TOBIT QUANTILE REGRESSION OF LEFT-CENSORED LONGITUDINAL DATA WITH INFORMATIVE OBSERVATION TIMES

Zhanfeng Wang¹, Jieli Ding², Liuquan Sun³ and Yaohua Wu¹

¹University of Science and Technology of China, ²Wuhan University and ³Chinese Academy of Sciences

Abstract: In many longitudinal studies, longitudinal responses are subject to left-censoring and may be correlated with observation times. In this article, we propose a Tobit quantile regression model for the analysis of left-censored longitudinal data with informative observation times and with the longitudinal responses allowed to depend on the past observation history. Estimating equation approaches are developed for parameter estimation, and the resulting estimators are shown to be consistent and asymptotically normal. A modified Majorize-Minimize algorithm is proposed to compute the proposed estimators. Simulation studies show that the proposed estimators perform well. An application to a data set from an AIDS clinical trial study is provided.

Key words and phrases: Bootstrap resampling, estimating equations, informative observation times, left-censored longitudinal data, Tobit quantile regression.

1. Introduction

Longitudinal data arise when subjects are followed up over a period of time. This occurs in such fields as medical follow-up studies, psychology, sociology, and observational investigations. Due to accuracies of measurement tools or mechanism, the longitudinal responses are often subject to a lower detection limit such that some responses are left censored. For example, in an HIV-RNA level study, viral load measurements are often subject to left censoring due to a lower limit of quantification (Hammer et al. (2002)). Other examples include the antibody concentration in blood serum (Moulton and Halsey (1995)) and the concentration of a pollutant in the environment (Singh and Nocerino (2002)). Such data are referred to as left-censored longitudinal data (Jacqmin-Gadda et al. (2000), Wang and Fygenson (2009)).

Several methods have been developed for analyzing left-censored longitudinal
data (Paxton et al. (1997); Hughes (1999); Wang and Fygenson (2009); Kobayashi and Kozumi (2012); Xiao et al. (2014)). For example, Paxton et al. (1997) proposed a multiple imputation approach and each censored residual was replaced by a random number drawn from the appropriate truncated normal distribution. Hughes (1999) and Jacqmin-Gadda et al. (2000) used likelihood-based methods while assuming a Gaussian distribution for both random effects and random errors (also see Lyles, Lyles and Taylor (2000); Thiébaut and Jacqmin-Gadda (2004)). Under no distributional assumption, Wang and Fygenson (2009) and Xiao et al. (2014) suggested a rank score test and a randomly weighting test for censored quantile regression models, respectively. Kobayashi and Kozumi (2012) developed Bayesian approaches for analyzing quantile regression models.

For these modeling approaches, a strong assumption is that the longitudinal responses and the observation times are independent given covariates.

In many applications, however, observation times are informative about the longitudinal responses. For example, the observation times may be hospitalization times of subjects, which are response variable-dependent in the study (Lin, Scharfstein and Rosenheck (2004); Sun et al. (2005)). Some methods have been proposed for situations where the longitudinal responses and the observation times are related (Sun, Sun and Liu (2007); Liu, Huang and O’Quigley (2008); Liang, Lu and Ying (2009); Sun et al. (2012); Chen, Tang and Zhou (2016)). For example, Lin, Scharfstein and Rosenheck (2004) proposed a class of inverse intensity-of-visit process-weighted estimators for a typical marginal regression model. Sun et al. (2005) suggested a conditional model where the longitudinal responses are assumed to depend on the past observation history. Sun, Sun and Liu (2007) and Liang, Lu and Ying (2009) presented some joint models for the longitudinal responses and the observation times via latent variables. Recently, Chen, Tang and Zhou (2016) considered a quantile regression method when the response variable depends on the past observation history. These methods primarily analyze longitudinal data with informative observation times in the absence of left censoring. To the best of our knowledge, there is no existing work considering the joint analysis of left-censored longitudinal data with informative observation times. The method of Chen, Tang and Zhou (2016) cannot be extended in a straightforward manner to deal with left-censored longitudinal data because there is a need for considering left-censored responses. In general, discarding censored measurements or ignoring them leads to biased inferences.

In this article, we propose joint modeling of left-censored longitudinal data with informative observation times. A Tobit quantile regression model is used for
the longitudinal responses with left censoring, and a nonhomogeneous Poisson
process is used for the observation times. The longitudinal responses are allowed
to depend on the past observation history. Estimating equation approaches are
developed for parameter estimation, and the resulting estimators are shown to
be consistent and asymptotically normal. The proposed objective function is nei-
ther differentiable nor convex due to left-censored responses, and the theoretical
and computational developments are challenging. A modified Majorize-Minimize
(MM) algorithm is used to handle the computational difficulty. The algorithm is
different from that of Chen, Tang and Zhou (2016).

The remainder of the paper is as follows. Section 2 describes joint models for
the longitudinal responses and the observation times. Section 3 proposes estimat-
ing procedures for regression parameters of interest. The asymptotic properties
of the proposed estimators are established, and the MM algorithm is presented.
Section 4 reports some results from simulation studies for evaluating the pro-
posed methods. An application to a HIV-1 RNA data set from an AIDS clinical
trial is provided in Section 5, and some concluding remarks are made in Section
6. Proofs are relegated to the Appendix.

2. Model Specification

Consider a longitudinal study involving \( n \) independent subjects. For the \( i \)th
subject, let \( Y_i^*(t) \) be the underlying response variable at time \( t \), and \( X_i(t) \) be the
\( p \)-dimensional vector of possibly time-dependent covariates. Due to limitations
of accuracy of measurement tool or mechanism, the response \( Y_i^*(t) \) is subject
to a lower bound \( d \). Without loss of generality, we assume that \( d = 0 \). Let
\( Y_i(t) = \max\{Y_i^*(t), 0\} \), and \( C_i \) be the follow-up or censoring time. If \( N_i^*(t) \) is the
counting process denoting the number of the observation times before or at time \( t \),
the process \( Y_i(t) \) is only observed at the jump points of \( N_i(t) = N_i^*(\min\{t, C_i\}) \).
The covariate histories \( \{X_i(t) : 0 \leq t \leq C_i\} \) \( (i = 1, ..., n) \) are assumed to be observed.

Define \( \mathcal{F}_{it} = \{N_i(u), 0 \leq u < t\} \). We start with the marginal regression
model

\[ Y_i^*(t) = \beta'X_i(t) + \alpha'\mathbf{H}(\mathcal{F}_{it}) + e_i(t), \quad i = 1, ..., n, \tag{2.1} \]

where \( \beta \) and \( \alpha \) are vectors of unknown regression parameters with dimensions \( p \)
and \( q \), respectively, \( \mathbf{H}(\cdot) \) is a vector of known functions on the counting process
\( N_i(t) \) up to time \( t- \), and \( e_i(t) \) is a measurement error process (e.g., Sun et al.
(2005)). We cannot observe the potential response \( Y_i^*(t) \) because of its nonneg-
ative constraint. Hence in view of (2.1), following Wang and Fygenson (2009) and for a given $0 < \tau < 1$, we consider the marginal regression model

$$Y_i(t) = \max\{0, \beta'X_i(t) + \alpha'H(F_i) + e_i(t)\}, \quad i = 1, \ldots, n,$$

(2.2)

where the $\tau$th quantile of $e_i(t)$ is assumed to be zero. Here, (2.2) is referred to as the Tobit quantile regression model for the longitudinal responses with left censoring (Wang and Fygenson (2009); Xiao et al. (2014)).

