

A quantile varying-coefficient regression approach to length-biased data modeling

Xuerong Chen*

School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, 611130, China

e-mail: chenxr522@foxmail.com

Alan T. K. Wan†

Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong

e-mail: alan.wan@cityu.edu.hk

and

Yong Zhou‡

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, 200433, China

e-mail: yzhou@amss.ac.cn

Abstract: Recent years have seen a growing body of literature on the analysis of length-biased data. Much of this literature adopts the accelerated failure time or proportional hazards models as the basis of study. The overwhelming majority of the existing work also assumes independence between the censoring variable and covariates. In this paper, we develop a varying-coefficient quantile regression approach to model length-biased data. Our approach does not only allow the direct estimation of the conditional quantiles of survival times based on a flexible model structure, but also has the important appeal of permitting dependence between the censoring variable and the covariates. We develop local linear estimators of the coefficients using a local inverse probability weighted estimating equation approach, and examine these estimators' asymptotic properties. Moreover, we develop a resampling method for computing the estimators' covariances. The small sample properties of the proposed methods are investigated in a simulation study. A real data example illustrates the application of the methods in practice.

MSC 2010 subject classifications: Primary 62G08, 60K35; secondary 62N02.

*Chen's work was supported by the Fundamental Research Funds for the central universities of China (JBK140507, JBK141111).

†Corresponding author. Wan's work was supported by a General Research Fund from the Hong Kong Research Grants Council (CityU-9042086).

‡Zhou's work was supported by National Natural Science Foundation of China (NSFC) (71271128), the State Key Program of National Natural Science Foundation of China (71331006), NCMIS, Key Laboratory of RCSDS, AMSS, CAS2008DP173182 and Shanghai First-class Discipline A and IRTSHUFE, and PCSIRT (IRT13077).

Keywords and phrases: Estimating equation, length-biased, local linear, prevalent cohort, quantile regression, resampling method, right-censored, varying-coefficient.

Received February 2014.

Contents

1	Introduction	2515
2	Model framework and a weighted estimating equation approach . . .	2517
2.1	Model framework	2517
2.2	A local linear inverse probability weighted estimator	2518
2.3	Covariate-dependent censoring	2519
3	Large sample properties and variance estimation	2521
3.1	Large sample properties	2521
3.2	Asymptotic variance estimation	2522
4	A simulation study	2522
5	A real data example	2526
	Appendix: Proofs of theorems	2529
	References	2537

1. Introduction

In prevalent sampling, when the probability of selecting an observation from the population is proportional to the time from the initiation event to the failure event, the sample is said to be length-biased. As the cohort can only include subjects that survive to the examination time, the observed time interval from initiation to failure of the subjects within the prevalent cohort is longer than that arising from the distribution of the general population. Length-biasedness is therefore due to the way the data are available and not by choice. In technical terms, length-biased sampling is usually understood to refer to the situation where the event initiation time follows a stationary Poisson process such that the failure time is left-truncated by a uniformly distributed truncation time. Length-biased data are typically also right-censored because in a prevalent cohort, the observed subjects are followed forward until death, withdrawal, or the termination of study, whichever is the earliest.

A large amount of work has been done on developing methods for estimating the unbiased distribution of the target population under length-biased sampling (e.g. [41, 42, 3, 4]). There is also a growing body of literature on regression modeling of the association between the length-biased failure times and risk factors. This literature invariably focuses on semi-parametric methods. Some better known examples include [39], who study the covariate effects on failure times using the semi-parametric transformation and accelerated failure time (AFT) models; [37] develop estimating equation methods for estimating covariate effects under the semi-parametric proportional hazards (PH) model, while

[35] propose a Buckley-James type estimator for estimating the covariate effects under the AFT model. It should be noted that these studies all assume independence of the censoring variable and the covariates; when the latter assumption does not hold, the validity of the proposed methods cannot be guaranteed. Moreover, the PH and AFT models are limited by their inability to examine the covariate effects on the shape of the survival distribution as they assume that the covariates only affect the location of the survival times ([32]). To overcome this difficulty, a substantial literature has investigated the use of quantile regression (QR) methods for modeling the conditional quantiles of survival times directly (e.g., [48, 46, 32, 28, 5, 47]). The majority of these studies account for data censoring and assume that censoring times are independent of either or both of the covariates and failure times. None of these studies consider length-biased sampling because they all implicitly or explicitly assume that the data are obtained from an incident follow-up study.

To date, [43] and [14] (WW and CZ hereafter) are the only existing studies that have considered QR for length-biased data. These two studies have much in common in terms of the approaches taken for parameter estimation, namely, methods based on inverse probability weighting. They share the common strength of allowing censoring to depend on the covariates, but have the same disadvantage of restricting the functional relationship between the survival time and the covariates to linearity. While similar in many respects, WW and CZ differ in the ways they model the dependence between the censoring variable and the covariates – CZ uses a semi-parametric PH model, while WW uses a method that involves replacing the integral of the conditional survival function of the censoring variable by the integral of a local Kaplan-Meier estimator. WW also develops a goodness of fit test which is unexplored in CZ.

In this paper we use a varying-coefficient QR approach to model length-biased survival data. This model framework is more flexible than the linear framework of WW and CZ. The varying-coefficient model has been widely applied since the seminal contributions of [15] and [26]. An important appeal of the varying-coefficient model is that by allowing the coefficients to vary as smooth functions of other variables, the curse of dimensionality can be avoided. Due to this advantage the varying-coefficient approach has experienced rapid development in theory and methodology in recent years. We refer to the articles by [13, 11] and [10] for novel applications of the varying-coefficient approach to time series analysis; [29, 23] and [21] for longitudinal data analysis; [22] and [7, 8, 9] for survival analysis; and [45] for functional linear regression. For more references, see [20] and [24]. The varying-coefficient approach to estimating conditional quantiles has been considered by [27, 30] and [12] use local polynomials to estimate conditional quantiles with varying coefficients, while [30] propose an estimation methodology based on polynomial splines. [18] also examine the connection between quantile autoregression and varying-coefficient models. To the best of our knowledge, no previous study under length-biased sampling, whether the focus is on conditional mean or quantile estimation, has considered the varying-coefficient approach, and the purpose of this study is to take steps in this direction.

Like other studies on length-biased sampling, we assume that the initiation times follow a stationary Poisson process (hereinafter referred to as the stationarity assumption), but unlike the majority of the existing studies, we do not require the censoring variable to be distributed independently of the covariates or failure times. We consider the latter an important strength that underlies our analysis besides the advantage offered by the flexibility of the varying-coefficient model. We adopt a local inverse probability weighting approach to estimation, and establish the consistency and asymptotic normality of the QR-based local linear estimators obtained by this estimation approach. Furthermore, a resampling method is developed for the computation of estimator variance.

The layout of this paper is as follows. The next section introduces the model framework, and develops the inverse probability weighting estimation procedure. Section 3 presents the main theoretical results on the consistency and asymptotic normality of estimators, along with the development of a resampling method for covariance estimation. Simulation findings on the properties of estimators in small samples are reported in Section 4. Section 5 considers an application of the proposed methods to real data, and the Appendix provides the proofs of our technical results.

2. Model framework and a weighted estimating equation approach

2.1. Model framework

Let \tilde{T} be the time from the initiation event to the failure event. The initiation and failure events could be, for example, the onset of a disease and death respectively. Further, let A , V and C be the times from the initiation event to recruitment, from recruitment to failure, and from recruitment to censoring respectively. When the data are length-biased, only \bar{T} , the subset of \tilde{T} such that $\tilde{T} > A$, can be observed. For this reason, A is also known as the truncation variable. The variable V is the residual survival time because $V = \bar{T} - A$, and C is the censoring variable. Note that A is common to the survival time $\bar{T} = A + V$ and the total censoring time $A + C$. Hence, the survival and censoring times are dependent. This also means that \bar{T} is subject to informative censoring. We assume that conditional on the covariates (\mathbf{X}, U) , C and (A, V) are independent and the stationarity assumption holds. Further, denote $f_{\tilde{T}}$ as the density function of \tilde{T} given $\mathbf{X} = \mathbf{x}$ and $\mathbf{U} = \mathbf{u}$. Then the conditional density function g of the length-biased data \bar{T} given $\bar{T} > A$ is

$$g(t|\mathbf{x}, \mathbf{u}) = \frac{t f_{\tilde{T}}(t|\mathbf{x}, \mathbf{u})}{\mu(\mathbf{x}, \mathbf{u})},$$

where $\mu(\mathbf{x}, \mathbf{u}) = \int_0^\infty s f_{\tilde{T}}(s|\mathbf{x}, \mathbf{u}) ds < \infty$.

Now, let $(Y_i, A_i, \delta_i, U_i, \mathbf{X}_i)$ be a random sample, where $Y_i = \min(\bar{T}_i, A_i + C_i)$, $\bar{T}_i = A_i + V_i$, $\delta_i = I(V_i \leq C_i)$, and U_i and \mathbf{X}_i are one-dimensional and p -dimensional covariates for individual i respectively, $i = 1, 2, \dots, n$. We assume

that \tilde{T} (or its transformation such as log) follows the varying-coefficient quantile regression model

$$\tilde{T}_i = \mathbf{a}(\mathbf{U}_i, \tau)^T \mathbf{X}_i + \epsilon_i, \quad (2.1)$$

where $\mathbf{a}(\mathbf{u}, \tau) = (a_1(\mathbf{u}, \tau), \dots, a_p(\mathbf{u}, \tau))^T$ is a $p \times 1$ vector of unknown coefficient functions, $\tau \in (0, 1)$ is a given quantile level, the random errors ϵ_i 's are independent, and covariates U_i and $X_i, i = 1, 2, \dots, n$, are i.i.d; furthermore, it is assumed that $P(\epsilon_i < 0 | \mathbf{X}_i, U_i) = \tau$. In the notation that follows, we will suppress the argument τ and write $\mathbf{a}(\mathbf{u}, \tau)$ as $\mathbf{a}(\mathbf{u})$ whenever there is no confusion. Also, we assume that $P(C > t_{\mathbf{x}, u} | \mathbf{X}, U) > 0$, where $0 < t_{\mathbf{x}, u} = \sup\{t : P(V \geq t | \mathbf{X}, U) > 0\}$, such that the support of C also covers that of V for any covariate value. We make this assumption to avoid the technical complications regarding the tail behavior of the limiting distribution. Additionally, in order for all of the regression parameters to be estimable, we restrict inference to the time interval $[0, t_0]$, where t_0 is chosen such that $\inf_{\mathbf{x}, u} P(V \geq t_0 | \mathbf{X}, U) > 0$.

Now, if the complete data that are neither length-biased nor right-censored can be observed, then the estimator of $\mathbf{a}(U) = (a_1(U), \dots, a_p(U))^T$ is the minimizer of the objective function

$$\sum_{i=1}^n \left\{ \rho_{\tau} \left(\tilde{T}_i - \mathbf{a}(U_i)^T \mathbf{X}_i \right) \right\},$$

where $\rho_{\tau}(\epsilon) := \epsilon(\tau - I(\epsilon < 0))$.

2.2. A local linear inverse probability weighted estimator

In this subsection, we will construct the local asymptotically unbiased estimating equation to derive the local linear estimator of the coefficient function $\mathbf{a}(u)$, taking into account the length-biased and right-censored characteristics of the data. Similar to the approach of [39], we obtain the local linear estimator of the unknown coefficient functions from the local inverse probability weighted estimating equation by redistributing the mass of the censored observations to the uncensored observations. The survival function of the censoring variable C is $S_C(t | \mathbf{W}) = P(C > t | \mathbf{W})$, where \mathbf{W} is (\mathbf{X}, U) or a subset of it.

