Annals of the Institute of Statistical Mathematics (2020) 72:265–295
https://doi.org/10.1007/s10463-018-0688-3

Semiparametric quantile regression with random censoring

Francesco Bravo[1]

Received: 16 May 2017 / Revised: 7 August 2018 / Published online: 10 September 2018

Abstract
This paper considers estimation and inference in semiparametric quantile regression models when the response variable is subject to random censoring. The paper considers both the cases of independent and dependent censoring and proposes three iterative estimators based on inverse probability weighting, where the weights are estimated from the censoring distribution using the Kaplan–Meier, a fully parametric and the conditional Kaplan–Meier estimators. The paper proposes a computationally simple resampling technique that can be used to approximate the finite sample distribution of the parametric estimator. The paper also considers inference for both the parametric and nonparametric components of the quantile regression model. Monte Carlo simulations show that the proposed estimators and test statistics have good finite sample properties. Finally, the paper contains a real data application, which illustrates the usefulness of the proposed methods.

Keywords Inverse probability of censoring · Local linear estimation · M-M algorithm

1 Introduction

Since its introduction as a generalization of the linear regression model, quantile regression (Bassett and Koenker 1978; Koenker and Bassett 1978) has been widely used in economics, finance, biostatistics and medical statistics—see Koenker (2005) for a review of applications. Compared to standard linear regression models, quantile

I am grateful to the Associate Editor and two Referees for useful comments and suggestions that improved considerably the paper. The usual disclaimer applies.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-018-0688-3) contains supplementary material, which is available to authorized users.

✉ Francesco Bravo
francesco.bravo@york.ac.uk
https://sites.google.com/a/york.ac.uk/francescobravo/

1 Department of Economics, University of York, York YO10 5DD, UK

regression models provide a more complete characterization of the conditional distribution of the responses given a set of covariates, being at the same time more robust to the presence of possible outliers. Nonparametric and semiparametric extensions to quantile regression have been considered by Chauduri (1991), Fan et al. (1994), He and Shi (1996), Chauduri et al. (1997), Yu and Jones (1998), He and Liang (2000), Lee (2003), Horowitz and Lee (2005), Cai and Xu (2008) and Cai and Xiao (2012), among many others.

All of the above results assume that the data are always observable. However, in many situations of empirical relevance, some of the responses are subject to censoring and ignoring this fact may give highly biased estimates (see, e.g., Koenker 2005). One important type of censoring is random censoring, which naturally arises in duration and survival analysis. Ying et al. (1995), Bang and Tsiatis (2002) and Zhou (2006) have considered censored median regression. Ying et al. (1995) proposed a simple estimation method, which, however, involves a complicated set of discontinuous estimating equations that can be difficult to solve. Bang and Tsiatis (2002) proposed a modified version of the least absolute deviation estimator (Bassett and Koenker 1978) that is similar to the one used by Koul et al. (1981) and is computationally easy, but potentially suffers from the well-known instability in the right tail of the Kaplan–Meier estimator. Zhou (2006) provided a simple modification of Bang and Tsiatis (2002) estimator that involves a convex function and a simple modification to the data that avoids the potential instability problem of the Kaplan–Meier estimator. All of these procedures are based on the assumption of unconditional independence between the censored response and the censoring variable itself, which is often restrictive. Indeed, as noted, for example, by Kalbfleisch and Prentice (2002), conditional independence (given the covariates) is often a more natural and appropriate assumption. Conditional independence was assumed by Peng and Huang (2008), Leng and Tong (2013) and Wang and Wang (2009) for quantile regressions. El Ghouch and van Keilegom (2009) and Xie et al. (2015) considered, respectively, nonparametric and varying coefficients quantile regressions. Peng and Huang (2008) used martingales techniques under an assumption of global linearity (that can be restrictive in practice) and suggested an L_1-type convex objective function to compute their estimator; Leng and Tong (2013) computed their estimator using linear programming based on a modification of Ying et al.'s (1995) estimating equations, while Wang and Wang (2009) proposed an estimator based on locally reweighting. El Ghouch and van Keilegom (2009) and Wang and Wang (2009) both used the M-M algorithm (Hunter and Lange 2000), which replaces the nonsmooth objective function used in the quantile estimation with an approximating function that can be majorized by a smooth (quadratic) function that can be easily minimized by standard iterative methods.

In this paper, we consider estimation of a semiparametric quantile regression model, where the response is subject to random censoring. We consider both unconditional and conditional independent censoring, but it is important to note that the proposed estimation method could be used also for certain type of informative ("induced dependent") censoring situations, such as the analysis of medical cost data and health outcome data (see, e.g., Bang and Tsiatis 2000; Lin 2000), and more generally in any situation where the process (data) of interest is increasing over time and its observations are stopped because of the occurrence of a terminal underlying event.

The estimation procedure that we propose is a weighted two (or three) step one, where the weights are given by the inverse probability of censoring. The inverse probability of censoring weighting approach has been used in survival analysis by Koul et al. (1981), Robins and Rotnitzky (1992), Bang and Tsiatis (2000) and Satten and Datta (2001), among many others. The first step is used to estimate locally all the unknown parameters of the model, whereas the second step is used to estimate the parametric component. As the second step requires undersmoothing, an additional third step can be used to re-estimate the nonparametric part of the model, should it be of interest.

In this paper, we make the following contributions: First, we consider three different estimators for the censoring distribution: the Kaplan–Meier estimator for independent censoring, a parametric estimator (e.g., Cox's (1972) maximum partial likelihood estimator or Breslow's (1972) probability of censoring estimator) and a nonparametric estimator—the conditional or local Kaplan–Meier estimator (Beran 1981) for the dependent censoring case. We derive the asymptotic distributions of the three resulting estimators of both the nonparametric and parametric components. Second, we propose a computationally simple resampling method that can be used to estimate the asymptotic variances of the estimators of the parametric component. Third, we consider inference both for the parametric and the nonparametric components of the model and propose test statistics that can be used to test both global and local hypotheses about the unknown parameters. Fourth, we use a Monte Carlo study to illustrate the finite sample properties of the proposed estimators and test statistics. Finally, we show the usefulness of the proposed method with a real data application.

The rest of the paper is structured as follows: Next section introduces the model and the estimators. Section 3 contains the main results, Sect. 4 introduces the resampling method and shows its consistency, whereas Sect. 5 first describes some details on the computational aspects of the proposed estimators and then reports the results of the Monte Carlo study. Section 6 contains an empirical application. All proofs and some additional results on a two-step version of the proposed estimators can be found in the online supplemental Appendix.

The following notation is used throughout the paper: "T" indicates transpose, a prime "′" and double prime "″" denote first and second derivative and for any vector v, $v^{\otimes 2} = vv^T$.

2 The model and the estimators

Consider a semiparametric quantile regression model

$$Q_{Y|X}(\tau|X) = \inf(t\colon \Pr(Y \leq t|X) \geq \tau) = X_1^T \beta_{0\tau} + \theta_{0\tau}(X_2), \tag{2.1}$$

where $\beta_{0\tau}$ is a k dimensional vector of unknown parameters, $X = \left[X_1^T, X_2\right]^{\mathrm{T}}$ and $\theta_{0\tau}(\cdot)$ is an unknown real-valued function, assumed to be twice continuously differentiable with derivatives $\theta'_{0\tau}(\cdot)$ and $\theta''_{0\tau}(\cdot)$. We note here that bearing in mind the curse of dimensionality, the results reported below can be readily modified to allow for X_2 to be vector valued. In this case, the convergence rate of the estimators of the nonparametric component would be slower as it depends on the dimension of X_2, whereas the convergence rate for the parametric estimators would not be affected under an appropriately strengthened version of the (standard) undersmoothing condition given in Theorems 3–5 (see Sect. 3.2).

We assume that the sample values of the response variable $(Y_i)_{i=1}^n$ are subject to random censoring; hence, the random sample we observe is $\left(Z_i, X_i', \delta_i\right)_{i=1}^n$ where $Z_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \le C_i)$ denotes the censoring indicator. Let $G_0(\cdot)$ denote the unknown survival distribution for both the independent and dependent cases of the censoring random variable C. We follow the same approach as that originally suggested for parametric median regression models by Bang and Tsiatis (2000) (see also Zhou 2006) and use inverse probability of censoring weighting based on the survival function of the censoring variable.

Let

$$Q_n(\beta, \theta, G) = \sum_{i=1}^n \frac{\delta_i}{G_0(\cdot)} \rho_\tau\left(Z_i - X_{1i}^T \beta_\tau - \theta_\tau(X_{2i})\right) \tag{2.2}$$

be the objective function, where $\rho_\tau(\cdot) = \cdot\,(\tau - I(\cdot < 0))$ denotes the check function.

Let $\widehat{G}(\cdot)$ denote a consistent estimator for $G_0(\cdot)$, which depends on the type of censoring and will be discussed in some detail at the end of this section, and let

$$\theta_{0\tau}(X_2) = \theta_{0\tau}(x_2) + \theta_{0\tau}'(x_2)(X_2 - x_2) := a_\tau + b_\tau(X_2 - x_2) \tag{2.3}$$

denote the local linear approximation to $\theta_{0\tau}(X_2)$.

The estimation procedure to estimate the unknown parameters $\beta_{0\tau}$ and $\theta_{0\tau}(\cdot)$ is the following:

Step 1 Estimate $\beta_{0\tau}$ and $\theta_{0\tau}(\cdot)$ locally using (2.3), that is

$$\widehat{\beta}_\tau^l, \widehat{a}_\tau^l, \widehat{b}_\tau^l = \arg\min_{a_\tau, b_\tau, \beta_\tau} \sum_{i=1}^n \frac{\delta_i}{\widehat{G}(\cdot)} \rho_\tau\left(Z_i - X_{1i}^T \beta_\tau - a_\tau - b_\tau(X_{2i} - x_2)\right) K_h(X_{2i} - x_2), \tag{2.4}$$

where $K_h(\cdot) = K(\cdot/h)$ is a kernel function and h is a bandwidth.

Step 2 Estimate $\beta_{0\tau}$ globally using

$$\widehat{\beta}_\tau = \arg\min_{\beta_\tau} \sum_{i=1}^n \frac{\delta_i}{\widehat{G}(\cdot)} \rho_\tau\left(Z_i - X_{1i}^T \beta_\tau - \widehat{\theta}_\tau^l(X_{2i})\right) \tag{2.5}$$

where $\widehat{\theta}_\tau^l(X_{2i}) = \widehat{a}_\tau^l$.

Step 3 Estimate $\theta_{0\tau}(\cdot)$ locally using

$$\widehat{a}_\tau, \widehat{b}_\tau = \arg\min_{a_\tau, b_\tau} \sum_{i=1}^n \frac{\delta_i}{\widehat{G}(\cdot)} \rho_\tau\left(Z_i - X_{1i}^T \widehat{\beta}_\tau - a_\tau - b_\tau(X_{2i} - x_2)\right) K_h(X_{2i} - x_2),$$

where $\widehat{\beta}_\tau$ is the estimate of Step 2.

The form of the estimator $\widehat{G}(\cdot)$ depends on the type of censoring. In the case of independent censoring, $\widehat{G}(\cdot) = \widehat{G}(Z_i)$ is the Kaplan–Meier estimator, which, as mentioned in Introduction, is well known to be unstable on the right tail of the survival distribution. To avoid this problem, we follow Zhou's (2006) suggestion and use a modification of the response. To be specific, since for any constant $L > \left(X_1^T \beta_{0\tau} + \theta_{0\tau}(X_2)\right)$, the τ-quantile of Y equals that of $\min(Y, L)$, we can replace the observations $(Z_i, Y_i, \delta_i)_{i=1}^n$ with $\left(Z_i^L, Y_i^L, \delta_i^L\right)_{i=1}^n$ where $Z_i^L = \min(Z_i, L)$, $Y_i^L = \min(Z_i, L)$ and $\delta_i^L = 1 - (1 - \delta_i) I(L > Z_i)$, and define both (2.4) and (2.5) in terms of $\left(Z_i^L, Y_i^L, \delta_i^L\right)_{i=1}^n$. The resulting estimators are more stable since $\widehat{G}\left(Z_i^L\right)$ is bounded from below by $\widehat{G}(L)$.

