Power-Transformed Linear Quantile Regression With Censored Data

Guosheng YIN, Donglin ZENG, and Hui LI

We propose a class of power-transformed linear quantile regression models for survival data subject to random censoring. The estimation procedure follows two sequential steps. First, for a given transformation parameter, we can easily obtain the estimates for the regression coefficients by minimizing a well-defined convex objective function. Second, we can estimate the transformation parameter based on a model discrepancy measure by constructing cumulative sum processes. We show that both the regression and transformation parameter estimates are strongly consistent and asymptotically normal. The variance–covariance matrix depends on the unknown density function of the error term, so we estimate the variance by the usual bootstrap approach. We examine the performance of the proposed method for finite sample sizes through simulation studies and illustrate it with a real data example.

KEY WORDS: Asymptotic normality; Box–Cox transformation; Empirical estimation; Median regression; Random censoring; Survival data; Transformation model.

1. INTRODUCTION

Linear regression models have been extensively studied for randomly censored survival data (e.g., see Prentice 1978; Buckley and James 1979; Ritov 1990; Tsiatis 1990; Wei, Ying, and Lin 1990; Lai and Ying 1991; Jin, Lin, Wei, and Ying 2003). In particular, the accelerated failure time (AFT) model is intuitively attractive by formulating a linear model between the logarithm of the failure time and covariates $\mathbf{Z}$,

$$\log(T) = \boldsymbol{\beta}_0^T \mathbf{Z} + \varepsilon, \tag{1}$$

where the error term ε typically has mean 0, but its distribution is unknown and unspecified. Extensive research has been carried out on the semiparametric least squares or rank estimation of model (1). The interpretation of $\boldsymbol{\beta}_0$ is straightforward, which explicitly characterizes prolonging or shortening the subject survival time. The usual AFT model is a mean-based regression model that may be viewed as a special case of the quantile regression model. The mean-AFT model in (1) gives mainly an overall quantification of patient survival but cannot characterize the local effects of covariates for lower or higher quantiles of the survival time. In the biomedical literature, the median survival time (which is known to be robust to outliers) often is used as a summary statistic. Likewise, in contrast to the mean-based model, quantile regression can give a more complete assessment of covariate effects at a properly chosen set of quantiles (e.g., Koenker and Bassett 1978; Portnoy and Koenker 1997; Yu, Lu, and Stander 2003; Koenker 2005). The regression parameters often are estimated by solving quantile-based estimating equations through linear programming or interior point methods, and the corresponding variances typically depend on the density function of the error terms. To avoid nonparametric functional density estimation, which may not be stable for small sample sizes, various resampling methods have been proposed for variance estimation (see, e.g., Parzen, Wei, and Ying 1994; Buchinsky 1995; Hahn 1995; Horowitz 1998; Bilias, Chen, and Ying 2000; Jin, Ying, and Wei 2001). In economics, a special type of censored data often arise, under which the maximum of zero and a latent response is observed. Quantile regression has been extensively investigated for such censored data (e.g., Powell 1986; Buchinsky and Hahn 1998; Khan and Powell 2001). Interest has been growing in quantile regression with randomly censored failure time data, and such models have become increasingly important in survival analysis (Ying, Jung, and Wei 1995; Lindgren 1997; Yang 1999; Koenker and Geling 2001; Honoré, Khan, and Powell 2002; Portnoy 2003). When the focus is on an earlier or later stage of the follow-up, conventional methods, such as the proportional hazards model (Cox 1972) or the AFT model, may not be very useful. Quantile regression, in contrast, can directly model the lower or higher quantile of interest to provide a natural assessment of covariate effects specific for those quantiles. This distinctive feature makes quantile regression very attractive for randomly censored survival data, because the failure times often are quite right-skewed, and the extreme survival times and the long-term treatment effect may be of the special interest.

For randomly censored survival data, Ying et al. (1995) proposed a novel median regression model under which the minimum dispersion test statistic was used for inferences because the variances of the parameter estimates depend on the unknown density function of the error. Yang (1999) studied median regression by constructing an empirically covariate-weighted cumulative hazard function. Honoré et al. (2002) adapted quantile regression estimators for the fixed censoring case to the random censoring scenario by considering the censoring value as missing when the latent dependent variable is uncensored. A key assumptions of the approaches of Ying et al. (1995) and Honoré et al. (2002) is the independence of the covariates and censoring, whereas that of Yang (1999) requires that the covariates and errors be independent. For medical cost data, Bang and Tsiatis (2002) studied a median regression model based on the inverse probability-weighted estimating equation. Portnoy (2003) proposed censored regression quantiles that allow censoring times to depend on covariates. Furthermore, Portnoy (2003) introduced an efficient computing algorithm based on the redistribute-to-the-right form of the Kaplan–Meier estimator (i.e., the recursively reweighted empirical survival function). Khan and Tamer (2007) studied a partial rank estimator (up to a scale of the true parameter)

Guosheng Yin is Associate Professor, Department of Biostatistics, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030 (E-mail: *gsyin@mdanderson.org*). Donglin Zeng is Associate Professor, Department of Biostatistics, The University of North Carolina, Chapel Hill, NC 27599 (E-mail: *dzeng@bios.unc.edu*). Hui Li is Assistant Professor, School of Mathematical Sciences, Beijing Normal University, Beijing, China. The authors thank the editor, associate editor, and two referees for their insightful comments, which led to substantial improvements in the article.

Journal of the American Statistical Association
September 2008, Vol. 103, No. 483, Theory and Methods
DOI 10.1198/016214508000000490

that accommodates covariate-dependent censoring but assumes homoscedastic errors. More recently, Khan and Tamer (2008) proposed a general method with weaker assumptions based on conditional moment inequalities that allows for conditional heteroscedasticity, covariate dependent censoring, and endogenous censoring.

Transformation models are flexible and typically include a broad class of modeling structures (e.g., Cheng, Wei, and Ying 1995; Cai, Tian, and Wei 2005; Khan and Tamer 2007). The Box–Cox transformation (Box and Cox 1964) is often used to improve the error normality in linear models. Particularly, for the failure time T, we define $H_\gamma(T) = (T^\gamma - 1)/\gamma$ for $\gamma > 0$ and $H_\gamma(T) = \log(T)$ for $\gamma = 0$. Transformation quantile regression has recently been proposed for complete data without any censoring, which may have great potential in various applications (Mu and He 2007). But when data are subject to random censoring, the theoretical and computational developments for quantile regression become much more involved and challenging. In this article we propose a power-transformed linear quantile regression model for randomly censored survival data. By imposing a monotonic transformation on the failure time, the equivariance property of quantiles is preserved. We establish the consistency and asymptotic normality of the transformation and regression parameter estimates. We derive the explicit form for the variance–covariance matrix of the estimators, and use a bootstrap resampling method to estimate the variance, to circumvent the nonparametric functional density estimation. We explore the cases with known transformation parameter γ and unknown γ that correspond to the conditional and unconditional regression. The variances for the regression coefficients are greatly inflated by the variation from estimating γ when we conduct a unconditional inference compared with conditional inference with a fixed γ.

The rest of the article is organized as follows. In Section 2 we propose the estimation procedure under the censored transformation quantile regression model. In Section 3 we establish the strong consistency and asymptotic normality of the parameter estimates. In Section 4 we examine the finite-sample properties using simulation studies, and in Section 5 we illustrate the proposed method with application to a lung cancer data set. We give concluding remarks in Section 6 and delineate the proofs of our theorems in the Appendix.

2. TRANSFORMED QUANTILE REGRESSION

For $i = 1, \dots, n$, we let T_i be the failure time for the ith subject and C_i be the censoring time, and we observe $X_i = \min(T_i, C_i)$ and the failure time indicator $\Delta_i = I(T_i \le C_i)$, where $I(\cdot)$ is the indicator function. We let $\mathbf{Z}_i$ be the corresponding $p \times 1$ vector of bounded covariates and assume that C_i is independent of T_i and $\mathbf{Z}_i$. The assumption of completely random censorship that we use here has been used in almost all of the quantile regression literature for censored data, whereas in a different regression setting from ours, Portnoy (2003) presented a very innovative way of relaxing this assumption to missing at random. For $i = 1, \dots, n$, $(X_i, \Delta_i, \mathbf{Z}_i)$ are independent and identically distributed (iid).

Let $H_\gamma(T_i)$ be the failure time under a monotonic transformation and let $\xi_\tau(\cdot|\mathbf{Z}_i)$ be the 100τth conditional quantile function for $0 < \tau < 1$. To enhance modeling flexibility, we propose the censored transformation linear quantile regression model in the form of

$$\xi_\tau(H_\gamma(T_i)|\mathbf{Z}_i) = \boldsymbol{\beta}_\tau^T \mathbf{Z}_i, \tag{2}$$

where $\xi_\tau(\varepsilon_{\tau i}^{(\gamma)}|\mathbf{Z}_i) = 0$ with $\varepsilon_{\tau i}^{(\gamma)} = H_\gamma(T_i) - \boldsymbol{\beta}_\tau^T \mathbf{Z}_i$. The conditional distribution of $\varepsilon_{\tau i}^{(\gamma)}$ is unspecified and may depend on $\mathbf{Z}_i$. If there is no censoring, then $\boldsymbol{\beta}_\tau$ can be estimated by minimizing $n^{-1}\sum_{i=1}^n \phi_\tau(H_\gamma(T_i) - \boldsymbol{\beta}_\tau^T \mathbf{Z}_i)$, where the "check function" is defined as $\phi_\tau(u) = u\{\tau - I(u < 0)\}$ (see Mu and He 2007). Moreover, the transformation parameter γ can be estimated by constructing a cusum process of residuals that performs better than other two-stage estimators (e.g., Chamberlain 1994; Buchinsky 1995).