Under the stationarity assumption (i.e., the initiation times follow a stationary Poisson process), from the results of [4], the joint distribution of (A, V) and (A, T) conditional on \mathbf{X} and U has the form

$$f_{A,V}(a, v | \mathbf{X} = \mathbf{x}, U = u) = f_{\tilde{T}}(a + v | \mathbf{x}, u) I(a > 0, v > 0) / \mu(\mathbf{x}, u).$$

Similar to [39], the probability of observing the failure data can be written as

$$\begin{aligned} & P(A = a, V = y - a, C \geq y - a | \mathbf{X} = \mathbf{x}, U = u) \\ &= f_{\tilde{T}}(y | \mathbf{x}, u) S_C(y - a | \mathbf{w}) / \mu(\mathbf{x}, u). \end{aligned} \quad (2.2)$$

Now, based on (2.2) and the joint distribution of (A, V) , we have

$$\begin{aligned} & E\left[\frac{\delta}{\pi_{0w}(Y|\mathbf{W})}\rho_\tau(Y - \mathbf{a}(U)^T\mathbf{X})\right] \\ &= E\left\{E\left[\left(\frac{\delta}{\pi_{0w}(Y|\mathbf{W})}\rho_\tau(Y - \mathbf{a}(U)^T\mathbf{X})\right)\middle|\mathbf{X}, U\right]\right\} \\ &= E\left\{\int_0^\infty \frac{1}{\mu(\mathbf{X}, U)}f_{\mathbf{X}, U}(t)\rho_\tau(T - \mathbf{a}(U)^T\mathbf{X})dt\right\} \\ &= E\left\{\frac{1}{\mu(\mathbf{X}, U)}E[\rho_\tau(\tilde{T} - \mathbf{a}(U)^T\mathbf{X})|\mathbf{X}, U]\right\}, \end{aligned}$$

where $\pi_{0w}(t|\mathbf{W}) = \int_0^t S_C(u|\mathbf{W})du$.

To proceed with the estimation of $\mathbf{a}(u)$, we approximate $a_j(U)$ as $a_j(U) \approx a_j(u_0) + a'_j(u_0)(U - u_0)$ for U in the neighborhood of a given point u_0 . Write $\mathbf{b}_0(u_0) = (a_1(u_0), \dots, a_p(u_0))^T$, $\mathbf{c}_0(u_0) = (a'_1(u_0), \dots, a'_p(u_0))^T$, $\mathbf{b} = (b_1, \dots, b_p)^T$, $\mathbf{c} = (c_1, \dots, c_p)^T$, $\beta_0(u_0) = (\mathbf{b}_0^T(u_0), \mathbf{c}_0^T(u_0))^T$, $\beta = (\mathbf{b}^T, \mathbf{c}^T)^T$, $\mathbf{H} = \text{diag}(1, h) \otimes \mathbf{I}_p$, and $\mathbf{X}_i^* = (\mathbf{X}_i^T, \mathbf{X}_i^T(U_i - u_0)/h)^T$. Furthermore, we let $\theta =: \mathbf{H}\beta = \mathbf{H}(\mathbf{b}^T, \mathbf{c}^T)^T$, $\theta_0 =: \mathbf{H}\beta_0(u_0) = \mathbf{H}(\mathbf{b}_0^T(u_0), \mathbf{c}_0^T(u_0))^T$, $K_h(\cdot) = K(\cdot/h)/h$, with $K(\cdot)$ being a kernel function and $h = h_n > 0$ a bandwidth. The estimator $\hat{\theta} =: \mathbf{H}\hat{\beta} = \mathbf{H}(\hat{\mathbf{b}}^T, \hat{\mathbf{c}}^T)^T$ of θ_0 is obtained by minimizing the loss function

$$L_n(\theta, \pi_{0w}) = \sum_{i=1}^n \frac{\delta_i K_h(U_i - u_0)}{\pi_{0w}(Y_i|\mathbf{W}_i)} \rho_\tau(Y_i - \theta^T \mathbf{X}_i^*), \tag{2.3}$$

and replacing $\pi_{0w}(t|\mathbf{W})$ in the solution to (2.3) by a consistent estimator $\hat{\pi}_w(t|\mathbf{W})$ of $\pi_{0w}(t|\mathbf{W})$.

The task of obtaining $\hat{\pi}_w(t|\mathbf{W})$ can be accomplished by seeking a consistent estimator of $S_C(t|\mathbf{W})$ since $\pi_{0w}(t|\mathbf{W}) = \int_0^t S_C(u|\mathbf{W})du$. When censoring depends on the covariates, $S_C(t|\mathbf{W})$ can be estimated semi-parametrically or non-parametrically. Semi-parametric methods are based on a regression model specified for the censoring time (e.g., the Cox proportional hazards model), while non-parametric methods usually involve using the local Kaplan-Meier estimator to estimate the survival function directly. When censoring is independent of the covariates, the same modeling method can be applied with $\pi_0(t) = \int_0^{t_0} S_C(u)du$ used as the weight function. In this paper, we adopt the semi-parametric method to estimate the weight function. More details are given in the next subsection.

2.3. Covariate-dependent censoring

We assume that the dependence of the censoring time on the covariates can be described by the Cox proportional hazards model

$$\lambda(t|\mathbf{W}_i) = \lambda_0(t) \exp(\alpha_0^T \mathbf{W}_i), \tag{2.4}$$

where $\lambda(t|\mathbf{W}) = \lim_{\Delta t \rightarrow 0} P[(t \leq C \leq t + \Delta t) | C \geq t, \mathbf{W}] / \Delta t$ is the hazard function, $\boldsymbol{\alpha}_0$ is a r -dimensional parameter, \mathbf{W}_i is a r -dimensional covariate vector, and $\lambda_0(t)$ is an unspecified baseline hazard function. The possibility of (2.4) being a mis-specified representation can be tested using the method proposed by [34].

The following gives the procedure for estimating the weight function $\pi_{0w}(Y|\mathbf{W})$. Now, recognizing that $\pi_{0w}(t|\mathbf{W}) = \int_0^t S_C(s|\mathbf{W})ds$, a consistent estimator of $\pi_{0w}(t|\mathbf{W})$ is $\hat{\pi}_w(t|\mathbf{W}) = \int_0^t \hat{S}_C(s|\mathbf{W})ds$, where

$$\hat{S}_C(s|\mathbf{W}) = \exp\{-\exp(\hat{\boldsymbol{\alpha}}^T \mathbf{W})\hat{\Lambda}_0(s)\}$$

is a commonly used consistent estimator of $S_C(s|\mathbf{W})$ (see, for example, [25]). Let $\hat{\boldsymbol{\alpha}}$ be the maximum partial likelihood estimator of $\boldsymbol{\alpha}_0$, obtained by solving the estimating equation

$$U_n(\boldsymbol{\alpha}) = \sum_{i=1}^n \int_0^{t_0} \{\mathbf{W}_i - \bar{\mathbf{W}}(t; \boldsymbol{\alpha})\} dN_i^C(t) = 0,$$

where $N_i^C = I(Y_i - A_i \leq t, \delta_i = 0)$ is a counting function, and $\bar{\mathbf{W}}(t; \boldsymbol{\alpha}) = S^{(1)}(t; \boldsymbol{\alpha}) / S^{(0)}(t; \boldsymbol{\alpha})$, with

$$S^{(k)}(t; \boldsymbol{\alpha}) = \frac{1}{n} \sum_{i=1}^n Y_i(t) \mathbf{W}_i^{\otimes k} \exp(\boldsymbol{\alpha}^T \mathbf{W}_i), \quad k = 0, 1, 2,$$

$Y_i(t) = I(Y_i - A_i \geq t)$, and $\mathbf{v}^{\otimes 0} = 1$, $\mathbf{v}^{\otimes 1} = \mathbf{v}$ and $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^T$ for any arbitrary vector \mathbf{v} . Further, let

$$\hat{\Lambda}_0(t) = \sum_{i=1}^n \int_0^t \frac{dN_i^C(u)}{\sum_{j=1}^n Y_j(u) \exp(\hat{\boldsymbol{\alpha}}^T \mathbf{W}_j)}$$

be the [6] estimator of the cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. See [33] for a discussion of Breslow estimator in survival analysis. Let

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n \int_0^{t_0} \left\{ \frac{S^{(2)}(t; \hat{\boldsymbol{\alpha}})}{S^{(0)}(t; \hat{\boldsymbol{\alpha}})} - \bar{\mathbf{W}}^{\otimes 2}(t; \hat{\boldsymbol{\alpha}}) \right\} dN_i^C(t),$$

and denote $\bar{\mathbf{w}}(t)$, $s^{(0)}(t; \boldsymbol{\alpha}_0)$ and Ω as the limits of $\bar{\mathbf{W}}(t; \boldsymbol{\alpha}_0)$, $S^{(0)}(t; \boldsymbol{\alpha}_0)$ and $\hat{\Omega}$ respectively.

These notations facilitate the derivation of the asymptotic martingale representation of $\pi_{0w}(Y|\mathbf{W}) - \hat{\pi}_w(Y|\mathbf{W})$, which will be used to prove the asymptotic normality of the local linear estimator. The latter is the minimizer of the loss function

$$L_n(\boldsymbol{\theta}, \hat{\pi}_w) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i K_h(U_i - u_0)}{\hat{\pi}_w(Y_i | \mathbf{W}_i)} \rho_\tau(Y_i - \boldsymbol{\theta}^T \mathbf{X}_i^*), \quad (2.5)$$

where $\hat{\pi}_w(t|\mathbf{W}) = \int_0^t \hat{S}_C(s|\mathbf{W})ds$.

3. Large sample properties and variance estimation

3.1. Large sample properties

This subsection is devoted to an analysis of the large sample properties of the local linear estimator. In order to apply the counting process and martingale theory, let us define the filtration

$$\mathcal{F}(t) = \sigma\{I(Y_i - A_i \leq u, \delta_i = 0), I(Y_i - A_i \geq u^+), \mathbf{X}_i, U_i, i = 1, 2, \dots, n, 0 \leq u \leq t\}.$$

Then the martingale with respect to the censoring variable is

$$M_i(t) = N_i^C(t) - \int_0^t I(Y_i - A_i \geq u) \exp\{\boldsymbol{\alpha}_0^T \mathbf{W}_i\} \lambda_0(u) du.$$

For notational simplicity, write $\eta_i(u, \mathbf{X}) = \mathbf{a}(u)^T \mathbf{X}_i$, $\mu_k = \int u^k K(u) du$, $\nu_k = \int u^k K^2(u) du$, $f_U(\cdot)$ as the marginal density of U , $f_{\epsilon_i}(\cdot | \mathbf{X}, U)$ as the conditional density of random error,

$$\begin{aligned} A_n(u_0) &= f_U(u_0) \begin{pmatrix} 1 \\ \mu_2 \end{pmatrix} \otimes \Gamma_n(u_0), A(u_0) = \lim_{n \rightarrow \infty} A_n(u_0), \\ \Gamma_n(u_0) &= \frac{1}{n} \sum_{i=1}^n E \left[\frac{\mathbf{X}_i \mathbf{X}_i^T}{\mu(\mathbf{X}_i, u_0)} f_{\epsilon_i}(0 | \mathbf{X}_i, U_i) \middle| U_i = u_0 \right], \end{aligned}$$

and

$$V = \lim_{n \rightarrow \infty} f_U(u_0) \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \frac{1}{n} \sum_{i=1}^n E \left\{ \frac{\mathbf{X}_i \mathbf{X}_i^T \phi_\tau^2(Y_i - \eta_i(u_0, \mathbf{X}))}{\mu(\mathbf{X}_i, u_0) \pi_{0w}(Y_i | \mathbf{W}_i)} \middle| U_i = u_0 \right\},$$

where $\phi_\tau(\epsilon) = I(\epsilon \leq 0) - \tau$.

