{\sum_{i=1}^{N} \hat{w}_{ig}^{(r)}}{N}. \tag{21}$$

If we let $S_{it} = j$, this implies that $\boldsymbol{\alpha}_{s_{it}} = \boldsymbol{\alpha}_j$ and the M-step updates of model parameters $\boldsymbol{\beta}$, $\mathbf{b}_g$, $\boldsymbol{\alpha}_j$, $\boldsymbol{\Sigma}$ and $\mathbf{D}$, are given in the following proposition.

Proposition 3 *The values of $\boldsymbol{\beta}$, $\mathbf{b}_g$, $\boldsymbol{\alpha}_j$, $\boldsymbol{\Sigma}$ and $\mathbf{D}$ maximizing Equation (18) are:*

$$\hat{\boldsymbol{\beta}}^{(r)} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \sum_{j=1}^{M} \hat{z}_{itjg}^{(r)} \hat{\bar{z}}_{itjg}^{(r)} \mathbf{X}'_{it} \mathbf{X}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \sum_{j=1}^{M} \hat{z}_{itjg}^{(r)} \hat{\bar{z}}_{itjg}^{(r)} \mathbf{X}'_{it} \tilde{\mathbf{Y}}_{it} \right.$$
$$\left. - \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \sum_{j=1}^{M} z_{itjg}^{(r)} \mathbf{X}'_{it} \tilde{\boldsymbol{\xi}}' \mathbf{D}^{(r-1)} \right), \tag{22}$$

where $\tilde{\mathbf{Y}}_{it} = \mathbf{Y}_{it} - \mathbf{Z}_{it} \hat{\mathbf{b}}_g^{(r-1)} - \mathbf{W}_{it} \hat{\boldsymbol{\alpha}}_j^{(r-1)}$.

$$\hat{\mathbf{b}}_g^{(r)} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{M} \hat{z}_{itjg}^{(r)} \hat{\bar{z}}_{itjg}^{(r)} \mathbf{Z}'_{it} \mathbf{Z}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{M} \hat{z}_{itjg}^{(r)} \hat{\bar{z}}_{itjg}^{(r)} \mathbf{Z}'_{it} \tilde{\mathbf{Y}}_{it} - \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{M} z_{itjg}^{(r)} \mathbf{Z}'_{it} \tilde{\boldsymbol{\xi}}' \mathbf{D}^{(r-1)} \right), \tag{23}$$

where $\tilde{\mathbf{Y}}_{it} = \mathbf{Y}_{it} - \mathbf{X}_{it} \hat{\boldsymbol{\beta}}^{(r)} - \mathbf{W}_{it} \hat{\boldsymbol{\alpha}}_j^{(r-1)}$.

$$\hat{\boldsymbol{\alpha}}_j^{(r)} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \hat{z}_{itjg}^{(r)} \hat{\bar{z}}_{itjg}^{(r)} \mathbf{W}'_{it} \mathbf{W}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \hat{z}_{itjg}^{(r)} \hat{\bar{z}}_{itjg}^{(r)} \mathbf{W}'_{it} \tilde{\mathbf{Y}}_{it} - \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} z_{itjg}^{(r)} \mathbf{W}'_{it} \tilde{\boldsymbol{\xi}}' \mathbf{D}^{(r-1)} \right), \tag{24}$$

where $\tilde{\mathbf{Y}}_{it} = \mathbf{Y}_{it} - \mathbf{X}_{it} \hat{\boldsymbol{\beta}}^{(r)} - \mathbf{Z}_{it} \hat{\mathbf{b}}_g^{(r)}$.

$$\hat{\boldsymbol{\Sigma}}^{(r)} = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \sum_{j=1}^{M} \hat{z}_{itjg}^{(r)} \hat{\bar{z}}_{itjg}^{(r)} \hat{\mathbf{D}}^{-1(r-1)} (\mathbf{Y}_{it} - \hat{\boldsymbol{\mu}}_{it}^{(r)})' (\mathbf{Y}_{it} - \hat{\boldsymbol{\mu}}_{it}^{(r)}) \hat{\mathbf{D}}^{-1(r-1)}$$

$$+ \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \sum_{j=1}^{M} \hat{z}_{itjg}^{(r)} \hat{\bar{c}}_{itjg}^{(r)} \tilde{\boldsymbol{\xi}} \tilde{\boldsymbol{\xi}}' - \frac{2}{\sum_{i=1}^{N} T_i} \hat{\mathbf{D}}^{-1(r-1)} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \sum_{j=1}^{M} \hat{z}_{itjg}^{(r)} (\mathbf{Y}_{it} - \hat{\boldsymbol{\mu}}_{it}^{(r)})' \tilde{\boldsymbol{\xi}}', \tag{25}$$

where $\hat{\boldsymbol{\mu}}_{it}^{(r)} = \mathbf{X}_{it} \hat{\boldsymbol{\beta}}^{(r)} + \mathbf{Z}_{it} \hat{\mathbf{b}}_g^{(r)} + \mathbf{W}_{it} \hat{\boldsymbol{\alpha}}_j^{(r)}$.
Finally, the elements d_j, $j = 1, \dots, p$ of the diagonal scale matrix $\mathbf{D}$ are estimated by:

$$\hat{d}_j^{(r)} = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \sum_{k=1}^{M} \hat{z}_{itkg}^{(r)} \rho_\tau (Y_{it}^{(j)} - \hat{\mu}_{it}^{(j)(r)}), \tag{26}$$

where $\rho_\tau(\cdot)$ is the quantile check function of Koenker and Bassett (1978):

$$\rho_\tau(u) = u(\tau - \mathbf{1}(u < 0)) \tag{27}$$

and $\hat{\mu}_{it}^{(j)(r)}$ is the jth element of the vector $\hat{\boldsymbol{\mu}}_{it}^{(r)}$.

The E- and M-steps are alternated until convergence, that is when $|\hat{\boldsymbol{\Phi}}_{\tau}^{(r)} - \hat{\boldsymbol{\Phi}}_{\tau}^{(r-1)}|$ is smaller than a predetermined threshold. In this paper, we set this convergence criterion equal to 10^{-6}.

Because both the number of components of the finite mixture and hidden states of the Markov chain are unknown a-priori, we select the optimal value of G and M using the BIC (Schwarz, 1978):

$$BIC_{(G,M)} = -2\ell(\hat{\boldsymbol{\Phi}}_{\tau}) + \log(N)\nu_f, \tag{28}$$

where $\ell(\hat{\boldsymbol{\Phi}}_{\tau})$ is the observed data log-likelihood in Equation (10) in correspondence of the ML estimate of $\boldsymbol{\Phi}_{\tau}$, N is the number of observed individuals and ν_f denotes the number of free model parameters in $\boldsymbol{\Phi}_{\tau}$. Following Marino et al. (2018), to avoid convergence to local maxima and better explore the parameter space, for fixed τ, G and M, we fit the QMHMM using a multiple random starts strategy with 50 different starting points and retain the solution corresponding to the maximum likelihood value. We then repeat this procedure for a grid of values of G and M, and select the best combination of the pair (G, M) corresponding to the lowest BIC value. The validity of the proposed EM algorithm and model selection procedure have been assessed using also a simulation exercise (see Appendix B).

Standard errors of model parameters are computed using a non-parametric block bootstrap. That is, by re-sampling individuals with replacement and retaining the corresponding sequence of measurements to preserve the within individual dependence structure (see Geraci & Bottai, 2014; Marino & Farcomeni, 2015; Marino et al., 2018 for example). We refit the model to H bootstrap samples and approximate the standard error of each model parameter with the square root of the variance of the matrix:

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\Phi}}_{\tau}) = \sqrt{\frac{1}{H-1}\sum_{h=1}^{H}(\hat{\boldsymbol{\Phi}}_{\tau}^{(h)} - \overline{\boldsymbol{\Phi}}_{\tau})(\hat{\boldsymbol{\Phi}}_{\tau}^{(h)} - \overline{\boldsymbol{\Phi}}_{\tau})'}, \tag{29}$$

where $\hat{\boldsymbol{\Phi}}_{\tau}^{(h)}$ is the set of parameter estimates for the hth bootstrap sample and $\overline{\boldsymbol{\Phi}}_{\tau}$ denotes the mean of the model parameters over bootstrap iterations. The standard errors are given by the diagonal elements of $\widehat{\text{Cov}}(\hat{\boldsymbol{\Phi}}_{\tau})$.

5 | ANALYSIS OF THE MILLENNIUM COHORT STUDY DATA

In this section, we analyse internalizing and externalizing data on disorders of children collected in the MCS dataset. We are interested in investigating the impact of environmental, parental and child factors across both the distributions of SDQ scores. In order to account for all the data features described in Section 2, we consider a bivariate QMHMM with time-varying random intercepts and constant random slopes specified for age to jointly model internalizing and externalizing disorders. We fitted the proposed model at quantile levels $\tau = (0.25, 0.25)$, $\tau = (0.50, 0.50)$, $\tau = (0.75, 0.75)$ and $\tau = (0.90, 0.90)$. Considering the 75th and 90th percentiles puts emphasis on children with more severe problems generally associated with higher levels of SDQ scores. We estimated the QMHMM for a varying number of hidden states ($M = 1, \ldots, 8$) and mixture components ($G = 1, \ldots, 8$) employing the multi-start strategy described in Section 4, and then selected the optimal value of the pair (G, M) corresponding to the lowest BIC value. Following Marino

et al. (2018), to enhance model interpretability and produce meaningful results, we retain only those solutions ensuring $\pi_g > 0.05$ for $g = 1, \ldots, G$ and $q_j > 0.05$ for $j = 1, \ldots, M$.

