A bias-adjusted estimator in quantile regression for clustered data

Battagliola, Maria Laura; Sørensen, Helle; Tolver, Anders; Staicu, Ana Maria

Published in:
Econometrics and Statistics

DOI:
10.1016/j.ecosta.2021.07.003

Publication date:
2022

Document version
Peer reviewed version

Document license:
CC BY-NC-ND

Citation for published version (APA):
A bias-adjusted estimator in quantile regression for clustered data

Maria Laura Battagliola\textsuperscript{1}, Helle Sørensen\textsuperscript{1,}\textsuperscript{*}, Anders Tolver\textsuperscript{1}, Ana-Maria Staicu\textsuperscript{2}

\textsuperscript{1}Department of Mathematical Sciences, University of Copenhagen, 
\textsuperscript{2}Department of Statistics, North Carolina State University, 
\textsuperscript{*}Corresponding author: helle@math.ku.dk

March 26, 2021

Abstract

The manuscript discusses how to incorporate random effects for quantile regression models for clustered data with focus on settings with many but small clusters. The paper has three contributions: (i) documenting that existing methods may lead to severely biased estimators for fixed effects parameters; (ii) proposing a new two-step estimation methodology where predictions of the random effects are first computed by a pseudo likelihood approach (the LQMM method) and then used as offsets in standard quantile regression; (iii) proposing a novel bootstrap sampling procedure in order to reduce bias of the two-step estimator and compute confidence intervals. The proposed estimation and associated inference is assessed numerically through rigorous simulation studies and applied to an AIDS Clinical Trial Group (ACTG) study.

Keywords: Linear quantile regression; Clustered data; Random effects; Bias-adjustment; Wild bootstrap; ACTG study

1 Introduction

Quantile regression has been introduced by Koenker and Bassett Jr (1978) as a way to describe the association between covariates and quantiles of the response distribution at pre-set quantile levels. See the comprehensive monographs by Koenker (2005) and Koenker et al. (2017) on quantile regression. In recent years, quantile regression has for example been employed in econometrics and finance (Bayer 2018, Wang et al. 2018b, Maciak 2021a,b). In this article we consider linear quantile regression for clustered data, such as longitudinal data, and discuss estimation approaches that properly account for the inherent dependence of the observations within the same cluster. Research in this area has been very active, especially in econometrics, but existing methods for quantile regression estimation are proved to be asymptotically consistent only when both the number of clusters and cluster size increase to infinity. This assumption is rather strong in practice, where the common scenario is that there are many clusters of moderate to small sizes. When the cluster size is small, numerical investigations show (see Figure 2) that the popular quantile regression estimators may exhibit severe bias, even if there are many clusters. This represents a gap in the literature, as data settings that involve many clusters of small to moderate sizes are ubiquitous in medicine and animal science, to name a few.

Existing approaches to account for dependence in parameter estimation of quantile regression for clustered (repeated measures) data treat the cluster-specific parameters either
as fixed or random. For example, Kato et al. (2012) and Galvao and Kato (2016) use cluster-specific intercepts and estimate them as fixed effects parameters together with the quantile regression parameters using the so-called fixed effects quantile regression (FE-QR) and fixed effects smoothed quantile regression (FE-SQR), respectively, while Galvao and Wang (2015) and Galvao et al. (2017) develop minimum-distance-based estimation for the same purpose. Some approaches consider shrinkage to deal with an increasing number of clusters, in the presence of cluster-specific parameters. Penalized quantile regression for longitudinal data is discussed by Koenker (2004), Lamarche (2010), Harding and Lamarche (2017) and Gu and Volgushev (2019). Canay (2011) proposes a two-step estimator, relying on mean regression estimates of cluster-specific intercepts, see also Besstremyannaya and Golovan (2019). Geraci and Bottai (2007) and Geraci and Bottai (2014) introduce a pseudo-likelihood approach, where a linear quantile mixed model (LQMM) with random cluster parameters is used as a working model, and Galarza et al. (2017) develop an EM-based estimation methodology for the LQMM framework. Abrevaya and Dahl (2008) discuss estimation in a model with correlated random effects (CRE), and Luo et al. (2012) consider a fully Bayesian quantile inference using Markov Chain Monte Carlo, to account for correlated random effects. We consider a frequentist perspective and propose a novel two-step estimation approach and associated inference that rely on the LQMM framework.

When the cluster-specific parameters are treated and estimated as fixed effects parameters, estimation suffers from what is known in the literature as the “incidental parameters problem” (Neyman and Scott, 1948; Lancaster, 2000): the number of (nuisance) parameters grows with the number of clusters, leading to inconsistent joint estimation, when the cluster size is small. Not surprisingly, only asymptotic scenarios where both the number of clusters and the cluster size increase to infinity have been studied (Koenker, 2004; Kato et al., 2012; Galvao and Kato, 2016; Canay, 2011; Besstremyannaya and Golovan, 2019). To bypass the issues caused by the incidental parameter problem, the cluster-specific parameters can be modeled as random effects; however, asymptotic properties are not studied for the LQMM-based estimator (Geraci and Bottai, 2007, 2014).

Different solutions have been suggested for bias-adjustment in the case of small clusters: Galvao and Kato (2016) introduce an analytical adjustment for FE-SQR based on asymptotic analysis, nonetheless the approach requires an optimal bandwidth selection, which is challenging in practice. The authors also adapt the half-panel jackknife method (Dhaene and Jochmans, 2015) to longitudinal quantile regression. We consider the use of half-panel jackknife for bias correction in our numerical investigation. Usually, bootstrap methods have been used for construction of confidence intervals in models with cluster-specific effects (Galvao and Montes-Rojas, 2015; Canay, 2011; Geraci and Bottai, 2014), and for marginal models (without cluster-specific effects), see for example Karlsson (2009) and Hagemann (2017). We introduce a non-standard bootstrap technique for both bias-adjustment and inference of quantile regression parameters, in the context of clustered (longitudinal) data.

This paper makes three main contributions. First, we numerically demonstrate that Koenker’s penalized estimator, Canay’s two-step estimator and the LQMM estimator can be severely biased when clusters are small or of moderate size. Although no papers have claimed the opposite, we are the first to raise this issue. Second, we propose a new estimation methodology and associated inference for the quantile regression parameters. The point estimator is computed in two steps: (i) an LQMM framework is used to predict the cluster-specific parameters; and (ii) the predictions are used as offsets in a standard quantile regression. The two-step estimator is furthermore adjusted for bias using bootstrap, and the third contribution is the novel combination of wild bootstrap and ordinary resampling, that reduces bias and allows to construct confidence intervals that have good coverage performance. Numerical studies show that the proposed estimator has considerably smaller bias than the existing competitors, when the cluster size is small.
The structure of the paper is as follows: we set up the model framework in Section 2. In Section 3, we summarize some of the existing estimation methods in quantile regression for repeated measures data and then present the proposed estimation method. The estimation method is evaluated numerically in a thorough simulation study in Section 4 (with additional results in the appendix) and applied to a clinical trial regarding HIV treatments in Section 5. The paper concludes with Section 6, which discusses the main findings.

2 Regression framework

Let \( (Y_{ij}, x_{ij})_{j=1}^{n_i} \) be the observed data for the \( i \)th cluster \((i = 1, \ldots, N)\), where \( x_{ij} \in \mathbb{R}^{p-1} \) is the vector of covariates corresponding to the \( j \)th observation of the \( i \)th cluster and \( Y_{ij} \in \mathbb{R} \) is the respective response. Here \( n_i \) denotes the cluster size and the responses are assumed independent across different clusters but expected to be correlated within the same cluster. Let \( \tau \in (0, 1) \) be a fixed quantile level of interest, and let \( Q_{Y_{ij}|x_{ij}}(\tau) \) be the \( \tau \)th quantile of the conditional distribution of \( Y_{ij} \) given \( x_{ij} \) for cluster \( i \). Consider a linear quantile regression model

\[
Q_{Y_{ij}|x_{ij}}(\tau) = X_{ij}^T \beta_\tau, \tag{2.1}
\]

where \( X_{ij}^T = (1, x_{ij}^T) \) and \( \beta_\tau = (\beta_{1,\tau}^\tau, \ldots, \beta_{p,\tau}^\tau) \) is an unknown vector regression parameter that quantifies the association between the covariates and the \( \tau \)-quantile of the response for cluster \( i \). Due to the definition of \( X_{ij}^T \), the first component of \( \beta_\tau \) is the intercept; by an abuse of notation we refer to \( X_{ij} \) as the vector of covariates.

This model formulation allows for cluster-level effects for every scalar component of \( X_{ij} \); an equivalent formulation is to represent the cluster-level effect as the sum of a population level effect and a cluster-specific deviation. Such formulation is standard in the mixed effects model representation [Laird and Ware 1982], and we adopt it here as well. As for mean regression, all covariates are not necessarily modeled with cluster-specific levels, and the selection of variables without cluster-specific effects can be based on interpretational as well as computational arguments. Without loss of generality, assume that only the first \( q \leq p \) components of \( X_{ij} \) have cluster-varying effects; denote by \( Z_{ij} \) the vector formed by the first \( q \) elements of \( X_{ij} \). The remaining \( n - q \) components of \( X_{ij} \) have only population level effect. The effects corresponding to \( Z_{ij} \) are used to account for the dependence of the observations within the same cluster; for example, Koenker (2004), Canay (2011), and Galvao and Kato (2017) used a random intercept only \((q = 1)\) to model this dependence.

Using the terminology from linear mixed effects we can re-write model (2.1) as

\[
Q_{Y_{ij}|x_{ij}}(\tau) = X_{ij}^T \beta_\tau + Z_{ij}^T u_\tau, \tag{2.2}
\]

by separating the quantile regression parameters that describe a population level effect, \( \beta_\tau = (\beta_{1,\tau}^\tau, \ldots, \beta_{q,\tau}^\tau) \), from the ones that describe cluster-specific deviations, \( u_\tau = (u_{1,\tau}^\tau, \ldots, u_{n,\tau}^\tau) \). Just like in linear mixed models, it is assumed that \( u_\tau \) are zero mean random quantities. Our primary interest lies in the estimation of \( \beta_\tau \) in situations with many clusters (large \( N \)) but modest cluster sizes (small \( n_i \)).

Let \( \mathbf{u}_\tau = (u_{1,\tau}, \ldots, u_{n,\tau}) \) denote the collection of (unobserved) cluster-specific parameters. Moreover, let \( \mathbf{Y} \) be the vector of the (observed) responses \( Y_{ij} \). Consider the loss function

\[
L(\beta_\tau, \mathbf{u}_\tau; \mathbf{Y}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \rho_\tau (Y_{ij} - X_{ij}^T \beta_\tau - Z_{ij}^T u_\tau), \tag{2.3}
\]

where \( \rho_\tau (v) = v(\tau - 1(v < 0)) \) is the check function [Koenker and Bassett Jr 1978]. If the values of the cluster-specific effects, \( u_\tau \), were observed, a natural estimator would be the
linear quantile regression estimator corresponding to the covariates \( X_{ij} \) and the modified responses \( Y_{ij} - Z_{ij}^T u_i^\tau \). We call this the oracle estimator,

\[
\hat{\beta}_\text{oracle} = \arg \min_{\beta^\tau} L(\beta^\tau, \mathbf{u}^\tau; Y);
\]  

(2.4)
evidently the estimator \( \hat{\beta}_\text{oracle} \) enjoys the asymptotic properties of a standard quantile regression estimator (Koenker, 2005). However, \( \hat{\beta}_\text{oracle} \) is an unattainable estimator, as \( u_i^\tau \)’s are not observed, and the question we consider in this paper concerns the effect of uncertainty in the cluster-specific effects on estimating the population level quantile regression parameter.

