Statistica Sinica Preprint No: SS-2016-0527R1	
Title	Tobit quantile regression of left-censored longitudinal data with informative observation times
Manuscript ID	SS-2016-0527R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0527
Complete List of Authors	Zhanfeng Wang Jieli Ding Liuquan Sun and Yaohua Wu
Corresponding Author	Liuquan Sun
E-mail	slq@amt.ac.cn
Notice: Accepted version subject to English editing.	

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Tobit quantile regression of left-censored longitudinal data with informative observation times

Zhanfeng Wang[1], Jieli Ding[2], Liuquan Sun[3] and Yaohua Wu[1]

[1]*Department of Statistics and Finance, The School of Management, University of Science and Technology of China, Hefei, Anhui, 230026, P.R.China, zfw@ustc.edu.cn, wuyh@ustc.edu.cn*

[2]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, 430072, P.R.China, jlding.math@whu.edu.cn*

[3]*Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, P.R.China, slq@amt.ac.cn*

Abstract: In many longitudinal studies, longitudinal responses are often 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 being 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, and frequently occur in many fields, such as medical follow-up studies, psychology, sociology and observational investigations. In practice, 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 or longitudinal data with left censoring (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 all these modeling approaches, one strong assumption is that the longitudinal responses and the observation times are independent given covariates.

In many applications, however, the longitudinal responses may be correlated with the

observation times, that is, the 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 et al. (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. All these methods primarily analyze longitudinal data with informative observation times in the absence of left censoring. However, to the best of our knowledge, there is no existing work considering the joint analysis of left-censored longitudinal data with informative observation times. Furthermore, 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 would lead to biased inferences. Hence, there is clearly a need for an analytical method that can directly model left-censored longitudinal data, taking informative observation times into account.

In this article, we propose joint modeling of left-censored longitudinal data with informative observation times. To be specific, a Tobit quantile regression model is used for the 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. In addition, unlike Chen, Tang and Zhou (2016), the proposed objective function is neither differentiable nor convex due to left-censored responses, and the theoretical and computational developments become much more involved and challenging. Furthermore, a modified Majorize-Minimize (MM) algorithm is used to handle the computational difficulty, and the algorithm procedure 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 estimating 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 proposed 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. All technical proofs are relegated to the Appendix.

2. Model Specification

Consider a longitudinal study involving n independent subjects. For the ith subject, let $Y_i^*(t)$ be the underlying response variable at time t, and $\mathbf{X}_i(t)$ 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 may assume that $d = 0$. Let $Y_i(t) = \max\{Y_i^*(t), 0\}$, and C_i be the follow-up or censoring time. In addition, let $N_i^*(t)$ be the counting process denoting the number of the observation times before or at time t. Then 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 $\{\mathbf{X}_i(t) : 0 \le t \le C_i\}$ $(i = 1, ..., n)$ are assumed to be observed.

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

model:

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

where β and α 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). In reality, we cannot observe the potential response $Y_i^*(t)$ because of its nonnegative constraint. Hence in view of (1), following Wang and Fygenson (2009) and for a given $0 < \tau < 1$, we consider the following marginal regression model:

$$Y_i(t) = \max\{0, \beta' \mathbf{X}_i(t) + \alpha' \mathbf{H}(\mathcal{F}_{it}) + e_i(t)\}, \quad i = 1, ..., n, \tag{2}$$

where the τth quantile of $e_i(t)$ is assumed to be zero. Here, model (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 $\mathbf{X}_i(t)$, $N_i^*(t)$ is a nonhomogeneous Poisson process with

$$E(dN_i^*(t)|\mathbf{X}_i(t)) = \exp(\gamma' \mathbf{X}_i(t)) d\Lambda_0(t), \quad i = 1, 2, ..., n, \tag{3}$$

where γ 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 notational convenience, models (2) and (3) assume the same set of covariates $X(t)$. The proposed estimation procedure can be extended in a straightforward manner to deal with different set of covariates for these two models.

In comparison to the common Tobit quantile regression models with longitudinal data, a main feature of model (2) is that it allows the response process $Y_i(t)$ to be correlated with the observation process $N_i^*(t)$. In particular, model (2) specifies that the process $Y_i(t)$ depends on the process $N_i^*(t)$ in a linear fashion through the function $\mathbf{H}$. When

$\alpha = 0$, model (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 model (2), the function $\mathbf{H}$ can be chosen according to practical situation. As discussed in Sun et al. (2005), a natural choice for $\mathbf{H}$ may be $\mathbf{H}(\mathcal{F}_{it}) = N_i(t-)$, which implies that all information about $Y_i(t)$ in $\mathcal{F}_{it}$ is given by the total number of observations. Another choice is $\mathbf{H}(\mathcal{F}_{it}) = N_i(t-) - N_i(t-s)$, which indicates that $Y_i(t)$ depends on $\mathcal{F}_{it}$ only through the number of observations in recent s time units. The function $\mathbf{H}$ also can be defined by some combination of the foregoing two choices if both the total and recent numbers of observations may include information about the response $Y_i(t)$. In what follows, we assume that conditional on the covariate $\mathbf{X}_i(t)$, 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 $\mathbf{X}_i(t)$. In this case, a natural and common method to estimate $\theta = (\beta', \alpha')'$ is to minimize the objection function

$$Q(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^L \rho_\tau\{Y_i(u) - (\theta'\mathbf{Z}_i(u))^+\}dN_i(u), \tag{4}$$

where L is the maximum follow-up time, $a^+ = \max\{0, a\}$, $\mathbf{Z}_i(u) = (\mathbf{X}_i'(u), \mathbf{H}'(\mathcal{F}_{iu}))'$, and $\rho_\tau(v) = v\{\tau - I(v < 0)\}$ is the quantile loss function (e.g., Wang and Fygenson, 2009). Note that the subgradient of $\rho_\tau\{Y_i(u) - (\theta'\mathbf{Z}_i(u))^+\}dN_i(u)$ with respect to θ is

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

Thus, minimizing (4) is equivalent to solving the following estimating equation:

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^L \mathbf{Z}_i(u)I\{\theta'\mathbf{Z}_i(u)>0\}\Big[I\{Y_i(u)-\theta'\mathbf{Z}_i(u)<0\}-\tau\Big]dN_i(u)=o_p(n^{-1/2}),$$

which can be actually obtained from the derivative of (4) with respect to θ.

