BIOMETRICS 61, 151–161
March 2005

Quantile Regression Models with Multivariate Failure Time Data

Guosheng Yin

Department of Biostatistics, M. D. Anderson Cancer Center, The University of Texas,
Houston, Texas 77030, U.S.A.
email: gsyin@mdanderson.org

and

Jianwen Cai

Department of Biostatistics, The University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina 27599, U.S.A.

SUMMARY. As an alternative to the mean regression model, the quantile regression model has been studied extensively with independent failure time data. However, due to natural or artificial clustering, it is common to encounter multivariate failure time data in biomedical research where the intracluster correlation needs to be accounted for appropriately. For right-censored correlated survival data, we investigate the quantile regression model and adapt an estimating equation approach for parameter estimation under the working independence assumption, as well as a weighted version for enhancing the efficiency. We show that the parameter estimates are consistent and asymptotically follow normal distributions. The variance estimation using asymptotic approximation involves nonparametric functional density estimation. We employ the bootstrap and perturbation resampling methods for the estimation of the variance–covariance matrix. We examine the proposed method for finite sample sizes through simulation studies, and illustrate it with data from a clinical trial on otitis media.

KEY WORDS: Bootstrap; Correlation; Estimating equations; Kaplan–Meier estimator; Perturbation; Regression quantile.

1. Introduction

Quantiles are useful summary statistics to characterize the survival experience of patients. Compared to the mean survival time, quantiles are more robust against outliers. Regression models based on a set of properly selected quantiles may give a global assessment of covariate effects. As an important alternative to the mean regression model for examining the covariate effects on a continuous outcome variable, semiparametric quantile regression models have been extensively studied (Koenker and Bassett, 1978; Jung, 1996; Portnoy and Koenker, 1997). The parameter estimation is based on estimating equations that can often be solved by the linear programming or iterative bisection methods. The variances of regression coefficient estimates typically depend on the density function of error terms. To avoid the nonparametric functional density estimation, a variety of resampling methods have been proposed to estimate the standard errors of the parameter estimates (Parzen, Wei, and Ying, 1994; Buchinsky, 1995; Hahn, 1995; Horowitz, 1998; Bilias, Chen, and Ying, 2000; Jin, Ying, and Wei, 2001). Recently, quantile regression has been generalized to model survival data (Ying, Jung, and Wei, 1995; Lindgren, 1997; Yang, 1999; Koenker and Geling, 2001; McKeague, Subramanian, and Sun, 2001; Tian and Wei, 2002; Portnoy, 2003). This is a natural extension since quantile-based regression is robust against outliers, and the survival times are usually highly right-skewed. In econometrics, much research has been conducted for the type I censoring cases, where the event is observed only if it occurs prior to some prespecified time, namely the "Tobit" model (Powell, 1984; Fitzenberger, 1997; Buchinsky and Hahn, 1998; Khan and Powell, 2001).

Ying et al. (1995) proposed the median regression model for independent failure time data. Their method involves minimizing discrete and nonsmooth functions, which may have multiple local minima. The standard optimization algorithms are not practically suitable for this minimization problem, while the simulated annealing algorithm (Lin and Geyer, 1992) was suggested. Since the limiting variance–covariance matrix of the parameter estimates depends on the unknown density function, the minimum dispersion test statistic was used for inference. McKeague et al. (2001) applied the missing information principle to replace the censored data by an estimator using the conditional expectation based on the observed data. Tian and Wei (2002) proposed an iterative method for the parameter estimation, and an algorithm to estimate the variance–covariance matrix by appropriately perturbing the estimating function (Jin et al., 2001). To deal with dependent censoring, Bang and Tsiatis (2002) studied a median regression model with univariate censored medical cost data, which may lead to more cost-effective therapies and more efficient medical intervention. In quantile regression analysis, one can do such regressions simultaneously for a set of properly selected quantiles. It provides a complete picture and may show different important covariate effects at different follow-up stages. Some covariates may have a significant effect at an early period of follow-up, and no effect later on, or vice versa. Especially in the situation where the sign of the effect may change over the duration of follow-up, the quantile regression model offers a natural analytic approach. For example, some prospective cohort studies have shown that men seem to have a lower mortality rate than women before 60 years of age and a higher mortality rate after 79 years of age after hospital discharge for myocardial infarction. Quantile regression, therefore, is a suitable method for these studies.

To ensure the validity of the aforementioned methods, a common and critical assumption is that the failure times are independent. However, in many biomedical studies, this assumption of independence may not hold due to natural or artificial clustering (e.g., dental or litter-matched mice studies). One interesting example is a clinical trial involving children with inflammation of the middle ear, otitis media (OM), which is one of the most common childhood infections (Le and Lindgren, 1996). A small tunnel (Eustachian tube) serves to equalize the air pressure between the middle ear and the outer side of the eardrum. Bacteria and viruses can enter the middle ear through the Eustachian tube (shorter in children than adults), which may cause the middle ear to be filled with fluid and sometimes pus, with an accompanying loss of hearing in that ear. Even temporary periods of hearing loss in young children (between 6 months and 6 years of age) can cause delays in speech, behavior, and language development, and learning. Inserting ventilating tubes into the infected ears has been shown to reduce the incidence of OM episodes and improve hearing as long as the tubes are in place and working. Every year, about 1 million children receive ventilating tubes in the United States. The aim of this clinical trial is to examine whether a medical treatment (prednisone and sulfamethoprim) prolongs the life of the ventilating tubes, where the failure of a tube is defined as the cessation of tube functioning (blocked) or tube extrusion. A straightforward application of the existing quantile regression approaches is not appropriate because the paired observations from the two ears of each child are clearly not independent.

For highly stratified correlated failure time data, extensive research has been carried out on hazard-based regression. Lee, Wei, and Amato (1992) proposed a multiplicative intensity model where they proposed an estimating equation under the working independence model and showed that the parameter estimates are consistent and asymptotically normally distributed. Lee, Wei, and Ying (1993) applied the linear regression model or the accelerated failure time (AFT) model to the clustered censored data. More recently, Cai, Wei, and Wilcox (2000) studied a class of linear transformation models, which includes the Cox proportional hazards model and the proportional odds model as two special cases.

In this article, we propose the quantile regression model for correlated failure time data. We adapt an estimating equation approach (Ying et al., 1995) for parameter estimation, derive the asymptotic theories, and outline an estimating procedure for a weighted version of quantile regression. For the estimation of the variance–covariance matrix, we employ and compare two different resampling methods. Our resampling schemes take a cluster instead of an individual observation as a sampling unit. In particular, with relatively large cluster sizes, we propose a two-stage version of resampling.

The rest of this article is organized as follows. In Section 2.1, we introduce the semiparametric quantile regression models for correlated survival data. In Section 2.2, we describe the estimating procedure based on an iterative algorithm. In Section 2.3, we derive several asymptotic properties for the proposed model. In Section 2.4, we propose a weighted version of the estimating equation by accounting for the within-cluster correlation. In Section 3, we study the bootstrap and perturbation resampling methods for variance estimation. In Section 4, we conduct simulation studies to investigate the performance of the quantile regression model with finite sample sizes. In Section 5, we apply the proposed method to the OM data for illustration. We provide the concluding remarks in Section 6, and outline the theorem proofs in the Appendix.

2. Quantile Regression Models

2.1 *Model Formulation*

Let T_{ik} $(i = 1, \ldots, n;\ k = 1, \ldots, K_i)$ be the failure time or its log transformation for the kth subject in the ith cluster, and let $\mathbf{Z}_{ik}$ be the corresponding $p \times 1$ vector of bounded covariates. Assume that the K_i's are bounded, i.e., for a constant $K < \infty$, $K_i \leq K$ for all $i = 1, \ldots, n$. Let C_{ik} be the censoring variable and $X_{ik} = T_{ik} \wedge C_{ik}$ be the observed time, where "$\wedge$" denotes the minimum and "$\vee$" denotes the maximum of the two values. The failure time indicator $\Delta_{ik} = 1$ if T_{ik} is observed, and $\Delta_{ik} = 0$ otherwise. Assume that C_{ik} is independent of both T_{ik} and $\mathbf{Z}_{ik}$. Within each cluster, $\{(T_{ik}, C_{ik}, \mathbf{Z}_{ik}),$ $k = 1, \ldots, K_i\}$ may be dependent but exchangeable. The cluster size may vary for different clusters by setting $X_{ik} = 0$ and $\Delta_{ik} = 0$ if the kth failure time is not applicable to the ith cluster. For technical reasons, we assume that potentially every cluster has K members. Define $\mathbf{T}_i = (T_{i1}, \ldots, T_{iK})'$, and similarly for $\mathbf{C}_i$ and $\mathbf{Z}_i$. We assume that $\{(\mathbf{T}_i, \mathbf{C}_i, \mathbf{Z}_i),$ $i = 1, \ldots, n\}$ are independent and identically distributed.

