

Working Paper M10/07

Methodology

M-Quantile And Expectile Random Effects Regression For Multilevel Data

N. Tzavidis, N. Salvati, M. Geraci, M. Bottai

Abstract

The analysis of hierarchically structured data is usually carried out by using random effects models. The primary goal of random effects regression is to model the expected value of the conditional distribution of an outcome variable given a set of explanatory variables while accounting for the dependence structure of hierarchical data. The expected value, however, may not offer a complete picture of this conditional distribution. In this paper we propose using linear M-quantile regression, to model other parts of the conditional distribution of the outcome variable given the covariates. The proposed random effects regression model extends M-quantile regression and can be viewed as an alternative to the quantile random effects model. Inference for estimators of the fixed and random effects parameters is discussed. The performance of the proposed methods is evaluated in a series of simulation studies. Finally, we present a case study where M-quantile and expectile random effects regression is employed for analyzing repeated measures data collected from a rotary pursuit tracking experiment.

M-quantile and Expectile Random Effects Regression for Multilevel Data

N. Tzavidis* N. Salvati† M. Geraci‡ M. Bottai§

Abstract

The analysis of hierarchically structured data, for example longitudinal or geographically clustered data, is usually carried out by using random effects models. The primary goal of random effects regression is to model the expected value of the conditional distribution of an outcome variable given a set of explanatory variables while accounting for the dependence structure of hierarchical data. The expected value, however, may not offer a complete picture of this conditional distribution. In this paper we propose using linear M-quantile regression, to model other parts of the conditional distribution of the outcome variable given the covariates, which includes random intercepts to account for the dependence of hierarchical data. The proposed random effects regression model extends M-quantile regression (Breckling and Chambers, 1988) and can be viewed as an alternative to the quantile random effects model (Geraci and Bottai, 2007). M-estimation is synonymous with outlier-robust estimation. Consequently, the proposed approach allows for robust estimation of both fixed and random effects. The proposed M-estimation framework also includes expectile regression as a special case. Expectile regression can potentially lead to efficiency gains when the use of outlier-robust estimation methods is not justified but there is still interest in modelling not only the centre but also other parts of the conditional distribution.

Fixed and random effects are estimated using maximum likelihood and inference for estimators of the fixed and random effects parameters is discussed. The performance of the proposed methods is evaluated in a series of simulation studies. Finally, we present a case study where M-quantile and expectile random effects regression is employed for analyzing repeated measures data collected from a rotary pursuit tracking experiment.

Keywords: influence function; linear mixed model; longitudinal data; M-estimation; robust estimation; quantile regression; repeated measures.

1 Introduction

Linear random effects regression is widely used in the analysis of hierarchical data, such as, longitudinal or geographically clustered data (Singer and Willett, 2003; Rabe-Hesketh

*Social Statistics and S3RI, University of Southampton; Social Statistics and Centre for Census and Survey Research, University of Manchester, Manchester M13 9PL, UK nikos.tzavidis@manchester.ac.uk

†Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa, Via Ridolfi 10, Pisa 56124, Italy salvati@ec.unipi.it

‡School of Cancer and Enabling Sciences, University of Manchester, Manchester M13 9PL, UK geraci@manchester.ac.uk

§Arnold School of Public Health, University of South Carolina, 800 Sumter Street Columbia, SC 29208; Unit of Biostatistics, IMM, Karolinska Institutes, Nobels Väg 13, 17177 Stockholm mbottai@mailbox.sc.edu

and Skrondal, 2008; Goldstein, 2003). The random effects regression models a location parameter, namely the expected value of the conditional distribution of an outcome variable given a set of covariates. In many situations, however, the focus of the analysis extends beyond modelling a location parameter and emphasis is also placed on modelling other parts of this conditional distribution. The idea of modelling the quantiles of a conditional distribution has a long history in statistics. The seminal paper by Koenker and Bassett (1978) is usually regarded as the first detailed development of quantile regression. This can be viewed as a generalization of median regression. In the same way, expectile regression (Newey and Powell, 1987) is a quantile-like generalization of mean regression. M-quantile regression (Breckling and Chambers, 1988) combines these two concepts within a common framework defined by a quantile-like generalization of regression based on influence functions (M-regression).

Let us start with a motivating example that we will fully analyse later in this paper. Data on reaction time and hand-eye coordination were collected on 108 members of the public who visited the Human Systems Integration Laboratory at Naval Postgraduate School in Monterey, California in October 1995. One experiment that demonstrates motor learning and hand-eye coordination is rotary pursuit tracking. The subject's task in this experiment is to maintain contact with a target spot using a metal wand. Trials were conducted for 15 seconds at a time, and the total contact time during the 15 seconds was recorded. Four trials were recorded for each subject. The age and sex of each subject were also recorded. The outcome variable is the total contact time and we are interested in investigating its association with age and sex. Two examples of research questions that we are interested in answering with this dataset are the following: (1) is the performance in motor learning and hand eye coordination affected by age and/or sex? and (2) if so, do the effects of age and/or sex depend on the level of performance itself? that is, are age and/or sex differences the same among those who perform better and those who perform worse? In analysing this dataset one must take into account its longitudinal structure and one way for doing so is by using a linear growth curve model. However, the target of a growth curve model in this case will be the expected value of the conditional distribution of contact time given the set of covariates. In this respect, a growth curve model can help in answering question 1 but not question 2. For answering question 2 we need a quantile regression model that can handle hierarchical data.

Extensions of quantile regression for modelling dependent data have been considered by Lipsitz et al. (1997), Koenker (2004), Karlsson (2008), Geraci and Bottai (2007), and Liu and Bottai (2009). More specifically, Lipsitz et al. (1997) and Karlsson (2008) propose marginal models targeting the overall trend, over subjects, for a given quantile. Koenker (2004) proposes the use of penalized quantile regression. Geraci and Bottai (2007) propose a conditional model for quantile regression with random intercepts that uses the asymmetric Laplace distribution (ALD) for modelling the conditional likelihood while the distribution of the random effects is assumed to be Gaussian. Liu and Bottai (2009) further extend the conditional model by Geraci and Bottai (2007) and propose a quantile regression model with multiple random effects. The use of the ALD is mainly driven by convenience as it provides a parametric link between maximum likelihood estimation and minimizing the sum of absolute deviations.

To the best of our knowledge the existing literature on M-quantile regression can not handle dependent data. Although approaches to robust estimation in random effects models have been proposed by a series of authors (Huggins, 1993; Huggins and Loesch, 1998; Richardson and Welsh, 1995; Welsh and Richardson, 1997), these approaches focus

only on modelling a location parameter at the centre of the conditional distribution rather than the entire conditional distribution.

In this article we propose the use of a linear M-quantile random effects (random intercepts) regression (MQRE) model for modelling the quantiles of the conditional distribution of hierarchically structured data. One of the main advantages of the M-estimation framework is that it easily allows for robust estimation of both fixed and random effects. Nevertheless, robust estimation is not always justified and can potentially lead to a loss of efficiency for the model parameters. Our approach is based on the use of influence functions that depend on a tuning constant, which can be used to trade robustness for efficiency. As the value of the tuning constant tends to zero, the proposed method converges to quantile regression while as the value of the tuning constant increases, it tends to expectile regression (ERE). Expectile regression can be used to model the entire conditional distribution of interest without employing robust estimation. This feature, not available in the quantile regression offers added flexibility.

The structure of the paper is as follows. In Section 2 we revise quantile and M-quantile regression models. Section 3 reviews random effects models and focuses specifically on robust estimation. In Section 4 we present the M-quantile and expectile random effects regression models (MQRE and ERE) and discuss maximum likelihood in detail. The estimation algorithms are presented, approaches to inference are provided, and a small scale simulation study is employed for assessing finite sample approximations. In Section 5 we evaluate the MQRE and ERE regression models using model-based simulation studies, under a range of data generating mechanisms. We further compare MQRE and ERE to the quantile random effects model (QRRE) (Geraci and Bottai, 2007) aware of the fact that this alternative model does not target identical distributional parameters. In Section 6 we present the results from an application of MQRE and ERE to repeated measures data collected from a rotary pursuit tracking experiment. Finally, in Section 7 we conclude the paper with some final remarks.

2 Quantile/M-quantile regression

The classical regression model summarises the behaviour of the mean of a random variable y at each point in a set of covariates x (Mosteller and Tukey, 1977). This summary provides a rather incomplete picture, in much the same way as the mean gives an incomplete picture of a distribution. Instead, quantile regression summarises the behaviour of different parts (e.g. quantiles) of the conditional distribution of y at each point in the set of the x 's.

In the linear case, quantile regression leads to a family of hyper-planes indexed by a real number $q \in (0, 1)$. For a given value of q , the corresponding model shows how the q th quantile of the conditional distribution of y varies with x . For example, if $q = 0.5$ the quantile regression hyperplane shows how the median of the conditional distribution changes with x . Similarly, for $q = 0.1$ the quantile regression hyperplane separates the lower 10% of the conditional distribution from the remaining 90%.

Suppose (\mathbf{x}_i^T, y_i) , $i = 1, \dots, n$ is an independent random sample from a population, where \mathbf{x}_i^T are row p -vectors of a known design matrix \mathbf{X} and y_i is a scalar response variable of a continuous random variable with unknown continuous cumulative distribution function F . A linear regression model for the q th conditional quantile of y_i given \mathbf{x}_i is

$$Q_{y_i}(q|\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_q. \quad (1)$$

An estimate of the q th regression parameter β_q is obtained by minimizing

$$\sum_{i=1}^n [|y_i - \mathbf{x}_i^T \beta_q| \{ (1-q)I(y_i - \mathbf{x}_i^T \beta_q \leq 0) + qI(y_i - \mathbf{x}_i^T \beta_q > 0) \}].$$

Solutions are usually obtained by linear programming methods (Koenker and D'Orey, 1987) and algorithms for fitting quantile regression are now available in standard statistical software for example, the library *quantreg* in R (R Development Core Team, 2004), the command *qreg* in Stata, and the procedure *quantreg* in SAS.

