Statistica Sinica **21** (2011), 949-971

CENSORED QUANTILE REGRESSION WITH COVARIATE MEASUREMENT ERRORS

Yanyuan Ma and Guosheng Yin

Texas A&M University and The University of Hong Kong

Abstract: We study censored quantile regression with covariates measured with errors. We propose a composite quantile objective function based on inverse censoring-probability weighting, and an averaging estimator to improve estimation efficiency. Our procedure can eliminate the bias in the naive estimator that is obtained by treating mismeasured covariates as error-free. Using a combination of martingale and quantile regression techniques, we show that the proposed estimators for the regression coefficients are consistent and asymptotically normal. We conducted simulation studies to examine the finite-sample properties of the new method, and demonstrated efficiency gain of the averaging estimator over the single quantile regression estimator. For illustration, we applied our model to a lung cancer study.

Key words and phrases: Averaging estimation, bootstrap, errors-in-variables problem, regression quantiles, semiparametric method, survival data.

1. Introduction

Linear regression models have been extensively studied for randomly censored survival data. In particular, the accelerated failure time (AFT) model is intuitively attractive and easily interpretable. It directly formulates a linear model between the logarithm of the failure time and covariates $\mathbf{X}$,

$$\log T = \boldsymbol{\beta}^T \mathbf{X} + \epsilon, \tag{1.1}$$

where the model error ϵ has a zero mean. Semiparametric estimation of model (1.1) is typically based on least squares or rank methods; see Prentice (1978), Buckley and James (1979), Ritov (1990), Tsiatis (1990), Wei, Ying, and Lin (1990), Lai and Ying (1991), and Jin et al. (2003), among others. The estimating functions are discrete and potentially have multiple roots, and furthermore, the variances are typically estimated using resampling methods due to their dependence on the density function of ϵ.

When covariate $\mathbf{X}$ is subject to measurement errors, naively considering the mismeasured covariate to be error-free would cause estimation bias. In classical linear measurement error models, the attenuation effect of the likelihood-based estimator is known due to measurement errors. We consider the measurement error in the form $\mathbf{W} = \mathbf{X} + \mathbf{U}$, where $\mathbf{W}$ is the observed surrogate, $\mathbf{X}$ is the true unobserved covariate, and $\mathbf{U}$ is the measurement error. A common practice is to assume that both the model error ϵ and the measurement error $\mathbf{U}$ are normal, and to specify either a ratio between the variances of ϵ, and $\mathbf{U}$, or the variance for one of the two errors. These specifications are needed for model identifiability and carrying out estimation via mean regression techniques, such as the total least-squares method. However, it is obvious that such an assumption is quite restrictive and would lead to bias if normality does not hold. For example, in the data set considered in Section 6, the normality assumption of the measurement error is not empirically supported by a Kolmogorov-Smirnov test, hence a less restrictive assumption on the error distribution is needed. See also Carroll et al. (2006) on various measurement error models and their effects on inference.

In contrast to the mean-based linear models, quantile regression serves as a robust alternative when the median or any other quantile of the model error ϵ is assumed to be zero (Koenker and Bassett (1978)). In quantile regression, model parameters are often estimated by solving quantile-based estimating equations through linear programming or interior point methods (Koenker (2005)), and the corresponding variances are typically estimated by resampling algorithms. For the fixed censoring case that is related to the Tobit model in economics, see the work of Powell (1984), Buchinsky and Hahn (1998), Fitzenberger (1997), and Khan and Powell (2001), among others. Recently, there has also been a growing interest in the application of quantile regression to randomly censored failure time data (Ying, Jung, and Wei (1995); Lindgren (1997); Yang (1999); Koenker and Geling (2001); Bang and Tsiatis (2002); Chernozhukov and Hong (2002); Portnoy (2003); and Peng and Huang (2008), among others).

In practice, one may be interested in an overall covariate effect for a certain range of quantiles instead of each specific quantile. A composite quantile regression model assumes that there exist common covariate effects in a range of quantiles such that the quantile levels only differ in terms of the intercept. From a more general regression perspective, composite quantile regression seeks to model a set of parallel regression curves, and thus it can be viewed as a compromise between a set of quantile regression curves with different intercepts and slopes and a single summary regression curve. In the measurement error model that we consider, composite quantile regression does not require extra assumptions, as the independence between the distribution for the model error ϵ and covariates already guarantees parallel regression curves. Intuitively, the composite quantile regression should provide estimation efficiency gain over a single quantile regression, see also Zou and Yuan (2008). In light of these considerations, we derive our estimator in the composite quantile regression framework, while noting that if only one quantile is concerned, the proposed method degenerates to traditional single quantile regression.

Very limited research has been carried out with measurement error problems in the context of quantile regression, mainly due to its difficult nature. An early attempt by Brown (1982) examined a median regression model and described the difficulty of parameter estimation. He and Liang (2000) proposed an innovative root-n consistent estimator in the context of linear and partially linear models using a quantile regression approach. Recently, Wei and Carroll (2009) studied a general quantile regression model with measurement errors, a substantial advance in this area. In this paper, we study the issue of covariate measurement errors in quantile regression with randomly censored data. We propose an inverse weighting estimation method and give its asymptotic properties. The derivation of the asymptotics requires techniques in martingale theory and in the treatment of quantile regression, which is highly nontrivial. Moreover, we propose a computing algorithm for composite quantile regression through an averaging estimation approach. The quantile averaging estimator is naturally motivated, numerically stable, and easy to implement. There exist several general approaches in measurement error models that could also be implemented in the quantile regression context, such as simulation extrapolation and regression calibration (Carroll et al. (2006)). However, both are approximate methods and do not produce consistent estimates in the quantile regression context.

2. Estimation

For $i = 1, \ldots, n$, let T_i be the failure time for the ith subject and C_i be the censoring time. We observe $Y_i = \min(T_i, C_i)$ and the censoring time indicator $\Delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function. Let $\mathbf{X}_i$ be the corresponding unobservable p-dimensional vector of bounded covariates. Instead, we observe the surrogate $\mathbf{W}_i$ and write $\mathbf{W}_i = \mathbf{X}_i + \mathbf{U}_i$, where $\mathbf{U}_i$ is the measurement error. We assume that the observed data $(\mathbf{W}_i, Y_i, \Delta_i)$ are independent and identically distributed (i.i.d.). Under this setup, the censored quantile regression model subject to measurement errors has the form

$$\begin{aligned} \log T_i &= \alpha + \boldsymbol{\beta}^T \mathbf{X}_i + \epsilon_i, \\ \mathbf{W}_i &= \mathbf{X}_i + \mathbf{U}_i, \end{aligned} \tag{2.1}$$

where we assume $(\epsilon_i, \mathbf{U}_i^T)^T \in \mathcal{R}^{p+1}$ are i.i.d. according to a spherically symmetric distribution and independent of $\mathbf{X}_i$. In particular, for any orthogonal matrix $\mathbf{D}$, $\mathbf{D}(\epsilon, \mathbf{U}^T)^T$ has the same distribution as $(\epsilon, \mathbf{U}^T)^T$. A spherically symmetric distribution accommodates the commonly used multivariate normal and multivariate t distributions as special cases, hence is a more relaxed assumption than the usual normal error assumption. A further extension would allow $\mathbf{M}(\epsilon, \mathbf{U}^T)^T$ to be spherically symmetric for some fixed matrix $\mathbf{M}$. However, ϵ and $\mathbf{U}$ are usually independent, and the effect of $\mathbf{M}$ on $\mathbf{U}$ can be equivalently captured through a reparametrization (see Appendix); for ease of exposition, we focus on the case in which $\mathbf{M}$ is the identity matrix.

If no censoring occurs and the true covariates $\mathbf{X}_i$ are observed, then at a preselected τth quantile, the main parameters $(\alpha_\tau, \boldsymbol{\beta})$ in the quantile regression model can be estimated through

$$\min_{\alpha_\tau, \boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(\log T_i - \alpha_\tau - \boldsymbol{\beta}^T \mathbf{X}_i), \tag{2.2}$$

where the usual "check function" $\rho_\tau(u) = u\{I(u \geq 0) - (1 - \tau)\}$.