One way to address the estimation problem is to treat \( u_i^\tau \)’s in (2.2) as fixed effects parameters and have them estimated jointly with \( \beta^\tau \) using a standard quantile regression framework. The FE-QR estimation of Kato et al. (2012) minimizes the loss function (2.3) with respect to both \( \beta^\tau \) and \( u^\tau \). With this approach, the number of parameters grows at the same rate as the number of clusters, so the estimator of \( \beta^\tau \) is only consistent in asymptotic scenarios where \( n_i \) grows faster than \( N \) (Kato et al., 2012).

We will instead pursue an approach to estimate \( \beta^\tau \), when \( u^\tau \)’s are treated as random. Similar to the generalized linear mixed effects framework, there are two interpretations of the covariates’ effects on the response distribution quantile. On one hand, we have the conditional perspective, following from the definition (2.2) that \( P(Y_{ij} \leq X_{ij}^T \beta^\tau + Z_{ij}^T u_i^\tau | X_{ij}, u_i^\tau) = \tau \), which states that \( \beta^\tau \) is the quantile regression parameter associated with the covariates \( X_{ij} \), conditional on the cluster-specific effects. On the other hand, we have the marginal perspective that \( P(Y_{ij} \leq X_{ij}^T \tilde{\beta}^\tau | X_{ij}) = \tau \), which describes the covariates’ effect on the \( \tau \)-quantile of the marginal distribution of \( Y_{ij} \). The two quantile regression parameters (\( \beta^\tau \) and \( \tilde{\beta}^\tau \)) are generally different in the same manner that a fixed effects parameter of a generalized linear mixed model has a different interpretation than its counterpart in a marginal or population average approach (Zeger et al., 1988; Neuhaus et al., 1991). The difference between the conditional and marginal quantile models is discussed more thoroughly in Reich et al. (2009), see also the simulation model in Section 4.

As a consequence, also pointed out in Koenker (2004), it is vital for the estimation of \( \beta^\tau \) of a conditional perspective that the cluster-specific parameters \( u_i^\tau \) are not ignored. Indeed, we illustrate in Section 4.2 that the simple marginal quantile regression estimator \( \hat{\beta}_\text{marg} = \arg \min_{\beta^\tau} L(\beta^\tau, 0; Y) = \arg \min_{\beta^\tau} \sum_{i,j} \rho_\tau(Y_{ij} - X_{ij}^T \beta^\tau) \) based on standard quantile regression (where all \( u_i^\tau \)s are replaced by zero) may be severely biased for \( \beta^\tau \).

The conditional perspective implies that

\[
P(Y_{ij} - Z_{ij}^T u_i^\tau \leq X_{ij}^T \beta^\tau | X_{ij}) = \tau,
\]

where the probability is taken with respect to the joint distribution of \( Y_{ij} \) and \( u_i^\tau \). Inspired by this equality, we propose to first predict the cluster-specific effects and then use these predictions as offset in a standard linear quantile regression model using a transformed response.

3 Estimation

3.1 Review of selected methods for estimation and bias-adjustment

Penalization of cluster-specific parameters

The model (2.2) was first introduced in the literature by Koenker (2004) in a simpler form, where the term \( Z_{ij}^T u_i^\tau \) is replaced by only a cluster-specific intercept, call it \( u_{i0} \), which is
assumed to be quantile-invariant. For fixed quantile level $\tau$, both the parameter $\beta^\tau$ and the cluster-specific intercepts, $u_{i0}$, are estimated by minimizing the penalized loss function

$$L(\beta^\tau, u_{i0}; Y) + \lambda \sum_{i=1}^{N} |u_{i0}|,$$

where $\lambda \geq 0$ is a regularization parameter. Koenker (2004) uses $\ell_1$ penalty in (3.1) due to its computational convenience; in our numerical investigation of the estimators in Section 4, we also use $\ell_2$ penalty and find minor differences. While (3.1) focuses on a single quantile level, Koenker (2004) describes the estimation of the quantile regression parameters simultaneously at multiple quantile levels, by introducing quantile-level weights and minimizing a weighted penalized likelihood.

The $\ell_1$-penalized estimator for $\beta^\tau$ is consistent and asymptotically normal, provided that $N^n/n \to 0$ for some $a > 0$ (where $n_i = n$); see Koenker (2004). Nonetheless, when the cluster size, $n_i$, is small the estimator may not enjoy these theoretical properties and can be seriously biased, especially for extreme quantile levels; see Section 4.

Canay’s two-step estimator

Canay (2011) assumes a cluster-specific intercept, $u_{i0}$, too, but considers a two-step procedure to estimate the linear quantile regression parameter $\beta^\tau$ of (2.2). First, $u_{i0}$ are estimated as part of the fixed parameters in a mean regression framework. Second, the quantile regression parameter $\beta^\tau$ is estimated using a standard quantile regression framework (Koenker and Bassett Jr, 1978) applied to adjusted responses $\tilde{Y}_{ij} = Y_{ij} - \hat{u}_{i0}$, where $\hat{u}_{i0}$ denotes the estimated cluster-specific effects from the previous step. Equivalently, $\beta^\tau$ is estimated by minimizing the loss function (2.3) with $Z_{0i} = 1$ and $u^\tau$ replaced by $\hat{u}_{i0}$, the vector containing the $\hat{u}_{i0}$s:

$$\hat{\beta}^\tau_{\text{Canay}} = \arg \min_{\beta^\tau} L(\beta^\tau, \hat{u}_{i0}; Y).$$

Canay (2011) and Besstremyannaya and Golovan (2019) discuss asymptotic properties for $\hat{\beta}^\tau_{\text{Canay}}$ in scenarios where both the number of clusters and cluster size increase.

The use of the mean regression in the first step is justified in Canay’s set-up because only intercepts are allowed to be cluster-specific, and the deviations from the average are assumed to be constant over quantile levels. In such case, the random effects correspond to location shifts; their estimation is quantile-invariant, which may be restrictive. Moreover, while treating $u_{i0}$s as fixed parameters as opposed to random may lead to negligible differences, in terms of estimation, for large clusters, the correct approach for small clusters is to treat them as random parameters. To address this issue, we propose a new quantile regression estimator in Section 3.2, which is inspired by Canay (2011).

Marginalization over random effects in a working model (LQMM)

Geraci and Bottai (2007, 2014) propose to embed the problem in a fully specified working model, a linear quantile mixed model (LQMM), using the duality between the quantile loss (check function) and the asymmetric Laplace distribution (ALD, Yu and Zhang (2005)). Specifically, assume $u_i \sim f(\cdot; \varphi)$ for some density $f$ that is parameterized by a scale parameter $\varphi$ and posit the following joint model for the responses $Y_{i,j}$s and the cluster-specific $u_i$s:

$$Y_{ij} | u_i, X_{ij} \overset{\text{ind}}{\sim} ALD(X_{ij}^T \beta^\tau + Z_{ij}^T u_i, \sigma, \tau), \quad j = 1, ..., n_i$$

$$u_i \overset{\text{ind}}{\sim} f(\cdot; \varphi),$$

(3.2)
for \( i = 1, \ldots, N \), where \( \sigma \) is a scale parameter for the residual distribution. The conditional \( \tau \)-quantile function associated to the working model is given by (2.2), and the conditional likelihood of \( Y_{ij} \) given \( X_{ij} \) and \( u_i \) takes the form (2.3); with \( u_i^* = u_i \).

Estimation of model parameters \((\beta^\tau, \sigma, \varphi)\) is based on maximizing the pseudo likelihood of \( Y \) obtained by integrating the joint density of \((Y_{i1}, \ldots, Y_{in}, u_i)\) with respect to the distribution of latent random effects \( u_i \). In practice, the random effects are assumed to be drawn either from a Gaussian distribution \( N(0, \varphi^2) \) or a Laplace distribution \( ALD(0, \varphi, 1/2) \), see Geraci (2014) for details about the computations. In the special case of random intercepts only, when the Laplace distribution is used for the cluster-specific parameters \( u_{i0s} \), maximizing the joint model (3.2) is equivalent to minimizing Koenker’s penalized loss function, while if the Gaussian distribution is used, then maximizing the joint model (3.2) is equivalent to minimizing the \( \ell_2 \)-penalized criterion. From this perspective, the tuning parameters using Koenker’s penalization approach are scale parameters in the joint model framework and thus can be estimated with increased computational efficiency. Finally, once the parameters \( \beta^\tau, \sigma \) and \( \varphi \) are estimated, the random effects can be predicted using best linear predictors (BLPs), see equation (12) in Geraci and Bottai (2014). These predictions are essential ingredients for the new estimator suggested in Section 3.2; note that the computed predictions vary with the level \( \tau \) even though \( u_i \) in the model (3.2) does not.

Geraci and Bottai (2007) and Geraci and Bottai (2014) do not discuss asymptotics for the LQMM estimator, but if the working assumptions are true (ALD for the within-cluster distribution and Gaussian or Laplace distribution for the random effects), then the LQMM estimator is the maximum likelihood estimator, and the usual asymptotic results hold. On the other hand, the bias of the LQMM estimator may be non-negligible, even when \( N \) is large, if the data generating process does not coincide with the working model. This will be illustrated in Section 4.2.

**Jackknife-based bias-adjustment for an existing estimator**

Since the estimators above show bias when used for clustered data, a bias reduction adjustment would be appropriate. There are various ways to do this; one approach to reduce the bias of an estimator is by using a jackknife bias-adjustment. The half-panel jackknife was first introduced in Dhaene and Jochmans (2015) as a method for bias correction for mean regression in longitudinal settings with many subjects and fixed panel size. Later, it was applied to the FE-SQR estimator for longitudinal quantile regression (Galvao and Kato, 2016); we describe it here for clustered data.

