M-quantile regression for multivariate longitudinal data with an application to the Millennium Cohort Study

Marco Alfò* Maria Francesca Marino† Maria Giovanna Ranalli‡
Nicola Salvati§ Nikos Tzavidis¶

Abstract

Motivated by the analysis of data from the UK Millennium Cohort Study on emotional and behavioural disorders, we develop an M-quantile regression model for multivariate longitudinal responses. M-quantile regression is an appealing alternative to standard regression models; it combines features of quantile and expectile regression and it may produce a detailed picture of the conditional response variable distribution, while ensuring robustness to outlying data. As we deal with multivariate data, we need to specify what it is meant by M-quantile in this context, and how the structure of dependence between univariate profiles may be accounted for. Here, we consider univariate (conditional) M-quantile regression models with outcome-specific random effects for each outcome. Dependence between outcomes is introduced by assuming that the random effects in the univariate models are dependent. The multivariate distribution of the random effects is left unspecified and estimated from the observed data. Adopting this approach, we are able to model dependence both within and between outcomes. We further discuss a suitable model parameterization to account for potential endogeneity of the observed covariates. An extended EM algorithm is defined to derive estimates under a maximum likelihood approach.

Keywords: Correlated random effects; Finite mixtures; Influence function; Multivariate responses; Nonparametric maximum likelihood; Robust regression.

1 Introduction

The analysis of longitudinal data may help obtain in-depth information on the evolution of a response of interest over time. In empirical applications, we need to account for

*Dipartimento di Scienze Statistiche, Sapienza Università di Roma

†Dipartimento di Statistica, Informatica, Applicazioni, Università degli Studi di Firenze. mariafrancesca.marino@unifi.it

‡Dipartimento di Scienze Politiche, Università degli Studi di Perugia

§Dipartimento di Economia e Management, Università di Pisa

¶Department of Social Statistics and Demography, Southampton Statistical Sciences Research Institute, University of Southampton

dependence between observations taken from the same unit at different time occasions. In a regression framework, this is often achieved by considering subject-specific random effects in the linear predictor; the corresponding model is known in the literature as random or mixed effect model, see Laird and Ware (1982) for early developments.

Recently, there has been an increasing interest in the application of quantile regression to longitudinal data to study how the effect of observed covariates changes across the range of the (conditional) response distribution, and to obtain a more detailed picture of the phenomenon of interest (Koenker, 2004). Geraci and Bottai (2007) propose a quantile regression model for longitudinal observations with subject-specific random intercepts having either a Gaussian or an Asymmetric Laplace density. Liu and Bottai (2009) and Geraci and Bottai (2014) further extend this proposal to general models with random intercepts and slopes. Alfò et al. (2017) consider finite mixtures of quantile regression models where the discrete distribution represents a nonparametric estimate of an unspecified, possibly continuous, distribution for the random effects. The time constant random effect model has been extended to time-varying subject-speficic intercepts by Farcomeni (2012); Marino et al. (2016) develop a mixed hidden Markov quantile regression model where both time-constant and time-varying random effects are considered. The interested reader may refer e.g. to Marino and Farcomeni (2015) for a review on quantile regression models for longitudinal data. Further, one may refer to Kulkarni et al. (2019) for a review on joint quantile regression models of multiple longitudinal responses.

M-quantile regression generalizes quantile regression by considering influence functions (Breckling and Chambers, 1988). Although M-quantiles have a less intuitive interpretation when compared to standard quantiles (Jones, 1994), M-quantile regression offers a number of specific advantages: (*i*) it allows for robust estimation; (*ii*) it can trade robustness for efficiency by modifying the tuning constant in the influence function; (*iii*) it offers computational stability as a wide range of continuous influence functions can be used (Tzavidis et al., 2016). Regression modeling beyond the mean of the response has found a lot of attention in the last years and expectile regression provides a quantile-like extension to mean regression (Newey and Powell, 1987; Kneib, 2013). See Waltrup et al. (2015) for a comparison of expectile regression with quantile regression. As expectile regression is based on an asymmetric (weighted) least squares estimate, flexible models and smoothing can be directly transferred from mean regression. In fact, extensions of expectile regression have been introduced that allow for smoothing (Schnabel and Eilers, 2009), semiparametric, and geoadditive modeling (Sobotka and Kneib, 2012), inference (Sobotka et al., 2013), Bayesian (Waldmann et al., 2017) and frequentist model selection (Spiegel et al., 2017). However, expectile regression is sensitive to outliers as much as mean regression. In this regard, M-quantiles provide a robustification of expectiles through the use of influence functions, so that M-quantile regression can also be seen as a generalization of expectile regression.

The extension of M-quantile regression to longitudinal data is quite recent. Tzavidis et al. (2016) propose a model with Gaussian subject-specific random intercepts and suggest the use of pseudo-BLUP equations (see e.g. Harville, 1976) to derive parameter estimates. Alfò et al. (2017) consider a discrete specification for the distribution of subject-specific random intercepts. As noted before, the likelihood function resembles that of a finite mixture, and the model is referred to as Finite Mixture of M-Quantile regression models

(FMMQ). This class of models has also been applied to handle unobserved heterogeneity in assessing the effect of meteorology and traffic on air quality data (Del Sarto et al., 2019).

In the analysis of multivariate longitudinal responses, the research interest may focus not only on defining a regression model for each response, but also on investigating (and interpreting) the structure of dependence between responses. The univariate approaches we have mentioned so far are not appropriate for this purpose, as they provide only partial (univariate) pictures of such a complex phenomenon. For an extensive discussion on models for fitting (at the mean) multivariate longitudinal data, see Verbeke et al. (2014). Here, the authors present a review on several approaches to model multiple outcomes measured repeatedly within a set of study participants. In particular, they focus on advantages and disadvantages of different families of models that can be distinguished based on whether or not latent variables are assumed for the time dimension and/or for the outcome dimension.

The idea of extending M-estimates and M-quantiles to multivariate settings dates back to Breckling and Chambers (1988), who discuss how to provide a robust technique for summarizing the distribution of multidimensional data. The definition is based on a simple generalization of the one-dimensional loss function for quantiles and M-quantiles. However, this approach does not produce intuitive results in certain situations. For example, the estimated quantiles may be outside the convex hull of the data (Breckling et al., 2001). An alternative definition, based on a multivariate generalization of the univariate estimating equations for quantiles and M-quantiles, is discussed in Breckling et al. (2001). In the quantile regression framework, the extension to multivariate responses is dealt by Petrella and Raponi (2019). They account for the association among several responses while studying the effect of observed predictors on different quantiles of the marginal (univariate), conditional distribution of the responses. Earlier references in the quantile regression context are Chakraborty (2003) and Hallin et al. (2010).

However, these proposals are designed for cross-sectional data only and, thus, they do not allow to model dependence between observations taken repeatedly from the same subject over time. To handle such a complex data structure, we propose to extend the univariate FMMQ introduced by Alfò et al. (2017) to the multivariate context. We define a set of univariate equations with outcome-specific random effects to model the association *within* individual profiles (same response recorded at different time occasions from the same subject); the multivariate distribution of the outcome-specific random effects accounts for the dependence *between* the individual univariate profiles (different responses from the same subject at the same occasion). The model is similar to the one proposed by Kulkarni et al. (2019) in the context of joint quantile regression models for multiple longitudinal data, apart from the different scale induced by the Asymmetric Laplace distribution they use and the M-estimation (and M-quantiles) we consider here. In this respect, a comparison between the two methods would not be well-grounded.

We also consider potential *endogeneity* of the observed covariates and show how the auxiliary regression approach by Mundlak (1978) can be simply adapted to the current M-quantile framework. Smith et al. (2015), Arellano and Bonhomme (2016), and Weidner and Moon (2017) discuss similar approaches in the quantile regression framework. To the best of our knowledge, the present manuscript represents the first attempt to account for both issues (multivariate dependence and endogeneity) in the context of M-quantile regression.

The proposed model is motivated by the analysis of data from the Millennium Cohort Study (MCS), a multi-disciplinary research project covering around 19,000 children who were born in the UK during 2000/02 (`http://www.cls.ioe.ac.uk`). Extending the analysis by Tzavidis et al. (2016), we build a *joint* model for emotional and behavioral disorders as a function of neighborhood and family risk factors, allowing for potential endogeneity issues through an auxiliary regression approach.

The paper is organized as follows. In Section 2, the Millennium Cohort Study is introduced and the data are described; in Sections 3 and 4, we present the M-quantile regression model for longitudinal responses and its extension to the multivariate framework. The analysis of the MCS data is discussed in Section 5. The last section contains concluding remarks and outlines potential future research agenda.

2 The Millennium Cohort Study data

The Millennium Cohort Study (MCS in the following) is a longitudinal study on the growth of around 19,000 children in the United Kingdom. It involves children living in the UK at nine months of age, who were born between September 1, 2000 and August 31, 2001 in England and Wales, or between November 23, 2000 and January 11, 2002 in Scotland and Northern Ireland, whose families were eligible to receive child benefits.

To better address the effect of social disadvantage on children outcomes, the study was designed to over-represent children with deprived backgrounds, with a specific focus on those areas in the country characterized by high concentration of ethnic minorities. In detail, for England, the population was grouped into three different strata: the first stratum, *ethnic minority*, includes children living in wards where the proportion of ethnic minorities was not less than 30% at the 1991 Census. The second stratum, *disadvantaged*, includes children living in wards, not in the first stratum, which fell into the poorest 25% wards based on the Child Poverty Index. The third stratum, *advantaged*, includes all other children. For Wales, Scotland, and Northern Ireland, the population was stratified into the *disadvantaged* and the *advantaged* strata only, due to the low presence of ethnic minorities. MCS wards were randomly selected within each stratum and country; then a list of all children turning nine months old during the survey window and living in the selected wards was populated. Overall, a cohort of 18,818 children was eligible and was followed for up to seven time periods. The first measurement took place when children were about 9 months old; subsequent measures were recorded at $3, 5, 7, 11, 14$, and 17 years of age. For further details, see e.g. Plewis et al. (2007)

Children's emotional and behavioral disorders were measured by means of the Strengths and Difficulties Questionnaire (SDQ); see Goodman (1997). This covers five different domains: emotional symptoms, peer problems, conduct problems, hyperactivity, and prosocial behavior. 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 (i-SDQ) is the sum of the scores to the items in the domains of emotional symptoms and peer problems. With 10 items, it ranges in the interval $[0, 20]$. The externalizing SDQ score (e-SDQ) is the sum of the scores to the items in the domains of conduct problem and hyperactivity; also in this case, with

10 items, it ranges in the interval $[0, 20]$. The two outcomes of interest were recorded for children who were, at least, 3 years old (wave 2 and more) only.