Under random censorship, we observe that $\Pr(X_i \ge H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T \mathbf{Z}_i)) = \tau G(H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T \mathbf{Z}_i))$, where $G(\cdot)$ is the survival function for the censoring time C_i and we drop the dependence on τ for notational brevity. For a fixed γ, the estimator for $\boldsymbol{\beta}$ can be obtained by solving the following estimating equation:

$$n^{-1}\sum_{i=1}^n \mathbf{Z}_i \left\{ \frac{I(H_\gamma(X_i) - \boldsymbol{\beta}(\gamma)^T \mathbf{Z}_i \ge 0)}{\widehat{G}(H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T \mathbf{Z}_i))} - \tau \right\} = 0, \tag{3}$$

where $\widehat{G}(\cdot)$ is the Kaplan–Meier estimator for the censoring times based on $\{(X_i, 1 - \Delta_i), i = 1, \dots, n\}$. In practice, if $\widehat{G}(H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T \mathbf{Z}_i)) = 0$, then we set $I(H_\gamma(X_i) - \boldsymbol{\beta}(\gamma)^T \mathbf{Z}_i \ge 0)/\widehat{G}(H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T \mathbf{Z}_i)) = 0$, as in the work of Ying et al. (1995). A limitation of (3) is that the censoring times must be independent of the covariates. To relax this assumption, a stratified Kaplan–Meier estimator can be constructed by categorizing the covariate values into groups and thus replacing $\widehat{G}(\cdot)$ by $\widehat{G}(\cdot|\mathbf{Z})$ in (3), which, however, may suffer from the high dimensionality of covariates. In practice, we can apply the log-rank test to examine the censoring time distribution across each covariate and use only $\widehat{G}(\cdot|\mathbf{Z})$ stratified by those statistically significant covariates.

Because of the discontinuity of the estimating function in (3), its solution may not exist. Instead, we can minimize the Euclidean norm of the estimating function; however, this is discontinuous and has no derivatives. Even if we implement the Nelder–Mead simplex algorithm, which does not require any derivatives or continuity of the target function, a unique minimum still is not guaranteed. To overcome the numerical difficulties, we use the following two-step procedure, which minimizes a convex function at each stage. For a fixed γ, we first obtain $\widehat{\boldsymbol{\alpha}}(\gamma)$ by minimizing

$$\begin{aligned}\Phi_{n0}(\boldsymbol{\alpha};\gamma) &= n^{-1}\sum_{i=1}^n \frac{\Delta_i}{\widehat{G}(X_i)}\big\{(H_\gamma(X_i) - \boldsymbol{\alpha}^T\mathbf{Z}_i) \\ &\quad \times I(H_\gamma(X_i) - \boldsymbol{\alpha}^T\mathbf{Z}_i \ge 0) - \tau(H_\gamma(X_i) - \boldsymbol{\alpha}^T\mathbf{Z}_i)\big\}.\end{aligned}$$

This is based on the inverse probability weighted estimating equation (see, e.g., Robins and Rotnitzky 1992; Robins 1996).

We take $\widehat{\boldsymbol{\alpha}}(\gamma)$ as an initial value and then obtain $\widehat{\boldsymbol{\beta}}(\gamma)$ by minimizing

$$\Phi_{n1}(\boldsymbol{\beta};\gamma) = n^{-1}\sum_{i=1}^{n}\left\{\frac{(H_\gamma(X_i)-\boldsymbol{\beta}^T\mathbf{Z}_i)I(H_\gamma(X_i)-\boldsymbol{\beta}^T\mathbf{Z}_i\geq 0)}{\widehat{G}(H_\gamma^{-1}(\widehat{\boldsymbol{\alpha}}(\gamma)^T\mathbf{Z}_i))} - \tau(H_\gamma(X_i)-\boldsymbol{\beta}^T\mathbf{Z}_i)\right\},$$

where we substitute $\widehat{\boldsymbol{\alpha}}(\gamma)$ for $\boldsymbol{\beta}(\gamma)$ in the denominator function $\widehat{G}(\cdot)$. We show that both $\widehat{\boldsymbol{\alpha}}(\gamma)$ and $\widehat{\boldsymbol{\beta}}(\gamma)$ are consistent estimators, but the numerical performance of $\widehat{\boldsymbol{\beta}}(\gamma)$ is typically better than that of $\widehat{\boldsymbol{\alpha}}(\gamma)$. Furthermore, because $\Phi_{n0}(\boldsymbol{\alpha};\gamma)$ and $\Phi_{n1}(\boldsymbol{\beta};\gamma)$ are both convex functions, the minimizers are unique.

To estimate γ, we define a discrepancy measure based on the cusum process that can distinguish the correct transformation from those misspecified ones. The cusum process based on residuals or score-type functions is very useful in modeling goodness-of-fit tests (e.g., He and Zhu 2003; Mu and He 2007). Let $\widehat{\gamma}$ be the estimator minimizing $R_n(\gamma)=\sum_{i=1}^n D_n(\mathbf{Z}_i,\gamma)^2$, where

$$D_n(\mathbf{z},\gamma)=\sum_{i=1}^n I(\mathbf{Z}_i\leq\mathbf{z})\left\{\frac{I(H_\gamma(X_i)-\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}_i\geq 0)}{\widehat{G}(H_\gamma^{-1}(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}_i))}-\tau\right\}$$

and $I(\mathbf{Z}_i\leq\mathbf{z})=I(Z_{i1}\leq z_1,\ldots,Z_{ip}\leq z_p)$. In the case with high-dimensional covariates, especially many categorical variables, minimization of $R_n(\gamma)$ may not be numerically stable. An alternative cusum process is given by

$$D_n^*(t,\gamma)=\sum_{i=1}^n I(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}_i\leq t) \times\left\{\frac{I(H_\gamma(X_i)-\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}_i\geq 0)}{\widehat{G}(H_\gamma^{-1}(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}_i))}-\tau\right\}.$$

Intuitively, if the transformation is correctly specified, then $D_n^*(t,\gamma)$ asymptotically converges to a mean-zero Gaussian process. Therefore, the transformation parameter γ can be consistently estimated by minimizing

$$R_n^*(\gamma)=\sum_{i=1}^n\int_0^\infty D_n^*(t,\gamma)^2\,dN_i(t),$$

where $N_i(t)=I(X_i\leq t)$ regardless of failure or censoring observations. Because both $R_n(\gamma)$ and $R_n^*(\gamma)$ are functions of a single unknown parameter γ, the standard grid search algorithm can be used.

3. LARGE–SAMPLE PROPERTIES

Let L be the end time of a study, and let $(\boldsymbol{\beta}_0,\gamma_0)$ be the true parameters. Throughout the derivations that follow, we assume that the following conditions hold:

(C.1) γ_0 belongs to a compact set Γ.

(C.2) With probability 1, $\mathbf{Z}$ is bounded, and if $H_\gamma^{-1}(\boldsymbol{\beta}^T\mathbf{Z})=H_{\gamma_0}^{-1}(\boldsymbol{\beta}_0^T\mathbf{Z})$, then $\boldsymbol{\beta}=\boldsymbol{\beta}_0$ and $\gamma=\gamma_0$.

(C.3) There exists a constant $\delta>0$ such that $\Pr(C\geq L)>\delta$. Moreover, the conditional density of T given $\mathbf{Z}$ is continuous and positive in its support, and the density of C is continuous in $[0,L)$.

(C.4) The 100τth quantile of $H_{\gamma_0}(T)$ given $\mathbf{Z}$ is unique with probability 1 and is strictly less than L.

(C.5) The transformation $H_\gamma(\cdot)$ is strictly increasing and twice-continuously differentiable in a neighborhood of γ_0.

Conditions (C.1) and (C.3) are standard in the context of survival analysis. Condition (C.2) guarantees the identifiability of the transformation and regression parameters. In particular, when H_γ is the Box–Cox transformation and $\mathbf{Z}$ contains one continuous covariate with a nonzero effect, this condition can be replaced by the linear independence of $\mathbf{Z}$. This is because both sides of the equality $H_\gamma^{-1}(\boldsymbol{\beta}^T\mathbf{Z})=H_{\gamma_0}^{-1}(\boldsymbol{\beta}_0^T\mathbf{Z})$ are analytic in the continuous covariate, so examining the behavior of the covariate at infinity gives $\gamma=\gamma_0$; thus (C.2) is implied by the condition that if $\boldsymbol{\beta}^T\mathbf{Z}=\boldsymbol{\beta}_0^T\mathbf{Z}$, then $\boldsymbol{\beta}=\boldsymbol{\beta}_0$. Condition (C.4) is needed because otherwise, the 100τth quantile would not be estimable from the data. Condition (C.5) ensures the unique parameterization of the transformation.

To facilitate the theoretical development, we introduce some necessary notations. We define $\boldsymbol{\alpha}(\gamma)$ as the minimizer of

$$\Phi_0(\boldsymbol{\alpha};\gamma)=E\{(H_\gamma(T)-\boldsymbol{\alpha}^T\mathbf{Z})I(H_\gamma(T)-\boldsymbol{\alpha}^T\mathbf{Z}\geq 0) - \tau(H_\gamma(T)-\boldsymbol{\alpha}^T\mathbf{Z})\}$$

and define $\boldsymbol{\beta}(\gamma)$ as the minimizer of

$$\Phi_1(\boldsymbol{\beta};\gamma)=E\left\{\frac{(H_\gamma(X)-\boldsymbol{\beta}^T\mathbf{Z})I(H_\gamma(X)-\boldsymbol{\beta}^T\mathbf{Z}\geq 0)}{G_0(H_\gamma^{-1}(\boldsymbol{\alpha}(\gamma)^T\mathbf{Z}))} - \tau(H_\gamma(X)-\boldsymbol{\beta}^T\mathbf{Z})\right\},$$

where $G_0(x)=\Pr(C>x)$ is the true survival function for the censoring distribution. Clearly, Φ_0 and Φ_1 are the limiting functions of Φ_{n0} and Φ_{n1}, and both are strictly convex. In addition, we define

$$\Psi_0(\mathbf{Z};\boldsymbol{\alpha},\gamma)=E\{I(H_\gamma(T)-\boldsymbol{\alpha}^T\mathbf{Z}\geq 0)-\tau|\mathbf{Z}\}$$

and

$$\Psi_1(\mathbf{Z};\boldsymbol{\alpha},\boldsymbol{\beta},\gamma)=E\left\{\frac{I(H_\gamma(X)-\boldsymbol{\beta}^T\mathbf{Z}\geq 0)}{G_0(H_\gamma^{-1}(\boldsymbol{\alpha}^T\mathbf{Z}))}-\tau\middle|\mathbf{Z}\right\}.$$