The theorems below show that the local linear estimator is consistent as well as asymptotically normal.

Theorem 3.1 (Consistency). *Assume that conditions (A₁)–(A₁₂) in the [Appendix](#) hold, and as $n \rightarrow \infty$, $h = h_n \rightarrow 0$, $nh \rightarrow \infty$, and $n^{2\delta-1}h \rightarrow \infty$ for $3/5 < \delta < 1 - s^{-1}$ with $s > 2$ being a positive integer. Then*

$$\mathbf{H}(\hat{\boldsymbol{\beta}}(u_0) - \boldsymbol{\beta}_0(u_0)) \xrightarrow{P} 0.$$

Theorem 3.2 (Asymptotic Normality). *Assume that conditions of Theorem 1 hold. Then*

$$\sqrt{nh} \left[\mathbf{H}(\hat{\boldsymbol{\beta}}(u_0) - \boldsymbol{\beta}_0(u_0)) - \frac{h^2}{2} \begin{pmatrix} \mu_2 \mathbf{a}''(u_0) \\ 0 \end{pmatrix} + o(h^2) \right] \xrightarrow{D} N(0, \Sigma),$$

where $\Sigma = A^{-1}(u_0)VA^{-1}(u_0)$, and $\mathbf{a}''(u)$ is the second derivative of coefficient function vector $\mathbf{a}(u)$.

3.2. Asymptotic variance estimation

It is difficult to estimate the asymptotic variance by the direct plug-in method as this method involves the computation of the estimator of an unknown density function which is cumbersome. For this reason we use an alternative bootstrap method based on Efron's non-parametric bootstrap ([16, 17]). It works as follows. Select B random samples each of size n with replacement from $(Y_i, A_i, \delta_i, U_i, \mathbf{X}_i), i = 1, 2, \dots, n$. Let the j -th bootstrapped sample be $(Y_i^{(j)}, A_i^{(j)}, \delta_i^{(j)}, U_i^{(j)}, \mathbf{X}_i^{(j)}), i = 1, 2, \dots, n, j = 1, 2, \dots, B$. For each bootstrapped sample, compute $\hat{\pi}_w^{(j)}(Y_i^{(j)} | \mathbf{W}_i^{(j)})$, and let $\hat{\boldsymbol{\theta}}^{(j)}$ be the solution to the following bootstrapped estimating equation:

$$\bar{L}_n(\boldsymbol{\theta}, \hat{\pi}_w^{(j)}) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i^{(j)} K_h(U_i^{(j)} - u_0)}{\hat{\pi}_w(Y_i^{(j)} | \mathbf{W}_i^{(j)})} \rho_\tau(Y_i^{(j)} - \boldsymbol{\theta}^T \mathbf{X}_i^{*(j)}). \quad (3.1)$$

Then we use the empirical variance of $(\hat{\boldsymbol{\theta}}^{(1)}, \dots, \hat{\boldsymbol{\theta}}^{(B)})$ as the estimator of the asymptotic variance of $\hat{\boldsymbol{\theta}}$. Note that local linear approximation has the effect of reducing the dimension of the original parameters $\mathbf{a}(u)$ to a finite one point-wisely, and transforming the estimation problem from being non-parametric to parametric. Along the lines of [16, 17], it is not difficult to show that this procedure can result in a consistent estimator of the variance. Clearly, the value B has to be reasonably large for this method to be effective.

4. A simulation study

In this section, we evaluate the finite sample performance of the proposed method by simulations. Our generation of length-biased data follows the method of [39]. The method works as follows. We generate independent pairs of (A_i, \tilde{T}_i) from $A \sim U(b, \omega)$, where b is an arbitrary constant, and ω is chosen such that it is larger than the upper bound of \tilde{T} for the stationarity assumption to hold. We then select the pairs such that $A_i < \tilde{T}_i$. For all cases in the study, we set the sample size n to 300, the number of replications to 500, and B , the number of bootstrapped samples for the estimation of the asymptotic variance, to 500.

Our experiment is based on the following data generating process:

$$\log \tilde{T}_i = a_1(U_i)X_{1i} + a_2(U_i)X_{2i} + (1 + \gamma X_{2i})\epsilon_i, \quad (4.1)$$

where $X_{1i} \sim N(0, 1)$ is a continuous covariate, X_{2i} is a binary (0,1) covariate with $P(X_{2i} = 1) = 0.5$, and $U_i \sim U(0, 3)$ is the effect modifier. We let the varying-coefficient functions be $a_1(U_i) = \sin(2U_i)$ and $a_2(U_i) = 0.5U_i(2 - U_i)$. The censoring variable C_i is assumed to depend on X_{2i} through the Cox proportional hazards model

$$\lambda(t) = \lambda_0(t) \exp(\alpha_0 X_{2i})$$

described in (2.4). Here, we let $\alpha_0 = -1$ and $\lambda_0(t) = 0.35$. Now, under model (4.1), the τ -conditional quantile of $\log \tilde{T}$ is

$$Q_\tau(\log \tilde{T} | X_{1i}, X_{2i}) = a_1(U_i, \tau)X_{1i} + a_2(U_i, \tau)X_{2i},$$

where $a_1(U_i, \tau) = a_1(U_i) + Q_\epsilon(\tau)$, $a_2(U_i, \tau) = a_2(U_i) + \gamma Q_\epsilon(\tau)$, and $Q_\epsilon(\tau)$ is τ -quantile of ϵ_i . We consider $\tau = 0.25, 0.50$ and 0.95 , and the following three cases of specifications of γ and ϵ_i : Case 1: $\gamma = 0$ and $\epsilon_i \sim N(0, 0.25)$; Case 2: $\gamma = 1$ and $\epsilon_i \sim U(-0.5, 0.5)$; Case 3: $\gamma = 1$ and $\epsilon_i \sim N(0, 0.25)$.

Case 1 corresponds to a homoscedastic model, while Cases 2 and 3 correspond to heteroscedastic models. Note that for the latter two cases, as $\gamma = 1$, the coefficient function $a_2(U, \tau)$ has different behaviour for different τ . The censoring ratios for Cases 1, 2 and 3 are about 27.5%, 26% and 35.7% respectively. We adopt the Epanechnikov kernel, $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ in all kernel estimation, and set the bandwidth h to 0.25, 0.15 and 0.3 when $\tau = 0.25, 0.50$ and 0.95 respectively.

Table 1 presents the results of variance estimation based on the QR method for the three quantiles at $u_0 = 0.5, 1.0, 1.5, 2.0, 2.5$. In the table, “*sd*” is the standard deviation of $\hat{a}_j(U)$ based on 500 replications, and “*se*” is the average of the standard errors of the 500 estimated standard deviations based on the resampling method. Hence “*sd*” may be viewed as an indicator of the efficiency of the QR estimates; it also provides the basis for evaluating the accuracy of the resampling method. Table 1 shows that *se* is always very close to *sd*, suggesting that the resampling method is effective. We also observe from Table 1 that the variances are generally not large, suggesting that our method performs well. There is also a general tendency for the estimators of the first coefficient to be more efficient than the corresponding estimators of the second coefficient.

Figures 1–3 provide the plots of the estimated coefficient functions of $a_1(U)$ and $a_2(U)$ for the three cases. In each figure, the true coefficient is shown by the black solid curve, the estimated functions for $\tau = 0.25, 0.5$ and 0.95 are represented by the green dashed-dotted, blue dotted and the green dashed curves respectively, and the 95% point-wise confidence intervals of the coefficients are shown for the case of $\tau = 0.5$ by the red dashed curves. The figures show that all the estimated functions are very close to the true coefficient function. As well, the true values are always enclosed by the 95% confidence intervals. The QR method thus appears to perform well.

Some insights on the effect of ignoring the length-biasedness of the data (i.e., treat the data as if it were only right censored) are provided in Figure 4. We use Case 1 as an illustration but the effects under the other two cases are similar. A comparison of Figure 4 with Figure 1 reveals that when length-biasedness is ignored, the three QR estimators of $a_2(U)$ are all biased and the 95% point-wise confidence interval based on the median estimator does not always contain the true coefficient. On the other hand, the effects of neglecting length-biased sampling appears to be minimal for $a_1(U)$.

We also compare the estimator’s performance under different bandwidths by the following mean square error (MSE) measure

$$\text{MSE} = \frac{1}{n_{\text{grid}}} \sum_{j=1}^2 \sum_{k=1}^{n_{\text{grid}}} [\hat{a}_j(u_k) - a_j(u_k)]^2,$$

TABLE 1. Variance estimation based on QR estimation

Case 1	$a_1(U)$						$a_2(U)$					
	$\tau = .25$		$\tau = .5$		$\tau = .95$		$\tau = .25$		$\tau = .5$		$\tau = .95$	
u_0	sd	se	sd	se	sd	se	sd	se	sd	se	sd	se
0.5	0.141	0.129	0.129	0.125	0.118	0.128	0.171	0.160	0.142	0.153	0.129	0.141
1.0	0.143	0.126	0.132	0.121	0.106	0.111	0.156	0.146	0.137	0.141	0.121	0.124
1.5	0.120	0.120	0.110	0.115	0.100	0.105	0.155	0.150	0.137	0.144	0.124	0.127
2.0	0.145	0.134	0.130	0.132	0.116	0.126	0.212	0.171	0.165	0.172	0.143	0.149
2.5	0.178	0.164	0.163	0.161	0.181	0.187	0.269	0.226	0.246	0.244	0.231	0.226
Case 2	$a_1(U)$						$a_2(U)$					
u_0	$\tau = .25$		$\tau = .5$		$\tau = .95$		$\tau = .25$		$\tau = .5$		$\tau = .95$	
0.5	0.114	0.113	0.124	0.116	0.073	0.074	0.196	0.184	0.198	0.197	0.073	0.079
1.0	0.118	0.109	0.128	0.117	0.064	0.068	0.190	0.167	0.193	0.188	0.065	0.072
1.5	0.093	0.096	0.107	0.108	0.040	0.049	0.185	0.169	0.201	0.194	0.059	0.064
2.0	0.126	0.115	0.133	0.123	0.063	0.074	0.228	0.171	0.237	0.222	0.082	0.088
2.5	0.147	0.142	0.149	0.145	0.125	0.118	0.294	0.249	0.301	0.284	0.143	0.157
Case 3	$a_1(U)$						$a_2(U)$					
u_0	$\tau = .25$		$\tau = .5$		$\tau = .95$		$\tau = .25$		$\tau = .5$		$\tau = .95$	
0.5	0.219	0.204	0.207	0.202	0.173	0.187	0.407	0.313	0.322	0.323	0.221	0.246
1.0	0.224	0.196	0.213	0.196	0.163	0.165	0.389	0.295	0.341	0.309	0.206	0.226
1.5	0.206	0.183	0.197	0.188	0.149	0.156	0.406	0.307	0.328	0.321	0.212	0.228
2.0	0.230	0.204	0.210	0.211	0.176	0.191	0.430	0.321	0.397	0.356	0.243	0.272
2.5	0.275	0.239	0.255	0.247	0.266	0.260	0.501	0.390	0.518	0.444	0.378	0.408