We now consider the case in which the observation times depend on the covariates through model (3). Motivated by the above equation define

$$M_i(t;\theta,\gamma,\Lambda_0)=\int_0^t I\{\theta'\mathbf{Z}_i(u)>0\}\Big[I\{Y_i(u)-\theta'\mathbf{Z}_i(u)<0\}dN_i(u) - \tau\xi_i(u)\exp\left(\gamma'\mathbf{X}_i(u)\right)d\Lambda_0(u)\Big],$$

where $\xi_i(t)=I(C_i\geq t)$. Let θ_0 and γ_0 be true values of θ and γ, respectively. Note that under models (2) and (3) and the assumptions,

$$E\Big[I\{\theta_0'\mathbf{Z}_i(u)>0\}I\{Y_i(u)-\theta_0'\mathbf{Z}_i(u)<0\}dN_i(u)|\mathbf{Z}_i(u),C_i\Big]=\tau\xi_i(u)\exp\left(\gamma_0'\mathbf{X}_i(u)\right)d\Lambda_0(u).$$

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 γ_0 and Λ_0, we can estimate θ_0 using the estimating equation:

$$U(\theta;\gamma_0,\Lambda_0)=\frac{1}{n}\sum_{i=1}^{n}\int_0^L \mathbf{Z}_i(u)dM_i(u;\theta,\gamma_0,\Lambda_0)=0. \tag{5}$$

Of course, in reality, γ_0 and Λ_0 are unknown. Using the approach of Lin et al. (2000), we propose the following estimating equation for γ_0:

$$\sum_{i=1}^{n}\int_0^L\{\mathbf{X}_i(u)-\bar{\mathbf{X}}(u;\gamma)\}dN_i(u)=0, \tag{6}$$

where $\bar{\mathbf{X}}(u;\gamma)=\mathbf{S}^{(1)}(u;\gamma)/\mathbf{S}^{(0)}(u;\gamma)$, and

$$\mathbf{S}^{(k)}(u;\gamma)=\frac{1}{n}\sum_{j=1}^{n}\xi_j(u)\mathbf{X}_j(u)^{\otimes k}\exp\left(\gamma'\mathbf{X}_j(u)\right),\quad k=0,1,2,$$

with $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}'$ for any vector $\mathbf{a}$. Let $\hat{\gamma}$ denote the solution to the above estimating equation. Then $\Lambda_0(t)$ can be consistently estimated by the Aalen-Breslow-type estimator

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

By replacing γ_0 and $\Lambda_0(t)$ with $\hat{\gamma}$ and $\hat{\Lambda}_0(t)$ in (5), we specify the following estimating function for θ_0:

$$U(\theta;\hat{\gamma},\hat{\Lambda}_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^L \mathbf{Z}_i(u)dM_i(u;\theta,\hat{\gamma},\hat{\Lambda}_0). \tag{7}$$

Since $U(\theta;\hat{\gamma},\hat{\Lambda}_0)$ is a discontinuous function of θ, 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 $\|\mathbf{a}\| = (\mathbf{a}'\mathbf{a})^{1/2}$.

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). Specifically, let

$$q_i(\theta) = \int_0^L \mathbf{Z}_i(u)I\{\theta'\mathbf{Z}_i(u) > 0\}\Big[I\{Y_i(u) - \theta'\mathbf{Z}_i(u) < 0\} - \tau\Big]dN_i(u),$$
$$G(\theta;\hat{\gamma}) = \frac{\tau}{n}\sum_{i=1}^{n}\int_0^L \Big[\mathbf{Z}_i(u)I\{\theta'\mathbf{Z}_i(u) > 0\} - \frac{\tilde{S}^{(1)}(u;\theta,\hat{\gamma})}{S^{(0)}(u;\hat{\gamma})}\Big]dN_i(u),$$

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

$$\tilde{S}^{(1)}(u;\theta,\gamma) = \frac{1}{n}\sum_{i=1}^{n}\xi_i(u)\mathbf{Z}_i(u)I\{\theta'\mathbf{Z}_i(u) > 0\}\exp\left(\gamma'\mathbf{X}_i(u)\right).$$

Then it can be checked that

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

Let θ^k be the kth iteration estimate of θ. By following Hunter and Lange (2000), the objection function $Q(\theta)$ can be replaced with a Majorizing function with a disturbance

constant $\varepsilon > 0$,

$$\widetilde{Q}_\varepsilon(\theta|\theta^k) = \frac{1}{n}\sum_{i=1}^{n}\int_0^L \frac{1}{4}\Big\{\frac{r_i(\theta)^2}{\varepsilon + |r_i(\theta^k)|} + (4\tau-2)r_i(\theta) + c_i\Big\}dN_i(u),$$

where $r_i(\theta) = Y_i(u) - (\theta'\mathbf{Z}_i(u))^+$. Since $(\theta'\mathbf{Z}_i(u))^+$ is not continuous with respect to θ, the function $\widetilde{Q}_\varepsilon(\theta|\theta^k)$ is still not smooth at θ. To this end, we use the following smoothing function to approximate t^+:

$$K_h(t) = \frac{t}{1+\exp(-t/h)},$$

where $h > 0$ is a known smoothing parameter. It is easy to validate that $K_h(t) \approx t^+$ for small enough h. By substituting $\tilde{r}_i(\theta) = Y_i(u) - K_h(\theta'\mathbf{Z}_i(u))$ for $r_i(\theta)$ in $\widetilde{Q}_\varepsilon(\theta|\theta^k)$, we obtain a smooth Majorizing function:

$$Q_\varepsilon(\theta|\theta^k) = \frac{1}{n}\sum_{i=1}^{n}\int_0^L \frac{1}{4}\Big\{\frac{\tilde{r}_i(\theta)^2}{\varepsilon + |\tilde{r}_i(\theta^k)|} + (4\tau-2)\tilde{r}_i(\theta) + c_i\Big\}dN_i(u),$$

where c_i satisfies

$$\frac{1}{4}\Big\{\frac{\tilde{r}_i(\theta^k)^2}{\varepsilon + |\tilde{r}_i(\theta^k)|} + (4\tau-2)\tilde{r}_i(\theta^k) + c_i\Big\} = \rho_\tau(r_i(\theta^k)) - \frac{\varepsilon}{2}\log(\varepsilon + |r_i(\theta^k)|).$$

