Computational Statistics and Data Analysis 88 (2015) 154–172

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Quantile regression methods with varying-coefficient models for censored data

Shangyu Xie [a,*], Alan T.K. Wan [b], Yong Zhou [c,d]

[a] RCAF and School of Banking and Finance, University of International Business and Economics, Beijing, China

[b] Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong

[c] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

[d] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China

ARTICLE INFO

Article history:
Received 11 July 2014
Received in revised form 12 February 2015
Accepted 14 February 2015
Available online 11 March 2015

Keywords:
Composite quantile regression
Covariate-dependent censoring
Inverse probability weighting
MM algorithm
Perturbation resampling

ABSTRACT

Considerable intellectual progress has been made to the development of various semi-parametric varying-coefficient models over the past ten to fifteen years. An important advantage of these models is that they avoid much of the curse of dimensionality problem as the nonparametric functions are restricted only to some variables. More recently, varying-coefficient methods have been applied to quantile regression modeling, but all previous studies assume that the data are fully observed. The main purpose of this paper is to develop a varying-coefficient approach to the estimation of regression quantiles under random data censoring. We use a weighted inverse probability approach to account for censoring, and propose a majorize–minimize type algorithm to optimize the non-smooth objective function. The asymptotic properties of the proposed estimator of the nonparametric functions are studied, and a resampling method is developed for obtaining the estimator of the sampling variance. An important aspect of our method is that it allows the censoring time to depend on the covariates. Additionally, we show that this varying-coefficient procedure can be further improved when implemented within a composite quantile regression framework. Composite quantile regression has recently gained considerable attention due to its ability to combine information across different quantile functions. We assess the finite sample properties of the proposed procedures in simulated studies. A real data application is also considered.

1. Introduction

Since the seminal work of Koenker and Bassett (1978), there has been an abundance of literature on various applications and theoretical extensions of quantile regression (QR). Regression quantiles have the important advantage over conditional mean regression of being able to directly estimate the effects of the covariates on quantiles other than the center of the distribution. It is also well-known that compared to the method of least-squares (LS), QR is more robust to outliers. QR has been extensively applied in economics, finance, biology, medicine, and many other disciplines. Recent empirical studies involving applications of QR can be found in Wheelock and Wilson (2008), Li et al. (2010), among others.

Although Koenker and Basset's (1978) conventional QR estimator is based on a linear parametric set-up, there has been a rapidly growing literature on the statistical theory and implementation of nonparametric and semiparametric QRs.

* Corresponding author.
E-mail address: xshyu@amss.ac.cn (S. Xie).

http://dx.doi.org/10.1016/j.csda.2015.02.011
0167-9473/

For example, Koenker et al. (1994) discussed quantile smoothing splines; Yu and Jones (1998) considered nonparametric regression quantile estimation by kernel weighted linear fitting; De Gooijer and Zerom (2003) modeled the conditional quantile of the response as a nonlinear additive function of the covariates; and Wei and He (2006) developed a global QR approach to conditional growth charts. One important class of nonparametric models that has gained considerable attention in recent years is the varying-coefficient approach proposed by Cleveland et al. (1991) and Hastie and Tibshirani (1993). The 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 important advantage the varying-coefficient approach has experienced rapid development in theory and methodology. We refer to the articles by Cai (2007) for novel adaptations of the varying-coefficient approach to time series analysis; Fan and Li (2004) for longitudinal data analysis; Fan et al. (2006) and Cai et al. (2007, 2008) for survival analysis, and Wu et al. (2010) for functional linear regression. For more references, see Fan and Zhang (2008). To the best of our knowledge, Honda (2004), Kim (2007) and Cai and Xu (2008) are the only existing studies that consider the varying-coefficient approach for conditional quantiles; Honda (2004) and Cai and Xu (2008) used local polynomials to estimate conditional quantiles with varying coefficients, while Kim (2007) proposed an estimation methodology based on polynomial splines.

In recent years, we have also seen the emergence of a parallel literature on censored QR for which the usual set-up is one where the dependent variable of interest cannot be completely observed due to censoring. Censored QR was first studied by Powell (1986) for fixed censoring that assumes known censoring times for all observations. The examination of censored QR under the assumption of random censoring with unknown censoring points was taken up in a series of studies by Lindgren (1997), Yang (1999), Honoré et al. (2002), Gannoun et al. (2005) and Chen (2010). Portnoy (2003) studied censored QR under the self-consistency principle for the Kaplan–Meier estimator and developed a recursively re-weighted estimation procedure. Peng and Huang (2008) proposed a martingale-based estimating equations approach for censored QR models. This approach was subsequently extended by Qian and Peng (2010) to the analysis of a partially functional QR. Huang (2010) proposed a procedure for estimating censored QR based on estimating integral equations. The preponderance of this literature emphasizes nonparametric estimation of the conditional quantiles. However, some of the methods proposed in these studies rely on very strong distributional assumptions, or have major computational and/or theoretical drawbacks. For example, Lindgren's (1997) method requires an iterative minimization procedure for which no theoretical justification has been provided; this method is also computationally cumbersome when the dimension of the problem is high. The approach of Yang (1999) is restricted only to i.i.d. errors, and involves solving some highly complicated non-linear equations that can lead to multiple solutions. Portnoy (2003) also noted the algorithmic complications and computational issues associated with his proposed procedure. With few exceptions, most existing studies assume that censoring is independent of the covariates although it is not uncommon in practice to find correlations between the censoring time in the dependent variable and the covariates. For example, probability of loan default is typically thought to be associated with the borrower's credit worthiness reflected in the covariates. Dependence of censoring on covariates in QR is considered in the work of Portnoy (2003), Peng and Huang (2008) and Huang (2010), but none of these studies consider a varying-coefficient approach.

The major objective of the current paper is to develop a varying-coefficient approach to the estimation of regression quantiles under random data censoring when censoring times depend on covariates. We propose a weighted estimating function approach (Robins et al., 1994), whereby the contribution to the estimating function from an uncensored observation is weighted by the inverse of the probability of its being fully observed (Bang and Tsiatis, 2000). Inverse probability weighting is a widely used approach in censored data studies and has been adopted and refined in many subsequent studies (e.g., Zhao and Tian, 2001, Bang and Tsiatis, 2002 and Wang et al., 2012). One difficulty with this approach, however, is that the resultant estimating functions are non-smooth, rendering the Newton–Raphson algorithm inapplicable in solving the estimating equations. To reconcile this difficulty, we draw on the majorize–minimize (MM) algorithm (Hunter and Lange, 2000), and adapt it for the current analysis. A perturbation resampling method is developed for obtaining the estimator of the sampling variance. When censoring depends on the covariates, we can model this dependency either parametrically through, for example, the proportional hazards or the additive Aalen models, or nonparametrically by the Kaplan–Meier (Kaplan and Meier, 1958) estimator. Another purpose of this paper is to show that our procedure can be further improved when implemented within a composite quantile regression (CQR) framework. CQR was first introduced by Zou and Yuan (2008) for estimating coefficients in a linear regression. Extensions to local polynomial regression and varying-coefficient models were undertaken more recently by Kai et al. (2010, 2011). CQR has the appealing strength of combining information from different QRs, and hence the potential to improve estimation efficiency. All of the above-mentioned studies have shown that CQR can yield substantially more efficient estimators than LS-based procedures, but none of these studies have allowed for the possibility of censored data. This paper makes some progress towards obtaining results for CQR under censored data. Although this work may be thought of as an extension of Kai et al. (2011), the extension being considered is no means straightforward. Indeed, a very different set of analytical techniques is needed for obtaining results under random censoring using an inverse probability weighting approach. The efficiency gains of CQR over LS and QR in finite samples are examined by simulations.

The remainder of this paper is organized as follows. In Section 2, we present the model framework, and the local polynomial fitting and inverse probability weighting mechanisms, and describe the adaptation of the MM algorithm to the present analysis. In Section 3, we examine the asymptotic properties of the estimators, and provide a perturbation resampling method for variance estimation. Section 4 develops a CQR approach in the context of the censored varying-coefficient QR model. In Section 5, we report results of simulation studies that examine the performance of the proposed methods in finite samples. In the same section, we also present an empirical application. Proofs of theorems are contained in an Appendix B and an online supplemental file.

2. Varying-coefficient quantile regression

2.1. Model specification

Let T be the response variable, $\mathbf{X} = (X_1, \ldots, X_p)^T$ and U be observed covariates, and $\mathbf{a}(U) = (a_1(U), \ldots, a_p(U))^T$ be the unknown coefficient functions capturing the effects of the covariates. For a given $\tau (0 < \tau < 1)$, the varying-coefficient model assumes that the conditional QR of T is expressed as

$$Q_\tau(T|\mathbf{X} = \mathbf{x}, U = u) = \mathbf{x}^T \mathbf{a}_\tau(u), \tag{2.1}$$

where $\mathbf{a}_\tau(u) = (a_{1,\tau}, \ldots, a_{p,\tau})^T$ is a vector of smooth varying-coefficient functions of u, and $a_{j,\tau}$'s, $j = 1, \ldots, p$, may depend on τ. This model allows the coefficients of $\mathbf{X}$ to change with U, the effect modifier. Here, we assume that U is a single variable, but in general, U can be a low-dimensional vector of variables. The objective is to obtain the τ-quantile of T, given $\mathbf{X}$ and U, through estimating $\mathbf{a}_\tau(u)$ nonparametrically. Because only low-dimensional functions are estimated, the curse of dimensionality problem can be avoided even if p is large. This is the thrust of the varying-coefficient approach. This framework also offers the benefit of interpretability as it permits one to explore how the regression coefficients vary over different values of the effect modifier. In general, the sign of T is unrestricted.

Previous studies of the varying-coefficient model for conditional quantiles by Honda (2004), Kim (2007) and Cai and Xu (2008) all assume that the sample values of T_i, $i = 1, 2, \ldots, n$, are fully observed in a sample with n observations. In this paper, we assume instead that T_i is subject to random right censoring. Let C_i be the censoring variable, $V_i = \min(T_i, C_i)$, and $\Delta_i = I(T_i \le C_i)$, where $I(.)$ is an indicator function. Due to censoring, we observe V_i and Δ_i instead of T_i. We assume, for analytical convenience, that $\{C_i\}_{i=1}^n$ are i.i.d., and $Pr\{C_i \ge l_{x,u}|x, u\} > 0$, where $0 < l_{x,u} = \inf\{t : Pr(T_i \ge t|x, u) = 0\} < \infty$, for any given values of $\{x, u\}$. By this assumption, the support of C_i also covers that of T_i conditional on the covariates $\{x, u\}$. Furthermore, inference is assumed to be restricted to the interval $[0, L]$ to ensure that all regression parameters are estimable, with L being chosen such that $\inf_{x,u} Pr\{T \ge L|x, u\} > 0$. We allow the distribution of C_i to depend on $\mathbf{X}$ and U, but conditional on $\mathbf{X}$ and U, T_i and C_i are assumed to be independent.

Let $\rho_\tau(y) = y[\tau - I(y < 0)]$ be the check loss function at $\tau \in (0, 1)$. The QR estimator of $\mathbf{a}_\tau(u)$ in (2.1) when data are fully observed can be obtained by minimizing the quantile loss function

$$\sum_{i=1}^n \rho_\tau\left(T_i - \mathbf{X}_i^T \mathbf{a}_\tau(U_i)\right). \tag{2.2}$$

Now, write $\phi_\tau(y) = \tau - I(y < 0)$. By noting that $E[\phi_\tau(T_i - \mathbf{X}_i^T \mathbf{a}_\tau(U_i))] = 0$, the minimizer of (2.2) is also the root to the estimating equation (Ying et al., 1995)

$$\sum_{i=1}^n \phi_\tau\left(T_i - \mathbf{X}_i^T \mathbf{a}_\tau(U_i)\right) \approx 0.$$

Hence it is reasonable to use $\sum_{i=1}^n \phi_\tau(T_i - \mathbf{X}_i^T \mathbf{a}_\tau(U_i))$ as the estimating function for $\mathbf{a}(u)$.

We propose an inverse probability weighting approach with local linear smoothing for estimating $\mathbf{a}(U)$. This approach accounts for the random censoring to which T is subject and $\mathbf{a}(U)'s$ being nonparametric functions. Now, let $\mathbf{a} = (a_1, \ldots, a_p)^T$ and $\mathbf{b} = (b_1, \ldots, b_p)^T = (a_1'(\cdot), \ldots, a_p'(\cdot))^T$ be vectors of real constants. Assume that $a_j(u)$ is twice continuously differentiable so that the function $a_j(\cdot)$ can be approximated locally by $a_j(u) \approx a_j + b_j(u - u_0)$, with u in the neighborhood of a given point u_0. Write $\boldsymbol{\beta} = (\mathbf{a}^T, \mathbf{b}^T)^T$, and let $\boldsymbol{\beta}_0(.) = (\mathbf{a}^T(.), \mathbf{a}'(.)^T)^T$ be the vector of the true parameter functions. If T_i's are fully observed with no censoring, the local quantile regression estimator $\widehat{\boldsymbol{\beta}} = (\widehat{\mathbf{a}}^T, \widehat{\mathbf{b}}^T)^T$ is obtained by minimizing

$$\frac{1}{n} \sum_{i=1}^n \rho_\tau(T_i - \mathbf{Z}_i^T \boldsymbol{\beta}) K\left(\frac{U_i - u_0}{h}\right), \tag{2.3}$$

where $\mathbf{Z}_i = (\mathbf{X}_i^T, \mathbf{X}_i^T(U_i - u_0))^T$, $K(\cdot)$ is a bounded (kernel) function, and $h = h_n > 0$ is a bandwidth parameter. With T_i now subject to random censoring, we adopt the approach of inverse probability weighting (Robins et al., 1994), whereby the contribution of an uncensored observation to the objective function is weighted by the inverse of the probability of being fully observed. Let $G(\cdot|\mathbf{X}, U)$ be the conditional survival distribution of the censoring variable C, conditional on $\{\mathbf{X}, U\}$. The estimating functions in (2.3) may be replaced by the inverse probability weighted estimating function

$$\frac{1}{n} \sum_{i=1}^n \frac{\Delta_i}{G(T_i|\mathbf{X}_i, U_i)} \rho_\tau(T_i - \mathbf{Z}_i^T \boldsymbol{\beta}) K\left(\frac{U_i - u_0}{h}\right). \tag{2.4}$$

It is worth emphasizing that although our analytical framework assumes the dependence of the survival function of C on $\{\mathbf{X}, U\}$, it also permits the case where C and the covariates are independent.