Quantile regression can be viewed as a generalization of median regression. In the same way, expectile regression (Newey and Powell, 1987) is a ‘quantile-like’ generalization of mean (i.e. standard) regression. M-quantile regression (Breckling and Chambers, 1988) integrates these concepts within a common framework defined by a ‘quantile-like’ generalization of regression based on influence functions (M-regression). The M-quantile of order q for the conditional density of y given the set of covariates x , $f(y|x)$, is defined as the solution $MQ_y(q|x; \psi)$ of the estimating equation $\int \psi_q(y - MQ)f(y|x)dy = 0$, where ψ_q denotes an asymmetric influence function, which is the derivative of an asymmetric loss function ρ_q . A linear M-quantile regression model y_i given \mathbf{x}_i is one where we assume that

$$MQ_{y_i}(q|\mathbf{x}_i; \psi) = \mathbf{x}_i^T \beta_q. \quad (2)$$

That is, we allow a different set of p regression parameters for each value of $q \in (0, 1)$. Estimates of β_q are obtained by minimizing

$$\sum_{i=1}^n \rho_q(y_i - \mathbf{x}_i^T \beta_q). \quad (3)$$

Different regression models can be defined as special cases of (3). In particular, using different specifications for the asymmetric loss function ρ_q we can obtain the expectile, M-quantile and quantile regression models as special cases. When ρ_q is the square loss function we obtain the expectile regression model if $q \neq 0.5$ (Newey and Powell, 1987) and the standard ordinary least squares regression if $q = 0.5$. When ρ_q is the Huber loss function we obtain the M-quantile regression model (Breckling and Chambers, 1988). Finally, when ρ_q is the loss function described by Koenker and Bassett (1978) we obtain quantile regression.

Setting the first derivative of (3) leads to the following estimating equations

$$\sum_{i=1}^n \psi_q(r_{iq}) \mathbf{x}_i = \mathbf{0}, \quad (4)$$

where $r_{iq} = y_i - \mathbf{x}_i^T \beta_q$, $\psi_q(r_{iq}) = 2\psi(s^{-1}r_{iq})\{qI(r_{iq} > 0) + (1-q)I(r_{iq} \leq 0)\}$ and $s > 0$ is a suitable estimate of scale. For example, in the case of robust regression, $s = \text{median}|r_{iq}|/0.6745$. Since the focus of our paper is on M-type estimation, one may use different influence functions such as the Huber or the Hampel influence functions. For the robust versions of our regression model, in this article we employ the Huber Proposal 2 influence function, $\psi(u) = uI(-c \leq u \leq c) + c \cdot \text{sgn}(u)$. Provided that the tuning constant c is strictly greater than zero, estimates of β_q are obtained using iterative weighted least squares (IWLS). The steps of the algorithm for fitting the M-quantile regression model (2) are as follows,

- 1 Start with initial estimates of β_q and s .
- 2 Estimate the residuals r_{iq} .
- 3 Define weights $w_{iq} = \psi_q(r_{iq})/r_{iq}$.
- 4 Update the estimate of β_q using weighted least squares regression with weights w_{iq} .
- 5 Iterate until convergence.

These steps can be implemented in **R** by a simple modification of the IWLS algorithm used for fitting M-regression with function *rlm* (Venables and Ripley, 2002, section 8.3).

The IWLS algorithm used to fit an M-quantile regression model guarantees convergence to a unique solution (Kokic et al., 1997) when a continuous monotone influence function (e.g. Huber Proposal 2 with $c > 0$) is used. The tuning constant c can be used to trade robustness for efficiency in the M-quantile regression fit, with increasing robustness/decreasing efficiency as we move towards quantile regression and decreasing robustness/increasing efficiency as we move towards expectile regression. The flexibility of M-quantile regression is of particular importance for the present paper as this will allow us to also define an expectile random effects regression model.

3 Models for hierarchically structured data

Suppose we have data on an outcome variable y and a set of covariates x for n individuals clustered within d groups. A popular approach for modelling hierarchically structured data is to use a random effects model. In the simplest case we can define a random intercepts model

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_j^T \boldsymbol{\gamma} + \epsilon_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, d, \quad (5)$$

where \mathbf{x}_{ij} is a vector of p auxiliary variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients and \mathbf{z}_j is a $d \times 1$ vector of group indicators used for defining the random part of the model. In addition, $\boldsymbol{\gamma}$ denotes a $d \times 1$ vector of group-specific random effects, ϵ_{ij} is the individual random effect, and we assume that $\boldsymbol{\gamma} \sim N(0, \sigma_\gamma^2)$, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. A popular approach for estimating the parameters of (5) is to employ maximum likelihood estimation. Assuming normality for the error components, $cov(\gamma_j, \gamma_{j'}) = 0$, for $j \neq j'$, and $\epsilon \perp \gamma$, the log-likelihood function is

$$l(\boldsymbol{\beta}, \sigma_\gamma^2, \sigma_\epsilon^2) = -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (6)$$

where \mathbf{y} is the $n \times 1$ response vector, $\mathbf{V} = \boldsymbol{\Sigma}_\epsilon + \mathbf{Z}\boldsymbol{\Sigma}_\gamma\mathbf{Z}^T$, $\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \mathbf{I}_n$, $\boldsymbol{\Sigma}_\gamma = \sigma_\gamma^2 \mathbf{I}_d$, \mathbf{Z} is an $n \times d$ matrix of known positive constants. Here and throughout the paper \mathbf{I}_n , \mathbf{I}_d are identity matrices of size n and d , respectively. Estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ_γ^2 , and σ_ϵ^2 are obtained by solving the estimating equations obtained by differentiating the log-likelihood with respect to the parameters and setting these derivatives equal to zero (Goldstein, 2003). It is easy to see that in (6) we assume a squared loss function. In practice, however, data may contain outliers that invalidate the Gaussian assumptions. If normality is still assumed, the estimated model parameters under (6) will be biased and inefficient (Richardson and Welsh, 1995). One approach for robustifying the random effects model against departures from normality is to use an alternative loss function in the log-likelihood function that grows at slower rate than the squared loss function. This is the approach followed by Huggins (1993), Huggins and Loesch (1998), Richardson and Welsh (1995), and Welsh and Richardson (1997). In particular, robust maximum likelihood estimation for the random

effects model is performed by maximizing the following log-likelihood function

$$l(\boldsymbol{\beta}, \sigma_\gamma^2, \sigma_\epsilon^2) = -\frac{K_1}{2} \log|\mathbf{V}| - \rho\{\mathbf{V}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\}, \quad (7)$$

where ρ is a loss function, ψ is its derivate and $K_1 = E[\boldsymbol{\epsilon}\psi(\boldsymbol{\epsilon})^T]$ is computed over the standard normal distribution. Robust estimates of $\boldsymbol{\beta}$, σ_γ^2 , and σ_ϵ^2 are obtained by solving the following estimating equations,

$$\begin{aligned} \mathbf{X}^T \mathbf{V}^{-1/2} \psi \left\{ \mathbf{V}^{-1/2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} &= \mathbf{0} \\ \frac{1}{2} \left\{ \mathbf{V}^{-1/2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}^T \mathbf{V}^{-1/2} \mathbf{Z} \mathbf{Z}^T \mathbf{V}^{-1/2} \psi \left\{ \mathbf{V}^{-1/2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} - \frac{K_1}{2} \text{tr} [\mathbf{V}^{-1} \mathbf{Z} \mathbf{Z}^T] &= 0 \\ \frac{1}{2} \left\{ \mathbf{V}^{-1/2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}^T \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \psi \left\{ \mathbf{V}^{-1/2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} - \frac{K_1}{2} \text{tr} [\mathbf{V}^{-1}] &= 0. \end{aligned}$$

This is the maximum likelihood approach proposed by Richardson and Welsh (1995). Robust estimates of the random effects can be obtained by solving the following estimating equation with respect to $\boldsymbol{\gamma}$ (Fellner, 1986),

$$\mathbf{Z}^T \boldsymbol{\Sigma}_\epsilon^{-1/2} \psi \left\{ \boldsymbol{\Sigma}_\epsilon^{-1/2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma}) \right\} - \boldsymbol{\Sigma}_\gamma^{-1/2} \psi \left\{ \boldsymbol{\Sigma}_\gamma^{-1/2} \boldsymbol{\gamma} \right\} = \mathbf{0}.$$

An alternative estimation approach that can potentially lead to more efficient estimates of the variance components when there is a small number of groups or groups with a small number of observations is the robust restricted maximum likelihood approach proposed by Richardson and Welsh (1995) (see also Staudenmayer et al., 2009).

4 M-quantile and expectile random effects regression

With log-likelihood functions (6) and (7) we obtain respectively estimates of the parameters of the random effects model (5) and of its robust version. In fact, with (6) and (7) the target is respectively the expectation and a location parameter, close to the median, of the conditional distribution of the outcome variable given a set of explanatory variables. As stated in Section 2, on many occasions we are not only interested in describing the relationship between y and x near the centre of this conditional distribution, but also at other parts of the conditional distribution. In this section we propose extending the use of asymmetric loss functions in the case of hierarchically structured data.