We randomly split the dataset into two sub-datasets, each containing half of the observations from every cluster. Denote the quantile regression estimator from the two sub-datasets by \( \hat{\beta}^\tau_1 \) and \( \hat{\beta}^\tau_2 \), respectively, and let \( \hat{\beta}^\tau \) be the estimator from the full dataset. Then, the half-panel jackknife estimator \( \hat{\beta}^\tau_{\text{jackknife}} \) is defined as

\[
\hat{\beta}^\tau_{\text{jackknife}} = \hat{\beta}^\tau - \left( \frac{1}{2} (\hat{\beta}^\tau_1 + \hat{\beta}^\tau_2) - \hat{\beta}^\tau \right) = 2 \hat{\beta}^\tau - \frac{1}{2} (\hat{\beta}^\tau_1 + \hat{\beta}^\tau_2).
\]  

To gain some intuition about the bias reduction of this estimator, assume that all clusters have equal size \( n \) and that the asymptotic bias of the initial estimator \( \hat{\beta}^\tau \) is of the form \( C/n + o(n^{-1}) \) for some constant \( C \). Then the asymptotic bias of the jackknife estimator \( \hat{\beta}^\tau_{\text{jackknife}} \) is of order \( o(n^{-1}) \), so the order of the bias is reduced. Nonetheless, empirical studies indicate that while the adjustment indeed reduces the bias, the resulting variance of the estimator is increased; see Galvao and Kato (2016).
3.2 Proposed quantile estimation with reduced bias

A new two-step estimator (unadjusted)

We propose to estimate the linear quantile regression parameter $\beta^\tau$ using a new approach, which is inspired by the LQMM estimation framework and Canay (2011). It consists of two steps:

Step 1: Use the LQMM framework to predict the cluster-specific random effects by the best linear predictors (BLPs) and center them; denote the centered prediction for cluster $i$ by $\tilde{u}_i^\tau$;

Step 2: Transform the responses to $\tilde{Y}_{ij} = Y_{ij} - Z_{ij}^T \tilde{u}_i^\tau$ and use the standard quantile regression framework for the new responses $\tilde{Y}_{ij}$ and covariates $X_{ij}$ to estimate $\beta^\tau$.

There are two key differences between the proposed approach and Canay (2011): 1) Canay estimates the cluster-specific effects using a mean regression framework, whereas we use a quantile regression model, and 2) Canay estimates the cluster-specific effects by treating them as fixed parameters; in contrast we view and estimate them as random parameters. We illustrate in Section 4 that these differences have a large impact in terms of the estimation quality of quantile regression parameters.

Figure 1 shows a comparison between true random effects ($x$-axis) and their predicted values ($y$-axis) for the first cluster from 200 simulated data sets representing the benchmark scenario in Section 4. The BLPs capture the variation among clusters quite well, but it is clear that some degree of shrinkage takes place as more extreme random effects are drawn towards zero.

The second step consists of standard quantile regression applied to $Y_{ij} - Z_{ij}^T \tilde{u}_i^\tau$; equivalently the quantile regression parameter is estimated by minimizing the loss function (2.3), with $u$ fixed at value $\tilde{u}_i^\tau$, the vector containing $\tilde{u}_i^\tau$'s:

$$\hat{\beta}_{\text{two-step}} = \arg\min_{\beta^\tau} L(\beta^\tau, \tilde{u}_i^\tau; Y).$$

Our two-step estimator turns out to have considerably smaller bias than the LQMM estimator; yet, the deviation between the true and estimated random effects introduces some bias. To bypass this issue, we propose a bias-corrected adjustment based on bootstrap as explained below. The second step can be carried out with standard software, which
typically provides standard errors for each component of the vector $\beta^\tau$. However, it is important to recognize that these uncertainty estimates are not necessarily reliable, as they only account for the sampling variability of $\hat\beta_{\text{two-step}}^\tau$ conditional on the random effects, not for the extra variation due to the uncertainty in predicting the random effects. We propose to use bootstrap to estimate the total variation of $\hat\beta_{\text{two-step}}^\tau$. We describe the bootstrap procedures used for bias-adjustment and estimation of variability in the following.

**Bootstrap sampling for bias-adjustment**

We propose a semi-parametric-type of bootstrap, which combines non-parametric bootstrap and wild bootstrap and relies on the linearity of the quantile regression model. Let $U = \{\tilde{u}_1^\tau, \ldots, \tilde{u}_N^\tau\}$ be the sample of predicted cluster-specific effects obtained with two-step estimation procedure and for each $i$ and $j$ denote the observed residuals by $\varepsilon_{ij} = Y_{ij} - X_{ij}^T \hat{\beta}_{\text{two-step}}^\tau - Z_{ij}^T \tilde{u}_i^\tau$.

We define the bootstrap sample as $\{(Y_{ij}^*, X_{ij}, Z_{ij})_{j=1}^{n_i}, u_i^\tau*\}_{i=1}^N$ where $u_i^\tau*$ are obtained by resampling with replacement from $U$ and $Y_{ij}^*$ is defined by

$$Y_{ij}^* = X_{ij}^T \hat{\beta}_{\text{two-step}}^\tau + Z_{ij}^T u_i^\tau* + \varepsilon_{ij}^*,$$

where $\varepsilon_{ij}*$ are attained by wild bootstrap; see Wu (1986) and Liu (1988) who introduced this method in the context of mean regression. Specifically, let $\varepsilon_{ij}^* = w_{ij} |\varepsilon_{ij}|$, where $w_{ij}$s are drawn independently from the following distribution:

$$w = \begin{cases} 2(1 - \tau), & \text{with probability } 1 - \tau \\ -2\tau, & \text{with probability } \tau \end{cases}$$

which has the $\tau$-quantile equal to 0. The idea of scaling the residuals by weights drawn from an asymmetric distribution was proposed by Feng et al. (2011); as Wang et al. (2018a) also recognized, the wild bootstrap captures asymmetry and homoscedasticity better than ordinary resampling of residuals. Notice that the coupling between covariates and residuals is maintained in the equation (3.4) in the sense that each residual is used to generate a bootstrap value for its own observation.

Bootstrap methods have been used for inference on quantile regression for longitudinal data. Most of the approaches rely on non-parametric resampling where complete clusters are sampled with replacement, by sampling the covariates and the outcomes jointly (Canay 2011; Kato et al. 2012; Galvao and Montes-Rojas 2015; Geraci and Bottai 2014; Karlsson 2009). This method is useful for evaluation of an estimator’s variation, and thus for computation of standard errors and confidence intervals. However, we expect such bootstrap estimators to be centered around the estimate from the observed data, and they would therefore not be useful for bias-adjustment. In contrast, our bootstrap procedure ensures that the resampled observations are generated from a distribution with $\hat{\beta}_{\text{two-step}}^\tau$ as the “true” parameter; therefore, we can measure bias as the deviation between $\hat{\beta}_{\text{two-step}}^\tau$ and the bootstrap estimates. Details are given below. Our proposed bootstrap method (abbreviated RW, for standard Resampling and Wild) is compared with resampling of complete clusters and two additional approaches in Section 4.

The RW bootstrap sampling procedure ensures that, conditional on the resampled random effects, the model assumption about the association between the covariates and the quantile at level $\tau$ is satisfied with $\beta^\tau = \hat{\beta}_{\text{two-step}}^\tau$ (obtained from the observed data). Furthermore, if the random effects were known then all observations were independent, and the distribution of the bootstrap estimators obtained with wild bootstrap would represent the sampling distribution of $\hat{\beta}_{\text{two-step}}^\tau$. Feng et al. (2011) and Wang et al. (2018a). However, due to the potential deviation between the working model in LQMM and the true data generating model, the empirical distribution of LQMM predictors of the random effects may
not fully represent the cluster-to-cluster variation, and since this variation is driving the bias, the proposed estimator does not completely remove the bias of the initial estimator asymptotically.

Once a bootstrap sample is available, the quantile regression estimator is obtained by using the proposed two-step estimation approach. At this part, information about the resampled cluster-specific effects are ignored; nonetheless these terms are used in a subsequent step, when we estimate the estimator’s variability. The bootstrap estimate of the quantile regression parameter is obtained by averaging the estimates in $B$ such bootstrap samples. If $\hat{\beta}_{\text{two-step}, b}$ denotes the $b$th bootstrap replicate then the overall bootstrap estimate of the quantile regression parameter is $\bar{\hat{\beta}}_{\text{two-step}} = \sum_{b=1}^{B} \hat{\beta}_{\text{two-step}, b} / B$. The deviation $\bar{\beta}_{\text{two-step}} - \hat{\beta}_{\text{two-step}}$ between the overall bootstrap estimate and the original estimate is regarded as an estimate of the bias, so an adjusted estimator (Efron and Tibshirani, 1993, Chapter 10.6) is defined by

$$\hat{\beta}_{\text{adj}} = \hat{\beta}_{\text{two-step}} - \left( \bar{\hat{\beta}}_{\text{two-step}} - \bar{\hat{\beta}}_{\text{two-step}} \right) = 2 \hat{\beta}_{\text{two-step}} - \bar{\hat{\beta}}_{\text{two-step}}.$$  

(3.6)

As illustrated by numerical studies, this quantile regression estimator has reduced bias compared to the (unadjusted) two-step estimator.

**Confidence intervals**

An important advantage of using a bootstrap-based estimator is that it allows to study the variability of the estimator, and we now discuss construction of the confidence intervals for the quantile regression parameter for each component $k$ of the $p$-dimensional parameter $\beta^\tau$. We consider two approaches: the first approach is based on the so-called basic bootstrap method to construct confidence intervals and the second approach capitalizes on the availability of the bootstrap sample of the cluster-specific effects, which is obtained at each step of the bootstrap procedure.

The basic bootstrap $100(1 - \alpha)\%$ confidence intervals (Davison and Hinkley, 1997, eq. 5.6) for $\beta^\tau_k$ are defined as

$$\left( 2 \hat{\beta}_{\text{two-step}, k}^\tau - \beta^\tau_{1-\alpha/2, k} ; 2 \hat{\beta}_{\text{two-step}, k}^\tau - \beta^\tau_{\alpha/2, k} \right), \quad k = 1, \ldots, p,$$

where $\beta^\tau_{\alpha/2, k}$ and $\beta^\tau_{1-\alpha/2, k}$ are the $\alpha/2$ and $(1 - \alpha/2)$ quantiles, respectively, in the bootstrap sample of $\hat{\beta}_{\text{two-step}, k}$.

The second approach to construct confidence intervals relies on a normal asymptotic distribution for the quantile regression estimator and the bootstrap-based estimate of the variance of the quantile regression estimator. However, in contrast to most bootstrap-based confidence intervals constructed this way, the bootstrap standard error alone, $SD_{\text{two-step}, k} = \sqrt{\sum_{b=1}^{B} (\hat{\beta}_{\text{two-step}, k, b}^\tau - \bar{\beta}_{\text{two-step}, k}^\tau)^2 / (B - 1)}$, fails to accurately quantify the full variability of the quantile regression estimator of $\beta^\tau$. This is due to the shrinkage phenomenon of the LQMM predicted cluster-specific effects, which is further perpetuated in the bootstrap samples of $u^\tau_i$s and incorporated in the bootstrap replicates $\hat{\beta}_{\text{two-step}, b}^\tau$.

