Cure Rate Quantile Regression for Censored Data With a Survival Fraction

Yuanshan WU and Guosheng YIN

Censored quantile regression offers a valuable complement to the traditional Cox proportional hazards model for survival analysis. Survival times tend to be right-skewed, particularly when there exists a substantial fraction of long-term survivors who are either cured or immune to the event of interest. For survival data with a cure possibility, we propose cure rate quantile regression under the common censoring scheme that survival times and censoring times are conditionally independent given the covariates. In a mixture formulation, we apply censored quantile regression to model the survival times of susceptible subjects and logistic regression to model the indicators of whether patients are susceptible. We develop two estimation methods using martingale-based equations: One approach fully uses all regression quantiles by iterating estimation between the cure rate and quantile regression parameters; and the other separates the two via a nonparametric kernel smoothing estimator. We establish the uniform consistency and weak convergence properties for the estimators obtained from both methods. The proposed model is evaluated through extensive simulation studies and illustrated with a bone marrow transplantation data example. Technical proofs of key theorems are given in Appendices A, B, and C, while those of lemmas and additional simulation studies on model misspecification and comparisons with other models are provided in the online Supplementary Materials A and B.

KEY WORDS: Cure rate model; Empirical process; Long-term survivor; Martingale; Random censoring; Regression quantile; Survival analysis; Volterra integral equation.

1. INTRODUCTION

In oncology clinical trials, it is often observed that a substantial proportion of subjects are cured and thus become risk-free of disease relapse. On the other hand, patients may never respond to the treatment under study due to drug resistance or disease status, and these patients are considered insusceptible. To accommodate the cured or insusceptible proportion of subjects, a cure fraction can be explicitly incorporated into survival models (Zeng, Yin, and Ibrahim 2006). A commonly used approach to modeling such survival data is the two-component mixture cure rate model, which assumes the underlying population is a mixture of susceptible and insusceptible subjects. All susceptible subjects would eventually experience the event of interest if the follow-up is sufficiently long, while the insusceptible subjects would never experience the event regardless of the length of the follow-up. As a result, one can separately model the survival distribution of the susceptible subjects and the insusceptible fraction of the population. Based on parametric models, Berkson and Gage (1952) proposed the exponential–logistic mixture model, and Farewell (1982, 1986) considered the Weibull–logistic mixture model for survival data with a cure fraction. In semiparametric settings, Kuk and Chen (1992) proposed to use the Cox proportional hazards (PH) model (Cox 1972) for the survival times of susceptible subjects and the logistic regression model for the cure indicator. The mixture Cox PH cure rate model was further investigated by Peng and Dear (2000), Sy and Taylor (2000), Fang, Li, and Sun (2005), and Lu (2008). Along similar lines, extensive research has been conducted with other semiparametric cure rate models. For example, Lu and Ying (2004) and Mao and Wang (2010) proposed transformation cure rate models based on transformed linear regression. Zhang and Peng (2009) studied an accelerated hazards cure rate model, in which the accelerated hazards model (Chen and Wang 2000) is adopted to fit the survival times of susceptible subjects. Lu (2010) further developed the accelerated failure time (AFT) mixture cure rate model through sieve maximum likelihood estimation (Zeng and Lin 2007).

All the aforementioned survival models with a cure fraction are essentially mean-based regression models, which mainly give an overall quantification for the central covariate effects. In contrast, quantile regression can directly model a series of (from lower to higher) quantiles of the survival times so as to provide a more complete assessment of covariate effects (Koenker and Bassett 1978; Koenker 2005). This distinctive feature of quantile regression makes it very attractive for typically right-skewed survival data, especially when a substantial cure fraction exists.

Extensive research has been carried out in censored quantile regression. Powell (1984, 1986) proposed to minimize the least absolute deviation to handle fixed censoring cases. For random censoring, Ying, Jung, and Wei (1995) modified quantile estimating equations by assuming independence between survival and censoring times. For medical cost data with informative censoring, Bang and Tsiatis (2002) developed a median regression method based on the inverse probability weighting scheme. Honoré, Khan, and Powell (2002) studied quantile regression with random censoring by assuming censoring times independent of both survival times and covariates. Likewise, under such completely random censorship, Yin, Zeng, and Li (2008) proposed a class of transformation quantile regression models for survival data.

Nevertheless, a more typical and realistic assumption is the conditional independence of survival and censoring times given the covariates. Yang (1999) proposed a median regression model

Yuanshan Wu is Assistant Professor, School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China (E-mail: *shan@whu.edu.cn*). Guosheng Yin is Associate Professor, Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong (E-mail: *gyin@hku.hk*). We thank the associate editor and three anonymous referees for their careful reading and many insightful suggestions, which strengthened the work immensely. Wu's research was supported in part by National Natural Science Foundation of China grant 11201350, and Yin's research was supported in part by a grant (784010) from the Research Grants Council of Hong Kong.

Journal of the American Statistical Association
December 2013, Vol. 108, No. 504, Theory and Methods
DOI: 10.1080/01621459.2013.837368

based on the weighted empirical survival function, which, however, requires homogeneous errors or the error distributions converging to a common one at a certain rate. Portnoy (2003) developed censored quantile regression by redistributing censored data to the right, and Neocleous, Vanden Branden, and Portnoy (2006) and Portnoy and Lin (2010) studied the asymptotic properties of the estimation procedure. By using the martingale structure of right-censored survival data, Peng and Huang (2008) proposed a martingale-based estimating equation. To relax the global linearity assumption for all regression quantiles, Wang and Wang (2009) proposed a locally weighted quantile regression approach for any particular quantile of interest, which adopts the redistribution-of-mass idea and employs nonparametric functional estimation for the local Kaplan–Meier estimator. For enhancing model flexibility, Qian and Peng (2010) studied the partially functional model with a mixture of quantile-variant and -invariant covariate effects. Huang (2010) proposed an estimating integral equation procedure for censored quantile regression, which allows for zero-density intervals and discontinuities in a distribution. The variances of regression quantiles typically involve density functions. To circumvent nonparametric functional estimation, a variety of resampling methods have been developed for variance estimation (Parzen, Wei, and Ying 1994; Buchinsky 1995; Hahn 1995; Buchinsky and Hahn 1998; Horowitz 1998; Bilias, Chen, and Ying 2000; Jin, Ying, and Wei 2001).

In the presence of a cure fraction, quantile regression is particularly appealing, because the population contains a fraction of insusceptible subjects with infinitely long survival times. The survival probability of insusceptible subjects is one, which may confound the real treatment effect if we directly fit the censored quantile regression model to all the subjects. To provide a "clean" assessment of covariate effects on the survival times of susceptible subjects, we propose mixture cure rate quantile regression to accommodate a survival fraction in the population. Under the usual conditional independent censoring, we adopt the censored quantile regression model by Peng and Huang (2008) to fit the survival times of susceptible subjects, and use the logistic regression to model the cure fraction. Both regression quantiles and cure rate parameters are estimated via martingale equations (Fleming and Harrington 1991). We develop two estimation methods for the cure rate parameters: one relies on the iteration between the cure rate parameters and the quantile regression coefficients, while the other separates them by employing the nonparametric kernel smoothing technique. The uniform consistency and weak convergence properties of the resultant estimators are established using the empirical process, kernel smoothing, and Volterra integral equation theories.

The rest of the article is organized as follows. In Section 2, we propose the cure rate quantile regression and two estimation procedures for survival data with a cure fraction. In Section 3, we establish the large-sample properties of the resultant estimators. We conduct simulation studies to evaluate the finite-sample performance of the proposed methods in Section 4 and illustrate our model with application to a bone marrow transplantation data example in Section 5. We conclude with some remarks in Section 6 and delineate the proofs of the theorems in Appendices A, B, and C. Additional simulation studies and technical derivations of useful lemmas are provided in the online Supplementary Materials A and B.

2. MODEL AND ESTIMATION

2.1 Cure Rate Quantile Regression

The mixture cure rate model assumes a decomposition of the failure time as

$$T = \eta T^* + (1-\eta)\infty,$$

where $T^* < \infty$ denotes the survival time of a susceptible subject, and the indicator η takes a value of 1 if a subject is susceptible, and 0 otherwise. Let C be the censoring time, let $\mathbf{Z}$ be a $(p+1)$-vector of covariates related to T^*, and let $\mathbf{W}$ be a $(q+1)$-vector of covariates associated with η. Both $\mathbf{Z}$ and $\mathbf{W}$ include 1 as an intercept, and they may share common components. The observed time is $X = T \wedge C$, the minimum of T and C, and let $\Delta = I(T \le C)$ be the censoring indicator. For $i = 1, \ldots, n$, $(X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i)$ are assumed to be independent and identically distributed (iid), and (T_i^*, η_i) and C_i are conditionally independent given covariates $\mathbf{Z}_i$ and $\mathbf{W}_i$.

Based on the logistic regression (Farewell 1982), we can model the susceptibility indicator η,

$$P(\eta = 1|\mathbf{W}) = \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}) = \frac{\exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})}{1+\exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})}. \tag{2.1}$$

For survival times T^*, we take the usual linear regression model

$$\log T^* = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z} + \epsilon, \tag{2.2}$$

where the error ϵ may depend on $\mathbf{Z}$. If we consider the mean regression with homogeneous errors, model (2.2) is known as the accelerated failure time model. In contrast, if we model a set of quantiles of the susceptible survival times, we can provide a more comprehensive assessment on the covariate effects, particularly when the errors are heteroscedastic. Given $\tau \in (0,1)$, $Q_{T^*}(\tau|\mathbf{Z}) = \inf\{t\colon P(T^* \le t|\mathbf{Z}) \ge \tau\}$ is the τth conditional quantile function, and the quantile regression model is given by

$$Q_{T^*}(\tau|\mathbf{Z}) = \exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}, \quad \tau \in (0,1), \tag{2.3}$$

where $\boldsymbol{\beta}(\tau)$ is an unknown $(p+1)$-vector of regression coefficients.

2.2 Estimation of $\boldsymbol{\beta}(\tau)$

We denote $F_{T^*}(t|\mathbf{Z}) = P(T^* \le t|\mathbf{Z})$, and

$$\begin{aligned}\Lambda_{T,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W}) &= -\log\{1 - P(T \le t|\mathbf{Z},\mathbf{W})\}\\ &= -\log\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_{T^*}(t|\mathbf{Z})\}.\end{aligned}$$

Let $N(t) = \Delta I(X \le t)$ and $M(t;\boldsymbol{\gamma}, F_{T^*}) = N(t) - \Lambda_{T,\boldsymbol{\gamma}}(t \wedge X|\mathbf{Z},\mathbf{W})$. Following the martingale formulation of censored quantile regression by Peng and Huang (2008), we replace t by the true conditional quantile $\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\}$ in $M(t;\boldsymbol{\gamma}, F_{T^*})$,

$$E\big(\mathbf{Z}\big[N(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\}) - \Lambda_{T,\boldsymbol{\gamma}_0}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\} \wedge X|\mathbf{Z},\mathbf{W})\big]\big) = \mathbf{0},$$

where $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\gamma}_0$ are the true parameter values. Under model (2.3), some algebraic manipulations yield

$$\Lambda_{T,\boldsymbol{\gamma}_0}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\} \wedge X|\mathbf{Z},\mathbf{W}) = \int_0^{\tau} I[X \geq \exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(u)\}]H_{\boldsymbol{\gamma}_0}(\mathrm{d}u|\mathbf{W}),$$

where $H_{\boldsymbol{\gamma}}(u|\mathbf{W}) = -\log\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})u\}$. Immediately, we have

$$E\left[\mathbf{Z}\left\{N(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\}) - \int_0^{\tau} I[X \geq \exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(u)\}]H_{\boldsymbol{\gamma}_0}(\mathrm{d}u|\mathbf{W})\right\}\right] = \mathbf{0},$$

which leads to the estimating function

$$\mathbf{U}_n(\boldsymbol{\beta},\tau;\boldsymbol{\gamma}) = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i\left\{N_i\left(\exp\left\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(\tau)\right\}\right) - \int_0^{\tau} I\left[X_i \geq \exp\left\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\right\}\right]H_{\boldsymbol{\gamma}}(\mathrm{d}u|\mathbf{W}_i)\right\}, \tag{2.4}$$

with $N_i(t) = \Delta_i I(X_i \leq t)$ for $i = 1,\ldots,n$.

Let L denote the duration of the study, and let $\tau_{\max}$ be a constant in (0, 1), which is the upper bound of the quantile levels that can be estimated. To ensure the identifiability for all regression quantiles below $\tau_{\max}$, we require $\tau_{\max}$ to be smaller than $\inf_{\mathbf{Z}} F_{T^*}(L|\mathbf{Z}=\mathbf{z})$. If we consider the conditional distribution of T given $\mathbf{Z}$ and $\mathbf{W}$, we have $F_{T,\boldsymbol{\gamma},\boldsymbol{\beta}}(t|\mathbf{Z},\mathbf{W}) = \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_{T^*,\boldsymbol{\beta}}(t|\mathbf{Z})$ for $t \leq L$. Suppose there exists another pair of $\boldsymbol{\gamma}^{\dagger}$ and $\boldsymbol{\beta}^{\dagger}(\tau)$ such that $F_{T,\boldsymbol{\gamma},\boldsymbol{\beta}}(t|\mathbf{Z},\mathbf{W}) = F_{T,\boldsymbol{\gamma}^{\dagger},\boldsymbol{\beta}^{\dagger}}(t|\mathbf{Z},\mathbf{W})$; that is, $\pi(\mathbf{W}^{\mathrm{T}}\boldsymbol{\gamma})/\pi(\mathbf{W}^{\mathrm{T}}\boldsymbol{\gamma}^{\dagger}) = F_{T^*,\boldsymbol{\beta}^{\dagger}}(t|\mathbf{Z})/F_{T^*,\boldsymbol{\beta}}(t|\mathbf{Z}) \equiv C^*_{\mathbf{Z}}$, where the constant $C^*_{\mathbf{Z}}$ does not depend on t (or τ). Using $F_{T^*,\boldsymbol{\beta}}(t|\mathbf{Z}) = F_{T^*,\boldsymbol{\beta}^{\dagger}}(t|\mathbf{Z})/C^*_{\mathbf{Z}}$ and model (2.3), we have $\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\} = \exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}^{\dagger}(\tau C^*_{\mathbf{Z}})\}$. Under condition C1 in Appendix A, $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}^{\dagger}(\tau C^*_{\mathbf{Z}})$ holds for $\tau \leq \tau_{\max}$, which implies $C^*_{\mathbf{Z}} \equiv C^*$, a constant that is not dependent on $\mathbf{Z}$. Furthermore, $\pi(\mathbf{W}^{\mathrm{T}}\boldsymbol{\gamma})/\pi(\mathbf{W}^{\mathrm{T}}\boldsymbol{\gamma}^{\dagger}) = C^*$ leads to $C^* = 1$ (Li, Taylor, and Sy 2001). Therefore, we conclude $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}^{\dagger}(\tau)$, and then $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{\dagger}$ follows by conditions C1 and C2 in Appendix A. Hence, both the cure rate parameter and regression quantiles below $\tau_{\max}$ are identifiable.

For a fixed $\boldsymbol{\gamma}$, let $\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma})$ denote the solution to $\mathbf{U}_n(\boldsymbol{\beta},\tau;\boldsymbol{\gamma}) = \mathbf{0}$ for $0 < \tau \leq \tau_{\max}$. The stochastic integration representation of $\mathbf{U}_n(\boldsymbol{\beta},\tau;\boldsymbol{\gamma})$ suggests a grid-based estimation procedure for $\boldsymbol{\beta}_0(\tau)$ as follows. We denote a partition over the interval $[0, \tau_{\max}]$ by $\mathcal{S}_{q_n} = \{0 \equiv \tau_0 < \tau_1 < \cdots < \tau_{q_n} \equiv \tau_{\max}\}$, where the number of grid points q_n is allowed to depend on n. The estimator of $\boldsymbol{\beta}_0(\tau)$ is defined as a right-continuous piecewise constant function that jumps only at grid points in $\mathcal{S}_{q_n}$. For a fixed $\boldsymbol{\gamma}$, the estimates $\widehat{\boldsymbol{\beta}}(\tau_j,\boldsymbol{\gamma})$, $j = 1,\ldots,q_n$, can be obtained sequentially by solving

$$n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i\left[N_i\left(\exp\left\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(\tau_j,\boldsymbol{\gamma})\right\}\right) - \sum_{k=0}^{j-1} I\left[X_i \geq \exp\left\{\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(\tau_k,\boldsymbol{\gamma})\right\}\right]\{H_{\boldsymbol{\gamma}}(\tau_{k+1}|\mathbf{W}_i) - H_{\boldsymbol{\gamma}}(\tau_k|\mathbf{W}_i)\}\right] = \mathbf{0},$$

where we set $\exp\{\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(\tau_0,\boldsymbol{\gamma})\} = 0$. Similar to the estimator by Fygenson and Ritov (1994) and Peng and Huang (2008), the proposed estimator $\widehat{\boldsymbol{\beta}}(\tau_j,\boldsymbol{\gamma})$ is equivalent to sequentially locating the minimizer of the L_1-type convex objective function:

$$\begin{aligned}\mathcal{L}_j(\mathbf{b}) = {}& \sum_{i=1}^{n}|\Delta_i \log X_i - \Delta_i \mathbf{b}^{\mathrm{T}}\mathbf{Z}_i| + \left|R^* - \mathbf{b}^{\mathrm{T}}\sum_{i=1}^{n}(-\Delta_i\mathbf{Z}_i)\right| \\ &+ \left|R^* - \mathbf{b}^{\mathrm{T}}\sum_{i=1}^{n}\left[2\mathbf{Z}_i\sum_{k=0}^{j-1} I\left[X_i \geq \exp\left\{\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(\tau_k,\boldsymbol{\gamma})\right\}\right] \times \{H_{\boldsymbol{\gamma}}(\tau_{k+1}|\mathbf{W}_i) - H_{\boldsymbol{\gamma}}(\tau_k|\mathbf{W}_i)\}\right]\right|\end{aligned} \tag{2.5}$$

for $j = 1,\ldots,q_n$, where R^* is a large number. In practice, we may set R^* as a constant greater than $10^3(p+1) \times \max\{\|\mathbf{Z}_i\|\colon 1 \leq i \leq n\}$, where $\|\mathbf{a}\|$ denotes the Euclidean norm for a vector $\mathbf{a}$.