Unlike Geraci and Bottai (2007), who assumed that the data follow an asymmetric Laplace distribution, here we remain within the M-estimation framework. Starting with (7) we note that one approach for extending the idea of asymmetric weighting of residuals in the case of hierarchical data is by defining the following modified Gaussian log-likelihood function

$$l(\boldsymbol{\beta}_q, \sigma_{\gamma_q}^2, \sigma_{\epsilon_q}^2) = -\frac{K_{1q}}{2} \log|\mathbf{V}_q| - \rho_q\{\mathbf{V}_q^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_q)\}, \quad (8)$$

where $\boldsymbol{\beta}_q$ is the $p \times 1$ vector of M-quantile regression coefficients, σ_{ϵ_q} and σ_{γ_q} are the quantile-specific variance components. Here $\rho_q(u) = 2\rho(u)(qI(u > 0) + (1 - q)I(u \leq 0))$ is a non-negative function and $\mathbf{r}_q = \mathbf{V}_q^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_q)$ is a vector of scaled residuals with components r_{ijq} , $\mathbf{V}_q = \boldsymbol{\Sigma}_{\epsilon_q} + \mathbf{Z}\boldsymbol{\Sigma}_{\gamma_q}\mathbf{Z}^T$, $\boldsymbol{\Sigma}_{\gamma_q} = \sigma_{\gamma_q}^2 \mathbf{I}_d$, $\boldsymbol{\Sigma}_{\epsilon_q} = \sigma_{\epsilon_q}^2 \mathbf{I}_n$.

On closer inspection, (6) and (7) can be obtained as special cases of (8) for specific choices of ρ_q and q . In particular, when $q = 0.5$ and ρ_q is the square loss function we obtain

(6) whereas when $q = 0.5$ and we use a loss function other than the square, for example the Huber loss function, we obtain (7). For q values other than 0.5 and for different choices of ρ the maximization of (8) will provide estimates of the fixed effects, β_q , and of the variance components, $\sigma_{\epsilon_q}, \sigma_{\gamma_q}$, which can then be used for obtaining the MQRE or the ERE fits. More specifically, using a square loss function in (8) at $q \neq 0.5$ results in an expectile random effects fit whereas using the Huber loss function in (8) results in an M-quantile random effects fit. The steps of the estimation algorithm are outlined in the next section. With function (8) we therefore extend the idea of asymmetric weighting – positive residuals are weighted by q and negative residuals by $(1 - q)$ – used in fitting the single-level M-quantile regression (Breckling and Chambers, 1988), to asymmetric weighting for obtaining a regression fit for the entire conditional distribution $f(y|x)$ accounting at the same time for the dependence structure in hierarchical data.

One characteristic of M-quantile random effects regression is that in addition to obtaining quantile-specific estimates of fixed effects, we now also obtain quantile-specific estimates of variance components. The easiest way of interpreting these quantile-specific variance components is by focusing on a regression model without covariates. Using a square loss function at $q = 0.5$, these variance components are the between and within group variance around the mean. Similarly, using an alternative loss function, e.g. the Huber one at $q = 0.5$, these variance components represent the between and within group variance around a location parameter that is close to the median. For $q \neq 0.5$ the quantile-specific variance components represent the between and within group variance around conditional locations other than the median.

4.1 Estimation algorithm

- 1 Start by assuming that $(\sigma_{\gamma_q}^2, \sigma_{\epsilon_q}^2)$ are known.
- 2 Given these variance components, form the covariance matrix \mathbf{V}_q , and estimate β_q by solving

$$\mathbf{X}^T \mathbf{V}_q^{-1/2} \psi_q \{ \mathbf{V}_q^{-1/2} (\mathbf{y} - \mathbf{X} \beta_q) \} = \mathbf{0} \quad (9)$$

using IWLS. In this case,

$$\hat{\beta}_q = \{ (\mathbf{V}_q^{-1/2} \mathbf{X})^T \mathbf{W} \mathbf{V}_q^{-1/2} \mathbf{X} \}^{-1} (\mathbf{V}_q^{-1/2} \mathbf{X})^T \mathbf{W} \mathbf{V}_q^{-1/2} \mathbf{y},$$

where \mathbf{W} is the $n \times n$ matrix of weights obtained for unit i in group j as $w_{ij} = \psi_q(r_{ijq})/r_{ijq}$.

- 3 Use the estimates of β_q to obtain estimates of the variance components. The ML estimates of the variance components are obtained by maximizing (8) with respect to $(\sigma_{\gamma_q}^2, \sigma_{\epsilon_q}^2)$. The corresponding score functions are:

$$\begin{aligned} \frac{1}{2} \left\{ \mathbf{V}_q^{-1/2} (\mathbf{y} - \mathbf{X} \beta_q) \right\}^T \mathbf{V}_q^{-1/2} \mathbf{Z} \mathbf{Z}^T \mathbf{V}_q^{-1/2} \psi_q \left\{ \mathbf{V}_q^{-1/2} (\mathbf{y} - \mathbf{X} \beta_q) \right\} \\ - \frac{K_{1q}}{2} \text{tr} [\mathbf{V}_q^{-1} \mathbf{Z} \mathbf{Z}^T] = 0 \end{aligned} \quad (10)$$

$$\begin{aligned} \frac{1}{2} \left\{ \mathbf{V}_q^{-1/2} (\mathbf{y} - \mathbf{X} \beta_q) \right\}^T \mathbf{V}_q^{-1/2} \mathbf{V}_q^{-1/2} \psi_q \left\{ \mathbf{V}_q^{-1/2} (\mathbf{y} - \mathbf{X} \beta_q) \right\} \\ - \frac{K_{1q}}{2} \text{tr} [\mathbf{V}_q^{-1}] = 0. \end{aligned} \quad (11)$$

For maximizing (8) we use the function `constrOptim` in R by supplying the log-likelihood function and the score functions.

- 4 Iterate steps 2 and 3 until convergence.
- 5 At convergence, estimates of the random effects at q th quantile fit are obtained by solving the following estimating equation with respect to γ_q

$$\mathbf{Z}^T \Sigma_{\epsilon_q}^{-1/2} \psi_q \{ \Sigma_{\epsilon_q}^{-1/2} (\mathbf{y} - \mathbf{X}\beta_q - \mathbf{Z}\gamma_q) \} - \Sigma_{\gamma_q}^{-1/2} \psi_q \{ \Sigma_{\gamma_q}^{-1/2} \gamma_q \} = \mathbf{0}. \quad (12)$$

An alternative estimation approach that includes an adjustment to allow for the loss of degrees of freedom incurred in estimating the fixed effects is the maximization of the restricted version of the modified Gaussian log-likelihood (8). It can be used when the number of groups or the number of observations within groups are small.

In practice the MQRE and ERE fits can be obtained by changing the value of the tuning constant c in the influence function. Using for example the Huber influence function, and setting this to the typical value $c = 1.345$ we obtain the MQRE regression fit while setting c equal to an arbitrary large value results in the ERE regression fit. With c equal to an arbitrary large value at $q = 0.5$ the ERE and the linear random effects (LRE) regression fits are equivalent.

4.2 Inference

Distribution theory for the parameters of the linear random effects model when the Gaussian assumptions hold has been studied by Hartley and Rao (1967) and Miller (1977). In the presence of outliers, however, the Gaussian assumptions are violated and large sample approximations are required. Huber (1967) proved consistency and asymptotic normality of maximum likelihood estimators when (i) the true distribution underlying the observations does not belong to the parametric family defining the ML estimators, and (ii) the second and higher derivatives of the likelihood function are not available. The work of Huber (1967) is specifically linked to robust estimation problems and his arguments are used by Welsh and Richardson (1997) for proposing approximate inference for the robust ML estimators of the linear random effects model. Since the log-likelihood function (8) is an extension of the log-likelihood function (7), a sketch of the evaluation of the stochastic behaviour of ML estimators $\hat{\boldsymbol{\theta}}_q = (\hat{\beta}_q^T, \hat{\sigma}_{\gamma_q}^2, \hat{\sigma}_{\epsilon_q}^2)^T$ is given in this section using the developments in Huber (1967) and Welsh and Richardson (1997). The estimators $\hat{\boldsymbol{\theta}}_q = (\hat{\beta}_q^T, \hat{\sigma}_{\gamma_q}^2, \hat{\sigma}_{\epsilon_q}^2)^T$ satisfy a set of estimating equations of the form

$$\sum_{j=1}^d \Phi_{qj}(\boldsymbol{\theta}_q) = \mathbf{0}, \quad (13)$$

where $\Phi_{qj}(\boldsymbol{\theta}_q) = \left(\Phi_{qj\beta_q}^T, \Phi_{qj\sigma_{\gamma_q}^2}, \Phi_{qj\sigma_{\epsilon_q}^2} \right)^T$, for particular choices of $\Phi_{qj}(\boldsymbol{\theta}_q)$. Under a general response distribution D , the estimator $\hat{\boldsymbol{\theta}}_q$ satisfying (13) is estimating a root $\boldsymbol{\theta}_q$ of

$$\sum_{j=1}^d E_D [\Phi_{qj}(\boldsymbol{\theta}_q)] = \mathbf{0}. \quad (14)$$

Provided that

$$-n^{-1} \sum_{j=1}^d E_D [\partial \Phi_{qj}(\boldsymbol{\theta}_q) / \partial \boldsymbol{\theta}_q] \rightarrow \mathbf{G}, \quad (15)$$

where the $(p+2) \times (p+2)$ matrix \mathbf{G} is positive definite, and

$$n^{-1} \sum_{j=1}^d E_D [\Phi_{qj}(\boldsymbol{\theta}_q)^T \Phi_{qj}(\boldsymbol{\theta}_q)] \rightarrow \mathbf{F}, \quad (16)$$

a Taylor series approximation which holds uniformly in a neighborhood of $\boldsymbol{\theta}_q$ is

$$\hat{\boldsymbol{\theta}}_q = \boldsymbol{\theta}_q + \mathbf{G}^{-1} n^{-1} \sum_{j=1}^d \Phi_{qj}(\boldsymbol{\theta}_q) + o_p(n^{-1/2}), \quad \text{as } n \rightarrow \infty. \quad (17)$$

