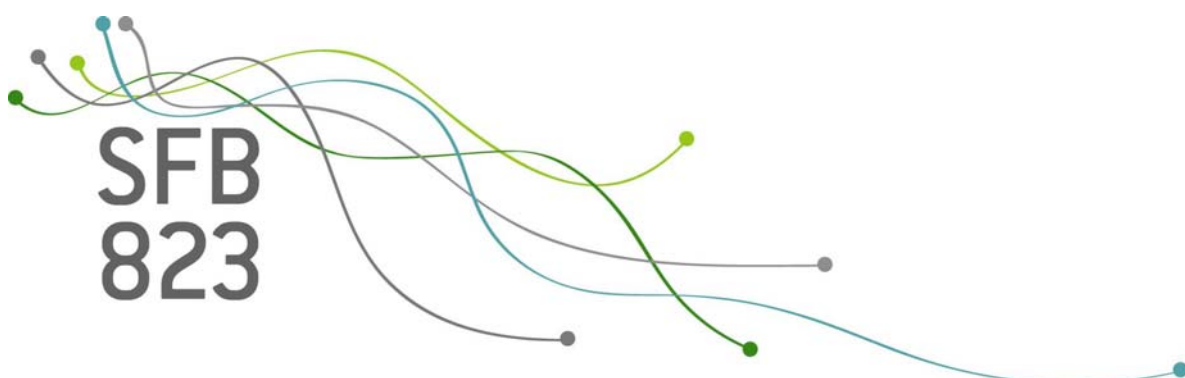# The quantile process under random censoring

Jens Wagener, Stanislav Volgushev, Holger Dette

The quantile process under random censoring

Jens Wagener, Stanislav Volgushev, Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: stanislav.volgushev@rub.de

February 22, 2012

Abstract

In this paper we discuss the asymptotical properties of quantile processes under random censoring. In contrast to most work in this area we prove weak convergence of an appropriately standardized quantile process under the assumption that the quantile regression model is only linear in the region, where the process is investigated. Additionally, we also discuss properties of the quantile process in sparse regression models including quantile processes obtained from the Lasso and adaptive Lasso. The results are derived by a combination of modern empirical process theory, classical martingale methods and a recent result of Kato (2009).

1 Introduction

Quantile regression was introduced by Koenker and Bassett (1978) and provides an important alternative to classical least squares analysis by focusing on conditional quantiles instead of condition mean and variance. Since its introduction it has found considerable attention in the literature because of its flexibility, easy interpretation and robustness properties [see Yu et al. (2003) or Koenker (2005) for some recent review of this field]. While most of the literature discusses properties and applications of quantile regression in the case of fully observed data, much less effort

has been spent to develop quantile regression analysis for censored data. Powell (1984), Powell (1986) and Newey and Powell (1990) proposed quantile regression methods in the case where all censoring variables are known [see also Fitzenberger (1997)]. Ying et al. (1995) introduced median regression in the presence of independent right censoring. This research was continued by Bang and Tsiatis (2002) and Zhou (2006), who derived various inverse-censoring-probability weighted methods for parameter estimation in median regression. However, none of those authors considered quantile processes. Portnoy (2003) and Portnoy and Lin (2010) avoided the rather strong assumption of unconditional independence between survival and censoring times by adopting the principle of self-consistency for the Kaplan-Meier estimate [see Efron (1967)]. An alternative quantile regression method for survival data subject to conditionally independent censoring was developed by Peng and Huang (2008). These authors proposed to use martingale based estimating equations and showed uniform consistency and weak convergence (of an appropriately standardized) quantile process. Wang and Wang (2009) pointed out that the methods of Portnoy (2003) and Peng and Huang (2008) require the conditional quantile curves at lower quantiles to be linear. In order to relax these assumptions they considered locally weighted censored quantile regression estimates that adopt the redistribution-of mass idea and employ a local re-weighting scheme. Closely related approaches were recently investigated by Leng and Tong (2012) and Tang et al. (2012). However, all methods mentioned above require non-parametric smoothing and are thus only of limited use for covariate dimensions larger than 3 or 4 due to the curse of dimensionality. Moreover, these authors did not address the problem of weak convergence of the quantile process.

If the dimension of the parameter is large compared to the sample size, estimating the parameters in quantile regression is intrinsically harder. More recently sparse estimation of quantile regression has found considerable interest in the literature [see Zou and Yuan (2008) or Wu and Liu (2009) among others]. On the other hand - to the best knowledge of the authors - there are only two recent papers, which discuss sparse estimation problems in the context of censored quantile regression [see Shows et al. (2010) and Wang et al. (2012)], and properties of the quantile process in this context have not been studied so far.

The present research has two main purposes. In the first part of the paper we provide an alternative analysis of the quantile process compared to the work of Peng and Huang (2008) and Portnoy and Lin (2010). In contrast to these references the methodology provided here only requires a linear quantile regression model in the region, where the properties of the quantile regression process are investigated. This greater flexibility in the modeling part comes with the price that our approach requires unconditional independence between survival and censoring

times. We also provide a proof for the consistency of a related resampling procedure. In the second part of this paper we investigate the properties of the quantile process in sparse regression models under random censoring, which has - to the best knowledge of the authors - not been studied before.

