Assessing dynamic covariate effects with survival data

Ying Cui[1] · Limin Peng[1]

Abstract

Dynamic (or varying) covariate effects often manifest meaningful physiological mechanisms underlying chronic diseases. However, a static view of covariate effects is typically adopted by standard approaches to evaluating disease prognostic factors, which can result in depreciation of some important disease markers. To address this issue, in this work, we take the perspective of globally concerned quantile regression, and propose a flexible testing framework suited to assess either constant or dynamic covariate effects. We study the powerful Kolmogorov–Smirnov (K–S) and Cramér–Von Mises (C–V) type test statistics and develop a simple resampling procedure to tackle their complicated limit distributions. We provide rigorous theoretical results, including the limit null distributions and consistency under a general class of alternative hypotheses of the proposed tests, as well as the justifications for the presented resampling procedure. Extensive simulation studies and a real data example demonstrate the utility of the new testing procedures and their advantages over existing approaches in assessing dynamic covariate effects.

Keywords Hypothesis testing · Globally concerned quantile regression · Testing consistency · Resampling

1 Introduction

Identifying useful prognostic factors is often of critical interests in chronic disease studies. When the disease outcome is captured by a time-to-event, a commonly used approach is to model the mechanism of a prognostic factor influencing the time-to-event outcome via a standard survival regression model and then test the corresponding covariate effects (see a review in Kleinbaum and Klein (2010) and Cox and Oakes (2018)). The standard survival regression models, such as the Cox proportional hazard

✉ Limin Peng
  lpeng@emory.edu

[1] Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA

(PH) regression model and the accelerated failure time (AFT) model, impose assumptions like the proportional hazards and the location-shift effects, which implicitly confine the prognostic factor of interest to be a static portent of disease progression.

There has been growing awareness that a prognostic factor may follow a dynamic association with a time-to-event disease outcome. Many reports in literature (Dickson et al. 1989; Thorogood et al. 1990; Verweij and van Houwelingen 1995; Bellera et al. 2010), for example) have suggested that postulating constant covariate effects, sometimes, is not adequate to reflect underlying physiological disease mechanisms, leading to distorted assessment of the prognostic factor. For example, an analysis of a dialysis dataset reported by Peng and Huang (2008) suggested that the severity of restless leg syndrome (RLS) symptoms may be prognostic of mortality for short-term dialysis survivors but not for long-term dialysis survivors. The standard tests based on the Cox PH model and the AFT model failed to detect such a dynamic effect.

Quantile regression (Koenker and Bassett 1978), which directly formulates covariate effects on quantile(s) of a response, confers a seminal venue to characterize a dynamic effect of a prognostic factor. Specifically, given a time-to-event outcome T and a covariate $\widetilde{Z}$ (which represents the prognostic factor of interest), a linear quantile regression model may assume,

$$Q_T(\tau|\widetilde{Z}) = \exp\{\mathbf{Z}^T \boldsymbol{\theta}_0(\tau)\}, \quad \tau \in \Delta, \tag{1}$$

where $\mathbf{Z} = (1, \widetilde{Z})^T$, $Q_T(\tau|\widetilde{Z}) \equiv \inf\{t : \Pr(T \leq t|\widetilde{Z}) \geq \tau\}$ denotes the τ-th conditional quantile of T given $\widetilde{Z}$, $\boldsymbol{\theta}_0(\tau) \equiv (\beta_0^{(0)}(\tau), \beta_0^{(1)}(\tau))^T$ is an unknown coefficient vector, and $\Delta \subseteq (0, 1)$ is a pre-specified set including the quantile levels of interest. The coefficient $\beta_0^{(1)}(\tau)$ represents the effect of $\tilde{Z}$ on the τ-th conditional quantile of T, and is allowed to change with τ. This implicates that the prognostic factor is permitted to have different effects across different segments of the distribution of the time-to event outcome.

Many authors have studied linear quantile regression with a time-to-event outcome (Powell 1986; Ying et al. 1995; Portnoy 2003; Zhou 2006; Peng and Huang 2008; Wang and Wang 2009; Huang 2010, for example). Most of the existing methods concern covariate effects on a single or multiple pre-specified quantile levels (e.g. Δ is a singleton set $\{0.5\}$), and, following the terminology of Zheng et al. (2015), are locally concerned. As discussed in Zheng et al. (2015), locally concerned quantile regression cannot inform of the covariate effect on quantiles other than the specifically targeted ones (e.g. median), and thus may miss important prognostic factors. Adopting the perspective of globally concerned quantile regression, one can simultaneously examine covariate effects over a continuum of quantile levels (e.g. Δ is an interval [0.1, 0.9]), and thus confer a more comprehensive assessment of a prognostic factor. However, powerful tests tailored to evaluate covariate effects under the perspective of globally concerned quantile regression have not been formally studied, partly owing to the associated inferential complexity.

In this work, we develop a new framework for evaluating a survival prognostic factor following the spirit of globally concerned quantile regression. As a proof of concept, we shall confine the scope of this work to the standard survival setting where

the time-to-event outcome T is subject to random censoring. Specifically, our proposal is to simultaneously assess the influence of the prognostic factor on a range of quantiles of T, indexed by a τ-interval, $[\tau_L, \tau_U] \subset (0, 1)$. As the key rationale, a significant prognostic factor is allowed to have a dynamic τ-varying effect, which may be non-zero throughout the whole τ-interval (i.e. full effect), or only over a part of the τ-interval (i.e. partial effect). Under this view, when model (1) with $\Delta = [\tau_L, \tau_U]$ holds, the task of identifying a prognostic factor reduces to testing the null hypothesis,

$$H_0 : \ \beta_0^{(1)}(\tau) = 0, \ \ \tau \in [\tau_L, \tau_U].$$

Moreover, without assuming any models, we may consider the null hypothesis formulated as,

$$H_0^* : \ Q_T(\tau|\widetilde{Z}) = Q_T(\tau) \ \text{for} \ \tau \in [\tau_L, \tau_U],$$

where $Q_T(\tau) = \inf\{t : \ \Pr(T \leq t) \geq \tau\}$, denoting the τ-th unconditional (or marginal) quantile of T. The null hypothesis H_0^* corresponds to the setting where $\widetilde{Z}$ has no influence on the conditional quantile of T at any quantile level between τ_L and τ_U.

It is remarkable that under mild regularity conditions, H_0^* implies that model (1) holds with $\Delta = [\tau_L, \tau_U]$ and $\beta_0^{(1)}(\tau) = 0$ for $\tau \in [\tau_L, \tau_U]$; on the other hand, model (1) with $\Delta = [\tau_L, \tau_U]$ and $\beta_0^{(1)}(\tau) = 0$ for $\tau \in [\tau_L, \tau_U]$ implies $Q_T(\tau|\widetilde{Z}) = Q_T(\tau)$ for $\tau \in [\tau_L, \tau_U]$; see Lemma 1 in the Appendix A. This finding sheds an important insight that a model-based test developed for H_0 may be used towards testing the model-free null hypothesis H_0^*. From an alternative view, this result suggests that the globally concerned quantile regression model (1) with $\Delta = [\tau_L, \tau_U]$ can be used as a working model to test H_0^*, which adopts the view that the effect of a prognostic factor can be assessed through contrasting the conditional versus unconditional quantiles of T.

Regarding H_0^*, we study two "omnibus" test statistics constructed based on the estimator of $\boldsymbol{\theta}_0(\tau)$ obtained under the working model (1) with $\Delta = [\tau_L, \tau_U]$. One test is a Kolmogorov–Smirnov (K–S) type test statistic defined upon the maximum "signal" strength (i.e. covariate effect) over τ's in $[\tau_L, \tau_U]$. The other one is a Cramér–Von-Mises (C–V) type test statistics based on the average "signal" strength over τ's in $[\tau_L, \tau_U]$. These two types of test statistics are known to be very sensitive to detect any departure from the null hypothesis H_0 under model (1). However, the analytic form of their limit null distributions are generally complex and sometimes intractable. This challenge is more intense in the quantile regression setting, where coefficient estimates do not have a closed form, and the corresponding asymptotic variance matrix involves unknown density functions (Koenker 2005). To overcome these difficulties, we propose to approximate the limit null distributions through a resampling procedure that perturbs the influence function associated with the adopted coefficient estimator under the working model (1), following similar strategies of Lin et al. (1993) and Li and Peng (2014). We derive a sample-based procedure to estimate the influence function without requiring the correct specification of model (1), thereby circumvents directly

evaluating the unknown density function via smoothing. The proposed resampling procedure is easy to implement and is shown to perform well even with realistic sample sizes. Moreover, we provide rigorous theoretical justifications for the proposed resampling procedure.

The rest of this paper is organized as follows. In Sect. 2, we first briefly review some existing results about the estimation of model (1), which we use as a working model for testing H_0^*. We then present the proposed test statistics along with their theoretical properties. A resampling procedure is developed to carry out inference regarding H_0 or H_0^* based on the proposed test statistics. We also discuss some computational strategies to help simplify or improve the implementation of the proposed method. In Sect. 3, we report extensive simulation studies conducted to evaluate the finite-sample performance of the proposed testing procedures. Our simulation results show that the proposed tests have accurate empirical sizes and can be much more powerful than benchmark methods when assessing a covariate with a dynamic effect. In Sect. 4, we further demonstrate the usefulness of the proposed testing procedures with a real data example. Concluding remarks and discussions are provided in Sect. 5.

2 The proposed testing procedures

2.1 Estimation of $\theta_0(\tau)$ under model (1)

As explained in Sect. 1, we propose to use globally concerned quantile regression as a vehicle to address the testing problem regarding the general null hypothesis H_0^*. The first step is to obtain an estimator of $\theta_0(\tau)$ (and thus $\beta_0^{(1)}(\tau)$) from fitting the working model (1) to the observed data. Here and hereafter, we shall set the Δ in model (1) as $\Delta = [\tau_L, \tau_U]$, which is a pre-specified interval within $(0, 1)$. Let C denote time to censoring, $X = \min(T, C)$, and $\delta = I(T \leq C)$. The observed data include n i.i.d. replicates of (X, δ, Z), denoted by $\{(X_i, \delta_i, Z_i)\}_{i=1}^n$.

To estimate $\theta_0(\tau)$ under model (1), we choose to adapt the existing results of Peng and Fine (2009) developed for competing risks data to the setting with randomly censored data. Compared to the other available estimators developed by Portnoy (2003) and Peng and Huang (2008), which require $\tau_L = 0$, the estimator derived from Peng and Fine (2009) is more robust to any realistic violation of the global linearity assumed by model (1) (Peng 2021). The influence function associated with Peng and Fine (2009)'s estimator also has a simpler form that can facilitate the development of the corresponding testing procedures.