Based on the smoothing objection function $Q_\varepsilon(\theta|\theta^k)$, the $(k+1)$-step estimating function can be constructed as

$$\widetilde{U}(\theta;\hat{\gamma},\hat{\Lambda}_0|\theta^k) = \frac{\partial Q_\varepsilon(\theta|\theta^k)}{\partial\theta} + G(\theta^k;\hat{\gamma}). \tag{8}$$

Then the $(k+1)$th iteration estimate θ^{k+1} is obtained by solving $\tilde{U}(\theta;\hat{\gamma},\hat{\Lambda}_0)|\theta^k) = 0$. Given an initial estimate θ^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 and in the numerical studies below, 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 θ^{k+1} as follows: $\theta^{k+1} = \theta^k - \Delta^{k+1}$, where

$$\Delta^{k+1} = \Big(\frac{\partial^2 Q_\varepsilon(\theta|\theta^k)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta^k}\Big)^{-1}\tilde{U}(\theta^k;\hat{\gamma},\hat{\Lambda}_0|\theta^k).$$

3.3. Properties

To establish the asymptotic properties of $\hat{\theta}$, define

$$\tilde{S}^{(2)}(u;\theta,\gamma) = \frac{1}{n}\sum_{i=1}^{n} \xi_i(u)\mathbf{Z}_i(u)\mathbf{X}_i(u)' I\{\theta'\mathbf{Z}_i(u) > 0\} \exp\left(\gamma'\mathbf{X}_i(u)\right),$$

and $\bar{\mathbf{Z}}(u;\gamma,\theta) = \tilde{S}^{(1)}(u;\theta,\gamma)/S^{(0)}(u;\gamma)$. Let $\tilde{s}^{(1)}(u)$, $\tilde{s}^{(2)}(u)$ 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 $S^{(k)}(u;\gamma_0)$ $(k = 0, 1, 2)$, respectively. Also, let $\bar{\mathbf{z}}(u) = \tilde{s}^{(1)}(u)/s^{(0)}(u)$ and $\bar{\mathbf{x}}(u) = s^{(1)}(u)/s^{(0)}(u)$. Define

$$\begin{aligned}
&M_i^*(t) = N_i(t) - \int_0^t \xi_i(u)\exp\left(\gamma_0'\mathbf{X}_i(u)\right)d\Lambda_0(u),\\
&\mathbf{A} = E\Big[\int_0^L \Big\{\frac{s^{(2)}(u)}{s^{(0)}(u)} - \bar{\mathbf{x}}(u)^{\otimes 2}\Big\}dN_i(u)\Big],\\
&\mathbf{B} = E\Big[\int_0^L \Big\{\frac{\tilde{s}^{(2)}(u)}{s^{(0)}(u)} - \frac{\tilde{s}^{(1)}(u)s^{(1)}(u)'}{s^{(0)}(u)^2}\Big\}dN_i(u)\Big],\\
&\mathbf{D} = E\Big[\int_0^L \mathbf{Z}_i(u)\mathbf{Z}_i(u)' I\{\theta_0'\mathbf{Z}_i(u) > 0\} f_{Y^*(u)|\{\mathbf{X}_i(u),\mathcal{F}_{iu}\}}(\theta_0'\mathbf{Z}_i(u))dN_i(u)\Big],
\end{aligned}$$

where $f_{Y^*(u)|\{\mathbf{X}_i(u),\mathcal{F}_{iu}\}}(y)$ is the density function of $Y^*(u)$ conditional on $\{\mathbf{X}_i(u), \mathcal{F}_{iu}\}$ for $u \in [0, L]$.

The consistency and the asymptotic normality of $\hat{\theta}$ are established in the following theorem with the proof 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 $\mathbf{\Sigma} = \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}$, where*

$$\begin{aligned}
\mathbf{V} = E\Big[&q_i(\theta_0) + \tau\int_0^L [\mathbf{Z}_i(u)I\{\theta_0'\mathbf{Z}_i(u) > 0\} - \bar{\mathbf{z}}(u)]dM_i^*(u)\\
&-\tau\mathbf{B}\mathbf{A}^{-1}\int_0^L \{\mathbf{X}_i(u) - \bar{\mathbf{x}}(u)\}dM_i^*(u)\Big]^{\otimes 2}.
\end{aligned}$$

The asymptotic variance of $\hat{\theta}$ may be consistently estimated by $\hat{\mathbf{\Sigma}}$, which can be obtained by the usual plug-in method. Note that $\mathbf{\Sigma}$ is of complicated form involving some nuisance parameters such as the conditional density function of $Y^*(u)$. Thus, it is difficult to estimate $\mathbf{\Sigma}$ directly. Here, we propose to use the bootstrap method to estimate the asymptotic variance of $\hat{\theta}$. The accuracy of the bootstrap method depends on the sample size and the number of bootstrap samples. A large number of bootstrap samples yield a high accuracy. In the following simulation studies with the sample size $n = 100$, we used 200 bootstrap samples and found the variance estimation fairly accurate. Of course, if the number of bootstrap samples is too large, the computation will be time-consuming.

4. Simulation Studies

Simulation studies were conducted to examine the finite sample performance of the proposed estimators. In the study, let $\mathbf{X}_i = (X_{i1}, X_{i2})'$, 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 $\mathbf{X}_i$, the observation times were generated from a nonhomogeneous Poisson process with

$$E(dN_i^*(t)|\mathbf{X}_i) = \exp(\gamma_1 X_{i1} + \gamma_2 X_{i2}) d\Lambda_0(t),$$

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 following 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. Note that α reflects the dependence among the response variable and the observation times. For example, $\alpha = 0$ implies that

the response variable and the observation times are independent, while $\alpha \neq 0$ means that the two related processes have nonzero correlations. For π and $e_i(t)$, we considered the following three cases:

S1. $\pi = 0$, $e_i(t)$ was generated from the standard normal distribution;

S2. $\pi = 1$, $e_i(t)$ was generated from the standard normal distribution;

S3. $\pi = 0$, $e_i(t)$ was generated from the standard Cauchy distribution.