Deriving the gradients $\nabla_{\boldsymbol{\alpha}}\Phi_0(\boldsymbol{\alpha};\gamma)=-E\{\mathbf{Z}\Psi_0(\mathbf{Z};\boldsymbol{\alpha},\gamma)\}$ and $\nabla_{\boldsymbol{\beta}}\Phi_1(\boldsymbol{\beta};\gamma)=-E\{\mathbf{Z}\Psi_1(\mathbf{Z};\boldsymbol{\alpha},\boldsymbol{\beta},\gamma)\}$ is straightforward. We use $\Psi_{0\gamma}$, $\Psi_{0\boldsymbol{\alpha}}$, and $\Psi_{0\boldsymbol{\beta}}$ to denote the row-vector gradients of Ψ_0 with respect to γ, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ evaluated at the true parameters $(\gamma_0,\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$. We define $(\Psi_{1\gamma},\Psi_{1\boldsymbol{\alpha}},\Psi_{1\boldsymbol{\beta}})$ as the gradients of Ψ_1 in the same manner. Note that $E(\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z}))=E(\mathbf{Z}^T\mathbf{Z}f(\boldsymbol{\beta}_0^T\mathbf{Z};\gamma_0))$, where $f(\cdot;\gamma_0)$ denotes the conditional density of $H_{\gamma_0}(T)$ given $\mathbf{Z}$. In condition (C.2), if we let $\gamma=\gamma_0$, then this implies that $\mathbf{Z}$ is linearly independent. This, combined with condition (C.3), implies that $E(\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z}))$ is positive definite. Similarly, $E(\mathbf{Z}\Psi_{0\boldsymbol{\beta}}(\mathbf{Z}))$ also is positive definite.

We first state a useful lemma and sketch its proof.

Lemma 1. $\boldsymbol{\alpha}(\gamma_0) = \boldsymbol{\beta}(\gamma_0) = \boldsymbol{\beta}_0$, and in a neighborhood of γ_0, it holds that

$$\boldsymbol{\alpha}(\gamma) = \boldsymbol{\beta}_0 - \left\{E(\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z}))\right\}^{-1} E(\mathbf{Z}\Psi_{0\gamma}(\mathbf{Z}))(\gamma - \gamma_0) + o(|\gamma - \gamma_0|)$$

and

$$\boldsymbol{\beta}(\gamma) = \boldsymbol{\beta}_0 - \left\{E(\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z}))\right\}^{-1}\left\{E(\mathbf{Z}\Psi_{1\gamma}(\mathbf{Z})) - E(\mathbf{Z}\Psi_{1\boldsymbol{\alpha}}(\mathbf{Z}))\left\{E(\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z}))\right\}^{-1} E(\mathbf{Z}\Psi_{0\gamma}(\mathbf{Z}))\right\} \times (\gamma - \gamma_0) + o(|\gamma - \gamma_0|).$$

Proof. Clearly, $\boldsymbol{\beta}_0$ satisfies both $\nabla_{\boldsymbol{\beta}}\Phi_0(\boldsymbol{\alpha}; \gamma_0) = 0$ and $\nabla_{\boldsymbol{\beta}}\Phi_1(\boldsymbol{\beta}; \gamma_0) = 0$. From the strict convexity of $\Phi_0(\boldsymbol{\alpha}; \gamma_0)$ and $\Phi_1(\boldsymbol{\beta}; \gamma_0)$, the first part of the conclusion should hold. By direct calculation and the positive density condition in (C.3), both $\Phi_0(\boldsymbol{\alpha}; \gamma_0)$ and $\Phi_1(\boldsymbol{\beta}; \gamma_0)$ have invertible Hessian matrixes at $\boldsymbol{\beta}_0$. The second part follows from the inverse mapping theorem.

In what follows, we lay out the asymptotic properties of the estimators.

Theorem 1. Under conditions (C.1)–(C.5), with probability 1,

$$|\widehat{\gamma} - \gamma_0| + \|\widehat{\boldsymbol{\alpha}}(\widehat{\gamma}) - \boldsymbol{\beta}_0\| + \|\widehat{\boldsymbol{\beta}}(\widehat{\gamma}) - \boldsymbol{\beta}_0\| \to 0.$$

Both the transformation and regression parameter estimates are strongly consistent. Because the initial value obtained from the inverse probability weighted equation, $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$, is consistent, this ensures that $\widehat{\boldsymbol{\beta}}(\widehat{\gamma})$ has well-behaved asymptotic properties.

Theorem 2. Under conditions (C.1)–(C.5),

$$\sqrt{n}\begin{pmatrix} \widehat{\gamma} - \gamma_0 \\ \widehat{\boldsymbol{\beta}}(\widehat{\gamma}) - \boldsymbol{\beta}_0 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathrm{N}(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is given at the end of the Appendix.

The proofs depend heavily on the modern empirical process theories (van der Vaart and Wellner 2000), which are outlined briefly in the Appendix.

4. SIMULATIONS

We conducted simulation studies to examine the finite-sample properties of the proposed methods. We considered the Box–Cox transformation linear quantile regression model

$$H_\gamma(T) = \frac{T^\gamma - 1}{\gamma} = \beta_0 + \beta_1 Z + \varepsilon, \tag{4}$$

where we took the true values of $\beta_0 = .2$, $\beta_1 = 1$, and $\gamma = 0, .5$ and 1. The covariate Z was generated from a uniform distribution, Unif[0, 2]. The error, ε, was simulated from a normal distribution with mean 0 and variance .25, N(0, .25). Censoring times were generated independently from uniform distributions to yield an approximate censoring rate of 20% or 40%. We took $\tau = .5$ to examine the median regression model, along with the sample size $n = 300$. For each configuration, we replicated 500 simulations. To obtain the standard errors (SEs) of the parameter estimates, we used the bootstrap method with 400 resampled data sets (Efron and Tibshirani 1993; Koenker 2005).

Table 1 summarizes the estimation results when γ is taken as an unknown parameter. We present the average of the parameter estimates over 500 simulations ($\widehat{\gamma}, \widehat{\beta}_0, \widehat{\beta}_1$), the sample standard deviation of the estimates (SD), the average of the estimated

Table 1. Estimation under the transformation quantile regression model with an unknown γ

		$\gamma = (0, .5, 1)$				$\beta_0 = .2$				$\beta_1 = 1$			
γ	c%	$\widehat{\gamma}$	SD	SE	CP%	$\widehat{\beta}_0$	SD	SE	CP%	$\widehat{\beta}_1$	SD	SE	CP%
Minimizing $R_n(\gamma)$													
0	0	.010	.261	.301	96.8	.182	.114	.121	95.4	1.057	.328	.399	97.0
	20	.027	.315	.359	96.4	.181	.130	.138	96.6	1.092	.409	.506	97.6
	40	.183	.417	.478	95.8	.148	.158	.188	97.4	1.295	.601	.793	98.6
.5	0	.503	.440	.470	95.6	.181	.119	.123	93.4	1.078	.427	.483	96.6
	20	.528	.513	.566	96.2	.177	.128	.144	93.6	1.123	.497	.615	97.2
	40	.560	.621	.673	95.8	.155	.151	.171	97.2	1.202	.640	.768	96.6
1	0	.979	.555	.620	96.2	.193	.116	.121	93.6	1.057	.432	.525	96.2
	20	.975	.692	.713	93.6	.185	.132	.141	94.2	1.103	.583	.640	94.8
	40	.982	.873	.816	91.4	.158	.167	.160	93.4	1.189	.751	.740	94.0
Minimizing $R_n^*(\gamma)$													
0	0	−.014	.278	.315	96.6	.193	.111	.117	93.4	1.028	.331	.393	95.4
	20	.024	.337	.379	96.6	.185	.125	.135	95.6	1.094	.427	.500	97.4
	40	.152	.401	.498	97.4	.158	.160	.166	95.6	1.249	.634	.699	98.8
.5	0	.468	.418	.478	98.0	.192	.115	.118	94.0	1.032	.386	.464	96.8
	20	.535	.530	.568	95.6	.172	.127	.137	93.6	1.137	.503	.589	96.2
	40	.519	.625	.691	94.8	.171	.141	.161	96.0	1.152	.586	.741	96.4
1	0	.937	.571	.625	95.8	.200	.110	.116	94.2	1.031	.442	.509	95.8
	20	1.001	.680	.715	95.0	.184	.127	.135	93.8	1.116	.543	.625	95.4
	40	.989	.836	.819	94.2	.176	.144	.149	94.6	1.158	.665	.707	95.4

Table 2. Estimation under the transformation quantile regression model when fixing $\gamma = 0$, .5, and 1

		$\beta_0 = .2$				$\beta_1 = 1$			
γ	$c\%$	$\widehat{\beta}_0$	SD	SE	CP%	$\widehat{\beta}_1$	SD	SE	CP%
0	0	.200	.073	.077	94.6	1.000	.062	.066	95.6
	20	.200	.080	.084	94.0	.997	.071	.076	95.0
	40	.229	.087	.090	93.8	.948	.078	.081	91.6
.5	0	.200	.073	.077	94.6	1.000	.062	.066	95.6
	20	.199	.081	.088	94.8	1.000	.073	.079	95.2
	40	.203	.094	.098	95.6	.997	.089	.092	95.4
1	0	.202	.069	.075	95.0	1.001	.060	.065	96.2
	20	.204	.080	.086	95.2	1.000	.074	.077	95.0
	40	.205	.095	.098	94.4	.998	.089	.092	95.2

SEs based on the bootstrap method, and the coverage probability of the 95% confidence intervals (CP%). We can see that the biases of the parameter estimates are small; the estimated SEs based on the bootstrap resampling method are reasonably close to the empirical SDs, and that the CP% generally match the nominal level. As the censoring percentage ($c\%$) increases, the biases and the variances of the estimators clearly increase, whereas the CP% is still maintained at around 95%. The numerical performance from minimizing $R_n^*(\gamma)$ is slightly better than that based on $R_n(\gamma)$ in terms of the estimation bias and variance.

