Fall 2021

Marginally Interpretable Models and Multilevel Models For Quantile Regression With Random-Effects

Nahid Sultana Sumi

Marginally Interpretable Models and Multilevel Models for Quantile Regression with random-effects

by

Nahid Sultana Sumi

Bachelor of Science
University of Dhaka, 2006

Master of Science
University of Dhaka, 2008

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Biostatistics

The Norman J. Arnold School of Public Health

University of South Carolina

2021

Accepted by:

Marco Geraci, Major Professor

James W Hardin, Committee Member

Alexander C McLain, Committee Member

Alessio Farcomeni, Committee Member

Tracey L. Weldon,, Interim Vice Provost and Dean of the Graduate School

Dedication

To My Late Father, Abdus Shahid and To My Mother, Nazmun Nahar.

To My Husband, Akhtar and Daughter, Alethea; their understanding, love and support are very much appreciated.

Acknowledgments

I acknowledge the Centre for Longitudinal Studies, Institute of Education for the use of Millennium Cohort Study (MCS) data; the UK Data Service for making them available; the MRC Centre of Epidemiology for Child Health (Grant reference G0400546), Institute of Child Health, University College London for creating the accelerometer data resource which was funded by the Wellcome Trust (grant reference 084686/Z/08/A). The institutions and funders acknowledged bear no responsibility for the analysis or interpretation of these data. The data is used in demonstration of applications of the methods proposed in Chapters 2, and 3 of this research work.

Abstract

The quantile regression model is an active area of statistical research that has received a lot of attention. This complements the most widely used statistical tool, that is, mean regression analysis. Quantile regression analysis It has become more flexible because of its properties that include no assumption on the distribution of the response variable, equivalent to monotone transformations, and robustness to outliers. However, regression analysis offers methodological challenges if the observations are not independent. Cluster, multilevel, and repeated measures (longitudinal data) designs introduce such dependence. The correlation between observations on the same units or clusters should be accounted for to obtain correct inferences. The mixed-effects model has been used to analyze such complex sampling designs. As an increased research interest in quantile regression, in the last few years, several approaches to quantile regression for dependent data have been proposed. These approaches consist of distribution-free and likelihood-based methods. The mixed-effects model for quantile regression in existing literature can be applied to 2-levels. However, in some situations, researchers may need to apply the quantile regression for the mixed-effects model where more than 2-levels are of interest (frequentist approach). Conditional modeling of quantiles may not be ideal when the focus is on the marginal quantile effects. A way to make inferences on marginal effects is to adjust estimates from a conditional model when the latter does not naturally lead to marginal interpretations. This is the case in quantile regression with random effects and interest often lies in marginal or population-averaged effects rather than conditional effects. Given the limitations in the existing body of the multilevel linear mixed-effects model, this re-

search work aims to present a mixed model for quantile regression with the potentials of filling the noted literature gaps.

In Chapter 2 of this dissertation, we presented a marginally interpretable model for quantile regression with random effects. We discussed the derivative free numerical integral algorithm with Gauss-Hermite quadrature to the proposed model parameter estimations. The performance of the parameter estimation methods is studied through statistical simulations for the proposed model.

Given the limitation in literature, we proposed a 3-level mixed-effects model for quantile regression in Chapter 3. The proposed model is an extension of 2-level mixed effects model available in literature. Simulation studies were performed to evaluate the proposed model. In addition to the extensive simulation studies, we have demonstrated applications of all proposed models using Millennium Cohort Study (MCS) data.

Preface

I would like to express sincere gratitude to my advisor Dr. Marco Geraci for his continuous guidance, encouragement, and support during my dissertation. I particularly, would like to appreciate Dr. James W Hardin, for his guidance and suggestions. I would like to thank other committee members Dr. Alexander C, Dr. Alesso Farcomeni for their insights and comments on this dissertation. It has been a pleasant experience to study in the Department of Epidemiology and Biostatistics at the University of South Carolina, Columbia. Finally, I would like to thank to my parents, my husband Akhtar, my daughter Alethea and all of my other family members, for their unconditional love, encouragement and support.

Table of Contents

List of Tables

Chapter 1

Introduction

1.1 Background

Linear regression is one of the most popular and useful statistical tools. It models an outcome variable as a function of other variables called covariates. Classical linear regression analysis studies how the mean of the conditional distributions of the outcome changes when covariates change. Mean regression cannot give a complete picture of the relationship between the outcome and the set of covariates when the relationship is more complex than location-shift, that is, when the covariates affect the scale and the shape of the outcome's distribution. Moreover, one may be interested in studying the entire distribution or the lower/upper tail of a distribution. A regression analysis that complements the mean regression analysis is quantile regression (QR).

QR is able to uncover complex relationships since the entire conditional distribution of the outcome is modeled as a function of the covariates. QR also provides a way to assess the relationship between the covariates and any quantile of the response variable distribution. The estimation of quantile regression requires no assumption on the distribution of the response variable (Koenker and Bassett 1978; Bassett and Koenker 1978). It also has the properties of equivalent to monotone transformations and robustness to outliers(Huber 1981). A lot of work has been done using this statistical tool since it is popularized by Koenker and Bassett (1978). However, in regression analysis, lack of independence between observations offers methodological challenges.

Correlated or dependent data arise in a wide variety of studies. A number of sampling designs such as cluster, multilevel and repeated measures (longitudinal data) introduce such dependence. The correlation between observations on same units or clusters should be accounted for to obtain correct inferences. To analyze data from complex sampling designs, mixed effects models (also known as multilevel or random-effects models) have been used frequently because of their flexibility.

Several approaches to QR for dependent data have been proposed in the last few years. Koenker (2004) estimated quantile functions with subject-specific fixed effects, where variability in the estimation process is controlled by some form of shrinkage that requires a suitable choice of penalization. Geraci and Bottai (2007) proposed linear QR model with a subject-specific random intercepts that accounts for within group correlation by using asymmetric Laplace density. Models that incorporate random slopes include Geraci and Bottai (2014). Other approaches to QR with correlated data are given by Jung (1996), Liu and Bottai (2009), Galvao and Montes-Rojas (2010), Yuan and Yin (2010), Kim and Yang (2011), Fu and Wang (2012), Farcomeni (2012), and Li, Dowling, and Chappell (2015). Currently there are no proposal for mixed effects QR models with more than 2 levels.

Conditional modeling of quantiles (with either fixed or random cluster-specific effects) may not be ideal when the focus is on the marginal quantile effects. Marginal modeling of quantiles of correlated responses provides alternative approach. For example, Wang and Zhu (2011) defined an empirical likelihood under the generalized estimating equation (GEE) framework. Jung (1996) preserved marginal effects by incorporating correlated errors in a quasi-likelihood model. Another way to make inference on marginal effects is to adjust estimates from a conditional model when the latter does not naturally lead to marginal interpretations. This is the case in generalized linear mixed model (GLMM) (Zeger, Liang, and Albert 1988; Lesaffre et al. 2000; Gory, Craigmile, and MacEachern 2016) and QR with random effects.

In the rest of this chapter we will give a brief description of mean and quantile mixed-effects models. We will briefly consider conditional and marginal interpretation in these models. Further we will present multilevel data from the Millennium Cohort Study (MCS), a UK population-based longitudinal, multipurpose survey.

1.2 GLMM

Generalized linear models (GLMs) represent a class of fixed effects regression models for several types of dependent variables (i.e., continuous, dichotomous, counts). Nelder and Wedderburn (1972) described how a collection of seemingly disparate statistical techniques could be unified, that is, the outcome is assumed to fall within the exponential family of distributions in 'generalized linear model' which are described in great detail by McCullagh and Nelder (1989). Linear regression, logistic regression, and Poisson regression are common Generalized linear models (GLMs). These are fixed effect models.

Fixed effects models, which assume that all observations are independent of each other, are not appropriate for analysis of correlated data structures such as clustered and/or longitudinal data. In clustered designs, subjects are nested within larger units, for example, schools, hospitals, and so on. In longitudinal designs, repeated observations are nested within subjects. These are often referred to as multilevel (Goldstein 1995) or hierarchical (Raudenbush and Bryk 2002) data in which the level-1 observations (subjects or repeated observations) are nested within the higher level-2 observations (clusters or subjects).A three-level design could have repeated observations (level-1) nested within subjects (level-2) who are nested within clusters (level-3). Higher levels are also possible. For analysis of such multilevel data, random cluster effects can be added into the regression model to account for the correlation of the data. The resulting model is a mixed model including the usual fixed effects for the regressors plus the random effects. Mixed models for continuous normal outcomes

have been extensively developed since the seminal paper by Laird and Ware (1982). For non-normal data, there have also been many developments. Many of these developments fall under the rubric of generalized linear mixed models (GLMMs), which extend GLMs by the inclusion of random effects in the predictor. The generalized linear mixed model (GLMM) involves fiixed effects and normally distributed random model effects as in the "traditional" mixed model, but the error distribution is more general. Errors may have any distribution belonging to the exponential family.

For a two-level nested designs, we consider n_i is the number of observations nested within each of $i = 1, 2, \ldots, M$ clusters. The data is in the form of $\left(\boldsymbol{x}_{ij}^T, \boldsymbol{z}_{ij}^T, y_{ij}\right)$, for $j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, M$, and $N = \sum_{i=1}^{M} n_i$. Here, $\boldsymbol{x}_{ij}^T$ is the jth row of a known $n_i \times p$ matrix $\boldsymbol{X}_i$, $\boldsymbol{z}_{ij}^T$ is the jth row of a known $n_i \times q$ matrix $\boldsymbol{Z}_i$, and y_{ij} is the jth observation of the response vector $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$ for the the ith cluster. The generalized linear mixed model is given by

$$
\begin{aligned}
f(\boldsymbol{y}_i|\boldsymbol{u}_i) &= \exp\left\{[\boldsymbol{y}_i\gamma_i - b(\gamma_i)]/\alpha - c(\boldsymbol{y}_i, \alpha)\right\} \\
E[\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{X}_i] &= \mu_i,\ g(\mu_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{u}_i
\end{aligned} \tag{1.1}
$$

where the distribution of the n_i dimensional response vector $\boldsymbol{y}_i$ for the ith group is from an exponential family; where γ_i and α are scalars, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots \beta_p)^T$. The $q \times 1$ random effects vector $\boldsymbol{u}_i$ are assumed to be distributed according to $\boldsymbol{u}_i \sim f(u_i)$. We assume independence between the clusters. We also assume that, conditionally on the random effects, observations within the same cluster are independent. A link function, $g(.)$ is applied to the conditional mean of $\boldsymbol{y}_i$ given $\boldsymbol{u}_i$ to obtain the conditional linear predictor. The linear predictor is assumed to consist of two components, the fixed effects portion, described by $\boldsymbol{X}_i\boldsymbol{\beta}$ and the random effects portion, $\boldsymbol{Z}_i\boldsymbol{u}_i$, for which a distribution is assigned to $\boldsymbol{u}_i$.

The random effects are usually assumed to be independently distributed as $N(0, \sigma_u^2)$ in case of a GLMM. The parameter σ_u^2 indicates the variance in the population distribution, and therefore the degree of heterogeneity of clusters. GLMMs are often

referred to as conditional models in contrast to the marginal generalized estimating equations (GEE) models because of the expectation of the conditional distribution of the outcome given the random effects. Different types of link functions are used for different types of outcomes. For example, we use a logistic link function for binary outcomes. For a count variable, we can use a log link function.

The single level random effect model in (1.1) can be extended to two (multiple) levels (Pinheiro and Bates 2006). In the case of two nested levels of random effects the response vectors at the innermost level of grouping are written $\boldsymbol{y}_{ij}$, $i = 1, 2, \ldots, M; j = 1, 2, \ldots, M_i$ where M is the number of first-level groups and M_i is the number of second-level groups within first-level group i. The length of $\boldsymbol{y}_{ij}$ is n_{ij} and $N = \sum_{i=1}^{M} \sum_{j=1}^{M_i} n_{ij}$. Here, $\boldsymbol{x}_{ijk}^T$ is the kth row of a known $n_{ij} \times p$ matrix $\boldsymbol{X}_{ij}$, using level-1 random effects $\boldsymbol{u}_i^{(1)}$ of length q_1 and level-2 random effects $\boldsymbol{u}_{ij}^{(2)}$ of length q_2 with corresponding matrices $\boldsymbol{Z}_{i,j}$ of size $n_{ij} \times q_1$, and $\boldsymbol{Z}_{i,j}$ of size $n_{ij} \times q_2$, we can write the link function as

$$g(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{i,j}\boldsymbol{u}_i^{(1)} + \boldsymbol{Z}_{ij}\boldsymbol{u}_{ij}^{(2)} \tag{1.2}$$

The level-1 random effects $\boldsymbol{u}_i^{(1)}$ are assumed to be independent for different i, the level-2 random effects $\boldsymbol{u}_{ij}^{(2)}$ are assumed to be independent for different i or j and to be independent of the level-1 random effects. The two-level GLMM can be easily extended to include multiple random effects known as multilevel GLMM. The multi levels random effects models follow the same general pattern as two level random effects models. For a k level nested design, we write the link function as

$$g(\mu_{ij\ldots k}) = \boldsymbol{X}_{ij\ldots k}\boldsymbol{\beta} + \boldsymbol{Z}_{i,j\ldots k}\boldsymbol{u}_i^{(1)} + \boldsymbol{Z}_{ij,\ldots k}\boldsymbol{u}_{ij}^{(2)} + \cdots + \boldsymbol{Z}_{ij,\ldots k}\boldsymbol{u}_{ij\ldots k}^{(k)} \tag{1.3}$$

where all the k random effects are assumed to be independent for k different levels and to be independent of each other.

In a population average model, responses of three level nested designs $\boldsymbol{y}_{ij}$ are modeled without explicitly modeling the differences between clusters. The dependency

induced by clustered data is dealt by assuming a form for within cluster covariance matrix and this form is assumed to be the same for all clusters.The population average models are essentially averaging data over heterogeneous clusters, and thus are also sometimes referred to as marginal models. Marginal models will be taken to refer to models that are not conditional on unobserved, random effects.

1.2.1 Interpretation of GLMM

The interpretation of GLMMs is similar to generalized linear models; however, there is an added complexity because of the random effects. The interpretation is based on cluster-specific or subject-specific. Although the cluster-specific model seems to provide the more unified approach, parameter interpretation in these models is difficult. The cluster-specific model pre-supposes the existence of latent risk groups indexed by $\boldsymbol{u}_i$ (see(1.1)), and parameter interpretation is with reference to these groups. No empirical verification of this statement can be available from the data unless the latent risk groups can be identified. In order to interpret we must keep the cluster-specific latent effect $\boldsymbol{u}_i$ the same in (1.1). Typically fixed effects parameters are interpreted by considering the effect for a unit change in the predictor variable holding all other predictors fixed. In cluster specific regression models, this includes also holding random effect constant. Under the mixed effects model the interpretation of regression coefficients is conditional on the value of the random effect. As for example, see the model below

$$E[Y|x, \boldsymbol{u}] = \exp(\beta_0 + u_{0i} + (\beta_1 + u_{1i})x_i)$$

e^{β_1} is the relative risk between two populations with the same $\boldsymbol{u}$ but whose x values differ by one unit, that is:

$$\exp(\beta_1) = \frac{E[Y|x, \boldsymbol{u}]}{E[Y|x-1, \boldsymbol{u}]}$$

An alternative interpretation is to say that it is the expected change between two "typical individuals", that is, individuals with random effects, $\boldsymbol{u} = \boldsymbol{0}$.

In contrast to cluster-specific models, the fixed effects parameters in population average models pertain to the average Level 1 unit in the whole population, see (1.1). The fixed effect parameters in this model are interpreted by considering the effects for a unit change in the predictor holding all other observed effects constant. This interpretation does not require that whether subjects are from the same cluster effects or not. Marginal models are obtained by integrating over the random effects. Suppose in a marginal model, we have

$$E[Y|x] = \exp(\beta_0 + \beta_1 x)$$

in which case e^{β_1} is the change in the average response when we increase x by 1 unit in the population under consideration.

1.3 Quantile Regression and Linear Quantile Mixed Models

Let us define quantile regression (QR) at first, which was introduced by Koenker and Bassett (1978), that estimates and conducts inference about conditional quantile functions. For any $0 < \tau < 1$, the τth quantile of a real valued random variable Y is defined as

$$Q(\tau) = \inf\{y : F(y) \geq \tau\},$$

where $F(y) = \text{Prob}(Y \leq y)$ is the distribution function of Y. If (x_i^T, y_i), $i = 1, 2, \ldots, n$ is an independent random sample from some population, where x_i is a known $p \times 1$ vector of regressors and y_i is a scalar response variable with conditional cumulative distribution function F_y and its shape is unspecified, then the linear conditional quantile function, $Q_{y_i}(\tau|x_i) = x_i^T \beta_\tau$, can be estimated by solving the optimization problem

$$\hat{\beta}_\tau = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(\boldsymbol{y}_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau \right) \tag{1.4}$$

where $\rho_\tau(r) = r\{\tau - I(r < 0)\}$ is the loss function with r being a real number and $I(.)$ is the indicator function. The solution $\hat{\beta}_\tau$ is usually called the τth regression quantile.

Asymmetric Laplace Distribution (ALD) provides a natural link between minimization of equation (1.4) and the maximum likelihood estimation. A random variable T follows an ALD if its corresponding probability density is given by

$$p(t) = \frac{\tau(\tau-1)}{\sigma_\tau} \exp\left\{-\frac{1}{\sigma_\tau}\rho_\tau(t-\mu_\tau)\right\}$$

where $\mu_\tau \in \mathbb{R}$ and $\sigma_\tau > 0$ are location and scale parameters, respectively. Let $\mu_i = \boldsymbol{x}_i^T\boldsymbol{\beta}$, then the likelihood of n independent observations is written as (for any given τ)

$$L \propto \frac{1}{\sigma^n} \exp\left\{-\sum_{i=1}^{n} \rho_\tau\Big(\frac{\boldsymbol{y}_i - \boldsymbol{x}_i^T\boldsymbol{\beta}}{\sigma}\Big)\right\} \tag{1.5}$$

The maximization of the likelihood in (1.5) with respect to the parameter β is equivalent to the minimization of the objective function in (1.4).

Linear mixed-effects models that include fixed effects to account for covariate effects and random effects to explain heterogeneity between the subjects provides efficient analysis of hierarchical or nested data. Quantile regression has also been extended to the linear mixed-effects model (Geraci and Bottai 2014). The parameters of quantile regression for the linear mixed-effects model achieve an unbiased estimation due to the fact that the correlation of the measurements from the same subjects are adequately accounted for (Geraci and Bottai 2007). Geraci and Bottai (2014) utilized a combination of Gaussian quadrature approximations and non-smooth optimization algorithms for the estimation of the fixed regression coefficients and of the random effects' covariance matrix, and this method was implemented in R package `lqmm` (Geraci 2014). Other available approaches have been reviewed by Marino and Farcomeni (2015).

