A Note On Self-Consistent Estimation of Censored Quantile Regression

by

Limin Peng

Technical Report 09-03
October 2009

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
1518 Clifton Road, N.E.
Emory University
Atlanta, Georgia 30322

Telephone: (404) 727-7701    FAX: (404) 727-1370
e-mail: lpeng@emory.edu
A Note On Self-Consistent Estimation of Censored Quantile Regression

Limin Peng

Abstract

Censored quantile regression has recently been studied under two major approaches: one is based on Portnoy (2003) that adopts self-consistent principle for handling censoring and the other one is based on Peng and Huang (2008) that utilizes the martingale structure of randomly censored data. Though numerical studies have suggested the close proximity between these two methods, their underlying connection has not yet been established. In this note, we propose to formulate self-consistent estimation of censored quantile regression through stochastic integral equations. This novel formulation not only renders a clear presentation of self-consistent estimation procedure but also aid in understanding the associated asymptotics, particularly uncovering the large sample equivalence between Portnoy’s approach and Peng-Huang’s approach. Furthermore, the proposed framework for censored regression quantiles is readily extended to other survival settings where the principle of self-consistency is applicable.

Key Words: Censoring; Martingale; Regression Quantile; Self-consistency; Stochastic integral equation.

1. INTRODUCTION

Quantile regression (Koenker and Bassett 1978) has arisen into a useful regression technique for survival data (i.e. time-to-event data). Compared to traditional survival regression
methods, including the Cox proportional hazards model and the accelerated failure time (AFT) model, quantile regression can accommodate a more general relationship between event time and covariates while providing straightforward interpretations.

Let $T$ and $C$ denote time to event and time to censoring, and let $\mathbf{Z}$ denote $p \times 1$ covariate vector with the first component set as 1. Without loss of generality, the censored regression quantile model investigated in this paper takes the form,

$$ Q_T(\tau | \mathbf{Z}) = \exp \{ \mathbf{Z}^T \beta_0(\tau) \}, \quad \tau \in (0, 1), $$

where $Q_T(\tau | \mathbf{Z}) = \inf \{ t : \Pr(T \leq t | \mathbf{Z}) \geq \tau \}$ denoting the conditional quantile function of $T$ given $\mathbf{Z}$ (with the same definition applied to any other random variable), and $\beta_0(\tau)$ is a $p \times 1$ vector of unknown regression coefficients representing covariate effects on $\tau$-th quantile of $\log T$. Model (1) adopts the standard random censoring mechanism. That is, $T$ and $C$ are assumed to be independent conditional on $\mathbf{Z}$.

It is worth noting that much previous work on quantile regression with randomly censored data can not address the regression quantile problem defined above. For example, early efforts by Powell (1984, 1986) require that all censoring times be known or fixed and so do the subsequent work by Fitzenberger (1997), Buchinsky and Hann (1998), among others. Other methods, such as those by Ying, Jung, and Wei (1995) and Honore, Khan, and Powell (2002), demand unconditional independence between $T$ and $C$, which is a stronger assumption than the standard random censorship, and thus can not be applied to compute censored regression quantiles in model (1).

Portnoy (2003) made the first attempt to tackle the censored regression quantile problem (1) by novelly employing the principle of self-consistency (Efron 1967). The resulting estimator reduces to the Kaplan-Meier estimator (Kaplan and Meier 1958) in one-sample case. The self-consistent algorithm was polished in Neocleous, Vanden Branden, and Portnoy (2006) in a form of grid-based estimation procedure. We hereafter refer the estimator obtained from Neocleous et al. (2006)’s algorithm to as Portnoy’s self-consistent estimator,
denoted by $\hat{\beta}_{PSC}(\tau)$. Asymptotics associated with $\hat{\beta}_{PSC}(\tau)$, however, have not yet been fully established.

Alternatively, Peng and Huang (2008) proposed to estimate $\beta_0(\tau)$ in model (1) by utilizing the martingale structure of randomly censored data. Their grid-based algorithm is clearly defined and was developed based on a set of monotone estimating equations. These estimating equations have appealing stochastic integral representations which greatly facilitate large-sample studies. Peng and Huang (2008) derived the close form for the limit process of their estimator, denoted by $\hat{\beta}_{PH}(\tau)$. In addition, $\hat{\beta}_{PH}(\tau)$ was shown to be closely related to the Nelson-Aalen estimator (Nelson 1972; Aalen 1978) when there is no covariate.

Both Portnoy (2003)’s and Peng and Huang (2008)’s methods have been implemented for R in the contributed package quantreg by Koenker (2008a). A comprehensive comparison was conducted by Koenker (2008b), and demonstrated very similar empirical performance between these two approaches, agreeing with the numerical studies reported in Peng and Huang (2008). Though the asymptotic equivalence between $\hat{\beta}_{PSC}(\tau)$ and $\hat{\beta}_{PH}(\tau)$ can be established in one-sample case given the fact that the Kaplan-Meier estimator and the Nelson-Aalen estimator are equivalent in the large sample sense, the connection between these two estimators remains unknown in general regression settings. The major challenge may be the lack of knowledge on the large-sample behavior of $\hat{\beta}_{PSC}(\tau)$.

In this paper, we propose a novel formulation of the self-consistent estimation of censored regression quantiles. This new formulation based on stochastic integral equation leads to several clearly defined estimation procedures, and suggests a reasonable modification of Neocleous et al (2006)’s algorithm that avoids the ambiguity with estimating $\beta_0(\cdot)$ at the lower tail. More importantly, the new representation of the self-consistent estimation provides a direct approach to the asymptotic theory of resulting estimators. As a result, we are able to established the asymptotic equivalence between $\hat{\beta}_{PH}(\tau)$ and self-consistent estimators including a modified $\hat{\beta}_{PSC}(\cdot)$ provided that grid size is of order $o(n^{-1/2})$. It is also
important to note that the proposed formulation of self-consistent estimation of censored quantile regression is readily adapted to settings with other types of censorship, for example, double censoring, or interval-censoring. Such extensions will be investigated in separate papers.

The rest of the paper will be organized as follows. In Section 2, we give a brief introduction of algorithms in Neocleous et al. (2006) and Peng and Huang (2008). We present our main results in Section 3 including the self-consistent stochastic integral equations for censored regression quantiles, the resulting estimators, and justifications for their asymptotic equivalence to \( \hat{\beta}_{PH}(\tau) \). Monte-Carlo simulations reported in Section 4 confirm our theoretical findings. A few concluding remarks are provided in Section 5.

2. TWO APPROACHES FOR CENSORED QUANTILE REGRESSION

Define \( \tilde{X} = T \wedge C \) and \( \delta = I(T \leq C) \), where \( \wedge \) is the minimum operator and \( I(\cdot) \) is the indicator function. The observed randomly censored data consist of \( n \) iid replicates of \((X, \delta, Z)\), denoted by \( \{X_i, \delta_i, Z_i\}, \ i = 1, \ldots, n \). We define \( X = \log \tilde{X} \) and accordingly \( X_i = \log \tilde{X}_i \).

2.1 Portnoy’s Self-Consistent Approach

We outline the grid-based algorithm presented in Neocleous et al. (2006). A grid of \( \tau \)-values is defined as \( 0 < \tau_1 < \tau_2 < \ldots < \tau_M < 1 \). Define \( m(\beta, i, k) = \max\{l : 1 \leq l \leq k, Z_i^T \beta(\tau_{l-1}) < X_i \leq Z_i^T \beta(\tau_l)\} \) if the set \( \{l : 1 \leq l \leq k, Z_i^T \beta(\tau_{l-1}) < X_i \leq Z_i^T \beta(\tau_l)\} \) is not empty, and \( m(\beta, i, k) = k + 1 \) otherwise. By this definition, \( m(\beta, i, 0) = 1 \).

1. Compute \( \hat{\beta}_{PSC}(\tau_1) \) fitting the uncensored quantile regression with data \( \{X_i\}_{i=1}^n \). It is assumed that all censored \( X_i \)'s are above the hyperplane determined by \( \hat{\beta}_{PSC}(\tau_1) \). Set \( k = 1 \).