2.3 Estimation of $\boldsymbol{\gamma}$: Iterative Approach

We extract the cure information to construct an estimating equation for $\boldsymbol{\gamma}$ (Lu and Ying 2004). The conditional probability that a subject with covariates $\mathbf{Z}$ and $\mathbf{W}$ belongs to the uncured group given that this subject is censored at X is $\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W})\{1-F_{0T^*}(X|\mathbf{Z})\}/\{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W})F_{0T^*}(X|\mathbf{Z})\}$, where F_{0T^*} is the true F_{T^*}. On the other hand, if a subject experiences an event, he/she must belong to the uncured group, and thus

$$P(\eta=1|\Delta,X,\mathbf{Z},\mathbf{W}) = \Delta + (1-\Delta)\frac{\pi\left(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}\right)\{1-F_{0T^*}(X|\mathbf{Z})\}}{1-\pi\left(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}\right)F_{0T^*}(X|\mathbf{Z})}.$$

By noting that $\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W})$ is the probability of $\eta = 1$ given $\mathbf{W}$, we have

$$E\left[\mathbf{W}\left\{P(\eta=1|\Delta,X,\mathbf{Z},\mathbf{W}) - P(\eta=1|\mathbf{W})\right\}\right] = \mathbf{0},$$

which leads to

$$\begin{aligned}\mathbf{S}_n(\boldsymbol{\gamma};F_{T^*}) &= n^{-1}\sum_{i=1}^{n}\frac{\mathbf{W}_i\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_{T^*}(X_i|\mathbf{Z}_i)} \times \left\{\Delta_i - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_{T^*}(X_i|\mathbf{Z}_i)\right\} \\ &= n^{-1}\sum_{i=1}^{n}\int_0^{L}\frac{\mathbf{W}_i\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_{T^*}(t|\mathbf{Z}_i)} \times \mathrm{d}M_i(t;\boldsymbol{\gamma},F_{T^*}).\end{aligned} \tag{2.6}$$

Similarly by replacing t with $\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}$ in (2.6), we can construct an estimating function for $\boldsymbol{\gamma}$,

$$\begin{aligned}\mathbf{R}_n(\boldsymbol{\gamma};\boldsymbol{\beta}(\cdot)) &= n^{-1}\sum_{i=1}^{n}\int_0^{\tau_{\max}}\frac{\mathbf{W}_i\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)u} \\ &\times \left[\mathrm{d}N_i(\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\}) - I[X_i \geq \exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\}]H_{\boldsymbol{\gamma}}(\mathrm{d}u|\mathbf{W}_i)\right].\end{aligned}$$

To solve $\mathbf{R}_n(\boldsymbol{\gamma};\boldsymbol{\beta}(\cdot)) = \mathbf{0}$, we need to know $\boldsymbol{\beta}(\cdot)$, thus leading to an iterative algorithm:

1. Choose an initial value $\widehat{\boldsymbol{\gamma}}^{(0)}$ for $\boldsymbol{\gamma}$.

2. At the mth iteration, set $\exp\{\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(\tau_0, \widehat{\boldsymbol{\gamma}}^{(m)})\} = 0$ and obtain $\widehat{\boldsymbol{\beta}}(\tau_j, \widehat{\boldsymbol{\gamma}}^{(m)})$, $j = 1, \ldots, q_n$, by sequentially minimizing (2.5).
3. Obtain $\widehat{\boldsymbol{\gamma}}^{(m+1)}$ by solving

$$n^{-1}\sum_{i=1}^{n}\sum_{k=0}^{q_n-1}\mathbf{W}_i\Bigg[\frac{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)(\tau_k+\tau_{k+1})/2}\Delta_i \times I\left[\exp\left\{\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(\tau_k, \widehat{\boldsymbol{\gamma}}^{(m)})\right\} \le X_i < \exp\left\{\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(\tau_{k+1}, \widehat{\boldsymbol{\gamma}}^{(m)})\right\}\right] - I\left[X_i \ge \exp\left\{\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(\tau_k, \widehat{\boldsymbol{\gamma}}^{(m)})\right\}\right]\int_{\tau_k}^{\tau_{k+1}}\frac{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)u} \times H_{\boldsymbol{\gamma}}(\mathrm{d}u|\mathbf{W}_i)\Bigg] = \mathbf{0}$$

   using the Newton–Raphson algorithm.
4. Repeat Steps 2 and 3 until a predetermined convergence criterion is met.

The resultant estimators are denoted by $\widehat{\boldsymbol{\gamma}}_I$ and $\widehat{\boldsymbol{\beta}}_I(\cdot) \equiv \widehat{\boldsymbol{\beta}}(\cdot, \widehat{\boldsymbol{\gamma}}_I)$. The initial value $\widehat{\boldsymbol{\gamma}}^{(0)}$ is obtained by treating all the censored subjects as cured and fitting logistic regression of Δ on $\mathbf{W}$. However, not only does the entanglement of $\widehat{\boldsymbol{\beta}}_I(\cdot)$ and $\widehat{\boldsymbol{\gamma}}_I$ make the derivations of their asymptotic properties challenging, but it also makes the computation intensive and unstable, and even causes nonconvergence sometimes. Mao and Wang (2010) discussed such nonconvergence issues in a class of proportional odds cure rate models.

2.4 Estimation of $\boldsymbol{\gamma}$: Nonparametric Approach

To avoid the difficulty arising from the entanglement of $\widehat{\boldsymbol{\beta}}_I(\cdot)$ and $\widehat{\boldsymbol{\gamma}}_I$, we propose an alternative nonparametric approach, which estimates $\boldsymbol{\gamma}_0$ separately from $\boldsymbol{\beta}_0(\cdot)$. Obviously, $\mathbf{S}_n(\boldsymbol{\gamma}; F_{T^*})$ is an unbiased estimating function for $\boldsymbol{\gamma}$ provided that the true F_{0T^*} were known. Following the locally weighted Kaplan–Meier estimator by Wang and Wang (2009), we take a local Nelson–Aalen type estimator for the cumulative hazard function $\Lambda_{T^*}(t|\mathbf{z})$ in the context of cure rate analysis,

$$\widehat{\Lambda}_{T^*}(t|\mathbf{z}) = \int_0^t \frac{\sum_{i=1}^n B_{ni}(\mathbf{z})\mathrm{d}N_i(u)}{\sum_{k=1}^n I(X_k \ge u)\omega_k(\widehat{\boldsymbol{\gamma}}, \widehat{\Lambda}_{T^*})B_{nk}(\mathbf{z})}, \quad (2.7)$$

where

$$\omega_k(\boldsymbol{\gamma}, \Lambda_{T^*}) = \Delta_k + (1-\Delta_k) \times \frac{\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_k)\exp\{-\Lambda_{T^*}(X_k|\mathbf{z})\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_k)+\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_k)\exp\{-\Lambda_{T^*}(X_k|\mathbf{z})\}} \quad (2.8)$$

and $B_{ni}(\mathbf{z})$ is a sequence of Nadaraya–Watson type weights,

$$B_{ni}(\mathbf{z}) = \frac{K_p\{(\mathbf{z}-\mathbf{Z}_i)/h_n\}}{\sum_{k=1}^n K_p\{(\mathbf{z}-\mathbf{Z}_k)/h_n\}}.$$

Here, $K_p(\cdot)$ is a p-variate kernel function and $h_n > 0$ is the bandwidth converging to zero as $n \to \infty$. For ease of exposition, we assume that, except for the intercept, the remaining p components of $\mathbf{Z}$ are continuous and thus adopt a multivariate product kernel $K_p(\mathbf{u}) = \prod_{i=1}^p K(u_i)$, where $K(\cdot)$ is a univariate kernel function and $\mathbf{u} \in \mathbb{R}^p$. As a result, we can obtain an estimator of $F_{0T^*}(t|\mathbf{z})$, $\widehat{F}_{T^*}(t|\mathbf{z}) = 1 - \exp\{-\widehat{\Lambda}_{T^*}(t|\mathbf{z})\}$. Sy and Taylor (2000) and Lu (2010) also considered an estimator similar to (2.7), but without the local weights $B_{ni}(\mathbf{z})$.

In the numerical algorithm, we first obtain the initial value $\widehat{\boldsymbol{\gamma}}^{(0)}$ as before, and then obtain $\widehat{\Lambda}_{T^*}^{(0)}(t|\mathbf{z})$ from (2.7) by taking all ω_k's to be one. Plugging $\widehat{\boldsymbol{\gamma}}^{(0)}$ and $\widehat{\Lambda}_{T^*}^{(0)}(t|\mathbf{z})$ into (2.8) leads to $\omega_k^{(0)}$. At the mth iteration, our algorithm for estimating $\boldsymbol{\gamma}_0$ and $\boldsymbol{\beta}_0(\cdot)$ proceeds as follows.

1. Plug $\omega_k^{(m)}$ into (2.7) and obtain $\widehat{\Lambda}_{T^*}^{(m+1)}(t|\mathbf{z})$.
2. Plug $\widehat{\Lambda}_{T^*}^{(m+1)}(t|\mathbf{z})$ into (2.6) and solve the resultant equation using the Newton–Raphson algorithm to obtain $\widehat{\boldsymbol{\gamma}}^{(m+1)}$.
3. Plug $\widehat{\boldsymbol{\gamma}}^{(m+1)}$ and $\widehat{\Lambda}_{T^*}^{(m+1)}(t|\mathbf{z})$ into (2.8) and obtain $\omega_k^{(m+1)}$.
4. Repeat Steps 1, 2, and 3 until a predetermined convergence criterion is met.

The resultant estimator is denoted by $\widehat{\boldsymbol{\gamma}}_N$, and we plug it into (2.5) to obtain the estimator $\{\widehat{\boldsymbol{\beta}}_N(\tau) \equiv \widehat{\boldsymbol{\beta}}(\tau, \widehat{\boldsymbol{\gamma}}_N)$: $\tau \in \mathcal{S}_{q_n}\}$. For identifiability and computational stability, we set $\widehat{\Lambda}_{T^*}(t|\mathbf{z}) = \infty$ if t is greater than the largest uncensored observation. This nonparametric approach separates the estimation for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}(\cdot)$ and, as a result, there would be no iterative step between (2.4) and (2.6). As shown in the simulation study, this numerically alleviates nonconvergence issues caused by the entanglement of $\widehat{\boldsymbol{\gamma}}_I$ and $\widehat{\boldsymbol{\beta}}_I(\cdot)$ in the iterative approach.

3. ASYMPTOTIC THEORY

3.1 Iterative Approach

Let $a_n = \max_{1\le j\le q_n}|\tau_j - \tau_{j-1}|$, and thus $q_n^{-1} = O(a_n)$ since $\tau_{\max}/a_n \le q_n$. Let ν be a constant, $\nu \in (0, \tau_{\max}]$.

Theorem 1. Under conditions C1–C4 and C5′ in Appendix A, if $a_n \to 0$, then $\widehat{\boldsymbol{\gamma}}_I \to \boldsymbol{\gamma}_0$ in probability and $\sup_{\tau\in[\nu,\tau_{\max}]}\|\widehat{\boldsymbol{\beta}}_I(\tau) - \boldsymbol{\beta}_0(\tau)\| \to 0$ in probability.

Theorem 2. Under conditions C1–C4 and C5′ in Appendix A, if $n^{1/2}a_n \to 0$, then $n^{1/2}(\widehat{\boldsymbol{\gamma}}_I - \boldsymbol{\gamma}_0)$ is asymptotically normal with mean zero and variance-covariance matrix $\boldsymbol{\Omega}_I^{-1}\mathbf{V}_I\boldsymbol{\Omega}_I^{-1}$, and $n^{1/2}\{\widehat{\boldsymbol{\beta}}_I(\tau) - \boldsymbol{\beta}_0(\tau)\}$ converges weakly to a zero-mean Gaussian process with variance-covariance matrix $\boldsymbol{\Sigma}_I(\tau, \tau') = E[\{\boldsymbol{\zeta}^{(1)}(\tau) + \boldsymbol{\zeta}^{(2)}(\tau)\}\{\boldsymbol{\zeta}^{(1)}(\tau') + \boldsymbol{\zeta}^{(2)}(\tau')\}^{\mathrm{T}}]$ for $\tau, \tau' \in [\nu, \tau_{\max}]$, where $\boldsymbol{\Omega}_I$, $\mathbf{V}_I$, $\boldsymbol{\zeta}^{(1)}(\tau)$, and $\boldsymbol{\zeta}^{(2)}(\tau)$ are given in Appendix B.

The derivations of the asymptotic properties for the proposed estimators are challenging due to the entanglement of $\widehat{\boldsymbol{\beta}}_I(\tau)$ and $\widehat{\boldsymbol{\gamma}}_I$. The proofs of the theorems mainly rely upon the stochastic integral representations and modern empirical process theory, which are deferred to Appendix B.

3.2 Nonparametric Approach

Under the nonparametric approach, we can also derive the uniform consistency and weak convergence properties for the proposed estimators, while both the conditions and derivations are different from those of the iterative approach.

Theorem 3. Under conditions C1–C8 in Appendix A, if $a_n \to 0$, then $\widehat{\boldsymbol{\gamma}}_N \to \boldsymbol{\gamma}_0$ in probability and $\sup_{\tau\in[\nu,\tau_{\max}]}\|\widehat{\boldsymbol{\beta}}_N(\tau) - \boldsymbol{\beta}_0(\tau)\| \to 0$ in probability.

Theorem 4. Under conditions C1–C7 and C8′ in Appendix A, if $n^{1/2}a_n \to 0$, then $n^{1/2}(\widehat{\boldsymbol{\gamma}}_N - \boldsymbol{\gamma}_0)$ is asymptotically normal with mean zero and variance-covariance matrix $\boldsymbol{\Omega}_N^{-1}\mathbf{V}_N\boldsymbol{\Omega}_N^{-1}$, and $n^{1/2}\{\widehat{\boldsymbol{\beta}}_N(\tau) - \boldsymbol{\beta}_0(\tau)\}$ converges weakly to a zero-mean Gaussian process with variance-covariance matrix $\boldsymbol{\Sigma}_N(\tau, \tau') =$

$E[\{\boldsymbol{\zeta}^{(1)}(\tau)+\boldsymbol{\zeta}^{(3)}(\tau)\}\{\boldsymbol{\zeta}^{(1)}(\tau')+\boldsymbol{\zeta}^{(3)}(\tau')\}^{\mathrm{T}}]$ for $\tau, \tau' \in [\nu, \tau_{\max}]$, where $\boldsymbol{\Omega}_N$, $\mathbf{V}_N$, and $\boldsymbol{\zeta}^{(3)}(\tau)$ are given in Appendix C.

We establish the asymptotic properties for $\widehat{\boldsymbol{\gamma}}_N$ by employing the results of kernel smoothing and Chen, Linton, and Van Keilegom (2003). Based on the expansion $\widehat{\boldsymbol{\beta}}_N(\tau)-\boldsymbol{\beta}_0(\tau)=\{\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0)-\boldsymbol{\beta}_0(\tau)\}+\{\partial\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0)/\partial\boldsymbol{\gamma}^{\mathrm{T}}\}(\widehat{\boldsymbol{\gamma}}_N-\boldsymbol{\gamma}_0)+o_P(\|\widehat{\boldsymbol{\gamma}}_N-\boldsymbol{\gamma}_0\|)$, we can derive the asymptotic properties for $\widehat{\boldsymbol{\beta}}_N(\tau)$. The proofs of Theorems 3 and 4 are deferred to Appendix C.