To compare the unconditional and conditional inferences under model (4), we carried out estimations conditioning on the fixed γ under the same setups. Table 2 shows that the estimates of β_0 and β_1 are much more stable when γ is fixed at the true values of 0, .5, and 1. The biases of the parameter estimates are negligible and increase as the $c\%$ increases. The SE based on the bootstrap method provides a reasonable approximation to the empirical SD, and the corresponding CP% closely matches the nominal level. The variances of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ decrease dramatically compared with those in Table 1. We thus conclude

Table 3. Comparison of the joint MSE when using $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ or $\widehat{\boldsymbol{\beta}}(\widehat{\gamma})$ as the final estimator

Final	Minimizing $R_n(\gamma)$		Minimizing $R_n^*(\gamma)$	
estimator	20%	40%	20%	40%
$\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$	2.658	5.397	2.546	4.498
$\widehat{\boldsymbol{\beta}}(\widehat{\gamma})$	2.459	3.963	2.513	3.943

that taking γ as an additional unknown parameter highly inflates the estimation variability for β_0 and β_1.

Note that $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ itself also is a consistent estimator of $\boldsymbol{\beta}_0$, whereas the main advantage of using $\widehat{\boldsymbol{\beta}}(\widehat{\gamma})$ as the final estimator is to improve the estimation efficiency. To examine the gain in efficiency, we computed the joint mean squared errors (MSEs) for $(\widehat{\gamma}, \widehat{\beta}_0, \widehat{\beta}_1)$, when taking $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ or $\widehat{\boldsymbol{\beta}}(\widehat{\gamma})$ as the final estimator. Table 3 presents the joint MSEs based on minimizing either $R_n(\gamma)$ or $R_n^*(\gamma)$ under censoring rates of 20% and 40%. When there is no censoring [i.e., $\widehat{G}(\cdot) = 1$], the results based on $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ and $\widehat{\boldsymbol{\beta}}(\widehat{\gamma})$ are exactly the same. We can see a trend toward improved efficiency using $\widehat{\boldsymbol{\beta}}(\widehat{\gamma})$ in contrast to $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ as the final estimator, and the gain in efficiency is more prominent when the censoring percentage is high.

One major attraction of quantile regression is that it can handle the skewed or heteroscedastic error distribution. Following this route, we considered model (4) with the skewed error. We took ε from a shifted chi-squared distribution with 1 degree of freedom and a median of 0, and also examined model (4) with a heteroscedastic error of εZ, where $\varepsilon \sim \mathrm{N}(0, .25)$ and $Z \sim \mathrm{Unif}[0, 2]$. The true parameters were $\gamma = .5$, $\beta_0 = -.5$, and $\beta_1 = 1$, whereas other model setups remained the same as before. As shown in Table 4, the proposed method can produce satisfactory estimation results with small biases, reasonable variance estimates, and 95% coverage probabilities. Moreover, we conducted the corresponding simulation studies with skewed or heteroscedastic errors while fixing $\gamma = .5$. Table 5 shows that the conditional estimation is more stable and accurate, and that the variance estimates are greatly reduced in these fixed γ scenarios.

Table 4. Estimation under the transformation quantile regression model with an unknown γ and skewed or heteroscedastic errors

		$\gamma = .5$				$\beta_0 = -.5$				$\beta_1 = 1$			
Error	$c\%$	$\widehat{\gamma}$	SD	SE	CP%	$\widehat{\beta}_0$	SD	SE	CP%	$\widehat{\beta}_1$	SD	SE	CP%
Minimizing $R_n(\gamma)$													
Skewed	0	.402	.554	.598	95.5	−.478	.155	.149	93.7	.984	.243	.262	94.9
	20	.422	.614	.652	95.1	−.475	.149	.154	93.9	.993	.257	.282	96.9
	40	.508	.704	.748	97.1	−.469	.158	.159	93.9	1.009	.299	.319	97.4
Heteroscedastic	0	.520	.272	.268	95.9	−.498	.023	.022	94.5	1.000	.057	.066	96.5
	20	.526	.305	.326	96.5	−.499	.026	.026	95.1	1.002	.066	.081	97.1
	40	.660	.361	.434	97.2	−.488	.028	.033	95.9	.998	.080	.112	98.2
Minimizing $R_n^*(\gamma)$													
Skewed	0	.413	.539	.600	96.5	−.474	.156	.155	92.9	.982	.240	.260	95.3
	20	.382	.630	.652	94.5	−.463	.159	.160	93.1	.973	.268	.280	94.3
	40	.453	.711	.750	96.3	−.459	.159	.163	94.5	.979	.285	.306	96.5
Heteroscedastic	0	.524	.251	.253	97.4	−.497	.023	.022	94.5	.998	.052	.060	97.4
	20	.539	.299	.301	96.7	−.497	.027	.026	94.1	1.001	.060	.071	97.1
	40	.629	.311	.382	97.6	−.487	.027	.030	92.0	.990	.073	.088	96.6

Table 5. Estimation under the transformation quantile regression model with a fixed $\gamma = .5$ and skewed or heteroscedastic errors

		$\beta_0 = -.5$				$\beta_1 = 1$			
Error	c%	$\widehat{\beta}_0$	SD	SE	CP%	$\widehat{\beta}_1$	SD	SE	CP%
Skewed	0	−.487	.127	.128	92.4	.994	.113	.111	94.6
	20	−.488	.138	.136	92.0	.993	.124	.122	93.0
	40	−.486	.140	.104	94.2	.967	.124	.125	93.8
Hetero-	0	−.501	.016	.019	98.4	1.000	.050	.054	95.8
scedastic	20	−.499	.019	.023	98.8	.998	.059	.064	95.8
	40	−.492	.024	.028	98.2	.967	.068	.073	92.0

5. EXAMPLE

As an illustration, we applied our model to a data set involving patients with non–small-cell lung cancer (NSCLC), the leading cause of cancer-related mortality in the United States. To determine an optimum duration of chemotherapy in the treatment of advanced NSCLC, a multicenter phase III clinical trial was initiated in 1998 to investigate four cycles of therapy versus continuous therapy (Socinski et al. 2002). In this trial, patients were randomized to one of two arms: four cycles of carboplatin and paclitaxel every 21 days (arm A) or continuous treatment with carboplatin/paclitaxel until progression (arm B). In this analysis, 223 patients were diagnosed with NSCLC, of whom 111 were randomized to arm A and 112 to arm B. The primary endpoint was survival, and the censoring rate was 31% caused by loss to follow-up. Figure 1 shows the Kaplan–Meier survival curves for patients in the two treatment arms. There seems to be no survival difference between the two arms; of particular note is the presence of several crossings of the two curves. The covariates included patient sex (0, male; 1, female), the logarithm of age at entry, and treatment status (0, arm A; 1, arm B). In this population, 63% of the patients were male, and the age at entry ranged from 32 to 82 years, a mean of 62 years. To examine the assumption of covariate-dependent censoring, we fit the Cox proportional hazards model to the censoring times with respect to covariates and found that all three covariate effects were not significant. Moreover, we conducted

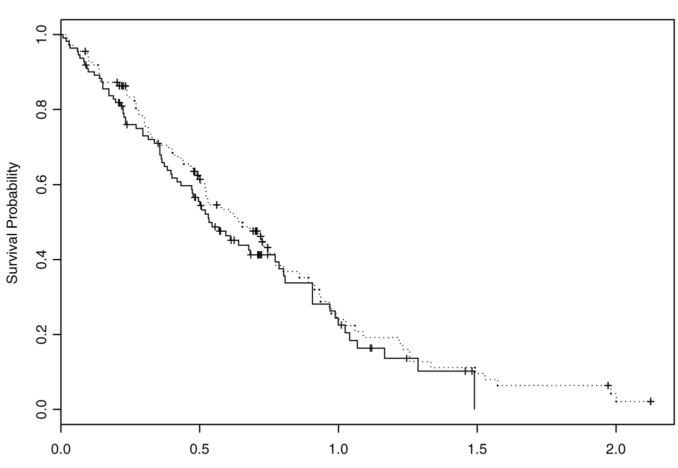

Figure 1. Estimated Kaplan–Meier survival curves for the NSCLC data (—, arm A: four cycles of treatment; ⋯⋯, arm B: continuous treatment).

the log-rank test for the censoring times by each covariate (with patient age stratified into several groups by different cutoffs). We found no censoring time difference across each covariate, providing empirical support for a common censoring distribution.

We applied the Box–Cox–transformed quantile regression model to the NSCLC data, taking the transformation parameter γ as unknown. We estimated γ by minimizing $R_n^*(\gamma)$ and took 400 bootstrap samples for the variance estimation. We specified the value of τ from .1 up to .8 in steps of .05 to cover the entire follow-up period. Because of the transformation on the failure time, the covariate effects obtained from each quantile regression have different scales. The marginal covariate effects in the original scale of the outcome are more useful (Koenker and Geling 2001; Mu and He 2007). Evaluated at a set of covariates $\mathbf{z}_0$, if the jth covariate is continuous, then its marginal effect is given by

$$\frac{\partial \xi_\tau(T|\mathbf{Z})}{\partial Z_j}\bigg|_{\mathbf{z}_0} = \begin{cases} \beta_{\tau,j}(\gamma_\tau \boldsymbol{\beta}_\tau^T \mathbf{z}_0 + 1)^{1/\gamma_\tau - 1}, & \gamma_\tau \neq 0 \\ \beta_{\tau,j} \exp(\boldsymbol{\beta}_\tau^T \mathbf{z}_0), & \gamma_\tau = 0. \end{cases}$$

For a discrete covariate, the marginal effect is $\xi_\tau(T|\mathbf{z}_{0(-j)}, Z_j = 1) - \xi_\tau(T|\mathbf{z}_{0(-j)}, Z_j = 0)$, where $\mathbf{z}_{0(-j)}$ is the rest of $\mathbf{z}_0$ without Z_j. We took $\mathbf{z}_0$ as specific covariates from a 50-year-old male patient in arm B. Figure 2 presents the estimates and the pointwise 95% confidence bands for the marginal covariate effects and the transformation parameter γ. Female patients survived significantly longer than males at the earlier follow-up, but this difference gradually decreased over time. Patient age did not appear to significantly affect survival. The patients in arm B seemed to have a slightly better rate of survival, as also shown in Figure 1, in which the survival curve of arm B lies close to that of arm A; however, there was no significant difference in survival between patients in these two arms. The estimate of the Box–Cox transformation parameter γ oscillates around the horizontal zero axis but tends to take a negative value for most of the regression quantiles. To compare our model with Portnoy's method, we first transformed the observations $H_{\widehat{\gamma}_\tau}(X_i)$ based on our estimator $\widehat{\gamma}_\tau$ for each τ, and then applied the censored regression quantiles (the R package "crq") developed by Portnoy (2003), the estimates of which are indicated by the dotted lines in Figure 2. We can see that the marginal effects of treatment are quite close between the two quantile regression models; however, the marginal effects of patient sex are quite different at the lower quantiles but are basically matched at the higher quantiles. The marginal effects of patient age are similar at most of the quantiles except those few above the 70th quantile.