For the observation process, we assume that conditioning on $X_i(t)$, $N_i^*(t)$ is a nonhomogeneous Poisson process with

$$E(dN_i^*(t)|X_i(t)) = \exp(\gamma'X_i(t))d\Lambda_0(t), \quad i = 1, 2, \ldots, n,$$

(2.3)

where $\gamma$ is a vector of unknown regression parameters, and $\Lambda_0(t)$ is an arbitrary nondecreasing function (e.g., Lin and Ying (2001); Sun et al. (2005)). For convenience, models (2.2) and (2.3) assume the same set of covariates $X(t)$. The proposed estimation procedure can be extended in a straightforward manner to deal with different sets of covariates.

In contrast with the common Tobit quantile regression models with longitudinal data, a main feature of (2.2) is that it allows the response process $Y_i(t)$ to be correlated with the observation process $N_i^*(t)$ and, in particular, in a linear fashion through the function $H$. When $\alpha = 0$, (2.2) reduces to the models studied by Wang and Fygenson (2009) and Xiao et al. (2014) for the case that the observation process has no information on the response process.

In (2), the function $H$ can be chosen according to practical matters. As in Sun et al. (2005), a natural choice for $H$ can be $H(F_i) = N_i(t-)$, which implies that all information about $Y_i(t)$ in $F_i$ is given by the total number of observations. Another choice is $H(F_i) = N_i(t-) - N_i(t-s)$, which indicates that $Y_i(t)$ depends on $F_i$ only through the number of observations in the last $s$ time units. The function $H$ also can be defined by some combination of the foregoing choices if both the total and recent numbers of observations include information about the response $Y_i(t)$. In what follows, we assume that conditional on the covariate $X_i(t)$, the censoring time $C_i$ is independent of $Y_i(t)$ and $N_i^*(t)$.

3. Estimation Procedures

3.1. Procedures

We first consider the special case of $\gamma = 0$, which implies that the observation times are independent of the covariate $X_i(t)$. Here a common method to estimate $\theta = (\beta', \alpha')'$ is to minimize the objection function
where \( L \) is the maximum follow-up time, \( a^+ = \max\{0, a\} \), \( \mathbf{Z}_i(u) = (\mathbf{X}_i'(u), \mathbf{H}'(\mathcal{F}_{ti}))' \), and \( \rho_r(v) = v\{v - I(v < 0)\} \) is the quantile loss function (e.g., [Wang and Fygenson (2009)]). The subgradient of \( \rho_r\{Y_i(u) - (\theta'\mathbf{Z}_i(u))^+\}dN_i(u) \) with respect to \( \theta \) is

\[
\mathbf{Z}_i(u)I\{\theta'\mathbf{Z}_i(u) > 0\}\left[I\{Y_i(u) - \theta'\mathbf{Z}_i(u) < 0\} - \tau\right]dN_i(u).
\]

Thus, minimizing (3.1) is equivalent to solving the estimating equation

\[
\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{L} \mathbf{Z}_i(u)I\{\theta'\mathbf{Z}_i(u) > 0\}\left[I\{Y_i(u) - \theta'\mathbf{Z}_i(u) < 0\} - \tau\right]dN_i(u) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{L} \mathbf{Z}_i(u)I\{\theta'\mathbf{Z}_i(u) > 0\}\left[I\{Y_i(u) - \theta'\mathbf{Z}_i(u) < 0\} - \tau\right]dN_i(u) = o_p(n^{-1/2}),
\]

which can be obtained from the derivative of (3.1) with respect to \( \theta \).

We now consider the case in which the observation times depend on the covariates through model (2.3). Motivated by the above, we take

\[
M_i(t; \theta, \gamma, \Lambda_0) = \int_{0}^{t} I\{\theta'\mathbf{Z}_i(u) > 0\}\left[I\{Y_i(u) - \theta'\mathbf{Z}_i(u) < 0\} - \tau\right]\xi_i(u)\exp(\gamma\mathbf{X}_i(u))d\Lambda_0(u),
\]

where \( \xi_i(t) = I(C_i \geq t) \). Let \( \theta_0 \) and \( \gamma_0 \) be true values of \( \theta \) and \( \gamma \), respectively. Under models (2.2) and (2.3) and the assumptions,

\[
E\left[I\{\theta_0'\mathbf{Z}_i(u) > 0\}I\{Y_i(u) - \theta_0'\mathbf{Z}_i(u) < 0\}dN_i(u)|\mathbf{Z}_i(u), C_i\right]
\]

\[
= \tau\xi_i(u)\exp(\gamma_0'\mathbf{X}_i(u))d\Lambda_0(u).
\]

Then it can be checked that

\[
E\{M_i(t; \theta_0, \gamma_0, \Lambda_0)|\mathbf{Z}_i(u)\} = 0,
\]

which implies that \( M_i(t; \theta_0, \gamma_0, \Lambda_0) \) is zero-mean stochastic process. Thus, for given \( \gamma_0 \) and \( \Lambda_0 \), we can estimate \( \theta_0 \) using the estimating equation

\[
U(\theta; \gamma_0, \Lambda_0) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{L} \mathbf{Z}_i(u) dM_i(u; \theta, \gamma_0, \Lambda_0) = 0.
\]

In reality, \( \gamma_0 \) and \( \Lambda_0 \) are unknown. Using the approach of [Lin et al. (2000)], we propose an estimating equation for \( \gamma_0 \):

\[
\sum_{i=1}^{n} \int_{0}^{L} \{\mathbf{X}_i(u) - \bar{\mathbf{X}}(u; \gamma)\}dN_i(u) = 0,
\]

where \( \bar{\mathbf{X}}(u; \gamma) = \mathbf{S}^{(1)}(u; \gamma)/\mathbf{S}^{(0)}(u; \gamma) \), and
where $a^\otimes 2 = aa'$ for any vector $a$. Let $\hat{\gamma}$ denote the solution to this estimating equation. Then $\Lambda_0(t)$ can be consistently estimated by the Aalen-Breslow-type estimator

$$\hat{\Lambda}_0(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{S^{(0)}(u; \hat{\gamma})}.$$ 

By replacing $\gamma_0$ and $\Lambda_0(t)$ with $\hat{\gamma}$ and $\hat{\Lambda}_0(t)$ in (3.2), we specify an estimating function for $\theta_0$:

$$U(\theta; \hat{\gamma}, \hat{\Lambda}_0) = \frac{1}{n} \sum_{i=1}^{n} \int_0^L Z_i(u) dM_i(u; \theta, \hat{\gamma}, \hat{\Lambda}_0).$$

Since $U(\theta; \hat{\gamma}, \hat{\Lambda}_0)$ is a discontinuous function of $\theta$, we define the estimator $\hat{\theta}$ as a zero-crossing of $U(\theta; \hat{\gamma}, \hat{\Lambda}_0)$ or as a minimizer of $\|U(\theta; \hat{\gamma}, \hat{\Lambda}_0)\|$, where $\|a\| = (a'a)^{1/2}$.