Bandwidth selection is often a critical part of nonparametric regression. In practice, we recommend a d-fold cross-validation method for choosing h_n. We randomly divide the data into d nonoverlapping and equal-sized subgroups. For the jth subgroup, $\mathcal{D}_j$, we fit the model using the data excluding subgroup j, $\mathcal{D}_{(-j)}$, and calculate the martingale residuals

$$\mathcal{M}_j^{\mathrm{CV}}(h)=\frac{1}{|\{i:\Delta_i=1 \text{ and } i\in\mathcal{D}_j\}|}\times\sum_{k\in\mathcal{D}_j}\int_0^L\{\mathcal{M}_{(-j)}^{\mathrm{CV}}(t,\mathbf{W}_k)\}^2 dN_k(t),$$

where $|A|$ denotes the cardinality of a set A, and

$$\mathcal{M}_{(-j)}^{\mathrm{CV}}(t,\mathbf{w})=\frac{1}{|\{i: i\in\mathcal{D}_{(-j)}\}|}\times\sum_{i\in\mathcal{D}_{(-j)}}\int_0^t\frac{I(\mathbf{W}_i\le\mathbf{w})\{1-\pi(\widehat{\boldsymbol{\gamma}}_{(-j)}^{\mathrm{T}}\mathbf{W}_i)\}}{1-\pi(\widehat{\boldsymbol{\gamma}}_{(-j)}^{\mathrm{T}}\mathbf{W}_i)\widehat{F}_{T^*(-j)}(u|\mathbf{Z}_i)}\times dM_i(u;\widehat{\boldsymbol{\gamma}}_{(-j)},\widehat{F}_{T^*(-j)}).$$

Here, both $\widehat{\boldsymbol{\gamma}}_{(-j)}$ and $\widehat{F}_{T^*(-j)}$ are estimated using the data from $\mathcal{D}_{(-j)}$. Finally, we choose the bandwidth that minimizes the total martingale residuals $\sum_{j=1}^d \mathcal{M}_j^{\mathrm{CV}}(h)$.

4. SIMULATION STUDY

We conducted extensive simulation studies to examine the finite-sample performance of our proposed methods. First, we generated the survival times T^* of the susceptible subjects from the log-transformed linear model with heteroscedastic errors,

$$\log T^* = b_0 Z + (1+Z)\epsilon,$$

where $(1+Z)\epsilon$ is the error term, and the true parameter value $b_0=-1$. The corresponding quantile regression model (2.3) given $\mathbf{Z}=(1,Z)^{\mathrm{T}}$ is

$$Q_{T^*}(\tau|\mathbf{Z})=\exp\{\beta_0(\tau)+\beta_1(\tau)Z\},$$

where $\beta_0(\tau)=Q_\epsilon(\tau)$, $\beta_1(\tau)=b_0+Q_\epsilon(\tau)$, and $Q_\epsilon(\tau)$ is the τth quantile of ϵ. The susceptibility indicator η was generated from the logistic regression model (2.1) with $\mathbf{W}=\mathbf{Z}$ and the true parameter values $\gamma_0=1$, and $\gamma_1=-0.5$. We simulated Z from Bernoulli(0.5), and took the censoring time $C=\widetilde{C}\wedge L$, where $\widetilde{C}$ was generated from Unif$(0, L+2)$ if $Z<0.5$, and from Unif$(1, L+2)$ otherwise. The study duration L was chosen to yield a censoring rate of 40%, and the cure rate was approximately 32%. We considered three types of error distributions: the standard normal distribution, the extreme value distribution with $F_\epsilon(x)=1-\exp(-e^x)$, and student's t-distribution with 2

Table 1. Simulation results for the iterative method under the cure rate quantile regression model with $\tau=(0.1,\ldots,0.6)$, heteroscedastic errors, and $Z\sim$ Bernoulli(0.5)

	$\beta_0(\tau)$ and γ_0				$\beta_1(\tau)$ and γ_1			
	True	Est	SE	MSE	True	Est	SE	MSE
$\boldsymbol{\beta}(\tau)$				Normal error				
0.1	−1.282	−1.311	0.206	0.043	−2.282	−2.374	0.478	0.236
0.2	−0.842	−0.876	0.167	0.029	−1.842	−1.946	0.394	0.166
0.3	−0.524	−0.575	0.156	0.027	−1.524	−1.670	0.362	0.152
0.4	−0.253	−0.316	0.150	0.026	−1.253	−1.422	0.336	0.141
0.5	0.000	−0.080	0.143	0.027	−1.000	−1.208	0.340	0.159
0.6	0.253	0.161	0.146	0.030	−0.747	−0.975	0.348	0.173
$\boldsymbol{\gamma}$	1.000	0.927	0.139	0.025	−0.500	−0.507	0.159	0.025
$\boldsymbol{\beta}(\tau)$				Extreme value error				
0.1	−2.250	−2.335	0.401	0.167	−3.250	−3.318	0.962	0.928
0.2	−1.500	−1.598	0.270	0.082	−2.500	−2.573	0.678	0.463
0.3	−1.031	−1.143	0.227	0.064	−2.031	−2.108	0.536	0.292
0.4	−0.672	−0.789	0.195	0.052	−1.672	−1.772	0.478	0.238
0.5	−0.367	−0.502	0.170	0.047	−1.367	−1.491	0.425	0.195
0.6	−0.087	−0.222	0.158	0.043	−1.087	−1.211	0.396	0.172
$\boldsymbol{\gamma}$	1.000	0.908	0.143	0.029	−0.500	−0.435	0.152	0.027
$\boldsymbol{\beta}(\tau)$				Student's $t_{(2)}$ error				
0.1	−1.886	−1.991	0.515	0.276	−2.886	−3.268	1.355	1.977
0.2	−1.061	−1.147	0.261	0.075	−2.061	−2.303	0.679	0.519
0.3	−0.617	−0.683	0.207	0.047	−1.617	−1.839	0.492	0.291
0.4	−0.289	−0.358	0.178	0.036	−1.289	−1.491	0.403	0.203
0.5	0.000	−0.086	0.164	0.034	−1.000	−1.242	0.382	0.204
0.6	0.289	0.194	0.167	0.037	−0.711	−0.994	0.390	0.231
$\boldsymbol{\gamma}$	1.000	0.942	0.145	0.024	−0.500	−0.531	0.149	0.023

NOTE: "Est" is the average value of the parameter estimates, "SE" is the sample standard error of the estimates, and "MSE" is the mean squared errors of the parameter estimates.

Table 2. Simulation results for the nonparametric method under the cure rate quantile regression model with $\tau = (0.1, \ldots, 0.6)$, heteroscedastic errors, and $Z \sim$ Bernoulli(0.5)

	$\beta_0(\tau)$ and γ_0						$\beta_1(\tau)$ and γ_1					
	True	Est	SE	ESE	CP	MSE	True	Est	SE	ESE	CP	MSE
$\boldsymbol{\beta}(\tau)$						Normal error						
0.1	−1.282	−1.286	0.200	0.210	0.939	0.040	−2.282	−2.316	0.468	0.520	0.952	0.220
0.2	−0.842	−0.852	0.174	0.176	0.949	0.030	−1.842	−1.890	0.402	0.420	0.940	0.164
0.3	−0.524	−0.528	0.158	0.161	0.936	0.025	−1.524	−1.584	0.378	0.383	0.939	0.146
0.4	−0.253	−0.258	0.155	0.153	0.928	0.024	−1.253	−1.332	0.358	0.365	0.939	0.134
0.5	0.000	−0.002	0.159	0.153	0.921	0.025	−1.000	−1.081	0.360	0.362	0.942	0.136
0.6	0.253	0.261	0.166	0.155	0.913	0.028	−0.747	−0.843	0.370	0.367	0.918	0.146
$\boldsymbol{\gamma}$	1.000	0.992	0.267	0.302	0.968	0.071	−0.500	−0.588	0.362	0.399	0.967	0.139
$\boldsymbol{\beta}(\tau)$						Extreme value error						
0.1	−2.250	−2.268	0.369	0.402	0.938	0.136	−3.250	−3.307	0.884	0.899	0.952	0.783
0.2	−1.500	−1.523	0.272	0.280	0.946	0.075	−2.500	−2.548	0.629	0.677	0.945	0.397
0.3	−1.031	−1.039	0.219	0.225	0.934	0.048	−2.031	−2.094	0.517	0.544	0.952	0.271
0.4	−0.672	−0.681	0.195	0.193	0.927	0.038	−1.672	−1.740	0.452	0.468	0.952	0.209
0.5	−0.366	−0.372	0.182	0.175	0.920	0.033	−1.367	−1.429	0.412	0.424	0.951	0.174
0.6	−0.087	−0.084	0.172	0.162	0.912	0.030	−1.087	−1.146	0.380	0.384	0.933	0.148
$\boldsymbol{\gamma}$	1.000	0.992	0.269	0.290	0.970	0.072	−0.500	−0.553	0.356	0.373	0.965	0.129
$\boldsymbol{\beta}(\tau)$						Student's $t_{(2)}$ error						
0.1	−1.886	−1.919	0.480	0.619	0.947	0.231	−2.886	−3.059	1.259	1.701	0.972	1.615
0.2	−1.061	−1.088	0.272	0.290	0.935	0.075	−2.061	−2.202	0.682	0.734	0.953	0.484
0.3	−0.617	−0.644	0.201	0.211	0.952	0.041	−1.617	−1.744	0.500	0.525	0.947	0.266
0.4	−0.289	−0.316	0.181	0.182	0.936	0.034	−1.289	−1.420	0.438	0.436	0.932	0.209
0.5	0.000	−0.028	0.172	0.172	0.943	0.030	−1.000	−1.165	0.430	0.405	0.907	0.212
0.6	0.289	0.257	0.176	0.176	0.922	0.032	−0.711	−0.915	0.430	0.406	0.898	0.226
$\boldsymbol{\gamma}$	1.000	0.883	0.251	0.289	0.916	0.077	−0.500	−0.579	0.332	0.387	0.967	0.117

NOTE: "Est" is the average value of the parameter estimates, "SE" is the sample standard error of the estimates, "ESE" is the average of the bootstrap estimated standard errors, "CP" is the coverage probability of the 95% confidence intervals using the bootstrap, and "MSE" is the mean squared errors of the parameter estimates.

degrees of freedom. For estimation of regression quantiles, we set $\tau_{\max} = 0.6$ and adopted an equally spaced grid over interval [0.02, 0.6] with a step size of 0.02. A general guidance on choosing $\tau_{\max}$ is to first estimate $\inf_{\mathbf{z}} F_{T^*}(L|\mathbf{Z} = \mathbf{z})$ using the nonparametric local kernel smoothing estimator $\widehat{F}_{T^*}(t|\mathbf{z})$. We initially set $\tau_{\max}$ to be close to $\min_{i=1,\ldots,n} \widehat{F}_{T^*}(L|\mathbf{Z} = \mathbf{z}_i)$, and then select the final $\tau_{\max}$ in an adaptive manner as follows. If all the regression quantiles over $[0, \tau_{\max}]$ can be estimated, we increase $\tau_{\max}$ by some small step size, for example, 0.05 or 0.1; otherwise, we decrease $\tau_{\max}$ slightly. By this trial-and-error way, we can push $\tau_{\max}$ to the largest value at which the model parameters can be identified. We examined both the iterative and nonparametric approaches, and we set the sample size $n = 200$ and replicated 1000 simulations.

Our simulation showed that the iterative approach was relatively unstable and sensitive to the initial values, which on average resulted in 28% nonconvergence cases among 1000 simulations (also see Mao and Wang 2010, p. 307). For estimation of $\boldsymbol{\gamma}$, $\mathbf{R}_n(\boldsymbol{\gamma}; \boldsymbol{\beta}(\cdot)) = \mathbf{0}$ would be automatically satisfied if $\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i) = 0$ or 1 (when the estimate of $\boldsymbol{\gamma}$ converges to a very small or very large value, for example, close to the boundaries of the parameter searching space). To facilitate the comparison, Table 1 summarizes the estimation results for those converged cases, which shows that after discarding those nonconvergence replications, the performance of the iterative approach is reasonable under the three different error distributions.

By contrast, the nonparametric approach separates the estimation of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}(\cdot)$, which is shown to result in better convergence. Under the three error distributions, only for the $t_{(2)}$ heteroscedastic error, there were about 0.5% nonconvergence cases among 1000 simulations. As to variance estimation, we generated 200 bootstrap samples for each simulated dataset. In particular, we generated n independent variates from Exp(1), and multiplied each term i in the estimating equations by the ith Exp(1) variate. Table 2 presents the simulation results using the nonparametric approach. Clearly, the biases are essentially negligible, the estimated standard errors (ESE) using the bootstrap method agree well with the sample standard errors (SE), and the coverage probabilities (CP) of the bootstrap confidence intervals are around the nominal level 95%. Comparing Tables 1 and 2, the MSEs are close between the iterative and nonparametric methods.

Second, we generated covariate Z from Unif(0, 1), while keeping the rest of the data generation scheme the same as before. The kernel smoothing procedure was used to estimate the nonparametric function $F_{0T^*}(\cdot|\mathbf{Z})$. As there was only one continuous covariate, we adopted the biquadratic kernel $K(u) = (15/16)(1 - u^2)^2 I(|u| < 1)$; for the bivariate case (as in online Supplementary Material A), we employed a fourth-order product kernel, $K_2(u_1, u_2) = K(u_1)K(u_2)$. We took the sample size $n = 400$ and selected the bandwidth h_n through the eight-fold cross-validation procedure. We conducted 500

Table 3. Simulation results for the nonparametric method under the cure rate quantile regression model with $\tau = (0.1, \ldots, 0.6)$, heteroscedastic errors, and $Z \sim$ Unif(0, 1)

	$\beta_0(\tau)$ and γ_0						$\beta_1(\tau)$ and γ_1					
	True	Est	SE	ESE	CP	MSE	True	Est	SE	ESE	CP	MSE
$\boldsymbol{\beta}(\tau)$						Normal error						
0.1	−1.282	−1.276	0.251	0.257	0.932	0.063	−2.282	−2.268	0.525	0.558	0.942	0.275
0.2	−0.842	−0.828	0.208	0.216	0.948	0.044	−1.842	−1.867	0.446	0.461	0.952	0.199
0.3	−0.524	−0.511	0.199	0.196	0.938	0.040	−1.524	−1.559	0.411	0.420	0.950	0.170
0.4	−0.253	−0.228	0.185	0.192	0.946	0.035	−1.253	−1.315	0.394	0.406	0.946	0.159
0.5	0.000	0.039	0.185	0.189	0.940	0.036	−1.000	−1.093	0.390	0.398	0.940	0.161
0.6	0.253	0.295	0.189	0.186	0.924	0.037	−0.747	−0.862	0.411	0.395	0.914	0.182
$\boldsymbol{\gamma}$	1.000	0.987	0.258	0.295	0.976	0.067	−0.500	−0.573	0.430	0.489	0.970	0.190
$\boldsymbol{\beta}(\tau)$						Extreme value error						
0.1	−2.250	−2.279	0.476	0.480	0.938	0.227	−3.250	−3.191	0.980	1.046	0.948	0.962
0.2	−1.500	−1.517	0.333	0.342	0.936	0.111	−2.500	−2.503	0.706	0.730	0.950	0.497
0.3	−1.031	−1.044	0.278	0.273	0.944	0.077	−2.031	−2.057	0.570	0.588	0.952	0.324
0.4	−0.672	−0.669	0.234	0.242	0.958	0.055	−1.672	−1.732	0.493	0.515	0.946	0.246
0.5	−0.366	−0.356	0.215	0.215	0.952	0.046	−1.367	−1.448	0.445	0.460	0.946	0.204
0.6	−0.087	−0.078	0.203	0.193	0.908	0.041	−1.087	−1.183	0.434	0.414	0.930	0.197
$\boldsymbol{\gamma}$	1.000	0.922	0.246	0.269	0.968	0.067	−0.500	−0.505	0.410	0.437	0.964	0.168
$\boldsymbol{\beta}(\tau)$						Student's $t_{(2)}$ error						
0.1	−1.886	−1.935	0.598	0.651	0.944	0.360	−2.886	−3.014	1.333	1.439	0.960	1.790
0.2	−1.061	−1.082	0.332	0.349	0.956	0.110	−2.061	−2.167	0.714	0.749	0.952	0.520
0.3	−0.617	−0.635	0.253	0.261	0.958	0.064	−1.617	−1.733	0.540	0.559	0.946	0.305
0.4	−0.289	−0.290	0.220	0.227	0.952	0.048	−1.289	−1.450	0.480	0.488	0.936	0.256
0.5	0.000	−0.006	0.211	0.213	0.938	0.044	−1.000	−1.176	0.453	0.456	0.934	0.236
0.6	0.289	0.279	0.217	0.216	0.940	0.047	−0.711	−0.930	0.478	0.464	0.908	0.276
$\boldsymbol{\gamma}$	1.000	0.875	0.253	0.308	0.930	0.079	−0.500	−0.568	0.411	0.511	0.974	0.173

NOTE: "Est" is the average value of the parameter estimates, "SE" is the sample standard error of the estimates, "ESE" is the average of the bootstrap estimated standard errors, "CP" is the coverage probability of the 95% confidence intervals using the bootstrap, and "MSE" is the mean squared errors of the parameter estimates.

replications, and for each simulated dataset 200 bootstrap samples were generated as before to estimate the variance. For τ from 0.1 to 0.6 in Table 3, the point estimates of $\beta_0(\tau)$ and $\beta_1(\tau)$ are close to the true values, and the variance estimation also performs well. The same is true for the cure rate parameters γ_0 and γ_1. Thus, we conclude that the proposed nonparametric method generally works well with continuous covariates and the bandwidth selection procedure is also viable.