One can then derive an estimator of $\boldsymbol{\beta}$ from (2.4) by replacing the unknown survival distribution $G(t|x, u)$ by a uniformly consistent estimator $\widehat{G}(t|x, u)$ for $\{x, u\}$ and $0 \leq t \leq L$. If censoring is independent of the covariates, $G(t)$ can be estimated by the Kaplan–Meier estimator with the roles of the censoring time C_i and survival time T_i reversed; that is, estimate $G(\cdot)$ by

$$\widehat{G}(t) = \prod_{s \leq t} \left\{ 1 - \frac{dN^c(s)}{Y(s)} \right\}, \tag{2.5}$$

where $N^c(s) = \sum_{i=1}^n I(V_i \leq s, \Delta_i = 0)$ and $Y(s) = \sum_{i=1}^n I(V_i \geq s)$. When censoring depends on the covariates, $G(t|x, u)$ can be estimated by the nonparametric version of the Kaplan–Meier estimator (Wang and Wang, 2009), or a model specified for the censoring time—for example, the Cox or the additive Aalen regression models.

In the following, let $\widehat{G}(t|x, u)$ be an estimator of $G(t|x, u)$. Assume that $\widehat{G}(t|x, u)$ is uniformly strongly consistent for x, u and t (with $0 \leq t \leq L$), as well as being regular, asymptotically linear with influence function Ψ; that is,

$$\widehat{G}(t|x, u) - G(t|x, u) = n^{-1} \sum_{i=1}^n \Psi(t, x, u, W_i) + o_p(n^{-1/2}), \tag{2.6}$$

where $W_i = \{V_i, \Delta_i, \mathbf{X}_i, U_i\}$. Note that the Kaplan–Meier estimator satisfies (2.6) whether or not censoring depends on the covariates (see Examples 5.5.2, 6.6.1. of Bickel et al., 1993 and Martinussen and Scheike, 2006). When the dependency of censoring on the covariates is modeled via the Cox model, (2.6) is satisfied by the maximum partial likelihood and the Breslow estimators of the probability of censoring (Bickel et al., 1993). It is also satisfied by the additive Aalen and the accelerated failure time and the linear transformation models (Martinussen and Scheike, 2006), provided that the model is correctly specified. In practice, it is often difficult to specify a parametric model that correctly describes the failure time T. An alternative is to use a nonparametric approach based on the Kaplan–Meier estimator. Specifically, we can estimate $G(t|x, u)$ by

$$\widehat{G}(t|x, u) = \prod_{i=1}^n \left\{ 1 - \frac{B_{ni}(x, u)}{\sum_{j=1}^n I_{\{V_j \geq V_i\}} B_{nj}(x, u)} \right\}^{\delta_i(t)}, \tag{2.7}$$

where $\delta_i(t) = I(V_i \leq t, \Delta_i = 0)$, and $B_{ni}(x, u)$ is a sequence of non-negative weights that sum to 1. When $B_{ni}(x, u) = 1/n$, $\widehat{G}(t|x, u)$ is simply the classical Kaplan–Meier estimator given in (2.5) of the survival function of C_i, $i = 1, \ldots n$. Here, we employ the common Nadaraya–Watson type weights:

$$B_{nj}(w) = \frac{L\left(\frac{w - w_j}{b_n}\right)}{\sum_{i=1}^n L\left(\frac{w - w_i}{b_n}\right)}, \tag{2.8}$$

where $L(\cdot)$ is a density kernel function, and $b_n \in \Re^+$ is a bandwidth that converges to zero as $n \to \infty$. It can be shown that (2.7) that uses (2.8) as weights satisfies (2.6) (Wang and Wang, 2009). Substituting $\widehat{G}(t|x, u)$ for $G(t|x, u)$ in (2.4) leads to

$$\frac{1}{n} \sum_{i=1}^n \frac{\Delta_i}{\widehat{G}(T_i)} \rho_\tau(T_i - \mathbf{Z}_i^T \boldsymbol{\beta}) K\left(\frac{U_i - u_0}{h}\right), \tag{2.9}$$

when censoring is independent of the covariates, or

$$\frac{1}{n} \sum_{i=1}^n \frac{\Delta_i}{\widehat{G}(T_i|\mathbf{X}_i, U_i)} \rho_\tau(T_i - \mathbf{Z}_i^T \boldsymbol{\beta}) K\left(\frac{U_i - u_0}{h}\right), \tag{2.10}$$

when censoring depends on the covariates.

Let $\widehat{\boldsymbol{\beta}}_{\tau k}$, $k = 1, 2$, be the estimator of $\boldsymbol{\beta}$ obtained by minimizing (2.9) and (2.10) respectively. Thus, the estimator of $(a_1(u_0), \ldots, a_p(u_0))^T$ is $\widehat{\mathbf{a}}_{\tau k} = e_p \widehat{\boldsymbol{\beta}}_{\tau k}$, where e_p is a $p \times 2p$ matrix, with the first p diagonal elements taking on a value of unity and all other elements zero.

2.2. MM algorithm

Our objective is to minimize (2.9) and (2.10). With this in mind, we have

$$L(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i K_i}{\widehat{G}(T_i|\mathbf{X}_i, U_i)}\rho_\tau(T_i - \mathbf{Z}_i^T\boldsymbol{\beta}),$$

where $K_i = K((U_i - u_0)/h)$, and $\widehat{G}(T_i|\mathbf{X}_i, U_i)$ can be either dependent or independent of the covariates $\{\mathbf{X}_i, U_i\}$; in the case of the latter, $\widehat{G}(T_i|\mathbf{X}_i, U_i)$ can be written as $\widehat{G}(T_i)$. Note that the traditional Newton–Raphson algorithm is inapplicable here due to the non-smoothness of these functions. We adopt the majorize–minimize (MM) algorithm, which is a widely used technique for optimizing non-smooth objective functions. This algorithm works by finding a surrogate function that minorizes or majorizes the objective function. Readers may consult Hunter and Lange (2000) for a general exposition of this algorithm.

To obtain a surrogate function, we follow the approach of Hunter and Lange (2000) that begins with $\rho_\tau(T_i - \mathbf{Z}_i^T\boldsymbol{\beta})$. Write $r_i = T_i - \mathbf{Z}_i^T\boldsymbol{\beta}$, and define, for $\bar{\delta} > 0$, the perturbation

$$\rho_\tau^{\bar{\delta}}(r) = \rho_\tau(r) - \frac{\bar{\delta}}{2}\ln(\bar{\delta} + |r|).$$

Then the sum

$$L_{\bar{\delta}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i K_i}{\widehat{G}(T_i|\mathbf{X}_i, U_i)}\rho_\tau^{\bar{\delta}}(r_i) \tag{2.11}$$

is approximately equal to $L(\boldsymbol{\beta})$. It turns out that for a given residual $r^j = r(\boldsymbol{\beta}^j) \equiv T - \mathbf{Z}^T\boldsymbol{\beta}^j$ at the jth iteration, $\rho_\tau^{\bar{\delta}}(r)$ is majorized at r^j through the quadratic function

$$\zeta_\tau^{\bar{\delta}}(r|r^j) = \frac{1}{4}\Big[\frac{r^2}{\bar{\delta} + |r^j|} + (4\tau - 2)r + c\Big],$$

where c is a constant such that $\zeta_\tau^{\bar{\delta}}(r^j|r^j) = \rho_\tau^{\bar{\delta}}(r^j)$. This result can be proved along the lines of Proposition A.2 of Hunter and Lange (2000). The MM algorithm then proceeds by minimizing the majorizer

$$Q_{\bar{\delta}}(\boldsymbol{\beta}|\boldsymbol{\beta}^j) = \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i K_i}{\widehat{G}(T_i|\mathbf{X}_i, U_i)}\zeta_\tau^{\bar{\delta}}(r_i|r^j), \tag{2.12}$$

and the value of $\boldsymbol{\beta}$ which minimizes $Q_{\bar{\delta}}(\boldsymbol{\beta}|\boldsymbol{\beta}^j)$ becomes $\boldsymbol{\beta}^{j+1}$ in the next iteration. This algorithm thus converts a non-smooth function $L(\boldsymbol{\beta})$ into a sequence of smooth functions $Q_{\bar{\delta}}(\boldsymbol{\beta}|\boldsymbol{\beta}^j)$. Traditional methods such as the Newton–Raphson method can then be implemented on the converted functions.

The above approach will be used in our simulation study and empirical examples.

3. Asymptotic properties and variance estimation

In this section, we prove that $\widehat{\boldsymbol{\beta}}_{\tau k}$, $k = 1, 2$, is consistent as well as asymptotically normal. We also develop an empirical estimator of the covariance matrix of $\widehat{\boldsymbol{\beta}}_{\tau k}$.

3.1. Asymptotic results

Recall that $\widehat{\boldsymbol{\beta}}_{\tau k} = (\widehat{\mathbf{a}}_{\tau k}^T, \widehat{\mathbf{b}}_{\tau k}^T)^T$. Let $\eta(u, \mathbf{X}) = \mathbf{X}^T\mathbf{a}(u)$, $\mu_l = \int u^l K(u)du$, $\nu_l = \int u^l K^2(u)du$, $l = 0, 1, \ldots, f_U(\cdot)$ be the marginal density of U, and $\mathbf{H} = diag(1, h) \otimes \mathbf{I}_p$, with $\otimes$ denoting the Kronecker product. Following Fleming and Harrington (1991), we define the intensity function of the random variable T as $\lambda(t) = \lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}P\{t \le T < t+\Delta t|T \ge t\}$. As discussed in Section 2.1, both $\widehat{G}(t)$ in (2.9) (under covariate-independent censoring) and $\widehat{G}(t|x, u)$ in (2.10) (under covariate-dependent censoring) satisfy (2.6). Thus, we have the following theorems:

Theorem 1 (*Consistency*). *Assume that $\{V_i, \mathbf{X}_i, U_i, \Delta_i\}$, $i = 1, 2, \ldots, n$, constitutes an i.i.d. multivariate random sample, and the censoring variable C_i is independent of T_i, but dependent on the covariates $\mathbf{X}_i$ and U_i. Assume also that the technical conditions in the Appendix are satisfied, and $h = h_n \to 0$ and $nh \to \infty$ as $n \to \infty$. Then we have, as $n \to \infty$,*

$$\widehat{\boldsymbol{\beta}}_{\tau k} \to \boldsymbol{\beta}_0$$

in probability, for $k = 1, 2$.

Theorem 2 (*Asymptotic Normality*). *Let $f_{\mathbf{x},U}(\cdot)$ and $F_{\mathbf{x},U}(\cdot)$ be the conditional density and cumulative distribution functions of T given $\mathbf{X}$ and U respectively, and $\lambda^c(\cdot)$ be the intensity function of the censoring variable C. Under the same set of assumptions as in Theorem 1, we have*

$$\sqrt{nh}\left[\mathbf{H}\{\widehat{\boldsymbol{\beta}}_{\tau k}(u_0)-\boldsymbol{\beta}_0(u_0)\}-\frac{h^2}{2(\mu_0\mu_2-\mu_1^2)}\begin{pmatrix}(\mu_2^2-\mu_1\mu_3)\mathbf{a}''(u_0)\\(\mu_0\mu_3-\mu_1\mu_2)\mathbf{a}''(u_0)\end{pmatrix}+o(h^2)\right]$$
$$\xrightarrow{\mathcal{L}} N(0,\tau(1-\tau)\mathbf{B}^{-1}\mathbf{A}_k\mathbf{B}^{-1}), \tag{3.1}$$

for $k=1,2$, where $\boldsymbol{\Gamma}(u_0)=E\left\{f_{\mathbf{x},U}(\eta(u_0,\mathbf{X}))\mathbf{X}\mathbf{X}^T|U=u_0\right\}$,

$$\mathbf{A}_1=f_U(u_0)\begin{pmatrix}\nu_0&\nu_1\\\nu_1&\nu_2\end{pmatrix}\otimes\left\{E[\mathbf{X}\mathbf{X}^T|U=u_0]+E\left[\varrho_1(U,\mathbf{X})\mathbf{X}\mathbf{X}^T|U=u_0\right]\right\},$$

$$\mathbf{A}_2=f_U(u_0)\begin{pmatrix}\nu_0&\nu_1\\\nu_1&\nu_2\end{pmatrix}\otimes\left\{E[\mathbf{X}\mathbf{X}^T|U=u_0]+E\left[\varrho_2(U,\mathbf{X})\mathbf{X}\mathbf{X}^T|U=u_0\right]\right\},$$

$$\mathbf{B}=f_U(u_0)\begin{pmatrix}\mu_0&\mu_1\\\mu_1&\mu_2\end{pmatrix}\otimes\boldsymbol{\Gamma}(u_0),$$

and

$$\varrho_1(U,\mathbf{X})=\frac{1}{\tau}\int_0^L\left[\tau+(1-\tau)F_{\mathbf{x},U}(s)\right]\frac{\lambda^c(s)}{G(s)}ds,\qquad \varrho_2(U,\mathbf{X})=\frac{1-G(T|\mathbf{X},U)}{G(T|\mathbf{X},U)}.$$

In particular, when $G(t|x,u)$ is estimated in accordance with (2.7) and (2.8),

$$\mathbf{A}_2=f_U(u_0)\begin{pmatrix}\nu_0&\nu_1\\\nu_1&\nu_2\end{pmatrix}\otimes\left\{E\left[\frac{\mathbf{X}\mathbf{X}^T}{G(T|\mathbf{X},U)}|U=u_0\right]+E\{\varrho_3(X,U)^{\otimes 2}\mathbf{X}\mathbf{X}^T|U=u_0\}\right\},$$

where $h(x,u)$ is the joint density function of $(\mathbf{X},U)$,

$$\varrho_3(\mathbf{X},U)=h(\mathbf{X},U)E\left[\frac{\varsigma(V,\Delta,T,\mathbf{X},U)\phi_\tau\{T-\mathbf{a}^T(U)\mathbf{X}\}}{G(T|\mathbf{X},U)}|\mathbf{X},U\right],$$

with

$$\varsigma(V_i,\Delta_i,t,x,u)=G(t|x,u)\left[\int_0^{\min(V_i,t)}\frac{g_0(s|x,u)ds}{G^2(s|x,u)(1-F_{\mathbf{x},u}(s|x,u))}+\frac{I(V_i\le t,\Delta_i=0)}{G(V_i|x,u)\{1-F_{\mathbf{x},u}(V_i|x,u)\}}\right],$$

where $g_0(s|x,u)$ is the first derivative of $G(s|x,u)$ with respect to s.