3.2. Implementation

Due to the complicated nature of $U(\theta; \hat{\gamma}, \hat{\Lambda}_0)$, it is apparently not possible to obtain $\hat{\theta}$ directly. To overcome this difficulty, we propose a modified MM algorithm (e.g., Ortega and Rheinboldt (1970, P. 253); Lange, Hunter and Yang (2000)). Let

$$q_i(\theta) = \int_0^L Z_i(u) I\{\theta'Z_i(u) > 0\} \left[ I\{Y_i(u) - \theta'Z_i(u) < 0\} - \tau \right] dN_i(u),$$

$$G(\theta; \hat{\gamma}) = \frac{\tau}{n} \sum_{i=1}^{n} \int_0^L \left[ Z_i(u) I\{\theta'Z_i(u) > 0\} - \frac{\tilde{S}^{(1)}(u; \theta, \hat{\gamma})}{S^{(0)}(u; \hat{\gamma})} \right] dN_i(u),$$

and $\partial Q(\theta)/\partial \theta = n^{-1} \sum_{i=1}^{n} q_i(\theta)$, where

$$\tilde{S}^{(1)}(u; \theta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} \xi_i(u) Z_i(u) I\{\theta'Z_i(u) > 0\} \exp(\gamma'X_i(u)).$$

Then it can be checked that

$$U(\theta; \hat{\gamma}, \hat{\Lambda}_0) = \frac{\partial Q(\theta)}{\partial \theta} + G(\theta; \hat{\gamma}).$$

Let $\theta_k$ be the $k$th iteration estimate of $\theta$. Following Hunter and Lange (2000), the objection function $Q(\theta)$ can be replaced with a Majorizing function with a disturbance constant $\varepsilon > 0$,
\[
\tilde{Q}_\varepsilon(\theta|\theta^k) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{L} \frac{1}{4} \left\{ \frac{r_i(\theta)^2}{\varepsilon + |r_i(\theta^k)|} + (4\tau - 2)r_i(\theta) + c_i \right\} dN_i(u),
\]
where \( r_i(\theta) = Y_i(u) - (\theta'Z_i(u))^+ \). Since \((\theta'Z_i(u))^+\) is not continuous with respect to \( \theta \), the function \( \tilde{Q}_\varepsilon(\theta|\theta^k) \) is still not smooth at \( \theta \). To this end, we use a smoothing function to approximate \( t^+ \)
\[
K_h(t) = \frac{t}{1 + \exp(-t/h)},
\]
where \( h > 0 \) is a known smoothing parameter. Here, for small enough \( h \), \( K_h(t) \approx t^+ \). By substituting \( \tilde{r}_i(\theta) = Y_i(u) - K_h(\theta'Z_i(u)) \) for \( r_i(\theta) \) in \( \tilde{Q}_\varepsilon(\theta|\theta^k) \), we obtain a smooth Majorizing function
\[
Q_\varepsilon(\theta|\theta^k) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{L} \frac{1}{4} \left\{ \frac{\tilde{r}_i(\theta)^2}{\varepsilon + |\tilde{r}_i(\theta^k)|} + (4\tau - 2)\tilde{r}_i(\theta) + c_i \right\} dN_i(u),
\]
where \( c_i \) satisfies
\[
\frac{1}{4} \left\{ \frac{\tilde{r}_i(\theta^k)^2}{\varepsilon + |\tilde{r}_i(\theta^k)|} + (4\tau - 2)\tilde{r}_i(\theta^k) + c_i \right\} = \rho_r(r_i(\theta^k)) - \frac{\varepsilon}{2} \log(\varepsilon + |r_i(\theta^k)|).
\]

Based on the smoothing objection function \( Q_\varepsilon(\theta|\theta^k) \), the \((k + 1)\)-step estimating function can be constructed as
\[
\tilde{U}(\theta; \hat{\gamma}, \hat{\Lambda}_0|\theta^k) = \frac{\partial Q_\varepsilon(\theta|\theta^k)}{\partial \theta} + G(\theta^k; \hat{\gamma}). \tag{3.5}
\]
Then the \((k + 1)\)th iteration estimate \( \theta^{k+1} \) is obtained by solving \( \tilde{U}(\theta; \hat{\gamma}, \hat{\Lambda}_0|\theta^k) = 0 \). Given an initial estimate \( \theta^0 \) and \( \varepsilon > 0 \), this iteration is continued until convergence and the estimate \( \hat{\theta} \) is obtained at convergence. For the convergence, several criteria can be applied; in our numerical studies, we used the absolute differences between the iterative estimates of the parameters.

**Remark 1.** In the above MM algorithm, we propose to use a one-step Newton-Raphson method to calculate \( \theta^{k+1} \) as \( \theta^{k+1} = \theta^k - \Delta^{k+1} \), where
\[
\Delta^{k+1} = \left( \frac{\partial^2 Q_\varepsilon(\theta|\theta^k)}{\partial \theta \partial \theta'} \right)_{\theta=\theta^k}^{-1} \tilde{U}(\theta^k; \hat{\gamma}, \hat{\Lambda}_0|\theta^k).
\]

### 3.3. Properties

To establish the asymptotic properties of \( \hat{\theta} \), let
\[
\tilde{S}^{(2)}(u; \theta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} \xi_i(u)Z_i(u)X_i(u)'I\{\theta'Z_i(u) > 0\} \exp(\gamma'X_i(u)),
\]
and \( \tilde{Z}(u; \gamma, \theta) = \tilde{S}^{(1)}(u; \theta, \gamma)/\tilde{S}^{(0)}(u; \gamma) \). Let \( \hat{s}^{(1)}(u), \hat{s}^{(2)}(u), \) and \( s^{(k)}(u) \) denote the limits of \( \tilde{S}^{(1)}(u; \theta_0, \gamma_0), \tilde{S}^{(2)}(u; \theta_0, \gamma_0), \) and \( S^{(k)}(u; \gamma_0) \) \( k = 0, 1, 2 \), respect-
tively. Also, let \( \tilde{z}(u) = \bar{s}^{(1)}(u)/s^{(0)}(u) \) and \( \bar{x}(u) = s^{(1)}(u)/s^{(0)}(u) \). Define