In addition to the proposed model, we compare our methodology with two well-known univariate alternatives for modelling longitudinal data: (i) the linear random effects model (LREM) for the mean with time-constant random intercepts; (ii) the LQMM of Geraci and Bottai (2014) with time-constant random intercepts, at quantile levels τ, 0.25, 0.50, 0.75 and 0.90. Specifically, the two models are estimated on SDQ_{Int} and SDQ_{Ext} scores independently. The reason why we consider the LREM is because it is a popular model for targeting the conditional expectation of the response given the explanatory variables. While it produces efficient results when the normality assumptions hold, the LREM could potentially miss out important information related to other parts of the distribution of the outcome. In this case, the conditional mean may not offer the best summary; by contrast, the LQMM with a random-intercept specification has a correlation structure that is simple to estimate while allowing for modelling the entire conditional distribution of the outcome. However, both models are using a univariate approach, which completely disregards the possible dependence between the SDQ scores, and assume time-constant random intercepts. In contrast, the proposed model allows to: (i) study the magnitude and direction of the dependence structure between the responses at different quantile levels of interest; (ii) give additional insight into the evolution of the SDQ scores over time and the presence of serial heterogeneity between subjects.

5.1 | Results

We start by commenting on the QMHMM results. Table 2 reports point estimates of model parameters for the two outcomes and standard errors (in parentheses) based on $B = 1000$ bootstrap re-samples. Parameter estimates are displayed in boldface when significant at the standard 5% level. As one can see, the model selection procedure described in Section 1 leads to an increasing number of mixture components G equal to 3, 5, 5 and 5, and a decreasing number of hidden states M equal to 5, 4, 3 and 3 at quantile levels (0.25, 0.25), (0.50, 0.50), (0.75, 0.75) and (0.90, 0.90) respectively. The chosen values for G and M confirm the presence of constant and serial latent heterogeneity in the data, and support the exploratory analysis of individual SDQ trajectories in Figure 2. Moreover, they allow us to classify children based on the intensity of mental and behavioural problems.

The second crucial finding is that the coefficient estimates vary with the quantile level τ and the effect of the covariates appears to be more pronounced at the right tail of the distribution of the responses. In particular, increasing adverse life events, socio-economic disadvantage, maternal depression and low maternal education are statistically associated with both SDQ scores and their impact increases when looking at the upper tail compared to the lower tail of the distribution. This indicates that an unstable and fragmented family environment has a greater adverse impact on children facing more problems. Regarding income, there is evidence that poorer children are more likely to suffer from both physical and mental health problems (Currie, 2009), hence the role of family income is likely to be concentrated at low incomes (see Fitzsimons et al., 2017). Moreover, maternal depression has a more pronounced effect at the top end where children display critical levels of adjustment problems than at the bottom end of the distribution (see Kiernan & Huerta, 2008). These considerations suggest that low socioeconomic status creates stress within the household, causing poorer child health. In relation to gender, males present lower internalizing problems at low quantiles compared to females, while the estimated

TABLE 2   Point estimates with standard errors in parentheses for different quantile levels

τth quantile	(0.25, 0.25) [G = 3, M = 5]		(0.50, 0.50) [G = 5, M = 4]		(0.75, 0.75) [G = 5, M = 3]		(0.90, 0.90) [G = 5, M = 3]	
Variable	SDQ_{Int}	SDQ_{Ext}	SDQ_{Int}	SDQ_{Ext}	SDQ_{Int}	SDQ_{Ext}	SDQ_{Int}	SDQ_{Ext}
Age year scal	**−0.053** (0.010)	**−0.289** (0.016)	**−0.051** (0.009)	**−0.440** (0.011)	−0.040 (0.021)	**−0.450** (0.031)	−0.015 (0.017)	**−0.414** (0.029)
Age2 year scal	**0.045** (0.005)	**0.193** (0.009)	**0.077** (0.011)	**0.208** (0.013)	**0.104** (0.018)	**0.287** (0.024)	**0.065** (0.014)	**0.233** (0.021)
ALE 11	**0.022** (0.008)	**0.036** (0.016)	**0.086** (0.018)	**0.113** (0.019)	**0.116** (0.039)	**0.205** (0.055)	**0.202** (0.031)	**0.284** (0.048)
SED 4	**0.070** (0.023)	**0.105** (0.045)	**0.175** (0.030)	**0.221** (0.030)	**0.201** (0.051)	**0.398** (0.076)	**0.217** (0.033)	**0.347** (0.059)
Kessm	**0.090** (0.009)	**0.143** (0.012)	**0.167** (0.009)	**0.189** (0.012)	**0.208** (0.018)	**0.299** (0.025)	**0.257** (0.012)	**0.379** (0.020)
Degree	**−0.350** (0.109)	**−0.894** (0.149)	**−0.526** (0.114)	**−1.482** (0.171)	**−0.703** (0.160)	**−1.267** (0.232)	**−0.856** (0.135)	**−1.530** (0.271)
GCSE	**−0.217** (0.110)	**−0.582** (0.150)	**−0.352** (0.113)	**−0.430** (0.166)	**−0.413** (0.149)	**−0.427** (0.213)	**−0.524** (0.132)	**−0.671** (0.258)
White	−0.090 (0.061)	−0.143 (0.097)	−0.075 (0.149)	0.059 (0.160)	−0.216 (0.203)	0.320 (0.303)	−0.218 (0.127)	0.125 (0.259)
Male	**−0.062** (0.016)	**0.793** (0.030)	0.027 (0.037)	**0.950** (0.045)	0.082 (0.080)	**0.944** (0.119)	**0.258** (0.053)	**1.308** (0.094)
IMD	**−0.022** (0.005)	**−0.036** (0.009)	**−0.025** (0.008)	**−0.027** (0.010)	−0.027 (0.020)	−0.045 (0.029)	**−0.054** (0.012)	**−0.079** (0.025)
Eth stratum	−0.043 (0.159)	−0.168 (0.174)	0.122 (0.193)	0.144 (0.220)	0.174 (0.241)	−0.025 (0.374)	−0.063 (0.136)	−0.345 (0.298)
Dis stratum	−0.003 (0.031)	0.003 (0.072)	**0.085** (0.042)	**0.106** (0.050)	0.156 (0.111)	**0.337** (0.171)	0.005 (0.064)	**0.203** (0.103)
α_1	**0.437** (0.001)	**1.040** (0.114)	**0.983** (0.013)	**4.769** (0.049)	**2.082** (0.016)	**2.954** (0.054)	**3.311** (0.005)	**5.078** (0.006)
α_2	**3.930** (0.051)	**4.140** (0.089)	**2.480** (0.007)	**8.329** (0.077)	**2.735** (0.119)	**7.769** (0.014)	**9.104** (0.069)	**9.919** (0.103)
α_3	**1.008** (0.005)	**6.362** (0.000)	**1.164** (0.016)	**1.954** (0.044)	**6.957** (0.091)	**6.347** (0.201)	**3.848** (0.056)	**10.541** (0.212)
α_4	**1.523** (0.130)	**1.281** (0.099)	**4.829** (0.008)	**4.670** (0.128)				
α_5	**0.477** (0.043)	**3.115** (0.013)						
r_{12}	**0.356** (0.016)		**0.211** (0.023)		**0.602** (0.021)		**0.738** (0.013)	
$\ell(\mathbf{\Phi}_\tau)$	−71,665.9		−73,173.9		−77,223.6		−80,517.3	
ν_f	69		64		55		55	
AIC	14,3470.0		14,6794.6		15,4557.2		16,1144.5	
BIC	14,3924.2		14,6914.3		15,4919.3		16,1506.6	

Note: Parameter estimates are displayed in boldface when significant at the standard 5% level.

effect is statistically significant and positive at $\tau = (0.9, 0.9)$. On the other hand, ethnicity and ethnic stratification variables do not appear to be associated with the responses at the 25th percentiles.

By looking at the fixed parameter estimates for the SDQ_{Ext}, we observe that, in contrast with internalizing scores, boys present more externalizing problems than girls (Flouri & Sarmadi, 2016) and the effect is more exacerbated in the right tail of the distribution. In general, girls are at lower risk of behavioural problems than boys which experience an increased risk for conduct and hyperactivity problems (Carona et al., 2014). Stressful life events, socio-economic disadvantage, maternal depression and maternal education are all significantly associated with internalizing scores. The effect of the covariates is not uniform across quantiles but it is more apparent as the quantile level increases. Moreover, the impact of such variables is more pronounced, across the distribution, on externalizing than on internalizing scores. This is consistent with other studies on behaviour disorders in child psychopathology claiming that poverty and material deprivation and education are more strongly related with children's externalizing problems compared with internalizing problems (see Costello et al., 2003; Dearing et al., 2006). Overall, point estimates of regression coefficients are consistent with child development theory, as well as with the results discussed in Tzavidis et al. (2016) and Alfò et al. (2021).