We give a details of the linear quantile mixed model (LQMM) here. The LQMM for two level clustered data (Koenker 2004; Geraci and Bottai 2014) can be written as follows

$$Q_{y_{ij}|u_i,x_{ij},z_{ij}}(\tau) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_\tau + \boldsymbol{z}_{ij}^T\boldsymbol{u}_{\tau,i} \tag{1.6}$$

for $\tau \in (0,1)$; $j = 1,2,\ldots,n_i$; $i = 1,2,\ldots,M$ and $\boldsymbol{\beta}_\tau = (\beta_{\tau,0},\beta_{\tau,1},\ldots\beta_{\tau,p})^T$. The $q\times 1$ random effects vector $\boldsymbol{u}_{\tau,i}$ are assumed to be distributed according to $p(\boldsymbol{u}_{\tau,i}|\boldsymbol{\Psi}_\tau)$ (no specific distribution), where $\boldsymbol{\Psi}_\tau$ is a $q\times q$ covariance matrix. The random effects vectors depends on τ through $\boldsymbol{\Psi}_\tau$ and are assumed to be independent for different i. The objective function to be minimized is given as

$$\sum_{i=1}^{M}\sum_{j=1}^{n_i}\rho_\tau\left(\boldsymbol{y}_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_\tau - \boldsymbol{z}_{ij}^T\boldsymbol{u}_{\tau,i}\right) \tag{1.7}$$

The minimization of (1.7) is equivalent to fitting a LQMM where the responses, conditionally on the random effects, is assumed to follow the asymmetric Laplace distribution (Geraci and Bottai 2007; Geraci and Bottai 2014). So, the responses y_{ij}, conditionally on the random effects, are independently distributed as ALD with location and scale parameters given by $\mu_{\tau,ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta}_\tau + \boldsymbol{z}_{ij}\boldsymbol{u}_{\tau,i}$ and σ_τ. The skew parameter τ is fixed and defines the quantile levels. The authors also assumed that the $q\times 1$ random effects vector $\boldsymbol{u}_{\tau,i}$, $i = 1,2,\ldots,M$ are assumed to be independent of model's error term with mean zero and $q\times q$ variance-covariance matrix $\boldsymbol{\Psi}_\tau$. Let us define $\boldsymbol{y} = \left(\boldsymbol{y}_1^T,\boldsymbol{y}_2^T,\ldots,\boldsymbol{y}_M^T\right)^T$; $\boldsymbol{u} = \left(\boldsymbol{u}_1^T,\boldsymbol{u}_2^T,\ldots,\boldsymbol{u}_M^T\right)^T$. According to Geraci and Bottai 2014, the joint density of $(\boldsymbol{y},\boldsymbol{u})$ based on M clusters for τth LQMM is given by

$$p(\boldsymbol{y},\boldsymbol{u}) = \left\{\frac{\tau(\tau-1)}{\sigma_\tau}\right\}^N\prod_{i=1}^{M}\exp\left\{-\frac{1}{\sigma_\tau}\sum_{j=1}^{n_i}\rho_\tau(y_{ij}-\mu_{\tau,ij})\right\}p(\boldsymbol{u}_{\tau,i})$$

A combination of Gaussian quadrature and non-smooth optimization is used for estimating LQMMs (Geraci and Bottai 2014).The approximated marginal (over the random effects) log-likelihood using the rule is given is

$$L(\boldsymbol{\beta}_\tau,\sigma_\tau,\boldsymbol{\Psi}|\boldsymbol{y}) = \sum_{i=1}^{M}\log\left\{\sum_{k_1=1}^{K}\cdots\sum_{k_q=1}^{K}p(\boldsymbol{y}_i|\boldsymbol{v}_{k_1,\ldots,k_q})\prod_{l=1}^{q}w_{k_l}\right\} \tag{1.8}$$

where v_{k_l} and w_{k_l}, $k_l = 1, \ldots, K; l = 1, \ldots, q$ denote the abscissas and weights of the Gaussian quadrature, respectively, with $v_{k_1,\ldots,k_q} = \left(v_{k_1}, \ldots, v_{k_q}\right)^T$. Prediction of the random effects for LQMMs was carried out via best linear prediction (BLP).

1.3.1 Interpretation of LQMM

Let us consider the following random intercept model as defined in Geraci and Bottai (2014),

$$y|u = \mu + Zu + \epsilon;\ u \sim N(0, \psi_u^2 I), \epsilon \sim N(0, \psi^2 I) \text{ and } u \perp \epsilon$$

where Z is a block-diagonal matrix with diagonal blocks given by vectors of ones and I be an identity matrix of appropriate dimensions. Here μ is the mean effect at the population level, ψ_u^2 is a measure of the dispersion of the cluster-specific random effects and proportional to the ICC, and ψ^2 is the error's variance. The scale parameter σ does not have, in general, a straightforward interpretation since the use of the Laplace distribution for the conditional response follows from the convenience of manipulating a likelihood rather than from the observation that the data is indeed Laplacian (Geraci and Bottai 2014). Here, since each cluster has a conditional distribution $N(\mu + u_i, \psi^2 I)$, so τth quantile is $q_{i.}(\tau) \equiv \mu + u_i + \psi\Phi^{-1}(\tau)$ for all $j = 1, \ldots n_i$, where Φ is the standard normal cumulative distribution function. The $y_{ij}|u_i$'s are independent. Thus, the τth sample quantile $\hat{q}_{i.}(\tau)$ is an estimator of $q_{i.}(\tau)$. So, for large n_i,

$$\hat{q}_{i.}(\tau) \sim N\Big(q_{i.}(\tau), \frac{\tau(1-\tau)}{n_i[p_{y_{i.}|u_{i.}} q_{i.}(\tau)]^2}\Big)$$

where p is continuous and $o < \tau < 1$. For M such clusters, the average is $\overline{q_{i.}}(\tau) = \frac{1}{M}\sum_{i=1}^{M} q_{i.}(\tau)$, and for large M the distribution of $\frac{1}{M}\sum_{i=1}^{M} \hat{q}_{i.}(\tau)$ follows normal with mean $\frac{1}{M}\sum_{i=1}^{M} q_{i.}(\tau)$ and variance $\frac{1}{M^2}\sum_{i=1}^{M} \frac{\tau(1-\tau)}{n_i[p_{y_{i.}|u_{i.}} q_{i.}(\tau)]^2}$. Thus, the fixed effects can be interpreted as a weighted average approximation of the population quantiles they target.

1.4 Millennium Cohort Study (MCS)

The Millennium Cohort Study (MCS) is an observational, multidisciplinary cohort study of children born in the United Kingdom (UK), which is nationally representative of the total UK population at baseline and conducted by the Centre for Longitudinal Studies (CLS) at the University of London. This ongoing study consists of a sample of all children born between September 2000 and January 2002, alive and living in England, Scotland, Wales, or Northern Ireland at age 9 months, and eligible to receive child benefit at that age (Plewis et al. 2007). In the first sweep known as MCS1 that occurred at age 9 months, 18,552 families (18,827 children) were recruited to the cohort and information was collected on 18,818 children. Subsequent sweeps have occurred when the children were aged 3 (MCS2), 5 (MCS3), 7 (MCS4), 11 (MCS5), 14 (MCS6), 17 (MCS7) and 22 (MCS8) years. On a broad range, information of demographic, behavioral, developmental, health, and parental socioeconomic characteristics were recorded at first sweep of MCS (Connelly and Platt 2014). Trained interviewers were required to interview the subjects with information provided by mothers. Ethical approval for data collection at each sweep of MCS and accelerometer studies has been granted. Approval for seasonal accelerometer and calibration studies was granted by the University College London Research Ethics Committee (Griffiths et al. 2013b). The MCS data are freely available to the researchers under standard access conditions through the UK Data Service (http://ukdataservice.ac.uk) and the MCS website provides detailed information on the study (http://www.cls.ioe.ac.uk/mcs). As a multidisciplinary study, MCS collects a rich range of information regarding the experiences and outcomes of the included children and their families. The MCS collects data on socio-demographic characteristics, child development, social stratification and family life.

Physical activity is health enhancing, and is associated with the well-being of people to lead a good life that has impact in mental, physical or psychological health.

Among UK children, Griffiths et al. (2013a) found that girls have lower physical activity than boys using the MCS4 data. Therefore, it is important to understand the factors that influence children's physical activity which may help to identify interventions to promote active lifestyle (Pouliou et al. 2015).

Physical activity levels (15 second sampling epochs) was recorded at 7 years using accelerometers (Actigraph GT1M, Pensacola, Florida) to measure activity reliably in children (Ott et al. 2000). Children were instructed to wear the accelerometers, on an elastic belt around the waist, during waking time for seven consecutive days, excluding bathing or during water time activities. After eliminating non-wear time and extreme readings, a valid data of > 10 hours on at least 2 days (Rich et al. 2013) resulting in a physical activity study sample were considered for further analyses. Data that were missing in physical activity data were not imputed. A detail of the data processing have been published by Griffiths et al. (2013a). The cut-points developed by Pulsford et al. (2011) were used to define sedentary, light, moderate, and vigorous physical activity. Moderate-to-vigorous physical activity (MVPA) was defined as $> 2,241$ accelerometer counts of activity per minute (Griffiths et al. 2013a; Griffiths et al. 2013b) and was reported in terms of number of minutes of activity per day.

1.5 Aims and Structure of Dissertation

This research finds a number of gaps in existing quantile regression (QR) for mixed effect models and aims to make some novel contribution which are expected to be instrumental in practical applications as well as to open windows for further developments. The aims and structure of this dissertation are outlined as follows.

In Chapter 2, we developed a marginally interpretable model for QR with random effects. Interest often lies in marginal or population-averaged effects rather than conditional effects. Using a LQMM (Geraci and Bottai 2014), one could obtain

marginal predictions, but directly modeling the marginal median or quantile using what is known as a marginal model is of greater interest. We discussed the derivative free numerical integral algorithm with Gauss-Hermite quadrature to the proposed model parameter estimations. Our proposed model were evaluated using statistical simulations. We demonstrated an application of the proposed model using MCS data.

In Chapter 3, we proposed a 3-level mixed-effects model for QR. The proposed model is an extension of 2-level mixed effects model (Geraci and Bottai 2014) available in literature. Estimation of parameters in our proposed model is carried out using a combination of a smoothing algorithm for QR and a second order Laplacian approximation for mixed models. We evaluated the proposed model using extensive statistical simulations. An application to MCS data were examined using the proposed model.

Chapter 4, the final chapter of this dissertation, presents an overall summary of the contributions as well as discusses some possible future research directions.

Chapter 2

Marginally Interpretable Models for QR with random-effects

2.1 Introduction

Quantile regression (QR), popularized by Koenker and Bassett (1978), has gradually become a well established statistical technique to assess the relationship between a response variable and one or more covariates. QR complements and improves the widely used classic mean regression analysis and characterizes the whole conditional distribution of a response variable given a set of covariates. Particularly, when researchers are more interested in the upper or lower tail of a distribution then QR is the best choice to perform the statistical analyses. Some properties of QR also make it readily available in a wide range of applications. These properties include robustness of outliers, no assumption on the distribution of error, and equivariance to monotone transformation. The immense use of QR has been noticed by the different researchers in a wide variety of areas because of its flexibility. Some of the areas includes, but not limited to, economics, medicine or public health, ecology and survival data analysis. QR has attracted considerable research interest in decades, and has been widely applied to independent data and time-to-event data (Huang et al. 2017). However, there are methodological challenges in regression analysis when the data are not independent.

Dependent or Correlated data arise frequently in statistical analyses. This occurs because of the grouping of subjects, such as, students within classrooms, or of the

repeated measurements on each subject over time, or of the multiple related outcome measures at one point in time. Dealing with such type of data introduce the correlation between observations on same units or clusters. To obtain correct inferences, these correlations should be accounted for. Mixed effects models, also known as multilevel or random-effects models, have been used frequently because of their flexibility to analyze data from such complex sampling designs.

The use of QR for longitudinal data has received increasing attention in many areas now. However, the extension of QR to longitudinal data or dependent data is limited. These type of models, mainly, known as conditional and marginal models. The approaches to the conditional QR may be distinguished by distribution free approaches (Koenker 2004; Lamarche 2010; Galvao and Montes-Rojas 2010; Galvao 2011) where QR with a large number of subject-specific fixed effects were panelized and likelihood based approaches (Geraci and Bottai 2007; Liu and Bottai 2009; Yuan and Yin 2010; Farcomeni 2012; Geraci and Bottai 2014) which are mainly based on the asymmetric Laplace density (ALD).

If the research interest lies on the marginal quantile effects then fitting conditional modeling of quantiles (with either fixed or random cluster-specific effects) may not be an ideal approach. In such situations, marginal modeling of quantiles of correlated responses provides a good alternative. The work is limited in marginal modeling for QR analysis. Jung (1996) proposed quasi-likelihood based estimating equations for median regression and preserved marginal effects by incorporating correlated errors in the quasi-likelihood model. Building on this approach, Lipsitz et al. (1997) described a weighted model in QR for longitudinal data. For nonlinear longitudinal data, Karlsson (2008) examined a weighted version of QR analysis. These methods are basically marginal models which capture the overall trend among all subjects for a given quantile. For example, Wang and Zhu (2011) defined an empirical likelihood under the generalized estimating equation (GEE) framework. Another way to make

inference on marginal effects is to adjust estimates from a conditional model when the latter does not naturally lead to marginal interpretations (Zeger, Liang, and Albert 1988; Lesaffre et al. 2000; Gory, Craigmile, and MacEachern 2016). This is also the case in QR with random effects.

Given the practical needs and the gap in the literature, in this study, we present a marginally interpretable model for QR for mixed effect models. This model will be a novel one for QR analysis. Obtaining marginal summaries from a linear quantile mixed model often requires evaluation of an analytically intractable integral. Our aim is to find a class of marginally interpretable linear quantile mixed models that lead to parameter estimates with a marginal interpretation while maintaining the desirable statistical properties of a conditionally-specified model. We discuss the maximum likelihood parameter estimation using the numerical integral algorithm with Gauss-Hermite quadrature. An application of the proposed model is illustrated by MCS data.

The rest of the chapter is organized as follows. The proposed model and parameter estimation algorithm asre presented in Section 2.2 and 2.3, respectively. Section 2.4 demonstrates an application, and finally, a discussion is presented in Section 2.5.

2.2 Marginally interpretable Models for QR with random-effects

The random intercept linear quantile mixed model (LQMM) for the τth quantile can be written as

$$Q_{y_{ij}|u_i}(\tau) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_\tau + u_{\tau,i} \tag{2.1}$$

where for $j = 1, 2, \ldots n_i; i = 1, 2, \ldots M$, $\boldsymbol{x_{ij}}$ is a $p \times 1$ covariate vector, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effect parameters, u_i is a (scalar) random effect and y_{ij} is the jth response in the ith cluster. To complete the specification of the model, Geraci and Bottai (2007) assume that the random effects u_i for $i = 1, 2, \ldots M$ are distributed as $f(u_i)$ symmetrically about 0. The conditional distribution of the outcome y_{ij} given

the random effect u_i, denoted by $f(y_{ij}|u_i)$, are independently distributed according to the asymmetric Laplace density (ALD) (Geraci and Bottai 2007)

$$f(y_{ij}|\boldsymbol{\beta}, u_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij} - (\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_\tau + u_{\tau,i})\Big)\right\} \tag{2.2}$$

and the specification of the τth conditional quantile model is given in (2.1), τ is fixed and known.

Interest often lies in marginal or population-averaged effects rather than conditional effects. Although one could obtain marginal predictions from a LQMM (Geraci and Bottai 2014), directly modeling the marginal median or quantile using what is known as a marginal model is of interest.

Our purpose is to develop a model with a direct marginal interpretation for the parameters. We begin with a structure for marginal model as

$$Q_{y_{ij}}(\tau) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_\tau \tag{2.3}$$

We then introduce an another model conditioning on a the latent variable

$$Q_{y_{ij}|u_i} = \Delta_{\tau,ij}(\beta, \sigma_u) + u_{\tau,i} \tag{2.4}$$

where, the response y_{ij} is conditionally independent given the random effects, u_i. $\Delta_{\tau,ij}(\beta, \sigma_u)$ is a function of both $\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_\tau$ and the random effect standard deviation, $\sigma_{u_\tau} = \sqrt{\text{var}(u_{\tau,i})}$. We refer the model defined by the Equations (2.3) and (2.4) as a marginally interpretable LQMM. We say that an LQMM is marginally interpretable if

$$\int \exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij} - \Delta_{\tau,ij} - u_{\tau,i}\Big)\right\} f(u_{\tau,i})du_{\tau,i} = \exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_\tau\Big)\right\} \tag{2.5}$$

for $j = 1, 2, \dots n_i$; $i = 1, 2, \dots M$, where $\rho_\tau(r) = r\Big\{\tau - I(r < 0)\Big\}$ is the loss function with $I(.)$ denoting indicator function. Here $\Delta_{\tau,ij}(\beta, \sigma_u)$ is defined as the solution to the integral in (2.5); the existence of which makes the LQMM model marginally interpretable.

An advantage of this model is that it can be more robust to miss-specification of random effects. The adjustment $\Delta_{\tau,ij}(\beta, \sigma_u)$ potentially depends on the fixed portion $(\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_\tau)$ of the model, and the parameters characterizing the random effects distribution $f(u_i)$. By including the adjustment in the LQMM, we have specified a formal statistical model for which the parameters $(\boldsymbol{\beta}_\tau)$ have a marginal interpretation for a specific τ. For specific quantiles (τ), the functional form of the adjustment is determined by the choice of random effects distribution, whereas its specific value depends on $\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_\tau$.

For a marginal quantile regression model for dependent data, each realization of a random effect is considered as a single value that applies to all units in a group of observations sharing that random effect. Such as, each random intercept in a conventional random intercepts quantile regression model corresponds to a shift in the τth quantile of the response, and for all units with the same random intercept the average of the quantile is shifted by the same amount. However, it may be necessary to associate each unit in the same group with a different value of the random effect in order to preserve the marginal quantile. Therefore, the idea that all units sharing the same random intercept are shifted by the same amount is not always applicable in a marginally interpretable LQMM model.

2.3 Estimation

The marginal likelihood for the ith cluster of a random intercept LQMM, $L_i\left(\boldsymbol{\beta}, \sigma, \sigma_u\right)$, given in (2.4) is obtained as (excluding the subscript τ)

$$\begin{aligned} L_i\left(\boldsymbol{\beta}, \sigma, \sigma_u\right) = &\prod_{j=1}^{n_i} \int f\left(y_{ij}|u_i\right) f(u_i)\, du_i \\ =&\sigma_{n_i}(\tau) \prod_{j=1}^{n_i} \int \exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij} - \Delta_{ij}(\beta, \sigma_u) - u_i\Big)\right\} f(u_i)\, du_i \end{aligned} \tag{2.6}$$

where $\sigma_{n_i}(\tau) = \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}}$. Let us denote the marginal log-likelihood by $l_i\left(\boldsymbol{\beta}, \sigma, \sigma_u\right) = \log L_i\left(\boldsymbol{\beta}, \sigma, \sigma_u\right)$ for $i = 1, 2, \ldots M$.

The integral in Equation (2.6) cannot be solved analytically, as a result, numerical methods are required for likelihood evaluation and parameter estimation. A growing literature exists on approaches to fit linear mixed effect quantile regression models. Geraci and Bottai (2007) used a Monte Carlo Expectation Maximization (MCEM) approach to maximize the likelihood function. In 2014, Geraci and Bottai proposed an LQMM for which the computational burden of the parameter estimation was reduced by using an approach based on a combination of Gaussian quadrature approximation and a non-smooth optimization process. To date, all of these algorithms have only address the conditionally specified models. In this work, to numerically approximate the marginally-specified likelihood function, we used K-point Gauss-Hermite quadrature algorithm. Details are provided in following.

Assuming normally distributed random random intercepts, i.e., $u_i \sim N(0, \sigma_u^2)$, the approximation of the integral in (2.6) by Gauss–Hermite quadrature is,

$$\begin{aligned}
&\tau^{n_i}(1-\tau)^{n_i}\sigma^{-n_i}\frac{\sigma_{u_i}^{-1}}{\sqrt{2\pi}}\prod_{j=1}^{n_i}\int \exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-u_i\Big)\right\}\times\exp\left(-\frac{u_i^2}{2\sigma_u^2}\right)du_i\\
&=\tau^{n_i}(1-\tau)^{n_i}\sigma^{-n_i}\frac{1}{\sqrt{2\pi}}\prod_{j=1}^{n_i}\int \exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v\Big)\right\}\times\exp\left(-\frac{v^2}{2}\right)dv\\
&\simeq\tau^{n_i}(1-\tau)^{n_i}\sigma^{-n_i}\prod_{j=1}^{n_i}\sum_{k=1}^{K}\exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k\Big)\right\}\times w_k
\end{aligned}$$

with nodes v_k, weight w_k, and K denoting the number of weights and nodes. We can then write the log-likelihood function of ith cluster as

$$l_i\left(\boldsymbol{\beta},\sigma,\sigma_u\right)=\log\left[\frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{-n_i}}\prod_{j=1}^{n_i}\sum_{k=1}^{K}\exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k\Big)\right\}\times w_k\right] \tag{2.7}$$

and the (marginal) log-likelihood for all M clusters is approximated by

$$l\left(\boldsymbol{\beta},\sigma,\sigma_u\right)=\sum_{i=1}^{M}\log\left[\frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{-n_i}}\prod_{j=1}^{n_i}\sum_{k=1}^{K}\exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k\Big)\right\}\times w_k\right] \tag{2.8}$$

To approximate the log-likelihood function, $\Delta_{ij}(\beta, \sigma_u)$ must be computed as a function of the marginal quantile parameters $\boldsymbol{\beta}$ and the random effect variance pa-

rameters σ_u. We have done so through numerical solution of the convolution function in Equation (2.5) using the K-point Gauss-Hermite quadrature algorithm. For given $\boldsymbol{x}_{ij}^T\boldsymbol{\beta}$ and σ_u, we used Newton-Raphson method to solve for the implied conditional parameter $\Delta_{ij}(\beta, \sigma_u)$. For this, we needed the derivatives of the convolution function of equation (2.5) with respect to $\Delta_{ij}(\beta, \sigma_u)$.