To study the impact of demographic and socio-economic factors on children disorders, we focus on the following covariates. ALE_{11} measures the number of potentially Adverse (stressful) Life Events experienced by the family in the priod between two consecutive waves. This variable is obtained by summing the responses to 11 items from the Adverse Life Event scale; see Tiet et al. (1998). SED_4 measures Socio-Economic Disadvantage and it is obtained by summing responses to four items on family poverty. KESSM reports maternal depression according to the Kessler scale (Kessler and Mroczek, 1992), with a range in $[0, 24]$, where higher values identify more severe depression symptoms. IMD measures neighborhood deprivation via the Index of Multiple Deprivation; it varies in the range $[1, 10]$ and lower values correspond to areas with higher deprivation. We also consider child age (Age), maternal education (no qualification, GCSE, and University degree), ethnicity (non-white/white), gender, and the stratification variable as covariates.

In this paper, we focus on $9,021$ children living in England, who participated in, at least, one of the waves $2, 3$, and 4 of the study and present complete covariate information. Using these eligibility criteria, data on $7,055$ children are available at the first time occasion, on $7,938$ children at occasion 2, and on $7,078$ children at occasion 3. Only $5,342$ children out of $9,021$ (59.22%) present complete information on i-0SDQ and e-SDQ scores, while the remaining ones (40.78%) have incomplete response information. In the following, we assume a MAR mechanism. A graphical representation of the available data is provided in the Supplementary Material (Section **??**).

Our main interest here is on analyzing the impact of neighborhood and family risk factors on children emotional and behavioral problems. We should remark two issues: (i) the effect of one or more risk factors may not be constant across the distribution of the SDQ scores; (ii) our interest relies on understanding the effect of observed covariates on (conditionally) higher SDQ scores associated with more problematic children. In what follows, we address both issues.

3 M-quantile regression for longitudinal data

M-quantile regression extends the ideas of M-estimation (Huber, 1964) to location parameters of a conditional response distribution. The M-quantile of order $q \in (0, 1)$ for the conditional density $f(y \mid \boldsymbol{x})$ is the solution to the following estimating equation:

$$\int \psi_q[y - MQ_q(y \mid \boldsymbol{x}; \psi)] f(y \mid x) dy = 0.$$

Here, $\psi_q(u) = 2\psi(u/\sigma_q)\{qI(u > 0) + (1 - q)I(u \leq 0)$ denotes an asymmetric influence function, that is, the first derivative of an asymmetric loss function $\rho_q(\cdot)$, and σ_q is a suitable scale parameter. When $\psi(\varepsilon) = \varepsilon$, we obtain the expectile of order q. In this sense, expectiles can be seen as a quantile-like generalization of the mean and, at the same time, M-quantiles can be seen as a robustification of expectiles through the influence function $\psi(\cdot)$. On the other hand, when $\psi(\varepsilon) = \text{sign}(\varepsilon)$, we obtain the standard quantile of order q.

In the regression context and for a given choice of ψ, Breckling and Chambers (1988) define a linear M-quantile regression model of order $q \in (0,1)$ as follows:

$$MQ_q(y_i \mid \boldsymbol{x}_i; \psi) = \boldsymbol{x}_i' \boldsymbol{\beta}_q, \quad i = 1, \ldots, n. \tag{1}$$

In this paper, we use the Huber influence function $\psi(\varepsilon) = \epsilon\, \mathrm{I}(-c \leq \varepsilon \leq c) + c\, \mathrm{sign}(\varepsilon)\, \mathrm{I}(|\varepsilon| > c)$, where c denotes a suitable tuning constant. Note that when $c \to \infty$, $\psi(\varepsilon) = \epsilon$ and M-quantile regression reverts to expectile regression, so that the latter can be seen as a particular case of the former. Therefore, by choosing a large enough value for c, we obtain expectile regression estimates. For a specified q and a continuous $\psi(\cdot)$, an estimate $\hat{\boldsymbol{\beta}}_q$ can be derived via a simple Iterative Re-Weighted Least Squares (IRWLS) algorithm. The asymptotic theory for $\hat{\boldsymbol{\beta}}_q$ follows directly from well-known results for M-estimation; see Section 2.2 of Breckling and Chambers (1988). Proofs of consistency for $\hat{\boldsymbol{\beta}}_q$ and the analytic expression for the corresponding asymptotic covariance matrix with stochastic regressors are discussed in Bianchi and Salvati (2015).

With longitudinal data, subject-specific random effects are used in the model specification to account for omitted covariates and describe dependence between observations from the same subject. To introduce notation, let Y_{it} denote a continuous longitudinal response and y_{it} the observed value for the i-th subject, $i = 1, \ldots, n$, at time occasion $t = 1, \ldots, T_i$. For a given $q \in (0,1)$, let $\boldsymbol{x}_{it} = (1, \ldots, x_{itp})'$ and $\boldsymbol{\beta}_q$ be a p-dimensional design vector and the associated vector of parameters, respectively. We consider that $s \geq 1$ covariates, denoted by $\boldsymbol{w}_{it} = (1, w_{it2}, \ldots, w_{its})' = (1, \tilde{\boldsymbol{w}}_{it})'$, are associated to a vector of subject-specific random effects $\boldsymbol{b}_{i,q} = (b_{i1,q}, \ldots, b_{is,q})'$. The individual-specific effects may represent random variation with respect to the corresponding elements in $\boldsymbol{\beta}_q$, and in this case the corresponding covariate terms are included in both $\boldsymbol{x}_{it}$ and $\boldsymbol{w}_{it}$, or rather be unconstrained, with the *global set* of covariates split between the two. In the following, we consider the former approach, so that $\boldsymbol{w}_{it} \subseteq \boldsymbol{x}_{it}$. The vector of subject-specific random effects $\boldsymbol{b}_{i,q}$ includes the intercept and, possibly, further slopes that are assumed to vary with individuals. The modeling structure is completed by the distribution of the random effects, conditional on the $(T_i \times p)$-dimensional matrix of individual covariates $\boldsymbol{X}_i$, denoted by $f_{b,q}(\boldsymbol{b}_{i,q} \mid \boldsymbol{X}_i; \boldsymbol{\Sigma}_q)$, where $\boldsymbol{\Sigma}_q$ identifies a (possibly M-quantile dependent) covariance matrix. Details on the specification of such a distribution are provided in the subsequent sections for the general case of multivariate longitudinal responses.

The M-quantile regression model of order q for Y_{it}, conditional on the observed covariates $\boldsymbol{x}_{it}$ and the subject-specific random effects $\boldsymbol{b}_{i,q}$, is defined by

$$MQ_q(y_{it} \mid \boldsymbol{x}_{it}, \boldsymbol{b}_{i,q}; \psi) = \boldsymbol{x}_{it}' \boldsymbol{\beta}_q + \boldsymbol{w}_{it}' \boldsymbol{b}_{i,q}. \tag{2}$$

Repeated measurements coming from the same subject are assumed to be independent conditional on $\boldsymbol{b}_{i,q}$ (local independence assumption) and estimates for $\boldsymbol{\beta}_q$ are obtained by solving the equation

$$\sum_{i=1}^{n} \sum_{t=1}^{T_i} \psi_q[y_{it} - MQ_q(y_{it} \mid \boldsymbol{x}_{it}, \boldsymbol{b}_{i,q}; \psi)] \boldsymbol{x}_{it} = \boldsymbol{0} \tag{3}$$

via an IRWLS algorithm. Estimation of model parameters can be cast in a maximum likelihood framework by specifying a parametric form for the conditional response distribution

associated to a specific influence function $\psi(\cdot)$. Here, our choice for the Huber influence function corresponds to the so-called Generalized Asymmetric Least Informative (GALI - Bianchi et al., 2018) density

$$f_q\left(y_{it} \mid \boldsymbol{b}_{i,q}, \boldsymbol{x}_{it}\right) = \left(\sigma_q B_q\right)^{-1} \exp\left\{-\rho_q\left[\frac{y_{it} - \mu_{it,q}}{\sigma_q}\right]\right\}, \tag{4}$$

where B_q is a normalizing constant ensuring the density integrates to one, σ_q is a scale parameter, $\mu_{it,q} = MQ_q(y_{it} \mid \mathbf{x}_{it}, \mathbf{b}_{i,q}; \psi)$ is a location parameter defined according to equation (2), and

$$\rho_q(u) = \frac{u^2}{2}\mathrm{I}(|u| \leq c) + \left(c|u| - \frac{c^2}{2}\right)\mathrm{I}(|u| > c).$$

Adopting a GALI conditional density, we obtain a likelihood equation of the type in (3). Based on the local independence assumption, the joint (conditional) density of the individual response vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT_i})'$ is

$$f_q\left(\boldsymbol{y}_i \mid \boldsymbol{b}_{i,q}, \boldsymbol{X}_i\right) = \prod_{t=1}^{T_i} f_q\left(y_{it} \mid \boldsymbol{b}_{i,q}, \boldsymbol{x}_{it}\right). \tag{5}$$

Given the modeling assumptions, the log-likelihood function for the observed data is obtained by integrating out the unobservables $\boldsymbol{b}_{i,q}, i = 1, \ldots, n$. That is,

$$\ell_q\left(\cdot\right) = \sum_{i=1}^{n} \log\left\{\int_{\mathcal{B}} f_q\left(\boldsymbol{y}_i \mid \boldsymbol{b}_{i,q}, \boldsymbol{X}_i\right) f_{b,q}(\boldsymbol{b}_{i,q} \mid \boldsymbol{X}_i; \boldsymbol{\Sigma}_q)\mathrm{d}\boldsymbol{b}_{i,q}\right\}, \tag{6}$$

As stated above, we remark that, by representing the conditional response distribution via the GALI density, we may cast standard estimation of M-quantile regression into a maximum likelihood context. This approach is similar to the one used in quantile regression modeling, where the Asymmetric Laplace Distribution (ALD - Yu and Moyeed, 2001) is considered for similar purposes.

4 Modeling multivariate longitudinal responses

In this section, we extend the approach defined for the univariate case to the analysis of an H-dimensional, longitudinal, outcome. Let us denote by Y_{ith} and $y_{ith}, h = 1, \ldots, H$, the h-th continuous longitudinal response and the corresponding observed value for the i-th subject, $i = 1, \ldots, n$, at time occasion $t = 1, \ldots, T_i$. Let $\boldsymbol{x}_{ith}$ be a p-dimensional design vector, and $\mathbf{X}_{ih}$ the corresponding $(T_i \times p)$-dimensional design matrix. We denote by $\boldsymbol{b}_{ih,q} = (b_{ih1,q}, \ldots, b_{ihs,q})'$ the s-dimensional vector of subject- and outcome-specific random effects associated to the vector of covariates $\boldsymbol{w}_{ith} \subseteq \boldsymbol{x}_{ith}$.