\[
M^*_i(t) = N_i(t) - \int_0^t \xi_i(u) \exp(\gamma_i X_i(u)) d\Lambda_0(u),
\]

\[
A = E\left[\int_0^\infty \left\{ \frac{s^{(2)}(u)}{s^{(0)}(u)} - \bar{x}(u) \right\} dN_i(u)\right],
\]

\[
B = E\left[\int_0^\infty \left\{ \frac{\tilde{s}^{(2)}(u)}{s^{(0)}(u)} - \frac{\tilde{s}^{(1)}(u)s^{(1)}(u)'}{s^{(0)}(u)^2} \right\} dN_i(u)\right],
\]

\[
D = E\left[\int_0^\infty \tilde{Z}_i(u)Z_i(u)' I\{\theta_i Z_i(u) > 0\} f_{Y^*(u)}(X_i(u), F_{i*}) dN_i(u)\right],
\]

where \( f_{Y^*(u)}(X_i(u), F_{i*}) \) is the density function of \( Y^*(u) \) conditional on \( \{X_i(u), F_{i*}\} \) for \( u \in [0, L] \).

The proof of the following is given in the Appendix.

**Theorem 1.** Under the regularity conditions (C1)-(C5) stated in the Appendix, \( \hat{\theta} \) is consistent, and \( n^{1/2}(\hat{\theta} - \theta_0) \) has an asymptotic normal distribution with mean zero and covariance matrix \( \Sigma = D^{-1}VD^{-1} \), where

\[
V = E\left[q_i(\theta_0) + \tau \int_0^\infty \left\{ \tilde{Z}_i(u) I\{\theta_i Z_i(u) > 0\} - \bar{z}(u) \right\} dM^*_i(u) \right.
\]

\[
- \tau BA^{-1} \int_0^\infty \left\{ X_i(u) - \bar{x}(u) \right\} dM^*_i(u) \] \( \otimes^2 \).

The asymptotic variance of \( \hat{\theta} \) can be consistently estimated by \( \hat{\Sigma} \), which can be obtained by the usual plug-in method. Note that \( \Sigma \) is of complicated form involving some nuisance parameters such as the conditional density function of \( Y^*(u) \). Thus, it is difficult to estimate \( \Sigma \) directly. Here, we propose to use the bootstrap method to estimate the asymptotic variance of \( \hat{\theta} \). In our simulation studies with the sample size \( n = 100 \), we used 200 bootstrap samples and found the variance estimation to be fairly accurate.

4. Simulation Studies

Simulation studies were conducted to examine the finite sample performance of the proposed estimators. In the study, we let \( X_i = (X_{i1}, X_{i2})' \), where \( X_{i1} \) was generated from the standard normal distribution, and \( X_{i2} \) was from a Bernoulli distribution with success probability 0.5. For given \( X_i \), the observation times were generated from a nonhomogeneous Poisson process with

\[
E(dN^*_i(t)|X_i) = \exp(\gamma_1 X_{i1} + \gamma_2 X_{i2}) d\Lambda_0(t),
\]

where \( \Lambda_0 = 0.5t, \gamma_1 = -0.5 \) and \( \gamma_2 = 1 \). The censoring time was generated
from a uniform distribution on \((\kappa/2, \kappa\)) with \(\kappa = 2\) or 4 representing the largest follow-up time.

For the response variable, we assumed that \(Y_i(t)\) was given by the Tobit quantile regression model

\[
Y_i(t) = \max\{0, \beta_1 X_{i1} + \beta_2 X_{i2} + \alpha N_i(t-) + (1 + \pi X_{i1}) e_i(t)\},
\]

where \(\beta_1 = -1\), \(\beta_2 = 1\), and \(\alpha = 0, 0.25,\) or 0.5. Here \(\alpha\) reflects the dependence between the response variable and the observation times. While \(\alpha = 0\) has the response variable and the observation times independent, \(\alpha \neq 0\) has the two processes with nonzero correlations. For \(\pi\) and \(e_i(t)\), we considered three cases. S1. \(\pi = 0\), \(e_i(t)\) was generated from the standard normal; S2. \(\pi = 1\), \(e_i(t)\) was generated from the standard normal; S3. \(\pi = 0\), \(e_i(t)\) was generated from the standard Cauchy.

For these three cases, the median of \(e_i(t)\) was zero, \(\tau = 0.5\). Our results are based on 1,000 replications with sample size \(n=100\), and final estimates were reached when the absolute difference of the estimates between two successive iterations was less than \(10^{-5}\). The asymptotic variance was estimated using the bootstrap method with 200 bootstrap samples. We found this to be adequate.

Tables 1 and 2 present the simulation results on the estimates of \(\theta = (\beta_1, \beta_2, \alpha)'\) for \(\kappa = 2\) and 4, respectively. In these tables, TQR stands for the proposed method, Bias is the sample mean of the estimate minus the true value, SE is the sampling standard error of the estimate, SEE is the sample mean of the standard error estimate, and CP is the empirical coverage probability of the 95% confidence interval based on the normal approximation. It can be seen from Tables 1 and 2 that the proposed method performed well for the situations considered here: the proposed estimators are virtually unbiased, the standard error estimates are close to the sampling standard errors, and the 95% empirical coverage probabilities are reasonable. The results are better with increasing \(\kappa\) or increasing the numbers of observations. In addition, the proposed estimators are robust to the cases of a heteroscedastic error distribution (S2) and a heavy-tailed error distribution (S3).