Under random censorship, we let $G(\cdot)$ be the survival function for the censoring time C_i, and $\hat{G}(\cdot)$ be the corresponding Kaplan-Meier estimator based on $\{(Y_i, 1 - \Delta_i), i = 1, \ldots, n\}$. Taking into account both the censoring and the measurement error in $\mathbf{W}$, we propose minimizing the inverse censoring-probability weighted objective function

$$\Psi_n(\alpha_\tau, \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\hat{G}(Y_i)} \rho_\tau \left(\frac{\log Y_i - \alpha_\tau - \boldsymbol{\beta}^T \mathbf{W}_i}{\sqrt{1 + |\boldsymbol{\beta}|^2}} \right) \tag{2.3}$$

with respect to $(\alpha_\tau, \boldsymbol{\beta})$. In contrast to the usual quantile regression models, $\boldsymbol{\beta}$ in (2.3) does not depend on τ, i.e., the slopes are the same regardless of the quantile level. If we denote the minimizer of $\Psi_n(\alpha_\tau, \boldsymbol{\beta})$ by $(\hat{\alpha}_\tau, \hat{\boldsymbol{\beta}})$, then $\hat{\alpha}_\tau$ is a consistent estimator of $\alpha_\tau = \alpha + q_\tau \sqrt{1 + |\boldsymbol{\beta}|^2}$, and $\hat{\boldsymbol{\beta}}$ is that of $\boldsymbol{\beta}$, where q_τ is the unique solution to $E\{\rho_\tau(\epsilon_i - q)\} = 0$, i.e., q_τ is the τth quantile of ϵ. For such inverse probability weighting techniques, see Robins and Rotnitzky (1992), Robins (1996), and Bang and Tsiatis (2000). The measurement error correction factor $\sqrt{1 + |\boldsymbol{\beta}|^2}$ is widely used in linear models with additive errors. The main intuition is the following. In the usual regression, one minimizes the vertical standardized distance $d\{(Y - \alpha - \boldsymbol{\beta}^T \mathbf{X})/\text{s.d.}(Y - \alpha - \boldsymbol{\beta}^T \mathbf{X})\}$ where d stands for a suitable distance measure and s.d. is the standard deviation, because only the vertical Y direction has errors. However, in the measurement error situation, errors also occur along the horizontal $\mathbf{X}$ direction, hence a distance containing both vertical and horizontal components should be favored. In fact, the minimization of the same standardized distance with $\mathbf{X}$ replaced by $\mathbf{W}$ automatically corrects for this. If we denote the variance of ϵ as σ_ϵ^2 and the variance-covariance matrix of $\mathbf{U}$ as $\boldsymbol{\Sigma}_U$, we have

$$\frac{(Y - \alpha - \boldsymbol{\beta}^T \mathbf{W})}{\text{s.d.}(Y - \alpha - \boldsymbol{\beta}^T \mathbf{W})} = \frac{(Y - \alpha - \boldsymbol{\beta}^T \mathbf{W})}{\sqrt{\sigma_\epsilon^2 + \boldsymbol{\beta}^T \boldsymbol{\Sigma}_U \boldsymbol{\beta}}},$$

which is proportional to $(Y - \alpha - \boldsymbol{\beta}^T \mathbf{W})/\sqrt{1 + |\boldsymbol{\beta}|^2}$ under the spherical symmetry assumption. This correction is first seen in Lindley (1947) for the L_2 distance and in He and Liang (2000) for the L_1 distance.

If there is no particular reason to favor any specific value of τ, we can use a set of different quantile levels simultaneously to estimate $(\alpha_\tau, \boldsymbol{\beta})$. Let Q represent the set of quantiles under consideration, $Q = \{\tau_1, \ldots, \tau_k\}$, $\tau_1 < \ldots < \tau_k$. Under composite quantile regression, we estimate $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ by minimizing

$$\Psi_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\hat{G}(Y_i)} \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log Y_i - \alpha_\tau - \boldsymbol{\beta}^T \mathbf{W}_i}{\sqrt{1 + |\boldsymbol{\beta}|^2}} \right), \tag{2.4}$$

where $\boldsymbol{\alpha}$ represents the concatenated vector of the α_τ's corresponding to different values of $\tau \in Q$, $\boldsymbol{\alpha} = (\alpha_{\tau_1}, \ldots, \alpha_{\tau_k})^T$. We include a quantile specific weight ω_τ to allow for weighing each regression quantile differently. In practice, a convenient choice of ω_τ is simply to use equal weights. More sophisticated weighting strategies may use weights proportional to the effective sample size, weights inversely proportional to the trace of the variance-covariance matrix of each quantile regression estimation, or weights that minimize the overall estimation variance. These more complex weighting schemes require iterative implementation of the quantile regression procedure, and typically induce numerical difficulties.

3. Asymptotic Properties

We use the notation

$$\begin{aligned}
t_i(\alpha_\tau, \boldsymbol{\beta}) &= \omega_\tau \frac{\partial}{\partial \alpha_\tau} \rho_\tau \left(\frac{\log T_i - \alpha_\tau - \boldsymbol{\beta}^T \mathbf{W}_i}{\sqrt{1 + |\boldsymbol{\beta}|^2}} \right), \\
\mathbf{r}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log T_i - \alpha_\tau - \boldsymbol{\beta}^T \mathbf{W}_i}{\sqrt{1 + |\boldsymbol{\beta}|^2}} \right), \\
\mathbf{t}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \{t_i(\alpha_{\tau_1}, \boldsymbol{\beta}), \ldots, t_i(\alpha_{\tau_k}, \boldsymbol{\beta})\}^T, \\
\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \{\mathbf{t}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})^T, \mathbf{r}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})^T\}^T.
\end{aligned}$$

With L as the study duration, we define a filtration $\mathcal{F}(u)$ to be the set of σ-algebras generated by $\sigma\{I(C_i \le t), t \le u; I(T_i \le y), \mathbf{W}_i, \tau \in Q, 0 \le y \le L, i = 1, \ldots, n\}$. We take the counting process for the censoring time to be $N_i^c(u) = I(Y_i \le u, \Delta_i = 0)$, the risk process as $R_i(u) = I(Y_i \ge u)$, and let $\lambda^c(u)$ be the hazard function for the censoring distribution. With the martingale $\mathcal{M}_i^c(u) = N_i^c(u) - \int_0^u \lambda^c(t) R_i(t) dt$, we write $\mathcal{M}^c(u) = \sum_{i=1}^n \mathcal{M}_i^c(u)$, $N^c(u) = \sum_{i=1}^n N_i^c(u)$, and $R(u) = \sum_{i=1}^n R_i(u)$. Note that $R(u) = n\hat{G}(u^-)\hat{S}(u^-)$, where $\hat{G}(u^-)$ is the left continuous version of the Kaplan-Meier estimator for the survival function of the censoring times, and $\hat{S}(u)$ is the Kaplan-Meier estimator for the survival function $S(u) = \Pr(T \ge u)$. We state the consistency and asymptotic convergence properties of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ in two theorems.

Theorem 1. *Assume that the covariates* $\mathbf{X}$ *belong to a finite set and, for any* $\tau \in Q$, $E\{\rho_\tau(\epsilon - q)\} = 0$ *has a unique solution* q_τ. *Assume the minimizer of* (2.4), $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$, *belongs to a compact set. Then as* $n \to \infty$, $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ *converges strongly to* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

The proof is briefly outlined in the Appendix. Let f denote the density of ϵ, and suppose $E(\epsilon^2) < \infty$.

Theorem 2. *Under the conditions of Theorem* 1, *assume further that* $E(\mathbf{X}) = 0$, $\boldsymbol{\Sigma}_X = E(\mathbf{X}\mathbf{X}^T)$ *is positive definite,* $f(q_\tau) > 0$, $f(q_\tau + \delta) - f(q_\tau) = o(1)$ *as* $\delta \to 0$ *for any* $\tau \in Q$. *Then, as* $n \to \infty$, $\sqrt{n}\{(\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\beta}}^T)^T - (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T\} \to N\{0, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^T\}$ *in distribution, where*

$$\begin{aligned}
\mathbf{A} &= \frac{\partial E\{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\}}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)}, \qquad (3.1)\\
\mathbf{B} &= \mathbf{B}_1 + \mathbf{B}_2 = E\{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})^{\otimes 2}\}\\
&\quad + E\left[\int_0^L \frac{\{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbf{F}(\mathbf{s}, u)\}^{\otimes 2}}{G(u)^2} \lambda^c(u) R_i(u) du\right],\\
\mathbf{F}(\mathbf{s}, u) &= \frac{1}{S(u)} E\{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) I(T_i \geq u)\}.
\end{aligned}$$

The explicit forms of $\mathbf{A}$ and $\mathbf{B}_1$ are given in the Appendix. We can approximate $\mathbf{A}$ and $\mathbf{B}_1$ by using the empirical sample averages evaluated at the parameter estimates,

$$\hat{\mathbf{A}} = \frac{\partial}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)} \left\{ n^{-1} \sum_{i=1}^n \frac{\Delta_i \mathbf{s}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})}{\hat{G}(Y_i)} \right\} \quad \text{and} \quad \hat{\mathbf{B}}_1 = n^{-1} \sum_{i=1}^n \left\{ \frac{\Delta_i \mathbf{s}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{\otimes 2}}{\hat{G}(Y_i)} \right\}.$$