However, the non-differentiability of the loss function $\rho_\tau(r)$ at the origin, $r = 0$, makes it difficult to apply the standard smooth and gradient based optimization algorithms. Zheng (2011) introduced a smooth function to approximate the loss function $\rho_\tau(r)$ and discussed that gradient descent algorithms using the approximation of $\rho_\tau(r)$ provides an efficient means of estimating the quantile regression parameters with higher prediction accuracy. In this work, we adopted the loss function approximation proposed by Zheng (2011) to facilitate smooth optimization of the log-likelihood as well as solving for the convolution equation (2.5). The smoothed approximation of loss function function from Zheng (2011) is defined as

$$S_{\tau,\alpha}(r) = \tau r + \alpha \log\Big(1 + \exp\Big(-\frac{r}{\alpha}\Big)\Big) \tag{2.9}$$

where $\alpha > 0$ is called the smoothing parameter and we calculate

$$\begin{aligned}\frac{d}{dr}S_{\tau,\alpha}(r) &= \tau + \alpha\frac{1}{1+\exp\Big(-\frac{r}{\alpha}\Big)}(-\frac{1}{\alpha})\exp\Big(-\frac{r}{\alpha}\Big) \\ &= \tau - \frac{1}{1+\exp\Big(\frac{r}{\alpha}\Big)}\end{aligned}$$

Also,

$$\begin{aligned}\frac{d}{dr}\exp\left\{-kS_{\tau,\alpha}(r)\right\} &= -k\exp\left\{-kS_{\tau,\alpha}(r)\right\}\frac{d}{dr}S_{\tau,\alpha}(r) \\ &= -k\exp\left\{-kS_{\tau,\alpha}(r)\right\}\left\{\tau - \frac{1}{1+\exp\Big(\frac{r}{\alpha}\Big)}\right\}\end{aligned}$$

Some properties of the smooth function include: (1) $S_{\tau,\alpha}(r)$ is a convex function for any $\alpha > 0$; (2) for any $\alpha > 0$ and any $r \in \mathbb{R}$, $0 < S_{\tau,\alpha}(r) - \rho_\tau(r) \leq \alpha \log 2$, therefore, $\lim_{\alpha\to 0^+} S_{\tau,\alpha}(r) = \rho_\tau(r)$; and (3) $S_{\tau,\alpha}(r)$ is always positive for all $r \in \mathbb{R}$.

To ensure smoothness of the approximation, the smoothing parameter α should be small but not too small (Zheng 2011). Through extensive simulation studies, Zheng (2011) suggested use of α value around 0.5 for better performance at all quantile levels.

Rewriting the equation (2.5) without τ subscript on the parameters, we obtain

$$\begin{aligned} &\frac{\tau(1-\tau)}{\sigma}\int \exp\left\{-\frac{1}{\sigma}\rho_\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k\Big)\right\}\phi(v)dv \\ &=\frac{\tau(1-\tau)}{\sigma}\exp\left\{-\frac{1}{\sigma}\rho_\tau\Big((y_{ij}-\boldsymbol{x}_{ij}^T\boldsymbol{\beta}\Big)\right\} \end{aligned} \tag{2.10}$$

where ϕ is the standard normal density. Applying the smooth function to ρ_τ in (2.10), we get

$$\begin{aligned} &\frac{\tau(1-\tau)}{\sigma}\int \exp\Bigg[-\frac{1}{\sigma}\Bigg[\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k\Big)+ \\ &\alpha\log\left(1+\exp\left(-\frac{y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k}{\alpha}\right)\right)\Bigg]\Bigg]\phi(v)dv \\ &=\frac{\tau(1-\tau)}{\sigma}\exp\left[-\frac{1}{\sigma}\left\{\tau\Big(y_{ij}-\boldsymbol{x}_{ij}^T\boldsymbol{\beta}\Big)+\alpha\log\left(1+\exp\left(-\frac{y_{ij}-\boldsymbol{x}_{ij}^T\boldsymbol{\beta}}{\alpha}\right)\right)\right\}\right] \end{aligned} \tag{2.11}$$

Given $\boldsymbol{\beta}$ and σ_u for a specific τ, the integral equation can be solved numerically for $\Delta_{ij}(\beta,\sigma_u)$. By introducing the marginally specified model, we allow a choice to focus on a model marginally or conditionally when using a latent variable formulation. The choice between marginal and conditional models can be determined by the scientific objectives of the analysis.

Now we summarize the likelihood estimation for the proposed marginally interpretable model. Evaluation of the likelihood function requires numerical integration over the distribution of u_i. For marginally interpretable models, it requires the calculation of the implicitly defined conditional quantile parameter $\Delta_{ij}(\beta,\sigma_u)$. Modification of the existing algorithm (Geraci and Bottai 2014) to fit our proposed model is possible if $\Delta_{ij}(\beta,\sigma_u)$ can be obtained as a function of $\boldsymbol{x}_{ij}^T\boldsymbol{\beta}$ and σ_u, for a fixed τ. This is achieved through numerical solution to the convolution equation (2.11) that

connect marginal and conditional quantile functions. The partial derivatives of $\boldsymbol{\beta}$ are also required and are obtained through implicit differentiation of the convolution equation.

2.3.1 Calculation of $\Delta_{ij}(\beta, \sigma_u)$ and Derivatives

To numerically evaluate the convolution equation, we used Gauss-Hermite quadrature. Given $\left[\Delta_{ij}(\beta, \sigma_u), \sigma_u\right]$, we use Newton-Raphson to solve for the implied conditional parameter $\Delta_{ij}(\beta, \sigma_u)$ for τth quantile. For this, we need the following by differentiating (2.11) with respect to $\Delta_{ij}(\beta, \sigma_u)$, using Leibniz integral rule for differentiation under the integration,

$$
\begin{aligned}
A_{ij} &= \frac{\partial}{\partial \Delta_{ij}(\beta, \sigma_u)} \frac{\tau(1-\tau)}{\sigma} \\
&\qquad \int \exp\left[-\frac{1}{\sigma} \left\{ \tau\Big(y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}\Big) + \alpha \log\left(1 + \exp\left(- \frac{y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}}{\alpha}\right)\right)\right\}\right] \\
&= \frac{\partial}{\partial \Delta_{ij}(\beta, \sigma_u)} \frac{\tau(1-\tau)}{\sigma} \int \exp\left[-\frac{1}{\sigma}\left[\tau\Big(y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k\Big) + \right.\right. \\
&\qquad \left.\left. \alpha \log\left(1 + \exp\left(- \frac{y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k}{\alpha}\right)\right)\right]\right] \phi(v) dv \\
&= -\frac{\tau(1-\tau)}{\sigma^2} \int \exp\left[-\frac{1}{\sigma}\left[\tau\Big(y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k\Big) + \right.\right. \\
&\qquad \left.\left. \alpha \log\left(1 + \exp\left(- \frac{y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k}{\alpha}\right)\right)\right]\right] \\
&\qquad \left[-\tau + \frac{1}{1 + \exp\left(\frac{y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k}{\alpha}\right)}\right] \phi(v) dv
\end{aligned} \tag{2.12}
$$

which we also obtain numerically.

For maximum likelihood estimation of marginally interpretable models, we need the derivatives of the deconvolution solution, $\Delta_{ij}(\beta, \sigma_u)$, with respect to $\boldsymbol{\beta}$ and σ_u for a specific τ.

The necessary derivatives can be obtained through implicit differentiation of the convolution equation. Let us define

$$\begin{aligned}
B_{ij} &= \frac{\partial}{\partial \boldsymbol{\beta}} \frac{\tau(1-\tau)}{\sigma} \exp\Bigg[- \frac{1}{\sigma} \Bigg[\tau\Big(y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}\Big) + \\
&\qquad\qquad \alpha \log\Bigg(1 + \exp\Bigg(- \frac{y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}}{\alpha}\Bigg)\Bigg)\Bigg]\Bigg] \\
&= -\frac{\tau(1-\tau)}{\sigma^2} \exp\Bigg[- \frac{1}{\sigma} \Bigg\{ \tau\Big(y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}\Big) + \alpha \log\Bigg(1 + \exp\Bigg(- \frac{y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}}{\alpha}\Bigg)\Bigg)\Bigg\}\Bigg] \\
&\qquad\qquad \Bigg[- \tau + \frac{1}{1 + \exp\Big(\frac{y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}}{\alpha}\Big)}\Bigg] \boldsymbol{x}_{ij}
\end{aligned} \tag{2.13}$$

And,

$$\begin{aligned}
C_{ij} &= \frac{\partial}{\partial \boldsymbol{\beta}} \frac{\tau(1-\tau)}{\sigma} \int \exp\Bigg[- \frac{1}{\sigma} \Bigg[\tau\Big(y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k\Big) + \\
&\qquad\qquad \alpha \log\Bigg(1 + \exp\Bigg(- \frac{y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k}{\alpha}\Bigg)\Bigg)\Bigg]\Bigg] \phi(v) dv \\
&= -\frac{\tau(1-\tau)}{\sigma^2} \int \exp\Bigg[- \frac{1}{\sigma} \Bigg[\tau\Big(y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k\Big) + \\
&\qquad\qquad \alpha \log\Bigg(1 + \exp\Bigg(- \frac{y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k}{\alpha}\Bigg)\Bigg)\Bigg]\Bigg] \\
&\qquad\qquad \Bigg[- \tau + \frac{1}{1 + \exp\Big(\frac{y_{ij} - \Delta_{ij}(\beta, \sigma_u) - \sigma_u v_k}{\alpha}\Big)}\Bigg] \frac{\partial \Delta_{ij}(\beta, \sigma_u)}{\partial \boldsymbol{\beta}} \phi(v) dv \\
&= A_{ij} \frac{\partial \Delta_{ij}(\beta, \sigma_u)}{\partial \boldsymbol{\beta}}
\end{aligned} \tag{2.14}$$

Using (2.13) and (2.14), if the equation defined in (2.11) is differentiated with respect $\boldsymbol{\beta}$ to then we obtain

$$A_{ij} \frac{\partial \Delta_{ij}(\beta, \sigma_u)}{\partial \boldsymbol{\beta}} = B_{ij}$$

from which we get

$$\frac{\partial \Delta_{ij}(\beta, \sigma_u)}{\partial \boldsymbol{\beta}} = \frac{B_{ij}}{A_{ij}}$$

Differentiating the equation (2.11) with respect to σ_u,

$$-\frac{\tau(1-\tau)}{\sigma^2}\int \exp\Bigg[-\frac{1}{\sigma}\Bigg[\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k\Big)+ \alpha\log\left(1+\exp\left(-\frac{y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k}{\alpha}\right)\right)\Bigg]\Bigg] \left[-\tau+\frac{1}{1+\exp\left(\frac{y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k}{\alpha}\right)}\right]\left(\frac{\partial\Delta_{ij}(\beta,\sigma_u)}{\partial\sigma_u}+v_k\right)\phi(v)dv=0$$

which results the following

$$\frac{\partial\Delta_{ij}(\beta,\sigma_u)}{\partial\sigma_u}=\frac{D_{ij}}{A_{ij}}$$

where,

$$D_{ij}=\frac{\tau(1-\tau)}{\sigma^2}\int \exp\Bigg[-\frac{1}{\sigma}\Bigg[\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k\Big)+ \alpha\log\left(1+\exp\left(-\frac{y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k}{\alpha}\right)\right)\Bigg]\Bigg] \left[-\tau+\frac{1}{1+\exp\left(\frac{y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k}{\alpha}\right)}\right]v_k\phi(v)dv$$

and A_{ij} is defined in (2.12).

2.3.2 Maximum Likelihood Estimation

For the marginally interpretable model with a random intercept, the log likelihood for all clusters can be approximated numerically using K-points Gauss-Hermite quadrature as

$$l\left(\boldsymbol{\beta},\sigma,\sigma_u\right)=\sum_{i=1}^{M}\log\left[\tau^{n_i}(1-\tau)^{n_i}\sigma^{-n_i}\sum_{k=1}^{K}\exp\left[-\frac{1}{\sigma}\rho_\tau\Big(y_{ij}-\Delta_{ij}(\beta,\sigma_u)-\sigma_u v_k\Big)\right]\times w_k\right]$$

for $(w_k, v_k)_{k=1}^{K}$.

Finally, to obtain the model parameters $(\boldsymbol{\beta},\sigma_u,\sigma)$, the numerically approximated log-likelihood function is to be maximize. In this work, we considered using the derivative-free Nealder–Mead algorithm with tolerance and maximum number of iterations set to 10^{-5} and 1000, respectively.

2.4 Application

The Millennium Cohort Study (MCS) is an observational, multidisciplinary cohort study of children born in the United Kingdom (UK), which is nationally representative of the UK population at baseline and conducted by the Centre for Longitudinal Studies (CLS) at the University of London. This ongoing study consists of a sample of all children born between September 2000 and January 2002, alive and living in England, Scotland, Wales, or Northern Ireland at age 9 months, and eligible to receive child benefit at that age (Plewis et al. 2007). In the first sweep known as MCS1 that occurred at age 9 months, 18,552 families (18,827 children) were recruited to the cohort and information was collected on 18,818 children. Subsequent sweeps have occurred when the children were aged 3 (MCS2), 5 (MCS3), 7 (MCS4), 11 (MCS5), 14 (MCS6), 17 (MCS7) and 22 (MCS8) years. On a broad range, information of demographic, behavioral, developmental, health, and parental socioeconomic characteristics were recorded at first sweep of MCS (Connelly and Platt 2014). Trained interviewers were required to interview the subjects with information provided by mothers. Ethical approval for data collection at each sweep of MCS and accelerometer studies has been granted and approval for seasonal accelerometer and calibration studies was granted by the University College London Research Ethics Committee (Griffiths et al. 2013b). The MCS data are freely available to the researchers under standard access conditions through the UK Data Service (http://ukdataservice.ac.uk) and the MCS website provides detailed information on the study (http://www.cls.ioe.ac.uk/mcs).

Physical activity is health enhancing, and is associated with the well-being of people that has impact in mental, physical or psychological health. Among UK children, Griffiths et al. (2013a) found that girls have lower physical activity than boys using the MCS4 data. Therefore, it is important to understand the factors that influence children's physical activity which may help to identify interventions to promote active lifestyle (Pouliou et al. 2015).

Physical activity levels (15 second sampling epochs) was recorded at 7 years using accelerometers (Actigraph GT1M, Pensacola, Florida) to measure activity reliably in children (Ott et al. 2000). Children were instructed to wear the accelerometers, on an elastic belt around the waist, during waking time for seven consecutive days, excluding bathing or during water time activities. After eliminating non-wear time and extreme readings, a valid data of ≥ 10 hours on at least 2 days (Rich et al. 2013) resulting in a physical activity study sample were considered for further analyses. Data that were missing in physical activity data were not imputed. A detail of the data processing have been published by Griffiths et al. (2013a). The cut-points developed by Pulsford et al. (2011) were used to define sedentary, light, moderate, and vigorous physical activity. Moderate-to-vigorous physical activity (MVPA) was defined as $> 2,241$ accelerometer counts of activity per minute (Griffiths et al. 2013a; Griffiths et al. 2013b) and was reported in terms of number of minutes of activity per day.

We examined the data from the MCS4 which took place for children aged 7. We considered a complete sample of 4,738 singleton children (2,321 boys, 2,417 girls) that were from only England. There are 253 electoral wards and we considered daily moderate to vigorous physical activity (dMVPA) as our outcome variable. Covariates that we considered for this study include: gender (binary, reference: boy), ethnic group (binary, reference: white), income quintiles (categorical, reference: fifth quintile), time spent reading for enjoyment (binary, reference: often), mode of transport to/from school (binary, reference: active), number of cars or vans owned (categorical, reference: two) and body mass index (BMI) of the child. We aggregated dMVPA by child. Using electoral ward as grouping factor, our objective is obtaining marginal interpretation of daily PA quantiles at the national level (that is, marginally with respect to administrative areas).

Table 2.1 Demographic and Socio-demographic information of children from the Millennium Cohort Study

Characteristics	Children (%)	Five number summary (Min, Q1, Median, Q3, Max)
Gender		
Boys	2321 (49.0)	
Girls	2417 (51.0)	
Ethnicity		
White	3846 (81.2)	
Other	892 (18.8)	
Income quintile		
1	744 (15.7)	
2	816 (17.2)	
3	964 (20.3)	
4	1070 (22.6)	
5	1144 (24.1)	
Reading for pleasure		
Often	4025 (85.0)	
Not often	713 (15.0)	
Transportation to/from school		
Active	2589 (54.6)	
Passive	2149 (45.4)	
Number of cars or vans owned		
0	480 (10.1)	
1	1781 (37.6)	
2	2211 (46.7)	
3 or more	266 (5.6)	
BMI (kg/m^2)		(10.9, 15.1, 16.1, 17.5, 44.6)
Weight(kg)		(13.2, 22.1, 24.5, 27.6, 63.4)
MVPA		(0.0, 12.9, 37.9, 59.7, 600.0)

Note: MVPA = Moderate to vigorous physical activity.

A summary of the data set is given in Table 2.1 where boys and girls of the sample are approximately equal (49% vs 51%). White children (81.2%) dominate the other ethnic grouped children (18.8%). Most of the children in the selected sample are in the fifth income quintile (24.1%), followed by fourth income quintile (22.6%), third income quintile (20.3%), second income quintile (17.2%) and first income quintile (15.7%). Most of the children read often for pleasure (85.0%). Children who go

schools by car or bus are fewer (45.4%) as compared to their peers (54.6%). About half of the children's family own two cars (46.7%). The five number summary for BMI, weights and daily MVPA are given in the Table 2.1 as well.

In Table 2.2 ($\tau = 0.10, 0.25$), Table 2.3 ($\tau = 0.50, 0.75$), and Table 2.4 ($\tau = 0.90$), we presented marginal (proposed) and conditional linear quantile mixed effect models to assess the relationship between daily moderate to vigorous physical activity (MVPA) and the covariates of interest for the 10th, 25th, 50th, 75th and 90th quantiles. Using child as grouping factor (there are 4,738 children), our objective is obtaining marginal interpretation of daily PA quantiles at population level. We also showed in Table 2.5 ($\tau = 0.10, 0.25$), Table 2.6 ($\tau = 0.50, 0.75$), and Table 2.7 ($\tau = 0.90$) that how the results compare to the quantile regressions, when observations for different children within a ward are assumed to be independent, and the unconditional quantile regressions (Firpo, Fortin, and Lemieux 2009). We estimated parameters of interest for all models and determined the Proportion of Negative Residuals (PNR) which is expected to be approximately equal to τ. Root Mean Square Error (RMSE) were estimated to measure the performance of the models. Standard errors were calculated using $R = 50$ bootstrap replications for all models.

The interpretation of marginally specified parameters does not depend on the assumption of within-subject variations. We found that conditionally specified model estimates are different than marginally interpretable model estimates in all quantile levels. As expected, PNR are approximately equal to the τ for all models. The fit statistics, such as, likelihoods and RMSE for marginal models are comparable with the corresponding conditional models at the quantile levels we considered (see the Tables). However, RMSE for unconditional quantile regressions are larger as compared to other models at all quantiles.

Table 2.2 Conditionally and Marginally specified linear quantile mixed effect model parameter estimates (standard error) for MVPA using MCS data ($\tau = 0.10, 0.25$).

