Computational Statistics and Data Analysis 56 (2012) 2526–2538

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Quantile regression for longitudinal data with a working correlation model

Liya Fu [a,b,*], You-Gan Wang [b]

[a] School of Mathematics and Statistics, Xian Jiaotong University, China

[b] Centre for Applications in Natural Resource Mathematics, School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia

ARTICLE INFO

Article history:
Received 27 September 2011
Received in revised form 2 February 2012
Accepted 5 February 2012
Available online 11 February 2012

Keywords:
Covariance estimate
Unbiased estimating functions
Exchangeable correlation structure
Independence working model
Induced smoothing method

ABSTRACT

This paper proposes a linear quantile regression analysis method for longitudinal data that combines the between- and within-subject estimating functions, which incorporates the correlations between repeated measurements. Therefore, the proposed method results in more efficient parameter estimation relative to the estimating functions based on an independence working model. To reduce computational burdens, the induced smoothing method is introduced to obtain parameter estimates and their variances. Under some regularity conditions, the estimators derived by the induced smoothing method are consistent and have asymptotically normal distributions. A number of simulation studies are carried out to evaluate the performance of the proposed method. The results indicate that the efficiency gain for the proposed method is substantial especially when strong within correlations exist. Finally, a dataset from the audiology growth research is used to illustrate the proposed methodology.

1. Introduction

Longitudinal data are very common in biological and medical research. Correlation may arise when data are measured repeatedly from the same subject. Various methods have been developed to analyze the longitudinal data, which can be only used to evaluate covariate effects on the mean of a response variable (Liang and Zeger, 1986; Qu et al., 2000; Jung and Ying, 2003). To give a global assessment about covariate effects on the distribution of the response variable, a quantile regression model is an important alternative (Koenker and Bassett, 1978). Each quantile regression characterizes a particular point of a distribution, and thus provides more complete description of the distribution. Furthermore, quantile regression is more robust against outliers and does not require specifying any error distribution. Therefore, it has gained much attention in recent years and has become a more widely used technique to describe the response distribution (Chen et al., 2004; Koenker, 2005; Geraci and Bottai, 2007; Reich et al., 2010).

For independent measurements, parameter estimation and statistical inference procedures for quantile regression have been developed by Bassett and Koenker (1978) and Koenker and Bassett (1978). For dependent measurements, it is difficult to describe and specify the underlying correlation structure, which may prevent practitioners from using the quantile regression. Jung (1996) first extended the quantile regression to longitudinal data and developed a quasi-likelihood method for median regression. Koenker (2004) considered a linear quantile regression with subject specified fixed effects and made inferences from a penalized likelihood. Geraci and Bottai (2007) introduced random intercepts to account for the within-subject correlations and presented a likelihood-based approach by assuming the response variable following an asymmetric

* Correspondence to: School of Mathematics and Statistics, Xian Jiaotong University, 710049 Xian, Shanxi, China.
E-mail addresses: fuly737@gmail.com (L. Fu), you-gan.wang@uq.edu.au (Y.-G. Wang).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.02.005

Laplace density. Liu and Bottai (2009) generalized the model by Geraci and Bottai (2007) to linear mixed-effects models with random regression coefficients and assumed random effects by a multivariate Laplace distribution. Farcomeni (2012) studied linear quantile regression with time-varying random effects and modeled the random effects by a first order latent Markov chain. Chen et al. (2004) and Yin and Cai (2005) proposed using the generalized estimating equations based on the independence working model to estimate parameters in linear quantile regression, which are simple and have desirable properties but could lead to a loss of efficiency for parameter estimation when strong correlations exist. In linear quantile regression, parameter estimates can be obtained through linear programming techniques or iterative bisection methods (Koenker, 2004). The variances of parameter estimates typically depend on the error distribution which is usually unknown. Various resampling approaches have been developed to estimate the variances of the parameter estimates (Parzen et al., 1994; Buchinsky, 1995; Bilias et al., 2000). However, intensive resampling may add computational burdens, while with very small perturbing random variables, the estimating functions may not be jittered enough to obtain good estimates (Yin and Cai, 2005).

In this paper, we consider the linear quantile regression model by Chen et al. (2004) and propose a combination of the between- and within-subject estimating functions for parameter estimation, which take into account the correlations and variation of the repeated measurements for subjects. The proposed method is robust to the error correlation structure and improves the efficiency of parameter estimators. Furthermore, the induced smoothing method (Brown and Wang, 2005) is used to obtain an estimate of the asymptotic covariance matrix of parameter estimators, which has been used for univariate quantile regression models (Wang et al., 2009; Pang et al., 2010). The induced smoothing method is simple and eliminates computational issues resulting from the unsmoothed estimating functions and intensive resampling. Under certain moderate conditions, the asymptotic properties of the proposed method are derived.

The remainder of this paper is organized as follows: In Section 2, a linear quantile regression model and two types of unbiased estimating functions for longitudinal data are briefly described. Then a combination of the between- and within-subject estimating functions are proposed. In Section 3, the induced smoothing method is introduced, and asymptotic properties for the induced estimating functions and parameter estimators are derived. In Section 4, extensive simulation studies are carried out to demonstrate the performance of the proposed method. In Section 5, the audiology growth data from the Iowa Cochlear Implant project (Gantz et al., 1988) is used to illustrate the proposed method.

2. Quantile regression models

Let y_{ik} be the kth measurement for the ith subject, where $k = 1, \ldots, n_i$; $i = 1, \ldots, N$. Let x_{ik} be the corresponding covariate vector. Suppose that measurements from different subjects are independent and those from the same subject are dependent. Assume that the 100τth percentile of y_{ik} is $x_{ik}^{\mathrm{T}}\beta$, where β is a $p \times 1$ unknown parameter vector. Consider the following model for the conditional quantile functions of the response y_{ik}:

$$Q_\tau(y_{ik}|x_{ik}) = x_{ik}^{\mathrm{T}}\beta_0,$$

where β_0 is the true value of β. Let $\epsilon_{ik} = y_{ik} - x_{ik}^{\mathrm{T}}\beta_0$, which is a continuous error term satisfying $p(\epsilon_{ik} \le 0) = \tau$ and with an unspecified density function $f_{ik}(\cdot)$. The median regression is obtained by taking $\tau = 0.5$. Here, what of interest is to find an efficient estimate of β for a particular τ.

2.1. An independence working model

Let $S_{ik} = \tau - I(y_{ik} - x_{ik}^{\mathrm{T}}\beta \le 0)$, where $I(\cdot)$ is an indicator function. Under the independence working model assumption, Chen et al. (2004) proposed using the following estimating functions to make inferences about β:

$$W_\tau(\beta) = \sum_{i=1}^{N}\sum_{k=1}^{n_i} x_{ik}S_{ik}. \tag{1}$$

The resulting estimates $\hat{\beta}_I$ from (1) can be also obtained by minimizing the objective function

$$L_\tau(\beta) = \sum_{i=1}^{N}\sum_{k=1}^{n_i} \rho_\tau(y_{ik} - x_{ik}^{\mathrm{T}}\beta), \tag{2}$$

where $\rho_\tau(u) = u\{\tau - I(u \le 0)\}$ (Koenker and Bassett, 1978). Koenker and D'Orey (1987) developed an efficient algorithm to optimize $L_\tau(\beta)$, which is available in statistical software R (package quantreg). The estimating functions $W_\tau(\beta)$ are based on the independence working model assumption, hence the efficiency of the parameter estimators $\hat{\beta}_I$ derived from $W_\tau(\beta)$ could be improved if the within correlations are incorporated.

2.2. Weighted estimating functions

It is difficult to specify and model the underlying correlation structure from the same subject. Assume that $p(\epsilon_{ik} \le 0, \epsilon_{il} \le 0)$ equals a constant δ, for any $k \ne l$, hence covariance matrix of $S_i = (S_{i1}, \ldots, S_{in_i})^{\mathrm{T}}$ is an exchangeable structure and given by

$$V_i = (\tau - \tau^2)[(1 - \gamma)I_{n_i} + \gamma J_{n_i}],$$

where γ is the correlation coefficient of S_{ik} and S_{il} and equals $(\delta-\tau^2)/(\tau-\tau^2)$, I_{n_i} is the $n_i \times n_i$ identity matrix, and J_{n_i} is an $n_i \times n_i$ matrix of 1s. Let $X_i = (x_{i1}, \ldots, x_{in_i})^{\mathrm{T}}$. Efficient parameter estimators could be obtained by incorporating an appropriate weighted function that accounts for the correlation and variation of the number of repeated measurements for each subject. According to Jung (1996), we consider the following weighted estimating functions based on the exchangeable correlation structure assumption:

$$U_\tau(\beta) = \sum_{i=1}^{N} X_i^{\mathrm{T}} V_i^{-1} S_i,$$

where V_i^{-1} is the inverse matrix of V_i written as

$$V_i^{-1} = \frac{1}{\tau-\tau^2}\left[\frac{J_{n_i}}{n_i[1+(n_i-1)\gamma]} + \frac{1}{1-\gamma}\left(I_{n_i} - \frac{1}{n_i}J_{n_i}\right)\right].$$

If there is no correlation, $\gamma = 0$ and $V_i^{-1} = 1/\{\tau(1-\tau)\}I_{n_i}$. In this case, $U_\tau(\beta)$ is equivalent to the estimating functions $W_\tau(\beta)$.