In order to highlight the practical relevance of the proposed methodology, we compare our findings with the parameter estimates of the univariate LREM and LQMM reported in Table 3. At first, we observe that the LQMM results are generally in line with our findings, except for $\tau = 0.25$. We notice that the estimated parameters for the SDQ_{Ext} are greater than those for the SDQ_{Int} scores and the effect of the considered covariates increases moving towards the right tail of the conditional distribution of each outcome, especially at the 90th percentile. Consistently with the QMHMM, mother's education and family environmental risk factors are important for predicting emotional and behavioural disorders in early childhood. Furthermore, males typically present higher SDQ_{Int} and SDQ_{Ext} problems than females whereas, contrary to the results obtained from the QMHMM, lower SDQ_{Int} scores are generally associated with white children. Point estimates of the LREM and the LQMM at the median are not similar due to the asymmetry in SDQ distributions. This is consistent with the graphical analysis in Figure 1, and highlights the importance of considering a quantile regression approach to assess the heterogeneous impact of risk factors across the distribution of children's psychopathologies. To further show that modelling the conditional mean is an unreasonable approach, Figure 3 presents normal probability plots of level 1 and level 2 residuals of the fitted LREM models. These reveal the presence of potentially influential observations in the data, indicate severe departures from the Gaussian assumption of the random-intercepts model for both SDQ outcomes and show that residuals are skewed.

In addition to that, both the LQMM and LREM analyse children's disorders by fitting two univariate models separately and hence, they disregard the possible association between the SDQ scores. In contrast, one of the main benefits of the proposed multivariate approach is the possibility to study the magnitude and direction of the dependence structure between the responses at different quantile levels of interest. Following Kotz et al. (2012), we can compute the correlation between SDQ scores using Equation (6) and understand whether their association structure becomes stronger for children with more pronounced problems. In particular, the estimated correlation coefficient, r_{12}, reported in Table 2 gives a measure of tail correlation and, consistently with the recent work of Alfò et al. (2021), it indicates that internalizing and externalizing disorders are positively associated and this association increases with the quantile level τ. From the median to the 90th percentile, the estimated coefficient rapidly increases from 0.211 to 0.738,

TABLE 3 Univariate linear random effects model (LREM) for the mean and linear quantile mixed model (LQMM) point estimates for internalizing and externalizing scores at the investigated quantile levels

| τth quantile	(0.25, 0.25) SDQ_{Int}	SDQ_{Ext}	(0.50, 0.50) SDQ_{Int}	SDQ_{Ext}	LREM- Mean SDQ_{Int}	SDQ_{Ext}	(0.75, 0.75) SDQ_{Int}	SDQ_{Ext}	(0.90, 0.90) SDQ_{Int}	SDQ_{Ext}
Intercept	**1.645** (0.171)	**3.298** (0.233)	**2.313** (0.154)	**3.942** (0.241)	**2.575** (0.149)	**4.314** (0.221)	**2.900** (0.174)	**4.438** (0.234)	**3.123** (0.175)	**4.954** (0.248)
Age year scal	−0.021 (0.017)	**−0.391** (0.012)	**−0.037** (0.008)	**−0.472** (0.011)	−0.002 (0.009)	**−0.450** (0.011)	**−0.036** (0.012)	**−0.488** (0.012)	−0.016 (0.020)	**−0.525** (0.025)
Age2 year scal	0.020 (0.015)	**0.175** (0.009)	**0.053** (0.006)	**0.238** (0.006)	**0.083** (0.007)	**0.252** (0.009)	**0.103** (0.008)	**0.298** (0.009)	**0.111** (0.013)	**0.312** (0.015)
ALE 11	0.026 (0.014)	**0.120** (0.026)	**0.055** (0.015)	**0.072** (0.030)	**0.095** (0.016)	**0.114** (0.021)	**0.105** (0.021)	**0.113** (0.029)	**0.106** (0.049)	**0.185** (0.040)
SED 4	**0.136** (0.068)	**0.110** (0.042)	**0.181** (0.036)	**0.185** (0.054)	**0.138** (0.024)	**0.200** (0.032)	**0.231** (0.048)	**0.259** (0.038)	**0.204** (0.048)	**0.285** (0.093)
Kessm	**0.122** (0.057)	**0.147** (0.014)	**0.146** (0.008)	**0.143** (0.026)	**0.161** (0.007)	**0.195** (0.009)	**0.187** (0.012)	**0.215** (0.016)	**0.230** (0.023)	**0.214** (0.028)
Degree	**−0.673** (0.178)	**−1.407** (0.171)	**−0.564** (0.111)	**−1.383** (0.162)	**−0.687** (0.099)	**−1.583** (0.148)	**−0.483** (0.131)	**−1.246** (0.160)	**−0.390** (0.136)	**−1.182** (0.184)
GCSE	**−0.597** (0.282)	**−0.640** (0.163)	**−0.392** (0.112)	**−0.391** (0.170)	**−0.372** (0.093)	**−0.680** (0.141)	−0.117 (0.118)	**−0.317** (0.158)	−0.043 (0.119)	−0.210 (0.187)
White	**−0.564** (0.165)	0.009 (0.145)	**−0.275** (0.110)	0.255 (0.157)	**−0.352** (0.091)	0.121 (0.138)	−0.132 (0.111)	**0.445** (0.146)	−0.015 (0.126)	**0.713** (0.181)
Male	0.012 (0.023)	**0.744** (0.093)	0.061 (0.039)	**0.958** (0.073)	**0.127** (0.049)	**0.938** (0.075)	**0.232** (0.083)	**1.082** (0.087)	**0.319** (0.103)	**1.164** (0.118)
IMD	−0.012 (0.018)	−0.027 (0.021)	**−0.033** (0.011)	−0.009 (0.021)	**−0.044** (0.010)	**−0.048** (0.014)	**−0.048** (0.018)	−0.006 (0.022)	−0.026 (0.022)	−0.054 (0.036)
Eth stratum	−0.013 (0.105)	−0.189 (0.196)	0.153 (0.119)	0.089 (0.191)	0.169 (0.112)	−0.016 (0.168)	**0.373** (0.139)	0.216 (0.203)	**0.541** (0.141)	0.281 (0.224)
Dis stratum	−0.014 (0.037)	0.058 (0.100)	0.039 (0.056)	0.095 (0.094)	0.070 (0.061)	**0.226** (0.092)	**0.227** (0.076)	**0.379** (0.093)	**0.272** (0.093)	**0.535** (0.108)
σ^2	1.136	4.578	1.986	5.708	2.215	5.907	3.527	7.062	4.432	8.040
$\ell(\mathbf{\Phi}_\tau)$	−32,991.1	−39,048.1	−34,088.9	−39,125.8	−34,799.8	−39,229.7	−36,086.5	−40,174.2	−37,454.3	−41,192.0
AIC	66,012.1	78,126.3	68,207.7	78,281.6	69,629.6	78,489.9	72,203.1	80,378.4	74,938.7	82,414.0
BIC	66,127.4	78,225.0	68,306.5	78,380.3	69,728.4	78,588.2	72,301.8	80,477.2	75,037.4	82,512.7