For conditional inferences when fixing $\gamma = 0$, Figure 3 presents the estimation results based on our method and Portnoy's method. For the covariate effects of sex and age, the parameter estimates at most of quantiles using Portnoy's method lie within the pointwise 95% confidence band of our method, except for the first few early quantiles ($\tau < .2$). For the treatment effect, Portnoy's estimates are completely covered by our 95% confidence band. The results show that female patients survived significantly longer than males up to the 70th quantile, whereas the significant survival difference eventually disappeared at the end of the follow-up period due to a larger variance. Evaluating

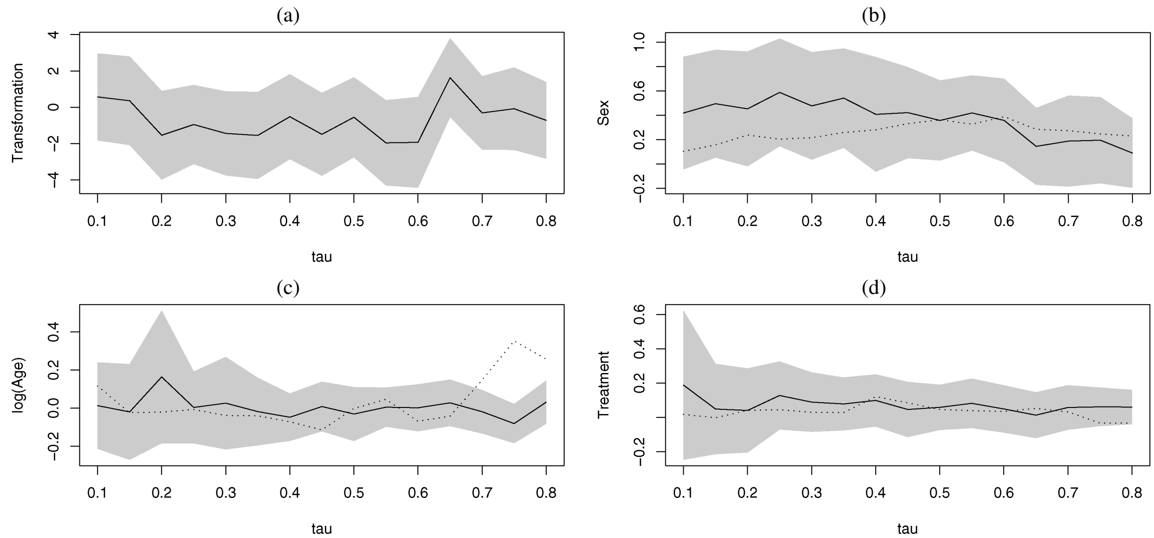

Figure 2. Marginal covariate effects for a 50-year-old male patient in arm B based on the quantile regression analysis of the NSCLC data with a parametric Box–Cox transformation (—, our proposed method; ······, Portnoy's method). (a) Transformation; (b) sex; (c) log(age); (d) treatment.

the covariates of age and treatment, we found no significant effects on survival for all of the regression quantiles. The overall trend is for younger patients and patients in arm B to have better survival rates. Moreover, we fit the usual Cox proportional hazards model to the same data; the estimates are summarized in Table 6. The Cox model gives an overall mean-based assessment of the covariate effects on survival for the entire follow-up period, which cannot distinguish the survival difference occurring mainly at the earlier or later stage of the trial. We can see that male patients had a significantly higher risk of death than females; however, treatment and patient age did not significantly affect the hazard. These results are generally consistent with those given by our quantile regression models. The quantile regression offers substantially more information, however,

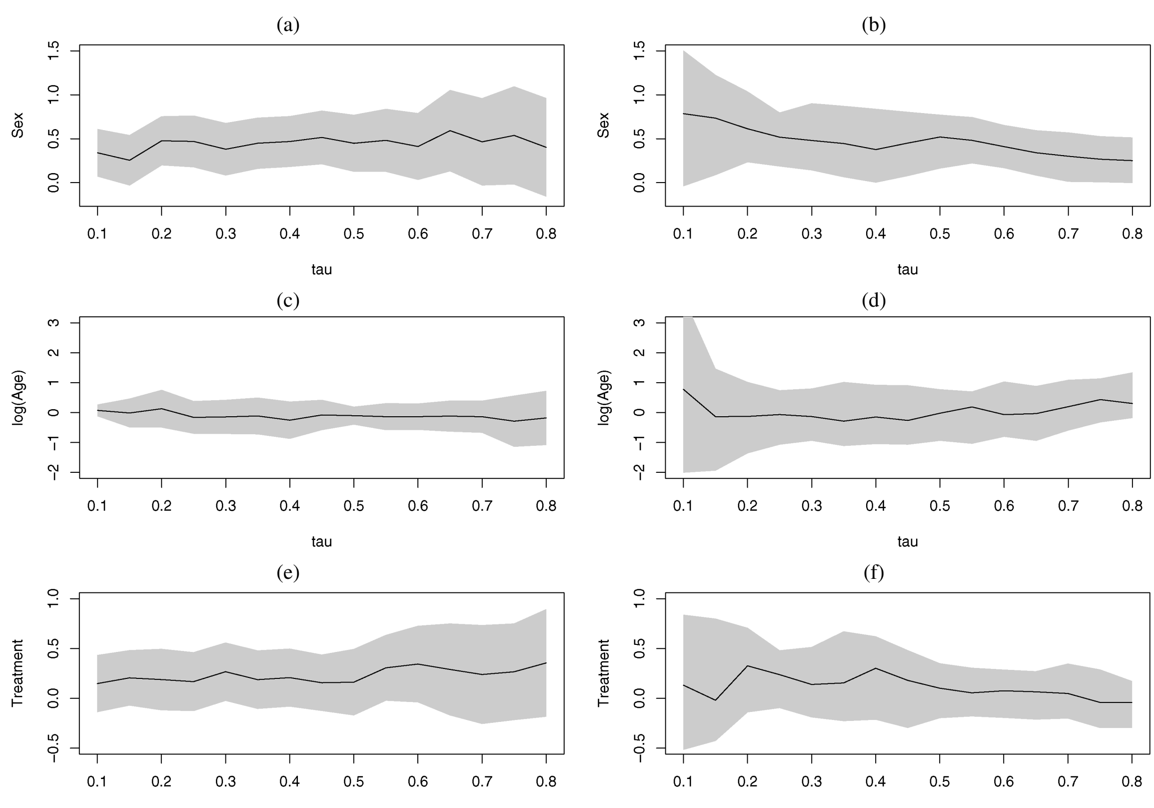

Figure 3. Quantile regression covariate effects for the NSCLC data when fixing the Box–Cox transformation parameter $\gamma = 0$. (a), (c), and (e) sex, log(age), and treatment for the proposed method. (b), (d), and (f) sex, log(age), and treatment for Portnoy's method.

Table 6. Estimates of the covariate effects for the NSCLC data under the Cox proportional hazards model

Covariate	Estimate	SE	p value
Sex	−.460	.174	.008
log(age)	−.211	.511	.680
Treatment	−.114	.168	.500

and thus provides a global picture of the relationship between the covariates and patient survival.

6. DISCUSSION

We have proposed a class of semiparametric transformation linear quantile regression models for randomly censored survival data. This class is broad, rich, and robust, and includes the well-known AFT mean-based model as a special case. We have provided an efficient two-stage algorithm for calculating the parameter estimates. At each stage, the estimation procedure minimizes a convex objective function that guarantees a unique solution. We have shown that the parameter estimators are strongly consistent and asymptotically normal. Numerical studies have demonstrated that the proposed method performs well for sample sizes of practical use.

Regarding the transformation $H_\gamma(T)$, we have studied both the conditional (with a known γ) and unconditional (with an unknown γ) inferences. Which approach is preferable in real applications is debatable. Each has its own advantages and disadvantages; for example, fixing $\gamma = 0$ yields the quantile-AFT model, which is popular and easily interpretable, and comparisons with other benchmarks (e.g., the Cox and Portnoy models) are straightforward, whereas the γ random case corresponds to a more flexible regression model under which the usual regression coefficients are not directly comparable due to different scales or transformations, and only the marginal covariate effects are meaningful, depending on subject-specific covariates. For the case with fixed γ, we can use a suitable model selection criterion to choose the model that best fits the data from a class of power-transformed quantile regression models. We then can interpret the estimates and inferences as conditioning on the given transformation. To account for variation in the estimation of γ, we can take the unconditional estimation by treating the transformation $H_\gamma(T)$, indexed by γ, as another model parameter. As a byproduct, the bootstrap resampling method automatically takes into account the variation from the estimation of γ, and thus provides correct inferences for the regression parameters.