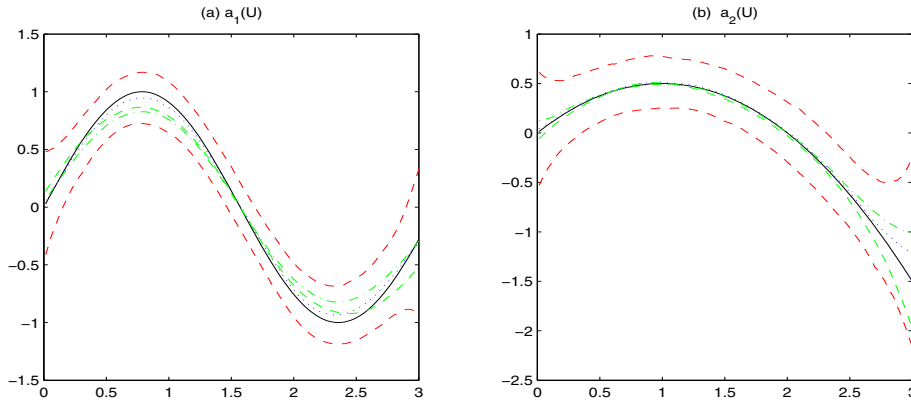


FIG 1. (Case 1) The estimated coefficient function for three quantiles: $\tau=0.25$ (green dashed dotted curve), $\tau=0.50$ (blue dotted curve) and $\tau=0.95$ (green dashed curve). The red dashed curves are the 95% pointwise confidence bands of the median regression curve. The black solid curve represents the true coefficient function.

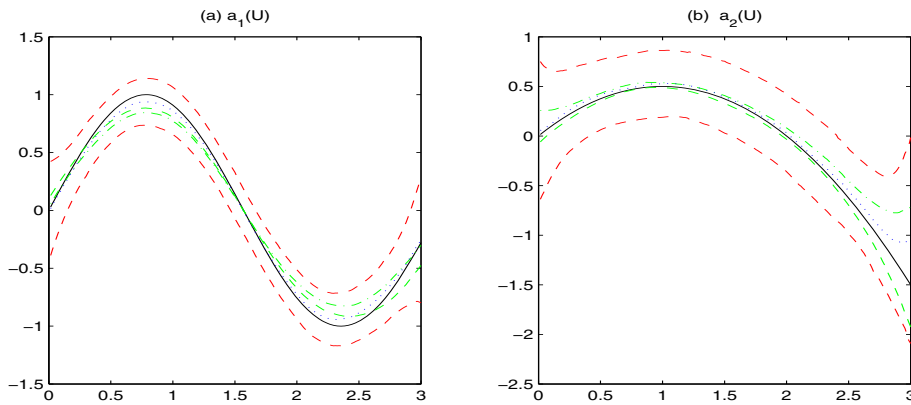


FIG 2. (Case 2) The estimated coefficient function for three quantiles: $\tau=0.25$ (green dashed dotted curve), $\tau=0.50$ (blue dotted curve) and $\tau=0.95$ (green dashed curve). The red dashed curves are the 95% pointwise confidence bands of the median regression curve. The black solid curve represents the true coefficient function.

where $u_k, k = 1, \dots, n_{\text{grid}}$, are the grid points at which the functions $a_j(\cdot)$'s are estimated.

Figure 5(a)–(c) depict the box-plots of the MSEs of the QR estimators for $\tau = 0.25, 0.5, 0.95$ under Case 1 (Boxes 1, 2 and 3), Case 2 (Boxes 4, 5 and 6) and Case 3 (Boxes 7, 8 and 9) based on 500 replications and bandwidth values of $h = 0.15, 0.2, 0.25$. From the figures, we observe that under Case 1, the results for estimating the median are the best, followed by those for the estimating the 25-th quantile, and the results for estimating the 95-th quantile are the worst. This is quite reasonable because under normal errors, data sparsity increases as one moves away from the center of the distribution. However, for Cases 2

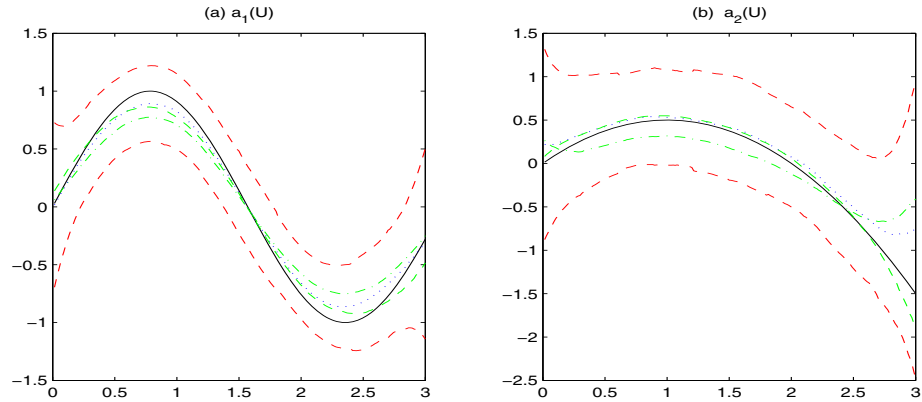


FIG 3. (*Case 3*) The estimated coefficient function for three quantiles: $\tau=0.25$ (green dashed dotted curve), $\tau=0.50$ (blue dotted curve) and $\tau=0.95$ (green dashed curve). The red dashed curves are the 95% pointwise confidence bands of the median regression curve. The black solid curve represents the true coefficient function.

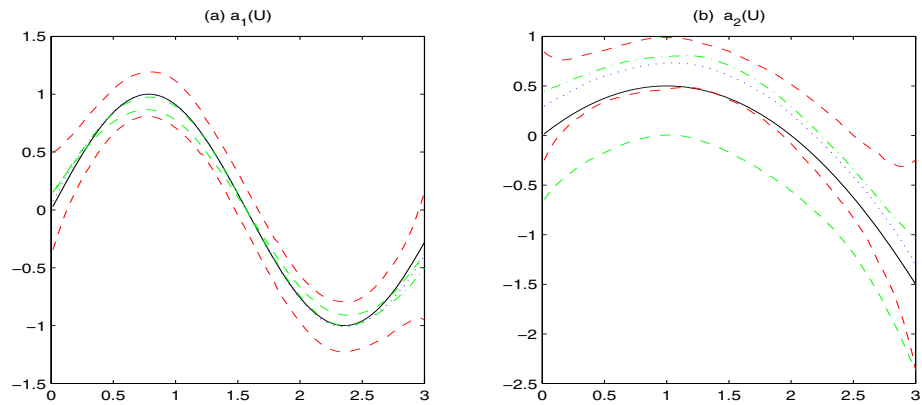


FIG 4. (*Case 1 when ignoring length-biased sampling*) The estimated coefficient function for three quantiles: $\tau=0.25$ (green dashed dotted curve), $\tau=0.50$ (blue dotted curve) and $\tau=0.95$ (green dashed curve). The red dashed curves are the 95% pointwise confidence bands of the median regression curve. The black solid curve represents the true coefficient function.

and 3, the box-plots show that the estimation results are better for the higher than the lower quantiles. This arises because length-biasedness is a special case of left truncatedness, and hence data become sparse at lower quantiles. The heteroscedastic structure of the data may also exacerbate the bad effect of left truncation on the quantile estimates.

5. A real data example

Our real data example is based on the Oscar data set that contains 1668 observations of actors and actresses between 1929 (the inaugural year of the Academy

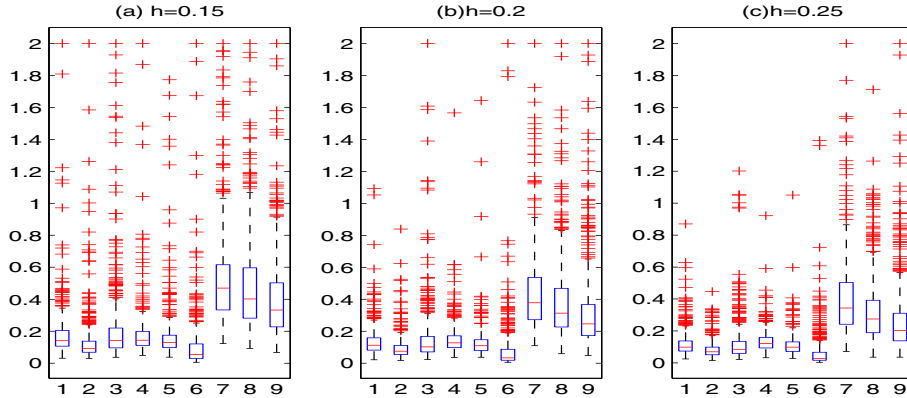


FIG 5. Box-plots of the MSEs of the QR estimators for $\tau = 0.25, 0.5, 0.95$ under Case 1 (Boxes 1, 2 and 3), Case 2 (Boxes 4, 5 and 6) and Case 3 (Boxes 7, 8 and 9).

Award (AA) and 2001 given in [38] and [40]. Among these observations of data are 238 individuals who won at least one AA, 528 individuals who received one or more nominations for AA but never won, and 902 individuals who never received any nomination during the said period. The data set contains personal as well as professional information of the actors and actresses including their gender, country of birth, years of birth and death, ethnicity, genre, the number of movies acted, and so on. The variable of interest is \tilde{T}_i , the age of individual i at death or in 2001 when the study ended. We are mainly interested in the difference between the life spans of AA winners and non-AA winners, and how the first AA nomination affects the subsequent life span of the nominee. Results of [1] stationary test indicate the subset of data that contains only those who have been nominated for AA (including the nominees who eventually won the award) would satisfy the stationary assumption. It means that the subset is a length-biased and right censored data set. The data are length-biased because only those individuals who survived to the first nominated year would be included in the subset, and right censored because some individuals of the study were still alive in 2001. This subset comprises 766 observations. Accordingly, we base our study on this subset. Our stationary test results also concur with the conclusion of [44] who used the same data set in a related study. The censoring rate is 57.31% as 329 of these 766 individuals died before the end of the study.

Our analysis is based on the quantile varying-coefficient model

$$Q(\tilde{T}_i|\mathbf{X}_i) = a_0(U_i, \tau) + \mathbf{a}^T(U_i, \tau)\mathbf{X}_i,$$

where \tilde{T} is the survival time defined above, $\mathbf{a}(u, \tau) = (a_1(u, \tau), \dots, a_7(u, \tau))^T$, U is the age of the nominee when (s)he was first nominated for AA, $\mathbf{X} = (X_1, \dots, X_8)$, with X_1, X_2, X_3, X_4, X_5 , and X_7 being indicator variables representing, respectively, gender (1 = male, 0 = female), country of birth (1 = U.S., 0 = others), ethnicity (1 = Caucasian, 0 = others), name change (1 = yes,

0 = no), genre (1 = drama, 0 = others), and AA winner (1 = yes, 0 = no), X_6 being the number of 4-star movies in which the actor or actress had acted, and X_8 being the number of nomination received for an AA.