Characteristics	$\tau = 0.10$		$\tau = 0.25$	
	Conditional	Marginal	Conditional	Marginal
Intercept	**3.658 (0.51)**	3.283 (8.60)	**29.35 (1.34)**	**27.07 (5.78)**
Gender				
Girls	-0.004 (0.04)	-0.046 (0.36)	**-1.129 (0.27)**	**-1.871 (0.62)**
Boys (Ref)				
Ethnicity				
Not white	0.002 (0.24)	0.011 (0.38)	**1.173 (0.27)**	**2.790 (0.62)**
White (Ref)				
Reading for Pleasure				
Not often	-0.002 (0.04)	-0.024 (0.33)	**-0.513 (0.19)**	0.107 (1.05)
Often (Ref)				
Transportation				
Passive	-0.039 (0.04)	-0.005 (0.30)	**-0.443 (0.11)**	**-1.484 (0.62)**
Active (Ref)				
Number of cars or van				
0	-0.042 (0.89)	-1.376 (0.91)	**0.459 (0.18)**	-0.271 (1.24)
1	-0.039 (0.03)	-0.001 (0.39)	**0.438 (0.06)**	0.204 (0.77)
$\geq$3	0.004 (0.20)	0.042 (0.71)	0.124 (0.35)	1.121 (1.34)
2 (Ref)				
BMI (kg/m^2)	-0.0003 (0.03)	-0.013 (0.53)	**-0.643 (0.07)**	**-0.748 (0.34)**
Heterogeneity (σ_u)				
Intercept	1.926	1.737	9.750	3.398
Scale parameter (σ)	4.334	4.355	9.436	9.900
Log-likelihood	-215000	-215092	-218288	-219362
PNR	0.12	0.10	0.29	0.25
RMSE	64.75	65.22	56.59	58.95

Note: PNR = Proportion of Negative Residuals, RMSE = Root Mean Square Error.

Table 2.3 Conditionally and Marginally specified linear quantile mixed effect model parameter estimates (standard error) for MVPA using MCS data, continued ($\tau = 0.50, 0.75$).

Characteristics	$\tau = 0.50$		$\tau = 0.75$	
	Conditional	Marginal	Conditional	Marginal
Intercept	**43.17 (2.98)**	**45.38 (1.85)**	**61.67 (2.65)**	**51.26 (2.91)**
Gender				
Girls	**-7.545 (0.63)**	**-7.562 (0.50)**	**-10.43 (0.71)**	**-11.54 (0.76)**
Boys (Ref)				
Ethnicity				
Not white	**-2.762 (0.74)**	**-1.891 (0.85)**	**-3.634 (1.15)**	**-4.296 (0.72)**
White (Ref)				
Reading for Pleasure				
Not often	1.806 (1.18)	1.386 (1.02)	1.968 (1.15)	**2.160 (0.86)**
Often (Ref)				
Transportation				
Passive	**-2.375 (0.69)**	**-2.227 (0.60)**	**-2.478 (0.63)**	**-2.387 (0.72)**
Active (Ref)				
Number of cars or van				
0	**3.014 (1.10)**	**4.704 (1.28)**	**5.562 (1.53)**	**7.006 (1.50)**
1	1.030 (0.79)	**1.951 (0.66)**	**2.925 (0.63)**	**2.911 (0.79)**
$\geq$3	-0.388 (1.47)	1.684 (1.23)	2.812 (2.31)	0.403 (1.68)
2 (Ref)				
BMI (kg/m^2)	0.080 (0.18)	-0.019 (0.12)	0.124 (0.16)	**0.864 (0.18)**
Heterogeneity (σ_u)				
Intercept	13.82	5.909	16.28	8.049
Scale parameter (σ)	13.14	13.76	11.99	12.53
Log-likelihood	-220874	-222315	-229892	-231742
PNR	0.52	0.51	0.74	0.75
RMSE	50.19	50.25	51.95	52.36

Note: PNR = Proportion of Negative Residuals, RMSE = Root Mean Square Error.

Table 2.4 Conditionally and Marginally specified linear quantile mixed effect model parameter estimates (standard error) for MVPA using MCS data, continued ($\tau = 0.90$).

Characteristics	$\tau = 0.90$	
	Conditional	Marginal
Intercept	**89.85 (6.65)**	**83.03 (7.39)**
Gender		
Girls	**-13.99 (1.79)**	**-14.48 (1.15)**
Boys (Ref)		
Ethnicity		
Not white	**-9.246 (2.51)**	**-5.630 (1.25)**
White (Ref)		
Reading for Pleasure		
Not often	0.820 (2.06)	**5.276 (1.54)**
Often (Ref)		
Transportation		
Passive	-2.612 (1.67)	**-3.114 (1.10)**
Active (Ref)		
Number of cars or van		
0	**13.53 (3.22)**	**11.92 (2.73)**
1	3.185 (1.93)	**4.044 (1.28)**
$\geq$3	-5.979 (3.29)	**2.272 (2.05)**
2 (Ref)		
BMI (kg/m^2)	0.020 (0.34)	0.309 (0.46)
Heterogeneity (σ_u)		
Intercept	25.95	7.950
Scale parameter (σ)	8.278	9.122
Log-likelihood	-246843	-249768
PNR	0.90	0.89
RMSE	62.87	62.53

Note: PNR = Proportion of Negative Residuals, RMSE = Root Mean Square Error.

Table 2.5 Quantile regression (QR) and unconditional quantile regression (UQR) parameter estimates (standard error) for MVPA using MCS data ($\tau = 0.10, 0.25$).

Characteristics	$\tau = 0.10$		$\tau = 0.25$	
	QR	UQR	QR	UQR
Intercept	**3.254 (0.54)**	**3.571 (0.05)**	**14.56 (2.01)**	**14.23 (0.11)**
Gender				
Girls	**-0.404 (0.18)**	**-0.445 (0.02)**	**-1.578, (0.51)**	**-1.325 (0.09)**
Boys (Ref)				
Ethnicity				
Not white	-0.174 (0.20)	**-0.209 (0.01)**	**2.481 (0.70)**	**1.934 (0.05)**
White (Ref)				
Reading for Pleasure				
Not often	-0.309 (0.21)	**-0.365 (0.02)**	-0.641 (0.97)	**-0.584 (0.04)**
Often (Ref)				
Transportation				
Passive	**-0.306 (0.13)**	**-0.408 (0.01)**	**-2.136 (0.60)**	**-1.906 (0.06)**
Active (Ref)				
Number of cars or van				
0	**-2.680 (0.24)**	**-1.773 (0.05)**	-1.344 (1.04)	**-1.085 (0.04)**
1	**-0.422 (0.15)**	**-0.476 (0.02)**	-0.251 (0.64)	**-0.281 (0.04)**
$\geq$3	0.233 (0.30)	**0.274 (0.03)**	0.561 (1.18)	**0.234 (0.10)**
2 (Ref)				
BMI (kg/m^2)	0.009 (0.03)	0.001 (0.003)	0.007 (0.12)	0.007 (0.006)

Table 2.6 Quantile regression (QR) and unconditional quantile regression (UQR) parameter estimates (standard error) for MVPA using MCS data, continued ($\tau = 0.50, 0.75$).

Characteristics	$\tau = 0.50$		$\tau = 0.75$	
	QR	UQR	QR	UQR
Intercept	**42.47 (1.09)**	**41.88 (0.08)**	**64.75 (1.51)**	**64.70 (0.18)**
Gender				
Girls	**-7.734 (0.40)**	**-7.842 (0.09)**	**-12.39 (0.52)**	**-12.45 (0.17)**
Boys (Ref)				
Ethnicity				
Not white	**-2.215 (0.40)**	**-2.607 (0.05)**	**-4.400 (0.61)**	**-4.119 (0.06)**
White (Ref)				
Reading for Pleasure				
Not often	0.484 (0.53)	**0.444 (0.05)**	0.984 (0.61)	**1.362 (0.09)**
Often (Ref)				
Transportation				
Passive	**-2.550 (0.35)**	**-2.797 (0.06)**	**-2.819 (0.33)**	**-3.119 (0.05)**
Active (Ref)				
Number of cars or van				
0	**3.641 (0.58)**	**3.478 (0.08)**	**6.373 (0.91)**	**5.651 (0.11)**
1	**1.290 (0.38)**	**1.305 (0.04)**	**3.049 (0.46)**	**2.524 (0.04)**
$\geq$3	0.636 (0.80)	**0.367 (0.05)**	0.848 (0.97)	**0.918 (0.07)**
2 (Ref)				
BMI (kg/m^2)	0.036 (0.06)	**0.053 (0.006)**	0.085 (0.08)	**0.112 (0.009)**

Table 2.7 Quantile regression (QR) and unconditional quantile regression (UQR) parameter estimates (standard error) for MVPA using MCS data, continued ($\tau = 0.90$).

Characteristics	$\tau = 0.90$	
	QR	UQR
Intercept	**89.02 (2.34)**	**87.69 (0.24)**
Gender		
Girls	**-15.29 (0.68)**	**-14.45 (0.28)**
Boys (Ref)		
Ethnicity		
Not white	**-5.429 (0.97)**	**-4.310 (0.11)**
White (Ref)		
Reading for Pleasure		
Not often	**3.256 (1.29)**	**3.251 (0.09)**
Often (Ref)		
Transportation		
Passive	**-3.127 (0.69)**	**-2.547 (0.06)**
Active (Ref)		
Number of cars or van		
0	**10.27 (1.78)**	**9.366 (0.18)**
1	**3.482 (0.86)**	**2.901 (0.08)**
≥ 3	0.663 (1.56)	0.401 (0.24)
2 (Ref)		
BMI (kg/m^2)	0.092 (0.13)	**0.214 (0.02)**

2.5 Discussion

Statistical analyses for dependent data are increasing in a wide range of studies. Such type of data introduces complex sampling designs that is analyzed by using mixed effects models, also known as multilevel or random effects models. The use of quantile regression (QR) for such complex designs has been increasing attention to many researchers now. However, the extension of QR analyses to dependent data is limited. A number of conditional models for QR with random effects are available in the literature (Koenker 2004; Geraci and Bottai 2007; Liu and Bottai 2009; Yuan and Yin 2010; Farcomeni 2012; Geraci and Bottai 2014). However, researchers may be interested in marginal modeling of quantiles for dependent or correlated responses. Some methods for marginal models are developed (Jung 1996; Lipsitz et al. 1997; Karlsson 2008; Wang and Zhu 2011) that capture the overall trend among all subjects

for a specific quantile. Moreover, to make inference on marginal effects is to adjust estimates from a conditional model when the latter does not naturally lead to marginal interpretations. This is the case in generalized linear mixed model (Zeger, Liang, and Albert 1988; Lesaffre et al. 2000; Gory, Craigmile, and MacEachern 2016) and quantile linear mixed effect model (Geraci and Bottai 2014).

In this chapter, we considered a random intercept linear quantile mixed model, where we adopted the structure for the marginal quantile as opposed to conditional quantile. We considered numerical quadrature (Geraci and Bottai 2014) and Nealder-Mead algorithms for the parameter estimations. We presented an application of our proposed model to Millennium Cohort Study (MCS).

Finally, our model is computationally intensive. For example, it took about two hours to fit a single model using the MCS dataset. However, a possible improvement in computing speed of the proposed algorithm is obtainable and is part of future research.

Chapter 3

Multilevel Models for Quantile Regression

3.1 Introduction

Regression analysis is one of the most widely used statistical tools to assess the relationship between an outcome variable and one or more explanatory variables. Depending on the nature of the outcome variable, different types of regression analyses are developed and used in practice. Mean regression analyses allow researchers to make inferences regarding the conditional mean and assume that the effect of an explanatory variable is the same for all individuals, all other explanatory variables being equal, regardless of the distribution of the outcome variable. However, nature of statistical association may not be the same for individuals who are below or above the average of the outcome variable distribution. For an example, birthweight is considered as a strong predictor of infant mortality and morbidity, also it has significant effect on the physical and the behavioral outcomes in later life of infants. As a result, it is important to predict low/high birthweight because babies born with low/high birthweight are more likely to have several health complications compared to the babies born with normal birthweight. From the literature review, it is found that the mean regression analyses are commonly used statistical methods to assess the effects of determinants on birthweigth (Tian and Chen 2006; Geraci 2016). Averaging out the stronger or weaker effects of outcome may not be ideal to describe a situation, especially, where lower or upper tail of the distribution of the outcome plays an important role.

A statistical tool that complements mean regression is Quantile Regression (QR) (Koenker and Bassett 1978). OR provides a framework within which the entire conditional distribution can be characterized. QR has been considered as a successful analytical method in many fields as this enables inferences about the relationship between covariates and any quantile of a response. This is more valuable in applications where the tails of the response distribution are of interest. QR has a number of good properties including no assumption on the distribution of the response variable, equivalent to monotone transformations and robustness to outliers (Huber 1981). A lot of work has been done using this statistical tool since the seminal work by Koenker and Bassett (1978). However, there is an ongoing area of research to incorporate multiple sources of variation from hierarchical designs by using QR methods. The breadth of work addressing dependence within the quantile regression framework is still limited.

Here, we are interested in correlated or dependent data analyses that arise in a wide variety of studies. A number of sampling designs such as cluster, multilevel and repeated measures (longitudinal data) introduce such dependence. The correlation between observations on same units or clusters should be accounted for to obtain appropriate inferences. To analyze data from complex sampling designs, mixed effects models (also known as multilevel or random-effects models) have been used frequently because of their flexibility.

Mixed-effects models are applied to data where the observations are grouped according to one or more levels of clustering or dependence. Mixed-effects models incorporate both fixed and random effects. Fixed effects describe relationships common to the entire population or of those units associated with repeatable levels of experimental factors. Random effects are associated with the variability in the impact of fixed effects on the outcome due to the one or more levels of clustering. A common assumption in mixed models is that random effects are normally distributed. This assumption is typically made for computational convenience, rather than based on

some theoretical justification as in linear models the integral of the conditional likelihood has a closed-form. As noted by Burr and Doss (2005), random effects, unlike error terms, cannot be checked as there are no residuals that are uniquely associated with the random effects. Misspecification of the random-effects distribution may lead to biased and unreliable results. Inferences about the random effects themselves are also more likely to be affected by misspecification of the random-effects distribution (Drikvandi, Verbeke, and Molenberghs 2017).

A specific issue that has not yet been adequately addressed in the case of QR is that when data are collected at multiple levels and inferences regarding each of those levels are of research interest. Multilevel mixed-effects models use random effects at multiple levels of clustering, are commonly used in such settings. Using quantile regression for multilevel models enables the identification of heterogeneous covariate effects for each level at different quantiles of the outcome, and also provide robust estimates for the random-effects distributions with heavy tails and outliers. Both distribution-free and likelihood based approaches have been proposed for quantile regression mixed-effects models. A quasi-likelihood method for median regression for dependent observations was developed by Jung (1996). Koenker (2004) used penalized least squares method to analyze the longitudinal data with quantile regression. Expectation-Maximization (EM) algorithm (Farcomeni 2012), Monte Carlo Expectation-Maximization (MCEM) algorithm (Geraci and Bottai 2007; Liu and Bottai 2009), and Bayesian approaches using by Markov chain Monte Carlo (MCMC) (Yuan and Yin 2010; Yue and Rue 2011; Luo, Lian, and Tian 2012; Waldmann et al. 2013) have been used for quantile regression analysis with dependent data. Geraci and Bottai (2014) extended their previous study (Geraci and Bottai 2007) by incorporating multiple random effects, and used a combination of Gaussian quadrature approximations and non-smooth optimization algorithms for the estimation of the parameters.

The approach that will be addressed here is an extension of Geraci and Bottai (2014)'s linear quantile mixed model (LQMM) focusing on inferences with data collected with more than 2-levels of clustering. Specifically, the goal is to extend QR approaches associated with 2-level random-effects models to 3-level random-effects models. Our approach is novel in terms of modeling and estimation in a frequentist framework. Ours is a complement of Bayesian approaches (Yu and Moyeed 2001; Reich, Bondell, and Wang 2010; Luo, Lian, and Tian 2012; Yang and He 2012; Huang, Chen, and Qiu 2017) that avoids introducing prior distributions on the fixed effects. We will provide an analytic form of the objective function to be optimized. In our approach, we propose using bootstrap at cluster levels and demonstrate the procedure has empirical coverage probabilities close to the nominal level. Estimation of parameters in our model is carried out using a combination of a smoothing algorithm for QR and a second order Laplacian approximation for mixed models.

Here, we are concerned with a QR function of continuous responses when data are from cluster designs. This research is motivated by a study on daily moderate to vigorous physical activity (dMVPA) of school going children. Daily physical activity (PA) provides wellbeing benefits for children. This study was conducted as part of a nationally representative of a cohort of children and their families in UK known as the Millennium Cohort Study (MCS). From MCS study, Griffiths et al. (2013a) showed that only half of UK 7-year old children achieve recommended levels of PA where girls are less active than boys. Having concern of the wellbeing of children, many national and international guidelines have been developed advocating the importance of PA (Aggio et al. 2017). Children and adolescents aged 5-17 years should spend at least 60 min per day in moderate to vigorous physical activity (MVPA), that helps to have additional health benefits (Active 2011). Therefore, it is important to identify the predictors of dMVPA. We will consider child (repeated measurements of daily average MVPA within child) and electoral ward (children within wards) as grouping factors.

In this paper, we assume that the random effects are normally distributed. Though a literature review showed that the incorrect specification of the random effects' distributions in mixed models may affect inferences about the random effects themselves (Drikvandi, Verbeke, and Molenberghs 2017), there are some disagreements as to whether such parametric assumptions are harmless or have important consequences on inference (McCulloch and Neuhaus 2011). The answer partly lies in the specific model and type of variables involved, as well as the target of the inference. In the context of QR with random effects, Geraci and Bottai (2014) found that parameter estimation in LQMMs was relatively robust to random effects with a heavy-tailed distribution or a distribution contaminated with outliers, although bias resulted when random effects followed a skewed distribution. The authors noted that distributions other than normal could be considered, although possibly at the cost of more involved technical developments. The approach by Alfò, Salvati, and Ranalli (2017) can be used to avoid parametric assumptions on the random effects. Assessing the impact of misspecification of the random effects' distributions here is particularly complicated due to the nature of the models and the computational complexity. We do not explore this issue here.

The rest of this paper is organized as follows. In Section 3.2, we introduce the standard linear QR model and also outline the LQMM approach. In Section 3.3, we introduce the multilevel linear quantile mixed-effects model and related inference. In Section 3.4, we carry out a simulation study to investigate the statistical and computational performance of the proposed methods. In Section 3.5, we consider an application of MCS data to the proposed models and and the final Section 3.6 presents a discussion.

3.2 Linear Quantile Mixed Models (LQMM)

Quantile regression estimates and conducts inference about conditional quantile functions (Koenker and Bassett 1978). For any $0 < \tau < 1$, the τth quantile of a real valued random variable Y is defined as

$$Q(\tau) = \inf\{y : F(y) \geq \tau\},$$

where $F(y) = \text{Prob}(Y \leq y)$ is the distribution function of Y. If (x_i^T, y_i), $i = 1, 2, \ldots, n$ is an independent random sample from some population, where $\boldsymbol{x}$ is a known $p \times 1$ vector of regressors and $\boldsymbol{y}$ is a scalar response variable with conditional cumulative distribution function F_y and its shape is unspecified, then the linear conditional quantile function, $Q_{y_i}(\tau|x_i) = x_i^T \beta_\tau$, can be estimated by solving the optimization problem

$$\hat{\beta}_\tau = \arg\min_{\beta \,\in\, \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(\boldsymbol{y}_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_\tau \right) \tag{3.1}$$

where $\rho_\tau(r) = r\{\tau - I(r < 0)\}$ is the loss function or check function with r being a real number and $I(.)$ is the indicator function. The solution $\hat{\beta}_\tau$ is usually called the τth regression quantile.

Asymmetric Laplace Distribution (ALD) provides a natural link between minimization of equation (3.1) and the maximum likelihood estimation. A random variable T follows an ALD if its corresponding probability density is given by

$$p(t) = \frac{\tau(\tau - 1)}{\sigma_\tau} \exp\left\{ -\frac{1}{\sigma_\tau} \rho_\tau(t - \mu_\tau) \right\}$$

where $\mu_\tau \in \mathbb{R}$ and $\sigma_\tau > 0$ are location and scale parameters, respectively. Let $\mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$, then the likelihood of n independent observations is written as (for any given τ)

$$L \propto \frac{1}{\sigma^n} \exp\left\{ -\sum_{i=1}^{n} \rho_\tau \Big(\frac{\boldsymbol{y}_i - \boldsymbol{x}_i^T \boldsymbol{\beta}}{\sigma} \Big) \right\} \tag{3.2}$$

the maximization of the likelihood in (3.2) with respect to the parameter β is equivalent to the minimization of the objective function in (3.1).

