Biostatistics (2013), **0**, 0, *pp.* 1–14
doi:10.1093/biostatistics/TSWLatexianTemp˙000042

Censored Quantile Regression with Recursive Partitioning Based Weights: "Supplementary Materials"

ANDREW WEY*,1, LAN WANG[2], KYLE RUDSER[1]

1. Division of Biostatistics, School of Public Heath, University of Minnesota

2. School of Statistics, University of Minnesota

1. Consistency

This section shows that recursive partitioning based weights lead to consistent estimation of regression quantiles. We first introduce some notation and regularity conditions, then show that survival trees are uniformly consistent for conditional survival functions on a certain support.

In order to clearly state the regularity conditions, some concepts from the tree literature are introduced. Consider the partition $Q^{(n)}$ of the covariate space, i.e., as produced by a tree, then $B_k^{(n)}$ is the k^{th} *box*, or terminal node, of $Q^{(n)}$ such that $\bigcup_k B_k^{(n)} = Q^{(n)}$. Now define the *mesh*, or diameter, of the box k as

$$D_n(k) = \sup\{||y - z||, \text{ such that y, z } \epsilon \ B_k^{(n)}\}, \tag{1.1}$$

where it is assumed that $B_k^{(n)}$ is contained within the support of $\vec{x}$ for all k. Define $\hat{F}(t|B_k^{(n)})$ as the within terminal node cumulative distribution estimator for all $\vec{x} \ \epsilon \ B_k^{(n)}$. We adopt the following conditions:

A1. For $\beta(\tau)$ in the neighborhood of $\beta_o(\tau)$, $E[\vec{x}\vec{x}f_T(\vec{x}\beta(\tau)|\vec{x})\{1 - F_C(\vec{x}\beta(\tau)|\vec{x})\}]$ is positive def-

inite.

A2. There exists a constant $K_{\vec{x}}$ such that $E[||\vec{x}||^3] \leqslant K_{\vec{x}}$. In addition, $\max_{1 \leqslant i \leqslant n} ||\vec{x}_i|| = O_p(n^{1/2}(\log n)^{-1})$, and $E[\vec{x}\vec{x}]$ is a positive definite $p \times p$ matrix.

A3. The conditional distribution functions $F_T(t|\vec{x})$ and $F_C(t|\vec{x})$ have first derivatives with respect to t, $f_T(t|\vec{x})$ and $f_C(t|\vec{x})$ respectively, which are uniformly bounded away from infinity. Also, $F_T(t|\vec{x})$ and $F_C(t|\vec{x})$ have bounded (uniformly in t) second-order partial derivatives with respect to $\vec{x}$.

A4. $E[|T|^r] < \infty$ for some $r > 1$ and $Q^{(n+1)}$ is a refinement of $Q^{(n)}$. Let $k(n)$ be a nondecreasing sequence of integers which approaches infinity, $n^{1/r} \log n/k(n)$, $D_n(\vec{x})$, and $I_{\hat{H}_n(B_k^{(n)}) < k(n)/n}$ approach 0 as $n \to \infty$, where $\hat{H}_n(B_k^{(n)})$ is the empirical probability of box k on the sample.

A5. Let $\xi > 0$ and for any fixed $t < \xi$, $\hat{F}(t|B_k^{(n)}) \to F(t|B_k)$, in probability.

Assumptions A1 and A2 are common to the censored quantile regression literature, see for example Wang and Wang (2009). Assumptions A3, A4 and A5 imply that the conditional distribution function converges uniformly in t. Assumption A4 ensures the conditions for theorems needed from Gordon and Olshen (1984) are satisfied. Assumption A5 requires that the within terminal node cumulative distribution estimator is pointwise consistent on a subset of the support, where ξ is typically chosen to ensure a well defined survival distribution.