We treat the relationship between the censoring variable C and X_2, X_3, X_5, X_7 and X_8 as independent as the correlation between them is found to be very weak. On the other hand, the correlation between C and X_1, X_4 and X_6 is stronger, and we model this relationship by the Cox model

$$\lambda(t|\tilde{\mathbf{X}}_i) = \lambda_0(t) \exp(\boldsymbol{\alpha}_0^T \tilde{\mathbf{X}}_i),$$

where $\tilde{\mathbf{X}} = (X_1, X_3, X_6)$. All estimated quantiles are based on the uniform kernel with bandwidth $h = 0.2 * (\max_i(U_i) - \min_i(U_i))$.

Figures 6 (a)–(i) present the various estimated functional coefficients for the three quantiles, $\tau = 0.25, 0.5$ and 0.95 , and the 95% pointwise confidence intervals based on the $\tau = 0.5$ quantile estimator. Figure 6(a) shows that the estimated intercept functions $a_0(U)$ are non-negative and increasing as U increases for all three quantiles. This means that those who received their first nomination at an older age tend to live longer than those nominated at a younger age. As shown in Figures 6(b) and (c), the estimated functions $a_1(U)$ and $a_2(U)$ have similar behavior; these functions mostly lie below 0 and tend to increase as U increases. This suggests that actresses generally live longer than actors, and those born in the U.S. tend to have shorter lives than their non-U.S. born counterparts, but the differences decrease as U increases. The difference in life span between the two genders is statistically significant only when the first nomination is received at an age younger than 26. Figure 6(d) shows that $a_3(U)$'s are nearly always above 0, and are close to 0 only when U is large, suggesting that Caucasian actors and actresses tend to live longer but this race advantage decreases as U , the age at which first nomination is received, increases. From Figure 6(e), the estimated functions $a_4(U)$'s are negative when $U < 50$, increase rapidly to about 0 when U is between 50 and 65, then decrease to under zero again for $U > 65$. This is an interesting result as it implies that other things being equal, name change leads to a shorter life. Figure 6(f) and (h) show that the coefficient functions $a_5(U)$ and $a_7(U)$ are both very close to 0, suggesting that there is no significant difference in life span between AA winners and non-winners, and between those acting for the drama-genre and non-drama-genre. The results of Figure 6(g) suggest that the actor or actress' life span has a positive association with the number of 4-star movies s(he) acted in, as $a_6(U)$'s are invariably non-negative. However, this positive association is statistically significant only when $U < 21.5$. Moreover, as U increases, all three $a_6(U)$ functions quickly decrease towards zero. Figure 6(i) shows that the three $a_8(U)$'s estimated functions generally do not share a common pattern, and the coefficient is not significantly different from zero. This suggests that other things being equal, there is no difference in life span between AA winners and non-winners. This is consistent with the conclusion of [40].

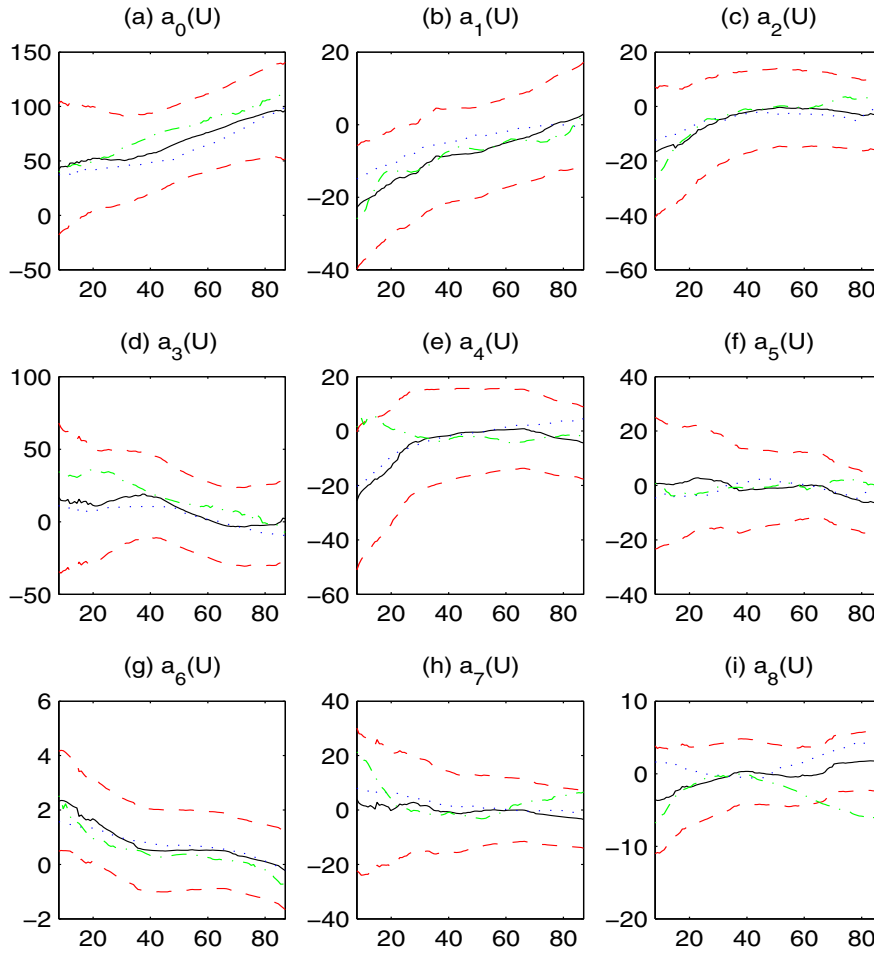


FIG 6. The estimated coefficient functions for the three quantiles: $\tau = 0.25$ (blue dotted curve), $\tau = 0.50$ (black solid curve), and $\tau = 0.95$ (green dash-dotted curve). The red dashed curve represents the 95% pointwise confidence intervals based on the $\tau = 0.5$ quantile estimator.

Appendix: Proofs of theorems

This section provides the proofs of the main results in the paper. For notational convenience, let us define

$$g_i(\boldsymbol{\theta}, \pi_w) = \frac{\delta_i \mathbf{X}_i^*}{\pi_w(Y_i | \mathbf{W}_i)} \phi_\tau(Y_i - \boldsymbol{\theta}^T \mathbf{X}_i^*) K_h(U_i - u_0), G_n(\boldsymbol{\theta}, \pi_w) = \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\theta}, \pi_w),$$

where $\phi_\tau(x) = I(x \leq 0) - \tau$. Let $\boldsymbol{\alpha}_0$ belong in a compact nuisance parameter space Θ , $\Theta \in R^{2p}$ be the compact parameter space of $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_0$ belong in the interior of Θ . Our proofs of results require the following technical conditions:

- (A₁) $a_j(u)$ is twice continuously differentiable in a neighborhood of u_0 , $j = 1, 2, \dots, p$.
- (A₂) The density function $f_U(u)$ of U is positive, and has a continuous second derivative on its bounded support \mathcal{U} .
- (A₃) The covariate \mathbf{X} has bounded support.
- (A₄) $K(\cdot)$ is a symmetric density function with bounded support, and satisfies the Lipschitz condition.
- (A₅) The conditional density function of the random errors $f_{\epsilon_i}(\cdot|\mathbf{X}, U)$, $i = 1, \dots, n$ are continuous, and bounded away from zero.
- (A₆) There exist $A(u)$ and $V(u)$ such that $A_n(u) \rightarrow A(u)$ and $V_n(u) \rightarrow V(u)$. Moreover, $A(u)$ is nonsingular and continuous in \mathcal{N}_{u_0} , and $V(u)$ is continuous in \mathcal{N}_{u_0} .
- (A₇) $\int_0^{t_0} \lambda_0(t) dt < \infty$.
- (A₈) Let \mathcal{N}_{α_0} be a neighborhood of α_0 , and $s^{(0)}$, $s^{(1)}$, and $s^{(2)}$ be scalar, vector and matrix functions respectively defined on $\mathcal{N}_{\alpha_0} \times [0, t_0]$ for $S^{(k)}(t; \alpha)$ (defined in Section 2.3). As $n \rightarrow \infty$,

$$\sup_{t \in [0, t_0], \alpha \in \mathcal{N}_{\alpha_0}} \|S^{(k)}(t; \alpha) - s^{(k)}(t; \alpha)\| \xrightarrow{P} 0,$$

for $k = 0, 1, 2$.

- (A₉) There exists $\delta > 0$ such that as $n \rightarrow \infty$,

$$n^{-\frac{1}{2}} \sup_{1 \leq i \leq n, 0 \leq t \leq t_0} |Z_i| Y_i(t) I_{\{\alpha_0^T \mathbf{W}_i > -\delta |\mathbf{W}_i|\}} \xrightarrow{P} 0.$$

- (A₁₀) Consider \mathcal{N}_{α_0} and $s^{(k)}$, $k = 0, 1, 2$, in (A₈). Define $\bar{\mathbf{w}} = s^{(1)}/s^{(0)}$, $v(t; \alpha) = \{s^{(2)}(t; \alpha)/s^{(0)}(t; \alpha) - \bar{\mathbf{w}}^{\otimes 2}(t; \alpha)\}$, and $\Omega = n^{-1} \sum_{i=1}^n \int_0^{t_0} v(t; \alpha) dN_i^C(t)$. For all $t \in [0, t_0]$, $\alpha \in \mathcal{N}_{\alpha_0}$,

$$\frac{\partial}{\partial \alpha} s^{(0)}(t; \alpha) = s^{(1)}(t; \alpha), \text{ and } \frac{\partial}{\partial \alpha^2} s^{(0)}(t; \alpha) = s^{(2)}(t; \alpha).$$

- (A₁₁) Consider \mathcal{N}_{α_0} and $s^{(k)}$ in (A₈). All of $s^{(k)}$'s, $k = 0, 1, 2$, are bounded, $s^{(0)}$ is bounded away from 0 on $\mathcal{N}_{\alpha_0} \times [0, t_0]$, and the family of $s^{(k)}(\cdot, t)$, $0 \leq t \leq t_0$, is an equicontinuous family at α_0 .
- (A₁₂) The matrix $\int_0^{t_0} v(t; \alpha) s^{(0)}(t; \alpha_0) \lambda_0(t) dt$ is positive definite.

Some interpretations of these conditions are in order. Conditions (A₁) and (A₂) are required for the smoothness of some of the functions used in the proofs. Conditions (A₃) and (A₄) are standard conditions for kernel analysis, while conditions (A₅)–(A₆) are commonly used for quantile regression. Note that (A₄) implies that $\mu_k = 0$ for any odd integer k and $\mu_0 = 1$. Conditions (A₇)–(A₁₂) are adopted from [2] – these conditions ensure the asymptotic normality and weak convergence of $\hat{\alpha}$ and $\hat{\Lambda}_0(t)$, the latter being the baseline cumulative hazard function of the Cox model in Section 2.3. Readers may consult [2] or [25] for detailed discussions of these conditions.

We introduce the following lemmas to facilitate the proofs of results.

Lemma 1. Suppose that $U_i, i = 1, 2, \dots, n$ are an i.i.d. random variables, $\mathbf{Z}_i, i = 1, 2, \dots, n$ are independent random variables. $E|\varphi(\mathbf{Z}_i, U)|^s < \infty$ and $\sup_{\mathbf{z}} \int |\varphi(\mathbf{z}, u)|^s f_i(\mathbf{z}, u) du < \infty$, where f_i denote the joint density of \mathbf{Z}_i and U_i for $i = 1, \dots, n$. Let K be a bounded positive function with a bounded support satisfying the Lipschitz condition. Given that $n^{2\delta-1}h \rightarrow \infty$ for some $\delta < 1 - s^{-1}$, we have

$$\sup_{u \in \mathcal{U}} \left| \frac{1}{n} \sum_{i=1}^n \left[K_h(U_i - u) \varphi(\mathbf{Z}_i, U_i) - E(K_h(U_i - u) \varphi(\mathbf{Z}_i, U_i)) \right] \right| = O_p(c_{n1}),$$

where $c_{n1} = (\frac{\log h^{-1}}{nh})^{1/2}$.