When the sample size n approaches infinity, by Theorem 1, $\widehat{\boldsymbol{\beta}}_{\tau k}$ converges to the true parameter value, and by Theorem 2, $\widehat{\boldsymbol{\beta}}_{\tau k}$ follows a normal distribution. The latter result is useful for constructing asymptotic confidence intervals of the unknowns. Note that the terms $E[\varrho_1(U,\mathbf{X})\mathbf{X}\mathbf{X}^T|U=u_0]$ in $\mathbf{A}_1$ and $E[\varrho_2(U,\mathbf{X})\mathbf{X}\mathbf{X}^T|U=u_0]$ in $\mathbf{A}_2$ arise as a result of censoring. If the data are uncensored and fully observed, the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{\tau k}$ reduces to a special case of the asymptotic variance expression derived by Cai and Xu (2008), who examined nonparametric quantile estimation for dynamic smooth coefficient models.

3.2. A resampling method for covariance estimation

Here, we employ a resampling method for computing the variance of $\widehat{\boldsymbol{\beta}}_{\tau k}$ to avoid the rather cumbersome direct computation of the variance. Our method generalizes the resampling method of Jin et al. (2001) by perturbing the minimand directly and repeatedly. To describe the method, let the perturbation estimating function be

$$\tilde{L}^*(\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^n\frac{\Delta_iK_i}{\widehat{G}^*(T_i|\mathbf{X}_i,U_i)}\rho_\tau(T_i-\mathbf{Z}_i^T\boldsymbol{\beta})\xi_i, \tag{3.2}$$

where ξ_i's are generated from an exponential distribution with mean and variance both equal to unity, and independently of $(V_i,\mathbf{X}_i,U_i,\Delta_i)_{i=1}^n$, $\widehat{G}^*$ is the perturbed version of $\widehat{G}$ (see (Yin and Cai, 2005)). This loss function is essentially a perturbed version of the original loss $L(\beta)$, with the perturbations caused by the random variables ξ_i, $i=1,\dots,n$. Note that the only random quantities in $\tilde{L}^*(\boldsymbol{\beta})$ are $\{\xi_i\}$'s, and $\tilde{L}^*(\boldsymbol{\beta})$ has the same mean and variance as $L(\boldsymbol{\beta})$. The MM algorithm previously described can be used to obtain the solution $\widehat{\boldsymbol{\beta}}_\tau^*$ for optimizing $\tilde{L}^*(\boldsymbol{\beta})$. It can be shown that $\sqrt{nh}[\mathbf{H}(\widehat{\boldsymbol{\beta}}_\tau^*-\widehat{\boldsymbol{\beta}}_\tau)]$ and $\sqrt{nh}[\mathbf{H}(\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{\beta}_0)]$ have the same asymptotic distribution (see Jin et al., 2006, or Yin and Cai, 2005). Based on the above results, the covariance matrix of $\widehat{\boldsymbol{\beta}}_\tau$ may be obtained from the empirical variance of $\widehat{\boldsymbol{\beta}}_\tau^*$.

Specifically, the steps of the perturbation resampling method are as follows:

Step 1: Treat the observed data $(V_i, \mathbf{X}_i, U_i, \Delta_i)_{i=1}^n$ as fixed, generate random observations of $\{\xi_i\}_{i=1}^n$ from an exponential distribution with mean and variance both equal to unity.

Step 2: Based on the sample of $\{\xi_i\}_{i=1}^n$ generated in Step 1, minimize $\tilde{L}^*(\boldsymbol{\beta})$ in (3.2) using the MM algorithm, and denote the estimator as $\widehat{\beta}_\tau^{*(1)}$.

Step 3: Repeat Steps 1 and 2 BS times and obtain $\widehat{\beta}_\tau^{*(m)}$, $m = 1, \ldots, BS$.

The empirical distribution of $\widehat{\boldsymbol{\beta}}_\tau^{*(m)}$ can then be used to approximate the distribution of $\widehat{\boldsymbol{\beta}}_\tau$. In particular, we use the empirical covariances of $\widehat{\boldsymbol{\beta}}_\tau^{*(m)}$ as estimates of the covariances of $\widehat{\boldsymbol{\beta}}_\tau$.

4. A varying-coefficient composite quantile regression approach

This section develops a varying-coefficient CQR approach to estimation under random censoring. As mentioned in Section 1, CQR was introduced in a series of recent papers by Zou and Yuan (2008) and Kai et al. (2010, 2011). It combines information from different QRs and has been shown to enjoy superior properties, both asymptotically and in finite samples, to LS-based estimators. Our development of CQR is based on the model

$$T = a_0(U) + \mathbf{X}^T\mathbf{a}(U) + \varepsilon \tag{4.1}$$

where $a_0(U)$ is a baseline function, $\mathbf{a}(U) = (a_1(U), \ldots, a_p(U))^T$ contains p unknown varying-coefficient functions, and ε follows a distribution F with mean zero. The zero mean assumption is reasonable as CQR is mainly intended to improve over the LS estimator. Kai et al. (2011) also made the same assumption in their study of CQR for the varying-coefficient partially linear model with completely observed data. Write $Q_\tau(T|\mathbf{X} = \mathbf{x}, U = u) = a_0(u) + c_\tau^* + \mathbf{x}^T\mathbf{a}(u)$, where $c_\tau^* = F^{-1}(\tau)$. Within this framework, the varying coefficients are common across all quantiles, and the same target quantile $\mathbf{a}(u)$ is estimated by different QR estimators $\widehat{\mathbf{a}}_\tau(u)$, each having the optimal rate of convergence.

Let $\{T_i, \mathbf{X}_i, U_i, i = 1, \ldots, n\}$ be an i.i.d. sample from model (4.1) and ε have a mean of zero, where T_i are censored observations, with $V_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \le C_i)$. For a given q, write $\tau_k = k/(q+1)$ for $k = 1, 2, \ldots, q$. The CQR procedure obtains estimates of $a_0(\cdot)$ and $\mathbf{a}(\cdot)$ by minimizing the loss function

$$\sum_{k=1}^{q}\Big[\sum_{i=1}^{n}\frac{\Delta_i}{G(T_i|\mathbf{X}_i, U_i)}\rho_{\tau_k}\{T_i - a_{0,k}(U_i) - \mathbf{X}_i^T\mathbf{a}(U_i)\}\Big]. \tag{4.2}$$

CQR combines information across multiple QRs, and imposes a single parameter on the slope. As the functional coefficients are approximated locally by linear models, substituting $\widehat{G}(T_i|\mathbf{X}_i, U_i)$ for $G(T_i|\mathbf{X}_i, U_i)$ in (4.2) leads to the locally weighted CQR loss function

$$\sum_{k=1}^{q}\Big[\sum_{i=1}^{n}\frac{\Delta_i}{\widehat{G}(T_i|\mathbf{X}_i, U_i)}\rho_{\tau_k}\big\{T_i - a_{0,k} - b_0(U_i - u) - \mathbf{X}_i^T\{\mathbf{a} + \mathbf{b}(U_i - u)\}\big\}K_h(U_i - u)\Big], \tag{4.3}$$

where $\mathbf{a}_0 = (a_{0,1}, \ldots, a_{0,K})^T$, $\mathbf{a} = (a_1, \ldots, a_p)^T$, $\mathbf{b} = (b_1, \ldots, b_p)^T$, $K(\cdot)$ is a kernel function, $K_h(\cdot) = K(\cdot/h)/h$, and h is a bandwidth. Let the minimizer of (4.3) be $(\widehat{\mathbf{a}}_0, \widehat{\mathbf{a}})$. The estimators of $a_0(u)$ and $\mathbf{a}(u)$ are then given by

$$\widehat{a}_0(u) = \frac{1}{q}\sum_{k=1}^{q}\widehat{a}_{0,k},$$

and

$$\widehat{\mathbf{a}}(u) = \widehat{\mathbf{a}}$$

respectively.

We now turn to the investigation of the asymptotic behavior of $\widehat{a}_0(u)$ and $\widehat{\mathbf{a}}(u)$. Let $f(\cdot)$ and $F(\cdot)$ be the density and cumulative distribution functions of the errors respectively. Write $c_k^* = F^{-1}(\tau_k)$ and let $\mathbf{l}$ be a $q \times 1$ identity vector, $\mathcal{C}$ be a $q \times q$ diagonal matrix with $\mathcal{C}_{jj} = f(c_j^*)$, $\mathcal{C}_1 = \mathcal{C}\mathbf{l}$, and $\mathcal{C}_0 = \mathbf{l}^T\mathcal{C}\mathbf{l}$, and

$$\Omega(u) = E\left[\begin{pmatrix}\mathcal{C} & \mathcal{C}_1\mathbf{X}^T\\ \mathbf{X}\mathcal{C}_1^T & \mathcal{C}_0\mathbf{X}\mathbf{X}^T\end{pmatrix}\Bigg|U = u\right].$$

Further, let $\tau_{kk'} = \tau_k \wedge \tau_{k'} - \tau_k\tau_{k'}$, and $\mathcal{T}$ be a $q \times q$ matrix with $\tau_{kk'}$ as the (k, k')-th element. Write $\mathcal{T}_1 = \mathcal{T}\mathbf{l}$, $\mathcal{T}_0 = \mathbf{l}^T\mathcal{T}\mathbf{l}$, and

$$\mathbf{\Sigma}_1(u) = E\left[\begin{pmatrix}\mathcal{T} & \mathcal{T}_1\mathbf{X}^T\\ \mathbf{X}\mathcal{T}_1^T & \mathcal{T}_0\mathbf{X}\mathbf{X}^T\end{pmatrix}\Bigg|U = u\right].$$

Let

$$\varrho_{kk'}(U, \mathbf{X}) = \int_0^L \left[\tau_{kk'} + (\tau_k\tau_{k'} - \tau_k - \tau_{k'} + 1)F_{X,U}(s)\right] \frac{\lambda^c(s)}{G(s)} ds,$$

and $\mathcal{K}$ be a $q \times q$ matrix with the (k, k') element being $\varrho_{kk'}(U, \mathbf{X})$. Write $\mathcal{K}_1 = \mathcal{K}\mathbf{l}$, $\mathcal{K}_0 = \mathbf{l}^T\mathcal{K}\mathbf{l}$,

$$\mathbf{\Sigma}_2(u) = E\left[\begin{pmatrix} \mathcal{K} & \mathcal{K}_1\mathbf{X}^T \\ \mathbf{X}\mathcal{K}_1^T & \mathcal{K}_0\mathbf{X}\mathbf{X}^T \end{pmatrix} \Big| U = u\right],$$

and

$$\mathbf{\Sigma}_3(u) = E\left[\varrho_2(\mathbf{X}, U)\begin{pmatrix} \mathcal{T} & \mathcal{T}_1\mathbf{X}^T \\ \mathbf{X}\mathcal{T}_1^T & \mathcal{T}_1\mathbf{X}\mathbf{X}^T \end{pmatrix} \Big| U = u\right],$$

where $\varrho_2(\mathbf{X}, U)$ is defined in Theorem 2.

Theorem 3. *Under the technical conditions given in the Appendix, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$\sqrt{nh}\left[\begin{pmatrix} \widehat{\mathbf{a}}_0 - \mathbf{a}_0(u) \\ \widehat{\mathbf{a}} - \mathbf{a}(u) \end{pmatrix} - \frac{\mu_2 h^2}{2}\begin{pmatrix} \mathbf{a}_0''(u) \\ \mathbf{a}''(u) \end{pmatrix} + o(h^2)\right] \xrightarrow{\mathcal{L}} N\left(0, \frac{\nu_0}{f_U(u)}\Omega^{-1}(u)\mathbf{\Sigma}(u)\Omega^{-1}(u)\right), \tag{4.4}$$

where $\mathbf{a}_0(u) = (a_0(u) + c_1^*, \ldots, a_0(u) + c_q^*)^T$, and

$$\mathbf{\Sigma}(u) = \mathbf{\Sigma}_1(u) + \mathbf{\Sigma}_2(u)$$

when censoring is independent of the covariates, or

$$\mathbf{\Sigma}(u) = \mathbf{\Sigma}_1(u) + \mathbf{\Sigma}_3(u)$$

when censoring is dependent of the covariates.