For a given M-quantile $q \in (0, 1)$, we assume that, conditional on $\boldsymbol{b}_{ih,q}$ and the observed design matrix $\mathbf{X}_{ih}$, measurements for the h-th response taken on subject i at different time

occasions are independent (local independence assumption). As a result, the conditional density for the individual response vector $\boldsymbol{y}_{ih} = (y_{i1h}, \ldots, y_{iT_ih})'$ is

$$f_q\left(\boldsymbol{y}_{ih} \mid \boldsymbol{b}_{ih,q}, \boldsymbol{X}_{ih}\right) = \prod_{t=1}^{T_i} f_q\left(y_{ith} \mid \boldsymbol{b}_{ih,q}, \boldsymbol{x}_{ith}\right), \quad h = 1, \ldots, H, \tag{7}$$

where $f_q\left(y_{ith} \mid \boldsymbol{b}_{ih,q}, \boldsymbol{x}_{ith}\right)$ denotes the GALI density, with scale $\sigma_{h,q}$ and location $\mu_{ith,q}$ defined by

$$\mu_{ith,q} = MQ_q(y_{ith} \mid \boldsymbol{x}_{ith}, \boldsymbol{b}_{ih,q}; \psi) = \boldsymbol{x}'_{ith}\boldsymbol{\beta}_{h,q} + \boldsymbol{w}'_{ith}\boldsymbol{b}_{ih,q}. \tag{8}$$

As in the univariate setting, $\boldsymbol{\beta}_{h,q}$ denotes the vector of constant effects for covariates $\boldsymbol{x}_{ith}$ on the (conditional) q-th M-quantile of Y_{ith}, while $\mathbf{b}_{ih,q}$ is the vector of subject-specific random effects capturing heterogeneity between units and modeling dependence between responses in $\boldsymbol{y}_{ih}, i = 1, \ldots, n$.

To account for association between multiple outcomes recorded on the same subject, we introduce a further assumption. In particular, we assume that, conditional on $\boldsymbol{b}_{i,q} = \{\boldsymbol{b}_{i1,q}, \ldots, \boldsymbol{b}_{iH,q}\}$, responses $\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iH}$ are independent. Therefore, the conditional density for the multivariate individual sequence $\boldsymbol{y}_i = \{\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iH}\}$ is:

$$f_q\left(\boldsymbol{y}_i \mid \boldsymbol{b}_{i,q}, \boldsymbol{X}_i\right) = \prod_{h=1}^{H} f_q\left(\boldsymbol{y}_{ih} \mid \boldsymbol{b}_{ih,q}, \boldsymbol{X}_{ih}\right), \tag{9}$$

where $\boldsymbol{X}_i$ collects all design matrices for subject i. Under the proposed specification, dependence between responses Y_{ith} and $Y_{ith'}$ is induced by dependence between the corresponding random effects $\boldsymbol{b}_{i,hq}$ and $\boldsymbol{b}_{i,h'q}$.

Since these latter effects are unobserved, one approach to obtain the observed data likelihood is based on integrating them out. According to equation (9), the individual contribution to the observed data log-likelihood for the q-th M-quantile is, therefore, given by

$$\ell_{i,q}\left(\cdot\right) = \log\left\{\int_{\mathcal{B}} f_q\left(\boldsymbol{y}_i \mid \boldsymbol{b}_{i,q}, \boldsymbol{X}_i\right) f_{b,q}(\boldsymbol{b}_{i,q} \mid \boldsymbol{X}_i; \boldsymbol{\Sigma}_q)\mathrm{d}\boldsymbol{b}_{i,q}\right\}, \tag{10}$$

while the observed data log-likelihood is, for n independent subjects,

$$\ell_q\left(\cdot\right) = \sum_{i=1}^{n} \ell_{i,q}\left(\cdot\right). \tag{11}$$

4.1 Handling potential endogeneity

It is worth noting that, often, individual-specific effects are included to account for potential omitted covariates, as they provide a simple and efficient way to model the impact of time-constant individual-specific features that are not included in the model. A standard assumption is that observed covariates are uncorrelated with the omitted ones (*exogeneity* assumption). That is, $f_{b,q}(\boldsymbol{b}_{i,q} \mid \boldsymbol{X}_i; \boldsymbol{\Sigma}_q) = f_{b,q}(\boldsymbol{b}_{i,q}; \boldsymbol{\Sigma}_q)$. Often, this assumption is relaxed for the less stringent assumption of *weak exogeneity* of the observed covariates: $\mathrm{E}(\boldsymbol{b}'_{ih,q}\boldsymbol{X}_{ih}) = \mathrm{cov}(\boldsymbol{b}_{ih,q}, \boldsymbol{X}_{ih}) = \mathbf{0}$.

However, in many circumstances, exogeneity does not hold and this issue (known as *endogeneity*) should be taken into consideration to obtain valid inference. The impact of endogeneity may be described by considering differences in the *within* and *between* effects of observed covariates on the conditional response distribution. The former measure the impact of dynamics in time-varying covariates on the temporal evolution of the response. The latter refer to the association between individual mean levels of the observed covariates and corresponding mean levels in the responses. Obviously, time-constant covariates can be associated to *between* effects only. According to Bartels (2008), when we consider the model specifications in equations (2) and (8) and assume $f_{b,q}(\boldsymbol{b}_{i,q} \mid \boldsymbol{X}_i; \boldsymbol{\Sigma}_q) = f_{b,q}(\boldsymbol{b}; \boldsymbol{\Sigma}_{q_{i,q}})$, we implicitly state that *within* and *between* effects for time-varying covariates are equal. However, when this is not the case, parameter estimates correspond to an uninterpretable weighted average of the two effects, and may simply reflect the impact of unobserved covariates as *mediated* by the observed ones; see Krishnakumar (2006); Neuhaus and Kalbfleisch (1998). Further, estimates of variance components may be severely biased; see Grilli and Rampichini (2011).

To handle potential endogeneity, we consider an auxiliary regression approach (Mundlak, 1978) in the current M-quantile specification:

$$\begin{cases} \boldsymbol{b}_{ih,q} = \mathrm{E}(\boldsymbol{b}_{ih,q} \mid \boldsymbol{X}_{ih}) + \boldsymbol{b}^*_{ih,q}, \\ \mathrm{E}(\boldsymbol{b}_{ih,q} \mid \boldsymbol{X}_{ih}) = \boldsymbol{\Lambda}_{h,q} \bar{\boldsymbol{x}}_{ih}, \end{cases}$$

where $\bar{\boldsymbol{x}}_{ih} = T_i^{-1} \sum_t \boldsymbol{x}_{ith}$ is the vector of individual means used as an (informal) *instrument*. We may note that the *residual* latent effects $\boldsymbol{b}^*_{ih,q}$ are now, at least approximately, free from $\boldsymbol{X}_{ih}$. Using such a parameterization, the M-quantile regression model in (8) can be written as

$$\begin{aligned} MQ_q(y_{ith} \mid \boldsymbol{x}_{ith}, \boldsymbol{b}_{ih,q}; \psi) &= \boldsymbol{x}'_{ith}\boldsymbol{\beta}_{h,q} + \boldsymbol{w}'_{ith}\boldsymbol{b}_{ih,q} \\ &= \boldsymbol{x}'_{ith}\boldsymbol{\beta}_{h,q} + \boldsymbol{w}'_{ith}\left[\boldsymbol{\Lambda}_{h,q}\bar{\boldsymbol{x}}_{ih} + \boldsymbol{b}^*_{ih,q}\right] \\ &= (\boldsymbol{x}_{ith} - \bar{\boldsymbol{x}}_{ih})'\boldsymbol{\beta}_{h,q} + \bar{\boldsymbol{x}}'_{ih}\left\{\boldsymbol{\beta}_{h,q} + \boldsymbol{\Lambda}'_{h,q}[1, \tilde{\boldsymbol{w}}_{ith}]\right\} + \boldsymbol{w}'_{ith}\boldsymbol{b}^*_{ih,q} \\ &= (\boldsymbol{x}_{ith} - \bar{\boldsymbol{x}}_{ih})'\boldsymbol{\beta}_{h,q} + \bar{\boldsymbol{x}}_{ih}\left[\boldsymbol{\beta}_{h,q} + \boldsymbol{\Lambda}^{(1)}_{h,q} + \boldsymbol{\Lambda}^{(2:s)}_{h,q}\tilde{\boldsymbol{w}}_{ith} +\right] + \boldsymbol{w}'_{ith}\boldsymbol{b}^*_{ih,q}, \end{aligned} \tag{12}$$

where $\boldsymbol{\Lambda}^{(1)}_{h,q}$ and $\boldsymbol{\Lambda}^{(2:s)}_{h,q}$ are used to denote the first and the remaining columns of $\boldsymbol{\Lambda}_{h,q}$, respectively, while $\boldsymbol{\beta}_{h,q}$ and $\boldsymbol{\delta}_{h,q} = \boldsymbol{\beta}_{h,q} + \boldsymbol{\Lambda}^{(1)}_{h,q}$ represent the *within* and the *between* effects for the observed covariates on the q-th M-quantile of the conditional response distribution, respectively. On the other hand, $\boldsymbol{b}^*_{ih,q}$ is a vector of subject-specific random effects capturing unobserved, individual-specific, heterogeneity uncorrelated with $\boldsymbol{X}_{ih}$. In the following, we simplify the notation and suppress the asterisk when referring to such a vector. The projection of the observed covariates onto the spaces spanned by the *mean* and the *deviation from the mean* is known in the literature as the QP decomposition. Further references on *endogeneity* in the context of mean and quantile regression are Bell and Jones (2015) and Abrevaya and Dahl (2008), among others.

4.2 The random effect distribution

In the context of univariate M-quantile regression for longitudinal data ($H = 1$), Alfò et al. (2017) propose to leave the random effect distribution unspecified and estimate it from the observed data by means of a NonParametric Maximum Likelihood (NPML - Aitkin, 1996, 1999) approach. This is known to lead to a discrete estimate defined over a finite number of locations (Laird, 1978; Lindsay, 1983a,b) and to approximate the likelihood function via a finite mixture.