To bypass this issue, we consider an adjustment. In this regard, denote by $SE_{\text{obs}, k}$ the estimated standard error of the $k$th component of $\hat{\beta}_{\text{two-step}}^\tau$ reported by the standard quantile regression (Koenker and Bassett Jr, 1978) with the cluster-specific effects set to the LQMM predicted values and using the accordingly transformed data (step 2 of our procedure). Recall that this quantity ignores the variability of the cluster-specific effects, and thus underestimates the true variability of the regression estimator. Fortunately, our bootstrap algorithm, by resampling from the empirical distribution of the predicted cluster-effects, allows us to track the variability of the regression estimator induced by
the uncertainty in predicting these effects. Let \( \hat{\beta}_{\text{oracle},b}^{*,*} \) denote the oracle-type quantile regression estimator based on the \( b \)th bootstrap sample, i.e. the \( Y_{ij}^{*,*} \)'s, and by using the "true" values of the cluster-specific effects, i.e. the \( u_{i}^{*,*} \)'s. As before, for each component \( k \) denote by

\[
\bar{\beta}_{\text{oracle},k}^{*,*} = \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_{\text{oracle},k,b}^{*,*},
\]

and its standard deviation

\[
SD_{\text{oracle},k}^{*,*} = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} (\hat{\beta}_{\text{oracle},k,b}^{*,*} - \bar{\beta}_{\text{oracle},k}^{*,*})^2}
\]

the mean and standard deviation, respectively, of the oracle-type quantile regression estimator.

We define the adjusted standard error of the \( k \)th component of the two-step quantile regression estimator as

\[
SE_{\text{adj},k} = \frac{SE_{\text{obs},k}}{SD_{\text{oracle},k}^{*,*}} k = 1, \ldots, p.
\]

Since both terms of the ratio are based on keeping the cluster-specific constant, the ratio is used to account for the shrinkage phenomenon. Another way to understand the adjusted standard error is to view it as a multiplicative factor to the standard error that is reported in our step 2, \( SE_{\text{obs},k}^{*,*} \); in this case the ratio \( SD_{\text{two-step},k}^{*,*}/SD_{\text{oracle},k}^{*,*} \) measures the extra variation of the quantile regression estimator due to estimation of the random cluster-specific effects.

The 100(1 - \( \alpha \))% confidence intervals for \( \beta_{k}^{*,*} \) based on the adjusted standard errors are computed as

\[
\hat{\beta}_{\text{adj},k}^{*,*} \pm q_{1-\alpha/2} \cdot SE_{\text{adj},k}, \quad (3.7)
\]

where \( q_{1-\alpha/2} \) is the \((1 - \alpha/2)\) quantile of \( N(0,1) \). These confidence intervals will later be referred to as SE-adjustment confidence intervals.

We summarize our procedures for estimation and inference in Algorithm 1.

3.3 Software

The two-step quantile regression estimator is computed using two different R (R Core Team, 2020) packages. For the first step, the LQMM estimation method is implemented by the \texttt{lqmm()} function from the package \texttt{lqmm} (Geraci, 2014; Geraci and Bottai, 2014). For the second step, we use standard quantile regression implemented by the function \texttt{rq()} from the \texttt{quantreg} package (Koenker, 2020). Bootstrap datasets are generated with standard sampling functions. An R function for the complete estimation and inference process is available from the corresponding author’s website.

4 Simulations

4.1 Data generating model

We consider a data generating model inspired by the simulation designs in Koenker (2004) and Geraci and Bottai (2014). Specifically,

\[
Y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + (1 + \gamma x_{ij}) e_{ij}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, n_i, \quad (4.1)
\]

where \( u_i \overset{iid}{\sim} N(0, \sigma_u^2) \), \( e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2) \), \( x_{ij} \) are uniformly distributed on \((0, 1)\) and \( \gamma \geq 0 \) is a homoscedasticity-departure parameter. Notice that \( 1 + \gamma x_{ij} \) is always positive. When \( \gamma \neq 0 \), the covariate has both a location shift and a scale effect (Koenker, 2004). In the homoscedastic case (i.e. \( \gamma = 0 \)), the correlation between observations from the same cluster is

\[
\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}.
\]

With a slight abuse of notation, we refer to this ratio as the interclass correlation coefficient (ICC) even when \( \gamma > 0 \).

Model (4.1) implies the following quantile regression model

\[
Q_{Y_{ij} \mid x_{ij}, u_i}(\tau) = \beta_0^{*,*} + \beta_1^{*,*} x_{ij} + u_i,
\]

where \( \beta_0^{*,*} \) and \( \beta_1^{*,*} \) are the estimated intercept and slope parameters, respectively.
Consider data \( \{(Y_{ij}, X_{ij}, Z_{ij})_{i=1}^{N_i} : i = 1, \ldots, N\} \);
Using LQMM framework, obtain the centered BLPs of the random effects: \( \{ \tilde{u}_i^\tau \} \);
Use data \( \{(Y_{ij}, X_{ij})_{i=1}^{N_i} : i = 1, \ldots, N\} \), where \( \tilde{Y}_{ij} = Y_{ij} - Z_{ij}^T \tilde{u}_i^\tau \) and get the (unadjusted) estimate, \( \hat{\beta}_{\text{two-step}}^\tau \), and its estimated standard error, \( \text{SE}_{\text{obs}} \);
For all \(i, j\) compute residuals as \( \varepsilon_{ij} = Y_{ij} - X_{ij}^T \hat{\beta}_{\text{two-step}}^\tau - Z_{ij}^T \tilde{u}_i^\tau \);

forall \( b = 1 : B \) do
\begin{itemize}
  \item Draw weights \( w_{ij} \) from the weight distribution \( (3.5) \);
  \item Use wild bootstrap on \( \varepsilon_{ij} : \varepsilon_{ij}^w = w_{ij} \varepsilon_{ij} \);
  \item Resample \( \tilde{u}_{i}^\tau \) with replacement to get \( \tilde{u}_{i}^{\tau,w} \);
  \item Construct the bootstrap sample: \( \{(Y_{ij}^{w,b}, X_{ij}, Z_{ij})_{i=1}^{N_i} : i = 1, \ldots, N\} \)
where \( Y_{ij}^{w,b} = Z_{ij}^T \tilde{u}_{i}^{\tau,w} + X_{ij}^T \hat{\beta}^{\tau,w}_{\text{two-step}} + \varepsilon_{ij}^{w} \);
  \item Use data \( \{(\tilde{Y}_{ij}^{w,b}, X_{ij})_{i=1}^{N_i} : i = 1, \ldots, N\} \), where \( \tilde{Y}_{ij}^{w,b} = Y_{ij}^{w,b} - Z_{ij}^T \tilde{u}_{i}^{\tau,w} \) and standard linear quantile regression estimation to get \( \hat{\beta}^{\tau,w}_{\text{oracle},b} \);
  \item Use data \( \{(Y_{ij}^{w,b}, X_{ij}, Z_{ij})_{i=1}^{N_i} : i = 1, \ldots, N\} \) and the proposed two-step estimation to get \( \hat{\beta}^{\tau,w}_{\text{two-step,b}} \);
\end{itemize}
Compute the two-step bootstrap mean, \( \bar{\beta}^{\tau,w}_{\text{two-step}} \);
For each component \( k = 1, \ldots, p \), calculate the standard deviation for the two-step and oracle estimators, \( \text{SD}_{\text{two-step},k} \) and \( \text{SD}_{\text{oracle},k} \), respectively;
For specified \( \alpha \), for each component \( k = 1, \ldots, p \) in part calculate:
\begin{itemize}
  \item \( 100(1-\alpha)\% \) basic confidence interval: \( 2\hat{\beta}_{\text{two-step},k}^{\tau} - \hat{\beta}_{1-\alpha/2,k}^{\tau} ; 2\hat{\beta}_{\text{two-step},k}^{\tau} - \hat{\beta}_{\alpha/2,k}^{\tau} \)
  \item \( 100(1-\alpha)\% \) SE adjusted confidence interval: \( \hat{\beta}_{\text{adj},k}^{\tau} \pm q_{1-\alpha/2} \cdot \text{SE}_{\text{adj},k} \), where
\end{itemize}
\[
\text{SE}_{\text{adj,k}} = \frac{\text{SD}_{\text{two-step,k}}}{\text{SE}_{\text{obs},k}} \frac{\text{SE}_{\text{oracle,k}}}{\text{SE}_{\text{obs},k}}
\]
\textbf{Algorithm 1:} Pseudo code for implementation of the bootstrap adjusted two-step estimator and related confidence intervals.

where \( \beta_0^\tau = \beta_0 + \sigma_u \Phi^{-1}(\tau) \) and \( \beta_1^\tau = \beta_1 + \gamma \sigma_u \Phi^{-1}(\tau) \), with \( \Phi \) denoting the cumulative distribution function for the \( N(0,1) \) distribution. In particular, the quantiles are of the same form as \( (2.2) \), with \( X_{ij} = (1, x_{ij}) \) and \( Z_{ij} = 1 \), and with \( u_i^\tau \) not depending on \( \tau \). When \( \gamma = 0 \) the slope parameter of the quantile is constant across \( \tau \), i.e., \( \beta_1^\tau = \beta_1 \), while the covariate effect differs between quantile levels when \( \gamma \neq 0 \). Irrespective of the choice of \( \gamma \), the regression parameter for the median, \( \beta_1^{0.5} \), does not depend on \( \gamma \), since \( \Phi^{-1}(0.5) = 0 \).

Notice that the data generating model implies that the marginal-type quantile at level \( \tau \) of \( Y_{ij} \) given \( x_{ij} \) (but not conditional on \( u_i \)) is given by
\[
\beta_0 + \beta_1 x_{ij} + \Phi^{-1}(\tau) \sqrt{\sigma_u^2 + (1 + \gamma x_{ij})^2 \sigma_e^2}.
\]

In the heteroscedastic setting \( (\gamma > 0) \) this expression is not linear in \( x_{ij} \), in contrast with \( (4.2) \), and a linear approximation has parameters that are different from \( \beta_0^\tau \) and \( \beta_1^\tau \). This shows that a marginal estimation approach aims at different parameters compared to those in \( (4.2) \).

We are going to compare our proposed estimators to the marginal estimator and the other estimation methods discussed in Section \( \S \). To implement the approaches we use the function \texttt{rq()} of the \texttt{quantreg} package [Koenker 2020] to perform standard quantile regression and the \texttt{lqmm()} function of the package \texttt{lqmm} [Geraci 2014] to perform LQMM. More specifically, we use Gauss-Hermite quadrature (option \texttt{lqmmType="normal"} in \texttt{lqmm})
with 15 quadrature points \((nK=15)\) and derivative-free optimisation \((\text{lqmmMethod}="\text{df}"\)). Quantile regression with \(\ell_1\) and \(\ell_2\) penalization and cross validation for selection of the penalty parameter is implemented in the function \(\text{cv.hqreg()}\) of the \text{hqreg} package \((\text{Yi} 2017)\). We use five-fold cross validation. Finally, we use \(B=100\) bootstrap replications for bias-adjustment, where applicable.