Now, let us consider data from two level nested designs. For simplicity, we consider n_i observations nested within each of $i = 1, 2, \ldots, M$ clusters. The data are in the form of $\left(\boldsymbol{x}_{ij}^T, \boldsymbol{z}_{ij}^T, y_{ij}\right)$, for $j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, M$, and $N = \sum_{i=1}^{M} n_i$. Here, $\boldsymbol{x}_{ij}^T$ is the jth row of a known $n_i \times p$ matrix $\boldsymbol{X}_i$, $\boldsymbol{z}_{ij}^T$ is the jth row of a known $n_i \times q$ matrix $\boldsymbol{Z}_i$, and y_{ij} is the jth observation of the response vector $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$ for the ith cluster.

The linear quantile mixed model (LQMM) for two level clustered data (Koenker 2004; Geraci and Bottai 2014) can be written as follows

$$Q_{y_{ij}|u_i, x_{ij}, z_{ij}}(\tau) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_\tau + \boldsymbol{z}_{ij}^T \boldsymbol{u}_{\tau,i} \tag{3.3}$$

for $\tau \in (0, 1)$; $j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, M$ and $\boldsymbol{\beta}_\tau = (\beta_{\tau,0}, \beta_{\tau,1}, \ldots \beta_{\tau,p})^T$. The $q \times 1$ random effects vector $\boldsymbol{u}_{\tau,i}$ are assumed to be distributed according to $p\left(\boldsymbol{u}_{\tau,i} | \boldsymbol{\Psi}_\tau\right)$ (no specific distribution), where $\boldsymbol{\Psi}_\tau$ is a $q \times q$ covariance matrix. The random effects vectors depends on τ through $\boldsymbol{\Psi}_\tau$ and are assumed to be independent for different i.

The parameters of quantile regression for the linear mixed-effects model achieve an unbiased estimation due to the fact that the correlation of the measurements from the same subjects are adequately accounted for (Geraci and Bottai 2007). Geraci and Bottai (2014) used a combination of Gaussian quadrature approximations and non-smooth optimization algorithms for the estimation of the fixed regression coefficients and of the random effects' covariance matrix, and this method was implemented in R package 'lqmm' (Geraci 2014). Other approaches are reviewed by Marino and Farcomeni (2015).

The objective function to be minimized is given as

$$\sum_{i=1}^{M} \sum_{j=1}^{n_i} \rho_\tau \left(\boldsymbol{y}_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_\tau - \boldsymbol{z}_{ij}^T \boldsymbol{u}_{\tau,i}\right) \tag{3.4}$$

The minimization of (3.4) is equivalent to fitting a LQMM where the responses, conditionally on the random effects, is assumed to follow the asymmetric Laplace distribution (Geraci and Bottai 2007; Geraci and Bottai 2014). So, the responses y_{ij}, conditionally on the random effects, are independently distributed as ALD with location and scale parameters given by $\mu_{\tau,ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta}_\tau + \boldsymbol{z}_{ij}\boldsymbol{u}_{\tau,i}$ and σ_τ. The skew parameter τ is fixed and defines the quantile levels. The authors also assumed that the $q \times 1$ random effects vector $\boldsymbol{u}_{\tau,i}$, $i = 1, 2, \ldots, M$ are assumed to be independent of model's error term with mean zero and $q \times q$ variance-covariance matrix $\boldsymbol{\Psi}_\tau$. Let us define $\boldsymbol{y} = \left(\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_M^T\right)^T$; $\boldsymbol{u} = \left(\boldsymbol{u}_1^T, \boldsymbol{u}_2^T, \ldots, \boldsymbol{u}_M^T\right)^T$. According to Geraci and Bottai (2014), the joint density of $(\boldsymbol{y}, \boldsymbol{u})$ based on M clusters for τth LQMM is given by

$$p(\boldsymbol{y}, \boldsymbol{u}) = \left\{\frac{\tau(\tau-1)}{\sigma_\tau}\right\}^N \prod_{i=1}^{M} \exp\left\{-\frac{1}{\sigma_\tau}\sum_{j=1}^{n_i} \rho_\tau(y_{ij} - \mu_{\tau,ij})\right\} p(\boldsymbol{u}_{\tau,i})$$

A combination of Gaussian quadrature and non-smooth optimization is used for estimating LQMMs (Geraci and Bottai 2014).The approximated marginal (over the random effects) log-likelihood using the rule is given is

$$L\left(\boldsymbol{\beta}_\tau, \sigma_\tau, \boldsymbol{\Psi}|\boldsymbol{y}\right) = \sum_{i=1}^{M} \log\left\{\sum_{k_1=1}^{K} \cdots \sum_{k_q=1}^{K} p(\boldsymbol{y}_i|\boldsymbol{v}_{k_1,\ldots,k_q}) \prod_{l=1}^{q} w_{k_l}\right\}, \tag{3.5}$$

where v_{k_l} and w_{k_l}, $k_l = 1, \ldots, K; l = 1, \ldots, q$ denote the abscissas and weights of the Gaussian quadrature, respectively, with $v_{k_1,\ldots,k_q} = \left(v_{k_1}, \ldots, v_{k_q}\right)^T$. Prediction of the random effects for LQMMs was carried out via best linear prediction (BLP).

3.3 Methods of Multilevel Linear Quantile Mixed Models

3.3.1 Notation

Let us consider data from three level nested cluster designs. For simplicity, we consider n_{ij} observations nested within each of $j = 1, 2, \ldots, M_i$ level-2 units that are, in turn, nested within each of $i = 1, 2, \ldots, M$ clusters (level-1 units). The data are in the form of $\left(\boldsymbol{x}_{ijk}^T, \boldsymbol{z}_{i,jk}^T, \boldsymbol{z}_{ij,k}^T, y_{ijk}\right)$, for $k = 1, 2, \ldots, n_{ij}$; $j = 1, 2, \ldots, M_i$;

$i = 1, 2, \ldots, M$, and $N = \sum_{i=1}^{M} \sum_{j=1}^{M_i} n_{ij}$. Here, $\boldsymbol{x}_{ijk}^T$ is the kth row of a known $n_{ij} \times p$ matrix $\boldsymbol{X}_{ij}$, $\boldsymbol{z}_{i,jk}^T$ is the kth row of a known $n_{ij} \times q_1$ matrix $\boldsymbol{Z}_{i,j}$, $\boldsymbol{z}_{ij,k}^T$ is the kth row of a known $n_{ij} \times q_2$ matrix $\boldsymbol{Z}_{ij}$ and y_{ijk} is the kth observation of the response vector $\boldsymbol{y}_{ij} = \left(y_{ij1}, y_{ij2}, \ldots, y_{ijn_{ij}}\right)^T$ for the jth level-2 unit within the ith cluster.

3.3.2 Model

We define the τth 3-level model for quantile regression as

$$Q_{y_{ijk}|u_i,u_{ij},x_{ijk},z_{i,jk},z_{ij,k}}(\tau) = \boldsymbol{x}_{ijk}^T \boldsymbol{\beta}_\tau + \boldsymbol{z}_{i,jk}^T \boldsymbol{u}_{\tau,i}^{(1)} + \boldsymbol{z}_{ij,k}^T \boldsymbol{u}_{\tau,ij}^{(2)}, \tag{3.6}$$

for the quantile $\tau \in (0,1)$; $k = 1, 2, \ldots, n_{ij}$; $j = 1, 2, \ldots, M_i$; $i = 1, 2, \ldots, M$ and $\boldsymbol{\beta}_\tau = (\beta_{\tau,0}, \beta_{\tau,1}, \ldots \beta_{\tau,p})^T$. The $q_1 \times 1$ level-1 random effects vector $\boldsymbol{u}_{\tau,i}^{(1)}$ are assumed to be distributed according to $p_1\left(\boldsymbol{u}_{\tau,i}^{(1)} | \boldsymbol{\Psi}_\tau^{(1)}\right)$, where $\boldsymbol{\Psi}_\tau^{(1)}$ is a $q_1 \times q_1$ covariance matrix; the $q_2 \times 1$ level-2 random effects vector $\boldsymbol{u}_{\tau,ij}^{(2)}$ are assumed to be distributed according to $p_2\left(\boldsymbol{u}_{\tau,ij}^{(2)} | \boldsymbol{\Psi}_\tau^{(2)}\right)$, where $\boldsymbol{\Psi}_\tau^{(2)}$ is a $q_2 \times q_2$ covariance matrix. The random effects vectors depends on τ through $\boldsymbol{\Psi}_\tau^{(1)}$ and $\boldsymbol{\Psi}_\tau^{(2)}$. The level-1 random effects vector $\boldsymbol{u}_{\tau,i}^{(1)}$ are assumed to be independent for different i, while the level-2 random effects vector $\boldsymbol{u}_{\tau,ij}^{(2)}$ are assumed to be independent for different i and j and to be independent of the first level (level-1) random effects. We write (3.6) for the jth level-2 unit within the ith level-1 unit in matrix form as

$$Q_{\boldsymbol{y}_{ij}|\boldsymbol{u}_i,\boldsymbol{u}_{ij},\boldsymbol{X}_{ij},\boldsymbol{Z}_{i,j},\boldsymbol{Z}_{ij}}(\tau) = \boldsymbol{X}_{ij} \boldsymbol{\beta}_\tau + \boldsymbol{Z}_{i,j} \boldsymbol{u}_{\tau,i}^{(1)} + \boldsymbol{Z}_{ij} \boldsymbol{u}_{\tau,ij}^{(2)} \tag{3.7}$$

where, $j = 1, 2, \ldots, M_i$; $i = 1, 2, \ldots, M$. The objective function is given as

$$\sum_{i=1}^{M} \sum_{j=1}^{M_i} \rho_\tau \left(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij} \boldsymbol{\beta}_\tau - \boldsymbol{Z}_{i,j} \boldsymbol{u}_{\tau,i}^{(1)} - \boldsymbol{Z}_{ij} \boldsymbol{u}_{\tau,ij}^{(2)}\right) \tag{3.8}$$

where $\rho_\tau(\boldsymbol{r}) = \sum_{j=1}^{n} r_j \{\tau - I(r_j < 0)\}$ for a vector $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$, where $I(.)$ is the indicator function.

3.3.3 Inference

The minimization of (3.8) is equivalent to fitting a LQMM when the response, conditionally on the random effects, is assumed to follow the asymmetric Laplace distribution (Geraci and Bottai 2007; Geraci and Bottai 2014). So, our response vector $\boldsymbol{y}_{ij}$, conditionally on the random effects, are independently distributed as asymmetric Laplace (AL) distribution with location and scale parameters given by $\boldsymbol{\mu}_{\tau,ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta}_\tau + \boldsymbol{Z}_i\boldsymbol{u}^{(1)}_{\tau,i} + \boldsymbol{Z}_{ij}\boldsymbol{u}^{(2)}_{\tau,ij}$ and σ_τ. The τ defines the quantile levels and is usually set as a priori.

For simplification of notations, let us define $\boldsymbol{y}_i = \left(\boldsymbol{y}^T_{i1}, \boldsymbol{y}^T_{i2}, \dots, \boldsymbol{y}^T_{iM_i}\right)^T$; $\boldsymbol{y} = \left(\boldsymbol{y}^T_1, \boldsymbol{y}^T_2, \dots, \boldsymbol{y}^T_M\right)^T$; $\boldsymbol{u}^{(1)} = \left(\{\boldsymbol{u}^{(1)}_1\}^T, \{\boldsymbol{u}^{(1)}_2\}^T, \dots, \{\boldsymbol{u}^{(1)}_M\}^T\right)^T$ $i = 1, 2, \dots, M$; and the second level random effects can be written as $\boldsymbol{u}^{(2)}_i = \left(\{\boldsymbol{u}^{(2)}_{i1}\}^T, \{\boldsymbol{u}^{(2)}_{i2}\}^T, \dots, \{\boldsymbol{u}^{(2)}_{iM_i}\}^T\right)^T$; $\boldsymbol{u}^{(2)} = \left(\{\boldsymbol{u}^{(2)}_1\}^T, \{\boldsymbol{u}^{(2)}_2\}^T, \dots, \{\boldsymbol{u}^{(2)}_M\}^T\right)^T$. We assume that random-effect vectors $\boldsymbol{u}^{(1)}_{\tau,i}$ and $\boldsymbol{u}^{(2)}_{\tau,ij}$ follow zero-mean multivariate Gaussian distribution with variance-covariance matrices $\boldsymbol{\Psi}^{(1)}_\tau$ and $\boldsymbol{\Psi}^{(2)}_\tau$, respectively. The parameters of the model are $\boldsymbol{\beta}$, σ, and $\boldsymbol{\Psi}^{(1)}$ and $\boldsymbol{\Psi}^{(2)}$ and the marginal likelihood for the jth level-2 unit within the ith level-1 unit is given as (excluding the subscript τ)

$$L\left(\boldsymbol{\beta}, \sigma, \boldsymbol{\Psi}^{(1)}, \boldsymbol{\Psi}^{(2)} | \boldsymbol{y}_{ij}\right) = \prod_{i=1}^{M} \int \prod_{j=1}^{M_i} \left[\int p\left(\boldsymbol{y}_{ij} | \boldsymbol{u}^{(1)}_i, \boldsymbol{u}^{(2)}_{ij}, \boldsymbol{\beta}, \sigma\right) p_2\left(\boldsymbol{u}^{(2)}_{ij} | \boldsymbol{\Psi}^{(2)}\right) d\boldsymbol{u}^{(2)}_{ij}\right] p_1\left(\boldsymbol{u}^{(1)}_i | \boldsymbol{\Psi}^{(1)}\right) d\boldsymbol{u}^{(1)}_i \quad (3.9)$$

Let $\boldsymbol{\theta}_\tau \equiv \left(\boldsymbol{\beta}^T_\tau, \boldsymbol{\xi}^T_{\tau,1}, \boldsymbol{\xi}^T_{\tau,2}\right)^T \in \mathbb{R}^{p+m_1+m_2}$ denote the parameter of interest, where $\boldsymbol{\xi}_{\tau,1}$ is unrestricted m_1 dimensional vector, $1 \leq m_1 \leq q_1(q_1+1)/2$, of non-redundant parameters in $\boldsymbol{\Psi}^{(1)}_\tau$, and $\boldsymbol{\xi}_{\tau,2}$ is unrestricted m_2 dimensional vector, $1 \leq m_2 \leq q_2(q_2+1)/2$, of non-redundant parameters in $\boldsymbol{\Psi}^{(2)}_\tau$ (Pinheiro and Bates 1996).

We can write the likelihood function as

$$L(\boldsymbol{\theta}_\tau, \boldsymbol{y}) = \prod_{i=1}^{M} \int_{\mathbb{R}^{q_1}} \prod_{j=1}^{M_i} \left[\int_{\mathbb{R}^{q_2}} \frac{\exp\left\{ -\frac{2\rho_\tau(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{\tau,ij}) + (\boldsymbol{u}_{\tau,ij}^{(2)})^T (\tilde{\boldsymbol{\Psi}}_\tau^{(2)})^{-1} \boldsymbol{u}_{\tau,ij}^{(2)}}{2\sigma_\tau} \right\}}{(2\pi\sigma_\tau)^{\frac{q_2}{2}} \sqrt{|\tilde{\boldsymbol{\Psi}}_\tau^{(2)}|}} \times \left\{ \frac{\tau(1-\tau)}{\sigma_\tau} \right\}^{n_{ij}} d\boldsymbol{u}_{\tau,ij}^{(2)} \right] \times \frac{\exp\left[-\frac{(\boldsymbol{u}_{\tau,i}^{(1)})^T (\tilde{\boldsymbol{\Psi}}_\tau^{(1)})^{-1} \boldsymbol{u}_{\tau,i}^{(1)}}{2\sigma_\tau} \right]}{(2\pi\sigma_\tau)^{\frac{q_1}{2}} \sqrt{|\tilde{\boldsymbol{\Psi}}_\tau^{(1)}|}} d\boldsymbol{u}_{\tau,i}^{(1)} \quad (3.10)$$

where $\tilde{\boldsymbol{\Psi}}_\tau^{(1)} = \frac{\boldsymbol{\Psi}_\tau^{(1)}}{\sigma_\tau}$ and $\tilde{\boldsymbol{\Psi}}_\tau^{(2)} = \frac{\boldsymbol{\Psi}_\tau^{(2)}}{\sigma_\tau}$. Then the log-likelihood function can be written as in following

$$l(\boldsymbol{\theta}_\tau, \boldsymbol{y}) = N \log \left[\frac{\tau(1-\tau)}{\sigma_\tau} \right] - \sum_{i=1}^{M} \frac{M_i}{2} \log |\tilde{\Psi}_\tau^{(2)}| - \frac{M}{2} \log |\tilde{\Psi}_\tau^{(1)}| + \sum_{i=1}^{M} \log \int_{\mathbb{R}^{q_1}} \prod_{j=1}^{M_i} \left[\int_{\mathbb{R}^{q_2}} \frac{\exp\left[-\frac{2\rho_\tau(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{\tau,ij}) + (\boldsymbol{u}_{\tau,ij}^{(2)})^T (\tilde{\boldsymbol{\Psi}}_\tau^{(2)})^{-1} \boldsymbol{u}_{\tau,ij}^{(2)}}{2\sigma_\tau} \right]}{(2\pi\sigma_\tau)^{\frac{q_2}{2}}} d\boldsymbol{u}_{\tau,ij}^{(2)} \right] \times \frac{\exp\left[-\frac{(\boldsymbol{u}_{\tau,i}^{(1)})^T (\tilde{\boldsymbol{\Psi}}_\tau^{(1)})^{-1} \boldsymbol{u}_{\tau,i}^{(1)}}{2\sigma_\tau} \right]}{(2\pi\sigma_\tau)^{\frac{q_1}{2}}} d\boldsymbol{u}_{\tau,i}^{(1)} \quad (3.11)$$

A double approximation is applied to the log-likelihood function to estimate the parameters of interest (Geraci 2018; Geraci 2019). The loss function $\rho_\tau(r)$ is first smoothed at the kink $r = 0$ and then the integral is solved using a Laplacian approximation to the loss function (Geraci 2018). We use the following smooth approximation (Madsen and Nielsen 1993; Chen 2007)

$$\kappa_{\omega,\tau}(r) = \begin{cases} r(\tau-1) - \frac{1}{2}(\tau-1)^2\omega & \text{if } r \le (\tau-1)\omega, \\ \frac{1}{2\omega} r^2 & \text{if } (\tau-1)\omega < r < \tau\omega, \\ r\tau - \frac{1}{2}(\tau)^2\omega & \text{if } r \ge \tau\omega, \end{cases} \quad (3.12)$$

where $r \in \mathbb{R}$ and $\omega > 0$ is a scalar "tuning" parameter.