The estimator of $\theta_0(\tau)$ adapted from Peng and Fine (2009)'s work, denoted by $\widehat{\theta}(\tau)$, is obtained as the solution to the following estimating equation:

$$S_n(b, \tau) = n^{-1/2} \sum_{i=1}^n Z_i \left[\frac{I(X_i \leq \exp\{Z_i^T b\})I(\delta_i = 1)}{\widehat{G}(X_i|Z_i)} - \tau \right] = 0, \qquad (2)$$

where $\widehat{G}(x|Z)$ is a reasonable estimator of $G(x|Z) \equiv \Pr(C \geq x|Z)$. For simplicity of illustration, in sequel, we shall assume C is independent of $\tilde{Z}$ and thus take $\widehat{G}(x|Z)$

as the Kaplan–Meier estimator of the marginal survival function of C, $\widehat{G}(x)$. As noted by Peng and Fine (2009), solving (2) can be formulated as a L_1-type minimization problem of the following convex objective function:

$$
U_n(\boldsymbol{b}, \tau) = \sum_{i=1}^{n} I(\delta_i = 1) \left| \frac{\log(X_i)}{\widehat{G}(X_i)} - \boldsymbol{b}^T \frac{\boldsymbol{Z}_i}{\widehat{G}(X_i)} \right| + \left| M - \boldsymbol{b}^T \sum_{l=1}^{n} \frac{-\boldsymbol{Z}_l I(\delta_l = 1)}{\widehat{G}(X_i)} \right|
$$
$$
+ \left| M - \boldsymbol{b}^T \sum_{k=1}^{n} (2\tau \boldsymbol{Z}_k) \right|.
$$

Here M is a sufficiently large number. This L_1-type minimization problem can be easily solved using the `rq()` function in the R package `quantreg` by Koenker (2022).

By the results of Peng and Fine (2009), the estimator $\widehat{\boldsymbol{\theta}}(\tau)$ enjoys desirable asymptotic properties. Specifically, under certain regularity conditions, we have (i) $\lim_{n \to \infty} \sup_{\tau \in [\tau_L, \tau_U]} ||\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)|| \to_p 0$; and (ii) $\sqrt{n}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\}$ converge weakly to a mean zero Gaussian process for $\tau \in [\tau_L, \tau_U]$ with covariance function $\boldsymbol{\Phi}(\tau', \tau) = E\{\boldsymbol{\xi}_1(\tau')\boldsymbol{\xi}_1(\tau)^T\}$. Here $\boldsymbol{\xi}_i(\tau)$ ($i = 1, \ldots, n$) are defined as

$$
\boldsymbol{\xi}_i(\tau) \equiv \{\boldsymbol{\xi}_i^{(0)}(\tau), \boldsymbol{\xi}_i^{(1)}(\tau)\}^T = \{A(\boldsymbol{\theta}_0(\tau))\}^{-1} \left\{ \boldsymbol{Z}_i \left(\frac{I(\log(X_i) \le \boldsymbol{Z}_i^T \boldsymbol{\theta}_0(\tau), \delta_i = 1)}{G(X_i)} - \tau \right) \right.
$$
$$
\left. - \int_0^{\infty} \boldsymbol{w}\{\boldsymbol{\theta}_0(\tau), s\} y(s)^{-1} dM_i^G(s) \right\},
$$

where $G(x) = \Pr(C > x)$, $A(\boldsymbol{b}) = E[\boldsymbol{Z}\boldsymbol{Z}^T f(\boldsymbol{Z}^T \boldsymbol{b}|\boldsymbol{Z})]$ with $f(t|\boldsymbol{Z})$ denoting the conditional density of X given $\boldsymbol{Z}$, $\boldsymbol{w}(\boldsymbol{b}, t) = E[\boldsymbol{Z}Y(t)I(X \le \exp\{\boldsymbol{Z}^T \boldsymbol{b}\})I(\delta = 1)G(X)^{-1}]$, and $M_i^G(t) = N_i^G(t) - \int_0^{\infty} Y_i(s)d\Lambda^G(t)$ with $N_i^G(t) = I(X_i \le t, \delta_i = 0)$, $Y_i(t) = I(X_i \ge t)$, $y(t) = \Pr(X \ge t)$, $\lambda^G(t) = \lim_{\Delta \to 0} P(C \in (t, t + \Delta)|C \ge t)/\Delta$, and $\Lambda^G(t) = \int_0^t \lambda^G(s)ds$. In addition, $n^{1/2}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\} \approx n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\xi}_i(\tau)$, where $\approx$ indicate asymptotical equivalence uniformly in $\tau \in [\tau_L, \tau_U]$. Consequently, $\boldsymbol{\xi}_i(\tau)$ is referred to as the influence function of $n^{1/2}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\}$.

Note that the variance estimation for $\widehat{\boldsymbol{\theta}}(\tau)$ is complicated by the involvement of the unknown density $f(t|\boldsymbol{Z})$ in the asymptotic covariance matrix $\boldsymbol{\Phi}(\tau', \tau)$. As justified by Peng and Fine (2009), a sample-based procedure that avoids smoothing-based density estimation can be used for variance estimation and is outlined below:

(1.a) Compute an consistent variance estimate for $\boldsymbol{S}_n(\boldsymbol{\theta}_0(\tau), \tau)$ given by

$$
\widehat{\boldsymbol{\Sigma}}(\tau, \tau) = n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i^{\otimes 2} \left(\frac{I[\log(X_i) \le \boldsymbol{Z}_i^T \widehat{\boldsymbol{\theta}}(\tau)), \delta_i = 1]}{\widehat{G}(X_i)} - \tau \right)^2
$$
$$
- n^{-1} \sum_{i=1}^{n} I(\delta_i = 0) \left(\sum_{j=1}^{n} \boldsymbol{Z}_j I(X_j \ge X_i) I[\log(X_j)]
$$

$$\leq Z_i^T \widehat{\boldsymbol{\theta}}(\tau), \delta_j = 1]\{\widehat{G}(X_j)\}^{-1} \Big/ \sum_{j=1}^{n} I(X_j \geq X_i) \Bigg)^{\otimes 2},$$

where for a vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T$.

(1.b) Find a symmetric and nonsingular matrix $\boldsymbol{E}_n(\tau) \equiv \{\boldsymbol{e}_{n,0}(\tau), \boldsymbol{e}_{n,1}(\tau)\}$ such that $\{\boldsymbol{E}_n(\tau)\}^2 = \widehat{\boldsymbol{\Sigma}}(\tau, \tau)$.

(1.c) Calculate $\boldsymbol{D}_n(\tau) = \{S_n^{-1}\{\boldsymbol{e}_{n,0}(\tau), \tau\} - \widehat{\boldsymbol{\theta}}(\tau), S_n^{-1}\{\boldsymbol{e}_{n,1}(\tau), \tau\} - \widehat{\boldsymbol{\theta}}(\tau)\}$, where $S_n^{-1}\{\boldsymbol{e}(\tau), \tau\}$ is the solution to the perturbed estimating equation $\boldsymbol{S}_n(\boldsymbol{b}, \tau) = \boldsymbol{e}(\tau)$.

(1.d) Obtain an estimate for the asymptotic variance of $\sqrt{n}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\}$ as $\boldsymbol{V}_n(\tau) \equiv n\boldsymbol{D}_n^{\otimes 2}(\tau)$.

Here $\boldsymbol{E}_n(\tau)$ can be computed with the eigenvalue eigenvector decomposition of $\widehat{\boldsymbol{\Sigma}}(\tau, \tau)$ using the R function `eigen()`. As another important remark, the above procedure ensures that the perturbation terms, $\boldsymbol{e}_{n,j}(\tau)$, $j = 1, 2$, have the desired asymptotic order. As a result, this procedure remains valid when $\boldsymbol{e}_{n,j}(\tau)$ in step (1.c) is replaced by $u \cdot \boldsymbol{e}_{n,j}(\tau)$ for some constant u. Based on our numerical experiences, incorporating some constant u can help stabilize variance estimation when sample size is small or τ is close to 0 or 1. Variance estimation based on the above procedure is found to have satisfactory finite sample performance based on some unreported simulation studies.

2.2 The proposed test statistics and theoretical properties

Express $\widehat{\boldsymbol{\theta}}(\tau) \equiv (\widehat{\beta}^{(0)}(\tau), \widehat{\beta}^{(1)}(\tau))'$ and let $\hat{\sigma}_n^{(1)}(\tau)$ denote the square root of the second diagonal element of $\boldsymbol{V}_n(\tau)$, which corresponds to the variance estimate for $\sqrt{n}\widehat{\beta}^{(1)}(\tau)$ under H_0^*. We propose to construct two "omnibus" test statistics based on $\widehat{\beta}^{(1)}(\tau)$ and $\hat{\sigma}_n^{(1)}(\tau)$:

$$\widehat{T}_{sup}^{(1)} = \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{\sqrt{n}\widehat{\beta}^{(1)}(\tau)}{\hat{\sigma}_n^{(1)}(\tau)} \right|,$$

and

$$\widehat{T}_{inte}^{(1)} = \int_{\tau_L}^{\tau_U} \left| \frac{\sqrt{n}\widehat{\beta}^{(1)}(\tau)}{\hat{\sigma}_n^{(1)}(\tau)} \right|^2 d\tau.$$

These two test statistics mimic the classic Kolmogorov–Smirnov (K–S) test statistic and Cramér–Von-Mises (C–V) test statistic for two-sample distribution comparisons (Darling 1957). Under model (1), $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ capture the maximum and average magnitude of the covariate effect over $\tau \in [\tau_L, \tau_U]$ respectively. By this design, both test statistics are sensitive to any type of departures from the null hypothesis H_0 and can be used to construct powerful tests for H_0.