Suppose that $\hat{\beta}_v$ is the resulting estimator from $U_\tau(\beta)$. Under some regularity conditions, $\hat{\beta}_v$ is a consistent estimator of β_0, $N^{-1/2}U_\tau(\beta_0) \to N(0, V)$, and $\sqrt{N}(\hat{\beta}_v - \beta_0) \to N(0, V_v)$, where $V = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} X_i^{\mathrm{T}} V_i^{-1}\mathrm{cov}(S_i)V_i^{-1}X_i$, and

$$V_v = \lim_{N\to\infty} N D_\tau^{-1}(\beta)\left(\sum_{i=1}^{N} X_i^{\mathrm{T}} V_i^{-1}\mathrm{cov}(S_i)V_i^{-1}X_i\right)\{D_\tau^{-1}(\beta)\}^{\mathrm{T}},$$

where $D_\tau(\beta) = \sum_{i=1}^{N} X_i^{\mathrm{T}} V_i^{-1}\Lambda_i X_i$, and Λ_i is an $n_i \times n_i$ diagonal matrix with the kth diagonal element $f_{ik}(0)$. The $\mathrm{cov}(S_i)$ is unknown and can be estimated empirically (Qu et al., 2000; Stoner and Leroux, 2002). These properties can be derived according to Jung (1996) and Yin and Cai (2005) and the proof is briefly given in Appendix.

2.3. A combination of the between- and within-subject estimating functions

Let $W_{wi} = (1-\gamma)^{-1}(I_{n_i} - J_{n_i}/n_i)$ and $W_{bi} = J_{n_i}/\{n_i(1+(n_i-1)\gamma)\}$. We can extract the following unbiased estimating functions from $U_\tau(\beta)$,

$$U_w(\beta) = \sum_{i=1}^{N} X_i^{\mathrm{T}} W_{wi} S_i = \frac{1}{1-\gamma}\sum_{i=1}^{N} X_i^{\mathrm{T}}(S_i - 1_{n_i}\bar{S}_{i+}),$$

$$U_b(\beta) = \sum_{i=1}^{N} X_i^{\mathrm{T}} W_{bi} S_i = \sum_{i=1}^{N} \frac{1}{1+(n_i-1)\gamma} X_i^{\mathrm{T}} 1_{n_i}\bar{S}_{i+},$$

where 1_{n_i} is an $n_i \times 1$ vector of 1s, and $\bar{S}_{i+} = \sum_{k=1}^{n_i} S_i/n_i$. The estimating functions $U_w(\beta)$ indicate the differences within a subject, and $U_b(\beta)$ capture the information from different subjects. To allow for the different precision of within-subject and between-subject comparisons, it is necessary to combine the estimating functions $U_w(\beta)$ and $U_b(\beta)$ for more efficient parameter estimation. Let

$$X_w = \sum_{i=1}^{N}\begin{pmatrix} X_i^{\mathrm{T}} W_{wi} X_i \\ X_i^{\mathrm{T}} W_{bi} X_i \end{pmatrix} \quad \text{and} \quad G_N(\beta) = \sum_{i=1}^{N}\begin{pmatrix} X_i^{\mathrm{T}} W_{wi} \\ X_i^{\mathrm{T}} W_{bi} \end{pmatrix} S_i.$$

According to Hansen (1982) and Heyde (1989), an optimal combination of $U_w(\beta)$ and $U_b(\beta)$ in terms of asymptotic variance of parameter estimators is given by

$$U_c(\beta) = X_w^{\mathrm{T}} V_G^{-1} G_N(\beta),$$

where

$$V_G = \mathrm{cov}\{G_N(\beta)\} = \sum_{i=1}^{N}\begin{pmatrix} X_i^{\mathrm{T}} W_{wi} \\ X_i^{\mathrm{T}} W_{bi} \end{pmatrix}\mathrm{cov}(S_i)\begin{pmatrix} X_i^{\mathrm{T}} W_{wi} \\ X_i^{\mathrm{T}} W_{bi} \end{pmatrix}^{\mathrm{T}}.$$

Suppose that $\hat{\beta}_c$ is the estimator derived from $U_c(\beta)$. Under some regularity conditions, $\sqrt{N}(\hat{\beta}_c - \beta_0)$ is asymptotically a normal distribution with mean zero and the asymptotic covariance matrix $(X_w^{\mathrm{T}} V_G^{-1} D_c)^{-1}(X_w^{\mathrm{T}} V_G^{-1} X_w)(D_c^{\mathrm{T}} V_G^{-1} X_w)^{-1}$, where $D_c = (D_w, D_b)^{\mathrm{T}}$, and D_w, D_b are the derivatives of the expected values of $U_w(\beta)$ and $U_b(\beta)$, respectively.

3. Parameter and covariance matrix estimation

It is difficult to estimate the covariance matrix of parameter estimators in a quantile regression model because it involves the error density functions which are usually unknown. Resampling and perturbing methods have been proposed to estimate the covariance matrix (Parzen et al., 1994; Yin and Cai, 2005). However, these methods often cause some numerical computational problems (Yin and Cai, 2005).

The induced smoothing method has been extended to the quantile regression with independent data (Wang et al., 2009; Pang et al., 2010). We now extend it to the quantile regression with repeated measurements. Assume $Z \sim N(0, I_p)$, and approximate $\hat{\beta}_v$ by $\beta + \Gamma^{1/2}Z$. The smoothed estimating functions of $U_\tau(\beta)$ can be naturally defined as $\tilde{U}_\tau(\beta) = E_Z[U_\tau(\beta + \Gamma^{1/2}Z)]$, where expectation is over Z. By some algebra calculations, we obtain

$$\tilde{U}_\tau(\beta) = \sum_{i=1}^{N} X_i^{\mathrm{T}} V_i^{-1} \tilde{S}_i,$$

where $\tilde{S}_i = (\tilde{S}_{i1}, \ldots, \tilde{S}_{in_i})^{\mathrm{T}}$, and $\tilde{S}_{ik} = \tau - 1 + \Phi(b_{ik}/\sigma_{ik})$, where $b_{ik} = y_{ik} - x_{ik}^{\mathrm{T}}\beta$ and $\sigma_{ik}^2 = x_{ik}^{\mathrm{T}}\Gamma x_{ik}$. The induced estimating functions $\tilde{U}_\tau(\beta)$ are smoothing functions of β, thus we can easily calculate $\partial\tilde{U}_\tau(\beta)/\partial\beta$ and use it as an approximation of D_τ. Let

$$\tilde{D}_\tau(\beta) = \frac{\partial \tilde{U}_\tau(\beta)}{\partial\beta} = -\sum_{i=1}^{N} X_i^{\mathrm{T}} V_i^{-1} \tilde{\Lambda}_i X_i,$$

where $\tilde{\Lambda}_i$ is an $n_i \times n_i$ diagonal matrix with the kth diagonal element $\sigma_{ik}^{-1}\phi(b_{ik}/\sigma_{ik})$. In general, the resulting estimator $\tilde{\beta}_v$ from $\tilde{U}_\tau(\beta)$ and its covariance matrix can be obtained by iteration. The estimation algorithm can be summarized as the following stepwise procedures.

Step 1. Let $\tilde{\beta}^0 = \hat{\beta}_I$, the estimator obtained from (2), and $\Gamma^0 = N^{-1}I_p$.