In Chen *et al.* (2003), "*Theorem 1*" states five sufficient conditions for consistency. Condition (1.3) is satisfied by assumptions while conditions (1.1), (1.2) and (1.5') are satisfied by identical arguments used by Wang and Wang (2009). Condition (1.4) requires us to show that survival tree estimators, denoted $\hat{F}_T(t|\vec{x})$, that use a general within node cumulative distribution estimator,

say $\hat{F}(t|B^{(n)})$, is uniformly consistent for $F_T(t|\vec{x})$. Define the following quantities as

$$F_Y(t|\vec{x}) = P_Y(Y \leqslant t|\vec{x}) = 1 - (1 - F_T(t|\vec{x}))(1 - F_C(t|\vec{x})),$$
$$F_{T1}(t|\vec{x}) = P_{T1}(T \leqslant t, \delta = 1|\vec{x}) = \int_0^t (1 - F_C(u|\vec{x}))f_T(u|\vec{x})du,$$

which can be thought as the observed distribution function and observed event distribution function, respectively. Consider the survival tree partitioning the covariate space into a set of 'boxes' denoted by B_N. Now define the distribution estimators for $F_Y(t|\vec{x})$ and $F_{T1}(t|\vec{x})$ as $\hat{F}_Y(t|\vec{x})$ and $\hat{F}_{T1}(t|\vec{x})$, respectively.

For fixed $t > 0$, define two types of convergence as described in Gordon and Olshen (1984),

(i) $E\{|\hat{F}_n(t|\vec{x}, \zeta_n) - F(t|\vec{x})||\zeta_n\} \to 0$, and

(ii) $\hat{F}_n(t|\vec{x}, \zeta_n) - F(t|\vec{x}) \to 0$, almost surely,

where ζ_n is the observed data, and the set $\Omega = \{\xi : F_Y(\xi|\vec{x}) < 1 - \delta\}$ for $\delta > 0$. Now we can concisely state the following lemma,

LEMMA 1.1 Let the conditional distribution functions $F_{T1}(t|\vec{x})$ and $F_C(t|\vec{x})$ be continuous. For all $s \leqslant \xi$, where $\xi > 0$ is a fixed time, $\hat{F}_Y(s|\vec{x})$ and $\hat{F}_{T1}(s|\vec{x})$ are recursive partitioning based estimators that are type (i) and/or (ii) consistent tree estimators for $F_Y(s|\vec{x})$ and $F_{T1}(s|\vec{x})$, respectively, for each single s. Then

$$\int_0^\xi \frac{d\hat{F}_{T1}(s|\vec{x})}{1 - \hat{F}_Y(s|\vec{x})} \to -\log\{1 - F_T(\xi|\vec{x})\},$$

almost surely on Ω.

This lemma is *Theorem 1* from Butler *et al.* (1989). Assumptions (A4) and (A5) combined with either *Theorem 3.6* or *Theorem 4.1* from Gordon and Olshen (1984) satisfy the conditions for *Lemma 1.1*. Now define "*Lemma 1*" from Breslow and Crowley (1974),

LEMMA 1.2 Let $N(t) = \Sigma_{i=1}^{N} I[Y_i \geqslant t]$ be the number of individuals still "at risk" at time t. Then with probability 1, for all $0 < t < max_{1 \leqslant i \leqslant N} Y_i$,

$$0 < -\log\{1 - \hat{F}_T(t|\vec{x})\} - \int_0^{\tau} \frac{d\hat{F}_{T1}(s|\vec{x})}{1 - \hat{F}_Y(s|\vec{x})} < \frac{N - N(t)}{N \cdot N(t)}.$$

Lemma 1.1 and *Lemma 1.2* imply that for all $t \, \epsilon \, \Omega$,

$$-\log\{1 - \hat{F}_T(t|\vec{x})\} \to -\log\{1 - F_T(t|\vec{x})\},$$

as $N(t) \to \infty$. This means that $\hat{F}_T(t|\vec{x}) \to F_T(t|\vec{x})$ at all the continuity points of F_T on Ω. Thus, by definition,

$$\hat{F}_T(t|\vec{x}) \to_D F_T(t|\vec{x}),$$

in distribution. By problem 1.6 of Ferguson (1996), convergence in law with previously stated conditions implies uniform convergence which satisfies condition (1.4) of Chen *et al.* (2003). This completes the proof for consistency.

2. Simulation Extensions

The simulation study presented in the paper illustrated the potential of tree based weighting compared to current methods. While designing the simulations, we ran into interesting issues regarding censored quantile regression in general, and situations that are better for using tree based weights for censored quantile regression. The paper introduced and summarized these issues, but we present and discuss them in more detail here. In particular, we first discuss the effect of the censoring distribution on the weights used in (2.3), then intuition is introduced for the better performance of tree based weights for quantiles of interest further from the marginal censoring rate. In Supplementary Section 2.3, smooth estimators of the conditional distribution function are considered, while the effect of mild non-linearity is investigated in Supplementary Section 2.4. Finally, results using a larger sample size are presented. All table references refer to

the Supplementary Materials unless stated otherwise.