Notes: Standard errors are in parentheses. Parameter estimates are displayed in boldface when significant at the standard 5% level.
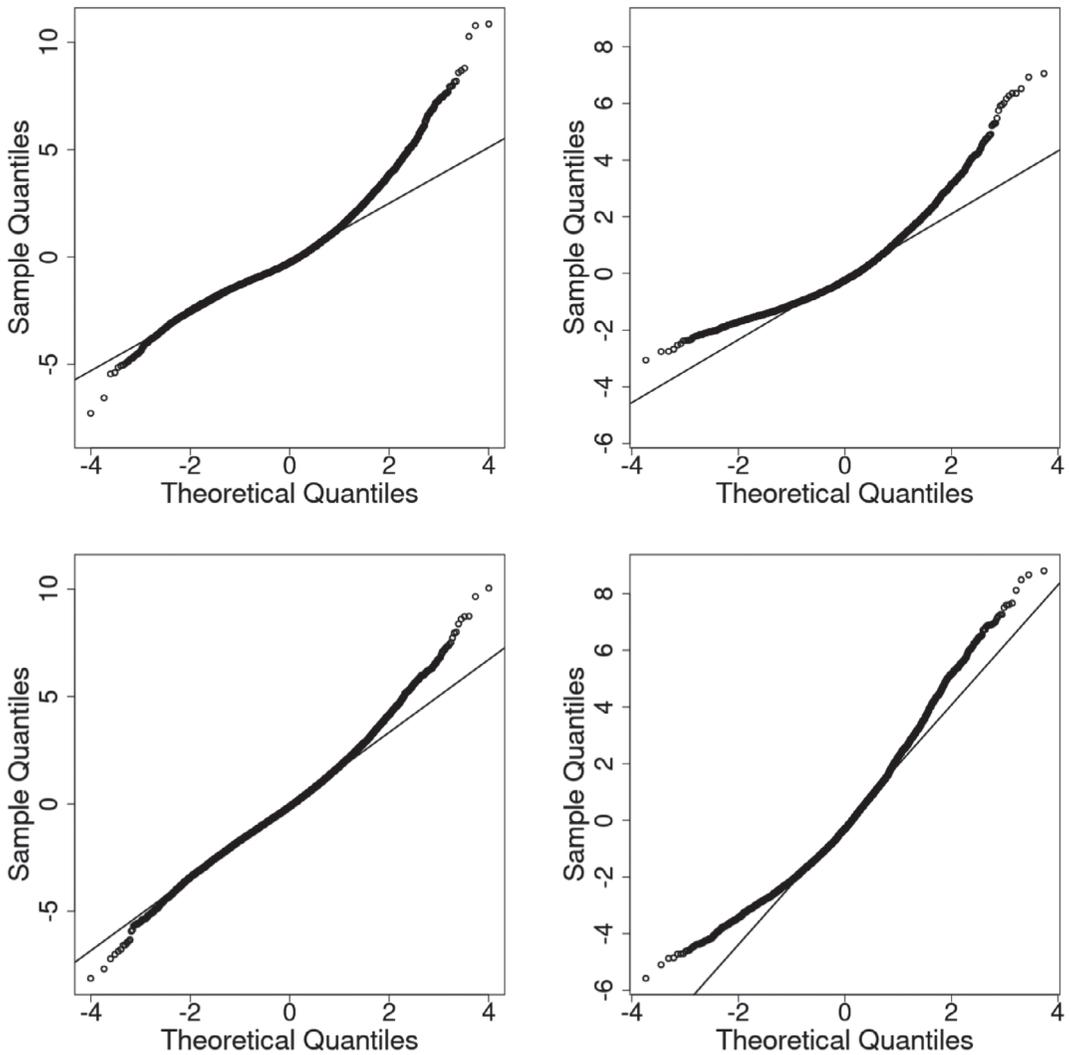
FIGURE 3 Normal probability plots of level 1 (first column) and level 2 (second column) residuals from the linear random effects model for Strengths and Difficulties Questionnaire internalizing (first row) and externalizing (second row) problems

which suggests that children with high levels of internalizing SDQ scores are more likely to experience or develop externalizing problems, and vice versa. Hence, children may present a constellation of symptoms comprised of both disorders which is aggravated in disadvantaged ones by the accumulation of risk factors. This finding is also in line with the existence of positive covariation among psychiatric diagnoses (see Cicchetti & Toth, 2014; Lilienfeld, 2003; Liu, 2004).

To further justify our empirical strategy, we conclude the analysis by reporting selected diagnostics for the fitted QMHMM. First, Figure 4 shows the estimated marginal cumulative density functions of the discrete random slopes for both SDQ outcomes. In both plots, it is clear that the estimated distribution functions depart substantially from the Gaussian distribution, having pronounced asymmetries. Hence, the underlying assumption of normally
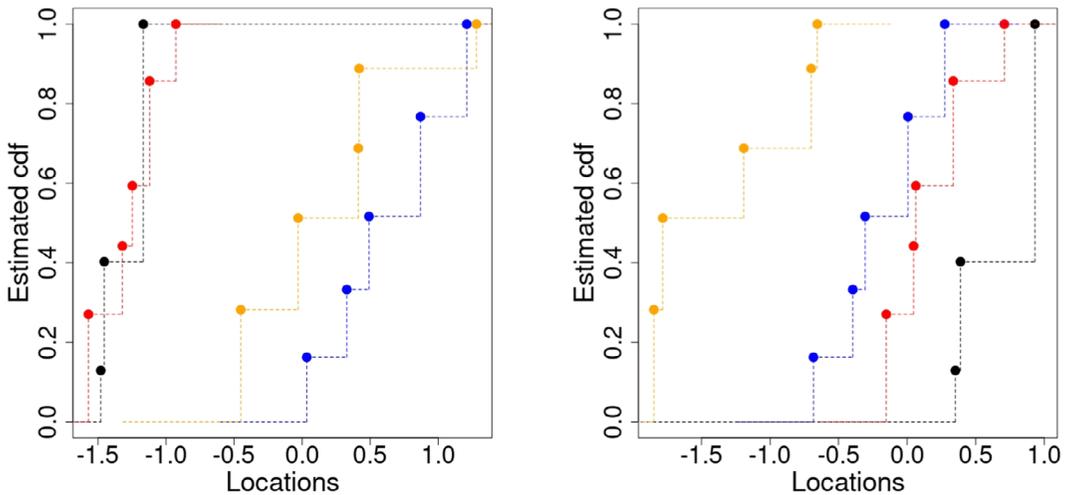
FIGURE 4 Estimated cumulative density function of the discrete random slopes for Strengths and Difficulties Questionnaire internalizing (left) and externalizing (right) problems scores, at the 0.25 (black), 0.50 (red), 0.75 (blue) and 0.90 (orange) quantile levels [Colour figure can be viewed at wileyonlinelibrary.com]

distributed random intercepts in the LQMM and LRE models is inappropriate. In contrast, the considered discrete mixture approach for the random effects distribution is more flexible and is able to accommodate possible departures from the Gaussianity assumption (Alfò et al., 2017).

Second, Tables 2 and 4 summarize the estimated random intercepts α, initial and transition probabilities of the hidden Markov chain. Inference about the hidden Markov process gives additional insight into the evolution of the SDQ scores over time and the serial heterogeneity between subjects. In this application, the states are not only a tool for modelling time dependence but also have a practical meaning. The transition matrices describe how, and how frequently, children move from low to high level of disorders and the random intercepts α correspond to different severities of disorders in children. At first, it is worth noting that the estimated state-dependent intercepts α, tend to increase when moving from lower to upper quantiles resulting in higher levels of children's disorders. For $\tau = (0.25, 0.25)$ (Panel A), the initial probability distribution defined on $\mathcal{S}$ is relatively uniformly distributed and the probability of not moving from states 1 and 4 is also very high, that is, $\hat{q}_{11} = 0.991$ and $\hat{q}_{44} = 0.993$ respectively. This implies that almost every child in the lower tail of the outcomes distributions starts and maintains low-level disorders over time. If any transition is observed, units tend to move towards states 1 and 4 with lower intercepts and a reduction in juvenile developmental disorders with temporary jumps to moderate values of disorders. For $\tau = (0.50, 0.50)$ (Panel B), by looking at the initial probabilities one can see that half of the units ($\hat{q}_3 = 0.415$) start the study with low values of emotional and behavioural disorders and transitions between states are unlikely. When $\tau = (0.75, 0.75)$ (Panel C), the majority of children start with moderate values of developmental difficulties ($\hat{q}_1 + \hat{q}_2 > 0.80$) and transitions to more severe disorders are more frequent. Finally, by looking at the right tail of the SDQ score distributions, $\tau = (0.90, 0.90)$ (Panel D), we conclude that, among those children with high problems, states 2 and 3 are associated with even higher incidence of disorders. Around 60% of children start the study in state 1 and 91% remain within the same class, with temporary changes towards more severe level of emotional and behavioural disorders.

T A B L E 4 Initial probabilities, **q**, and transition probabilities, **Q**, estimates for different quantiles

	1	2	3	4	5
Panel A: $\tau = (0.25, 0.25)$					
q	0.201 (0.017)	0.111 (0.013)	0.269 (0.017)	0.165 (0.017)	0.254 (0.020)
1	0.991 (0.016)	0.000 (0.000)	0.000 (0.000)	0.009 (0.004)	0.000 (0.029)
2	0.000 (0.002)	0.883 (0.044)	0.007 (0.019)	0.110 (0.008)	0.000 (0.012)
3	0.000 (0.000)	0.089 (0.012)	0.723 (0.032)	0.000 (0.000)	0.187 (0.014)
4	0.003 (0.015)	0.004 (0.042)	0.000 (0.001)	0.993 (0.009)	0.000 (0.013)
5	0.096 (0.003)	0.016 (0.009)	0.027 (0.029)	0.029 (0.000)	0.832 (0.034)
Panel B: $\tau = (0.50, 0.50)$					
q	0.287 (0.019)	0.160 (0.016)	0.415 (0.018)	0.138 (0.014)	
1	0.863 (0.036)	0.034 (0.034)	0.071 (0.007)	0.032 (0.024)	
2	0.053 (0.017)	0.850 (0.038)	0.000 (0.000)	0.096 (0.013)	
3	0.006 (0.031)	0.000 (0.000)	0.967 (0.013)	0.027 (0.034)	
4	0.049 (0.015)	0.015 (0.024)	0.067 (0.011)	0.868 (0.036)	
Panel C: $\tau = (0.75, 0.75)$					
q	0.541 (0.021)	0.293 (0.017)	0.165 (0.015)		
1	0.942 (0.011)	0.006 (0.020)	0.052 (0.029)		
2	0.072 (0.004)	0.825 (0.024)	0.103 (0.020)		
3	0.185 (0.010)	0.109 (0.015)	0.707 (0.031)		
Panel D: $\tau = (0.90, 0.90)$					
q	0.606 (0.015)	0.152 (0.011)	0.242 (0.014)		
1	0.913 (0.008)	0.067 (0.025)	0.019 (0.019)		
2	0.292 (0.007)	0.605 (0.026)	0.103 (0.012)		
3	0.147 (0.004)	0.099 (0.015)	0.754 (0.022)		

6 | CONCLUSIONS

Longitudinal data allow us to understand the evolution of a certain phenomenon over time. In this context, it becomes of crucial importance to determine an appropriate modelling framework to assess the effects of unobserved factors and hidden heterogeneity which can be either time invariant or time varying; ignoring these factors may induce bias and lead to invalid conclusions. Moreover, the literature on this topic which is traditionally focused on the conditional mean, might not provide a good summary of the response distribution. To account for the complex data structure, this work generalizes the multivariate quantile approach of Petrella and Raponi (2019) for the analysis of multivariate longitudinal data by combining the features of quantile regression and MHMMs (Altman, 2007). The proposed model allows for the quantile-specific effects to be quantified and also permits joint modelling of several outcomes. The model further allows for different sources of heterogeneity to be distinguished, that is, between individual heterogeneity and time heterogeneity are modelled through the state-specific effects. In order to avoid possibly

misleading inferences caused by erroneous assumption on the random effects distribution, we rely on the NPML estimation theory and we approximate this distribution by a multivariate discrete latent variable.