The remaining part of the paper is organized as follows. In Section 2 we introduce the quantile regression process. We derive a representation of this process in terms of a sum of independent identically distributed random variables which is used to prove weak convergence of the quantile process. Because the limiting processes depend on certain properties of the data generating process we introduce a resampling procedure and prove its consistency in Section 2.2.

In Section 3 we investigate the penalized quantile process under random censoring. In particular we show that for the Lasso the estimators of the non-zero parameters converge weakly to Gaussian processes and the zero parameters are estimated as exactly zero with positive probability for all quantiles in the region of investigation. We also prove that the adaptive Lasso penalty yields consistent model selection and the corresponding quantile process converges weakly to a Gaussian process. In particular this method possesses the 'oracle property' in the sense of Fan and Li (2001).

Finally, all technical details are deferred to an appendix [see Section 4].

2 The unpenalized quantile process estimator

We consider independent identically distributed random variables of the form $\{(T_i, C_i, X_i, \delta_i)\}_{i=1,\dots,n}$. We assume that only the data $\{(Y_i, X_i, \delta_i)\}_{i=1,\dots,n}$ can be observed, where $Y_i = \min\{C_i, T_i\}$, T_i denotes the survival time, C_i is a censoring time independent of T_i, $\delta_i = I\{Y_i = C_i\}$, $I\{\cdot\}$ is the indicator function, and X_i is a $(p-1)$ dimensional vector of random covariates not including an intercept. We are interested in statistical inference about the conditional distribution $P^{T|X}$ and, following Shows et al. (2010), assume the following parametric form of the conditional quantile of $L = \log(T)$

(2.1) $\qquad Q_L(\tau|X) = Q_L(\tau|X) = \inf\{t : P(L \leq t|X) \geq \tau\} = \alpha_0(\tau) + X^t \tilde{\beta}_0(\tau),$

where $\tau \in [0, 1]$. With the notation $Z_i = (1, X_i^t)^t$ and $\beta_0(\tau) = (\alpha_0(\tau), \tilde{\beta}_0(\tau)^t)^t$ model (2.1) can be alternatively written as

(2.2) $\qquad\qquad\qquad L = \log(T) = Z^t \beta_0(\tau) + \varepsilon^{(\tau)},$

where the errors $\varepsilon^{(\tau)}$ have τ-quantile zero. Thus model (2.1) is a generalization of the well known accelerated failure time (AFT) model [see e.g. Miller (1976); Buckley and James (1979); Louis

3

(1981)], which assumes independence of X and ε in (2.2) and places additional restrictions on the intercept. In this section, we provide the properties of an estimator of the regression quantile process $\beta_0(\cdot) = \{\beta_0(\tau)\}_{\tau \in [\tau_L, \tau_U]}$ for some interval $[\tau_L, \tau_U] \subset (0, 1)$. To achieve this goal we adopt the approach of Bang and Tsiatis (2002) who proposed to solve the equation

$$(2.3) \qquad \sum_{i=1}^{n} \frac{\delta_i}{\hat{G}(Y_i)} Z_i \left(I\{L_i \leq Z_i^t \beta\} - \frac{1}{2} \right) \approx 0$$

with respect to β in order to estimate the conditional median of L_i given X_i. Here $\hat{G}$ denotes the Kaplan-Meier estimator based on the data $\{Y_i, 1 - \delta_i\}$, that is the Kaplan-Meier estimator of the survival function of the censoring variables C_i. Note that solving (2.3) is equivalent to solving the minimization problem

$$(2.4) \qquad \text{argmin}_\beta \sum_{i=1}^{n} \frac{\delta_i}{\hat{G}(Y_i)} |L_i - Z_i^t \beta|.$$

A generalization to quantiles other than the median is now straightforward, it suffices to replace the absolute value in (2.4) by the check function $\rho_\tau(x) = x(\tau - I\{x \leq 0\})$. Note that the Kaplan-Meier estimator of the survival function of the censoring times C_i appears in the denominator of the above minimization problem. It is well known that this estimator can be very unstable near the right tail of the survival and censoring distribution. In fact, for many combinations of survival and censoring distributions it does not even converge with the optimal rate $1/\sqrt{n}$ uniformly on the interval $[0, \max_i Y_i]$, see Chen and Lo (1997) and the references therein for further details. This suggests that in order to stabilize the performance of the estimator, one should take extra care when considering $\hat{G}(Y_i)$ for values of Y_i that are close to the largest observation. A simple remedy was recently proposed by Zhou (2006) in the context of median regression, who exploited the fact that, for any constant M such that $M > \exp(Z^t \beta_0(\tau))$, the τ-quantile of random variable T equals the τ-quantile of the random variable $\min\{T, M\}$. For this reason, Zhou (2006) suggested to replace the original observations $Y_1, \ldots, Y_n$ by artificially censored variables on the left hand side of (2.3). To be precise, this author proposed to choose a constant $M > \exp(Z_i^t \beta_0(1/2))$ for all $i = 1, \ldots, n$ and to minimize (2.4), where the observations Y_i, δ_i and L_i are replaced by the quantities $Y_i^M = \min\{Y_i, M\}, T_i^M = \min\{Y_i, M\}, \delta_i^M = I\{Y_i^M \leq C_i\} = 1 - (1 - \delta_i)I\{M > Y_i\}$ and $L_i^M = \min\{L_i, \log M\}$, respectively. For arbitrary quantiles $\tau \in (0, 1)$ this approach yields the estimator