For comparison, we considered the method of Wang and Fygenson (2009) (denoted by WF). They studied model (2) with \(\alpha = 0\). Under the same setups, the comparison results on estimation of \(\beta_1\) and \(\beta_2\) are given in Tables 1 and 2. When the observation times are noninformative, the WF estimators are unbiased. Here, the methods provide reasonable and comparable estimates. The variances of our method are only slightly larger than those of WF, because they utilize the assumption of the independent observation times in their estimation. When
<table>
<thead>
<tr>
<th>Case</th>
<th>Method</th>
<th>Bias</th>
<th>SE</th>
<th>SEE</th>
<th>CP</th>
<th>Bias</th>
<th>SE</th>
<th>SEE</th>
<th>CP</th>
<th>Bias</th>
<th>SE</th>
<th>SEE</th>
<th>CP</th>
<th>Bias</th>
<th>SE</th>
<th>SEE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>S1</td>
<td>TQR</td>
<td>0.0321 0.0763 0.0677 0.0722 0.0543 0.0573 0.0677 0.0722 0.0543 0.0573</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>TQR</td>
<td>0.0021 0.0672 0.0677 0.0677 0.0204 0.0204 0.0204 0.0204 0.0204 0.0204</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>WF</td>
<td>0.0000 0.0687 0.0672 0.0672 0.0392 0.0392 0.0392 0.0392 0.0392 0.0392</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>S1</td>
<td>TQR</td>
<td>0.0023 0.0667 0.0674 0.0573 0.0204 0.0204 0.0204 0.0204 0.0204 0.0204</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>TQR</td>
<td>0.0331 0.1007 0.1007 0.1007 0.0384 0.0384 0.0384 0.0384 0.0384 0.0384</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>WF</td>
<td>0.0000 0.0667 0.0672 0.0672 0.0392 0.0392 0.0392 0.0392 0.0392 0.0392</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>S1</td>
<td>TQR</td>
<td>0.0023 0.0667 0.0674 0.0573 0.0204 0.0204 0.0204 0.0204 0.0204 0.0204</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>TQR</td>
<td>0.0331 0.1007 0.1007 0.1007 0.0384 0.0384 0.0384 0.0384 0.0384 0.0384</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>WF</td>
<td>0.0000 0.0667 0.0672 0.0672 0.0392 0.0392 0.0392 0.0392 0.0392 0.0392</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Simulation results for the estimation of $\theta$ with $k = 2$. 
Table 2. Simulation results for the estimation of $\theta$ with $\kappa = 4$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Case</th>
<th>Method</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bias</td>
<td>SE</td>
<td>SEE</td>
</tr>
<tr>
<td>0</td>
<td>S1</td>
<td>TQR</td>
<td>0.0058</td>
<td>0.0573</td>
<td>0.0557</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>-0.0026</td>
<td>0.0486</td>
<td>0.0474</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>TQR</td>
<td>0.0080</td>
<td>0.0464</td>
<td>0.0477</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>0.0008</td>
<td>0.0443</td>
<td>0.0456</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>TQR</td>
<td>0.0125</td>
<td>0.0698</td>
<td>0.0706</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>0.0006</td>
<td>0.0589</td>
<td>0.0590</td>
</tr>
<tr>
<td>0.25</td>
<td>S1</td>
<td>TQR</td>
<td>0.0092</td>
<td>0.0543</td>
<td>0.0531</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>-0.6434</td>
<td>0.3827</td>
<td>0.1356</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>TQR</td>
<td>0.0065</td>
<td>0.0465</td>
<td>0.0476</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>-0.6216</td>
<td>0.1983</td>
<td>0.1559</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>TQR</td>
<td>0.0078</td>
<td>0.0678</td>
<td>0.0685</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>-0.6945</td>
<td>0.1902</td>
<td>0.1615</td>
</tr>
<tr>
<td>0.5</td>
<td>S1</td>
<td>TQR</td>
<td>0.0084</td>
<td>0.0529</td>
<td>0.0529</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>-1.2499</td>
<td>0.3164</td>
<td>0.2495</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>TQR</td>
<td>0.0057</td>
<td>0.0475</td>
<td>0.0475</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>-1.2386</td>
<td>0.3440</td>
<td>0.2746</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>TQR</td>
<td>0.0146</td>
<td>0.0649</td>
<td>0.0675</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WF</td>
<td>-1.3325</td>
<td>0.3414</td>
<td>0.2825</td>
</tr>
</tbody>
</table>
Table 3. Simulation results for the estimation of $\theta$ with $\kappa = 4$ for $\tau = 0.25$ and 0.75.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\alpha$</th>
<th>S4</th>
<th>S5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0 $\beta_1$</td>
<td>0.0027</td>
<td>0.0567</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>-0.0071</td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.0009</td>
<td>0.0087</td>
</tr>
<tr>
<td>0.75</td>
<td>0 $\beta_1$</td>
<td>0.0120</td>
<td>0.1211</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>-0.0101</td>
<td>0.0849</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.0010</td>
<td>0.0183</td>
</tr>
<tr>
<td>0.25</td>
<td>0 $\beta_1$</td>
<td>0.0289</td>
<td>0.1365</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>-0.0038</td>
<td>0.1018</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.0085</td>
<td>0.0269</td>
</tr>
</tbody>
</table>

the independent assumption is violated, WF can lead to large bias and yield improper coverage probabilities.

We conducted simulation studies to examine the performance of the proposed method for the cases of $\tau = 0.25$ and 0.75, with $\kappa = 4$ and $\alpha = 0$ or 0.25. For $\tau = 0.25$, we considered two situations for the error term: S4. $e_i(t)$ normal with mean $-\Phi^{-1}(0.25)$ and variance 1; S5. $e_i(t)$ a Cauchy with location parameter 1 and scale parameter 1. For $\tau = 0.75$, we considered two situations: S6. $e_i(t)$ normal with mean $-\Phi^{-1}(0.75)$ and variance 1; S7. $e_i(t)$ Cauchy with location parameter $-1$ and scale parameter 1. Here, different mean and location parameters guarantee that the $\tau$th quantile of $e_i(t)$ is zero. For all these situations, we took $\pi = 0$, and all other setups were the same as before.

The simulation results are summarized in Table 3. It can be seen there that the proposed method still performed reasonably well for the cases of $\tau = 0.25$ and 0.75: the proposed estimators have small biases, reasonable variance estimates, and the empirical coverage probabilities. We also considered other setups and obtained similar results.

We conducted some simulation studies to compare the proposed method with the naive method that replaces the left censored responses by half of the detection limit and applied the method of Chen, Tang and Zhou (2016) to the imputation data. We considered the case S2 in Table 1, where $X_{i1}$ was standard normal and $X_{i2}$ was normal with mean 0.2 and variance 1. The results for other cases are...
Table 4. Comparison results on the estimation of $\theta$ for our method and a naïve method with $\kappa = 2$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>TQR</th>
<th>Naïve</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>Bias</td>
</tr>
<tr>
<td>0</td>
<td>0.0244</td>
<td>0.0854</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>-0.0843</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.0438</td>
</tr>
<tr>
<td>0.25</td>
<td>$\beta_1$</td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>-0.0876</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.0524</td>
</tr>
<tr>
<td>0.50</td>
<td>$\beta_1$</td>
<td>0.0359</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>-0.0931</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.0518</td>
</tr>
</tbody>
</table>

5. An Application

We applied the proposed method to the HIV-RNA level data from an AIDS clinical trial study (Hammer et al. (2002); Sun and Wu (2005); Wang and Fygeson (2009)). In this study, some subjects received a single protease inhibitor (PI), while others received a double-PI antiretroviral regimens in treating HIV-infected patients. HIV-1 RNA levels in plasma (viral load) was measured repeatedly during the follow-up. The scheduled visits for the measurements were at weeks 0, 2, 4, 8, 16, and 24. However, the actual visit times varied around the scheduled visiting times, and the numbers of measurements differed, which indicates that the observation times might be informative about the viral load. A total of 481 patients was enrolled in the study. The numbers of patients with 1 to 6 visits were 10, 11, 13, 29, 69, and 349, respectively. Due to technical limitations, about 22% of measurements were censored from below at 200 copies/ml. Some patients had prior antiviral treatment with non-nucleoside analogue reverse transcriptase inhibitors (NNRTI) and others did not have prior NNRTI treatment. The prior treatment experience was considered to be a factor that would affect the antiviral response to the antiretroviral regimens. We focused on the effects of the prior NNRTI treatment and the PI treatment on the HIV viral load response.