Let $\boldsymbol{r}_{ij} = \boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta} - \boldsymbol{Z}_{i,j}\boldsymbol{u}_i^{(1)} - \boldsymbol{Z}_{ij}\boldsymbol{u}_{ij}^{(2)}$ be the $n_{ij} \times 1$ vector of residuals for the jth level-2 within ith level-1 cluster with generic element r_{ijk}, and define the corresponding sign vector $\boldsymbol{s}_{ij} = \left(s_{ij1}, s_{ij2}, \ldots, s_{ijn_{ij}}\right)^T$ with

$$s_{ijk} = \begin{cases} -1 & \text{if } r_{ijk} \leq (\tau - 1)\omega, \\ 0 & \text{if } (\tau - 1)\omega < r_{ijk} < \tau\omega, \\ 1 & \text{if } r_{ijk} \geq \tau\omega \end{cases} \tag{3.13}$$

We apply the smooth approximation given in (3.12) to the elements of $\rho_\tau(r_{ijk})$ and write

$$\kappa_{\omega,\tau}\left(\boldsymbol{r}_{ij}\right) = \sum_{i=1}^{M}\sum_{j=1}^{M_i} \frac{1}{2}\left(\|\boldsymbol{r}_{ij}\|^2_{\boldsymbol{A}_{ij}} + \boldsymbol{b}_{ij}^T\boldsymbol{r}_{ij} + \boldsymbol{c}_{ij}^T\mathbf{1}_{n_{ij}}\right), \tag{3.14}$$

where $\boldsymbol{A}_{ij}$ is an $n_{ij} \times n_{ij}$ diagonal matrix with diagonal elements $\{\boldsymbol{A}_{ij}\}_{kk} = \left(1 - s^2_{ijk}\right)/\omega$, b_{ijk} and c_{ijk} are two $n_{ij} \times 1$ vectors with elements

$$b_{ijk} = s_{ijk}\left[(2\tau - 1)s_{ijk} + 1\right]$$

and

$$c_{ijk} = \frac{1}{2}\left[(1 - 2\tau)\omega s_{ijk} - (1 - 2\tau + 2\tau^2)\,\omega s^2_{ijk}\right],$$

respectively. Using $\boldsymbol{r_i} = \left(\boldsymbol{r}_{i1}^T, \boldsymbol{r}_{i2}^T, \ldots, \boldsymbol{r}_{iM_i}^T\right)^T$, $\boldsymbol{r} = \left(\boldsymbol{r}_1^T, \boldsymbol{r}_2^T, \ldots, \boldsymbol{r}_M^T\right)^T$, $\boldsymbol{A}_i = \bigoplus_{j=1}^{M_i}\boldsymbol{A}_{ij}$, $\boldsymbol{A} = \bigoplus_{i=1}^{M}\boldsymbol{A}_i$, $\boldsymbol{b_i} = \left(\boldsymbol{b}_{i1}^T, \boldsymbol{b}_{i2}^T, \ldots, \boldsymbol{b}_{iM_i}^T\right)^T$, $\boldsymbol{b} = \left(\boldsymbol{b}_1^T, \boldsymbol{b}_2^T, \ldots, \boldsymbol{b}_M^T\right)^T$, $\boldsymbol{c_i} = \left(\boldsymbol{c}_{i1}^T, \boldsymbol{c}_{i2}^T, \ldots, \boldsymbol{c}_{iM_i}^T\right)^T$, $\boldsymbol{c} = \left(\boldsymbol{c}_1^T, \boldsymbol{c}_2^T, \ldots, \boldsymbol{c}_M^T\right)^T$, $\boldsymbol{w}_\tau = \left((\boldsymbol{u}_\tau^{(2)})^T, (\boldsymbol{u}_\tau^{(1)})^T\right)^T$, $\tilde{\Phi}_\tau = (\mathbf{1}_{\sum_{i=1}^M M_i} \otimes \tilde{\Psi}_\tau^{(2)}) \bigoplus (\mathbf{1}_M \otimes \tilde{\Psi}_\tau^{(1)})$, $\boldsymbol{Z}^{(1)} = \bigoplus_{i=1}^{M}\boldsymbol{Z}_{i,j}$, $\boldsymbol{Z}^{(2)} = \bigoplus_{i=1}^{M}\left(\bigoplus_{j=1}^{M_i}\boldsymbol{Z}_{ij}\right)$ and $\boldsymbol{G} = [\boldsymbol{Z}^{(2)}\ \boldsymbol{Z}^{(1)}]$, we now define the smoothed function as follows

$$\begin{aligned} h(\boldsymbol{\theta}_\tau, \boldsymbol{y}, \boldsymbol{u}_\tau^{(1)}, \boldsymbol{u}_\tau^{(2)}) &= \sum_{i=1}^{M}\left[\sum_{j=1}^{M_i}\left\{\boldsymbol{r}_{ij}^T\boldsymbol{A}_{ij}\boldsymbol{r}_{ij} + \boldsymbol{b}_{ij}^T\boldsymbol{r}_{ij} + \boldsymbol{c}_{ij}^T\mathbf{1}_{n_{ij}} + (\boldsymbol{u}_{\tau,ij}^{(2)})^T(\tilde{\boldsymbol{\Psi}}_\tau^{(2)})^{-1}\boldsymbol{u}_{\tau,ij}^{(2)}\right\} + \right. \\ &\qquad \left. (\boldsymbol{u}_{\tau,i}^{(1)})^T(\tilde{\boldsymbol{\Psi}}_\tau^{(1)})^{-1}\boldsymbol{u}_{\tau,i}^{(1)}\right] \\ &= \boldsymbol{r}^T\boldsymbol{A}\boldsymbol{r} + \boldsymbol{b}^T\boldsymbol{r} + \boldsymbol{c}^T\mathbf{1}_N + \boldsymbol{w}_\tau^T\tilde{\Phi}_\tau^{-1}\boldsymbol{w}_\tau \end{aligned} \tag{3.15}$$

We replace the function ρ_τ in the log-likelihood (3.11) with $\kappa_{\omega,\tau}$ to obtain a smoothed likelihood. The smoothed version of log-likelihood then can be written as

$$l(\boldsymbol{\theta}_\tau, \boldsymbol{y}, \boldsymbol{w}) = N \log \frac{\tau(1-\tau)}{\sigma_\tau} - \frac{1}{2} \log \mid \tilde{\Phi}_\tau \mid + \\ \log \int_{\mathbb{R}^{\sum_{i=i}^{M} M_i q_2 + M q_1}} \frac{\exp\left[-\frac{1}{2\sigma_\tau} h(\boldsymbol{\theta}_\tau, \boldsymbol{y}, \boldsymbol{u}_\tau^{(1)}, \boldsymbol{u}_\tau^{(2)})\right]}{(2\pi\sigma_\tau)^{\frac{\sum_{i=i}^{M} M_i q_2 + M q_1}{2}}} d\boldsymbol{w}_\tau \quad (3.16)$$

Since h is differentiable with respect to $\boldsymbol{w}_\tau$, we can further derive the following quantities

$$\begin{aligned}
\frac{\partial h}{\partial \boldsymbol{u}_{\tau,ij}^{(2)}} &= -\boldsymbol{Z}_{ij}^T \left(2\boldsymbol{A}_{ij}\boldsymbol{r}_{ij} + \boldsymbol{b}_{ij}\right) + 2(\tilde{\boldsymbol{\Psi}}_\tau^{(2)})^{-1}\boldsymbol{u}_{\tau,ij}^{(2)} \\
\frac{\partial h}{\partial \boldsymbol{u}_{\tau,i}^{(1)}} &= -\sum_{j=1}^{M_i} \boldsymbol{Z}_{i,j}^T \left(2\boldsymbol{A}_{ij}\boldsymbol{r}_{ij} + \boldsymbol{b}_{ij}\right) + 2(\tilde{\boldsymbol{\Psi}}_\tau^{(1)})^{-1}\boldsymbol{u}_{\tau,i}^{(1)} \\
\frac{\partial h}{\partial \boldsymbol{u}_{\tau,ij}^{(2)} \left(\boldsymbol{u}_{\tau,ij}^{(2)}\right)^T} &= 2\left(\boldsymbol{Z}_{ij}^T \boldsymbol{A}_{ij} \boldsymbol{Z}_{ij} + (\tilde{\boldsymbol{\Psi}}_\tau^{(2)})^{-1}\right) \\
\frac{\partial^2 h}{\partial \boldsymbol{u}_{\tau,i}^{(1)} \left(\boldsymbol{u}_{\tau,i}^{(1)}\right)^T} &= \sum_{j=1}^{M_i} 2\left(\boldsymbol{Z}_{i,j}^T \boldsymbol{A}_{ij} \boldsymbol{Z}_{i,j}\right) + 2(\tilde{\boldsymbol{\Psi}}_\tau^{(1)})^{-1} \\
\frac{\partial^2 h}{\partial \boldsymbol{u}_{\tau,ij}^{(2)} \left(\boldsymbol{u}_{\tau,ij'}^{(2)}\right)^T} &= \mathbf{O}_{q_2 \times q_2}, \ j \neq j' \\
\frac{\partial^2 h}{\partial \boldsymbol{u}_{\tau,i} \left(\boldsymbol{u}_{\tau,i'}^{(1)}\right)^T} &= \mathbf{O}_{q_1 \times q_1} \ i \neq i' \\
\frac{\partial^2 h}{\partial \boldsymbol{u}_{\tau,ij}^{(2)} \left(\boldsymbol{u}_{\tau,i}^{(1)}\right)^T} &= 2\boldsymbol{Z}_{i,j}^T \boldsymbol{A}_{ij} \boldsymbol{Z}_{ij}
\end{aligned}$$

Considering all available clusters, the above derivatives can be written as in the following

$$\frac{\partial h}{\partial \boldsymbol{w}_\tau} = -\boldsymbol{G}^T \left(2\boldsymbol{A}\boldsymbol{r} + \boldsymbol{b}\right) + 2\tilde{\boldsymbol{\Phi}}_\tau^{-1}\boldsymbol{w}_\tau \quad (3.17)$$

$$\frac{\partial h}{\partial \boldsymbol{w}_\tau \boldsymbol{w}_\tau^T} = 2\left(\boldsymbol{G}^T \boldsymbol{A} \boldsymbol{G} + \tilde{\boldsymbol{\Phi}}_\tau^{-1}\right) \quad (3.18)$$

Let

$$\hat{\boldsymbol{w}}_\tau \equiv \left(\hat{\boldsymbol{u}}_\tau^{(2)}, \hat{\boldsymbol{u}}_\tau^{(1)}\right) = \underset{\boldsymbol{u}_\tau^{(2)}, \boldsymbol{u}_\tau^{(1)}}{\operatorname{argmin}}\, h(\boldsymbol{\theta}_\tau, \boldsymbol{y}, \boldsymbol{u}_\tau^{(1)}, \boldsymbol{u}_\tau^{(2)}) \tag{3.19}$$

be the conditional mode of $\hat{\boldsymbol{w}}_\tau$. A second order approximation of h around $\hat{\boldsymbol{w}}_\tau$ is given by

$$h(\boldsymbol{\theta}_\tau, \boldsymbol{y}, \boldsymbol{u}_\tau^{(1)}, \boldsymbol{u}_\tau^{(2)}) \approx h_0 + \dot{\boldsymbol{h}}^T (\boldsymbol{w}_\tau - \hat{\boldsymbol{w}}_\tau) + (\boldsymbol{w}_\tau - \hat{\boldsymbol{w}}_\tau)^T \ddot{\boldsymbol{H}} (\boldsymbol{w}_\tau - \hat{\boldsymbol{w}}_\tau)$$

where $\ddot{\boldsymbol{H}}$ is Hessian matrix and h_0 is the term of order 0 of the above Taylor-series expansion around the mode $\hat{\boldsymbol{w}}_\tau$. Applying the above second-order Taylor expansion (Pinheiro and Bates 2006) to the resulting exponent, we can write the Laplace approximation of smoothed log-likelihood function as

$$\begin{aligned}
l_{LA}(\boldsymbol{\theta}_\tau, \boldsymbol{y}, \hat{\boldsymbol{w}}_\tau) &= N \log \frac{\tau(1-\tau)}{\sigma_\tau} - \frac{1}{2} \log \mid \tilde{\Phi}_\tau \mid + \\
&\qquad \log \int_{\mathbb{R}^{\sum_{i=i}^{M} M_i q_2 + M q_1}} \frac{\exp\left[-\frac{1}{2\sigma_\tau}\{h_0 + (\boldsymbol{w}_\tau - \hat{\boldsymbol{w}}_\tau)^T \ddot{\boldsymbol{H}} (\boldsymbol{w}_\tau - \hat{\boldsymbol{w}}_\tau)\}\right]}{(2\pi\sigma_\tau)^{\frac{\sum_{i=i}^{M} M_i q_2 + M q_1}{2}}} d\boldsymbol{w}_\tau \\
&= N \log \frac{\tau(1-\tau)}{\sigma_\tau} - \frac{1}{2} \log \mid \tilde{\Phi}_\tau \mid - \frac{1}{2\sigma_\tau} h_0 + \log \left[\frac{(2\pi\sigma_\tau)^{-\frac{\sum_{i=i}^{M} M_i q_2 + M q_1}{2}}}{\sqrt{|\ddot{\boldsymbol{H}}^{-1}|}} \right. \\
&\qquad \left. \sqrt{|\ddot{\boldsymbol{H}}^{-1}|}\ \exp\{-\frac{1}{2\sigma_\tau} (\boldsymbol{w}_\tau - \hat{\boldsymbol{w}}_\tau)^T \ddot{\boldsymbol{H}} (\boldsymbol{w}_\tau - \hat{\boldsymbol{w}}_\tau)\} \right] d\boldsymbol{w}_\tau \\
&= N \log \frac{\tau(1-\tau)}{\sigma_\tau} - \frac{1}{2} \log \mid \tilde{\Phi}_\tau \mid - \frac{1}{2\sigma_\tau} h_0 + \log \sqrt{|\ddot{\boldsymbol{H}}^{-1}|} \\
&= N \log \frac{\tau(1-\tau)}{\sigma_\tau} - \frac{1}{2} \{\log |\tilde{\Phi}_\tau \ddot{\boldsymbol{H}}| + \sigma_\tau^{-1} h_0\} \qquad (3.20)
\end{aligned}$$

where $\tilde{\Phi}_\tau$ is the scaled variance-covariance matrix of $\boldsymbol{w}_\tau = \left((\boldsymbol{u}_\tau^{(2)})^T, (\boldsymbol{u}_\tau^{(1)})^T\right)^T$.

To maximize (3.20) with respect to $\boldsymbol{\theta}_\tau$, we can use a general optimizer such the Nelder-Mead or the Broyden–Fletcher–Goldfarb– Shanno (BFGS) algorithms (Geraci 2018). For a given $\boldsymbol{\theta}_\tau$, σ_τ can be estimated by optimizing (3.20) with respect to σ_τ

$$\frac{\partial l_{LA}}{\partial \sigma_\tau} = -\frac{N}{\sigma_\tau} - \frac{1}{2}\left(-\frac{1}{\sigma_\tau^2} h_0\right)$$

and setting $\frac{\partial l_{LA}}{\partial \sigma_\tau} = 0$ we get

$$\hat{\sigma}_\tau = (2N)^{-1} h_0$$

Finally, for a given value of ω, the mode $\boldsymbol{w}_\tau$ can be obtained using the Newton-Raphson (Pinheiro and Chao 2006) algorithm with the use of the equations given by (3.17) and (3.18).

Estimation of the parameters from (3.20) is carried out iteratively. The algorithm requires setting the starting value of $\boldsymbol{\theta}_\tau$, σ_τ and the tuning parameter ω, the tolerance for the change in the log-likelihood, and the maximum number of iterations respectively. Also, the modes of the random effects can be obtained by setting the equating (3.17) to 0 and then solving for $\boldsymbol{w}_\tau$. This leads to the following system of equations as in following

$$2\ddot{\boldsymbol{H}}\boldsymbol{w}_\tau = \boldsymbol{G}^T\left(2\boldsymbol{A}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_\tau\right) + \boldsymbol{b}\right) \tag{3.21}$$

As the right-hand side of (3.21) depends on $\boldsymbol{w}_\tau$ through $\boldsymbol{A}$ and $\boldsymbol{b}$, an estimate $\hat{\boldsymbol{w}}_\tau$ is obtained iteratively.

At each iteration, the parameter ω is reduced by a given factor (for example, by half). At convergence, the value of ω should be small, ideally, since the approximation of $\kappa_{\omega,\tau}(r)$ to the loss function ρ_τ improves as ω decreases, i.e. $\kappa_{\omega,\tau}(r) \to \rho_\tau(r)$ for $\omega \to 0$. Set T as the maximum number of iterations and the factor $0 < \gamma < 1$ for reducing the tuning parameter ω at each iteration. The algorithm is described as follows

1. Initial values are obtained for $t = 0$ as

    i. Obtain an estimate of $\boldsymbol{\beta}_\tau^{(0)}$ using the multilevel linear mixed effects model.

    ii. Obtain an estimate for $\boldsymbol{\xi}_{\tau,1}^{(0)}$, $\boldsymbol{\xi}_{\tau,2}^{(0)}$ from the fitted model in (i).

    iii. Obtain en estimate for $\sigma_\tau^{(0)}$. This can be estimated as the mean of the absolute least square residuals from step 1(i).

    iv. Provide a starting value for $\omega^{(0)}$.

    v. Use initial values of $\boldsymbol{\beta}_\tau^{(0)}$, $\boldsymbol{\xi}_{\tau,1}^{(0)}$, $\boldsymbol{\xi}_{\tau,2}^{(0)}$ and $\sigma_\tau^{(0)}$ in (3.19) to obatin $\boldsymbol{w}_\tau^{(\mathbf{0})}$

2. For $t < T$

   i. Update $\boldsymbol{\theta}_{\tau}^{(t)}$ by minimizing (3.20)

   ii. If the change in (3.20) is smaller than a given tolerance then return $\boldsymbol{\theta}_{\tau}^{(t+1)}$ else set $\boldsymbol{\theta}_{\tau}^{(t+1)} = \boldsymbol{\theta}_{\tau}^{(t)}$; $\omega^{(t+1)} = \gamma\omega^{(t)}$; $t + 1 = t$.

   iii. Repeat 2(i) and 2(ii).

3. Update $\sigma_{\tau}^{(t)}$ and $\boldsymbol{w}_{\tau}^{(t)}$

To obtain consistent estimators of bias and standard errors, resampling methods are used. Among a number of general resampling approaches, we considered bootstrap. As the hierarchical data structure should be considered for re-sampling for multilevel models (Leeden, Meijer, and Busing 2008), we considered multilevel bootstrapping here. We considered cases bootstrap approach where entire cases were re-sampled and data structure generation was maintained. To select each bootstrap sample from the original sample, we first obtained a with replacement sample of the highest level clusters, and we again sampled the second highest level units with replacement in the selected highest level clusters. Finally, all lowest level units within each selected second level clusters were considered to form the bootstrap sample.

3.4 Simulation

We carried out an extensive simulation study. Data were generated according to the following linear mixed model

$$y_{ijk} = \left(\beta_0 + u_i^{(1)} + u_{ij}^{(2)}\right) + \left(\beta_1 + v_i^{(1)} + v_{ij}^{(2)}\right) x_{1ijk} + \beta_2 x_{2ijk} + (1 + \zeta x_{1ijk})\, e_{ijk}$$

$k = 1, 2, \ldots, n_{ij}$; $j = 1, 2, \ldots, M_i$; $i = 1, 2, \ldots, M$, where $\boldsymbol{\beta} = (2.0, 8.0, 5.0)^T$, $u_i^{(1)}$ and $v_i^{(1)}$ are cluster specific level 1 random effects, $u_{ij}^{(2)}$ and $v_{ij}^{(2)}$ are cluster specific level 2 random effects within level 1, $x_{1ijk} = \delta_i + \lambda_{ij} + \nu_{ijk}$, $\delta_i \sim N(0, 1)$, $\lambda_{ij} \sim N(0, 1)$, $\nu_{ijk} \sim N(0, 1)$, x_{2ijk} is a dichotomous variable generated from a binomial distribution

with success probability 0.4. In all heteroscedastic models ($\zeta \neq 0$), δ_i, λ_{ij} and ν_{ijk} were generated as standard uniforms to ensure a positive scale parameter and to avoid quantile-crossings. We considered 5 different data generating scenarios. A summary of these simulations scenarios is given in Table 3.1. In all cases, data were generated 500 times, independently.

The error was generated from either a normal distribution with mean 0 and variance 5, a t-distribution with 3 degrees of freedom or a χ^2-distribution with 2 degrees of freedom. For each model, a balanced data set was generated according to 2 sample size combinations of $(M, M_i, n_{ij}) \in \{(10, 10, 10), (20, 10, 10)\}$.

All models were estimated for $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ using Nelder–Mead algorithm to maximize the approximated log-likelihood with tolerance parameter and maximum number of iterations set to 10^{-5} and 500, respectively. The modal random effects were estimated using a Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with gradient calculated as in (3.17). The fixed and random-effects parameters were initialized using the estimates from a multilevel linear mixed effects model and the mean of the absolute least squares residuals was used as initial value for σ_τ. The tuning parameter ω was set at half the standard deviation of response variable and subsequently halved at each iteration.

For each scenario in Table 3.1, as a measure of performance, we assessed the relative bias and the coverage rate of different quantile estimator's confidence intervals at 90% nominal level for different data scenarios. The standard errors were calculated using $R = 200$ bootstrap replications as described at the end of the section 3.3.3. We also determined the Proportion of Negative Residuals (PNR) which is expected to be approximately equal to τ. All summary measures were averaged over all replications.

The results are shown in the following Tables. Overall, the performance of the proposed model estimator was satisfactory, except in some specific cases. However, relative biases for intercepts were observed higher in the lower quantile of location-

shift symmetric, location-shift symmetric with $cov(u, v) > 0$ data sets for both combination of sample sizes. For location-shift heavy tailed sets, only intercept at 10th quantile for small sample size $(10, 10, 10)$ shows higher relative bias. Heteroscedasticity symmetric data sets have severely high relative biases at both upper and lower quantiles for both sample size combinations. In all cases, binary covariate provided zero relative biases for almost all of the scenarios of data sets.

