Statistica Sinica **24** (2014), 855-870
doi:http://dx.doi.org/10.5705/ss.2011.195

CENSORED QUANTILE REGRESSION WITH VARYING COEFFICIENTS

Guosheng Yin, Donglin Zeng, and Hui Li

The University of Hong Kong, The University of North Carolina at Chapel Hill and Beijing Normal University

Abstract: We propose a varying-coefficient quantile regression model for survival data subject to random censoring. Motivated by the work of Yang (1999), quantile-based moments are constructed using covariate-weighted empirical cumulative hazard functions. We estimate regression parameters based on the generalized method of moments. The proposed estimators are shown to be consistent and asymptotically normal. We examine the proposed method with finite sample sizes through simulation studies, and illustrate it with a Richter's syndrome study.

Key words and phrases: Generalized method of moments, local polynomial, regression quantiles, semiparametric models, random censoring, survival data.

1. Introduction

The median survival time is often used as a summary statistic to characterize patient survival. In contrast to the mean survival time, the median is more robust to outliers or extreme observations. The linear regression, also known as the accelerated failure time model, is a mean-based regression approach to covariate analysis. It formulates a linear model between the logarithm of the failure time T and covariates $\mathbf{Z}$ in the form

$$\log(T) = \boldsymbol{\beta}_0^T \mathbf{Z} + \varepsilon. \tag{1.1}$$

If the distribution of the error ε is not specified, model (1.1) is semiparametric, for which estimation methods are typically based on the least squares or rank estimators (Buckley and James (1979); Tsiatis (1990); Wei, Ying, and Lin (1990); and Jin et al. (2003)). The mean-based regression model quantifies the central effects of covariates, but may not capture the full distributional impact of covariates with heterogeneous effects. By contrast, when a properly chosen set of quantiles is modeled simultaneously, we can obtain a global assessment of covariate effects (Koenker and Bassett (1978); Koenker (2005)). In quantile regression, the model parameters are estimated by minimizing a quantile-based objective function. The corresponding variances are typically estimated through resampling methods to avoid nonparametric functional estimation of the error's density function (Parzen, Wei, and Ying (1994); Horowitz (1998); Bilias, Chen and Ying (2000); Jin, Ying, and Wei (2001); and He and Hu (2002)). Yu, Lu, and Stander (2003) provided a general coverage of various applications of quantile regression.

Censored quantile regression, particularly the so-called Tobit model, has been investigated for fixed-censoring data (Powell (1984); Buchinsky and Hahn (1998)). For random censoring cases, Ying, Jung, and Wei (1995) proposed quantile regression for randomly censored failure time data under the assumption of independence between covariates and censoring. Lindgren (1997) studied generalized L_1 minimization under censored quantile regression. Yang (1999) derived an estimating equation approach to censored median regression based on the covariate-weighted cumulative hazard function. Portnoy (2003) relaxed the independence condition between covariates and censoring times for censored quantile regression by redistributing weights of censored data to the right. Peng and Huang (2008) developed an estimation method for censored quantile regression based on martingale properties and minimization of a sequence of L_1-type convex functions. Wang and Wang (2009) proposed redistributing the censored data to the right by using the local Kaplan-Meier estimator. Based on conditional moment inequalities, Khan and Tamer (2009) further relaxed model assumptions in quantile regression.

On the other hand, varying-coefficient models characterize the trends of covariate effects over time or some exposure variable (Hastie and Tibshirani (1993)), while limited research has been conducted in quantile regression with varying coefficients. Yu and Jones (1998) proposed nonparametric regression quantiles using kernel weighted local linear fitting. Honda (2004) studied the local L_1 estimation with varying coefficients through local polynomial expansions. Kim (2007) proposed spline-based quantile regression and a Rao-score model fit test in contrast to linear quantile regression. Cai and Xu (2008) investigated dynamic quantile estimation for time series data. Neocleous and Portnoy (2009) studied partially linear censored quantile regression with B-splines. Qian and Peng (2010) developed censored quantile regression by incorporating partially functional effects.

By extending the work of Yang (1999), we propose a varying-coefficient quantile regression method with randomly censored survival data. We take a local polynomial expansion (Fan and Gijbels (1996)), and incorporate a kernel function to the empirical cumulative hazard function. The quantile-based estimating equations can be viewed as moment conditions in the generalized method of moments (GMM) framework (Hansen (1982); Hansen, Heaton, and Yaron (1996)). In contrast to likelihood-based approaches, the moments of quantile regression can be constructed in a relatively straightforward way, and we can combine the available moments and minimize the GMM objective function to estimate regression quantiles.

The rest of this article is organized as follows. In Section 2, we propose the estimation procedure under the censored varying-coefficient quantile regression model. In Section 3, we establish the consistency and asymptotic normality of the parameter estimates. We examine the finite sample properties of the proposed method using simulation studies in Section 4, and illustrate it with a Richter's syndrome study in Section 5. We give concluding remarks in Section 6 and delineate the proofs of the theorems in the supplementary material on the journal's website.

2. Varying-coefficient Quantile Regression

Let T_i be the failure time, and let C_i be the censoring time for the ith subject, $i = 1, \ldots, n$. The associated covariates are denoted by a p-vector $\mathbf{Z}_i$ and a scalar W_i (an exposure variable that may interact with $\mathbf{Z}_i$ in a nonlinear way). We observe $X_i = \min(T_i, C_i)$, and the censoring indicator $\Delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function. We assume that T_i is conditionally independent of C_i given the covariates, and $(X_i, \Delta_i, \mathbf{Z}_i, W_i)$ are independent and identically distributed (i.i.d.).

Let $q_\tau(\cdot|\mathbf{Z}_i, W_i)$ be the conditional τ-quantile function given covariates $\mathbf{Z}_i$ and W_i, for $0 < \tau < 1$. To accommodate nonlinear interactions between covariates, we propose the varying-coefficient quantile regression model in the form

$$q_\tau(\log(T_i)|\mathbf{Z}_i, W_i) = \boldsymbol{\beta}_\tau(W_i)^T\mathbf{Z}_i, \tag{2.1}$$

where the first component of $\mathbf{Z}_i$ is 1 corresponding to the main effect of W_i. The error term is $\epsilon_i = \log(T_i) - \boldsymbol{\beta}_\tau(W_i)^T\mathbf{Z}_i$, so $q_\tau(\epsilon_i|\mathbf{Z}_i, W_i) = 0$. We assume independence between ϵ_i and $\mathbf{Z}_i$, while ϵ_i may depend on W_i; this allows for possibly heteroscedastic errors to some extent.