2. Given \( \hat{\beta}_{PSC}(\tau_l) \) \( (l \leq k) \), obtain \( \hat{\beta}_{PSC}(\tau_{k+1}) \) by minimizing the following weighted check
\[ \sum_{\delta_i=1} \rho_{\tau_i}(X_i - Z_i^T b) + \sum_{\delta_i=0} \{\hat{w}_{k+1,i} \rho_{\tau_i}(X_i - Z_i^T b) + (1 - \hat{w}_{k+1,i}) \rho_{\tau_i}(X^* - Z_i^T b) \}, \quad (2) \]

where \( \hat{w}_{k+1,i} = (\tau_{k+1} - \tau_m(\hat{\beta}_{PSC,i,k}))/(1 - \tau_m(\hat{\beta}_{PSC,i,k})) \), \( \rho_{\tau_i}(u) = u\{\tau - I(u < 0)\} \) and \( X^* \) is an extremely large value.

3. Replace \( k \) by \( k + 1 \).

4. Repeat steps 2-3 until \( k > M \) or only censored observations remain above \( Z_i^T \hat{\beta}_{PSC}(\tau_{k-1}) \).

It has been noticed that there involves some ambiguity in determining \( \hat{\beta}_{PSC}(\tau_1) \) when there is a censored \( X \) lying below the hyperplane \( Z_i^T \hat{\beta}_{PSC}(\tau_1) \). A detailed discussion on this issue can be found in the Appendix of Portnoy (2003).

2.2 Peng and Huang (2008)’s Approach

The estimator proposed by Peng and Huang (2008), \( \hat{\beta}_{PH}(\tau) \), is defined as a cadlag solution to the estimating equation

\[ n^{1/2} S_n^{(PH)}(\beta, \tau) \equiv n^{-1/2} \sum_{i=1}^n Z_i \left[ N_i\{Z_i^T \beta(\tau)\} - \int_0^\tau Y_i\{Z_i^T \beta(u)\} dH(u) \right] = 0, \quad (3) \]

where \( N_i(x) = I(X_i \leq x, \delta_i = 1) \), \( Y_i(x) = I(X_i \geq x) \), and \( H(x) = -\log(1 - x) \). A grid of \( \tau \)-values is defined as \( 0 = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_M < \tau_U \), where \( \tau_U \) is a deterministic constant subject to some identifiability constraint. Without further mentioning, such defined grid will be adopted throughout the rest of the paper, even for \( \hat{\beta}_{PSC}(\tau) \). The algorithm for obtaining \( \hat{\beta}_{PH}(\tau) \) is well defined from \( \tau_1 \) to \( \tau_M \) as follows.

1. Set \( \exp\{Z_i^T \hat{\beta}_{PH}(\tau_0)\} = 0 \) for all \( i \) by the definition of \( Q_T(\tau|Z) \). Set \( k = 0 \).

2. Given \( \exp\{Z_i^T \hat{\beta}_{PH}(\tau_l)\} = 0 \) for \( l \leq k \), obtain \( \hat{\beta}_{PH}(\tau_{k+1}) \) as the minimizer of the
following $L_1$-type convex objective function:

$$
l_{k+1}(h) = \sum_{i=1}^{n} \left| \delta_i X_i - \delta_i h^\top Z_i \right| + \left| X^* - h^\top \sum_{i=1}^{n} (-\delta_i Z_i) \right| + \left| X^* - h^\top \sum_{r=1}^{n} \left( 2Z_r \sum_{l=0}^{k} I[X_r \geq Z_r^\top \beta_{\hat{PH}}(\tau_l)] H(\tau_{l+1}) - H(\tau_l) \right) \right|,
$$

where $X^*$ is an extremely large value.

3. Replace $k$ by $k + 1$ and repeat step 2 until $k = M$ or no feasible solution can be found for minimizing $l_k(h)$.

3. MAIN RESULTS

3.1 Stochastic Integral Formulation of Self-Consistent Estimation of Censored Regression Quantiles

Define $R_i(x) = I(X_i \leq x, \delta_i = 0)$. Without considering covariates, according to Efron (1967), the self-consistent estimating equation for the distribution function of $\log T$, $F_{\log T}(t) \equiv \Pr(\log T \leq t)$, is given by

$$
F_{\log T}(t) = n^{-1} \sum_{i=1}^{n} \left\{ N_i(t) + R_i(t) \frac{F_{\log T}(t) - F_{\log T}(X_i)}{1 - F_{\log T}(X_i)} \right\}. 
$$

(4)

The right hand side (RHS) of equation (4) can be expressed as a stochastic integral. Specifically,

RHS of (4) $= n^{-1} \sum_{i=1}^{n} \left\{ N_i(t) + R_i(t) \int_{0}^{t} \frac{F_{\log T}(t) - F_{\log T}(u)}{1 - F_{\log T}(u)} dR_i(u) \right\}$

$= n^{-1} \sum_{i=1}^{n} \left[ N_i(t) + R_i(t) \{1 - F_{\log T}(t)\} \int_{0}^{t} \frac{R_i(u)}{(1 - F_{\log T}(u))^2} dF_{\log T}(u) \right],
$

where the second equality follows by applying stochastic integral by parts assuming the continuity of $F_{\log T}(u)$. This renders an alternative form of the self-consistent estimating equation for $F_{\log T}(t)$,

$$
F_{\log T}(t) = n^{-1} \sum_{i=1}^{n} \left[ N_i(t) + R_i(t) \{1 - F_{\log T}(t)\} \int_{0}^{t} \frac{R_i(u)}{(1 - F_{\log T}(u))^2} dF_{\log T}(u) \right].
$$

(5)
In regression set-up, note that model (1) is equivalent to
\[ Q_{\log T}(\tau | Z) = Z^{\top} \beta_0(\tau), \quad \tau \in (0, 1). \]

With \( t \) replaced by \( Z_i^{\top} \beta(\tau) \), equation (5) naturally evolves into an estimating equation for \( \beta_0(\tau) \):
\[ n^{1/2} S_n^{(SC)}(\beta, \tau) = 0, \]
where
\[ S_n^{(SC)}(\beta, \tau) = n^{-1} \sum_{i=1}^{n} Z_i \left[ \frac{1}{2} \left( N_i \{ Z_i^{\top} \beta(\tau) \} + R_i \{ Z_i^{\top} \beta(\tau) \} (1 - \tau) \right) \right. \]

The stochastic integral equation (6) entails a simple and clearly-defined estimation procedure for \( \beta_0(\tau) \). The solution to (6) can be approximated by mimicking Euler’s method, say the forward version, for first order differential equation. Specifically, let \( \hat{\beta}_{SC}(-) \) denote a solution to (6), which is cadlag and only jumps at the \( \tau \)-grid, \( 0 = \tau_0 < \tau_1 < \ldots < \tau_M = \tau_U < 1 \). Setting \( \exp \{ Z_i \hat{\beta}_{SC}(0) \} = 0 \) for all \( i \), one can obtain \( \hat{\beta}_{SC}(\tau_{k+1}) \) by sequentially solving the following equation for \( b \) for \( k = 0, \ldots, M - 1 \):
\[ n^{-1/2} \sum_{i=1}^{n} Z_i \left( N_i \{ Z_i^{\top} b \} + R_i \{ Z_i^{\top} b \} \left[ \sum_{l=0}^{k} R_i \{ Z_i^{\top} \hat{\beta}_{SC}(\tau_l) \} \left( \frac{1 - \tau_{k+1}}{1 - \tau_{l+1}} - \frac{1 - \tau_{k+1}}{1 - \tau_l} \right) \right] - \tau_{k+1} \right) = 0. \]