In the case of dependent censoring, $\widehat{G}(\cdot)$ can be a parametric or a nonparametric estimator. In the former case, $\widehat{G}(\cdot) = G_{\widehat{\gamma}}(Z_i|X_i)$, where $G_{\gamma_0}(Z_i|X_i) =: G_0(Z_i|X_i)$ is a parametric specification indexed by the unknown finite-dimensional parameter γ_0 and $\widehat{\gamma}$ is the maximum likelihood estimator of γ_0. In the latter case, $\widehat{G}(\cdot) = \widehat{G}(Z_i|X_i)$ is the local Kaplan–Meier estimator

$$\widehat{G}(z|x) = \prod_{i=1}^{n} \left(1 - \frac{\omega_i(x)}{\sum_{j=1}^{n} I\left(Z_j \geq Z_i\right) \omega_j(x)}\right)^{I(Z_i \leq z, \delta_i = 0)}, \tag{2.6}$$

where $\omega_i(x)$ is a sequence of nonnegative weights such that $\sum_{i=1}^{n} \omega_i(x) = 1$. In the remaining part of the paper, we use the Nadaraya–Watson weights

$$\omega_i(x) = L\left(\frac{X_i - x}{b}\right) \Big/ \sum_{j=1}^{n} L\left(\frac{X_j - x}{b}\right),$$

where $L(\cdot)$ is a kernel function and b is another bandwidth. Note that to avoid the curse of dimensionality, we consider only $\widehat{G}(Z_i|X_{2i})$; that is, we assume that the response and the censoring variable are conditionally independent given only the covariate X_2. To relax this assumption, one could use the same dimension reduction approach as that suggested by Li and Patilea (2017). They assume that the response and censoring variable are conditionally independent of all of the covariates given the index $X^T\alpha_0$, where α_0 is an unknown parameter that can be estimated at the parametric rate. Then, at least in principle, their asymptotic representation of the conditional Kaplan–Meier estimator $\widehat{G}(z|\widehat{\alpha}, v)$, where

$$\widehat{G}(z|\alpha, v) = \prod_{i=1}^{n} \left(1 - \frac{\omega_i(\alpha, v)}{\sum_{j=1}^{n} I\left(Z_j \geq Z_i\right) \omega_j(\alpha, v)}\right)^{I(Z_i \leq z, \delta_i = 0)}$$

and

$$\omega_i(\alpha, v) = L\left(\frac{X_i^T\alpha - v}{b}\right) \Big/ \sum_{j=1}^{n} L\left(\frac{X_j^T\alpha - v}{b}\right),$$

could be used to obtain an extension of Theorem 5 given below that would not rely on the conditional independence of the response and censoring variable given only the covariate X_2. We leave this possibility for future communications.

3 Asymptotic results

3.1 Nonparametric component

In this section, we obtain the asymptotic distribution of the local estimator (2.4) defined in Step 1. Let $\kappa_j = \int t^j K(t)\,dt$ and $\upsilon_j = \int t^j K^2(t)\,dt$ and assume that:

A1 (i) $\{Y_i, X_i\}_{i=1}^n$ is an i.i.d. sample from the joint distribution $F_{Y,X}(\cdot)$ of Y and X, $\{C_i\}_{i=1}^n$ is an i.i.d. sample from a distribution with survival function $G_0(\cdot)$, (ii) *either* Y is independent of C *or* Y and C are conditionally independent given X or given X_2,

A2 (i) the conditional distribution of $\varepsilon = Y - X_1^T \beta_{0\tau} - \theta_{0\tau}(X_2)$ given X, $F_{\varepsilon|X}(\cdot)$ is such that $F_{\varepsilon|X}(0|x) = \tau$ for all $x \in \mathcal{X}_1 \times \mathcal{X}_2$, (ii) the conditional density $f_{\varepsilon|X}(\cdot|x)$ is uniformly bounded and positive in a neighborhood of 0 for all $x \in \mathcal{X}_1 \times \mathcal{X}_2$, (iii) the marginal density of X_2 $f_{X_2}(x)$ is continuous and positive at $x = x_2$, (iv) X_1 and X_2 have bounded support $\mathcal{X}_1 \times \mathcal{X}_2$,

A3 The kernel functions $K(\cdot)$ and $L(\cdot)$ are symmetric with bounded support and bounded first derivatives,

A4 (i) $\theta''_\tau(x)$ is continuous at $x = x_2$, (ii) the matrix $\Sigma(x_2)$ defined in (3.1) is nonsingular for all $x_2 \in \mathcal{X}_2$,

A5 (i) $G_0(\cdot)$ has a uniformly bounded density $g_0(\cdot)$ and there exists a constant C such that $G_0(Y \geq C) > 0$, *or* (ii) the conditional distribution $G_\gamma(\cdot|X)$ has conditional density $g_\gamma(\cdot|X)$ uniformly bounded in a neighborhood of γ_0, and the maximum likelihood estimator $\widehat{\gamma}$ satisfies $n^{1/2}(\widehat{\gamma} - \gamma_0) = O_p(1)$, *or* (iii) there exists a constant C such that $\sup_{x_2 \in \mathcal{X}_2} G_0(Y \geq C | X_2 = x_2) > 0$ and the conditional distribution $G(\cdot|X_2)$ has conditional density $g(\cdot|x)$ uniformly bounded in a neighborhood of $x = x_2 \in \mathcal{X}_2$.

The above regularity conditions are fairly standard: A1(ii) and A5(i)–(iii) are commonly used in survival analysis; in particular, A5(i) ensures the uniform consistency of the Kaplan–Meier estimator for all $c \leq C$ and similarly A5(iii) ensures the uniform consistency of the local Kaplan–Meier estimator. A2(iii)–A2(iv), A3 and A4 are commonly used in semiparametric estimation and finally A2(i), A2(ii) are standard assumption in quantile regression; see, for example, Koenker (2005).

Theorem 1 *Under assumptions* A1–A5 *and for* $nh \to \infty$, $nhb^4 \to 0$, $h \log n/b \to 0$

$$(nh)^{1/2} \begin{bmatrix} \widehat{\beta}^l_\tau - \beta_{0\tau} \\ \widehat{\theta}^l_\tau(x_2) - \theta_{0\tau}(x_2) \end{bmatrix} - B(x_2) \Big] \xrightarrow{d} N\left(0, \Sigma_1(x_2)^{-1} \Sigma_{1G}(x_2) \Sigma_1(x_2)^{-1}\right),$$

where

$$B(x_2) = \frac{h^2}{2} f_{X_2}(x_2)\,\theta''_{0\tau}(x_2)\,\Sigma_1(x_2)^{-1} E\left\{\kappa_2 f_{\varepsilon|X}(0|X)\begin{bmatrix} X_1^T & 1\end{bmatrix}^{\mathrm{T}} | X_2 = x_2\right\},$$
$$\Sigma_1(x_2) = f_{X_2}(x_2)\,E\left\{f_{\varepsilon|X}(0|X)\begin{bmatrix} X_1^T & 1\end{bmatrix}^{\mathrm{T}\otimes 2} | X_2 = x_2\right\},$$
$$\Sigma_{1G}(x_2) = f_{X_2}(x_2)\,E\left\{\frac{\tau(1-\tau)\,v_0}{G_0(\cdot)}\begin{bmatrix} X_1^{\otimes 2} & X_1 \\ X_1^T & 1\end{bmatrix} | X_2 = x_2\right\},$$

where $G_0(\cdot)$ *is* $G_0(Z)$ *or* $G_{\gamma_0}(Z|X)$ *or* $G_0(Z|X_2)$.

Theorem 1 shows that the asymptotic variance of the inverse probability of censoring weighted local estimator depends on the unknown distribution of censoring but not on the type of censoring. The asymptotic variance is larger than the corresponding one with uncensored responses, but is typical for nonparametric estimators with inverse probability of censoring weighting and more generally with synthetic type of responses (see, e.g., Fan and Gijbels 1994). Note also that in case of dependent censoring estimated nonparametrically without the bandwidth assumption $nhb^4 \to 0$, the bias term $B(x_2)$ would feature an extra term of order $O_p\left(b^2\right)$, which might dominate the mean squared error.

For the local estimator of Step 3, we have the following result:

Theorem 2 *Under the same assumptions of Theorem 1 and for any* $n^{1/2}\left(\widehat{\beta}_\tau - \beta_{0\tau}\right) = O_p(1)$

$$(nh)^{1/2}\left[\left(\widehat{\theta}_\tau(x_2) - \theta_{0\tau}(x_2)\right) - e_{K+1}^T B(x_2)\right]$$
$$\xrightarrow{d} N\left(0, e_{K+1}^T \Sigma_1(x_2)^{-1}\,\Sigma_{1G}(x_2)\,\Sigma_1(x_2)^{-1} e_{K+1}\right),$$

where $e_{K+1} = \left[0_K^T, 1\right]^{\mathrm{T}}$.

3.2 Parametric component

In this section, we obtain the asymptotic distribution of the global estimator (2.5) defined in Step 2. We first consider the case of independent censoring so that the global estimator for $\beta_{0\tau}$ is defined as

$$\widehat{\beta}_\tau = \arg\min_{\beta_\tau} \sum_{i=1}^{n} \frac{\delta_i^L}{\widehat{G}\left(Z_i^L\right)} \rho_\tau\left(Z_i^L - X_{1i}^T\beta_\tau - \widehat{\theta}_\tau^l(X_{2i})\right).$$

Let

$$\varphi(X_i) = E\left[f_{\varepsilon|X}(0|X)\,X_1\left[0^T, 1, 0\right] | X_2 = X_{2i}\right]\Sigma(X_{2i})^{-1}\left[X_{1i}^T, 1, 0\right]^{\mathrm{T}},$$
$$\rho'_\tau(\cdot) = (\tau - I(\cdot < 0)),$$

where $\Sigma(\cdot)$ is defined as

$$\Sigma(\cdot) = E\left\{ f_{\varepsilon|X}(0|X) \begin{bmatrix} X_1^{\otimes 2} & X_1 & 0 \\ X_1^T & 1 & 0 \\ 0 & 0 & \kappa_2 \end{bmatrix} |X_2 = \cdot \right\}, \tag{3.1}$$

and assume that

A6 $E\left(f_{\varepsilon|X}(0|X)\, X_1^{\otimes 2}\right) := \Sigma_2$ is nonsingular.

Theorem 3 *Under assumptions A1–A5(i) and A6 for* $nh \to \infty$ *and* $nh^4 \to 0$

$$n^{1/2}\left(\widehat{\beta}_\tau - \beta_{0\tau}\right) \xrightarrow{d} N\left(0, \Sigma_2^{-1}\Sigma_{2km}\Sigma_2^{-1}\right),$$

where

$$\Sigma_{2km} = E\left[\left((X_1 - \varphi(X))\,\rho'_\tau(\varepsilon)\right)^{\otimes 2}\right] + E\left[\int_0^L \left((X_1 - \varphi(X))\,\rho'_\tau(\varepsilon)\right.\right.$$
$$\left.\left.\frac{E\left[(X_1 - \varphi(X))\,\rho'_\tau(\varepsilon)\, I(Z \geq u)\right]}{S(u)} I(Z > u)\right]^{\otimes 2} \frac{\lambda_0(u)}{G_0(u)} du\right],$$

$\lambda_0(u)$ *is the hazard function for the censoring distribution and* $S(u) = \Pr(Y \geq u)$.

In the case of dependent censoring with the censoring distribution estimated parametrically, the global estimator for $\beta_{0\tau}$ is defined as

$$\widehat{\beta}_\tau = \arg\min_{\beta_\tau} \sum_{i=1}^{n} \frac{\delta_i}{G_{\widehat{\gamma}}(Z_i|X_i)} \rho_\tau\left(Z_i - X_{1i}^T\beta_\tau - \widehat{\theta}_\tau^l(X_{2i})\right).$$

We assume that

A6′ (i) $E\left(f_{\varepsilon|X}(0|X)\, X_1^{\otimes 2}\right) := \Sigma_2$ is nonsingular, (ii) the parametric estimator $G_{\widehat{\gamma}}(\cdot|\cdot)$ admits the following linear representation

$$G_{\widehat{\gamma}}(Z_i|x) - G_0(Z|x) = \frac{1}{n}\sum_{i=1}^{n} \psi_{\gamma_0}(W_i, x) + o_p\left(n^{-1/2}\right), \tag{3.2}$$

where $W_i = \left[Z_i, \delta_i, X_i^T\right]$.

Note that (3.2) is satisfied by both Cox's (1975) maximum partial likelihood and Breslow's (1972) estimators of the probability of censoring.