Let $\mathcal{W}$ denote the support of the exposure variable W. By Taylor's series expansion, for each chosen $w_0 \in \mathcal{W}$ we have that

$$\boldsymbol{\beta}_\tau(w) \approx \boldsymbol{\beta}_\tau(w_0) + \boldsymbol{\beta}_\tau^{[1]}(w_0)(w - w_0) + \cdots + \boldsymbol{\beta}_\tau^{[r]}(w_0)(w - w_0)^r,$$

where

$$\boldsymbol{\beta}_\tau^{[k]}(w_0) = \frac{1}{k!}\frac{\partial^k \boldsymbol{\beta}_\tau(w)}{\partial w^k}\Big|_{w_0}, \quad k = 1, \ldots, r.$$

We reparameterize with

$$\boldsymbol{\xi}_\tau(w_0) = \{\boldsymbol{\beta}_\tau(w_0)^T, \boldsymbol{\beta}_\tau^{[1]}(w_0)^T, \ldots, \boldsymbol{\beta}_\tau^{[r]}(w_0)^T\}^T,$$
$$\mathbf{Z}_i^* = \{\mathbf{Z}_i^T, \mathbf{Z}_i^T(W_i - w_0), \ldots, \mathbf{Z}_i^T(W_i - w_0)^r\}^T.$$

Let $K(\cdot)$ be a kernel density function, h_n be a bandwidth, and $K_{h_n}(\cdot) = K(\cdot/h_n)$. After the local polynomial expansion, we can compute the τ-quantile residuals

$$e_i(\boldsymbol{\xi}) = \log(X_i) - \boldsymbol{\xi}_\tau(w_0)^T \mathbf{Z}_i^*.$$

For notational brevity, we drop the dependence of $\boldsymbol{\xi}_\tau(w_0)$ on w_0 and τ whenever doing so causes no ambiguity. Using the jth covariate as a weight, the local empirical cumulative hazard function is

$$\widehat{\Lambda}_j(t, \boldsymbol{\xi}) = \sum_{e_i(\boldsymbol{\xi}) \le t} \frac{K_{h_n}(W_i - w_0) Z_{i,j}^* \Delta_i}{\sum_{k=1}^n K_{h_n}(W_k - w_0) Z_{k,j}^* I(e_k(\boldsymbol{\xi}) \ge e_i(\boldsymbol{\xi}))},$$

where $Z_{i,j}^*$ is the jth component of $\mathbf{Z}_i^*$, $j = 1, \ldots, (r+1)p$. The τ-quantile estimating equations can be constructed as

$$\widehat{\Lambda}_j(0, \boldsymbol{\xi}) = -\log(1-\tau), \quad j = 1, \ldots, (r+1)p.$$

This leads to $(r+1)p$ moment conditions that can be concatenated as

$$\mathbf{U}_n(\boldsymbol{\xi}) = n^{-1} \sum_{i=1}^n \mathbf{u}_i(\boldsymbol{\xi}),$$

where

$$\mathbf{u}_i(\boldsymbol{\xi}) = \begin{pmatrix} \dfrac{n K_{h_n}(W_i - w_0) Z_{i,1}^* \Delta_i I(e_i(\boldsymbol{\xi}) \le 0)}{\sum_{j=1}^n K_{h_n}(W_j - w_0) Z_{j,1}^* I(e_j(\boldsymbol{\xi}) \ge e_i(\boldsymbol{\xi}))} + \log(1-\tau) \\ \vdots \\ \dfrac{n K_{h_n}(W_i - w_0) Z_{i,(r+1)p}^* \Delta_i I(e_i(\boldsymbol{\xi}) \le 0)}{\sum_{j=1}^n K_{h_n}(W_j - w_0) Z_{j,(r+1)p}^* I(e_j(\boldsymbol{\xi}) \ge e_i(\boldsymbol{\xi}))} + \log(1-\tau) \end{pmatrix}.$$

In the GMM framework (Hansen (1982)), $\boldsymbol{\xi}_\tau(w_0)$ is estimated by minimizing the weighted quadratic function

$$Q_n(\boldsymbol{\xi}) = \mathbf{U}_n(\boldsymbol{\xi})^T \boldsymbol{\Omega}_n(\boldsymbol{\xi})^{-1} \mathbf{U}_n(\boldsymbol{\xi}), \tag{2.2}$$

where $\boldsymbol{\Omega}_n(\boldsymbol{\xi}) = n^{-1} \sum_{i=1}^n \mathbf{u}_i(\boldsymbol{\xi}) \mathbf{u}_i(\boldsymbol{\xi})^T - \mathbf{U}_n(\boldsymbol{\xi}) \mathbf{U}_n(\boldsymbol{\xi})^T$.

3. Asymptotic Theories

Let $\widehat{\boldsymbol{\xi}}(w_0)$ be the minimizer of $Q_n(\boldsymbol{\xi})$ in (2.2), and let $\boldsymbol{\xi}_0(w_0)$ denote the true parameter. The conditions needed for developing the asymptotic properties of $\widehat{\boldsymbol{\xi}}(w_0)$ are as follows:

(C1) $(\mathbf{Z}, W)$ has a bounded support.

(C2) The joint density of $(T, C, \mathbf{Z}, W)$ is twice-continuously differentiable and is bounded away from zero on its support.

(C3) $P(C \geq c_0|\mathbf{Z}, W) = P(C = c_0|\mathbf{Z}, W) > 0$, where c_0 is the study duration.

(C4) $K(\cdot)$ is a symmetric density function and has a finite $(2r+2)$-moment.

(C5) w_0 is an interior point of the support of W, $w_0 \in \mathcal{W}$.

(C6) $\boldsymbol{\beta}_0(w_0)$ is $(r+1)$-continuously differentiable at w_0 with $r \geq 0$.

(C7) $h_n \to 0$, and $nh_n \to \infty$.