### 4.2 Comparison of estimation methods

#### Overall comparison for a benchmark scenario

In the model \((4.2)\), we consider true (mean) parameters \(\beta_0 = \beta_1 = 1\), homoscedasticity departure parameter \(\gamma = 0.4\), variances \(\sigma_u^2 = \sigma_e^2 = 1\), and thus \(\text{ICC} = 0.5\). The main focus is on the quantile level \(\tau = 0.1\) that is somewhat extreme; then true parameter values amount to \(\beta_0^* = -0.281\) and \(\beta_1^* = 0.487\). Define the “benchmark scenario” by the case with \(N = 500\) clusters of size \(n_i = 6\) \((i = 1, ..., N)\); we use this scenario to study the performance of the estimators in the situation with \(N \gg n_i\).

Figure 2 shows the boxplots of the bias for \(\beta_0^*\) (left) and \(\beta_1^*\) (right) corresponding to quantile levels \(\tau = 0.5\) (top) and \(\tau = 0.1\) (bottom), based on 200 Monte Carlo simulations. We compare the proposed two-step estimator and its adjusted version \((\text{twostep and adj}, \text{respectively})\), the estimator from \[\text{Canay (2011)}\ (canay), the LQMM estimator \((\text{lqmm})\) and its jackknife-based adjustment \((\text{jackknife})\), the estimators arising from penalized quantile regression, both with \(\ell_1\) and \(\ell_2\) penalties \((\text{l1pen and l2pen}, \text{respectively})\), the marginal estimator arising from standard quantile regression \((\text{marg})\), and the estimator from \((2.4)\) where the actual random effects are used in the computations \((\text{oracle})\). The oracle estimator is unfeasible in practice, but is used as a reference to study the effect of random effects being latent.

All nine estimators have similar distributions for \(\tau = 0.5\), except the jackknife-adjusted estimator, which has slightly larger variation for both parameters. The results are more interesting for \(\tau = 0.1\). Focusing first on the methods developed in this paper, the unadjusted two-step estimator has a smaller bias (component-wise) than the other estimators studied; yet, there is still some bias left compared to the oracle estimator. The bias-adjusted estimator, on the other hand, has a very small bias (for each component) and variance that is slightly larger than that of the oracle estimator, but comparable to the other competitors.

The estimator proposed by \[\text{Canay (2011)}\ has a comparable bias to the other estimators when it comes to the slope, but it shows positive (but small) bias for the intercept. The variance is small for both components of the quantile regression parameters. Results for the LQMM estimators and the estimators from \[\text{Koenker (2004)}\ based on \(\ell_1\) penalisation are similar and show a small bias for both components. The estimator based on \(\ell_2\) penalisation has the same properties for the slope, but has a larger bias for the intercept. The jackknife-based adjustment of the LQMM estimator reduces the bias for the slope parameter, but not for the intercept, and generally, it has large variation.

As expected, the standard quantile regression estimator, which completely ignores the cluster structure, leads to increased bias. The bias is particularly severe for the intercept, whereas the bias for the slope is comparable to that of Canay’s estimator, the LQMM estimator, and the penalization-based estimators. This is interesting, as it indicates that these latter estimators effectively estimate the slope coefficient in (a linearized version of) a marginal quantile model rather than in the conditional quantile model.

Additional simulation results are included in the appendix; Tables \[4\] show results for settings where \((N, n_i)\) differ from the benchmark scenario, and for quantile levels \(\tau = 0.1, 0.5\). The conclusions from Figure 2 are confirmed; in particular an advantage of the proposed estimators is observed for \(\tau = 0.1\) (Tables \[5\] and \[7\]). In passing, we note that the \(\ell_1\)-penalized estimator is preferable to the \(\ell_2\)-penalized estimator in all settings, and
Figure 2: Bias for different estimators of $\beta^0_\tau$ (left) and $\beta^1_\tau$ (right) for 200 datasets from the benchmark scenario. The quantile level is 0.5 (top) and 0.1 (bottom). The true parameter values are $\beta^0_0 = \beta^0_1 = 1$ and $\beta^0_0 = -0.281$, $\beta^0_1 = 0.487$, respectively.

that the jackknife estimator reduces bias for $\hat{\beta}^1_1$ but increases bias for $\hat{\beta}^0_0$ and has larger variance. For those reasons we do not study the $\ell_2$-penalized and the jackknife estimators any further. The remaining estimators are discussed in more detail in the next section.

The average computing time per simulated dataset for the bootstrap-adjusted two-step estimator was 18.83 seconds. By comparison, the computation time for the LQMM estimator was 0.15 seconds. The difference reflects the additional $B = 100$ iterations involving LQMM estimation and the construction of the confidence intervals that are required by the proposed method. The average computation time for Canay’s estimator was 0.58 seconds. The average computation time for the $\ell_1$-penalized estimator was 72.29 seconds, partly due to the cross-validation step. The computation time for the $\ell_2$-penalized estimator was close to that of the $\ell_1$-penalized, and computations for the jackknife adjusted estimator took about three times longer than computations for LQMM. Computations were run on a commodity PC with 2.9 GHz Dual–Core Intel Core i5 processor 5287U.
Bias for LQMM, $\ell_1$-penalized, and Canay’s estimator for extreme quantile levels

For quantile level 0.1, the bias of the LQMM, $\ell_1$-penalized, $\ell_2$-penalized and Canay’s estimators in the bottom of Figure 2 is quite large. This flaw is reported for Canay’s estimator in a simulation study with $N$ much larger than $n_i$ and varying quantile levels (Canay, 2011); however, to the best of our knowledge, the bias has not been documented thoroughly in the literature for the other estimators. The $\ell_1$-penalized estimation is carried out in Koenker (2004) for a simulation model similar to ours, but only for the median ($\tau = 0.5$) where all estimators are unbiased. LQMM estimation is analyzed in Geraci and Bottai (2014) in many simulation scenarios with good overall performance, but the dependence on bias of sample size ($N$ and $n_i$) is not studied in the presence of heteroscedasticity.

Figure 3 shows boxplots of the bias for the LQMM, the $\ell_1$-penalized, and Canay’s estimator for various number of clusters, $N$, cluster sizes, $n_i$, and at different quantile levels, $\tau$; results are based on 200 replications. We vary one factor at a time, while keeping the others fixed at their benchmark values ($N = 500$, $n_i = 6$, $\tau = 0.1$). As a consequence, the benchmark scenario appears in each panel. The top plots show the results for the intercept, while the bottom row shows results for the slope.

Generally, the magnitude of the bias decreases as the number of observations per cluster increases for fixed $N$ (central panels): this confirms the existing asymptotic results (Koenker, 2004; Canay, 2011). However, when the cluster size, $n_i$, is fixed (left most panels), there is non-negligible bias for these estimators, as the sample size, $N$, increases. The results are valid for both parameter components, but in particular for the slope (bottom panel). In other words, the estimators are not consistent for $\beta_{\tau 1}$ in the asymptotic scenario with a fixed (and small) number of repeated measurements and increasing the number of clusters. The bias behavior is worse for quantile levels closer to the boundaries, $\tau = 0.1$ or $\tau = 0.9$, than for levels closer to the median, $\tau = 0.5$ (right panels).

The three methods are comparable for estimation of $\beta_{\tau 1}$ whereas there are subtle differences for $\beta_{\tau 0}$: LQMM and $\ell_1$-penalized estimators behave similarly, except for small values of $N$; Canay’s estimator has bias of opposite sign and of smaller size as well as smaller variation compared to the two other methods. Further simulation scenarios are presented in Tables 4 and 5 in the appendix, showing similar results.

4.3 Performance of the proposed estimators

Bias and variation

We now turn to a more detailed study of our proposed estimators. Figure 4 has the same structure as Figure 3 but now includes the oracle estimator (as an infeasible point of reference), the LQMM estimator (as a representative of the existing methods, cf. Figure 3 and as starting point of our two-step procedure), and the unadjusted and adjusted two-step estimators. Results are based on 1000 replications. The benchmark scenario ($N = 500$, $n_i = 6$, $\tau = 0.1$) was also considered in Figure 2 but notice that the results of Figure 4 summarize performance in 1000 simulations, while only 200 simulations were considered in Figure 2 due to the increased computational burden required by some of the alternative methods.

For the slope quantile regression parameter, $\beta_{\tau 1}$ (bottom panels), the bias is reduced for the two-step estimator compared to the LQMM estimator and is almost completely removed in all scenarios for the bias-adjusted estimator. The variability is only slightly larger than the variability of the oracle estimator. For the intercept quantile regression parameter, $\beta_{\tau 0}$, the bias is considerably reduced for the proposed two-step estimators compared to the LQMM estimator when the cluster size is small (top left panel). For large clusters
the unadjusted two-step estimator shows the best performance in terms of both bias and variance (top central panel).

Results for more combinations of $N$, $n_i$, and $\tau$ are reported in the appendix. For the median, $\tau = 0.5$ (Table 6), all three estimators are unbiased and show similar variability. For $\tau = 0.1$ (Table 7), the situation is more complex. Nonetheless, the proposed two-step estimators (without adjustment) yields a smaller RMSE than the LQMM counterpart. Consider the estimation of the slope parameter $\beta_1$: all estimators seem to show similar variability, however the two-step estimators indicate a considerably improved bias behavior compared to the LQMM estimator. The numerical studies show that the cluster size has a larger impact on estimation performance than the number of clusters; compare the RMSE when the number of observations is kept fixed to say 3000 composed by 1) $N = 1000$ clusters of size $n_i = 3$ and 2) $N = 500$ clusters of size $n_i = 6$.

Figure 5 compares the two-step estimators with the oracle and LQMM for three extra
Figure 4: Boxplots for estimators of $\beta_0^\tau$ (top) and $\beta_1^\tau$ (bottom) for 1000 datasets from the oracle estimator, the LQMM estimator and the adjusted two-step estimator with varying $N$ (left panels), $n_i$ (middle panels) and $\tau$ (right panels). Parameters that do not vary are kept fixed at benchmark values: $N = 500$, $n_i = 6$, $\tau = 0.1$. Scenarios that have larger heteroscedasticity ($\gamma = 1$) or larger within-cluster relative variance ($\sigma_u^2 = 1.5$, $\sigma_e^2 = 0.5$ yielding ICC = 0.75), or larger total variation ($\sigma_u^2 = \sigma_e^2 = 1.5$) compared to the benchmark scenario. All other simulation parameters are kept fixed to the values from the benchmark setting. The changed parameter settings have larger impact on the distribution of the LQMM estimator than on the distribution of the two-step estimators. In particular, the two-step estimation results in improved bias performance compared to the LQMM estimator, irrespective of the setting.

Confidence intervals and comparison of bootstrap strategies

Next, we turn to evaluating the proposed bootstrap scheme for decreasing the estimator’s bias and construction of confidence intervals. We compare the proposed mixture of standard
Figure 5: Boxplots of the estimates of $\beta^0_\tau$ (left) and $\beta^1_\tau$ (right) obtained using oracle method, LQMM, and the two-step estimators with and without adjustment for two-step estimation for the benchmark scenario and scenarios with larger homoscedasticity, larger ICC, and larger variance. All the other simulation factors are kept constant to their values of the benchmark scenario. Results are based on 200 simulations.

and wild resampling (denoted by RW) with other types of data resampling, with respect to bias-adjustment in estimating the parameters, as well as the actual coverage and average length of the confidence intervals.