In this paper, we extend this approach to the multivariate context. For a given $q \in (0, 1)$ level, we assume that the (possibly continuous) distribution of the random effects is approximated by a discrete distribution defined over a finite set of multivariate locations $\{\boldsymbol{\zeta}_{1,q}, \ldots, \boldsymbol{\zeta}_{K_q,q}\}$, with $\boldsymbol{\zeta}_{k,q} = (\boldsymbol{\zeta}_{k1,q}, \ldots, \boldsymbol{\zeta}_{kH,q})'$, $\boldsymbol{\zeta}_{kh,q} \in \mathbb{R}^s$, and masses

$$\pi_{k,q} = \Pr(\boldsymbol{b}_{i,q} = \boldsymbol{\zeta}_{k,q}) = \Pr\left\{(\boldsymbol{b}_{i1,q} = \boldsymbol{\zeta}_{k1,q}, \ldots, \boldsymbol{b}_{iH,q} = \boldsymbol{\zeta}_{kH,q})\right\},$$

where $\pi_{k,q} \geq 0, \forall k = 1, \ldots, K_q$ and $\sum_k \pi_{k,q} = 1$.

The association between multiple responses coming from the same subject arises through the common latent structure. Since locations may vary with the specific response, the proposed approach directly allows for negative association between the different outcomes, thus overcoming a drawback of standard shared parameters models, where $\boldsymbol{b}_{ih,q} = \boldsymbol{b}_{i,q}, \forall h = 1, \ldots, H$, see e.g. Wu and Carroll (1988) and Wu and Bailey (1989).

For a given M-quantile level $q \in (0, 1)$, conditional on the membership to the k-th component of the finite mixture (where $\boldsymbol{b}_{i,q} = \boldsymbol{\zeta}_{k,q}$) and the set of observed covariates $\boldsymbol{x}_{ith}$, we assume that responses Y_{ith} follow a GALI distribution with location parameter

$$\begin{aligned}\mu_{ithk,q} &= MQ_q(y_{ith} \mid \boldsymbol{x}_{ith}, \boldsymbol{\zeta}_{kh,q}; \psi) = \\ &= (\boldsymbol{x}_{ith} - \bar{\boldsymbol{x}}_{ih})'\boldsymbol{\beta}_{h,q} + \bar{\boldsymbol{x}}'_{ih}\boldsymbol{\delta}_{h,q} + \boldsymbol{\Lambda}^{(2:s)}_{h,q}\tilde{\boldsymbol{w}}_{ith} + \boldsymbol{w}'_{ith}\boldsymbol{\zeta}_{kh,q}. \qquad (13)\end{aligned}$$

The observed data likelihood in equation (11) can therefore be written as

$$\ell_q(\cdot) = \sum_{i=1}^{n} \log\left\{\sum_{k=1}^{K_q} f_q\left(\boldsymbol{y}_i \mid \boldsymbol{\zeta}_{k,q}, \boldsymbol{X}_i\right)\pi_{k,q}\right\}, \qquad (14)$$

where, due to local independence,

$$f_q\left(\boldsymbol{y}_i \mid \boldsymbol{\zeta}_{k,q}, \boldsymbol{X}_i\right) = \prod_{h=1}^{H}\prod_{t=1}^{T_i} f_q(y_{ith} \mid \boldsymbol{\zeta}_{kh,q}, \boldsymbol{x}_{ith}),$$

and $f_q(y_{ith} \mid \boldsymbol{\zeta}_{kh,q}, \boldsymbol{x}_{ith})$ is the GALI density with scale $\sigma_{h,q}$ and location parameter $\mu_{ithk,q}$ defined by equation (13). Under the proposed specification, as for the univariate case, the likelihood function resembles that of a finite mixture; for this reason, we will refer to it as multivariate Finite Mixture of M-Quantile regression models (mFMMQ). The computational details of the EM algorithm for maximum likelihood estimation are reported in Section **??** of the Supplementary Material. There, we also discuss the procedure to derive the estimate for the covariance matrix of model parameter estimates and the corresponding standard errors, as well as that for selecting the optimal number of mixture components

K_q^*.

While the discrete nature of $f_{b,q}(\boldsymbol{b}_i; \boldsymbol{\Sigma}_q)$ may seem unappealing, it is worth highlighting that most approximation techniques used to deal with mixed effect models based on a parametric distribution of the random effects are exactly of the type in equation (14). The only substantial difference is that, under the proposed framework, locations $\boldsymbol{\zeta}_{k,q}$ and masses $\pi_{k,q}$ are directly estimated from the observed data, rather than being fixed a priori. This implies that, if the standard zero mean, multivariate, Gaussian assumption was reasonable, the NPML estimate of the random effect distribution would have symmetric, bell-shaped, margins centered around zero. This clearly would come at the cost of a limited loss in efficiency of parameter estimates. Based on such considerations, the proposal can be seen as a flexible approach to model dependence among multivariate longitudinal data, which does not rely on parametric assumptions and, therefore, is more in line with the spirit of M-quantile regression.

4.3 Studying the association between responses

As stated before, one of the main advantages of the proposed specification is that we may explicitly study the association between multivariate responses. Potential research questions entail the direction of such a dependence and the corresponding magnitude. As regards the former, we may be interested in understanding whether the outcomes are likely to occur jointly or, rather, if they move in the opposite direction. Furthermore, it can be relevant to understand whether specific values of the responses are more or less likely to jointly occur, that is understanding whether association among outcomes varies with q. In this sense, postulating a random effect distribution depending on the M-quantile level allows us to provide an answer to such questions.

More in detail, to study the association among multiple outcomes, we may rely on the estimated covariance matrix of the random effects. Under the proposed modeling approach, this matrix can be estimated as follows

$$\hat{\boldsymbol{\Sigma}}_q = \left\{ \hat{\boldsymbol{\Sigma}}_{h,h',q} \right\}_{h,h'=1,\ldots,H},$$

where

$$\hat{\boldsymbol{\Sigma}}_{h,h',q} = \sum_{k=1}^{K_q} \left(\hat{\boldsymbol{\zeta}}_{kh,q} - \hat{\bar{\boldsymbol{\zeta}}}_{h,q} \right) \left(\hat{\boldsymbol{\zeta}}_{kh',q} - \hat{\bar{\boldsymbol{\zeta}}}_{h,q} \right)' \hat{\pi}_{k,q},$$

and $\hat{\bar{\boldsymbol{\zeta}}}_{h,q} = \sum_k \hat{\boldsymbol{\zeta}}_{kh,q} \hat{\pi}_{k,q}$.

Block diagonal elements in $\hat{\boldsymbol{\Sigma}}_q$ correspond to the covariance matrix for the random effects in a given response equation and, as usual, provide information on the impact that sources of unobserved heterogeneity have on the (conditional) distribution of that response. On the other hand, off-block diagonal elements correspond to covariances between random effects in different response equations. To give an example, when $s = 1$ (i.e. when a random intercept model is considered), the quantity

$$\hat{\sigma}_{hh',q} = \sum_{k=1}^{K_q} \left(\hat{\zeta}_{kh,q} - \hat{\bar{\zeta}}_{h,q} \right) \left(\hat{\zeta}_{kh',q} - \hat{\bar{\zeta}}_{h',q} \right) \hat{\pi}_{k,q}$$

provides an indirect measure of association between Y_{ith} and $Y_{ith'}$ for a given M-quantile level q.

5 Analysis of the Millennium Cohort Study data

In this section, we analyze data from the MCS to understand how individual covariates, especially those associated with neighborhood and family risk factors, influence children's emotional and behavioral disorders, measured by means of the i-SDQ and the e-SDQ scores, respectively.

To start, we analyze MCS data via the univariate FMMQ approach by Alfò et al. (2017). Here, the potential dependence between internalizing and externalizing scores is not taken into account. A second step of analysis involves the proposed mFMMQ specification, with the aim of deriving insight on the association between the two outcomes under investigation. Both analyses are based on the assumption of *exogeneity* of observed covariates. In the following, we refer to such models as the univariate and the multivariate *pooled* FMMQ approaches, respectively, and denote them by uFMMQ$_p$ and mFMMQ$_p$. Last, we consider the auxiliary regression approach based on the QP decomposition discussed in Section 4.1. We refer to such a model as the multivariate *within-between* FMMQ and denote it as mFMMQ$_{wb}$. The latter analysis is meant to avoid possible bias in the parameter estimates due to *endogeneity* and to provide more reliable estimates of the effects of children socio-economic conditions on the evolution of children's psychopathology over time.

We focus on the right tail of the response distribution and estimate the model for $q = \{0.50, 0.75, 0.90\}$. These levels are generally associated with more severe problems (conditional on the observed covariates and the random effects). Results from the above analyses are reported in Sections 5.1-5.3.

We must notice that M-quantile regression is designed to deal with continuous data, while the SDQ scores are indeed discrete. However, the sum over 20 different items and the structure of the observed covariates provide sufficient variability to justify the proposed approach. A potential alternative could be that of considering a power transformation of the response (such as the logarithmic one). However, the literature on this topic mainly concerns models that estimate the conditional mean. There are only a few contributions on the use of power transformations in the context of linear quantile regression (Mu and He, 2007); these exploit the equivariance property typical of this class of models. However, as we model M-quantiles, this appealing property does not hold any longer, making such an approach not really appropriate.

5.1 The univariate FMMQ with *pooled* effects

For a given M-quantile level $q \in (0, 1)$ and a given response $h = 1, 2$, we consider the following univariate model specifications:

$$\begin{cases} MQ_q(\text{i-SDQ}_{it} \mid \boldsymbol{x}_{it}, \zeta_{k_1 1,q}; \psi) = \zeta_{k_1 1,q} + \boldsymbol{x}'_{1it}\boldsymbol{\beta}_{11,q} + \boldsymbol{x}'_{2i}\boldsymbol{\beta}_{21,q}, \\ MQ_q(\text{e-SDQ}_{it} \mid \boldsymbol{x}_{it}, \zeta_{k_2 2,q}; \psi) = \zeta_{k_2 2,q} + \boldsymbol{x}'_{1it}\boldsymbol{\beta}_{12,q} + \boldsymbol{x}'_{2i}\boldsymbol{\beta}_{22,q}, \end{cases} \tag{15}$$

for $t = 1, 2, 3$, $i = 1, \ldots, n$, with $n = 9,021$, $k_1 = 1, \ldots, K_{1,q}$, and $k_2 = 1, \ldots, K_{2,q}$. In the model above, $\boldsymbol{x}_{1it}$ and $\boldsymbol{x}_{2i}$ denote the vectors of time-varying and time-constant covariates, respectively. In particular, the former includes ALE_{11}, SED_4, KESSM, IMD and the age variable (centered around the mean). In Tzavidis et al. (2016), the relationship between age and each of the outcomes of interest is modeled using a linear and a quadratic term, by this providing evidence of a more complex than a linear effect of such a covariate. Here, to allow for more flexibility and in order to estimate the shape of such relationships from the data, we introduce a polynomial spline with B-spline basis functions with 4 degrees of freedom. We have selected the degree of the spline for $q = 0.50$ using BIC and then we have kept the degree fixed at the other values of q for ease of comparison. We have tested also the other continuous covariates for nonlinearities, but we have found no evidence. On the other hand, $\boldsymbol{x}_{2i}$ includes maternal education (reference = no qualification), ethnicity (reference = non-white), gender (reference = female) and stratification (reference = advantaged stratum). To complete the model, we consider the following finite mixture probabilities

$$\begin{cases} \pi_{k_1 1,q} = \Pr(b_{i1,q} = \zeta_{k_1 1,q}), & k_1 = 1, \ldots, K_{1,q}, \\ \pi_{k_2 2,q} = \Pr(b_{i2,q} = \zeta_{k_2 2,q}), & k_2 = 1, \ldots, K_{2,q}, \end{cases}$$

for the two model equations.