2.1 *Reweighting Effect*

A small simulation study was performed to evaluate the effect of the censoring location on bias and mean squared error (MSE), and the proportion of reweighted observations that are reweighted for each situation. These simulations are univariate versions of the ones presented in section 4 of the main paper. In particular, we introduce non-linearity in all quantiles except the quantile of interest through a quadratic error term. In particular,

$$t_i = 2 + x_i + (\frac{1}{2} + c \cdot (x_i - \frac{1}{2})^2)(N(0,1) - \Phi^{-1}(\tau)),$$
$$x_i \sim N(0,1).$$

We consider two types of censoring: an unconditionally independent form of censoring and a conditionally independent (on x_i) form of censoring. Let $c_{i,1}$ and $c_{i,2}$ be the censoring times for unconditionally independent censoring and conditionally independent censoring, respectively,

$$c_{i,1} \sim Unif(a_1(s,l), b_1(s,l)),$$
$$c_{i,2} \sim Unif(a_2(s,l), b_2(s,l)) \times (1 + (x_i - \frac{1}{2})^2),$$

where $a_k(s,l)$ and $b_k(s,l)$ are chosen to ensure 35% censoring, and s and l denote the severity of non-linearity and location of censoring, respectively.

The results (Table 1) suggest that the bias of Portnoy's estimator does not change with the censoring distribution when the linearity assumption in all quantiles is satisfied, but the bias can vary radically when non-linearity is present and the differences increase with the degree of non-linearity. Note that the bias is typically the least when the censoring is 'middle' located. There is limited intuition for this observation. Although, there the censored observations that require reweighting, i.e., $F_T(c_i|\vec{x}_i) < \tau_o$, is probably clearer when the censoring is not concentrated around the quantile of interest.

2.2 *Higher Censoring Effect on Median*

Simulations similar to the ones described in Section 4 are presented in Tables 2 and 3 for the median. The only difference is that $a(C_k, E_l)$ was chosen to ensure 35% instead of 25% censoring. The results show that the bias for the tree based estimator increases with increased censoring. This increased bias is similar or slightly larger than the Portnoy and/or PH estimator's bias, but the three methods perform similarly for mild non-linearity. For severe non-linearity, the tree based estimator performs better than the Portnoy and PH estimators.

This observation makes intuitive sense. In particular, our stopping rules strictly require that the quantile of interest is defined by setting the minimum number of events to the number at risk within a node times the quantile of interest. When the marginal censoring level is close to the quantile of interest, then the set of potential splits is substantially reduced. The survival tree will become closer to an univariate Kaplan-Meier when the number of potential splits is small. A potential solution would set the minimum number of events and a direct condition that requires the quantile of interest to be defined.

2.3 *Using Smooth Local Kaplan-Meier Estimators*

As pointed out by a reviewer, smooth estimators, e.g., hazard regression with splines, may be more efficient at estimating the conditional distribution function. This idea is close in spirit to the Wang and Wang (2009) (WW) estimator that uses kernel estimators to smooth the Kaplan-Meier estimator across covariates. We implemented the WW estimator for the two variable simulations presented in the paper (Section 4, Table 1). A product kernel, i.e., $K(x_1, x_2) = K_1(\frac{x_1 - x_{1,o}}{h_1}) \times K_2(\frac{x_2 - x_{2,o}}{h_2})$, was used with bandwidths determined on one simulated dataset with five-fold cross-validation.

Table 4 presents similar results when the global linearity assumption is satisfied, but the results are underwhelming when the global linearity assumption is violated. Since the bandwidth

was determined in an ad-hoc fashion that did not adapt to each simulation, these results could almost surely be improved. Unfortunately, the computational requirements to search for tuning parameters across multiple dimensions are too intensive to do across every simulation iteration. Additionally, the three variable scenario considered in the paper presents substantially issues for the WW estimator due to the discrete covariate.

Wey (2011) conducted several similar simulations that included the WW estimator; a univariate scenario is introduced here

$$
\begin{aligned}
t_i &= 2 + x_i + (\frac{1}{5} - 6 \cdot (x_i - \frac{1}{2})^2)(N(0,1) - \Phi^{-1}(\tau)), \\
x_i &\sim N(0,1), \\
c_i &\sim Unif(0,8).
\end{aligned}
$$

Only consider the median, set $n = 200$, minimum at risk is 8 and minimum events the number of observations at risk in a node times the quantile of interest. As Table 5 shows, the tree based estimator does substantially better than the Portnoy and PH estimators.