Proof. This Lemma is similar to Lemma 7.1 of [20]. Hence the result follows an argument similar to that of Lemma 7.1 of [20]. \square

Lemma 2. Assume that $(A_7)-(A_{12})$ hold. Then we have

$$\begin{aligned} & \frac{\pi_{0w}(Y|\mathbf{W}) - \hat{\pi}_w(Y|\mathbf{W})}{\pi_{0w}(Y|\mathbf{W})} \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(\boldsymbol{\alpha}_0^T \mathbf{W})}{\pi_{0w}(Y|\mathbf{W}) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^Y S_C(u|\mathbf{W}) du dM_i(s) \\ & \quad + \frac{B^T \Omega^{-1}}{\pi_{0w}(Y|\mathbf{W})} \frac{1}{n} \sum_{i=1}^n \int_0^{t_0} \{ \mathbf{W}_i - \bar{\mathbf{w}}(s) \} dM_i(s) + o_p(n^{-\frac{1}{2}}). \end{aligned}$$

Proof. By [25] p. 299, we have

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = \Omega^{-1} \frac{1}{n} \sum_{i=1}^n \int_0^{t_0} \{ \mathbf{W}_i - \bar{\mathbf{w}}(t) \} dM_i(t) + o_p(n^{-1/2}),$$

and

$$\hat{\Lambda}_0(t) - \Lambda_0(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{dM_i(u)}{s^{(0)}(s; \boldsymbol{\alpha}_0)} - \int_0^t \bar{\mathbf{w}}^T(s) d\Lambda_0(s) (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + o_p(n^{-1/2}).$$

It is straightforward to obtain Lemma 2 by these martingale representations and the Functional Delta method. \square

Lemma 3. Assume that $(A_7)-(A_{12})$ hold. Then we have

$$G_n(\boldsymbol{\theta}, \hat{\pi}_w) = \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\theta}, \pi_{0w}) + o_p((nh)^{-\frac{1}{2}}).$$

Proof. By noting that

$$\begin{aligned} & G_n(\boldsymbol{\theta}, \hat{\pi}_w) \\ &= \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\theta}, \pi_{0w}) \left\{ 1 + \frac{\pi_{0w}(Y_i|\mathbf{W}_i) - \hat{\pi}_w(Y_i|\mathbf{W}_i)}{\pi_{0w}(Y_i|\mathbf{W}_i)} (1 + o_p(1)) \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\theta}, \pi_{0w}) \left[\frac{1}{n} \sum_{j=1}^n \int_0^{t_0} \frac{\exp(\boldsymbol{\alpha}_0^T \mathbf{W}_i) \int_s^{Y_i} S_C(u|\mathbf{W}_i) du}{\pi_{0w}(Y_i|\mathbf{W}_i) s^{(0)}(s; \boldsymbol{\alpha}_0)} dM_j(s) \right] \\
&\quad + \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\theta}, \pi_{0w}) \left[\frac{B_i^T \Omega^{-1}}{\pi_{0w}(Y_i|\mathbf{W}_i)} \frac{1}{n} \sum_{j=1}^n \int_0^{t_0} \{\mathbf{W}_j - \bar{\mathbf{w}}(s)\} dM_j(s) \right] \\
&\quad + \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\theta}, \pi_{0w}) + o_p((nh)^{-\frac{1}{2}}) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_0^{t_0} \frac{g_i(\boldsymbol{\theta}, \pi_{0w}) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_i)}{\pi_{0w}(Y_i|\mathbf{W}_i) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^{Y_i} S_C(u|\mathbf{W}_i) dudM_j(s) \\
&\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{g_i(\boldsymbol{\theta}, \pi_{0w}) B_i^T \Omega^{-1}}{\pi_{0w}(Y_i|\mathbf{W}_i)} \int_0^{t_0} \{\mathbf{W}_j - \bar{\mathbf{w}}(s)\} dM_j(s) \\
&\quad + \frac{1}{n} \sum_{i=1}^n g_i(\boldsymbol{\theta}, \pi_{0w}) + o_p((nh)^{-\frac{1}{2}}) \\
&=: I_{n1} + I_{n2} + I_{n3} + o_p((nh)^{-\frac{1}{2}}),
\end{aligned}$$

where

$$B_i = \int_0^{Y_i} S_C(u|\mathbf{W}_i) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_i) \int_0^u (\mathbf{W}_i - \bar{\mathbf{w}}(s)) d\Lambda_0(s) du.$$

To prove this lemma, we only need to prove $I_{n2} = o_p((nh)^{-\frac{1}{2}})$ and $I_{n3} = o_p((nh)^{-\frac{1}{2}})$. To prove the former, let $S_i = \{Y_i, A_i, \delta_i, \mathbf{X}_i, U_i\}$, and define

$$\begin{aligned}
\mathcal{K}(S_i, S_j) &= \frac{1}{2} \int_0^{t_0} \frac{g_i(\boldsymbol{\theta}, \pi_{0w}) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_i)}{\pi_{0w}(Y_i|\mathbf{W}_i) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^{Y_i} S_C(u|\mathbf{W}_i) dudM_j(s) \\
&\quad + \frac{1}{2} \int_0^{t_0} \frac{g_j(\boldsymbol{\theta}, \pi_{0w}) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_j)}{\pi_{0w}(Y_j|\mathbf{W}_j) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^{Y_j} S_C(u|\mathbf{W}_j) dudM_i(s).
\end{aligned}$$

It is readily seen that $\mathcal{K}(S_i, S_j)$ is a symmetric function. We then have

$$I_{n2} = \frac{1}{n^2} \sum_{i=1}^n \mathcal{K}(S_i, S_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{K}(S_i, S_j) =: I_{n21} + I_{n22}.$$

It is straightforward to show that $E\mathcal{K}(S_i, S_i) = 0$ and $Var(I_{n21}) = O(n^{-3})$, and thus $I_{n21} = o_p((nh)^{-\frac{1}{2}})$. Hence, it suffices to consider the U-statistic I_{n22} only. Let $\mathcal{K}(S_i) = E[\mathcal{K}(S_i, S_j)|S_i]$, we have

$$\begin{aligned}
\mathcal{K}(S_i) &= \frac{1}{2} \int_0^{t_0} \frac{g_i(\boldsymbol{\theta}, \pi_{0w}) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_i)}{\pi_{0w}(Y_i|\mathbf{W}_i) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^{Y_i} S_C(u|\mathbf{W}_i) du E[dM_j(s)] \\
&\quad + \frac{1}{2} \int_0^{t_0} E \left\{ \frac{g_j(\boldsymbol{\theta}, \pi_{0w}) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_j)}{\pi_{0w}(Y_j|\mathbf{W}_j) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^{Y_j} S_C(u|\mathbf{W}_j) du \right\} dM_i(s)
\end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2} \int_0^{t_0} \left\{ f_U(u_0) \begin{pmatrix} 1 \\ \mu_1 \end{pmatrix} \right. \\
 &\quad \left. \otimes \frac{1}{n} \sum_{j=1}^n E \left[\frac{\mathbf{X}_j \phi_\tau(\epsilon_j) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_j) \int_s^{Y_j} S_C(u | \mathbf{W}_j) du}{\mu(\mathbf{X}_i, U_i) \pi_{0w}(Y_j | \mathbf{W}_j) s^{(0)}(s; \boldsymbol{\alpha}_0)} \middle| U_j \right] \right\} dM_i(s),
 \end{aligned}$$

which implies that the U-projection of I_{n22} is

$$\hat{I}_{n22} = \frac{2}{n} \sum_{i=1}^n \mathcal{K}(S_i) = \frac{1}{n} \sum_{i=1}^n \int_0^{t_0} D_1(\boldsymbol{\theta}, s) dM_i(s),$$

where

$$\begin{aligned}
 D_1(\boldsymbol{\theta}, s) &= f_U(u_0) \begin{pmatrix} 1 \\ \mu_1 \end{pmatrix} \\
 &\quad \otimes \frac{1}{n} \sum_{j=1}^n E \left\{ \frac{\mathbf{X}_j \phi_\tau(\epsilon_j) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_j) \int_s^{Y_j} S_C(u | \mathbf{W}_j) du}{\mu(\mathbf{X}_j, U_j) \pi_{0w}(Y_j | \mathbf{W}_j) s^{(0)}(s; \boldsymbol{\alpha}_0)} \middle| U_j \right\}.
 \end{aligned}$$

Note that $E\mathcal{K}(S_i, S_j) = E\mathcal{K}(S_i) = 0$. Recognizing that S_i independent of S_j for any $i \neq j$, by some calculations, we obtain

$$\begin{aligned}
 &Var(\mathcal{K}(S_i, S_j)) \\
 &= \frac{1}{4} E \left\{ \int_0^{t_0} \left[\frac{g_i(\boldsymbol{\theta}, \pi_{0w}) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_i)}{\pi_{0w}(Y_i | \mathbf{W}_i) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^{Y_i} S_C(u | \mathbf{W}_i) du \right]^{\otimes 2} \right. \\
 &\quad \left. \times I(Y_j - A_j \geq s) \exp\{\boldsymbol{\alpha}_0^T \mathbf{W}_j\} \lambda_0(s) ds \right\} \\
 &\quad + \frac{1}{4} E \left\{ \int_0^{t_0} \left[\frac{g_j(\boldsymbol{\theta}, \pi_{0w}) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_j)}{\pi_{0w}(Y_j | \mathbf{W}_j) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^{Y_j} S_C(u | \mathbf{W}_j) du \right]^{\otimes 2} \right. \\
 &\quad \left. \times I(Y_i - A_i \geq s) \exp\{\boldsymbol{\alpha}_0^T \mathbf{W}_i\} \lambda_0(s) ds \right\} \\
 &= \frac{1}{2} \int_0^{t_0} E \left\{ \left[\frac{g_i(\boldsymbol{\theta}, \pi_{0w}) \exp(\boldsymbol{\alpha}_0^T \mathbf{W}_i)}{\pi_{0w}(Y_i | \mathbf{W}_i) s^{(0)}(s; \boldsymbol{\alpha}_0)} \int_s^{Y_i} S_C(u | \mathbf{W}_i) du \right]^{\otimes 2} \right\} \\
 &\quad \times E \left\{ I(Y_j - A_j \geq s) \exp\{\boldsymbol{\alpha}_0^T \mathbf{W}_j\} \right\} \lambda_0(s) ds \\
 &= \frac{1}{2h} \int_0^{t_0} f_U(u_0) \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \\
 &\quad i \otimes E \left[\frac{\mathbf{X}_i \mathbf{X}_i^T \phi_\tau^2(\epsilon_i) \exp(2\boldsymbol{\alpha}_0^T \mathbf{W}_i)}{\mu(\mathbf{X}_i, U_i) \pi_{0w}^3(Y_i | \mathbf{W}_i) [s^{(0)}(s; \boldsymbol{\alpha}_0)]^2} \left(\int_s^{Y_i} S_C(u | \mathbf{W}_i) du \right)^2 \middle| U_i \right] \\
 &\quad \times E \left\{ S_C(s | W_j) \exp\{\boldsymbol{\alpha}_0^T \mathbf{W}_j\} \int_s^\infty \int_0^{Y_j} \frac{f_T(a + v | \mathbf{X}_j, U_j)}{\mu(\mathbf{X}_j, U_j)} da dv \right\} \lambda_0(s) ds.
 \end{aligned}$$