It is easy to see that equation (7) is monotone (Fygenson and Ritov 1994). Simple algebraic manipulations show that the estimating function in (7) is the minus subgradient of the following weighted check function:
\[ n^{-1/2} \left[ \sum_{\delta_i=1}^{n} \rho_{\tau} (X_i - Z_i^{\top} b) + \sum_{\delta_i=0} \{ \tilde{w}_{k+1,i} \rho_{\tau} (X_i - Z_i^{\top} b) + (1 - \tilde{w}_{k+1,i}) \rho_{\tau} (X^* - Z_i^{\top} b) \} \right], \]
where \( \tilde{w}_{k+1,i} = \sum_{l=0}^{k} R_i \{ Z_i^{\top} \hat{\beta}_{SC}(\tau_l) \} \left( \frac{1 - \tau_{k+1}}{1 - \tau_{l+1}} - \frac{1 - \tau_{k+1}}{1 - \tau_l} \right), k = 0, \ldots, M - 1. \)

As shown above, the self-consistent estimator \( \hat{\beta}_{SC}(\cdot) \) is well defined and can be obtained in a sequential manner like Portnoy’s self-consistent estimator. Its connection with \( \hat{\beta}_{PH}(\cdot) \) and \( \hat{\beta}_{PSC}(\cdot) \) are elaborated in the next two subsections.
3.2 Asymptotic equivalence between $\hat{\beta}_{PH}(\cdot)$ and $\hat{\beta}_{SC}(\cdot)$

The stochastic integral representation of (6) greatly facilitates exploiting the relationship between $\hat{\beta}_{PH}(\cdot)$ and $\hat{\beta}_{SC}(\cdot)$. We establish the asymptotic equivalence between $\hat{\beta}_{PH}(\cdot)$ and $\hat{\beta}_{SC}(\cdot)$ stated in the following proposition:

**Proposition 1.** Suppose model (1) holds. Assuming conditions required by Theorem 2 of Peng and Huang (2008) and condition (D1) in Appendix are satisfied,

$$\sup_{\tau \in [\nu, \nu_T]} \| n^{1/2} \{ \hat{\beta}_{PH}(\tau) - \hat{\beta}_{SC}(\tau) \} \| \rightarrow_p 0,$$

(8)

for any $\nu \in (0, \nu_U)$.

The proof is sketched as follows. First note that $Y_i(t) = 1 - N_i(t-) - R_i(t-)$. By the definition of $\hat{\beta}_{PH}(\cdot)$, we have

$$\sum_{i=1}^{n} Z_i N_i \{ Z_i^\top \hat{\beta}_{PH}(\tau) \} = - \int_{0}^{\tau} \left[ \sum_{i=1}^{n} Z_i N_i \{ Z_i^\top \hat{\beta}_{PH}(u-) \} \right] dH(u)$$

$$+ \sum_{i=1}^{n} \left[ Z_i H(\tau) - \int_{0}^{\tau} Z_i R_i \{ Z_i^\top \hat{\beta}_{PH}(u-) \} dH(u) \right] + \nu_n(\tau),$$

where $\nu_n(\tau)$ is cadlag and satisfies that $n^{-1/2} \nu_n(\tau) \overset{a}{\rightarrow} 0$. Here and hereafter, $\overset{a}{=}$ means asymptotic equivalence uniformly in $\tau \in (0, \tau_U)$. Viewing the above equality as a stochastic integral equation for $\sum_{i=1}^{n} Z_i N_i \{ Z_i^\top \hat{\beta}_{PH}(\tau) \}$, we get from Theorem II.6.3 of Andersen, Borgan, Gill, and Keiding (1998) that

$$\sum_{i=1}^{n} Z_i N_i \{ Z_i^\top \hat{\beta}_{PH}(\tau) \} = \int_{0}^{\tau} \left( \sum_{i=1}^{n} Z_i \left[ dH(s) - R_i \{ Z_i^\top \hat{\beta}_{PH}(s) \} dH(s) \right] + d\nu_n(s) \right) \Pi_{(s, \tau)} \{ 1 - dH(u) \}.$$

Since $\Pi_{(s, \tau)} \{ 1 - dH(u) \} = \exp \{ -H(u)^{s \rightarrow \tau} \} = (1 - \tau)/(1 - s)$, it follows that

$$\sum_{i=1}^{n} Z_i N_i \{ Z_i^\top \hat{\beta}_{PH}(\tau) \} = \sum_{i=1}^{n} Z_i \int_{0}^{\tau} \left[ \frac{1 - \tau}{1 - s} \cdot \frac{ds}{1 - s} - \frac{1 - \tau}{(1 - s)^2} R_i \{ Z_i^\top \hat{\beta}_{PH}(s) \} ds \right] + \hat{\nu}_n(\tau)$$

$$= \sum_{i=1}^{n} Z_i \left[ \tau - (1 - \tau) \int_{0}^{\tau} R_i \{ Z_i^\top \hat{\beta}_{PH}(s) \} \frac{ds}{(1 - s)^2} \right] + \hat{\nu}_n(\tau),$$

where $n^{-1/2} \hat{\nu}_n(\tau) \overset{a}{\rightarrow} 0$. Then Lemma A1 (in Appendix) implies that for any $\hat{\beta}(\tau)$ that satisfies

$$n^{1/2} S_n^{(OSC)}(\beta, \tau) \equiv n^{-1/2} \sum_{i=1}^{n} Z_i \left( N_i \{ Z_i^\top \beta(\tau) \} - \left[ \tau - (1 - \tau) \int_{0}^{\tau} R_i \{ Z_i^\top \beta(s) \} \frac{ds}{(1 - s)^2} \right] \right) \overset{a}{=},$$

(9)
we have \( \sup_{\tau \in [\nu, \tau_U]} \| \hat{\beta}_{PH}(\tau) - \tilde{\beta}(\tau) \| \to_p 0 \) for any \( 0 < \nu < \tau_U \).

It is important to note that the LHS of (9) has the same form that the self-consistent estimating function in (6) takes except for the absence of \( R_i \{ Z_i^T \tilde{\beta}(\tau) \} \) before the term \( \int_0^\tau \frac{R_i \{ Z_i^T \tilde{\beta}(s) \}}{(1-s)^2} ds \). Define \( \hat{\beta}_{OSC}(\tau) \) as a cadlag process obtained as follows: (i) set \( \exp \{ Z_i^T \hat{\beta}_{OSC}(\tau) \} = 0 \) for all \( i \); (ii) for \( k = 0, \ldots, M - 1 \), sequentially compute \( \hat{\beta}_{OSC}(\tau_{k+1}) \) as the minimizer of the \( L_1 \)-type convex objective function,

\[
\ell_{k+1}(h) = \sum_{i=1}^n \left| \delta_i X_i - \delta_i h^T Z_i \right| + \left| X^* - h^T \sum_{l=1}^n (-\delta_l Z_l) \right| + \left| X^* - 2Z r \sum_{r+1}^n (\tau_{k+1} - \hat{w}_{k+1,r}) \right|.
\]

It is easily seen that (9) is satisfied with \( \beta(\cdot) = \hat{\beta}_{OSC}(\cdot) \) when the grid size, \( \min_{j=1, \ldots, M}(\tau_j - \tau_{j-1}) \), is of order \( o(n^{-1/2}) \). Therefore, \( \sup_{\tau \in [\nu, \tau_U]} \| n^{1/2} \{ \hat{\beta}_{PH}(\tau) - \hat{\beta}_{OSC}(\tau) \} \| \to_p 0 \) for any \( \nu \in (0, \tau_U) \).

The estimator \( \hat{\beta}_{OSC}(\tau) \) serves as a bridge to connect \( \hat{\beta}_{SC}(\tau) \) and \( \hat{\beta}_{PH}(\tau) \). First, it follows from Lemma A2 (in Appendix) that \( n^{1/2} S_n^{(SC)}(\hat{\beta}_{OSC}, \tau) \Rightarrow 0 \). This, coupled with Lemma A1, implies \( \sup_{\tau \in [\nu, \tau_U]} \| n^{1/2} \{ \hat{\beta}_{OSC}(\tau) - \hat{\beta}_{SC}(\tau) \} \| \to_p 0 \) and \( \sup_{\tau \in [\nu, \tau_U]} \| n^{1/2} \{ \hat{\beta}_{OSC}(\tau) - \hat{\beta}_{PH}(\tau) \} \| \to_p 0 \) for any \( \nu \in (0, \tau_U) \), and consequently (9) holds.