Theorem 4 *Under assumptions A1–A4, A5(ii) and A6′ for* $nh \to \infty$ *and* $nh^4 \to 0$

$$n^{1/2}\left(\widehat{\beta}_\tau - \beta_{0\tau}\right) \xrightarrow{d} N\left(0, \Sigma_2^{-1}\Sigma_{2p}\Sigma_2^{-1}\right),$$

where

$$\Sigma_{2p} = E\left[\frac{(X_1 - \varphi(X))\,\rho'_\tau(\varepsilon)}{G_0(Z|X)} - E\left((X_{12} - \varphi(X_{12}, X_{22}))\,\rho'_\tau(\varepsilon_2)\,\frac{\psi_{\gamma_0}(W_1, X_{12}, X_{22})}{G_0(Z_1|X_{12}, X_{22})}|W_1\right)\right]^{\otimes 2}.$$

In the case of dependent censoring with the censoring distribution estimated non-parametrically, the global estimator for $\beta_{0\tau}$ is defined as

$$\widehat{\beta}_\tau = \arg\min_{\beta_\tau} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(Z_i|X_{2i})} \rho_\tau\left(Z_i - X_{1i}^T\beta_\tau - \widehat{\theta}_\tau^l(X_{2i})\right),$$

where $\widehat{G}(\cdot|\cdot)$ is the local Kaplan–Meier defined in (2.6). Let

$$\psi(Z, \delta, t, u) = \int_0^{\min(Z,t)} -\frac{g_0(s|u)\,\mathrm{d}s}{G_0(s|u)^2(1 - F(s|u))} + \frac{(1-\delta)\,I(Z \le t)}{G_0(Z|u)(1 - F(Z|u))};$$

Theorem 5 *Under assumptions A1–A4, A5(iii) and A6 for* $nh \to \infty$, $nh^4 \to 0$, $nb^3 \to \infty$ *and* $nb^4 \to 0$

$$n^{1/2}\left(\widehat{\beta}_\tau - \beta_{0\tau}\right) \xrightarrow{d} N\left(0, \Sigma_2^{-1}\Sigma_{2np}\Sigma_2^{-1}\right),$$

where

$$\Sigma_{2np} = E\left[\frac{(X_1 - \varphi(X))\,\rho'_\tau(\varepsilon)}{G_0(Z|X_2)} - E\left[f_{X_2}(X_2)\,\frac{\psi(Z, \delta, Y, X_2)(X_1 - \varphi(X))\,\rho'_\tau(\varepsilon)}{G_0(Z|X_2)}|X_2\right]\right]^{\otimes 2}.$$

3.3 Resampling

The asymptotic variances of the estimators of Theorems 3–5 are rather complicated to estimate, so in this section we suggest a resampling technique that has been previously used by Su and Wei (1991), Jin et al. (2001), Zhou (2006) and Xie et al. (2015), among others. We generate B random samples $\{\xi_i\}_{i=1}^n$ from the random variable ξ with $E(\xi) = 1$ and $\mathrm{Var}(\xi) = 1$ and compute

$$\widehat{\beta}_\tau^* = \arg\min_{\beta_\tau} \sum_{i=1}^{n} \frac{\delta_i\xi_i}{\widehat{G}_\xi(\cdot)} \rho_\tau\left(Z_i - X_{1i}^T\beta_\tau - \widehat{\theta}_\tau^l(X_{2i})\right),$$

where in the case of independent censoring the Z_i's and δ_i's are replaced by the Z_i^L's and δ_i^L's and $\widehat{G}_\xi(\cdot)$ corresponds to the perturbed version of the three different estimators of $G_0(\cdot)$. To be specific in the case of independent censoring, $\widehat{G}_\xi(\cdot)$ corresponds to the perturbed Kaplan–Meier estimator $\widehat{G}_\xi(Z_i)$, where

$$\widehat{G}_\xi(z) = \prod_{i=1}^{n}\left(1 - \frac{\mathrm{d}N_\xi(z)}{Y_\xi(z)}\right)$$

and $N_\xi(z) = \xi_i I\left(Z_i^L \leq z, \delta_i = 0\right)$, $Y_\xi(u) = \sum_{i=1}^{n} \xi_i I\left(Z_i^L \geq u\right)$. In the case of dependent censoring, $\widehat{G}_\xi(Z_i|X_i)$ is either $G_{\widehat{\gamma}\xi}(Z_i|X_i)$ with

$$G_{\widehat{\gamma}\xi}(Z_i|x) - G_0(Z|x) = \frac{1}{n}\sum_{i=1}^{n}\xi_i\psi_{\gamma_0}(W_i, x) + o_p\left(n^{-1/2}\right),$$

or

$$\widehat{G}_\xi(Z_i|x) = \prod_{i=1}^{n}\left(1 - \frac{\omega_i(x)}{\sum_{j=1}^{n}\xi_i I\left(Z_j \geq Z_i\right)\omega_j(x)}\right)^{I\xi_i(Z_i \leq z, \delta_i = 0)}.$$

Theorem 6 *Under the same assumptions of Theorems 3–5, conditionally on* $\left(Z_i, \delta_i, X_i^T\right)_{i=1}^{n}$

$$n^{1/2}\left(\widehat{\beta}_\tau^* - \widehat{\beta}_\tau\right) \xrightarrow{d} N\left(0, \Sigma_2^{-1}\Sigma_{2*}\Sigma_2^{-1}\right),$$

where Σ_{2*} is either Σ_{2km} or Σ_{2p} or Σ_{2np}, defined, respectively, in Theorems 3–5.

Theorem 6 shows that the proposed resampling technique consistently estimates the distributions of the various estimators proposed in Sect. 3.2. In particular, we can use the asymptotic variance–covariance matrices of $\widehat{\beta}_\tau^*$ to obtain confidence intervals for $\beta_{0\tau}$ using the normal approximation and test statistical hypotheses on β_τ using the χ^2 approximation and the delta method—see Sect. 4 for further details.

3.4 Extension: partially linear varying coefficients models

The results of the previous sections can be readily extended to semiparametric models containing varying coefficients (see, e.g., Fan and Huang 2005). To be specific, let

$$Q_{Y|X}(\tau|X) = \inf(t\colon \Pr(Y \leq t|X) \geq \tau) = X_1^T\beta_{0\tau} + X_3^T\theta_{0\tau}(X_2), \tag{3.3}$$

where X_3 is a p-dimensional vector of additional covariates and, as in the previous sections, X_2 is assumed to be univariate. Then, the same iterative estimation of Sect. 2 based on the inverse probability of censoring weighting and the local approximation

$$\theta_{0\tau}(X_2) = \theta_{0\tau}(x_2) + \theta_{0\tau}'(x_2)(X_2 - x_2) := a_\tau + b_\tau(X_2 - x_2),$$

where now both a_τ and b_τ are p-dimensional vectors, can be used to estimate $\beta_{0\tau}$ and $\theta_{0\tau}(\cdot)$.

Theorem 7 *Under assumptions A1–A5 (with* $X = \left[X_1^T, X_2, X_3^T\right]^{\mathrm{T}}$ *and* X_3 *with bounded support* $\mathcal{X}_3$*) and for* $nh \to \infty$, $nhb_1^4 \to 0$, $h \log n / b_1 \to 0$

$$(nh)^{1/2}\left[\begin{matrix}\widehat{\beta}_\tau^l - \beta_{0\tau}\\ \widehat{\theta}_\tau^l(x_2) - \theta_{0\tau}(x_2)\end{matrix} - B(x_2)\right] \xrightarrow{d} N\left(0, \Omega_1(x_2)^{-1}\Omega_{1G}(x_2)\Omega_1(x_2)^{-1}\right),$$

where

$$B(x_2) = \frac{h^2}{2} f_{X_2}(x_2)\Omega_1(x_2)^{-1} E\left\{\kappa_2 f_{\varepsilon|X}(0|X)\begin{bmatrix}X_1X_3^T\\ X_3^{\otimes 2}\end{bmatrix}|X_2 = x_2\right\}\theta_{0\tau}''(x_2),$$

$$\Omega_1(x_2) = f_{X_2}(x_2) E\left\{f_{\varepsilon|X}(0|X)\begin{bmatrix}X_1\\ X_3\end{bmatrix}^{\otimes 2}|X_2 = x_2\right\},$$

$$\Omega_{1G}(x_2) = f_{X_2}(x_2) E\left\{\frac{\tau(1-\tau)v_0}{G_0(\cdot)}\begin{bmatrix}X_1\\ X_3\end{bmatrix}^{\otimes 2}|X_2 = x_2\right\}$$

and $G_0(\cdot)$ *is* $G_0(Z)$ *or* $G_{\gamma_0}(Z|X)$ *or* $G_0(Z|X_2)$.

Let $S = \left[O_{pk}, I_p, O_p\right]$ denote a selection matrix, where O_{pk} is a $p \times k$ matrix of zeroes, I_p is the identity matrix of order p and O_p is a $p \times p$ matrix of zeroes and let

$$\xi(X_i) = E\left[f_{\varepsilon|X}(0|X)X_1X_3^T|X_2 = X_{2i}\right]S\Omega(X_{2i})^{-1}\left[X_{1i}^T, X_{3i}^T, 0^T\right]^{\mathrm{T}},$$

where $\Omega(\cdot)$ is defined as

$$\Omega(\cdot) = E\left\{f_{\varepsilon|X}(0|X)\begin{bmatrix}X_1^{\otimes 2} & X_1X_3^T & 0\\ X_1X_3^T & X_3^{\otimes 2} & 0\\ 0 & 0 & \kappa_2 X_3^{\otimes 2}\end{bmatrix}|X_2 = \cdot\right\}$$

The following two theorems are direct generalizations of Theorems 3–6 to the partially linear varying coefficient model (3.3).

Theorem 8 *Under assumptions A1–A5 (with* $X = \left[X_1^T, X_2, X_3^T\right]^{\mathrm{T}}$ *and* X_3 *with bounded support* $\mathcal{X}_3$*) and A6 for* $nh \to \infty$, $nh^4 \to 0$, $nb_1^3 \to \infty$ *and* $nb_1^4 \to 0$

$$n^{1/2}\left(\widehat{\beta}_\tau - \beta_{0\tau}\right) \xrightarrow{d} N\left(0, \Sigma_2^{-1}\Omega_{2*}\Sigma_2^{-1}\right)$$

where Ω_{2*} is either Ω_{2km} or Ω_{2p} or Ω_{2np}, and

$$\begin{aligned}\Omega_{2km} =& E\left[\left(\left(X_1-\xi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)\right)^{\otimes 2}\right]+E\left[\int_0^L\left(X_1-\xi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)\right.\\&\left.-\frac{E\left[\left(X_1-\xi\left(X\right)\right)I\left(Z\geq u\right)\right]}{S\left(u\right)}I\left(Z>u\right)\right]^{\otimes 2}\frac{\lambda_0\left(u\right)}{G_0\left(u\right)}du\Bigg],\\ \Omega_{2p} =& E\left\{\frac{\left(X_{11}-\xi\left(X\right)\right)\rho_\tau'\left(\varepsilon_1\right)}{G_0\left(Z_1|X_1\right)}\right.\\&\left.-E\left[\left(X_{12}-\xi\left(X_{12},X_{22}\right)\right)\rho_\tau'\left(\varepsilon_2\right)\frac{\psi_{\gamma_0}\left(W_1,X_{12},X_{22}\right)}{G_0\left(Z_1|X_{12},X_{22}\right)}|W_1\right]\right\}^{\otimes 2}.\\ \Omega_{2np} =& E\left[\frac{\left(X_1-\xi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)}{G_0\left(Z|X_2\right)}\right.\\&\left.-E\left[f_{X_2}\left(X_2\right)\frac{\psi\left(Z,\delta,Y,X_2\right)\left(X_1-\xi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)}{G_0\left(Z|X_2\right)}|X_2\right]\right]^{\otimes 2}.\end{aligned}$$

Theorem 9 *Under the same assumptions of Theorem 8, conditionally on $\left(Z_i,\delta_i,X_i^T\right)_{i=1}^n$*

$$n^{1/2}\left(\widehat{\beta}_\tau^*-\widehat{\beta}_\tau\right)\xrightarrow{d}N\left(0,\Sigma_2^{-1}\Omega_{2*}\Sigma_2^{-1}\right),$$

where Ω_{2*} is either Ω_{2km} or Ω_{2p} or Ω_{2np}, given in Theorem 8.

4 Inference

The results of the previous section can be used to test statistical hypotheses about both the parametric and nonparametric components β_τ and $\theta_\tau(\cdot)$. First, Theorem 7 can be used to construct Wald statistics to test local hypotheses about $\theta_\tau(\cdot)$. To investigate the asymptotic properties of such statistics, we consider the following local hypothesis with a Pitman drift

$$H_n\colon R\theta_{0\tau}\left(x_2^*\right)=r_\tau\left(x_2^*\right)+\gamma_{\tau n}\left(x_2^*\right) \tag{4.1}$$

for some fixed $x_2^*\in\mathcal{X}_2$, where R is an $l\times p$ matrix of constants and $\gamma_{\tau n}(\cdot)$ is a bounded continuous function that may depend on n. Let

$$W_l\left(x_2^*\right)=(nh)\left(R\left(\widehat{\theta}_\tau\left(x_2^*\right)-r_\tau\left(x_2^*\right)\right)\right)^T\left(R\widehat{\Omega}_{1G\theta_\tau}\left(x_2^*\right)R^T\right)^{-1}R\left(\widehat{\theta}_\tau\left(x_2^*\right)-r_\tau\left(x_2^*\right)\right)$$

denote the local Wald statistic, where

$$\widehat{\Omega}_{1G\theta_\tau}\left(x_2^*\right)=\left[O_{pk},I_p\right]\widehat{\Omega}_1\left(x_2^*\right)^{-1}\widehat{\Omega}_{1\widehat{G}}\left(x_2^*\right)\widehat{\Omega}_1\left(x_2^*\right)^{-1}\left[O_{pk}^T,I_p\right]^{\mathrm{T}},$$

$$\widehat{\Omega}_1\left(x_2\right)=\widehat{f}_{X_2}\left(x_2\right)\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{G}\left(\cdot\right)}\widehat{f}_{\widehat{\varepsilon}|X}\left(0|X_i\right)\begin{bmatrix}X_{1i}\\X_{3i}\end{bmatrix}^{\otimes 2}K_h\left(X_{2i}-x_2\right),$$

$$\widehat{\Omega}_{1G}\left(x_2\right)=\widehat{f}_{X_2}\left(x_2\right)\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i\tau\left(1-\tau\right)v_0}{\widehat{G}\left(\cdot\right)^2}\begin{bmatrix}X_{1i}\\X_{3i}\end{bmatrix}^{\otimes 2}K_h\left(X_{2i}-x_2\right),$$

$\widehat{f}_{X_2}\left(\cdot\right)$, $\widehat{f}_{\widehat{\varepsilon}|X}\left(\cdot\right)$ are kernel estimates of $f_{X_2}\left(\cdot\right)$, $f_{\varepsilon|X}\left(\cdot\right)$ and $\widehat{G}\left(\cdot\right)$ is any of the three estimators described in Sect. 2 for $G_0\left(\cdot\right)$.