As illustrated in the real data application, the proposed method models simultaneously the quantiles of children's emotional and behavioural disorders as a function of demographic and socio-economics risk factors. The results show that behavioural and emotional difficulties are mainly affected by the family poverty conditions and mother's characteristics. Such effects are much stronger in the upper tail of the response distribution, that is, for those children experiencing more severe internalizing and externalizing problems. In addition, the analysis reveals moderate levels of codependency between internalizing and externalizing disorders, that cannot be detected by univariate models.

The methodology can be further extended to allow for a non-homogeneous hidden Markov process where transition probabilities are allowed to depend on covariates. Finally, the hidden Markov chain implicitly assumes that the sojourn time, that is, the number of consecutive time points that the process spends in a given state, is geometrically distributed. As a further generalization of this work one may consider a semi-Markov process which is designed to relax this condition by allowing the sojourn time to be modelled by more flexible distributions.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://urldefense.com/v3/__http://www.cls.ioe.ac.uk__;!!N11eV2iwtfs!s7vhACgBsnIamFuRrItB5EicSzWAyp9h86KEo7gs7tiIK2yVDpumqSvzQYIkanc6xQKVJYRDUafRBSWDLQ$.

ORCID

Luca Merlo https://orcid.org/0000-0001-5267-5390

REFERENCES

Agresti, A., Caffo, B. & Ohman-Strickland, P. (2004) Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis*, 47(3), 639–653.

Ahn, J.V., Sera, F., Cummins, S. & Flouri, E. (2018) Associations between objectively measured physical activity and later mental health outcomes in children: findings from the UK Millennium Cohort Study. *Journal of Epidemiology and Community Health*, 72(2), 94–100.

Aitkin, M. & Alfó, M. (1998) Regression models for binary longitudinal responses. *Statistics and Computing*, 8(4), 289–307.

Aitkin, M. & Alfò, M. (2003) Longitudinal analysis of repeated binary data using autoregressive and random effect modelling. *Statistical Modelling*, 3(4), 291–303.

Akaike, H. (1998) Information theory and an extension of the maximum likelihood principle. In: Parzen, E., Tanabe, K., Kitagawa, G. Eds. *Selected papers of Hirotugu Akaike*. Springer Series in Statistics. New York, NY: Springer, pp. 199–213.

Alfò, M. & Aitkin, M. (2000) Random coefficient models for binary longitudinal responses with attrition. *Statistics and Computing*, 10(4), 279–287.

Alfò, M. & Maruotti, A. (2010) Two-part regression models for longitudinal zero-inflated count data. *Canadian Journal of Statistics*, 38(2), 197–216.

Alfò, M., Salvati, N. & Ranallli, M.G. (2017) Finite mixtures of quantile and M-quantile regression models. *Statistics and Computing*, 27(2), 547–570.

Alfò, M., Marino, M.F., Ranalli, M.G., Salvati, N. & Tzavidis, N. (2021) M-quantile regression for multivariate longitudinal data with an application to the Millennium Cohort Study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(1), 122–146.

Altman, R.M. (2007) 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.

Bartolucci, F. & Farcomeni, A. (2009) A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, 104(486), 816–831.

Becker, A., Hagenberg, N., Roessner, V., Woerner, W. & Rothenberger, A. (2004) Evaluation of the self-reported SDQ in a clinical setting: do self-reports tell us more than ratings by adult informants? *European Child & Adolescent Psychiatry*, 13(2), ii17–ii24.

Bell, S.L., Audrey, S., Gunnell, D., Cooper, A. & Campbell, R. (2019) The relationship between physical activity, mental wellbeing and symptoms of mental health disorder in adolescents: a cohort study. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1), 138.

Böhning, D. (1999) *Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others*, vol. 81. Hertfordshire: CRC Press.

Bradley, R.H. & Corwyn, R.F. (2002) Socioeconomic status and child development. *Annual Review of Psychology*, 53(1), 371–399.

Breckling, J. & Chambers, R. (1988) M-quantiles. *Biometrika*, 75(4), 761–771.

Carona, C., Moreira, H., Silva, N., Crespo, C. & Canavarro, M.C. (2014) Social support and adaptation outcomes in children and adolescents with cerebral palsy. *Disability and Rehabilitation*, 36(7), 584–592.

Chakraborty, B. (2003) On multivariate quantile regression. *Journal of Statistical Planning and Inference*, 110(1–2), 109–132.

Charlier, I., Paindaveine, D. & Saracco, J. (2020) Multiple-output quantile regression through optimal quantization. *Scandinavian Journal of Statistics*, 47(1), 250–278.

Chavas, J.-P. (2018) On multivariate quantile regression analysis. *Statistical Methods & Applications*, 27(3), 365–384.

Cicchetti, D. & Toth, S.L. (2014) A developmental perspective on internalizing and externalizing disorders. *Internalizing and Externalizing Expression of Dysfunction*, 2, 1–19.

Connelly, R. & Platt, L. (2014) Cohort profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology*, 43(6), 1719–1725.

Costello, E.J., Compton, S.N., Keeler, G. & Angold, A. (2003) Relationships between poverty and psychopathology: a natural experiment. *Journal of the American Medical Association*, 290(15), 2023–2029.

Crowther, M.J., Look, M.P. & Riley, R.D. (2014) Multilevel mixed effects parametric survival models using adaptive Gauss–Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Statistics in Medicine*, 33(22), 3844–3858.

Currie, J. (2009) Healthy, wealthy, and wise: socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47(1), 87–122.

Dearing, E., McCartney, K. & Taylor, B.A. (2006) Within-child associations between family income and externalizing and internalizing problems. *Developmental Psychology*, 42(2), 237.

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

Diggle, P., Diggle, P.J., Heagerty, P., Liang, K.-Y., Heagerty, P.J., Zeger, S. et al. (2002) *Analysis of longitudinal data*. Oxford: Oxford University Press.

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

Farcomeni, A. & Viviani, S. (2015) Longitudinal quantile regression in the presence of informative dropout through longitudinal-survival joint modeling. *Statistics in Medicine*, 34(7), 1199–1213.

Fitzsimons, E., Goodman, A., Kelly, E. & Smith, J.P. (2017) Poverty dynamics and parental mental health: determinants of childhood mental health in the UK. *Social Science & Medicine*, 175, 43–51.

Flouri, E. & Sarmadi, Z. (2016) Prosocial behavior and childhood trajectories of internalizing and externalizing problems: the role of neighborhood and school contexts. *Developmental Psychology*, 52(2), 253.

Flouri, E., Tzavidis, N. & Kallis, C. (2010) Area and family effects on the psychopathology of the Millennium Cohort Study children and their older siblings. *Journal of Child Psychology and Psychiatry*, 51(2), 152–161.

Geraci, M. & Bottai, M. (2014) Linear quantile mixed models. *Statistics and Computing*, 24(3), 461–479.

Goldstein, H. (2011) *Multilevel statistical models*, vol. 922. John Wiley & Sons.

Goodman, R. (1997) The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586.

Goodman, A. & Goodman, R. (2009) Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4), 400–403.

Goodman, A. & Goodman, R. (2011) Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *Journal of Child Psychology and Psychiatry*, 52(1), 100–108.

Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer, H. (2003) Using the strengths and difficulties questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry*, 15(1–2), 166–172.

Goodnight, J.A., Lahey, B.B., Van Hulle, C.A., Rodgers, J.L., Rathouz, P.J., Waldman, I.D. et al. (2012) A quasi-experimental analysis of the influence of neighborhood disadvantage on child and adolescent conduct problems. *Journal of Abnormal Psychology*, 121(1), 95.

Griffiths, L.J., Dezateux, C. & Hill, A. (2011) Is obesity associated with emotional and behavioural problems in children? Findings from the Millennium Cohort Study. *International Journal of Pediatric Obesity*, 6(supplement 3), e423–432.