We further investigated the performance of the proposed nonparametric method under the misspecification of the logistic model for the cure indicator or that of the quantile regression model for survival times of the susceptible subjects. Meanwhile, we made comparisons with the AFT cure rate method and the Cox PH cure rate method, and also explored a hypothetical situation where the distribution function $F_{0T^*}(\cdot|\mathbf{Z})$ were known. In addition, we assessed the behavior of the proposed nonparametric method when two continuous covariates were involved, for which we adopted a bivariate product kernel for nonparametric kernel smoothing. The detailed simulation results are presented in online Supplementary Material A.

5. BONE MARROW TRANSPLANTATION STUDY

As an illustration, we applied the cure rate quantile regression model to a bone marrow transplantation (BMT) study (Szydlo et al. 1997). This study was conducted between the years 1985 and 1991 and involved 1715 leukemia patients. To minimize potential side effects, the transplanted stem cells should match the patient's own stem cells as closely as possible. The matching is usually based on the proteins on the surface of the cells, namely human leukemia-associated antigens (HLA). Patients were treated with the BMT from either HLA-identical siblings, HLA-matched, or mismatched unrelated donors. The primary endpoint was the time to cancer relapse or death while in remission, recorded in months. From the clinical perspective, patients may be considered as "cured" if they could survive the risk of the graft-versus-host disease (GVHD), which is a reaction of donated stem cells against patients' own tissues. If there is a strong immune reaction of GVHD, patients are likely to die soon after the transplantation. The censoring rate in this BMT study was 49.4%, and Figure 1 shows the Kaplan–Meier survival curves for the three donor groups. After approximately five years follow-up, there appears to be a stable plateau for each arm, which indicates a possible cure fraction in this patient population.

In our analysis, the covariates of interest included three donor types: HLA-identical siblings (71.4%), HLA-matched unrelated (22.3%), and mismatched unrelated donors (6.3%); three disease types: acute lymphoblastic leukemia (ALL, 31.3%), acute myelogenous leukemia (AML, 19.8%), and chronic myelogenous leukemia (CML, 48.9%); and the Karnofsky score (taking a value of 1 if the Karnofsky score $\geq$ 90 and 0 otherwise; 80.6% patients with the Karnofsky score = 1). For the two three-level categorical variables, we created two indicator variables by

Table 4. Analysis results of the BMT data fitted by the Cox proportional hazards cure rate model, the accelerated failure time (AFT) cure rate model, and the cure rate quantile regression (QR) model

		Survival model parameters			Cure rate parameters		
Cure model	Covariate	Est	SE	p-value	Est	SE	p-value
Cox	Intercept	–	–	–	1.925	0.365	< 0.001
	HLA-match	0.424	0.091	< 0.001	1.535	0.320	< 0.001
	Mismatched	0.610	0.127	< 0.001	1.547	0.759	0.041
	AML	0.087	0.095	0.364	0.393	0.258	0.128
	CML	−0.110	0.090	0.217	−0.067	0.203	0.740
	Karnofsky	−0.441	0.087	< 0.001	−1.287	0.302	< 0.001
AFT	Intercept	–	–	–	0.939	0.177	< 0.001
	HLA-match	−0.162	0.080	0.042	1.304	0.169	< 0.001
	Mismatched	−0.258	0.123	0.036	1.840	0.359	< 0.001
	AML	0.067	0.101	0.508	0.437	0.168	0.009
	CML	0.057	0.091	0.532	−0.216	0.151	0.154
	Karnofsky	0.286	0.092	0.002	−0.981	0.174	< 0.001
QR	Intercept	–	–	–	0.827	0.376	0.028
	HLA-match	–	–	–	1.249	0.297	< 0.001
	Mismatched	–	–	–	1.425	0.507	0.005
	AML	–	–	–	0.376	0.217	0.084
	CML	–	–	–	−0.163	0.182	0.371
	Karnofsky	–	–	–	−0.967	0.357	0.007

taking the HLA-identical sibling and ALL as the baseline, respectively. We took $\mathbf{Z} = \mathbf{W}$, and 200 bootstrap samples for variance estimation. Figure 2 displays the quantile regression estimates of covariate effects and the corresponding 95% pointwise confidence intervals for $\tau \in [0.1, 0.7]$ with a step size of 0.01. As expected, patients treated with the BMT from HLA-matched unrelated donors and mismatched unrelated donors had significantly worse survival than those with HLA-sibling donors, particularly for higher quantiles of survival times. Patients with higher Karnosky scores survived longer across all the quantiles. There was no significant difference in survival across

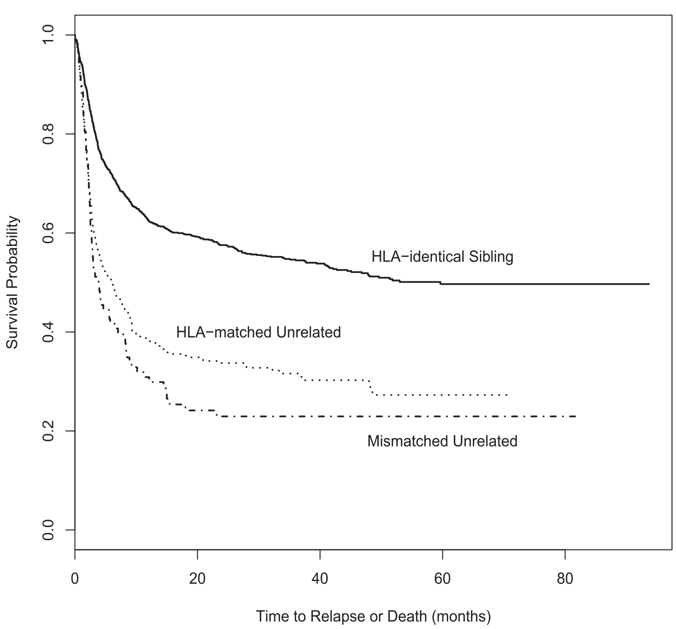

Figure 1. Kaplan–Meier survival curves stratified by the donor type in the BMT data.

the three disease types. Table 4 presents the estimates for the cure rate parameters in the logistic regression, from which one can see that patients treated with the BMT from HLA-identical siblings or with higher Karnofsky scores were more likely to be long-term survivors. For comparison, we also analyzed the BMT data using the Cox PH cure rate model and the AFT cure rate model, for which the results are also shown in Table 4. Similar conclusions can be drawn, while these traditional hazard-based and mean-based methods can only provide overall quantification of covariate effects. In contrast, our proposed cure rate quantile regression method can provide global trends of covariate effects along different quantiles of survival times. For example, for lower quantiles there appeared to be no survival difference between different donor types, while for higher quantiles the survival difference became larger, especially for the mismatched unrelated donor group.

6. CONCLUDING REMARKS

We have proposed cure rate quantile regression for censored data with a survival fraction under the usual censoring assumption, that is, survival times and censoring times are conditionally independent given covariates. Under the global linearity assumption, the estimate for the regression quantile at a particular τ requires the availability of the estimates for all regression quantiles below τ. If such linearity only holds at one specific quantile level τ, the work by Wang and Wang (2009) may be extended along this direction. Identifiability is an inherent and subtle issue in censored quantile regression. Regression quantiles with τ close to 1 may not be identifiable due to a lack of event information in the upper tail. Thus, we confine the estimation of regression quantiles to the quantile levels below $\tau_{\max}$. In principle, $\tau_{\max}$ should satisfy the identifiability condition C4 in Appendix A, while in practice the selection of $\tau_{\max}$ needs to be settled in an adaptive manner as described in Section 4.

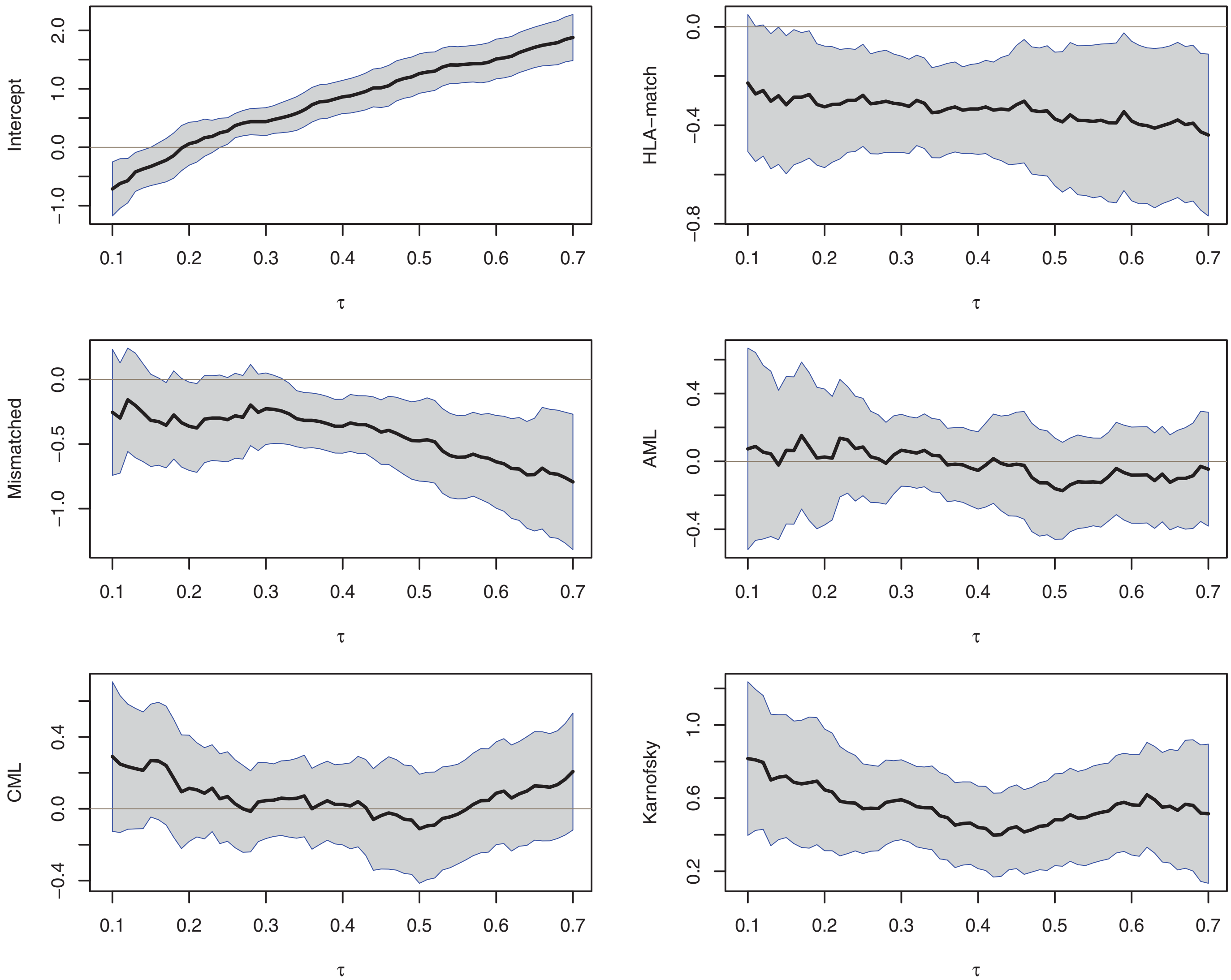

Figure 2. Estimated covariate effects for the BMT data and the corresponding 95% pointwise confidence intervals using the nonparametric method under the cure rate quantile regression model. The online version of this figure is in color.

Our additional simulation studies in online Supplementary Material A shows that the proposed nonparametric approach is still feasible when two continuous covariates are involved in $\mathbf{Z}$. However, the nonparametric estimation for $F_{0T^*}(\cdot|\mathbf{Z})$ would deteriorate as the dimension of $\mathbf{Z}$ increases. If there exist many (continuous) covariates, we suggest to first perform dimension reduction as in the article by Wang, Zhou, and Li (2013), and then take more cautious exploration and analysis of the data, for example, initially fit the model with each covariate one by one, and then add important covariates one at a time into the model. In survival analysis, model fitting and checking are often based on martingale residuals (Barlow and Prentice 1988; Lin, Wei, and Ying 1993). The forms of the limiting processes of the cumulative martingale residuals constructed from the estimating Equations (2.4) and (2.6) are complicated because the nonparametric estimation for $F_{0T^*}(\cdot|\mathbf{Z})$ is involved, for which further research is warranted. It is also possible to extend the proposed method to partially linear models and a mixture of quantile-variant and -invariant effects to enhance the flexibility and efficiency (Neocleous and Portnoy 2009; Qian and Peng 2010).

APPENDIX A: CONDITIONS AND LEMMAS

We first introduce the notation: $F_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W}) = P(X \le t|\mathbf{Z},\mathbf{W})$, $\bar{F}_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W}) = 1 - F_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W})$, $\widetilde{F}_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W}) = P(X \le t, \Delta = 1|\mathbf{Z},\mathbf{W})$, $\bar{f}_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W}) = -f_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W}) = -\mathrm{d}F_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W})/\mathrm{d}t$, $\widetilde{f}_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W}) = \mathrm{d}\widetilde{F}_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W})/\mathrm{d}t$, $F_C(t|\mathbf{Z},\mathbf{W}) = P(C \le t|\mathbf{Z},\mathbf{W})$, and $\bar{F}_C(t|\mathbf{Z},\mathbf{W}) = 1 - F_C(t|\mathbf{Z},\mathbf{W})$, and then impose the conditions as follows.

C1. Let $\mathcal{K}$ be a compact subset of $\mathbb{R}^{q+1}$ that contains $\boldsymbol{\gamma}_0$ as its interior point. With probability 1, both $\mathbf{Z}$ and $\mathbf{W}$ are bounded, and $E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}) > 0$ and $E(\mathbf{W}\mathbf{W}^{\mathrm{T}}) > 0$.

C2. Each component of $E[\mathbf{Z}\bar{F}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\}|\mathbf{Z},\mathbf{W})\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\tau\}^{-1}]$ and $E\{\mathbf{Z}N(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\})\}$ is a Lipschitz function of $\tau \in (0, \tau_{\max}]$ for every $\boldsymbol{\gamma} \in \mathcal{K}$. Both $f_{X,\boldsymbol{\gamma}}(t|\mathbf{z},\mathbf{w})$ and $\widetilde{f}_{X,\boldsymbol{\gamma}}(t|\mathbf{z},\mathbf{w})$ are uniformly bounded away from zero for $\boldsymbol{\gamma} \in \mathcal{K}$, $t \in (0, L]$, $\mathbf{z} \in \mathcal{Z}$, and $\mathbf{w} \in \mathcal{W}$, where $\mathcal{Z}$ and $\mathcal{W}$ are the domains of covariates $\mathbf{Z}$ and $\mathbf{W}$, respectively. In addition,

$\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})$ is uniformly bounded away from 0 and 1 for $\mathbf{w} \in \mathcal{W}$ and $\boldsymbol{\gamma} \in \mathcal{K}$.

C3. Each component of

$$E\bigg[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\bar{f}_{X,\boldsymbol{\gamma}}(\exp(\mathbf{Z}^{\mathrm{T}}\mathbf{b})|\mathbf{Z},\mathbf{W})\exp(\mathbf{Z}^{\mathrm{T}}\mathbf{b})\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}) \times\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\tau\}^{-1}\bigg]\bigg[E\{\mathbf{Z}\mathbf{Z}^{\mathrm{T}} \times \widetilde{f}_{X,\boldsymbol{\gamma}}(\exp(\mathbf{Z}^{\mathrm{T}}\mathbf{b})|\mathbf{Z},\mathbf{W})\exp(\mathbf{Z}^{\mathrm{T}}\mathbf{b})\}\bigg]^{-1}$$

is uniformly bounded in $\mathbf{b} \in \mathcal{B}(d_0)$ and $\boldsymbol{\gamma} \in \mathcal{K}$, where

$$\mathcal{B}(d) = \left\{\mathbf{b} \in \mathbb{R}^{p+1}: \inf_{\tau\in(0,\tau_{\max}],\boldsymbol{\gamma}\in\mathcal{K}} \|E[\mathbf{Z}N(\exp(\mathbf{Z}^{\mathrm{T}}\mathbf{b})) - \mathbf{Z}N(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\})]\| \le d\right\}$$

and d_0 is chosen such that $\mathcal{B}(d_0)$ contains $\{\boldsymbol{\beta}_0(\tau): \tau \in (0,\tau_{\max}]\}$.