Following Sun and Wu (2005) and Sun, Sun and Zhou (2013), we used a
The bootstrap method with 200 bootstrap samples. The analysis results of profile of the treatments’ effects. The asymptotic variance was estimated using the single-PI treatment.

The results show that the patients with the prior NNRTI treatment tend to have higher viral load than those without the prior NNRTI treatment, and the double-PI treatment, and the protease inhibitor treatment have significant effects on the HIV viral load at several \( \tau \)'s. In particular, the patients with the prior NNRTI treatment have higher viral load than those without the prior NNRTI treatment, and the double-

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>TQR</th>
<th>WF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>1.1163 (0.0814)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>1.0239 (0.1192)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-2.0876 (0.3753)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.0993 (0.0490)</td>
<td>0.0427</td>
</tr>
</tbody>
</table>

Sun and Wu (2005) showed that the effect of \( X^*_{2i} \) on the viral load response was linear over time, we took \( X^*_{2i}(t) = X^*_{2i}t \). Let \( Y_i(t) \) be the observed \( \log_{10} \) (viral load response). We considered the Tobit quantile regression model

\[
Y_i(t) = \max\{\log_{10}(200), \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}(t) + \alpha N_i(t-) + \epsilon_i(t)\}, \tag{5.1}
\]

and allowed that the actual visit times can be described by model \([2.3]\) with \( X_i(t) = (X_{i1}, X_{i2}(t))' \) and \( \gamma = (\gamma_1, \gamma_2)' \). For \([2.3]\), we got \( \hat{\gamma}_1 = 0.0327 \) and \( \hat{\gamma}_2 = -0.0683 \), with estimated standard errors of 0.0160 and 0.0450, respectively. The results show that the patients with the prior NNRTI treatment tend to have more visit times than those without the NNRTI treatment, whereas the patients with the double-PI treatment are likely to have less visit times than those with the single-PI treatment.

We fitted \([5.1]\) at different quantiles with \( \tau \) at 0.5, 0.6 and 0.7 to obtain a profile of the treatments’ effects. The asymptotic variance was estimated using the bootstrap method with 200 bootstrap samples. The analysis results of \( \beta \) and \( \alpha \) are summarized in Table 5. The results imply that the prior NNRTI treatment and the protease inhibitor treatment have significant effects on the HIV viral load at several \( \tau \)'s. In particular, the patients with the prior NNRTI treatment have higher viral load than those without the prior NNRTI treatment, and the double-

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>TQR</th>
<th>WF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>1.3991 (0.0784)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.9073 (0.1413)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-1.8635 (0.3696)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.2038 (0.0346)</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>TQR</th>
<th>WF</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>1.7414 (0.0799)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.5888 (0.1127)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-2.0445 (0.3219)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.3295 (0.0361)</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
Table 6. Sensitivity analysis to the choice of $H(F_{it})$ for the HIV-1 RNA data with $\tau = 0.7$.

<table>
<thead>
<tr>
<th>$H(F_{it})$</th>
<th>Est</th>
<th>SEE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log{N_i(t-) + 1}$</td>
<td>$\beta_0$</td>
<td>1.8936</td>
<td>0.0825</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>0.6682</td>
<td>0.1219</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>-1.2867</td>
<td>0.2491</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.3750</td>
<td>0.0797</td>
</tr>
<tr>
<td>$N_i(t-)^{1/2}$</td>
<td>$\beta_0$</td>
<td>1.9404</td>
<td>0.0837</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>0.6619</td>
<td>0.1435</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>-1.1303</td>
<td>0.2662</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>0.2200</td>
<td>0.0611</td>
</tr>
</tbody>
</table>

PI treatment can reduce HIV viral load compared to the single-PI treatment. All $\hat{\alpha}$’s are positive for three different $\tau$ values, indicating that the HIV viral load response and the actual visit times are positively correlated.

For comparison, Table 5 gives the results of the WF method assuming the observation times noninformative. It can be seen that the WF estimators are nearly the same as our proposed estimators for $\beta_1$ at three quantile levels. However, the WF method overestimate slightly $\beta_0$ and underestimates significantly $\beta_2$, especially at the quantile level $\tau = 0.6$ and $0.7$. Thus, after adjusting for informative visit times, the estimate for the double-PI treatment effect becomes much larger.

We performed a sensitivity analysis to the choice of the function $H(F_{it})$ for the data, and replaced $N_i(t-)$ at (5.1) with $\log\{N_i(t-) + 1\}$ or $N_i(t-)^{1/2}$. The other setups were the same as in Table 5 with $\tau = 0.7$. The results are reported in Table 6. Tables 5 and 6 show results that are similar for the three choices of $H(F_{it})$, and that the conclusions are consistent.

6. Concluding Remarks

In this article, we proposed a Tobit quantile regression model for the analysis of left-censored longitudinal data in the presence of informative observation times, where the longitudinal responses are allowed to depend on the past observation history. Estimating equation approaches were proposed to obtain consistent and asymptotically normal estimators, and the Majorize-Minimize algorithm was used to compute the proposed estimators. The simulation results suggest that the proposed estimation approach performs well, and is robust to the cases of the heteroscedastic error and heavy-tailed error distributions. An application
to the HIV-1 RNA data set from an AIDS clinical trial has been provided to illustrate our method.

We have assumed that the observation process is a nonhomogeneous Poisson process. It would be interesting to extend the proposed estimation procedure to deal with a counting process. In addition, we have used the multiplicative intensity (Cox) model for the observation process. Other competing models, such as the additive intensity (Aalen) model, the accelerated failure time model (Lin, Wei and Ying (1998)) and the semiparametric transformation models (Zeng and Lin (2006)), may be used as well. It would be worthwhile to investigate the potential bias due to misspecification for each of these models.

The proposed method relies on the assumption that the censoring time is independent of both the longitudinal responses and the observation times conditional on the covariates. In some applications, this assumption may be violated, especially when censoring is caused by informative dropouts such as death (Wang, Qin and Chiang (2001)). Some methods have been developed for the analysis of longitudinal data in the presence of informative observation and censoring times (e.g., Sun et al. (2012)). It would be useful to extend the existing and proposed methods to analyze left-censored longitudinal data with informative observation and censoring times. This is a challenging problem and requires further research efforts.

Acknowledgment

The authors thank the Co-Editor, an associate editor, and a referee for their insightful comments and suggestions that greatly improved the article. This research was partly supported by the National Natural Science Foundation of China (Grant Nos. 11231010, 11690015, 11671310 and 11471302), Key Laboratory of RCSDS, CAS (No. 2008DP173182) and BCMIIS.