For each response h and each M-quantile level $q \in (0, 1)$, we consider a varying number of mixture components ($K_{1,q} = K_{2,q} = 2, \ldots, 15$) and a multi-start strategy. A first deterministic solution is obtained by setting component probabilities to $\pi_{k_h h,q} = 1/K_{h,q}$, $k_h = 1, \ldots, K_{h,q}$, and fixed parameters in the longitudinal data models equal to the estimates from the corresponding homogeneous linear models. Component-specific random intercepts are then obtained by adding $K_{h,q}$ Gaussian quadrature locations to the corresponding (fixed) effect from the linear model above. For each value $K_{h,q}$, we derive $d(K_{h,q} - 1)$ random starting solutions from the deterministic ones by randomly perturbing model parameters ($d = 3$). For given h and q, the solution corresponding to the highest log-likelihood value is retained as the optimal one. To identify the optimal number of mixture components $K^{\star}_{h,q}$, we consider the BIC index; this has led us to select $K^{\star}_{1,q} = \{4, 4, 2\}$ for i-SDQ and $K^{\star}_{2,q} = \{9, 3, 2\}$ for e-SDQ, for $q = \{0.50, 0.75, 0.90\}$ respectively.

We report in Table 1 the estimates for fixed model parameters in equation (15). By looking at these results, we notice that the estimated parameters for the e-SDQ have stronger magnitude than those associated to the i-SDQ scores. Mother's education has a significant effect on the evolution of children's emotional and behavioral disorders; the higher the educational level, the smaller the SDQ scores and the stronger the absolute magnitude of these estimates as we move towards higher M-quantiles. Lower i-SDQ scores are generally observed for white children, with an effect that becomes stronger moving towards the right tail of the (conditional) response distribution. On the other hand, e-SDQ scores do not seem to be influenced by race, but at $q = 0.9$, with a positive and significant effect for such a variable. Males typically present higher internalizing and externalizing problems, while the stratification variable doesn't play a significant role, except when focusing on internalizing disorders and looking at the center of the distribution. As expected, children belonging to the disadvantaged or to the ethnic stratum present higher i-SDQ scores. As regards the effect of age, both scores reduce until children reach the mean age (about 5 years) and show another peak around the age of 7 during

Table 1: MCS data. uFMMQ$_p$ specification. Fixed parameter estimates for i-SDQ and e-SDQ scores at different M-quantiles.

	i-SDQ						e-SDQ					
	$q = 0.50$		$q = 0.75$		$q = 0.90$		$q = 0.50$		$q = 0.75$		$q = 0.90$	
	Est	Se	Est	Se	Est	Se	Est	Se	Est	Se	Est	Se
Intercept	2.990	0.269	3.859	0.270	4.714	0.336	6.443	0.285	7.136	0.390	7.599	0.524
Degree	-0.752	0.070	-0.915	0.099	-1.035	0.138	-1.349	0.104	-1.436	0.173	-1.794	0.206
Gcse	-0.473	0.068	-0.585	0.100	-0.615	0.133	-0.608	0.100	-0.629	0.139	-0.777	0.186
White	-0.300	0.059	-0.378	0.080	-0.506	0.151	0.076	0.082	0.255	0.162	0.468	0.195
Male	0.067	0.031	0.159	0.048	0.365	0.076	0.754	0.053	0.920	0.088	1.193	0.116
Ethnic St.	0.212	0.077	0.174	0.103	0.182	0.176	-0.074	0.107	-0.211	0.185	-0.109	0.238
Disadv. St.	0.103	0.039	0.068	0.064	0.048	0.100	0.119	0.067	0.185	0.118	0.283	0.151
BS(age)_1	0.395	0.491	0.527	0.638	-0.178	1.072	0.982	0.727	1.368	0.888	2.282	1.462
BS(age)_2	-1.713	0.322	-2.060	0.422	-1.668	0.701	-6.212	0.459	-6.984	0.553	-7.549	0.915
BS(age)_3	1.026	0.400	1.315	0.530	1.244	0.881	1.273	0.565	1.797	0.689	2.443	1.157
BS(age)_4	-0.400	0.139	-0.405	0.185	-0.248	0.308	-3.116	0.179	-3.335	0.223	-3.087	0.356
ALE_{11}	0.079	0.013	0.117	0.019	0.208	0.031	0.126	0.018	0.176	0.026	0.265	0.038
SED_4	0.075	0.018	0.084	0.025	0.112	0.038	0.126	0.027	0.185	0.040	0.313	0.050
KESSM	0.145	0.006	0.185	0.010	0.259	0.013	0.182	0.008	0.220	0.013	0.261	0.017
IMD	-0.024	0.007	-0.043	0.010	-0.076	0.018	-0.058	0.011	-0.077	0.018	-0.067	0.023

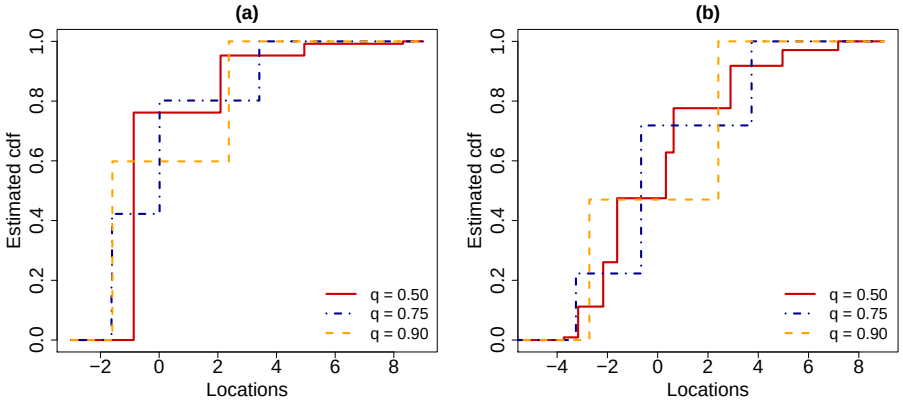
primary school and then decrease again. The effect of age is much stronger for e-SDQ rather than for i-SDQ, but this difference decreases as we move towards higher values of q. Figure 4 displays these effects for the multivariate within-between FMMQ model. For reasons of space, we have not included such a plot for the uFMMQ$_p$ specification, but the main features are similar.

Family and neighborhood risk factors play a central role in explaining the evolution of emotional and behavioral disorders over time. Adverse life events, family poverty measured by SED_4, and maternal depression are all positively associated with both i-SDQ and e-SDQ. The worse the socio-economic conditions, the higher the scores, and such effects are more pronounced at the right tail of the conditional response distribution. For IMD, we observe a negative and significant effect, whose absolute magnitude generally increases with q, especially when looking at the internalizing outcome. That is, higher scores are observed for children living in more deprived areas. For a clearer interpretation of these parameters, we report a graphical representation of the estimates for varying M-quantile levels in the Supplementary Material (see Figure **??**).

The results discussed so far are generally in line with those reported by Tzavidis et al. (2016). To allow for comparison, we have fitted the univariate FMMQ model on the same set of covariates used in that paper – i.e. using a quadratic term to model the effect of age instead of the B-spline – and we have observed a slight improvement in terms of efficiency. This is likely due to the higher flexibility of the finite mixture approach, where the subject-specific random intercepts are not constrained to a specific parametric form. Results are not reported for reasons of space, but they are available from the Authors upon request.

To conclude the analysis, we report in Figure 1 the cumulative density function of the random intercepts. From this figure, it is evident that, for both outcomes, the estimated

Figure 1: MCS data. uFMMQ$_p$ specification. Estimated cumulative density function of the discrete random intercepts for i-SDQ (a) and e-SDQ (b) scores at different M-quantiles.



distribution is quite far from symmetry and/or unimodality, thus making the Gaussian assumption rather inappropriate in this application. We also observe that locations for the e-SDQ score have higher variability than those for the i-SDQ. Last, looking at Figure 1, we notice that the probability of locations with a relatively larger value increases as we move from $q = 0.50$ to $q = 0.90$, thus highlighting how higher M-quantile levels identify (conditionally) more severe disorders.

5.2 The multivariate FMMQ with *pooled* effects

In this section, we extend the previous analysis and consider the proposed multivariate specification with *pooled* effects only. We considered the same linear predictors reported in equation (15). However, in this case, we explicitly model the dependence between children's internalizing and externalizing disorders by tying the discrete random intercepts in the two equations via the common prior probabilities

$$\pi_{k,q} = \Pr(b_{i1,q} = \zeta_{k1,q}, b_{i2,q} = \zeta_{k2,q})$$

For each M-quantile $q \in \{0.50, 0.75, 0.90\}$, we have run the estimation algorithm for a varying number of mixture components ($K_q = 2, \ldots, 15$) and considered the same multi-start strategy described in Section 5.1. For each q, we have selected the optimal model via the BIC index; this has lead us to select $K_q^\star = \{9, 5, 2\}$ components for $q = \{0.50, 0.75, 0.90\}$, respectively.