Step 2. Given $\tilde{\beta}^{k-1}$ and Γ^{k-1} from the $k-1$ step, update $\hat{\delta}^{k-1}$, by

$$\hat{\delta}^{k-1} = \frac{\sum_{i=1}^{N}\sum_{k=1}^{n_i}\sum_{l\neq k}^{n_i} I(\hat{\epsilon}_{ik} \leq 0, \hat{\epsilon}_{il} \leq 0)}{\sum_{i=1}^{N} n_i(n_i - 1)},$$

where $\hat{\epsilon}_{ik} = y_{ik} - x_{ik}^{\mathrm{T}}\hat{\beta}^{k-1}$, and then update $\tilde{\beta}^k$ and Γ^k by

$$\tilde{\beta}^k = \tilde{\beta}^{k-1} + \{-\tilde{D}_\tau(\tilde{\beta}^{k-1}, \Gamma^{k-1})\}^{-1}\tilde{U}_\tau(\hat{\delta}^{k-1}, \tilde{\beta}^{k-1}, \Gamma^{k-1}),$$
$$\Gamma^k = \tilde{D}_\tau^{-1}(\tilde{\beta}^k, \Gamma^{k-1})V(\hat{\delta}^{k-1}, \tilde{\beta}^k)\tilde{D}_\tau^{-1}(\tilde{\beta}^k, \Gamma^{k-1}).$$

Step 3. Repeat the above iteration Step 2 until convergence.

The finial values of $\tilde{\beta}$ and Γ will be taken as the smoothed estimators of β and its covariance matrix, respectively. The following theorems require regularity conditions which are listed in the Appendix.

Theorem 1. *Under some regularity conditions,* $N^{-1/2}\{\tilde{U}_\tau(\beta) - U_\tau(\beta)\} = o_p(1)$.

Theorem 2. *Under some regularity conditions, the smoothing estimator* $\tilde{\beta}_v \to \beta_0$ *in probability, and* $\sqrt{N}(\tilde{\beta}_v - \beta_0)$ *converges in distribution to* $N(0, V_v)$.

Theorem 1 indicates smoothed and unsmoothed estimating functions are asymptotically equivalent uniformly in β. Basically, this means the induced smoothing method hardly changes the estimating functions except making direct differentiation possible. Theorem 2 demonstrates that the limiting distributions of the smoothed estimators coincide with the unsmoothed estimators. The proofs for these two theorems are given in Appendix. The same procedures can be used to smooth $U_c(\beta)$ and obtain an estimate $\tilde{\beta}_c$ and its covariance estimate.

4. Simulation studies

To investigate the performance of the proposed method, we carry out extensive simulation studies in this section. We generate data from the following linear model:

$$y_{ik} = x_{ik}\beta + \epsilon_{ik}, \quad k = 1, \ldots, n_i, \; i = 1, \ldots, 100,$$

where the true value of $\beta = 1$, x_{ik} are sampled from the uniform distribution $U(0, 1)$. In all the simulations, we use $\epsilon_{ik} = \xi + e_{ik}$, where ξ is a shift to ensure $p(\epsilon_{ik} \leq 0) = \tau$. Five different cases are considered for e_{ik}:

Case (1). Independence models, assume that $e_{i1}, \ldots, e_{in_i}$ are independent identically distributed, and we consider three different distributions, $N(0, 1)$, a t-distribution with three degrees of freedom (t_3), and a Chi-squared distribution with two degrees of freedom (χ_2^2).

Case (2). Normal distribution, assume that $e_i = (e_{i1}, \ldots, e_{in_i})^{\mathrm{T}}$ follow a multivariate normal distribution with mean zero, unit standard deviation, and an exchangeable correlation matrix, $\Sigma_e(\rho)$, with a correlation parameter ρ.

Case (3). Heteroscedastic normal, assume that $e_{ik} = \alpha_i + \eta_{ik}$, where random effects $\alpha_i \sim N(0, 1)$, and error terms $\eta_{ik} \sim x_{ik}N(0, \sigma^2)$, in which σ equals 1 or 0.5.

Table 1

Biases (Bias) and relative efficiencies (Eff) to the independence estimator $\hat{\beta}_I$ of the weighted estimator $\hat{\beta}_v$, the proposed estimator $\hat{\beta}_c$, and an estimator $\hat{\beta}_m$ from a linear mixed effects model at three different quantiles for Case (1).

	$N(0, 1)$		t_3		χ^2_2	
$\tau = 0.25$ $n_i = 3$						
	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.004	1.000	−0.002	1.000	0.006	1.000
$\hat{\beta}_v$	0.002	1.094	−0.008	1.077	0.011	1.084
$\hat{\beta}_c$	0.006	1.080	−0.004	1.061	0.013	1.066
$\tau = 0.25$ $n_i \sim U(2, 10)$	$N(0, 1)$		t_3		χ^2_2	
	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	−0.001	1.000	−0.002	1.000	−0.003	1.000
$\hat{\beta}_v$	0.002	1.070	−0.001	1.048	0.008	1.024
$\hat{\beta}_c$	0.001	1.069	−0.002	1.058	0.007	1.014
$\tau = 0.50$ $n_i = 3$	$N(0, 1)$		t_3		χ^2_2	
	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.001	1.000	−0.004	1.000	−0.003	1.000
$\hat{\beta}_v$	0.003	1.072	0.001	1.092	0.011	1.081
$\hat{\beta}_c$	0.000	1.076	−0.004	1.092	0.004	1.104
$\hat{\beta}_m$	−0.003	1.562	−0.002	0.636	−0.031	0.257
$\tau = 0.50$ $n_i \sim U(2, 10)$	$N(0, 1)$		t_3		χ^2_2	
	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.001	1.000	−0.003	1.000	0.004	1.000
$\hat{\beta}_v$	0.009	0.984	0.006	0.967	0.024	0.961
$\hat{\beta}_c$	0.003	1.019	−0.000	1.053	0.015	1.046
$\hat{\beta}_m$	−0.001	1.498	−0.003	0.613	0.009	0.268
$\tau = 0.75$ $n_i = 3$	$N(0, 1)$		t_3		χ^2_2	
	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.001	1.000	0.004	1.000	0.003	1.000
$\hat{\beta}_v$	0.019	0.908	0.028	0.790	0.061	0.740
$\hat{\beta}_c$	0.000	1.067	0.005	1.068	0.015	1.045
$\tau = 0.75$ $n_i \sim U(2, 10)$	$e_{ik} \sim N(0, 1)$		t_3		χ^2_2	
	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	−0.004	1.000	0.002	1.000	0.001	1.000
$\hat{\beta}_v$	0.013	0.792	0.026	0.759	0.060	0.771
$\hat{\beta}_c$	−0.001	1.012	0.008	1.046	0.018	1.056

Case (4). Nonnormal distribution, suppose that e_i follow a multivariate t distribution with three degrees of freedom (T_3) and AR(1) correlation matrix, $\Sigma_a(\rho)$, with a correlation parameter ρ.

Case (5). Mixed distribution, assume that $e_{ik} = \alpha_i + \eta_{ik}$, in which error terms η_{ik} follow χ^2_2. For random effects α_i, we consider $N(0, 1)$ and χ^2_2 distributions.

For cases (2) and (3), the correlation coefficient ρ takes values of 0.3 and 0.7. For the number of repeated measurements n_i, we consider balanced design with $n_i = 3$ and unbalanced design where n_i is an integer number randomly generated between 2 and 10 with equal probability. For each case, we estimate the regression parameter β for three different quantile functions, $\tau = (0.25, 0.5, 0.75)$, and 1000 simulation studies are carried out. For comparison, we also provide the estimates from a linear mixed effects model. The simulation results are provided in Tables 1–4.

As we can see, when correlations do not exist (case (1)), biases and efficiencies of the proposed estimator ($\hat{\beta}_c$) are comparable with those of the independence estimator ($\hat{\beta}_I$) (Table 1), and these two estimators are more efficient than the ($\hat{\beta}_v$) obtained from the weighted estimating functions.

Table 2
Biases (Bias) and relative efficiencies (Eff) to the independence estimator $\hat{\beta}_I$ of the weighted estimator $\hat{\beta}_v$, and the proposed estimator $\hat{\beta}_c$ at $\tau = 0.25$ quantile for the Cases 2–5.