Without assuming model (1), we can also show that $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ provide valid tests for H_0^* and have power approaching one under a general class of alternative hypotheses as specified in Theorem 2. The key insight is that even when model (1) does not hold, $\widehat{\theta}(\tau)$ may still converge in probability to a deterministic function $\widetilde{\theta}(\tau) \equiv (\widetilde{\beta}^{(0)}(\tau), \widetilde{\beta}^{(1)}(\tau))'$ that is the solution to $\boldsymbol{\mu}(\boldsymbol{b}, \tau) \equiv E[\boldsymbol{Z}\{I(\log T \leq \boldsymbol{Z}^T \boldsymbol{b}) - \tau\}] = 0$. It is easy to see that $\widetilde{\theta}(\tau) = \theta_0(\tau)$ under model (1). By Lemma 1, it follows that under H_0^*, $\widetilde{\beta}^{(1)}(\tau) = 0$ for $\tau \in [\tau_L, \tau_U]$. As detailed in Theorems A1–A2 in Appendix A, under certain regularity conditions, we further have $\lim_{n \to \infty} \sup_{\tau \in [\tau_L, \tau_U]} ||\widehat{\theta}(\tau) - \widetilde{\theta}(\tau)|| \to_p 0$, and $\sqrt{n}\{\widehat{\theta}(\tau) - \widetilde{\theta}(\tau)\}$ converge weakly to a mean zero Gaussian process for $\tau \in [\tau_L, \tau_U]$ with covariance function $\widetilde{\boldsymbol{\Phi}}(\tau', \tau) = E\{\widetilde{\boldsymbol{\xi}}_1(\tau')\widetilde{\boldsymbol{\xi}}_1(\tau)^T\}$, where $\widetilde{\boldsymbol{\xi}}_i(\tau)$ $(i = 1, \ldots, n)$ are defined as

$$\widetilde{\boldsymbol{\xi}}_i(\tau) \equiv \{\widetilde{\xi}_i^{(0)}(\tau), \widetilde{\xi}_i^{(1)}(\tau)\}^T = \{A(\widetilde{\theta}(\tau))\}^{-1} \left\{ \boldsymbol{Z}_i \left(\frac{I(\log(X_i) \leq \boldsymbol{Z}_i^T \widetilde{\theta}(\tau), \delta_i = 1)}{G(X_i)} - \tau \right) \right.$$
$$\left. - \int_0^\infty \boldsymbol{w}\{\widetilde{\theta}(\tau), s\} y(s)^{-1} dM_i^G(s) \right\}.$$

A useful by-product from the proof of Theorem A2 is that

$$n^{1/2}\{\widehat{\theta}(\tau) - \widetilde{\theta}(\tau)\} \approx n^{-1/2} \sum_{i=1}^n \widetilde{\boldsymbol{\xi}}_i(\tau), \tag{3}$$

We can prove these results by adapting the arguments of Peng and Fine (2009) which utilize model assumption (1) only through using its implication $\boldsymbol{\mu}(\theta_0, \tau) = 0$ for $\tau \in [\tau_L, \tau_U]$. This provides the critical justification for why $\widehat{\beta}^{(1)}(\tau)$ can be used to test H_0^* even when model (1) does not hold. The sample-based procedure reviewed in Sect. 2.1 is still applicable to estimate the asymptotic covariance matrix $\widetilde{\boldsymbol{\Phi}}(\tau', \tau)$.

In Theorems 1 and 2, we establish useful asymptotic properties of $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ without assuming model (1). Specifically, in Theorem 1, we provide the limit distributions of the proposed test statistics under the null hypothesis H_0^*:

Theorem 1 *Assuming the regularity conditions (C1)–(C5) in the Appendix hold, under the null hypothesis H_0 or H_0^*, we have*

$$\widehat{T}_{sup}^{(1)} = \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{n^{1/2}\widehat{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right| \to_d \sup\{|\mathcal{X}^{(1)}(\tau)|, \tau \in [\tau_L, \tau_U]\}$$

$$\widehat{T}_{inte}^{(1)} = \int_{\tau_L}^{\tau_U} \left| \frac{n^{1/2}\widehat{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right|^2 d\tau \to_d \int_{\tau_L}^{\tau_U} \{\mathcal{X}^{(1)}(\tau)\}^2 d\tau,$$

where $\mathcal{X}^{(1)}(\tau)$ is a mean zero Gaussian process defined in Appendix C.

We also investigate the asymptotic behavior of the proposed test statistics under a general class of alternative hypotheses. The findings are stated in Theorem 2.

Theorem 2 *Assuming the regularity conditions (C1)–(C5) in the Appendix hold,*

(A) $\widehat{T}_{sup}^{(1)}$ *is consistent against the alternative hypothesis*

$$H_{a,1} : \quad \sup_{\tau \in [\tau_L, \tau_U]} \left| \widetilde{\beta}^{(1)}(\tau) \right| > 0.$$

(B) $\widehat{T}_{inte}^{(1)}$ *is consistent against the alternative hypothesis:*

$$H_{a,2} : \quad \int_{\tau_L}^{\tau_U} \{ \widetilde{\beta}^{(1)}(\tau) \}^2 d\tau > 0.$$

The results of Theorem 2 indicate that the test statistics have power approaching to 1 (as n goes to ∞) under alternative cases subject to very mild constraints. Given the smoothness of $\widetilde{\beta}^{(1)}(\cdot)$, a general scenario that ensures the consistency of both $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ can be described as

$\widetilde{H}_a$: There exists an interval $[\tau_1, \tau_2] \subseteq [\tau_L, \tau_U]$ such that $|\widetilde{\beta}_0^{(1)}(\tau)| > 0$ for $\tau \in [\tau_1, \tau_2]$.

This suggests that the proposed tests are powerful to identify a significant prognostic factor even when it only influences a segment of the outcome distribution, not necessarily the whole outcome distribution. This feature is conceptually appealing for handling a dynamic covariate effect, which may not have similar effect strength across different quantiles. The detailed proofs for Theorems 1 and 2 can be found in Appendix C.

2.3 The proposed resampling procedure to obtain *p* values

The results in Theorem 1 suggest that $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$, like the classic K–S test statistic and C–V test statistic, have complex, non-standard limit null distributions. This motivates us to develop a resampling-based procedure to approximate their limit null distributions and obtain the corresponding p values for testing H_0^*.

Our key strategy is to approximate the distribution of $n^{1/2} \{ \widehat{\beta}^{(1)}(\tau) - \widetilde{\beta}_0^{(1)}(\tau) \}$, which reduces to $n^{1/2} \widehat{\beta}^{(1)}(\tau)$ under H_0, through perturbing the influence function $\widetilde{\xi}_i^{(1)}(\tau)$, which is the second component of $\widetilde{\xi}_i(\tau)$. Similar ideas were used by other authors, for example, Lin et al. (1993) and Li and Peng (2014). The core justification of our proposal is provided by equation (3), which suggests that $n^{-1/2} \sum_{i=1}^{n} \widetilde{\xi}_i^{(1)}(\tau) \iota_i / \widehat{\sigma}_n^{(1)}(\tau)$ may be used to approximate $\sqrt{n} \widehat{\beta}^{(1)}(\tau) / \widehat{\sigma}_n^{(1)}(\tau)$, where $\{\iota_i\}_{i=1}^{n}$ are i.i.d. standard normal variates.

Specifically, we take the following steps:

(2.a) Generate B independent sets of $\{\iota_i^b\}_{i=1}^{n}$, where $\{\iota_i^b\}_{i=1}^{n}$ are independent random variables from a standard normal distribution and $b = 1, 2, \ldots, B$.

(2.b) Compute the estimates for the influence function $\widetilde{\xi}_i^{(1)}(\tau)$ as the second component of

$$\widehat{\xi}_i(\tau) = \{\widehat{A}(\widehat{\theta}(\tau))\}^{-1}\left\{ \mathbf{Z}_i(\frac{I[\log(X_i) \leq \mathbf{Z}_i^T\widehat{\theta}(\tau)), \delta_i = 1]}{\widehat{G}(X_i)} - \tau) - I(\delta_i = 0) \right.$$

$$\left. \times \frac{\sum_{j=1}^n \mathbf{Z}_j I(X_j \geq X_i)I[\log(X_j) \leq \mathbf{Z}_j^T\widehat{\theta}(\tau), \delta_j = 1]\{\widehat{G}(X_j)\}^{-1}}{\sum_{j=1}^n I(X_j \geq X_i)} \right\},$$

where $\widehat{A}\{\widehat{\theta}(\tau)\}^{-1} = n^{1/2}\mathbf{D}_n(\tau)\mathbf{E}_n(\tau)^{-1}$.

(2.c) For $b = 1, \ldots, B$, calculate

$$\widehat{T}_{sup,b}^{(1)} = \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{n^{-1/2}\sum_{i=1}^n \widehat{\xi}_i^{(1)}(\tau)\iota_i^b}{\widehat{\sigma}_n^{(1)}(\tau)} \right| \quad \text{and}$$

$$\widehat{T}_{inte,b}^{(1)} = \int_{\tau_L}^{\tau_U} \left| \frac{n^{-1/2}\sum_{i=1}^n \widehat{\xi}_i^{(1)}(\tau)\iota_i^b}{\widehat{\sigma}_n^{(1)}(\tau)} \right|^2 d\tau,$$

where $\widehat{\xi}_i^{(1)}(\tau)$ is the second component of $\widehat{\xi}_i(\tau)$.

(2.d) The p values based on $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ are calculated respectively as

$$p_{sup}^{(1)} = \sum_{b=1}^B I(\widehat{T}_{sup,b}^{(1)} > \widehat{T}_{sup}^{(1)})/B \quad \text{and} \quad p_{inte}^{(1)} = \sum_{b=1}^B I(\widehat{T}_{inte,b}^{(1)} > \widehat{T}_{inte}^{(1)})/B.$$

The resampling procedure presented above is easy to implement without involving smoothing. The rigorous theoretical justification for the presented resampling procedure is provided in Appendix D.

2.4 Some computational considerations

Note that $\widehat{\beta}^{(1)}(\tau)$ and $\widehat{\sigma}_n^{(1)}(\tau)$ are piecewise constant; thus an exact calculation of the supremum or integration involved in $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ is possible. Alternatively, we may follow the recommendation of Zheng et al. (2015) to compute $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ based on a simpler piecewise-constant approximation of $\widehat{R}(\tau) \equiv \widehat{\beta}^{(1)}(\tau)/\widehat{\sigma}_n^{(1)}(\tau)$ on a predetermined fine τ-grid, $\mathcal{G} \equiv \tau_L = \tau_1 < \tau_2 < \cdots < \tau_{N_*} = \tau_U$, with the grid size $\max_{1 \leq l \leq N_*-1}(\tau_{l+1} - \tau_l) = o(n^{-1/2})$. In this case, the proposed test statistics can be calculated as

$$\widehat{T}_{sup}^{(1)} = \sqrt{n}\max\{\widehat{R}(\tau_l) : 1 \leq l \leq N_*\}, \quad \widehat{T}_{inte}^{(1)} = \sum_{l=1}^{N_*-1} n\{\widehat{R}(\tau_l)\}^2(\tau_{l+1} - \tau_l). \quad (4)$$

When n is not large, the sample-based variance estimation (i.e. the computation of $\widehat{\sigma}_n^{(1)}(\tau)$) sometimes is not stable. Our remedy is to replace the $\mathbf{e}_{n,j}(\tau)$ in step (1.c)

(see Sect. 2.1) with $u \cdot \boldsymbol{e}_{n,j}(\tau)$, where u is a pre-specified constant. We develop the following algorithm to determine a good choice of the adjusting constant u among a set of candidate values, $\mathcal{U} = \{1, 2, \ldots, U\}$.

(3.a) For each $u \in \mathcal{U}$, calculate $\widehat{R}(\tau; u) \equiv \widehat{\beta}^{(1)}(\tau)/\widehat{\sigma}_n^{(1)}(\tau; u)$ for $\tau \in \mathcal{G}$, where $\widehat{\sigma}_n^{(1)}(\tau; u)$ is the $\widehat{\sigma}_n^{(1)}(\tau)$ computed with the adjusting constant u.

(3.b) For each $u \in \mathcal{U}$, calculate $\widehat{R}^*(u) = \max_{\tau \in \mathcal{G}} \widehat{R}(\tau; u)$ and $\widehat{R}^\dagger(u) = \text{median}_{\tau \in \mathcal{G}} \widehat{R}(\tau; u)$.