Theorem 4. *Under the technical conditions given in the Appendix, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then*

$$\sqrt{nh}\left(\widehat{a}_0(u) - a_0(u) - \frac{1}{q}\sum_{k=1}^{q} c_k - \frac{\mu_2 h^2}{2}a_0''(u)\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{\nu_0}{f_U(u)}\frac{1}{q^2}\mathbf{l}^T\left[\Omega^{-1}(u)\mathbf{\Sigma}(u)\Omega^{-1}(u)\right]_{11}\mathbf{l}\right), \tag{4.5}$$

and

$$\sqrt{nh}\left(\widehat{\mathbf{a}}(u) - \mathbf{a}(u) - \frac{\mu_2 h^2}{2}\mathbf{a}''(u)\right) \xrightarrow{\mathcal{L}} N\left(0, \frac{\nu_0}{f_U(u)}\left[\Omega^{-1}(u)\mathbf{\Sigma}(u)\Omega^{-1}(u)\right]_{22}\right), \tag{4.6}$$

where $[\cdot]_{11}$ and $[\cdot]_{22}$ represent the upper-left $q \times q$ and lower-right $p \times p$ submatrices respectively.

Theorem 4 is a direct consequence of Theorem 3.

Remark 1. CQR estimators are more efficient asymptotically and more stable and robust than LS-based estimators. This is not unexpected because unlike LS which uses only information in the mean function, CQR combines information across different quantiles.

Remark 2. The estimator $\widehat{a}_0(u)$ converges to a quantity that equals the sum of $a_0(u)$ and the mean of the uniform quantiles of the error distribution. Clearly, when the error distribution is symmetric, $\widehat{a}_0(u)$ is unbiased. When the error distribution is asymmetric, the bias of $\widehat{a}_0(u)$ converges to the mean of the errors, which actually tends to zero as q increases.

Remark 3. Within the expressions of the asymptotic variances given in Theorems 3 and 4, the term $\Omega^{-1}(u)\mathbf{\Sigma}_1(u)\Omega^{-1}(u)$ is the asymptotic variance when T_i's are fully observed, while $\Omega^{-1}(u)\mathbf{\Sigma}_2(u)\Omega^{-1}(u)$ arises due to censoring. When the data are uncensored, our results reduce to those given in Theorem 3.3 of Kai et al. (2011), who considered CQR for semiparametric varying-coefficient partially linear models.

Remark 4. Due to censoring, CQR does not possess the same asymptotic efficiency as demonstrated in Kai et al. (2011). The performance of CQR in the present context will be examined by simulations in Section 5.

5. Simulation experiments and a real data example

5.1. Simulation experiments

In this subsection, we assess the finite sample performance of the proposed methods by two simulated examples. In Example 1, the model comprises coefficient functions that vary with the quantile level τ. Within this context, we examine the cases of censoring being independent and dependent of the covariates. In Example 2, we compare the CQR estimator with the single-τ QR and LS estimators. For convenience purpose, we refer to the varying-coefficient QR method described in Section 2 simply as the QR method, and the varying-coefficient CQR method described in Section 5 as the CQR method.

Example 1. Our experiment is based on the following data generating process using a sample of 500 observations:

$$T = a_1(U)X_1 + a_2(U)X_2 + (0.5 * a_2(U)X_2)\varepsilon,$$

where $a_1(U) = (4 - 3\cos((U - 0.5)\pi/2))/4$, $a_2(U) = 0.5U(3 - U) + 1$, and U, X_1, X_2 and ε follow the $U[0, 2]$, $N(1, 1)$, $U[0.5, 1.5]$ and $N(0, 1)$ distributions respectively. We examine the following two cases of censoring mechanisms:

Case I: the covariate-independent censoring variable follows the $U[0, C]$ distribution, with the constant $C = 10$ so that about 30% of the sample data are censored.

Case II: the covariate-dependent censoring variable C_i is modeled by the Cox proportional hazard model

$$\lambda(t|X_1) = c_0 \exp(\alpha_0 X_1), \tag{5.1}$$

with $\alpha_0 = 0.25$ and $c_0 = 0.1$, such that about 30% of the data are censored. We handle the covariate dependency of censoring by two methods—Method A estimates the survival function $G(\cdot|X_1)$ based on the parametric model (5.1) with unknown parameters α_0 and c_0; Method B is based on the nonparametric Kaplan–Meier estimator (2.7) and (2.8). We use a bandwidth of 0.25 when computing the Kaplan–Meier estimator.

The quantile regression model is $Q_\tau(T|\mathbf{X}, U) = a_1(U)X_1 + (1 + 0.5\Phi^{-1}(\tau))a_2(U)X_2$, where $\Phi^{-1}(\tau)$ is the τ-quantile of the $N(0, 1)$ distribution. Specifically, $\Phi^{-1}(\tau) = 0$ when $\tau = 0.5$. Our simulations are based on 500 replications. We use the Epanechnikov kernel function $K(u) = \frac{3}{4}(1 - u^2)I(|u| \le 1)$ for local linear smoothing, and apply the aforementioned MM algorithm and resampling method to estimate the regression coefficients and compute the variances of the estimated coefficient functions. The number of bootstrap samples used for calculating the sample variance is set to 500.

The steps of the MM algorithm are described as follows:

Step 1: Set the initial values of the unknowns to $\theta^0 = (0, 0, 0, 0)^T$, and let $j = 0$.
Step 2: Let (2.12) be the surrogate function of the objective function $L(\theta)$.
Step 3: Minimize the surrogate function $Q_\varepsilon(\theta|\theta^0)$ by the Newton–Raphson method, and denote the estimator as θ^1.
Step 4: Repeat Steps 2 and 3, until $|\theta^j - \theta^{j-1}| < \Delta$, where $\Delta = 10^{-6}$. Let θ^j be the final estimator.

We choose values of ε in $Q_\varepsilon(\theta|\theta^j)$ that satisfies the condition $\varepsilon n|\ln(\varepsilon)| = \Delta$, as suggested by Hunter and Lange (2000). We set the number of replications to 500.

Figs 1(a)–(d) provide the plots of the estimated coefficient functions of $a_1(U)$ and $a_2(U)$ based on 500 replications. In each figure, the true coefficient function is shown by the blue solid curve, the estimated functions for $\tau = 0.25$, 0.5, and 0.75 are represented by the black dotted, red dashed, and green dashed–dotted curves respectively, and the 95% point-wise confidence intervals of the coefficients without bias corrections are shown for the case of $\tau = 0.50$ by the thick red dashed curves. When estimating the coefficient functions, the optimal bandwidth h_{opt} is obtained by minimizing the average median error under $\tau = 0.5$ (see Cai et al., 2000 for details). This results in the following average optimal bandwidths based on 500 replications: $\bar{h}_{opt} = 0.3210$ for Case I and $\bar{h}_{opt} = 0.3063$ for Case II under Method A. 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 under both covariate-independent and covariate-dependent censoring.

Table 1 presents the results of variance estimation based on the QR method under Case I, covariate-independent censoring, and Case II, covariate-dependent censoring, for quantiles $\tau = 0.25, 0.5$ and 0.75 at $u_0 = 0.2, 0.5, 1.0, 1.5$ and 1.8, which correspond to the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of U respectively. For convenience purpose, we set $h = 0.20$. In the table, SD is the standard deviation of $\widehat{a}_j(u_0)$ based on 500 replications, and SE is the average of standard deviations of the 500 estimated standard errors based on perturbation resampling method. Hence SD may be viewed as the true standard errors providing the basis for evaluating the accuracy of the perturbation resampling method. Table 1 reveals that in all cases, the SD and SE values are very close, suggesting that the resampling method works well. Generally speaking, the difference between SD and SE is smaller when $\tau = 0.50$ than when $\tau = 0.25$ or $\tau = 0.75$, and when $0.5 \le u_0 \le 1.5$ than when $u_0 = 0.2$ or $u_0 = 1.8$, *ceteris paribus*. It is also observed that in terms of SE, Method A is the better method of the two in the great majority of cases; in terms of SD, the two methods are close without either one being the clear favorite. Having said that, the fact that Method B yields very similar results to Method A reaffirms that the robustness of the nonparametric approach upon which Method B is based.

Table 1
Variance estimation based on QR under Example 1.

	u_0	$\tau = 0.25$				$\tau = 0.50$				$\tau = 0.75$			
		$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$	
		SE	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*
Case I	0.20	0.107	0.122	0.163	0.185	0.095	0.120	0.140	0.153	0.113	0.127	0.179	0.192
	0.50	0.138	0.158	0.203	0.240	0.128	0.157	0.157	0.181	0.147	0.170	0.223	0.259
	1.00	0.171	0.201	0.274	0.301	0.167	0.201	0.206	0.230	0.193	0.221	0.299	0.332
	1.50	0.197	0.232	0.279	0.336	0.183	0.229	0.216	0.248	0.218	0.235	0.316	0.342
	1.80	0.191	0.225	0.284	0.316	0.184	0.225	0.254	0.270	0.220	0.237	0.314	0.336
Case II	0.20	0.113	0.146	0.172	0.202	0.101	0.098	0.140	0.152	0.107	0.141	0.170	0.198
(A)	0.50	0.134	0.174	0.204	0.253	0.129	0.128	0.131	0.192	0.174	0.179	0.220	0.260
	1.00	0.182	0.235	0.272	0.317	0.164	0.167	0.231	0.211	0.181	0.242	0.270	0.331
	1.50	0.209	0.271	0.281	0.351	0.192	0.165	0.254	0.270	0.211	0.265	0.290	0.344
	1.80	0.228	0.275	0.290	0.335	0.205	0.179	0.272	0.228	0.217	0.269	0.259	0.339
Case II	0.20	0.109	0.134	0.166	0.189	0.108	0.131	0.160	0.191	0.128	0.137	0.177	0.199
(B)	0.50	0.144	0.167	0.208	0.241	0.140	0.168	0.205	0.248	0.147	0.176	0.230	0.251
	1.00	0.182	0.217	0.264	0.308	0.180	0.219	0.259	0.310	0.195	0.230	0.286	0.326
	1.50	0.208	0.250	0.306	0.335	0.219	0.254	0.283	0.342	0.246	0.261	0.334	0.356
	1.80	0.218	0.251	0.281	0.333	0.209	0.249	0.255	0.325	0.226	0.262	0.274	0.350

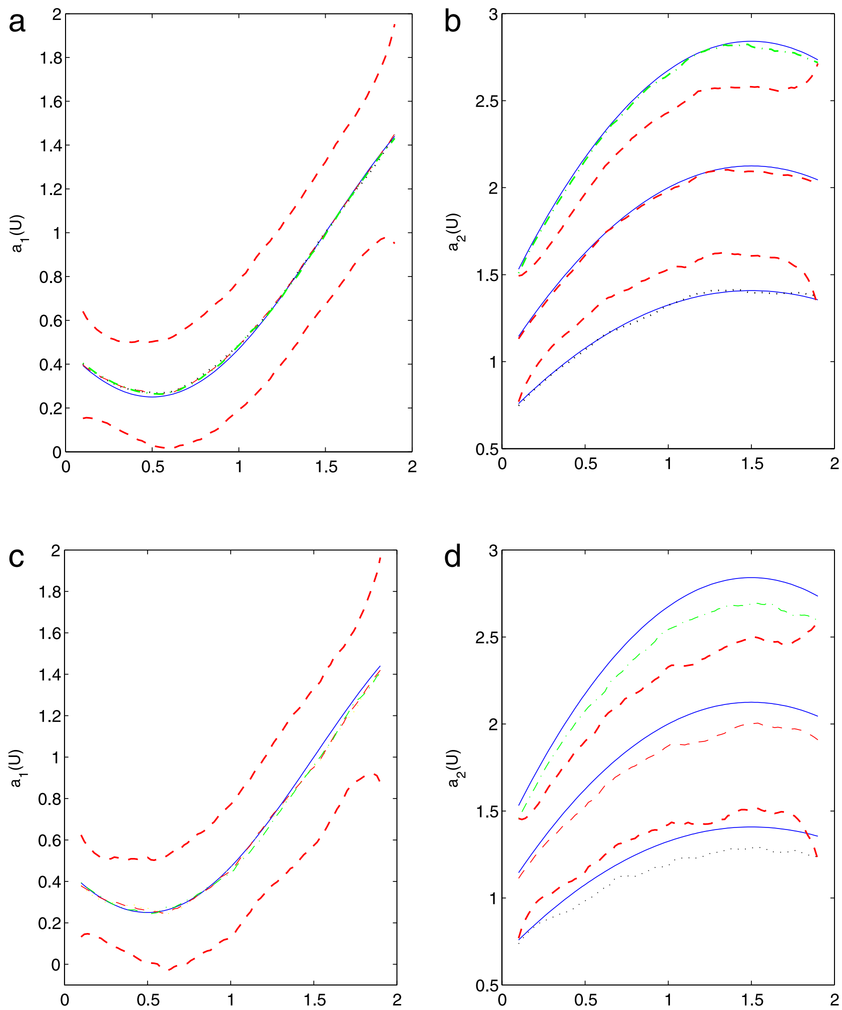


Fig. 1. The estimated coefficient functions for Example 1 under Case I (Fig. 1(a) and (b)) and Case II based on Method A (Fig. 1(c) and (d)) for three quantiles: $\tau = 0.25$ (black dotted curve), $\tau = 0.50$ (red dashed curve), and $\tau = 0.75$ (green dashed–dotted curve) based on 500 replications. The red thick dashed curves are the 95% point-wise confidence intervals based on the $\tau = 0.50$ quantile estimator. The blue solid curve represents the true coefficient function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Variance estimation under Case I: covariate-independent censoring of Example 2 based on QR (upper panel) and CQR (lower panel).