(C8) $nh_n^{2r+3} = O(1)$.

Conditions (C1)−(C3) are standard assumptions in the context of survival analysis. The smoothness conditions for $\boldsymbol{\beta}_0(\cdot)$ and the kernel function in (C4)−(C6) allow us to carry out estimation using rth order local polynomials at w_0. The bandwidth condition (C7) controls the variability of the local polynomial estimator, while (C8) implies that the bias of the proposed estimator can be up to $\sqrt{nh_n^{2r+3}}$-order. From the last two conditions, one choice of h_n is $n^{-1/(2r+3)}$.

Theorem 1. *If* (C1)−(C7) *hold,* $\widehat{\boldsymbol{\xi}}(w_0)$ *is a uniformly consistent estimator of* $\boldsymbol{\xi}_0(w_0)$ *for* $w_0 \in \mathcal{W}$.

Set $\mu_k = \int u^k K(u)du$ and $\mathbf{H} = \{(\mu_{k+j})_{(r+1)\times(r+1)} \times \text{Diag}(1, h_n, \ldots, h_n^r)\} \otimes \mathbf{I}_{p\times p}$, where $(\mu_{k+j})_{(r+1)\times(r+1)}$ is a $(r+1)\times(r+1)$ dimensional matrix with elements μ_{k+j} $(k, j = 0, \ldots, r)$. Here, $\otimes$ is the Kronecker product, and $\mathbf{I}_{p\times p}$ is the p-dimensional identity matrix.

Theorem 2. *If* (C1)−(C8) *hold, for each interior point* $w_0 \in \mathcal{W}$,

$$\sqrt{nh_n}\left\{\mathbf{H}\big(\widehat{\boldsymbol{\xi}}(w_0) - \boldsymbol{\xi}_0(w_0)\big) - h_n^{r+1}\mathbf{D}^{-1}\mathbf{b}\right\} \xrightarrow{\mathcal{D}} N\big(\mathbf{0}, \boldsymbol{\Sigma}\big),$$

where $\mathbf{D}$, $\mathbf{b}$, *and* $\boldsymbol{\Sigma}$ *are given in the supplementary material.*

The proofs of the theorems are outlined in the supplementary material; they rely on empirical process theory (van der Vaart and Wellner (1996); Kosorok (2008)). As the variance of $\widehat{\boldsymbol{\xi}}(w_0)$ depends on the density of the error, we use the bootstrap procedure to estimate the standard errors of the parameter estimates so as to avoid nonparametric functional estimation.

The first p components of $\boldsymbol{\xi}(w_0)$ correspond to $\boldsymbol{\beta}(w_0)$. Take $\beta_j(w_0)$ to be the jth element of $\boldsymbol{\beta}(w_0)$. The optimal bandwidth can be obtained by minimizing the weighted mean squared error

$$\int \left\{\text{Bias}\big(\hat{\beta}_j(w)\big)^2 + \text{Var}\big(\hat{\beta}_j(w)\big)\right\}\Psi(w)dw,$$

where $\Psi(\cdot)$ is a nonnegative and integrable weight function. From Theorem 2, the weighted mean squared error is given by

$$\int \left\{ h_n^{2r+2} \phi_j^2(w) + \frac{1}{nh_n} \sigma_{jj}(w) \right\} \Psi(w) dw,$$

where $\phi_j(\cdot)$ is the jth element of the vector $\mathbf{D}^{-1}\mathbf{b}$ and $\sigma_{jj}(\cdot)$ is the jth diagonal element of $\boldsymbol{\Sigma}$. As a result, we can obtain an optimal h_n by minimizing the overall mean squared error

$$\sum_{j=1}^{p} \int \left\{ \mathrm{Bias}\big(\hat{\beta}_j(w)\big)^2 + \mathrm{Var}\big(\hat{\beta}_j(w)\big) \right\} \Psi(w) dw.$$

This leads to

$$h_n = n^{-1/(2r+3)} \left\{ \frac{\sum_{j=1}^{p} \int \sigma_{jj}(w) \Psi(w) dw}{(2r+2) \sum_{j=1}^{p} \int \phi_j^2(w) \Psi(w) dw} \right\}^{1/(2r+3)}.$$

In practice, we can use a K-fold cross-validation approach to selecting the bandwidth (Hoover et al. (1998)). We divide the data into K equal-sized subgroups, denoted by D_k, $k = 1, \ldots, K$. For the data excluding D_k, we fit the τ-quantile regression model to obtain the parameter estimates $\hat{\boldsymbol{\beta}}_{(-k)}(W_i)$. We then estimate the residual for each subject belonging to D_k, $\hat{e}_i = \log X_i - \widehat{\beta}_{(-k)}(W_i)^T \mathbf{Z}_i$ for $i \in D_k$, and construct the empirical cumulative hazard function based on $\hat{e}_i$,

$$\tilde{\Lambda}_k(0) = \sum_{\hat{e}_i \le 0, i \in D_k} \frac{\Delta_i}{\sum_{j \in D_k} I(\hat{e}_j \ge \hat{e}_i)}.$$

We find the optimal bandwidth by minimizing $\sum_{k=1}^{K} |\tilde{\Lambda}_k(0) + \log(1-\tau)|$.

4. Simulation Studies

4.1. Homogeneous error

To examine the finite sample property of the proposed method, we conducted extensive simulation studies. We first considered a model with homogeneous errors,

$$\log(T) = \beta_0(W) + \beta_1(W) Z_1 + \beta_2(W) Z_2 + \epsilon, \tag{4.1}$$

where $\beta_0(W) = 0.5$, $\beta_1(W) = W^2$, and $\beta_2(W) = \cos(3W)$. The covariates W, Z_1, and Z_2 were Unif$(-1, 1)$, Unif$(0, 1)$, and $N(0, 1)$, respectively; the error ϵ was $N(0, 1)$. Censoring times were independently generated from a uniform distribution to yield a censoring rate of 25%. We partitioned the range of W, $[-1, 1]$, into 20 equal intervals to evaluate the coefficient functions. As the GMM

Table 1. Estimation of the regression coefficients and the corresponding derivatives under model (4.1) with 25% censoring.