(3.c) For each $u \in \mathcal{U}$, calculate $\widetilde{R}(u) = \max_{\tau \in \mathcal{G}} \max\{\boldsymbol{V}_n(\tau; u)\} - \min_{\tau \in \mathcal{G}} \min\{\boldsymbol{V}_n(\tau; u)\}$, where $\boldsymbol{V}_n(\tau; u)$ is $\boldsymbol{V}_n(\tau)$ computed with the adjusting constant u. Here, for a matrix $\boldsymbol{A}$, $\max(\boldsymbol{A})$ (or $\min(\boldsymbol{A})$) denotes the largest (or the smallest) component of the matrix $\boldsymbol{A}$.

(3.d) Assign a large positive value to $A^{[0]}$ and $B^{[0]}$, say 10^5. Set $k = 1$ and $u^{[0]} = U + 1$.

  (i) If $\widehat{R}^*(k) - \widehat{R}^\dagger(k) < A^{[k-1]}$ and $\widetilde{R}(k) < B^{[k-1]}$, then let $A^{[k]} = \widehat{R}^*(k) - \widehat{R}^\dagger(k)$, $B^{[k]} = \widetilde{R}(k)$, and $u^{[k]} = k$. Otherwise, let $A^{[k]} = A^{[k-1]}$, $B^{[k]} = B^{[k-1]}$ and $u^{[k]} = u^{[k-1]}$.

 (ii) Increase k by 1 and go back to (i) until $k > U$.

(3.e) If $u^{[U]} < U + 1$, then choose u as $u^{[U]}$. Otherwise, no appropriate u can be selected from $\mathcal{U}$.

By this algorithm, we provide an empirical strategy to select u based on two estimation instability measures: (A) $\widehat{R}^*(k) - \widehat{R}^\dagger(k)$, which reflects the spread of $\widehat{R}(\tau) \equiv \widehat{\beta}^{(1)}(\tau)/\widehat{\sigma}_n^{(1)}(\tau)$ over τ given $u = k$; (B) $\widetilde{R}(k)$, which measures the maximum fluctuation of the estimated variance matrices across τ given $u = k$. It is clear that both measures would be large when unstable variance estimation occurs. Our algorithm first compares them with pre-specified initial values, $A^{[0]}$ and $B^{[0]}$, to rule out the occurrence of obviously outlying estimates of $\widehat{R}(\tau)$ or $\widehat{\sigma}_n^{(1)}(\tau)$. Once these two measures are found to meet the stability criteria set by the initial values with some $u \in \mathcal{U}$, the algorithm will proceed to check if other u's can yield smaller values of the instability measures. The output from this algorithm is either the value of u that produces the smallest instability measures, or an error message indicating that none of the constants in $\mathcal{U}$ can lead to stable estimation required by the proposed testing procedure. Based on our numerical experiences, setting $\mathcal{U} = \{1, 2, \ldots, 6\}$, which corresponds to $U = 6$, works well for small sample sizes such as 200 or 400. In a rare case where this algorithm fails to identify an appropriate u, we recommend adaptively increasing the value of U until an appropriate u can be identified. Our extensive numerical experiences suggest that incorporating the adjusting constant u selected by this algorithm results in good and stable numerical performance of the proposed tests. The algorithm can be easily generalized to allow $\mathcal{U}$ to include non-integer values.

3 Simulation studies

We conduct extensive simulation studies to investigate the finite-sample performance of the proposed resampling-based testing procedures. To simulate randomly censored

data, we consider six setups where T and $\widetilde{Z}$ follow different relationships. In all setups, we generate $\widetilde{Z}$ from $Uniform(0, 1)$ and generate censoring time C from $Uniform(U_L, U_U)$, where U_L and U_U are properly specified to produce 15% or 30% censoring. Let $\Phi(\cdot)$ denote the cumulative distribution function of the standard normal distribution. The six simulation set-ups are described as follows.

(I) Setup I: Generate T such that $Q_\tau\{\log(T)\} = \Phi^{-1}(\tau)$. Set $(U_L, U_U) = (2, 3.8)$ to produce 15% censoring, and set $(U_L, U_U) = (1, 2.5)$ to produce 30% censoring.

(II) Setup II: Generate T such that $Q_\tau\{\log(T)\} = 0.2X + \Phi^{-1}(\tau)$. Set $(U_L, U_U) = (2.5, 3.9)$ to produce 15% censoring and set $(U_L, U_U) = (1.2, 2.8)$ to produce 30% censoring.

(III) Setup III: Generate T such that $Q_\tau\{\log(T)\} = 0.5X + \Phi^{-1}(\tau)$. Set $(U_L, U_U) = (2.7, 4.9)$ to produce 15% censoring, and set $(U_L, U_U) = (1.5, 3)$ to produce 30% censoring.

(IV) Setup IV: Generate T such that $Q_\tau\{\log(T)\} = l_4(\tau)X + \Phi^{-1}(\tau)$, where $l_4(\tau)$ is as plotted in Fig. 1. Set $(U_L, U_U) = (2, 3.9)$ to produce 15% censoring, and set $(U_L, U_U) = (1, 2.5)$ to produce 30% censoring.

(V) Setup V: Generate T such that $Q_\tau\{\log(T)\} = l_5(\tau)X + \Phi^{-1}(\tau)$, where $l_5(\tau)$ is as plotted in Fig. 1. Set $(U_L, U_U) = (5.2, 6.5)$ to produce 15% censoring, and set $(U_L, U_U) = (1.5, 3.5)$ to produce 30% censoring.

(VI) Setup VI: Generate T such that $Q_\tau\{\log(T)\} = l_6(\tau)X + \Phi^{-1}(\tau)$, where $l_6(\tau)$ is as plotted in Fig. 1. Set $(U_L, U_U) = (3.5, 5.5)$ to produce 15% censoring, and set $(U_L, U_U) = (1.1, 3.5)$ to produce 30% censoring.

Under all setups, model (1) holds for $\tau \in (0, 1)$ and thus for $\tau \in [0.1, 0.6]$, a pre-specified τ-interval of interest $[\tau_L, \tau_U]$. In Fig. 1, we plot the true coefficient function $\beta_0^{(1)}(\tau)$ for each setup. It is easy to see that setup (I) represents a null case, where $\widetilde{Z}$ has no effect on any quantile of T. Setup (II) and (III) are two setups where $\widetilde{Z}$ has nonzero constant effects over all $\tau \in [0.1, 0.6]$. The constant effect in setup (II) has a magnitude of 0.2, which is smaller than that in setup (III), which is 0.5. In setups (IV), (V), and (VII), $\widetilde{Z}$ has a dynamic effect varying across different τ's. More specifically, $\widetilde{Z}$ has a partial effect over the τ-interval $[0.1, 0.49]$ in setup (IV). In setup (V), the magnitude of $\widetilde{Z}$'s effect is symmetric around 0.5, while the sign of the effect is opposite for $\tau < 0.5$ and for $\tau > 0.5$, and the effect equals 0 at $\tau = 0.5$. In setup (VI), the τ-varying effect pattern of $\widetilde{Z}$ is similar to that in setup (V) except that there is a small interval around 0.5 where $\widetilde{Z}$ has no effect in setup(VI).

We compare the proposed method with the Wald test based on the Cox PH model, denoted by "CPH (Wald)", as well as the Wald test based on the locally concerned quantile regression that focuses on $\tau = 0.4, 0.5$, or 0.6, denoted by "CQR (Wald)". To implement CQR (Wald), we adopt Peng and Huang (2008)'s estimates with variance estimated by bootstrapping. The resampling size used for both CQR (Wald) and the proposed testing procedures is set as 2500. In the sequel, we shall refer the testing procedures based on $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ respectively to as GST and GIT. For all the methods, we consider sample sizes 200, 400, and 800. We set $\mathcal{U} = \{1, \ldots, 6\}$ when implementing the algorithm for selecting the constant u.
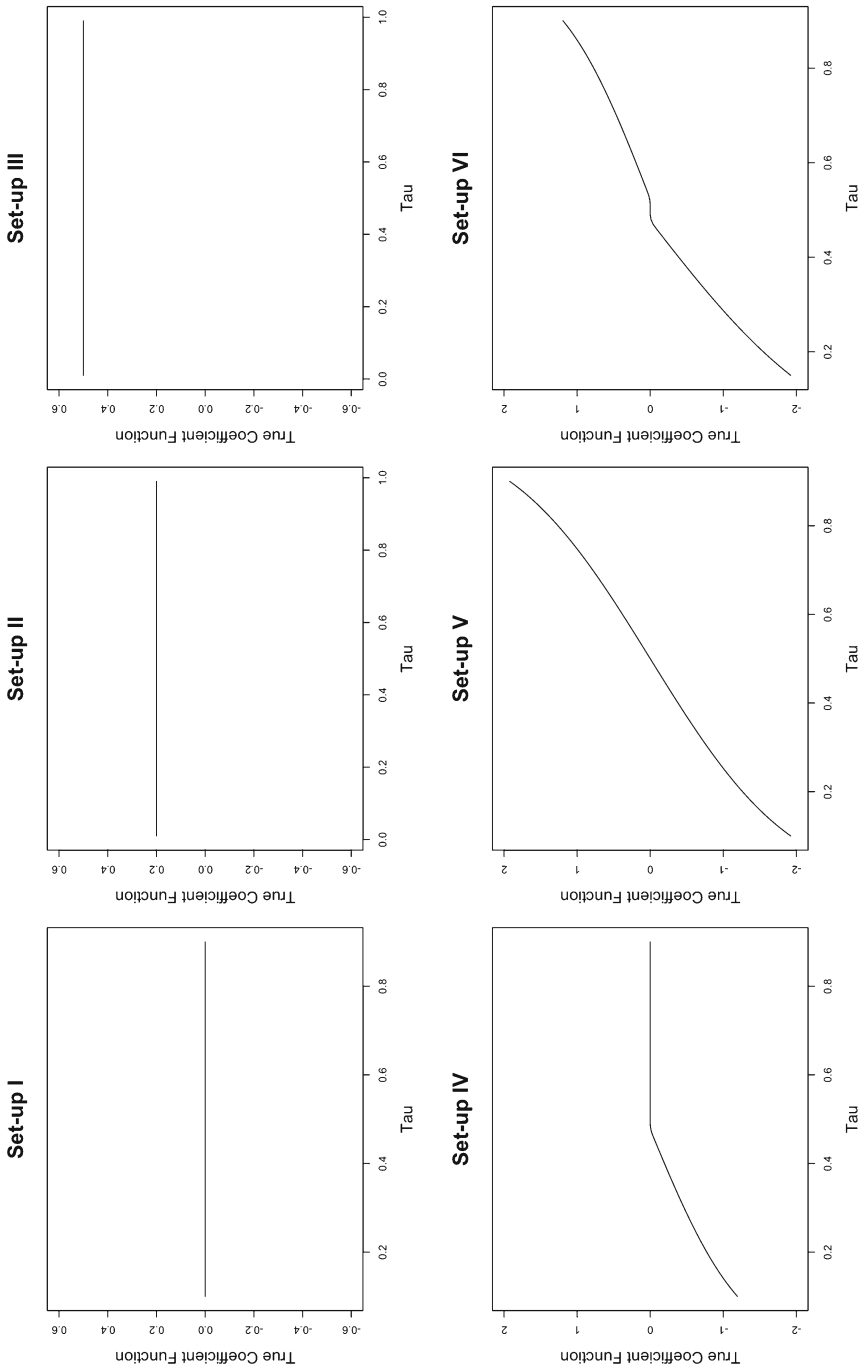
Fig. 1 The true coefficient function for all simulation set-ups

In Table 1, we report the empirical rejection rates based on 1000 simulations. The results in setup I show that the proposed GIT, and the existing tests, CQR (Wald) and CPH (Wald), have empirical sizes quite close to the nominal level 0.05. The proposed GST yields relatively larger empirical type I errors as compared to the other tests. The empirical size of GST equals 0.1 when the sample size is 200 but decreases to 0.077 when the sample size increases to 800. Such an anti-conservative behavior of GST is not surprising because the K–S type test statistic is defined based on the largest value of $\widehat{\beta}^{(1)}(\tau)/\widehat{\sigma}_n^{(1)}(\tau)$ over $\tau \in [0.1, 0.6]$, which is more sensitive to a possible outlying value of $\widehat{\sigma}_n^{(1)}(\tau)$ at some τ.