Coverage rates and PNR are also reported in Tables. Coverage rates were close to the nominal level in most cases. However, in case of heteroscedasticity, as shown in Table 3.6, coverage rates in the tails were deviated a lot from the nominal level. Further investigations may be necessary to explore the possible reasons for such deviation. As expected, PNR are approximately equal to the τ in all scenarios.

Table 3.1 Simulation study scenarios. Distributions: $\boldsymbol{N}_g(\mathbf{0}, \boldsymbol{\Psi})$, g variate normal with mean vector $(0_1, ..., 0_g)^T$ and $g \times g$ variance-covariance matrix Ψ; t_g, t with g degrees of freedom; χ^2_g, chi-square with g degrees of freedom

Model Description	Sample Size M, M_i, n_{ij}	$u^{(1)}, v^{(1)}$ $\boldsymbol{N}_2(\mathbf{0}, \boldsymbol{\Psi}^{(1)})$	$u^{(2)}, v^{(2)}$ $\boldsymbol{N}_2(\mathbf{0}, \boldsymbol{\Psi}^{(2)})$	e	ζ
Location-shift symmetric	10, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 0 \\ 0 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.64 \end{bmatrix}$	$N(0,5)$	0
Location-shift symmetric with $cov(u,v) > 0$	10, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 1.5 \\ 1.5 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 1.5 \\ 1.5 & 0.64 \end{bmatrix}$	$N(0,5)$	0
Location-shift heavy-tailed	10, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 0 \\ 0 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.64 \end{bmatrix}$	t_3	0
Location-shift asymmetric	10, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 0 \\ 0 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.64 \end{bmatrix}$	χ^2_2	0
Heteroscedastic symmetric	10, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 0 \\ 0 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.64 \end{bmatrix}$	$N(0,5)$	0.25
Location-shift symmetric	20, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 0 \\ 0 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.64 \end{bmatrix}$	$N(0,5)$	0
Location-shift symmetric with $cov(u,v) > 0$	20, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 1.5 \\ 1.5 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 1.5 \\ 1.5 & 0.64 \end{bmatrix}$	$N(0,5)$	0
Location-shift heavy-tailed	20, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 0 \\ 0 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.64 \end{bmatrix}$	t_3	0
Location-shift asymmetric	20, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 0 \\ 0 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.64 \end{bmatrix}$	χ^2_2	0
Heteroscedastic symmetric	20, 10, 10	$\boldsymbol{\Psi}^{(1)} = \begin{bmatrix} 4.0 & 0 \\ 0 & 2.25 \end{bmatrix}$	$\boldsymbol{\Psi}^{(2)} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.64 \end{bmatrix}$	$N(0,5)$	0.25

Table 3.2 Results on parameter estimation performance from analysis of 500 simulated location-shift symmetric datasets at nominal 90% levels.

	$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.50$		
	Intercept	X_1	X_2	Intercept	X_1	X_2	Intercept	X_1	X_2
$(M, M_i, n_{ij}) = (10, 10, 10)$									
Relative Bias	-0.355	0.002	0.002	0.308	-0.004	0.002	0.011	-0.006	0.000
Coverage in %	83.6	87.0	96.6	86.0	86.8	97.2	86.8	86.0	97.2
PNR	0.101			0.251			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	-0.029	-0.008	-0.001	-0.051	-0.014	0.000			
Coverage in %	87.0	86.4	97.2	85.2	85.6	96.6			
PNR	0.750			0.899					
$(M, M_i, n_{ij}) = (20, 10, 10)$									
Relative Bias	-0.284	0.008	0.000	0.221	0.003	0.001	-0.008	0.001	0.001
Coverage in %	85.9	86.7	96.2	89.0	88.4	98.4	90.0	90.2	97.8
PNR	0.100			0.250			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	-0.040	-0.001	0.001	-0.059	-0.006	0.001			
Coverage in %	87.1	89.0	97.6	84.1	89.2	97.0			
PNR	0.750			0.900					

Table 3.3 Results on parameter estimation performance from analysis of 500 simulated Location-shift symmetric with $cov(u,v) > 0$ datasets at nominal 90% levels.

	$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.50$		
	Intercept	X_1	X_2	Intercept	X_1	X_2	Intercept	X_1	X_2
$(M, M_i, n_{ij}) = (10, 10, 10)$									
Relative Bias	-0.374	0.005	0.001	0.339	-0.001	0.001	0.019	-0.003	0.000
Coverage in %	83.7	88.2	96.2	85.5	89.4	96.6	85.7	89.4	97.6
PNR	0.100			0.251			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	-0.025	-0.005	-0.001	-0.048	-0.010	0.000			
Coverage in %	87.1	89.6	96.2	85.5	89.2	97.0			
PNR	0.749			0.899					
$(M, M_i, n_{ij}) = (20, 10, 10)$									
Relative Bias	-0.322	0.006	0.000	0.305	0.001	-0.002	0.009	-0.001	0.000
Coverage in %	83.4	88.2	98.0	89.0	89.0	96.6	86.6	88.6	97.4
PNR	0.100			0.250			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	-0.031	-0.002	0.000	-0.050	-0.007	0.001			
Coverage in %	85.6	89.0	96.2	81.6	88.2	97.4			
PNR	0.750			0.900					

Table 3.4 Results on parameter estimation performance from analysis of 500 simulated location-shift heavy-tailed datasets at nominal 90% levels.

	$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.50$		
	Intercept	X_1	X_2	Intercept	X_1	X_2	Intercept	X_1	X_2
$(M, M_i, n_{ij}) = (10, 10, 10)$									
Relative Bias	0.253	-0.002	-0.001	0.026	0.002	-0.002	0.019	0.001	-0.001
Coverage in %	87.1	83.9	98.0	89.4	87.8	96.0	88.2	87.8	96.2
PNR	0.101			0.250			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	0.018	0.000	-0.001	-0.008	-0.002	-0.001			
Coverage in %	87.6	88.2	95.6	88.8	88.2	98.0			
PNR	0.750			0.899					
$(M, M_i, n_{ij}) = (20, 10, 10)$									
Relative Bias	0.099	0.004	-0.001	-0.024	0.001	-0.001	-0.008	0.001	-0.001
Coverage in %	90.1	89.1	97.3	89.7	89.3	97.3	89.3	89.3	96.3
PNR	0.100			0.250			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	0.000	0.000	-0.001	-0.020	-0.002	0.001			
Coverage in %	89.7	88.7	96.3	89.9	88.1	97.3			
PNR	0.750			0.900					

Table 3.5 Results on parameter estimation performance from analysis of 500 simulated location-shift asymmetric datasets at nominal 90% levels.

	$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.50$		
	Intercept	X_1	X_2	Intercept	X_1	X_2	Intercept	X_1	X_2
$(M, M_i, n_{ij}) = (10, 10, 10)$									
Relative Bias	0.015	0.001	0.000	0.096	-0.003	0.001	0.069	-0.005	-0.001
Coverage in %	84.8	85.2	96.0	83.0	85.0	96.8	83.8	85.8	96.8
PNR	0.101			0.251			0.501		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	0.009	-0.005	0.000	-0.043	-0.010	0.001			
Coverage in %	86.6	85.4	96.8	84.8	87.0	97.2			
PNR	0.750			0.899					
$(M, M_i, n_{ij}) = (20, 10, 10)$									
Relative Bias	0.002	0.003	-0.001	0.089	0.001	-0.001	0.064	-0.001	0.000
Coverage in %	87.7	88.3	96.4	83.1	88.3	97.2	83.7	87.7	97.0
PNR	0.100			0.250			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	0.005	-0.002	0.001	-0.042	-0.006	0.000			
Coverage in %	88.3	88.5	96.4	82.5	88.5	96.2			
PNR	0.750			0.900					

Table 3.6 Results on parameter estimation performance from analysis of 500 simulated heteroscedastic symmetric datasets at nominal 90% levels.

	$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.50$		
	Intercept	X_1	X_2	Intercept	X_1	X_2	Intercept	X_1	X_2
$(M, M_i, n_{ij}) = (10, 10, 10)$									
Relative Bias	-0.888	0.106	-0.005	0.630	0.047	-0.001	0.002	0.001	-0.001
Coverage in %	86.2	77.6	95.8	91.4	89.2	96.4	91.4	90.6	96.6
PNR	0.101			0.251			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	-0.086	-0.039	-0.002	-0.167	-0.081	-0.003			
Coverage in %	89.2	84.6	96.2	83.4	80.4	95.8			
PNR	0.749			0.899					
$(M, M_i, n_{ij}) = (20, 10, 10)$									
Relative Bias	-0.932	0.117	-0.001	0.684	0.051	-0.003	0.000	0.001	-0.004
Coverage in %	79.8	62.2	97.6	90.8	84.4	98.2	94.4	93.4	98.0
PNR	0.101			0.250			0.500		
	$\tau = 0.75$			$\tau = 0.90$					
	Intercept	X_1	X_2	Intercept	X_1	X_2			
Relative Bias	-0.101	-0.044	-0.002	-0.169	-0.096	-0.003			
Coverage in %	90.4	84.6	98.0	78.8	63.6	97.2			
PNR	0.750			0.900					

3.5 Application

The Millennium Cohort Study (MCS) is an observational, multidisciplinary cohort study of children born in the United Kingdom (UK), which is nationally representative of the total UK population at baseline and conducted by the Centre for Longitudinal Studies (CLS) at the University of London. This is an ongoing study consist of a sample of all children born between September 2000 and January 2002, alive and living in England, Scotland, Wales, or Northern Ireland at age 9 months, and eligible to receive child benefit at that age (Plewis et al. 2007). In the first sweep known as MCS1 that occurred at age 9 months, 18,552 families (18,827 children) were recruited to the cohort and information was collected on 18,818 children. Subsequent sweeps have occurred when the children were aged 3 (MCS2), 5 (MCS3), 7 (MCS4), 11 (MCS5), 14 (MCS6), 17 (MCS7) and 22 (MCS8) years. On a broad range, information of demographic, behavioral, developmental, health, and parental socioeconomic characteristics were recorded at first sweep of MCS (Connelly and Platt 2014). Trained interviewers were required to interview the subjects with information provided by mothers. Ethical approval for data collection at each sweep of MCS and accelerometer studies has been granted and approval for seasonal accelerometer and calibration studies was granted by the University College London Research Ethics Committee (Griffiths et al. 2013b). The MCS data are freely available to the researchers under standard access conditions through the UK Data Service (http://ukdataservice.ac.uk) and the MCS website provides detailed information on the study (`http://www.cls.ioe.ac.uk/mcs`).

Physical activity levels (15 second sampling epochs) was recorded at 7 years using accelerometers (Actigraph GT1M, Pensacola, Florida) to measure activity reliably in children (Ott et al. 2000). Children were instructed to wear the accelerometers, on an elastic belt around the waist, during waking time for seven consecutive days, excluding bathing or during water time activities. After eliminating non-wear time

and extreme readings, a valid data of ≥ 10 hours on at least 2 days (Rich et al. 2013) resulting in a physical activity study sample were considered for further analyses. Data that were missing in physical activity data were not imputed. A detail of the data processing have been published by Griffiths et al. (2013a). The cut-points developed by Pulsford et al. (2011) were used to define sedentary, light, moderate, and vigorous physical activity. Moderate-to-vigorous physical activity (MVPA) was defined as $> 2,241$ accelerometer counts of activity per minute (Griffiths et al. 2013a; Griffiths et al. 2013b) and was reported in terms of number of minutes of activity per day.

Physical activity is health enhancing, and is associated with the well-being of people to lead a good life that has impact in mental, physical or psychological health. Among the UK children, from an analysis of MCS4 data, Griffiths et al. 2013a found that girls have lower physical activity than boys. It is therefore may be of interest to understand the factors, among others, that influence children's physical activity which may help to identify interventions to promote active lifestyle (Pouliou et al. 2015). We examined physical activity data from the MCS4 which took place at age 7. Our sample comprised 4,738 singleton children (2,321 boys, 2,417 girls) with complete data for England only. There are 253 electoral wards in our data. We considered average time per day in MVPA as an outcome variable. Several covariates that we considered for this study include: sex (binary, reference: boy), ethnic group (binary, reference: white), income quintiles (categorical, reference: fifth quintile), time spent reading for enjoyment (binary, reference: often), mode of transport to/from school (binary, reference: active), number of cars or vans owned (categorical, reference: two), body mass index (BMI) and weight (in kg) of the child. A summary of the data set is given in Table 3.7 where boys and girls of the sample are approximately equal (49% vs 51%). White children (81.2%) dominate the other ethnic grouped children (18.8%). Most of the children in the selected sample are in the fifth income quintile (24.1%),

followed by fourth income quintile (22.6%), third income quintile (20.3%), second income quintile (17.2%) and first income quintile (15.7%). Most of the children read often for pleasure (85.0%). Children who go schools by car or bus are fewer (45.4%) as compared to their peers (54.6%). About half of the children's family own two cars (46.7%). The five number summary for BMI, weights and daily MVPA are given in the Table 3.7 as well.

Table 3.7 Demographic and Socio-demographic information of children from the Millennium Cohort Study

Characteristics	Children (%)	Five number summary (Min, Q1, Median, Q3, Max)
Gender		
Boys	2321 (49.0)	
Girls	2417 (51.0)	
Ethnicity		
White	3846 (81.2)	
Other	892 (18.8)	
Income quintile		
1	744 (15.7)	
2	816 (17.2)	
3	964 (20.3)	
4	1070 (22.6)	
5	1144 (24.1)	
Reading for pleasure		
Often	4025 (85.0)	
Not often	713 (15.0)	
Transportation to/from school		
Active	2589 (54.6)	
Passive	2149 (45.4)	
Number of cars or vans owned		
0	480 (10.1)	
1	1781 (37.6)	
2	2211 (46.7)	
3 or more	266 (5.6)	
BMI (kg/m^2)		(10.9, 15.1, 16.1, 17.5, 44.6)
Weight (kg)		(13.2, 22.1, 24.5, 27.6, 63.4)
MVPA		(0.0, 12.9, 37.9, 59.7, 600.0)

Note: MVPA = Moderate to vigorous physical activity.

We considered child (repeated measurements of daily average MVPA within child) and electoral ward (children within wards) as grouping factors from MCS data. The results are provided in table 3.8 for $\tau \in (0.10, 0.25, 0.50, 0.75, 0.90)$. The τth 3-level quantile linear mixed effect model for MCS data, using the similar notaion of the model (3.6), was specified as

$$Q^*_{y_{ijk}|u_i,u_{ij},x_{ijk}}(\tau) = \beta_{\tau,0} + \beta_{\tau,1}\text{gender}_{k,1} + \beta_{\tau,2}\text{ethnicity}_{k,1} + \beta_{\tau,3}\text{income}_{k,1} + \beta_{\tau,4}\text{income}_{k,2}$$
$$+ \beta_{\tau,5}\text{income}_{k,3} + \beta_{\tau,6}\text{income}_{k,4} + \beta_{\tau,7}\text{reading}_{k,1} + \beta_{\tau,8}\text{transport}_{k,1} + \beta_{\tau,9}\text{cars}_{k,0} + \beta_{\tau,10}\text{cars}_{k,1}$$
$$+ \beta_{\tau,11}\text{cars}_{k,3} + \beta_{\tau,12}\text{bmi} + \beta_{\tau,13}\text{weight} + {u^1}_{\tau,i} + {u^2}_{\tau,ij} \quad (3.22)$$

We considered electoral wards as the first level random effect $({u^1}_{\tau,i})$ and children within wards as the second level random effect $({u^2}_{\tau,ij})$. We fitted model for random intercepts only at both levels. The intercept of (3.22) can be interpreted as the τth quantile function of average daily moderate to vigorous physical activity for a boy of white ethnicity with no specific BMI or weight living in a household in the highest income quintile and with two cars, who reads often (at least once or twice a week) and walks or bikes from/to school. Estimates of the fixed effects and confidence intervals from the proposed model are reported in Table 3.8. Standard errors were obtained using the bootstrap approach described in Section 3.4 with R = 100 bootstrap replications.

Some of the finding are consistent with those from previous analyses of Griffiths et al. (2013a) and Geraci (2018), that is, girls are less active than their peers; reading frequently during the week is negatively associated with activity.

We also compared the estimates from our model to LQMM (Geraci and Bottai 2014) and multilevel linear mixed model (LME), that focused on the central part of the distribution, provided in Table 3.9 and 3.10 respectively. There is larger differences across quantiles in activity levels for girls and children of ethnicity other than white than LQMM and LME. Activity is higher in children from less affluent

households at the tail quantiles. However, the estimates of the coefficients for income are statistical non-significant at the 95% level. The effects associated with reading does not seem to be important. Children who go schools by car or bus have a statistically significant lower physical activity as compared to their peers and there is a large difference across quantile than LQMM and LME. In contrast, there are marked differences between children living in households with two vehicles (reference) and those with none, the latter being substantially more active, specially for 75th and 90th quantiles. The effects associated with BMI does not seem to be important, statistically. However, in the upper quantile, it is seen that increasing weight results more physical activity and it is statistically significant.

Table 3.8 Estimated fixed effects and their confidence intervals (CI) from the multilevel linear quantile mixed model for MCS physical activity data.

Characteristics	β (95%CI)				
	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Intercept	**17.14 (9.55, 24.72)**	**13.31 (2.77, 23.85)**	**46.93 (38.17, 55.69)**	**67.25 (59.32, 75.18)**	**81.89 (68.61, 95.18)**
Gender					
Girls	-0.24 (-0.86, 0.39)	-1.67 (-3.48, 0.14)	**-8.13 (-9.55, -6.71)**	**-11.97 (-13.79, -10.16)**	**-14.67 (-17.19, -12.16)**
Boys	Ref	Ref	Ref	Ref	Ref
Ethnicity					
Not white	0.40 (-0.68, 1.47)	**2.99 (0.67, 5.32)**	-1.63 (-3.58, 0.33)	**-4.49 (-7.00, -1.98)**	**-5.92 (-9.60, -2.23)**
White	Ref	Ref	Ref	Ref	Ref
Income quintile					
1	0.81 (-0.64, 2.25)	-2.73 (-6.11, 0.65)	-0.15 (-2.96, 2.66)	-0.27 (-3.70, 3.16)	4.71 (-1.26, 10.68)
2	1.18 (-0.21, 2.56)	1.77 (-1.39, 4.93)	-0.10 (-2.36, 2.17)	1.14 (-1.99, 4.27)	3.95 (-1.28, 9.17)
3	**1.29 (0.07, 2.51)**	0.79 (-1.35, 2.92)	-0.11 (-2.39, 2.18)	0.39 (-2.52, 3.30)	3.03 (-1.31, 7.37)
4	0.22 (-0.91, 1.35)	2.25 (-0.30, 4.80)	-0.21 (-2.24, 1.82)	-0.21 (-2.79, 2.38)	1.53 (-2.02, 5.09)
5	Ref	Ref	Ref	Ref	Ref
Reading for pleasure					
Not often	0.53 (-0.52, 1.58)	-1.60 (-4.87, 1.67)	0.99 (-1.18, 3.16)	1.47 (-1.27, 4.20)	3.70 (-0.50, 7.91)
Often	Ref	Ref	Ref	Ref	Ref
Transportation					
Passive	-0.08 (-0.83, 0.66)	**-2.35 (-4.06, -0.63)**	**-2.75 (-4.20, -1.31)**	**-2.89 (-4.76, -1.03)**	**-3.29 (-6.55, -0.02)**
Active	Ref	Ref	Ref	Ref	Ref
Number of cars or van					
0	-1.22 (-2.67, 0.23)	2.67 (-1.29, 6.64)	**4.23 (0.85, 7.60)**	**6.85 (2.77, 10.93)**	**7.18 (0.45, 13.90)**
1	-0.44 (-1.28, 0.40)	1.53 (-0.47, 3.52)	1.44 (-0.16, 3.04)	**2.80 (0.80, 4.80)**	2.31 (-0.88, 5.50)
≥ 3	1.14 (-1.28, 0.40)	3.17 (-0.77, 7.10)	1.56 (-1.98, 5.10)	1.27 (-2.92, 5.46)	1.31 (-4.31, 6.93)
2	Ref	Ref	Ref	Ref	Ref
BMI (kg/m^2)	**-1.39 (-2.42, -0.35)**	0.02 (-1.23, 1.28)	-0.12 (-1.52, 1.29)	**-1.24 (-2.19, -0.29)**	0.15 (-1.62, 1.92)
Weight (kg)	0.33 (-0.12, 0.78)	-0.02 (-0.59, 0.55)	-0.07 (-0.75, 0.62)	**0.75 (0.32, 1.19)**	0.16 (-0.66, 0.98)

Table 3.9 Estimated fixed effects and their confidence intervals (CI) from the multilevel linear mixed effects model for MCS physical activity data.