u_0	$\tau = 0.25$				$\tau = 0.50$				$\tau = 0.75$			
	$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$	
	SE	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*
0.20	0.167	0.224	0.188	0.225	0.116	0.150	0.126	0.158	0.202	0.234	0.207	0.237
0.50	0.139	0.153	0.154	0.157	0.119	0.121	0.123	0.134	0.176	0.177	0.188	0.191
1.00	0.164	0.160	0.165	0.163	0.117	0.118	0.133	0.125	0.180	0.177	0.188	0.187
1.50	0.165	0.152	0.170	0.178	0.128	0.144	0.144	0.131	0.187	0.182	0.190	0.196
1.80	0.206	0.242	0.218	0.251	0.135	0.176	0.140	0.186	0.214	0.267	0.232	0.275
u_0	$q = 5$				$q = 9$				$q = 19$			
	$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$	
	SE	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*
0.20	0.099	0.128	0.103	0.129	0.098	0.119	0.103	0.134	0.093	0.127	0.097	0.128
0.50	0.103	0.097	0.110	0.108	0.095	0.100	0.106	0.109	0.101	0.099	0.109	0.106
1.00	0.105	0.100	0.116	0.111	0.101	0.099	0.114	0.115	0.100	0.102	0.112	0.110
1.50	0.108	0.102	0.115	0.119	0.108	0.100	0.121	0.110	0.108	0.102	0.115	0.114
1.80	0.114	0.140	0.125	0.150	0.110	0.153	0.120	0.141	0.105	0.144	0.117	0.149

Table 3
Variance estimation under Case II: covariate-dependent censoring of Example 2 based on QR (upper panel) and CQR (lower panel).

	u_0	$\tau = 0.25$				$\tau = 0.50$				$\tau = 0.75$			
		$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$	
		SE	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*
A	0.20	0.179	0.242	0.196	0.243	0.118	0.159	0.124	0.169	0.195	0.222	0.204	0.236
	0.50	0.159	0.167	0.182	0.185	0.117	0.129	0.136	0.122	0.164	0.160	0.176	0.180
	1.00	0.169	0.188	0.168	0.179	0.122	0.125	0.131	0.136	0.171	0.163	0.176	0.185
	1.50	0.173	0.192	0.190	0.191	0.137	0.139	0.142	0.136	0.183	0.204	0.176	0.187
	1.80	0.219	0.261	0.268	0.228	0.137	0.183	0.135	0.180	0.214	0.279	0.218	0.279
B	0.20	0.179	0.227	0.173	0.219	0.124	0.158	0.124	0.162	0.206	0.245	0.196	0.266
	0.50	0.147	0.160	0.165	0.156	0.123	0.123	0.136	0.129	0.186	0.171	0.184	0.191
	1.00	0.164	0.169	0.172	0.172	0.132	0.131	0.146	0.142	0.191	0.193	0.189	0.188
	1.50	0.172	0.172	0.171	0.174	0.147	0.139	0.154	0.137	0.210	0.204	0.206	0.200
	1.80	0.220	0.244	0.209	0.257	0.152	0.207	0.142	0.192	0.251	0.285	0.206	0.196
	u_0	$q = 5$				$q = 9$				$q = 19$			
		$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$		$a_1(U)$		$a_2(U)$	
		SE	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*	*SE*	*SD*
A	0.20	0.101	0.119	0.109	0.135	0.098	0.128	0.101	0.130	0.095	0.133	0.098	0.132
	0.50	0.094	0.105	0.099	0.106	0.098	0.096	0.110	0.106	0.096	0.095	0.106	0.108
	1.00	0.106	0.103	0.118	0.113	0.103	0.108	0.110	0.111	0.102	0.105	0.114	0.111
	1.50	0.113	0.118	0.114	0.114	0.113	0.113	0.117	0.113	0.111	0.113	0.113	0.108
	1.80	0.120	0.150	0.117	0.143	0.122	0.154	0.117	0.153	0.119	0.154	0.114	0.143
B	0.20	0.104	0.131	0.106	0.127	0.102	0.134	0.105	0.136	0.104	0.137	0.105	0.132
	0.50	0.106	0.107	0.110	0.108	0.100	0.103	0.106	0.109	0.105	0.112	0.104	0.106
	1.00	0.116	0.113	0.116	0.111	0.113	0.103	0.111	0.115	0.113	0.107	0.117	0.105
	1.50	0.122	0.119	0.115	0.115	0.123	0.114	0.118	0.117	0.117	0.113	0.126	0.121
	1.80	0.131	0.146	0.123	0.157	0.126	0.161	0.127	0.159	0.130	0.160	0.124	0.144

Example 2. Our experiment is based on the following data generating process using a sample of 500 observations:

$$T = a_1(U)X_1 + a_2(U)X_2 + \varepsilon,$$

where $a_1(U) = (4 - 3\cos((U - 0.5)\pi/2))/4$, $a_2(U) = 0.5U(3 - U) + 1$, U and ε follow the $U[0, 2]$ and $N(0, 1)$ distributions, while observations of X_1 and X_2 are generated from the $N(1, 1)$ distribution. We consider the same censoring set-up, Cases I and II, as in Example 1.

The QR model is $Q_\tau(T|\mathbf{X}, U) = a_1(U)X_1 + a_2(U)X_2 + \Phi^{-1}(\tau)$, where $\Phi^{-1}(\tau)$ is the τ-quantile of the $N(0, 1)$ distribution. Specifically, $\Phi^{-1}(\tau) = 0$ when $\tau = 0.5$. Our simulations are based on 500 replications. We apply the aforementioned MM algorithm and resampling method for estimating the regression coefficients and computing the variances of the estimated coefficient functions. We consider $\tau = 0.25, 0.50$ and 0.75 for QR estimation. To examine the effect of varying values of q on the CQR estimator, we consider $q = 5, 9, 19$. The number of bootstrap samples used for calculating the sample variance is set to 500.

Tables 2 and 3 present the results of QR-based and CQR-based variance estimation under covariate-independent and covariate-dependent censoring with $\tau = 0.25, 0.50, 0.75$ at $u_0 = 0.2, 0.5, 1.0, 1.5$ and 1.8, which correspond to the 10th,

Table 4
RASE comparisons for Example 2 under covariate-independent censoring (s.d. in parentheses).

	Normal	Logistic	Cauchy	t_3	Log-Normal	Mixture
CQR_5	0.922(0.138)	1.089(0.267)	106,216(2,248,001)	1.641(0.745)	4.713(2.199)	5.011(2.839)
CQR_9	0.948(0.118)	1.103(0.235)	108,902(2,307,314)	1.624(0.691)	5.012(2.257)	5.010(2.790)
CQR_{19}	0.959(0.108)	1.107(0.220)	103,052(2,179,230)	1.620(0.671)	5.195(2.247)	4.924(2.713)
$QR_{0.25}$	0.776(0.320)	0.909(0.444)	62,719(1,348,593)	1.311(0.841)	11.076(8.259)	3.505(1.987)
$QR_{0.50}$	0.661(0.205)	0.887(0.376)	100,924(2,039,274)	1.463(0.832)	3.027(1.723)	4.205(2.721)
$QR_{0.75}$	0.676(0.243)	0.666(0.311)	37,603 (732,521)	0.997(0.534)	0.592(0.274)	2.649(1.731)

25th, 50th, 75th, and 90th percentiles of the distribution of U respectively. A comparison of the QR-based and CQR-based results under normal errors reveals that for a given u_0, the SD and SE of an estimated coefficient function based on CQR for any of the three choices of q are smaller than the corresponding SD and SE for any τ. This is not surprising as the CQR combines information across multiple quantiles, and thus improves the estimates of $a_j(u_0)$. Also, the effect of varying values of q on the SD and SE produced by CQR appears to be minimal. As well, Methods A and B produce very similar results for the covariate-dependent censoring case.

Fig. 2(a)–(d) provide plots of $a_1(U)$ and $a_2(U)$ based on CQR estimation with $q = 9$ and QR estimation with $\tau = 0.5$ based on 500 replications. In each figure, the true coefficient function is represented by the blue solid curve, and the estimated function based on QR (CQR) and the corresponding 95% confidence intervals are represented by the three red dashed (black dashed–dotted) curves. Under covariate-independent and covariate-dependent censoring, the average optimal bandwidths are $\bar{h}_{opt} = 0.2901$ and $\bar{h}_{opt} = 0.3173$ respectively. As Methods A and B for handling covariate-dependent censoring produce very similar results, our results reported under covariate-dependent censoring are based on Method A only. We observe from the figures that the estimated CQR and QR functions are very close to the true coefficient function. Also, at most points, the QR-based confidence interval is wider than the CQR-based confidence interval. This is expected as CQR generally results in smaller estimator variance. Generally, the results are similar whether or not censoring is dependent on the covariates.

A comparison of the efficiency of QR and CQR estimators with the LS estimator is also performed. Following Kai et al. (2011), our evaluation is based on the average squared errors (ASE), defined as

$$ASE = \left\{ \frac{1}{n_{grid}} \sum_{j=1}^{2} \sum_{k=1}^{n_{grid}} \{\widehat{a}_j(u_k) - a_j(u_k)\}^2 \right\},$$

where $\{u_k : k = 1, \ldots, n_{grid}\}$ is a set of grid points uniformly placed on [0,2] with $n_{grid} = 200$. The following ratio of average squared errors (RASE) is the ratio of the ASE of the LS estimator to that of the QR or the CQR estimator:

$$RASE(\widehat{g}) = \frac{ASE(\widehat{g}_{LS})}{ASE(\widehat{g})}$$

where $\widehat{g}$ is either the QR or the CQR estimator, and $\widehat{g}_{LS}$ is the LS estimator. The estimator $\widehat{g}$ is superior to the LS estimator if $RASE(\widehat{g})$ exceeds one, and vice versa. As in Kai et al. (2011), when comparing QR and CQR with LS, we expand the set of error distributions considered to include the $N(0, 1)$, Logistic, standard Cauchy, Log-Normal, Student's t with three degrees of freedom, and the mixture of normal $0.9N(0, 1) + 0.1N(0, 10^2)$ distributions. For the Logistic and Log-Normal distributions, we adjust the means to make them equal to zero.

Table 4 gives the means and standard deviations of RASEs of the three estimators under these six error distributions under covariate-independent censoring based on 500 replications. We observe from the table that when the errors are normally distributed, the LS estimator has an advantage over both the QR and CQR estimators, but the situation is generally reversed under other distributions with the CQR frequently results in the smallest ASE and LS the largest. While the effects of varying values of q on the CQR are minimal, the performance of the QR appears to depend strongly on τ. It is also observed that QR can sometimes have an advantage over CQR, although CQR is always the better of the two methods under Normal, Logistic, and Student's t errors, and usually outperforms QR under other error distributions. The QR estimator has superior performance to the LS estimator except under Normal and Logistic errors. Under the Cauchy distribution for which the variance is infinite, the LS estimator performs especially poorly but the CQR estimator still performs well.

Table 5 provides the corresponding results under covariate-dependent censoring based on Method A. As far as the comparisons between the QR, CQR and LS are concerned, all of the general comments above under the covariate-independent censoring also apply to covariate-dependent censoring.

5.2. A real data example

Our real data example is based on the nursing home data given in Morris et al. (1994). These data, containing $n = 1601$ observations from thirty six for-profit nursing homes in San Diego, California, were collected as part of a study by the National Center for Health Services Research to examine the extent to which financial incentives impacted patient care in nursing homes between 1980 and 1982. Of these 36 nursing homes, 18 received higher per diem payments for admitting Medicaid patients and bonuses if the patients' prognosis improved. The same data have been used by Paul (2007) and Fan et al. (2006).