Along the line of the AFT model (Lee et al., 1993), the quantile regression model substitutes the $100(1 - \tau)$th percentile of failure times, $\xi_\tau(T_{ik} \mid \mathbf{Z}_{ik})$, for the traditional mean,

$$\xi_\tau(T_{ik} \mid \mathbf{Z}_{ik}) = \boldsymbol{\beta}'_{0\tau}\mathbf{Z}_{ik}, \tag{1}$$

where $\tau \in (0, 1)$ and $\boldsymbol{\beta}_{0\tau}$ is the true coefficient vector. Let $\epsilon_{\tau,ik} = T_{ik} - \boldsymbol{\beta}'_{0\tau}\mathbf{Z}_{ik}$, then $\xi_\tau(\epsilon_{\tau,ik} \mid \mathbf{Z}_{ik}) = 0$. The conditional distribution of $\epsilon_{\tau,ik}$ is completely unspecified and may depend on the covariate vector $\mathbf{Z}_{ik}$. For correlated data without any censoring, under the working independence assumption (Liang and Zeger, 1986), we propose to estimate $\boldsymbol{\beta}_{0\tau}$ by minimizing $n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\phi_\tau(T_{ik} - \boldsymbol{\beta}'_\tau\mathbf{Z}_{ik})$, where $\phi_\tau(u) = u\{I(u \geq 0) - \tau\}$. The estimating equation for $\boldsymbol{\beta}_{0\tau}$ is thus given by $n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\mathbf{Z}_{ik}\{I(T_{ik} - \boldsymbol{\beta}'_\tau\mathbf{Z}_{ik} \geq 0) - \tau\} = 0$. For the special case of median regression, with $\tau = 1/2$, it reduces to the least absolute deviation estimation.

For correlated survival data, let $G(t)$ be the common survival function of the censoring variable and $\hat{G}(t)$

be the corresponding Kaplan–Meier estimator based on $\{(X_{ik}, 1-\Delta_{ik}), k=1,\ldots,K_i; i=1,\ldots,n\}$, $\hat{G}(t) = \prod_{u\leq t}\{1-d\bar{N}_G(u)/\bar{Y}(u)\}$, where the counting process $N_{Gik}(t) = I(X_{ik}\leq t, \Delta_{ik}=0)$, $\bar{Y}(t) = \sum_{i=1}^n\sum_{k=1}^{K_i} I(X_{ik}\geq t)$, $\bar{N}_G(t) = \sum_{i=1}^n\sum_{k=1}^{K_i} N_{Gik}(t)$, and $d\bar{N}_G(t) = \bar{N}_G(t) - \bar{N}_G(t-)$. The usual Nelson–Aalen estimator for the cumulative hazard of the censoring time is defined as $\hat{\Lambda}_G(t) = \int_0^t d\bar{N}_G(u)/\bar{Y}(u)$. Ying and Wei (1994) proved the consistency of $\hat{G}(t)$ under the ϕ-mixing condition. Observe the fact that conditional on $\mathbf{Z}_{ik}$, $\Pr(X_{ik} - \boldsymbol{\beta}_{0\tau}'\mathbf{Z}_{ik}\geq 0) = G(\boldsymbol{\beta}_{0\tau}'\mathbf{Z}_{ik})\tau$. Under the working independence correlation matrix, we propose the estimating function for $\boldsymbol{\beta}_{0\tau}$ as

$$\mathbf{S}_n(\boldsymbol{\beta}_\tau) = n^{-1}\sum_{i=1}^n \mathbf{Z}_i'\hat{\mathbf{e}}_i(\boldsymbol{\beta}_\tau, \mathbf{X}_i, \mathbf{Z}_i), \tag{2}$$

where $\mathbf{Z}_i = (\mathbf{Z}_{i1},\ldots,\mathbf{Z}_{iK})'$, $\mathbf{X}_i = (X_{i1},\ldots,X_{iK})'$, $\hat{\mathbf{e}}_i = (\hat{e}_{i1},\ldots,\hat{e}_{iK})'$, and

$$\hat{e}_{ik}(\boldsymbol{\beta}_\tau, X_{ik}, \mathbf{Z}_{ik}) = \frac{I(X_{ik} - \boldsymbol{\beta}_\tau'\mathbf{Z}_{ik}\geq 0)}{\hat{G}(\boldsymbol{\beta}_\tau'\mathbf{Z}_{ik})} - \tau.$$

In practice, if $\hat{G}(\boldsymbol{\beta}_\tau'\mathbf{Z}_{ik}) = 0$, we set $I(X_{ik} - \boldsymbol{\beta}_\tau'\mathbf{Z}_{ik}\geq 0)/\hat{G}(\boldsymbol{\beta}_\tau'\mathbf{Z}_{ik}) = 0$.

2.2 *Estimating Procedures*

Due to the discontinuity of $\mathbf{S}_n(\boldsymbol{\beta}_\tau)$, a unique solution in $\mathbf{S}_n(\boldsymbol{\beta}_\tau) = 0$ is usually unavailable. One may obtain a solution $\hat{\boldsymbol{\beta}}_\tau$ by minimizing the Euclidean norm $\|\mathbf{S}_n(\boldsymbol{\beta}_\tau)\|$. A numerical challenge arises for the minimization of the nonsmooth estimating function that cannot be achieved via standard optimization techniques, such as the Newton–Raphson algorithm. We adapt an efficient iterative procedure (Tian and Wei, 2002) using the Nelder–Mead simplex algorithm to estimate the parameters.

For notational simplicity and clarity, we suppress the subindex "τ" such that $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}$ denote the true parameter and the corresponding estimator, respectively. To obtain a root to the equation $\mathbf{S}_n(\boldsymbol{\beta}) = 0$, the iterative algorithm is described below.

Step 1: Obtain an initial value $\hat{\boldsymbol{\beta}}_{[0]}$ by minimizing a convex function of $\boldsymbol{\beta}$,

$$n^{-1}\sum_{i=1}^n\sum_{k=1}^{K_i}\frac{\Delta_{ik}}{\hat{G}(X_{ik})}\phi_\tau(X_{ik} - \boldsymbol{\beta}'\mathbf{Z}_{ik}). \tag{3}$$

Step 2: Update the estimating function by replacing $\boldsymbol{\beta}$ in $\hat{G}(\boldsymbol{\beta}'\mathbf{Z}_{ik})$ with $\hat{\boldsymbol{\beta}}_{[0]}$,

$$\mathbf{S}_{[1]}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n\sum_{k=1}^{K_i}\mathbf{Z}_{ik}\left\{\frac{I(X_{ik} - \boldsymbol{\beta}'\mathbf{Z}_{ik}\geq 0)}{\hat{G}\big(\hat{\boldsymbol{\beta}}_{[0]}'Z_{ik}\big)} - \tau\right\}.$$

A root $\hat{\boldsymbol{\beta}}_{[1]}$ of $\mathbf{S}_{[1]}(\boldsymbol{\beta}) = 0$ can be obtained by minimizing the convex function of $\boldsymbol{\beta}$,

$$C_{[1]}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n\sum_{k=1}^{K_i}\left\{\frac{(X_{ik} - \boldsymbol{\beta}'\mathbf{Z}_{ik})I(X_{ik} - \boldsymbol{\beta}'\mathbf{Z}_{ik}\geq 0)}{\hat{G}\big(\hat{\boldsymbol{\beta}}_{[0]}'\mathbf{Z}_{ik}\big)} - (X_{ik} - \boldsymbol{\beta}'\mathbf{Z}_{ik})\tau\right\}.$$

The Nelder–Mead simplex algorithm can be used here, which directly evaluates the target function at each point of the simplex without requiring any derivatives or continuity properties.

Step 3: Replace $\hat{\boldsymbol{\beta}}_{[0]}$ in $\hat{G}(\cdot)$ with $\hat{\boldsymbol{\beta}}_{[1]}$, and minimize the convex function $C_{[2]}(\boldsymbol{\beta})$ to obtain its minimizer $\hat{\boldsymbol{\beta}}_{[2]}$. Then go back to Step 2, and continue this procedure until the prescribed convergence criteria are met.

Note that (3) is basically the inverse probability weighted estimating equation which is reminiscent of the Horvitz–Thompson estimator (1952). Finally, we have a sequence of well-defined $\hat{\boldsymbol{\beta}}_{[m]}(m=0,\ldots,M)$. Following arguments similar to those in Appendix I of Tian and Wei (2002) for independent failure time data, it can be shown that $\|\hat{\boldsymbol{\beta}}_{[0]} - \boldsymbol{\beta}_0\| = o(n^{-1/2+\epsilon})$ a.s. with $\epsilon > 0$, and for $m = 1,\ldots,M$, $\|\hat{\boldsymbol{\beta}}_{[m]} - \boldsymbol{\beta}_0\| = o(n^{-1/2+\epsilon})$ a.s. by the method of induction.

2.3 *Asymptotic Properties*

Now, we study the asymptotic theories under the regularity conditions as given in the Appendix. Let $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{[M]}$, and the strong consistency of $\hat{\boldsymbol{\beta}}$ is given in the following.

THEOREM 1: *As $n\to\infty, \hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}_0$ almost surely.*

Let $M_{Gik}(t) = N_{Gik}(t) - I(X_{ik}\geq t)\Lambda_G(t)$ be the censoring time martingale with respect to the marginal filtration $\mathcal{F}_{ik}(t) = \sigma\{(N_{ik}(s), Y_{ik}(s+), \mathbf{Z}_{ik}), 0\leq s\leq t\}$. However, $\bar{M}_G(t) = \sum_{i=1}^n\sum_{k=1}^{K_i} M_{Gik}(t)$ is not a martingale due to the intraclass correlation. The asymptotic normality of $n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0)$ can be proved using the empirical process techniques and the multivariate central limit theorem (CLT).