$$(2.5) \qquad \hat{\beta}(\tau) = \text{argmin}_\beta \sum_{i=1}^{n} \frac{\delta_i^M}{\hat{G}(Y_i^M)} \rho_\tau(L_i^M - Z_i^t \beta).$$

Note that $\hat{G}$ still denotes the estimator based on the untransformed data, and that only the values of this estimator at points from $[0, M]$ enter the minimization above. Assumptions (A4)

4

and (A5) made below guarantee that $\hat{G}$ is well-behaved on this interval [see also Csörgö and Horváth (1983) for more details]. Note also that the minimization problem in (2.5) is convex for each τ and so a minimizer can easily be computed. In order to guarantee that the τ-quantile of L_i is not affected by the artificial censoring, we require $M > \exp(\beta_0(\tau)^t Z_i)$ for all quantiles τ that are of interest, i.e. we fix a constant $\tau_U < 1$ and restrict our attention to the interval $[\tau_L, \tau_U]$ after requiring $M > \exp(Z_i^t \beta_0(\tau_U))$.

Remark 2.1 Bang and Tsiatis (2002) and Shows et al. (2010) both implicitly apply some sort of data transformation similar to the one described above without mentioning this explicitly. In the paper of Bang and Tsiatis (2002), the constant L (see their discussion in Section 2, in particular the last paragraph) implicitly plays a role similar to that of M as can be seen from their asymptotic derivations. Similarly, Shows et al. (2010) mention a quantity τ which they call "the maximum follow-up", see Condition 3 in their Appendix A and the subsequent theoretical developments. Unfortunately, from the discussion given in the papers it is not completely clear how they treat the data beyond L and τ, respectively. However, we would like to point out that the proofs given in the two papers cited above will only work if some truncation similar to the one proposed by Zhou (2006) is applied.

2.1 Asymptotic analysis

In order to formulate the first main result which states the asymptotic behavior of the estimators defined in (2.5) as a process indexed by $\tau \in [\tau_L, \tau_U]$ we introduce some notation. We denote the survival function of the censoring variable C_i by G_0 and the conditional distribution function of $T_i|Z_i$ by $F(\cdot|Z_i)$. We assume that $F(y|Z_i)$ is continuously differentiable with respect to y and that the corresponding density $f(y|Z_i) = \partial/\partial y F(y|Z_i)$ is uniformly bounded. Further let the following assumptions hold.

(A1) There exists a constant C_Z such that $\|Z_1\| \leq C_Z$ almost surely (here $\|\cdot\|$ denotes the L_2-norm on $\mathbb{R}^p$).

(A2) There exist constants $0 < \tau_L < \tau_U$ and $0 < C_\Sigma^{(L)} < C_\Sigma^{(U)} < \infty$ such that

$$\inf_{\tau \in [\tau_L, \tau_U]} \lambda_{\min}(\Sigma(\tau)) \geq C_\Sigma^{(L)}$$

and

$$\sup_{\tau \in [\tau_L, \tau_U]} \lambda_{\max}(\Sigma(\tau)) \leq C_\Sigma^{(U)}$$

for all $\tau \in [\tau_L, \tau_U]$, where $\Sigma(\tau) = \mathrm{E}\left[f(Z_1^t \beta_0(\tau)|X_1)Z_1 Z_1^t\right]$ and $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ denote the minimal and maximal eigenvalues of a matrix M, respectively.

(A3) For all $\tau \in [\tau_L, \tau_U]$ the vector $\beta_0(\tau)$ is an interior point of some bounded convex set B and there exists a constant M_0 such that $P(Y > M_0) > 0$ and $\beta^t Z \leq M_0$ a.s. for all $\beta \in B$. The mapping $\tau \mapsto \beta_0(\tau)$ is continuous.

(A4) G_0 has a uniformly bounded density g_0 and there exists a constant $C_G > 0$ such that $G_0(T_1 \wedge M) \geq C_G$ almost surely.

(A5) There exists a constant $C_M > 0$ such that $P(Y_1 < M) \leq 1 - C_M$.

Remark 2.2 Assumptions (A1)-(A5) are appropriately modified versions of (1)-(4) in appendix 1 of Zhou (2006). The modifications are needed since we consider process asymptotics while Zhou (2006) only discussed pointwise results.