### 3.3 Connection between \( \hat{\beta}_{SC}(\cdot) \) and \( \hat{\beta}_{PSC}(\cdot) \)

Define \( A_i(\beta, \tau) = \{ u : 0 \leq u < \tau, Z_i^T \beta(u-) \leq X_i \leq Z_i^T \beta(u) \} \) and \( \psi_i(\beta, \tau) = \sup \{ A_i(\beta, \tau) \} \cdot I(A_i(\beta, \tau) \text{ is not empty}) + \tau \cdot I(A_i(\beta, \tau) \text{ is empty}) \). It is easy to see that \( \psi_i(\hat{\beta}_s, \tau_{k+1}) = m(\hat{\beta}_s, i, k) \), where the subscript \( s \) can be “PH”, “SC”, “OSC”, or “PSC”.

9
By Lemma A3 in the Appendix,

\[ 0 \overset{a}{=} n^{1/2} S_n^{(SC)} (\widehat{\beta}_{SC}, \tau) \]
\[ = n^{-1/2} \sum_{i=1}^{n} Z_i \left[ N_i \{ Z_i^T \widehat{\beta}_{SC} (\tau) \} \right. \]
\[ + R_i \{ Z_i^T \widehat{\beta}_{SC} (\tau) \} (1 - \tau) \int_{\psi_i(\widehat{\beta}_{SC}, \tau)}^{\tau} \frac{R_i \{ Z_i^T \widehat{\beta}_{SC} (u) \}}{(1 - u)^2} du - \tau \]
\[ = n^{-1/2} \sum_{i=1}^{n} Z_i \left[ N_i \{ Z_i^T \widehat{\beta}_{SC} (\tau) \} \right. \]
\[ + R_i \{ Z_i^T \widehat{\beta}_{SC} (\tau) \} (1 - \tau) \left( \frac{1}{1 - \tau} - \frac{1}{1 - \psi_i(\widehat{\beta}_{SC}, \tau)} \right) - \tau \]
\[ = n^{-1/2} \sum_{i=1}^{n} Z_i \left[ N_i \{ Z_i^T \widehat{\beta}_{SC} (\tau) \} + R_i \{ Z_i^T \widehat{\beta}_{SC} (\tau) \} \frac{\tau - \psi_i(\widehat{\beta}_{SC}, \tau)}{1 - \psi_i(\widehat{\beta}_{SC}, \tau)} - \tau \right]. \]

Note that the second \( a \) can be replaced by \( = \) if \( Z_i^T \widehat{\beta}_{SC} (\tau) \) are increasing in \( \tau \) for all \( i \). In this case, Lemma A3 would not be needed.

Let \( S_n^{(MSC)} (\beta, \tau) = n^{-1} \sum_{i=1}^{n} Z_i \left[ N_i \{ Z_i^T \beta (\tau) \} + R_i \{ Z_i^T \beta (\tau) \} \frac{\tau - \psi_i(\beta, \tau)}{1 - \psi_i(\beta, \tau)} - \tau \right] \). The above result shows that \( n^{1/2} S_n^{(MSC)} (\widehat{\beta}_{SC}, \tau) \overset{a}{=} 0 \). Define \( \widehat{\beta}_{MSC}(\cdot) \) as a cadlag approximation of a solution to \( S_n^{(MSC)} (\beta, \tau) = 0 \). Specifically, \( \widehat{\beta}_{MSC}(\cdot) \) jumps only on \( 0 = \tau_0 < \tau_1 < \ldots < \tau_M = \tau_U \) with \( \exp \{ Z_i^T \widehat{\beta}_{MSC} (0) \} \) = 0 and \( \widehat{\beta}_{MSC}(\tau_{k+1}) \), given \( \{ \widehat{\beta}_{MSC}(\tau_i) \}_{i=0}^{k} \), obtained as the solution to

\[ n^{-1/2} \sum_{i=1}^{n} Z_i \left[ N_i (Z_i^T b) + R_i (Z_i^T b) \left( \frac{\tau_{k+1} - \tau_{m \widehat{\beta}_{MSC}, i, k}}{1 - \tau_{m \widehat{\beta}_{MSC}, i, k}} \right) - \tau \right] = 0. \] (10)

When the grid size is \( o(n^{-1/2}) \), we have \( n^{1/2} S_n^{(MSC)} (\widehat{\beta}_{MSC}, \tau) \overset{a}{=} 0 \). The asymptotic equivalence between \( \widehat{\beta}_{SC}(\cdot) \) and \( \widehat{\beta}_{MSC}(\cdot) \) then follows from Lemma A1. Therefore, we have Proposition 2:

**Proposition 2.** Assuming conditions required by Theorem 2 of Peng and Huang (2008) and condition (D1) in Appendix are satisfied,

\[ \sup_{\tau \in [\nu, \tau_U]} \| n^{1/2} \{ \widehat{\beta}_{SC}(\tau) - \widehat{\beta}_{MSC}(\tau) \} \| \overset{p}{\rightarrow} 0, \]

for any \( \nu \in (0, \tau_U) \).
It is remarkable that $\hat{\beta}_{MSC}(\cdot)$ greatly resembles Portnoy’s estimator $\hat{\beta}_{PSC}(\cdot)$. Note that the estimating function in (10) is the minus subgradient of the following weighted check function,

$$n^{-1/2} \left[ \sum_{\delta_i=1} \rho_\tau(X_i - Z_i^T b) + \sum_{\delta_i=0} \{ w^*_k, i \rho_\tau(X_i - Z_i^T b) + (1 - w^*_k, i) \rho_\tau(X^* - Z_i^T b) \} \right],$$

where $w^*_k, i = (\tau_{k+1} - \tau_m(\hat{\beta}_{MSC,i,k}))/\{1 - \tau_m(\hat{\beta}_{MSC,i,k})\}$, $k = 0, \ldots, M - 1$. The weights for $\hat{\beta}_{PSC}(\cdot)$, $\hat{w}_k, i$, and the weights for $\hat{\beta}_{MSC}(\cdot)$, $w^*_k, i$, have nearly identical definitions except for the weights used in computation of $\hat{\beta}_{PSC}(\tau_1)$ and $\hat{\beta}_{MSC}(\tau_1)$. The algorithm in Portnoy (2003) or Neocleous et al (2006) forces all $\hat{w}_1, i$’s equal to 1 when obtaining $\hat{\beta}_{PSC}(\tau_1)$ by uncensored regression quantiles. This may incur complications if there are censored $X_i$’s less than $Z_i^T \hat{\beta}_{PSC}(\tau_1)$. In contrast, the weights for $\hat{\beta}_{MSC}(\tau_1)$, $w^*_1, i$’s, are set as 0 on the basis of estimating equation (6) without any restriction on $\hat{\beta}_{MSC}(\tau_1)$. By these arguments, $\hat{\beta}_{MSC}(\cdot)$ may be viewed as a modified, perhaps corrected, version of Portnoy’s self-consistent estimator, $\hat{\beta}_{PSC}(\cdot)$. This new version of Portnoy’s estimator is asymptotically equivalent to $\hat{\beta}_{sc}(\cdot)$, and thus Peng and Huang (2008)’s estimator, $\hat{\beta}_{PH}(\cdot)$.