Case (2): $e_i \sim N(0, \Sigma_e(\rho))$, and $\Sigma_e(\rho) = (1-\rho)I + \rho 11^T$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\rho = 0.3$		$\rho = 0.7$		$\rho = 0.3$		$\rho = 0.7$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.006	1.000	0.002	1.000	0.001	1.000	−0.014	1.000
$\hat{\beta}_v$	0.009	1.100	0.019	1.286	0.012	1.187	0.014	1.980
$\hat{\beta}_c$	0.006	1.106	−0.002	1.362	0.003	1.212	−0.007	2.254
Case (3): $e_{ik} = \alpha_i + \eta_{ik}, \alpha_i \sim N(0, 1), \eta_{ik} \sim x_{ik}N(0, \sigma^2)$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\sigma = 1$		$\sigma = 2$		$\sigma = 1$		$\sigma = 2$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	−0.006	1.000	0.0163	1.000	0.007	1.000	0.005	1.000
$\hat{\beta}_v$	0.039	1.134	0.047	0.765	0.044	1.744	0.059	0.841
$\hat{\beta}_c$	−0.004	1.415	0.012	1.032	0.003	2.315	0.002	1.261
Case (4): $e_i \sim T_3(0, \Sigma_a(\rho))$, and $\Sigma_a(\rho) = (\rho^{\lvert k-l \rvert})$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\rho = 0.3$		$\rho = 0.7$		$\rho = 0.3$		$\rho = 0.7$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	−0.000	1.000	−0.004	1.000	−0.002	1.000	−0.003	1.000
$\hat{\beta}_v$	−0.009	1.066	0.016	1.243	0.005	1.152	0.011	1.435
$\hat{\beta}_c$	−0.002	1.102	−0.010	1.323	−0.001	1.156	−0.003	1.476
Case (5): $e_{ik} = \alpha_i + \eta_{ik}, \eta_{ik} \sim \chi_2^2$								
	$\alpha_i \sim N(0, 1)$				$\alpha_i \sim \chi_2^2$			
	$n_i = 3$		$n_i \sim U(2, 10)$		$n_i = 3$		$n_i \sim U(2, 10)$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	−0.005	1.000	0.000	1.000	0.015	1.000	0.012	1.000
$\hat{\beta}_v$	0.011	1.183	0.020	1.396	0.036	1.110	0.034	1.406
$\hat{\beta}_c$	−0.001	1.181	0.001	1.486	0.020	1.168	0.011	1.531

When the correlation exists (cases 2–5), the results show that the proposed estimator has small bias and performs better than the other quantile methods across the four simulation cases. Furthermore, its efficiency increases as the correlation (ρ) increases. Even when the correlation structure is not correctly specified (cases (3) and (4)), the proposed method outperforms other quantile methods. Compared with the estimator $\hat{\beta}_v$, the independence estimator and the proposed estimator have smaller bias. When random effect is the standard norm distributed and the error term is χ_2^2 distributed, all the estimators are more efficient at the lower quantiles.

In the simulation studies, we also compare the median regression method ($\tau = 0.5$) with the linear mixed effect model (LME) (the second panel in Tables 1 and 3). When the error distribution is normal, the LME and the quantile method have comparable bias, but the efficiency of LME is higher than that of the quantile methods, as one expected. However, when the error distribution has heavy tails (t-distribution and χ_2^2), the quantile methods outperforms the LME method. These agree with the expectation that the median regression performs better than the mean regression when the error distribution is with heavy tails, and is less efficient when the distribution is normal.

5. Audiology growth data

In this section, the proposed method is illustrated by the audiology growth data reported by Gantz et al. (1988) and successively analyzed by Núñez-Anton and Woodwoth (1994). The data consist of repeated measurements of percent correct scores on a sentence test administered under audition-only conditions to subjects wearing two different cochlear implants, referred to here as A and B. A total of 23 subjects were randomly assigned to group A and 21 subjects were in group B. The electrode array was implanted five to six weeks prior to being electrically connected to the external speech processor. Since the subjects were profoundly, bilaterally deaf, the baseline values for the sentence test were all zeros. After the connection, the measurements were scheduled at 1, 9, 18, and 30 months. There was some variation in actual follow-up times, therefore measurement times were an unbalanced design. Gantz et al. (1988) considered the quadratic model

$$y_{ik} = \beta_0 + \beta_1 t_{ik} + \beta_2 t_{ik}^2 + \beta_3 G_i + \beta_4 G_i * t_{ik} + \beta_5 G_i * t_{ik}^2 + \epsilon_{ik},$$

Table 3
Biases (Bias) and relative efficiencies (Eff) to the independence estimator $\hat{\beta}_I$ of the weighted estimator $\hat{\beta}_v$, the proposed estimator $\hat{\beta}_c$, and an estimator $\hat{\beta}_m$ from a linear mixed effects model at $\tau = 0.50$ quantile for all Cases 2–5.

Case (2): $e_i \sim N(0, \Sigma_e(\rho))$, and $\Sigma_e(\rho) = (1-\rho)I + \rho 11^{\mathrm{T}}$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\rho = 0.3$		$\rho = 0.7$		$\rho = 0.3$		$\rho = 0.7$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	−0.005	1.000	−0.001	1.000	0.001	1.000	0.007	1.000
$\hat{\beta}_v$	0.008	0.963	0.029	1.225	0.018	1.087	0.025	1.807
$\hat{\beta}_c$	−0.004	1.074	0.002	1.432	0.003	1.200	0.004	2.009
$\hat{\beta}_m$	−0.002	1.563	0.002	2.461	0.001	1.655	0.001	4.215
Case (3): $e_{ik} = \alpha_i + \eta_{ik}, \alpha_i \sim N(0, 1), \eta_{ik} \sim x_{ik}N(0, \sigma^2)$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\sigma = 1$		$\sigma = 2$		$\sigma = 1$		$\sigma = 2$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.002	1.000	0.004	1.000	−0.001	1.000	0.003	1.000
$\hat{\beta}_v$	0.058	1.013	0.059	0.712	0.049	1.622	0.066	0.817
$\hat{\beta}_c$	0.004	1.459	0.006	1.057	0.002	2.297	0.004	1.405
$\hat{\beta}_m$	0.005	2.434	0.005	1.446	−0.002	4.459	−0.003	1.794
Case (4): $e_i \sim T_3(0, \Sigma_a(\rho))$, and $\Sigma_a(\rho) = (\rho^{\lvert k-l \rvert})$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\rho = 0.3$		$\rho = 0.7$		$\rho = 0.3$		$\rho = 0.7$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.003	1.000	0.009	1.000	0.000	1.000	−0.008	1.000
$\hat{\beta}_v$	0.014	1.063	0.029	1.133	0.014	1.019	0.013	1.319
$\hat{\beta}_c$	0.004	1.098	0.004	1.275	0.004	1.055	−0.003	1.387
$\hat{\beta}_m$	0.005	0.532	−0.002	0.877	0.003	0.632	−0.006	0.889
Case (5): $e_{ik} = \alpha_i + \eta_{ik}, \eta_{ik} \sim \chi_2^2$								
	$\alpha_i \sim N(0, 1)$				$\alpha_i \sim \chi_2^2$			
	$n_i = 3$		$n_i \sim U(2, 10)$		$n_i = 3$		$n_i \sim U(2, 10)$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	−0.001	1.000	−0.001	1.000	0.008	1.000	0.032	1.000
$\hat{\beta}_v$	0.026	1.054	0.037	1.154	0.077	1.022	0.078	1.424
$\hat{\beta}_c$	0.002	1.101	0.007	1.283	0.023	1.253	0.025	1.710
$\hat{\beta}_m$	0.001	0.358	−0.002	0.596	−0.015	0.735	−0.005	1.636

where y_{ik} is the percent correct score for the ith subject at time j, t_{ik} is the measurement time, and G_i takes 1 if the ith subject belonged to group B and −1 if the subject was in group A, and ϵ_{ik} is an error term. Gantz et al. (1988) eliminated measurements on subjects who never reached 5% understanding. Assume that the error terms follow a multivariate normal distribution, Gantz et al. (1988) obtained parameter estimates by the restricted maximum likelihood method. The estimate of the correlation coefficient is 0.94 indicating a high correlation between the repeated measurements.

Since removing the measurements could cause a loss of information, we therefore use all the measurements for analysis. Fig. 1 indicates that the audiology performances for most of the subjects in groups A and B were improved over time. Fig. 2 shows the median of the percent correct scores over time for two groups. Three possible outliers are present. Fig. 3 reports the estimates of the marginal density of the response in the entire sample, in group A only, and in group B only. It can be seen that the shapes of group B and the entire sample densities are asymmetric. The mean is highly influenced by unusually large (small) measurements which may exist in skewed populations. The median of the percent correct scores is not sensitive to extreme values and might be more informative than the mean for a skewed distribution. We therefore consider the median regression

$$Q_{0.5}(y_{ik}|x_{ik}) = \beta_0 + \beta_1 t_{ik} + \beta_2 t_{ik}^2 + \beta_3 G_i + \beta_4 G_i * t_{ik} + \beta_5 G_i * t_{ik}^2.$$

The resulting parameter estimates and 95% confidence intervals are 11.35 (0.59, 22.11) for β_0, 2.83 (1.59, 4.07) for β_1, −0.56 (−0.10, −0.02) for β_2, −7.13 (−17.89, 3.64) for β_3, −0.092 (−1.33, 1.14) for β_4, and 0.004 (−0.034, 0.04) for β_5. Hence, the differences between the two cochlear implant types are not statistically significant (β_3 is insignificant at 5% level). The audiologic performance becomes better after the first few months of use, and then they are more consistent over time (β_1 and β_2 are significant). The average percent correct score in group A reached the maximum value at 34.3 month, and the scores in group B reached the maximum value at 26.2 month.