For the above three cases, the median of $e_i(t)$ was zero, that is, we took $\tau = 0.5$. The results presented below are based on 1000 replications with sample size n=100, and final estimates were reached at convergence, i.e., the absolute difference of the estimates between two successive iterations is less than 10^{-5}. The asymptotic variance was estimated using the bootstrap method with 200 bootstrap samples, which are found 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. Specifically, the proposed estimators are virtually unbiased, the standard error estimates are practically close to the sampling standard errors, and the 95% empirical coverage probabilities are reasonable. The results become better with increasing κ 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 also considered the method of Wang and Fygenson (2009) (denoted by WF), who studied model (2) with $\alpha = 0$, and assumed that the observation times are noninformative without modeling the observation times. Under the same setup as

above, the comparison results on estimation of β_1 and β_2 are given in Tables 1 and 2. As expected, when the observation times are noninformative (i.e., $\alpha = 0$), the WF estimators are unbiased. Under such situation, both methods provide reasonable and comparable estimates, and the variances of our method are only slightly larger than those of the WF's method. This is because the latter utilizes the assumption of the independent observation times in estimation. However, when such independent assumption is violated (i.e., $\alpha \neq 0$), the WF's method may lead to large bias, and yields improper coverage probabilities.

Furthermore, 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. $e_i(t)$ followed a normal distribution with mean $-\Phi^{-1}(0.25)$ and variance 1;

S5. $e_i(t)$ was generated from a Cauchy distribution with location parameter 1 and scale parameter 1.

For $\tau = 0.75$, we considered the following two situations:

S6. $e_i(t)$ followed a normal distribution with mean $-\Phi^{-1}(0.75)$ and variance 1;

S7. $e_i(t)$ was generated from a Cauchy distribution with location parameter -1 and scale parameter 1.

Here, different mean and location parameters guarantee that the τth quantile of $e_i(t)$ equals to zero. For all the above 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 from Table 3 that the proposed method still performed reasonably well for the cases of $\tau = 0.25$ and 0.75. That is, the proposed estimators have small biases, reasonable variance estimates and the empirical coverage probabilities. We also considered other setups and obtained similar results.

Finally, we conducted some simulation studies to compare the proposed method with a naïve method, which replaces the left censored responses by half of the detection limit, and applies the method of Chen, Tang and Zhou (2016) to the imputation data. We just considered the case S2 in Table 1, where X_{i1} was generated from the standard normal distribution, and X_{i2} was from a normal distribution with mean 0.2 and variance 1. The results for other cases are similar. Table 4 gives the comparison results on estimation of $\theta = (\beta_1, \beta_2, \alpha)'$, and Naive stands for the naïve method in the table. Simulation results show that the naïve method may lead to biases even when $\alpha = 0$. In addition, the naïve method tends to cause an inflated SEE, and yields improper coverage probabilities.

5. An Application

For illustration purposes, 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 Fygenson, 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 of individuals may vary around the scheduled visiting times, and the numbers of measurements were also different, which indicates that the observation times may be informative about the viral load. A total of 481 patients was enrolled in this study. The numbers of patients with 1 to 6 times visiting hospital are 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 is considered to be a factor that affects the antiviral response to the antiretroviral regimens. Here, we focus 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 use the following transformation of the actual visit times $t = \log_{10}(\text{day of the actual visit} + 40) - \log_{10}(33)$. For subject i, let X_{i1} be the indicator of the prior antiviral treatment with NNRTI, with 1 for having had NNRTI and 0 for having not received NNRTI. Also let $X^*_{2i} = 1$ denote the patients who received a double-PI treatment and $X^*_{2i} = 0$ for patients who received a single-PI treatment. Since Sun and Wu (2005) showed that the effect of X^*_{2i} on the viral load response was linear over time, we defined $X_{2i}(t) = X^*_{2i}t$. Let $Y_i(t)$ stand for the observed $\log_{10}$(viral load response). We considered the following Tobit quantile regression model in this study:

$$Y_i(t) = \max\{\log_{10}(200),\ \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}(t) + \alpha N_i(t-) + e_i(t)\}, \tag{9}$$

and the actual visit times can be described by model (3) with $\mathbf{X}_i(t) = (X_{i1}, X_{i2}(t))'$ and $\gamma = (\gamma_1, \gamma_2)'$. For model (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 model (9) at different quantiles with τ taking $0.5, 0.6$ 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 β and α 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 τ's. In particular, the patients with the prior NNRTI treatment have higher viral load than those without the prior NNRTI treatment, and the double-PI treatment can reduce HIV viral load compared to the single-PI treatment. In addition, all $\hat{\alpha}$'s are positive for three different τ values, indicating that the HIV viral load response and the actual visit times

are positively correlated. That is, the patients who visited more often have higher HIV viral load.

For comparison, Table 5 also gives the results of the WF's method assuming that the observation times were noninformative. It can be seen that the WF estimators are nearly the same as our proposed estimators for β_1 at three different quantile levels. However, the WF's method overestimate slightly β_0 and underestimate significantly β_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 also we performed the sensitivity analysis to the choice of the function $\mathbf{H}(\mathcal{F}_{it})$ for the data, and replaced $N_i(t-)$ in model (9) with $\log\{N_i(t-)+1\}$ or $N_i(t-)^{1/2}$, respectively. The other setups were the same as in Table 5 with $\tau = 0.7$. The analysis results are reported in Table 6. It can be seen from Tables 5 and 6 that these results are similar for the three choices of $\mathbf{H}(\mathcal{F}_{it})$, and 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.

Here we assume that the observation process is a nonhomogeneous Poisson process. However, 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.

Note that 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, however, this noninformative censoring assumption might 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.