4. SIMULATIONS

Simulation studies have been conducted to compare Peng and Huang’s estimator $\hat{\beta}_{PH}(\cdot)$ and the four types of self-consistent estimators, including $\hat{\beta}_{sc}(\cdot)$, $\hat{\beta}_{OSC}(\cdot)$, $\hat{\beta}_{MSC}(\cdot)$, and $\hat{\beta}_{PSC}(\cdot)$. For brevity of presentation, we only report results from two scenarios: (I) log-linear model with iid errors; (ii) log-linear model with heteroscedastic errors. The configurations follow those used in Peng and Huang (2008) with 25% censoring. Specifically, for scenario (I), $T$ follows the model $\log T = 0.5Z_1 - 0.5Z_2 + \epsilon$, where $\epsilon \sim$ extreme value distribution, $Z_1 \sim Unif(0,1)$, and $Z_2 \sim Bernoulli(.5)$. The censoring distribution is $Unif(0.1I(Z_2 = 1), 3.8)$. For scenario (II), $T$ is generated from the model $\log T = 0.5Z_1 - 0.5Z_2 \xi + \epsilon$, where $\xi \sim exponential(1)$, $\epsilon \sim N(0,1)$, $Z_1 \sim Unif(0,1)$, and $Z_2 \sim Bernoulli(.5)$. The censoring time $C \sim Unif(0.3I(Z_2 = 1), 5.2)$. Under each configuration, we generate 1000 simulated
datasets of sample size $n = 200$.

In Table 1, we present the empirical biases (Bias) and empirical variances (Var) of all five estimators under comparison. As expected, all estimators have small empirical biases, and their empirical variances are very similar, especially those of $\hat{\beta}_{OSC}(\tau)$, $\hat{\beta}_{SC}(\tau)$ and $\hat{\beta}_{PSC}(\tau)$. We also examine the difference among these estimators based on each simulated dataset. Selecting $\hat{\beta}_{SC}(\tau)$ as the reference, we present in Table 2 the empirical 25th percentiles (Diff25) and empirical 75th percentile (Diff75) of $\hat{\beta}_{PH}(\tau) - \hat{\beta}_{SC}(\tau)$, $\hat{\beta}_{OSC}(\tau) - \hat{\beta}_{SC}(\tau)$, $\hat{\beta}_{MSC}(\tau) - \hat{\beta}_{SC}(\tau)$, and $\hat{\beta}_{PSC}(\tau) - \hat{\beta}_{SC}(\tau)$. Results in Table 2 confirm the observation from Table 1 that the three new versions of self-consistent estimator presented in Section 3, $\hat{\beta}_{OSC}(\cdot)$, $\hat{\beta}_{SC}(\cdot)$, and $\hat{\beta}_{MSC}(\cdot)$, are in close proximity; in over 50% of simulations they coincide with each other. It is interesting to note that Portnoy’s estimator $\hat{\beta}_{PSC}$ seems to have relatively larger deviation from $\hat{\beta}_{SC}$ than does $\hat{\beta}_{PH}$ at small $\tau$’s. This is likely due to its ambiguous estimation of $\beta_0(\tau_1)$. The deviation of Peng and Huang (2008)’s estimator $\hat{\beta}_{PH}$ from self-consistent estimators appears to rise as $\tau$ increases. This phenomenon is not surprising and shares the same spirit as the increasing cumulative error of Euler’s solution to ordinary differential equation. In summary, our simulation results provide empirical evidence for the asymptotic equivalence among $\hat{\beta}_{PH}(\tau)$, $\hat{\beta}_{SC}(\tau)$, $\hat{\beta}_{OSC}(\tau)$, $\hat{\beta}_{MSC}(\tau)$, and $\hat{\beta}_{PSC}(\tau)$, in addition to the theoretical arguments given in Section 3.

[Table 1 about here.]

[Table 2 about here.]

5. REMARK

The principle of self-consistency has been widely adopted in survival analysis as an intuitive way to handle missing information due to censoring and/or truncation. Examples include estimating survival function with randomly censored data (Efron 1967), doubly censored data (Turnbull 1974), and interval-censored data (Turnbull 1976). In this paper, we present
a novel representation of self-consistent estimation of censored regression quantiles based
on stochastic integral equation. The new formulation allows us to carry out self-consistent
estimation in a sequential manner without involving iterations. It also help clear up several
issues with Portnoy’s self-consistent estimator of censored regression quantiles, such as the
ambiguity in defining the first regression quantile and its large sample properties when grid
size is of order \( o(n^{-1/2}) \). Moreover, on the basis of the presented stochastic integral equation,
we are able to establish the connection between two major approaches for censored quantile
regression, Portnoy (2003) and Peng and Huang (2008). The proposed framework can also
be extended to other settings with more complex survival data. Such an endeavor is expected
to receive a similar benefit to that for randomly censoring case.

6. APPENDIX

Define \( F_X(x|Z) = \Pr(X \leq x|Z) \), \( F_{X,1}(x|Z) = \Pr(X \leq x, \delta = 1|Z) \), \( F_{\log T}(x|Z) = \Pr(\log T \leq x|Z) \),
\( f_X(x|Z) = dF_X(x|Z)/dx \), \( f_{X,1}(x|Z) = dF_{X,1}(x|Z)/dx \), and \( f_{\log T}(x|Z) = dF_{\log T}(x|Z)/dx \). Let \( Z \) denote the domain of \( Z \).

Regularity conditions include conditions (C1)-(C6) in Peng and Huang (2008), and

(D1): (i) \( \sup_{x,z \in Z} f_X(x|z) \) is bounded; (ii) \( \sup_{x,z \in Z} f_{\log T}(x|z) \) is bounded; (iii) Each com-
ponent of \( E[Z^{\otimes 2} f_X(Z^b|Z)] (E[Z^{\otimes 2} f_{X,1}(Z^b|Z)])^{-1} \) is uniformly bounded in \( b \in B(d_0) \),
where \( B(d_0) \) is the same as \( B(d_0) \) defined in Peng and Huang (2008).

Lemma A1: Suppose that model (1) holds. Let \( \hat{\beta}(-) \) and \( \tilde{\beta}(-) \) be cadlag processes that
only jump at \( 0 < \tau_1 < \tau_2 < \ldots < \tau_M < \tau_U \). Assuming that conditions required by
Theorem 2 of Peng and Huang (2008) and condition (D1) are satisfied, if \( n^{1/2} S_n^{(u)}(\tilde{\beta}, \tau) \rightarrow^d 0 \)
and \( n^{1/2} S_n^{(u)}(\hat{\beta}, \tau) \rightarrow^d 0 \), then

\[
\sup_{\tau \in [\nu, \tau_U]} \| \hat{\beta}(\tau) - \beta_0(\tau) \| \rightarrow_p 0, \quad \sup_{\tau \in [\nu, \tau_U]} \| \tilde{\beta}(\tau) - \beta_0(\tau) \| \rightarrow_p 0,
\]

13
and

\[ \sup_{\tau \in [\nu, \tau_V]} \| \hat{\beta}(\tau) - \beta(\tau) \| \rightarrow_p 0 \]

for any \( \nu \in (0, \tau_V) \), where the superscript \( u \) can be “PH”, “SC”, “OSC”, or “MSC”.

**Proof.** In case that \( u \) represents “PH”, Lemma A1 follows immediately from the proofs of Theorems 1-2 in Peng and Huang (2008). Very similar arguments can be applied to prove the case with \( u \) representing “SC”, “OSC” and “MSC” and hence are omitted here.

**Remark 1:** Lemma A1 states that a deviation of \( o(1) \) (uniformly in \( \tau \)) in the estimating function \( n^{1/2} S_n^{(PH)}, n^{1/2} S_n^{(SC)}, n^{1/2} S_n^{(OSC)}, \) or \( n^{1/2} S_n^{(MSC)} \) does not affect the uniform consistency and change the limit process of the resulting estimator.