Table 4
Biases (Bias) and relative efficiencies (Eff) to the independence estimator $\hat{\beta}_I$ of the weighted estimator $\hat{\beta}_v$, and the proposed estimator $\hat{\beta}_c$ at $\tau = 0.75$ quantile for Cases 2–5.

Case (2): $e_i \sim N(0, \Sigma_e(\rho))$, and $\Sigma_e(\rho) = (1-\rho)I + \rho 11^{\mathrm{T}}$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\rho = 0.3$		$\rho = 0.7$		$\rho = 0.3$		$\rho = 0.7$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.000	1.000	−0.002	1.000	−0.005	1.000	0.005	1.000
$\hat{\beta}_v$	0.026	0.979	0.056	1.272	0.027	1.061	0.045	1.520
$\hat{\beta}_c$	−0.000	1.102	0.003	1.328	0.001	1.217	0.001	2.008
Case (3): $e_{ik} = \alpha_i + \eta_{ik}, \alpha_i \sim N(0, 1), \eta_{ik} \sim x_{ik}N(0, \sigma^2)$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\sigma = 1$		$\sigma = 2$		$\sigma = 1$		$\sigma = 2$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	−0.010	1.000	0.004	1.000	−0.009	1.000	−0.001	1.000
$\hat{\beta}_v$	0.073	0.798	0.130	0.418	0.065	1.183	0.145	0.469
$\hat{\beta}_c$	−0.004	1.346	0.003	1.057	−0.002	2.124	0.010	1.381
Case (4): $e_i \sim T_3(0, \Sigma_a(\rho))$, and $\Sigma_a(\rho) = (\rho^{\lvert k-l \rvert})$								
	Balanced design $n_i = 3$				Unbalanced design $n_i \sim U(2, 10)$			
	$\rho = 0.3$		$\rho = 0.7$		$\rho = 0.3$		$\rho = 0.7$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.009	1.000	0.007	1.000	0.001	1.000	0.008	1.000
$\hat{\beta}_v$	0.052	0.727	0.085	0.805	0.041	0.744	0.060	0.997
$\hat{\beta}_c$	0.012	1.067	0.015	1.415	0.009	1.075	0.022	1.329
Case (5): $e_{ik} = \alpha_i + \eta_{ik}, \eta_{ik} \sim \chi^2_2$								
	$\alpha_i \sim N(0, 1)$				$\alpha_i \sim \chi^2_2$			
	$n_i = 3$		$n_i \sim U(2, 10)$		$n_i = 3$		$n_i \sim U(2, 10)$	
	Bias	Eff	Bias	Eff	Bias	Eff	Bias	Eff
$\hat{\beta}_I$	0.002	1.000	−0.005	1.000	0.040	1.000	0.007	1.000
$\hat{\beta}_v$	0.079	0.766	0.068	0.791	0.076	1.071	0.114	1.158
$\hat{\beta}_c$	0.014	1.096	0.007	1.054	0.064	1.204	0.004	1.449

6. Discussion

In this paper, we have studied linear quantile regression models for longitudinal data, which provide an appealing alternative way of studying covariate effects. To account for the within correlation and capture varying cluster sizes, a combination of the between- and within-subject estimating functions is proposed to estimate regression parameters. Núñez-Anton et al. (1999) considered nonlinear model with longitudinal data with an unknown link function which is estimated by kernel smoothing. In our paper, the link function is known and we use the induced smooth method to estimate the asymptotic covariance of the mean parameter estimates. This approach can bypass density estimation of the errors and reduce computational burden caused by the unsmooth estimating functions. The extensive simulation studies indicate that the proposed method performs well even when the correlation structure is misspecified.

It is worth mentioning that when each covariate is a subject-level covariate (Neuhaus and Kalbfleish, 1998), $X_{ik} = X_{il} = z_i$, that is z_i does not change within each subject or cluster, but it may differ among subjects, then the estimating functions $U_w(\beta) = 0$, and $U_b(\beta)$ becomes identical to the weighted estimating function $U_\tau(\beta)$. This is because

$$
\begin{aligned}
U_b(\beta) &= \frac{1}{\tau(1-\tau)} \sum_{i=1}^{N} z_i 1_i^{\mathrm{T}} \frac{1_{n_i} 1_{n_i}^{\mathrm{T}}}{n_i[1 + (n_i - 1)\gamma]} S_i \\
&= \frac{1}{\tau(1-\tau)} \sum_{i=1}^{N} \frac{1}{[1 + (n_i - 1)\gamma]} z_i 1_i^{\mathrm{T}} S_i \\
&= \frac{1}{\tau(1-\tau)} \sum_{i=1}^{N} \frac{1}{[1 + (n_i - 1)\gamma]} X_i^{\mathrm{T}} S_i \\
&= U_\tau(\beta).
\end{aligned}
$$

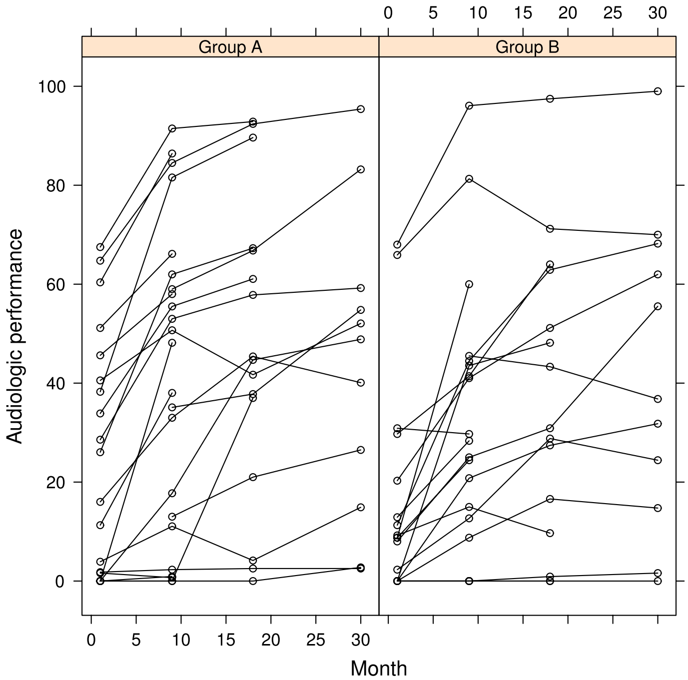

Fig. 1. The percent correct score versus the measurement time for groups A and B.

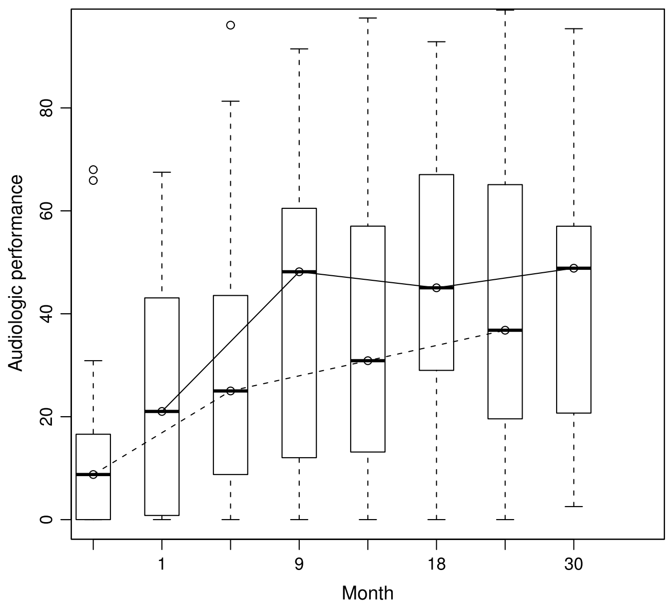

Fig. 2. Boxplot of percent correct scores for group A (solid line) and group B (dashed line).