We report in Table 2 the fixed parameter estimates for demographic and socio-economic covariates. Based on these results, we may infer similar conclusions to those obtained through the univariate specification described in Section 5.1. However, a by-product of the proposed multivariate approach is the possibility to study the dependence structure between the outcomes of interest. This allows us to understand whether internalizing and externalizing disorders are related or not and investigate the nature of such a dependence. Do internalizing and externalizing symptoms occur jointly? Does one disorder exclude the other? As stated above, to answer such questions, we may rely on the estimated covariance between the random effects in the equations for the i-SDQ and the e-SDQ scores. We

Table 2: MCS data. mFMMQ$_p$ specification. Fixed parameter estimates for i-SDQ and e-SDQ scores at different M-quantiles.

	i-SDQ						e-SDQ					
	$q = 0.50$		$q = 0.75$		$q = 0.90$		$q = 0.50$		$q = 0.75$		$q = 0.90$	
	Est	Se	Est	Se	Est	Se	Est	Se	Est	Se	Est	Se
Intercept	3.097	0.165	4.077	0.231	4.745	0.385	6.376	0.279	7.290	0.300	7.637	0.533
Degree	-0.696	0.077	-0.851	0.101	-1.109	0.152	-1.242	0.139	-1.523	0.156	-1.724	0.224
Gcse	-0.422	0.070	-0.536	0.095	-0.635	0.147	-0.535	0.137	-0.699	0.138	-0.757	0.201
White	-0.335	0.068	-0.464	0.087	-0.431	0.141	0.107	0.091	0.263	0.132	0.417	0.203
Male	0.031	0.039	0.172	0.051	0.358	0.085	0.732	0.062	0.946	0.086	1.261	0.135
Ethnic St.	0.201	0.076	0.186	0.117	0.233	0.162	-0.061	0.121	-0.133	0.158	-0.078	0.226
Disadv. St.	0.074	0.044	0.068	0.061	0.049	0.102	0.098	0.077	0.220	0.101	0.309	0.164
BS(age)_1	0.294	0.526	0.208	0.686	0.381	1.057	1.008	0.771	1.216	0.896	2.686	1.460
BS(age)_2	-1.643	0.349	-1.946	0.449	-2.084	0.695	-6.188	0.483	-6.883	0.560	-7.755	0.873
BS(age)_3	0.971	0.446	1.303	0.561	1.898	0.889	1.204	0.596	1.646	0.698	2.568	1.098
BS(age)_4	-0.433	0.142	-0.496	0.199	-0.313	0.331	-3.081	0.184	-3.352	0.228	-2.984	0.360
ALE_{11}	0.075	0.013	0.120	0.020	0.237	0.031	0.120	0.020	0.148	0.025	0.289	0.040
SED_4	0.053	0.021	0.067	0.027	0.151	0.039	0.107	0.031	0.131	0.037	0.321	0.054
KESSM	0.135	0.007	0.193	0.010	0.275	0.014	0.169	0.009	0.193	0.011	0.274	0.018
IMD	-0.027	0.007	-0.058	0.011	-0.089	0.017	-0.049	0.013	-0.063	0.017	-0.081	0.026

have $\hat{\sigma}_{12,0.50} = 2.260, \hat{\sigma}_{12,0.75} = 2.815$, and $\hat{\sigma}_{12,0.90} = 3.635$, which suggest the presence of a positive association between the two scores. Furthermore, by focusing on the magnitude of $\hat{\sigma}_{12,q}$, we observe that such association increases with q. This translates into a higher chance for children to jointly have high i-SDQ and e-SDQ scores, especially when more severe symptoms are present (right tail of the distribution). The estimated covariance and correlation matrix of the random parameters at different M-quantile levels under the mFMMQ$_p$ specification are reported in the Supplementary Material; see Table **??**.

5.3 The multivariate FMMQ with auxiliary regression

To conclude the analysis, we consider the multivariate FMMQ based on the QP decomposition. As stated before, the aim is that of providing further insight into the MCS data and separating the effect of covariates' dynamics from effect associated to the corresponding mean levels. These are time-constant and may be possibly correlated with the random intercepts introduced in the model. For each M-quantile level $q \in \{0.50, 0.75, 0.90\}$, we considered the following equations:

$$\begin{cases} MQ_q(\text{i-SDQ}_{it} \mid \boldsymbol{x}_{it}, \zeta_{k1,q}; \psi) = \zeta_{k1,q} + \boldsymbol{x}'_{1it}\boldsymbol{\beta}_{11,q} + (\boldsymbol{x}_{2it} - \bar{\boldsymbol{x}}_{2i})'\boldsymbol{\beta}_{21,q} + \bar{\boldsymbol{x}}'_{2i}\boldsymbol{\delta}_{1,q} + \boldsymbol{x}'_{3i}\boldsymbol{\beta}_{31,q}, \\ MQ_q(\text{e-SDQ}_{it} \mid \boldsymbol{x}_{it}, \zeta_{k2,q}; \psi) = \zeta_{k2,q} + \boldsymbol{x}'_{1it}\boldsymbol{\beta}_{12,q} + (\boldsymbol{x}_{2it} - \bar{\boldsymbol{x}}_{2i})'\boldsymbol{\beta}_{22,q} + \bar{\boldsymbol{x}}'_{2i}\boldsymbol{\delta}_{2,q} + \boldsymbol{x}'_{3i}\boldsymbol{\beta}_{32,q}. \end{cases} \tag{16}$$

In this case, $\boldsymbol{x}_{1it}$ includes the set of B-spline basis functions for the variable age (centered around the overall mean), while $\boldsymbol{x}_{i2t}$ includes the remaining time-varying covariates: ALE_{11}, SED_4, KESSM, and IMD. These are centered around their individual means, $\bar{\boldsymbol{x}}_{2i}$. Last, $\boldsymbol{x}_{3i}$ denotes the vector of time-constant covariates (maternal education, ethnicity, gender, and the stratification variable). As stated in Section 4.1, the fixed parameters

in equation (16) have a different interpretation when compared to those discussed in the previous sections. In particular, the parameters $\boldsymbol{\beta}_{1h,q}$ and $\boldsymbol{\beta}_{2h,q}$ provide a measure of the direct effect that changes in individual covariates have on the dynamics of the SDQ scores. On the other hand, the parameters $\boldsymbol{\delta}_{h,q}$ and $\boldsymbol{\beta}_{3h,q}$ measure the impact of differences between children's global conditions (across all years of observation) on the M-quantile levels of the outcomes under investigation.

As before, we have run the estimation algorithm for varying K_q, that is $K_q = 2, \ldots, 15$, and considered a multi-start strategy to avoid local maxima. The optimal solution has been identified via the BIC index, which has led us to select a model with $K_q^\star = \{7, 4, 2\}$ components for $q = \{0.50, 0.75, 0.90\}$, respectively. Table 3 reports parameter estimates under the mFMMQ$_{wb}$ specification for the internalizing and the externalizing scores in the top and the bottom panel, respectively.

As expected, when looking at the estimated effects of time-constant covariates (maternal education, ethnicity, gender, and stratification), we derive similar conclusions to those obtained from the mFMMQ$_p$ approach discussed in Section 5.2. However, when focusing on the effect of time-varying covariates, it is evident that the QP decomposition provides a deeper understanding of the phenomenon of interest. We report in Figures 2-3 the estimated *between* (upper panel) and *within* (bottom panel) effects for the (time-varying) covariates associated to children's socio-economic conditions on the i-SDQ and the e-SDQ scores, respectively. By looking at these figures, we observe that the magnitude of the *within* effects estimated under mFMMQ$_{wb}$ approach are generally smaller than those obtained when considering *pooled* effects only (see Table 2) for both i-SDQ and e-SDQ scores. On the contrary, *between* parameters seem to have a higher impact on the distribution of the SDQ scores. That is, overall socio-economic conditions explain a higher portion of the SDQ variability than the corresponding dynamics. These findings highlight the potential dependence between sources of unobserved heterogeneity and the observed covariates, an issue which is quite common when dealing with observational studies and that makes our proposal an interesting modeling approach for analyzing the MCS data. Figure 4 shows the estimated shapes of the effect of age on the two scores at different levels of q. As noted when discussing the univariate models, both scores reduce until children reach the mean age (about 5 years) and show another peak around the age of 7 and then decrease again. This pattern is significant in particular for e-SDQ. In fact, the effect of age is much stronger for e-SDQ rather than for i-SDQ, but this difference decreases as we move towards higher M-quantile levels.

Last, parameter estimates for the mFMMQ$_{wb}$ specification leave substantially unchanged the inferential conclusions on the association between i-SDQ and e-SDQ scores discussed in Section 5.2. That is, children are likely to experience internalizing and externalizing disorders that are coherent and related to each other, especially when these are more severe. The estimated covariance and correlation matrix for the random effects at different M-quantile levels under the mFMMQ$_{wb}$ specification are reported in the Supplementary Material (Table **??**).

To conclude, we report in Table 5.3 the BIC values associated to the optimal model specifications in terms of number of mixture components for mFMMQ$_p$ and mFMMQ$_{wb}$, for the three M-quantile levels under investigation. Obviously, as it is typically done with nested modes, standard penalized likelihood criteria such as the BIC can also be exploited

Table 3: MCS data. mFMMQ$_{wb}$ specification. Fixed parameter estimates for i-SDQ (top panel) and e-SDQ scores (bottom panel) at different M-quantiles.

i-SDQ												
	$q = 0.50$				$q = 0.75$				$q = 0.90$			
	Between		Within		Between		Within		Between		Within	
	Est	Se	Est	Se	Est	Se	Est	Se	Est	Se	Est	Se
Intercept	2.858	0.185			3.841	0.227			4.370	0.359		
Degree	-0.630	0.072			-0.785	0.098			-1.038	0.141		
Gcse	-0.384	0.069			-0.511	0.093			-0.614	0.131		
White	-0.317	0.060			-0.436	0.087			-0.406	0.132		
Male	0.038	0.034			0.159	0.050			0.355	0.084		
Ethnic St.	0.195	0.078			0.158	0.115			0.184	0.169		
Disadv. St.	0.070	0.044			0.042	0.063			-0.023	0.111		
BS(age)$_1$			0.239	0.500			-0.090	0.662			0.222	1.020
BS(age)$_2$			-1.616	0.326			-1.771	0.437			-2.092	0.665
BS(age)$_3$			0.903	0.406			1.018	0.550			1.788	0.835
BS(age)$_4$			-0.467	0.144			-0.577	0.198			-0.548	0.312
ALE$_{11}$	0.101	0.023	0.054	0.016	0.190	0.036	0.077	0.022	0.395	0.055	0.107	0.033
SED$_4$	0.083	0.024	-0.043	0.031	0.084	0.036	-0.043	0.040	0.150	0.049	-0.047	0.058
Kessm	0.175	0.009	0.084	0.009	0.255	0.012	0.102	0.012	0.350	0.019	0.130	0.017
IMD	-0.022	0.008	-0.004	0.018	-0.058	0.012	-0.008	0.024	-0.089	0.021	-0.002	0.036
e-SDQ												
	$q = 0.50$				$q = 0.75$				$q = 0.90$			
	Between		Within		Between		Within		Between		Within	
	Est	Se	Est	Se	Est	Se	Est	Se	Est	Se	Est	Se
Intercept	6.042	0.308			6.861	0.355			7.006	0.508		
Degree	-1.174	0.138			-1.436	0.164			-1.622	0.224		
Gcse	-0.497	0.128			-0.668	0.149			-0.710	0.190		
White	0.130	0.106			0.273	0.146			0.409	0.191		
Male	0.720	0.062			0.932	0.088			1.279	0.129		
Ethnic St.	-0.089	0.140			-0.163	0.175			-0.107	0.226		
Disadv. St.	0.086	0.080			0.175	0.109			0.236	0.166		
BS(age)$_1$			0.917	0.769			1.085	0.882			2.822	1.436
BS(age)$_2$			-6.136	0.481			-6.863	0.560			-8.019	0.855
BS(age)$_3$			1.090	0.594			1.560	0.699			2.683	1.064
BS(age)$_4$			-3.121	0.188			-3.450	0.229			-3.206	0.343
ALE$_{11}$	0.181	0.042	0.080	0.022	0.283	0.059	0.082	0.025	0.539	0.079	0.078	0.038
x	0.167	0.052	-0.010	0.040	0.169	0.056	0.004	0.046	0.345	0.072	0.029	0.065
Kessm	0.221	0.014	0.116	0.011	0.263	0.019	0.131	0.012	0.343	0.024	0.166	0.018
IMD	-0.040	0.015	-0.026	0.024	-0.057	0.020	-0.035	0.027	-0.073	0.031	-0.018	0.044