**Lemma A2:** Suppose that model (1) holds. Assuming that conditions required by Theorem 2 of Peng and Huang (2008) and condition (D1) are satisfied,

\[ \sup_{\tau \in (0, \tau_V]} \left| \sum_{i=1}^{n} Z_i I\{X_i > \hat{Z}_i^T \beta_{OSC}(\tau)\} \int_0^\tau \frac{R_i\{\hat{Z}_i^T \beta_{OSC}(u)\}}{(1-s)^2} ds \right| \rightarrow_p 0. \quad \text{(A.1)} \]

**Proof:** First, it is easy to note that the LHS of (A.1) is bounded above by \( c_1 n^{-1/2} \sum_{i=1}^{n} I\{X_i > \hat{Z}_i^T \beta_{OSC}(\tau)\} \int_0^\tau \frac{R_i\{Z_i^T \beta_{OSC}(u)\}}{(1-u)^2} du \), and thus by

\[ c_1 \cdot \int_0^\tau \left[ n^{-1/2} \sum_{i=1}^{n} I\{Z_i^T \beta_{OSC}(\tau) < X_i \leq Z_i^T \beta_{OSC}(u)\} \right] du, \]

where \( c_1 \) is the upper bound for \( \| Z \| \). Then (A.1) holds if

\[ \sup_{0 < u < \tau \leq \tau_V} n^{-1/2} \sum_{i=1}^{n} \frac{I\{Z_i^T \beta_{OSC}(\tau) < X_i \leq Z_i^T \beta_{OSC}(u)\}}{(1-u)^2} = o_p(1). \quad \text{(A.2)} \]

Let \( \varphi(\beta, u, \tau) = \Pr\{Z^T \beta(\tau) < X \leq Z^T \beta(u)\} \). To prove (A.2), it suffices to show that \( \sup_{0 < u < \tau \leq \tau_V} \varphi(\beta_{OSC}, u, \tau) = o_p(1) \). Define \( \mu(b) = E\{Z I(X \leq Z^T b)\} \), \( \phi(b) = E\{Z I(X \leq Z^T b)\} \), and \( \phi_1(b) = E\{I(X \leq Z^T b)\} \). By Peng and Huang (2008) (see the proof for Theorem 2) and the asymptotic equivalence between \( \hat{\beta}_{PH}(\cdot) \) and \( \hat{\beta}_{OSC}(\cdot) \), we have \( \sup_{\tau \in [0, \tau_V]} \| \mu(\hat{\beta}_{OSC}(\tau)) - \mu(\beta_0(\tau)) \| \rightarrow_p 0 \). By condition D1 (iii), we then have \( \sup_{\tau \in [0, \tau_V]} \| \phi(\hat{\beta}_{OSC}(\tau)) - \phi(\beta_0(\tau)) \| \rightarrow_p 0 \).
For any $\vartheta > 0$, we can find some $\nu_0$ such that $\sup_{\tau \in [0, \tau_U]} |\phi_1\{\beta_0(\tau)\}| \leq \vartheta/8$ because $\phi_1\{\beta_0(0)\} = 0$ and $\phi_1\{\beta_0(\tau)\}$ is Lipschitz-continuous in $\tau$. Given the uniform convergence of $\phi\{\hat{\beta}_{OSC}(\tau)\}$ to $\phi\{\beta_0(\tau)\}$, there exists $N_{\vartheta, \xi, 1} > 0$ such that for $n \geq N_{\vartheta, \xi, 1},$

$$
\Pr\left( \sup_{\tau \in [0, \tau_U]} |\phi_1\{\hat{\beta}_{OSC}(\tau)\} - \phi_1\{\beta_0(\tau)\}| > \vartheta/8 \right) < \xi/3.
$$

When $\sup_{\tau \in [0, \tau_U]} |\phi_1\{\hat{\beta}_{OSC}(\tau)\} - \phi_1\{\beta_0(\tau)\}| \leq \vartheta/8$, we have $\sup_{0 < \tau \leq \tau_U, \tau < \nu_0} |\varphi\{\hat{\beta}_{OSC}, u, \tau\}| \leq \sup_{\tau \in [0, \nu_0]} |\phi_1\{\hat{\beta}_{OSC}(\tau)\}| < \vartheta/4$.

Since $\sup_{\tau \in [\nu, \tau_U]} \|\hat{\beta}_{OSC}(\tau) - \beta_0(\tau)\| \to_p 0$ for any $\nu \in (0, \tau_U)$ by Theorem 2 of Peng and Huang (2008), there exists $N_{\vartheta, \xi, 2}$ such that for $n \geq N_{\vartheta, \xi, 2},$

$$
\Pr\left( \sup_{\tau \in [\nu, \tau_U]} \|\hat{\beta}_{OSC}(\tau) - \beta_0(\tau)\| > \vartheta/2c_2 \right) < \xi/3,
$$

where $c_2$ is a postive constant that bounds $\sup_{x, z} f_X(x | z)$ from above and its existence is guaranteed by condition D1 (i). Note that $\sup_{\tau \in [\nu, \tau_U]} \|\hat{\beta}_{OSC}(\tau) - \beta_0(\tau)\| \leq \vartheta/(2c_2)$ implies that

$$
\sup_{\nu_0 \leq u < \tau \leq \tau_U} \varphi(\hat{\beta}_{OSC}, u, \tau)
\leq \sup_{\nu_0 \leq u < \tau \leq \tau_U} \Pr(X \leq Z^T\beta_0(u) + \frac{\vartheta}{2c_2}, X > Z^T\beta_0(\tau) - \frac{\vartheta}{2c_2})
\leq \sup_{\nu_0 \leq u \leq \tau_U} \Pr(Z^T\beta_0(u) - \frac{\vartheta}{2c_2} < X \leq Z^T\beta_0(u) + \frac{\vartheta}{2c_2}) \leq \vartheta/2.
$$

Therefore, for $n \geq \max(N_{\vartheta, \xi, 1}, N_{\vartheta, \xi, 1}),$

$$
\Pr\left( \sup_{0 < \tau \leq \tau_U} \varphi(\hat{\beta}_{OSC}, u, \tau) > \vartheta \right) \leq \Pr\left( \sup_{\tau \in [0, \tau_U]} |\phi_1\{\hat{\beta}_{OSC}(\tau)\} - \phi_1\{\beta_0(\tau)\}| > \vartheta/8 \right)
+ \Pr\left( \sup_{\tau \in [\nu_0, \tau_U]} \|\hat{\beta}_{OSC}(\tau) - \beta_0(\tau)\| > \frac{\vartheta}{2c_2} \right) < \xi.
$$

This completes the proof of Lemma A2.
Lemma A.3: Suppose model (1) holds. Assuming that conditions required by Theorem 2 of Peng and Huang (2008) and condition (D1) are satisfied,

$$\sup_{\tau \in (0, \tau_U)} \left\| n^{-1/2} \sum_{i=1}^{n} Z_i R_i \{ Z_i^T \bar{\beta}_{SC}(\tau) \}(1 - \tau) \int_{0}^{\psi_i(\bar{\beta}_{SC}, \tau)} \frac{R_i \{ Z_i^T \bar{\beta}_{SC}(s) \}}{(1 - s)^2} ds \right\| \to_p 0.$$  

Proof: Under model (1), it holds that $F_{\log T} \{ Z_i^T \beta_0(\tau) | Z_i \} = \tau$ for $\tau \in (0, 1)$. This implies that $d \{ Z_i^T \beta_0(\tau) \}/d\tau = 1/f_{\log T} \{ Z_i^T \beta_0(\tau) | Z_i \}$. By condition D1(ii), $f_{\log T} \{ Z_i^T \beta_0(\tau) | Z_i \} < \infty$ for some $c_3$ for all $i$ and $\tau \in (0, 1)$. Therefore, for any given $0 < \Delta < \tau_U/(2c_3)$, $| Z_i^T \beta_0(\tau) - Z_i^T \beta_0(\tau') | \geq \Delta$ for all $i$ and $| \tau - \tau' | \geq 2\Delta c_3$.