w_0	h_n	True	$\widehat{\beta}(w_0)$	SD	SE	CP(%)	True	$\widehat{\beta}'(w_0)$	SD	SE	CP(%)
		$\beta_0(w) = 0.5$					$\beta_0'(w) = 0$				
-0.5	0.06	0.5	0.592	0.464	0.520	96.8	0	-0.113	2.208	2.461	96.8
	0.08		0.547	0.391	0.454	97.2		-0.008	1.595	1.792	97.2
0	0.06	0.5	0.534	0.439	0.506	97.2	0	0.270	1.886	2.205	97.0
	0.08		0.521	0.389	0.432	96.0		0.037	1.403	1.578	96.6
0.5	0.06	0.5	0.523	0.498	0.538	96.6	0	0.414	1.945	2.323	97.4
	0.08		0.483	0.419	0.455	96.2		0.208	1.350	1.616	98.0
		$\beta_1(w) = w^2$					$\beta_1'(w) = 2w$				
-0.5	0.06	0.25	0.262	0.800	0.886	98.0	-1.0	-0.232	3.394	3.730	97.4
	0.08		0.334	0.706	0.780	97.6		-0.075	2.585	2.869	96.6
0	0.06	0	0.162	0.780	0.872	97.2	0	0.257	2.926	2.331	97.2
	0.08		0.132	0.683	0.754	97.2		0.475	2.288	2.481	96.2
0.5	0.06	0.25	0.434	0.934	0.913	97.8	1.0	0.748	3.088	3.431	96.4
	0.08		0.485	0.818	0.782	98.8		0.954	2.094	2.507	97.6
		$\beta_2(w) = \cos(3w)$					$\beta_2'(w) = -3\sin(3w)$				
-0.5	0.06	0.071	0.141	0.477	0.506	94.8	2.992	2.530	1.576	1.816	96.4
	0.08		0.210	0.515	0.464	95.4		2.465	1.012	1.234	96.2
0	0.06	1.000	0.917	0.380	0.460	97.0	0	0.353	1.545	1.826	98.2
	0.08		0.912	0.385	0.420	95.2		0.274	1.030	1.197	97.2
0.5	0.06	0.071	0.171	0.561	0.521	91.6	-2.992	-2.059	1.640	2.012	95.4
	0.08		0.130	0.474	0.485	95.6		-2.095	1.005	1.282	93.4

SD is the standard deviation, SE is the estimated standard error using the bootstrap method averaged over 500 simulations, CP(%) is the 95% confidence interval coverage probability.

objective function in (2.2) is complicated and highly nonlinear with respect to the parameters, we applied the Nelder and Mead (1965) simplex algorithm to minimize the quadratic function. We took a local linear expansion with $r = 1$, chose the Gaussian kernel function and explored the bandwidths $h_n = 0.06$ and 0.08. The sample size was $n = 200$, and we took 400 bootstrap samples for variance estimation. For each configuration, we replicated 500 simulations.

Table 1 summarizes the estimation results for $w_0 = -0.5$, 0, and 0.5. We present the average of the varying-coefficient estimates, the standard deviation (SD), the average of the estimated standard errors (SE) based on the bootstrap method, and the coverage probability of the 95% confidence interval (CP%). We also provide the estimates for the corresponding derivatives of the varying coefficients. One can see that the estimation bias is small, the bootstrap variance estimate provides a fairly good approximation to the variability of the estimators, and the coverage probability reasonably matches the nominal level. Under