THEOREM 2: *As $n\to\infty$, $n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0)$ is normal with mean zero and variance–covariance matrix $\boldsymbol{\Gamma} = \mathrm{E}(\boldsymbol{\xi}_1\boldsymbol{\xi}_1')$, where $\boldsymbol{\xi}_i$ $(i=1,\ldots,n)$ is given by*

$$\boldsymbol{\xi}_i = \sum_{k=1}^{K_i}\left[\mathbf{Z}_{ik}\left\{\frac{I(X_{ik}\geq\boldsymbol{\beta}_0'\mathbf{Z}_{ik})}{G(\boldsymbol{\beta}_0'\mathbf{Z}_{ik})} - \tau\right\} + \tau\int_0^\infty\frac{\mathbf{q}(t)}{\pi(t)}\,dM_{Gik}(t)\right], \tag{4}$$

and $\mathbf{q}(t) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n\sum_{k=1}^{K_i}\mathbf{Z}_{ik}I(\boldsymbol{\beta}_0'\mathbf{Z}_{ik}\geq t)$, $\pi(t) = \lim_{n\to\infty} n^{-1}\bar{Y}(t)$.

A consistent estimator $\hat{\boldsymbol{\Gamma}}$ can be obtained by replacing $G(\cdot)$, $\boldsymbol{\beta}$, $\pi(t)$, and $\Lambda_G(t)$ with their empirical counterparts $\hat{G}(\cdot)$, $\hat{\boldsymbol{\beta}}$, $n^{-1}\bar{Y}(t)$, and $\hat{\Lambda}_G(t)$, respectively, i.e., $\hat{\boldsymbol{\Gamma}} = n^{-1}\sum_{i=1}^n\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i'$ where

$$\hat{\boldsymbol{\xi}}_i = \sum_{k=1}^{K_i}\left[\mathbf{Z}_{ik}\left\{\frac{I(X_{ik}\geq\hat{\boldsymbol{\beta}}'\mathbf{Z}_{ik})}{\hat{G}(\hat{\boldsymbol{\beta}}'\mathbf{Z}_{ik})} - \tau\right\} + \tau\int_0^\infty\frac{\sum_{j=1}^n\sum_{l=1}^{K_j}\mathbf{Z}_{jl}I(\hat{\boldsymbol{\beta}}'\mathbf{Z}_{jl}\geq t)}{\sum_{j=1}^n\sum_{l=1}^{K_j}I(X_{jl}\geq t)} \times\{dN_{Gik}(t) - I(X_{ik}\geq t)\,d\hat{\Lambda}_G(t)\}\right].$$

By proving the linearity property of $\mathbf{S}_n(\boldsymbol{\beta})$ in a local neighborhood of $\boldsymbol{\beta}_0$ and applying Taylor's series expansion, we conclude the following theorem.

THEOREM 3: *As* $n \to \infty$, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *is normally distributed with mean zero and a sandwich variance–covariance matrix* $\mathbf{A}^{-1}\boldsymbol{\Gamma}\mathbf{A}^{-1}$, *where* $\mathbf{A} = -\zeta E[f(0\,|\,\mathbf{Z}_{11})\mathbf{Z}_{11}\mathbf{Z}_{11}']$, $\zeta = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} K_i$, *and* $f(\cdot\,|\,\mathbf{Z}_{ik})$ *is the density of* ϵ_{ik} *conditional on* $\mathbf{Z}_{ik}$.

We outline the proofs of the three theorems in the Appendix. Based on the quantile regression model, we can estimate or predict the quantiles of the survival time for a given subject, and construct the confidence interval accordingly. However, the variance of the parameter estimate depends on the density of errors. The density could be estimated by some nonparametric smoothing techniques, but it might not be reliable for sample sizes of practical use, especially with correlated error terms. In Section 3, we propose to employ two resampling methods for the estimation of the variance–covariance matrix, i.e., a blockwise bootstrap method (Künsch, 1989) and a cluster-based perturbation procedure.

2.4 *Weighted Estimating Equations*

Efficiency might be gained by incorporating an appropriate weight function to account for the intraclass correlation. Analogous to Cai and Prentice (1995, 1997), we consider the weighted estimating equations,

$$\mathbf{S}_n^W(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} \mathbf{Z}_i'\mathbf{W}\hat{\mathbf{e}}_i(\boldsymbol{\beta}, \mathbf{X}_i, \mathbf{Z}_i),$$

where the $K \times K$ weight matrix $\mathbf{W}$ is included in an attempt to enhance the efficiency of the estimation of $\boldsymbol{\beta}$ by accounting for the correlation within clusters. Setting $\mathbf{S}_n^W(\boldsymbol{\beta}) = 0$, we have

$$n^{-1}\sum_{i=1}^{n}\begin{bmatrix}\mathbf{Z}_{i1}'\\ \vdots\\ \mathbf{Z}_{iK}'\end{bmatrix}'\begin{bmatrix}W_{11} & \cdots & W_{1K}\\ \vdots & \ddots & \vdots\\ W_{K1} & \cdots & W_{KK}\end{bmatrix}\begin{bmatrix}\frac{I(X_{i1}-\beta'\mathbf{Z}_{i1}\geq 0)}{\hat{G}(\beta'\mathbf{Z}_{i1})} - \tau\\ \vdots\\ \frac{I(X_{iK}-\beta'\mathbf{Z}_{iK}\geq 0)}{\hat{G}(\beta'\mathbf{Z}_{iK})} - \tau\end{bmatrix} = 0.$$

To facilitate the computation, we split the weight matrix $\mathbf{W}$ into an identity matrix $\mathbf{I}$ plus the difference $\mathbf{W} - \mathbf{I}$, and by rearranging terms, we have

$$n^{-1}\sum_{i=1}^{n}\begin{bmatrix}\mathbf{Z}_{i1}'\\ \vdots\\ \mathbf{Z}_{iK}'\end{bmatrix}'\begin{bmatrix}\frac{I(X_{i1}-\beta'\mathbf{Z}_{i1}\geq 0)}{\hat{G}(\beta'\mathbf{Z}_{i1})} - \tau\\ \vdots\\ \frac{I(X_{iK}-\beta'\mathbf{Z}_{iK}\geq 0)}{\hat{G}(\beta'\mathbf{Z}_{iK})} - \tau\end{bmatrix}$$

$$= n^{-1}\sum_{i=1}^{n}\begin{bmatrix}\mathbf{Z}_{i1}'\\ \vdots\\ \mathbf{Z}_{iK}'\end{bmatrix}'\begin{bmatrix}1-W_{11} & \cdots & -W_{1K}\\ \vdots & \ddots & \vdots\\ -W_{K1} & \cdots & 1-W_{KK}\end{bmatrix}$$

$$\times\begin{bmatrix}\frac{I(X_{i1}-\beta'\mathbf{Z}_{i1}\geq 0)}{\hat{G}(\beta'\mathbf{Z}_{i1})} - \tau\\ \vdots\\ \frac{I(X_{iK}-\beta'\mathbf{Z}_{iK}\geq 0)}{\hat{G}(\beta'\mathbf{Z}_{iK})} - \tau\end{bmatrix}.$$

Denote the right-hand side of the above equation by $\mathbf{U}(\boldsymbol{\beta})$, and $\mathbf{U}(\boldsymbol{\beta}) = 0$ if $\mathbf{W} = \mathbf{I}$. To obtain a root to the equation $\mathbf{S}_n^W(\boldsymbol{\beta}) = 0$, we devise a modified iterative algorithm, as described below.

Step 1: Obtain an initial value $\hat{\boldsymbol{\beta}}_{[0]}$ as in the unweighted version.

Step 2: Update $\mathbf{S}_{[1]}^W(\boldsymbol{\beta})$ by replacing $\boldsymbol{\beta}$ in $\hat{G}(\boldsymbol{\beta}'\mathbf{Z}_{ik})$ and $\boldsymbol{\beta}$ in $\mathbf{U}(\boldsymbol{\beta})$ with $\hat{\boldsymbol{\beta}}_{[0]}$,

$$\mathbf{S}_{[1]}^W(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\mathbf{Z}_{ik}\left\{\frac{I(X_{ik}-\boldsymbol{\beta}'\mathbf{Z}_{ik}\geq 0)}{\hat{G}\left(\hat{\boldsymbol{\beta}}_{[0]}'Z_{ik}\right)} - \tau\right\} - \mathbf{U}\left(\hat{\boldsymbol{\beta}}_{[0]}\right).$$

A root $\hat{\boldsymbol{\beta}}_{[1]}$ of $\mathbf{S}_{[1]}^W(\boldsymbol{\beta}) = 0$ can be equivalently obtained by minimizing the convex function of $\boldsymbol{\beta}$,

$$C_{[1]}^W(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\left\{\frac{(X_{ik}-\boldsymbol{\beta}'\mathbf{Z}_{ik})I(X_{ik}-\boldsymbol{\beta}'\mathbf{Z}_{ik}\geq 0)}{\hat{G}\left(\hat{\boldsymbol{\beta}}_{[0]}'\mathbf{Z}_{ik}\right)} - (X_{ik}-\boldsymbol{\beta}'\mathbf{Z}_{ik})\tau\right\} + \boldsymbol{\beta}'\mathbf{U}\left(\hat{\boldsymbol{\beta}}_{[0]}\right).$$

Step 3: Replace $\hat{\boldsymbol{\beta}}_{[0]}$ in $\hat{G}(\cdot)$ and $\hat{\boldsymbol{\beta}}_{[0]}$ in $\mathbf{U}(\cdot)$ with $\hat{\boldsymbol{\beta}}_{[1]}$, and obtain the corresponding convex function $C_{[2]}^W(\boldsymbol{\beta})$ and its minimizer $\hat{\boldsymbol{\beta}}_{[2]}$. Go back to Step 2, and continue until convergence.