Proposition 10 *Under the assumptions of Theorem 7, if* $\operatorname{rank}\left(R\right)=l$ $(l\leq p)$ *and* $nh^5\rightarrow 0$*, then under* (4.1) *(i) for* $(nh)^{1/2}\gamma_{\tau n}\left(x_2^*\right)\rightarrow\gamma_\tau\left(x_2^*\right)>0$ *(for some* $\left\|\gamma_\tau\left(x_2^*\right)\right\|<\infty$*)*

$$W_l\left(x_2^*\right)\xrightarrow{d}\chi^2\left(\kappa,l\right),$$

where $\chi^2\left(\kappa,l\right)$ *is a noncentral Chi-squared distribution with* l *degrees of freedom and noncentrality parameter*

$$\kappa=f_{X_2}\left(x_2^*\right)\gamma_\tau\left(x_2^*\right)^T\left(R\Omega_{1G\theta_\tau}\left(x_2^*\right)R^T\right)^{-1}\gamma_\tau\left(x_2^*\right);$$

(ii) for $(nh)^{1/2}\gamma_{\tau n}\left(x_2^*\right)\rightarrow\infty$*,*

$$W_l\left(x_2^*\right)\xrightarrow{p}\infty.$$

Proposition 10 shows that the proposed test has power against local Pitman-type alternatives and is consistent against any fixed alternatives of the form $\gamma_{\tau n}\left(\cdot\right)=\gamma_\tau\left(\cdot\right)$. Under the null hypothesis H_0: $R\theta_{0\tau}\left(x_2^*\right)=r_\tau\left(x_2^*\right)$, the proposition can be used to construct confidence regions for $R\theta\left(x_2^*\right)$ with nominal coverage $1-\alpha$ that is for $\Pr\left(\chi^2\left(l\right)\leq c_\alpha\right)=1-\alpha$ and $C_\alpha\left(x_2^*\right)=\Pr\left(r\left(x_2^*\right)|W_l\left(x_2^*\right)\leq c_\alpha\right)$,

$$\Pr\left(r\left(x_2^*\right)\in C_\alpha\left(x_2^*\right)\right)=1-\alpha+o\left(1\right).$$

Proposition 10 can also be used to test the important hypothesis of constancy of the varying coefficients $\theta_\tau\left(\cdot\right)$, corresponding to

$$H_0\colon\theta_{0\tau}\left(x_2^*\right)=\theta_{0\tau}. \tag{4.2}$$

The test can be easily implemented by considering the restricted quantile regression model

$$Q_{Y|X}\left(\tau|X\right)=\inf\left(t\colon\Pr\left(Y\leq t|X\right)\geq\tau\right)=X_1^T\beta_{0\tau}+X_3^T\theta_{0\tau}. \tag{4.3}$$

Let $\breve{\theta}_\tau$ denote the quantile estimator of $\theta_{0\tau}$ in (4.3), and note that under the null hypothesis (4.2) and assumptions A1–A3 (only for the kernel $L(\cdot)$), A5, A6 for $E\left(f_{\varepsilon|X}(0|X)\left[X_1^T, X_3^T\right]^{T\otimes 2}\right)$ and $nb^4 \to 0$, it is possible to show that $n^{1/2}\left(\breve{\theta}_\tau - \theta_{0\tau}\right) = O_p(1)$. Hence,

$$(nh)^{1/2}\left(\widehat{\theta}_\tau\left(x_2^*\right) - \breve{\theta}_\tau\right) = (nh)^{1/2}\left(\widehat{\theta}_\tau\left(x_2^*\right) - \theta_{0\tau}\right) + o_p(1),$$

and by Proposition 10

$$W_c\left(x_2^*\right) = nh\left(\widehat{\theta}_\tau\left(x_2^*\right) - \theta_{0\tau}\right)^T \Omega_{1G\theta_\tau}\left(x_2^*\right)^{-1}\left(\widehat{\theta}_\tau\left(x_2^*\right) - \theta_{0\tau}\right) \xrightarrow{d} \chi^2(p). \quad (4.4)$$

It is important to note that the test statistics $W_l\left(x_2^*\right)$ and $W_c\left(x_2^*\right)$ are asymptotically valid at a single point x_2^*. If one wants to consider them over a fixed range of values of x_2^*, say $\left\{x_{2j}^*\right\}_{j=1}^m$, then the test statistics $\max_j W_l\left(x_{2j}^*\right)$ and $\max_j W_c\left(x_{2j}^*\right)$ $(j = 1, \ldots, m)$ can be used instead, as the following proposition shows.

Proposition 11 *Under the assumptions of Proposition 10*, (*i*)

$$\max_{1\le j\le m} W_l\left(x_{2j}^*\right) \xrightarrow{d} \max_j \chi_j^2\left(\kappa_j, l\right),$$

where

$$\kappa_j = f_{X_2}\left(x_{2j}^*\right)\gamma_\tau\left(x_{2j}^*\right)^T\left(R\Omega_{1G\theta_\tau}\left(x_{2j}^*\right)R^T\right)^{-1}\gamma_\tau\left(x_{2j}^*\right),$$

or (*ii*)

$$\max_{1\le j\le m} W_l\left(x_{2j}^*\right) \xrightarrow{p} \infty.$$

Note that the distribution of the test statistic in Proposition 11 is nonstandard, since it involves the maximum of m independent noncentral Chi-squared distributions. However, under the null hypothesis $R\theta_{0\tau}\left(x_2^*\right) = r_\tau\left(x_2^*\right)$, the test statistic is asymptotic distribution free; that is, it does not depend on any nuisance parameters; hence, its distribution can be evaluated numerically or easily simulated.

Finally, we consider inference on the parametric component β_τ; let

$$H_n\colon R\beta_{0\tau} = r_\tau + \gamma_{\tau n}, \quad (4.5)$$

where R is an $l \times k$ matrix of constants and $\gamma_{\tau n}(\cdot)$ is a bounded continuous function that may depend on n. Let

$$W = n\left(R\left(\widehat{\beta}_\tau - r_\tau\right)\right)^T\left(R\widehat{\Omega}_{2*\beta_\tau}R^T\right)^{-1}R\left(\widehat{\beta}_\tau - r_\tau\right) \;\; j = 2, 3$$

denote the Wald statistic for (4.5), where $\widehat{\Omega}_{2*\beta_\tau} = \widehat{\Sigma}_2^{-1}\widehat{\Omega}_{2*}\widehat{\Sigma}_2^{-1}$ and $\widehat{\Omega}_{2*}$ are consistent estimators of Ω_{2km}, Ω_{2p} and Ω_{2np} defined in Theorem 8.

Proposition 12 *Under the assumptions of Theorem 8, if* rank $(R) = l$ $(l \leq k)$ *and* $nh^5 \to 0$*, then under* (4.5) (*i*) *for* $n^{1/2}\gamma_{\tau n} \to \gamma_\tau > 0$ (*for some* $\|\gamma_\tau\| < \infty$)

$$W \overset{d}{\to} \chi^2(\kappa, l),$$

where $\chi^2(\kappa, l)$ *is a noncentral Chi-squared distribution with* l *degrees of freedom and noncentrality parameter* $\kappa = \gamma_\tau^T \left(R\Omega_{2*\beta_\tau} R^T\right)^{-1} \gamma_\tau$*; (ii) for* $n^{1/2}\gamma_{\tau n} \to \infty$*,*

$$W \overset{p}{\to} \infty.$$

5 Simulation study

We first discuss some computational aspects of the proposed estimators and describe how to use the M-M algorithm to estimate the unknown parameters. Let $\varepsilon_{(k)} =: Z - X_1^T\beta_{\tau(k)} - X_3^T\theta_{\tau(k)}(X_2)$ denote the kth iterate in finding the minimum of the objective function, and let

$$\varsigma_\tau\left(\varepsilon|\varepsilon_{(k)}\right) = \frac{1}{4}\left[\frac{\varepsilon^2}{\epsilon + \left|\varepsilon_{(k)}\right|} + (4\tau - 2)\,\varepsilon + c_{(k)}\right]$$

denote the so-called surrogate function, where the constant $c_{(k)}$ is such that $\varsigma\left(\varepsilon_{(k)}|\varepsilon_{(k)}\right)$ is equal to $\rho_\tau\left(\varepsilon_{(k)}\right)$ and $0 < \epsilon \leq 1$ is a tuning parameter to be selected. Then, since $\varsigma\left(\varepsilon|\varepsilon_{(k)}\right) \geq \rho_\tau(\varepsilon)$ for all ε, the unknown parameters can be estimated by minimizing both the local and the global majorizer objective functions

$$\sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\cdot)}\varsigma_\tau\left(\varepsilon_i|\varepsilon_{i(k)}\right) K_h(X_{2i} - x_2), \quad \sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(\cdot)}\varsigma_\tau\left(\widehat{\varepsilon}_i^l|\widehat{\varepsilon}_{i(k)}^l\right),$$

where $\widehat{G}(\cdot)$ is any of the three estimators of $G_0(\cdot)$ and $\widehat{\varepsilon}^l = Z - X_1^T\beta_\tau - X_3^T\widehat{\theta}_\tau^l(X_2)$. As in Hunter and Lange (2000), we use the Gauss–Newton algorithm with direction

$$\Delta_{(k)}(x_2) = -\left[X(x_2)^T W\left(\delta, \widehat{G}(\cdot), \varepsilon_{(k)}, K\right) X(x_2)\right]^{-1} X(x_2)^T d\left(\delta, \widehat{G}(\cdot), \varepsilon, K\right),$$
$$\Delta_{(k)} = -\left[X_1^T W\left(\delta, \widehat{G}(\cdot), \varepsilon_{(k)}\right) X_1\right]^{-1} X_1^T d\left(\delta, \widehat{G}(\cdot), \varepsilon\right),$$

where $X\left(x_2\right)$ is an $n \times (k+2p)$ matrix containing the k, p and p covariates X_{1i}^T, X_{3i}^T and $X_{3i}^T\left(X_{2i}-x_2\right)$ $(i=1,\ldots,n)$,

$$W\left(\delta, \widehat{G}(\cdot), \varepsilon_{(k)}, K\right) = \operatorname{diag}\left[\frac{\delta_1}{\widehat{G}(\cdot)} \frac{1}{\epsilon + \varepsilon_{1(k)}} K_h\left(X_{21}-x_2\right), \ldots, \frac{\delta_n}{\widehat{G}(\cdot)} \frac{1}{\epsilon + \varepsilon_{n(k)}} K_h\left(X_{2n}-x_2\right)\right]^{\mathrm{T}},$$

$$d\left(\delta, \widehat{G}(\cdot), \varepsilon, K\right) = \left[\left(1-2\tau-\frac{\varepsilon_1}{\epsilon+\varepsilon_1}\right) K_h\left(X_{21}-x_2\right), \ldots, \left(1-2\tau-\frac{\varepsilon_n}{\epsilon+\varepsilon_n}\right) K_h\left(X_{2n}-x_2\right)\right]^{\mathrm{T}},$$

and $W\left(\delta, \widehat{G}(\cdot), \varepsilon_{(k)}\right)$ and $d\left(\delta, \widehat{G}(\cdot), \varepsilon\right)$ defined similarly.

The implementation of the M-M algorithm involves the following steps:

1. Set $k=0$, choose the initial values $\left[\beta_\tau^{0T}, a_\tau^{0T}, b_\tau^{0T}\right]^{\mathrm{T}}$ and set $\epsilon n\left|\ln \epsilon\right| = \delta$, with $\delta = 10^{-6}$,
2. Define $\left[\beta_\tau^{k+1T}, a_\tau^{k+1T}, b_\tau^{k+1T}\right]^{\mathrm{T}} = \left[\beta_\tau^{kT}, a_\tau^{kT}, b_\tau^{kT}\right]^{\mathrm{T}} + \Delta_{(k)}\left(x_2\right)/2^k$,
3. Iterate until $\left\|\left[\beta_\tau^{k+1T}, a_\tau^{k+1T}, b_\tau^{k+1T}\right]^{\mathrm{T}} - \left[\beta_\tau^{kT}, a_\tau^{kT}, b_\tau^{kT}\right]^{\mathrm{T}}\right\| < \delta$.