For the formulation of our main statements we introduce the quantities

$$(2.6) \qquad M_n(s) = \sum_{i=1}^{n}\left[(1-\delta_i)I\{Y_i \leq s\} - \int_0^s I\{Y_i \geq x\}d\Lambda_C(x)\right],$$

$$(2.7) \qquad H_\tau(s) = \frac{\mathrm{E}\left[Z_1 I\{T_1^M \geq s\}(\tau - I\{\varepsilon_1^{(\tau)} \leq 0\})\right]}{P(Y_1 \geq s)}.$$

where $\Lambda_C(t) := -\log G_0(t)$ denotes the cumulative hazard function of the censoring variables. Throughout this paper $l^\infty([\tau_L, \tau_U])$ denotes the space of bounded functions on the interval $[\tau_L, \tau_U]$ equipped with the supremum norm.

Theorem 2.3 *If the assumptions (A1)-(A5) are satisfied the stochastic expansion*

$$(2.8) \qquad \sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau)) = -\Sigma(\tau)^{-1}W_n(\tau) + r_n(\tau)$$

holds, where the remainder satisfies $\sup_{\tau \in [\tau_L, \tau_U]} \|r_n(\tau)\| = o_P(1)$ *and*

$$(2.9) \qquad W_n(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i^M Z_i}{G_0(Y_i^M)}(I\{\varepsilon_i^{(\tau)} \leq 0\} - \tau) + \frac{1}{\sqrt{n}}\int_0^M H_\tau(s)dM_n(s).$$

Further the weak convergence

$$(2.10) \qquad \sqrt{n}(\hat{\beta}(\cdot) - \beta_0(\cdot)) \xrightarrow{w} -\Sigma(\cdot)^{-1}W(\cdot)$$

6

in the space $l^\infty([\tau_L, \tau_U])^p$ holds. Here W is a centered Gaussian process with covariance function $V(\tau_1, \tau_2) = V_1(\tau_1, \tau_2) - V_2(\tau_1, \tau_2)$, where

$$V_1(\tau_1, \tau_2) = (\min\{\tau_1, \tau_2\} - \tau_1\tau_2)\mathrm{E}\left[Z_1 Z_1^t / G_0(T_i^M)\right]$$

and

$$V_2(\tau_1, \tau_2) = \int_0^M H_{\tau_1}(s) H_{\tau_2}(s)^t P(Y_1 \geq s) d\Lambda_C(s),$$

respectively.

Remark 2.4 Zhou (2006) proposed an alternative data adapted artificial censoring. Here the constant M is replaced by the quantities $\hat{M}_i = Z_i^t \hat{\beta}(\tau) + c_0$ $(i = 1, \ldots, n)$, where $\hat{\beta}(\tau)$ is the estimator defined above, c_0 is a positive constant and the quantities Y_i^M, δ_i^M, L_i^M in (2.5) are replaced by $Y_i^{\hat{M}_i}, \delta_i^{\hat{M}_i}, L_i^{\hat{M}_i}$, respectively. If we combine the arguments given in the proof of Theorem 2.3 and in Appendix 2 of Zhou (2006) we also obtain weak convergence of the estimator based on this data adapted artificial censoring (centered by $\beta_0(\tau)$ and scaled by $\sqrt{n}$) to a Gaussian process. Essentially all processes appearing in the proof of Theorem 2.3 have to be additionally indexed in the parameter appearing in the definition of the artificial censoring variable and additional arguments (similar to those already presented in the proof of Theorem 2.3) are needed to prove the Donsker property of those indexing classes.

2.2 Resampling procedures

The variance functions of the estimators in the last sections are rather complicated and involve unknown quantities such as the conditional density of the conditional distribution $P^{T|Z}$. Because these quantities are hard to estimate in practice statistical inference based on the estimators discussed in the previous section is very complicated. In order to provide a solution to this problem we suggest a resampling scheme very similar to that in Jin et al. (2001) and Zhou (2006). We generate B samples $\{\xi_i : i = 1, \ldots, n\}$ of independent identically distributed random variables independent of the data and for each of these samples calculate $\beta^*(\tau)$ as the minimizer of

$$\sum_{i=1}^n \xi_i \frac{\delta_i^M}{\hat{G}_\xi(Y_i^M)} \rho_\tau(L_i^M - Z_i^t \beta),$$

where

$$\hat{G}_\xi(y) = \prod_{s \leq y} \left(1 - \frac{\Delta N_\xi(s)}{\bar{Y}_\xi(s)}\right)$$

and

$$N_\xi(s) = \sum_{i=1}^{n}(1-\delta_i)\xi_i I\{Y_i \le s\}, \qquad \bar{Y}_\xi(s) = \sum_{i=1}^{n}\xi_i I\{Y_i \ge s\}.$$

The following result shows that for suitable choices of ξ_i the distribution of the process $\sqrt{n}(\hat{\beta}(\cdot) - \beta_0(\cdot))$ can be approximated by that of $\sqrt{n}(\beta^*(\cdot) - \hat{\beta}(\cdot))$. As a consequence, distributional characteristics of $\sqrt{n}(\hat{\beta}(\cdot) - \beta_0(\cdot))$ can be estimated by sample analogons of $\sqrt{n}(\beta^*(\cdot) - \hat{\beta}(\cdot))$ calculated on the basis of the B samples $\{\xi_i : i = 1, \ldots, n\}$.