When the quantile effect of $\widetilde{Z}$ is constant over τ (i.e. setups (II) and (III)), we note that in setup (II) where the effect size (i.e. magnitude of the constant effect) is relatively small, CPH (Wald) has lower empirical power as compared to the proposed GIT and GST, and the power improvement associated with the proposed GIT and GST is more evident with the smaller sample size 200. In setup (III), where the effect size is larger, CPH (Wald) still generally has lower empirical power compared to the proposed tests but its empirical power becomes comparable to that of GIT when the sample size is large (i.e. $n = 800$). These observations suggest that even in the trivial constant effect cases, the proposed tests can outperform the traditional Cox regression based tests in data scenarios with small effect sizes or sample sizes. In both setups (II) and (III), the locally concerned CQR (Wald) consistently yields lower empirical power than the proposed globally concerned GIT and GST. This reflects the power benefit resulted from integrating information on covariate effects on different quantiles as in GST and GIT, rather than focusing on the covariate effect on a single quantile as in CQR (Wald).

In setups (IV), (V), and (VI), the effect of $\widetilde{Z}$ is τ-varying, reflecting its dynamic association with T. In these cases, CPH (Wald), which assumes a constant covariate effect, can have poor power to detect the dynamic effect of $\widetilde{Z}$ (e.g. 8.3% empirical power in setup (VI) with $n = 800$ in the presence of 30% censoring), while the proposed GST and GIT may yield much higher power (e.g. >99% power in setup (VI) with $n = 800$ in the presence of 30% censoring). The locally concerned CQR (Wald) can have higher power than CPH (Wald) when the targeted quantile level is within the τ-region where $\beta_0^{(1)}(\tau)$ is non-zero. When the targeted quantile level is outside the τ-region with non-zero effect, such as $\tau = 0.6$ in setup (IV) or $\tau = 0.5$ in setups (V) and (VI), the CQR (Wald) has even poorer power compared to CPH (Wald). This is well expected because these cases may serve as the null cases for the locally concerned CQR (Wald). This confirms that CQR (Wald) is inadequate to capture the meaningful effect of $\widetilde{Z}$ that is manifested at non-targeted quantiles.

We compare the simulation results across settings that are only differed by the censoring distribution. For each relationship between $\widetilde{Z}$ and T specified by setups (I)-(VI), we consider three different censoring distributions to yield 0%, 15%, and 30% censoring. The results for settings with 15% and 30% censoring are presented in Table 1 and the results based on uncensored data are presented in Table 3 in Appendix E. From our comparisons, we find that quantile regression based tests, including GST, GIT and CQR (Wald), demonstrate small variations in empirical powers as the censoring rate (or distribution) changes. In cases with a constant covariate effect, the Cox regression based test, CPH (Wald), also has similar performance among settings with different

Table 1 Empirical rejection rate based on 1000 simulations

Set-up	n	Proposed test		CQR (Wald)			CPH (Wald)
		GST	GIT	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	
15% censoring							
I	200	0.100	0.073	0.066	0.062	0.057	0.049
	400	0.091	0.078	0.072	0.072	0.066	0.051
	800	0.077	0.055	0.064	0.063	0.059	0.061
II	200	0.234	0.167	0.117	0.131	0.117	0.115
	400	0.275	0.214	0.155	0.153	0.150	0.178
	800	0.410	0.362	0.277	0.265	0.247	0.322
III	200	0.566	0.485	0.359	0.401	0.360	0.450
	400	0.786	0.772	0.585	0.592	0.576	0.722
	800	0.957	0.957	0.873	0.887	0.865	0.960
IV	200	0.377	0.254	0.097	0.060	0.053	0.063
	400	0.652	0.478	0.116	0.065	0.063	0.067
	800	0.939	0.816	0.148	0.047	0.058	0.090
V	200	0.653	0.464	0.143	0.070	0.118	0.260
	400	0.937	0.827	0.208	0.071	0.153	0.458
	800	0.999	0.993	0.291	0.053	0.279	0.757
VI	200	0.731	0.552	0.149	0.062	0.086	0.125
	400	0.971	0.896	0.198	0.055	0.095	0.201
	800	1.000	0.995	0.260	0.033	0.142	0.364
30% censoring							
I	200	0.171	0.095	0.062	0.060	0.048	0.047
	400	0.110	0.085	0.069	0.074	0.065	0.056
	800	0.066	0.052	0.063	0.059	0.050	0.038
II	200	0.302	0.186	0.115	0.122	0.105	0.122
	400	0.305	0.221	0.152	0.156	0.138	0.188
	800	0.411	0.359	0.277	0.259	0.245	0.298
III	200	0.681	0.539	0.360	0.393	0.322	0.432
	400	0.828	0.791	0.585	0.590	0.534	0.703
	800	0.959	0.957	0.874	0.877	0.855	0.952
IV	200	0.440	0.271	0.101	0.061	0.044	0.056
	400	0.668	0.480	0.115	0.065	0.062	0.085
	800	0.947	0.804	0.150	0.048	0.046	0.089
V	200	0.799	0.573	0.135	0.069	0.103	0.092
	400	0.960	0.846	0.206	0.068	0.135	0.140
	800	1.000	0.993	0.292	0.054	0.282	0.211
VI	200	0.803	0.587	0.148	0.063	0.077	0.053
	400	0.978	0.903	0.199	0.052	0.082	0.064
	800	1.000	0.995	0.263	0.033	0.141	0.083

censoring rates. However, in setup (V), where the covariate effect is not constant over τ, CPH (Wald) has reasonably good power when there is no censoring or only 15% censoring, but its performance deteriorates considerably when the censoring rate is increased to 30%. We have a similar observation for CPH (Wald) in setup (VI). A reasonable interpretation of these observations is that the capacity to detect a dynamic effect can be weakened by incorrectly assuming a constant proportional hazard effect and can be further attenuated by the missing data from censoring.

We also investigate whether the proposed tests are sensitive to the choice of $\mathcal{U}$. We conduct additional simulation studies with $\mathcal{U}$ set as $\{1, \ldots, 3\}$, $\{1, \ldots, 6\}$, and $\{1, \ldots, 12\}$ for the six set-ups with 15% censoring. The results are summarized in Table 4 in the Appendix. From this table, we note that GIT is quite robust to the change in $\mathcal{U}$, while GST demonstrates more variations across different choices of $\mathcal{U}$. Another observation is that GIT becomes less sensitive to the change in $\mathcal{U}$ when the sample size becomes larger. A possible explanation for these results is similar to that for the observed anti-conservative behavior of GST. That is, GST, by its construction, is sensitive to any outlying value of $\widehat{\sigma}_n^{(1)}(\tau)$ with $\tau \in [\tau_L, \tau_U]$, which is more likely to occur when the sample size is not large.

Aligning with the definitions of the proposed tests, the simulation results suggest that GST, as compared to GIT, is more sensitive to detect a departure from the null hypothesis, yielding higher power. This observation is also consistent with the anti-conservative behavior of GST observed in the null cases, which is reflected by empirical sizes notably greater than 0.05. With a smaller sample size, such as $n = 200$, GST can produce quite elevated type I errors, while GIT yields more reasonable empirical sizes. Therefore, in practice, one may need to exercise caution for applying GST to a small dataset, for which we recommend using GIT instead.

In summary, our simulation results demonstrate the proposed testing procedures have robust satisfactory performance for detecting a covariate of either a constant or dynamic effect. The new tests tend to exhibit greater advantages over benchmark approaches when the covariate presents a dynamic effect, or the covariate has a constant effect but of a small magnitude.

4 Real data analysis

To illustrate the utility of the proposed testing framework, we apply our method to investigate the prognostic factors for dialysis survival based on a dataset collected from a cohort of 191 incident dialysis patients (Kutner et al. 2002). In this dataset, time to death is censored in about 35% of dialysis patients due to either renal transplantation or end of the study as of December 31, 2005. In our analysis, we consider six potential prognostic factors (or covariates), which include age in years (AGE), indicator of reporting fish consumption over the first year of dialysis (FISHH), the indicator for baseline HD dialysis modality (BHDPD); whether the patient has severe symptoms of restless leg syndrome or not (BLEGS); whether or not education level is equal or higher than college (HIEDU); and the indicator of being in the black race group (BLACK). In our analyses, we standardize AGE by subtracting the sample mean and then dividing the resulting quantity by the sample standard deviation.

As a part of exploratory analyses, we check the proportional hazard assumption for each covariate based on Grambsch and Therneau (1994)'s method, using the R function `cox.zph()` in the R package `survival`. The p values corresponding to AGE, FISHH, BHDPD, BLEGS, HIEDU and BLACK are 0.43, 0.63, 0.55, 0.0006, 0.047 and 0.0004, respectively. These results suggest that the proportional hazard assumption may be violated for BLEGS, HIEDU and BLACK.

We fit model (1) for time to death (i.e. T) with each covariate separately. We set $[\tau_L, \tau_U]$ as [0.1, 0.6] for FISHH, BLGES, HIEDU, and BLACK, but set $[\tau_L, \tau_U]$ as [0.1, 0.54] and [0.1, 0.49] respectively for AGE and BHDPD. This is because the estimation of $\beta_0^{(1)}(\tau)$ based on Peng and Fine (2009) does not converge for some τ's larger than 0.54 and 0.49 when $\tilde{Z}$ is AGE or BHDPD. Figure 2 presents the estimated coefficients with the pointwise 95% confidence interval across $\tau \in [\tau_L, \tau_U]$. It is suggested by Fig. 2 that AGE and BLACK have strong and persistent effects across all or most quantiles of time to death, implying an apparent survival advantage for younger or black patients. For each of the rest covariates, FISHH, BHDPD, BLEGS, or HIEDU, we note a partial effect pattern. For example, FISHH and BLEGS may only impact some lower quantiles of the survival time. BHDPD and HIEDU may only have quantile effects in the τ-intervals, [0.15, 0.3] and [0.3, 0.4], respectively. These observations suggest the presence of dynamic covariate effects as well as the need to appropriately accommodate such dynamic covariate effects.