Hallin, M., Paindaveine, D., Šiman, M., Wei, Y., Serfling, R., Zuo, Y. et al. (2010) Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *The Annals of Statistics*, 38(2), 635–703.

Joshi, H. & Fitzsimons, E. (2016) The Millennium Cohort Study: the making of a multi-purpose resource for social science and policy. *Longitudinal and Life Course Studies*, 7(4), 409–430.

Kiernan, K.E. & Huerta, M.C. (2008) Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood. *The British Journal of Sociology*, 59(4), 783–806.

Koenker, R. (2005) *Quantile regression*. Cambridge: Cambridge University Press.

Koenker, R. & Bassett, G. (1978) Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46(1), 33–50.

Koenker, R., Chernozhukov, V., He, X. & Peng, L. (2017) *Handbook of quantile regression*. Boca Raton, FL: CRC Press.

Kong, L. & Mizera, I. (2012) Quantile tomography: using quantiles with multivariate data. *Statistica Sinica*, 22, 1589–1610.

Kotz, S., Kozubowski, T. & Podgorski, K. (2012) *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Berlin/Heidelberg: Springer Science & Business Media.

Kulkarni, H., Biswas, J. & Das, K. (2019) A joint quantile regression model for multiple longitudinal outcomes. *AStA Advances in Statistical Analysis*, 103(4), 453–473.

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

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

Lilienfeld, S.O. (2003) Comorbidity between and within childhood externalizing and internalizing disorders: reflections and directions. *Journal of Abnormal Child Psychology*, 31(3), 285–291.

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

Lindsey, J.K. (1999) *Models for repeated measurements*. Oxford: Oxford University Press.

Liu, J. (2004) Childhood externalizing behavior: theory and implications. *Journal of Child and Adolescent Psychiatric Nursing*, 17(3), 93–103.

Liu, Y. & Bottai, M. (2009) Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics*, 5(1), 1–24.

Marino, M.F. & Farcomeni, A. (2015) Linear quantile regression models for longitudinal experiments: an overview. *Metron*, 73(2), 229–247.

Marino, M.F., Tzavidis, N. & Alfò, M. (2018) Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequences. *Statistical Methods in Medical Research*, 27(7), 2231–2246.

Maruotti, A. (2011) Mixed hidden Markov models for longitudinal data: an overview. *International Statistical Review*, 79(3), 427–454.

Mathai, J., Anderson, P. & Bourne, A. (2004) Comparing psychiatric diagnoses generated by the strengths and difficulties questionnaire with diagnoses made by clinicians. *Australian & New Zealand Journal of Psychiatry*, 38(8), 639–643.

McCulloch, C.E. (1997) Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437), 162–170.

McMunn, A., Kelly, Y., Cable, N. & Bartley, M. (2012) Maternal employment and child socio-emotional behaviour in the UK: longitudinal evidence from the UK Millennium Cohort Study. *Journal of Epidemiology and Community Health*, 66(7), e19.

Merlo, L., Petrella, L. & Raponi, V. (2021) Forecasting VaR and ES using a joint quantile regression and its implications in portfolio allocation. *Journal of Banking & Finance*, 133, 106248.

Merlo, L., Maruotti, A. & Petrella, L. (2021) Two-part quantile regression models for semi-continuous longitudinal data: a finite mixture approach. *Statistical Modelling*, 1471082X21993603.

Neuhaus, J.M., McCulloch, C.E. & Boylan, R. (2013) Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercepts and slopes. *Statistics in Medicine*, 32(14), 2419–2429.

Petrella, L. & Raponi, V. (2019) Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. *Journal of Multivariate Analysis*, 173, 70–84.

Pinheiro, J.C. & Chao, E.C. (2006) Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1), 58–81.

Platt, R., Williams, S.R. & Ginsburg, G.S. (2016) Stressful life events and child anxiety: examining parent and child mediators. *Child Psychiatry & Human Development*, 47(1), 23–34.

Plewis, I., Calderwood, L., Hawkes, D., Hughes, G. & Joshi, H. (2007) *Millennium cohort study: technical report on sampling*. London: Centre for Longitudinal Studies.

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2005) Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128(2), 301–323.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.

Stolfi, P., Bernardi, M. & Petrella, L. (2018) The sparse method of simulated quantiles: an application to portfolio optimization. *Statistica Neerlandica*, 72(3), 375–398.

Tiet, Q.Q., Bird, H.R., Davies, M., Hoven, C., Cohen, P., Jensen, P.S. et al. (1998) Adverse life events and resilience. *Journal of the American Academy of Child & Adolescent Psychiatry*, 37(11), 1191–1200.

Trentacosta, C.J., Hyde, L.W., Shaw, D.S., Dishion, T.J., Gardner, F. & Wilson, M. (2008) The relations among cumulative risk, parenting, and behavior problems during early childhood. *Journal of Child Psychology and Psychiatry*, 49(11), 1211–1219.

Tzavidis, N., Salvati, N., Schmid, T., Flouri, E. & Midouhas, E. (2016) Longitudinal analysis of the strengths and difficulties questionnaire scores of the Millennium Cohort Study children in England using M-quantile random-effects regression. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179, 427–452.

Van der Meer, M., Dixon, A. & Rose, D. (2008) Parent and child agreement on reports of problem behaviour obtained from a screening questionnaire, the SDQ. *European Child & Adolescent Psychiatry*, 17(8), 491–497.

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

Wickham, S., Whitehead, M., Taylor-Robinson, D. & Barr, B. (2017) The effect of a transition into poverty on child and maternal mental health: a longitudinal analysis of the UK Millennium Cohort Study. *The Lancet Public Health*, 2(3), e141–e148.

Yu, K. & Zhang, J. (2005) A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics—Theory and Methods*, 34(9–10), 1867–1879.

APPENDIX A

Proof of Proposition 1 Under the constraints imposed on $\tilde{\xi}$ and Λ, the representation in Equation (5) implies that:

$$\mathbf{Y} \mid \tilde{C} = \tilde{c} \sim \mathcal{N}_p(\boldsymbol{\mu} + \mathbf{D}\tilde{\xi}\tilde{c}, \tilde{c}\mathbf{D}\Sigma\mathbf{D}), \quad \tilde{C} \sim \exp(1). \tag{A1}$$

This implies that the joint density function of $\mathbf{Y}$ and $\tilde{C}$ is:

$$f_{\mathbf{Y},\tilde{C}}(\mathbf{y},\tilde{c}) = \frac{\exp\left\{(\mathbf{y}-\boldsymbol{\mu})'\mathbf{D}^{-1}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\xi}}\right\}}{(2\pi)^{p/2}|\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}|^{1/2}} \left(\tilde{c}^{-p/2}\exp\left\{-\frac{1}{2}\frac{\tilde{m}}{\tilde{c}} - \frac{1}{2}\tilde{c}(\tilde{d}+2)\right\}\right). \tag{A2}$$

Then, the complete log-likelihood function (up to additive constant terms) can be written as follows:

$$\log \ell_c(\boldsymbol{\Phi}_\tau \mid \mathbf{y},\mathbf{x},\tilde{\mathbf{c}},\mathbf{s},\mathbf{b}) = \sum_{i=1}^{N}\left\{\sum_{g=1}^{G} w_{ig}\log\pi_g + \sum_{j=1}^{M} u_{i1j}\log q_j + \sum_{t=2}^{T_i}\sum_{j=1}^{M}\sum_{k=1}^{M} v_{itjk}\log q_{jk}\right.$$
$$\left. + \sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg}\log f_{\mathbf{Y},\tilde{C}}(\mathbf{y}_{it},\tilde{c}_{it}\mid\mathbf{x}_{it},S_{it}=j,\mathbf{b}_g)\right\}. \tag{A3}$$

By substituting Equations (A2) in (A3), we obtain:

$$\ell_c(\boldsymbol{\Phi}_\tau) = \sum_{i=1}^{N}\left\{\sum_{g=1}^{G} w_{ig}\log\pi_g + \sum_{j=1}^{M} u_{i1j}\log q_j + \sum_{t=2}^{T_i}\sum_{j=1}^{M}\sum_{k=1}^{M} v_{itjk}\log q_{jk}\right.$$
$$-\frac{1}{2}T_i\log|\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}| + \sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg}(\mathbf{Y}_{it}-\boldsymbol{\mu}_{it})'\mathbf{D}^{-1}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\xi}}$$
$$-\frac{1}{2}\sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg}\frac{1}{\tilde{C}_{itjg}}(\mathbf{Y}_{it}-\boldsymbol{\mu}_{it})'(\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})^{-1}(\mathbf{Y}_{it}-\boldsymbol{\mu}_{it})$$
$$-\frac{1}{2}\tilde{\boldsymbol{\xi}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\xi}}\sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg}\tilde{C}_{itjg}\right\}. \tag{A4}$$