We also note that

$$n^{-1/2} \sum_{i=1}^{n} Z_i R_i \{ Z_i^T \bar{\beta}_{SC}(\tau) \}(1 - \tau) \int_{0}^{\psi_i(\bar{\beta}_{SC}, \tau)} \frac{R_i \{ Z_i^T \bar{\beta}_{SC}(s) \}}{(1 - s)^2} ds$$

$$= n^{-1/2} \sum_{i=1}^{n} Z_i R_i \{ Z_i^T \bar{\beta}_{SC}(\tau) \}(1 - \tau) I \{ \psi_i(\bar{\beta}_{SC}, \tau) < \Delta c_3 \} \int_{0}^{\psi_i(\bar{\beta}_{SC}, \tau)} \frac{R_i \{ Z_i^T \bar{\beta}_{SC}(s) \}}{(1 - s)^2} ds$$

$$+ n^{-1/2} \sum_{i=1}^{n} Z_i R_i \{ Z_i^T \bar{\beta}_{SC}(\tau) \}(1 - \tau) I \{ \psi_i(\bar{\beta}_{SC}, \tau) \geq \Delta c_3, \psi_i(\beta_0, \tau) > 2\Delta c_3 \}$$

$$\int_{\psi_i(\beta_0, \tau)}^{\psi_i(\bar{\beta}_{SC}, \tau)} \frac{R_i \{ Z_i^T \bar{\beta}_{SC}(s) \}}{(1 - s)^2} ds$$

$$+ n^{-1/2} \sum_{i=1}^{n} Z_i R_i \{ Z_i^T \bar{\beta}_{SC}(\tau) \}(1 - \tau) I \{ \psi_i(\bar{\beta}_{SC}, \tau) \geq \Delta c_3, \psi_i(\beta_0, \tau) \leq 2\Delta c_3, \tau \leq 2\Delta c_3 \}$$

$$\int_{\psi_i(\beta_0, \tau)}^{\psi_i(\bar{\beta}_{SC}, \tau)} \frac{R_i \{ Z_i^T \bar{\beta}_{SC}(s) \}}{(1 - s)^2} ds$$

$$+ n^{-1/2} \sum_{i=1}^{n} Z_i R_i \{ Z_i^T \bar{\beta}_{SC}(\tau) \}(1 - \tau) I \{ \psi_i(\bar{\beta}_{SC}, \tau) \geq \Delta c_3, \psi_i(\beta_0, \tau) \leq 2\Delta c_3, \tau > 2\Delta c_3 \}$$

$$\int_{\psi_i(\beta_0, \tau)}^{\psi_i(\bar{\beta}_{SC}, \tau)} \frac{R_i \{ Z_i^T \bar{\beta}_{SC}(s) \}}{(1 - s)^2} ds$$

$$+ n^{-1/2} \sum_{i=1}^{n} Z_i R_i \{ Z_i^T \bar{\beta}_{SC}(\tau) \}(1 - \tau) I \{ \psi_i(\bar{\beta}_{SC}, \tau) \geq \Delta c_3 \} \int_{0}^{\psi_i(\beta_0, \tau)} \frac{R_i \{ Z_i^T \bar{\beta}_{SC}(s) \}}{(1 - s)^2} ds$$

$$\equiv A_{1,n}(\tau) + A_{2,n}(\tau) + A_{3,n}(\tau) + A_{4,n}(\tau) + A_{5,n}(\tau) \quad (A.3)$$
It is easy to see that
\[
\| A_{5,n}(\tau) \| \leq n^{-1/2} \sum_{i=1}^{n} \| Z_i \| \int_0^{\psi_i(\beta_0, \tau)} \frac{R_i \{ Z_i^T \hat{\beta}_{SC}(s) \}}{(1-s)^2} ds
\]
\[
\leq c_1 \int_0^{\tau} \left[ n^{-1/2} \sum_{i=1}^{n} \frac{I \{ Z_i^T \beta_0(0) \leq X_i \leq Z_i^T \hat{\beta}_{SC}(s) \} }{(1-s)^2} \right] ds
\]
Using similar arguments to those for (A.2), we can show that
\[
\sup_{\tau \in (0, \tau_U]} n^{-1/2} \sum_{i=1}^{n} \frac{I \{ Z_i^T \beta_0(0) \leq X_i \leq Z_i^T \hat{\beta}_{SC}(s) \} }{(1-s)^2} \rightarrow_p 0,
\]
and it follows that \( \sup_{\tau \in (0, \tau_U]} \| A_{5,n}(\tau) \| = o_p(1) \).

Now it remains to show \( \sup_{\tau \in (0, \tau_U]} \| \sum_{k=1}^{4} A_{k,n}(\tau) \| = o_p(1) \). We first consider the situation where \( \sup_{i, \tau \in [\Delta_3, \tau_U]} | Z_i^T \hat{\beta}_{SC}(\tau) - Z_i^T \beta_0(\tau) | \leq \Delta / 2 \). In this case, we can show that
\[
\sup_{i, \tau \in (0, \tau_U]} | I \{ \psi_i(\beta_0, \tau) > 2\Delta_3, \psi_i(\hat{\beta}, \tau) \geq \Delta_3 \} \{ \psi_i(\beta_0, \tau) - \psi_i(\hat{\beta}_{SC}, \tau) \} | \leq \Delta_3.
\]
Note that \( I \{ \psi_i(\beta_0, \tau) > 2\Delta_3 \} = 1 \) implies \( \tau > 2\Delta_3 \). When \( Z_i^T \beta_0 \{ \psi_i(\beta_0, \tau) \} = X_i < \tau \), we have for \( \tau_U > \tau \geq \psi_i(\beta_0, \tau) + \Delta_3 \),
\[
Z_i^T \beta_{SC} (\tau) \geq Z_i^T \beta_0 (\tau) - \Delta / 2 \geq Z_i^T \beta_0 \{ \psi_i(\beta_0, \tau) \} + \Delta - \Delta / 2 = X_i + \Delta / 2,
\]
and
\[
Z_i^T \beta_{SC} \{ \psi_i(\beta_0, \tau) - \Delta_3 \} \leq Z_i^T \beta_0 \{ \psi_i(\beta_0, \tau) - \Delta_3 \} + \Delta / 2
\]
\[
\leq Z_i^T \beta_0 \{ \psi_i(\beta_0, \tau) \} - \Delta + \Delta / 2 = X_i - \Delta / 2.
\]
This implies \( | \{ \psi_i(\beta_0, \tau) - \psi_i(\hat{\beta}_{SC}, \tau) \} | \leq \Delta_3 \). When \( \psi_i(\beta_0, \tau) = \tau < \tau_U \), we have \( Z_i^T \beta_0 (\tau) < X_i \). Then for \( u \in [\Delta_3, \tau - \Delta_3] \),
\[
Z_i^T \beta_{SC} (u) \leq Z_i^T \beta_0 (u) + \Delta / 2 \leq Z_i^T \beta_0 (\tau) - \Delta + \Delta / 2 < X_i - \Delta / 2
\]
This implies that either \( \tau - \Delta_3 \leq \psi(\beta_{SC}, \tau) \leq \tau \) or \( \psi(\hat{\beta}_{SC}, \tau) < \Delta_3 \), and hence \( I \{ \psi(\hat{\beta}, \tau) \geq \Delta_3 \} \{ \psi_i(\beta_0, \tau) - \psi_i(\hat{\beta}_{SC}, \tau) \} \leq \Delta_3 \). Therefore, our claim holds, that is, \( \sup_{i, \tau \in (0, \tau_U]} | I \{ \psi_i(\beta_0, \tau) > 2\Delta_3, \psi_i(\hat{\beta}, \tau) \geq \Delta_3 \} \{ \psi_i(\beta_0, \tau) - \psi_i(\hat{\beta}_{SC}, \tau) \} | \leq \Delta_3 \).
By this result and (A.3), we get, when sup$_{i, \tau \in [\Delta c_3, \tau_U]} |Z_i^\prime \hat{\beta}_{SC}(\tau) - Z_i^\prime \beta_0(\tau)| \leq \Delta/2,$

$$\|A_{1,n}(\tau) + A_{2,n}(\tau) + A_{3,n}(\tau) + A_{4,n}(\tau)\|$$

$$\leq n^{-1/2} \sum_{i=1}^{n} \|Z_i\| \cdot [\Delta c_3 + \Delta c_3 + 2\Delta c_3 + I\{X_i \leq Z_i^\prime \beta_0(2\Delta c_3)\}] = B_{1,n}$$