In this case, we propose using $U_\tau(\beta)$ to obtain parameter estimates. Furthermore, for subject-level covariates and balanced designs ($n_i = n$), it can be seen that the correlation coefficient γ and cluster sizes n_i do not affect the estimate of β when an exchangeable working model is used, therefore, $U_\tau(\beta)$ is equivalent to the estimating functions based on the independence working model.

Yin and Cai (2005) considered quantile regression models with multivariate failure time data and employed the bootstrap and perturbation resampling methods for the parameter covariance estimation. Assume that the link function is known, Karlsson (2007) examined a weighted version of the quantile regression estimator to the case of nonlinear regression with longitudinal data. The proposed method can be extended to linear quantile regression models with multivariate failure time data and to the nonlinear quantile regression, which will be considered in our future work.

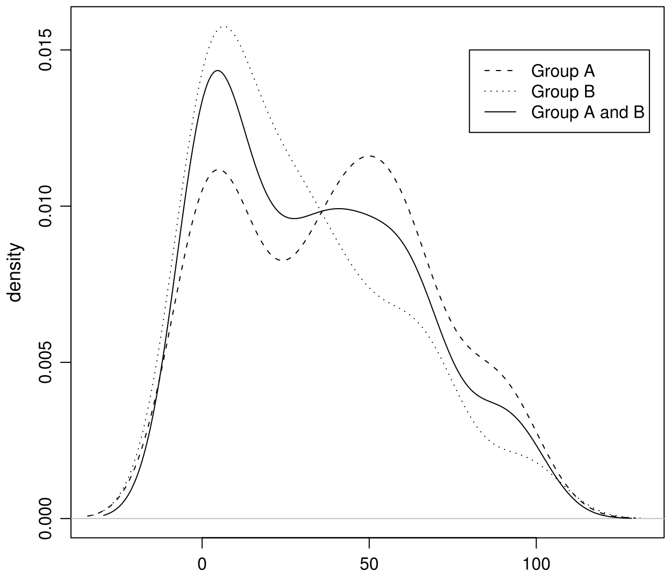

Fig. 3. The kernel density plot of the percent correct scores for the entire sample (solid line), for group A only (dashed line), and for group B only (dotted line).

Acknowledgments

The authors wish to thank the Associate Editor and two anonymous referees for their constructive comments. Dr. Liya Fu's postdoctoral research was supported by the Centre for Applications in Natural Resource Mathematics (CARM), School of Mathematics and Physics, The University of Queensland, Australia.

Appendix

In this appendix, we impose the following regularity conditions for establishing the consistency and asymptotic normality of $\hat{\beta}_v$ derived from the estimating functions $U_\tau(\beta)$ and prove Theorems 1 and 2.

A1. The cumulative distribution function $F_{ik}(\cdot)$ is absolutely continuous, with continuous density function $f_{ik}(\cdot)$ uniformly bounded, and $f'_{ik}(\cdot)$ exists and is uniformly bounded.
A2. The true value β_0 is in the interior of a bounded convex region **B**.
A3. For each i, the number of repeated measurements n_i is bounded and X_i satisfies the following conditions:
  (a) For any positive definite matrix W_i, $N^{-1}\sum_{i=1}^N X_i^{\mathrm{T}} W_i \Lambda_i X_i$ converges to a positive definite matrix; where Λ_i is an $n_i \times n_i$ diagonal matrix with the kth diagonal element $f_{ik}(0)$.
  (b) $\sup_i \|X_i\| < +\infty$, where $\|\cdot\|$ denotes the Euclidean norm.
A4. Matrix Γ is positive definite and $\Gamma = O(1/N)$.
A5. Matrix $\tilde{D}_\tau(\beta)$ is positive definite.

Proof of the consistency and asymptotic normality of $\hat{\beta}_v$

Let $H_i^{\mathrm{T}} = X_i^{\mathrm{T}} V_i^{-1}$, then $U_\tau(\beta) = \sum_{i=1}^N H_i^{\mathrm{T}} S_i$. Let $\bar{U}_\tau(\beta) = \sum_{i=1}^N H_i^{\mathrm{T}} P_i$, where $P_i = (\tau - p(y_{i1} - x_{ik}^{\mathrm{T}}\beta \le 0), \ldots, \tau - p(y_{in_i} - x_{in_i}^{\mathrm{T}}\beta \le 0))^{\mathrm{T}}$. We can obtain

$$
\begin{aligned}
N^{-1}\{U_\tau(\beta) - \bar{U}_\tau(\beta)\} &= N^{-1}\sum_{i=1}^N H_i^{\mathrm{T}}(S_i - P_i) \\
&= N^{-1}\sum_{i=1}^N H_i^{\mathrm{T}} \begin{pmatrix} p(y_{i1} - x_{ik}^{\mathrm{T}}\beta \le 0) - I(y_{i1} - x_{ik}^{\mathrm{T}}\beta \le 0) \\ \cdot \\ \cdot \\ \cdot \\ p(y_{in_i} - x_{in_i}^{\mathrm{T}}\beta \le 0) - I(y_{in_i} - x_{in_i}^{\mathrm{T}}\beta \le 0) \end{pmatrix} \\
&= N^{-1}\sum_{i=1}^N \sum_{k=1}^{n_i} h_{ik}[p(y_{ik} - x_{ik}^{\mathrm{T}}\beta \le 0) - I(y_{ik} - x_{ik}^{\mathrm{T}}\beta \le 0)],
\end{aligned}
$$

where h_{ik} is a $p \times 1$ vector and $(h_{i1}, \ldots, h_{in_i}) = H_i^{\mathrm{T}}$. Under condition A3 and from the uniform strong law of large numbers (Pollard, 1990),

$$\sup_{\beta\in\mathbf{B}} \left| N^{-1} \sum_{i=1}^{N} \sum_{k=1}^{n_i} h_{ik}[p(y_{ik} - x_{ik}^{\mathrm{T}}\beta \le 0) - I(y_{ik} - x_{ik}^{\mathrm{T}}\beta \le 0)] \right| \to o(N^{-1/2+\epsilon}) \quad \text{a.s..}$$

Therefore, $\sup_{\beta\in\mathbf{B}} \|N^{-1}\{U_\tau(\beta) - \bar{U}_\tau(\beta)\}\| = o(N^{-1/2+\epsilon})$ a.s.. Now focusing on $\bar{U}_\tau(\beta)$,

$$D_N(\beta_0) = \frac{1}{N} \frac{\partial \bar{U}_\tau(\beta)}{\partial \beta}\bigg|_{\beta=\beta_0} = -\frac{1}{N}\sum_{i=1}^{N} H_i^{\mathrm{T}} \Lambda_i X_i.$$

Under condition A3, $D_N(\beta_0)$ is negative definite. Since $p(y_{ik} - x_{ik}^{\mathrm{T}}\beta_0 \le 0) = \tau$, β_0 is the unique solution of the equation $\bar{U}_\tau(\beta) = 0$. Due to $\hat{\beta}_\upsilon$ is the solution of the equation $U_\tau(\beta) = 0$, hence $\hat{\beta}_\upsilon \to \beta_0$ as $N \to +\infty$.

Because S_i are independent random variables with mean zero, and $\mathrm{var}(N^{-1/2}U_\tau(\beta_0)) = N^{-1}\sum_{i=1}^{N} X_i^{\mathrm{T}} V_i^{-1}\mathrm{cov}(S_i)V_i^{-1}X_i$. The multivariate central limit theorem implies that $U_\tau(\beta_0) \to N(0, V)$. For any β satisfying $\|\beta - \beta_0\| < cN^{-1/3}$,

$$\begin{aligned} U_\tau(\beta) - U_\tau(\beta_0) &= \sum_{i=1}^{N} H_i^{\mathrm{T}}(\beta)S_i(\beta) - \sum_{i=1}^{N} H_i^{\mathrm{T}}(\beta_0)S_i(\beta_0) \\ &= \sum_{i=1}^{N} H_i^{\mathrm{T}}(\beta)\{S_i(\beta) - S_i(\beta_0)\} + \sum_{i=1}^{N}\{H_i(\beta) - H_i(\beta_0)\}^{\mathrm{T}} S_i(\beta_0). \end{aligned}$$

According to Lemma 3 in Jung (1996), the first term

$$\begin{aligned} \sum_{i=1}^{N} H_i^{\mathrm{T}}(\beta)\{S_i(\beta) - S_i(\beta_0)\} &= \sum_{i=1}^{N} H_i^{\mathrm{T}}(\beta)P_i(\beta) + \sum_{i=1}^{N} H_i^{\mathrm{T}}(\beta)\{S_i(\beta) - S_i(\beta_0) - P_i(\beta)\} \\ &= \bar{U}(\beta) + o_p(N^{1/2}). \end{aligned}$$