Under the regularity conditions described by Huber (1967), it follows that

$$n^{1/2} (\hat{\boldsymbol{\theta}}_q - \boldsymbol{\theta}_q) \xrightarrow{D} N(\mathbf{0}, \mathbf{G}^{-1} \mathbf{F} (\mathbf{G}^{-1})^T), \quad \text{as } n \rightarrow \infty. \quad (18)$$

The components of the information matrix \mathbf{G} and the variance of the normalized score functions \mathbf{F} for ML estimators are given in Appendix. The covariance matrix can be consistently estimated by $\hat{\mathbf{G}}^{-1} \hat{\mathbf{F}} (\hat{\mathbf{G}}^{-1})^T$ where the matrices $\hat{\mathbf{G}}$ and $\hat{\mathbf{F}}$ are evaluated at $\hat{\boldsymbol{\theta}}_q$. The variance of $\hat{\boldsymbol{\beta}}_q$ can be also estimated following Street et al. (1988),

$$\hat{V}(\hat{\boldsymbol{\beta}}_q) = \frac{(n-p)^{-1} \sum_{j=1}^d \sum_{i=1}^{n_j} \psi_q^2(\hat{r}_{ijq})}{\left[n^{-1} \sum_{j=1}^d \sum_{i=1}^{n_j} \psi_q'(\hat{r}_{ijq}) \right]^2} (\mathbf{X}^T \hat{\mathbf{V}}_q^{-1} \mathbf{X})^{-1} \quad (19)$$

where \hat{r}_{ijq} is the i th component of the vector $\hat{\mathbf{r}}_q = \hat{\mathbf{V}}_q^{-1/2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_q)$, $\psi_q^2(\hat{r}_{ijq}) = \hat{w}_{ijq}^2 \hat{r}_{ijq}^2$, $\psi_q'(\hat{r}_{ijq}) = \{qI(0 < \hat{r}_{ijq} \leq c) + (1-q)I(c < \hat{r}_{ijq} \leq 0)\}$ and \hat{w}_{ijq} is the final weight in the IWLS process. An approximate confidence interval of level $(1-\alpha)$ for the $\boldsymbol{\beta}_q$ fixed effects is

$$\hat{\boldsymbol{\beta}}_q \pm z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{\boldsymbol{\beta}}_q)}, \quad (20)$$

where $z_{(1-\alpha/2)}$ denotes the $(1-\alpha/2)$ quantile of the standard normal distribution.

The variance of $(\hat{\sigma}_{\gamma_q}^2, \hat{\sigma}_{\epsilon_q}^2)$ is also obtained from (18). Following Pinheiro and Bates (2000), an approximate confidence interval of level $(1-\alpha)$ for the variance components at q is obtained using

$$\left(\hat{\sigma}_{\gamma_q}^2 \exp \left\{ -z_{(1-\alpha/2)} \frac{1}{\hat{\sigma}_{\gamma_q}^2} n^{-1/2} \sqrt{\hat{V}(\hat{\sigma}_{\gamma_q}^2)} \right\}, \hat{\sigma}_{\gamma_q}^2 \exp \left\{ +z_{(1-\alpha/2)} \frac{1}{\hat{\sigma}_{\gamma_q}^2} n^{-1/2} \sqrt{\hat{V}(\hat{\sigma}_{\gamma_q}^2)} \right\} \right) \quad (21)$$

and

$$\left(\hat{\sigma}_{\epsilon_q}^2 \exp \left\{ -z_{(1-\alpha/2)} \frac{1}{\hat{\sigma}_{\epsilon_q}^2} n^{-1/2} \sqrt{\hat{V}(\hat{\sigma}_{\epsilon_q}^2)} \right\}, \hat{\sigma}_{\epsilon_q}^2 \exp \left\{ +z_{(1-\alpha/2)} \frac{1}{\hat{\sigma}_{\epsilon_q}^2} n^{-1/2} \sqrt{\hat{V}(\hat{\sigma}_{\epsilon_q}^2)} \right\} \right). \quad (22)$$

An alternative approach for computing the variance of the estimated parameters and for constructing confidence intervals is by using parametric bootstrap. A recent example of using parametric bootstrap in the case of the robust random effects model is given by Sinha and Rao (2009). The drawback of using bootstrap in the case of MQRE and ERE regression is the computation time required for performing a large number of bootstrap replicates. We have performed some limited empirical assessment of the parametric bootstrap procedure and the results appear to be close to those obtained from the analytic approximations presented earlier on in this section.

4.3 Evaluating the asymptotic approximations

Before evaluating the performance of MQRE and of ERE regression, in this section we assess the large sample approximations outlined in Section 4.2. To this end, we design a small scale Monte-Carlo simulation study. Data are generated under the following location-shift model

$$y_{ij} = 100 + 2x_{1ij} + \gamma_j + \varepsilon_{ij}, i = 1, \dots, 4, j = 1, \dots, 100,$$

where the values of $x_1 \sim U[0, 15]$. Different distributions have been considered for the error terms ε_{ij} (level 1) and γ_j (level 2). In particular, we consider two scenarios for generating the error terms:

[0, 0] - No outliers: $\gamma \sim N(0, 16)$ and $\varepsilon \sim N(0, 36)$.

$[\varepsilon, \gamma]$ - Outliers in both hierarchical levels: $\gamma \sim N(0, 16)$ for $j = 1, \dots, 90$, and $\gamma \sim N(0, 225)$ for $j = 91, \dots, 100$; $\varepsilon \sim \delta N(0, 36) + (1 - \delta)N(0, 1100)$ where $\delta \sim \text{Bernoulli}(0.9)$.

The tuning constant c is set to 1.345 for MQRE and to 100 for ERE. For this simulation study, we replicate $R = 1000$ datasets. Since we are interested in inference under the correct model, we present results only for the ERE under scenario [0,0] and results only for MQRE under scenario $[\varepsilon, \gamma]$. The complete tables are not reported here but are available from the authors upon request. To start with, Table 1 presents results on how well the variance of the fixed effects and of the variance components is estimated. For each scenario and for each estimator $\hat{\theta}$, at $q = 0.25$ and $q = 0.5$, Table 1 reports

- (a) The Monte-Carlo variance,

$$S^2(\hat{\theta}) = R^{-1} \sum_{r=1}^R (\hat{\theta}^{(r)} - \bar{\theta})^2,$$

where $\hat{\theta}^{(r)}$ is the estimated parameter at quantile q for the r th replication and $\bar{\theta} = R^{-1} \sum_{r=1}^R \hat{\theta}^{(r)}$.

- (b) The estimated variance of $\hat{\beta}_q$, $\hat{\sigma}_{\gamma_q}^2$ and $\hat{\sigma}_{\varepsilon_q}^2$ averaged over the Monte-Carlo replications.
- (c) The coverage rate of nominal 95 per cent confidence intervals and their mean length. The coverage of these intervals is defined by the number of times the interval, defined by the estimate of the parameter plus or minus twice its estimated variance, contains the ‘true’ population parameter.

Under scenario [0, 0], the asymptotic variance of the estimators of ERE at $q = 0.25$ and $q = 0.5$ provide a good approximation to the true variances. In this case, there are no outliers and hence there is no reason to employ robust estimation. Turning now to the results for scenario $[\varepsilon, \gamma]$, we note that the approximation to the true variance of the estimated parameters of MQRE is overall satisfactory although there is a noticeable underestimation of the variance of the level 2 variance component. These results can potentially improve as the number of observations within groups and the number of groups increases.

We now focus on the construction of confidence intervals. Figure 1 presents normal probability plots of the estimates of the fixed effects and of the variance components under the two scenarios. Under scenario [0, 0] we note that for all target parameters a normal approximation is reasonable. Under scenario $[\varepsilon, \gamma]$, a normal approximation appears to be reasonable for the fixed effects and for the level 1 variance component. For the level 2 variance component there are some departures from normality but these are not very

severe. This is supported by the fact that the coverage rates of normal confidence intervals for the level 2 variance component are about 92% for both $q = 0.25$ and $q = 0.5$.

[Table 1 about here.]

[Figure 1 about here.]

5 Simulation Study

In this section we report Monte-Carlo simulation results that we carried out for assessing the performance of MQRE and ERE at two quantiles, $q = 0.25$ and $q = 0.5$. Data are generated under the following location-shift model

$$y_{ij} = 100 + 2x_{1ij} + 3x_{2ij} + \gamma_j + \varepsilon_{ij}, i = 1, \dots, 6, j = 1, \dots, 100,$$

where the values of $x_1 \sim U[0, 15]$, and the values of x_2 are set equal to the values $x_{1j} = 1, x_{2j} = 2, \dots, x_{6j} = 6$, and are kept constant throughout the simulations. The level 1 and level 2 error terms γ_j and ε_{ij} are independently generated according to four scenarios:

$[0, 0]$ - No outliers: $\varepsilon \sim N(0, 36)$ and $\gamma \sim N(0, 16)$.