Note that $\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$ contains T_i, thus we have

$$\begin{aligned}
\mathbf{B}_2 &= E\left[\int_0^L \frac{\{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbf{F}(\mathbf{s}, u)\}^{\otimes 2}}{G(u)^2} \lambda^c(u) I(T_i \geq u) I(C_i \geq u) du\right]\\
&= E\Bigg\{\int_0^L \frac{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})^{\otimes 2} - \mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{F}(\mathbf{s}, u)^T - \mathbf{F}(\mathbf{s}, u)\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})^T + \mathbf{F}(\mathbf{s}, u)^{\otimes 2}}{G(u)}\\
&\quad \times \lambda^c(u) I(T_i \geq u) du\Bigg\}\\
&= \int_0^L \frac{S(u)\{\mathbf{F}(\mathbf{s}^{\otimes 2}, u) - \mathbf{F}(\mathbf{s}, u)^{\otimes 2}\}}{G(u)} \lambda^c(u) du\\
&= E\left[\int_0^L \frac{\{\mathbf{F}(\mathbf{s}^{\otimes 2}, u) - \mathbf{F}(\mathbf{s}, u)^{\otimes 2}\}}{G(u)} \lambda^c(u) I(T_i \geq u) du\right]
\end{aligned}$$

$$= E\left[\int_0^L \frac{\{\mathbf{F}(\mathbf{s}^{\otimes 2}, u) - \mathbf{F}(\mathbf{s}, u)^{\otimes 2}\}}{G(u)^2} \lambda^c(u) R_i(u) du\right],$$

where $\mathbf{F}(\mathbf{s}^{\otimes 2}, u) = (1/S(u))E\left\{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})^{\otimes 2} I(T_i \geq u)\right\}$. We can thus approximate $\mathbf{B}_2$ by using

$$n^{-1} \sum_{i=1}^{n} \int_0^L \frac{dN_i^c(u)}{\hat{G}(u)^2} \{\hat{\mathbf{F}}(\mathbf{s}^{\otimes 2}, u) - \hat{\mathbf{F}}(\mathbf{s}, u)^{\otimes 2}\}$$
$$= n^{-1} \sum_{i=1}^{n} \frac{1 - \Delta_i}{\hat{G}(Y_i)^2} \{\hat{\mathbf{F}}(\mathbf{s}^{\otimes 2}, Y_i) - \hat{\mathbf{F}}(\mathbf{s}, Y_i)^{\otimes 2}\},$$

where $\hat{\mathbf{F}}(\mathbf{s}, u)$ is the empirical estimate of $\mathbf{F}$ evaluated at $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$,

$$\hat{\mathbf{F}}(\mathbf{s}, u) = \frac{1}{n\hat{S}(u^-)} \sum_{i=1}^{n} \frac{\Delta_i \mathbf{s}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) I(Y_i \geq u)}{\hat{G}(Y_i)},$$

and $\hat{\mathbf{F}}(\mathbf{s}^{\otimes 2}, u)$ represents $\hat{\mathbf{F}}(\mathbf{s}, u)$ but with $\mathbf{s}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ replaced by $\mathbf{s}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{\otimes 2}$. However, the final asymptotic variance does not simplify, and it does not exhibit a clear separation between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

If some covariates in the regression model, say $\mathbf{Z}$, are precisely measured, we then need to modify the objective function to

$$\Psi_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\hat{G}(Y_i)} \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log Y_i - \alpha_\tau - \boldsymbol{\gamma}^T \mathbf{Z}_i - \boldsymbol{\beta}^T \mathbf{W}_i}{\sqrt{1 + |\boldsymbol{\beta}|^2}}\right). \quad (3.2)$$

The minimizer $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ of (3.2) is consistent, and similar arguments can be applied to derive its large-sample properties. If there exists a validation set of size n_1 where the true covariates, say $(\mathbf{X}_1, \ldots, \mathbf{X}_{n_1})$ are observed, then the objective function is modified to

$$\Psi_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = n^{-1} \sum_{i=1}^{n_1} \frac{\Delta_i}{\hat{G}(Y_i)} \sum_{\tau \in Q} \omega_\tau \rho_\tau (\log Y_i - \alpha_\tau - \boldsymbol{\gamma}^T \mathbf{Z}_i - \boldsymbol{\beta}^T \mathbf{X}_i)$$
$$+ n^{-1} \sum_{i=n_1+1}^{n} \frac{\Delta_i}{\hat{G}(Y_i)} \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log Y_i - \alpha_\tau - \boldsymbol{\gamma}^T \mathbf{Z}_i - \boldsymbol{\beta}^T \mathbf{W}_i}{\sqrt{1 + |\boldsymbol{\beta}|^2}}\right).$$

4. Averaging Estimation

In the median regression with $\tau = 1/2$, $q_\tau = 0$, we can directly obtain consistent estimators for both α and $\boldsymbol{\beta}$. However, for τ other than $1/2$, although $\hat{\alpha}_\tau$ is a consistent estimate of α_τ, it has a bias of $q_\tau \sqrt{1 + |\boldsymbol{\beta}|^2}$ as an estimator of α. To correct for the bias, a simple procedure is to also include $1-\tau$ in Q, and then average the resulting estimators to obtain $\hat{\alpha} = (\hat{\alpha}_\tau + \hat{\alpha}_{1-\tau})/2$, since $q_\tau + q_{1-\tau} = 0$ for spherically symmetric distributions.

The composite quantile regression model (2.4) explicitly uses a set of quantile values and constructs the objective function by summing the individual quantile regression models. Intuitively, it may yield a more efficient estimator than, say, the usual median regression. However, when several quantiles are modeled simultaneously in (2.4), more unknown parameters are introduced in $\boldsymbol{\alpha}$. Minimizing a single objective function involving many unknown parameters typically imposes numerical difficulties and thus the efficiency gain is often not observed in practice. Furthermore, when the bootstrap method is used to estimate the variance, the result might be unstable under the composite model (2.4). Thus, we propose to first take into consideration different regression quantiles separately, and then combine the parameter estimates by an averaging estimation approach. That is, we carry out a single quantile regression for each chosen quantile $\tau \in Q$ and average the resulting parameter estimates. This algorithm is straightforward and inherits the model averaging estimation flavor (for example, Hoeting et al. (1999); Yang (2003)).

The intuition behind the averaging estimation can be explained as follows. When carrying out a τth quantile regression, our focus is the covariate effect on the τth quantile of the survival time. The observations close to the τth quantile are more relevant, because a small shift in the quantile value, say from τ to τ^*, would result in dramatic changes in the contributions of the nearby data to the objective function; the data lying between the τth and τ^*th quantiles would flip their signs in the estimation function. However, such a quantile shift would not have much influence on other observations, because their ranks stay the same and their specific values are not essential. Thus when several quantile levels are modeled simultaneously, observations contribute to the estimation differently according to the fitted quantiles, and the information in the data can be used more effectively when the parameter estimates are combined across these quantiles. For a better understanding, we compare median and mean regression. Median regression is directly affected by the observations close to the regression line, while it is less sensitive to observations that are far away; mean regression depends equally on all the data, and even a single outlier pulls the regression curve away. As a tradeoff, mean regression typically yields a more efficient estimator than a median regression. Based on a collection of quantiles, we allow several different clusters of data around preselected regression quantiles to play a more important role in the estimation, hence provide a better use of the data and produce a more efficient estimator for the common slope. It can be viewed as a middle-ground estimator between median and mean regression.

5. Simulation

We conducted simulation studies to examine the performance of the proposed estimator in quantile regression. We generated data from model (2.1) with two true covariates X_1 and X_2, where X_1 and X_2 were simulated from uniform distributions on $[0, 1]$ and $[0, 2]$, respectively. We set the true parameter values to be $\alpha = -1.5$, $\beta_1 = 0.5$, and $\beta_2 = 0.5$. The common distribution of the model error ϵ and the measurement errors (U_1, U_2) was normal with a zero mean and a standard deviation of 0.2. We took the observed mismeasured covariates $W_1 = X_1 + U_1$ and $W_2 = X_2 + U_2$. The censoring time was generated from a mixture of a point mass at infinity and an exponential distribution with mean 1/3. By adjusting the mixing percentage, we could achieve a censoring level that was either light ($20 \sim 25\%$) or heavy ($35 \sim 40\%$). For each data realization, we carried out the median regression by taking $\tau = 1/2$. We estimated the variance of the estimator using the bootstrap method with 1,000 bootstrap samples. The asymptotic approximation of the variance given in Theorem 2 typically requires a very large sample size to exhibit its relevance due to its dependence on the hazard of ϵ, while the bootstrap performs well for sample sizes of practical use. In fact, the bootstrap often outperforms asymptotic approximation in variance estimation, and this is typical in the quantile regression framework. We conducted 1,000 simulations for different sample sizes, $n = 100, 200$, and 300.