Theorem 2.5 *Let assumptions (A1)-(A5) hold. Moreover, assume the random variables ξ_i are independent of the sample, i.i.d. and have expectation and variance equal to one. Then $\sqrt{n}(\beta^*(\cdot) - \hat{\beta}(\cdot))$ converges weakly to $-\Sigma(\cdot)^{-1}W(\cdot)$ conditional on (Z, δ, T) in outer probability.*

Remark 2.6 As noted by Kosorok (2008) on page 19, weak convergence on a metric space is equivalent to convergence in the bounded Lipschitz metric. Following this reference, let $\mathrm{E}_\xi[X]$ denote the expectation of the random variable X with respect to the ξ-variables and BL_1 the space of all functions $h : l^\infty([\tau_L, \tau_U]^d) \to [0,1]$ which are Lipschitz continuous with Lipschitz constant 1. For random elements X_n, X of $l^\infty([\tau_L, \tau_U]^d)$ (depending on ξ, Z, δ, Y) we call the sequence $(X_n)_{n\in\mathbb{N}}$ *weakly convergent to X conditional on (Z, δ, Y) in outer probability* if

$$\sup_{h\in BL_1} |\mathrm{E}_\xi[h(X_n)] - \mathrm{E}[h(X)]| \xrightarrow{P*} 0.$$

and if $\mathrm{E}_\xi[h(X_n)^*] - \mathrm{E}_\xi[h(X_n)_*] \xrightarrow{P*} 0$ for all $h \in BL_1$ where $h(X_n)^*$ and $h(X)_*$ denote measurable majorants and minorants of $h(X_n)$ and $h(X)$ with respect to (ξ, Z, δ, Y).

3   Penalized quantile process estimators

In this section we investigate the asymptotic properties of penalized quantile processes. First, let us consider the classical lasso penalization, i.e.

$$(3.1) \qquad \hat{\beta}^L(\tau) = \mathrm{argmin}_\beta \sum_{i=1}^{n} \frac{\delta_i^M}{\hat{G}(Y_i^M)} \rho_\tau(L_i^M - Z_i^t\beta) + \lambda_n\|\tilde{\beta}\|_1,$$

where $\tilde{\beta}$ denotes the last $(p-1)$ components of β and λ_n is a tuning parameter depending on the sample size. To the best of our knowledge, an analysis of the asymptotic properties of the lasso penalty in combination with censored quantiles is new and has so far not been considered even in a point-wise sense. The next theorem will show that the estimator defined in (3.1) performs

both consistent parameter estimation and model selection as usual for lasso penalized estimators. Thus it can be used to identify the components of the vector Z_i, which influences the conditional quantile function of L_i if the dimension of β_0 is large. In addition to the assumptions of the last section we require for all $j = 2, \ldots, p$:

(A6) If $\beta_{0,j}(\tau) \neq 0$ for some $\tau \in [\tau_L, \tau_U]$ then $\beta_{0,j}(\tau) \neq 0$ for all $\tau \in [\tau_L, \tau_U]$,

where $\beta_{0,j}$ denotes the j-th component of β_0. Assumption (A6) excludes scenarios in which the j-th covariate has an impact on some quantiles but no impact on other quantiles in the region of interest. It is required because otherwise the limiting process derived below is not necessarily continuous in τ anymore. Without loss of generality we assume that $\beta_{0,2}, \ldots, \beta_{0,q} \neq 0$ and $\beta_{0,q+1}, \ldots, \beta_{0,p} = 0$ for some $q \leq p$.

Theorem 3.1 *If the assumptions (A1)-(A6) are satisfied and if there exist a $\lambda_0 \in [0, \infty)$ such that $\lambda_n / \sqrt{n} \to \lambda_0$, then the weak convergence*

$$(3.2) \quad \sqrt{n}(\hat{\beta}^L(\cdot) - \beta_0(\cdot)) \xrightarrow{w} \mathrm{argmin}_u u^t W(\cdot) + \frac{1}{2} u^t \Sigma(\cdot) u + \lambda_0 \sum_{j=2}^{q} u_j \mathrm{sgn}(\beta_{0,j}(\cdot)) + \lambda_0 \sum_{j=q+1}^{p} |u_j|$$

in the space $l^\infty([\tau_L, \tau_U])^p$ holds, where the process W is defined in Theorem 2.3.