As mentioned by Cai and Prentice (1995, 1997), the efficiency gain would not be high by incorporating a weight function for many correlated failure time analyses, unless dependencies are very strong, sample sizes are large, and censoring is not severe. Because the stochastic dependence among multivariate failure times is often not very strong in practice, we focus on estimates based on the estimating equation with the identity weight matrix.

3. Resampling Methods for Variance Estimation

We describe and compare the bootstrap and perturbation resampling methods to estimate the variance–covariance matrix. We consider the one-stage and two-stage resampling schemes according to the cluster sizes. For the one-stage bootstrap method, we take each cluster as the sampling unit, and thus the bootstrap sample is composed of a simple random sample of clusters with replacement such that the intracluster correlation structure is preserved. Specifically, we draw a simple random sample $\{(\mathbf{X}_i^*, \boldsymbol{\Delta}_i^*, \mathbf{Z}_i^*), i = 1, \ldots, n\}$ with replacement from the original data $\{(\mathbf{X}_i, \boldsymbol{\Delta}_i, \mathbf{Z}_i), i = 1, \ldots, n\}$ with equal probability $1/n$. For the two-stage version, we first sample clusters, then within each selected cluster, randomly sample individuals with replacement. We repeat the resampling procedure B times and estimate the variance–covariance matrix based on the sample statistics of $(\hat{\boldsymbol{\beta}}_1^*, \ldots, \hat{\boldsymbol{\beta}}_B^*)$.

We generalize the perturbation method (Jin et al., 2001) for our model setting. Let $(U_1, \ldots, U_n)$ be a simple random sample from a positive random variable U with mean

1 and variance 1, e.g., Exp(1). With the martingale notation (Fleming and Harrington, 1991),

$$\hat{G}(t) \approx G(t) - G(t)\sum_{i=1}^{n}\sum_{k=1}^{K_i}\left\{\frac{I(X_{ik} \le t)(1-\Delta_{ik})}{\bar{Y}(X_{ik})} - \int_0^t \frac{I(X_{ik} \ge u)\, d\Lambda_G(u)}{\bar{Y}(u)}\right\}.$$

By perturbing the ith cluster by U_i, and replacing $G(t)$ and $\Lambda_G(t)$ by $\hat{G}(t)$ and $\hat{\Lambda}_G(t)$, the perturbed version of $\hat{G}(t)$ is then defined as

$$\hat{G}^*(t) = \hat{G}(t) - \hat{G}(t)\sum_{i=1}^{n}\left[\sum_{k=1}^{K_i}\left\{\frac{I(X_{ik} \le t)(1-\Delta_{ik})}{\bar{Y}(X_{ik})} - \int_0^t \frac{I(X_{ik} \ge u)\, d\hat{\Lambda}_G(u)}{\bar{Y}(u)}\right\}\right]U_i.$$

The corresponding perturbed version of the estimating function is

$$\mathbf{S}_n^*(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\mathbf{Z}_{ik}\left\{\frac{I(X_{ik} - \boldsymbol{\beta}'\mathbf{Z}_{ik} \ge 0)}{\hat{G}^*(\boldsymbol{\beta}'\mathbf{Z}_{ik})} - \tau\right\}U_i. \quad (5)$$

The iterative method proposed in Section 2.2 can be used to obtain the solution $\hat{\boldsymbol{\beta}}^*$ of $\mathbf{S}_n^*(\boldsymbol{\beta}) = 0$. It can be shown that, for some constant $\tilde{t}, \sup_{0 \le t \le \tilde{t}} |\hat{G}^*(t) - G(t)| = o(n^{-1/2+\epsilon})$ almost surely. Note that the U_i's are independent of the data $\{(X_{ik}, \Delta_{ik}, \mathbf{Z}_{ik}), k = 1, \ldots, K_i;\ i = 1, \ldots, n\}$, and thus conditional on the observed data, the only random variable in (5) is U_i. We can write $n^{1/2}\mathbf{S}_n^*(\hat{\boldsymbol{\beta}})$ as a sum of n independent random vectors. By the Linderberg–Feller CLT, the conditional distribution of $n^{1/2}\mathbf{S}_n^*(\hat{\boldsymbol{\beta}})$, given the observed data, is asymptotically normal with mean zero, and the conditional covariance function of $n^{1/2}\mathbf{S}_n^*(\hat{\boldsymbol{\beta}})$ converges to that of $n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0)$. Furthermore, by Taylor's series expansion, it can be shown that the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})$ is equivalent to that of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$.

We examine the following two versions: one-stage perturbation, in which we only generate cluster-level perturbing random variables to jitter or shift the estimating function of the whole cluster at the same level, such that the intracluster correlation is kept intact; and two-stage perturbation, where we first perturb the cluster-level estimating function, then within each cluster (conditional on the perturbed cluster), we perturb each individual estimating function. By generating many independent sets of random samples of $(U_1, \ldots, U_n)$, the covariance matrix of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ can be estimated

Table 1

Parameter estimation bias, standard deviation (SD), standard error (SE), and 95% confidence interval coverage rate (CR) with cluster size of 2

					Bootstrap		Perturb				Bootstrap		Perturb	
n	c%	ρ	$\Delta\beta_0$	SD	SE	CR	SE	CR	$\Delta\beta_1$	SD	SE	CR	SE	CR
25	20	0.2	0.022	0.299	0.292	93.4	0.288	91.8	−0.031	0.418	0.425	94.2	0.419	93.4
		0.5	0.018	0.294	0.294	93.4	0.289	92.8	−0.030	0.373	0.397	95.0	0.390	94.4
		0.8	0.017	0.292	0.291	94.2	0.286	92.0	−0.028	0.320	0.353	96.4	0.347	96.2
	40	0.2	0.026	0.338	0.343	93.6	0.320	91.2	−0.028	0.504	0.542	93.4	0.515	92.6
		0.5	0.029	0.341	0.345	92.4	0.323	90.4	−0.034	0.480	0.520	94.2	0.496	93.0
		0.8	0.039	0.348	0.341	92.2	0.321	91.4	−0.040	0.452	0.494	96.4	0.469	95.2
	60	0.2	0.022	0.378	0.423	94.0	0.364	89.6	−0.018	0.642	0.703	96.4	0.619	91.6
		0.5	0.037	0.399	0.423	92.8	0.368	89.2	−0.033	0.643	0.686	95.6	0.610	91.8
		0.8	0.029	0.406	0.416	92.4	0.361	87.6	−0.033	0.620	0.665	95.8	0.591	91.8
50	20	0.2	0.014	0.199	0.206	93.0	0.202	92.4	−0.001	0.290	0.300	94.6	0.297	94.4
		0.5	0.020	0.206	0.206	92.6	0.203	93.4	−0.006	0.272	0.278	93.8	0.276	93.6
		0.8	0.010	0.202	0.206	93.4	0.201	92.6	0.004	0.232	0.248	95.2	0.246	95.2
	40	0.2	0.020	0.238	0.241	92.4	0.230	89.6	−0.014	0.368	0.379	93.6	0.365	93.0
		0.5	0.024	0.242	0.242	92.8	0.231	92.0	−0.021	0.350	0.365	95.0	0.351	93.8
		0.8	0.015	0.229	0.241	92.6	0.230	91.6	−0.010	0.323	0.345	94.6	0.333	93.8
	60	0.2	0.020	0.291	0.299	91.8	0.270	90.8	−0.017	0.470	0.505	95.0	0.462	93.4
		0.5	0.030	0.289	0.298	93.6	0.268	90.8	−0.031	0.450	0.497	95.6	0.452	93.4
		0.8	0.018	0.277	0.301	92.8	0.269	90.8	−0.020	0.417	0.484	96.4	0.440	95.4
100	20	0.2	−0.004	0.140	0.145	95.0	0.143	93.2	0.006	0.197	0.211	94.4	0.208	95.0
		0.5	−0.002	0.141	0.145	93.6	0.144	92.0	0.004	0.183	0.196	94.0	0.194	93.8
		0.8	−0.004	0.142	0.145	94.2	0.144	93.4	0.007	0.160	0.174	95.2	0.173	96.0
	40	0.2	−0.005	0.163	0.170	93.4	0.165	93.0	−0.001	0.252	0.268	94.6	0.260	94.0
		0.5	0.000	0.163	0.172	94.0	0.166	93.0	−0.004	0.238	0.257	95.6	0.249	94.6
		0.8	−0.002	0.164	0.172	94.2	0.167	93.8	−0.001	0.218	0.242	96.2	0.235	95.8
	60	0.2	−0.005	0.198	0.209	93.4	0.192	90.4	0.008	0.330	0.353	93.8	0.328	92.4
		0.5	−0.006	0.198	0.211	92.6	0.193	90.6	0.011	0.324	0.347	95.4	0.321	93.8
		0.8	−0.007	0.208	0.210	91.0	0.194	89.4	0.013	0.318	0.336	94.2	0.314	94.0

using the standard sample statistics. When the cluster sizes are relatively large, the two-stage resampling approaches are expected to account for both the between-cluster variation and the within-cluster variation, while the one-stage resampling methods are expected to work better with small cluster sizes of 2–5.