2.4 *Effect of Mild Non-Linearity*

The scenarios presented in the paper possess severe non-linearity, but simulations with mild non-linearity displayed similar results. Considering the simulation scenario considered in the paper, the error term that induces mild non-linearity is $\frac{3}{2} + 2 \cdot (x_{i,1} - \frac{1}{2})^2$ while the censoring distribution is $C \sim (\frac{3}{10} + (x_{i,1} - \frac{1}{2})^2) \times Unif(-3.75, a(\Omega_k, E_l))$. The results (Table 6) indicate that the tree based estimator possesses less bias when estimating the median, while the tree based estimator performs similar to the Portnoy and Peng, Huang estimators for $\tau = 0.25$.

2.5 *Large Sample Simulations*

The simulations presented in Section 4 are extended with twice the sample size ($N = 800$). Tables 7 and 8 present the results for the two and three variable simulation scenarios, respectively. Generally, the tree based estimator experiences a small reduction in bias, particularly for scenarios possessing non-linearity. The Portnoy and PH estimators continue to experience significant difficulties with the non-linearity scenario, while the three estimators are nearly equivalent for the linear in all quantiles scenario. Additionally, the proposed estimator possesses better coverage probabilities than the Portnoy and PH estimators. Overall, the tree based estimator continues to perform better than the Portnoy and PH estimators when the global linearity assumption is violated.

3. Extending the PBC Analysis

This section focuses on two things: describing how to use the bagged trees to estimate the conditional quantile function, then analyze the validity of the linearity assumption in all quantiles up to, and including, the quantile of interest using the bagged survival tree.

A reviewer pointed out that bagged survival trees can be used to non-parametrically estimate the conditional quantile function. It is important to point out that without the stopping rules considered in this paper, then the estimated quantile function from traditional bagged survival trees are not guaranteed to be defined (i.e., ill-defined survival function). This non-parametric estimate can only be used for prediction. If we want to determine covariate associations with the event distribution T, then additional steps are required. Following Rudser *et al.* (2012), quantile time (for τ) for each observation in the data set is predicted. Let the predicted time for the i^{th} subject be $q_{\vec{x}_i}(\tau)$. Then the marginal linearity of the covariates is checked by using an univariate smoother (without controlling for other covariates). By using a smoother on $\vec{q}_{\vec{x}}(\tau)$, we obtain the

marginal quantile function for x as

$$\hat{Q}_T(\tau|x) = \hat{g}_\tau(x),$$

where g is a smooth function and x is a single covariate being checked (e.g., $\hat{g}$ kernel smoothes across $\vec{q}_{\vec{x}}(\tau)$).

To non-parametrically estimate the conditional quantile function for the PBC data set, local linear regression is used to marginally estimate g for age, bilirubin, and prothrombin time. The bandwidth is chosen using generalized cross-validation. Figure 3 presents the results for three quantiles up to $\tau = 0.25$. There exists obvious non-linearity for bilirubin and prothrombin time. Attenuated non-linearity remains for bilirubin for $\tau = 0.25$, while prothrombin time is nearly linear by τ. Age remains linear throughout the quantiles considered. These results suggest that the Portnoy (2003) and Peng and Huang (2008) estimators may be inappropriate for the PBC data set due to non-linearity in all quantiles. Some care is advised as the marginal interpretation is a significant limitation of this approach.

References

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics* **2**, 437–453.

Butler, Jeffrey H., Gilpin, Elizabeth A., Gordon, Louis and Olshen, Richard A. (1989). Tree-structured survival analysis, ii. *Technical Report* 133, Division of Biostatistics, Stanford University, Stanford University.

Chen, Xiaohong, Linton, Oliver and Van Keilegom, Ingrid. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71**, 1591–1608.

Ferguson, Thomas S. (1996). *A course in large sample theory*. Chapman and Hall/CRC.

Gordon, Louis and Olshen, Richard A. (1984). Almost surely consistent nonparametric regression from recursive partitioning schemes. *Journal of Multivariate Analysis* **15**, 147–163.