Remark 3.2 Define $\Sigma^{(11)}(\tau)$ as the upper left $q \times q$ block and $\Sigma^{(21)}(\tau)$ as the lower left $(p-q) \times q$ block of $\Sigma(\tau)$. Further denote by $W^{(1)}(\tau)$ the first q and by $W^{(2)}(\tau)$ the last $(p-q)$ components of $W(\tau)$. Then the Karush-Kuhn-Tucker conditions directly yield that $(u^{(1)}(\tau)^t, 0_{p-q}^t)^t = (u_1(\tau), \ldots, u_q(\tau), 0, \ldots, 0)^t$ minimizes $V(\cdot, \tau)$ if and only if

$$u^{(1)}(\tau) = -\left(\Sigma^{(1)}(\tau)\right)^{-1} \left(W^{(1)}(\tau) - \lambda_0 (0, \mathrm{sgn}(\beta_{0,2}(\tau)), \ldots, \mathrm{sgn}(\beta_{0,q}(\tau)))^t\right)$$

and

$$-\lambda_0 \mathbb{1}_{p-q} < \Sigma^{(21)}(\tau) u^{(1)}(\tau) - W^{(2)}(\tau) < \lambda_0 \mathbb{1}_{p-q},$$

where the inequalities above are understood component-wise and $\mathbb{1}_{p-q}$ denotes a $(p-q)$ vector with all entries given by 1. Thus the estimators of the non-zero parameters are asymptotically Gaussian processes and the zero parameters are estimated as exactly zero with positive probability for all $\tau \in [\tau_L, \tau_U]$ if $\lambda_0 > 0$.

Another popular penalization is the adaptive Lasso penalty proposed by Zou (2006). This penalty is also the one which was analyzed by Shows et al. (2010) in a point-wise sense. The estimator is now defined by

$$(3.3) \quad \hat{\beta}^{AL}(\tau) = \mathrm{argmin}_\beta \sum_{i=1}^{n} \frac{\delta_i^M}{\hat{G}(Y_i^M)} \rho_\tau(L_i^M - Z_i^t \beta) + \lambda_n \sum_{j=2}^{p} \left| \frac{\beta_j}{\bar{\beta}_j(\tau)} \right|,$$

where $\bar{\beta}(\tau)$ denotes a preliminary estimator of $\beta_0(\tau)$. We will show that the adaptive Lasso penalty yields consistent model selection and asymptotic normality in our framework, that is it has the so called 'oracle property' in the sense of Fan and Li (2001). Additionally to the assumptions (A1)-(A6) stated in the last sections we require the following.

(A7) The preliminary estimator $\bar{\beta}(\tau)$ satisfies $\sup_{\tau \in [\tau_j, \tau_u]} \|\bar{\beta}(\tau) - \beta_0(\tau)\| = O_p(n^{-1/2})$.

Note that assumption (A7) is satisfied for the unpenalized estimator $\hat{\beta}(\tau)$ defined in (2.5) because of Theorem 2.3. In the next Theorem we denote by $\beta^{(1)}$ the q- dimensional vector obtained from the first p components of the q-dimensional vector β.

Theorem 3.3 *If the assumptions (A1)-(A7) are satisfied and if $\lambda_n/\sqrt{n} \to 0, \lambda_n \to \infty$ then the weak convergence*

$$(3.4) \qquad \sqrt{n}(\hat{\beta}^{AL,(1)}(\cdot) - \beta_0^{(1)}(\cdot)) \xrightarrow{w} - \left(\Sigma^{(11)}(\cdot)\right)^{-1} W^{(1)}(\cdot)$$

in the space $l^\infty([\tau_L, \tau_U])^q$ holds, where $\Sigma^{(11)}(\tau)$ denotes the upper left $q \times q$ block of $\Sigma(\tau)$ and $W^{(1)}$ the vector of the first q components of the process W which is defined in Theorem 2.3. Further

$$(3.5) \qquad P \left(\sup_{\tau \in [\tau_L, \tau_U], q+1 \leq j \leq p} |\hat{\beta}_j^{AL}(\tau)| = 0 \right) \to 1.$$

Theorem 3.3 states that the adaptive lasso penalization of censored quantile processes yields estimators of non-zero components which have the same asymptotic distribution as the estimators constructed in a smaller model and that the zero components are estimated as zero uniformly over quantiles. This property can be described as a 'process oracle' closely related to the classical 'point-wise oracle' property in the sense of Fan and Li (2001).

4 Appendix: proofs

Remark 4.1 In the following discussion we will sometimes write $X_n = o_P(1)$ or $X_n = O_P(1)$ for non-measurable mappings $X_n : \Omega \to \mathcal{X}$, where $(\Omega, \mathcal{A}, P)$ denotes a probability space and $\mathcal{X}$ is a subset of $\mathbb{R}^p$. This means that X_n converges to zero in outer probability for $n \to \infty$ or X_n is stochastically bounded in outer probability for $n \to \infty$, respectively.