Appendix: Proof of Theorem 1

To study the asymptotic distribution of \( \hat{\beta} \), we assume that the true values \( \beta_0 \) and \( \gamma_0 \) are interior points of compact parameter spaces \( \Theta \) and \( \tilde{\Theta} \), respectively. We need some regularity conditions:

(C1) \{\( Y^*_i(t), N^*_i(\cdot), C_i, X_i(\cdot) \), \( i = 1, 2, ..., n \), are independent and identically distributed.

(C2) \( N_i(L) \) is bounded almost surely, and \( P(C_i \geq L) > 0 \).

(C3) The functions \( X_i(\cdot) \) and \( H(\cdot) \) are continuous and right continuous, respec-
tively, and have bounded variation on $[0, L]$.

(C4) For any $u \in [0, L]$, the conditional density function $f_{Y^*(u)}(\{X_i(u), s(u)\})$ is uniformly continuous and bounded.

(C5) $A$ and $D$ are nonsingular.

**Lemma 1.** Under the assumptions of Theorem 1, we have that for any positive $\varepsilon_n = o(1)$,

\[
\sup_{\|\theta - \theta_0\| \leq \varepsilon_n} \|U(\theta; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta_0; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta; \gamma_0, \Lambda_0)\| = o_p(n^{-1/2}),
\]

where $U(\theta; \gamma_0, \Lambda_0) = E\{U(\theta; \gamma_0, \Lambda_0)\}$.

**Proof.** Let $I_i(u, \theta) = I\{\theta'Z_i(u) > 0\}$. Then we have

\[
\sup_{\|\theta - \theta_0\| \leq \varepsilon_n} \|U(\theta; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta_0; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta; \gamma_0, \Lambda_0)\| \leq I_1 + I_2,
\]

where

\[
I_1 = \sup_{\|\theta - \theta_0\| \leq \varepsilon_n} \|U(\theta; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta; \gamma_0, \Lambda_0) + U(\theta_0; \gamma_0, \Lambda_0) - U(\theta_0; \hat{\gamma}, \hat{\Lambda}_0)\|,
\]
\[
I_2 = \sup_{\|\theta - \theta_0\| \leq \varepsilon_n} \|U(\theta; \gamma_0, \Lambda_0) - U(\theta; \gamma_0, \Lambda_0) - U(\theta_0; \gamma_0, \Lambda_0)\|.
\]

For $I_1$, we obtain

\[
I_1 = r \sup_{\|\theta - \theta_0\| \leq \varepsilon_n} \left\| \frac{1}{n} \sum_{i=1}^{n} \left[ \int_0^L Z_i(u)\xi_i(u)\{I_i(u, \theta) - I_i(u, \theta_0)\} \right. \left. \exp (\hat{\gamma}'X_i(u))d\hat{\Lambda}_0(u) \right. \right. \right. \right. \right. \]
\[
\left. \left. \left. \left. \left. \left. \left. - \exp (\gamma_0'X_i(u))d\Lambda_0(u) \right) \right) \right) \right) \right) \right) \right) \right) \right).
\]

Under (C1)-(C3) and (C5), it follows from Lin et al. (2000) and Sun and Wu (2005) that

\[
\hat{\gamma} - \gamma_0 = n^{-1}A^{-1} \sum_{i=1}^{n} \int_0^L \{X_i(t) - X(t; \gamma_0)\}dM_i^*(t) + O_p(n^{-1/2}) \tag{A.1}
\]

and, uniformly in $t \in [0, L],$

\[
\hat{\Lambda}_0(t) - \Lambda_0(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{dM_i^*(u)}{s(u)} - \int_0^t X(u; \gamma_0)'d\Lambda_0(u)(\hat{\gamma} - \gamma_0) + O_p(n^{-1/2}). \tag{A.2}
\]

In addition, we obtain that when $\theta \rightarrow \theta_0$, $I_i(u, \theta) - I_i(u, \theta_0) = O_p(1)$ for any $1 \leq i \leq n$ and $u \in (0, L)$. Then using (A.1), (A.2), a Taylor expansion and Lemma 1 of Lin et al. (2000), we get that $I_1 = O_p(n^{-1/2})$. Let

\[
r_i(\theta) = \int_0^L Z_i(u)I_i(u, \theta) \left[I\{Y_i(u) - \theta'Z_i(u) < 0\}\right]dN_i(u)
\]
where \( Z \) is a Euclidean class with a square-integrable envelope and \( r_i(\theta) \) is \( L_2(P) \) continuous at \( \theta_0 \).

In view of (C3), it follows from Lemma 22 (ii) in Nolan and Pollard (1987) and Lemmas 2.14 and 2.15 in Pakes and Pollard (1989) that \( \{Z_i(u)I_i(u, \theta)I\{Y_i(u) - \theta'Z_i(u) < 0\}, \ \theta \in \Theta\} \) and \( \{Z_i(u)I_i(u, \theta)\tau_{i}(u)\exp(\gamma_0'X_i(u)), \ \theta \in \Theta\} \) are Euclidean with constant envelope. Thus, following Lemma 5 of Sherman (1994) and Lemma 2.14 (i) in Pakes and Pollard (1989), we get that \( \{r_i(\theta), \ \theta \in \Theta\} \) is a Euclidean class with a square-integrable envelope.

To prove that \( r_{ij}(\theta) \) is \( L_2(P) \) continuous at \( \theta_0 \), let \( Z_{ij} \) and \( r_{ij}(\theta) \) be the \( j \)th components of \( Z_i \) and \( r_i(\theta) \), \( j = 1, ..., p + q \). It can be checked that

\[
E\{r_{ij}(\theta) - r_{ij}(\theta_0)\}^2 \leq 2(I_3 + I_4),
\]

where

\[
I_3 = E\left[ \int_0^L Z_{ij}(u)\left[ I_i(u, \theta)I\{Y_i(u) - \theta'Z_i(u) < 0\} \right. \right. \\
- I_i(u, \theta_0)I\{Y_i(u) - \theta_0'Z_i(u) < 0\}] dN_i(u) \left. \right]^2,
\]

\[
I_4 = \tau^2 E\left[ \int_0^L Z_{ij}(u)\xi_i(u)\left[ I_i(u, \theta) - I_i(u, \theta_0) \right] \exp(\gamma_0'X_i(u))d\Lambda_0(u) \right]^2.
\]

For \( I_3 \), it follows from the Dominated Convergence Theorem that

\[
I_3 \leq
\]

\[
E\left[ \int_0^L Z_{ij}(u)\left\{ F_{Y^*(u)}(x_i(u), \mathcal{F}_iu) (\theta'Z_i(u))I_i(u, \theta) + F_{Y^*(u)}(x_i(u), \mathcal{F}_iu) (\theta_0'Z_i(u))I_i(u, \theta_0) \\
- 2F_{Y^*(u)}(x_i(u), \mathcal{F}_iu) (\min\{\theta'Z_i(u), \theta_0'Z_i(u)\})I_i\left( \min\{\theta'Z_i(u), \theta_0'Z_i(u)\} > 0\right) \right\} dN_i(u) \right]
\]

\[
\rightarrow 0 \ \text{as} \ \theta \rightarrow \theta_0,
\]

where \( F_{Y^*(u)}(x_i(u), \mathcal{F}_iu) (y) \) is the cumulative distribution function of \( Y^*(u) \) conditional on \( \{X_i(u), \mathcal{F}_iu\} \). Similarly, we get that \( I_4 \) tends to 0 as \( \theta \rightarrow \theta_0 \). Thus, it follows from (A.3) that

\[
E\{r_{ij}(\theta) - r_{ij}(\theta_0)\}^2 \rightarrow 0, \ \text{as} \ \theta \rightarrow \theta_0.
\]
That is, \( r_{ij}(\theta) \) is \( L_2(P) \) continuous at \( \theta_0 \). This completes the proof of Lemma 1.