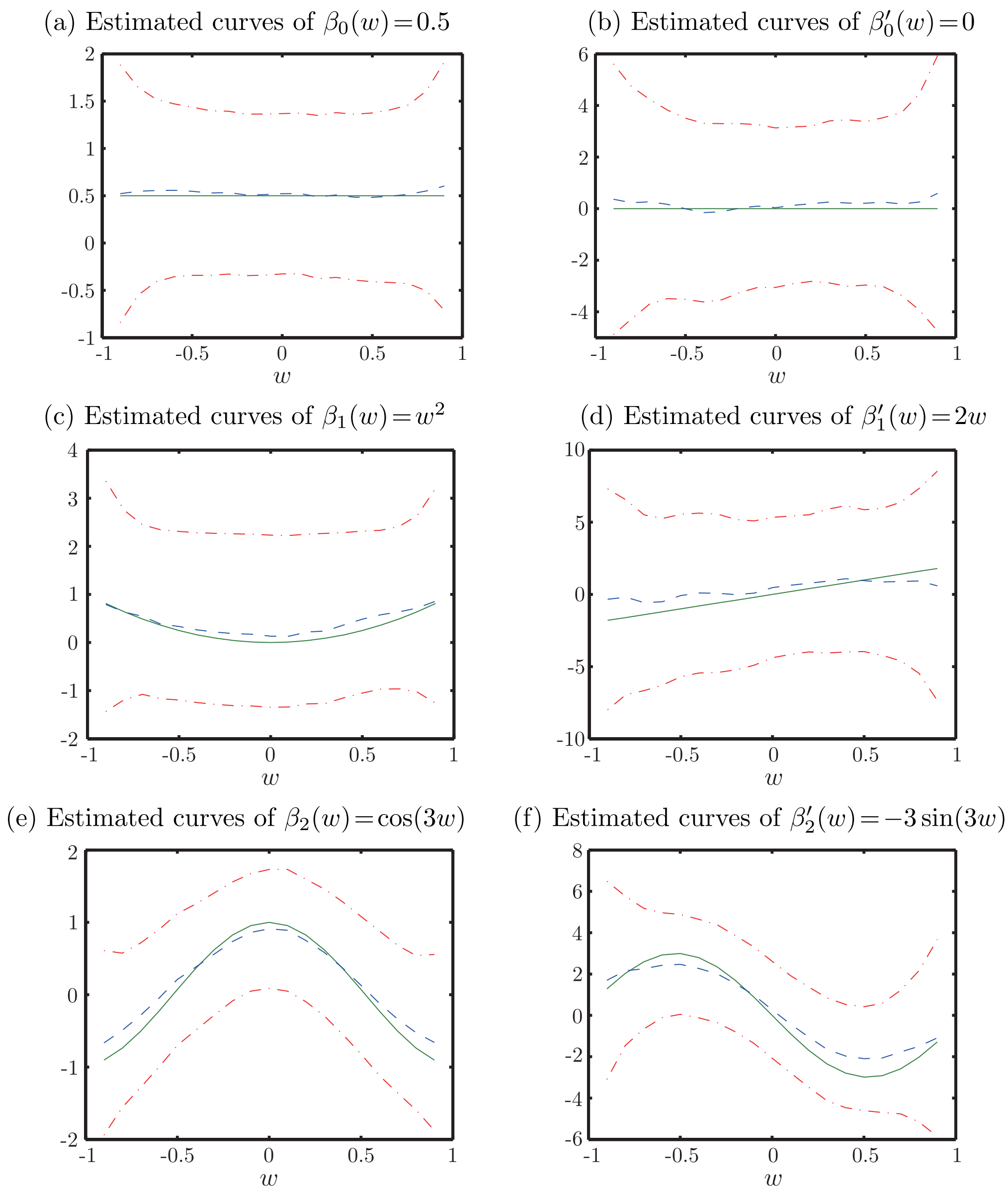

Figure 1. Estimation results under model (4.1) with $n = 200$, 25% censoring and $h_n = 0.08$. (a)$-$(f): solid lines are the true functions and dashed lines are the estimates of varying coefficients coupled with the pointwise 95% confidence intervals.

different bandwidths, the simulation results were similar, indicating to a certain extent the robustness of the method with respect to the bandwidth. Figure 1 shows both the estimated varying-coefficient functions and their derivatives with

Table 2. Estimation of the regression coefficients under model (4.2) with varying and constant coefficients.

			$\beta_0(w) = w^2$				$\beta_1(w) = \sin(3w)$				$\gamma = 0.5$			
n	$c\%$	w_0	Bias	SD	SE	CP(%)	Bias	SD	SE	CP(%)	Bias	SD	SE	CP(%)
200	0	-0.7	-0.024	0.182	0.226	97.0	0.128	0.268	0.333	95.4	0.004	0.089	0.093	96.2
		0	-0.060	0.161	0.180	94.0	-0.008	0.253	0.276	94.0				
		0.7	0.022	0.189	0.232	97.0	-0.100	0.307	0.314	94.4				
	20	-0.7	-0.019	0.193	0.254	98.4	0.127	0.307	0.368	96.6	0.023	0.100	0.103	91.6
		0	0.051	0.165	0.196	96.6	-0.011	0.262	0.305	97.6				
		0.7	0.050	0.239	0.285	97.2	-0.144	0.343	0.387	95.0				
	40	-0.7	0.020	0.252	0.299	97.8	0.115	0.399	0.440	96.0	0.034	0.113	0.126	96.0
		0	0.072	0.186	0.234	96.4	-0.025	0.284	0.347	97.4				
		0.7	0.121	0.318	0.405	96.6	-0.078	0.438	0.498	96.6				
400	0	-0.7	0.034	0.131	0.144	95.6	0.086	0.198	0.225	95.4	0.001	0.059	0.062	95.8
		0	0.035	0.123	0.137	94.8	0.009	0.182	0.213	97.4				
		0.7	0.021	0.130	0.147	96.4	-0.078	0.204	0.226	94.4				
	20	-0.7	0.034	0.141	0.162	97.2	0.082	0.223	0.252	95.6	-0.001	0.061	0.069	96.8
		0	0.039	0.131	0.147	95.4	0.008	0.202	0.231	97.4				
		0.7	0.033	0.152	0.191	98.2	-0.080	0.243	0.280	96.6				
	40	-0.7	0.026	0.157	0.197	98.2	0.087	0.248	0.296	95.4	0.010	0.082	0.087	96.2
		0	0.048	0.143	0.172	96.4	-0.001	0.231	0.265	96.8				
		0.7	0.134	0.342	0.353	93.4	-0.068	0.330	0.390	97.8				

SD is the standard deviation, SE is the estimated standard error using the bootstrap method averaged over 500 simulations, CP(%) is the 95% confidence interval coverage probability.

a bandwidth of $h_n = 0.08$ coupled with the 95% confidence intervals. The estimated curves match reasonably with the true functions, while the estimates for the derivatives of the varying coefficients are generally not as good as those for the functions themselves.

To acknowledge that some covariate effects are varying while others are constant, we took

$$\log(T) = \beta_0(W) + \beta_1(W)Z_1 + \gamma Z_2 + \epsilon, \tag{4.2}$$

where $\beta_0(W) = W^2$, $\beta_1(W) = \sin(3W)$, and $\gamma = 0.5$. The covariate W was Unif$(-1, 1)$, Z_1 was Unif$(0, 1)$, Z_2 was Bernoulli(0.5), and the error ϵ was $N(0, 0.25)$. We took the sample sizes $n = 200$ and 400, and the censoring rates $c\% = 0$, 20%, and 40%. A simple way to estimate the constant coefficient γ is to first estimate $\gamma(w_0)$, and then take an average over all the chosen w_0. Table 2 shows that the estimates are quite accurate in general when the model involves both varying and constant coefficients. The bias is small, the standard errors reasonably characterize the variation of the estimates, and the coverage probabilities are around 95%.

For comparison, we also implemented the B-spline method in Neocleous and Portnoy (2009), for which we chose three knots corresponding to the $25\%, 50\%$

Table 3. Comparison of the proposed method with the B-spline method of Neocleous and Portnoy (2009) in terms of mean squared errors ($\times 10^{-2}$) with a censoring rate of 20%, and $h_n = 0.06, 0.10$, and 0.14.

	MSE			MSE	
Kernel (h_n)	$\beta_0(w_0)$	$\beta_1(w_0)$	B-spline	$\beta_0(w_0)$	$\beta_1(w_0)$
0.06	3.806	3.615	Linear	3.714	10.862
0.10	9.380	3.151	Quadratic	4.541	14.199
0.14	22.875	4.873	Cubic	5.267	17.399

and 75% quantiles and explored piecewise linear, quadratic, and cubic spline terms. Table 3 shows that the mean squared errors are comparable between the two methods.