C4. Assume that $\inf_{\tau\in[\nu,\tau_{\max}]}\mathrm{Eigmin}\big\{E[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\widetilde{f}_{X,\boldsymbol{\gamma}_0}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}]\big\} > 0$ for any $\nu \in (0,\tau_{\max}]$ and $\boldsymbol{\beta}(\tau) \in \mathcal{B}(d_0)$, where $\mathrm{Eigmin}\{\cdot\}$ denotes the minimum eigenvalue of a matrix.

C5. For $\boldsymbol{\gamma}$ in the neighborhood of $\boldsymbol{\gamma}_0$, the matrix $\Gamma_1(\boldsymbol{\gamma}, F_{0T^*})$ defined in Appendix C is negative definite.

C5′. For $\boldsymbol{\gamma}$ in the neighborhood of $\boldsymbol{\gamma}_0$, the matrix $\Psi(\boldsymbol{\beta}_0(\cdot),\tau_{\max};\boldsymbol{\gamma})$ defined in Appendix B is negative definite.

C6. The kernel function $K(\cdot)$ has a compact support in $\mathbb{R}$, and it is an ℓth order kernel satisfying that $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$, $\int_{\mathbb{R}} K^2(u)\mathrm{d}u < \infty$, $\int_{\mathbb{R}} u^j K(u)\mathrm{d}u = 0$ for $1 \le j < \ell$, and $\int_{\mathbb{R}} |u|^{\ell} K(u)\mathrm{d}u < \infty$. Moreover, it is Lipschitz continuous of order ℓ, with $\ell \ge 2$.

C7. The first ℓ partial derivatives of the density function of $\mathbf{Z}$, $f_{\mathbf{Z}}(\mathbf{z})$, with respect to $\mathbf{z}$ are uniformly bounded for $\mathbf{z} \in \mathcal{Z}$, and $f_{0T^*}(t|\mathbf{z})$ and $f_C(t|\mathbf{z},\mathbf{w})$ are bounded (uniformly in t, $\mathbf{z}$, and $\mathbf{w}$) away from infinity, and the first ℓ partial derivatives of $f_{0T^*}(t|\mathbf{z})$ and $f_C(t|\mathbf{z},\mathbf{w})$ with respect to $\mathbf{z}$ or $\mathbf{w}$ are uniformly bounded in $t \in (0, L]$, $\mathbf{z} \in \mathcal{Z}$, and $\mathbf{w} \in \mathcal{W}$, where $f_{0T^*}(t|\mathbf{z}) = \mathrm{d}F_{0T^*}(t|\mathbf{z})/\mathrm{d}t$ and $f_C(t|\mathbf{z},\mathbf{w}) = \mathrm{d}F_C(t|\mathbf{z},\mathbf{w})/\mathrm{d}t$.

C8. The bandwidth $h_n = O(n^{-\upsilon})$, where $0 < \upsilon < \min(1/p, 1/\ell)$.

C8′. The bandwidth $h_n = O(n^{-\upsilon})$, where $1/(2\ell) < \upsilon < 1/(3p)$ and $\ell > 3p/2$.

Conditions C1–C3 impose the boundedness assumptions of covariates and the density functions associated with the observed data, which are common in censored quantile regression; condition C4 is the positive definite assumption for the inference of the quantile regression parameter. Conditions C5 and C5′ are the positive definite assumptions of the "Hessian" matrices of the cure rate parameters for the nonparametric and iterative approaches, respectively. Condition C7 allows to apply the Taylor expansion to determine the order of convergence of the estimators. For ease of exposition, we assume continuous $\mathbf{Z}$, while if discrete covariates are involved, all proofs are essentially the same except for replacing probability density functions by probability mass functions and integration by summation. Condition C8 on the bandwidth is needed to obtain the consistency of the proposed estimators and C8′ is a strengthened version for establishing their weak convergence properties. Due to the dependency of ℓ on p, a higher-order kernel function is needed to control the bias for a larger p (Wang, Zhou, and Li 2013).

Lemma A.1. Under conditions C1–C4, if $a_n \to 0$,

$$\sup_{\tau\in[\nu,\tau_{\max}],\boldsymbol{\gamma}\in\mathcal{K}} \|\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}) - \boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})\| \xrightarrow{P} 0,$$

where $\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})$ is the zero-crossing of $E\{\mathbf{U}_n(\boldsymbol{\beta},\tau;\boldsymbol{\gamma})\}$.

Lemma A.2. Under conditions C1–C4, for any sequence $\{\widetilde{\boldsymbol{\beta}}_n(\tau): \tau \in (0,\tau_{\max}]\}$, if $\sup_{\tau\in(0,\tau_{\max}]}\|\boldsymbol{\mu}(\widetilde{\boldsymbol{\beta}}_n(\tau),\boldsymbol{\gamma}_0) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)\| \xrightarrow{P} 0$, where

$$\boldsymbol{\mu}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma}) = E\left[\mathbf{Z}N(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\})\right] = E\left[\mathbf{Z}\widetilde{F}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})\right],$$

then we have

$$\sup_{\tau\in(0,\tau_{\max}]}\bigg\| n^{-1/2}\sum_{i=1}^n \mathbf{Z}_i\big[N_i(\exp\{\mathbf{Z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_n(\tau)\}) - N_i(\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\})\big] - n^{1/2}\{\boldsymbol{\mu}(\widetilde{\boldsymbol{\beta}}_n(\tau),\boldsymbol{\gamma}_0) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)\}\bigg\| \xrightarrow{P} 0; \quad \text{(A.1)}$$

if $\sup_{\tau\in(0,\tau_{\max}]}\|\widetilde{\boldsymbol{\mu}}(\widetilde{\boldsymbol{\beta}}_n(\tau),\boldsymbol{\gamma}_0) - \widetilde{\boldsymbol{\mu}}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)\| \xrightarrow{P} 0$, where

$$\begin{aligned}\widetilde{\boldsymbol{\mu}}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma}) &= E\left[\mathbf{Z}I(X \ge \exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\})\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})/\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\tau\}\right]\\ &= E\left[\mathbf{Z}\bar{F}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})/\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\tau\}\right],\end{aligned}$$

then we have

$$\sup_{\tau\in(0,\tau_{\max}]}\bigg\| n^{-1/2}\sum_{i=1}^n \mathbf{Z}_i\big\{I[X_i \ge \exp\{\mathbf{Z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_n(\tau)\}] - I[X_i \ge \exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_0(\tau)\}]\big\} \times \frac{\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}_i)}{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}_i)\tau} - n^{1/2}\{\widetilde{\boldsymbol{\mu}}(\widetilde{\boldsymbol{\beta}}_n(\tau),\boldsymbol{\gamma}_0) - \widetilde{\boldsymbol{\mu}}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)\}\bigg\| \xrightarrow{P} 0. \quad \text{(A.2)}$$

Lemma A.3. Under conditions C1 and C6–C8,

$$\sup_{t\in[0,L],\mathbf{z}\in\mathcal{Z}} |\widetilde{\Lambda}_{T^*}(t|\mathbf{z}) - \Lambda_{0T^*}(t|\mathbf{z})| = O_P(r_n)$$

and

$$\sup_{t\in[0,L],\mathbf{z}\in\mathcal{Z}} |\widetilde{F}_{T^*}(t|\mathbf{z}) - F_{0T^*}(t|\mathbf{z})| = O_P(r_n),$$

where

$$\widetilde{\Lambda}_{T^*}(t|\mathbf{z}) = \int_0^t \frac{\sum_{i=1}^n B_{ni}(\mathbf{z})\mathrm{d}N_i(u)}{\sum_{k=1}^n I(X_k \ge u)\omega_k(\boldsymbol{\gamma}_0,\Lambda_{0T^*})B_{nk}(\mathbf{z})},$$

$r_n = \max\{(\log n/(nh_n^p))^{1/2}, h_n^{\ell}\}$, $\widetilde{F}_{T^*}(t|\mathbf{z}) = 1-\exp\{-\widetilde{\Lambda}_{T^*}(t|\mathbf{z})\}$, and $\Lambda_{0T^*}(t|\mathbf{z}) = -\log\{1-F_{0T^*}(t|\mathbf{z})\}$.

Furthermore, for any $\boldsymbol{\gamma}^{\dagger} \in \mathcal{K}$, we can define $\widetilde{F}^{\dagger}_{T^*}(t|\mathbf{z})$ and $F^{\dagger}_{0T^*}(t|\mathbf{z})$ by replacing $\boldsymbol{\gamma}_0$ with $\boldsymbol{\gamma}^{\dagger}$ in their counterparts, respectively. Note that $F^{\dagger}_{0T^*}(t|\mathbf{z})$ is the conditional distribution function of T^* at $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{\dagger}$. Then, we have $\sup_{\boldsymbol{\gamma}^{\dagger}\in\mathcal{K},t\in[0,L],\mathbf{z}\in\mathcal{Z}} |\widetilde{F}^{\dagger}_{T^*}(t|\mathbf{z}) - F^{\dagger}_{0T^*}(t|\mathbf{z})| = O_P(r_n)$.

The detailed proofs of the three lemmas are given in online Supplementary Material B.

APPENDIX B: PROOFS OF THEOREMS 1 AND 2

Proof of Theorem 1.

We first prove the consistency of $\widehat{\boldsymbol{\gamma}}_I$. Denote $\mathbf{U}(\boldsymbol{\beta},\tau;\boldsymbol{\gamma}) = E\{\mathbf{U}_n(\boldsymbol{\beta},\tau;\boldsymbol{\gamma})\}$, then $\mathbf{U}(\boldsymbol{\beta}_0,\tau;\boldsymbol{\gamma}_0) = \mathbf{0}$. For a given $\boldsymbol{\gamma} \in \mathcal{K}$, $\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})$ is the solution of equation $\mathbf{U}(\boldsymbol{\beta},\tau;\boldsymbol{\gamma}) = \mathbf{0}$. Following theorem 2.7.5, the preservation of the Donsker property by van der Vaart and Wellner (1996), and the boundedness of $\mathbf{W}$ under condition C1, the class of functions

$$\left\{\int_0^{\tau_{\max}} \frac{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})u}\bigg[\mathrm{d}N(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(u)\}) - I[X \ge \exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(u)\}]H_{\boldsymbol{\gamma}}(\mathrm{d}u|\mathbf{W})\bigg]: \boldsymbol{\gamma} \in \mathcal{K}, \boldsymbol{\beta}(\tau) \in \mathcal{B}(d), \tau \in [\nu,\tau_{\max}]\right\}$$

is Donsker. Thus, together with Lemma A.1, we have $\sup_{\boldsymbol{\gamma}\in\mathcal{K}} \|\mathbf{R}_n(\boldsymbol{\gamma};\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma})) - \mathbf{R}(\boldsymbol{\gamma};\boldsymbol{\beta}^*(\cdot,\boldsymbol{\gamma}))\| \to 0$ almost surely, where $\mathbf{R}(\boldsymbol{\gamma};\boldsymbol{\beta}^*(\cdot,\boldsymbol{\gamma})) = E\{\mathbf{R}_n(\boldsymbol{\gamma};\boldsymbol{\beta}^*(\cdot,\boldsymbol{\gamma}))\}$. It follows immediately that $\mathbf{R}(\boldsymbol{\gamma}_0;\boldsymbol{\beta}^*(\cdot,\boldsymbol{\gamma}_0)) = \mathbf{0}$. Condition C5′ ensures that $\boldsymbol{\gamma}_0$ is the unique zero-crossing of $\mathbf{R}(\boldsymbol{\gamma};\boldsymbol{\beta}^*(\cdot,\boldsymbol{\gamma}))$ in the neighborhood of $\boldsymbol{\gamma}_0$. Thus, $\widehat{\boldsymbol{\gamma}}_I$, which solves $\mathbf{R}_n(\boldsymbol{\gamma};\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma})) = \mathbf{0}$, converges to $\boldsymbol{\gamma}_0$ in probability, as $n\to\infty$.

We prove the uniform consistency of $\widehat{\boldsymbol{\beta}}_I(\tau) \equiv \widehat{\boldsymbol{\beta}}(\tau,\widehat{\boldsymbol{\gamma}}_I)$ over $\tau\in[\nu,\tau_{\max}]$. Note that $\boldsymbol{\beta}_0(\tau) = \boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma}_0)$ and $\|\widehat{\boldsymbol{\beta}}_I(\tau) - \boldsymbol{\beta}_0(\tau)\| \le \|\widehat{\boldsymbol{\beta}}(\tau,\widehat{\boldsymbol{\gamma}}_I) - \boldsymbol{\beta}^*(\tau,\widehat{\boldsymbol{\gamma}}_I)\| + \|\boldsymbol{\beta}^*(\tau,\widehat{\boldsymbol{\gamma}}_I) - \boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma}_0)\|$. For a large n, we have $\sup_{\tau\in[\nu,\tau_{\max}]}\|\widehat{\boldsymbol{\beta}}(\tau,\widehat{\boldsymbol{\gamma}}_I) - \boldsymbol{\beta}^*(\tau,\widehat{\boldsymbol{\gamma}}_I)\| \xrightarrow{P} 0$ from the consistency of $\widehat{\boldsymbol{\gamma}}_I$ and Lemma A.1. Hence, it suffices to show that $\sup_{\tau\in[\nu,\tau_{\max}]}\|\boldsymbol{\beta}^*(\tau,\widehat{\boldsymbol{\gamma}}_I) - \boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma}_0)\| \xrightarrow{P} 0$. By the definition of $\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})$, we have for every τ and $\boldsymbol{\gamma}$,

$$
\begin{aligned}
&E\left[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\widetilde{f}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})\}|\mathbf{Z},\mathbf{W})\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})\}\right]\\
&\quad\times\frac{\partial\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}}\\
&=\int_0^\tau E\left[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\bar{f}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}^*(u,\boldsymbol{\gamma})\}|\mathbf{Z},\mathbf{W})\right.\\
&\quad\left.\times\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}^*(u,\boldsymbol{\gamma})\}\frac{\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})u}\right]\frac{\partial\boldsymbol{\beta}^*(u,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}}\mathrm{d}u\\
&\quad+\int_0^\tau E\left[\mathbf{Z}\mathbf{W}^{\mathrm{T}}\left[\bar{F}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}^*(u,\boldsymbol{\gamma})\}|\mathbf{Z},\mathbf{W})\right.\right.\\
&\quad\times\{2-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})u\}+\bar{F}_C(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}^*(u,\boldsymbol{\gamma})\}|\mathbf{Z},\mathbf{W})\\
&\quad\left.\left.\times\{\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})u-1\}\right]\frac{\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}}{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})u\}^2}\right]\mathrm{d}u.
\end{aligned}
$$

Observing that $\bar{F}_{X,\boldsymbol{\gamma}}(t|\mathbf{Z},\mathbf{W}) = \{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_{T^*}(t|\mathbf{Z})\}\bar{F}_C(t|\mathbf{Z},\mathbf{W})$, we have

$$
\begin{aligned}
&\frac{\partial\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\\
&=\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}\int_0^\tau \mathbf{D}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\left\{\frac{\partial\boldsymbol{\beta}^*(u,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\right\}\mathrm{d}u\\
&\quad+\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}\int_0^\tau \mathbf{C}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\mathrm{d}u,
\end{aligned}\tag{B.1}
$$

where

$$
\begin{aligned}
&\mathbf{B}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})\\
&\quad=E\left[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\widetilde{f}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}\right],\\
&\mathbf{C}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})\\
&\quad=E\left[\mathbf{Z}\mathbf{W}^{\mathrm{T}}\bar{F}_C(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\right.\\
&\quad\left.\times\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}\right],\\
&\mathbf{D}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})\\
&\quad=E\left[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\bar{f}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}\right.\\
&\quad\left.\times\frac{\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\tau}\right].
\end{aligned}
$$

Note that (B.1) is the Volterra integral equation of the second type and one solution of (B.1) can be expressed as

$$
\begin{aligned}
&\frac{\partial\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\\
&=\int_0^\tau \mathbf{A}(\tau,u)\mathbf{B}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)^{-1}\int_0^u \mathbf{C}(\boldsymbol{\beta}_0(s),\boldsymbol{\gamma}_0)\mathrm{d}s\,\mathrm{d}u\\
&\quad+\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}\int_0^\tau \mathbf{C}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\mathrm{d}u,
\end{aligned}
$$

where

$$
\begin{aligned}
&\mathbf{A}(\tau,u)\\
&\quad=\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}\int_0^u \mathbf{C}(\boldsymbol{\beta}_0(s),\boldsymbol{\gamma}_0)\mathrm{d}s\\
&\quad\times\exp\left\{\int_u^\tau \mathbf{B}(\boldsymbol{\beta}_0(s),\boldsymbol{\gamma}_0)^{-1}\int_0^s \mathbf{C}(\boldsymbol{\beta}_0(v),\boldsymbol{\gamma}_0)\mathrm{d}v\mathrm{d}s\right\}.
\end{aligned}
$$

Under conditions C2–C4, we have that $\partial\boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^{\mathrm{T}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}$ is bounded uniformly in $\tau\in[\nu,\tau_{\max}]$. Together with the consistency of $\widehat{\boldsymbol{\gamma}}_I$, and via the Taylor expansion, we obtain that $\sup_{\tau\in[\nu,\tau_{\max}]}\|\boldsymbol{\beta}^*(\tau,\widehat{\boldsymbol{\gamma}}_I) - \boldsymbol{\beta}^*(\tau,\boldsymbol{\gamma}_0)\| \xrightarrow{P} 0$, which completes the proof of the uniform consistency that $\sup_{\tau\in[\nu,\tau_{\max}]}\|\widehat{\boldsymbol{\beta}}_I(\tau) - \boldsymbol{\beta}_0(\tau)\| \xrightarrow{P} 0$.