The simulation results, presented in Table 1, show that the parameter estimates were approximately unbiased, and that the biases decreased as the sample size increased. The variance estimates using the bootstrap method were reasonably close to the empirical variances, and the variance increased as the censoring increased. The corresponding coverage probabilities of the 95% confidence intervals were close to the nominal level. We also implemented two naive estimators where the measurement error was completely ignored in one and the censoring was completely ignored in the other. Thus, in one estimator, we treated $\mathbf{W}$ as $\mathbf{X}$, and disregarded the $\sqrt{1 + |\boldsymbol{\beta}|^2}$ term from the denominator of the check function. In the other estimator, we ignored $\hat{G}(\cdot)$ in the objective function. As shown in the lower panels of Table 1, both naive estimators were biased. When ignoring measurement errors, as the sample size increased, the bootstrap estimated standard error decreased while the bias stayed approximately the same, and thus the 95% coverage probability deteriorated. When ignoring the censoring, the bias was more visible when the censoring rate was relatively high. The bias in the latter estimator was less than that seen when ignoring measurement errors. This is intuitive since the censoring time was simulated from a mixture of a point mass at infinity and an exponential distribution, the missingness caused by censoring was close to the case of missing completely at random.

Table 1. Simulation results using median regression ($\tau = 1/2$) with measurement errors. The true parameter values are in parentheses.

		20 ∼ 25% censoring			35 ∼ 40% censoring		
n	Estimate	$\alpha(-1.5)$	$\beta_1(0.5)$	$\beta_2(0.5)$	$\alpha(-1.5)$	$\beta_1(0.5)$	$\beta_2(0.5)$
				Proposed estimator			
100	Mean	−1.5047	0.5090	0.4997	−1.5082	0.5105	0.5026
	Emp sd	0.1000	0.1392	0.0635	0.1102	0.1624	0.0762
	Est se	0.1143	0.1654	0.0723	0.1351	0.2029	0.0846
	95% cv	0.9660	0.9710	0.9660	0.9700	0.9700	0.9690
200	Mean	−1.4976	0.4999	0.4988	−1.5053	0.5080	0.5019
	Emp sd	0.0692	0.1007	0.0456	0.0751	0.1113	0.0518
	Est se	0.0739	0.1067	0.0479	0.0820	0.1207	0.0546
	95% cv	0.9510	0.9560	0.9460	0.9550	0.9640	0.9490
300	Mean	−1.4998	0.5004	0.4995	−1.5029	0.5059	0.4993
	Emp sd	0.0556	0.0804	0.0367	0.0623	0.0911	0.0408
	Est se	0.0594	0.0860	0.0384	0.0654	0.0970	0.0432
	95% cv	0.9500	0.9610	0.9600	0.9510	0.9570	0.9550
				Naive estimator ignoring measurement errors			
100	Mean	−1.3654	0.3410	0.4449	−1.3694	0.3427	0.4486
	Emp sd	0.0797	0.0914	0.0536	0.0852	0.1068	0.0642
	Est se	0.0866	0.1040	0.0600	0.0967	0.1187	0.0683
	95% cv	0.6430	0.6630	0.8420	0.7280	0.7340	0.8690
200	Mean	−1.3644	0.3401	0.4458	−1.3710	0.3452	0.4483
	Emp sd	0.0557	0.0674	0.0388	0.0617	0.0764	0.0442
	Est se	0.0589	0.0703	0.0409	0.0654	0.0804	0.0465
	95% cv	0.3560	0.3670	0.7210	0.4830	0.4850	0.8000
300	Mean	−1.3666	0.3393	0.4469	−1.3675	0.3414	0.4459
	Emp sd	0.0452	0.0536	0.0326	0.0506	0.0633	0.0348
	Est se	0.0475	0.0573	0.0330	0.0529	0.0648	0.0375
	95% cv	0.1880	0.1980	0.6210	0.2870	0.3190	0.6870
				Naive estimator ignoring censoring			
100	Mean	−1.5120	0.5121	0.5003	−1.5208	0.5131	0.5039
	Emp sd	0.1009	0.1410	0.0633	0.1134	0.1628	0.0697
	Est se	0.1138	0.1647	0.0722	0.1335	0.2005	0.0849
	95% cv	0.9620	0.9670	0.9690	0.9700	0.9720	0.9680
200	Mean	−1.5032	0.5002	0.4996	−1.5141	0.5074	0.5004
	Emp sd	0.0694	0.1007	0.0458	0.0782	0.1159	0.0500
	Est se	0.0738	0.1068	0.0480	0.0834	0.1247	0.0540
	95% cv	0.9530	0.9570	0.9390	0.9600	0.9610	0.9610
300	Mean	−1.5054	0.5005	0.5000	−1.5119	0.5016	0.5001
	Emp sd	0.0562	0.0805	0.0368	0.0633	0.0925	0.0411
	Est se	0.0594	0.0861	0.0384	0.0669	0.0988	0.0439
	95% cv	0.9520	0.9590	0.9580	0.9630	0.9500	0.9530

Empirical standard deviation (Emp sd), average of the estimated standard errors (Est se) and coverage probability of 95% confidence intervals (95% cv).

Table 2. Simulation results using τ =(1/4, 1/2, 3/4) quantile regression with measurement errors.

		20 $\sim$ 25% censoring			35 $\sim$ 40% censoring		
n	Estimate	$\alpha(-1.5)$	$\beta_1(0.5)$	$\beta_2(0.5)$	$\alpha(-1.5)$	$\beta_1(0.5)$	$\beta_2(0.5)$
		Proposed estimator					
100	Mean	−1.5015	0.5071	0.4991	−1.5070	0.5156	0.5012
	Emp sd	0.0886	0.1243	0.0557	0.0977	0.1436	0.0669
	Est se	0.1027	0.1471	0.0617	0.1244	0.1827	0.0731
	95% cv	0.9670	0.9640	0.9610	0.9640	0.9670	0.9550
200	Mean	−1.4986	0.5026	0.4986	−1.5052	0.5080	0.5017
	Emp sd	0.0603	0.0867	0.0388	0.0665	0.0969	0.0448
	Est se	0.0647	0.0923	0.0409	0.0724	0.1052	0.0464
	95% cv	0.9590	0.9510	0.9570	0.9580	0.9730	0.9500
300	Mean	−1.5018	0.5028	0.5008	−1.4997	0.5031	0.4984
	Emp sd	0.0505	0.0712	0.0319	0.0556	0.0806	0.0350
	Est se	0.0522	0.0750	0.0329	0.0574	0.0835	0.0369
	95% cv	0.9440	0.9600	0.9620	0.9550	0.9510	0.9540
		Naive estimator ignoring measurement errors					
100	Mean	−1.3653	0.3397	0.4453	−1.3685	0.3434	0.4466
	Emp sd	0.0689	0.0796	0.0477	0.0732	0.0922	0.0555
	Est se	0.0723	0.0868	0.0501	0.0811	0.0992	0.0570
	95% cv	0.5160	0.5370	0.7940	0.6350	0.6310	0.8260
200	Mean	−1.3644	0.3400	0.4453	−1.3699	0.3432	0.4482
	Emp sd	0.0485	0.0583	0.0339	0.0531	0.0649	0.0379
	Est se	0.0498	0.0599	0.0346	0.0549	0.0675	0.0388
	95% cv	0.2210	0.2420	0.6260	0.3270	0.3630	0.7180
300	Mean	−1.3655	0.3373	0.4467	−1.3657	0.3403	0.4453
	Emp sd	0.0384	0.0455	0.0272	0.0429	0.0531	0.0301
	Est se	0.0401	0.0485	0.0280	0.0441	0.0543	0.0312
	95% cv	0.0770	0.0730	0.5190	0.1460	0.1700	0.5780
		Naive estimator ignoring censoring					
100	Mean	−1.5086	0.5097	0.4995	−1.5259	0.5202	0.5069
	Emp sd	0.0893	0.1263	0.0552	0.1031	0.1443	0.0620
	Est se	0.1047	0.1510	0.0623	0.1251	0.1872	0.0745
	95% cv	0.9600	0.9650	0.9640	0.9720	0.9750	0.9720
200	Mean	−1.5045	0.5032	0.4993	−1.5186	0.5132	0.5028
	Emp sd	0.0603	0.0871	0.0392	0.0714	0.1065	0.0430
	Est se	0.0650	0.0930	0.0410	0.0739	0.1087	0.0465
	95% cv	0.9580	0.9500	0.9580	0.9598	0.9656	0.9541
300	Mean	−1.5073	0.5037	0.5013	−1.5140	0.5051	0.5008
	Emp sd	0.0500	0.0699	0.0319	0.0561	0.0825	0.0352
	Est se	0.0523	0.0755	0.0330	0.0587	0.0858	0.0374
	95% cv	0.9529	0.9609	0.9632	0.9450	0.9450	0.9570