Acknowledgements

The authors thank the Co-Editor, Zhiliang Ying, 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

In order to study the asymptotic distribution of $\hat{\beta}$, we assume that the true values β_0

and γ_0 are two interior points of compact parameter spaces Θ and $\tilde{\Theta}$, respectively, and need the following regularity conditions:

(C1) $\{Y_i^*(t), N_i^*(\cdot), C_i, \mathbf{X}_i(\cdot)\},\ i = 1, 2, ..., n$, are independent and identically distributed.

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

(C3) The functions $\mathbf{X}_i(\cdot)$ and $H(\cdot)$ are continuous and right continuous, respectively, and have bounded variation on $[0, L]$.

(C4) For any $u \in [0, L]$, the conditional density function $f_{Y^*(u)|\{\mathbf{X}_i(u),\mathcal{F}_{iu}\}}(\cdot)$ is uniformly continuous and bounded.

(C5) $\mathbf{A}$ and $\mathbf{D}$ are nonsingular.

The following lemma is useful in proving Theorem 1.

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

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

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

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

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

where

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

$$I_2 = \sup_{||\theta-\theta_0||\leq\varepsilon_n} ||U(\theta; \gamma_0, \Lambda_0) - \mathcal{U}(\theta; \gamma_0, \Lambda_0) - U(\theta_0; \gamma_0, \Lambda_0)||.$$

For I_1, we obtain

$$I_1 = \tau \sup_{||\theta-\theta_0||\leq\varepsilon_n} \Big|\Big| \frac{1}{n} \sum_{i=1}^{n} \Big[\int_0^L \mathbf{Z}_i(u)\xi_i(u)\big\{I_i(u, \theta) - I_i(u, \theta_0)\big\}\big\{ \exp(\hat{\gamma}' \mathbf{X}_i(u)) d\hat{\Lambda}_0(u) - \exp(\gamma_0' \mathbf{X}_i(u)) d\Lambda_0(u)\big\}\Big]\Big|\Big|.$$

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

$$\hat{\gamma} - \gamma_0 = n^{-1}\mathbf{A}^{-1}\sum_{i=1}^{n}\int_0^L \{\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\gamma_0)\}dM_i^*(t) + o_p(n^{-1/2}), \tag{A.1}$$

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

$$\hat{\Lambda}_0(t) - \Lambda_0(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{dM_i^*(u)}{s^{(0)}(u)} - \int_0^t \bar{\mathbf{X}}(u;\gamma_0)' d\Lambda_0(u)(\hat{\gamma} - \gamma_0) + o_p(n^{-1/2}). \tag{A.2}$$

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

$$r_i(\theta) = \int_0^L \mathbf{Z}_i(u) I_i(u,\theta)\big[I\{Y_i(u) - \theta'\mathbf{Z}_i(u) < 0\}dN_i(u) - \tau\xi_i(u)\exp\left(\gamma_0'\mathbf{X}_i(u)\right)d\Lambda_0(u)\big].$$

Thus,

$$I_2 = \sup_{||\theta-\theta_0||\le\varepsilon_n} \Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\big[r_i(\theta) - E\{r_i(\theta)\}\big] - \frac{1}{n}\sum_{i=1}^{n}\big[r_i(\theta_0) - E\{r_i(\theta_0)\}\big]\Big|\Big|.$$

Hence by Lemma 2.17 in Pakes and Pollard (1989), to prove $I_2 = o_p(n^{-1/2})$, it suffices to prove that $\{r_i(\theta),\ \theta \in \Theta\}$ is a Euclidean class with a square-integrable envelope and $r_i(\theta)$ is $L_2(P)$ continuous at θ_0.

In view of Condition (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 $\{\mathbf{Z}_i(u)I_i(u,\theta)I\{Y_i(u) - \theta'\mathbf{Z}_i(u) < 0\},\ \theta \in \Theta\}$ and $\{\mathbf{Z}_i(u)I_i(u,\theta)\tau\xi_i(u)\exp\left(\gamma_0'\mathbf{X}_i(u)\right),\ \theta \in \Theta\}$ are Euclidean with constant envelope. Thus, following Lemma 5 of Sherman (1994) and Lemmas 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 θ_0, let Z_{ij} and $r_{ij}(\theta)$ be the jth components of $\mathbf{Z}_i$ and $r_i(\theta)$, $j = 1, ..., p+q$. It can be checked that

$$E\{r_{ij}(\theta) - r_{ij}(\theta_0)\}^2 \le 2(I_3 + I_4), \tag{A.3}$$

where

$$
\begin{aligned}
I_3 =& E\Big[\int_0^L Z_{ij}(u)\big[I_i(u,\theta)I\{Y_i(u)-\theta'\mathbf{Z}_i(u)<0\} \\
& - I_i(u,\theta_0)I\{Y_i(u)-\theta_0'\mathbf{Z}_i(u)<0\}\big]dN_i(u)\Big]^2, \\
I_4 =& \tau^2 E\Big[\int_0^L Z_{ij}(u)\xi_i(u)\{I_i(u,\theta)-I_i(u,\theta_0)\}\exp\left(\gamma_0'\mathbf{X}_i(u)\right)d\Lambda_0(u)\Big]^2.
\end{aligned}
$$

For I_3, it follows from the dominated convergence theorem that

$$
\begin{aligned}
I_3 \leq & E\Big[\int_0^L Z_{ij}(u)^2\Big\{F_{Y^*(u)|\{\mathbf{X}_i(u),\mathcal{F}_{iu}\}}(\theta'\mathbf{Z}_i(u))I_i(u,\theta)+F_{Y^*(u)|\{\mathbf{X}_i(u),\mathcal{F}_{iu}\}}(\theta_0'\mathbf{Z}_i(u))I_i(u,\theta_0) \\
& - 2F_{Y^*(u)|\{\mathbf{X}_i(u),\mathcal{F}_{iu}\}}(\min\{\theta'\mathbf{Z}_i(u),\theta_0'\mathbf{Z}_i(u)\})I(\min\{\theta'\mathbf{Z}_i(u),\theta_0'\mathbf{Z}_i(u)\}>0)\Big\}dN_i(u)\Big] \\
& \longrightarrow 0 \quad \text{as} \quad \theta\to\theta_0,
\end{aligned}
$$

where $F_{Y^*(u)|\{\mathbf{X}_i(u),\mathcal{F}_{iu}\}}(y)$ is the cumulative distribution function of $Y^*(u)$ conditional on $\{\mathbf{X}_i(u),\mathcal{F}_{iu}\}$. Similarly, we get that I_4 tends to 0 as $\theta\to\theta_0$. Thus, it follows from (A.3) that

$$
E\{r_{ij}(\theta)-r_{ij}(\theta_0)\}^2\longrightarrow 0, \quad \text{as} \quad \theta\to\theta_0.
$$