It can be shown that

$$E\{I_{n22} - \hat{I}_{n22}\}^{\otimes 2} = \frac{Var(\mathcal{K}(S_i, S_j))}{n(n-1)} + O(n^{-2}).$$

Using these results, it is readily seen that $E\{I_{n22} - \hat{I}_{n22}\}^{\otimes 2} = O((n^2h)^{-1})$. In other words, $\sqrt{n}I_{n22}$ and its projection $\sqrt{n}\hat{I}_{n22}$ are asymptotically mean square equivalent. Hence, $I_{n22} = O_p(n^{-1/2}) = o_p((nh)^{-1/2})$. This implies $I_{n2} = O_p(n^{-1/2}) = o_p((nh)^{-1/2})$. By a similar argument, we can show that $I_{n3} = O_p(n^{-1/2}) = o_p((nh)^{-1/2})$. This completes the proof of this lemma. \square

Now, let $r_i(u_0) = \mathbf{X}_i^T(\mathbf{a}(U_i) - \mathbf{b}_0(u_0) - \mathbf{c}_0(u_0)(U_i - u_0))$, $\eta_i = I(\epsilon_i \leq 0) - \tau$, $\boldsymbol{\theta}^* = \sqrt{nh}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, and $\Delta_i = \mathbf{X}_i^{*T}\boldsymbol{\theta}^*/\sqrt{nh}$. Note that minimizing $L_n(\boldsymbol{\theta}, \hat{\pi}_w)$ is equivalent to minimizing

$$\tilde{L}(\boldsymbol{\theta}, \hat{\pi}_w) = hL_n(\boldsymbol{\theta}, \hat{\pi}_w) - hL_n(\boldsymbol{\theta}_0, \hat{\pi}_w),$$

because $L_n(\boldsymbol{\theta}_0, \hat{\pi}_w)$, where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$, does not depend on $\boldsymbol{\theta}$.

Lemma 4. Assume that (A₁)–(A₅) hold, and $h = h_n \rightarrow 0, nh \rightarrow \infty$ as $n \rightarrow \infty$. We have

$$\tilde{L}_n(\boldsymbol{\theta}, \hat{\pi}_w) = \sqrt{nh}G_n^T(\boldsymbol{\theta}_0, \hat{\pi}_w)\boldsymbol{\theta}^* + \frac{1}{2}\boldsymbol{\theta}^{*T}A(u_0)\boldsymbol{\theta}^* + o_p(1),$$

uniformly for $\boldsymbol{\theta} \in \Theta$.

Proof. Note that $Y_i - \boldsymbol{\theta}^T\mathbf{X}_i^* = \epsilon_i + r_i(u_0) - \Delta_i$, and $Y_i - \boldsymbol{\theta}_0^T\mathbf{X}_i^* = \epsilon_i + r_i(u_0)$. Thus,

$$\begin{aligned} \tilde{L}_n(\boldsymbol{\theta}, \hat{\pi}_w) &= h[L_n(\boldsymbol{\theta}, \hat{\pi}_w) - L_n(\boldsymbol{\theta}_0, \hat{\pi}_w)] \\ &= \sum_{i=1}^n \frac{\delta_i K((U_i - u_0)/h)}{\hat{\pi}_w(Y_i|\mathbf{W}_i)} [\rho_\tau(\epsilon_i + r_i(u_0) - \Delta_i) - \rho_\tau(\epsilon_i + r_i(u_0))]. \end{aligned}$$

Using the identity ([31])

$$\rho_\tau(x - y) - \rho_\tau(x) = y\{I(x \leq 0) - \tau\} + \int_0^y \{I(x \leq t) - I(x \leq 0)\}dt,$$

we can write

$$\begin{aligned} \tilde{L}_n(\boldsymbol{\theta}, \hat{\pi}_w) &= \sum_{i=1}^n \frac{\delta_i K((U_i - u_0)/h)}{\hat{\pi}_w(Y_i|\mathbf{W}_i)} \left\{ \Delta_i [I(\epsilon_i + r_i(u_0)) \leq 0] - \tau \right\} \\ &\quad + \sum_{i=1}^n \frac{\delta_i K((U_i - u_0)/h)}{\hat{\pi}_w(Y_i|\mathbf{W}_i)} \left\{ \int_0^{\Delta_i} [I(\epsilon_i + r_i(u_0) \leq t) - I(\epsilon_i + r_i(u_0) \leq 0)] dt \right\} \\ &=: J_0 + J_1. \end{aligned}$$

From the definition of $G_n(\cdot, \cdot)$, we have

$$J_0 = \sqrt{nh}G_n^T(\boldsymbol{\theta}_0, \hat{\pi}_w)\boldsymbol{\theta}^* + o_p(1).$$

Now, write $J_1 = E(J_1) + R_n(\boldsymbol{\theta})$, where

$$\begin{aligned} & EJ_1 \\ &= \frac{n}{n} \sum_{i=1}^n E \left\{ \frac{\delta_i K((U_i - u_0)/h)}{\hat{\pi}_{0w}(Y_i | \mathbf{W}_i)} \int_0^{\Delta_i} [I(\epsilon_i + r_i(u_0) \leq t) - I(\epsilon_i + r_i(u_0) \leq 0)] dt \right\} \\ &= \frac{n}{n} \sum_{i=1}^n E \left\{ \frac{\delta_i K((U_i - u_0)/h)}{\pi_{0w}(Y_i | \mathbf{W}_i)} \int_0^{\Delta_i} [I(\epsilon_i + r_i(u_0) \leq t) - I(\epsilon_i + r_i(u_0) \leq 0)] dt \right\} \\ &\quad + o_p(1). \end{aligned}$$

By the joint distribution of (A, V) , note that $\pi_{0w}(t | \mathbf{W}) = \int_0^t S_C(u | \mathbf{W}) du$, we have

$$\begin{aligned} EJ_1 &= \frac{n}{n} \sum_{i=1}^n E \left\{ E \left[\frac{\delta_i K((U_i - u_0)/h)}{\pi_{0w}(Y_i | \mathbf{W}_i)} \int_0^{\Delta_i} [I(\epsilon_i + r_i(u_0) \leq t) \right. \right. \\ &\quad \left. \left. - I(\epsilon_i + r_i(u_0) \leq 0)] dt \middle| \mathbf{X}_i, U_i \right] \right\} + o_p(1) \\ &= \frac{n}{n} \sum_{i=1}^n E \left\{ \left[\int_0^\infty \int_0^y \frac{f_{\bar{T}_i}(y | \mathbf{x}, u) S_C(y - a | \mathbf{w})}{\pi_{0w}(y | \mathbf{w})} \int_0^{\Delta_i} [I(\epsilon_i + r_i(u_0) \leq t) \right. \right. \\ &\quad \left. \left. - I(\epsilon_i + r_i(u_0) \leq 0)] dt da dy \right] \times \frac{K((U_i - u_0)/h)}{\mu(\mathbf{X}_i, U_i)} \right\} + o_p(1) \\ &= \frac{n}{n} \sum_{i=1}^n E \left\{ \frac{K((U_i - u_0)/h)}{\mu(\mathbf{X}_i, U_i)} E \left[\int_0^{\Delta_i} [I(\epsilon_i + r_i(u_0) \leq t) \right. \right. \\ &\quad \left. \left. - I(\epsilon_i + r_i(u_0) \leq 0)] dt \middle| \mathbf{X}_i, U_i \right] \right\} + o_p(1) \\ &= \frac{1}{2} \boldsymbol{\theta}^{*T} \frac{1}{n} \sum_{i=1}^n E \left\{ \frac{K_h(U_i - u_0)}{\mu(\mathbf{X}_i, U_i)} f_{\epsilon_i}(0 | \mathbf{X}_i, U_i) \mathbf{X}_i^* \mathbf{X}_i^{*T} \right\} \boldsymbol{\theta}^* + o_p(1) \\ &=: \frac{1}{2} \boldsymbol{\theta}^{*T} A_n^*(u_0) \boldsymbol{\theta}^* + o_p(1), \tag{A.1} \end{aligned}$$

and $R_n(\boldsymbol{\theta}) = J_1 - E(J_1)$. By Lemma 1, it can be readily shown that $|R_n(\boldsymbol{\theta})| = O_p(\frac{\log h^{-1}}{nh})^{1/2} = o_p(1)$. Note that $\mu_1 = 0$ and for any integer k ,

$$E \left[K_h\left(\frac{U_i - u_0}{h}\right) \left(\frac{U_i - u_0}{h}\right)^k \right] = f_U(u_0) \mu_k + h f'_U(u_0) \mu_{k+1} + O(h^2) = f_U(u_0) \mu_k + o(1),$$

where f'_U is the first derivative of f_U . Note that $\mu_0 = 1$ and $\mu_1 = 0$. Hence we have

$$\begin{aligned} & A_n^*(u_0) \\ &= \frac{1}{n} \sum_{i=1}^n E \left\{ E \left[\frac{K_h(U_i - u)}{\mu(\mathbf{X}_i, U_i)} f_{\epsilon_i}(0 | \mathbf{X}_i, U_i) \mathbf{X}_i^* \mathbf{X}_i^{*T} \right] \middle| U_i \right\} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{n} \sum_{i=1}^n E \left\{ K_h(U_i - u) \left(\begin{array}{cc} 1 & \left(\frac{U_i - u_0}{h}\right) \\ \left(\frac{U_i - u_0}{h}\right) & \left(\frac{U_i - u_0}{h}\right)^2 \end{array} \right) \otimes E \left[\frac{f_{\epsilon_i}(0|\mathbf{X}_i, U_i)}{\mu(\mathbf{X}_i, U_i)} \mathbf{X}_i \mathbf{X}_i^T \right] \middle| U_i \right\} \\
 &= f_U(u_0) \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \frac{1}{n} \sum_{i=1}^n E \left[\frac{\mathbf{X}_i \mathbf{X}_i^T}{\mu(\mathbf{X}_i, U_i)} f_{\epsilon_i}(0|\mathbf{X}_i, U_i) \middle| U_i = u_0 \right] + o(1) \\
 &= f_U(u_0) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \otimes \Gamma_n(u_0) + o(1). \tag{A.2}
 \end{aligned}$$

Combining (A.1) and (A.2), we obtain

$$J_1 = \frac{1}{2} \boldsymbol{\theta}^{*T} A(u_0) \boldsymbol{\theta}^* + o_p(1).$$

It is readily seen that $\sqrt{nh}G_n(\boldsymbol{\theta}_0, \hat{\pi}_w)$ has a bounded second moment and is thus stochastically bounded. As the convex function $\tilde{L}_n(\boldsymbol{\theta}, \hat{\pi}_w)$ converges in probability to the convex function $(\sqrt{nh}G_n(\boldsymbol{\theta}_0, \hat{\pi}_w))^T \boldsymbol{\theta}^* + \frac{1}{2} \boldsymbol{\theta}^{*T} A(u_0) \boldsymbol{\theta}^*$, by the Convexity Lemma ([36]), for any compact set K , $\sup_{\boldsymbol{\theta} \in K} |R_n(\boldsymbol{\theta})| = o_p(1)$. Hence we have $\sup_{\boldsymbol{\theta} \in \Theta} |R_n(\boldsymbol{\theta})| = o_p(1)$. This implies that the quadratic approximation to the convex function $\tilde{L}_n(\boldsymbol{\theta}, \hat{\pi}_w)$ holds uniformly for $\boldsymbol{\theta}$ in the compact set Θ . This completes the proof of this Lemma. \square