Peng, Limin and Huang, Yijian. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association* **103**, 637–649.

Portnoy, Stephen. (2003). Censored regression quantiles. *Journal of the American Statistical Association* **98**, 1001–1012.

Rudser, Kyle D., LeBlanc, Michael L. and Emerson, Scott S. (2012). Distribution-free inference on contrasts of arbitrary summary measures of survival. *Statistics in Medicine* **31**, 1722–1737.

Wang, Huixia Judy and Wang, Lan. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association* **103**, 1117–1128.

Wey, Andrew. (2011). Recursive Partitioning Weights for Censored Quantile Regression [Master's Thesis]. University of Minnesota, Twin Cities.

		Independent Censoring			Conditionally Independent Censoring		
Non-Linearity	Metric	Late	Mid	Early	Late	Mid	Early
No Non-Linearity	Bias	0.03	0.04	0.04	0.04	0.04	0.04
	MSE	0.66	0.78	0.83	0.68	0.85	0.87
	Total Percent (Re-weighted)	10.2	24.8	27.4	16.5	23.3	18.4
	Relative Percent (Re-weighted)	31.1	72.2	83.3	47.0	80.0	83.7
Mild Non-Linearity	Bias	-0.11	-0.06	0.12	0.06	0.02	0.09
	MSE	0.18	0.23	0.32	0.23	0.25	0.31
	Total Percent (Re-weighted)	8.1	19.7	23.6	16.9	23.0	27.1
	Relative Percent (Re-weighted)	25.5	58.8	90.9	49.2	66.8	91.4
Severe Non-Linearity	Bias	-0.11	-0.09	0.32	0.17	0.02	0.26
	MSE	0.67	0.86	1.44	0.96	1.00	1.33
	Total Percent (Re-weighted)	7.4	17.1	28.3	6.3	18.1	29.6
	Relative Percent (Re-weighted)	22.5	51.4	91.8	20.0	54.3	86.7

Table 1. *Effect of Censoring Locations:* $N = 400$*,* $N_{SIM} = 2500$*, censoring* $= 35\%$*,* $\beta_0 = 2$*,* $\beta_1 = 1$*. Recall only censored observations can be reweighted.*

□

			No Non-Linearity					Severe Non-Linearity				
Quantile	Variable	Method	Bias	MSE	Cov.	ECL	Power	Bias	MSE	Cov.	ECL	Power
$\tau = 0.5$	Variable 1 $\beta_1 = 1$	Portnoy	0.00	0.03	0.96	0.74	1.00	0.10	0.56	0.95	2.94	0.32
		PH	0.00	0.04	0.96	0.76	1.00	-0.08	0.57	0.95	2.99	0.23
		TW	-0.01	0.03	0.97	0.75	1.00	0.11	0.56	0.96	3.00	0.33
	Variable 2 $\beta_2 = -2$	Portnoy	0.00	0.05	0.96	0.88	1.00	-0.24	0.24	0.94	1.84	1.00
		PH	0.00	0.05	0.97	0.89	1.00	-0.29	0.29	0.93	1.95	1.00
		TW	0.04	0.05	0.97	0.89	1.00	-0.13	0.25	0.97	2.04	1.00

Table 2. *Simulation Scenario for High Censoring (2 covariates):* $N = 400$, $N_{SIM} = 2500$, *censoring* $= 35\%$, $\beta_0 = 2$, $\beta_1 = 1$, $\beta_2 = -2$, 300 *bootstrap replicates,* 95% *nominal coverage with ECL representing the average CI width.*

			No Non-Linearity					Severe Non-Linearity				
Quantile	Variable	Method	Bias	MSE	Cov.	ECL	Power	Bias	MSE	Cov.	ECL	Power
$\tau = 0.5$	Variable 1 $\beta_1 = 1$	Portnoy	-0.01	0.04	0.96	0.77	1.00	0.12	0.62	0.95	3.06	0.30
		PH	-0.01	0.04	0.96	0.78	1.00	-0.05	0.61	0.96	3.12	0.22
		TW	-0.01	0.04	0.97	0.78	1.00	0.10	0.61	0.95	3.13	0.30
	Variable 2 $\beta_2 = -2$	Portnoy	-0.01	0.05	0.96	0.90	1.00	-0.24	0.28	0.94	2.00	0.99
		PH	-0.01	0.05	0.96	0.91	1.00	-0.28	0.33	0.93	2.11	0.99
		TW	0.01	0.05	0.97	0.91	1.00	-0.17	0.29	0.96	2.19	0.99
	Variable 3 $\beta_3 = 1$	Portnoy	0.00	0.05	0.97	0.93	0.99	0.11	0.23	0.96	2.01	0.62
		PH	0.00	0.05	0.97	0.94	0.99	0.13	0.26	0.96	2.14	0.60
		TW	0.00	0.05	0.98	0.93	0.99	0.08	0.23	0.97	2.09	0.59