Resample random effects and residuals (RRR) A bootstrap sample takes the form
$\{(Y_{ij}^{*eb}, X_{ij}, Z_{ij})_{j=1}^{n_i}, u_{i}^{\tau, eb}\}_{i=1}^{N}$ where $Y_{ij}^{*eb} = Z_{ij}^{T}u_{i}^{\tau, eb} + X_{ij}^{T}\hat{\beta}_{\text{two-step}}^{\tau} + \varepsilon_{ij}^{*eb}$, with $\varepsilon_{ij}^{*eb}$ obtained from a standard sampling with replacement procedure from the observed residuals, $\{\varepsilon_{ij}\}_{i,j}$, and $u_{i}^{\tau, eb}$ is sampled from $\mathcal{U}$. In contrast to RW sampling, there is no coupling between covariates and residuals. Carpenter et al. (2003) has proposed the method for mean regression for multilevel data. Notice that residuals could also be sampled cluster-wise in order to maintain within-cluster dependence not accounted for by the random effect, but we do not consider this.

Resample clusters (RC) The clusters are sampled with replacement in a completely non-parametric way. More specifically, $i_1^*, \ldots, i_N^*$ are sampled with replacement from $\{1, \ldots, N\}$, and a bootstrap dataset consists of $(Y_{i_1^*j}^*, X_{i_1^*j}, Z_{i_1^*j}) = (Y_{i_1^*j}, X_{i_1^*j}, Z_{i_1^*j})$, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$. Within-cluster dependence is maintained because complete clusters are sampled. The method, also known in the literature as cross-sectional resampling (Galvao and Montes-Rojas, 2015), is used by Canay (2011) and Geraci and Bottai (2014) to construct confidence intervals. Karlsson (2009) uses RC in an attempt to correct for estimation bias in a nonlinear quantile regression for longitudinal data, using a marginal perspective, but experienced limited gain.

Cluster-wise wild bootstrap (CW) The idea is to use wild bootstrap for the sum of random effects and error terms. Specifically, let $r_{ij} = Y_{ij} - X_{ij}^{T}\hat{\beta}_{\text{two-step}}^{\tau}$ be the residuals corresponding to the two-step estimation, and let $w_{ij}$ be a random sample from (3.5). The bootstrap sample is $\{(Y_{ij}^{*eb}, X_{ij}, Z_{ij})_{j=1}^{n_i}\}_{i=1}^{N}$, where $Y_{ij}^{*eb} = X_{ij}^{T}\hat{\beta}_{\text{two-step}}^{\tau} + w_{ij}r_{ij}$. In contrast to the residuals $\varepsilon_{ij}$ used for RW, $r_{ij}$ are defined without subtraction of
Table 1: Bias and coverage rates of 95% confidence intervals for the adjusted two-step method for different bootstrap schemes (RW, CW, RC, RRR) for 1000 datasets. Basic confidence intervals are used for all bootstrap schemes, whereas SE-adjusted confidence intervals are only defined for RW and RRR. Cluster size is fixed at $n_i = 6$ and the quantile level is $\tau = 0.1$.

<table>
<thead>
<tr>
<th>N</th>
<th>Bias</th>
<th>Coverage, basic</th>
<th>Coverage, SE-adj.</th>
<th>Bias</th>
<th>Coverage, basic</th>
<th>Coverage, SE-adj.</th>
<th>Bias</th>
<th>Coverage, basic</th>
<th>Coverage, SE-adj.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-0.01 -0.03</td>
<td>0.87 0.90 0.88</td>
<td>0.10</td>
<td>0.01 0.05 0.05</td>
<td>-0.03</td>
<td></td>
<td>-0.02 -0.03</td>
<td>0.89 0.90 0.90</td>
<td>0.02 0.06 0.06</td>
</tr>
<tr>
<td>100</td>
<td>-0.02 -0.03</td>
<td>0.95 0.93</td>
<td></td>
<td>0.95 0.96</td>
<td></td>
<td></td>
<td>-0.02 -0.03</td>
<td>0.94 0.95</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>-0.02 -0.03</td>
<td>0.96 0.91</td>
<td>0.91</td>
<td>0.93 0.92</td>
<td></td>
<td></td>
<td>-0.03 -0.02</td>
<td>0.89 0.86 0.88</td>
<td>0.05 0.06 0.07</td>
</tr>
<tr>
<td>1000</td>
<td>-0.03 -0.02</td>
<td>0.95 0.89</td>
<td>0.86</td>
<td>0.93 0.90</td>
<td></td>
<td></td>
<td>-0.03 -0.02</td>
<td>0.89 0.88 0.89</td>
<td>0.31</td>
</tr>
</tbody>
</table>

predicted random effects (often referred to as “level zero residuals”). Also, same weight $w_i$ is used for all the observations within cluster $i$ in order to preserve dependence within clusters. This resampling scheme is used by Modugno and Giannerini (2015) in the context of multilevel models for mean regression, but does not appear to have been used for quantile regression.

The RW and RRR sampling schemes use bootstrap to approximate the joint distribution of $(u^*_i, Y_{ij})$, whereas the other two bootstrap methods approximate the distribution of $Y_{ij}$ only. As RC- and CW-based approaches do not involve generation of random effects, SE-adjustment confidence intervals are only applicable for RW and RRR. The bias-adjusted estimator and basic confidence intervals, on the other hand, can be computed for any of the four bootstrap schemes.

Table 1 shows bias and actual coverage rates for confidence intervals with an intended level of 95%. We employ the benchmark scenario, except for a varying number of clusters (same simulated data as in the left part of Figure 4). Results are based on 1000 simulated datasets. SE-adjustment confidence intervals generated with the RW bootstrap method give the best coverage rates, close to the nominal 95% in all scenarios. Basic confidence intervals with RW bootstrap are also good for $\beta_0^*$ when $N$ is large, whereas coverage rates are below 0.90 for $\beta_1^*$. RRR and RW produce similar coverage rates for $\beta_1^*$, but no bias-adjustment is obtained with RRR (bias is equivalent to bias for the unadjusted two-step estimator, not reported). For $\beta_0^*$ the coverage rates are slightly smaller for RRR compared to RW. As expected, bootstrap method RC gives no bias reduction, neither for $\beta_0^*$ nor $\beta_1^*$, and coverage rates are consequently never above 0.90. The CW bootstrap method has poor coverage rates. For $\beta_1^*$ the main reason is that the adjusted estimator has large variability which is not properly taken into account, whereas the explanation for $\beta_0^*$ is that CW introduces a large bias such that the confidence interval is located far from the true value.

In summary, the semi-parametric bootstrap sampling methods using the additive model structure for the quantiles (RW and RRR) with SE-adjusted confidence intervals show the best coverage properties. Nonetheless, the proposed two-step with RW-based adjustment results in the greatest bias reduction.
Table 2: Bias, coverage rates of 95% confidence intervals and average length of confidence intervals for our adjusted two-step method as well as LQMM and Canay’s methods for 1000 datasets. Cluster size is fixed at \( n_i = 6 \) and the quantile level is \( \tau = 0.1 \).

### 4.4 Additional simulation studies

At the suggestion of an anonymous reviewer, we further investigate the proposed method when the errors \( e_{ij} \) are generated from a non-Gaussian distributions. Specifically, we use a scaled \( t_3 \)-distribution and an \( \text{ALD}(0, \sigma_0, \tau_0) \) with \( \tau_0 = 0.1 \) and \( \sigma_0 = \frac{(1-\tau_0)\tau_0}{\sqrt{1-2\tau_0+2\tau_0}} = 0.09939 \).

Both distributions are scaled to have unit variance in order to make fair the comparison with the standard normal errors scenarios considered previously. When sampling from the ALD distribution, we consider both the benchmark scenario and a departure from it, corresponding to \( \gamma = 0 \). Notice that the true values of \( \beta_{\tau 0} \) and \( \beta_{\tau 1} \) change compared to the standard normal case. The results are shown in Table 8 in the appendix and should be compared to the relevant scenarios in Table 7.

In the case of scaled \( t \)-distributed errors, the bias is reduced for the two-step estimator, compared to the LQMM estimator, but it is not completely removed. The RW bootstrap correction reduces the bias even further for \( \beta_{\tau 1} \), but surprisingly it increases the bias for \( \beta_{\tau 0} \). This may be due to the inflated residuals that are obtained with the wild bootstrap scheme, as they can be large in the situation of heavy-tailed errors, and therefore have large impact on the estimation of bias for the intercept.

In the case of heteroscedastic ALD errors (\( \gamma > 0 \)), the bias of the LQMM estimator for \( \beta_{\tau 0} \) is reduced considerably compared to the Gaussian case (Table 7). The estimators’ variability is also reduced in this setting, in spite of the error variance remaining fixed, because quantiles are generally estimated with higher precision when the model is ALD.
than when it is Gaussian. The two-step estimator and the adjusted two-step estimator have almost the same distributions as the LQMM estimator. For estimating the intercept, the performance of the proposed estimators is superior to that of the LQMM, in terms of reduced bias and variability.

When the errors come from a homoscedastic ALD ($\gamma = 0$), the working distribution for the LQMM estimation approach coincides with the data generating mechanism. As expected, the LQMM estimator of $\beta_1$ has a very good performance: no bias and small variance. The two-step estimators are also unbiased, but have slightly larger variance. For estimating the intercept parameter, surprisingly, the LQMM estimator shows a behavior comparable to the heteroscedastic ALD case; in contrast the two-step estimators have a much smaller bias and variance.

Finally, we also consider a quantile regression model involving both a random intercept and a random slope. To be specific, the data are generated from the model $Y_{ij} = \beta_0 + u_i + (\beta_1 + v_i)x_{ij} + (1 + \gamma x_{ij})e_{ij}$, where $u_i$ is generated as described in (4.1) and $v_i \overset{iid}{\sim} N(0, \sigma^2_v)$. Out of the existing methods, only LQMM allows to incorporate random slopes in the quantile regression; thus we compare the results of the two-step estimation with LQMM solely. Table 3 shows the results. We see that irrespective of the sample size or cluster size, the two-step estimation without adjustment improves or maintains the RMSE compared to LQMM estimation. The adjusted two-step estimator generally shows the smallest bias, but at the expense of increased variability; for the estimation of the intercept parameter in the case of $n_i = 12$ the unadjusted two-step estimator has the smallest bias and variance.