Lemma 5. Assume that (A1)–(A5) hold, and $h = h_n \rightarrow 0, nh \rightarrow \infty$ as $n \rightarrow \infty$. For $\hat{\boldsymbol{\theta}}^* = \sqrt{nh}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, we have

$$\hat{\boldsymbol{\theta}}^* = -A^{-1}(u_0)\sqrt{nh}G_n(\boldsymbol{\theta}_0, \hat{\pi}_w) + o_p(1).$$

Proof. Note that $\hat{\boldsymbol{\theta}}^*$ is the minimizer of $\tilde{L}_n(\boldsymbol{\theta}, \hat{\pi}_w)$. By Lemma 4 and similar argument of Theorem 2 of [19], the result of this Lemma is readily obtained. \square

Proof of Theorem 3.1. By Lemmas 4 and 5 and the Convexity Lemma ([36]), it can be readily shown that $\hat{\boldsymbol{\theta}}^* = O_P(1)$, meaning that $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p((nh)^{-1/2})$. Hence

$$\mathbf{H}(\hat{\boldsymbol{\beta}}(u_0) - \boldsymbol{\beta}_0(u_0)) \xrightarrow{P} 0.$$

This completes the proof of this theorem. \square

Proof of Theorem 3.2. Note that, by Lemma 3 and Lemma 5, we have

$$\begin{aligned}
 \sqrt{nh}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= -A^{-1}(u_0)\sqrt{nh}G_n(\boldsymbol{\theta}_0, \hat{\pi}_w) + o_p(1) \\
 &= -A^{-1}(u_0)\sqrt{\frac{h}{n}} \sum_{i=1}^n g_i(\boldsymbol{\theta}_0, \pi_{0w}) + o_p(1).
 \end{aligned}$$

Therefore, to obtain the desired asymptotic normality result by the Central Limit Theorem, it suffices to examine only the mean of $g_i(\boldsymbol{\theta}_0, \pi_{0w})$ and covariance of $\sqrt{\frac{h}{n}} \sum_{i=1}^n g_i(\boldsymbol{\theta}_0, \pi_{0w})$. Now, denote

$$g_i(\mathbf{a}(U), \pi_{0w}) = \frac{\delta_i \mathbf{X}_i^* K_h(U_i - u_0)}{\pi_{0w}(Y_i | \mathbf{W}_i)} (I(Y_i - \mathbf{a}(U_i)^T \mathbf{X}_i < 0) - \tau).$$

As $Eg_i(\mathbf{a}(U), \pi_{0w}) = 0$ and $\mu_3 = 0$, we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n Eg_i(\boldsymbol{\theta}_0, \pi_{0w}) \\ &= \frac{1}{n} \sum_{i=1}^n E \left\{ g_i(\boldsymbol{\theta}_0, \pi_{0w}) - g_i(\mathbf{a}(U), \pi_{0w}) \right\} \\ &= \frac{1}{n} \sum_{i=1}^n E \left\{ \frac{\delta \mathbf{X}_i^* K_h(U_i - u_0)}{\pi_{0w}(Y_i | \mathbf{W}_i)} \left[I(Y_i - \boldsymbol{\theta}_0^T \mathbf{X}_i^* < 0) - I(Y_i - \mathbf{a}(U_i)^T \mathbf{X}_i < 0) \right] \right\} \\ &= \frac{h^2}{2} f_U(u_0) \Gamma_n(u_0) \begin{pmatrix} \mu_2 \mathbf{a}''(u_0) \\ \mu_3 \mathbf{a}''(u_0) \end{pmatrix} (1 + o(1)) \\ &= \frac{h^2}{2} f_U(u_0) \Gamma_n(u_0) \begin{pmatrix} \mu_2 \mathbf{a}''(u_0) \\ 0 \end{pmatrix} (1 + o(1)). \end{aligned}$$

By some calculations, we obtain

$$\text{Var} \left[\sqrt{\frac{h}{n}} \sum_{i=1}^n g_i(\boldsymbol{\theta}_0, \pi_{0w}) \right] = V + o(1).$$

Thus, by the Central Limit Theorem, we can show that the local linear estimator is asymptotically normal. This completes the proof of Theorem 3.2. \square

References

- [1] ADDONA, V. and WOLFSON, D. B. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Anal.* **12** 267–284. [MR2328577](#)
- [2] ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. [MR0673646](#)
- [3] ASGHARIAN, M., M'LAN, C. M. and WOLFSON, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *J. Amer. Statist. Assoc.* **97** 201–209. [MR1947280](#)
- [4] ASGHARIAN, M. and WOLFSON, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Ann. Statist.* **33** 2109–2131. [MR2211081](#)
- [5] BANG, H. and TSIATIS, A. A. (2002). Median regression with censored cost data. *Biometrics* **58** 643–649. [MR1926117](#)
- [6] BRESLOW, N. E. (1972). Discussion of the paper by D. R. Cox. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34** 216–217.
- [7] CAI, J., FAN, J., JIANG, J. and ZHOU, H. (2007). Partially linear hazard regression for multivariate survival data. *J. Amer. Statist. Assoc.* **102** 538–551. [MR2370851](#)

- [8] CAI, J., FAN, J., JIANG, J. and ZHOU, H. (2008). Partially linear hazard regression with varying coefficients for multivariate survival data. *J. R. Stat. Soc. Ser. B Stat. Methodol* **70** 141–158. [MR2412635](#)
- [9] CAI, J., FAN, J., ZHOU, H. and ZHOU, Y. (2007a). Hazard models with varying coefficient for multivariate failure time data. *Ann. Statist.* **35** 324–354. [MR2332278](#)
- [10] CAI, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* **136** 163–188. [MR2328589](#)
- [11] CAI, Z., FAN, J. and YAO, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **95** 941–956. [MR1804449](#)
- [12] CAI, Z. and XU, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Amer. Statist. Assoc.* **103** 1595–1608. [MR2504207](#)
- [13] CHEN, R. and TSAY, R. S. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88** 298–308. [MR1212492](#)
- [14] CHEN, X. and ZHOU, Y. (2012). Quantile regression for right-censored and length-biased Data. *Acta Mathematicae Applicatae Sinica (English Series)* **28** 443–462. [MR2930999](#)
- [15] CLEVELAND, W. S., GROSS, E. and SHHYU, W. M. (1992). Local regression models. In *Statistical Models in S* (eds. J. M. Chambers and T. J. Hastie), Pacific Grove, CA: Wadsworth and Brooks, 309–376.
- [16] EFRON, B. (1979). Bootstrap methods: Another Look at the Jackknife. *Ann. Statist.* **7** 1–26. [MR0515681](#)
- [17] EFRON, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **72** 316–319. [MR0624333](#)
- [18] FAN, J. and FAN, Y. (2006). Comment on “Quantile Autoregression”. *J. Amer. Statist. Assoc.* **101** 991–994. [MR2324110](#)
- [19] FAN, J., HECKMAN, N. and WAND, M. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.* **90** 141–150. [MR1325121](#)
- [20] FAN, J. and HUANG, T. (2005). Profile likelihood inferences on semiparametric varying coefficient partially linear models. *Bernoulli* **11** 1031–1057. [MR2189080](#)
- [21] FAN, J. and LI, R. Z. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J. Amer. Statist. Assoc.* **99** 710–723. [MR2090905](#)
- [22] FAN, J., LIN, H. and ZHOU, Y. (2006). Local partial-likelihood estimation for lifetime data. *Ann. Statist.* **34** 290–325. [MR2275243](#)
- [23] FAN, J. and ZHANG, J. T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B Stat. Methodol* **62** 303–322. [MR1749541](#)
- [24] FAN, J. and ZHANG, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface* **1** 179–195. [MR2425354](#)
- [25] FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley: New York. [MR1100924](#)

- [26] HASTIE, T. and TIBSHIRANI, R. (1993). Varying-coefficient models. *J. R. Stat. Soc. Ser. B Stat. Methodol* **55** 757–796. [MR1229881](#)
- [27] HONDA, T. (2004). Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference* **121** 113–125. [MR2027718](#)
- [28] HONORÉ, B., KHAN, S. and POWELL, J. L. (2002). Quantile regression under random censoring. *J. Econometrics* **109** 67–105. [MR1899693](#)
- [29] HOOVER, D. R., RICE, J. A., WU, C. O. and YANG, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. [MR1666699](#)
- [30] KIM, M. O. (2007). Quantile regression with varying-coefficients. *Ann. Statist.* **35** 92–108. [MR2332270](#)
- [31] KNIGHT, K. (1998). Limiting distribution for L_1 regression estimator under general conditions. *Ann. Statist.* **26** 755–770. [MR1626024](#)
- [32] KOENKER, R. and GELING, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *J. Amer. Statist. Assoc* **96** 458–468. [MR1939348](#)
- [33] LIN, D. Y. (2007). On the Breslow estimator. *Lifetime Data Anal.* **13** 471–480. [MR2416534](#)
- [34] LIN, D. Y., WEI, L. J. and YING, Z. L. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80** 557–572. [MR1248021](#)
- [35] NING, J., QIN, J. and SHEN, Y. (2011). Buckley-James-type estimator with right-censored and length-biased Data. *Biometrics* **67** 1369–1378. [MR2872387](#)
- [36] POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199. [MR1128411](#)
- [37] QIN, J. and SHEN, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox Model. *Biometrics* **66** 382–392. [MR2758818](#)
- [38] REDELMEIER, D. A. and SINGH, S. M. (2001). Survival in academy award winning actors and actresses. *Ann. Intern. Med.* **134** 955–962.
- [39] SHEN, Y., NING, J. and QIN, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *J. Amer. Statist. Assoc.* **104** 1192–1202. [MR2750244](#)
- [40] SYLVESTRE, M. P., HUSZTI, E. and HANLEY, J. A. (2006). Do Oscar winners live longer than less successful peers? A reanalysis of the evidence. *Ann. Intern. Med.* **145** 361–363.
- [41] VARDI, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10** 616–620. [MR0653536](#)
- [42] VARDI, Y. (1985). Empirical distribution in selection bias models. *Ann. Statist.* **13** 178–203. [MR0773161](#)
- [43] WANG, H. and WANG, L. (2011). Quantile regression analysis of length-biased survival data. *Stat.* **3** 31–47.
- [44] WOLKEWITZ, M., ALLIGNOL, A., SCHUMACHER, M. and BEYERSMANN, J. (2010). Two pitfalls in survival analyses of time-dependent expo-

- sure: A case study in a cohort of Oscar nominees. *The Amer. Statistician* **64** 205–211. [MR2757164](#)
- [45] WU, Y., FAN, J. Q. and MÜLLER, H. (2010). Varying-coefficient functional linear regression. *Bernoulli* **16** 730–758. [MR2730646](#)
- [46] YANG, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *J. Amer. Statist. Assoc.* **94** 137–145. [MR1689219](#)
- [47] YIN, G., ZENG, D. and LI, H. (2008). Power-transformed linear quantile regression with censored data. *J. Amer. Statist. Assoc.* **103** 1214–1224. [MR2462894](#)
- [48] YING, Z., JUNG, S. H. and WEI, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90** 174–184. [MR1325125](#)