To illustrate the averaging estimation procedure, we performed a $\tau = (1/4, 1/2, 3/4)$ quantile regression on the same 1,000 simulated data sets. We first fit the three individual quantile regression models corresponding to each value of τ. Then, we took the average of the parameter estimates: $\bar{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_{\tau=1/4} + \hat{\boldsymbol{\beta}}_{\tau=1/2} + \hat{\boldsymbol{\beta}}_{\tau=3/4})/3$ and $\bar{\alpha} = (\hat{\alpha}_{\tau=1/4} + \hat{\alpha}_{\tau=1/2} + \hat{\alpha}_{\tau=3/4})/3$ as the final estimators. From the results in Table 2, we can see that the biases of the averaging estimators were negligible, and were generally smaller than those under the median regression in Table 1. The bootstrap variance estimation performed well and the estimated variance approached the sample empirical variance with the increasing sample size. The coverage probabilities of the 95% confidence intervals were accurate. Of special interest, the variance estimates under the three-quantile averaging estimation method decreased approximately 10% for all the scenarios, compared to those under the median regression. We further applied a $\tau = (1/6, 1/3, 1/2, 2/3, 5/6)$ quantile regression to the same data sets, and computed the five-quantile averaging estimators. Here the improvement on estimation efficiency was not as significant; see Table 3. Although one could enlarge Q, the composite regression quantiles would become more saturated and further gain in efficiency would be less visible. Since averaging more quantile regression estimates inevitably requires more intensive computation, we recommend using either one, three, or five quantiles. For comparison, we also present the naive estimators for the three- and five-quantile averaging estimators in the lower panels of Tables 2 and 3, respectively. Overall, the naive estimators performed much worse than the proposed estimators.

6. Example

We applied the proposed method to a lung cancer biomarker expression study at M.D. Anderson Cancer Center. The objective of the study was to evaluate the effects of the risk factors on patient disease-free survival (DFS). In our analysis, there were 309 patients, and the covariates of interest were histology (adenocarcinoma=1, 62%; squamous=0, 38%), patient age (ranging from 34 to 90 with a mean of 66 years), sex (female=1, 53%; male=0, 47%), and biomarker expression. The estimated Kaplan-Meier survival curves stratified by histology are shown in Figure 1. We can see that the survival curves cross, which often suggests the violation of the proportional hazards assumption.

For each patient, there were two reading scores of biomarker expression from two different junior research fellows in consideration of the possible measurement errors. In addition, there was a random validation set of 28 biomarker expression scores obtained from the reading by a senior pathologist, whose readings were considered error-free. We used the average score of the readings from the

Table 3. Simulation results using τ =(1/6, 1/3, 1/2, 2/3, 5/6) quantile regression with measurement errors.

		20 $\sim$ 25% censoring			35 $\sim$ 40% censoring		
n	Estimate	$\alpha(-1.5)$	$\beta_1(0.5)$	$\beta_2(0.5)$	$\alpha(-1.5)$	$\beta_1(0.5)$	$\beta_2(0.5)$
		Proposed estimator					
100	Mean	−1.5035	0.5094	0.4993	−1.5062	0.5157	0.5006
	Emp sd	0.0876	0.1218	0.0551	0.0968	0.1425	0.0646
	Est se	0.1078	0.1683	0.0626	0.1411	0.2277	0.0791
	95% cv	0.9610	0.9730	0.9650	0.9760	0.9750	0.9680
200	Mean	−1.4990	0.5034	0.4988	−1.5061	0.5081	0.5020
	Emp sd	0.0595	0.0854	0.0377	0.0671	0.0948	0.0436
	Est se	0.0636	0.0928	0.0395	0.0730	0.1116	0.0453
	95% cv	0.9590	0.9550	0.9580	0.9540	0.9690	0.9460
300	Mean	−1.5018	0.5026	0.5005	−1.4998	0.5023	0.4983
	Emp sd	0.0488	0.0694	0.0308	0.0540	0.0782	0.0339
	Est se	0.0509	0.0733	0.0317	0.0564	0.0830	0.0356
	95% cv	0.9530	0.9640	0.9590	0.9550	0.9560	0.9560
		Naive estimator ignoring measurement errors					
100	Mean	−1.3662	0.3408	0.4454	−1.3689	0.3440	0.4468
	Emp sd	0.0679	0.0773	0.0463	0.0702	0.0897	0.0530
	Est se	0.0693	0.0832	0.0479	0.0775	0.0949	0.0545
	95% cv	0.5050	0.5170	0.7870	0.6130	0.6290	0.8240
200	Mean	−1.3645	0.3402	0.4452	−1.3690	0.3411	0.4483
	Emp sd	0.0469	0.0564	0.0331	0.0512	0.0635	0.0368
	Est se	0.0477	0.0574	0.0331	0.0527	0.0647	0.0372
	95% cv	0.1790	0.2010	0.6160	0.2880	0.3240	0.6920
300	Mean	−1.3655	0.3379	0.4466	−1.3656	0.3395	0.4453
	Emp sd	0.0374	0.0437	0.0263	0.0420	0.0522	0.0291
	Est se	0.0386	0.0467	0.0268	0.0425	0.0523	0.0299
	95% cv	0.0670	0.0610	0.4710	0.1230	0.1470	0.5370
		Naive estimator ignoring censoring					
100	Mean	−1.5094	0.5098	0.5001	−1.5250	0.5199	0.5064
	Emp sd	0.0876	0.1232	0.0556	0.1014	0.1406	0.0604
	Est se	0.1045	0.1599	0.0617	0.1224	0.1761	0.0704
	95% cv	0.9560	0.9720	0.9650	0.9670	0.9730	0.9660
200	Mean	−1.5052	0.5036	0.4998	−1.5146	0.5078	0.5009
	Emp sd	0.0594	0.0851	0.0379	0.0693	0.0995	0.0418
	Est se	0.0632	0.0909	0.0394	0.0706	0.1031	0.0444
	95% cv	0.9520	0.9500	0.9560	0.9610	0.9620	0.9620
300	Mean	−1.5077	0.5032	0.5012	−1.5127	0.5037	0.5003
	Emp sd	0.0488	0.0695	0.0307	0.0545	0.0778	0.0337
	Est se	0.0507	0.0729	0.0318	0.0567	0.0820	0.0359
	95% cv	0.9510	0.9630	0.9610	0.9550	0.9540	0.9590

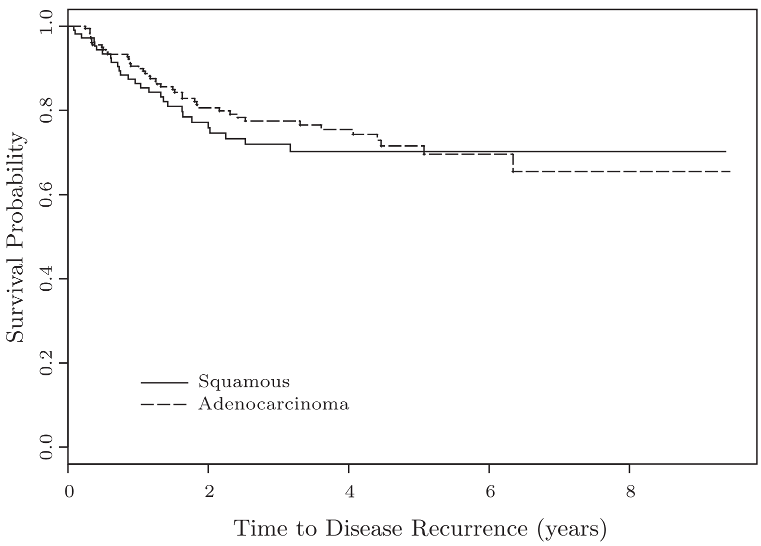

Figure 1. Estimated Kaplan-Meier survival curves for the lung cancer data stratified by tumor histology.

two fellows if there was no true reading. We computed the variance of the measurement error from the two separate measurements and found var(U)=0.0489. From a preliminary analysis using only the validation data set, we obtained the model error variance, var(ϵ)=0.5750. Therefore, to meet the spherical symmetry assumption, we scaled the response and error-free covariates by dividing by $\sqrt{\mathrm{var}(\epsilon)/\mathrm{var}(U)} = 3.4272$. We also tested the normal assumption on the errors obtained from the 28 validated observations, and found that it was not supported. However, there appeared to be no strong evidence against the symmetry assumption. Taking into account both errors in the biomarker measurement and in the model, we thus adopted a spherical symmetric distribution assumption after the scaling.