Let $\theta(\Delta) = E(\|Z_i\| \cdot [4\Delta c_3 + I\{X_i \leq Z_i^\prime \beta_0(2\Delta c_3)\}])$ and $\sigma^2(\Delta) = \text{var}(\|Z_i\| \cdot [2\Delta c_3 + I\{X_i \leq Z_i^\prime \beta_0(2\Delta c_3)\}]).$ It follows from $\lim_{\Delta \to 0} \text{Pr}(X \leq Z^\prime \beta_0(\Delta c_3)) = 0$ that $\lim_{\Delta \to 0} \theta(\Delta) = 0$ and $\lim_{\Delta \to 0} \sigma(\Delta) = 0.$ Then any $\rho > 0$ and $\xi > 0,$ we can find $\Delta \in (0, \tau_{U/2})$ such that $z_{1-\xi/3} \sigma(\Delta) + \theta(\Delta) < \rho,$ where $z_{1-\xi}$ denotes the $100(1-\xi)$th percentile of a standard normal distribution. By the Central Limit Theorem, for $n >$ some $N_{\rho,\xi,1},$ we have $\text{Pr}(B_{1,n} > z_{1-\xi/3} \sigma(\Delta) + \theta(\Delta)) < 2\xi/3.$ Because sup$_{\tau \in [\Delta c_3, \tau_U]} \|\hat{\beta}_{SC}(\tau) - \beta_0(\tau)\| \to p 0$ and $\|Z_i\| < c_1$ for all $i,$ we get $\text{Pr}(\sup_{i, \tau \in [\Delta c_3, \tau_U]} |Z_i^\prime \hat{\beta}_{SC}(\tau) - Z_i^\prime \beta_0(\tau)| > \Delta/2) < \xi/3$ for $n >$ some $N_{\rho,\xi,2}.$ Therefore, for $n > \max(N_{\rho,\xi,1}, N_{\rho,\xi,2}),$

$$\text{Pr}(\sup_{\tau \in (0, \tau_U]} \|\sum_{k=1}^{4} A_{k,n}(\tau)\| > \rho) \leq \text{Pr}(\sup_{\tau \in (0, \tau_U]} \|\sum_{k=1}^{4} A_{k,n}(\tau)\| > z_{1-\xi/3} \sigma(\Delta) + \theta(\Delta))$$

$$\leq \text{Pr}(\sup_{i, \tau \in (0, \tau_U]} |Z_i^\prime \hat{\beta}_{SC}(\tau) - Z_i^\prime \beta_0(\tau)| > \Delta/2) + \text{Pr}(B_{1,n} > z_{1-\xi/3} \sigma(\Delta) + \theta(\Delta)) < \xi.$$

This proves sup$_{\tau \in (0, \tau_U]} \|\sum_{k=1}^{4} A_{k,n}(\tau)\| \to p 0$ and thus completes the proof for Lemma A3.

REFERENCES


Table 1: Empirical biases ($\times 10^3$) and empirical variances ($\times 10^3$) of $\hat{\beta}_{PH}(\tau)$, $\hat{\beta}_{SC}(\tau)$, $\hat{\beta}_{OSC}(\tau)$, $\hat{\beta}_{MSC}(\tau)$, and $\hat{\beta}_{PSC}(\tau)$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\hat{\beta}_{PH}$ Bias</th>
<th>$\hat{\beta}_{OSC}$ Var</th>
<th>$\hat{\beta}_{SC}$ Bias</th>
<th>$\hat{\beta}_{MSC}$ Var</th>
<th>$\hat{\beta}_{PSC}$ Bias</th>
<th>$\hat{\beta}_{PSC}$ Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I) Log linear model with iid errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1 $\hat{\beta}^{(1)}$</td>
<td>3 250 12 251 12 251 12 252</td>
<td>71 272</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>0 588 6 594 6 594 5 597</td>
<td>2 635</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>30 188 33 188 34 189 34 188</td>
<td>33 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3 $\hat{\beta}^{(1)}$</td>
<td>2 87 15 88 15 88 15 88</td>
<td>39 89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>24 204 27 206 27 206 27 206</td>
<td>28 206</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>1 70 1 70 1 70 1 70 1 70</td>
<td>4 72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 $\hat{\beta}^{(1)}$</td>
<td>2 55 13 54 13 54 13 54</td>
<td>32 55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>22 139 24 140 24 140 25 140 25 140 25 140 25 140 25 137</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>5 49 4 50 4 50 4 50 4 50 4 50</td>
<td>3 51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7 $\hat{\beta}^{(1)}$</td>
<td>23 54 2 53 1 53 1 53 1 53 1 53 1 53 1 53 1 53 1 53</td>
<td>21 53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>12 131 14 130 14 130 15 131 15 131 15 131 15 131 15 131</td>
<td>8 130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>10 46 9 45 9 45 9 45 9 45 9 45</td>
<td>3 43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(II) Log linear model with heteroscedastic errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1 $\hat{\beta}^{(1)}$</td>
<td>13 88 9 88 9 88 9 88</td>
<td>26 92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>1 211 3 212 3 212 3 212</td>
<td>4 223</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>1 74 0 74 0 74 0 74 1 74 0 74</td>
<td>2 75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3 $\hat{\beta}^{(1)}$</td>
<td>9 49 1 49 1 49 1 49 1 49 1 49 1 49 1 49</td>
<td>17 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>1 114 2 115 2 115 2 115 2 115</td>
<td>1 117</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>5 43 4 43 4 43 4 43 4 43 4 43</td>
<td>4 43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 $\hat{\beta}^{(1)}$</td>
<td>29 50 12 49 12 49 12 49 12 49</td>
<td>5 49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>18 117 17 116 17 117 17 117</td>
<td>19 114</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>8 40 8 40 8 40 8 40 8 40</td>
<td>5 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7 $\hat{\beta}^{(1)}$</td>
<td>43 68 14 68 14 68 14 68 15 68</td>
<td>7 65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>10 156 5 156 4 156 4 156 4 156</td>
<td>6 152</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>14 51 14 50 15 50 15 50</td>
<td>12 48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\hat{\beta}^{(i)}$: the i-th component of $\hat{\beta}_{PH}$, $\hat{\beta}_{SC}$, $\hat{\beta}_{OSC}$, $\hat{\beta}_{MSC}$, or $\hat{\beta}_{PSC}$. 
Table 2: Empirical 25th percentiles ($\times 10^3$) and empirical 75th percentiles ($\times 10^3$) of $\hat{\beta}_{PH} - \hat{\beta}_{SC}$, $\hat{\beta}_{OSC} - \hat{\beta}_{SC}$, $\hat{\beta}_{MSC} - \hat{\beta}_{SC}$, and $\hat{\beta}_{PSC} - \hat{\beta}_{SC}$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\hat{\beta}<em>{PH} - \hat{\beta}</em>{SC}$ Diff25</th>
<th>$\hat{\beta}<em>{OSC} - \hat{\beta}</em>{SC}$ Diff25</th>
<th>$\hat{\beta}<em>{MSC} - \hat{\beta}</em>{SC}$ Diff25</th>
<th>$\hat{\beta}<em>{PSC} - \hat{\beta}</em>{SC}$ Diff25</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I) Log linear model with iid errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1 $\hat{\beta}^{(1)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.3 $\hat{\beta}^{(1)}$</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5 $\hat{\beta}^{(1)}$</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>-6</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.7 $\hat{\beta}^{(1)}$</td>
<td>0</td>
<td>39</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>-18</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>-13</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(II) Log linear model with heteroscedastic errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1 $\hat{\beta}^{(1)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.3 $\hat{\beta}^{(1)}$</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5 $\hat{\beta}^{(1)}$</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>-4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>-7</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.7 $\hat{\beta}^{(1)}$</td>
<td>0</td>
<td>48</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(2)}$</td>
<td>-27</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\hat{\beta}^{(3)}$</td>
<td>-14</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$\hat{\beta}^{(i)}$: the $i$th component of $\hat{\beta}_{PH}$, $\hat{\beta}_{SC}$, $\hat{\beta}_{OSC}$, $\hat{\beta}_{MSC}$, or $\hat{\beta}_{PSC}$. 

22