$[\varepsilon, 0]$ - Outliers only at the individual level (level 1): $\varepsilon \sim \delta N(0, 36) + (1 - \delta)N(0, 1100)$ and $\gamma \sim N(0, 16)$, where $\delta \sim \text{Bern}(0.9)$.

$[0, \gamma]$ - Outliers only at the group level (level 2): $\varepsilon \sim N(0, 36)$, $\gamma \sim N(0, 16)$ for $j = 1, \dots, 90$, and $\gamma \sim N(0, 225)$ for $j = 91, \dots, 100$.

$[\varepsilon, \gamma]$ - Outliers in both hierarchical levels: error terms are generated as above but with contamination at both levels.

Each scenario is independently replicated $R = 500$ times. Under scenario $[0, 0]$ the assumptions of the random effects model (5) are valid. Scenarios $[\varepsilon, 0]$, $[0, \gamma]$ and $[\varepsilon, \gamma]$ define situations under which the presence of outliers is likely and hence the Gaussian assumptions of model (5) are violated. The aim of this simulation study is two-fold. First, we assess the ability of MQRE and ERE to account for the dependence structure of hierarchical data. Second, we compare MQRE to the quantile random effects (QRRE) model proposed by Geraci and Bottai (2007). The tuning constant c is set at the same values used for the simulation study described in Section 4.3.

Starting with the first aim, we compare the MQRE and the linear M -quantile (MQ) regression model (see Section 2), for which we also use the Huber Proposal 2 influence function with $c = 1.345$. Although both MQRE and MQ are robust to outliers, we expect that MQRE will perform better than the single level M -quantile model when clustering is present. At $q = 0.5$ MQRE is compared with the linear random effects (LRE) model (5). We expect that when outliers are present, MQRE will perform better than the LRE. For other quantiles we also compare the MQRE and ERE models. In this case ERE replaces LRE and we expect that when outliers are present that MQRE will be superior. For comparing the different methods we mainly focus on the fixed effects parameters. For each regression parameter performance is assessed using the following indicators:

- (a) Average Relative Bias (ARB) defined as

$$ARB(\hat{\theta}) = R^{-1} \sum_{r=1}^R \frac{\hat{\theta}^{(r)} - \theta}{\theta} \times 100,$$

where $\hat{\theta}^{(r)}$ is the estimated parameter at quantile q for the r th replication and θ is the corresponding ‘true’ value of this parameter. The empirical values of each parameter are computed previously with 10,000 Monte Carlo replicates to ensure better accuracy, since they are the reference values.

(b) Relative Efficiencies (EFF) defined as

$$EFF(\hat{\theta}) = \frac{S_{model}^2(\hat{\theta})}{S_{MQRE}^2(\hat{\theta})}$$

where $S^2(\hat{\theta}) = R^{-1} \sum_{r=1}^R (\hat{\theta}^{(r)} - \bar{\theta})^2$ and $\bar{\theta} = R^{-1} \sum_{r=1}^R \hat{\theta}^{(r)}$.

This efficiency measure is also used for comparing estimates of the variance components obtained from the MQRE and LRE at $q = 0.5$ or the ERE at $q = 0.25$.

Table 2 reports the simulation results for estimators of the fixed effects under the different approaches. Under the scenario $[0,0]$ and quantile $q = 0.5$, the estimators of the fixed effects from LRE are more efficient than the corresponding estimators from MQ. This is expected because the LRE correctly models the two level structure present in the synthetic population data. The estimators of the fixed effects of the LRE model are also more efficient than the corresponding estimators from the MQRE. Under this scenario there is no reason to employ outlier-robust estimation. Doing so results in paying a premium that is reflected in the lower efficiency of the MQRE regression estimators. At $q = 0.25$, the estimators of the fixed effects of the ERE are also more efficient than the corresponding estimators of MQ and MQRE. This demonstrates the ability of the ERE to extend the LRE model for modelling other quantiles.

The superior performance of MQRE is demonstrated in scenarios $[\varepsilon, 0]$ and $[\varepsilon, \gamma]$ where outliers exist either at level 1, or at both hierarchical levels. In particular, in most cases the estimators of the fixed effects from MQRE are more efficient than the corresponding estimators from MQ or from LRE/ERE. These results provide evidence that using MQRE regression protects against outlying values and it accounts for the dependence structure of hierarchical data when modelling the conditional quantiles. Finally, it appears that having outliers only at level 2 (scenario $[0, \gamma]$) does not have a severe effect on the efficiency of the estimators of the fixed effects.

Table 2 also reports the efficiency of the estimators of the variance components and the average over simulations under the four scenarios. We note that under contamination there are large gains in efficiency when using MQRE instead of ERE at $q = 0.25$ or LRE at $q = 0.5$. In contrast to this, under the $[0,0]$ ERE and LRE perform better than MQRE as expected. In this case the robustness offered by MQRE is unnecessary and this is reflected in the lower efficiency of the estimators of the variance components.

[Table 2 about here.]

Having assessed the performance of MQRE and ERE, in the second part of this section we wish to compare the MQRE to the QRRE model (Geraci and Bottai, 2007). A direct comparison is difficult because MQRE and ERE give M-quantile regression fits while on the other hand QRRE aims at modelling ordinary quantiles. We therefore limit our comparison to the median, the only parameter for which the two approaches are comparable. We use two indicators, the Mean Average Squared Error (MASE) and the Mean Absolute Deviation Error (MADE) respectively defined by

$$MASE = (Rn)^{-1} \sum_{i=1}^n \sum_{r=1}^R (\hat{y}_{iq}^{(r)} - y_{iq}^{(r)})^2$$

$$MADE = (Rn)^{-1} \sum_{i=1}^n \sum_{r=1}^R |\hat{y}_{iq}^{(r)} - y_{iq}^{(r)}|,$$

where $\hat{y}_{iq}^{(r)}$ is the predicted value of unit i at quantile q for the r th replication and $y_{iq}^{(r)}$ is the corresponding true value. The results of these experiments are presented in Table 3, which shows MASE and MADE expressed as ratios to, respectively, MASE and MADE obtained for MQRE. Note that at the median ERE and LRE models are equivalent. As expected, LRE works well in scenario $[0, 0]$. In the presence of outliers, MQRE is overall performing best and also appears to perform a bit better than the QRRE. This result can be partially explained by the lack of robustness of the QRRE model to contamination of the random effects. Also, the Monte Carlo approach adopted by Geraci and Bottai (2007) for estimating the parameters of QRRE introduces additional variability to the estimation process which can affect, on some occasions, the overall convergence of the estimation algorithm. Alternative estimation algorithms based on numerical integration techniques are currently being investigated. Finally, outliers at level 2 do not appear to have a significant effect. These results indicate that in comparison to competitor models, MQRE performs very well and that both the MQRE and QRRE provide reasonable quantile random effects regression fits. As part of our empirical investigations we have also produced results for the penalized quantile regression model (Koenker, 2004). These results are not reported here, but in line with Geraci and Bottai (2007) we also find that the QRRE model performs better than the penalized quantile regression model.

[Table 3 about here.]

6 A Case Study: Analysis of rotary pursuit tracking experiment data

Discovery Day is a day set aside by the United States Naval Postgraduate School in Monterey, California, to invite the general public into its laboratories. On Discovery Day, 21 October 1995, data on reaction time and hand-eye coordination were collected on 108 members of the public who visited the Human Systems Integration Laboratory. One experiment that demonstrates motor learning and hand-eye coordination is rotary pursuit tracking. The equipment used has a rotating disk with a 3/4 inches target spot. The subject's task is to maintain contact with the target spot with a metal wand. The target spot on the circle tracker keeps constant speed in a circular path. The target spot on the box tracker has varying speeds as it traverses the box, making the task potentially more difficult. Trials were conducted for 15 seconds at a time, and the total contact time during the 15 seconds was recorded. Four trials were recorded for each of 108 subjects thus giving $n = 432$ observations in all. The age and sex of each subject were also recorded. The outcome variable is the total contact time and we are interested in investigating its association with age, sex, and shape.

Having each trial been recorded for all individuals, measurements of different trials pertaining the same subject could be, in general, correlated. Therefore, appropriate methods that account for the dependence structure in the data must be employed. Ignoring the

longitudinal structure, in fact, could lead to misleading results. We assess how this would impact on our analysis by estimating the following simplified regressions (intercept-only models): (a) MQRE regression ($c = 1.345$), (b) ERE regression ($c = 100$), (c) a single level M-quantile regression ($c = 1.345$), and (d) a single level expectile regression ($c = 100$). MQRE and ERE include subject-specific random effects (random intercepts) while the single level regression models ignore the fact that for each subject we have repeated observations. The results are reported in Table 4. We see that failing to account for the repeated measures design has an impact on the variance of the intercept term. In particular, the variance of the intercept terms of the single level regression models is notably smaller than the corresponding variance of MQRE and ERE. The results for ERE at $q = 0.5$ are identical to those produced by the *lme* function in R and the results of the single level expectile regression at $q = 0.5$ are identical to those produced by function *lm* in R. This confirms that our algorithm reproduces the results of standard software.

[Table 4 about here.]