To evaluate each potential prognostic factor considered, we apply the proposed testing procedures, GST and GIT, along with the benchmark methods, CPH (Wald) and CQR (Wald), as described in Sect. 3. Table 2 summarizes the p values obtained from different methods for evaluating each covariate. We note that all tests consistently suggest a strong effect of AGE or BLACK on the survival time. The locally concerned quantile regression tests, CQR (Wald), reveal τ-varying effects of FISHH, BHDPD, BLEGS, and HIEDU. For example, BLEGS may significantly influence the 10th and 20th quantiles of the survival time but not the 30th, 40th, 50th, 60th of quantiles. HIEDU may also have a partial effect, influencing some quantiles, such as the 30th and 40th quantiles, but not the other quantiles. The classic Cox regression based test, CPH (Wald), however, fails to capture the partial effects of BLEGS and HIEDU. The p values for testing the effect of BLEGS and HIEDU based on CPH (Wald) are 0.35 and 0.25 respectively. This is possibly caused by imposing a restrictive static view on how a covariate can influence the survival time. In contrast, the proposed GIT and GST, through simultaneously examining covariate effects at quantile levels $[\tau_L, \tau_U]$, are able to detect the partial effect of BLEGS, with small p values ≤ 0.001 and to suggest a trend toward the association between HIEDU and the survival time, with marginal p values 0.01 and 0.09. The proposed GIT and GST also provide some evidence for the dynamic prognostic value of FISHH and BHDPD for dialysis survival. For example, as suggested by CQR (Wald), fish consumption in the first year may benefit dialysis patients with shorter survival time but may manifest little effect on the long-term survival. In general, our analysis results are consistent with the analyses of Peng and Huang (2008) based on multivariate censored quantile regression model. This example demonstrates the good practical utility of the proposed methods when varying covariate effects are present.
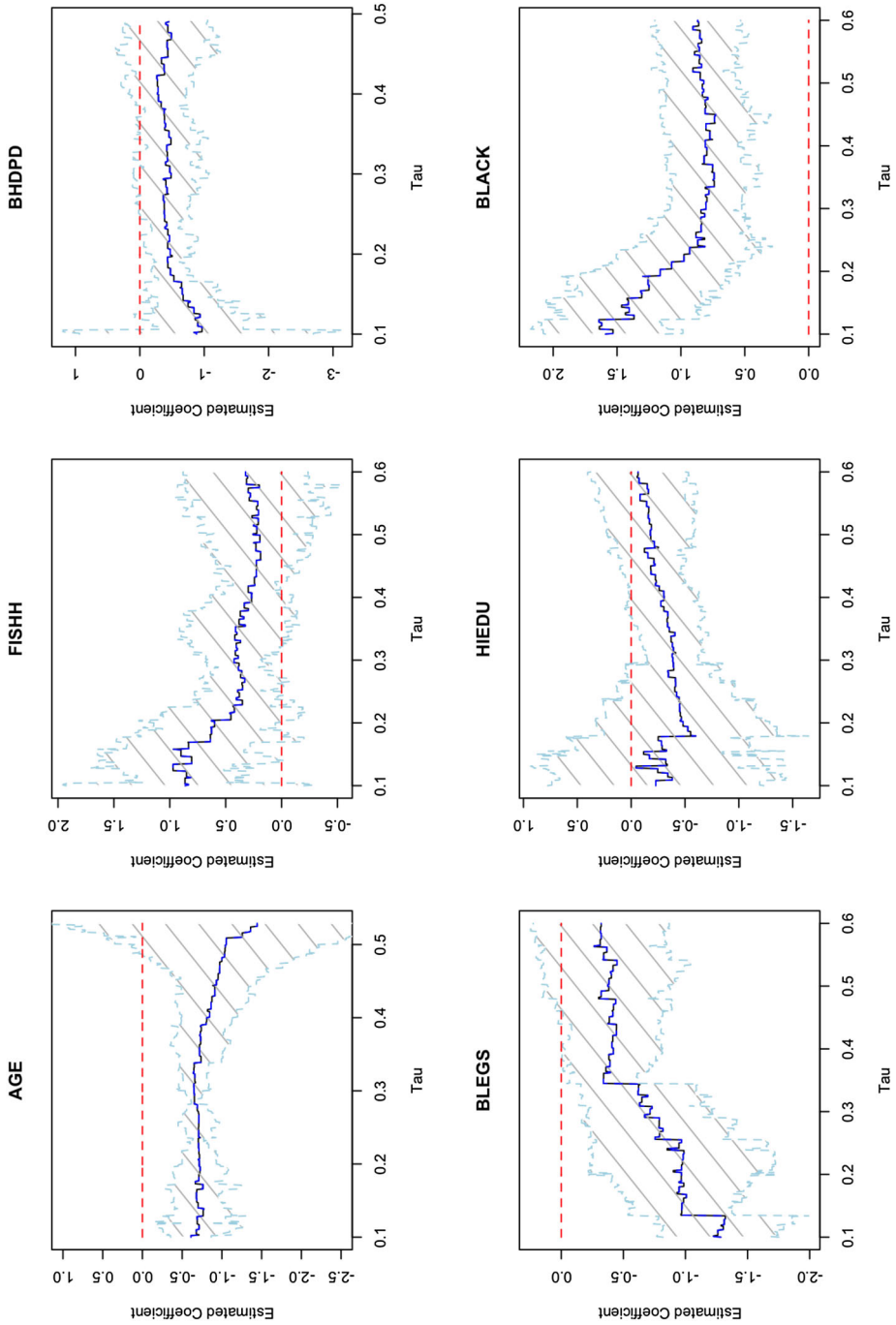
Fig. 2 The estimated coefficient with the 95% confidence interval for the covariates based on the censored quantile regression model on the dialysis data

Table 2 A summary of p values for each covariate with different methods

Covariate	Proposed test		CQR (Wald)						CPH (Wald)
	GST	GIT	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	
AGE	<0.001	<0.001	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
FISHH	<0.001	0.018	0.037	0.055	0.036	0.214	0.473	0.316	0.026
BHDPD	<0.001	0.005	0.090	0.021	0.152	0.228	0.229	0.030	0.008
BLEGS	<0.001	0.001	<0.001	0.001	0.062	0.082	0.091	0.507	0.349
HIEDU	0.013	0.093	0.596	0.137	0.003	0.032	0.068	0.241	0.245
BLACK	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

5 Discussion

In this paper, we develop a new testing framework for evaluating a survival prognostic factor. The main thrust of the new framework lies in its flexibility of accommodating a dynamic covariate effect, which is achieved through adapting the spirit of globally concerned quantile regression. Our testing procedures are conveniently developed based on existing results on fitting a working quantile regression model with randomly censored data. It is important to note that the validity of the testing procedures does not require that the working model is the true model. Moreover, the proposed methods can be readily extended to handle more complex survival outcomes, such as time to event subject to competing risks.

As suggested by one referee, we would like to point out that $Q_T(\tau|\widetilde{Z}) = Q_T(\tau)$ for $\tau \in (0, 1)$ implies the statistical independence between T and $\widetilde{Z}$. Nevertheless, in this work, we confine our attention to H_0^* with τ_U less than 1. This is because right censoring typically precludes the information on the upper tail of the distribution of T, and thus $Q_T(\tau)$ or $Q_T(\tau|\widetilde{Z})$ can become non-identifiable as τ approaches 1. The null hypothesis H_0^* entails a weaker version of the independence between T and $\tilde{Z}$ that can be better assessed with right censored data. Rejecting H_0^* can provide evidence for the dependence between T and $\tilde{Z}$, while accepting H_0^* may not sufficiently indicate the independence between T and $\tilde{Z}$.

Another commendable extension of this work is to generalize the current null hypothesis and testing procedures to permit evaluating multiple prognostic factors simultaneously. This work also lays a key foundation for developing a nonparametric screening method for helping identify useful prognostic factors among a large number of candidates. These extensions will be reported in separate work.

Appendix A: Lemma 1 and its proof

Lemma 1 *Suppose the conditional distribution function of T given $\widetilde{Z} = \tilde{z}$ is continuous and strictly monotone for all possible values of $\tilde{z}$. Then $Q_T(\tau|\widetilde{Z}) = Q_T(\tau)$ for $\tau \in [\tau_L, \tau_U]$ is equivalent to model (1) holds with $\Delta = [\tau_L, \tau_U]$ and $\beta_0^{(1)}(\tau) = 0$ for $\tau \in [\tau_L, \tau_U]$.*

Proof for Lemma 1 Suppose we have $Q_T(\tau|\widetilde{Z}) = Q_T(\tau)$ for $\tau \in [\tau_L, \tau_U]$. It is clear that for $\tau \in [\tau_L, \tau_U]$, we can write $Q_T(\tau|\widetilde{Z}) = \exp\{Z^T\theta_0(\tau)\}$ with $\theta_0(\tau) = (\log Q_T(\tau), 0)^T$. This means that model (1) holds with $\Delta = [\tau_L, \tau_U]$ and $\beta_0^{(1)}(\tau) = 0$ for $\tau \in [\tau_L, \tau_U]$.

Suppose model (1) holds with $\Delta = [\tau_L, \tau_U]$ and $\beta_0^{(1)}(\tau) = 0$ for $\tau \in [\tau_L, \tau_U]$. This means, $Q_T(\tau|\widetilde{Z}) = \exp\{\beta_0^{(0)}(\tau)\}$ for $\tau \in [\tau_L, \tau_U]$. Given that the conditional distribution function of T given $\widetilde{Z}$ is continuous and strictly monotone, it follows from the definition of $Q_T(\tau|\widetilde{Z})$ that $\Pr(T \leq \exp\{\beta_0^{(0)}(\tau)\}|\widetilde{Z}) = \tau$ for $\tau \in [\tau_L, \tau_U]$. Taking expectation on both sides of this equality with respect to $\widetilde{Z}$, we then get $\Pr(T \leq \exp\{\beta_0^{(0)}(\tau)\}) = \tau$ for $\tau \in [\tau_L, \tau_U]$. Given the continuity and strict monotonicity of the distribution function of T, which is implied by the continuity and strict monotonicity of the conditional distribution function of T given $\widetilde{Z}$, this implies that $\exp\{\beta_0^{(0)}(\tau)\} = Q_T(\tau)$. Thus, $Q_T(\tau|\widetilde{Z}) = Q_T(\tau)$ for $\tau \in [\tau_L, \tau_U]$. This completes the proof of Lemma 1.

Appendix B: Asymptotic properties of $\widehat{\theta}$ without assuming model (1)

We assume the following regularity conditions:

(C1) There exist a constant v such that $P(C = v) > 0$ and $P(C > v) = 0$.

(C2) $\widetilde{Z}$ is uniformly bounded, i.e. $\sup_i |\widetilde{Z}_i| < \infty$.