We took 20,000 bootstrap samples for the variance estimation and the analysis results on the original data scale are shown in Table 4. Across the three models, we can see that patient age had a significant effect on DFS, where older patients appeared to survive longer, possibly due to the slower reproduction of the cancer cells. For $\tau = 1/2$, there was no survival difference between the adenocarcinoma and squamous histology groups; the histology effect was strengthened by simultaneously modeling three or five regression quantiles, showing that patients with adenocarcinoma had better survival. There was no significant difference in survival between male and female patients, and the biomarker did not affect the survival. When the measurement error was ignored, using a naive estimator, we obtained slightly different estimates although the scientific conclusions drawn were the same. We also explored our estimator by ignoring the censoring and obtained quite different results; especially, the significance of histology was not detected (see Table 4).

For comparison, we applied the corrected score method under the Cox (1972) model to the lung cancer data (Nakamura (1990)). The model parameter estimates and standard errors for histology, log(age), sex, and biomarker expression are -0.1589 (0.2615), 0.7723 (0.6695), 0.3568 (0.2470) and -0.1003 (0.2215), respectively. Due to the violation of the proportional hazards assumption with crossing survival curves, the Cox model did not detect any risk factor that significantly affected survival.

7. Conclusion

We have proposed a censored quantile regression model with covariate measurement errors. Further extensions to nonlinear regression is possible along the lines of Wei and Carroll (2009). Considering the marginal distribution of the model error, the composite quantile regression model is natural. Our estimation procedure can be viewed as an alternative implementation of the composite quantile regression, which takes an averaging estimation approach after fitting each separate quantile regression model. The proposed method successfully eliminates the estimation bias caused by the covariate measurement errors. The efficiency gain of the averaging estimation over the usual single quantile regression estimator can be substantial, while the numerical computation is straightforward. Efficiency can also be improved through better handling of censored/missing data by augmenting the inverse probability weighted objective function (Robins, Rotnitzky and Zhao (1994)). The augmented terms consist of the censored cases, and thus the estimator uses the data more efficiently. The drawback of such augmentation approach is its computational complexity. Particularly, to achieve efficiency improvement, we need to estimate the expectation of the event time conditional on the censoring and covariates information at all the times for each censored observation, see Bang and Tsiatis (2002). Due to the numerical instability this augmentation may incur, we do not pursue this further. Although alternative approaches to handling the censored data are available, such as the imputation procedure of Buckley and James (1979), these methods are not applicable when covariates are measured with errors. This is because all these methods require calculating the survival function of model residuals, while in the presence of measurement errors, residuals cannot be obtained even if all the model parameters are known.

Acknowledgement

Ma's research is partially supported by a US NSF grant. Yin's research was partially supported by a grant from the Research Grants Council of Hong Kong.

Table 4. Analysis of the lung cancer data using censored quantile regression, with the biomarker expression measured with errors.

Estimate	Intercept	Histology	log(age)	Sex	Expression
		Proposed estimator			
		$\tau = 1/2$			
Est	−2.7645	0.1517	2.4946	0.0127	−0.1264
Est se	0.5111	0.0988	0.4420	0.1090	0.0863
p-value	<0.0001	0.1246	<0.0001	0.9075	0.1428
		$\tau = (1/4, 1/2, 3/4)$			
Est	−2.5102	0.1766	2.2222	−0.0067	−0.1117
Est se	0.3510	0.0847	0.2955	0.0900	0.0771
p-value	<0.0001	0.0370	<0.0001	0.9407	0.1471
		$\tau = (1/6, 1/3, 1/2, 2/3, 5/6)$			
Est	−2.7638	0.1789	2.3880	−0.0189	−0.0838
Est se	0.2580	0.0804	0.2281	0.0837	0.0721
p-value	<0.0001	0.0261	<0.0001	0.8212	0.2449
		Naive estimator ignoring measurement errors			
		$\tau = 1/2$			
Est	−2.7077	0.1518	2.4476	0.0100	−0.1238
Est se	0.5290	0.0984	0.4500	0.1085	0.0797
p-value	<0.0001	0.1229	<0.0001	0.9265	0.1202
		$\tau = (1/4, 1/2, 3/4)$			
Est	−2.4414	0.1920	2.1354	0.0091	−0.0947
Est se	0.3851	0.0844	0.3207	0.0894	0.0695
p-value	<0.0001	0.0230	<0.0001	0.9189	0.1729
		$\tau = (1/6, 1/3, 1/2, 2/3, 5/6)$			
Est	−2.7858	0.1670	2.3925	−0.0065	−0.0736
Est se	0.2685	0.0796	0.2325	0.0828	0.0645
p-value	<0.0001	0.0360	<0.0001	0.9378	0.2538
		Naive estimator ignoring censoring			
		$\tau = 1/2$			
Est	−1.4722	0.0571	1.2462	−0.0313	−0.0117
Est se	0.2798	0.0779	0.2402	0.0782	0.0742
p-value	<0.0001	0.4630	<0.0001	0.6892	0.8749
		$\tau = (1/4, 1/2, 3/4)$			
Est	−2.0254	0.1028	1.6952	−0.0440	−0.0354
Est se	0.2865	0.0729	0.2253	0.0761	0.0593
p-value	<0.0001	0.1581	<0.0001	0.5629	0.5509
		$\tau = (1/6, 1/3, 1/2, 2/3, 5/6)$			
Est	−2.1931	0.1015	1.8575	−0.0520	−0.0470
Est se	0.2104	0.0708	0.1713	0.0732	0.0567
p-value	<0.0001	0.1517	<0.0001	0.4777	0.4073

Appendix

Verification of using reparametrization to resolve non-spherical symmetry.

Assume $\mathbf{M}(\epsilon, \mathbf{U}^T)^T$ is spherically symmetric. We write

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{pmatrix} = \begin{pmatrix} M_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}$$

and note that $\log T = \alpha + \boldsymbol{\beta}^T \mathbf{X} + \epsilon, \mathbf{W} = \mathbf{X} + \mathbf{U}$, can be equivalently written as

$$\mathbf{M} \begin{pmatrix} \log T \\ \mathbf{W} \end{pmatrix} = \mathbf{M} \begin{pmatrix} \alpha + \boldsymbol{\beta}^T \mathbf{X} \\ \mathbf{X} \end{pmatrix} + \mathbf{M} \begin{pmatrix} \epsilon \\ \mathbf{U} \end{pmatrix}.$$

Thus, we can perform the proposed method on $\mathbf{M}(\log Y_i, \mathbf{W}_i^T)^T$ to obtain estimators for $(a, \mathbf{b})$, where $(a, \mathbf{b})$ satisfy

$$a + \mathbf{b}^T \mathbf{M}_2 \begin{pmatrix} \alpha + \boldsymbol{\beta}^T \mathbf{X} \\ \mathbf{X} \end{pmatrix} = \mathbf{M}_1 \begin{pmatrix} \alpha + \boldsymbol{\beta}^T \mathbf{X} \\ \mathbf{X} \end{pmatrix}$$

for any $\mathbf{X}$. We can subsequently solve for $(\alpha, \boldsymbol{\beta})$ to obtain $\hat{\alpha} = (M_{11} - \mathbf{M}_{21}^T \hat{\mathbf{b}})^{-1} \hat{a}$, $\hat{\boldsymbol{\beta}} = (M_{11} - \mathbf{M}_{21}^T \hat{\mathbf{b}})^{-1} (\mathbf{M}_{22}^T \hat{\mathbf{b}} - \mathbf{M}_{12})$ and use the delta method to further make inference on $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$ based on that of $(\hat{a}, \hat{\mathbf{b}})$.