**Proof of Theorem 1.** To prove the consistency of \( \hat{\theta} \), it suffices to verify Conditions (i), (ii) and (iii) of Corollary 3.2 in \cite{PakesPollard1989}. By the definition of \( \hat{\theta} \), we have

\[
\| U(\hat{\theta}; \hat{\gamma}, \hat{\Lambda}_0) \| \leq o_p(1) + \inf_{\theta \in \Theta} \| U(\theta; \hat{\gamma}, \hat{\Lambda}_0) \|. \tag{A.4}
\]

It follows from (C5) that for each \( \delta > 0, \) \( \inf_{\|\theta - \theta_0\| > \delta} \| U(\theta; \gamma_0, \Lambda_0) \| > 0. \) Thus, Conditions (i) and (ii) of Corollary 3.2 hold. If

\[
I_5 = \sup_{\theta \in \Theta} \| U(\theta; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta; \gamma_0, \Lambda_0) \|,
\]

\( I_5 \leq I_6 + I_7, \) where

\[
I_6 = \sup_{\theta \in \Theta} \| U(\theta; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta; \gamma_0, \Lambda_0) \|, \quad I_7 = \sup_{\theta \in \Theta} \| U(\theta; \gamma_0, \Lambda_0) - U(\theta; \gamma_0, \Lambda_0) \|.
\]

Obviously, \( I_7 = \sup_{\theta \in \Theta} \| n^{-1} \sum_{i=1}^n \{ r_i(\theta) - E(r_i(\theta)) \} \|. \) Following similar arguments as in the proof of Lemma 1, we obtain that \( \{ r_i(\theta), \theta \in \Theta \} \) is a Euclidean class with integrable envelope. Thus, it follows from Lemma 2.8 in \cite{PakesPollard1989} that \( I_7 = o_p(1). \) It can be checked that \( I_6 \leq I_8 + I_9, \) where

\[
I_8 = \sup_{\theta \in \Theta} \left\| \frac{\tau}{n} \sum_{i=1}^n \int_0^L Z_i(u) I_i(u, \theta) \xi_i(u) \exp(\hat{\gamma}'X_i(u)) d\{ \hat{\Lambda}_0(u) - \Lambda_0(u) \} \right\|
\]

\[
I_9 = \sup_{\theta \in \Theta} \left\| \frac{\tau}{n} \sum_{i=1}^n \int_0^L Z_i(u) I_i(u, \theta) \xi_i(u) \{ \exp(\hat{\gamma}'X_i(u)) - \exp(\gamma_0'X_i(u)) \} d\Lambda_0(u) \right\|
\]

In view of (A.2), the Functional Central Limit Theorem \cite{Pollard1990} and the Continuous Mapping Theorem imply that \( I_8 = o_p(1). \) It follows from a Taylor expansion that \( I_9 = o_p(1). \) Thus, \( I_5 = o_p(1), \) which is Condition (iii) of Corollary 3.2 in \cite{PakesPollard1989}. It then follows that \( \hat{\theta} \) converges in probability to \( \theta_0. \)

To prove the asymptotic normality of \( \hat{\theta} \), it is sufficient to verify Conditions (i)-(v) of Theorem 3.3 in \cite{PakesPollard1989}. In view of (C3) and (C5), using (A.4) and Lemma 1, we can show that Conditions (i)-(iii) and (v) of Theorem 3.3 in \cite{PakesPollard1989} holds:

(i) \( \| U(\hat{\theta}; \hat{\gamma}, \hat{\Lambda}_0) \| \leq o_p(1) + \inf_{\theta \in \Theta} \| U(\theta; \hat{\gamma}, \hat{\Lambda}_0) \|; \)

(ii) \( U(\theta; \gamma_0, \Lambda_0) \) is differentiable at \( \theta_0 \) with the derivative matrix \( D \) of full rank;

(iii) for every sequence \( \{ \epsilon_n \} \) of positive numbers that converges to zero,

\[
\sup_{\| \theta - \theta_0 \| \leq \epsilon_n} \| U(\theta; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta_0; \hat{\gamma}, \hat{\Lambda}_0) - U(\theta; \gamma_0, \Lambda_0) \| = o_p(n^{-1/2});
\]
(v) $\theta_0$ is an interior point of $\Theta$.

To verify Condition (iv), it can be shown that by (A.1) and (A.2),

$$n^{1/2}U(\theta_0; \hat{\gamma}, \hat{\Lambda}_0) = n^{-1/2} \sum_{i=1}^{n} q_i(\theta_0)$$

$$+ \tau n^{-1/2} \sum_{i=1}^{n} \int_{0}^{L} \{Z_i(u)I_i(u, \theta_0) - \bar{z}(u)\} dM_i^*(u)$$

$$- \tau \mathbf{B} \mathbf{A}^{-1} n^{-1/2} \sum_{i=1}^{n} \int_{0}^{L} \{X_i(u) - \bar{x}(u)\} dM_i^*(u) + o_p(1),$$

which implies by the Multivariate Central Limit Theorem that

$$n^{1/2}U(\theta_0; \hat{\gamma}, \hat{\Lambda}_0) \rightarrow N(0, \mathbf{V}) \quad \text{in distribution}. \quad (A.5)$$

Thus, Condition (iv) of Theorem 3.3 in Pakes and Pollard (1989) holds, and it follows that $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $\Sigma$ defined in Theorem 1.

References


Wang, H. and Fygenson, M. (2009). Inference for censored quantile regression models in longi-


Department of Statistics and Finance, The School of Management, University of Science and Technology of China, Hefei, Anhui 230026, P.R.China.

E-mail: zfw@ustc.edu.cn

School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P.R.China.

E-mail: jlding.math@whu.edu.cn

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, P.R.China.

E-mail: slq@amt.ac.cn

Department of Statistics and Finance, The School of Management, University of Science and Technology of China, Hefei, Anhui, 230026, P.R.China.

E-mail: wuyh@ustc.edu.cn

(Received December 2016; accepted January 2017)