4.2. Heteroscedastic error

As quantile regression is known to be most suitable for heteroscedastic errors, we also examined heterogeneity induced by covariate-dependent errors. We simulated failure times from the model

$$\log(T) = \beta_0(W) + \beta_1(W)Z + \epsilon, \tag{4.3}$$

where $\beta_0(W) = 2W$ and $\beta_1(W) = \sin(3W)$. We generated W as $N(0, 1)$, Z as Bernoulli(0.5), and the error ϵ as $W \times N(0, 1)$ for $Z = 1$ and as $W \times N(0, 0.7)$ for $Z = 0$. Censoring times were independently uniform and yielded an approximate censoring rate of 20%. We partitioned the range of W into 18 equal intervals to locally evaluate the coefficients at each chosen w_0. We took a local linear expansion with $r = 1$, chose the Gaussian kernel function and considered the bandwidths $h_n = 0.16, 0.18$, and 0.20. We took the sample size $n = 200$ and, for each configuration, we replicated 500 simulations. To handle heteroscedastic errors, we stratified the estimating equation by Z, and minimized $Q_n(\boldsymbol{\xi}|Z = 0)$ and $Q_n(\boldsymbol{\xi}|Z = 1)$, respectively. From the simulation results summarized in Table 4, one can see that the biases of the estimates are small, the standard errors are close to the standard deviations, and the coverage probabilities are reasonable. For $h_n = 0.18$, we exhibit the estimated curves for $\tau = 0.5$ in Figure 2, and those for $\tau = 0.25$ in Figure 3. Under (4.3), the computing time using I5-2430M CPU (2.40GHz and 8GB RAM) with FORTRAN PowerStation 4.0 was approximately 1 minute per simulation.

5. Example

We applied the proposed model to a data set from a leukemia study conducted at M. D. Anderson Cancer Center (Tsimberidou et al. (2006)). The study involved 130 evaluable patients, diagnosed with Richter's syndrome via

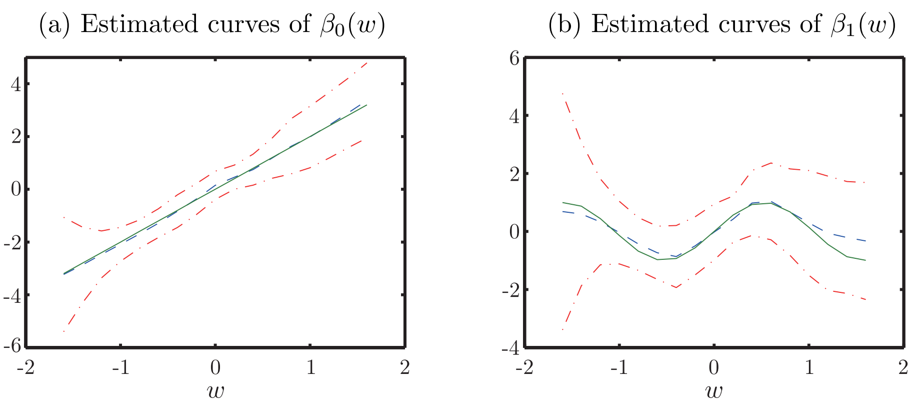

Figure 2. Estimation results under model (4.3) with $n = 200$, $\tau = 0.5$, 20% censoring and $h_n = 0.18$. (a)−(b): solid lines are the true functions and dashed lines are the estimates of the varying coefficients coupled with the pointwise 95% confidence intervals.

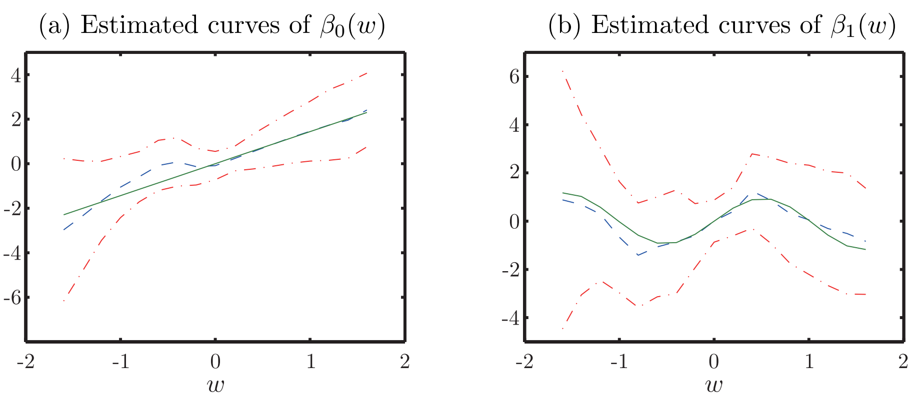

Figure 3. Estimation results under model (4.3) with $n = 200$, $\tau = 0.25$, 20% censoring and $h_n = 0.18$. (a)−(b): solid lines are the true functions and dashed lines are the estimates of the varying coefficients coupled with the pointwise 95% confidence intervals.

biopsy or fine-needle aspiration, who had been examined between January 1975 and June 2005. Richter's syndrome (RS) is a rare and aggressive type of acute adult leukemia that often results from a transformation of chronic lymphocytic leukemia into diffuse large cell lymphoma; it is usually fatal within a short period of time. In this study, patients were treated by chemoimmunotherapy with rituximab or chemotherapy alone. Figure 4 exhibits the Kaplan–Meier survival curves for the 130 patients with RS. We can see a sharp change-point in the survival curve around two years of follow-up, which would typically cause the violation of the usual proportional hazards assumption (Cox (1972)). The censoring rate of the data was approximately 12%.

Table 4. Estimation of the regression coefficients under model (4.3) with a censoring rate of 20%, and $h_n = 0.16, 0.18$, and 0.2.