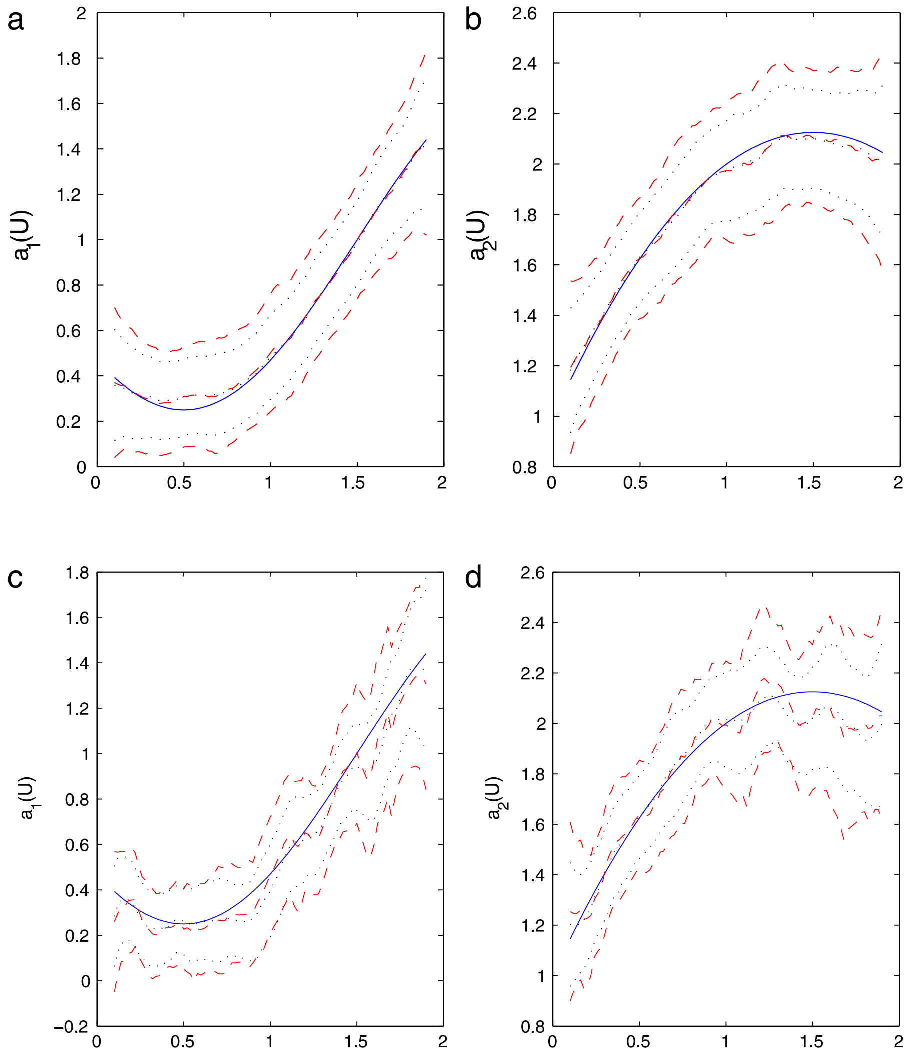


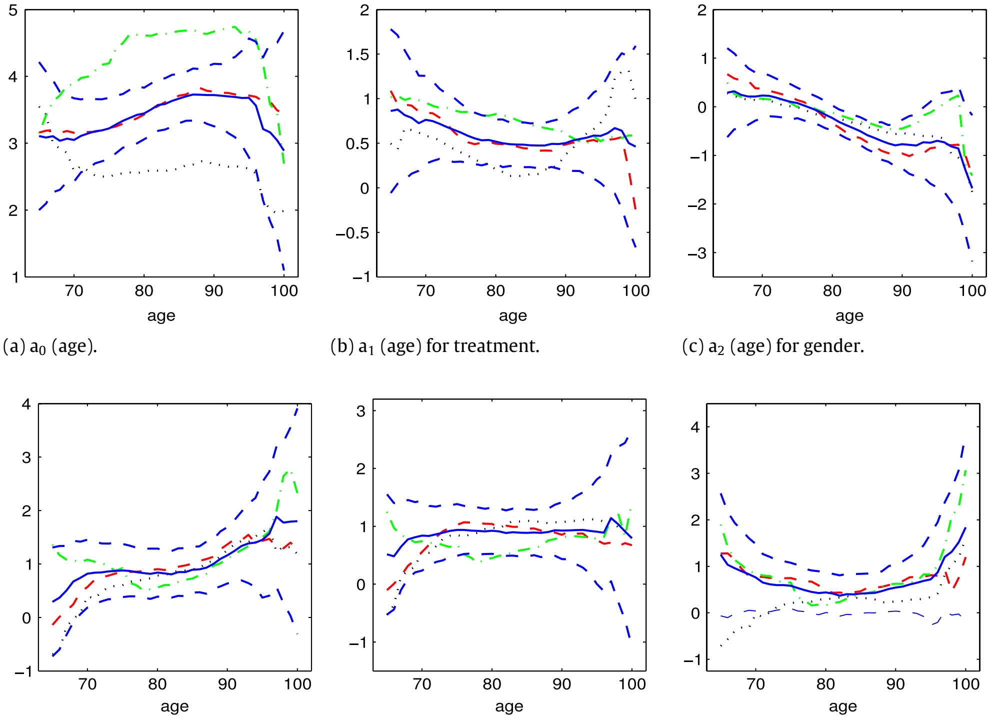
Fig. 2. The estimated coefficient functions and confidence intervals for Example 2 under Case I (Fig. 2(a) and (b)) and Case II based on Method A (Fig. 2(c) and (d)) for QR with $\tau = 0.5$ and CQR with $q = 9$ based on 500 replications. The blue solid curve represents true coefficient function, and the three red dashed (black dashed–dotted) curves represent the QR (CQR) estimated coefficient functions and the corresponding 95% confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
RASE comparisons for Example 2 under covariate-dependent censoring (s.d. in parentheses).

	Normal	Logistic	Cauchy	t_3	Log-Normal	Mixture
CQR_5	0.929(0.119)	1.075(0.168)	13,035(153,606)	1.861(2.926)	5.624(1.904)	6.975(3.455)
CQR_9	0.950(0.102)	1.082(0.149)	12,642(151,155)	1.839(2.792)	5.895(1.992)	6.791(3.218)
CQR_{19}	0.961(0.095)	1.086(0.140)	12,447(149,534)	1.827(2.735)	6.043(2.006)	6.614(3.100)
$QR_{0.25}$	0.758(0.268)	0.782(0.271)	4,809(63,254)	1.285(1.798)	11.457(5.741)	4.253(2.243)
$QR_{0.50}$	0.716(0.192)	0.967(0.289)	19,556(235,923)	1.783(2.574)	3.497(1.384)	6.412(3.456)
$QR_{0.75}$	0.723(0.230)	0.777(0.260)	15,798(234,831)	1.497(2.732)	0.779(0.227)	5.279(2.995)

We take the natural logarithm of the duration of stay in days, T, which is right-censored for 20% of patients, as the response variable and fit the data by varying-coefficient QR model

$$Q_\tau(T|\mathbf{X}, U) = a_{0,\tau}(U) + \mathbf{a}_\tau(U)\mathbf{X} = a_{0,\tau}(U) + \sum_{j=1}^{6} a_{j,\tau}(U)x_j, \quad (5.2)$$



(a) a_0 (age). (b) a_1 (age) for treatment. (c) a_2 (age) for gender.

(d) a_4 (age) for health1. (e) a_5 (age) for health2. (f) a_6 (age) for health3.

Fig. 3. The estimated coefficient functions for the real data example. The QR estimates at $\tau = 0.25$, $\tau = 0.50$ and $\tau = 0.75$ are represented by the black dotted, red dashed and green dashed–dotted curves respectively. The blue solid curve represents the CQR estimate with $q = 5$ and blue dashed curves represent the 95% confidence interval corresponding to the CQR estimate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where x_1 is a treatment indicator that takes on 1 if the patient is treated at a nursing home and 0 otherwise, $x_2 = 1$ for males and 0 for females; $x_3 = 1$ if married and 0 otherwise, x_4, x_5 and x_6 are health status variables that take on 1 for very healthy, reasonably healthy, and unhealthy respectively, and 0 otherwise, and $U = \min\{\text{age}, 100\}$ is the effect modifier, with *age* ranging from 65 to 104. This specification allows interactions between the age of the patient and the other covariates.

Our estimation is based on the Epanechnikov kernel, with a bandwidth of $h_{opt} = 13$ chosen based on $\tau = 0.50$. The estimation results are practically identical whether one treats C as dependent or independent of the covariates as the correlations between C and the covariates are all very weak. Our initial estimation suggests that all covariates except x_3 are significant; specifically, the 95% confidence bands of the coefficient functions $a_3(U)$ for all τ values considered contain zero across all values of U. We subsequently remove $a_3(U)$ from the model and the reports reported are based on the corresponding reduced model.

Figs. 3(a)–(f) present the estimated QR functions of the remaining six coefficient functions for $\tau = 0.25, 0.50, 0.75$, and the estimated CQR coefficient function for $q = 9$ and its corresponding 95% confidence bands. For ease of readability of the figures, we write U, x_1, x_2, x_4, x_5 and x_6 as *age, treatment, gender, health1, health2* and *health3* respectively in the figures. Fig. 3(a) shows that the estimated $a_0(U)$ is positive across all quantiles and it generally increases with age for age <90. This is unsurprising as the odds of a patient staying in a nursing home are expected to increase as the patient ages, but a significant portion of the very old patients (e.g., older than 90 years of age) will likely be staying in hospitals, hence the estimated $a_0(U)$'s mostly decrease as U increases beyond 90. Fig. 3(b) shows that a patient who receives medical treatments generally lives longer than a patient who does not receive treatment, but this positive impact tends to dwindle as the patient ages. Fig. 3(c) shows that after the age of 81, female patients tend to live longer than male patients. As well, Fig. 3(d)–(f) indicate that the three categories of health status generally have positive effects on T, with the effects being stronger for patients who are very healthy (Fig. 3(d)) and reasonably healthy (Fig. 3(e)), but substantially weaker for those who are unhealthy (Fig. 3(f)).

Acknowledgments

Xie's work was supported by National Natural Science Foundation of China (No. 71203025) and the special research project of University of International Business and Economics (TS3-03). Wan's work was supported by a Strategic Grant from City University of Hong Kong (No. 7008134) and a General Research Fund from the Hong Kong Research Grants Council (CityU: 9042086/11302914). Zhou's work was supported by National Natural Science Foundation of China (NSFC)

(71271128), NCMIS, Key Laboratory of RCSDS, AMSS, CAS (2008DP173182) and Shanghai First-class Discipline A and IRTSHUFE, PCSIRT (IRT13077). Xie and Zhou's work is partially supported by the State Key Program of National Natural Science Foundation of China (No. 71331006). We thank the associate editor and two referees for helpful comments on an earlier version of this paper. All remaining errors are ours.

Appendix

Our proof of results require the following technical assumptions:

(A.1) $K(.)$ is a bounded non-negative symmetric function with a bounded support $[-M, M]$.
(A.2) $a_j(u)$ is twice continuously differentiable in $u \in \mathcal{U},\ j = 0, 1, \ldots, p$.
(A.3) The distribution of U has the positive and continuous density $f_U(u)$ for all $u \in \mathcal{U}$, where $\mathcal{U}$ is the bounded support of random variable U.
(A.4) For the QR procedure:
  (i) $f_{\mathbf{x},U}(\cdot)$ is bounded away from zero and has a continuous and bounded derivative.
  (ii) $\mathbf{A}_1$, $\mathbf{A}_2$ and $\Gamma(u)$ defined in Theorem 2 are continuously differentiable at $u \in \mathcal{U}$, and $\Gamma(u)$ is nonsingular for all $u \in \mathcal{U}$.
(A.5) For the CQR procedure:
  (i) $f(\cdot)$ is bounded away from zero and has a continuous and uniformly bounded derivative.
  (ii) $\mathbf{\Sigma}(u)$ and $\Omega(u)$ defined in Theorem 3 are continuously differentiable at $u \in \mathcal{U}$, and $\Omega(u)$ is nonsingular for all $u \in \mathcal{U}$.

Remark 5. These assumptions are common for quantile regression and the varying coefficient models. Assumption (A.1) is a very mild condition on kernel functions satisfied by many kernels including the Epanechnikov kernel. Assumption (A.2) is necessary for local linear estimators as the second derivative of $a_j(u)$ impacts the bias (see Theorems 2–4). Assumption (A.3) is related to the localized behavior around $u \in \mathcal{U}$; specifically, if $f_U(u_0) = 0$ or $f_U(u_0)$ is very sparse at some $u_0 \in \mathcal{U}$, the function $a_j(u_0)$ cannot be estimated. Assumptions (A.4) and (A.5), which are similar to Assumption (C6) of Kai et al. (2011), ensure that $\mathbf{B}$ defined in Theorem 2 and $\Omega(u)$ defined in Theorem 3 are invertible for all $u \in \mathcal{U}$.

In the following, we provide a sketch of the proofs. The detailed proofs are available from the online supplementary file (see Appendix B).

Recall that $\mathbf{Z}_i = (\mathbf{X}_i^T, \mathbf{X}_i^T(U_i-u_0))^T, \mathbf{Z}_i^* = (\mathbf{X}_i^T, \mathbf{X}_i^T(U_i-u_0)/h)^T, K_i = K((U_i-u_0)/h), \phi_\tau(y) = \tau - I(y<0)$, and $\boldsymbol{\beta}_0(u_0) = (a_1(u_0), \ldots, a_p(u_0), a_1'(u_0), \ldots, a_p'(u_0))^T = (\mathbf{a}^T, \mathbf{b}^T)^T$. Write $\bar{\eta}(u_0, U_i, \mathbf{X}_i) = \sum_{j=1}^p (a_j(u_0) + a_j'(u_0)(U_i - u_0))X_{ij} = \boldsymbol{\beta}_0^T\mathbf{Z}_i$, and $\eta(u, \mathbf{X}_i) = \sum_{j=1}^p a_j(u)X_{ij}$. An application of the Taylor series expansion of $a_j(u)$ in the neighborhood of $|U_i - u_0| < Mh$ yields

$$\eta(U_i, \mathbf{X}_i) = \bar{\eta}(u_0, U_i, \mathbf{X}_i) + \frac{h^2}{2}\sum_{j=1}^p a_j''(u_0)X_{ij}\left(\frac{U_i-u_0}{h}\right)^2 + o_p(h^2). \tag{A.1}$$

For technical convenience, we reparameterize the inverse probability weighted estimator as

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{\tau k} &= \sqrt{nh}(\widehat{a}_1^k(u_0) - a_1(u_0), \ldots, \widehat{a}_p^k(u_0) - a_p(u_0), h(\widehat{b}_1^k(u_0) - b_1(u_0)), \ldots, h(\widehat{b}_p^k(u_0) - b_p(u_0)))^T \\ &= \sqrt{nh}((\widehat{\mathbf{a}}_{\tau_k}(u_0) - \mathbf{a}(u_0))^T, h(\widehat{\mathbf{b}}_{\tau_k}(u_0) - \mathbf{b}(u_0))^T)^T \\ &= \sqrt{nh}\mathbf{H}(\widehat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_0),\end{aligned}$$

where $\mathbf{H} = diag(1, h) \otimes \mathbf{I}_p$, with $\otimes$ denoting the Kronecker product. Then

$$\sum_{j=1}^p (\widehat{a}_j^k + \widehat{b}_j^k(U_i - u_0))X_{ij} = \boldsymbol{\beta}_0^T\mathbf{Z}_i + \widehat{\boldsymbol{\theta}}_{\tau k}^T\mathbf{Z}_i^*/\sqrt{nh}.$$

Now, let $\widehat{\boldsymbol{\theta}}_{\tau k}$ be the minimizer for (2.9) and (2.10). Then $\widehat{\boldsymbol{\theta}}_{\tau 1}$ and $\widehat{\boldsymbol{\theta}}_{\tau 2}$ minimize

$$Q_{n1}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\Delta_i K_i}{\widehat{G}(T_i)}\left[\rho_\tau\left(T_i^* - \frac{\boldsymbol{\theta}^T\mathbf{Z}_i^*}{\sqrt{nh}}\right) - \rho_\tau(T_i^*)\right]$$

and

$$Q_{n2}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\Delta_i K_i}{\widehat{G}(T_i|\mathbf{X}_i, U_i)}\left[\rho_\tau\left(T_i^* - \frac{\boldsymbol{\theta}^T\mathbf{Z}_i^*}{\sqrt{nh}}\right) - \rho_\tau(T_i^*)\right]$$

respectively, where $T_i^* = T_i - \boldsymbol{\beta}_0^T\mathbf{Z}_i = T_i - \bar{\eta}(u_0, U_i, \mathbf{X}_i)$.