4.1 Proof of Theorem 2.3:

Define the function that

$$f_n(u, \tau, G) = \sum_{i=1}^{n} \frac{\delta_i^M}{G(Y_i^M)} \rho_\tau \left(L_i^M - \frac{Z_i^t u}{\sqrt{n}} - Z_i^t \beta_0(\tau) \right),$$

then it follows by definition (2.5) that

$$\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau)) = \operatorname{argmin}_u f_n(u, \tau, \hat{G}) - f_n(0, \tau, \hat{G}).$$

In the following we will derive an asymptotic representation of $f_n(u, \tau, \hat{G}) - f_n(0, \tau, \hat{G})$ and then apply Theorem 2 of Kato (2009) in order to obtain the representation of $\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau))$ which is asserted in Theorem 2.3. For this purpose we use the decomposition

(4.1) $$f_n(u, \tau, \hat{G}) - f_n(0, \tau, \hat{G}) = f_n(u, \tau, G_0) - f_n(0, \tau, G_0) + R_n$$

where

$$R_n = f_n(0, \tau, G_0) - f_n(0, \tau, \hat{G}) + f_n(u, \tau, \hat{G}) - f_n(u, \tau, G_0).$$

We first establish a stochastic expansion of the terms on the right hand side of (4.1). Using Knight's identity [see Knight (1998)]

(4.2) $$\rho_\tau(x - y) - \rho_\tau(x) = -y(\tau - I\{x \le 0\}) + \int_0^y (I\{x \le s\} - I\{x \le 0\}) ds$$

direct calculations yield

(4.3) $$f_n(u, \tau, G_0) - f_n(0, \tau, G_0) = \frac{1}{2} u^t \Sigma(\tau) u + \frac{u^t}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i^M Z_i}{G_0(Y_i^M)} (I\{\varepsilon_i^{(\tau)} \le 0\} - \tau) + o_P(1)$$

for all $\tau \in [\tau_L, \tau_U]$, $u \in \mathbb{R}^p$. The statement in (4.3) relies on an application of the CLT and the lemma of Slutsky to the term $\sum_{i=1}^{n}(S_i(\tau) - \mathrm{E}\,[S_i(\tau)]) + n\mathrm{E}\,[S_1(\tau)]$, where

$$S_i(\tau) = \int_0^{Z_i^t u/\sqrt{n}} I\{\varepsilon_i^{(\tau)} \le s\} - I\{\varepsilon_i^{(\tau)} \le 0\} ds = \left| \frac{Z_i^t u}{\sqrt{n}} - \varepsilon_i^{(\tau)} \right| I\{|\varepsilon_i^{(\tau)}| \le |n^{-1/2} Z_i^t u|\}.$$

Note that a straightforward calculation shows $n\mathrm{E}\,[S_1(\tau)] = \frac{1}{2} u^t \Sigma(\tau) u + o(1)$ uniformly with respect to τ. Now uniformity of the approximation in (4.3) with respect to τ can be obtained in the following way. Observe that $\sum_{i=1}^{n}(S_i(\tau) - \mathrm{E}\,[S_i(\tau)])$ can be interpreted as empirical process indexed by a class of functions that is contained in the set

$$\mathcal{H} = \{(z, t) \mapsto f(z, t) g(z, t) | f \in \mathcal{F}, g \in \mathcal{G}\},$$

11

where the classes of functions $\mathcal{F}, \mathcal{G}$ are defined by

$$\mathcal{F} = \{(z,t) \mapsto \min(|z^t u - c \log t + cz^t \beta|, 2\|u\| C_Z) : \ \beta \in B, \ c \geq 0\}$$

and

$$\mathcal{G} = \{(z,t) \mapsto I\{|c \log t - cz^t \beta| \leq |z^t u|\} : \ \beta \in B, \ c \geq 0\}.$$

Combining Lemmas 2.6.15, 2.6.18 and 2.6.19 of van der Vaart and Wellner (1996) yields that for each u the classes $\mathcal{F}$ and $\mathcal{G}$ are uniformly bounded VC-major classes. Now we obtain from Lemma 2.6.20 of the same reference that $\mathcal{H}$ is also a uniformly bounded VC-major class and Theorem 2.6.14 of van der Vaart and Wellner (1996) yields that this class is Donsker. Thus the stochastic approximation in (4.3) holds uniformly with respect to $\tau \in [\tau_L, \tau_U]$ by an application of the continuous mapping theorem.