Proof of Theorem 2.

We first establish the weak convergence of $\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}_0)$. Note that $\partial\boldsymbol{\mu}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})/\partial\boldsymbol{\beta}(\tau)^{\mathrm{T}} = \mathbf{B}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})$ and $\partial\widetilde{\boldsymbol{\mu}}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})/\partial\boldsymbol{\beta}(\tau)^{\mathrm{T}} = \mathbf{D}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})$. Let $o_{\mathcal{I}}(1)$ denote a term that converges uniformly in $\tau\in\mathcal{I}$ to zero in probability. Because $\mathbf{U}_n(\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}),\tau_j;\boldsymbol{\gamma}) = \mathbf{0}$ for $j = 1,\ldots,q_n$,

$$
\begin{aligned}
&\sup_{\tau\in[\tau_j,\tau_{j+1}],\boldsymbol{\gamma}\in\mathcal{K}} n^{1/2}\|\mathbf{U}_n(\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}),\tau;\boldsymbol{\gamma})\|\\
&\quad=\sup_{\tau\in[\tau_j,\tau_{j+1}],\boldsymbol{\gamma}\in\mathcal{K}} n^{1/2}\|\mathbf{U}_n(\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}),\tau;\boldsymbol{\gamma})\\
&\quad\quad-\mathbf{U}_n(\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}),\tau_j;\boldsymbol{\gamma})\|\\
&\quad=\sup_{\tau\in[\tau_j,\tau_{j+1}],\boldsymbol{\gamma}\in\mathcal{K}} n^{1/2}\left\|\int_{\tau_j}^\tau n^{-1}\sum_{i=1}^n \mathbf{Z}_i I[X_i\right.\\
&\quad\quad\left.\ge\exp\{\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(u,\boldsymbol{\gamma})\}]H_{\boldsymbol{\gamma}}(\mathrm{d}u|\mathbf{W}_i)\right\|\\
&\quad\le n^{1/2}a_nc_0\to 0,
\end{aligned}\tag{B.2}
$$

under $n^{1/2}a_n\to 0$, where c_0 is a constant such that $\sup_i\|\mathbf{Z}_i\|\sup_{i,\boldsymbol{\gamma}}\exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i) < c_0$. Lemma A.1 and condition C2 immediately imply that $\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0),\boldsymbol{\gamma}_0)$ and $\widetilde{\boldsymbol{\mu}}(\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0),\boldsymbol{\gamma}_0)$ converge uniformly to $\boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)$ and $\widetilde{\boldsymbol{\mu}}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)$, respectively. By (A.1), (A.2), and (B.2), we have

$$
\begin{aligned}
&-n^{1/2}\mathbf{U}_n(\boldsymbol{\beta}_0,\tau;\boldsymbol{\gamma}_0)\\
&\quad=-\int_0^\tau n^{1/2}\{\widetilde{\boldsymbol{\mu}}(\widehat{\boldsymbol{\beta}}(u,\boldsymbol{\gamma}_0),\boldsymbol{\gamma}_0)-\widetilde{\boldsymbol{\mu}}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\}\mathrm{d}u\\
&\quad\quad+n^{1/2}\{\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0),\boldsymbol{\gamma}_0)-\boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)\}+o_{[\nu,\tau_{\max}]}(1)\\
&\quad=-\int_0^\tau \mathbf{D}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\mathbf{B}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)^{-1}n^{1/2}\{\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}(u,\boldsymbol{\gamma}_0),\boldsymbol{\gamma}_0)\\
&\quad\quad-\boldsymbol{\mu}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\}\mathrm{d}u+n^{1/2}\{\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0),\boldsymbol{\gamma}_0)\\
&\quad\quad-\boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)\}+o_{[\nu,\tau_{\max}]}(1).
\end{aligned}
$$

It follows from the production integration theory (Gill and Johansen 1990; Andersen et al. 1993) that

$$
\begin{aligned}
&n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0)-\boldsymbol{\beta}_0(\tau)\}\\
&\quad=\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}\phi\left\{-n^{1/2}\mathbf{U}_n(\boldsymbol{\beta}_0,\tau;\boldsymbol{\gamma}_0)\right\}+o_{[\nu,\tau_{\max}]}(1)\\
&\quad=\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}n^{-1/2}\sum_{i=1}^n\phi\{-\xi_i(\boldsymbol{\beta}_0(\cdot),\tau;\boldsymbol{\gamma}_0)\}\\
&\quad\quad+o_{[\nu,\tau_{\max}]}(1),
\end{aligned}\tag{B.3}
$$

where ϕ is a linear map from $\mathcal{F}$ to $\mathcal{F}$ such that for $g\in\mathcal{F}$,

$$
\phi(g)(\cdot)=\int_0^{\cdot}\mathcal{I}(s,\cdot)g(\mathrm{d}s),
$$

$\mathcal{I}(s,t)=\prod_{u\in(s,t]}[\mathbf{I}_{p+1}+\mathbf{D}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\mathbf{B}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)^{-1}\mathrm{d}u]$, $\mathbf{I}_{p+1}$ is the identity matrix of size $p+1$, $\mathcal{F}=\{g\colon[0,\tau_{\max}]\mapsto\mathbb{R}^{p+1},\ g$ is a left-continuous function with right limit and $g(0)=\mathbf{0}\}$, and

$$\xi_i(\boldsymbol{\beta}(\cdot),\tau;\boldsymbol{\gamma})=\mathbf{Z}_i\Bigg[N_i(\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(\tau)\})-\int_0^\tau I[X_i\ge\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\}]H_{\boldsymbol{\gamma}}(\mathrm{d}u|\mathbf{W}_i)\Bigg].$$

Under conditions C1 and C2, $\{\xi(\boldsymbol{\beta}_0(\cdot),\tau;\boldsymbol{\gamma}_0):\tau\in[\nu,\tau_{\max}]\}$ is a Donsker class and thus $-n^{1/2}\mathbf{U}_n(\boldsymbol{\beta}_0,\tau;\boldsymbol{\gamma}_0)$ converges weakly to a zero-mean Gaussian process. Since ϕ is a linear operator, the limiting process of $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0)-\boldsymbol{\beta}_0(\tau)\}$ is also a zero-mean Gaussian process. The continuous mapping theorem (van der Vaart and Wellner 1996) implies that $n^{1/2}\sup_{\tau\in[\nu,\tau_{\max}]}\|\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0)-\boldsymbol{\beta}_0(\tau)\|=O_P(1)$.

We prove the asymptotic normality of $\widehat{\boldsymbol{\gamma}}_I$. Via integration by parts, we have

$$\begin{aligned}\mathbf{R}_n(\boldsymbol{\gamma};\boldsymbol{\beta}(\cdot))&=n^{-1}\sum_{i=1}^n\Bigg[N_i(\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(\tau_{\max})\})\frac{\mathbf{W}_i\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\tau_{\max}}\\&\quad-\int_0^{\tau_{\max}}\{N_i(\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\})+I[X_i\ge\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\}]\}\\&\quad\times\frac{\mathbf{W}_i\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)}{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)u\}^2}\mathrm{d}u\Bigg].\end{aligned}$$

Using the similar arguments in the proofs of Lemma A.2, the consistency of $\widehat{\boldsymbol{\gamma}}_I$, the uniform consistency of $\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}_0)$ over $\tau\in[\nu,\tau_{\max}]$, and the Taylor expansion, we have

$$\begin{aligned}n^{1/2}\mathbf{R}_n(\boldsymbol{\gamma}_0;\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}_0))&=n^{1/2}\mathbf{R}_n(\boldsymbol{\gamma}_0;\boldsymbol{\beta}_0(\cdot))+\widetilde{\mathbf{B}}(\boldsymbol{\beta}_0(\tau_{\max}),\boldsymbol{\gamma}_0)n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau_{\max},\boldsymbol{\gamma}_0)\\&\quad-\boldsymbol{\beta}_0(\tau_{\max})\}-\int_0^{\tau_{\max}}\widetilde{\mathbf{D}}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)n^{1/2}\{\widehat{\boldsymbol{\beta}}(u,\boldsymbol{\gamma}_0)\\&\quad-\boldsymbol{\beta}_0(u)\}\mathrm{d}u+o_{[\nu,\tau_{\max}]}(1),\end{aligned}\tag{B.4}$$

where

$$\begin{aligned}\widetilde{\mathbf{B}}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})&=E\Bigg[\mathbf{W}\mathbf{Z}^{\mathrm{T}}\widetilde{f}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}\\&\quad\times\frac{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\tau}\Bigg],\\\widetilde{\mathbf{D}}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})&=E\Bigg[\mathbf{W}\mathbf{Z}^{\mathrm{T}}[\widetilde{f}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})\\&\quad+\bar{f}_{X,\boldsymbol{\gamma}}(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}|\mathbf{Z},\mathbf{W})]\\&\quad\times\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau)\}\frac{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})}{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\tau\}^2}\Bigg].\end{aligned}$$

If we plug (B.3) into Equation (B.4), we have

$$n^{1/2}\mathbf{R}_n(\boldsymbol{\gamma}_0;\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}_0))=n^{-1/2}\sum_{i=1}^n\Phi_i(\boldsymbol{\beta}_0(\cdot),\tau_{\max};\boldsymbol{\gamma}_0)+o_{[\nu,\tau_{\max}]}(1),\tag{B.5}$$

where

$$\begin{aligned}\Phi_i(\boldsymbol{\beta}(\cdot),\tau;\boldsymbol{\gamma})&=\int_0^\tau\frac{\mathbf{W}_i\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)u}\Bigg[\mathrm{d}N_i(\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\})\\&\quad-I[X_i\ge\exp\{\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\}]H_{\boldsymbol{\gamma}}(\mathrm{d}u|\mathbf{W}_i)\Bigg]\\&\quad+\widetilde{\mathbf{B}}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})\mathbf{B}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})^{-1}\phi\{-\xi_i(\boldsymbol{\beta}(\cdot),\tau;\boldsymbol{\gamma})\}\\&\quad-\int_0^\tau\widetilde{\mathbf{D}}(\boldsymbol{\beta}(u),\boldsymbol{\gamma})\mathbf{B}(\boldsymbol{\beta}(u),\boldsymbol{\gamma})^{-1}\phi\{-\xi_i(\boldsymbol{\beta}(\cdot),u;\boldsymbol{\gamma})\}\mathrm{d}u.\end{aligned}$$

Using the definition of $\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma})$, we also have that

$$\begin{aligned}\frac{\partial\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}&=\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}\int_0^\tau\mathbf{D}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\left\{\frac{\partial\widehat{\boldsymbol{\beta}}(u,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\right\}\mathrm{d}u\\&\quad+\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}\int_0^\tau\mathbf{C}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\mathrm{d}u+o_{[\nu,\tau_{\max}]}(1).\end{aligned}$$

In view of this integral equation, its solution takes one form as

$$\frac{\partial\widehat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}=\mathbf{B}(\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0)^{-1}\psi\left\{\int_0^\tau\mathbf{C}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\mathrm{d}u\right\}+o_{[\nu,\tau_{\max}]}(1),\tag{B.6}$$

where ψ is a linear map from $\mathcal{G}$ to $\mathcal{G}$ such that for $g\in\mathcal{G}$,

$$\psi(g)(\cdot)=\int_0^\cdot\mathcal{J}(s,\cdot)g(\mathrm{d}s),\qquad\mathcal{J}(s,t)=\prod_{u\in(s,t]}[\mathbf{I}_{p+1}+\mathbf{D}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\mathrm{d}u],$$

$\mathcal{G}=\{g:[0,\tau_{\max}]\mapsto\mathbb{R}_{(p+1)\times(q+1)},\ g$ is a left-continuous function with right limit and $g(0)=\mathbf{0}_{(p+1)\times(q+1)}\}$, and $\mathbb{R}_{(p+1)\times(q+1)}$ denotes the set of all matrices with $p+1$ rows and $q+1$ columns.

Using the chain rule and (B.6), we have

$$\begin{aligned}\frac{\partial\mathbf{R}(\boldsymbol{\gamma};\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}))}{\partial\boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}&=\widetilde{\mathbf{B}}(\boldsymbol{\beta}_0(\tau_{\max}),\boldsymbol{\gamma}_0)\frac{\partial\widehat{\boldsymbol{\beta}}(\tau_{\max},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\\&\quad-\int_0^{\tau_{\max}}\widetilde{\mathbf{C}}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\mathrm{d}u-\int_0^{\tau_{\max}}\widetilde{\mathbf{D}}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0)\\&\quad\times\left\{\frac{\partial\widehat{\boldsymbol{\beta}}(u,\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\right\}\mathrm{d}u+o_{[\nu,\tau_{\max}]}(1)\\&=\Psi(\boldsymbol{\beta}_0(\cdot),\tau_{\max};\boldsymbol{\gamma}_0)+o_{[\nu,\tau_{\max}]}(1),\end{aligned}\tag{B.7}$$

where

$$\begin{aligned}\widetilde{\mathbf{C}}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})&=E\Bigg[\left(\frac{\mathbf{W}\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})u}\right)^{\otimes2}\\&\quad\times\bar{F}_C(\exp\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(u)\}|\mathbf{Z},\mathbf{W})\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\Bigg],\\\Psi(\boldsymbol{\beta}(\cdot),\tau;\boldsymbol{\gamma})&=\widetilde{\mathbf{B}}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})\mathbf{B}(\boldsymbol{\beta}(\tau),\boldsymbol{\gamma})^{-1}\psi\left\{\int_0^\tau\mathbf{C}(\boldsymbol{\beta}(u),\boldsymbol{\gamma})\mathrm{d}u\right\}\\&\quad-\int_0^\tau\widetilde{\mathbf{C}}(\boldsymbol{\beta}(u),\boldsymbol{\gamma})\mathrm{d}u-\int_0^\tau\widetilde{\mathbf{D}}(\boldsymbol{\beta}(u),\boldsymbol{\gamma})\mathbf{B}(\boldsymbol{\beta}(u),\boldsymbol{\gamma})^{-1}\\&\quad\times\psi\left\{\int_0^u\mathbf{C}(\boldsymbol{\beta}(s),\boldsymbol{\gamma})\mathrm{d}s\right\}\mathrm{d}u,\end{aligned}$$

and $\mathbf{a}^{\otimes2}=\mathbf{a}\mathbf{a}^{\mathrm{T}}$ for a column vector $\mathbf{a}$.