As initial values $\left[\beta_\tau^{0T}, a_\tau^{0T}, b_\tau^{0T}\right]^{\mathrm{T}}$, we choose $\left[0^T, 0^T, 0^T\right]^{\mathrm{T}}$, as the Monte Carlo results presented below seem to suggest that the algorithm is not sensitive to the initial values.[1] The algorithm is very quick, with convergence achieved after few iterations (typically four or five) and each iteration taking between 10 and 15 s on average. Next, we discuss how to choose the bandwidth h. As mentioned by El Ghouch and van Keilegom (2009), the problem of optimally choosing the bandwidth in censored semiparametric quantile regression models is still an open one. Here, we propose a twofold method, which consists of computing for a random subset of the sample—the

[1] In the simulations below, we tried as starting values the following alternative specifications: $\left[\beta_\tau^{0T}, a_\tau^{0T}, b_\tau^{0T}\right]^{\mathrm{T}} = \left[\widehat{\beta}_{q\tau}^T, \widehat{\theta}_{q\tau}^T, 0^T\right]^{\mathrm{T}}$, where $\widehat{\beta}_{q\tau}$ and $\widehat{\theta}_{q\tau}$ are defined as $\widehat{\beta}_{q\tau}, \widehat{\theta}_{q\tau} = \arg\min_{\beta_\tau, \theta_\tau} \sum_{i=1}^n \frac{\delta_i}{\widehat{G}(\cdot)} \rho_\tau\left(Z_i - X_{1i}^T\beta_\tau - X_{3i}^T\theta_\tau\right)$, that is $\widehat{\beta}_{q\tau}$ and $\widehat{\theta}_{q\tau}$ are the estimators of a parametric quantile regression, $\left[\beta_\tau^{0T}, a_\tau^{0T}, b_\tau^{0T}\right]^{\mathrm{T}} = \left[\widehat{\beta}_{q\tau}^T, \widehat{a}_{f\tau}^T, \widehat{b}_{f\tau}^T\right]^{\mathrm{T}}$, where $\widehat{a}_{f\tau}$ and $\widehat{b}_{f\tau}$ are defined as $\widehat{a}_{f\tau}, \widehat{b}_{f\tau} = \arg\min_{a_\tau, b_\tau} \sum_{i=1}^n \frac{\delta_i}{\widehat{G}(\cdot)} \rho_\tau\left(Z_i - X_{1i}^T\widehat{\beta}_{q\tau} - X_{3i}^T\left(a_{\tau f} - b_{\tau f}\left(X_{2i}-x_{2f}\right)\right)\right)$ $K_h\left(X_{2i}-x_{2f}\right)$, where x_{2f} is a chosen point in the support of X_{2i} and the minimization is carried out using the Nelder–Mead algorithm, and finally $\left[\beta_\tau^{0T}, a_\tau^{0T}, b_\tau^{0T}\right]^{\mathrm{T}}$ are chosen as independent draws from a uniform distribution on $(-2, 2)$. All of the above initial values resulted in final estimators with biases and/or IMSEs that were very close (with maximum difference at the second decimal place) to those reported in Tables 1, 2, 3, 4, 5, 6 and 7 in the paper.

training set—S_t with $0 < t < 1$

$$\left[\beta_\tau^{-tT}, a_\tau^{-tT}, b_\tau^{-tT}\right]^{\mathrm{T}}(h) = \arg\min_{\beta_\tau, a_\tau, b_\tau} \sum_{i \in S_t} \frac{\delta_i}{\widehat{G}(\cdot)} \varsigma_\tau\left(\varepsilon_i | \varepsilon_{i(k)}\right) K_h\left(X_{2i} - x_2\right),$$

$$\widehat{\beta}_\tau^{-t}(h) = \arg\min_{\beta_\tau} \sum_{i \in S_\tau} \frac{\delta_i}{\widehat{G}(\cdot)} \varsigma_\tau\left(\widehat{\varepsilon}_i^{-t} | \widehat{\varepsilon}_{i(k)}^{-t}\right),$$

where $\widehat{\varepsilon}_i^{-t} = Z_i - X_{1i}^T \beta_\tau^{-t} - X_{3i}^T \widehat{\theta}_\tau^{-t}(X_{2i})$ and then using the remaining part of the sample S_{1-t}—the validation set—to select h as

$$\widehat{h} = \arg\min_h \sum_{i \in S_{1-t}} \frac{\delta_i}{\widehat{G}(\cdot)} \varsigma_\tau\left(\widehat{\varepsilon}_i^{-t}(h) | \widehat{\varepsilon}_{i(k)}^{-t}(h)\right). \tag{5.1}$$

In the simulations, 80% of the censored observations and 80% of the uncensored observations are used as the training set and the remaining 20% of the observations are used as the validation set. In this way, both the training and validation sets contain the original proportion of censored data.[2]

We consider the following semiparametric specifications

$$Y_i = \beta_{00\tau} + X_{11i}\beta_{10\tau} + X_{12i}\beta_{20\tau} + \sin(2\pi X_{2i}) + X_{11i}^{1/2}\varepsilon_{i\tau} \quad i = 1, \ldots, n, \tag{5.2}$$

$$Y_i = \beta_{00\tau} + X_{11i}\beta_{10\tau} + X_{12i}\beta_{20\tau} + X_{3i}^T\left[\cos(\pi X_{2i}), X_{2i}^2\right]^{\mathrm{T}} + \varepsilon_{i\tau} \quad i = 1, \ldots, n \tag{5.3}$$

where X_{11i}, X_{12i} and X_{2i} are generated independently from, respectively, a uniform distribution on (0, 2), a Bernoulli distribution with probability of success $p = 1/2$ and a uniform distribution on (0, 1), $X_{3i}^T = [X_{31i}, X_{32i}]$ are jointly normal with mean zero, variance 1 and correlation coefficient 0.5 and the unobservable error term $\varepsilon_{i\tau}$ has zero τ quantile. In the simulations, we specify the unknown parameter vector as $\beta_{0\tau} = [\beta_{00\tau}, \beta_{10\tau}, \beta_{20\tau}]^{\mathrm{T}} = [1, 2, 1/2]^{\mathrm{T}}$, use the Epanechnikov kernel and consider two sample sizes: $n = 100$ and $n = 400$.

We first consider the case of independent censoring; in this case, the censoring variables $\{C_i\}_{i=1}^n$ are generated from a normal distribution $N(c, 1)$, where the constant c is chosen to obtain two levels of censoring, a low one at 15% and a medium one at 45%, whereas the artificial censoring variable L is chosen so that $\widehat{G}(L) = 0.01$. Tables 1 and 2 report, respectively, the bias and the standard error for the three estimators $\widehat{\beta}_\tau = \left[\widehat{\beta}_{0\tau}, \widehat{\beta}_{1\tau}, \widehat{\beta}_{2\tau}\right]^{\mathrm{T}}$ at the five quantiles $\tau = (0.10, 0.25, 0.5, 0.75, 0.90)$ for the semiparametric quantile regressions (5.2) and (5.3) using 1000 replications and two specifications for the distribution of $\varepsilon_{i\tau}$: a standard normal ($N(0, 1)$) and a Chi-squared distribution with four degrees of freedom ($\chi^2(4)$). The standard errors are calculated using the resampling technique of Sect. 3.3 with the number of replications B set to

[2] I am grateful to one referee for suggesting this procedure.

Table 1 Bias and standard errors (SE) for the semiparametric quantile regression (5.2) with independent censoring

τ	$C = 15$[a]						$C = 45$[a]					
	$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$		$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$	
	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$n = 100$	$N(0, 1)$						$N(0, 1)$					
0.10	.012	.161	.025	.183	.024	.128	.018	.169	.029	.191	.027	.153
0.25	.011	.159	.024	.184	.022	.128	.018	.168	.028	.190	.026	.153
0.50	.012	.160	.025	.184	.022	.129	.020	.171	.030	.192	.026	.154
0.75	.013	.164	.025	.186	.023	.131	.021	.171	.031	.193	.027	.155
0.90	.014	.170	.027	.191	.025	.138	.022	.175	.031	.198	.028	.159
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.014	.169	.032	.201	.021	.155	.020	.211	.036	.217	.033	.189
0.25	.014	.168	.031	.201	.019	.156	.019	.210	.035	.216	.033	.188
0.50	.016	.170	.032	.203	.022	.157	.022	.212	.036	.217	.035	.188
0.75	.018	.172	.033	.203	.033	.159	.023	.214	.036	.218	.035	.190
0.90	.018	.178	.035	.209	.035	.163	.024	.219	.036	.223	.037	.195
$n = 400$	$N(0, 1)$						$N(0, 1)$					
0.10	.009	.111	.023	.121	.024	.092	.017	.124	.026	.127	.023	.104
0.25	.009	.111	.022	.120	.022	.093	.017	.125	.025	.127	.021	.103
0.50	.010	.113	.023	.123	.023	.094	.018	.127	.025	.129	.022	.105
0.75	.011	.115	.024	.124	.024	.095	.020	.131	.026	.133	.024	.106
0.90	.012	.121	.026	.129	.024	.103	.022	.138	.026	.140	.025	.111
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.013	.136	.031	.150	.024	.115	.018	.156	.030	.156	.022	.114
0.25	.012	.136	.029	.150	.025	.116	.017	.156	.030	.157	.023	.114
0.50	.013	.1376	.029	.151	.025	.116	.019	.158	.032	.158	.023	.115
0.75	.015	.137	.030	.153	.027	.117	.021	.158	.032	.160	.024	.118
0.90	.015	.142	.031	.159	.028	.121	.022	.162	.033	.165	.026	.123

[a] Percentage of censoring

500 and the random variables ξ_i generated from an exponential distribution with mean 1.

For the dependent censoring case, we assume a Cox proportional hazard model with $C_i = \exp(X_{2i}\gamma_0)$, where γ_0 is chosen to obtain the same level of censoring as that of the independent censoring case, namely 15% and 45%. We use Breslow's (1972) estimator to estimate $G(Z_i|X_{2i}) = G_{\gamma_0}(Z_i|X_{2i})$ parametrically, whereas we use the Epanechnikov kernel to compute the weights $\omega_i(\cdot)$ in the local Kaplan–Meier estimator (2.6) for the nonparametric estimation of $G(Z_i|X_{2i})$. Tables 3 and 4 report, respectively, the bias and the standard error for the three estimators $\widehat{\beta}_\tau = \left[\widehat{\beta}_{0\tau}, \widehat{\beta}_{1\tau}, \widehat{\beta}_{2\tau}\right]^{\mathrm{T}}$ and the five quantiles $\tau = (0.10, 0.25, 0.5, 0.75, 0.90)$ for the semiparametric quantile regressions (5.2) and (5.3) and the same two specifications of the distribution of $\varepsilon_{i\tau}$

Table 2 Bias and standard errors (SE) for the semiparametric quantile regression (5.3) with independent censoring

τ	$C = 15$[a]						$C = 45$[a]					
	$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$		$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$	
	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$n = 100$	$N(0, 1)$						$N(0, 1)$					
0.10	.015	.160	.027	.192	.021	.132	.020	.172	.031	.194	.030	.160
0.25	.014	.160	.027	.193	.020	.132	.020	.172	.031	.195	.030	.160
0.50	.015	.161	.028	.193	.021	.134	.021	.173	.031	.195	.030	.162
0.75	.015	.163	.029	.197	.022	.135	.022	.175	.032	.198	.031	.162
0.90	.016	.167	.030	.204	.023	.139	.022	.180	.034	.204	.033	.165
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.016	.181	.034	.215	.029	.160	.024	.207	.037	.231	.038	.188
0.25	.015	.181	.034	.216	.029	.160	.023	.208	.036	.231	.038	.189
0.50	.017	.183	.035	.216	.030	.162	.024	.210	.036	.233	.038	.191
0.75	.017	.185	.038	.218	.030	.164	.026	.212	.038	.235	.040	.191
0.90	.019	.190	.038	.223	.032	.168	.027	.218	.039	.239	.041	.194
$n = 400$	$N(0, 1)$						$N(0, 1)$					
0.10	.011	.120	.024	.122	.026	.109	.019	.130	.027	.140	.028	.116
0.25	.010	.120	.022	.121	.024	.102	.018	.131	.026	.138	.028	.116
0.50	.011	.121	.022	.121	.025	.103	.019	.132	.026	.139	.029	.118
0.75	.013	.122	.023	.123	.025	.105	.020	.134	.028	.141	.030	.118
0.90	.013	.126	.023	.127	.025	.108	.020	.138	.028	.148	.032	.121
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.014	.132	.028	.152	.026	.120	.021	.154	.029	.160	.028	.118
0.25	.013	.133	.029	.148	.025	.119	.020	.154	.029	.159	.028	.119
0.50	.014	.135	.030	.150	.027	.120	.020	.156	.030	.161	.029	.121
0.75	.014	.136	.031	.151	.028	.122	.021	.157	.031	.164	.030	.122
0.90	.014	.141	.032	.156	.028	.128	.021	.163	.032	.169	.032	.125

[a] Percentage of censoring

used in Tables 1 and 2. Tables 5 and 6 report the same results using the local Kaplan–Meier. As with Tables 1 and 2, Tables 3, 4, 5 and 6 are based on 1000 replications with the standard errors calculated using $B = 500$ replications with ξ_i generated from an exponential distribution with mean 1.