<table>
<thead>
<tr>
<th>N</th>
<th>$n_i$</th>
<th>lqmm Bias</th>
<th>two-step Bias</th>
<th>adj (RW) Bias</th>
<th>lqmm SD</th>
<th>two-step SD</th>
<th>adj (RW) SD</th>
<th>lqmm RMSE</th>
<th>two-step RMSE</th>
<th>adj (RW) RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>6</td>
<td>-0.03</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.15</td>
<td>0.14</td>
<td>0.04</td>
<td>0.14</td>
<td>0.09</td>
<td>0.17</td>
</tr>
<tr>
<td>1000</td>
<td>6</td>
<td>0.13</td>
<td>0.09</td>
<td>0.12</td>
<td>0.18</td>
<td>0.17</td>
<td>0.24</td>
<td>0.08</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>500</td>
<td>12</td>
<td>-0.07</td>
<td>-0.02</td>
<td>-0.05</td>
<td>0.06</td>
<td>0.10</td>
<td>0.04</td>
<td>0.12</td>
<td>0.07</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Table 3: Bias, standard deviation, and RMSE for the LQMM estimator (lqmm), the two-step estimator (two-step), and bootstrap-adjusted two-step estimator (adj) where bootstrap samples are generated with the RW method, and we consider the model with random intercept as well as random slope. The quantile level is $\tau = 0.1$, and results are from 200 replications.

5 Data application

AIDS Clinical Trial Group (ACTG) Study 193A ([Henry et al., 1998](#)) is a randomized and double-blinded study of patients affected by AIDS at severe immune suppression stage, with CD4 counts of less than 50 cells/mm$^3$. There are 1309 patients, who were assigned to one of four treatments, namely: 600 mg of zidovudine daily alternating monthly with 400 mg of didanosine (double treatment 1); 600 mg of zidovudine as well as 2.25 mg of zalcitabine, both daily (double treatment 2); 600 mg of zidovudine as well as 400 mg of didanosine, both daily (double treatment 3); the combination of 600 mg of zidovudine, 400 mg of didanosine and 400 mg of nevirapine, all of them daily (triple treatment). The CD4 counts were recorded at a baseline visit and at the follow-up visits during the subsequent
40 weeks. The measurements were intended to be taken every eight weeks, but occasionally there were dropouts or skipped medical appointments; see Figure 6. After excluding the subjects with a single measurement (baseline), there are $N = 1187$ subjects remaining in the study; their number of repeated measurements, $n_i$, varies between two and nine with a median of four. The data has been previously used as an illustrative application for mean regression frameworks in Fitzmaurice et al. (2012) and it is available at the associated webpage (https://content.sph.harvard.edu/fitzmaur/ala2e/).

![Figure 6: Transformed CD4 counts for 200 patients, showing the records of 50 random subjects from each of the four treatment groups. Observations from the same patients are connected with lines.](image)

Our aim is to study the progression of the infection under the four treatment regimes for patients at different stages of immune suppression. Since CD4 counts are proxies for the stage of suppression—with lower CD4 counts corresponding to later stages—this can be obtained by studying the time trend for each treatment at different quantile levels. More specifically, an effective treatment reduces the decrease in CD4 counts, yielding a time trend closer to zero than a less effective treatment, and the effect may be different for early-stage
patients (corresponding to high quantile levels) than late-state patients (corresponding to low quantile levels). Figure 6 shows that subjects tend to have low or high CD4 counts throughout, suggesting incorporation of subject-specific intercepts in the model.

As it is common in the literature, we log-transform the observed values and denote by \( Y_{ij} \) the \( \log(\text{CD4 count} + 1) \) for patient \( i \) at the \( j \)th hospital visit and by \( t_{ij} \) the time of the \( j \)th visit, which is recorded by the number of weeks since the patient’s baseline visit. We use dummy variables \( \text{Treat}_h \) \((h = 1, \ldots, 4)\) to indicate the assigned treatment, where \( \text{Treat}_1 \) corresponds to the triple therapy, and \( \text{Treat}_2, \text{Treat}_3 \) and \( \text{Treat}_4 \) correspond to the three double treatments. We account for age at baseline (variable \( \text{Age} \)) and sex (variable \( \text{Sex} \), zero for females and one for males) as well. For simplicity of notation, we collect covariates relative to the \( i \)th patient at the \( j \)th follow-up visits into \( X_{ij} \) such that \( X_{ij}^T = (\text{Treat}_{1,i}, \text{Treat}_{2,i}, \text{Treat}_{3,i}, \text{Treat}_{4,i}, \text{Age}_i, \text{Sex}_i, t_{ij}) \). To study the time-varying effect of treatment at quantile level \( \tau \) of the response, let \( u_{ij}^\tau \) be a subject-specific random effect associated with the quantile level \( \tau \) and posit the following linear quantile regression model:

\[
Q_{Y_{ij}|X_{ij},u_{ij}^\tau}(\tau) = \sum_{h=1}^{4} \beta_{1,h}^\tau \cdot \text{Treat}_{h,i} + \sum_{h=1}^{4} \beta_{2,h}^\tau \cdot \text{Treat}_{h,i} \cdot t_{ij} + \beta_2^\tau \cdot \text{Age}_i + \beta_3^\tau \cdot \text{Sex}_i + u_{ij}^\tau. \tag{5.1}
\]

The slope parameters \( \beta_{1,1}^\tau, \ldots, \beta_{1,4}^\tau \) describe the behavior of CD4 counts over time, conditional on subject, and represent the main object of interest. As our interest is in the time varying effect of each treatment we are using the so-called “explicit parameterization”; as a result, the model specification does not require a common intercept parameter. Estimation and inference are carried out using the proposed two-step estimation with adjustment; the results are compared with LQMM.

The estimated slope parameters for each treatment in part are plotted in Figure 7 for varying quantile levels. The left panels show the two-step estimates with adjustment and the corresponding 95% confidence intervals for quantile levels \( \tau \in \{0.1, 0.15, \ldots, 0.9\} \) (separate analyses). We used 100 RW bootstrap samples for the computations. The top panels concern the triple treatment: since the confidence band, corresponding to the two-step estimator, includes zero at all the quantile levels, it indicates that this therapy maintains an almost constant CD4 count during the study for subjects at any stage of their condition. For the other three treatments the situation is different. As depicted in the remaining panels, the two-step estimated coefficients \( \hat{\beta}_{1,2}^\tau, \hat{\beta}_{1,3}^\tau \) and \( \hat{\beta}_{1,4}^\tau \) are negative and significant at all the quantile levels, indicating that patients treated with either one of the double therapies must expect to see their CD4 count decrease over time. Notice that there is a slight increase in the estimated \( \hat{\beta}_{1,2}^\tau \) over quantile levels, which indicates that double treatment 1 makes the CD4 counts decrease faster for patients in the most severe conditions (lower quantile levels), whereas double treatments 2 and 3 appear to have more homogenous effects across patient groups.

In order to compare the treatments more directly we consider contrasts of the form \( \hat{\beta}_{1,h}^\tau - \hat{\beta}_{1,1}^\tau \), which describe the difference in the effects between each double treatment and the triple treatment at quantile level \( \tau \). The middle panels in Figure 7 show the estimated contrasts and the corresponding 95% confidence intervals. Except for a single quantile level for double treatment 3, confidence intervals exclude zero, showing that the triple therapy is the most efficient treatment for patients in all infection stages. Fitzmaurice et al. (2012) reported similar results for the mean.

For comparison, the LQMM estimates and confidence intervals for the contrasts are shown in the right panels of Figure 7. Confidence intervals are based on 100 RC bootstrap samples. LQMM estimates are in the same range as the adjusted two-step estimates, albeit in general closer to zero. Moreover, the confidence bands are much wider, implying that the LQMM method does not find evidence for significant treatment differences for double treatments 2 and 3. This should not be surprising, since our numerical investigation showed
that LQMM confidence intervals are wider (and coverage lower) than those corresponding to the adjusted two-step estimator, when the number of subjects is much larger than the number of repeated measurements; recall Table 1.

Figure 7: Estimated coefficients and 95% confidence bands at varying $\tau$ for model (5.1). The left panels show results for slope coefficients $\beta^T_{1,h}$ ($h = 1, \ldots, 4$, adjusted two-step method) whereas the central and right panels show results for contrasts with triple therapy as reference (adjusted two-step method in the centre, LQMM to the right).

While these results are interesting, we acknowledge one aspect of the data that our analysis does not account for: missing data. Out of the 1187 patients in the study, only 795 of them have measurements past the 30th week since their baseline. Missing data is not uncommon in ACTG studies and previous quantile regression analyses with longitudinal data have approached the problem by incorporating weights into the estimating equations (Lipsitz et al. 1997), employing hierarchical Bayesian models (Huang and Chen 2016; Feng et al. 2011), or by considering a linear quantile mixed hidden Markov model with a missing data indicator (Marino et al. 2018). Incorporation of such methods falls beyond the scope of this paper, but could be an interesting avenue for future research.

6 Discussion

We have identified a gap in the literature concerning mixed effects models for quantile regression for clustered data: existing estimation methods may yield severely biased estimators for fixed effects parameters in situations with many, but small clusters. In this paper, we propose a new estimation method that relies on predicted random effects computed by using an LQMM working framework (in particular, at the quantile level of interest), standard quantile regression with offsets, and a bias-adjustment by means of a novel bootstrap
sampling technique. In the simulation study, the proposed estimator shows considerably smaller bias compared to the available competitors, especially in situations with small clusters. The RW adjustment appears to be particularly beneficial for estimating slope parameters, while the results are less clear for the intercept and could be studied further. The two-step estimation procedure may be seen as the onset in an iterative procedure alternating between estimation of the regression parameters for fixed random effects and prediction of random effects for fixed regression parameters. An ALD working model with random effects only (no fixed effects) can be used in the second step, and this requires minor modifications of the current implementation of the \texttt{lqmm()} function.

Hitherto, the literature for quantile regression for clustered data has focused on studying asymptotics for increasing both the number of clusters and the cluster size (Koenker 2004; Kato et al. 2012; Canay 2011; Bessremyannaya and Golovan 2019). In such case, the cluster-specific parameters are asymptotically “eliminated” as stated by Canay (2011) or “concentrated out” as stated by Kato et al. (2012) and act as known quantities for the asymptotics of $\beta^\tau$. On the other hand, the theoretical study of the estimators is inherently challenging, when cluster size is fixed, and only the number of clusters increases to infinity. Results from (generalized) linear mixed models do not carry over for primarily two reasons. First, the criterion functions constructed from the check function is not differentiable. Second, the distributional assumptions are typically held to the minimum and focus on the relationship between the covariates and the quantile of interest. In particular, Geraci and Bottai (2007, 2014) do not mention any attempts to derive asymptotic results for the LQMM estimator and rely on bootstrap methods for inference. Neither do we provide asymptotic results for our estimators, nor claim that bias is removed asymptotically. The main difficulty lies in the prediction accuracy of the random effect predictors, which are used as one of the main ingredients in the bootstrap sampling procedure. If the predicted random effects do not accurately capture the variation of the cluster-specific random effects, then the estimated bias may not represent the bias of the unadjusted estimator. Therefore, when we are neither assuming an increasing cluster size nor considering a specific data generating model, then it is difficult to prove asymptotic results for our estimators, and we leave this for future research.