APPENDIX: PROOFS

Proof of Theorem 1

For any fixed probability sample such that conditions (C.1)–(C.5) hold, because Γ is compact, by choosing a subsequence indexed by n, we assume that $\widehat{\gamma} \to \gamma^*$. We now show that $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ is bounded. Otherwise, for a subsequence, still denoted by n, $\|\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})\| \to \infty$. Define

$$\widehat{\boldsymbol{\alpha}}^* = \left(1 - \frac{1}{\|\widehat{\boldsymbol{\alpha}}(\widehat{\gamma}) - \boldsymbol{\alpha}(\gamma^*)\|}\right)\boldsymbol{\alpha}(\gamma^*) + \frac{1}{\|\widehat{\boldsymbol{\alpha}}(\widehat{\gamma}) - \boldsymbol{\alpha}(\gamma^*)\|}\widehat{\boldsymbol{\alpha}}(\widehat{\gamma}).$$

Note that $\widehat{\boldsymbol{\alpha}}^*$ is bounded and that its distance from $\boldsymbol{\alpha}(\gamma^*)$ is 1. We can further choose a subsequence to assume that $\widehat{\boldsymbol{\alpha}}^*$ has a limit $\boldsymbol{\alpha}^*$. By the convexity, $\Phi_{0n}(\widehat{\boldsymbol{\alpha}}^*; \widehat{\gamma}) \geq \Phi_{0n}(\boldsymbol{\alpha}(\gamma^*); \widehat{\gamma})$. Because the class of functions

$$\big\{(H_\gamma(X) - \boldsymbol{\alpha}^T\mathbf{Z})I(H_\gamma(X) - \boldsymbol{\alpha}^T\mathbf{Z} \geq 0) - \tau(H_\gamma(X) - \boldsymbol{\alpha}^T\mathbf{Z}) : \gamma \in \Gamma, \|\boldsymbol{\alpha} - \boldsymbol{\alpha}(\gamma^*)\| \leq 1\big\}$$

is the Glivenko–Cantelli class and $\widehat{G}(t) \to G_0(t)$ uniformly in $t \in [0, L]$, we conclude that as $n \to \infty$, $\Phi_0(\boldsymbol{\alpha}^*; \gamma^*) \geq \Phi_0(\boldsymbol{\alpha}(\gamma^*); \gamma^*)$. This is a contradiction, because $\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}(\gamma^*)\| = 1$.

Because $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ is bounded, by choosing a further subsequence, we assume that $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ converges to some $\boldsymbol{\alpha}^*$. After taking the limit of inequality $\Phi_{0n}(\widehat{\boldsymbol{\alpha}}(\widehat{\gamma}); \widehat{\gamma}) \geq \Phi_{0n}(\boldsymbol{\alpha}(\gamma^*); \widehat{\gamma})$, we conclude that $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}(\gamma^*)$. Similarly, we can show that $\widehat{\boldsymbol{\beta}}(\widehat{\gamma}) \to \boldsymbol{\beta}(\gamma^*)$.

The result holds if we can show that $\gamma^* = \gamma_0$ and $\boldsymbol{\alpha}(\gamma^*) = \boldsymbol{\beta}(\gamma^*) = \boldsymbol{\beta}_0$. As the covariates are bounded, by the Glivenko–Cantelli theorem, for any compact set $\mathcal{B}$,

$$\sup_{\mathbf{z},\boldsymbol{\beta}\in\mathcal{B},\gamma\in\Gamma}\left| n^{-1}\sum_{i=1}^n I(\mathbf{Z}_i \leq \mathbf{z})\left\{\frac{I(H_\gamma(X_i) - \boldsymbol{\beta}^T\mathbf{Z}_i \geq 0)}{\widehat{G}(H_\gamma^{-1}(\boldsymbol{\beta}^T\mathbf{Z}_i))} - \tau\right\} - E\big[I(\mathbf{Z} \leq \mathbf{z})\big\{I(H_\gamma(T) - \boldsymbol{\beta}^T\mathbf{Z} \geq 0) - \tau\big\}\big]\right| \xrightarrow{\text{a.s.}} 0.$$

Thus

$$n^{-3}R_n(\widehat{\gamma}) \to E\Big(E\big[I(\mathbf{Z} \leq \mathbf{z})\big\{I(H_{\gamma^*}(T) - \boldsymbol{\beta}(\gamma^*)^T\mathbf{Z} \geq 0) - \tau\big\}\big]^2\big|_{\mathbf{z}=\mathbf{Z}}\Big).$$

Because $n^{-3}R_n(\gamma_0) \to 0$ and $n^{-3}R_n(\gamma_0) \geq n^{-3}R_n(\widehat{\gamma})$, we conclude that

$$E\big[I(\mathbf{Z} \leq \mathbf{z})\big\{I(H_{\gamma^*}(T) - \boldsymbol{\beta}(\gamma^*)^T\mathbf{Z} \geq 0) - \tau\big\}\big|_{\mathbf{z}=\mathbf{Z}}\big] = 0;$$

that is, for any $\mathbf{Z}$, $\Pr(H_{\gamma^*}(T) \geq \boldsymbol{\beta}(\gamma^*)^T\mathbf{Z}|\mathbf{Z}) = \tau$. From condition (C.4), this gives

$$\boldsymbol{\beta}(\gamma^*)^T\mathbf{Z} = H_{\gamma^*}\big(H_{\gamma_0}^{-1}(\boldsymbol{\beta}_0^T\mathbf{Z})\big).$$

From condition (C.2), we obtain that $\gamma^* = \gamma_0$. This completes the proof.

Proof of Theorem 2

Clearly, $\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})$ is the solution to the equation of

$$n^{-1}\sum_{i=1}^n \frac{\Delta_i}{\widehat{G}(X_i)}\mathbf{Z}_i\big\{I(H_{\widehat{\gamma}}(X_i) - \boldsymbol{\alpha}^T\mathbf{Z}_i \geq 0) - \tau\big\} = 0.$$

From the uniform expansion (Fleming and Harrington 1991),

$$\sqrt{n}(\widehat{G}(t) - G_0(t)) = n^{-1/2}\sum_{i=1}^n V(X_i, \Delta_i; t) + o_p(1), \tag{A.1}$$

where $V(X_i, \Delta_i; t)$ is the influence function for the Kaplan–Meier estimator for the censoring distribution, that is,

$$V(X_i, \Delta_i; t) = -G_0(t)\int_0^t \frac{\widehat{G}(u-)\,dM_i(u)}{G_0(u)\sum_{i=1}^n I(X_i \geq u)},$$

where $M_i(t)$ is the martingale for the censoring time. Noting the Donsker property of the class

$$\left\{\frac{\Delta}{G(X)}\mathbf{Z}\big\{I(H_\gamma(X) - \boldsymbol{\beta}^T\mathbf{Z} \geq 0) - \tau\big\} : G \text{ is decreasing bounded away from 0}, \gamma \text{ and } \boldsymbol{\beta} \text{ are in the neighborhood of } \gamma_0 \text{ and } \boldsymbol{\beta}_0\right\},$$

we obtain that

$$-\sqrt{n}E\big[\mathbf{Z}\big\{I(H_{\widehat{\gamma}}(T)-\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})^T\mathbf{Z}\ge 0)-\tau\big\}\big] = \sqrt{n}(\mathbf{P}_n-\mathbf{P})Q_0(X,\mathbf{Z},\Delta)+o_p(1),$$

where $\mathbf{P}_n$ denotes the empirical measure, $\mathbf{P}$ is the expectation, and

$$Q_0(X,\mathbf{Z},\Delta) = \frac{\Delta}{G_0(X)}\mathbf{Z}\big\{I\big(H_{\gamma_0}(X)-\boldsymbol{\beta}_0^T\mathbf{Z}\ge 0\big)-\tau\big\} - \widetilde{E}\bigg\{\frac{\widetilde{\Delta}}{G_0(\widetilde{X})^2}\widetilde{\mathbf{Z}}\big(I\big(H_{\gamma_0}(\widetilde{X})-\boldsymbol{\beta}_0^T\widetilde{\mathbf{Z}}\ge 0\big)-\tau\big)V(X,\Delta;\widetilde{X})\bigg\},$$

where $(\widetilde{X},\widetilde{\mathbf{Z}},\widetilde{\Delta})$ are iid copies of $(X,\mathbf{Z},\Delta)$ and $\widetilde{E}(\cdot)$ takes the expectation with respect to $(\widetilde{X},\widetilde{\mathbf{Z}},\widetilde{\Delta})$ only. Therefore, after Taylor expansion of the left side, we have that

$$\begin{aligned}&-\sqrt{n}\big[E\{\mathbf{Z}\Psi_{0\gamma}(\mathbf{Z})\}(\widehat{\gamma}-\gamma_0)+E\{\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})\}(\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})-\boldsymbol{\beta}_0)\big]\\&\quad=\sqrt{n}(\mathbf{P}_n-\mathbf{P})Q_0(X,\mathbf{Z},\Delta)\\&\qquad+o_p\big(1+\sqrt{n}|\widehat{\gamma}-\gamma_0|+\sqrt{n}\|\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})-\boldsymbol{\beta}_0\|\big).\end{aligned}\tag{A.2}$$

Similarly, we obtain that

$$-\sqrt{n}E\bigg[\mathbf{Z}\bigg\{\frac{I(H_{\widehat{\gamma}}(X)-\widehat{\boldsymbol{\beta}}(\widehat{\gamma})^T\mathbf{Z}\ge 0)}{G_0(H_{\widehat{\gamma}}^{-1}(\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})^T\mathbf{Z}))}-\tau\bigg\}\bigg] = \sqrt{n}(\mathbf{P}_n-\mathbf{P})Q_1(X,\mathbf{Z},\Delta)+o_p(1),$$

where

$$Q_1(X,\mathbf{Z},\Delta) = \mathbf{Z}\bigg\{\frac{I(H_{\gamma_0}(X)-\boldsymbol{\beta}_0^T\mathbf{Z}\ge 0)}{G_0(H_{\gamma_0}^{-1}(\boldsymbol{\beta}_0^T\mathbf{Z}))}-\tau\bigg\} - \widetilde{E}\bigg\{\widetilde{\mathbf{Z}}\frac{I(H_{\gamma_0}(\widetilde{X})-\boldsymbol{\beta}_0^T\widetilde{\mathbf{Z}}\ge 0)}{G_0(H_{\gamma_0}^{-1}(\boldsymbol{\beta}_0^T\widetilde{\mathbf{Z}}))^2}V\big(X,\Delta;H_{\gamma_0}^{-1}(\boldsymbol{\beta}_0^T\widetilde{\mathbf{Z}})\big)\bigg\}.$$

After Taylor expansion of the left side, we have that

$$\begin{aligned}&-\sqrt{n}E\{\mathbf{Z}\Psi_{1\gamma}(\mathbf{Z})\}(\widehat{\gamma}-\gamma_0)-\sqrt{n}E\{\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z})\}(\widehat{\boldsymbol{\beta}}(\widehat{\gamma})-\boldsymbol{\beta}_0)\\&\qquad-\sqrt{n}E\{\mathbf{Z}\Psi_{1\boldsymbol{\alpha}}(\mathbf{Z})\}(\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})-\boldsymbol{\beta}_0)\\&\quad=\sqrt{n}(\mathbf{P}_n-\mathbf{P})Q_1(X,\mathbf{Z},\Delta)\\&\qquad+o_p\big(1+\sqrt{n}|\widehat{\gamma}-\gamma_0|+\sqrt{n}\|\widehat{\boldsymbol{\beta}}(\widehat{\gamma})-\boldsymbol{\beta}_0\|\\&\qquad+\sqrt{n}\|\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})-\boldsymbol{\beta}_0\|\big).\end{aligned}\tag{A.3}$$