4. Simulation Studies

To investigate the finite sample properties of the proposed methodologies, we carried out extensive simulation studies. Without loss of generality, we focus on the median regression models with $\tau = 0.5$. The model for generating the failure time is given by

$$T_{ik} = \beta_0 + \beta_1 Z_{ik} + \epsilon_{ik}, \quad (k = 1, \ldots, K;\ i = 1, \ldots, n), \quad (6)$$

where the true values of the intercept $\beta_0 = 2$, the slope $\beta_1 = 1$, and the errors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iK})'$ follow a multivariate normal distribution with mean zero and a compound symmetric variance–covariance matrix $\boldsymbol{\Sigma}$. We define $\boldsymbol{\Sigma} = \rho\sigma^2\mathbf{1}\mathbf{1}' + (1-\rho)\sigma^2\mathbf{I}$, where $\mathbf{1}$ is a vector of one and $\mathbf{I}$ is the $K \times K$ identity matrix. We choose the variance $\sigma^2 = 1$ and the correlation coefficient $\rho = 0.2$, 0.5, and 0.8, and the cluster size $K = 2$, 5, and 20. For $K = 2$ or 5, the covariate Z_{ik} takes a value of 0 or 1, with the probability of 1/2 (individual-level covariate), while for $K = 20$, $(Z_{i1}, \ldots, Z_{iK})$ take the same values of 0 or 1 (cluster-level covariate). This is to mimic community randomized studies, where the individuals within a community receive the same treatment assignment. The number of clusters is 25, 50, or 100 for the cluster size of 2 or 5, and 25 for the cluster size of 20. To avoid generating negative failure times, one may take a log transformation of T_{ik} in (6). The censoring times are generated from an exponential distribution, $\text{Exp}(\lambda)$, with λ being prespecified as 0.09, 0.22, and 0.4 for achieving censoring proportions of 20%, 40%, and 60%.

For each configuration, we conducted 500 simulations and for each data realization we sampled 120 sets for both the bootstrap and perturbation methods to estimate the standard errors. The results are summarized in Tables 1, 2, and 3 for the cluster sizes of 2, 5, and 20, respectively. In these tables, we present the number of clusters (n), the censoring percentage (c%), the intracluster correlation (ρ), the empirical biases of the estimates of the intercept and slope ($\Delta\beta_0$ and $\Delta\beta_1$), the sample standard deviation of the 500 estimates (SD), the averaged estimated standard errors (SE), and the corresponding coverage rates of nominal 95% confidence intervals (CR).

The point estimates of the regression coefficients are approximately unbiased and approach the true values as the sample size increases. Compared to the SDs, the SEs using

Table 2

Parameter estimation bias, standard deviation (SD), standard error (SE), and 95% confidence interval coverage rate (CR) with cluster size of 5

					Bootstrap		Perturb				Bootstrap		Perturb	
n	c%	ρ	$\Delta\beta_0$	SD	SE	CR	SE	CR	$\Delta\beta_1$	SD	SE	CR	SE	CR
25	20	0.2	0.013	0.194	0.204	94.2	0.198	94.0	0.010	0.260	0.284	95.8	0.274	95.0
		0.5	0.016	0.222	0.224	92.8	0.219	92.8	0.004	0.255	0.284	95.4	0.274	94.6
		0.8	0.012	0.250	0.250	92.4	0.246	92.0	0.008	0.256	0.286	96.8	0.278	95.8
	40	0.2	0.013	0.228	0.234	93.0	0.220	91.6	0.006	0.346	0.351	93.8	0.332	92.0
		0.5	0.017	0.253	0.251	94.0	0.238	91.0	0.001	0.342	0.352	94.0	0.332	92.8
		0.8	0.022	0.276	0.275	92.2	0.262	92.6	−0.002	0.336	0.355	94.2	0.335	93.8
	60	0.2	0.020	0.272	0.277	92.8	0.244	89.4	−0.012	0.438	0.457	94.4	0.407	90.4
		0.5	0.023	0.286	0.291	91.8	0.258	88.2	0.004	0.442	0.454	93.0	0.408	90.0
		0.8	0.015	0.304	0.314	92.6	0.279	90.6	0.023	0.444	0.459	95.0	0.414	91.6
50	20	0.2	0.008	0.143	0.143	93.8	0.140	92.2	0.004	0.188	0.198	95.0	0.194	95.2
		0.5	0.014	0.157	0.161	94.0	0.158	93.2	−0.007	0.187	0.198	95.6	0.195	94.6
		0.8	0.018	0.174	0.177	95.0	0.174	93.6	−0.011	0.184	0.201	96.4	0.197	96.0
	40	0.2	0.007	0.164	0.161	93.4	0.155	92.2	0.010	0.240	0.244	93.4	0.237	92.8
		0.5	0.010	0.177	0.178	93.6	0.172	92.0	−0.003	0.234	0.243	94.6	0.236	94.4
		0.8	0.017	0.190	0.193	94.6	0.186	92.8	−0.004	0.231	0.247	95.6	0.241	95.4
	60	0.2	0.004	0.195	0.193	91.6	0.178	89.8	0.005	0.320	0.316	93.8	0.296	91.2
		0.5	0.010	0.209	0.207	93.4	0.191	91.0	0.000	0.318	0.318	93.8	0.297	91.8
		0.8	0.021	0.230	0.221	92.2	0.205	90.0	−0.006	0.318	0.321	94.8	0.301	93.0
100	20	0.2	0.001	0.097	0.099	94.2	0.098	94.6	0.005	0.138	0.138	96.0	0.137	94.2
		0.5	0.003	0.106	0.111	95.6	0.110	95.6	0.003	0.134	0.138	95.8	0.138	95.4
		0.8	0.009	0.120	0.126	95.2	0.125	94.6	−0.006	0.139	0.139	94.0	0.139	94.0
	40	0.2	0.002	0.111	0.113	95.6	0.110	93.2	0.003	0.170	0.172	94.2	0.168	92.6
		0.5	0.004	0.119	0.124	94.4	0.121	95.0	0.000	0.168	0.172	95.4	0.169	95.2
		0.8	0.010	0.137	0.138	94.6	0.134	93.8	−0.005	0.174	0.173	95.0	0.170	94.2
	60	0.2	0.003	0.128	0.136	95.2	0.127	94.4	0.008	0.215	0.222	94.4	0.211	93.4
		0.5	0.006	0.138	0.145	94.2	0.136	92.8	−0.001	0.210	0.221	94.8	0.210	92.8
		0.8	0.010	0.154	0.156	94.0	0.146	91.4	−0.004	0.217	0.223	94.2	0.212	93.4

Table 3

Two-stage parameter estimation bias, standard deviation (SD), standard error (SE), and 95% confidence interval coverage rate (CR) with $n = 25$ clusters of size 20

				Bootstrap				Perturb			
				One-stage		Two-stage		One-stage		Two-stage	
$c\%$	ρ	$\Delta\beta$	SD	SE	CR	SE	CR	SE	CR	SE	CR
				Intercept							
20	0.2	−0.005	0.160	0.150	91.4	0.173	94.4	0.141	88.6	0.187	96.4
	0.5	−0.004	0.223	0.215	91.8	0.229	93.0	0.205	89.8	0.237	94.2
	0.8	−0.005	0.278	0.276	91.8	0.283	91.8	0.267	91.6	0.291	93.0
40	0.2	−0.009	0.170	0.161	90.8	0.191	94.8	0.148	88.4	0.205	96.2
	0.5	−0.008	0.232	0.222	91.4	0.242	93.6	0.209	89.6	0.250	94.2
	0.8	−0.008	0.286	0.281	92.2	0.293	93.6	0.267	91.6	0.297	93.0
60	0.2	−0.011	0.189	0.176	91.2	0.217	96.6	0.156	88.2	0.229	96.8
	0.5	−0.009	0.250	0.236	92.0	0.265	94.4	0.211	88.0	0.266	94.4
	0.8	−0.006	0.302	0.291	92.4	0.311	94.0	0.264	89.2	0.305	93.2
				Slope							
20	0.2	0.014	0.226	0.223	93.0	0.260	96.4	0.209	90.6	0.279	97.6
	0.5	0.018	0.307	0.319	94.0	0.341	95.4	0.302	92.6	0.353	96.2
	0.8	0.016	0.385	0.414	94.2	0.427	94.8	0.397	93.6	0.437	94.4
40	0.2	0.018	0.246	0.245	94.0	0.297	96.8	0.225	91.2	0.320	98.4
	0.5	0.024	0.324	0.335	93.6	0.370	95.0	0.311	91.6	0.383	96.4
	0.8	0.022	0.399	0.426	93.4	0.449	95.2	0.399	91.2	0.456	95.4
60	0.2	0.021	0.287	0.284	93.4	0.358	97.8	0.253	90.8	0.382	98.4
	0.5	0.021	0.357	0.363	94.0	0.420	97.2	0.325	91.2	0.429	98.0
	0.8	0.022	0.430	0.449	93.2	0.491	95.4	0.402	89.2	0.485	96.4

the bootstrap or perturbation method provide good estimates of the variability. As expected, the SE decreases with the increase of the sample size and increases with the increase of the censoring percentage. The 95% confidence interval coverage rates based on the bootstrap method are satisfactory when the censoring is not heavy, while the perturbation method shows relative undercoverage, especially when the number of clusters is small or the censoring is heavy. As the number of clusters increases, with light and moderate censoring, the coverage rates using the perturbation method improve and approach 95%. With the larger cluster size ($K = 20$), we applied the two-stage resampling procedures to estimate the variance as shown in Table 3. The one-stage resampling method seems to have lower coverage rates and the two-stage methods may lead to some improvement. These results indicate that the second stage of bootstrapping or perturbing could be helpful to account for the within-cluster variation in the situations with relatively large clusters.