Table 3. *Simulation Scenario for High Censoring (3 covariates):* $N = 400$, $N_{SIM} = 2500$, *censoring* $= 35\%$, $\beta_0 = 2$, $\beta_1 = 1$, $\beta_2 = -2$, $\beta_3 = 1$, 300 *bootstrap replicates,* 95% *nominal coverage with ECL representing the average CI width.*

			No Non-Linearity					Severe Non-Linearity				
Quantile	Variable	Method	Bias	MSE	Cov.	ECL	Power	Bias	MSE	Cov.	ECL	Power
$\tau = 0.25$	Variable 1	WW	-0.04	0.04	0.96	0.79	1.00	0.24	0.64	0.95	3.02	0.34
	Variable 2	WW	0.09	0.06	0.94	0.93	1.00	0.23	0.23	0.93	1.80	0.99
$\tau = 0.5$	Variable 1	WW	-0.03	0.03	0.96	0.69	1.00	0.13	0.48	0.95	2.71	0.39
	Variable 2	WW	0.06	0.04	0.95	0.80	1.00	0.14	0.13	0.94	1.41	1.00

Table 4. *Simulation Scenario for the Wang and Wang (2009) estimator (2 covariates):* $N = 400$, $N_{SIM} = 2500$, $\beta_0 = 2$, $\beta_1 = 1$, $\beta_2 = -2$, 300 *bootstrap replicates,* 95% *nominal coverage with ECL representing the average CI width. Censoring is approximately* 25% *and* 35% *for* $\tau = 0.5$ *and* $\tau = 0.25$*, respectively.*

Method	Bias	MSE	Cov.	ECL	Power
Portnoy	-0.10	0.49	0.95	2.75	0.23
PH	-0.40	0.63	0.91	2.73	0.14
TW	-0.10	0.32	0.95	2.25	0.35
WW	0.00	0.33	0.96	2.30	0.45

Table 5. *Motivating Univariate Simulation Scenario:* $N = 200$, $N_{SIM} = 2500$, *censoring* $= 36\%$, $\beta_0 = 2$, $\beta_1 = 1$, 300 *bootstrap replicates,* 95% *nominal coverage with ECL representing the average CI width.*

			Two Variable Scenario					Three Variable Scenario				
Quantile	Variable	Method	Bias	MSE	Cov.	ECL	Power	Bias	MSE	Cov.	ECL	Power
0.25	Variable 1	Portnoy	0.07	0.15	0.96	1.51	0.81	0.06	0.15	0.95	1.54	0.78
	$\beta_1 = 1$	PH	0.00	0.14	0.96	1.50	0.76	-0.01	0.15	0.95	1.53	0.73
		TW	-0.03	0.13	0.96	1.47	0.76	-0.05	0.14	0.96	1.51	0.72
	Variable 2	Portnoy	-0.07	0.09	0.96	1.20	1.00	-0.06	0.09	0.97	1.26	1.00
	$\beta_2 = -2$	PH	-0.09	0.09	0.96	1.20	1.00	-0.08	0.09	0.97	1.26	1.00
		TW	0.03	0.09	0.97	1.21	1.00	0.02	0.09	0.97	1.29	1.00
	Variable 3	Portnoy						0.03	0.09	0.97	1.27	0.91
	$\beta_3 = 1$	PH						0.04	0.09	0.96	1.27	0.92
		TW						-0.09	0.09	0.97	1.23	0.88
0.5	Variable 1	Portnoy	0.05	0.11	0.96	1.32	0.89	0.05	0.11	0.95	1.33	0.88
	$\beta_1 = 1$	PH	-0.01	0.11	0.96	1.33	0.84	-0.02	0.11	0.96	1.34	0.82
		TW	0.02	0.11	0.96	1.33	0.86	0.01	0.11	0.96	1.34	0.86
	Variable 2	Portnoy	-0.05	0.05	0.97	0.95	1.00	-0.04	0.06	0.95	0.98	1.00
	$\beta_2 = -2$	PH	-0.06	0.06	0.97	0.97	1.00	-0.05	0.06	0.95	1.00	1.00
		TW	-0.02	0.06	0.97	0.98	1.00	-0.02	0.06	0.96	1.02	1.00
	Variable 3	Portnoy						0.01	0.06	0.97	1.00	0.98
	$\beta_3 = 1$	PH						0.01	0.06	0.97	1.03	0.98
		TW						0.01	0.06	0.97	1.03	0.98