The results of Tables 1, 2, 3, 4, 5 and 6 suggest that the proposed estimators perform well with reasonable sample sizes: The biases are statistically insignificant, and the standard errors are getting smaller as the sample size increases. As expected, the standard errors increase with the level of censoring, especially at the 0.90 quantile, which can be explained by the fact that right censoring affects the higher quantiles of the conditional distribution of the responses. Finally, between the three estimators of the survival function G_0, those based on the local Kaplan–Meier estimator seem to be characterized by a slightly larger bias and standard error, which can be explained by

Table 3 Bias and standard errors (SE) for the semiparametric quantile regression (5.2) with dependent censoring and Breslow's (1972) estimator

τ	$C = 15$[a]						$C = 45$[a]					
	$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$		$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$	
	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$n = 100$	$N(0, 1)$						$N(0, 1)$					
0.10	.009	.155	.031	.177	.023	.131	.021	.165	.026	.190	.027	.146
0.25	.010	.157	.030	.177	.023	.132	.021	.165	.026	.191	.028	.146
0.50	.011	.159	.031	.179	.024	.133	.023	.168	.026	.191	.029	.148
0.75	.013	.160	.032	.181	.025	.135	.023	.170	.030	.192	.030	.150
0.90	.014	.166	.031	.187	.025	.139	.024	.175	.030	.196	.031	.155
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.015	.162	.031	.190	.028	.148	.022	.200	.038	.216	.030	.185
0.25	.014	.162	.032	.191	.028	.149	.023	.199	.037	.217	.030	.185
0.50	.014	.164	.033	.193	.029	.151	.024	.201	.037	.218	.031	.185
0.75	.016	.165	.035	.194	.030	.152	.025	.202	.038	.218	.031	.186
0.90	.017	.169	.035	.198	.030	.158	.025	.208	.038	.221	.031	.190
$n = 400$	$N(0, 1)$						$N(0, 1)$					
0.10	.008	.108	.024	.115	.020	.095	.018	.120	.026	.131	.025	.101
0.25	.008	.109	.024	.115	.020	.096	.018	.120	.024	.126	.024	.101
0.50	.009	.111	.025	.117	.020	.096	.020	.122	.024	.128	.024	.102
0.75	.011	.112	.025	.119	.021	.099	.021	.124	.025	.130	.026	.104
0.90	.014	.117	.026	.125	.023	.104	.024	.123	.026	.135	.027	.110
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.011	.138	.028	.141	.021	.112	.021	.153	.034	.148	.028	.118
0.25	.012	.138	.026	.141	.022	.112	.020	.153	.032	.148	.026	.118
0.50	.012	.140	.026	.142	.022	.113	.021	.155	.034	.150	.026	.117
0.75	.014	.141	.029	.145	.024	.115	.025	.158	.035	.150	.028	.120
0.90	.014	.149	.030	.148	.025	.117	.026	.162	.038	.155	.028	.124

[a] Percentage of censoring

the fact that the local Kaplan–Meier estimator has a slightly higher integrated mean squared error compared to that of the Kaplan–Meier and Breslow's (1972) estimators; see Table 7 and the comments after it.

Figure 1 shows the five quantiles $\tau = (0.10, 0.25, 0.5, 0.75, 0.90)$ estimates for the nonparametric component—estimated with the global estimates $\widehat{\beta}_\tau$ replacing $\beta_{0\tau}$—in the case of the semiparametric quantile regression (5.2) with normal unobservable errors, censoring level at 15% and sample size $n = 100$.

Figure 2 shows the five quantiles $\tau = (0.10, 0.25, 0.5, 0.75, 0.90)$ estimates for the first nonparametric component of the semiparametric quantile regression (5.3) with Chi-squared unobservable errors, censoring level at 15% and sample size $n = 100$.

To measure the performance of the estimators $\widehat{\theta}_\tau(\cdot)$ for the nonparametric components, we use the (empirical) integrated mean squared error (IMSE) as in De Backer

Table 4 Bias and standard errors (SE) for the semiparametric quantile regression (5.3) with dependent censoring and Breslow's (1972) estimator

τ	$C = 15$[a]						$C = 45$[a]					
	$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$		$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$	
	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$n = 100$	$N(0, 1)$						$N(0, 1)$					
0.10	.011	.161	.035	.180	.024	.130	.023	.168	.035	.201	.027	.151
0.25	.011	.161	.034	.180	.024	.131	.023	.169	.035	.201	.027	.151
0.50	.011	.164	.034	.181	.025	.130	.026	.171	.035	.203	.028	.153
0.75	.013	.166	.036	.183	.027	.132	.027	.172	.037	.205	.030	.155
0.90	.013	.169	.037	.188	.028	.136	.028	.176	.038	.210	.031	.159
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.013	.168	.034	.201	.025	.148	.024	.195	.038	.215	.031	.169
0.25	.013	.168	.034	.201	.025	.148	.024	.195	.038	.215	.031	.169
0.50	.013	.171	.035	.203	.026	.151	.025	.197	.039	.217	.032	.172
0.75	.015	.173	.036	.205	.027	.153	.026	.199	.040	.218	.032	.174
0.90	.017	.179	.036	.210	.028	.158	.027	.204	.041	.223	.032	.180
$n = 400$	$N(0, 1)$						$N(0, 1)$					
0.10	.009	.117	.026	.121	.020	.103	.019	.121	.025	.133	.026	.106
0.25	.009	.116	.025	.121	.020	.103	.018	.121	.026	.133	.025	.106
0.50	.010	.115	.025	.123	.020	.105	.019	.123	.027	.135	.026	.108
0.75	.012	.119	.027	.125	.021	.106	.020	.125	.028	.136	.026	.109
0.90	.015	.123	.028	.127	.022	.111	.021	.128	.029	.141	.028	.114
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.012	.139	.027	.140	.019	.112	.020	.156	.029	.147	.026	.122
0.25	.010	.140	.026	.140	.020	.112	.020	.156	.029	.148	.027	.121
0.50	.011	.142	.026	.142	.020	.114	.021	.158	.029	.149	.028	.121
0.75	.012	.143	.028	.144	.021	.115	.023	.160	.030	.152	.029	.122
0.90	.013	.147	.029	.149	.023	.120	.024	.164	.031	.156	.030	.126

[a] Percentage of censoring

et al. (2017), which is given by

$$\text{IMSE}\left(\widehat{\theta}_{\tau}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{M}\sum_{j=1}^{M}\left(\widehat{\theta}_{\tau j}\left(X_{2i}\right) - \theta_{0\tau}\left(X_{2i}\right)\right)\right),$$

where we take $M = 100$ and $(X_{2i})_{i=1}^{N}$ with $N = 20$ are randomly generated from a uniform distribution on $(0, 1)$. Table 7 reports the IMSE for the estimator considered in Fig. 1 (at 15% and 45% censoring level and both unobservable errors ε_{τ} specifications).

Table 7 shows that among the three different estimators of $G_0(\cdot)$, those based on the local Kaplan–Meier estimator are typically characterized by a larger IMSE. This result is not surprising, though: Firstly, the local Kaplan–Meier estimator is the

Table 5 Bias and standard errors (SE) for the semiparametric quantile regression (5.2) with dependent censoring and local Kaplan–Meier estimator

τ	$C = 15^{a}$						$C = 45^{a}$					
	$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$		$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$	
	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$n = 100$	$N(0, 1)$						$N(0, 1)$					
0.10	.010	.161	.026	.185	.026	.134	.022	.174	.028	.194	.030	.155
0.25	.011	.161	.027	.186	.025	.134	.022	.174	.028	.194	.029	.156
0.50	.011	.162	.028	.188	.026	.136	.024	.176	.028	.195	.029	.156
0.75	.013	.168	.029	.190	.026	.136	.024	.178	.030	.197	.031	.158
0.90	.014	.173	.030	.194	.027	.140	.025	.182	.031	.203	.033	.161
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.014	.171	.027	.198	.027	.147	.026	.199	.032	.220	.030	.188
0.25	.013	.171	.028	.198	.027	.148	.025	.201	.032	.221	.031	.188
0.50	.013	.172	.029	.199	.027	.150	.027	.203	.033	.223	.032	.189
0.75	.015	.176	.031	.200	.029	.151	.027	.205	.035	.225	.034	.191
0.90	.016	.179	.033	.205	.029	.155	.028	.209	.036	.229	.035	.195
$n = 400$	$N(0, 1)$						$N(0, 1)$					
0.10	.008	.118	.023	.128	.021	.099	.019	.135	.027	.136	.025	.109
0.25	.009	.119	.022	.129	.020	.099	.018	.135	.026	.135	.026	.110
0.50	.010	.121	.023	.130	.021	.101	.019	.136	.027	.134	.026	.111
0.75	.011	.121	.024	.131	.022	.102	.020	.137	.029	.137	.027	.111
0.90	.012	.125	.025	.135	.023	.106	.022	.141	.029	.141	.029	.115
	$\chi^2(4)$						$\chi^2(4)$					
0.10	.012	.135	.023	.159	.025	.118	.020	.158	.026	.167	.027	.129
0.25	.012	.136	.024	.159	.025	.118	.020	.158	.027	.167	.026	.129
0.50	.012	.137	.024	.161	.025	.120	.022	.159	.028	.168	.025	.131
0.75	.014	.139	.026	.160	.028	.121	.025	.160	.029	.170	.026	.133
0.90	.015	.142	.027	.166	.029	.125	.026	.164	.030	.174	.028	.137

[a] Percentage of censoring

only one depending on a bandwidth (b) and its choice has some bearings on the performance of the quantile estimators of the nonparametric component. There are few methods available to optimally choose the bandwidth b, but they are either not easy to implement (see, e.g., Van Keilegom et al. 2001) or require a bootstrap approach (see, e.g., Li and Datta 2001). Here, we use the simple ad hoc selection method based on $\widehat{b} = 2\left|\widehat{\sigma}_{Z,X_2}\right| n^{-1/5}$, where $\widehat{\sigma}_{Z,X_2}$ is the sample covariance between Z_i and X_{2i}.[3] Secondly, and perhaps more importantly, the performance of the semiparametric quantile estimator based on the local Kaplan–Meier estimator is compared to that based on

[3] To assess the sensitivity of the IMSE to this choice of b, we considered two alternative bandwidths, $\widehat{b}_1 = \widehat{b}/4$ and $\widehat{b}_2 = 4\widehat{b}$, and computed the corresponding IMSE's. The results of the simulations indicated that the IMSE's of the resulting quantile estimators were still larger than those based on Breslow's (1972) estimator.

Table 6 Bias and standard errors (SE) for the semiparametric quantile regression (5.3) with dependent censoring and local Kaplan–Meier estimator

τ	$C = 15$[a]						$C = 45$[a]					
	$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$		$\beta_{0\tau}$		$\beta_{1\tau}$		$\beta_{2\tau}$	
	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
$n = 100$	N (0, 1)						N (0, 1)					
0.10	.010	.174	.030	.194	.025	.138	.024	.188	.031	.208	.031	.158
0.25	.011	.174	.029	.194	.026	.139	.024	.189	.031	.208	.030	.157
0.50	.013	.176	.030	.196	.027	.141	.026	.190	.031	.210	.029	.159
0.75	.014	.177	.032	.199	.030	.143	.028	.192	.033	.211	.030	.161
0.90	.015	.181	.032	.203	.030	.146	.029	.195	.035	.214	.033	.165
	χ^2 (4)						χ^2 (4)					
0.10	.014	.181	.032	.211	.027	.151	.025	.199	.034	.219	.032	.173
0.25	.014	.181	.032	.212	.027	.152	.024	.199	.034	.220	.032	.174
0.50	.015	.183	.034	.213	.028	.153	.025	.201	.035	.223	.033	.175
0.75	.017	.184	.034	.214	.029	.155	.027	.202	.037	.223	.034	.175
0.90	.018	.188	.035	.218	.030	.158	.028	.205	.037	.227	.035	.179
$n = 400$	N (0, 1)						N (0, 1)					
0.10	.007	.125	.022	.142	.022	.111	.019	.141	.024	.161	.022	.122
0.25	.008	.125	.023	.143	.021	.111	.020	.141	.025	.161	.023	.123
0.50	.010	.125	.024	.145	.022	.113	.021	.152	.027	.162	.024	.125
0.75	.011	.128	.026	.147	.023	.115	.023	.153	.027	.162	.027	.128
0.90	.012	.132	.027	.149	.024	.118	.023	.155	.028	.168	.028	.133
	χ^2 (4)						χ^2 (4)					
0.10	.011	.141	.026	.148	.021	.115	.021	.159	.027	.151	.022	.130
0.25	.010	.142	.025	.148	.022	.116	.021	.160	.028	.151	.024	.131
0.50	.010	.141	.025	.149	.022	.116	.023	.161	.030	.154	.025	.132
0.75	.013	.145	.028	.150	.025	.119	.024	.163	.033	.155	.028	.135
0.90	.013	.149	.029	.154	.026	.122	.024	.167	.034	.159	.029	.138

[a] Percentage of censoring

Breslow's (1972) estimator. The dependent censoring mechanism is fully parametric; hence, an estimator based on maximum likelihood will always be more accurate (in terms of IMSE) than a nonparametric one. It is also important to note that the local Kaplan–Meier estimator is robust to misspecification as opposed to Breslow's (1972) estimator, which is an important feature in applied research, especially in situations where a parametric specification of the survival distribution seems questionable.