(C3) (i) $\widetilde{\theta}(\tau)$ is Lipschitz continuous for $\tau \in [\tau_L, \tau_U]$; (ii) $f(y|z)$ is bounded above uniformly in y and z, where $f(y|z)$ denotes the conditional density of X given $Z = z$.

(C4) For some $\rho_0 > 0$ and $c_0 > 0$, $\inf_{b \in \mathcal{B}(\rho_0)} \text{eigmin} A(b) \geq c_0$, where $\mathcal{B}(\rho) = \{b \in R^2 : \inf_{\tau \in [\tau_L, \tau_U]} ||b - \widetilde{\theta}(\tau)|| \leq \rho\}$ and $A(b) = E[ZZ^T f(Z^T b|Z)]$. Here $|| \cdot ||$ is the Euclidean norm and $\text{eigmin} A(b)$ represents the minimal eigenvalue of $A(b)$.

Condition (C1) is adopted to simplify the theoretical arguments to ensure that $\widehat{G}(\cdot)$ is consistent for $G(\cdot)$. This condition is usually satisfied in studies subject to administrative censoring. Condition (C2) imposes covariate boundedness. Condition (C3) assumes that the limit coefficient process is smooth and the conditional density distribution is bounded and smooth. Condition (C4) requires that the asymptotic limit of $U_n(b, \tau)$ is strictly convex in a neighborhood of $\widetilde{\theta}(\tau)$ for $\tau \in [\tau_L, \tau_U]$, implying the uniqueness of the solution to $\mu(b, \tau) \equiv E\{ZI(\log T \leq Z^T b) - \tau\} = 0$. This plays a critical role in establishing the uniform convergence of $\widehat{\theta}(\tau)$ to $\widetilde{\theta}(\tau)$.

Theorem A1 *Under regularity conditions (C1)–(C4), we have*

$$\lim_{n \to \infty} \sup_{\tau \in [\tau_L, \tau_U]} ||\widehat{\theta}(\tau) - \widetilde{\theta}(\tau)|| \to_p 0.$$

Theorem A2 *Under regularity conditions (C1)–(C4), we have $\sqrt{n}(\widehat{\theta}(\tau) - \widetilde{\theta}(\tau))$ converge weakly to a mean zero Gaussian process for $\tau \in [\tau_L, \tau_U]$ with covariance*

$$\widetilde{\Phi}(\tau', \tau) = E\{\widetilde{\xi}_1(\tau')\widetilde{\xi}_1(\tau)^T\}.$$

The proofs of Theorems A1 and A2 closely resemble the proofs in Peng and Fine (2009) and thus are omitted.

Appendix C: Proofs of Theorems 1 and 2

We assume one additional regularity condition:

(C5) $\inf_{\tau \in [\tau_L, \tau_U]} \sigma^{(1)}(\tau) > 0$, where $\{\sigma^{(1)}(\tau)\}^2$ is the second diagonal element of $\widetilde{\boldsymbol{\Phi}}(\tau, \tau)$.

Proof of Theorem 1 Following the lines of Peng and Fine (2009), we can show that the sample-based variance estimation procedure presented in Sect. 2.1 provides consistent variance estimation, which implies $\sup_{\tau \in (\tau_L, \tau_U]} |\widehat{\sigma}_n^{(1)}(\tau) - \sigma^{(1)}(\tau)| \to_p 0$.

Note that under the null hypothesis H_0^*, we have $\widetilde{\beta}^{(1)}(\tau) = 0$ and consequently,

$$
n^{1/2}\widehat{R}(\tau) = \frac{n^{1/2}\{\widehat{\beta}^{(1)}(\tau) - \widetilde{\beta}^{(1)}(\tau)\}}{\widehat{\sigma}_n^{(1)}(\tau)} = \frac{n^{1/2}\{\widehat{\beta}(\tau) - \widetilde{\beta}^{(1)}(\tau)\}}{\sigma^{(1)}(\tau)} \left(\frac{\sigma^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} - 1 \right)
$$
$$
+ \frac{n^{1/2}\{\widehat{\beta}^{(1)}(\tau) - \widetilde{\beta}^{(1)}(\tau)\}}{\sigma^{(1)}(\tau)}. \tag{5}
$$

By Theorem A2, $n^{1/2}\{\widehat{\beta}^{(1)}(\tau) - \widetilde{\beta}^{(1)}(\tau)\}/\sigma^{(1)}(\tau)$ converges weakly to a mean zero Gaussian process $\mathcal{X}^{(1)}(\tau)$ with covariance process

$$
\widetilde{\Phi}^{(1)}(\tau, \tau') = \frac{\widetilde{\boldsymbol{\Phi}}^{(2,2)}(\tau, \tau')}{\sigma^{(1)}(\tau)\sigma^{(1)}(\tau')},
$$

where $\widetilde{\boldsymbol{\Phi}}^{(2,2)}(\tau, \tau')$ denotes the element in the second row and the second column of $\widetilde{\boldsymbol{\Phi}}(\tau, \tau')$. In addition, condition (C5) and $\sup_{\tau \in (\tau_L, \tau_U]} |\widehat{\sigma}_n^{(1)}(\tau) - \sigma^{(1)}(\tau)| \to_p 0$ imply $\sup_{\tau \in (\tau_L, \tau_U]} \left| \frac{\sigma^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} - 1 \right| \to_p 0$. Applying the result of Theorem A2 and the Slutsky's Theorem (line 11 of Example 1.4.7 in Boucheron et al. (2013)) to (5), we then get $n^{1/2}\widehat{R}(\tau) \to_d \mathcal{X}^{(1)}(\tau)$ in $l^\infty(\mathcal{F}_T)$, where $l^\infty(S)$ is the collection of all bounded functions $f : S \mapsto R$ for any index set S and $\mathcal{F}_T = \{\frac{\widetilde{\xi}_1^{(1)}(c, \tau)}{\sigma^{(1)}(\tau)}, c \in R^2, \tau \in [\tau_L, \tau_U]\}$. Then, by the extended continuous mapping theorem (Theorem 1.11.1 in van der Vaart et al. (1996)), we can establish the limiting null distribution for $\widehat{T}_{sup}^{(1)}$ and $\widehat{T}_{inte}^{(1)}$ as

$$
\widehat{T}_{sup}^{(1)} = \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{n^{1/2}\widehat{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right| = \sup_{\tau \in [\tau_L, \tau_U]} \left| n^{1/2}\widehat{R}(\tau) \right| \to_d \sup\{|\mathcal{X}^{(1)}(\tau)|, \tau \in [\tau_L, \tau_U]\},
$$

$$
\widehat{T}_{inte}^{(1)} = \int_{\tau_L}^{\tau_U} \left| \frac{n^{1/2}\widehat{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right|^2 d\tau = \int_{\tau_L}^{\tau_U} \left| n^{1/2}\widehat{R}(\tau) \right|^2 d\tau \to_d \int_{\tau_L}^{\tau_U} \{\mathcal{X}^{(1)}(\tau)\}^2 d\tau.
$$

This completes the proof of Theorem 1. □

Proof for Theorem 2 We first investigate the asymptotic limit of $\widehat{T}_{sup}^{(1)}$ under the alternative hypothesis $H_{a,1}$. Simple algebra shows that

$$
\widehat{T}_{sup}^{(1)} = \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{n^{1/2} \widehat{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right| = \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{n^{1/2} \widetilde{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} + \frac{n^{1/2} (\widehat{\beta}^{(1)}(\tau) - \widetilde{\beta}^{(1)}(\tau))}{\widehat{\sigma}_n^{(1)}(\tau)} \right|
$$

$$
\geq \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{n^{1/2} \widetilde{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right| - \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{n^{1/2} (\widehat{\beta}^{(1)}(\tau) - \widetilde{\beta}^{(1)}(\tau))}{\widehat{\sigma}_n^{(1)}(\tau)} \right| \equiv \widehat{T}_{sup,1}^{(1)} - \widehat{T}_{sup,2}^{(1)}.
$$

By the extended continuous mapping theorem, we can show that the $\widehat{T}_{sup,2}^{(1)}$ converges in distribution to $\sup_{\tau \in [\tau_L, \tau_U]} |\mathcal{X}^{(1)}(\tau)|$ and thus is $O_p(1)$. At the same time, given $\sup_{\tau \in (\tau_L, \tau_U]} |\widehat{\sigma}_n^{(1)}(\tau) - \sigma^{(1)}(\tau)| \to_p 0$, under condition (C5), we get $n^{-1/2} \widehat{T}_{sup,1}^{(1)} \to_p v_0$, where $v_0 = \sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{\widetilde{\beta}^{(1)}(\tau)}{\sigma^{(1)}(\tau)} \right|$.

Under the alternative hypothesis $H_{a,1}$ and condition (C5), we have $v_0 > 0$, and hence $P(n^{-1/2} \widehat{T}_{sup,1}^{(1)} > v_0/2) \to P(v_0 > v_0/2) = 1$ as $n \to \infty$. Furthermore, for any $a > 0$, we have $n^{-1/2} \widehat{T}_{sup,2}^{(1)} + a \cdot n^{-1/2} = o_p(1)$, which implies $P(n^{-1/2} \widehat{T}_{sup,2}^{(1)} + a \cdot n^{-1/2} > v_0/2) \to 0$ as $n \to \infty$. Note that

$$
P(\widehat{T}_{sup}^{(1)} > a) \geq P(n^{-1/2} \widehat{T}_{sup,1}^{(1)} > n^{-1/2} \widehat{T}_{sup,2}^{(1)} + a \cdot n^{-1/2})
$$

$$
\geq P(n^{-1/2} \widehat{T}_{sup,1}^{(1)} > v_0/2) - P(n^{-1/2} \widehat{T}_{sup,2}^{(1)} + a \cdot n^{-1/2} > v_0/2).
$$

It then follows that $P(\widehat{T}_{sup}^{(1)} > a) \to 1$ as $n \to \infty$ under the alternative hypothesis $H_{a,1}$. This immediately implies that $\widehat{T}_{sup}^{(1)}$ is a consistent test against $H_{a,1}$ because $P(\widehat{T}_{sup}^{(1)} > C_{sup,\alpha}) \to 1$ as $n \to \infty$ given $H_{a,1}$ holds, where $C_{sup,\alpha}$ denotes the α-level critical value determined upon the limit null distribution of $\widehat{T}_{sup}^{(1)}$, which is greater than 0.