Mean regression models for longitudinal data often incorporate more complex within-subject dependence structures than the one modeled by random intercepts alone (compound symmetry). Similar attempts do not seem to exist for quantile regression. The two-step estimator is not readily modified to take a serial dependence into account, but the RW bootstrap sampling could be easily adapted such as by sampling the weights for wild bootstrap at the subject level rather than at the measurement level. Moreover, longitudinal studies may involve drop-outs and occasional missing data, with data not missing at random, and how to incorporate such missingness in quantile regression in an appropriate way remains an open research problem.

One direction that the proposed methodology opens up is to consider quantile regression for time series data (one long series rather than many shorter series), see Xiao (2017). In such case, the quantile model would be $Q_{Y_t | X_t}(\tau) = X_t^T \beta^\tau + u^\tau_t$ where $Y_t$ and $X_t$ denote the response and covariate, respectively, at time $t$ ($t = 1, \ldots, T$), and $\{u^\tau_t\}_{t=1,\ldots,T}$ is a latent series which describes (random) fluctuations of quantiles over time. Another direction is to extend the approach to multi-level data with multiple levels of nested random effects or data with several, but non-nested random effects. The ideas behind the methods from this paper (existing as well as our proposed method) would carry over to such situations, but a rigorous investigation of this extension is left for future research.
Acknowledgements

The project was partly funded by the Danish Research Council (DFF grant 7014-00221). We would like to thank the associate editor and two anonymous reviewers for suggestions and comments that improved the paper.

References


A Appendix

The appendix contains additional numerical results from the simulation study with data generated from model (4.1). The results are discussed in the main text. Tables 4 and 5 compare various existing approaches when both the number of clusters and the cluster size vary; other simulation parameters are specified by their level at the benchmark scenario. Estimation is carried out for quantile levels $\tau = 0.5$ (Table 4) and $\tau = 0.1$ (Table 5), respectively, with results based on 200 replications. It is not possible to compute the jackknife estimator when $n_i = 3$ because clusters cannot be split into two subsets with several observations per cluster. Furthermore, in the scenario with $N = 1000$, $n_i = 12$ and $\tau = 0.1$ there were convergence problems for the $\ell_1$-penalized estimator for two datasets, and the results for this estimator are based on the remaining 198 replications. Table 6 and Table 7 have the same structure as described above and consider the same scenarios; they evaluate the performance of the LQMM estimator and our two proposed methods in 1000 replications. Notice the difference in the number of replications; as mentioned in Section 4.2 it is due to the computational burden of some of the traditional estimators. Finally, Table 8 summarizes the results for the case when the error terms in (4.1) are either sampled from a scaled $t$-distribution in the benchmark scenario, from an ALD distribution in the benchmark scenario or an ALD distribution when $\gamma = 0$. Results correspond to the quantile level $\tau = 0.1$ and are based on 200 replications.
<table>
<thead>
<tr>
<th>N</th>
<th>n</th>
<th>Bias</th>
<th>Standard Deviation (SD)</th>
<th>RMSE</th>
<th>Bias</th>
<th>Standard Deviation (SD)</th>
<th>RMSE</th>
<th>Bias</th>
<th>Standard Deviation (SD)</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>3</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.07</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.07</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.07</td>
<td>0.07</td>
<td></td>
<td>0.01</td>
<td>0.07</td>
<td></td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.07</td>
<td>0.07</td>
<td></td>
<td>0.01</td>
<td>0.07</td>
<td></td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.07</td>
<td>0.07</td>
<td></td>
<td>0.01</td>
<td>0.07</td>
<td></td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>1000</td>
<td>6</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.05</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.05</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td></td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td></td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td></td>
<td>0.01</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 4: Bias, standard deviation, and RMSE for the oracle, Canay’s, the jackknife, the $\ell_1$-penalized, the $\ell_2$-penalized and the marginal estimators. The quantile level is $\tau = 0.5$, and results are based on 200 replications.
Table 5: Bias, standard deviation, and RMSE for the oracle, Canay’s, the jackknife, the \( \ell_1 \)-penalized, the \( \ell_2 \)-penalized and the marginal estimators. The quantile level is \( \tau = 0.1 \), and results are based on 200 replications.
Table 6: Bias, standard deviation, and RMSE for the LQMM estimator (lqmm), the two-step estimator (two-step), and bootstrap-adjusted two-step estimator (adj) where bootstrap samples are generated with the RW method. The quantile level is \( \tau = 0.5 \), and results are based on 1000 replications.

<table>
<thead>
<tr>
<th>N</th>
<th>( n_i )</th>
<th>lqmm</th>
<th>( \hat{\beta}_0 )</th>
<th>two-step</th>
<th>adj (RW)</th>
<th>lqmm</th>
<th>( \hat{\beta}_1 )</th>
<th>two-step</th>
<th>adj (RW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>3</td>
<td>Bias</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>Bias</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>RMSE</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td></td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>Bias</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>Bias</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
<td>RMSE</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
<td></td>
<td>0.10</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>500</td>
<td>6</td>
<td>Bias</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>Bias</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
<td>RMSE</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.07</td>
<td>0.06</td>
<td>0.07</td>
<td></td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>1000</td>
<td>6</td>
<td>Bias</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>Bias</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>RMSE</td>
<td>0.06</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.06</td>
<td>0.05</td>
<td>0.07</td>
<td></td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 7: Bias, standard deviation, and RMSE for the LQMM estimator (lqmm), the two-step estimator (two-step), and bootstrap-adjusted two-step estimator (adj) where bootstrap samples are generated with the RW method. The quantile level is \( \tau = 0.1 \), and results are based on 1000 replications.

<table>
<thead>
<tr>
<th>N</th>
<th>( n_i )</th>
<th>lqmm</th>
<th>( \hat{\beta}_0 )</th>
<th>two-step</th>
<th>adj (RW)</th>
<th>lqmm</th>
<th>( \hat{\beta}_1 )</th>
<th>two-step</th>
<th>adj (RW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>3</td>
<td>Bias</td>
<td>0.02</td>
<td>0.09</td>
<td>0.06</td>
<td>RMSE</td>
<td>0.16</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.16</td>
<td>0.11</td>
<td>0.14</td>
<td>RMSE</td>
<td>0.16</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.16</td>
<td>0.15</td>
<td>0.15</td>
<td></td>
<td>0.21</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>Bias</td>
<td>0.04</td>
<td>0.09</td>
<td>0.05</td>
<td>RMSE</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.11</td>
<td>0.08</td>
<td>0.10</td>
<td>RMSE</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
<td>0.30</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>500</td>
<td>6</td>
<td>Bias</td>
<td>-0.07</td>
<td>0.03</td>
<td>-0.02</td>
<td>RMSE</td>
<td>-0.07</td>
<td>0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.13</td>
<td>0.08</td>
<td>0.10</td>
<td>RMSE</td>
<td>0.15</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.15</td>
<td>0.09</td>
<td>0.10</td>
<td></td>
<td>0.21</td>
<td>0.15</td>
<td>0.16</td>
</tr>
<tr>
<td>1000</td>
<td>6</td>
<td>Bias</td>
<td>-0.05</td>
<td>0.03</td>
<td>-0.03</td>
<td>RMSE</td>
<td>-0.05</td>
<td>0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.10</td>
<td>0.06</td>
<td>0.07</td>
<td>RMSE</td>
<td>0.11</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.11</td>
<td>0.07</td>
<td>0.08</td>
<td></td>
<td>0.18</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>500</td>
<td>12</td>
<td>Bias</td>
<td>-0.06</td>
<td>0.01</td>
<td>-0.05</td>
<td>RMSE</td>
<td>-0.06</td>
<td>0.01</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.12</td>
<td>0.07</td>
<td>0.07</td>
<td>RMSE</td>
<td>0.13</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.13</td>
<td>0.07</td>
<td>0.09</td>
<td></td>
<td>0.12</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td>1000</td>
<td>12</td>
<td>Bias</td>
<td>-0.05</td>
<td>0.01</td>
<td>-0.05</td>
<td>RMSE</td>
<td>-0.05</td>
<td>0.01</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.09</td>
<td>0.05</td>
<td>0.05</td>
<td>RMSE</td>
<td>0.11</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.11</td>
<td>0.05</td>
<td>0.07</td>
<td></td>
<td>0.10</td>
<td>0.07</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Table 8: Bias, standard deviation, and RMSE for the LQMM estimator (lqmm), the two-step estimator (two-step), and bootstrap-adjusted two-step estimator (adj) where bootstrap samples are generated with the RW method. The residuals are sampled from a scaled $t_3$ when $\gamma = 0.4$ (top part), and from an ALD when either $\gamma = 0.4$ (central part) or $\gamma = 0$ (bottom part). The quantile level is $\tau = 0.1$, and results are based 200 replications.

<table>
<thead>
<tr>
<th>$e_{ij}$</th>
<th>$N$</th>
<th>$n_i$</th>
<th>lqmm $\beta_0$</th>
<th>two-step $\beta_0$</th>
<th>adj (RW) $\beta_0$</th>
<th>lqmm $\beta_1$</th>
<th>two-step $\beta_1$</th>
<th>adj (RW) $\beta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_3$</td>
<td>500</td>
<td>6</td>
<td>Bias -0.27</td>
<td>-0.04</td>
<td>-0.15</td>
<td>0.14</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.14</td>
<td>0.08</td>
<td>0.10</td>
<td>0.12</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.30</td>
<td>0.09</td>
<td>0.18</td>
<td>0.18</td>
<td>0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>$t_3$</td>
<td>1000</td>
<td>6</td>
<td>Bias -0.25</td>
<td>-0.05</td>
<td>-0.16</td>
<td>0.15</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.11</td>
<td>0.06</td>
<td>0.07</td>
<td>0.09</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.28</td>
<td>0.07</td>
<td>0.17</td>
<td>0.17</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>$t_3$</td>
<td>500</td>
<td>12</td>
<td>Bias -0.18</td>
<td>-0.05</td>
<td>-0.15</td>
<td>0.10</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.13</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.22</td>
<td>0.08</td>
<td>0.17</td>
<td>0.13</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td>ALD</td>
<td>500</td>
<td>6</td>
<td>Bias -0.12</td>
<td>-0.10</td>
<td>-0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.13</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.17</td>
<td>0.12</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>ALD</td>
<td>1000</td>
<td>6</td>
<td>Bias -0.08</td>
<td>-0.10</td>
<td>-0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.10</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.12</td>
<td>0.11</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>ALD</td>
<td>500</td>
<td>12</td>
<td>Bias -0.08</td>
<td>-0.07</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.15</td>
<td>0.09</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>ALD</td>
<td>500</td>
<td>6</td>
<td>Bias -0.13</td>
<td>-0.06</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\gamma = 0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.14</td>
<td>0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>ALD</td>
<td>1000</td>
<td>6</td>
<td>Bias -0.07</td>
<td>-0.07</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\gamma = 0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.13</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>ALD</td>
<td>500</td>
<td>12</td>
<td>Bias -0.07</td>
<td>-0.05</td>
<td>-0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\gamma = 0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.14</td>
<td>0.07</td>
<td>0.06</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>