5. An Example

As an illustration, we apply the proposed method to the OM dataset to investigate the effectiveness of the medical treatment in prolonging the life of the ventilating tubes. The dataset consists of information on 78 children, ranging in age from 6 months to 8 years who had chronic OM with effusion (the most common form). Through randomization, 40 children were assigned to medical treatment and 38 served as controls. All subjects underwent therapeutic myringotomy for tympanostomy tube placement (intubation). Children were examined before and 2 weeks after surgery and every 3 months thereafter. Censoring was caused by loss to follow-up, and the censoring rate for these OM data is about 7.7%. The primary endpoint is the cessation of tube functioning (failure). Figure 1 shows the Kaplan–Meier survival curves for the treatment and control groups.

We took $\tau = (.1, .25, .5, .75, .9)$ and simultaneously applied the five quantile regression models. Variance estimation was based on 1000 resampling samples for both bootstrap and perturbation methods. In Table 4, $\hat{\beta}_1$ refers to the

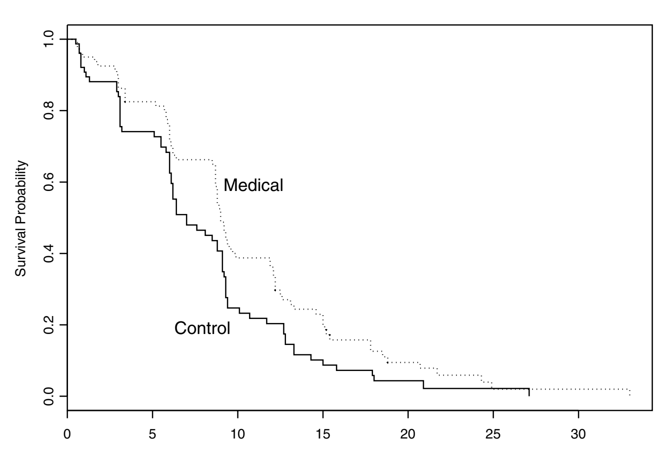

Figure 1. Kaplan–Meier curves for the duration of ventilating tubes in the otitis media study.

Table 4
Parameter estimation, standard error (SE), and test statistic (z) under different τ's, for the otitis media data

		Bootstrap		Perturb			Bootstrap		Perturb	
τ	$\hat{\beta}_0$	SE	z	SE	z	$\hat{\beta}_1$	SE	z	SE	z
0.1	15.000	2.325	6.452	2.165	6.929	3.800	3.473	1.094	3.281	1.158
0.25	9.400	1.415	6.644	1.449	6.487	3.900	1.985	1.965	1.988	1.961
0.5	6.400	0.986	6.491	0.990	6.463	2.800	1.221	2.293	1.278	2.190
0.75	3.100	1.127	2.750	1.112	2.788	2.900	1.578	1.838	1.654	1.754
0.9	1.100	0.814	1.351	0.820	1.342	1.900	1.456	1.305	1.473	1.290

estimated medical treatment effect and $z = \hat{\beta}_1/\widehat{\mathrm{SE}}(\hat{\beta}_1)$ is the Wald test statistic. It shows that there is no statistically significant treatment effect at the $\alpha = .05$ significance level with $\tau = .1$, .75, and .9, but with $\tau = .25$ and .5, the survival time of the treatment group is substantially longer than that of the control group. The variance of $\hat{\beta}_1$ is bigger when τ is near 0 because there is insufficient information for estimating the treatment effect for small τ. Figure 2 presents global pictures of $\hat{\beta}_{0,\tau}$ and $\hat{\beta}_{1,\tau}$ with respect to τ. The 95% confidence intervals based on the bootstrap and the perturbation methods are very close, which are narrower in the middle. Quantiles of the survival time and their confidence intervals can be estimated for each group. In particular, the median survival time for the treatment group is 9.2 months with the 95% confidence interval of (7.8, 10.6), while that for the control group is 6.4 months with (4.5, 8.3).

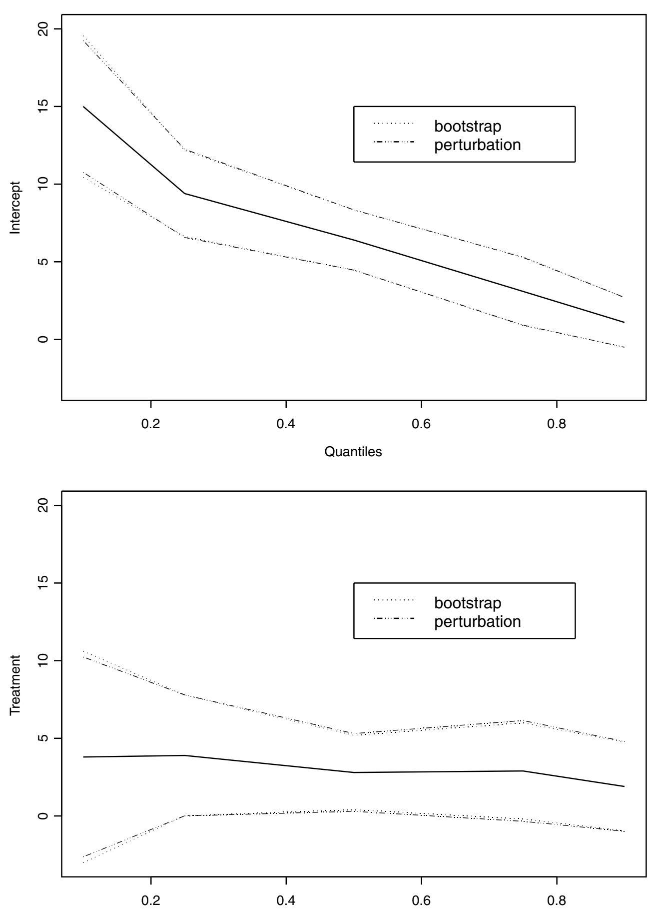

Figure 2. Parameter estimates and corresponding 95% confidence intervals versus quantiles.

6. Remarks

For correlated failure time data, we have studied quantile regression models, which provide a global assessment of covariate effects. We derived several large sample properties of the parameter estimates, including the strong consistency and the asymptotic normality. The regression parameters can be easily estimated through the iterative algorithm, while the variances can be obtained using the bootstrap method or by properly perturbing the estimating function. The bootstrap method is more straightforward to implement in practice, and the perturbation method requires choosing a suitable random variable U with mean 1 and variance 1. For the two-stage perturbation, it is more difficult to control the range of perturbing values. With very large perturbing random variables, we may have some numerical computational problems, while with very small perturbing random variables, the estimating function may not be jittered enough to yield good estimates. Another interesting resampling method (Parzen et al., 1994) may work here as well, which is to sample from the asymptotic distribution of $n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0)$, and obtain $\hat{\boldsymbol{\beta}}$ by solving $n^{1/2}\ \mathbf{S}_n(\boldsymbol{\beta}) = \mathbf{s}$, where $\mathbf{s}$ is a $p \times 1$ random vector from $N_p(0, \hat{\boldsymbol{\Gamma}})$. This resampling method is similar to the iterative estimating procedure in Section 2.4 in the spirit that both methods fix the right-hand side to a constant vector and solve the equation backward for $\hat{\boldsymbol{\beta}}$.

We can allow the dependence between the covariate vector $\mathbf{Z}$ and the censoring variable C by categorizing the covariate $\mathbf{Z}$ into several possible values. With the stratification over $\mathbf{Z}$, we could replace the censoring time Kaplan–Meier estimator $\hat{G}(t)$ by $\hat{G}(t \mid \mathbf{z})$, where the conditional survival function $G(t \mid \mathbf{z}) = \Pr(C \geq t \mid \mathbf{Z} = \mathbf{z})$. However, this method may not be practical for high-dimensional $\mathbf{Z}$ or for $\mathbf{Z}$ with many categories. To avoid the curse of dimensionality, an alternative is to model the censoring times with the covariates $\mathbf{Z}$ by the Cox-type regression model (Lee et al., 1992) and estimate $\hat{G}(\cdot)$ based on the fitted model.

The estimator obtained by minimizing the inverse weighted estimating function (3) is consistent, which is used as an initial value for (2) in our case. Because the difference between two adjacent estimators decreases stochastically as iteration increases, we take $\hat{\beta}_{[M]}$ as our final estimator in order to reduce the empirical bias for the finite sample size. As pointed out by a referee, it would be more flexible and might improve the convergence by augmenting the weight matrix through $\mathbf{W} = \text{diag}(c_1, \ldots, c_K) + \{\mathbf{W} - \text{diag}(c_1, \ldots, c_K)\}$ with appropriately chosen $(c_1, \ldots, c_K)$.