		$\beta_0(w) = 2w$					$\beta_1(w) = \sin(3w)$				
w_0	h	True	$\widehat{\beta}_0(w_0)$	SD	SE	CP(%)	True	$\widehat{\beta}_1(w_0)$	SD	SE	CP(%)
-1.2	0.16	-2.4	-2.459	0.471	0.453	96.6	0.443	0.307	0.484	0.714	98.6
	0.18		-2.463	0.339	0.479	96.8		0.284	0.528	0.744	99.4
	0.20		-2.517	0.464	0.483	95.4		0.285	0.498	0.792	98.4
-0.8	0.16	-1.6	-1.660	0.402	0.259	97.0	-0.675	-0.487	0.361	0.500	97.8
	0.18		-1.682	0.287	0.283	95.2		-0.462	0.367	0.519	97.2
	0.20		-1.700	0.428	0.314	93.6		-0.419	0.460	0.555	96.4
-0.4	0.16	-0.8	-0.811	0.315	0.303	97.6	-0.932	-0.854	0.405	0.485	98.2
	0.18		-0.811	0.411	0.319	97.2		-0.885	0.527	0.551	98.4
	0.20		-0.833	0.423	0.327	96.6		-0.900	0.476	0.568	97.8
0	0.16	0	0.180	0.501	0.278	96.0	0	-0.058	0.722	0.473	99.2
	0.18		0.184	0.532	0.265	94.4		0.023	0.672	0.471	97.0
	0.20		0.158	0.554	0.240	95.0		-0.012	0.639	0.454	97.8
0.4	0.16	0.8	0.779	0.323	0.302	98.0	0.932	0.937	0.506	0.557	98.2
	0.18		0.775	0.336	0.317	97.2		0.996	0.639	0.606	98.4
	0.20		0.755	0.339	0.298	97.2		1.016	0.632	0.613	97.6
0.8	0.16	1.6	1.612	0.504	0.467	97.6	0.675	0.776	0.767	0.771	97.8
	0.18		1.646	0.734	0.505	95.8		0.700	0.667	0.774	98.4
	0.20		1.643	0.766	0.625	97.0		0.711	0.656	0.805	97.4
1.2	0.16	2.4	2.502	0.612	0.674	96.2	-0.443	-0.016	0.804	0.956	92.2
	0.18		2.438	0.492	0.637	96.8		-0.017	0.936	0.988	93.8
	0.20		2.409	0.448	0.644	97.4		-0.057	0.795	0.972	93.4

SD is the standard deviation, SE is the estimated standard error using the bootstrap method averaged over 500 simulations, CP(%) is the 95% confidence interval coverage probability.

In the varying-coefficient quantile regression model, we included three covariates: treatment (1 if using chemotherapy alone, and 0 if using chemoimmunotherapy with rituximab), age (ranging from 29 to 77 years with a median of 60 years), and sex (1 if male, and 0 if female). We were interested in characterizing the nonlinear interactions between patient age and other covariates and how they affected the quantiles of patient survival times. With $\tau = 0.25$, 0.5, and 0.75, we fit the proposed varying-coefficient quantile regression model

$$q_\tau(\log(T)|\mathbf{Z}, W) = \beta_0(W) + \beta_1(W)Z_{\text{treatment}} + \beta_2(W)Z_{\text{sex}},$$

where W is the logarithm of patient age. In this analysis, we divided the data into five groups of 26 observations each. Based on the five-fold cross-validation procedure described earlier, the bandwidth $h_n = 0.19$ appeared to be a reasonable choice. We partitioned the range of W into 20 equal intervals. The estimated coefficient functions and the corresponding 95% confidence intervals

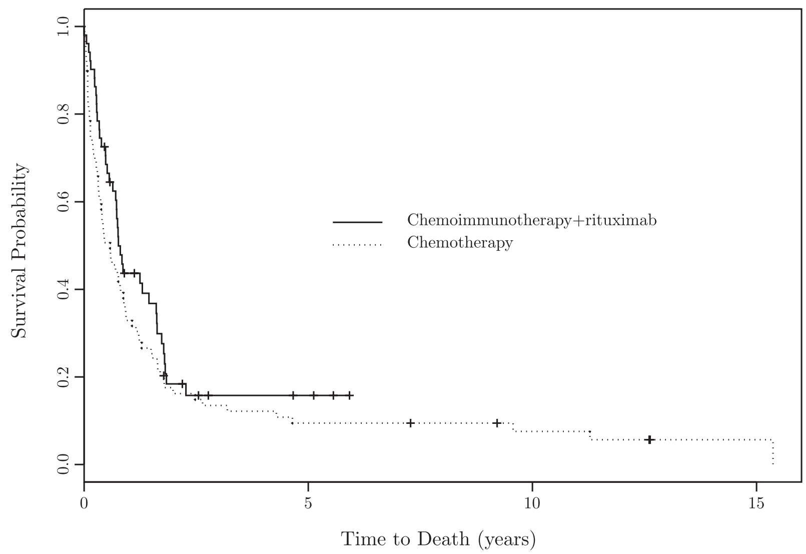

Figure 4. Estimated Kaplan–Meier survival curves stratified by treatment for the Richter's syndrome data.

are given in Figure 5. In the median regression, there was a trend that treatment with chemoimmunotherapy and rituximab improved patient survival while such difference diminished as patient age increased. This trend appeared to be similar for the other two conditional quantiles as well. With regard to the covariate effect of patient sex, we did not find any quantile difference between male and female patients for $\tau = 0.5$. For $\tau = 0.25$, it appeared that younger male patients had better survival than younger female patients, while older male patients had worse survival than older female patients. For $\tau = 0.75$, male patients seemed to have better survival than female regardless of their ages. Generally, all these findings are not statistically significant, they only exhibit some trends in patients' survival with respect to different covariates.

6. Discussion

We have proposed censored quantile regression with varying coefficients by adopting the estimation method of Yang (1999) due to the simplicity of constructing the covariate-weighted cumulative hazard function. By taking the local polynomial expansion of varying-coefficient functions, we construct the kernel-based moments such that the estimation procedure can be naturally integrated with the GMM. Although the GMM involves complicated nonlinear estimating equations, our numerical studies have shown that the estimators perform reasonably well with finite sample sizes. The proposed method allows the error to depend on the exposure variable W, and thus can handle heteroscedastic errors to