The second term

$$\sum_{i=1}^{N}\{H_i(\beta) - H_i(\beta_0)\}^{\mathrm{T}} S_i(\beta_0) = \sum_{i=1}^{N}\sum_{k=1}^{n_i}(h_{ik}(\beta) - h_{ik}(\beta_0))[p(y_{ik} - x_{ik}^{\mathrm{T}}\beta_0 \le 0) - I(y_{ik} - x_{ik}^{\mathrm{T}}\beta_0 \le 0)].$$

By term of Pollard (1990), $\sum_{i=1}^{N}\{H_i(\beta) - H_i(\beta_0)\}^{\mathrm{T}} S_i(\beta_0) = o_p(N^{1/2+\epsilon})$. Therefore, $U_\tau(\beta) - U_\tau(\beta_0) = \bar{U}(\beta) + o_p(N^{1/2})$. By Taylor's expansion of $\bar{U}(\beta)$, we can obtain

$$\frac{1}{\sqrt{N}}\{U_\tau(\beta) - U_\tau(\beta_0)\} = \frac{1}{N}\frac{\partial \bar{U}_\tau(\beta)}{\partial \beta}\bigg|_{\beta=\beta_0} \sqrt{N}(\beta - \beta_0) + o_p(1).$$

Because $U_\tau(\hat{\beta}) = 0$ and $\hat{\beta}$ is in the $N^{-1/3}$ neighborhood of β_0, we have

$$\sqrt{N}(\hat{\beta} - \beta_0) = -D_N^{-1}(\beta_0)\frac{1}{\sqrt{N}}U_\tau(\beta_0) + o_p(1).$$

Therefore $\sqrt{N}(\hat{\beta} - \beta_0) \to N(0, V_\upsilon)$, where $V_\upsilon = \lim_{N\to\infty} D_N^{-1}(\beta_0)V\{D_N^{-1}(\beta_0)\}^{\mathrm{T}}$.

Proof of Theorem 1 *and* Theorem 2

Proof of Theorem 1. Because $\tilde{S}_{ik} - S_{ik} = \mathrm{sgn}(-d_{ik})\Phi(-|d_{ik}|)$, where $\mathrm{sgn}(\cdot)$ is the sign function and $d_{ik} = \epsilon_{ik}/\sigma_{ik}$. We can obtain

$$\begin{aligned} N^{-1/2}\{\tilde{U}_\tau(\beta) - U_\tau(\beta)\} &= N^{-1/2}\sum_{i=1}^{N} X_i^{\mathrm{T}} V_i^{-1} \begin{pmatrix} \mathrm{sgn}(-d_{i1})\Phi(-|d_{i1}|) \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{sgn}(-d_{in_i})\Phi(-|d_{in_i}|) \end{pmatrix} \\ &= N^{-1/2}\sum_{i=1}^{N}\sum_{k=1}^{n_i} z_{ik}\,\mathrm{sgn}(-d_{ik})\Phi(-|d_{ik}|), \end{aligned}$$

where z_{ik} is the kth column of $X_i^{\mathrm{T}}V_i^{-1}$. Because

$$\begin{aligned} E(\tilde{S}_{ik} - S_{ik}) &= \int_{-\infty}^{+\infty} \mathrm{sgn}(-d_{ik})\Phi(-|d_{ik}|)f_{ik}(\epsilon)\mathrm{d}\epsilon \\ &= \int_{-\infty}^{+\infty} \Phi(-|\epsilon|/\sigma_{ik})\{2I(\epsilon \le 0) - 1\}f_{ik}(\epsilon)\mathrm{d}\epsilon \\ &= \sigma_{ik}\int_{-\infty}^{+\infty} \Phi(-|t|)\{2I(t \le 0) - 1\}[f_{ik}(0) + f'_{ik}(\zeta(t))\sigma_{ik}t]\mathrm{d}t, \end{aligned}$$

where $\zeta(t)$ is between 0 and $\sigma_{ik}t$. Since $\int_{-\infty}^{+\infty} \Phi(-|t|)\{2I(t \le 0) - 1\}\mathrm{d}t = 0$, we have $\sigma_{ik}\int_{-\infty}^{+\infty} \Phi(-|t|)\{2I(t \le 0) - 1\} f_{ik}(0)\mathrm{d}t = 0$. Because $\int_{-\infty}^{+\infty} |t|\Phi(-|t|)\mathrm{d}t = 1/2$, and by condition A1, there exists a constant M such that $\sup_{ik} |f'_{ik}(\zeta(t))| \le M$. Therefore

$$\begin{aligned} |E(\tilde{S}_{ik} - S_{ik})| &\le \sigma_{ik}^2 \int_{-\infty}^{+\infty} |t|\Phi(-|t|)|f'_{ik}(\zeta(t))|\mathrm{d}t \\ &\le M\sigma_{ik}^2/2. \end{aligned}$$

Under conditions A3 and A4, as $N \to \infty$,

$$\|N^{-1/2}E\{\tilde{U}_\tau(\beta) - U_\tau(\beta)\}\| \le N^{-1/2}\sup_{i,k}|z_{ik}|\sum_{i=1}^{N} M\sigma_{ik}^2/2 = o(1).$$

In addition,

$$N^{-1}\mathrm{var}\{\tilde{U}_\tau(\beta) - U_\tau(\beta)\} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{var}\left\{\sum_{k=1}^{n_i} z_{ik}\mathrm{sgn}(-d_{ik})\Phi(-|d_{ik}|)\right\}.$$

By Cauchy–Schwartz inequality,

$$N^{-1}\mathrm{var}\{\tilde{U}_\tau(\beta) - U_\tau(\beta)\} \le \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{n_i} z_{ik}z_{ik}^{\mathrm{T}}\mathrm{var}(\tilde{S}_{ik} - S_{ik}) + \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{n_i}\sum_{k'\ne k}^{n_i} z_{ik}z_{ik'}^{\mathrm{T}}\sqrt{\mathrm{var}(\tilde{S}_{ik} - S_{ik})\mathrm{var}(\tilde{S}_{ik'} - S_{ik'})}.$$

For each $k = 1, \ldots, n_i$,

$$\begin{aligned} \mathrm{var}(\tilde{S}_{ik} - S_{ik}) &\le E(\tilde{S}_{ik} - S_{ik})^2 = \int_{-\infty}^{+\infty} \{\mathrm{sgn}(-d_{ik})\Phi(-|d_{ik}|)\}^2 f_{ik}(\epsilon)\mathrm{d}\epsilon \\ &= \sigma_{ik}\int_{-\infty}^{+\infty} \Phi^2(-|t|)f_{ik}(\sigma_{ik}t)\mathrm{d}t \\ &= \sigma_{ik}\int_{|t|>\Delta} \Phi^2(-|t|)f_{ik}(\sigma_{ik}t)\mathrm{d}t + \sigma_{ik}\int_{|t|\le\Delta} \Phi^2(-|t|)f_{ik}(\sigma_{ik}t)\mathrm{d}t \\ &\le \Phi^2(-\Delta) + \sigma_{ik}\Delta f_{ik}(\zeta), \end{aligned}$$

where Δ is a positive value, and ζ lies between $(-\sigma_{ik}\Delta, \sigma_{ik}\Delta)$. Let $\Delta = N^{1/3}$. Under condition A4, since $\sigma_{ik} = O(N^{-1/2})$, then $\sigma_{ik}\Delta = O(N^{-1/6})$. As $N \to \infty$, both $\Phi^2(-\Delta)$ and $\sigma_{ik}\Delta f_{ik}(\zeta)$ go to 0. By conditions A2 and A3, it is easy to obtain $N^{-1}\mathrm{var}\{\tilde{U}_\tau(\beta) - U_\tau(\beta)\} = o(1)$. Therefore, we have $N^{-1/2}\{\tilde{U}_\tau(\beta) - U_\tau(\beta)\} \to 0$ as $N \to +\infty$ for any β. □

Proof of Theorem 2. According to Theorem 1 and $\sup_{\beta\in\mathbf{B}} \|N^{-1}\{U_\tau(\beta) - \bar{U}_\tau(\beta)\}\| = o(N^{-1/2+\epsilon})$ a.s., and by the triangle inequality, we can obtain $\sup_{\beta\in\mathbf{B}} \|N^{-1}\{\tilde{U}_\tau(\beta) - \bar{U}_\tau(\beta)\}\| = o(N^{-1/2+\epsilon})$. Since β_0 is the unique solution of equation $\bar{U}(\beta) = 0$. This together with the definite of $\tilde{\beta}$ implies $\tilde{\beta} \to \beta_0$ as $N \to +\infty$.