Let $\mathcal{F}(s)$ be the set of σ-algebras defined by

$$\sigma\{I(C_i \le t), t \le s; I(T_i \le v), \mathbf{X}_i, U_i, 0 \le v < \infty, i = 1, 2, \ldots, n\}.$$

Consider the counting process of the number of individuals censored over time. The martingale of this counting process is

$$\mu_i^c(s) = N_i^c(s) - \int_0^s \lambda^c(t) Y_i(t) dt,$$

where $N_i^c(s) = I(V_i \le s, \Delta_i = 0)$, and $Y_i(s) = I(V_i \ge s)$.

Lemma A.1. *Suppose that the technical conditions* (A.1) *through* (A.4) *are satisfied, and $h = h_n \to 0$ and $nh \to \infty$ as $n \to \infty$. Then*

$$\begin{aligned} Q_{n1}(\boldsymbol{\theta}) &= -\frac{1}{2}\boldsymbol{\theta}^T \mathbf{B}\boldsymbol{\theta} - \frac{1}{\sqrt{nh}} \sum_{i=1}^n \phi_\tau(T_i^*) \mathbf{Z}_i^{*T} K_i \boldsymbol{\theta} \\ &\quad + \frac{1}{\sqrt{nh}} \sum_{i=1}^n \left\{ \int_0^L \frac{d\mu_i^c(s)}{G(s)} [\phi_\tau(T_i^*)\mathbf{Z}_i^{*T} K_i - H_\phi(0, s)] \right\} \boldsymbol{\theta} + R_n(\boldsymbol{\theta}), \end{aligned} \tag{A.2}$$

where $\sup_{\theta \in \Theta} |R_n(\boldsymbol{\theta})| = o_p(1)$, *for any compact* Θ, *and*

$$H_\phi(\boldsymbol{\theta}, s) = \frac{1}{S(s)} E\left\{ \phi_\tau(T_i^* - \boldsymbol{\theta}^T \mathbf{Z}_i^* / \sqrt{nh}) \mathbf{Z}_i^* K_i I(T_i \ge s) \right\}.$$

Proof. See the online supplementary file (see Appendix B).

Lemma A.2. *Suppose that the technical conditions* (A.1) *through* (A.4) *are satisfied, and $h = h_n \to 0$ and $nh \to \infty$ as $n \to \infty$. Then*

$$\widehat{\boldsymbol{\theta}}_{\tau 1} = -\mathbf{B}^{-1} \frac{1}{\sqrt{nh}} \sum_{i=1}^n \left[\phi_\tau(T_i^*) \mathbf{Z}_i^* K_i - \int_0^L \frac{d\mu_i^c(s)}{G(s)} \left\{ \phi_\tau(T_i^*) \mathbf{Z}_i^* K_i - H_\phi(0, s) \right\} \right] + o_p(1), \tag{A.3}$$

uniformly for any $\boldsymbol{\theta} \in \Theta$, a compact set, and

$$|\mathbf{H}(\widehat{\boldsymbol{\beta}}_{\tau 1} - \boldsymbol{\beta})| \xrightarrow{P} 0.$$

Proof of Lemma A.2. The proof of Lemma A.2 is straightforward using the convexity lemma of Pollard (1991) and Lemma A.1.

Proof of Theorem 1. Theorem 1 is an immediate consequence of Lemma A.2 and Theorem 2.

Proof of Theorem 2. We first prove the case that censoring is independent of covariates. By the asymptotic representation of (A.3), we only have to prove the asymptotic normality of $\ell_n(\boldsymbol{\beta})$, which is

$$\begin{aligned} \ell_n(\boldsymbol{\beta}) &= \frac{1}{nh} \sum_{i=1}^n \phi_\tau(T_i^*) \mathbf{Z}_i^* K_i - \frac{1}{nh} \sum_{i=1}^n \int_0^L \frac{d\mu_i^c(s)}{G(s)} \{\phi_\tau(T_i^*) \mathbf{Z}_i^* K_i - H_\phi(0, s)\} \\ &= W_n + II_2, \end{aligned}$$

where $W_n = \frac{1}{nh} \sum_{i=1}^n \phi_\tau(T_i^*) \mathbf{Z}_i^* K_i$. It can be proven that

$$\sqrt{nh} \left[W_n - \frac{h^2}{2} f_U(u_0) \begin{pmatrix} \mu_2 \mathbf{a}''(u_0) \\ \mu_3 \mathbf{a}''(u_0) \end{pmatrix} \otimes \Gamma(u_0) + o(h^2) \right] \xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)\mathbf{A}).$$

Further, the martingale representation theorem can be used to show that II_2 is asymptotically normal with $EII_2 = 0$, and

$$Var(II_2) = \frac{\tau(1-\tau)}{nh} f_U(u_0) E\{\varrho_1(U, \mathbf{X}) \mathbf{X}\mathbf{X}^T | U = u_0\} \otimes \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \{1 + o(1)\}. \tag{A.4}$$

Obviously, W_n and II are uncorrelated. This completes the proof of Theorem 2 when censoring is independent of the covariates.

To prove Theorem 2 when censoring depends on covariates, recall that

$$Q_{n2}(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\Delta_i K_i}{\widehat{G}(T_i | \mathbf{X}_i, U_i)} \left[\rho_\tau\left(T_i^* - \frac{\boldsymbol{\theta}^T \mathbf{Z}_i^*}{\sqrt{nh}} \right) - \rho_\tau(T_i^*) \right].$$

Similar to previous arguments, it can be shown that the following asymptotic representation holds:

$$Q_{n2}(\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{\theta}^T \mathbf{B}\boldsymbol{\theta} - \frac{1}{\sqrt{nh}} \sum_{i=1}^n \left\{ \frac{\Delta_i K_i}{G(T_i | \mathbf{X}_i, U_i)} \phi_\tau(T_i^*) \mathbf{Z}_i^{*T} K_i \right\} \boldsymbol{\theta} + o_p(1). \tag{A.5}$$

From the convexity lemma of Pollard (1991), the minimizer $\widehat{\boldsymbol{\theta}}_{\tau 2}$ of the convex function $Q_{n2}(\boldsymbol{\theta})$ can be expressed as

$$\widehat{\boldsymbol{\theta}}_{\tau 2} = -\mathbf{B}^{-1}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{\Delta_i}{G(T_i|\mathbf{X}_i, U_i)}\phi_\tau(T_i^*)K_i\mathbf{Z}_i^*. \tag{A.6}$$

Similar to the proof of the asymptotic normality of W_n, and by some tedious calculations, we can prove that $\widehat{\boldsymbol{\theta}}_{\tau 2} = \sqrt{nh}\mathbf{H}(\widehat{\boldsymbol{\beta}}_{\tau 2}(u_0) - \boldsymbol{\beta}(u_0))$ has the following asymptotic property:

$$\sqrt{nh}\left[\mathbf{H}\{\widehat{\boldsymbol{\beta}}_{\tau 2}(u_0) - \boldsymbol{\beta}_0(u_0)\} - \frac{h^2}{2(\mu_0\mu_2 - \mu_1^2)}\begin{pmatrix}(\mu_2^2 - \mu_1\mu_3)\mathbf{a}''(u_0)\\(\mu_0\mu_3 - \mu_1\mu_2)\mathbf{a}''(u_0)\end{pmatrix} + o(h^2)\right]$$
$$\xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)\mathbf{B}^{-1}\mathbf{A}_2\mathbf{B}^{-1}). \tag{A.7}$$

In the following, we derive the asymptotic distribution when $\widehat{G}(t|x, u)$ is estimated by (2.7) and (2.8). We assume that the true survival distribution G of the censoring variable is G_0, and true value of $\boldsymbol{\theta}$ is 0. Denote

$$m_i(\boldsymbol{\theta}, G) = \frac{\Delta_i\mathbf{Z}_i^*}{G(T_i|\mathbf{X}_i, U_i)}\phi_\tau\left(T_i^* - \frac{\boldsymbol{\theta}^T\mathbf{Z}_i^*}{\sqrt{nh}}\right)K_i/h,$$
$$M_n(\boldsymbol{\theta}, G) = \frac{1}{n}\sum_{i=1}^{n}m_i(\boldsymbol{\theta}, G),$$
$$M(\boldsymbol{\theta}, G) = E(M_n(\boldsymbol{\theta}, G)) = Em_i(\boldsymbol{\theta}, G),$$

and

$$\begin{aligned}M(\boldsymbol{\theta}, G) &= E(M_n(\boldsymbol{\theta}, G)) = Em_i(\boldsymbol{\theta}, G)\\ &= E\{E(m_i(\boldsymbol{\theta}, G)|\mathbf{X}_i, U_i)\}\\ &= E\left\{\frac{G_0(T_i|\mathbf{X}_i, U_i)}{G(T_i|\mathbf{X}_i, U_i)}\phi\left(T_i^* - \frac{\boldsymbol{\theta}^T\mathbf{Z}_i^*}{\sqrt{nh}}\right)\mathbf{Z}_i^*K_i/hv\right\}.\end{aligned}$$

It is easy to show that

$$\Gamma_1(0, G_0) = \left.\frac{\partial M(\boldsymbol{\theta}, G)}{\partial\boldsymbol{\theta}}\right|_{\theta=0, G=G_0} = -\frac{1}{\sqrt{nh}}f_U(u_0)\begin{pmatrix}\mu_0 & \mu_1\\ \mu_1 & \mu_2\end{pmatrix}\otimes\boldsymbol{\Gamma}(u_0)\{1 + o(h)\}.$$

By Theorem 2 of Chen et al. (2003) and the results of Gonzalez-Manteiga and Cadarso-Suarez (1994), we obtain

$$\begin{aligned}\Gamma_2(\boldsymbol{\theta}_0, G_0)[\widehat{G} - G_0] &= \lim_{\tau\to 0}\frac{M(\boldsymbol{\theta}_0, G_0 + \tau(\widehat{G} - G_0)) - M(\boldsymbol{\theta}_0, G_0)}{\tau}\\ &= \frac{1}{n}\sum_{i=1}^{n}\varphi(\boldsymbol{\theta}_0, G_0, W_i) + o_p(1),\end{aligned}$$

where

$$\begin{aligned}\varphi(\boldsymbol{\theta}_0, G_0, W_i) &= \frac{1}{h}K_ih(\mathbf{X}_i, U_i)\mathbf{Z}_i^*\int_0^\infty\frac{\varsigma(V_i, \Delta_i, t, \mathbf{X}_i, U_i)}{G_0(t|\mathbf{X}_i, U_i)}\phi_\tau\left(t^* - \frac{\boldsymbol{\theta}_0^Tz^*}{\sqrt{nh}}\right)f(t|\mathbf{X}_i, U_i)dt\\ &= \frac{1}{h}K_ih(\mathbf{X}_i, U_i)\mathbf{Z}_i^*E\left\{\left.\frac{\varsigma(V_i, \Delta_i, T_i, \mathbf{X}_i, U_i)}{G_0(T_i|\mathbf{X}_i, U_i)}\phi_\tau(T_i^*)\right|\mathbf{X}_i, U_i\right\},\end{aligned}$$
$$\varsigma(V_i, \Delta_i, t, x, u) = G_0(t|x, u)\left[\int_0^{\min(V_i,t)}\frac{g_0(s|x, u)ds}{G_0^2(s|x, u)(1 - F_{\mathbf{x},u}(s|x, u))} + \frac{I(V_i\le t, \Delta_i = 0)}{G_0(V_i|x, u)\{1 - F_{\mathbf{x},u}(V_i|x, u)\}}\right] \tag{A.8}$$

and $g_0(s|x, u)$ is the function of the first derivative of $G(s|x, u)$ with respect to s.