Proof of Theorem 1. From the convergence property of the Kaplan-Meier estimator (Fleming and Harrington (1991)), for parameters $(\boldsymbol{a}, \mathbf{b})$ in a compact set, we have

$$\begin{aligned}
\Psi_n(\boldsymbol{a}, \mathbf{b}) &= n^{-1} \sum_{i=1}^n \frac{\Delta_i}{\hat{G}(Y_i)} \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log Y_i - a_\tau - \mathbf{b}^T \mathbf{W}_i}{\sqrt{1 + |\mathbf{b}|^2}} \right) \\
&= n^{-1} \sum_{i=1}^n \frac{\Delta_i}{G(Y_i)} \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log Y_i - a_\tau - \mathbf{b}^T \mathbf{W}_i}{\sqrt{1 + |\mathbf{b}|^2}} \right) \\
&\quad + n^{-1} \sum_{i=1}^n \frac{\Delta_i \{G(Y_i) - \hat{G}(Y_i)\}}{\hat{G}(Y_i) G(Y_i)} \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log Y_i - a_\tau - \mathbf{b}^T \mathbf{W}_i}{\sqrt{1 + |\mathbf{b}|^2}} \right) \\
&= E\left[E\left\{ \frac{\Delta_i}{G(Y_i)} \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log Y_i - a_\tau - \mathbf{b}^T \mathbf{W}_i}{\sqrt{1 + |\mathbf{b}|^2}} \right) \big| \mathbf{W}_i, Y_i \right\} \right] + o_p(1) \\
&= E \sum_{\tau \in Q} \omega_\tau \rho_\tau \left(\frac{\log T_i - a_\tau - \mathbf{b}^T \mathbf{W}_i}{\sqrt{1 + |\mathbf{b}|^2}} \right) + o_p(1) \\
&= E \sum_{\tau \in Q} \omega_\tau \rho_\tau \left\{ \frac{\epsilon_i - \mathbf{b}^T \mathbf{U}_i + (\alpha - a_\tau) + (\boldsymbol{\beta} - \mathbf{b})^T \mathbf{X}_i}{\sqrt{1 + |\mathbf{b}|^2}} \right\} + o_p(1) \\
&= E \sum_{\tau \in Q} \omega_\tau \rho_\tau \left\{ \epsilon_1 + \frac{\alpha - a_\tau + (\boldsymbol{\beta} - \mathbf{b})^T \mathbf{X}_i}{\sqrt{1 + |\mathbf{b}|^2}} \right\} + o_p(1)
\end{aligned}$$

uniformly in $(\boldsymbol{a}, \mathbf{b})$. Here, we base the arguments on the spherical symmetry property of $(\epsilon_i, \mathbf{U}_i^T)^T$, specifically $(\epsilon_i - \mathbf{b}^T\mathbf{U}_i)/\sqrt{1+|\mathbf{b}|^2}$ and ϵ_1 have the same distribution and both are independent of $\mathbf{X}_i$. Thus, we have now obtained that $\Psi_n(\boldsymbol{a}, \mathbf{b})$ converges to $E\sum_{\tau\in Q}\omega_\tau\rho_\tau\left\{\epsilon_1 + (\alpha - a_\tau + (\boldsymbol{\beta}-\mathbf{b})^T\mathbf{X}_i)/(\sqrt{1+|\mathbf{b}|^2})\right\}$. Since for $\tau \in Q$, $E\{\rho_\tau(\epsilon_1 - q)\} = 0$ has a unique solution q_τ, we have that $E\rho_\tau\left\{\epsilon_1 + (\alpha - a_\tau + (\boldsymbol{\beta}-\mathbf{b})^T\mathbf{X}_i)(\sqrt{1+|\mathbf{b}|^2})\right\}$ has a unique solution at $\mathbf{b} = \boldsymbol{\beta}$ and $a_\tau = \alpha + q_\tau\sqrt{1+|\mathbf{b}|^2} = \alpha_\tau$. Now we consider any sequence of the minimizers $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$. First, $(\hat{\alpha}_\tau - \alpha_\tau)/\sqrt{1+|\hat{\boldsymbol{\beta}}|^2}$ is bounded for any τ, otherwise, $\Psi_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ would be unbounded (note that ρ_τ are non-negative functions). We assume a subsequence of $\left\{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})/\sqrt{1+|\hat{\boldsymbol{\beta}}|^2}, (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sqrt{1+|\hat{\boldsymbol{\beta}}|^2}\right\}$ converges to $(\boldsymbol{a}_c, \mathbf{b}_c)$. Then, due to the minimization property,

$$E\sum_{\tau\in Q}\omega_\tau\rho_\tau\left(\epsilon_1 + \frac{\alpha - a_\tau + (\boldsymbol{\beta}-\mathbf{b})^T\mathbf{X}_i}{\sqrt{1+|\mathbf{b}|^2}}\right) \le E\sum_{\tau\in Q}\omega_\tau\rho_\tau\left(\epsilon_1 - c_\tau\right)$$

for an arbitrary set of c_τ's. Particularly, we let $c_\tau = q_\tau$; because of the uniqueness, we have $\boldsymbol{a}_c = \mathbf{b}_c = 0$, and the result is shown.

Proof of Theorem 2. Note that, except for the noise caused by discontinuity at the $O_p(1)$ terms, the residuals are exactly zero, the first order derivative of $\Psi_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in (2.4) is zero at the minimum. Thus, we take the derivative of $\Psi_n(\boldsymbol{\alpha}, \boldsymbol{\beta})$ with respect to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and evaluate it at $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$; as long as the regression curve passes only finitely many observations, we have

$$\begin{aligned}
O_p(1) = n\frac{\partial\Psi_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T} &= \sum_{i=1}^n \frac{\Delta_i}{\hat{G}(Y_i)}\mathbf{s}_i(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \\
&= \sum_{i=1}^n \frac{\Delta_i}{\hat{G}(Y_i)}\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \sum_{i=1}^n \frac{\partial}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)}E\left\{\frac{\Delta_i\mathbf{s}_i(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)}{\hat{G}(Y_i)}\right\}\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\end{pmatrix} \\
&= \sum_{i=1}^n \mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \sum_{i=1}^n\left\{\frac{\Delta_i}{G(Y_i)} - 1\right\}\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \sum_{i=1}^n \frac{\Delta_i\{G(Y_i) - \hat{G}(Y_i)\}}{G(Y_i)\hat{G}(Y_i)}\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\
&\quad + n\left[\frac{\partial E\{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\}}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)} + o_p(1)\right]\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\end{pmatrix}, \qquad \text{(A.1)}
\end{aligned}$$

where $(\boldsymbol{\alpha}^{*T}, \boldsymbol{\beta}^{*T})^T$ lies on the line segment between $(\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\beta}}^T)^T$ and $(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$. Then (A.1) holds uniformly for $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ in a compact support. From the definition of $\mathcal{F}(u)$, $N^c(u)$, $R(u)$, $\mathcal{M}^c(u)$, $G(u)$, and $S(u)$, using a martingale integral representation (Gill (1980, p.37)), we have

$$\frac{G(t)-\hat{G}(t)}{G(t)} = \int_0^t \frac{\hat{G}(u^-)}{G(u)} \frac{d\mathcal{M}^c(u)}{R(u)} = \int_0^L \frac{I(t \geq u)\hat{G}(u^-)}{G(u)} \frac{d\mathcal{M}^c(u)}{R(u)}$$
$$= \int_0^L \frac{I(t \geq u)}{nG(u)\hat{S}(u^-)} d\mathcal{M}^c(u).$$

Let $\tilde{\mathbf{F}}(\mathbf{s}, u)$ be $\hat{\mathbf{F}}(\mathbf{s}, u)$ evaluated at $(\boldsymbol{\alpha}, \boldsymbol{\beta})$,

$$\tilde{\mathbf{F}}(\mathbf{s}, u) = \frac{1}{n\hat{S}(u^-)} \sum_{i=1}^n \frac{\Delta_i \mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) I(Y_i \geq u)}{\hat{G}(Y_i)},$$

we obtain

$$n^{-1/2} \sum_{i=1}^n \frac{\Delta_i \{G(Y_i) - \hat{G}(Y_i)\}}{G(Y_i)\hat{G}(Y_i)} \mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$$
$$= n^{-1/2} \sum_{i=1}^n \frac{\Delta_i \mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\hat{G}(Y_i)} \int_0^L \frac{I(Y_i \geq u)}{nG(u)\hat{S}(u^-)} d\mathcal{M}^c(u)$$
$$= n^{-1/2} \int_0^L \frac{\tilde{\mathbf{F}}(\mathbf{s}, u)}{G(u)} d\mathcal{M}^c(u) = n^{-1/2} \sum_{i=1}^n \int_0^L \frac{\mathbf{F}(\mathbf{s}, u)}{G(u)} d\mathcal{M}_i{}^c(u) + o_p(1).$$

Using the property (Robins and Rotnitzky (1992, p.313))

$$\frac{\Delta_i}{G(Y_i)} = 1 - \int_0^L \frac{d\mathcal{M}_i^c(u)}{G(u)},$$

we have

$$n^{-1/2} \sum_{i=1}^n \left\{ \frac{\Delta_i}{G(Y_i)} - 1 \right\} \mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = -n^{-1/2} \sum_{i=1}^n \int_0^L \frac{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})}{G(u)} d\mathcal{M}_i^c(u).$$

Inserting this into (A.1), we have

$$-\left[\frac{\partial E\{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\}}{\partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)} + o_p(1) \right] \left\{ n^{1/2} \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \end{pmatrix} \right\}$$
$$= n^{-1/2} \sum_{i=1}^n \mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) - n^{-1/2} \sum_{i=1}^n \int_0^L \frac{\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbf{F}(\mathbf{s}, u)}{G(u)} d\mathcal{M}_i^c(u) + o_p(1).$$

Because $\mathbf{s}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is $\mathcal{F}(0)$ measurable, the two terms on the right side here are uncorrelated. Thus, we have that $n^{1/2}(\hat{\boldsymbol{\alpha}}^T - \boldsymbol{\alpha}^T, \hat{\boldsymbol{\beta}}^T - \boldsymbol{\beta}^T)^T \sim N(0, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^T)$, where $\mathbf{A}$ and $\mathbf{B}$ are given in (3.1). Note that the form of $\mathbf{B}_2$ is a result of the Martingale Central Limit Theorem.