Characteristics	β (95%CI)				
	Linear quantile mixed model				
	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Intercept	**38.45 (33.18, 43.72)**	**45.72 (41.12, 50.32)**	**46.98 (42.33, 51.62)**	**48.41 (43.67, 53.15)**	**49.83 (44.53, 55.14)**
Gender					
Girls	-1.28 (-2.27, 0.20)	**-5.04 (-6.21, -3.86)**	**-7.59 (-8.56, -6.62)**	**-9.24 (-10.36, -8.11**)	**-10.47 (-12.29, -8.64)**
Boys	Ref	Ref	Ref	Ref	Ref
Ethnicity					
Not white	0.24 (-0.94, 1.42)	-1.18 (-2.55, 0.19)	**-2.41 (-3.74, -1.08**)	**-2.86 (-4.43, -1.29)**	**-3.22 (-5.61, 0.84)**
White	Ref	Ref	Ref	Ref	Ref
Income quintile					
1	**-2.47 (-4.43, -0.51)**	-1.23 (-3.35, 0.88)	-0.84 (-2.95, 1.26)	-0.34 (-2.47, 1.80)	0.29 (-2.23, 2.81)
2	-1.56 (-3.40, 0.29)	-0.15 (-2.28, 1.98)	0.14 (-2.00, 2.27)	0.49 (-1.72, 2.70)	1.01 (-1.81, 3.84)
3	-1.30 (-2.74, 0.14)	-0.73 (-2.38, 0.91)	-0.12 (-1.78, 1.53)	0.27 (-1.57, 2.12)	1.08 (-2.05, 4.21)
4	-1.12 (-2.36, 0.12)	-0.18 (-1.72, 1.35)	-0.32 (-1.86, 1.22)	-0.17 (-1.76, 1.41)	-0.22 (-2.13, 1.69)
5	Ref	Ref	Ref	Ref	Ref
Reading for pleasure					
Not often	-1.20 (-2.68, 0.28)	-0.27 (-1.91, 1.38)	0.46 (-1.08, 2.00)	1.44 (-0.41, 3.29)	**2.78 (0.49, 5.07)**)
Often	Ref	Ref	Ref	Ref	Ref
Transportation					
Passive	**-1.05 (-1.79, -0.30)**	**-2.09 (-3.16, -1.02)**	**-2.45 (-3.49, -1.40)**	**-2.06 (-3.14, -0.99)**	**-1.97 (-3.77, -0.18)**
Active	Ref	Ref	Ref	Ref	Ref
Number of cars or van					
0	1.40 (-0.99, 3.78)	4.29 (1.76, 6.82)	**4.70 (2.36, 7.03)**	**5.46 (3.02, 7.89)**	**6.54 (3.67, 9.42)**
1	-0.16 (-1.26, 0.95)	0.89 (-0.46, 2.23)	1.58 (0.14, 3.02)	**2.29 (0.75, 3.84)**	**2.14 (0.21, 4.08)**
≥ 3	0.02 (-2.06, 2.09)	-0.28 (-2.54, 1.98)	-0.24 (-2.58, 2.11)	-0.37 (-2.98, 2.23)	-0.52 (-3.34, 2.30)
2	Ref	Ref	Ref	Ref	Ref
BMI (kg/m^2)	**-2.75 (-3.29, -2.21)**	**-2.33 (-2.97, -1.69)**	**-0.61 (-1.17, -0.05)**	**1.13 (0.48, 1.78)**	**3.08 (1.81, 4.35)**
Weight(kg)	**0.56 (0.33, 0.78)**	**0.57 (0.23, 0.91)**	**0.29 (0.03, 0.55)**	-0.14 (-0.47, 0.18)	-0.62 (-1.40, 0.16)

Table 3.10 Estimated fixed effects and their confidence intervals (CI) from the multilevel linear mixed effects model for MCS physical activity data.

Characteristics	Multilevel mixed model, β (95%CI)
Intercept	**47.45 (43.03, 51.86)**
Gender	
Girls	**-7.09 (-8.22, -5.96)**
Boys	Ref
Ethnicity	
Not white	**-2.10 (-3.63, -0.58)**
White	Ref
Income quintile	
1	-0.32 (-2.38, 1.75)
2	0.23 (-1.64, 2.09)
3	-0.32 (-2.01, 1.37)
4	-0.05 (-1.66, 1.57)
5	Ref
Reading for pleasure	
Not often	0.75 (-0.85, 2.34)
Often	Ref
Transportation	
Passive	**-1.94 (-3.10, -0.78)**
Active	Ref
Number of cars or van	
0	**5.22 (2.96, 7.47)**
1	**1.41 (0.10, 2.72)**
≥ 3	-0.51 (-2.97, 1.95)
2	Ref
BMI (kg/m^2)	-0.03 (-0.54, 0.47)
Weight(kg)	0.70 (-0.17, 0.31)

3.6 Discussion

We have developed a novel model for quantile regression with more than 2 levels when data are clustered. Ours has unique features as compared to alternative approaches, as we considered frequentist method for estimation of parameters from the mixed effect model. As shown in a simulation study, the performance of our model was satisfactory. This type of model can be used by the researches who mainly interest in analysis of 3-level data for different quantiles. We conducted a simulation study

that illustrates the utility of estimating parameters for a 3-level quantile mixed effect model. For large type of data sets as those illustrated in MCS physical activity analysis, where multiple levels are of interest, this model can be used.

Standard error calculation in our proposed model is facilitated by bootstrap. However, the bootstrap takes time for large data sets in computation. Further research is needed to develop accurate 'sampling-free' approximations of standard errors.

Finally, parameter estimates in our proposed model are obtained by the optimization algorithm rather than being calculated post hoc as shown in the estimation for LQMM based on linear quadrature (Geraci and Bottai 2014). However, parameter estimations is computationally intensive. A possible improvement in computing speed of our proposed algorithm is needed and is part of future research.

Chapter 4

Conclusion

To study the entire distribution or the tails of a distribution, a regression analysis that complements the mean regression analysis is known as quantile regression (QR). The use of QR is popularized by Koenker and Bassett (1978). QR has a number of properties, such as, no assumption on the distribution of the response variable is required (Koenker and Bassett 1978; Bassett and Koenker 1978), it is equivalent to monotone transformations and robustness to outliers (Huber 1981). Because of its flexibility in use, QR has become an interesting statistical tool for many researchers in a wide range of areas. Since then many statistical analysis for independent data using QR have been developed. However, lack of independence between observations offers methodological challenges in regression analysis. It is important to analyze correlated or dependent data that arise in a wide variety of studies. A number of sampling designs such as cluster, multilevel and repeated measures (longitudinal data) introduce such dependence. The correlation between observations on same units or clusters should be accounted for to obtain correct inferences. To analyze data from complex sampling designs, mixed effects models (also known as multilevel or random-effects models) have been used frequently.

Several approaches to QR for dependent data have been proposed in the last few years. Koenker (2004) estimated quantile functions with subject-specific fixed effects, where variability in the estimation process is controlled by some form of shrinkage that requires a suitable choice of penalization. Geraci and Bottai (2007) proposed linear QR model with a subject-specific random intercepts that accounts for within group

correlation by using asymmetric Laplace density. Models that incorporate random slopes include Geraci and Bottai (2014). Other approaches to QR with correlated data are given by Jung (1996), Liu and Bottai (2009), Galvao and Montes-Rojas (2010), Yuan and Yin (2010), Kim and Yang (2011), Fu and Wang (2012), Farcomeni (2012), and Li, Dowling, and Chappell (2015). Currently there are no proposal for mixed effects QR models with more than 2 levels.

Conditional modeling of quantiles (with either fixed or random cluster-specific effects) may not be ideal when the focus is on the marginal quantile effects. Marginal modeling of quantiles of correlated responses provides alternative approach. For example, Wang and Zhu (2011) defined an empirical likelihood under the generalized estimating equation (GEE) framework. Jung (1996) preserved marginal effects by incorporating correlated errors in a quasi-likelihood model. Another way to make inference on marginal effects is to adjust estimates from a conditional model when the latter does not naturally lead to marginal interpretations. This is the case in generalized linear mixed model (GLMM) (Zeger, Liang, and Albert 1988; Lesaffre et al. 2000; Gory, Craigmile, and MacEachern 2016) and QR with random effects.

Given the limitations in the literature, in this dissertation, we presented efforts providing linear quantile mixed effects model extension and marginally interpretable model for QR mixed effects model with the potentials of filling the gaps.In this concluding chapter, we summarize the specific contributions made, and some avenue of future research will also be discussed.

In Chapter 2 of this research work, we presented a marginally interpretable model for QR with random effects. An advantage of this model is that it can be more robust to miss-specification of random effects. We discussed the derivative free numerical integral algorithm with Gauss-Hermite quadrature to the proposed model parameter estimations. Our proposed model were evaluated using statistical simulations. The standard errors were calculated using bootstrap replications with replacement.

In Chapter 3, we proposed a 3-level mixed-effects model for QR. The proposed model is an extension of 2-level linear mixed effects model (Geraci and Bottai 2014). Estimation of parameters in our proposed model is carried out using a combination of a smoothing algorithm for QR and a second order Laplacian approximation for mixed models. We evaluated the proposed model using extensive statistical simulations. Bootstrap replications with replacement for cluster levels are used to obtain standard errors.

In both cases of proposed model, an application to Millennium Cohort Study (MCS) data were examined using the proposed model.

While proposed models in Chapters 2 and Chapter 3 offer new approaches than the existing approaches, and widely applicable solutions to analyze the data in respective concerns, we also discussed the specific limitations of them. The main limitations of both proposed models are that they are computationally intensive and time consuming. A further limitation of these models is that we considered these models only for continuous response.

Both of our proposed models in Chapter 2 and Chapter 3 leaves and opens avenues for future research. For both models, one can investigate the approaches to speed up the computation. Further research may also be needed to extend the proposed models. As a part of our future research for these projects, we will make R packages.

This dissertation has identified and offered solutions to certain gaps in literature concerning analyses of complex sample designs, such as dependent data analysis for quantile regression. Despite the certain limitations of the proposed models presented in this research works, we expect to make valuable additions to the existing body of related literature.

Bibliography

Active, Start (2011). "Stay Active–A report on physical activity for health from the four home countries' Chief Medical Officers (England". In: *Scotland, Wales and Northern Ireland*.

Aggio, Daniel et al. (2017). "Correlates of children's independent outdoor play: Cross-sectional analyses from the Millennium Cohort Study". In: *Preventive medicine reports* 8, pp. 10–14.

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

Bassett, G and R Koenker (1978). "Asymptotic theory of least absolute error regression". In: *Journal of the American Statistical Association* 73.363, pp. 618–622.

Burr, D and H Doss (2005). "A Bayesian semiparametric model for random-effects meta-analysis". In: *Journal of the American Statistical Association* 100.469, pp. 242–251.

Chen, C (2007). "A finite smoothing algorithm for quantile regression". In: *Journal of Computational and Graphical Statistics* 16.1, pp. 136–164.

Connelly, Roxanne and Lucinda Platt (2014). "Cohort profile: UK millennium Cohort study (MCS)". In: *International journal of epidemiology* 43.6, pp. 1719–1725.

Drikvandi, R, G Verbeke, and G Molenberghs (2017). "Diagnosing misspecification of the random-effects distribution in mixed models". In: *Biometrics* 73.1, pp. 63–71.

Farcomeni, A (2012). "Quantile regression for longitudinal data based on latent Markov subject-specific parameters". In: *Statistics and Computing* 22.1, pp. 141–152.

Firpo, S, NM Fortin, and T Lemieux (2009). "Unconditional quantile regressions". In: *Econometrica* 77.3, pp. 953–973.

Fu, Liya and You-Gan Wang (2012). "Quantile regression for longitudinal data with a working correlation model". In: *Computational Statistics & Data Analysis* 56.8, pp. 2526–2538.

Galvao, Antonio F (2011). "Quantile regression for dynamic panel data with fixed effects". In: *Journal of Econometrics* 164.1, pp. 142–157.

Galvao, Antonio F and Gabriel V Montes-Rojas (2010). "Penalized quantile regression for dynamic panel data". In: *Journal of Statistical Planning and Inference* 140.11, pp. 3476–3497.

Geraci, M (2014). "Linear quantile mixed models: the lqmm package for Laplace quantile regression". In: *Journal of Statistical Software* 57.1, pp. 1–29.

— (2016). "Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants". In: *Statistical Methods in Medical Research* 25.4, pp. 1393–1421.

— (2018). "Additive quantile regression for clustered data with an application to children's physical activity". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.

— (2019). "Modelling and estimation of nonlinear quantile regression with clustered data". In: *Computational statistics & data analysis* 136, pp. 30–46.

Geraci, M and M Bottai (2007). "Quantile regression for longitudinal data using the asymmetric Laplace distribution". In: *Biostatistics* 8.1, pp. 140–154.

— (2014). "Linear quantile mixed models". In: *Statistics and computing* 24.3, pp. 461–479.

Goldstein, H (1995). *Multilevel statistical models*. 2nd. Hodder Arnold.

Gory, Jeffrey J, Peter F Craigmile, and Steven N MacEachern (2016). "Marginally interpretable generalized linear mixed models". In: *arXiv preprint arXiv:1610.01526*.

Griffiths, L et al. (2013a). "How active are our children? Findings from the Millennium Cohort Study". In: *BMJ open* 3.8.

Griffiths, L et al. (2013b). "Technical report on the enhancement of Millennium Cohort Study data with accelerometer-derived measures of physical activity and sedentary behaviour in seven year olds". In: *London, UK*.

Huang, Q et al. (2017). "Quantile regression models and their applications: a review". In: *J Biom Biostat* 8.10.4172, pp. 2155–6180.

Huang, Y, J Chen, and H Qiu (2017). “Bayesian quantile regression for nonlinear mixed-effects joint models for longitudinal data in the presence of mismeasured covariate errors”. In: *Journal of Biopharmaceutical Statistics* 27.5, pp. 741–755.

Huber, PJ (1981). *Robust Statistics. NY.* John Wiley and Sons.

Jung, Sin-Ho (1996). “Quasi-likelihood for median regression models”. In: *Journal of the american statistical association* 91.433, pp. 251–257.

Karlsson, A (2008). “Nonlinear Quantile Regression Estimation of Longitudinal Data”. In: *Communications in statistics. Simulation and computation* 37.1-2, pp. 114–131.

Kim, Mi-Ok and Y Yang (2011). “Semiparametric approach to a random effects quantile regression model”. In: *Journal of the American Statistical Association* 106.496, pp. 1405–1417.

Koenker, R (2004). “Quantile regression for longitudinal data”. In: *Journal of Multivariate Analysis* 91.1, pp. 74–89.

Koenker, R and G Bassett (1978). “Regression Quantiles”. In: *Econometrica* 46.1, pp. 33–50.

Laird, Nan M and James H Ware (1982). “Random-effects models for longitudinal data”. In: *Biometrics*, pp. 963–974.

Lamarche, Carlos (2010). “Robust penalized quantile regression estimation for panel data”. In: *Journal of Econometrics* 157.2, pp. 396–408.

Leeden, Rien Van der, Erik Meijer, and Frank MTA Busing (2008). “Resampling multilevel models”. In: *Handbook of multilevel analysis*. Springer, pp. 401–433.

Lesaffre, E et al. (2000). “Marginalized multilevel models and likelihood inference-Comments & Rejoinder”. In: *Statistical science* 15.1, pp. 20–26.

Li, Chenxi, N Maritza Dowling, and Rick Chappell (2015). “Quantile regression with a change-point model for longitudinal data: An application to the study of cognitive changes in preclinical alzheimer’s disease”. In: *Biometrics* 71.3, pp. 625–635.

Lipsitz, Stuart R et al. (1997). “Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46.4, pp. 463–476.

Liu, Y and M Bottai (2009). “Mixed-effects models for conditional quantiles with longitudinal data”. In: *The International Journal of Biostatistics* 5.1.

Luo, Y, H Lian, and M Tian (2012). “Bayesian quantile regression for longitudinal data models”. In: *Journal of Statistical Computation and Simulation* 82.11, pp. 1635–1649.

Madsen, Kaj and Hans B Nielsen (1993). “A finite smoothing algorithm for linear l_1 estimation”. In: *SIAM Journal on Optimization* 3.2, pp. 223–235.

Marino, Maria F and A Farcomeni (2015). “Linear quantile regression models for longitudinal experiments: an overview”. In: *Metron* 73.2, pp. 229–247.

McCullagh, P and John A Nelder (1989). *Generalized linear models*. Chapman and Hill.

McCulloch, CE. and JM. Neuhaus (2011). “Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter”. In: *Statistical Science* 26.3. ISSN: 0883-4237.

Nelder, John A and Robert WM Wedderburn (1972). “Generalized linear models”. In: *Journal of the Royal Statistical Society: Series A (General)* 135.3, pp. 370–384.

Ott, AE et al. (2000). “The use of uniaxial and triaxial accelerometers to measure children’s “free-play” physical activity”. In: *Pediatric Exercise Science* 12.4, pp. 360–370.

Pinheiro, J and D Bates (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

Pinheiro, José C and Douglas M Bates (1996). “Unconstrained parametrizations for variance-covariance matrices”. In: *Statistics and computing* 6.3, pp. 289–296.

Pinheiro, José C and Edward C Chao (2006). “Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models”. In: *Journal of Computational and Graphical Statistics* 15.1, pp. 58–81.

Plewis, I et al. (2007). “Millennium cohort study: technical report on sampling”. In: *London: Centre for Longitudinal Study, Institute of Education*.

Pouliou, T et al. (2015). “Environmental influences on children’s physical activity”. In: *Journal of Epidemiology & Community Health* 69.1, pp. 77–85.

Pulsford, RM et al. (2011). "Actigraph accelerometer-defined boundaries for sedentary behaviour and physical activity intensities in 7 year old children". In: *PloS one* 6.8, e21822.

Raudenbush, Stephen W and Anthony S Bryk (2002). *Hierarchical linear models: Applications and data analysis methods*. Vol. 1. SAGE.

Reich, Brian J, Howard D Bondell, and Huixia J Wang (2010). "Flexible Bayesian quantile regression for independent and clustered data". In: *Biostatistics* 11.2, pp. 337–352.

Rich, C et al. (2013). "Quality control methods in accelerometer data processing: defining minimum wear time". In: *PloS one* 8.6, e67206.

Tian, M and G Chen (2006). "Hierarchical linear regression models for conditional quantiles". In: *Science in China Series A: Mathematics* 49.12, pp. 1800–1815.

Waldmann, E et al. (2013). "Bayesian semiparametric additive quantile regression". In: *Statistical Modelling* 13.3, pp. 223–252.

Wang, Huixia J and Z Zhu (2011). "Empirical likelihood for quantile regression models with longitudinal data". In: *Journal of statistical planning and inference* 141.4, pp. 1603–1615.

Yang, Y and X He (2012). "Bayesian empirical likelihood for quantile regression". In: *The Annals of Statistics* 40.2, pp. 1102–1131.

Yu, K and RA Moyeed (2001). "Bayesian quantile regression". In: *Statistics & Probability Letters* 54.4, pp. 437–447.

Yuan, Y and G Yin (2010). "Bayesian quantile regression for longitudinal studies with nonignorable missing data". In: *Biometrics* 66.1, pp. 105–114.

Yue, Yu R and H Rue (2011). "Bayesian inference for additive mixed quantile regression models". In: *Computational Statistics & Data Analysis* 55.1, pp. 84–96.

Zeger, Scott L, Kung-Yee Liang, and Paul S Albert (1988). "Models for longitudinal data: a generalized estimating equation approach". In: *Biometrics*, pp. 1049–1060.

Zheng, Songfeng (2011). "Gradient descent algorithms for quantile regression with smooth approximation". In: *International Journal of Machine Learning and Cybernetics* 2.3, pp. 191–207.