Before proving the normality of $\tilde{\beta}$, we first prove $N^{-1}\{\tilde{D}_\tau(\beta_0) - D_\tau(\beta_0)\} \xrightarrow{p} 0$. Let $\bar{x}_{i+} = \sum_{k=1}^{n_i} x_{ik}/n_i$. It is easy to obtain

$$E\{\tilde{D}(\beta_0)\} - D(\beta_0) = \frac{1}{(1-\gamma)}\sum_{i=1}^{N}\sum_{k=1}^{n_i}\left\{x_{ik}x_{ik}^{\mathrm{T}} - \frac{n_i\gamma}{1+(n_i-1)\gamma}\bar{x}_{i+}x_{ik}^{\mathrm{T}}\right\}\left\{\sigma_{ik}^{-1}E\phi\left(\frac{\epsilon_{ik}}{\sigma_{ik}}\right) - f_{ik}(0)\right\}.$$

Because

$$\begin{aligned}\left|\sigma_{ik}^{-1}E\phi\left(\frac{\epsilon_{ik}}{\sigma_{ik}}\right)-f_{ik}(0)\right| &= \left|\sigma_{ik}^{-1}\int_{-\infty}^{+\infty}\phi\left(\frac{\epsilon}{\sigma_{ik}}\right)f_{ik}(\epsilon)\mathrm{d}\epsilon - f_{ik}(0)\right| \\ &= \left|\int_{-\infty}^{+\infty}\phi(t)\{f_{ik}(0)+\sigma_{ik}tf_{ik}(\xi_t)\}\mathrm{d}t - f_{ik}(0)\right| \\ &= \left|\sigma_{ik}\int_{-\infty}^{+\infty}\phi(t)tf_{ik}(\xi_t)\mathrm{d}t\right| \\ &\le \sigma_{ik}\int_{-\infty}^{+\infty}|\phi(t)tf_{ik}(\xi_t)|\mathrm{d}t,\end{aligned}$$

where ξ_t lies between 0 and $\sigma_{ik}t$. By condition A1, $f_{ik}(\cdot)$ is uniformly bounded, hence there exists a constant M satisfying $f_{ik}(\xi_t) \le M$, and by condition A4, we can have

$$\left|\sigma_{ik}^{-1}E\phi\left(\frac{\epsilon_{ik}}{\sigma_{ik}}\right)-f_{ik}(0)\right| \le \sqrt{\frac{2}{\pi}}\sigma_{ik}M \to 0.$$

We therefore can obtain $|N^{-1}\{\tilde{D}_\tau(\beta_0)-D_\tau(\beta_0)\}| \to 0$. The strong law of large number implies $N^{-1}\tilde{D}_\tau(\beta_0) \to E\{N^{-1}\tilde{D}_\tau(\beta_0)\}$. Using the triangle inequality,

$$|N^{-1}\{\tilde{D}_\tau(\beta_0)-D_\tau(\beta_0)\}| \le |N^{-1}\{\tilde{D}_\tau(\beta_0)-E\tilde{D}_\tau(\beta_0)\}| + |N^{-1}\{E\tilde{D}_\tau(\beta_0)-D_\tau(\beta_0)\}| \to o(1).$$

By Taylor series expansion of $\tilde{U}_\tau(\beta)$ around β_0, we have

$$\tilde{U}_\tau(\beta) = \tilde{U}_\tau(\beta_0) + \tilde{D}_\tau(\beta^*)(\beta-\beta_0),$$

where β^* lies between β and β_0. Therefore, $\sqrt{N}(\beta-\beta_0) = -N\tilde{D}_\tau^{-1}(\beta^*)N^{-1/2}\tilde{U}_\tau(\beta_0)$. Let $\beta = \tilde{\beta}$. Because $\tilde{\beta} \to \beta_0$, we therefore obtain $\beta^* \to \beta_0$ and $\tilde{D}_\tau^{-1}(\beta^*) \to \tilde{D}_\tau^{-1}(\beta_0)$. By Theorem 1 and $N^{-1}\{\tilde{D}_\tau(\beta_0)-D_\tau(\beta_0)\} \xrightarrow{p} 0$, we have

$$\sqrt{N}(\tilde{\beta}-\beta_0) = -ND_\tau^{-1}(\beta_0)N^{-1/2}U_\tau(\beta_0) + o_p(1).$$

By the normality of $N^{-1/2}U_\tau(\beta_0)$, we can derive the normality of $\sqrt{N}(\tilde{\beta}-\beta_0)$. □

References

Bassett Jr., G., Koenker, R., 1978. Asymptotic theory of least absolute error regression. Journal of the American Statistics and Association 73, 618–622.
Bilias, Y., Chen, S., Ying, Z., 2000. Simple resampling methods for censored regression quantiles. Journal of Econometrics 99, 373–386.
Brown, B.M., Wang, Y.-G., 2005. Standard errors and covariance matrices for smoothed rank estimators. Biometrika 92, 149–158.
Buchinsky, M., 1995. Estimating the asymptotic covariance matrix for quantile regression models. A Monte Carlo study. Journal of Econometrics 68, 303–338.
Chen, L., Wei, L.J., Parzen, M.I., 2004, Quantile regression for correlated observations. In: Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data, vol. 179, pp. 51–70.
Farcomeni, A., 2012. Quantile regression for longitudinal data based on laent Markov subject-specific parameters. Statistics & Computing 22, 141–152.
Gantz, B.J., Tyler, R.S., Knutson, J.F., Woodworth, G., Abbas, P., McCabe, B.F., Hinrichs, J., Tye-Murray, N., Lansing, C., Kuk, F., et al., 1988. Evaluation of five different cochlear implant designs: audiologic assessment and predictors of performance. Laryngoscope 98, 1100–1106.
Geraci, M., Bottai, M., 2007. Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics 8, 140–154.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Heyde, C.C., 1989. Statistical Data Analysis and Inference. Elsevier, Amsterdam.
Jung, S.H., 1996. Quasi-likelihood for median regression models. Journal of the American Statistical Association 91, 251–257.
Jung, S.H., Ying, Z., 2003. Rank-based regression with repeated measurements data. Biometrika 90, 732–740.
Karlsson, A., 2007. Nonlinear quantile regression estimation of longitudinal data. Communications in Statistics—Simulation and Computation 37, 114–131.
Koenker, R., 2005. Quantile Regression. Cambridge University Press.
Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89.
Koenker, R., Bassett Jr., G., 1978. Regression quantiles. Econometrica 84, 33–50.
Koenker, R., D'Orey, V., 1987. Computing regression quantiles. Applied Statistics 36, 383–393.
Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Liu, Y., Bottai, M., 2009. Mixed-effects models for conditional quantiles with longitudinal data. International Journal of Biostatistics 5, Article 28.
Neuhaus, J.M., Kalbfleish, J.D., 1998. Between- and within-cluster covariate effects in the analysis of clustered data. Biometrics 54, 638–645.
Núñez-Anton, V., Rodríguez-Póo, J.M., Vieu, P., 1999. Longitudinal data with nonstationary errors: a nonparametric three-stage approach. Test 8, 201–231.
Núñez-Anton, V., Woodwoth, G.G., 1994. Analysis of longitudinal data with unequally spaced observations and time-dependent correlated errors. Biometrics 50, 445–456.
Pang, L., Lu, W., Wang, H.J., 2010. Variance estimation in censored quantile regression via induced smoothing. Computational Statistics and Data Analysis doi:10.1016/j.csda.2010.10.018.
Parzen, M.I., Wei, L.J., Ying, Z., 1994. A resampling method based on pivotal estimating functions. Biometrika 81, 341–350.
Pollard, D., 1990. Empirical Processes: Theories and Applications. Institute of Mathematical Statistics, Hayward, California.
Qu, A., Lindsay, B.G., Li, B., 2000. Improving generalised estimating equations using quadratic inference functions. Biometrika 87, 823–836.
Reich, B.J., Bondell, H.D., Wang, H.J., 2010. Flexible Bayesian quantile regression for independent and clustered data. Biostatistics 11, 337–352.
Stoner, J.A., Leroux, B.G., 2002. Analysis of clustered data: a combined estimating equations approach. Biometrika 89, 567–578.
Wang, Y.-G., Shao, Q., Zhu, M., 2009. Quantile regression without the curse of unsmoothness. Computational Statistics and Data Analysis 53, 3696–3705.
Yin, G., Cai, J., 2005. Quantile regression models with multivariate failure time data. Biometrics 61, 151–161.