In the quantile regression model, it is difficult to incorporate time-varying covariates. In heavy censoring cases, the model may not be well defined if the corresponding quantile does not exist. Let ξ_τ be the τth quantile, $\xi_\tau = \inf\{t : F(t) \geq \tau\}$, with a natural estimator $\hat{\xi}_\tau = \inf\{t : \hat{F}(t) \geq \tau\}$, where $\hat{F}(t) = 1 - \hat{S}(t)$ and $\hat{S}(t)$ is the Kaplan–Meier estimator of the survival function. If the $\hat{S}(t)$ curve does not drop below τ ($\hat{\xi}_\tau$ does not exist), the τth quantile regression is not readily applicable.

Acknowledgements

We would like to thank the editor, an associate editor, and a referee for their critical reading of the manuscript and constructive comments, which led to great improvement of the earlier versions. We thank Lu Tian and Donglin Zeng for very helpful discussions. This work was partially supported by the National Institutes of Health grant R01 HL-57444.

Résumé

En tant qu'alternative à la classique modélisation de la moyenne par régression, la modélisation des quantiles a été largement étudiée (surtout celle de la médiane) dans le cas de données de survie indépendantes. Cependant, les corrélations naturelles ou artificielles entre données font qu'il est courant de rencontrer, en recherche biomédicale, des données de survie multivariées pour lesquelles une corrélation intra-groupe doit être ajustée. Pour des données de survie corrélées et censurées à droite, nous explorons ici la modélisation des quantiles et adaptons une approche par équations d'estimation destinée à estimer les paramètres sous hypothèse d'indépendance. Nous adaptons également une version pondérée de cette même approche, qui en améliore l'efficacité. Nous montrons que les estimateurs des paramètres sont robustes et suivent asymptotiquement une distribution normale. L'estimation de la variance, obtenue par approximation asymptotique, utilise une estimation non paramétrique de la densité, cependant que la matrice de variance-covariance est estimée par bootstrap et méthodes de ré-échantillonnage perturbé. Des simulations montrent que l'approche proposée convient aux échantillons de taille finie. Cette nouvelle démarche est appliquée, pour illustration, à des données issues d'un essai clinique dans l'otite.

References

Bang, H. and Tsiatis, A. A. (2002). Median regression with censored cost data. *Biometrics* **58,** 643–649.

Bilias, Y., Chen, S., and Ying, Z. (2000). Simple resampling methods for censored regression quantiles. *Journal of Econometrics* **99,** 373–386.

Bilias, Y., Gu, M., and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Annals of Statistics* **25,** 662–682.

Buchinsky, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models. A Monte Carlo study. *Journal of Econometrics* **68,** 303–338.

Buchinsky, M. and Hahn, J. Y. (1998). An alternative estimator for censored quantile regression. *Econometrica* **66,** 653–671.

Cai, J. and Prentice, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* **82,** 151–164.

Cai, J. and Prentice, R. L. (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis* **3,** 197–213.

Cai, T., Wei, L. J., and Wilcox, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika* **87,** 867–878.

Fitzenberger, B. (1997). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics* **82,** 235–287.

Fleming, T. R. and Harrington, D. (1991). *Counting Processes and Survival Analysis.* New York: Wiley.

Hahn, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory* **11,** 105–121.

Horowitz, J. (1998). Bootstrap methods for median regression model. *Econometrica* **66,** 1327–1352.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47,** 663–685.

Jin, Z., Ying, Z., and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88,** 381–390.

Jung, S. H. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91,** 251–257.

Khan, S. and Powell, L. J. (2001). Two-step estimation of semiparametric censored regression models. *Journal of Econometrics* **103,** 73–110.

Koenker, R. and Bassett, G. J. (1978). Regression quantiles. *Econometrica* **46,** 33–50.

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of the American Statistical Association* **96,** 458–468.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17,** 1217–1241.

Le, C. T. and Lindgren, B. R. (1996). Duration of ventilating tubes: A test for comparing two clustered samples of censored data. *Biometrics* **52,** 328–334.

Lee, E. W., Wei, L. J., and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, J. P. Klein and P. K. Goel (eds), 237–247. Dordrecht: Kluwer Academic Publishers.

Lee, E. W., Wei, L. J., and Ying Z. (1993). Linear regression analysis for highly stratified failure time data. *Journal of the American Statistical Association* **88,** 557–565.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lin, D. Y. and Geyer, C. J. (1992). Computational methods for semiparametric linear regression with censored data.

Journal of Computational and Graphical Statistics **1,** 77–90.

Lindgren, A. (1997). Quantile regression with censored data using generalized L_1 minimization. *Computational Statistics and Data Analysis* **23,** 509–524.

McKeague, I., Subramanian, S., and Sun, Y. (2001). Median regression analysis for censored data. *Journal of Nonparametric Statistics* **13,** 709–727.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81,** 341–350.

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

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association* **98,** 1001–1012.

Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimations. *Statistical Science* **12,** 279–296.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression. *Journal of Econometrics* **25,** 303–325.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

Tian, L. and Wei, L. J. (2002). Quantile regression under general random censorship, unpublished manuscript.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer-Verlag.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard function. *Journal of the American Statistical Association* **94,** 137–145.

Ying, Z. and Wei, L. J. (1994). The Kaplan–Meier estimate for dependent failure time observations. *Journal of Multivariate Analysis* **50,** 17–29.

Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90,** 178–184.

Received August 2003. *Revised February* 2004.
Accepted April 2004.

Appendix

All the probability statements regarding T, X, as well as the density function $f(\cdot)$, and the cumulative distribution function (c.d.f.) $F(\cdot)$ are conditioning on $\mathbf{Z}$. For ease of exposition, we suppress the notation of conditional on $\mathbf{Z}$, i.e., $(\cdot \mid \mathbf{Z})$, throughout. We assume the following regularity conditions: (i) The true value of $\boldsymbol{\beta}_0$ lies in the interior of a bounded convex region $\mathbf{B}$; (ii) there exists a constant $\tilde{t}$ such that $\Pr(X \geq \tilde{t}) > 0$ and $\boldsymbol{\beta}'\mathbf{Z} \leq \tilde{t}$ with probability 1, for all $\boldsymbol{\beta} \in \mathbf{B}$; (iii) the covariate vector $\mathbf{Z}$ is bounded and the density functions of failure and censoring times are uniformly bounded; and (iv) the matrix $\mathrm{E}[f(0)\mathbf{Z}\mathbf{Z}']$ is positive definite, where $f(\cdot)$ is the density function of errors.

Proof of Theorem 1: Let $F(\cdot)$ be the c.d.f. of ϵ_{ik} and we define

$$\tilde{\mathbf{S}}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \left[(1-\tau) - F\{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{Z}_{ik}\}\right].$$

Thus, we have

$$\mathbf{S}_n(\boldsymbol{\beta}) - \tilde{\mathbf{S}}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \left\{ \frac{I(X_{ik} \geq \boldsymbol{\beta}'\mathbf{Z}_{ik})}{\hat{G}(\boldsymbol{\beta}'\mathbf{Z}_{ik})} - \Pr(T_{ik} \geq \boldsymbol{\beta}'\mathbf{Z}_{ik}) \right\}.$$

Because data from different clusters are independent (Ying and Wei, 1994), as $n \to \infty$, for some $\epsilon > 0$,

$$\sup_{0 \leq t \leq \tilde{t}} |\hat{G}(t) - G(t)| = o(n^{-1/2+\epsilon}) \quad \text{a.s.} \tag{A.1}$$

Hence, with the boundness of $\mathbf{Z}$ and the cluster sizes, we can show that, almost surely,

$$\mathbf{S}_n(\boldsymbol{\beta}) - \tilde{\mathbf{S}}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \frac{I(X_{ik} \geq \boldsymbol{\beta}'\mathbf{Z}_{ik}) - \Pr(X_{ik} \geq \boldsymbol{\beta}'\mathbf{Z}_{ik})}{G(\boldsymbol{\beta}'\mathbf{Z}_{ik})} + o(n^{-1/2+\epsilon}).$$

It follows from the uniform law of large numbers (Pollard, 1990, p. 41) that

$$\sup_{\boldsymbol{\beta} \in \mathbf{B}} \left| n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \frac{I(X_{ik} \geq \boldsymbol{\beta}'\mathbf{Z}_{ik}) - \Pr(X_{ik} \geq \boldsymbol{\beta}'\mathbf{Z}_{ik})}{G(\boldsymbol{\beta}'\mathbf{Z}_{ik})} \right| \to 0 \quad \text{a.s.}$$

We define an indicator variable $\eta_{ik} = 1$ if the ith cluster has a kth member, and 0 otherwise. Let $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iK})'$ and assume $(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n)$ are i.i.d. With K being the maximum of the cluster sizes, we thus have,

$$\mathbf{S}_n(\boldsymbol{\beta}) - \tilde{\mathbf{S}}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \frac{\eta_{ik}\mathbf{Z}_{ik}\{I(X_{ik} \geq \boldsymbol{\beta}'\mathbf{Z}_{ik}) - \Pr(X_{ik} \geq \boldsymbol{\beta}'\mathbf{Z}_{ik})\}}{G(\boldsymbol{\beta}'\mathbf{Z}_{ik})} + o(n^{-1/2+\epsilon}) \quad \text{a.s.}$$

Define the class of functions $(k = 1, \ldots, K)$,

$$\mathcal{F}_k = \left\{ \frac{\eta_k \mathbf{Z}_k\{I(X_k \geq \boldsymbol{\beta}'\mathbf{Z}_k) - \Pr(X_k \geq \boldsymbol{\beta}'\mathbf{Z}_k)\}}{G(\boldsymbol{\beta}'\mathbf{Z}_k)}, \; G(\cdot) \text{ is bounded away from } 0, \boldsymbol{\beta} \in \mathbf{B} \right\}.$$

By Example 2.11.16 (van der Vaart and Wellner, 1996, p. 215–216) and the Donsker preservation theorem, $\mathcal{F}_k$ is a Donsker class, and hence $\sup_{\boldsymbol{\beta} \in \mathbf{B}} \|\mathbf{S}_n(\boldsymbol{\beta}) - \tilde{\mathbf{S}}_n(\boldsymbol{\beta})\| = o(n^{-1/2+\epsilon})$ a.s., as $n \to \infty$.