We now analyse the data by estimating MQRE and ERE regression fits at $q \in \{0.25, 0.5, 0.75\}$. In the fixed part of the linear predictor we include trial, sex (female=1, male=0), age group (2 to 8, 9 to 11, 12 to 28, 29 to 38 and 39 to 52 years) of the individuals and an indicator variable for the shape of the tracker (box=1, circle=0). The random part is defined by a subject-specific intercept. Estimation is performed by using the maximum likelihood estimation approach (see Section 4.1) and the Huber Proposal 2 influence function with $c = 1.345$ (MQRE) and $c = 100$ (ERE). The results are reported in Table 5 and Figure 2. Figure 2 shows box-plots of the observed contact time for females and males. In addition, it plots the estimated quantile lines of the contact time by age group for males and females and for the different shapes of tracker by averaging the predictions under the different M-quantile and expectile fits over trials.

Examining Figure 2, we detect the asymmetry of the distribution of contact time, since the estimated quantile lines at $q = 0.25$ and $q = 0.5$ are closer than the estimated quantile lines at $q = 0.5$ and $q = 0.75$. The positive asymmetry can be also observed in the fixed effects estimates reported in Table 5. The estimated intercept at $q = 0.5$ is 1.49 for the MQRE and 1.57 for ERE. These two numbers are respectively estimates of the median and mean of contact time at the first trial for the reference subject i.e. a male, in age group 2 to 8 years that uses a circle tracker. Moreover, the estimates of the between subjects variance component appear to be increasing with q with the estimates under the MQRE model being smaller than the corresponding ERE estimates, which may indicate the presence of outliers.

With MQRE and ERE we are also able to examine the impact of the different covariates across quantiles. The effect of trial is positive since subjects' performance improves over time. In addition, the impact of trial appears to be constant across quantiles. Also, subjects who use a circle as a tracker have higher contact times than those that use a box as a tracker. Taking into account the standard errors of these estimates, the effect of shape appears to be constant across quantiles i.e. individuals with higher contact times face the same difficulties in using a box as a tracker as individuals with lower contact times. Overall, males have better contact times than females. This gender gap, also illustrated by Figure 2, appears to be more evident in the middle rather than at the top end of the distribution. In other words, the contact time is similar between the best-performing females and males. Finally, age appears to have a non-linear effect. The contact time increases until age 28, is stable between 28 and 38 years, and then decreases after 38 years (see Figure 2) with the

the positive effect of age on contact time being more pronounced at the top end rather than at the lower end of the distribution.

[Figure 2 about here.]

[Table 5 about here.]

7 Discussion

In this paper we propose the extension of M-quantile and expectile regression to M-quantile and expectile random effects regression. Our proposed method combines M-quantile and expectile regression for the analysis of dependent data. It inherits robustness to outliers from the former and efficiency from the latter and permits to control directly the trade-off between the two by means of a tuning constant. As illustrated in the real-data example, the proposed approaches to modelling conditional quantiles may prove a useful alternative to current approaches to the analysis of hierarchical data. In a simulation study specifically designed to evaluate the impact of outliers on the robustness of the estimates, the proposed methods appeared to perform consistently well across all scenarios considered. One limitation of MQRE and ERE is that they allow for a very specific correlation structure. More specifically, in this paper we considered regression models with random intercepts, which is equivalent to assuming a uniform or exchangeable correlation structure. Although this structure may be adequate in many applications with hierarchical data, it may be unsatisfactory for others. Future work will extend the proposed approaches for handling more complex correlation structures including random coefficient models.

Appendix

In this Appendix we derive the information matrix and the normalized score functions for obtaining the variance-covariance matrix of $\boldsymbol{\theta}_q = (\boldsymbol{\beta}_q^T, \sigma_{\gamma_q}^2, \sigma_{\epsilon_q}^2)^T$. The information matrix $\mathbf{G}_n(\boldsymbol{\theta}_q)$ has components

$$\mathbf{G}_{n\boldsymbol{\beta}_q\boldsymbol{\beta}_q}(\boldsymbol{\theta}_q) = n^{-1} \sum_{j=1}^d \mathbf{X}_j^T \mathbf{V}_{qj}^{-1/2} \boldsymbol{\Psi}'_{qj} \mathbf{V}_{qj}^{-1/2} \mathbf{X}_j, \quad (\text{A-1})$$

with $\boldsymbol{\Psi}'_{qj}$ is a $n_j \times n_j$ diagonal matrix with the i th component equal to $2(1-q)I(-c \leq r_{ijq} < 0) + 2qI(0 < r_{ijq} \leq c)$,

$$\begin{aligned} \mathbf{g}_{n\boldsymbol{\beta}_q\boldsymbol{\tau}_{kq}}(\boldsymbol{\theta}_q) &= \frac{1}{2n} \sum_{j=1}^d \left\{ \mathbf{X}_j^T \mathbf{V}_{qj}^{-1/2} \frac{\partial \mathbf{V}_j}{\partial \tau_{kq}} \psi_q \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\} \right. \\ &\quad \left. + \mathbf{X}_j^T \mathbf{V}_{qj}^{-1/2} \boldsymbol{\Psi}'_{qj} \mathbf{V}_{qj}^{-1} \frac{\partial \mathbf{V}_j}{\partial \tau_{kq}} \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\} \end{aligned} \quad (\text{A-2})$$

where $\boldsymbol{\tau}_q = (\sigma_{\gamma_q}^2, \sigma_{\epsilon_q}^2)^T$,

$$g_{n\boldsymbol{\tau}_{kq}\boldsymbol{\tau}_{lq}}(\boldsymbol{\theta}_q) = \frac{1}{2n} \sum_{j=1}^d \left\{ (3/2) \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\}^T \mathbf{V}_{qj}^{-1/2} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{kq}} \mathbf{V}_{qj}^{-1} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{lq}} \mathbf{V}_{qj}^{-1/2} \right.$$

$$\begin{aligned} & \psi_q \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\} + (1/2) \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\}^T \mathbf{V}_{qj}^{-1/2} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{k_q}} \mathbf{V}_{qj}^{-1/2} \boldsymbol{\Psi}'_{qj} \\ & \mathbf{V}_{qj}^{-1} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{l_q}} \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \left\} - K_{1q} \text{tr} \left[\mathbf{V}_{qj}^{-1} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{k_q}} \mathbf{V}_{qj}^{-1} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{l_q}} \right], \end{aligned} \quad (\text{A-3})$$

where $\mathbf{G}_{n\boldsymbol{\beta}_q\boldsymbol{\beta}_q}(\boldsymbol{\theta}_q)$ is a $p \times p$ matrix, $\mathbf{g}_{n\boldsymbol{\beta}_q\tau_{k_q}}(\boldsymbol{\theta}_q)$ is $p \times 1$ vectors, and $g_{n\tau_{k_q}\tau_{l_q}}(\boldsymbol{\theta}_q)$ is scalar. The matrix $\mathbf{G}_n(\boldsymbol{\theta}_q)$ of size $(p+2) \times (p+2)$ can be expressed as

$$\mathbf{G}_n(\boldsymbol{\theta}_q) = \begin{bmatrix} \mathbf{G}_{n\boldsymbol{\beta}_q\boldsymbol{\beta}_q}(\boldsymbol{\theta}_q) & \begin{bmatrix} \mathbf{g}_{n\boldsymbol{\beta}_q\sigma_{\gamma_q}^2}(\boldsymbol{\theta}_q) \\ \mathbf{g}_{n\boldsymbol{\beta}_q\sigma_{\epsilon_q}^2}(\boldsymbol{\theta}_q) \end{bmatrix} \\ \begin{bmatrix} \mathbf{g}_{n\boldsymbol{\beta}_q\sigma_{\gamma_q}^2}(\boldsymbol{\theta}_q) \\ \mathbf{g}_{n\boldsymbol{\beta}_q\sigma_{\epsilon_q}^2}(\boldsymbol{\theta}_q) \end{bmatrix}^T & \begin{bmatrix} g_{n\sigma_{\gamma_q}^2\sigma_{\gamma_q}^2}(\boldsymbol{\theta}_q) & g_{n\sigma_{\gamma_q}^2\sigma_{\epsilon_q}^2}(\boldsymbol{\theta}_q) \\ g_{n\sigma_{\epsilon_q}^2\sigma_{\gamma_q}^2}(\boldsymbol{\theta}_q) & g_{n\sigma_{\epsilon_q}^2\sigma_{\epsilon_q}^2}(\boldsymbol{\theta}_q) \end{bmatrix} \end{bmatrix}. \quad (\text{A-4})$$

The variance-covariance matrix $\mathbf{F}_n(\boldsymbol{\theta}_q)$ of the normalized score functions is

$$\begin{aligned} \mathbf{F}_{n\boldsymbol{\beta}_q\boldsymbol{\beta}_q}(\boldsymbol{\theta}_q) &= n^{-1} \sum_{j=1}^d \left\{ \mathbf{X}_j^T \mathbf{V}_{qj}^{-1/2} E \left[\psi_q \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\} \psi_q \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\}^T \right] \right. \\ & \quad \left. \boldsymbol{\Psi}'_{qj} \mathbf{V}_{qj}^{-1/2} \mathbf{X}_j \right\}, \end{aligned} \quad (\text{A-5})$$

$$\begin{aligned} \mathbf{f}_{n\boldsymbol{\beta}_q\tau_{k_q}}(\boldsymbol{\theta}_q) &= \frac{1}{2n} \sum_{j=1}^d \left\{ \mathbf{X}_j^T \mathbf{V}_{qj}^{-1/2} E \left[\psi_q \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\} \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\}^T \right] \right. \\ & \quad \left. \mathbf{V}_{qj}^{-1/2} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{k_q}} \mathbf{V}_{qj}^{-1/2} \psi_q \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\} \right\} \end{aligned} \quad (\text{A-6})$$

$$\begin{aligned} f_{n\tau_{k_q}\tau_{l_q}}(\boldsymbol{\theta}_q) &= \frac{1}{2n} \sum_{j=1}^d E \left\{ \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\}^T \mathbf{V}_{qj}^{-1/2} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{k_q}} \mathbf{V}_{qj}^{-1/2} \psi_q \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\} \right. \\ & \quad \left. - K_{1q} \text{tr} \left[\mathbf{V}_{qj}^{-1} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{k_q}} \right] \right\} \left\{ \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\}^T \mathbf{V}_{qj}^{-1/2} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{l_q}} \mathbf{V}_{qj}^{-1/2} \right. \\ & \quad \left. \psi_q \left\{ \mathbf{V}_{qj}^{-1/2} (\mathbf{y}_j - \mathbf{X}_j^T \boldsymbol{\beta}_q) \right\} - K_{1q} \text{tr} \left[\mathbf{V}_{qj}^{-1} \frac{\partial \mathbf{V}_{qj}}{\partial \tau_{l_q}} \right] \right\} \end{aligned} \quad (\text{A-7})$$