∎

Proof of Proposition 2   The E-step of the EM algorithm considers the conditional expectation of the complete log-likelihood function given the observed data and the current parameter estimates $\hat{\boldsymbol{\Phi}}_\tau^{(r-1)}$. The conditional expectations of w_{ig}, u_{itj}, v_{itjk} and z_{itjg} can be computed using standard arguments in the HMM literature as shown in Equation (17). To compute the conditional expectation of $\tilde{C}$ and $\tilde{C}^{-1}$, in the E-step of the EM algorithm, $\tilde{C}$ is treated as an additional latent variable and, hence, not observable. Using the joint distribution of $\mathbf{Y}$ and $\tilde{C}$ derived in Equation (A2) and the MAL density of $\mathbf{Y}$ given in Equation (3), we have that:

$$f_{\tilde{C}}(\tilde{C}\mid\mathbf{Y}=\mathbf{y}) = \frac{f_{\tilde{C},\mathbf{Y}}(\tilde{c},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{\tilde{c}^{-p/2}\left(\frac{2+\tilde{d}}{\tilde{m}}\right)^{\nu/2}\exp\left\{-\frac{\tilde{m}}{2\tilde{c}}-\frac{\tilde{c}(2+\tilde{d})}{2}\right\}}{2K_\nu\left(\sqrt{(2+\tilde{d})\tilde{m}}\right)}, \tag{A5}$$

which corresponds to a generalized inverse Gaussian (GIG) distribution with parameters $\nu, 2+\tilde{d}, \tilde{m}_i$, that is,[1]

$$f_{\tilde{C}}(\tilde{C}\mid\mathbf{Y}=\mathbf{y}) \sim \mathrm{GIG}(\nu,\tilde{d}+2,\tilde{m}). \tag{A6}$$

It follows that

$$\mathbb{E}[\tilde{C}\mid\cdot] = \left(\frac{\hat{\tilde{m}}}{2+\hat{\tilde{d}}}\right)^{\frac{1}{2}}\frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)}{K_\nu\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)} \tag{A7}$$

---

[1]The pdf of a GIG(p,a,b) distribution is defined as $f_{GIG}(x;p,a,b) = \frac{\left(\frac{a}{b}\right)^{p/2}}{2K_p(\sqrt{ab})}x^{p-1}e^{-\frac{1}{2}(ax+bx^{-1})}$, with $a>0$, $b>0$ and $p\in\mathcal{R}$.

and

$$\mathbb{E}[\tilde{C}^{-1} \,|\, \cdot\,] = \left(\frac{2 + \hat{\hat{d}}}{\hat{\hat{m}}} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2 + \hat{\hat{d}})\hat{\hat{m}}} \right)}{K_{\nu}\left(\sqrt{(2 + \hat{\hat{d}})\hat{\hat{m}}} \right)} - \frac{2\nu}{\hat{\hat{m}}}. \tag{A8}$$

Denoting the two conditional expectations in (A7) and (A8) by $\hat{\hat{c}}$ and $\hat{\hat{z}}$ respectively, concludes the proof. ∎

Proof of Proposition 3 Imposing the first-order conditions on (18) with respect to each component of the set $\boldsymbol{\Phi}_\tau$, gives the parameter estimates in Equations (21), (22) and (25). However, there is not closed formula solution to update the elements of the scale matrix **D**; hence, the M-step update requires using numerical optimization techniques to maximize Equation (18). A considerable disadvantage of this procedure is the necessary high computational effort which could be very time-consuming. For this reason, we utilize a simpler estimator for the scale parameters d_j, $j = 1, \dots, p$ which follows directly from the fact that all marginals of the MAL distribution are univariate AL distributions (see Marino et al., 2018; Yu & Zhang, 2005):

$$\hat{d}_j = \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \sum_{k=1}^M \hat{z}_{itkg} \rho_\tau (Y_{it}^{(j)} - \hat{\mu}_{it}^{(j)}). \tag{A9}$$

∎

APPENDIX B

In this appendix we conduct a simulation study to evaluate the finite sample properties of the proposed method and show that the introduced methodology represents a valid procedure to estimate the quantile regression coefficients. This simulation exercise addresses the following questions. First, we consider different distributional choices for the error term to study the performance of the model in the presence of non-Gaussian errors. Second, we evaluate the robustness of the non-parametric approach to non-Gaussian distributions for the subject-specific, random coefficients. Finally, we analyse the performance of penalized likelihood criteria in selecting the optimal number of mixture components G and hidden states M.

We consider two sample sizes $N = (100, 200)$ and two longitudinal lengths $T_i = T = (5, 10)$, for all $i = 1, \dots, N$, for a continuous response variable of dimension $p = 2$ and two explanatory variables $X_{it}^{(1)} \sim \mathcal{N}(0, 1)$ and $X_{it}^{(2)} \sim \text{Ber}(0.5)$. The observations are generated from a two state homogeneous Markov chain, that is, $M = 2$, using the following data generating process:

$$\mathbf{Y}_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{Z}_{it}\mathbf{b}_i + \mathbf{W}_{it}\boldsymbol{\alpha}_{S_{it}} + \varepsilon_{it}. \tag{A10}$$

Regarding the hidden Markov chain, the simulation scheme is similar to the one adopted by Marino et al. (2018). The true values of the fixed, $\boldsymbol{\beta}$, state dependent parameters, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_M)$ and the initial probabilities, $\mathbf{q}$, and transition probabilities, $\mathbf{Q}$, are given by, respectively:

$$\boldsymbol{\beta} = \begin{pmatrix} 2 & -0.8 \\ -1.4 & 3.0 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} 5 & -2 \\ -5 & 2 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}, \quad \mathbf{q} = \begin{pmatrix} 0.7 & 0.3 \end{pmatrix}. \tag{A11}$$

We consider a time-varying random intercept by setting $\mathbf{W}_{it} = \mathbf{1}$ and a random slope $\mathbf{Z}_{it} = X_{it}^{(1)}$. Hence, $\mathbf{b}_i$ are time-constant random slopes that capture individual departures from the marginal effect $\boldsymbol{\beta}$. For each sample size, two different simulation scenarios for the error distributions and for the random coefficients distributions are considered:

1. $(\mathcal{N} - \mathcal{N})$: $\mathbf{b}_i$ represent i.i.d. draws from a standard bivariate Gaussian with variance–covariance matrix, $\boldsymbol{\Omega} = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$ and the error terms, ε_{it}, are generated from a bivariate Normal random variable with zero mean vector and variance–covariance matrix equal to $\tilde{\boldsymbol{\Omega}}$;
2. $(\mathcal{T} - \mathcal{T})$: $\mathbf{b}_i$ are sampled from a bivariate Student t with three degrees of freedom, centred around zero and scale matrix $\boldsymbol{\Omega} = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$ while, ε_{it} are generated from a bivariate Student t distribution with three degrees of freedom, zero mean and scale matrix $\tilde{\boldsymbol{\Omega}}$.

Each simulation scenario is repeated twice by generating the errors ε_{it} with low $\left(\tilde{\boldsymbol{\Omega}} = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} \right)$ and high $\left(\tilde{\boldsymbol{\Omega}} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \right)$ correlation between the responses of unit i at a given time t.

To fit the proposed model, we consider a varying number of mixture components $G = (2, \ldots, 10)$ and retained the model with the lowest BIC value. We analyse three different quantile levels: in the first case, we assume $\boldsymbol{\tau} = (0.50, 0.50)$; in the second one, we set $\boldsymbol{\tau} = (0.25, 0.25)$ and in the third, we set $\boldsymbol{\tau} = (0.75, 0.75)$. For each model, we carry out $B = 250$ Monte Carlo replications and report the following indicators. The average relative bias (ARB) defined as:

$$ARB(\hat{\theta}_\tau) = \frac{1}{B} \sum_{b=1}^{B} \frac{(\hat{\theta}_\tau^{(b)} - \theta_\tau)}{\theta_\tau} \times 100, \tag{A12}$$

where $\hat{\theta}_\tau^{(b)}$ is the estimated parameter at quantile level $\boldsymbol{\tau}$ for the bth replication and θ_τ is the corresponding 'true' value of the parameter. Second, the root mean square error (RMSE) of model parameters averaged across the B simulations:

$$RMSE(\hat{\theta}_\tau) = \sqrt{\frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}_\tau^{(b)} - \theta_\tau)^2}. \tag{A13}$$

Tables A1 and A2 report the results for the fixed parameters $\boldsymbol{\beta}$ and state-specific coefficients $\boldsymbol{\alpha}$.