Table 4: MCS data. BIC values for mFMMQ$_p$ and mFMMQ$_{wb}$ at different M-quantiles.

	$q = 0.50$	$q = 0.75$	$q = 0.90$
mFMMQ$_p$	203505.0	214423.8	237968.8
mFMMQ$_{wb}$	203392.2	214211.3	237434.1

Figure 2: MCS data. mFMMQ$_{wb}$ specification. Estimates of the *between* (upper panel) and the *within* effects (bottom panel) for i-SDQ scores at different M-quantiles. Panels (a)-(e): ALE; panels (b)-(f): SED; panels (c)-(g): Kessm; panels (d)-(h): IMD.

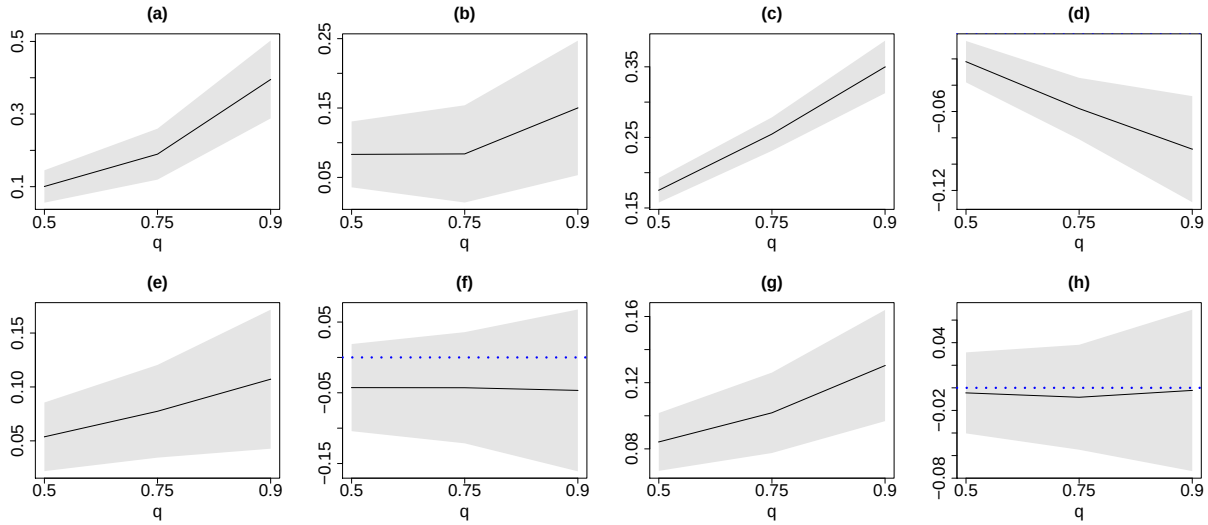


Figure 3: MCS data. mFMMQ$_{wb}$ specification. Estimates of the *between* (upper panel) and the *within* effects (bottom panel) for e-SDQ scores at different M-quantiles. Panels (a)-(e): ALE; panels (b)-(f): SED; panels (c)-(g): Kessm; panels (d)-(h): IMD.

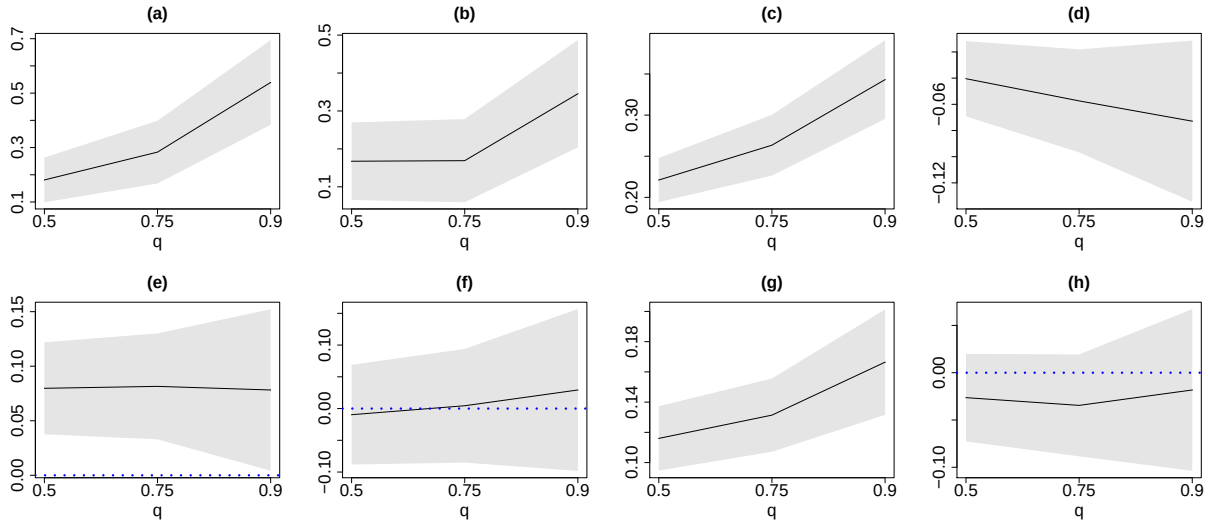
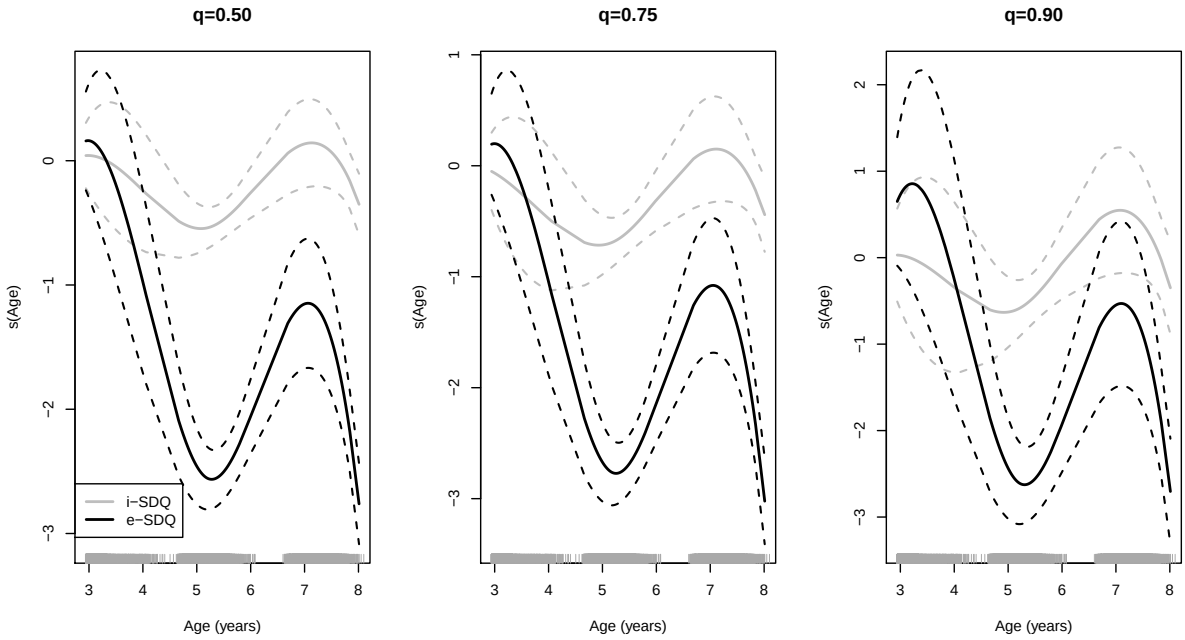


Figure 4: MCS data. mFMMQ$_{wb}$ specification. Estimates of the effect of age for i-SDQ and e-SDQ scores at different M-quantiles. Dashed lines represent 95% confidence bounds.



to identify the optimal specification in terms of explanatory variables to include in the linear predictor. By looking at the above table, it is evident that mFMMQ$_{wb}$ provides a better fitting to the data than mFMMQ$_p$ and in this sense should be preferred. Clearly, a similar strategy cannot be adopted to chose among the univariate (mFMMQ) and the multivariate specifications, as the former is not nested within the latter. In this sense, from our perspective, the choice should be driven by the need (or not) of studying the association between outcomes. If researchers think this is an important aspect to take into account and/or there is the suspect that the outcomes are associated, than the multivariate approach represents a natural way of proceeding. On the other side, if studying association is not a central matter and/or it is difficult to presume association between outcomes, then the univariate approach represents the simpler and preferable modeling strategy.

6 Conclusions

In this paper, we propose an M-quantile regression model for multivariate, continuous, longitudinal data by extending the finite mixture of M-quantile regression models proposed by Alfò et al. (2017) to a multivariate context. Correlated, subject-specific, random effects are used to account for dependence within the same response and association between responses observed on the same subject. We exploit the proposed mFMMQ regression model based on subject- and outcome-specific random intercepts only to analyze data on internalizing and externalizing SDQ scores from the Millennium Cohort Study. Clearly, a more complex specification, based on random intercepts and slopes could have been adopted, but for ease of interpretation, we decided not to follow this root. In this ap-

plication, we also handle potential endogeneity of the observed covariates by defining an auxiliary regression model in the spirit of Mundlak (1978).