The matrix $\mathbf{F}_n(\boldsymbol{\theta}_q)$ of size $(p+2) \times (p+2)$ can be expressed as

$$\mathbf{F}_n(\boldsymbol{\theta}_q) = \begin{bmatrix} \mathbf{F}_{n\boldsymbol{\beta}_q\boldsymbol{\beta}_q}(\boldsymbol{\theta}_q) & \begin{bmatrix} \mathbf{f}_{n\boldsymbol{\beta}_q\sigma_{\gamma_q}^2}(\boldsymbol{\theta}_q) \\ \mathbf{f}_{n\boldsymbol{\beta}_q\sigma_{\epsilon_q}^2}(\boldsymbol{\theta}_q) \end{bmatrix} \\ \begin{bmatrix} \mathbf{f}_{n\boldsymbol{\beta}_q\sigma_{\gamma_q}^2}(\boldsymbol{\theta}_q) \\ \mathbf{f}_{n\boldsymbol{\beta}_q\sigma_{\epsilon_q}^2}(\boldsymbol{\theta}_q) \end{bmatrix}^T & \begin{bmatrix} f_{n\sigma_{\gamma_q}^2\sigma_{\gamma_q}^2}(\boldsymbol{\theta}_q) & f_{n\sigma_{\gamma_q}^2\sigma_{\epsilon_q}^2}(\boldsymbol{\theta}_q) \\ f_{n\sigma_{\epsilon_q}^2\sigma_{\gamma_q}^2}(\boldsymbol{\theta}_q) & f_{n\sigma_{\epsilon_q}^2\sigma_{\epsilon_q}^2}(\boldsymbol{\theta}_q) \end{bmatrix} \end{bmatrix}. \quad (\text{A-8})$$

References

- Breckling, J. and Chambers, R. (1988). M -quantiles. *Biometrika* **75**, 761–771.
- Davis, C.S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in medicine*, **10**, 1959–1980.
- Fellner, W.H. (1986). Robust Estimation of Variance Components. *Technometrics*, **28**, 51–60.
- Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**, 140–54.
- Goldstein, H. (2003). *Multilevel Statistical Models*. Wiley, NewYork.
- Hartley, H. O. and J. N. K. Rao (1967). Maximum-likelihood estimation for the mixed analysis of variance model, *Biometrika*, **54**, 93–108.
- Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 221–233.
- Huber, P.J. (1981). *Robust Statistics*. John Wiley & Sons, NewYork.
- Huggins, R.M. (1993). A Robust Approach to the Analysis of Repeated Measures. *Biometrics*, **49**, 255–268.
- Huggins, R.M. and Loesch, D.Z. (1998). On the Analysis of Mixed Longitudinal Growth Data. *Biometrics*, **54**, 583–595.
- Karlsson, A. (2008). Nonlinear quantile regression estimation of longitudinal data. *Communications in Statistics - Simulation and Computation*, **37**, 114–131.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33–55.
- Koenker, R. and D'Orey, V. (1987). Computing regression quantiles. *Applied Statistics*, **36**, 383–393.
- Koenker, R. and Hallock, K.F. (2001). Quantile Regression: An Introduction. *Journal of Economic Perspectives*, **15**, 143–156.
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74–89.
- Kokic, P., Chambers, R., Breckling, J. and Beare, S. (1997). A measure of production performance. *J. Bus. Econom. Statist.*, **10**, 419–435.
- Jung, S. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, **91**, 251–257.
- Liu, Y. and Bottai, M. (2009). Mixed-Effects Models for Conditional Quantiles with Longitudinal Data. *The International Journal of Biostatistics*, **5**, 1, Art. 28.
- Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G. and Zhao, L.P. (1997). Quantile regression methods for longitudinal data with drop-outs: application to cd4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **46**, 463–476.
- Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model analysis of variance. *Ann. Statist.*, **5**, 746–762.
- Mosteller, F. and Tukey, J. (1977). *Data Analysis and Regression*. Addison-Wesley.
- Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.
- Petterson, H.D. and Nabumgoomu, F. (1992). REML and the analysis of a series of crop variety trials. *Proceedings of the 16th International Biometric Conference*, Biometric Society, Washington, DC.

- Pinheiro, J. and Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, Germany.
- Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*. STATA Press, USA.
- Richardson, A.M. and Welsh, A.H. (1995). Robust Estimation in the Mixed Linear Model. *Biometrics*, **51**, 1429–1439.
- R Development Core Team (2004). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL: <http://www.R-project.org>.
- Singer, J.D. and Willett, J.B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, New York.
- Sinha, S.K. and Rao, J.N.K. (2009). Robust Small Area Estimation. *Canadian Journal of Statistics*, **37**, 381–399.
- Staudenmayer, J., Lake, E.E. and Wand, M.P. (2009). Robustness for general design mixed models using the t-distribution. *Statistical Modelling*, **9**, 235–255.
- Street, J.O., Carroll, R.J. and Ruppert, D. (1988). A Note on Computing Robust Regression Estimates Via Iteratively Reweighted Least Squares. *The American Statistician*, **42**, 152–154.
- Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*. Springer, New York.
- Welsh, A.H. and Richardson, A.M. (1997). Approaches to the Robust Estimation of Mixed Models. G.S. Maddala and C.R. Rao, eds., *Handbook of Statistics*, **15**, Chapter 13.

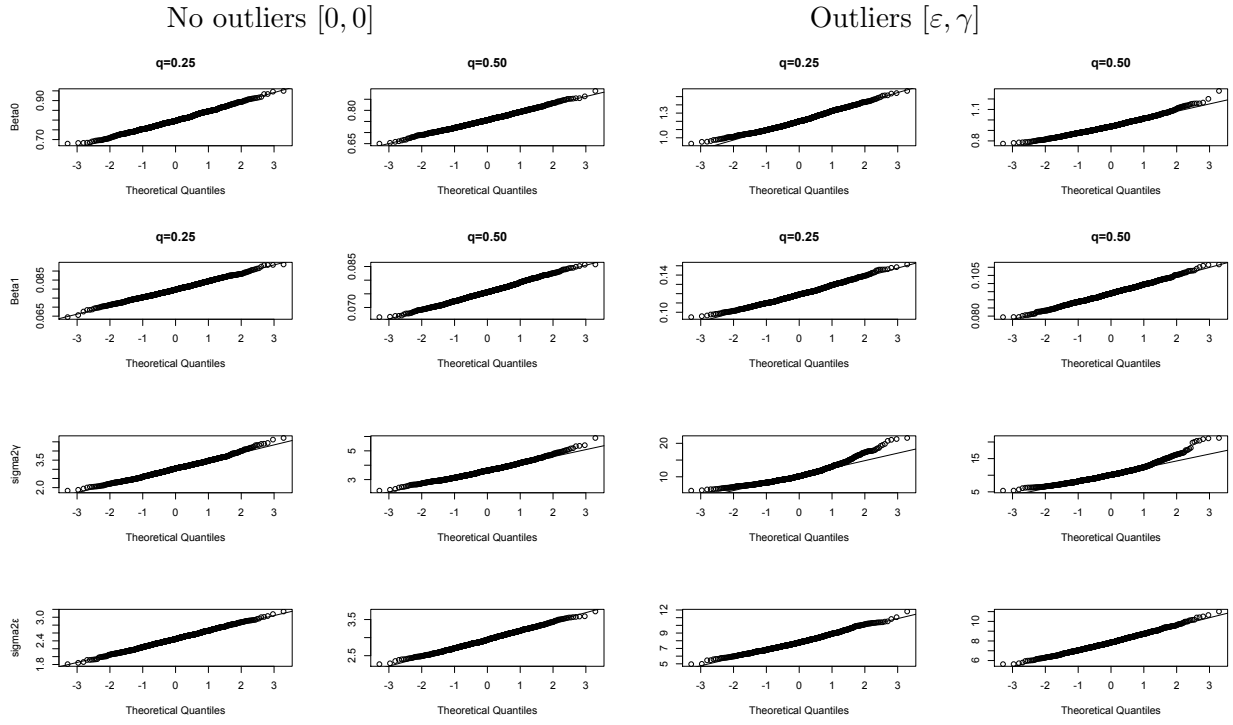
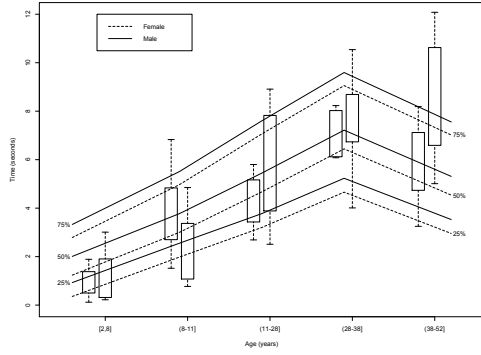
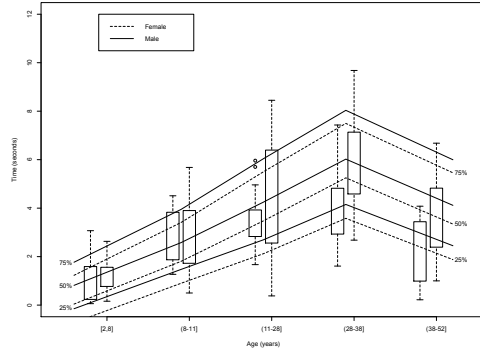