Table 6. *Simulation Scenario for Mild Non-Linearity (2 and 3 covariates):* $N = 400$, $N_{SIM} = 2500$, *censoring* $= 25\%$, $\beta_0 = 2$, $\beta_1 = 1$, $\beta_2 = -2$, $\beta_3 = 1$, 300 *bootstrap replicates,* 95% *nominal coverage with ECL representing the average CI width.*

			No Non-Linearity					Non-Linearity				
Quantile	Variable	Method	Bias	MSE	Cov.	ECL	Power	Bias	MSE	Cov.	ECL	Power
0.25	Variable 1	Portnoy	0.00	0.02	0.96	0.58	1.00	0.13	0.35	0.94	2.27	0.51
	$\beta_1 = 1$	PH	0.00	0.02	0.96	0.59	1.00	-0.02	0.33	0.95	2.28	0.39
		TW	-0.03	0.02	0.96	0.58	1.00	0.01	0.31	0.95	2.22	0.46
	Variable 2	Portnoy	0.00	0.03	0.96	0.68	1.00	-0.19	0.15	0.93	1.38	1.00
	$\beta_2 = -2$	PH	0.00	0.03	0.96	0.69	1.00	-0.23	0.17	0.92	1.42	1.00
		TW	0.07	0.03	0.95	0.68	1.00	0.05	0.11	0.97	1.37	1.00
0.5	Variable 1	Portnoy	0.00	0.02	0.95	0.50	1.00	0.09	0.27	0.94	1.99	0.58
	$\beta_1 = 1$	PH	0.00	0.02	0.95	0.51	1.00	-0.04	0.27	0.94	2.01	0.48
		TW	0.00	0.02	0.95	0.50	1.00	0.04	0.28	0.94	2.03	0.56
	Variable 2	Portnoy	0.00	0.02	0.95	0.58	1.00	-0.13	0.09	0.92	1.05	1.00
	$\beta_2 = -2$	PH	0.00	0.02	0.96	0.59	1.00	-0.15	0.09	0.92	1.09	1.00
		TW	0.01	0.02	0.96	0.59	1.00	-0.02	0.07	0.96	1.07	1.00

Table 7. *First simulation scenario:* $N = 800$, $N_{SIM} = 2500$, *censoring is* 45% *and* 25% *for* $\tau = 0.25$ *and* $\tau = 0.5$, *respectively,* $\beta_0 = 2$, $\beta_1 = 1$, $\beta_2 = -2$, 300 *bootstrap replicates,* 95% *nominal coverage with ECL representing the average CI width.*