Note that $m_i(\boldsymbol{\theta}_0, G_0)$ and $\phi(\boldsymbol{\theta}_0, G_0, W_i)$ are two sequences of i.i.d. variables with mean zero and finite variance, and they are uncorrelated. Hence, it follows from the Central Limit Theorem that

$$\begin{aligned}\sqrt{nh}\left\{M_n(\boldsymbol{\theta}_0, G_0) + \Gamma_2(\boldsymbol{\theta}_0, G_0)[\widehat{G} - G_0]\right\} &= \sqrt{nh}\cdot\left\{\frac{1}{n}\sum_{i=1}^{n}m_i(\boldsymbol{\theta}_0, G_0) + \frac{1}{n}\varphi(\boldsymbol{\theta}_0, G_0, W_i)\right\} + o_p(1)\\ &\xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)\mathbf{A}_2),\end{aligned}$$

where

$$\mathbf{A}_2 = f_U(u_0)\begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix} \otimes \left\{E\Big[\frac{\mathbf{X}\mathbf{X}^T}{G(T|\mathbf{X},U)}\Big|U=u_0\Big] + E\{\varrho_3(X,U)^{\otimes 2}\mathbf{X}\mathbf{X}^T|U=u_0\}\right\}.$$

Now, by Theorem 2 of Chen et al. (2003), $\widehat{\boldsymbol{\theta}}_{\tau 2}$ has an asymptotically normal distribution with mean 0 and variance–covariance $\tau(1-\tau)\mathbf{B}^{-1}\mathbf{A}_2\mathbf{B}^{-1}$. This completes the proof of Theorem 2. A detailed proof is given in the online supplementary file (see Appendix B). □

Proof of Theorem 3. We only provide the proof for the case where censoring is independent of the covariates. Similar methods can be used to prove the theorem when censoring depends on the covariates. Let $\eta_{i,k} = I(\varepsilon_i \le c_k^*) - \tau_k$ and $\eta_{i,k}^* = I(\varepsilon_i \le c_k^* - r_i(u)) - \tau_k$, where $r_i(u) = a_0(U_i) - a_0(u) - a_0'(u)(U_i - u) + \mathbf{X}_i^T\{\mathbf{a}(U_i) - \mathbf{a}(u) - \mathbf{a}'(u)(U_i - u)\}$. Also, let $\widehat{\boldsymbol{\theta}}^* = \sqrt{nh}\{\widehat{a}_{0,1} - a_0(u) - c_1^*, \ldots, \widehat{a}_{0,q} - a_0(u) - c_q^*, \{\widehat{\mathbf{a}} - \mathbf{a}(u)\}^T, h\{\widehat{b}_0 - a_0'(u)\}, h\{\widehat{\mathbf{b}} - \mathbf{a}'(u)\}^T\}^T$ and $\mathbf{X}_{i,k}^*(u) = (\mathbf{e}_k^T, \mathbf{X}_i^T, (U_i - u)/h, \mathbf{X}_i^T(U_i - u)/h)^T$, where $\mathbf{e}_k$ is a q-vector with an element of 1 at the kth position and 0 elsewhere.

We will first show that $\widehat{\boldsymbol{\theta}}^*$ can be expressed as

$$\widehat{\boldsymbol{\theta}}^* = -f_U^{-1}(u)\mathcal{D}(u)^{-1}\mathbf{W}^*(u) + o_p(1), \tag{A.9}$$

where $\mathcal{D}(u) = diag(\Omega(u), \mu_2\mathcal{C}_0\mathbf{B}_2(u))$, $\mathbf{B}_2(u) = E\big[(1,\mathbf{X}^T)^T(1,\mathbf{X}^T)|U=u\big]$, and

$$\mathbf{W}^*(u) = \frac{1}{\sqrt{nh}}\sum_{k=1}^{q}\sum_{i=1}^{n}\left[K_i(u)\eta_{i,k}^*\mathbf{X}_{i,k}^* - \int_0^L \frac{d\mu_i^c(s)}{G(s)}\left\{K_i(u)\eta_{i,k}^*\mathbf{X}_{i,k}^* - \frac{1}{S(s)}E[K_i(u)\eta_{i,k}^*\mathbf{X}_{i,k}^*I(T_i\ge s)]\right\}\right].$$

This will be followed by a demonstration of the asymptotic normality of $(\widehat{\mathbf{a}}_0, \widehat{\mathbf{a}})$ by showing the asymptotic normality of $\mathbf{W}^*(u)$.

Let

$$W_{n,1}^*(u) = \frac{1}{\sqrt{nh}}\sum_{k=1}^{q}\sum_{i=1}^{n}K_i(u)\eta_{i,k}^*(\mathbf{e}_k^T, \mathbf{X}_i^T)^T.$$

Using Theorem 3.1 of Kai et al. (2011), we have

$$W_{n,1}^* - E(W_{n,1}^*) \xrightarrow{\mathcal{L}} N(0, f_U(u)\nu_0\boldsymbol{\Sigma}_1(u)), \tag{A.10}$$

where

$$\frac{1}{\sqrt{nh}}E(W_{n,1}^*) = -\frac{\mu_2h^2}{2}f_U(u)\Omega(u)\begin{pmatrix}\mathbf{a}_0''(u)\\ \mathbf{a}''(u)\end{pmatrix} + o_p(h^2).$$

Let

$$W_{n,2}^* = \frac{1}{\sqrt{nh}}\sum_{k=1}^{q}\sum_{i=1}^{n}\int_0^L\frac{d\mu_i^c(s)}{G(s)}\left\{K_i(u)\eta_{i,k}^*(\mathbf{e}_k^T, \mathbf{X}_i^T)^T - \frac{1}{S(s)}E[K_i(u)\eta_{i,k}^*(\mathbf{e}_k^T, \mathbf{X}_i^T)^TI(T_i\ge s)]\right\}.$$

Applying Theorem 2, and arguments similar to those applied to $W_{n,1}^*$, we have

$$W_{n,2}^* \xrightarrow{\mathcal{L}} N(0, f_U(u)\nu_0\boldsymbol{\Sigma}_2(u)). \tag{A.11}$$

Combining (A.10) and (A.11), we obtain

$$\sqrt{nh}\left[\begin{pmatrix}\widehat{\mathbf{a}}_0 - \mathbf{a}_0(u)\\ \widehat{\mathbf{a}} - \mathbf{a}(u)\end{pmatrix} - \frac{\mu_2h^2}{2}\begin{pmatrix}\mathbf{a}_0''(u)\\ \mathbf{a}''(u)\end{pmatrix} + o(h^2)\right] \xrightarrow{\mathcal{L}} N\left(0, \frac{\nu_0}{f_U(u)}\Omega^{-1}(u)(\boldsymbol{\Sigma}_1(u) + \boldsymbol{\Sigma}_2(u))\Omega^{-1}(u)\right).$$

That implies that the first part of Theorem 3 holds. By the same argument as that applied to proving the second part of Theorem 2, we can show that the second part of Theorem 3 also holds. This completes the proof of Theorem 3. A detailed proof is given in the online supplementary file (see Appendix B). □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2015.02.011.

References

Bang, H., Tsiatis, A.A., 2000. Estimating medical costs with censored data. Biometrika 87, 329–343.
Bang, H., Tsiatis, A.A., 2002. Median regression with censored cost data. Biometrics 58, 643–649.
Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A., 1993. Efficient and Adaptive Inference in Semiparametric models. Johns Hopkins University Press, Baltimore.
Cai, Z., 2007. Trending time varying coefficient time series models with serially correlated errors. J. Econometrics 136, 163–188.
Cai, J., Fan, J., Jiang, J., Zhou, H., 2008. Partially linear hazard regression with varying-coefficients for multivariate survival data. J. R. Stat. Soc. Ser. B 70, 141–158.
Cai, J., Fan, J., Yao, Q., 2000. Functional coefficient regression models for nonlinear time series. J. Amer. Statist. Assoc. 95, 941–956.
Cai, J., Fan, J., Zhou, H., Zhou, Y., 2007. Marginal hazard models with varying-coefficients for multivariate failure time data. Ann. Statist. 35, 324–354.
Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. J. Amer. Statist. Assoc. 103, 1595–1608.
Chen, S., 2010. An integrated maximum score estimator for a generalized censored quantile regression model. J. Econometrics 155, 90–98.
Chen, X., Linton, O., Van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71, 1591–1608.
Cleveland, W.S., Gross, E., Shhyu, W.M., 1991. Local regression models. In: Chambers, J.M., Hastie, T.J. (Eds.), Statistical Models in S. Wadsworth and Brooks, Pacific Grove, pp. 309–376.
De Gooijer, J.G., Zerom, D., 2003. On additive conditional quantiles with high-dimensional covariates. J. Amer. Statist. Assoc. 98, 135–146.
Fan, J., Li, R., 2004. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. J. Amer. Statist. Assoc. 99, 710–723.
Fan, J., Lin, H., Zhou, Y., 2006. Local partial-likelihood estimation for lifetime data. Ann. Statist. 34, 290–325.
Fan, J., Zhang, W., 2008. Statistical methods with varying coefficient models. Stat. Interface 1, 179–195.
Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York.
Gannoun, A., Saracco, J., Yuan, A., Bonney, G.E., 2005. Non-parametric quantile regression with censored data. Scand. J. Statist. 32, 527–550.
Gonzalez-Manteiga, W., Cadarso-Suarez, C., 1994. Asymptotic properties of a generalized Kaplan–Meier estimator with some applications. J. Nonparametr. Stat. 4, 65–78.
Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. J. R. Stat. Soc. Ser. B 55, 757–796.
Honda, T., 2004. Quantile regression in varying coefficient models. J. Statist. Plann. Inference 121, 113–125.
Honoré, B., Khan, S., Powell, J.L., 2002. Quantile regression under random censoring. J. Econometrics 109, 67–105.
Huang, Y., 2010. Quantile calculus and censored regression. Ann. Statist. 38, 1670–1637.
Hunter, D.R., Lange, K., 2000. Quantile regression via an MM algorithm. J. Comput. Graph. Statist. 9, 60–77.
Jin, Z., Lin, D., Ying, Z., 2006. On least-square regression with censored data. Biometrika 93, 147–161.
Jin, Z., Ying, Z., Wei, L.J., 2001. A simple resampling method by perturbing the minimand. Biometrika 88, 381–390.
Kai, B., Li, R., Zou, H., 2010. Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regresion. J. R. Stat. Soc. Ser. B 72, 49–69.
Kai, B., Li, R., Zou, H., 2011. New efficient estimation and variable selection methods for semiparametric varying-coefficinet partially linear models. Ann. Statist. 39, 305–332.
Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observation. J. Amer. Statist. Assoc. 53, 457–481.
Kim, M.O., 2007. Quantile regression with varying-coefficients. Ann. Statist. 35, 92–108.
Koenker, R., Bassett, G.W., 1978. Regression quantiles. Econometrica 46, 33–50.
Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. Biometrika 81, 673–680.
Li, Y., Graubard, B.I., Korn, E.L., 2010. Application of nonparametric quantile regression to body mass index percentile curves from survey data. Stat. Med. 29, 558–572.
Lindgren, A., 1997. Quantile regression with censored data using generalized L_1 minimization. Comput. Statist. Data Anal. 23, 509–524.
Martinussen, T., Scheike, T.H., 2006. Dynamic Regression Models for Survival Data. Springer, New York.
Morris, C.N., Norton, E.C., Zhou, X.H., 1994. Parametric duration analysis of nursing home usage. In: Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L., Greenhouse, J. (Eds.), Case Studies in Biometry. Wiley, New York, pp. 231–248.
Paul, G., 2007. On robustness and model flexibility in survival analysis: transformed hazard models an average effects. Biometrics 63, 69–77.
Peng, L., Huang, Y., 2008. Survival analysis with quantile regression models. J. Amer. Statist. Assoc. 103, 637–649.
Pollard, D., 1991. Asymptotics for least absolute deviation regression estimator. Econometric Theory 7, 186–199.
Portnoy, S., 2003. Censored regression quantiles. J. Amer. Statist. Assoc. 98, 1001–1012.
Powell, J.L., 1986. Censored regression quantiles. J. Econometrics 32, 143–155.
Qian, J., Peng, L., 2010. Censored quantile regression with partially functional effects. Biometrika 97, 839–850.
Robins, J.M., Rotnitzky, A., Zhao, P., Claeskens, G., 1994. Estimation of regression coefficients when some observations are not always observed. J. Amer. Statist. Assoc. 89, 846–866.
Wang, Y., Garcia, T.P., Ma, Y., 2012. Nonparametric estimation for censored mixture data with application to the cooperative Huntington's observation research trial. J. Amer. Statist. Assoc. 107, 1324–1338.
Wang, H.J., Wang, L., 2009. Locally weighted censored quantile regression. J. Amer. Statist. Assoc. 104, 1117–1128.
Wei, Y., He, X., 2006. Conditional growth charts. Ann. Statist. 34, 2069–2097.
Wheelock, D.C., Wilson, P.W., 2008. Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations. J. Econometrics 145, 209–225.
Wu, Y., Fan, J., Müller, H.G., 2010. Varying-coefficient functional linear regression. Bernoulli 16, 730–758.
Yang, S., 1999. Censored median regression using weighted empirical survival and hazard functions. J. Amer. Statist. Assoc. 94, 137–146.
Yin, G., Cai, J., 2005. Quantile regression models with multivariate failure time data. Biometrics 61, 151–161.
Ying, Z., Jung, S., Wei, L., 1995. Survival analysis with median regression models. J. Amer. Statist. Assoc. 90, 178–184.
Yu, K., Jones, M.C., 1998. Local linear quantile regression. J. Amer. Statist. Assoc. 8, 228–237.
Zhao, H.W., Tian, L.L., 2001. On estimating medical cost and incremental cost-effectiveness ratios with censored data. Biometrics 2001, 1002–1008.
Zou, H., Yuan, M., 2008. Composite quantile regression and the oracle model selection theory. Ann. Statist. 36, 1108–1126.