As can be noted, the proposed model under the normal and the Student t error distributions is able to recover the true fixed parameters and state-dependent intercept values for both low (Panels A) and high (Panels B) degree of dependence. Not surprisingly, the bias effect is quite small when we analyse the median levels (see columns 1 and 4). As the quantile levels become more extreme (see columns 2, 3, 5 and 6), the ARB slightly increases but it still remains reasonably small. Such a differences are due to the reduced amount of information in the tails of the distribution. However, both the ARB and the RMSE tend to decrease with increasing sample sizes and number of measurement occasions. Also, under the scenario where $\mathbf{b}_i \sim \mathcal{T}_2(\mathbf{0}, \boldsymbol{\Omega})$ and $\varepsilon_{it} \sim \mathcal{T}_2(\mathbf{0}, \tilde{\boldsymbol{\Omega}})$, the heavier tails of the Student t contribute to higher ARB and RMSE especially at the 25th and 75th percentiles. Concerning the hidden process, it is worth noting that we observe sensible differences in

TABLE A1 Average relative bias (ARB) and root mean square error (RMSE) (in brackets) for longitudinal and state-parameter estimates with a sample size $N = 100$ and length of longitudinal sequences $T = 5$

τ	$(\mathcal{N} - \mathcal{N})$			$(\mathcal{T} - \mathcal{T})$		
	(0.50, 0.50)	**(0.25, 0.25)**	**(0.75, 0.75)**	**(0.50, 0.50)**	**(0.25, 0.25)**	**(0.75, 0.75)**
Panel A: $\rho_{12} = 0.3$						
β_{11}	0.386 (0.114)	0.316 (0.110)	0.470 (0.116)	−0.310 (0.168)	−0.325 (0.175)	−0.214 (0.174)
β_{12}	0.716 (0.122)	1.317 (0.123)	0.912 (0.125)	0.951 (0.158)	0.313 (0.166)	0.871 (0.173)
β_{21}	−0.281 (0.071)	0.942 (0.081)	−2.603 (0.091)	0.546 (0.075)	2.137 (0.102)	−1.752 (0.100)
β_{22}	−0.019 (0.073)	−0.652 (0.091)	0.838 (0.089)	0.044 (0.083)	−1.354 (0.111)	1.163 (0.103)
α_{11}	−0.028 (0.057)	−0.252 (0.063)	0.017 (0.070)	0.033 (0.058)	−0.201 (0.078)	0.535 (0.091)
α_{12}	0.397 (0.052)	0.387 (0.069)	0.042 (0.062)	−0.180 (0.068)	1.215 (0.082)	−1.459 (0.093)
α_{21}	−0.165 (0.074)	−0.034 (0.082)	−0.096 (0.080)	0.179 (0.081)	0.763 (0.113)	0.066 (0.097)
α_{22}	−0.120 (0.075)	−0.057 (0.083)	0.293 (0.084)	0.011 (0.078)	−1.862 (0.108)	0.795 (0.095)
Panel B: $\rho_{12} = 0.8$						
β_{11}	0.281 (0.113)	0.178 (0.111)	0.294 (0.116)	−0.339 (0.179)	−0.668 (0.185)	−0.655 (0.188)
β_{12}	0.951 (0.126)	1.170 (0.127)	0.700 (0.128)	1.413 (0.165)	1.876 (0.180)	1.946 (0.173)
β_{21}	−0.522 (0.073)	2.041 (0.087)	−3.086 (0.091)	0.282 (0.075)	3.756 (0.108)	−3.297 (0.113)
β_{22}	0.175 (0.074)	−1.205 (0.094)	1.502 (0.094)	−0.106 (0.081)	−2.019 (0.119)	1.717 (0.119)
α_{11}	−0.054 (0.055)	−0.438 (0.073)	0.168 (0.068)	0.012 (0.063)	−0.695 (0.094)	0.813 (0.097)
α_{12}	0.302 (0.052)	1.255 (0.078)	−0.434 (0.067)	0.050 (0.068)	1.997 (0.093)	−2.190 (0.099)
α_{21}	−0.031 (0.071)	0.170 (0.083)	−0.383 (0.077)	0.028 (0.086)	1.068 (0.121)	−0.582 (0.118)
α_{22}	−0.211 (0.076)	−0.457 (0.081)	1.068 (0.081)	0.012 (0.082)	−2.349 (0.115)	1.873 (0.112)

terms of efficiency for the state dependent parameters $\boldsymbol{\alpha}$. Given the true values of $\mathbf{Q}$ and $\mathbf{q}$, most of the units are in the first state of the latent Markov chain, sharing the common intercept value $\boldsymbol{\alpha}_1$. Hence, the intercept corresponding to the second state $\boldsymbol{\alpha}_2$ is estimated with lower precision due to lack of transitions from one state to the other. However, when the number of repeated measurements increases, we observe more frequent transitions towards the second state with the effect of reducing the RMSE. Again, such difference is more evident in the tails of the distribution. These findings are generally consistent with the ones in Marino et al. (2018).

To evaluate the performance of the model selection procedure described in Section 4, we considered the same simulation experiment with $N = 200$, $T = 10$, $M = 2$, $\tau = (0.50, 0.50)$ and $B = 100$. Following Marino et al. (2018), for each of the simulated datasets we fit the QMHMM for $G = (2, \ldots, 8)$ and $M = (2, \ldots, 4)$, and select the optimal value of the pair (G, M) by using the AIC (Akaike, 1998) and BIC in Equation (28). Because the time-constant random slopes $\mathbf{b}_i$ in Equation (A10) are generated from continuous distributions, we only report the distribution of absolute frequencies of the hidden states M selected by the two penalized likelihood criteria. Table A3 summarizes the results.

As one can see, the BIC works well and outperform the AIC, with an average of correctly identified number of hidden states of more than the 80% across all simulation scenarios and levels of correlation. Furthermore, regardless of the distributional assumptions on the random slopes or on the error terms, the BIC captures the serial heterogeneity in the data in a more parsimonious manner compared to the AIC, hence offering easier interpretation about unobserved heterogeneity.

TABLE A2 Average relative bias (ARB) and root mean square error (RMSE) (in brackets) for longitudinal and state-parameter estimates with a sample size $N = 200$ and length of longitudinal sequences $T = 10$

τ	$(\mathcal{N} - \mathcal{N})$			$(\mathcal{T} - \mathcal{T})$		
	(0.50, 0.50)	**(0.25, 0.25)**	**(0.75, 0.75)**	**(0.50, 0.50)**	**(0.25, 0.25)**	**(0.75, 0.75)**
Panel A: $\rho_{12} = 0.3$						
β_{11}	0.173 (0.077)	0.008 (0.075)	−0.004 (0.075)	0.098 (0.118)	−0.076 (0.120)	−0.010 (0.120)
β_{12}	−0.432 (0.076)	−0.273 (0.076)	0.031 (0.076)	−1.486 (0.107)	−0.503 (0.109)	−1.429 (0.106)
β_{21}	−0.291 (0.042)	1.947 (0.056)	−2.187 (0.055)	−0.108 (0.044)	1.628 (0.060)	−2.230 (0.066)
β_{22}	0.066 (0.036)	−1.013 (0.056)	1.052 (0.055)	−0.098 (0.043)	−1.165 (0.069)	1.280 (0.070)
α_{11}	−0.096 (0.032)	−0.197 (0.041)	0.241 (0.041)	0.030 (0.035)	−0.278 (0.047)	0.551 (0.057)
α_{12}	0.116 (0.032)	0.409 (0.041)	−0.657 (0.039)	0.062 (0.036)	1.425 (0.059)	−1.398 (0.063)
α_{21}	−0.038 (0.037)	0.262 (0.052)	−0.296 (0.047)	0.061 (0.039)	0.666 (0.067)	−0.349 (0.058)
α_{22}	−0.040 (0.038)	−0.862 (0.047)	0.372 (0.042)	−0.021 (0.038)	−1.695 (0.067)	1.568 (0.070)
Panel B: $\rho_{12} = 0.8$						
β_{11}	0.040 (0.075)	−0.028 (0.078)	−0.028 (0.079)	0.280 (0.123)	−0.113 (0.125)	−0.055 (0.119)
β_{12}	−0.117 (0.075)	0.151 (0.080)	0.156 (0.077)	−1.511 (0.115)	−0.685 (0.112)	−0.718 (0.116)
β_{21}	−0.385 (0.044)	3.246 (0.070)	−3.284 (0.070)	−0.048 (0.051)	4.132 (0.087)	−4.411 (0.085)
β_{22}	0.080 (0.045)	−1.532 (0.072)	1.500 (0.070)	−0.036 (0.051)	−1.968 (0.088)	2.047 (0.089)
α_{11}	−0.040 (0.035)	−0.703 (0.058)	0.594 (0.052)	0.031 (0.037)	−0.993 (0.075)	0.942 (0.077)
α_{12}	0.159 (0.036)	1.878 (0.059)	−1.637 (0.055)	−0.006 (0.036)	2.839 (0.080)	−2.511 (0.080)
α_{21}	−0.028 (0.039)	0.666 (0.058)	−0.702 (0.062)	0.054 (0.042)	1.031 (0.084)	−0.950 (0.077)
α_{22}	0.045 (0.037)	−1.822 (0.057)	1.933 (0.061)	−0.074 (0.044)	−2.590 (0.084)	2.615 (0.081)

TABLE A3 Absolute frequency distributions of the selected hidden states M via AIC and BIC, with a sample size $N = 200$ and length of longitudinal sequences $T = 10$, over $B = 100$ Monte Carlo replications

Correlation Scenario	$\rho_{12} = 0.3$				$\rho_{12} = 0.8$			
	$(\mathcal{N} - \mathcal{N})$		$(\mathcal{T} - \mathcal{T})$		$(\mathcal{N} - \mathcal{N})$		$(\mathcal{T} - \mathcal{T})$	
	AIC	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
M								
2	20	67	79	92	33	68	92	96
3	34	22	19	8	36	25	6	4
4	46	11	2	0	31	7	2	0