			No Non-Linearity					Non-Linearity				
Quantile	Variable	Method	Bias	MSE	Cov.	ECL	Power	Bias	MSE	Cov.	ECL	Power
	Variable 1	Portnoy	0.00	0.02	0.96	0.59	1.00	0.14	0.38	0.94	2.34	0.48
	$\beta_1 = 1$	PH	0.00	0.02	0.96	0.60	1.00	-0.02	0.35	0.95	2.34	0.38
		TW	-0.02	0.02	0.97	0.60	1.00	0.00	0.34	0.95	2.32	0.41
	Variable 2	Portnoy	0.00	0.03	0.96	0.69	1.00	-0.15	0.14	0.94	1.46	1.00
0.25	$\beta_2 = -2$	PH	0.00	0.03	0.97	0.70	1.00	-0.19	0.16	0.94	1.50	1.00
		TW	0.05	0.03	0.96	0.70	1.00	0.05	0.12	0.97	1.45	1.00
	Variable 3	Portnoy	0.00	0.03	0.96	0.71	1.00	0.09	0.13	0.96	1.45	0.88
	$\beta_3 = 1$	PH	0.00	0.03	0.96	0.72	1.00	0.11	0.14	0.96	1.49	0.88
		TW	-0.07	0.03	0.95	0.70	1.00	-0.13	0.11	0.96	1.32	0.80
	Variable 1	Portnoy	0.00	0.02	0.96	0.51	1.00	0.09	0.29	0.95	2.05	0.55
	$\beta_1 = 1$	PH	0.00	0.02	0.96	0.52	1.00	-0.05	0.28	0.95	2.07	0.44
		TW	0.00	0.02	0.96	0.51	1.00	0.03	0.28	0.96	2.08	0.52
	Variable 2	Portnoy	0.00	0.02	0.95	0.59	1.00	-0.11	0.08	0.95	1.11	1.00
0.5	$\beta_2 = -2$	PH	0.00	0.02	0.95	0.60	1.00	-0.13	0.09	0.95	1.15	1.00
		TW	0.01	0.02	0.96	0.60	1.00	-0.02	0.08	0.96	1.14	1.00
	Variable 3	Portnoy	0.00	0.02	0.95	0.61	1.00	0.05	0.08	0.96	1.12	0.96
	$\beta_3 = 1$	PH	0.00	0.02	0.95	0.62	1.00	0.06	0.08	0.96	1.16	0.95
		TW	0.01	0.02	0.96	0.62	1.00	0.01	0.07	0.96	1.13	0.95

Table 8. *Second simulation scenario:* $N = 800$*,* $N_{SIM} = 2500$*, censoring is* 45% *and* 25% *for* $\tau = 0.25$ *and* $\tau = 0.5$*, respectively,* $\beta_0 = 2$*,* $\beta_1 = 1$*,* $\beta_2 = -2$*,* $\beta_3 = 1$*,* 300 *bootstrap replicates,* 95% *nominal coverage with ECL representing the average CI width.*

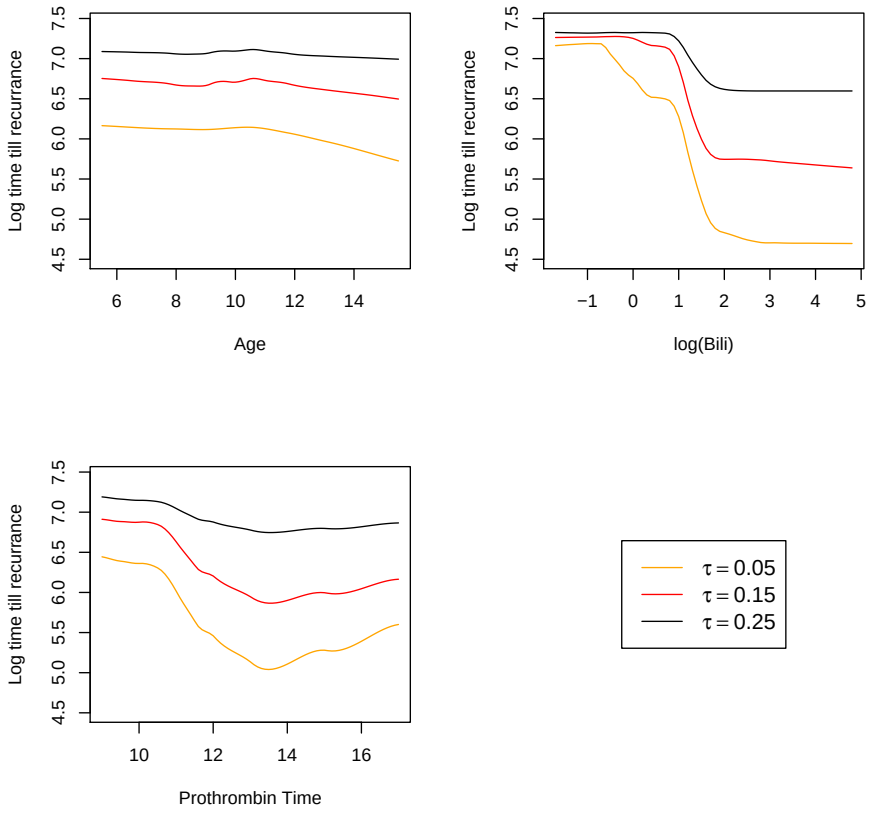

Fig. 1. The marginal quantile relationships for age, log(bilirubin), and prothrombin time. The quantile functions were estimated using the bagged survival trees ($bagN = 10$).