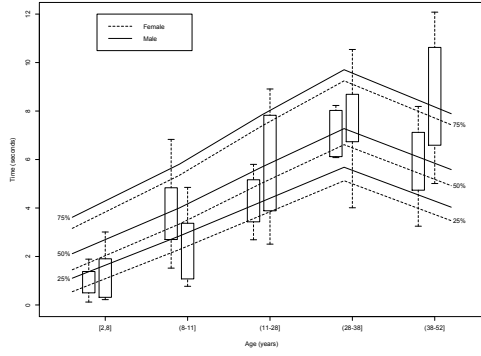
Figure 1: Normal Q–Q plots of estimates of β_q and $\sigma_{\gamma_q}^2, \sigma_{\varepsilon_q}^2$ for quantiles $q = 0.25, 0.5$. Estimated values are obtained from ERE with $c = 100$ (scenario $[0, 0]$) and MQRE with $c = 1.345$ (scenario $[\varepsilon, \gamma]$).



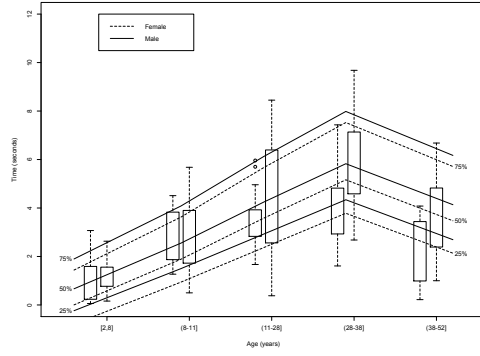
(a)



(b)



(c)



(d)

Figure 2: M-quantile fitted lines for circle (a) and box (b) tracker and expectile fitted lines for circle (c) and box (d) tracker.

Table 1: Coverage rates and average length of 95% confidence intervals, empirical standard errors and estimated standard errors of $\hat{\beta}_q$ and $\hat{\sigma}_{\gamma_q}^2, \hat{\sigma}_{\epsilon_q}^2$ for $q = 0.25, 0.5$ using MQRE (scenario with outliers) and ERE (scenario with no outliers). The results are based on 1000 Monte Carlo replications for each of the two scenarios.

Estimator	Coverage(%)	Mean length	Empirical s.e.	Estimated s.e.
No outliers				
$q = 0.25$				
$\hat{\beta}_0$	92.8	3.164	0.888	0.807
$\hat{\beta}_1$	95.1	0.313	0.079	0.079
$\hat{\sigma}_{\gamma_q}^2$	90.4	11.719	3.246	2.989
$\hat{\sigma}_{\epsilon_q}^2$	93.5	9.564	2.549	2.440
$q = 0.5$				
$\hat{\beta}_0$	95.0	2.966	0.767	0.756
$\hat{\beta}_1$	94.7	0.296	0.075	0.075
$\hat{\sigma}_{\gamma_q}^2$	93.0	14.063	3.666	3.587
$\hat{\sigma}_{\epsilon_q}^2$	94.0	11.494	2.911	2.935
Outliers				
$q = 0.25$				
$\hat{\beta}_0$	94.0	4.728	1.247	1.206
$\hat{\beta}_1$	95.3	0.468	0.120	0.119
$\hat{\sigma}_{\gamma_q}^2$	92.3	41.129	11.709	10.492
$\hat{\sigma}_{\epsilon_q}^2$	91.0	30.605	8.957	7.807
$q = 0.5$				
$\hat{\beta}_0$	92.7	3.698	1.037	0.943
$\hat{\beta}_1$	93.0	0.369	0.101	0.0942
$\hat{\sigma}_{\gamma_q}^2$	92.5	40.113	11.383	10.233
$\hat{\sigma}_{\epsilon_q}^2$	93.5	30.757	7.954	7.846

Table 2: Values of bias (ARB), efficiency (EFF), and average of point estimates over simulations of fixed effects and variance components under the four data generating scenarios and the alternative regression fits: MQRE, MQ and ERE at $q = 0.25$. The LRE estimates are also reported for comparisons with MQRE and MQ at $q = 0.5$. The results are based on 500 Monte Carlo replications for each of the four scenarios.

$q = 0.25$														$q = 0.5$																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
$\hat{\beta}_0$				$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\sigma}_u^2$				$\hat{\sigma}_\varepsilon^2$				$\hat{\beta}_0$				$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\sigma}_u^2$				$\hat{\sigma}_\varepsilon^2$																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
ARB	EFF	β_0	ARB	EFF	β_1	ARB	EFF	β_2	ARB	EFF	β_2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	β_2	ARB	EFF	β_2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	σ_u^2	ARB	EFF	

Table 3: MASE and MADE values for MQRE, LRE, and QRRE at $q = 0.5$ and for the different data generating scenarios. The results are based on 500 Monte Carlo replications for each of the four scenarios. The MASE and MADE of MQRE are equal to 1.00.

	MASE	MADE
	[0, 0]	
MQRE	1.00	1.00
QRRE	1.15	1.07
LRE	0.93	0.97
	[ε , 0]	
MQRE	1.00	1.00
QRRE	1.00	1.00
LRE	1.57	1.24
	[0, γ]	
MQRE	1.00	1.00
QRRE	1.25	1.12
LRE	1.00	1.00
	[ε , γ]	
MQRE	1.00	1.00
QRRE	1.23	1.12
LRE	2.20	1.44

Table 4: Assessing the effect of the longitudinal structure by comparing the standard errors of the MQRE and ERE random intercepts regression without covariates with those from the corresponding single level regression models.

	Constant	$\hat{\sigma}_u^2$	$\hat{\sigma}_\varepsilon^2$
Quantile	MQRE		
$q = 0.25$	1.67 (0.18)	2.92 (0.60)	0.55 (0.06)
$q = 0.50$	3.40 (0.23)	5.44 (0.95)	0.72 (0.07)
$q = 0.75$	5.34 (0.31)	5.62 (0.98)	0.76 (0.08)
M-quantile single level			
$q = 0.25$	2.37 (0.10)	-	-
$q = 0.50$	3.49 (0.12)	-	-
$q = 0.75$	4.82 (0.15)	-	-
ERE			
$q = 0.25$	2.06 (0.18)	3.42 (0.49)	0.65 (0.05)
$q = 0.50$	3.65 (0.23)	5.57 (0.78)	0.78 (0.06)
$q = 0.75$	5.37 (0.27)	5.04 (0.71)	0.71 (0.06)
Expectile single level			
$q = 0.25$	2.59 (0.10)	-	-
$q = 0.50$	3.64 (0.12)	-	-
$q = 0.75$	4.85 (0.15)	-	-

Table 5: Parameter estimates and corresponding standard error estimates in parentheses for the rotary pursuit tracking experiment data. Estimates of the MQRE and ERE are obtained using ML estimation. The baseline age group is (2,8).

	Constant	Trial	Gender	Shape	Age1	Age2	Age3	Age4	$\hat{\sigma}_u^2$	$\hat{\sigma}_\varepsilon^2$
Quantile	MQRE									
$q = 0.25$	0.43 (0.41)	0.33 (0.03)	-0.57 (0.32)	-1.08 (0.33)	1.62 (0.47)	2.84 (0.48)	4.30 (0.49)	2.60 (0.49)	1.53 (0.32)	0.44 (0.04)
$q = 0.50$	1.49 (0.37)	0.35 (0.03)	-0.77 (0.30)	-1.19 (0.30)	1.76 (0.43)	3.43 (0.44)	5.20 (0.45)	3.30 (0.45)	1.97 (0.36)	0.50 (0.05)
$q = 0.75$	2.80 (0.53)	0.35 (0.04)	-0.53 (0.42)	-1.56 (0.43)	2.18 (0.61)	4.27 (0.63)	6.26 (0.63)	4.22 (0.64)	2.31 (0.46)	0.47 (0.05)
	ERE									
$q = 0.25$	0.58 (0.38)	0.35 (0.03)	-0.56 (0.30)	-1.34 (0.31)	1.73 (0.44)	3.15 (0.45)	4.58 (0.45)	2.93 (0.45)	1.59 (0.23)	0.47 (0.04)
$q = 0.50$	1.57 (0.39)	0.36 (0.03)	-0.66 (0.31)	-1.44 (0.31)	1.88 (0.45)	3.57 (0.46)	5.16 (0.46)	3.47 (0.47)	2.11 (0.31)	0.56 (0.04)
$q = 0.75$	3.07 (0.49)	0.36 (0.03)	-0.45 (0.39)	-1.72 (0.40)	2.19 (0.57)	4.22 (0.58)	6.08 (0.58)	4.27 (0.59)	2.47 (0.35)	0.45 (0.03)