From (A.2) and (A.3), we cancel the term $(\widehat{\boldsymbol{\alpha}}(\widehat{\gamma})-\boldsymbol{\beta}_0)$ on the left side and obtain

$$\begin{aligned}&\sqrt{n}\big[-E\{\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z})\}\mathbf{A}_1\big](\widehat{\gamma}-\gamma_0)+\sqrt{n}E\{\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z})\}(\widehat{\boldsymbol{\beta}}(\widehat{\gamma})-\boldsymbol{\beta}_0)\\&\quad=\sqrt{n}(\mathbf{P}_n-\mathbf{P})\big[E\{\mathbf{Z}\Psi_{1\boldsymbol{\alpha}}(\mathbf{Z})\}E\{\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})\}^{-1}Q_0(X,\mathbf{Z},\Delta)\\&\qquad+Q_1(X,\mathbf{Z},\Delta)\big]+o_p(1),\end{aligned}$$

where

$$\mathbf{A}_1=-\big[E\{\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z})\}\big]^{-1}\times\big[E\{\mathbf{Z}\Psi_{1\gamma}(\mathbf{Z})\}-E\{\mathbf{Z}\Psi_{1\boldsymbol{\alpha}}(\mathbf{Z})\}E\{\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})\}^{-1}E\{\mathbf{Z}\Psi_{0\gamma}(\mathbf{Z})\}\big].$$

On the other hand, we note that the class of functions

$$\bigg\{I(\mathbf{Z}\le\mathbf{z})\bigg(\frac{I(H_\gamma(X)-\boldsymbol{\beta}^T\mathbf{Z}\ge 0)}{G(H_\gamma^{-1}(\boldsymbol{\beta}^T\mathbf{Z}))}-\tau\bigg): \gamma \text{ is in a neighborhood of } \gamma_0,\ \boldsymbol{\beta} \text{ is in a neighborhood of } \boldsymbol{\beta}_0,\ G \text{ is a nonincreasing function in } [0,L] \text{ and bounded away from } 0,\ \mathbf{z}\in R^p\bigg\}$$

is a Donsker class. Thus

$$\begin{aligned}\sup_{\mathbf{z},\gamma\in\Gamma}\bigg|&n^{-1}D_n(\mathbf{z},\gamma)\\&-E\bigg[I(\mathbf{Z}\le\mathbf{z})\bigg(\frac{I(H_\gamma(X)-\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}\ge 0)}{\widehat{G}(H_\gamma^{-1}(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}))}-\tau\bigg)\bigg]\\&-(\mathbf{P}_n-\mathbf{P})\bigg[I(\mathbf{Z}\le\mathbf{z})\bigg(\frac{I(H_\gamma(X)-\boldsymbol{\beta}(\gamma)^T\mathbf{Z}\ge 0)}{G_0(H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T\mathbf{Z}))}-\tau\bigg)\bigg]\bigg|\\&=o_p\big(n^{-1/2}\big).\end{aligned}$$

In addition,

$$\begin{aligned}&E\bigg[I(\mathbf{Z}\le\mathbf{z})\bigg(\frac{I(H_\gamma(X)-\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}\ge 0)}{\widehat{G}(H_\gamma^{-1}(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}))}-\tau\bigg)\bigg]\\&\quad=E\bigg[I(\mathbf{Z}\le\mathbf{z})\bigg(\frac{I(H_\gamma(X)-\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}\ge 0)}{G_0(H_\gamma^{-1}(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}))}-\tau\bigg)\bigg]\\&\qquad-(1+o_p(1))E\bigg[I(\mathbf{Z}\le\mathbf{z})\frac{I(H_\gamma(X)-\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}\ge 0)}{G_0(H_\gamma^{-1}(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z}))^2}\\&\qquad\times\big\{\widehat{G}\big(H_\gamma^{-1}(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z})\big)-G_0\big(H_\gamma^{-1}(\widehat{\boldsymbol{\beta}}(\gamma)^T\mathbf{Z})\big)\big\}\bigg].\end{aligned}$$

After Taylor expansion of the first term at $\boldsymbol{\beta}(\gamma)$ and using the expansion of $\widehat{G}$ in (A.1), we obtain that, uniformly in $\mathbf{z}$ and γ in a neighborhood of γ_0,

$$\begin{aligned}n^{-1}D_n(\mathbf{z},\gamma)&=E\big[I(\mathbf{Z}\le\mathbf{z})\big\{I(H_\gamma(T)-\boldsymbol{\beta}(\gamma)^T\mathbf{Z}\ge 0)-\tau\big\}\big]\\&\quad+(\mathbf{P}_n-\mathbf{P})Q_2(X,\mathbf{Z},\Delta;\gamma,\mathbf{z})\\&\quad+E[I(\mathbf{Z}\le\mathbf{z})\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})](\widehat{\boldsymbol{\beta}}(\gamma)-\boldsymbol{\beta}(\gamma))\\&\quad+o_p\big(\|\widehat{\boldsymbol{\beta}}(\gamma)-\boldsymbol{\beta}(\gamma)\|+|\gamma-\gamma_0|\big)+o_p\big(n^{-1/2}\big),\end{aligned}$$

where

$$\begin{aligned}&Q_2(X,\mathbf{Z},\Delta;\gamma,\mathbf{z})\\&\quad=I(\mathbf{Z}\le\mathbf{z})\bigg\{\frac{I(H_\gamma(X)-\boldsymbol{\beta}(\gamma)^T\mathbf{Z}\ge 0)}{G_0(H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T\mathbf{Z}))}-\tau\bigg\}\\&\qquad-\widetilde{E}\bigg\{I(\widetilde{\mathbf{Z}}\le\mathbf{z})\frac{I(H_\gamma(\widetilde{X})-\boldsymbol{\beta}(\gamma)^T\widetilde{\mathbf{Z}}\ge 0)}{G_0(H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T\widetilde{\mathbf{Z}}))}\\&\qquad\times V\big(\mathbf{X},\Delta;H_\gamma^{-1}(\boldsymbol{\beta}(\gamma)^T\widetilde{\mathbf{Z}})\big)\bigg\}.\end{aligned}$$

As a result,

$$\begin{aligned}&n^{-3}R_n(\gamma)\\&\quad=E\big\{\big(E\big[I(\mathbf{Z}\le\mathbf{z})\big\{I(H_\gamma(T)-\boldsymbol{\beta}(\gamma)^T\mathbf{Z}\ge 0)-\tau\big\}\big]\big)^2\big|_{\mathbf{z}=\mathbf{Z}}\big\}\\&\qquad+2E\big(E\big[I(\mathbf{Z}\le\mathbf{z})\big\{I(H_\gamma(T)-\boldsymbol{\beta}(\gamma)^T\mathbf{Z}\ge 0)-\tau\big\}\big]\\&\qquad\times(\mathbf{P}_n-\mathbf{P})Q_2(X,\mathbf{Z},\Delta;\gamma,\mathbf{z})\big|_{\mathbf{z}=\mathbf{Z}}\big)\\&\qquad+2E\big(E\big[I(\mathbf{Z}\le\mathbf{z})\big\{I\big(H_\gamma(T)-\boldsymbol{\beta}(\gamma)^T\mathbf{Z}\ge 0\big)-\tau\big\}\big]\\&\qquad\times E\{I(\mathbf{Z}\le\mathbf{z})\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})\}(\widehat{\boldsymbol{\beta}}(\gamma)-\boldsymbol{\beta}(\gamma))\big|_{\mathbf{z}=\mathbf{Z}}\big)\\&\qquad+o_p(n^{-1})+o_p\big(\|\widehat{\boldsymbol{\beta}}(\gamma)-\boldsymbol{\beta}(\gamma)\|^2+|\gamma-\gamma_0|^2\big).\end{aligned}$$

From Lemma 1, we note that

$$E\big[I(\mathbf{Z}\le \mathbf{z})\big\{I(H_\gamma(T)-\boldsymbol{\beta}(\gamma)^T\mathbf{Z}\ge 0)-\tau\big\}\big] = B(\mathbf{z})(\gamma-\gamma_0)+o(|\gamma-\gamma_0|)$$

in a neighborhood of γ_0, where

$$B(\mathbf{z}) = E\{I(\mathbf{Z}\le \mathbf{z})\Psi_{0\gamma}(\mathbf{Z})\}+\mathbf{A}_2(\mathbf{z})^T\mathbf{A}_1,$$

with $\mathbf{A}_2(\mathbf{z}) = E\{I(\mathbf{Z}\le \mathbf{z})\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})^T\}$. Therefore, $n^{-3}R_n(\gamma)$ has a quadratic expansion near γ_0 as

$$\begin{aligned} &o_p\big(|\gamma-\gamma_0|^2+n^{-1}+\|\widehat{\boldsymbol{\beta}}(\gamma)-\boldsymbol{\beta}(\gamma)\|^2\big)+E\{B(\mathbf{Z})^2\}(\gamma-\gamma_0)^2\\ &\quad+2(\gamma-\gamma_0)\big[(\mathbf{P}_n-\mathbf{P})\widetilde{E}\{B(\widetilde{\mathbf{Z}})Q_2(X,\mathbf{Z},\Delta;\gamma_0,\widetilde{\mathbf{Z}})\}\\ &\quad+E\{B(\mathbf{Z})\mathbf{A}_2(\mathbf{Z})^T\}(\widehat{\boldsymbol{\beta}}(\gamma)-\boldsymbol{\beta}(\gamma))\big].\end{aligned}$$