Finally, we investigate the finite sample properties of the test statistic of Proposition 11. We consider the semiparametric quantile regression (5.2) with the null hypothesis

$$H_0 \colon \theta_{0\tau}\left(x_{2*j}\right) = \theta_{0\tau} = 1, \tag{5.4}$$

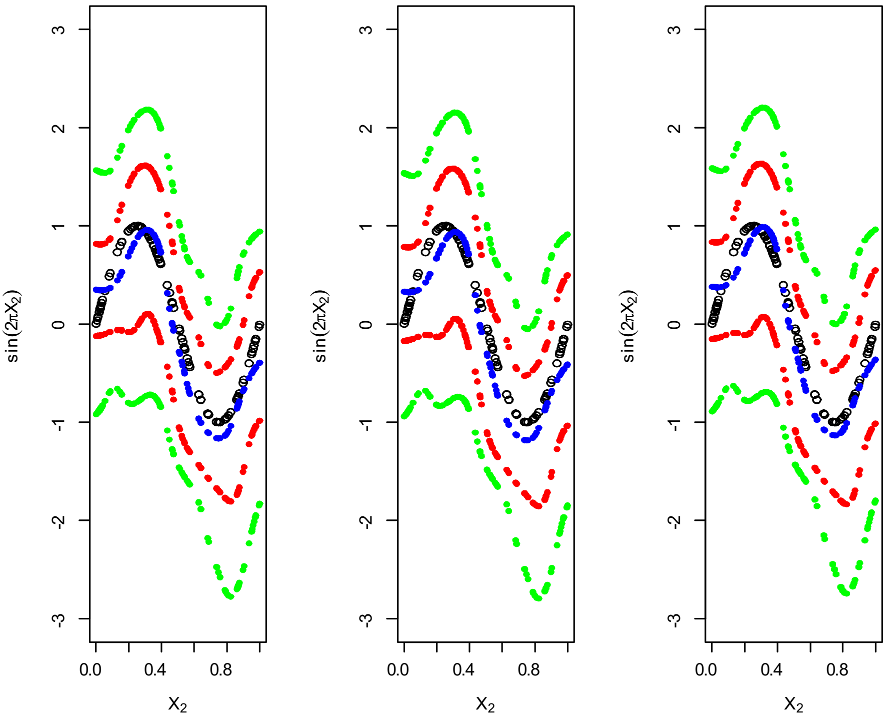

Fig. 1 Quantile estimates (full circles) for $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$ of the unknown nonparametric component $\theta_{0\tau}(X_{2i}) = \sin(2\pi X_{2i})$ (empty circle). Left panel Kaplan–Meier estimator, center panel Breslow's (1972) estimator and right panel local Kaplan–Meier estimator of the unknown survival distribution G_0

versus the sequence of alternative hypotheses indexed by $\delta = [0, 0.2, 0.4, 0.6, 0.8, 1, 1.2]$

$$H_1\colon 1 + \delta\left(\theta_{0\tau}\left(x_{2*j}\right) - 1\right). \tag{5.5}$$

Table 8 reports the finite sample size (corresponding to $\delta = 0$) for $x_{2*j} = 0.1j$ and $j = 1, \ldots, 8$ at a 0.10 and 0.05 nominal level for the semiparametric quantile regression (5.2) with the three estimators of $G_0(\cdot)$, level of censoring at 15% and 2 sample sizes $n = 100$ and $n = 400$, using 5000 replications and bandwidth fixed at $h = h^{\text{ave}}$ where h^{ave} is the average of the bandwidths used to obtain Tables 1, 3 and 5. The critical values of the nonstandard distribution given in Proposition 11 are calculated using 10^5 simulations and are [3.365, 4.779] and [3.044, 4.345] for $n = 100$ and $n = 400$, respectively.

Figure 3 (and its magnified version—at the lower and upper values of δ—Fig. 4) shows the size-adjusted finite sample power of the test statistic $\max_j W_l\left(x^*_{2j}\right)$ of Proposition 11 under the alternative hypothesis (5.5) for the semiparametric quantile regression (5.2) with the three estimators of $G_0(\cdot)$, the unobservable errors $\chi^2(4)$ and $n = 100$, computed using 1000 replications for each value of δ. Figure 3 shows that the test statistic has good power properties for the three estimators of $G_0(\cdot)$, although

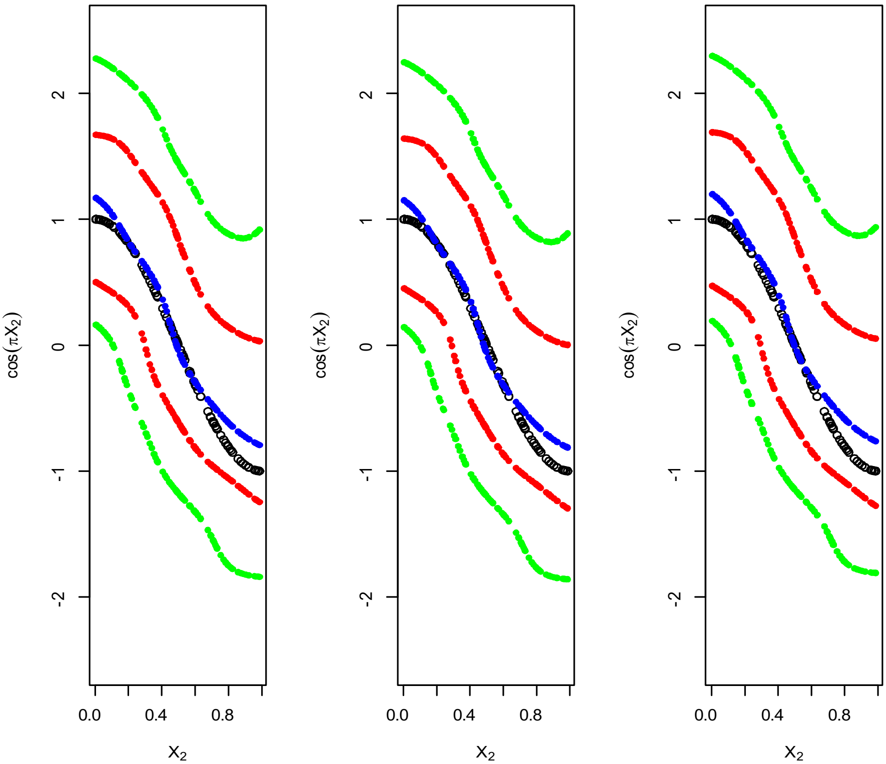

Fig. 2 Quantile estimates (full circles) for $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$ of the unknown nonparametric component $\theta_{0\tau}\ (X_{2i}) = \cos\ (\pi X_{2i})$ (empty circle). Left panel Kaplan–Meier estimator, center panel Breslow's (1972) estimator and right panel local Kaplan–Meier estimator of the unknown survival distribution G_0

the power is slightly lower at the lower and upper quantile, as Fig. 4 shows. Results for the other cases are similar and hence are not reported.

6 Empirical application

We illustrate the applicability of the proposed method by considering the same lung cancer study used by Ying et al. (1995). In this clinical study, 121 patients with limited-stage lung cancer were randomly assigned to two groups (A and B) in which the sequencing of the standard therapy based on etoposide (E) and cisplatin (P) is reversed: group A, P followed by E and group B, E followed by P. At the time of the study, there was no loss to follow-up and each death time was either observed or administratively censored, so that the censoring variable does not depend on the covariates, which are the treatment indicator and the patient's entry age. Let Y_i $(i = 1, \ldots, 121)$ denote the base 10 logarithm of the ith patient failure time with a censoring proportion of about 19%. To investigate the age-adjusted treatment difference, we consider the following

Table 7 IMSE for the estimator of $\theta_{\tau 0}(\cdot)$ in the censored semiparametric quantile regression (5.2)

τ	$C = 15$[a]			$C = 45$[a]		
	$\widehat{G}(\cdot)$[b]	$\widehat{G}(\cdot)$[c]	$\widehat{G}(\cdot)$[d]	$\widehat{G}(\cdot)$[b]	$\widehat{G}(\cdot)$[c]	$\widehat{G}(\cdot)$[d]
	$N(0, 1)$			$N(0, 1)$		
0.10	.142	.140	.146	.148	.145	.152
0.25	.142	.140	.145	.146	.144	.154
0.50	.140	.141	.144	.144	.141	.153
0.75	.146	.144	.150	.149	.151	.158
0.90	.157	.153	.158	.161	.155	.168
	$\chi^2(4)$			$\chi^2(4)$		
0.10	.146	.142	.150	.151	.148	.157
0.25	.145	.142	.151	.151	.150	.157
0.50	.145	.141	.152	.150	.151	.155
0.75	.148	.146	.155	.152	.151	.160
0.90	.161	.157	.161	.162	.163	.170

[a] Percentage of censoring
[b] Kaplan–Meier estimator
[c] Breslow's (1972) estimator
[d] Local Kaplan–Meier estimator

semiparametric quantile regression model

$$Q_{Y_i|X_i}(\tau|X_i) = \beta_{00\tau} + X_{1i}\beta_{10\tau} + \theta_{0\tau}(X_{2i}), \tag{6.1}$$

where $X_{1i} = 0$ if the ith patient is in group A and 1 otherwise, and X_{2i} is the patient's entry age. We assume independent censoring as at the time of the study there was no loss of follow-up, so that each death time was either observed or administratively censored. Thus, the censoring variable does not depend on the covariates (see also Ying et al. 1995). Table 8 reports the estimates of $\beta_{00\tau}$ and $\beta_{10\tau}$ with the 95% confidence intervals for the three quantiles $\tau = (0.25, 0.5, 0.75)$ based on $B = 500$ resampled data with ξ_i generated from an exponential distribution with mean 1, whereas Fig. 5 shows the estimates of $\theta_{0\tau}(\cdot)$ again for the three quantiles $\tau = (0.25, 0.5, 0.75)$.

The results of Table 9 show that for patients with the same entry age, the first quartile and median survival time of group A are larger than that of group B. This is consistent with the findings of Ying et al. (1995) and Zhou (2006), which reported estimates and confidence intervals, respectively, of $\widehat{\beta}_{1\tau} = -0.163\,(-0.388, -0.35)$ and $\widehat{\beta}_{1\tau} = -0.171\,(-0.335, -0.007)$ for $\tau = 0.5$. However, at the third quartile, groups A and B do not show any statistically different survival times (t-statistic is equal to -0.146 with p value for the one-sided alternative of 0.442). Further statistical analysis shows that the survival times of the two groups become statistically insignificant at $\tau = 0.64$ with $\widehat{\beta}_{1\tau} = 0.084$, t-statistic equals 0.674 and associated p value of 0.749; furthermore, at $\tau = 0.92$, we find that the survival time of group B becomes longer than that of group A, since $\widehat{\beta}_{1\tau} = 0.144$ with a t-statistic equal to 2.66 and associated p value for the one-sided alternative equal to 0.004.