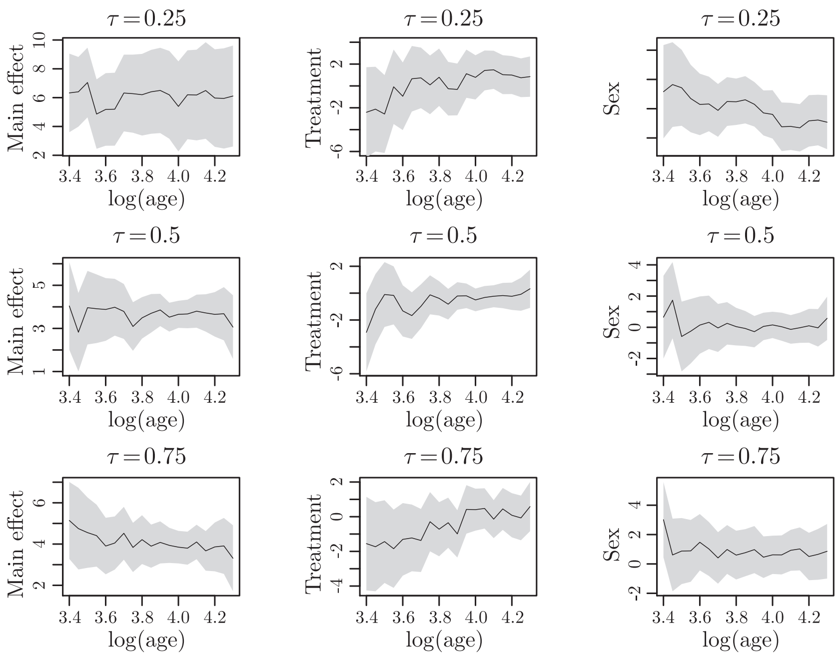

Figure 5. Analysis of the Richter's syndrome data based on the proposed varying-coefficient quantile regression. The curve is the estimated varying coefficients and the shaded area is the pointwise 95% confidence interval.

some extent. We may also allow the error to depend on covariate $\mathbf{Z}$ by stratification for discrete $\mathbf{Z}$, or by local kernel estimation for continuous $\mathbf{Z}$. Nevertheless, if the dimension of $\mathbf{Z}$ is high, estimation can be challenging. The partially linear quantile regression model in Neocleous and Portnoy (2009) can accommodate general heteroscedasticty, while to carry out median regression analysis, all the regression quantiles below the median must be calculated first.

In censored quantile regression, it is known that the estimation of upper quantiles may not be stable due to identifiability issues. In general, τ should be smaller than $\inf_{\mathbf{Z},w_0} P(T \leq c_0|\mathbf{Z}=\mathbf{z}, W=w_0)$, where c_0 is the study end time, so that we would have data to estimate the τ-quantile. For varying-coefficient models, it is important to determine whether covariate effects are varying or constant over the exposure variable W. Toward this goal, some model goodness-of-fit procedures might be considered (He and Zhu (2003)), and automatic discovery procedures (Zhang, Cheng, and Liu (2011)) also warrant further research.

Acknowledgement

We would like to thank the two anonymous referees, an associate editor, and Editor Qiwei Yao for many insightful suggestions that strengthened the work immensely. We also thank Professor Don Ylvisaker for his enormous editorial help. This work is partially supported by a grant (Grant No. 784010) from the Research Grants Council of Hong Kong (G. Yin), and a grant (Grant No. 11201031) from National Natural Science Foundation of China (H. Li).

References

Bilias, Y., Chen, S. and Ying, Z. (2000). Simple resampling methods for censored regression quantiles. *J. Econometrics* **99**, 373-386.

Buchinsky, M. and Hahn, J. Y. (1998). An alternative estimator for censored quantile regression. *Econometrica* **66**, 653-671.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Amer. Statist. Assoc.*, 1595-1608.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall, London.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029-1054.

Hansen, L. P., Heaton, J. and Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *J. Bus. Econom. Statist.* **14**, 262-280.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55**, 757-796.

He, X. and Hu, F. (2002). Markov chain marginal bootstrap. *J. Amer. Statist. Assoc.* **97**, 783-795.

He, X. and Zhu, L.-X. (2003). A lack-of-fit test for quantile regression. *J. Amer. Statist. Assoc.* **98**, 1013-1022.

Honda, T. (2004). Quantile regression in varying coefficient models. *J. Statist. Plann. Inference* **121**, 113-125.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809-822.

Horowitz, J. (1998). Bootstrap methods for median regression model. *Econometrica* **66**, 1327-1352.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341-353.

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

Khan, S. and Tamer, E. (2009). Inference on endogenously censored regression models using conditional moment inequalities. *J. Econometrics* **152**, 104-119.

Kim, M-O. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35**, 92-108.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press.

Koenker, R. and Bassett, G. J. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

Lindgren, A. (1997). Quantile regression with censored data using generalized L_1 minimization. *Comput. Statist. Data Anal.* **23**, 509-524.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer J.* **7**, 308-313.

Neocleous, T. and Portnoy, S. (2009). Partially linear censored quantile regression. *Lifetime Data Analysis* **15**, 357-378.

Parzen, M. I., Wei, L. J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341-350.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.* **103**, 637-649.

Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98**, 1001-1012.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression. *J. Econometrics* **25**, 303-325.

Qian, J. and Peng, L. (2010). Censored quantile regression with partially functional effects. *Biometrika* **97**, 839-850.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354-372.

Tsimberidou, A. M., O'Brien, S., Khouri, I., Giles, et al. F. G., Kantarjian, H. M., Champlin, R., Wen, S., Do, K-A, Smith, S. C., Lerner, S., Freireich, E. J. and Keating, M. J. (2006). Clinical outcomes and prognostic factors in patients with Richter's syndrome treated with chemotherapy or chemoimmunotherapy with or without stem-cell transplantation. *J. Clinical Oncology* **24**, 2343-2351.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104**, 1117-1128.

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845-851.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard function. *J. Amer. Statist. Assoc.* **94**, 137-145.

Ying, Z., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90**, 178-184.

Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: Applications and current research areas. *J. Roy. Statist. Soc. Ser. D* **52**, 331-350.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228-237.

Zhang, H. H., Cheng, G. and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *J. Amer. Statist. Assoc.* **106**, 1099-1112.

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong.
E-mail: gyin@hku.hk

Department of Biostatistics, Gillings School of Global Public Health, CB#7420, The University of North Carolina at Chapel Hill, 27599-7420.
E-mail: dzeng@bios.unc.edu

School of Mathematical Sciences, Beijing Normal University, No. 19, XinJieKouWai St., HaiDian District,Beijing 100875, P. R. China.
E-mail: li_hui@bnu.edu.cn

(Received August 2011; accepted April 2013)