Now, focusing on $\tilde{\mathbf{S}}_n(\boldsymbol{\beta})$,

$$\mathbf{A}_n(\boldsymbol{\beta}) = \frac{\partial \tilde{\mathbf{S}}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} f\{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{Z}_{ik}\}\mathbf{Z}_{ik}\mathbf{Z}_{ik}'.$$

By the Kolmogorov strong law of large numbers (SLLN), $\mathbf{A}_n(\boldsymbol{\beta}) \to -\zeta E[f\{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{Z}_{11}\}\mathbf{Z}_{11}\mathbf{Z}_{11}']$, a.s., and thus $\mathbf{A}_n(\boldsymbol{\beta}_0) \to \mathbf{A}$, which is negative definite.

It follows from $F(0) = 1 - \tau$ that $\tilde{\mathbf{S}}_n(\boldsymbol{\beta}_0) = 0$. By the standard inverse function theorem and coupled with the negative definiteness of $\mathbf{A}$, it entails the strong consistency of $\hat{\boldsymbol{\beta}}$.

Proof of Theorem 2: By adding and subtracting the same term $\mathbf{Z}_{ik} I(X_{ik} \geq \boldsymbol{\beta}_0' \mathbf{Z}_{ik})/G(\boldsymbol{\beta}_0' \mathbf{Z}_{ik})$,

$$\mathbf{S}_n(\boldsymbol{\beta}_0) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \left\{ \frac{I(X_{ik} \geq \boldsymbol{\beta}_0' \mathbf{Z}_{ik})}{G(\boldsymbol{\beta}_0' \mathbf{Z}_{ik})} - \tau \right\} - \int_0^\infty \frac{\hat{G}(t) - G(t)}{\hat{G}(t) G(t)} \, d\mathbf{Q}(t),$$

where $\mathbf{Q}(t) = n^{-1} \sum_{i=1}^n \sum_{k=1}^{K_i} I\{\boldsymbol{\beta}_0' \mathbf{Z}_{ik} \leq t \wedge X_{ik}\} \mathbf{Z}_{ik}$.

In terms of the martingale representation (Fleming and Harrington, 1991),

$$\frac{\hat{G}(t) - G(t)}{G(t)} = - \int_0^t \frac{\hat{G}(u-)}{G(u)} \left\{ \frac{d\bar{N}_G(u)}{\bar{Y}(u)} - d\Lambda_G(u) \right\}, \quad \text{(A.2)}$$

which is asymptotically equivalent to $-\int_0^t d\bar{M}_G(u)/\bar{Y}(u)$ with $d\bar{M}_G(t) = d\bar{N}_G(t) - \bar{Y}(t)\, d\Lambda_G(t)$.

If we plug in (A.2) and interchange the integrals, the second term of $\mathbf{S}_n(\boldsymbol{\beta}_0)$ is asymptotically equivalent to

$$\int_0^\infty \frac{1}{\hat{G}(t)} \int_0^t \frac{d\bar{M}_G(u)}{\bar{Y}(u)} \, d\mathbf{Q}(u) = \int_0^\infty \frac{d\bar{M}_G(u)}{\bar{Y}(u)} \int_u^\infty \frac{d\mathbf{Q}(t)}{\hat{G}(t)}.$$

With the definition of $\mathbf{Q}$(t),

$$\int_u^\infty \frac{d\mathbf{Q}(t)}{\hat{G}(t)} = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \frac{\mathbf{Z}_{ik}}{G(\boldsymbol{\beta}_0' \mathbf{Z}_{ik})} I(\boldsymbol{\beta}_0' \mathbf{Z}_{ik} \geq u) \times I(T_{ik} \geq \boldsymbol{\beta}_0' \mathbf{Z}_{ik}) I(C_{ik} \geq \boldsymbol{\beta}_0' \mathbf{Z}_{ik}) + o_p(n^{-1/2+\epsilon}),$$

which converges almost surely to $\mathbf{q}(u)\tau$ by the Kolmogorov SLLN.

Using the functional CLT (Pollard, 1990, Theorem 10.6), we can show that $n^{-1/2} \sum_{i=1}^n \sum_{k=1}^{K_i} M_{Gik}(t)$ converges in distribution to a zero-mean Gaussian process with continuous sample paths. It follows from the Skorohod strong embedding (Shorack and Wellner, 1986) and the Lemma A.3 in Bilias, Gu, and Ying (1997) that, in probability,

$$n^{-1/2} \int_0^\infty \left\{ \int_u^\infty \frac{d\mathbf{Q}(t)}{\hat{G}(t)} \frac{1}{n^{-1}\bar{Y}(u)} - \frac{\mathbf{q}(u)\tau}{\pi(u)} \right\} d\bar{M}_G(u) \to 0.$$

Finally, the statistic $n^{1/2} \mathbf{S}_n(\boldsymbol{\beta}_0)$ is asymptotically equivalent to $n^{-1/2} \sum_{i=1}^n \boldsymbol{\xi}_i$, where $\boldsymbol{\xi}_i$ is given by (4). By the multivariate CLT, the random vector $n^{1/2} \mathbf{S}_n(\boldsymbol{\beta}_0)$ converges to a normal distribution with mean zero and a variance–covariance matrix $\boldsymbol{\Gamma} = E(\boldsymbol{\xi}_1 \boldsymbol{\xi}_1')$.

Proof of Theorem 3: Given any fixed constant c, for all $\boldsymbol{\beta}$ in $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < cn^{-1/3}$, based on Lemma 1 of Ying et al. (1995),

$$\begin{aligned}
\mathbf{S}_n(\boldsymbol{\beta}) = \mathbf{S}_n(\boldsymbol{\beta}_0) &+ n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \left\{ \frac{I(X_{ik} \geq \boldsymbol{\beta}' \mathbf{Z}_{ik})}{G(\boldsymbol{\beta}' \mathbf{Z}_{ik})} - \tau \right\} \\
&- n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \left\{ \frac{I(X_{ik} \geq \boldsymbol{\beta}_0' \mathbf{Z}_{ik})}{G(\boldsymbol{\beta}_0' \mathbf{Z}_{ik})} - \tau \right\} \\
&+ n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \{ I(X_{ik} \geq \boldsymbol{\beta}' \mathbf{Z}_{ik}) - I(X_{ik} \geq \boldsymbol{\beta}_0' \mathbf{Z}_{ik}) \} \\
&\times \left\{ \frac{1}{\hat{G}(\boldsymbol{\beta}_0' \mathbf{Z}_{ik})} - \frac{1}{G(\boldsymbol{\beta}_0' \mathbf{Z}_{ik})} \right\} + o_p(n^{-1/2}).
\end{aligned}$$

By the lemma in Jung (1996, p. 252) and (A.1), it can be shown that the fourth term on the right-hand side of $\mathbf{S}_n(\boldsymbol{\beta})$ is $o_p(n^{-1/2})$ and

$$\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \left\{ \frac{I(X_{ik} \geq \boldsymbol{\beta}' \mathbf{Z}_{ik})}{G(\boldsymbol{\beta}' \mathbf{Z}_{ik})} - \tau \right\} \\
&\quad - n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \mathbf{Z}_{ik} \left\{ \frac{I(X_{ik} \geq \boldsymbol{\beta}_0' \mathbf{Z}_{ik})}{G(\boldsymbol{\beta}_0' \mathbf{Z}_{ik})} - \tau \right\} \\
&= \tilde{\mathbf{S}}_n(\boldsymbol{\beta}) + o_p(n^{-1/2}).
\end{aligned}$$

Hence, we have $\mathbf{S}_n(\boldsymbol{\beta}) = \mathbf{S}_n(\boldsymbol{\beta}_0) + \tilde{\mathbf{S}}_n(\boldsymbol{\beta}) + o_p(n^{-1/2})$. Taking Taylor's expansion of $\tilde{\mathbf{S}}_n(\boldsymbol{\beta})$ at $\boldsymbol{\beta}_0$ and noting that $\tilde{\mathbf{S}}_n(\boldsymbol{\beta}_0) = 0$ and $\partial \tilde{\mathbf{S}}_n(\boldsymbol{\beta}_0)/\partial \boldsymbol{\beta} \to \mathbf{A}$, we have

$$\mathbf{S}_n(\boldsymbol{\beta}) = \mathbf{S}_n(\boldsymbol{\beta}_0) + \mathbf{A}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(n^{-1/2} \vee \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|),$$

which entails that $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a normal distribution with mean zero and variance $\mathbf{A}^{-1} \boldsymbol{\Gamma} \mathbf{A}^{-1}$ in any $n^{-1/3}$-neighborhood of $\boldsymbol{\beta}_0$.