That is, $r_{ij}(\theta)$ is $L_2(P)$ continuous at θ_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 Pakes and Pollard (1989). First note that by the definition of $\hat{\theta}$, we have

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

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

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

Then $I_5 \leq I_6 + I_7$, where

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

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 Pakes and Pollard (1989) that $I_7 = o_p(1)$. It can be checked that $I_6 \leq I_8 + I_9$, where

$$I_8 = \sup_{\theta\in\Theta} \Big|\Big| \frac{\tau}{n}\sum_{i=1}^n \int_0^L \mathbf{Z}_i(u) I_i(u,\theta)\xi_i(u) \exp(\hat{\gamma}'\mathbf{X}_i(u)) d\big\{\hat{\Lambda}_0(u) - \Lambda_0(u)\big\} \Big|\Big|,$$

$$I_9 = \sup_{\theta\in\Theta} \Big|\Big| \frac{\tau}{n}\sum_{i=1}^n \int_0^L \mathbf{Z}_i(u) I_i(u,\theta)\xi_i(u) \big\{ \exp(\hat{\gamma}'\mathbf{X}_i(u)) - \exp(\gamma_0'\mathbf{X}_i(u))\big\} d\Lambda_0(u) \Big|\Big|.$$

In view of (A.2), the functional central limit theorem (Pollard, 1990) and the continuous mapping theorem imply that $I_8 = o_p(1)$. It follows from the Taylor expansion that $I_9 = o_p(1)$. Thus, $I_5 = o_p(1)$, which is Condition (iii) of Corollary 3.2 in Pakes and Pollard (1989). It then follows that $\hat{\theta}$ converges in probability to θ_0.

To prove the asymptotic normality of $\hat{\theta}$, it is sufficient to verify Conditions (i)-(v) of Theorem 3.3 in Pakes and Pollard (1989). In view of Conditions (C3) and (C5), using (A.4) and Lemma 1, we can show that Conditions (i)-(iii) and (v) of Theorem 3.3 in Pakes and Pollard (1989) holds, i.e.,

(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) $\mathcal{U}(\theta;\gamma_0,\Lambda_0)$ is differentiable at θ_0 with the derivative matrix D of full rank;

(iii) For every sequence $\{\varepsilon_n\}$ of positive numbers that converges to zero,

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

(v) θ_0 is an interior point of Θ.

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

$$n^{1/2}U(\theta_0;\hat{\gamma},\hat{\Lambda}_0) = n^{-1/2}\sum_{i=1}^{n} q_i(\theta_0) + \tau n^{-1/2}\sum_{i=1}^{n}\int_0^L \big\{\mathbf{Z}_i(u)I_i(u,\theta_0) - \bar{\mathbf{z}}(u)\big\}dM_i^*(u)$$
$$- \tau\mathbf{B}\mathbf{A}^{-1}n^{-1/2}\sum_{i=1}^{n}\int_0^L \big\{\mathbf{X}_i(u) - \bar{\mathbf{x}}(u)\big\}dM_i^*(u) + o_p(1),$$

which implies by the multivariate central limit theorem that

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

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

References

Chen, X., Tang, N., Zhou, Y. (2016). Quantile regression of longitudinal data with informative observation times. *Journal of Multivariate Analysis* **144**, 176-188.

Hammer, S., Vaida, F., Bennett, K., Holohan M., Sheiner, L., Eron J., Wheat L., Mitsuyasu, R., Gulick, R., Valentine F., Aberg J., Rogers, M., Karol, C., Saah, A., Lewis, R., Bessen, L., Brosgart, C., DeGruttola, V. and Mellors, J. (2002). Dual vs single protease inhibitor therapy following antiretroviral treatment failure: A randomized trial. *The Journal of the American Medical Association* **288**, 169–180.

Hughes, J. (1999). Links mixed effects models with censored data with application to HIV RNA levels. *Biometrics* **55**, 625–629.

Hunter, D. R. and Lange, K. (2000). Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics* **9**, 60–77.

Jacqmin-Gadda, H., Thiébaut, R., Chéne, G. and Commenges, D. (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics* **1**, 355–368.

Kobayashi, G. and Kozumi, H. (2012). Bayesian analysis of quantile regression for censored dynamic panel data. *Computational Statistics* **27**, 359–380.

Lange, K., Hunter, D. R. and Yang, I. (2000). Optimization transfer using surrogate objective functions (with discussion). *Journal of Computational and Graphical Statistics* **9**, 1–59.

Liang, Y., Lu, W. and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics* **65**, 377–384.

Lin, D., Wei, L., Yang, I. and Ying, Z. (2000). Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. *Journal of Royal Statistical Society, Ser. B* **62**, 711–730.

Lin, D., Wei, L. and Ying Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605–618.

Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **96**, 103–126.

Lin, H., Scharfstein, D. O. and Rosenheck, D. O. (2004), Analysis of longitudinal data with irregular outcome-Dependent follow-up. *Journal of the Royal Statistical Society, Series B* **66**, 791–813.

Liu, L., Huang, X. and O'Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* **64**, 950–958.

Lyles, R. H., Lyles, C. M. and Taylor, D. J. (2000). Random regression models for human immunodeficiency virus ribonucleic acid data subject to left censoring and informative dropouts. *Journal of the Royal Statistical Society, Series C* **49**, 485–497.

Moulton, L. H. and Halsey, N. A. (1995). A mixture model with detection limits for regression analysis of antibody response on vaccine. *Biometrics*, **51**, 1570–1578.

Nolan, D. and Pollard, D. (1987). U-processes: Rates of convergence. *The Annals of Statistics* **15**, 780–799.