Table 8 Finite sample sizes of the test statistic $\max_j W_l\left(x_{2j}^*\right)$ for the null hypothesis (5.4) in the censored semiparametric quantile regression (5.2)

τ	$C = 15$[a]						$C = 45$[a]					
	$\widehat{G}(\cdot)$[b]		$\widehat{G}(\cdot)$[c]		$\widehat{G}(\cdot)$[d]		$\widehat{G}(\cdot)$[b]		$\widehat{G}(\cdot)$[c]		$\widehat{G}(\cdot)$[d]	
	.100	.050	.100	.050	.100	.050	.100	.050	.100	.050	.100	.050
$n = 100$	$N(0, 1)$						$N(0, 1)$					
0.10	.119	.055	.114	.054	.123	.054	.121	.056	.124	.056	.124	.060
0.25	.118	.055	.115	.055	.122	.056	.122	.056	.123	.057	.125	.060
0.50	.120	.056	.117	.056	.125	.056	.123	.058	.126	.059	.126	.062
0.75	.119	.059	.118	.059	.128	.060	.124	.059	.129	.061	.130	.063
0.90	.121	.061	.119	.062	.129	.063	.125	.062	.128	.062	.132	.065
$n = 400$												
0.10	.111	.050	.111	.053	.117	.052	.120	.051	.118	.052	.120	.057
0.25	.112	.051	.110	.052	.118	.052	.120	.052	.119	.052	.120	.058
0.50	.113	.052	.109	.049	.119	.054	.121	.054	.120	.055	.121	.056
0.75	.118	.054	.112	.053	.121	.057	.122	.055	.121	.056	.122	.059
0.90	.119	.058	.113	.055	.122	.058	.123	.056	.122	.057	.123	.060
$n = 100$	$\chi^2(4)$						$\chi^2(4)$					
0.10	.120	.055	.114	.060	.125	.062	.123	.060	.124	.061	.127	.061
0.25	.121	.056	.114	.059	.126	.062	.124	.059	.125	.060	.128	.062
0.50	.120	.056	.115	.059	.128	.063	.125	.061	.125	.063	.130	.063
0.75	.121	.058	.117	.062	.131	.066	.128	.063	.127	.065	.131	.065
0.90	.123	.059	.118	.063	.132	.067	.128	.065	.128	.066	.132	.067
$n = 400$												
0.10	.114	.057	.112	.054	.118	.055	.119	.053	.120	.055	.121	.057
0.25	.115	.055	.112	.055	.119	.055	.119	.054	.120	.055	.122	.057
0.50	.115	.056	.114	.056	.120	.056	.120	.055	.121	.057	.123	.058
0.75	.118	.056	.115	.057	.121	.059	.123	.057	.121	.058	.125	.060
0.90	.120	.057	.116	.058	.122	.060	.124	.059	.123	.060	.128	.063

[a] Percentage of censoring
[b] Kaplan–Meier estimator
[c] Breslow's (1972) estimator
[d] Local Kaplan–Meier estimator

Table 9 Estimates and confidence intervals for the lung cancer study

	$\beta_{0\tau}$		$\beta_{1\tau}$	
$\tau = 0.25$	2.992	(2.362, 3.351)	−0.180	(−0.287, −0.057)
$\tau = 0.50$	2.913	(2.412, 3.401)	−0.113	(−0.234, −0.012)
$\tau = 0.75$	2.764	(2.121, 3.123)	−0.014	(−0.182, 0.083)

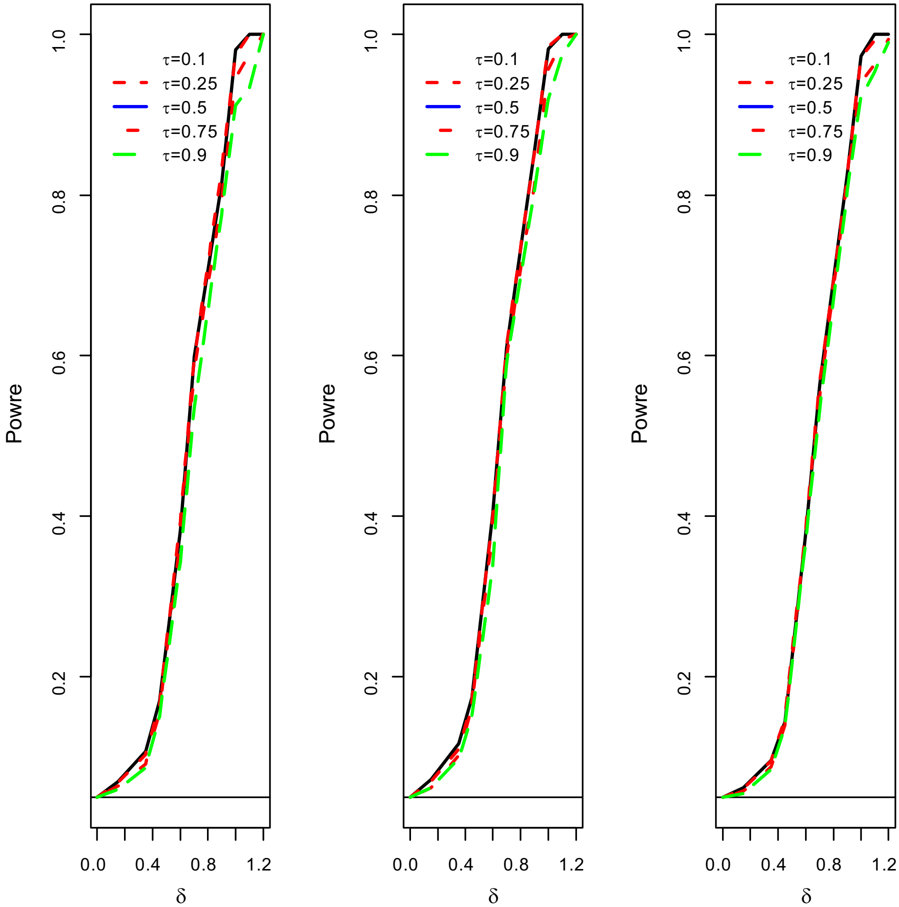

Fig. 3 Finite sample power for the test statistic max $_j W_l\left(x_{2j}^*\right)$ for (5.5) with χ^2 (4) errors and $n = 100$. Left panel Kaplan–Meier estimator, center panel Breslow's (1972) estimator and right panel local Kaplan–Meier estimator of the unknown survival distribution G_0

Finally, we test for the constancy of $\theta_{0\tau}(X_{2i})$ using the $\max_j W_c\left(x_{2j}^*\right)$ statistic evaluated at $x_2^* = [40, 44, 48, 52, 59, 64, 70, 74]$ (i.e., $j = 8$). The sample values of $\max_j W_c\left(x_{2j}^*\right)$ for $\tau = (0.25, 0.5, 0.75)$ are, respectively, 7.31, 6.96, 7.04 with corresponding p values of 0.019, 0.022 and 0.021; hence, the null hypothesis of constancy is rejected at the 0.05 nominal level. Taken together, these results indicate the usefulness of the semiparametric methods for quantile regression proposed in this paper.

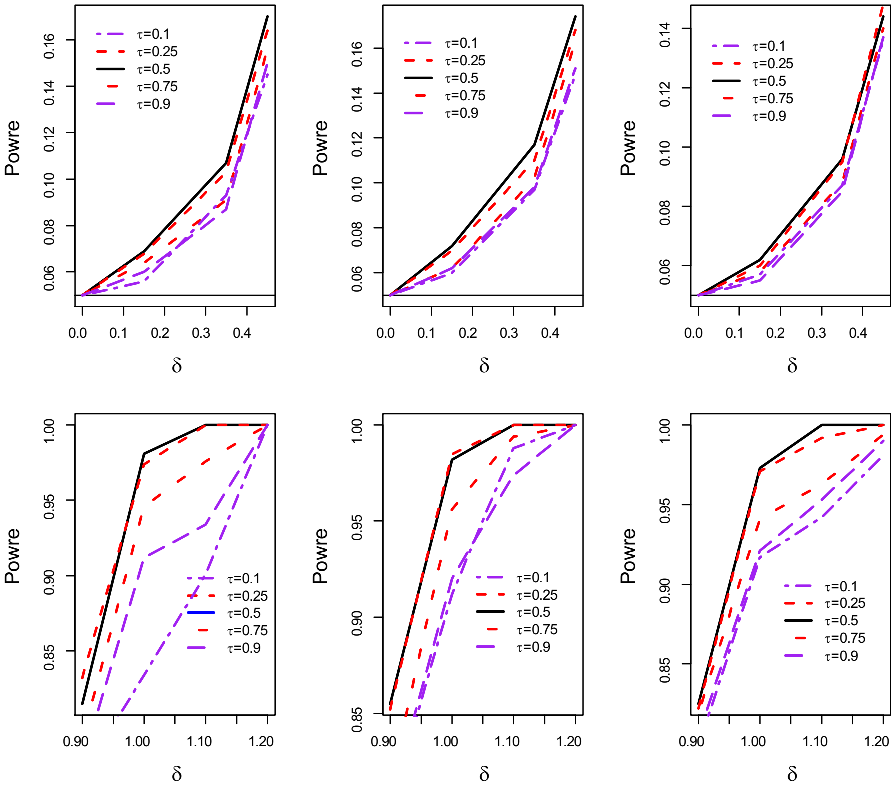

Fig. 4 Magnified version of Fig. 3 for $0 \leq \delta \leq 0.4$ and $0.9 \leq \delta \leq 1.2$

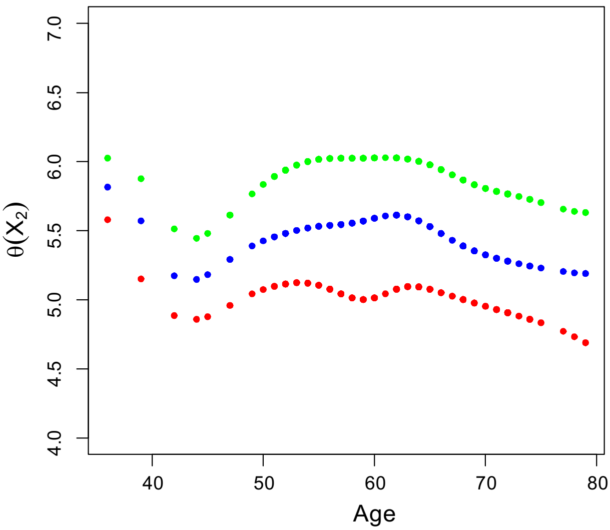

Fig. 5 Nonparametric quantile estimates for $\tau = (0.25, 0.50, 0.75)$ of the age function

7 Conclusions

In this paper, we propose a general method to estimate the unknown parameters in semi-parametric quantile regression models when the response variable is subject to random censoring. The method is based on the inverse probability of censoring weighting and can accommodate the cases of independent and dependent censoring. The paper also proposes test statistics that can be used to test local linear hypotheses (including those of constancy) of the nonparametric component. A Monte Carlo study shows that the resulting estimators and test statistics perform well in finite samples, whereas an empirical application illustrates the practical usefulness of the semiparametric methods proposed in this paper.

References

Bang, H., Tsiatis, A. (2000). Estimating medical costs with censored data. *Biometrika*, *87*, 329–343.

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

Bassett, G., Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, *73*, 618–622.

Beran, R. (1981). Nonparametric regression with randomly censored survival data, Technical Report, University of California, Berkeley.

Breslow, N. (1972). Discussion of a paper by D. R. Cox. *Journal of the Royal Statistical Society*, *34*, 261–217.

Cai, Z., Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, *167*, 413–425.

Cai, Z., Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, *103*, 1596–1608.

Chauduri, P. (1991). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, *39*, 246–269.

Chauduri, P., Doksum, K., Samarov, A. (1997). On average derivative quantile regression. *Annals of Statistics*, *25*, 715–744.

Cox, D. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, *34*, 187–220.

Cox, D. (1975). Partial likelihood. *Biometrika*, *62*, 269–272.

De Backer, M., El Ghouch, A., van Keilegom, I. (2017). Semiparametric copula quantile regression for complete or censored data. *Electronic Journal of Statistics*, *11*, 1660–1698.

El Ghouch, A., van Keilegom, I. (2009). Local linear quantile regression with dependent censored data. *Statistica Sinica*, *19*, 1621–1640.

Fan, J., Gijbels, I. (1994). Censored regression: Local linear approximations and their applications. *Journal of the American Statistical Association*, *89*, 560–569.

Fan, J., Hu, T., Truong, Y. (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics*, *21*, 433–446.

Fan, J., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, *11*, 1031–1057.

He, X., Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica*, *10*, 129–140.

He, X., Shi, P. (1996). Bivariate tensor product b-splines in a partially linear regression. *Journal of Multivariate Analysis*, *58*, 162–181.

Horowitz, J., Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, *100*, 1238–1249.

Hunter, D., Lange, K. (2000). Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics*, *9*, 60–77.

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

Kalbfleisch, J., Prentice, R. (2002). *The statistical analysis of failure data*. New York: Wiley.
Koenker, R. (2005). *Quantile regression*. Cambridge: Cambridge University Press.
Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.
Koul, H., Susarla, V., Ryzin, J. V. (1981). Regression analysis of randomly right censored data. *Annals of Statistics*, *9*, 1276–1288.
Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory*, *19*, 1–31.
Leng, C., Tong, X. (2013). A quantile regression estimator for censored data. *Bernoulli*, *19*, 344–361.
Li, G., Datta, S. (2001). A bootstrap approach to nonparametric regression for right censored data. *Annals of the Institute of Statistical Mathematics*, *53*, 708–729.
Li, W., Patilea, V. (2017). A dimension reduction approach for conditional Kaplan–Meier estimators, *Test* forthcoming.
Lin, Y. (2000). Linear regression analysis of censored medical costs. *Biostatistics*, *1*, 35–47.
Peng, L., Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, *103*, 637–649.
Robins, J., Rotnitzky, A. (1992). *Recovery information and adjustment for dependent censoring using surrogate markers in AIDS epidemiology-methodological issues* (pp. 297–331). Boston: Birkhauser.
Satten, G., Datta, S. (2001). The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average. *American Statistician*, *55*, 207–210.
Su, J., Wei, L. (1991). A lack of fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association*, *86*, 420–426.
Van Keilegom, I., Akritas, M., Veraverbeke, N. (2001). Estimation of the conditional distribution in regression with censored data: A comparative study. *Computational Statistics and Data Analysis*, *35*, 487–501.
Wang, H., Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, *104*, 1117–1128.
Xie, S., Wan, A., Zhou, Y. (2015). Quantile regression methods with varying-coefficient models for censored data. *Computational Statistics and Data Analysis 88*, 154–172.
Ying, Z., Jung, S., Wei, L. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association*, *90*, 178–184.
Yu, K., Jones, M. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, *93*, 228–237.
Zhou, L. (2006). A simple censored median regression estimator. *Statistica Sinica*, *16*, 1043–1058.