Next, we consider $\widehat{T}_{inte}^{(1)}$ under the alternative hypothesis $H_{a,2}$. Write $\widehat{T}_{inte}^{(1)}$ as

$$
\widehat{T}_{inte}^{(1)} = \int_{\tau_L}^{\tau_U} \left| \frac{n^{1/2} \widehat{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right|^2 d\tau = \int_{\tau_L}^{\tau_U} \left| \frac{n^{1/2} \widetilde{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} - \frac{n^{1/2} (\widetilde{\beta}^{(1)}(\tau) - \widehat{\beta}^{(1)}(\tau))}{\widehat{\sigma}_n^{(1)}(\tau)} \right|^2 d\tau
$$

$$
\geq \int_{\tau_L}^{\tau_U} \left| \frac{n^{1/2} \widetilde{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right|^2 d\tau - \int_{\tau_L}^{\tau_U} 2 \left| \frac{n^{1/2} \widetilde{\beta}^{(1)}(\tau)}{\widehat{\sigma}_n^{(1)}(\tau)} \right| \cdot \left| \frac{n^{1/2} (\widetilde{\beta}^{(1)}(\tau) - \widehat{\beta}^{(1)}(\tau))}{\widehat{\sigma}_n^{(1)}(\tau)} \right| d\tau
$$

$$
\equiv \widehat{T}_{inte,1}^{(1)} - \widehat{T}_{inte,2}^{(1)}.
$$

By the continuous mapping theorem, combined with $\sup_{\tau \in (\tau_L, \tau_U]} |\widehat{\sigma}_n^{(1)}(\tau) - \sigma^{(1)}(\tau)| \to_p 0$ and condition (C5), we get $n^{-1} \widehat{T}_{inte,1}^{(1)} \to_p v_0^*$, where $v_0^* = \int_{\tau_L}^{\tau_U} \left| \frac{\widetilde{\beta}^{(1)}(\tau)}{\sigma^{(1)}(\tau)} \right|^2 d\tau$, and

$$n^{-1/2}\widehat{T}^{(1)}_{inte,2} \to_d \int_{\tau_L}^{\tau_U} 2 \left| \frac{\widetilde{\beta}^{(1)}(\tau)}{\sigma^{(1)}(\tau)} \right| \cdot \{\mathcal{X}^{(1)}(\tau)\} d\tau$$

and thus $O_p(1)$. By condition (C5), the alternative hypothesis $H_{a,2}$ implies $v_0^* > 0$. Then following the same arguments for showing $P(\widehat{T}^{(1)}_{sup} > a) \to 1$ for any $a > 0$ based on the results that $n^{-1/2}\widehat{T}^{(1)}_{sup,1} \to_p v_0 > 0$ and $\widehat{T}^{(1)}_{sup,2} = O_p(1)$, we can prove that $P(n^{-1/2}\widehat{T}^{(1)}_{inte} > a) \to 1$ as $n \to \infty$ for any $a > 0$ under $H_{a,2}$. This implies that $P(\widehat{T}^{(1)}_{inte} > a) \to 1$ as $n \to \infty$ for any $a > 0$ under $H_{a,2}$. Therefore, $\widehat{T}^{(1)}_{inte}$ is a consistent test against the alternative hypothesis $H_{a,2}$. $\qquad\square$

Appendix D: Justification for the proposed resampling procedure

Given the observed data denoted by $\{\boldsymbol{O}_i\}_{i=1}^n \equiv \{(X_i, \delta_i, \tilde{Z}_i)\}_{i=1}^n$, since $\{\iota_i^b\}_{i=1}^n$ are i.i.d. standard normal random variables, we have

$$E\left\{ \frac{n^{-1/2}\sum_{i=1}^n \widehat{\xi}_i^{(1)}(\tau)\iota_i^b}{\widehat{\sigma}_n^{(1)}(\tau)} \cdot \frac{n^{-1/2}\sum_{i=1}^n \widehat{\xi}_i^{(1)}(\tau')\iota_i^b}{\widehat{\sigma}_n^{(1)}(\tau')} \middle| \{\boldsymbol{O}_i\}_{i=1}^n \right\}$$

$$= n^{-1}\sum_{i=1}^n \frac{\widehat{\xi}_i^{(1)}(\tau)\widehat{\xi}_i^{(1)}(\tau')}{\widehat{\sigma}_n^{(1)}(\tau)\widehat{\sigma}_n^{(1)}(\tau')} \to_p \widetilde{\Phi}^{(1)}(\tau, \tau').$$

By the arguments of Lin et al. (1993), the distribution of $n^{-1/2}\sum_{i=1}^n \widehat{\xi}_i^{(1)}(\tau)\iota_i^b/\widehat{\sigma}_n^{(1)}(\tau)$ converges weakly to $\mathcal{X}^{(1)}(\tau)$, the same limit as that of $n^{1/2}\{\widehat{\beta}^{(1)}(\tau) - \widetilde{\beta}^{(1)}(\tau)\}/\widehat{\sigma}_n^{(1)}(\tau)$, for almost all realizations of $\{\boldsymbol{O}_i\}_{i=1}^n$. Applying the extended continuous mapping theorem as in the proof of Theorem 2, we have that under H_0^*, the conditional distribution of $\widehat{T}^{(1)}_{sup,b}$ (or $\widehat{T}^{(1)}_{inte,b}$) given the observed data is asymptotically equivalent to the unconditional distributions of $T^{(1)}_{sup,b}$ (or $T^{(1)}_{inte,b}$). This justifies using the resampling procedure in Sect. 2.3 to obtain the p values of the proposed tests.

Appendix E: Additional simulation results

See Tables 3 and 4.

Table 3 Empirical rejection rate for the uncensored case based on 1000 simulations

Set-up	n	Proposed test		CQR (Wald)			CPH (Wald)
		GST	GIT	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	
I	200	0.098	0.070	0.055	0.052	0.056	0.048
	400	0.093	0.075	0.069	0.064	0.060	0.047
	800	0.076	0.058	0.053	0.053	0.048	0.061
II	200	0.215	0.156	0.104	0.108	0.108	0.121
	400	0.275	0.216	0.162	0.156	0.139	0.183
	800	0.420	0.372	0.276	0.265	0.238	0.328
III	200	0.541	0.478	0.344	0.374	0.337	0.456
	400	0.790	0.771	0.589	0.595	0.590	0.745
	800	0.958	0.961	0.883	0.886	0.873	0.963
IV	200	0.378	0.250	0.074	0.045	0.049	0.060
	400	0.656	0.476	0.101	0.056	0.055	0.049
	800	0.935	0.808	0.118	0.034	0.045	0.085
V	200	0.618	0.452	0.106	0.057	0.121	0.428
	400	0.939	0.828	0.169	0.071	0.165	0.737
	800	1.000	0.994	0.255	0.041	0.313	0.968
VI	200	0.729	0.543	0.095	0.047	0.088	0.228
	400	0.971	0.898	0.154	0.048	0.097	0.446
	800	1.000	0.995	0.243	0.020	0.154	0.756

Table 4 Empirical rejection rate for the proposed test with different choices of $\mathcal{U}$ on the six set-ups subject to 15% censoring based on 1000 simulations

Set-up	n	$\mathcal{U} = \{1, \ldots, 3\}$		$\mathcal{U} = \{1, \ldots, 6\}$		$\mathcal{U} = \{1, \ldots, 12\}$	
		GST	GIT	GST	GIT	GST	GIT
I	200	0.128	0.074	0.091	0.067	0.092	0.065
	400	0.126	0.079	0.086	0.067	0.081	0.063
	800	0.112	0.060	0.080	0.059	0.072	0.058
II	200	0.287	0.178	0.228	0.161	0.225	0.158
	400	0.359	0.231	0.283	0.218	0.256	0.206
	800	0.472	0.379	0.415	0.369	0.376	0.361
III	200	0.666	0.549	0.593	0.510	0.585	0.513
	400	0.841	0.789	0.779	0.773	0.761	0.757
	800	0.975	0.964	0.956	0.956	0.940	0.957
IV	200	0.427	0.257	0.364	0.242	0.362	0.243
	400	0.702	0.490	0.666	0.470	0.649	0.471
	800	0.952	0.808	0.942	0.811	0.936	0.808
V	200	0.695	0.478	0.649	0.452	0.649	0.446
	400	0.962	0.850	0.948	0.831	0.944	0.827
	800	1.000	0.994	1.000	0.991	1.000	0.993
VI	200	0.768	0.558	0.723	0.534	0.726	0.535
	400	0.981	0.902	0.971	0.894	0.970	0.892
	800	1.000	0.997	1.000	0.997	1.000	0.998

References

Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pélissier S (2010) Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med Res Methodol 10:1–12

Boucheron S, Lugosi G, Massart P (2013) Concentration inequalities: a nonasymptotic theory of independence. Oxford University Press, Oxford

Cox DR, Oakes D (2018) Analysis of survival data. Chapman and Hall, London

Darling DA (1957) The Kolmogorov–Smirnov, Cramer–von Mises tests. Ann Math Stat 28:823–838

Dickson ER, Grambsch PM, Fleming TR, Fisher LD, Langworthy A (1989) Prognosis in primary biliary cirrhosis: model for decision making. Hepatology 10:1–7

Grambsch PM, Therneau TM (1994) Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 81:515–526

Huang Y (2010) Quantile calculus and censored regression. Ann Stat 38:1607–1637

Kleinbaum DG, Klein M (2010) Survival analysis, vol 3. Springer, Berlin

Koenker R (2005) Quantile regression, vol 38. Cambridge University Press, Cambridge

Koenker R (2022) quantreg: quantile regression. R package version 5.87

Koenker R, Bassett G (1978) Regression quantiles. Econom J Econom Soc 46:33–50

Kutner NG, Clow PW, Zhang R, Aviles X (2002) Association of fish intake and survival in a cohort of incident dialysis patients. Am J Kidney Dis 39:1018–1024

Li R, Peng L (2014) Varying coefficient subdistribution regression for left-truncated semi-competing risks data. J Multivar Anal 131:65–78

Lin DY, Wei L-J, Ying Z (1993) Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika 80:557–572

Peng L (2021) Quantile regression for survival data. Annu Rev Stat Its Appl 8:413–437

Peng L, Fine JP (2009) Competing risks quantile regression. J Am Stat Assoc 104:1440–1453

Peng L, Huang Y (2008) Survival analysis with quantile regression models. J Am Stat Assoc 103:637–649

Portnoy S (2003) Censored quantile regression. J Am Stat Assoc 98:1001–1012

Powell JL (1986) Censored regression quantiles. J Econom 32:143–155

Thorogood J, Persijn G, Schreuder GM, D'amaro J, Zantvoort F, Van Houwelingen J, Van Rood J (1990) The effect of HLA matching on kidney graft survival in separate posttransplantation intervals. Transplantation 50:146–150

van der Vaart A, van der Vaart A, van der Vaart A, Wellner J (1996) Weak convergence and empirical processes: with applications to statistics. Springer series in statistics. Springer, Berlin

Verweij PJ, van Houwelingen HC (1995) Time-dependent effects of fixed covariates in Cox regression. Biometrics 51:1550–1556

Wang H, Wang L (2009) Locally weighted censored quantile regression. J Am Stat Assoc 104:1117–1128

Ying Z, Jung S, Wei L (1995) Survival analysis with median regression models. J Am Stat Assoc 90:178–184

Zheng Q, Peng L, He X (2015) Globally adaptive quantile regression with ultra-high dimensional data. Ann Stat 43:2225

Zhou L (2006) A simple censored median regression estimator. Stat Sin 16:1043–1058