Ortega, J. M. and Rheinboldt, W. C. (1970), *Iterative Solution of Nonlinear Equations in Several Variables*. Orlando: Academic Press.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

Paxton, W., Coombs, R., McElrath, M., Keefer, M., Hughes, J., Sinangil, F., Chernoff, D., Demeter, L., Williams B. and Corey L. (1997). Longitudinal analysis of quantitative virologic measures in human immunodeficiency virus-Infected subjects with $\geq$400 CD4 lymphocytes: Implications for applying measurements to individual patients. *Journal of Infectious Diseases* **175**, 247–254.

Pollard, D. (1990), *Empirical Processes: Theory and Applications*. Hayward, CA: Institute of Mathematical Statistics.

Sherman, R. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *The Annals of Statistics* **22**, 439–459.

Singh, A. and Nocerino, J. M. (2002). Robust estimation of mean and variance using environmental data sets with below detection limit observations. *Chemometrics and Intelligent Laboratory Systems*. **60**, 69–86.

Sun, J., Park, D., Sun, L. and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association 100*, 882–889.

Sun, J., Sun, L. and Liu, D. (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association* **102**, 1397–1406.

Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association*, 107, 688–700.

Sun, Y., Sun, L. and Zhou, J. (2013). Profile local linear estimation of generalized semiparametric regression model for longitudinal data. *Lifetime Data Analysis* **19**, 317–349.

Sun, Y. and Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics* **32**, 21–47.

Thiébaut, R. and Jacqmin-Gadda, H. (2004). Mixed models for longitudinal left-censored repeated measures. *Computer Methods and Programs in Biomedicine* **74**, 255–260.

Wang, H. and Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *The Annals of Statistics* **37**, 756–781.

Wang, M. C., Qin, J. and Chiang, C. T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association 96*, 1057–1065.

Xiao, L., Hou, B., Wang, Z. and Wu, Y. (2014). Random weighting approximation for Tobit regression models with longitudinal data. *Computational Statistics and Data Analysis*, **79**, 235–247.

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93**, 627–640.

Table 1: Simulation results for the estimation of θ with $\kappa = 2$.

			β_1				β_2				α			
α	Case	Method	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0	S1	TQR	0.0321	0.0793	0.0767	0.902	−0.0109	0.0543	0.0553	0.941	0.0111	0.0253	0.0259	0.925
		WF	0.0021	0.0674	0.0668	0.939	−0.0007	0.0403	0.0387	0.921	–	–	–	–
	S2	TQR	0.0187	0.0692	0.0672	0.922	0.0028	0.0424	0.0413	0.934	0.0035	0.0097	0.0110	0.972
		WF	0.0003	0.0760	0.0638	0.929	−0.0012	0.0530	0.0361	0.933	–	–	–	–
	S3	TQR	0.0373	0.1003	0.1007	0.917	−0.0114	0.0722	0.0728	0.944	0.0135	0.0342	0.0355	0.945
		WF	0.0000	0.0861	0.0844	0.926	0.0012	0.0512	0.0496	0.934	–	–	–	–
0.25	S1	TQR	0.0274	0.0747	0.0748	0.924	−0.0121	0.0542	0.0547	0.930	0.0105	0.0253	0.0260	0.935
		WF	−0.3511	0.1341	0.1077	0.106	0.3669	0.0529	0.0501	0.000	–	–	–	–
	S2	TQR	0.0233	0.0662	0.0674	0.929	0.0038	0.0414	0.0418	0.940	0.0042	0.0102	0.0116	0.968
		WF	−0.3076	0.1389	0.1216	0.265	0.3534	0.0606	0.0571	0.000	–	–	–	–
	S3	TQR	0.0400	0.0910	0.0994	0.943	−0.0106	0.0696	0.0719	0.952	0.0127	0.0338	0.0353	0.946
		WF	−0.3593	0.1582	0.1355	0.234	0.3588	0.1174	0.066	0.002	–	–	–	–
0.5	S1	TQR	0.0345	0.0711	0.0745	0.915	−0.0121	0.0539	0.0548	0.943	0.0126	0.0263	0.0262	0.909
		WF	−0.6435	0.1978	0.1631	0.024	0.6898	0.0692	0.0689	0.000	–	–	–	–
	S2	TQR	0.0233	0.0686	0.0670	0.925	0.0039	0.0410	0.0414	0.935	0.0040	0.0101	0.0116	0.973
		WF	−0.5887	0.2263	0.1848	0.116	0.6798	0.0822	0.0781	0.001	–	–	–	–
	S3	TQR	0.0408	0.0947	0.0974	0.919	−0.0117	0.0716	0.0722	0.946	0.0142	0.0344	0.0357	0.945
		WF	−0.6766	0.2467	0.2087	0.087	0.7025	0.0896	0.0916	0.001	–	–	–	–

Table 2: Simulation results for the estimation of θ with $\kappa = 4$.

			β_1				β_2				α			
α	Case	Method	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0	S1	TQR	0.0058	0.0573	0.0557	0.921	−0.0024	0.0383	0.0405	0.954	0.0012	0.0088	0.0095	0.954
		WF	−0.0026	0.0486	0.0474	0.928	−0.0009	0.0279	0.0275	0.939	–	–	–	–
	S2	TQR	0.0080	0.0464	0.0477	0.936	0.0022	0.0283	0.0295	0.949	0.0004	0.0033	0.0037	0.977
		WF	0.0008	0.0443	0.0456	0.935	0.0002	0.0249	0.0258	0.943	–	–	–	–
	S3	TQR	0.0125	0.0698	0.0706	0.943	−0.0019	0.0498	0.0519	0.948	0.0023	0.0117	0.0124	0.951
		WF	0.0006	0.0589	0.0590	0.937	0.0024	0.0349	0.0345	0.929	–	–	–	–
0.25	S1	TQR	0.0092	0.0543	0.0531	0.933	0.0001	0.0393	0.0399	0.947	0.0012	0.0096	0.0094	0.932
		WF	−0.6434	0.3827	0.1356	0.009	0.7055	0.6302	0.0517	0.000	–	–	–	–
	S2	TQR	0.0065	0.0465	0.0476	0.934	0.0019	0.0282	0.0294	0.943	0.0003	0.0033	0.0037	0.980
		WF	−0.6216	0.1983	0.1559	0.024	0.7114	0.0621	0.0583	0.000	–	–	–	–
	S3	TQR	0.0078	0.0678	0.0685	0.944	−0.0052	0.0512	0.0506	0.944	0.0018	0.0122	0.0120	0.939
		WF	−0.6945	0.1902	0.1615	0.007	0.7300	0.0641	0.0645	0.000	–	–	–	–
0.5	S1	TQR	0.0084	0.0529	0.0529	0.933	−0.0005	0.0402	0.0400	0.945	0.0010	0.0094	0.0093	0.940
		WF	−1.2499	0.3164	0.2495	0.012	1.3974	0.0928	0.0880	0.000	–	–	–	–
	S2	TQR	0.0057	0.0475	0.0475	0.932	0.0006	0.0301	0.0294	0.936	0.0007	0.0035	0.0037	0.970
		WF	−1.2386	0.3440	0.2746	0.008	1.3885	0.0930	0.0942	0.001	–	–	–	–
	S3	TQR	0.0146	0.0649	0.0675	0.944	−0.0012	0.0506	0.0507	0.939	0.0019	0.0115	0.0120	0.949
		WF	−1.3325	0.3414	0.2825	0.005	1.4262	0.1019	0.1060	0.000	–	–	–	–