Because $\widehat{\gamma}$ minimizes $R_n(\gamma)$, we obtain that

$$\begin{aligned} &E\{B(\mathbf{Z})^2\}(\widehat{\gamma}-\gamma_0)+E\{B(\mathbf{Z})\mathbf{A}_2(\mathbf{Z})^T\}(\widehat{\boldsymbol{\beta}}(\widehat{\gamma})-\boldsymbol{\beta}(\widehat{\gamma}))\\ &\quad=-(\mathbf{P}_n-\mathbf{P})\widetilde{E}\{B(\widetilde{\mathbf{Z}})Q_2(X,\mathbf{Z},\Delta;\gamma_0,\widetilde{\mathbf{Z}})\}\\ &\qquad+o_p\big(|\widehat{\gamma}-\gamma_0|+n^{-1/2}+\|\widehat{\boldsymbol{\beta}}(\widehat{\gamma})-\boldsymbol{\beta}(\widehat{\gamma})\|\big).\end{aligned}$$

Using the expansion of $\boldsymbol{\beta}(\widehat{\gamma})=\boldsymbol{\beta}_0+\mathbf{A}_1(\widehat{\gamma}-\gamma_0)+o(|\widehat{\gamma}-\gamma_0|)$, after some algebraic manipulations and defining $A_3(\mathbf{z})=E\{I(\mathbf{Z}\le \mathbf{z})\Psi_{0\gamma}(\mathbf{Z})\}$, we have that

$$\begin{aligned} &\sqrt{n}\big[E\{A_3(\mathbf{Z})^2\}-\mathbf{A}_1^TE\{\mathbf{A}_2(\mathbf{Z})A_3(\mathbf{Z})\}\big](\widehat{\gamma}-\gamma_0)\\ &\quad-\sqrt{n}\big[E\{A_3(\mathbf{Z})\mathbf{A}_2(\mathbf{Z})^T\}-\mathbf{A}_1^TE\{\mathbf{A}_2(\mathbf{Z})\mathbf{A}_2(\mathbf{Z})^T\}\big]\\ &\quad\times(\widehat{\boldsymbol{\beta}}(\widehat{\gamma})-\boldsymbol{\beta}_0)\\ &=-\sqrt{n}(\mathbf{P}_n-\mathbf{P})\widetilde{E}\{B(\widetilde{\mathbf{Z}})Q_2(X,\mathbf{Z},\Delta;\gamma_0,\widetilde{\mathbf{Z}})\}\\ &\quad+o_p\big(\sqrt{n}|\widehat{\gamma}-\gamma_0|+1+\sqrt{n}\|\widehat{\boldsymbol{\beta}}(\widehat{\gamma})-\boldsymbol{\beta}_0\|\big). \qquad \text{(A.4)}\end{aligned}$$

Finally, asymptotic normality follows if we can show that the coefficient matrix is nonsingular. We note that the coefficient matrix

$$\boldsymbol{\Sigma}_0=\begin{pmatrix} -E\{\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z})\}\mathbf{A}_1 & E\{\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z})\}\\ E\{A_3(\mathbf{Z})^2\}-\mathbf{A}_1^TE\{\mathbf{A}_2(\mathbf{Z})A_3(\mathbf{Z})\} & E\{A_3(\mathbf{Z})\mathbf{A}_2(\mathbf{Z})^T\}-\mathbf{A}_1^TE\{\mathbf{A}_2(\mathbf{Z})\mathbf{A}_2(\mathbf{Z})^T\}\end{pmatrix}$$

has the same rank as

$$\begin{pmatrix} 0 & -E\{\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})\}E\{\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z})\}\\ E\{A_3(\mathbf{Z})^2\}+\mathbf{A}_1^TE\{\mathbf{A}_2(\mathbf{Z})^{\otimes 2}\}\mathbf{A}_1 & E\{A_3(\mathbf{Z})\mathbf{A}_2(\mathbf{Z})^T\}-\mathbf{A}_1^TE\{\mathbf{A}_2(\mathbf{Z})^{\otimes 2}\}\end{pmatrix}.$$

The latter has full rank because $E\{\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})\}E\{\mathbf{Z}\Psi_{1\boldsymbol{\beta}}(\mathbf{Z})\}$ is positive definite by some straightforward calculations. Thus the asymptotic covariance is given by

$$\begin{aligned}\boldsymbol{\Sigma}&=\boldsymbol{\Sigma}_0^{-1}\\ &\quad\times E\left[\begin{pmatrix} E\{\mathbf{Z}\Psi_{1\boldsymbol{\alpha}}(\mathbf{Z})\}E\{\mathbf{Z}\Psi_{0\boldsymbol{\alpha}}(\mathbf{Z})\}^{-1}Q_0(X,\mathbf{Z},\Delta)+Q_1(X,\mathbf{Z},\Delta)\\ \widetilde{E}\{B(\widetilde{\mathbf{Z}})Q_2(X,\mathbf{Z},\Delta;\gamma_0,\widetilde{\mathbf{Z}})\}\end{pmatrix}^{\otimes 2}\right]\\ &\quad\times(\boldsymbol{\Sigma}_0^{-1})^T.\end{aligned}$$

Although we consider only the situation in which γ is estimated by minimizing $R_n(\gamma)$, the same arguments also apply to the case corresponding to $R_n^*(\gamma)$, whereas (A.4) will be different, reflecting the use of $I(\widehat{\boldsymbol{\beta}}(\widehat{\gamma})^T\mathbf{Z}\le t)$ instead of $I(\mathbf{Z}\le \mathbf{z})$.

[Received March 2007. Revised April 2008.]

REFERENCES

Bang, H., and Tsiatis, A. A. (2002), "Median Regression With Censored Cost Data," *Biometrics*, 58, 643–649.

Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 26, 211–252.

Bilias, Y., Chen, S., and Ying, Z. (2000), "Simple Resampling Methods for Censored Regression Quantiles," *Journal of Econometrics*, 99, 373–386.

Buchinsky, M. (1995), "Quantile Regression, Box–Cox Transformation Model, and the U.S. Wage Structure, 1963–1987," *Journal of Econometrics*, 65, 109–154.

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, T., Tian, L., and Wei, L. J. (2005), "Semiparametric Box–Cox Power Transformation Models for Censored Survival Observations," *Biometrika*, 92, 619–632.

Chamberlain, G. (1994), "Quantile Regression, Censoring and the Structure of Wages," in *Advances in Econometrics*, ed. C. Sims, New York: Cambridge University Press, pp. 171–209.

Cheng, S. C., Wei, L. J., and Ying, Z. (1995), "Analysis of Transformation Models With Censored Data," *Biometrika*, 82, 835–845.

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

Efron, B., and Tibishirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Fleming, T. R., and Harrington, D. (1991), *Counting Processes and Survival Analysis*, New York: Wiley.

Hahn, J. (1995), "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 11, 105–121.

He, X., and Zhu, L. X. (2003), "A Lack-of-Fit Test for Quantile Regression," *Journal of the American Statistical Association*, 98, 1013–1022.

Honoré, B., Khan, S., and Powell, J. L. (2002), "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 109, 67–105.

Horowitz, J. (1998), "Bootstrap Methods for the 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 Powell, L. J. (2001), "Two-Step Estimation of Semiparametric Censored Regression Models," *Journal of Econometrics*, 103, 73–110.

Khan, S., and Tamer, E. (2007), "Partial Rank Estimation of Duration Models With General Forms of Censoring," *Journal of Econometrics*, 136, 251–280.

——— (2008), "Inference on Randomly Censored Regression Models Using Conditional Moment Inequalities," manuscript, Duke University.

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

Koenker, R., and Bassett, G. J. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and Geling, O. (2001), "Reappraising Medfly Longevity: A Quantile Regression Survival Analysis," *Journal of the American Statistical Association*, 96, 458–468.

Lai, T. L., and Ying, Z. (1991), "Rank Regression Methods for Left-Truncated and Right-Censored Data," *The Annals of Statistics*, 19, 531–556.

Lindgren, A. (1997), "Quantile Regression With Censored Data Using Generalized L_1 Minimization," *Computational Statistics & Data Analysis*, 23, 509–524.

Mu, Y., and He, X. (2007), "Power Transformation Toward a Linear Regression Quantile," *Journal of the American Statistical Association*, 102, 269–279.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994), "A Resampling Method Based on Pivotal Estimating Functions," *Biometrika*, 81, 341–350.

Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012.

Portnoy, S., and Koenker, R. (1997), "The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimations," *Statistical Science*, 12, 279–296.

Powell, J. L. (1986), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

Prentice, R. L. (1978), "Linear Rank Tests With Right-Censored Data," *Biometrika*, 65, 167–180.

Ritov, Y. (1990), "Estimation in a Linear Regression Model With Censored Data," *The Annals of Statistics*, 18, 303–328.

Robins, J. M. (1996), "Locally Efficient Median Regression With Random Censoring and Surrogate Markers," in *Lifetime Data: Models in Reliability and Survival Analysis*, eds. N. P. Jewell, A. C. Kimber, M.-L. T. Lee, G. A. Whitmore, Kluwer Academic, pp. 263–274.

Robins, J. M., and Rotnitzky, A. (1992), "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers," in *AIDS Epidemiology: Methodological Issues*, eds. N. Jewell, K. Dietz, and V. Farewell, Boston: Birkhäuser, pp. 297–331.

Socinski, M. A., Schell, M. J., Peterman, A., Bakri, K., Yates, S., Gitten, R., Unger, P., Lee, J., Lee, J. H., Tynan, M., Moore, M., and Kies, M. S. (2002), "Phase III Trial Comparing a Defined Duration of Therapy versus Continuous Therapy Followed by Second-Line Therapy in Advanced-Stage IIIB/IV Non–Small-Cell Lung Cancer," *Journal of Clinical Oncology*, 20, 1335–1343.

Tsiatis, A. A. (1990), "Estimating Regression Parameters Using Linear Rank Tests for Censored Data," *The Annals of Statistics*, 18, 354–372.

van der Vaart, A. W., and Wellner, J. A. (2000), *Weak Convergence and Empirical Processes*, New York: Springer.

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," *Journal of the American Statistical Association*, 94, 137–145.

Ying, Z., Jung, S. H., and Wei, L. J. (1995), "Survival Analysis With Median Regression Models," *Journal of the American Statistical Association*, 90, 178–184.

Yu, K., Lu, Z., and Stander, J. (2003), "Quantile Regression: Applications and Current Research Areas," *Journal of the Royal Statistical Society*, Ser. D, 52, 331–350.