The results of the MCS data analysis are in line with those discussed by Tzavidis et al. (2016). Together with a more flexible specification of the random effect distribution, some further insights are provided and these better characterize children behavioral and emotional disorders in terms of static levels and dynamic changes of socio-economic conditions. The proposed analysis provides evidence of two interesting aspects with respect to those provided by previous analysis: (i) children likely experience both internalizing and externalizing disorders, especially the more severe ones; (ii) behavioral and emotional disorders are mainly affected by overall children socio-economic conditions rather than by the corresponding variations over time.

Simulation results reported in the Supplementary Material prove the reliability of these findings, in terms of bias and efficiency. Here, it is also shown that such good properties also hold when model assumptions are not fulfilled and, above all, when the standard Gaussian assumption for the random effects does not hold, as for the present application. With respect to point (i) above, we may further highlight some important aspects of the proposed approach. First, it considers a potentially varying strength in the association between the analyzed outcomes, as the estimate of the random effect distribution is (M-)quantile-specific. This could be particularly appropriate to those phenomena that experience so-called *tail dependences*; see, e.g., Venter (1997). Moreover, the proposed approach may be used to define a classification of the analyzed individuals based on the estimated posterior distribution of component membership; this may be of great help for the prediction of the outcome(s) of interest in a more efficient way, when compared to parametric approaches; see Neuhaus et al. (2013).

The proposal may be extended in a number of directions. First, we may consider time-varying random parameters in a hidden Markov model perspective, to capture time-varying sources of unobserved heterogeneity. A further step may be to separately model dependence between and within outcomes, with the aim of enhancing model flexibility. Furthermore, in the spirit of quantile regression for discrete outcomes, we may also extend the proposed mFMMQ approach to deal with non-continuous responses, such as counts. Quantile regression for cross-sectional count data has been developed in the literature. It relies on the equivariance property of quantiles to monotone transformations and is based on the use of a jittering approach. In the present framework, two issues need to be addressed: (i) the use of M-quantiles means that properties holding for quantile regression are no longer valid; (ii) the presence of random effects makes the problem more demanding. A starting point could be the approach for count data developed by Tzavidis et al. (2015) and Dreassi et al. (2014) for M-quantile regression based on robust generalized linear models that uses quasi-likelihood.

Finally, when several conditional quantiles or M-quantiles are estimated, two or more functions can potentially 'cross over' at some point in the space defined by the covariates. This is called quantile crossing and may be due to model misspecification, collinearity, or to the presence of outlying values. The problem occurs because each conditional M-quantile function is independently estimated, i.e. without enforcing the property that at each value of $\boldsymbol{x}$, the M-quantiles of y are ordered by q. Also the multivariate FMMQ can suffer from M-quantile crossing problems. He (1997) propose a simple way of building this

restriction into fitted quantile regression lines by a-posteriori restricting them relative to the median regression line. This approach has been adapted to M-quantile regression by Pratesi et al. (2009) and Salvati et al. (2012). Alternatively, Frumento and Salvati (2020) impose a parametric structure that can stabilize the behavior of the estimated regression coefficients, especially in the tails, and alleviate the M-quantile crossing problem. Finally, in the context of quantile regression, Schnabel and Eilers (2013) define a surface, called a quantile sheet, on the domain of the independent variable and the probability q that is monotonically increasing when using moderate or large amounts of smoothing in the direction of q. All these solutions could be adapted to the multivariate FMMQ and explored to overcome the issue.

Acknowledgments

The work of Marino, Ranalli and Salvati has been developed under the support of the project *PRIN-SURWEY* (grant 2012F42NS8, Italy). Salvati gratefully acknowledge support by the projects *InGRID-2 Integrating Research Infrastructure for European expertise on Inclusive Growth from data to policy* (Horizon 2020 research and innovation programme, grant 730998) and *Progetto di Ricerca di Ateneo: "From survey-based to register-based statistics: a paradigm shift using latent variable models"* (grant PRA2018-9). The work of Alfò has been supported by the project *"Mixture and latent variable models for causal inference and analysis of socio-economic data"* (grant RBFR12SHVV, FIRB - Futuro in Ricerca).

Supplementary Materials

The Supplementary Material available for this paper at the Journal website includes the computational details for ML estimation and for the estimation of the covariance matrix of model parameter estimates. Some further insights into the MCS data analysis are also provided, together with the results of an intensive simulation study. Last, a computationally efficient algorithm for estimation and inference developed in `R` language from the authors is made available as part of the online Supplementary Material.

References

Abrevaya, J. and Dahl, C. (2008). The effect of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business and Economic Statistics*, 26:379–397.

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

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

Alfò, M., Salvati, N., and Ranalli, M. G. (2017). Finite mixtures of quantiles and M-quantile models. *Statistics and Computing*, 27:547–570.

Arellano, M. and Bonhomme, S. (2016). Nonlinear panel data estimation via quantile regressions. *Econometrics Journal*, forthcoming:C61–C94.

Bartels, B. (2008). Beyond fixed versus random effects: a framework for improving substantive and statistical analysis of panel, time-series cross-sectional, and multilevel data. *The Society for Political Methodology*, pages 1–43.

Bell, A. and Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, 3(1):133–153.

Bianchi, A., Fabrizi, E., Salvati, N., and Tzavidis, N. (2018). Estimation and testing in m-quantile regression with applications to small area estimation. *International Statistical Review*, 86:541–570.

Bianchi, A. and Salvati, N. (2015). Asymptotic properties and variance estimators of the M-quantile regression coefficients estimators. *Communications in Statistics - Theory and Methods*, 44:2416–2429.

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

Breckling, J., Kokic, P., and Lübke, O. (2001). A note on multivariate M-quantiles. *Statistics and probability letters*, 55:39–44.

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

Del Sarto, S., Marino, M. F., Ranalli, M. G., and Salvati, N. (2019). Using finite mixtures of M-quantile regression models to handle unobserved heterogeneity in assessing the effect of meteorology and traffic on air quality. *Stochastic Environmental Research and Risk Assessment*, pages 1–15.

Dreassi, E., Ranalli, M. G., and Salvati, N. (2014). Semiparametric M-quantile regression for count data. *Statistical Methods in Medical Research*, 23(6, SI):591–610.

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

Frumento, P. and Salvati, N. (2020). Parametric modelling of m-quantile regression coefficient functions with application to small area estimation. *Journal of the Royal Statistical Society. Series A*, 183.

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

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

Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *Journal of child psychology and psychiatry*, 38:581–586.

Grilli, L. and Rampichini, C. (2011). The role of sample cluster means in multilevel models: A view on endogeneity and measurement error issues. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 7(4):121.

Hallin, M., Paindaveine, D., and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From l 1 optimization to halfspace depth. *The Annals of Statistics*, 38:635–669.

Harville, D. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics*, 4:384–395.

He, X. (1997). Quantile curves without crossing. *American Statistician*, 51.

Huber, P. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.*, 35:73–101.

Jones, M. C. (1994). Expectiles and M-quantiles are quantiles. *Statistics and Probability Letters*, 20:149–153.

Kessler, R. and Mroczek, D. (1992). An update of the development of mental health screening scales for the us national health interview study. *Ann Arbor: University of Michigan, Survey Research Center of the Institute for Social Research*.

Kneib, T. (2013). Beyond mean regression. *Statistical Modelling*, 13:275–303.

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

Krishnakumar, J. (2006). Time invariant variables and panel data models: A generalised frisch–waugh theorem and its implications. *Contributions to Economic Analysis*, 274:119–132.

Kulkarni, H., Biswas, J., and Das, K. (2019). A joint quantile regression model for multiple longitudinal outcomes. *AStA Adv Stat Anal*, 103:453–473.

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

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

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

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

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

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

Marino, M. F., Tzavidis, N., and Alfó, M. (2016). Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequences. *Statistical Methods in Medical Research*. doi: 10.1177/0962280216678433.

Mu, Y. and He, X. (2007). Power transformation toward a linear regression quantile. *Journal of the American Statistical Association*, 102(477):269–279.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46:69–85.

Neuhaus, J. M. and Kalbfleisch, J. D. (1998). Between-and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, pages 638–645.

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

Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847.

Petrella, L. and 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.

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

Pratesi, M., Ranalli, M. G., and Salvati, N. (2009). Nonparametric M-quantile regression using penalised splines. *Journal of Nonparametric Statistics*, 21(3):287–304.

Salvati, N., Tzavidis, N., Pratesi, M., and Chambers, R. (2012). Small area estimation via m-quantile geographically weighted regression. *Test*, 21.

Schnabel, S. and Eilers, P. (2009). Optimal expectile smoothing. *Computational Statistics & Data Analysis*, 53:4168–4177.

Schnabel, S. K. and Eilers, P. H. (2013). Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, 97:77–87.

Smith, L., Fuentes, M., Gordon-Larse, P., and Reich, B. (2015). Quantile regression for mixed models with an application to examine blood pressure trends in China. *Annals of Applied Statistics*, 9:1226–1246.

Sobotka, F., Kauermann, G., Schulze Waltrup, L., and Kneib, T. (2013). On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, 23:135–148.

Sobotka, F. and Kneib, T. (2012). Geoadditive expectile regression. *Computational Statistics & Data Analysis*, 56:755–767.

Spiegel, E., Sobotka, F., and Kneib, T. (2017). Model selection in semiparametric expectile regression. *Electron. J. Statist.*, 11(2):3008–3038.

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

Tzavidis, N., Ranalli, M. G., Salvati, N., Dreassi, E., and Chambers, R. (2015). Robust small area prediction for counts. *Statistical Methods in Medical Research*, 24(3, SI):373–395.

Tzavidis, N., Salvati, N., Schmid, T., Flouri, E., and 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.

Venter, G. (1997). Tails of copulas. In *Proceedings of the Casualty Actuarial Society, vol. 89*. Graphos. pages 68–113.

Verbeke, G., Fieuws, S., Molenberghs, G., and Davidian, M. (2014). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*, 23:49–52.

Waldmann, E., Sobotka, F., and Kneib, T. (2017). Bayesian regularisation in geoadditive expectile regression. *Statistics and Computing*, 27(6):1539–1553.

Waltrup, L. S., Sobotka, F., Kneib, T., and Kauermann, G. (2015). Expectile and quantile regression?david and goliath? *Statistical Modelling*, 15(5):433–456.

Weidner, M. and Moon, H.-R. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33:158–195.

Wu, M. and Bailey, K. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear models. *Biometrics*, 45:939–955.

Wu, M. and Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44:175–188.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, 54:437–447.