It follows from the consistency of $\widehat{\boldsymbol{\gamma}}_I$, (B.5), (B.7), condition C5′, and the Taylor expansion that

$$\begin{aligned}n^{1/2}(\widehat{\boldsymbol{\gamma}}_I-\boldsymbol{\gamma}_0)&=-\left\{\frac{\partial\mathbf{R}(\boldsymbol{\gamma};\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}))}{\partial\boldsymbol{\gamma}}\Big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}\right\}^{-1}n^{1/2}\mathbf{R}_n(\boldsymbol{\gamma}_0;\widehat{\boldsymbol{\beta}}(\cdot,\boldsymbol{\gamma}_0))\\&\quad+o_{[\nu,\tau_{\max}]}(1)\\&=-n^{-1/2}\sum_{i=1}^n\Psi(\boldsymbol{\beta}_0(\cdot),\tau_{\max};\boldsymbol{\gamma}_0)^{-1}\Phi_i(\boldsymbol{\beta}_0(\cdot),\tau_{\max};\boldsymbol{\gamma}_0)\\&\quad+o_{[\nu,\tau_{\max}]}(1),\end{aligned}\tag{B.8}$$

which is approximated by the scaled summation of iid random vectors. It can then be shown to converge in distribution to a Gaussian random vector with mean zero and variance-covariance matrix $\boldsymbol{\Omega}_I^{-1}\mathbf{V}_I\boldsymbol{\Omega}_I^{-1}$, where

$$\boldsymbol{\Omega}_I = \Psi(\boldsymbol{\beta}_0(\cdot), \tau_{\max}; \boldsymbol{\gamma}_0) \text{ and } \mathbf{V}_I = E\{\Phi_i(\boldsymbol{\beta}_0(\cdot), \tau_{\max}; \boldsymbol{\gamma}_0)\}^{\otimes 2}.$$

We establish the weak convergence of $\widehat{\boldsymbol{\beta}}_I(\cdot) \equiv \widehat{\boldsymbol{\beta}}(\cdot, \widehat{\boldsymbol{\gamma}}_I)$. It follows from (B.3), (B.6), and (B.8) that

$$\begin{aligned} n^{1/2}&\{\widehat{\boldsymbol{\beta}}_I(\tau) - \boldsymbol{\beta}_0(\tau)\} \\ &= n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau, \boldsymbol{\gamma}_0) - \boldsymbol{\beta}_0(\tau)\} + n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau, \widehat{\boldsymbol{\gamma}}_I) - \widehat{\boldsymbol{\beta}}(\tau, \boldsymbol{\gamma}_0)\} \\ &= n^{-1/2}\sum_{i=1}^{n}\{\boldsymbol{\zeta}_i^{(1)}(\tau) + \boldsymbol{\zeta}_i^{(2)}(\tau)\} + o_{[\nu,\tau_{\max}]}(1), \end{aligned} \tag{B.9}$$

where

$$\begin{aligned} \boldsymbol{\zeta}_i^{(1)}(\tau) &= \mathbf{B}(\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}_0)^{-1}\phi\{-\xi_i(\boldsymbol{\beta}_0(\cdot), \tau; \boldsymbol{\gamma}_0)\}, \\ \boldsymbol{\zeta}_i^{(2)}(\tau) &= -\mathbf{B}(\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}_0)^{-1}\psi\left\{\int_0^{\tau}\mathbf{C}(\boldsymbol{\beta}_0(u), \boldsymbol{\gamma}_0)\mathrm{d}u\right\} \\ &\qquad \times \boldsymbol{\Omega}_I^{-1}\Phi_i(\boldsymbol{\beta}_0(\cdot), \tau_{\max}; \boldsymbol{\gamma}_0). \end{aligned}$$

The class of $\{\phi\{-\xi(\boldsymbol{\beta}_0(\cdot), \tau; \boldsymbol{\gamma}_0)\}$: $\tau \in [\nu, \tau_{\max}]\}$ is Donsker, and the class of $\{\boldsymbol{\zeta}^{(1)}(\tau)$: $\tau \in [\nu, \tau_{\max}]\}$ is also Donsker under condition C3 by using the preservation property of Donsker class (van der Vaart and Wellner 1996). Similarly, the Lipschitz property of function $\int_0^{\cdot}\mathbf{C}(\boldsymbol{\beta}_0(u), \boldsymbol{\gamma}_0)\mathrm{d}u$ from $[\nu, \tau_{\max}]$ to $\mathbb{R}_{(p+1)\times(q+1)}$, the linearity of operator ψ, and together with condition C3 yield that $\{\boldsymbol{\zeta}^{(2)}(\tau)$: $\tau \in [\nu, \tau_{\max}]\}$ is a Donsker class. It follows from the preservation property of Donsker class again that $\{\boldsymbol{\zeta}^{(1)}(\tau) + \boldsymbol{\zeta}^{(2)}(\tau)$: $\tau \in [\nu, \tau_{\max}]\}$ is a Donsker class. Hence, $n^{-1/2}\sum_{i=1}^{n}\{\boldsymbol{\zeta}_i^{(1)}(\tau) + \boldsymbol{\zeta}_i^{(2)}(\tau)\}$ converges weakly to a Gaussian process over $\tau \in [\nu, \tau_{\max}]$ with mean zero and variance-covariance matrix function $\boldsymbol{\Sigma}_I(\tau, \tau') = E[\{\boldsymbol{\zeta}^{(1)}(\tau) + \boldsymbol{\zeta}^{(2)}(\tau)\}\{\boldsymbol{\zeta}^{(1)}(\tau') + \boldsymbol{\zeta}^{(2)}(\tau')\}^{\mathrm{T}}]$ for $\tau, \tau' \in [\nu, \tau_{\max}]$. The weak convergence of $n^{1/2}\{\widehat{\boldsymbol{\beta}}_I(\tau) - \boldsymbol{\beta}_0(\tau)\}$ over $\tau \in [\nu, \tau_{\max}]$ follows immediately using equation (B.9), which completes the proof of Theorem 2.

APPENDIX C: PROOFS OF THEOREMS 3 AND 4

Proof of Theorem 3.

We first prove the consistency of $\widehat{\boldsymbol{\gamma}}_N$ by verifying each conditions in Theorem 1 by Chen, Linton, and Van Keilegom (2003). Rewrite $\mathbf{S}_n(\boldsymbol{\gamma}, F_{T^*}) = n^{-1}\sum_{i=1}^{n}\mathbf{s}_i(\boldsymbol{\gamma}, F_{T^*})$, where for $i = 1, \ldots, n$,

$$\begin{aligned} \mathbf{s}_i&(\boldsymbol{\gamma}, F_{T^*}) \\ &= \frac{\mathbf{W}_i\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_{T^*}(X_i|\mathbf{Z}_i)} \\ &\qquad \times \left\{\Delta_i - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_{T^*}(X_i|\mathbf{Z}_i)\right\} \\ &= \int_0^L \frac{\mathbf{W}_i\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_{T^*}(t|\mathbf{Z}_i)}\mathrm{d}M_i(t; \boldsymbol{\gamma}, F_{T^*}). \end{aligned}$$

Let $\mathbf{S}(\boldsymbol{\gamma}, F_{T^*}) = E\{\mathbf{s}_i(\boldsymbol{\gamma}, F_{T^*})\}$. As $M_i(t; \boldsymbol{\gamma}, F_{T^*})$ is a martingale, $\mathbf{S}(\boldsymbol{\gamma}_0, F_{0T^*}) = \mathbf{0}$. Because $\widehat{\boldsymbol{\gamma}}_N$ solves $\mathbf{S}_n(\boldsymbol{\gamma}, \widehat{F}_{T^*}) = \mathbf{0}$, we have $\|\mathbf{S}_n(\widehat{\boldsymbol{\gamma}}_N, \widehat{F}_{T^*})\| = o_P(1)$, which leads to condition (1.1) by Chen, Linton, and Van Keilegom (2003).

For any $\delta > 0$ and $\boldsymbol{\gamma} \in \mathcal{K}$, we have $\inf_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|>\delta}\|\mathbf{S}(\boldsymbol{\gamma}, F_{0T^*})\| = \inf_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|>\delta}\|\Gamma_1(\boldsymbol{\gamma}^*, F_{0T^*})\ (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)\|$, which is positive under condition C5. Here, $\boldsymbol{\gamma}^*$ is between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_0$, and

$$\begin{aligned} \boldsymbol{\Gamma}_1&(\boldsymbol{\gamma}, F_{T^*}) \\ &\equiv \frac{\partial\mathbf{S}(\boldsymbol{\gamma}, F_{T^*})}{\partial\boldsymbol{\gamma}} \\ &= -E\Bigg[\int_0^L \frac{\mathbf{W}^{\otimes 2}\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}\{1 - F_{T^*}(t|\mathbf{Z})\}}{\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_{T^*}(t|\mathbf{Z})\}^2} \\ &\qquad \times \mathrm{d}M(t; \boldsymbol{\gamma}, F_{T^*})\Bigg] \\ &\quad - E\Bigg[\int_0^L \left(\frac{\mathbf{W}\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}}{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_{T^*}(t|\mathbf{Z})}\right)^{\otimes 2} I(X \geq t) \\ &\qquad \times \mathrm{d}\Lambda_{T,\boldsymbol{\gamma}}(t|\mathbf{Z}, \mathbf{W})\Bigg]. \end{aligned} \tag{C.1}$$

Thus, condition (1.2) by Chen, Linton, and Van Keilegom (2003) is verified.

Under conditions C1 and C2, there exists a constant c^* such that

$$\begin{aligned} &\left|\frac{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})}{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})F_{T^*}(t|\mathbf{z})} - \frac{1 - \pi(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{w})}{1 - \pi(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{w})\widetilde{F}_{T^*}(t|\mathbf{z})}\right| \\ &\quad \leq 2\frac{|\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w}) - \pi(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{w})| + |F_{T^*}(t|\mathbf{z}) - \widetilde{F}_{T^*}(t|\mathbf{z})|}{\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})F_{T^*}(t|\mathbf{z})\}\{1 - \pi(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{w})\widetilde{F}_{T^*}(t|\mathbf{z})\}} \\ &\quad \leq 2\frac{\|\mathbf{w}\|\|\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}\| + \sup_{t,\mathbf{z}}|F_{T^*}(t|\mathbf{z}) - \widetilde{F}_{T^*}(t|\mathbf{z})|}{\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})F_{T^*}(t|\mathbf{z})\}\{1 - \pi(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{w})\widetilde{F}_{T^*}(t|\mathbf{z})\}} \\ &\quad \leq c^*\|\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}\| + c^*\sup_{t,\mathbf{z}}|F_{T^*}(t|\mathbf{z}) - \widetilde{F}_{T^*}(t|\mathbf{z})| \end{aligned}$$

for any $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \mathcal{K}$, and $F_{T^*}(\cdot|\mathbf{z})$ and $\widetilde{F}_{T^*}(\cdot|\mathbf{z})$ in the class of nonparametric conditional distribution functions. Thus, under condition C1, every element of the first part of $\mathbf{s}_i(\boldsymbol{\gamma}, F_{T^*})$, $\mathbf{W}_i\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}\Delta_i/\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_{T^*}(X_i|\mathbf{Z}_i)\}$, is a Lipschitz function of $(\boldsymbol{\gamma}, F_{T^*})$. Similar argument can be used to show that every element of the remaining part of $\mathbf{s}_i(\boldsymbol{\gamma}, F_{T^*})$ is also Lipschitz. Thus, condition (1.3) is satisfied immediately. Furthermore, it follows from theorem 2.7.11 by van der Vaart and Wellner (1996) that the class of $\mathbf{s}_i(\boldsymbol{\gamma}, F_{T^*})$ over the parametric space is a Glivenko–Cantelli class. Thus, condition (1.5′) follows directly. In addition, condition (1.4) is also satisfied by Lemma A.3. Therefore, it follows from theorem 1 by Chen, Linton, and Van Keilegom (2003) that $\widehat{\boldsymbol{\gamma}}_N \xrightarrow{P} \boldsymbol{\gamma}_0$.

The uniform consistency of $\widehat{\boldsymbol{\beta}}_N(\tau) \equiv \widehat{\boldsymbol{\beta}}(\tau, \widehat{\boldsymbol{\gamma}}_N)$ over $\tau \in [\nu, \tau_{\max}]$ follows directly from the proof of Theorem 1 by replacing $\widehat{\boldsymbol{\gamma}}_I$ with $\widehat{\boldsymbol{\gamma}}_N$.

Proof of Theorem 4.

We first prove the asymptotic normality of $\widehat{\boldsymbol{\gamma}}_N$ by verifying conditions (2.1)–(2.4), (2.5′), and (2.6) in theorem 2 by Chen, Linton, and Van Keilegom (2003). It is clear that condition (2.1) is satisfied by the definition of $\widehat{\boldsymbol{\gamma}}_N$.

It can be seen from (C.1) that $\boldsymbol{\Gamma}_1(\boldsymbol{\gamma}, F_{0T^*})$ is continuous with respect to $\boldsymbol{\gamma}$ and

$$\begin{aligned} \boldsymbol{\Omega}_N &\equiv \boldsymbol{\Gamma}_1(\boldsymbol{\gamma}_0, F_{0T^*}) \\ &= -E\left(\int_0^L \frac{\mathbf{W}\{1 - \pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W})\}}{1 - \pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W})F_{0T^*}(t|\mathbf{Z})}\mathrm{d}M(t; \boldsymbol{\gamma}_0, F_{0T^*})\right)^{\otimes 2}, \end{aligned}$$

which is negative definite under condition C5. Thus, condition (2.2) is satisfied.

For all $\boldsymbol{\gamma} \in \mathcal{K}$, the functional derivative of $\mathbf{S}(\boldsymbol{\gamma}, F_{T^*})$ at F_{0T^*} along the direction $F_{T^*} - F_{0T^*}$ is

$$
\begin{aligned}
&\boldsymbol{\Gamma}_2(\boldsymbol{\gamma}, F_{0T^*})[F_{T^*} - F_{0T^*}] \\
&= E\Bigg[\int_0^L \mathbf{W}\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\} \\
&\quad\times\Bigg\{\frac{F_{T^*}(u|\mathbf{Z}) - F_{0T^*}(u|\mathbf{Z})}{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_{0T^*}(u|\mathbf{Z})\}^2}\mathrm{d}N(u) \\
&\quad - I(X \geq u)\mathrm{d}\left(\frac{F_{T^*}(u|\mathbf{Z}) - F_{0T^*}(u|\mathbf{Z})}{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_{0T^*}(u|\mathbf{Z})\}^2}\right)\Bigg\}\Bigg] \\
&= \int_{\mathbb{R}^p}\int_0^L \mathbf{w}\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})\} \\
&\quad\times\Bigg[\frac{F_{T^*}(u|\mathbf{z}) - F_{0T^*}(u|\mathbf{z})}{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})F_{0T^*}(u|\mathbf{z})\}^2}E\{\mathrm{d}N(u)|\mathbf{Z}=\mathbf{z}\} \\
&\quad - E\{I(X\geq u)|\mathbf{Z}=\mathbf{z}\}\mathrm{d}\left(\frac{F_{T^*}(u|\mathbf{z}) - F_{0T^*}(u|\mathbf{z})}{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})F_{0T^*}(u|\mathbf{z})\}^2}\right)\Bigg] \\
&\quad\times f_{\mathbf{Z}}(\mathbf{z})\mathrm{d}\mathbf{z}.
\end{aligned} \tag{C.2}
$$

It follows from the Taylor expansion that condition (2.3) holds under conditions C1 and C5 and Lemma A.1. Condition (2.4) is also satisfied directly from Lemma A.3.

By the Lipschitz continuity of $\mathbf{s}_i(\boldsymbol{\gamma}, F_{T^*})$ as shown in the proof of Theorem 3, $\mathbf{s}_i(\boldsymbol{\gamma}, F_{T^*})$ satisfies the Hölder continuity condition (3.1) in theorem 3 by Chen, Linton, and Van Keilegom (2003). Obviously, condition (3.2) is satisfied due to the continuity of $\mathbf{s}_i(\boldsymbol{\gamma}, F_{T^*})$ and condition (3.3) is satisfied by remark 3 (ii) of their article. Therefore, condition (2.5′) holds by applying theorem 3 by Chen, Linton, and Van Keilegom (2003).