Calculation of $\mathbf{A}$ *and* $\mathbf{B}_1$ *in Theorem* 2.

Write

$$\xi_{i\tau} = \frac{\log Y_i - a - \mathbf{b}^T\mathbf{W}_i}{\sqrt{1+|\mathbf{b}|^2}}$$
$$= \frac{\epsilon_i - \mathbf{b}^T\mathbf{U}_i}{\sqrt{1+|\mathbf{b}|^2}} - \left\{ \frac{\sqrt{1+|\boldsymbol{\beta}|^2}}{\sqrt{1+|\mathbf{b}|^2}} q_\tau + \frac{a-\alpha_\tau}{\sqrt{1+|\mathbf{b}|^2}} + \frac{(\mathbf{b}-\boldsymbol{\beta})^T\mathbf{X}_i}{\sqrt{1+|\mathbf{b}|^2}} \right\} = \epsilon_1 - c_{i\tau}.$$

It can be verified that

$$\mathbf{s}_i(\boldsymbol{a},\mathbf{b}) = \begin{bmatrix} \frac{\omega_{\tau_1}\{I(\xi_{i\tau_1}\geq 0)-1+\tau_1\}}{-\sqrt{1+|\mathbf{b}|^2}} \\ \vdots \\ \frac{\omega_{\tau_k}\{I(\xi_{i\tau_k}\geq 0)-1+\tau_k\}}{-\sqrt{1+|\mathbf{b}|^2}} \\ \sum_{\tau\in Q}\omega_\tau \frac{I(\xi_{i\tau}\geq 0)-1+\tau}{-\sqrt{1+|\mathbf{b}|^2}} \left(\mathbf{W}_i + \frac{\xi_{i\tau}\mathbf{b}}{\sqrt{1+|\mathbf{b}|^2}}\right) \end{bmatrix},$$

$$E\{\mathbf{s}_i(\boldsymbol{a},\mathbf{b})\} = \begin{pmatrix} \frac{\omega_{\tau_1}[\tau_1 - E\{F(c_{i\tau_1})\}]}{-\sqrt{1+|\mathbf{b}|^2}} \\ \vdots \\ \frac{\omega_{\tau_k}[\tau_k - E\{F(c_{i\tau_k})\}]}{-\sqrt{1+|\mathbf{b}|^2}} \\ \sum_{\tau\in Q}\omega_\tau \left[\frac{E\{\mathbf{X}_iF(c_{i\tau})\}}{\sqrt{1+|\mathbf{b}|^2}} + \frac{\mathbf{b}E(c_{i\tau})\tau - \mathbf{b}E\{c_{i\tau}F(c_{i\tau})\}}{1+|\mathbf{b}|^2}\right] \end{pmatrix},$$

where $F(\cdot)$ is the cumulative distribution function of ϵ_1. Take the derivative of $E\{\mathbf{s}_i(\boldsymbol{a},\mathbf{b})\}$ with respect to $(\boldsymbol{a},\mathbf{b})$ and evaluate it at $(\boldsymbol{\alpha},\boldsymbol{\beta})$, calculate the expectation of $\mathbf{s}_i(\boldsymbol{\alpha},\boldsymbol{\beta})^{\otimes 2}$, denoting the variance of ϵ_1 as σ_ϵ^2, to get

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{21}^T \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{21}^T \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix},$$

where

$$\mathbf{A}_{11} = \frac{1}{1+|\boldsymbol{\beta}|^2}\mathrm{diag}\left\{\omega_{\tau_1}f(q_{\tau_1}),\ldots,\omega_{\tau_k}f(q_{\tau_k})\right\},$$
$$\mathbf{A}_{21} = \frac{-\boldsymbol{\beta}}{(1+|\boldsymbol{\beta}|^2)^{3/2}}\left\{\omega_{\tau_1}q_{\tau_1}f(q_{\tau_1}),\ldots,\omega_{\tau_k}q_{\tau_k}f(q_{\tau_k})\right\},$$
$$\mathbf{A}_{22} = \sum_{\tau\in Q}\frac{\omega_\tau f(q_\tau)}{1+|\boldsymbol{\beta}|^2}\left\{\boldsymbol{\Sigma}_X + \frac{q_\tau^2\boldsymbol{\beta}\boldsymbol{\beta}^T}{1+|\boldsymbol{\beta}|^2}\right\},$$

$$\mathbf{B}_{11} = \frac{1}{1+|\boldsymbol{\beta}|^2}\left\{\begin{matrix} \omega_1^2\tau_1(1-\tau_1) & \omega_1\omega_2\tau_1(1-\tau_2) & \cdots & \omega_1\omega_k\tau_1(1-\tau_k) \\ \omega_1\omega_2\tau_1(1-\tau_2) & \omega_2^2\tau_2(1-\tau_2) & \cdots & \omega_2\omega_k\tau_2(1-\tau_k) \\ & & \vdots & \\ \omega_1\omega_k\tau_1(1-\tau_k) & \omega_2\omega_k\tau_2(1-\tau_k) & \cdots & \omega_k^2\tau_k(1-\tau_k) \end{matrix}\right\},$$

$$\mathbf{B}_{21} = \frac{-\boldsymbol{\beta}}{(1+|\boldsymbol{\beta}|^2)^{3/2}}\Big[\omega_{\tau_1}\sum_{\tau}\omega_\tau q_\tau(\tau_1-\tau\tau_1),\ldots,$$
$$\omega_{\tau_j}\sum_{\tau\in Q}\omega_\tau q_\tau\{\min(\tau,\tau_j)-\tau\tau_j\},\ldots,\omega_{\tau_k}\sum_{\tau}\omega_\tau q_\tau(\tau-\tau\tau_k)\Big]$$

$$\mathbf{B}_{22} = \sum_{\tau,\tau'\in Q}\omega_\tau\omega_{\tau'}\frac{\{\min(\tau,\tau')-\tau\tau'\}}{1+|\boldsymbol{\beta}|^2}$$
$$\times\left\{\boldsymbol{\Sigma}_X + \frac{q_\tau q_{\tau'}}{1+|\boldsymbol{\beta}|^2}\boldsymbol{\beta}\boldsymbol{\beta}^T + (1+|\boldsymbol{\beta}|^2)\sigma_\epsilon^2(\mathbf{I}+\boldsymbol{\beta}\boldsymbol{\beta}^T)^{-2}\right\}.$$

References

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

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

Brown, M. L. (1982). Robust line estimation with error in both variables. *J. Amer. Statist. Assoc.* **77**, 71-79. Correction in **78**, 1008.

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.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Second Edition. CRC Press, London.

Chernozhukov, V. and Hong, H. (2002). Three-step censored quantile regression and extramarital affairs, *J. Amer. Statist. Assoc.* **97**, 872-882.

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

Fitzenberger, B. (1997). A guide to censored quantile regressions. In *Handbook of Statistics, Robust Inference* (Edited by G. S. Maddala and C. R. Rao). North-Holland, Amsterdam.

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

Gill, R. D. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts **124**. Mathematisch Centrum, Amsterdam.

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

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statist. Sci.* **14**, 382-401.

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.

Khan, S. and Powell L. J. (2001). Two-step estimation of semiparametric censored regression models. *J. Econometrics* **103**, 73-110.

Koenker, R. (2005). *Quantile Regression.* 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. *J. Amer. Statist. Assoc.* **96**, 458-468.

Lai, T. L. and Ying, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531-556.

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

Lindley, D. B. (1947). Regression lines and the functional relationship. *J. Roy. Statist. Soc. Ser. B* **9**, 219-244.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika* **77**, 127-137.

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.

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. *Ann. Statist.* **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.* (Edited by N. P. Jewell et al.), 263-274. Kluwer Academic Publishers.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology - Methodological Issues.* (Jewell N., Dietz K., Farewell V. eds.). 297-331, Birkh 酶ser, Boston.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

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

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.

Wei, Y. and Carroll, R. J. (2009). Quantile regression with measurement error. *J. Amer. Statist. Assoc.* **104**, 1129-1143.

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

Yang, Y. (2003). Regression with multiple candidate models: Selecting or mixing? *Statist. Sinica* **13**, 783-809.

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

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36**, 1108-1126.

Department of Statistics, Texas A&M University, College Station, TX 77845, USA.

E-mail: ma@stat.tamu.edu

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: gyin@hku.hk

(Received April 2009; accepted October 2009)