Table 3: Simulation results for the estimation of θ with $\kappa = 4$ for $\tau = 0.25$ and 0.75.

			S4				S5			
τ	α		Bias	SE	SEE	CP	Bias	SE	SEE	CP
0.25	0	β_1	0.0027	0.0567	0.0576	0.943	0.0068	0.0565	0.0563	0.939
		β_2	−0.0071	0.0393	0.0416	0.941	−0.0072	0.0402	0.0408	0.934
		α	0.0009	0.0087	0.0092	0.952	0.0011	0.0090	0.0091	0.935
	0.25	β_1	0.0168	0.0679	0.0665	0.923	0.0123	0.0636	0.0639	0.927
		β_2	0.0026	0.0535	0.0510	0.940	−0.0006	0.0499	0.0509	0.958
		α	0.0026	0.0133	0.0133	0.940	0.0029	0.0126	0.0133	0.952
			S6				S7			
			Bias	SE	SEE	CP	Bias	SE	SEE	CP
0.75	0	β_1	0.0120	0.1211	0.1174	0.934	0.0048	0.1119	0.1145	0.946
		β_2	−0.0101	0.0849	0.0859	0.941	−0.0100	0.0873	0.0847	0.939
		α	0.0010	0.0183	0.0189	0.945	0.0006	0.0191	0.0189	0.938
	0.25	β_1	0.0289	0.1365	0.1340	0.918	0.0279	0.1276	0.128	0.929
		β_2	−0.0038	0.1018	0.1036	0.954	−0.0007	0.1001	0.1037	0.953
		α	0.0085	0.0269	0.0280	0.959	0.0075	0.0267	0.0282	0.964

Table 4: Comparison results on the estimation of θ for our method and a naïve method with $\kappa = 2$.

		TQR				Naive			
α		Bias	SE	SEE	CP	Bias	SE	SEE	CP
0	β_1	0.0244	0.0854	0.0887	0.968	0.2113	0.1604	0.2652	0.849
	β_2	-0.0843	0.1062	0.1355	0.913	-0.1495	0.1937	0.2446	0.919
	α	0.0438	0.0742	0.0861	0.977	0.1057	0.1490	0.1826	0.956
0.25	β_1	0.0393	0.0898	0.0980	0.958	0.2169	0.1829	0.2836	0.851
	β_2	-0.0876	0.1108	0.1412	0.927	-0.1627	0.2070	0.2632	0.901
	α	0.0524	0.0822	0.0974	0.974	0.1173	0.1668	0.2039	0.946
0.50	β_1	0.0359	0.0852	0.0924	0.962	0.2013	0.1641	0.2597	0.867
	β_2	-0.0931	0.1106	0.1445	0.917	-0.1610	0.1937	0.2518	0.900
	α	0.0518	0.0746	0.0908	0.977	0.1097	0.1464	0.1829	0.958

Table 5: Analysis results for the HIV-1 RNA data.

		TQR			WF		
τ		Est	SEE	p-value	Est	SEE	p-value
0.5	β_0	1.1163	0.0814	< 0.0001	1.2183	0.1133	< 0.0001
	β_1	1.0239	0.1192	< 0.0001	1.0509	0.1212	< 0.0001
	β_2	-2.0876	0.3753	< 0.0001	-1.8211	0.2754	< 0.0001
	α	0.0993	0.0490	0.0427	$-$	$-$	$-$
0.6	β_0	1.3991	0.0784	< 0.0001	1.5888	0.0947	< 0.0001
	β_1	0.9073	0.1413	< 0.0001	0.8937	0.1432	< 0.0001
	β_2	-1.8635	0.3696	< 0.0001	-1.0741	0.3428	0.0017
	α	0.2038	0.0346	< 0.0001	$-$	$-$	$-$
0.7	β_0	1.7414	0.0799	< 0.0001	2.0793	0.0802	< 0.0001
	β_1	0.6888	0.1127	< 0.0001	0.6765	0.1218	< 0.0001
	β_2	-2.0445	0.3219	< 0.0001	-0.7603	0.2355	0.0012
	α	0.3295	0.0361	< 0.0001	$-$	$-$	$-$

Table 6: Sensitivity analysis to the choice of $\mathbf{H}(\mathcal{F}_{it})$ for the HIV-1 RNA data with $\tau = 0.7$.

$\mathbf{H}(\mathcal{F}_{it})$		Est	SEE	p-value
$\log\{N_i(t-)+1\}$	β_0	1.8936	0.0825	< 0.0001
	β_1	0.6682	0.1219	< 0.0001
	β_2	-1.2867	0.2491	< 0.0001
	α	0.3750	0.0797	< 0.0001
$N_i(t-)^{1/2}$	β_0	1.9404	0.0837	< 0.0001
	β_1	0.6619	0.1435	< 0.0001
	β_2	-1.1303	0.2662	< 0.0001
	α	0.2200	0.0611	0.0004