It follows from theorem 2.3 by Liang, Una-Alvarez, and Iglesias-Perez (2012) that under conditions C1, C2, C5–C7, and C8′ that

$$
\begin{aligned}
&\widehat{F}_{T^*}(t|\mathbf{z}) - F_{0T^*}(t|\mathbf{z}) \\
&= \frac{1}{nh_n^p f_{\mathbf{Z}}(\mathbf{z})}\sum_{i=1}^n K_p\left(\frac{\mathbf{z}-\mathbf{Z}_i}{h_n}\right)\xi(X_i, \Delta_i, t, \mathbf{z}, \mathbf{w}) \\
&\quad + O_P(\alpha_n),
\end{aligned} \tag{C.3}
$$

where

$$
\begin{aligned}
&\xi(X_i, \Delta_i, t, \mathbf{z}, \mathbf{w}) \\
&= \{1-F_{0T^*}(t|\mathbf{z})\}\left[\int_0^t \frac{1}{\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{w})\{1-F_{0T^*}(u|\mathbf{z})\}\bar{F}_C(u|\mathbf{z},\mathbf{w})}\right. \\
&\quad\left.\times\left\{\mathrm{d}N_i(u) - \frac{I(X_i\geq u)\omega_{0i}\mathrm{d}F_{0T^*}(u|\mathbf{z})}{1-F_{0T^*}(u|\mathbf{z})}\right\}\right]
\end{aligned}
$$

and $\alpha_n = h_n^\ell + \{\log n/(nh_n^p)\}^{3/4}$. Note that for $i = 1, \ldots, n$,

$$
\begin{aligned}
&E\left\{K_p\left(\frac{\mathbf{z}-\mathbf{Z}_i}{h_n}\right)\xi(X_i, \Delta_i, t, \mathbf{z}, \mathbf{w})\right\} \\
&= \int_{\mathbb{R}^p} E\left\{K_p\left(\frac{\mathbf{z}-\mathbf{Z}_i}{h_n}\right)\xi(X_i, \Delta_i, t, \mathbf{z}, \mathbf{w})|\mathbf{Z}_i = \mathbf{x}\right\} \\
&\quad\times f_{\mathbf{Z}}(\mathbf{x})\mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^p} K_p\left(\frac{\mathbf{z}-\mathbf{x}}{h_n}\right)\int_0^t \frac{1-F_{0T^*}(t|\mathbf{z})}{\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{w})\{1-F_{0T^*}(u|\mathbf{z})\}\bar{F}_C(u|\mathbf{z},\mathbf{w})} \\
&\quad\times E\left\{\mathrm{d}N_i(u) - \frac{I(X_i\geq u)\omega_{0i}\mathrm{d}F_{0T^*}(u|\mathbf{z})}{1-F_{0T^*}(u|\mathbf{z})}\middle|\mathbf{Z}_i = \mathbf{x}\right\} \\
&\quad\times f_{\mathbf{Z}}(\mathbf{x})\mathrm{d}\mathbf{x} \\
&= 0,
\end{aligned}
$$

where $\omega_{0i} = \omega_i(\boldsymbol{\gamma}_0, \Lambda_{0T^*})$. Plugging (C.3) into (C.2) and then using the standard change of variables and the Taylor expansion argument, we obtain that

$$
\begin{aligned}
&\boldsymbol{\Gamma}_2(\boldsymbol{\gamma}_0, F_{0T^*})[\widehat{F}_{T^*} - F_{0T^*}] \\
&= \int_{\mathbb{R}^p}\int_0^L \mathbf{w}\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{w})\{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{w})\} \\
&\quad\times\left[\frac{\widehat{F}_{T^*}(u|\mathbf{z}) - F_{0T^*}(u|\mathbf{z})}{\{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{w})F_{0T^*}(u|\mathbf{z})\}^2}E\{\mathrm{d}N(u)|\mathbf{Z}=\mathbf{z}\}\right. \\
&\quad\left. - E\{I(X\geq u)|\mathbf{Z}=\mathbf{z}\}\mathrm{d}\left(\frac{\widehat{F}_{T^*}(u|\mathbf{z}) - F_{0T^*}(u|\mathbf{z})}{\{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{w})F_{0T^*}(u|\mathbf{z})\}^2}\right)\right] \\
&\quad\times f_{\mathbf{Z}}(\mathbf{z})\mathrm{d}\mathbf{z} \\
&= n^{-1}\sum_{i=1}^n \boldsymbol{\eta}(X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i) + O_P(\alpha_n),
\end{aligned}
$$

where

$$
\begin{aligned}
&\boldsymbol{\eta}(X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i) \\
&= \int_0^L \mathbf{W}_i\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}_i)\{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}_i)\} \\
&\quad\times\left[\frac{\xi(X_i, \Delta_i, u, \mathbf{Z}_i, \mathbf{W}_i)\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}_i)\bar{F}_C(u|\mathbf{Z}_i, \mathbf{W}_i)}{\{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}_i)F_{0T^*}(u|\mathbf{Z}_i)\}^2}\right. \\
&\quad\times \mathrm{d}F_{0T^*}(u|\mathbf{Z}_i) \\
&\quad - \{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}_i)F_{0T^*}(u|\mathbf{Z}_i)\}\bar{F}_C(u|\mathbf{Z}_i, \mathbf{W}_i) \\
&\quad\left.\times\ \mathrm{d}\left(\frac{\xi(X_i, \Delta_i, u, \mathbf{Z}_i, \mathbf{W}_i)}{\{1-\pi(\boldsymbol{\gamma}_0^{\mathrm{T}}\mathbf{W}_i)F_{0T^*}(u|\mathbf{Z}_i)\}^2}\right)\right]
\end{aligned}
$$

is a random vector with mean zero due to $E\{\mathrm{d}\xi(X_i, \Delta_i, u, \mathbf{Z}_i, \mathbf{W}_i)|\mathbf{Z}_i, \mathbf{W}_i\} = 0$. We further note that $E\|\boldsymbol{\eta}(X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i)\|^2 < \infty$ under conditions C1 and C2 and $O_P(\alpha_n)n^{1/2} = o_P(1)$ under condition C8′. Hence, we have

$$
\begin{aligned}
&n^{1/2}\{\mathbf{S}_n(\boldsymbol{\gamma}_0, F_{0T^*}) + \Gamma_2(\boldsymbol{\gamma}_0, F_{0T^*})[\widehat{F}_{T^*} - F_{0T^*}]\} \\
&= n^{-1/2}\sum_{i=1}^n\{\mathbf{s}_i(\boldsymbol{\gamma}_0, F_{0T^*}) + \boldsymbol{\eta}(X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i)\} \\
&\quad + o_P(1),
\end{aligned}
$$

which converges in distribution to a zero-mean normal random vector with variance-covariance matrix

$$
\mathbf{V}_N \equiv E\{\mathbf{s}_1(\boldsymbol{\gamma}_0, F_{0T^*}) + \boldsymbol{\eta}(X_1, \Delta_1, \mathbf{Z}_1, \mathbf{W}_1)\}^{\otimes 2}.
$$

Thus, condition (2.6) is verified. Therefore, it follows from theorem 2 by Chen, Linton, and Van Keilegom (2003) that $n^{1/2}(\widehat{\boldsymbol{\gamma}}_N - \boldsymbol{\gamma}_0)$ converges in distribution to a normal random vector with mean zero and variance-covariance matrix $\boldsymbol{\Omega}_N^{-1}\mathbf{V}_N\boldsymbol{\Omega}_N^{-1}$. In other words, we have

$$
\begin{aligned}
&n^{1/2}(\widehat{\boldsymbol{\gamma}}_N - \boldsymbol{\gamma}_0) \\
&= -n^{-1/2}\sum_{i=1}^n \Omega_N^{-1}\{\mathbf{s}_i(\boldsymbol{\gamma}_0, F_{0T^*}) + \boldsymbol{\eta}(X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i)\} \\
&\quad + o_P(1).
\end{aligned} \tag{C.4}
$$

We prove the weak convergence of $\widehat{\boldsymbol{\beta}}_N(\cdot) \equiv \widehat{\boldsymbol{\beta}}(\cdot, \widehat{\boldsymbol{\gamma}}_N)$. It follows from (B.3), (B.6), and (C.4) that

$$
\begin{aligned}
&n^{1/2}\{\widehat{\boldsymbol{\beta}}_N(\tau) - \boldsymbol{\beta}_0(\tau)\} \\
&= n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau, \boldsymbol{\gamma}_0) - \boldsymbol{\beta}_0(\tau)\} + n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau, \widehat{\boldsymbol{\gamma}}_N) - \widehat{\boldsymbol{\beta}}(\tau, \boldsymbol{\gamma}_0)\} \\
&= n^{-1/2}\sum_{i=1}^n\{\boldsymbol{\zeta}_i^{(1)}(\tau) + \boldsymbol{\zeta}_i^{(3)}(\tau)\} + o_{[\nu, \tau_{\max}]}(1),
\end{aligned} \tag{C.5}
$$

where

$$
\begin{aligned}
\boldsymbol{\zeta}_i^{(3)}(\tau) &= -\mathbf{B}(\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}_0)^{-1}\psi\left\{\int_0^\tau \mathbf{C}(\boldsymbol{\beta}_0(u), \boldsymbol{\gamma}_0)\mathrm{d}u\right\} \\
&\quad\times \boldsymbol{\Omega}_N^{-1}\{\mathbf{s}_i(\boldsymbol{\gamma}_0, F_{0T^*}) + \boldsymbol{\eta}(X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i)\}.
\end{aligned}
$$

Likewise, we have that $n^{-1/2}\sum_{i=1}^n\{\boldsymbol{\zeta}_i^{(1)}(\tau) + \boldsymbol{\zeta}_i^{(3)}(\tau)\}$ converges weakly to a Gaussian process over $\tau \in [\nu, \tau_{\max}]$ with mean zero and variance-covariance matrix function $\boldsymbol{\Sigma}_N(\tau, \tau') = E[\{\boldsymbol{\zeta}^{(1)}(\tau) + \boldsymbol{\zeta}^{(3)}(\tau)\}\{\boldsymbol{\zeta}^{(1)}(\tau') + \boldsymbol{\zeta}^{(3)}(\tau')\}^{\mathrm{T}}]$ for $\tau, \tau' \in [\nu, \tau_{\max}]$. The weak convergence of $n^{1/2}\{\widehat{\boldsymbol{\beta}}_N(\tau) - \boldsymbol{\beta}_0(\tau)\}$ over $\tau \in [\nu, \tau_{\max}]$ follows immediately using equation (C.5), which completes the proof of Theorem 4.

SUPPLEMENTARY MATERIALS

The supplementary materials contain additional simulation results and the proofs of the lemmas.

[Received September 2011. Revised August 2013.]

REFERENCES

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer. [1527]

Bang, H., and Tsiatis, A. A. (2002), "Median Regression With Censored Cost Data," *Biometrics*, 58, 643–649. [1517]

Barlow, W. E., and Prentice, R. L. (1988), "Residuals for Relative Risk Regression," *Biometrika*, 75, 65–74. [1525]

Berkson, J., and Gage, R. P. (1952), "Survival Curve for Cancer Patients Following Treatment," *Journal of the American Statistical Association*, 47, 501–515. [1517]

Bilias, Y., Chen, S., and Ying, Z. (2000), "Simple Resampling Methods for Censored Regression Quantiles," *Journal of Econometrics*, 99, 373–386. [1518]

Buchinsky, M. (1995), "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study," *Journal of Econometrics*, 68, 303–338. [1518]

Buchinsky, M., and Hahn, J. Y. (1998), "An Alternative Estimator for the Censored Quantile Regression," *Econometrica*, 66, 653–671. [1518]

Chen, X., Linton, O., and Van Keilegom, I. (2003), "Estimation of Semiparametric Models When the Criterion Function is Not Smooth," *Econometrica*, 71, 1591–1608. [1521,1529,1530]

Chen, Y. Q., and Wang, M. C. (2000), "Analysis of Accelerated Hazards Models," *Journal of the American Statistical Association*, 95, 608–618. [1517]

Cox, D. R. (1972), "Regression Models and Life-Tables" (with discussion), *Journal of the Royal Statistical Society,* Series B, 34, 187–220. [1517]

Fang, H., Li, G., and Sun, J. (2005), "Maximum Likelihood Estimation in a Semiparametric Logistic/Proportional-Hazards Mixture Model," *Scandinavian Journal of Statistics*, 32, 59–75. [1517]

Farewell, V. T. (1982), "The Use of Mixture Models for the Analysis of Survival Data With Long-Term Survivors," *Biometrics*, 38, 1041–1046. [1517,1518]

——— (1986), "Mixture Models in Survival Analysis: Are They Worth the Risk?" *The Canadian Journal of Statistics*, 14, 257–262. [1517]

Fleming, T. R., and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, New York: Wiley. [1518]

Fygenson, M., and Ritov, Y. (1994), "Monotone Estimating Equations for Censored Data," *The Annals of Statistics*, 22, 732–746. [1519]

Gill, R. D., and Johansen, S. (1990), "A Survey of Product-Integration With a View Toward Application in Survival Analysis," *The Annals of Statistics*, 18, 1501–1555. [1527]

Hahn, J. (1995), "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 11, 105–121. [1518]

Honoré, B., Khan, S., and Powell, J. L. (2002), "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 109, 67–105. [1517]

Horowitz, J. (1998), "Bootstrap Methods for Median Regression Model," *Econometrica*, 66, 1327–1352. [1518]

Huang, Y. (2010), "Quantile Calculus and Censored Regression," *The Annals of Statistics*, 38, 1607–1637. [1518]

Jin, Z., Ying, Z., and Wei, L. J. (2001), "A Simple Resampling Method by Perturbing the Minimand," *Biometrika*, 88, 381–390. [1518]

Koenker, R. (2005), *Quantile Regression*, Cambridge: Cambridge University Press. [1517]

Koenker, R., and Bassett, G. J. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50. [1517]

Kuk, A. Y. C., and Chen, C.-H. (1992), "A Mixture Model Combining Logistic Regression With Proportional Hazards Regression," *Biometrika*, 79, 531–541. [1517]

Li, C.-S., Taylor, J. M. G., and Sy, J. P. (2001), "Identifiability of Cure Models," *Statistics and Probability Letters*, 54, 389–395. [1519]

Liang, H.-Y., Una-Alvarez, J., and Iglesias-Perez, M. C. (2012), "Asymptotic Properties of Conditional Distribution Estimator With Truncated, Censored and Dependent Data," *Test*, 21, 790–810. [1530]

Lin, D. Y., Wei, L. J., and Ying, Z. (1993), "Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals," *Biometrika*, 80, 557–572. [1525]

Lu, W. (2008), "Maximum Likelihood Estimation in the Proportional Hazards Cure Model," *Annals of the Institute of Statistical Mathematics*, 60, 545–574. [1517]

——— (2010), "Efficient Estimation for an Accelerated Failure Time Model With a Cure Fraction," *Statistica Sinica*, 20, 661–674. [1517,1520]

Lu, W., and Ying, Z. (2004), "On Semiparametric Transformation Cure Models," *Biometrika*, 91, 331–343. [1517,1519]

Mao, M., and Wang, J. L. (2010), "Semiparametric Efficient Estimation for a Class of Generalized Proportional Odds Cure Models," *Journal of the American Statistical Association*, 105, 302–311. [1517,1520,1522]

Neocleous, T., and Portnoy, S. (2009), "Partially Linear Censored Quantile Regression," *Lifetime Data Analysis*, 15, 357–378. [1525]

Neocleous, T., Vanden Branden, K., and Portnoy, S. (2006), "Correction to Censored Regression Quantiles," *Journal of the American Statistical Association*, 101, 860–861. [1518]

Parzen, M. I., Wei, L. J., and Ying, Z. (1994), "A Resampling Method Based on Pivotal Estimating Functions," *Biometrika*, 81, 341–350. [1518]

Peng, L., and Huang, Y. (2008), "Survival Analysis With Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649. [1518,1519]

Peng, Y., and Dear, K. B. G. (2000), "A Nonparametric Mixture Model for Cure Rate Estimation," *Biometrics*, 56, 237–243. [1517]

Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012. [1518]

Portnoy, S., and Lin, G. (2010), "Asymptotics for Censored Regression Quantiles," *Journal of Nonparametric Statistics*, 22, 115–130. [1518]

Powell, J. L. (1984), "Least Absolute Deviations Estimation for the Censored Regression," *Journal of Econometrics*, 25, 303–325. [1517]

Powell, J. L. (1986), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155. [1517]

Qian, J., and Peng, L. (2010), "Censored Quantile Regression With Partially Functional Effects," *Biometrika*, 97, 839–850. [1518,1525]

Sy, J. P., and Taylor, J. M. G. (2000), "Estimation in a Cox Proportional Hazards Cure Model," *Biometrics*, 56, 227–236. [1517,1520]

Szydlo, R., Goldman, J. M., Klein, J. P., Gale, R. P., Ash, R. C., Bach, F. H., Bradley, B. A., Casper, J. T., Flomenberg, N., Gajewski, J. L., Gluckman, E., Henslee-Downey, P. J., Hows, J. M., Jacobsen, N., Kolb, H. J., Lowenberg, B., Masaoka, T., Rowlings, P. A., Sondel, P. M., van Bekkum, D. W., van Rood, J. J., Vowels, M. R., Zhang, M. J., and Horowitz, M. M. (1997), "Results of Allogeneic Bone Marrow Transplants for Leukemia Using Donors Other Than HLA-Identical Siblings," *Journal of Clinical Oncology*, 15, 1767–1777. [1523]

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer. [1526,1528,1529]

Wang, H., and Wang, L. (2009), "Locally Weighted Censored Quantile Regression," *Journal of the American Statistical Association*, 104, 1117–1128. [1518,1520,1524]

Wang, H., Zhou, J., and Li, Y. (2013), "Variable Selection for Censored Quantile Regression," *Statistica Sinica*, 23, 145–167. [1525,1526]

Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Survival and Hazard Function," *Journal of the American Statistical Association*, 94, 137–145. [1517]

Yin, G., Zeng, D., and Li, H. (2008), "Power-Transformed Linear Quantile Regression With Censored Data," *Journal of the American Statistical Association*, 103, 1214–1224. [1517]

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. [1517]

Zeng, D., and Lin, D. Y. (2007), "Efficient Estimation for the Accelerated Failure Time Model," *Journal of the American Statistical Association*, 102, 1387–1396. [1517]

Zeng, D., Yin, G., and Ibrahim, J. (2006), "Semiparametric Transformation Models for Survival Data With a Cure Fraction," *Journal of the American Statistical Association*, 101, 670–684. [1517]

Zhang, J., and Peng, Y. (2009), "Accelerated Hazards Mixture Cure Model," *Lifetime Data Analysis*, 15, 455–467. [1517]