Next we will derive an approximation of the remaining four terms on the right hand side of (4.1). Again using Knight's identity (4.2) and the rate of the uniform convergence of the Kaplan-Meier estimator given for example in Csörgö and Horváth (1983) we obtain that the second term in (4.1) equals

$$(4.4) \qquad\qquad R_n = -u^t P_1(\tau, \hat{G}) + P_2(\tau, \hat{G}) + o_P(1/\sqrt{n}),$$

where

$$P_1(\tau, \hat{G}) = \frac{-u^t}{\sqrt{n}} \sum_{i=1}^{n} \delta_i^M \frac{G_0(T_i^M) - \hat{G}(T_i^M)}{G_0^2(T_i^M)} Z_i(\tau - I\{\varepsilon_i^{(\tau)} \leq 0\})$$

$$P_2(\tau, \hat{G}) = \sum_{i=1}^{n} \delta_i^M \frac{G_0(T_i^M) - \hat{G}(T_i^M)}{G_0^2(T_i^M)} \int_0^{Z_i^t u/\sqrt{n}} (I\{\varepsilon_i^{(\tau)} \leq s\} - I\{\varepsilon_i^{(\tau)} \leq 0\}) ds + o_P(1/\sqrt{n})$$

uniformly with respect to τ for fixed u. We investigate the term P_1 first and show that it can be approximated by a martingale transformation. Note that by Example 2.6.21, Lemma 2.6.15, 2.6.18 and Example 2.10.8 of van der Vaart and Wellner (1996) the class of functions

$$\mathcal{F} = \left\{ (\delta, z, t) \mapsto \delta \frac{G_0(t) - G(t)}{G_0^2(t)} z(\tau - I\{\log t \leq z^t \beta\}) : \ G : \mathbb{R} \to [0,1] \text{ increasing}, \ \tau \in [0,1], \ \beta \in B \right\}$$

is Donsker. Thus the process $P_1(\tau, G) - \mathrm{E}[P_1(\tau, G)]$ (indexed by the class $\mathcal{F}$) converges weakly to a centered Gaussian process with variance function $V(\tau, G)$. Direct calculations yield

$$\mathrm{E}[P_1(\tau, G)] = \sqrt{n} \mathrm{E}\left[\frac{G_0(T_1^M) - G(T_1^M)}{G_0(T_1^M)} Z_1(\tau - I\{\varepsilon_1^{(\tau)} \leq 0\}) \right]$$

and

$$V(\tau, G) = \mathrm{E}\left[\frac{(G_0(T_1^M) - G(T_1^M))^2}{G_0(T_1^M)^3} Z_1 Z_1^t (\tau - I\{\varepsilon_1^{(\tau)} \le 0\})^2\right]$$
$$- \mathrm{E}\left[\frac{G_0(T_1^M) - G(T_1^M)}{G_0(T_1^M)} Z_1(\tau - I\{\varepsilon_1^{(\tau)} \le 0\})\right] \mathrm{E}\left[\frac{G_0(T_1^M) - G(T_1^M)}{G_0(T_1^M)} Z_1(\tau - I\{\varepsilon_1^{(\tau)} \le 0\})\right]^t.$$

Using the uniform almost sure convergence of $\hat{G}$ to G_0 we obtain that

$$P_1(\tau, \hat{G}) = \mathrm{E}\left[P_1(\tau, \hat{G}) | \hat{G}\right] + o_P(1),$$

where the approximation holds uniformly with respect to τ and the notation $E[P_1(\tau, \hat{G}) | \hat{G}]$ means that we treat $\hat{G}$ as fixed function. Theorem 3.2.3 in Fleming and Harrington (1991) now yields

$$E[P_1(\tau, \hat{G}) | \hat{G}] = \frac{1}{\sqrt{n}} \int_{\mathbb{R}^p} \int_0^\infty \int_0^{\min\{t, M\}} \frac{\hat{G}(s-)}{G_0(s)} \frac{1}{\frac{1}{n}\sum_{k=1}^n I\{Y_k \ge s\}}$$
$$\times z(\tau - I\{\min\{\log t, \log M\} \le z^t \beta_0(\tau)\}) f(t|z) dM_n(s) dt dF_Z(z)$$
$$= \frac{1}{\sqrt{n}} \int_0^M \frac{\hat{G}(s-)}{G_0(s)} \frac{1}{\frac{1}{n}\sum_{k=1}^n I\{Y_k \ge s\}} \mathrm{E}\left[Z_1(\tau - I\{\varepsilon_1^{(\tau)} \le 0\}) I\{T_1^M \ge s\}\right] dM_n(s),$$

where F_Z denotes the distribution function of Z. This yields the the approximation

$$(4.5) \qquad P_1(\tau, \hat{G}) = \int_0^M H_\tau^{(n)}(s) dM_n(s) + o_P(1)$$

uniformly with respect to $\tau \in [\tau_L, \tau_U]$, where the process $H_\tau^{(n)}$ is defined by

$$(4.6) \qquad H_\tau^{(n)}(s) = \frac{1}{\sqrt{n}} \frac{\hat{G}(s-)}{G_0(s)} \frac{1}{\frac{1}{n}\sum_{k=1}^n I\{Y_k \ge s\}} \mathrm{E}\left[Z_1(\tau - I\{\varepsilon_1^{(\tau)} \le 0\}) I\{T_1^M \ge s\}\right].$$

For the term P_2 we obtain by the Cauchy-Schwarz inequality

$$(4.7) \quad |P_2(\tau, \hat{G})|^2 \le \sum_{i=1}^n \frac{(G_0(T_i^M) - \hat{G}(T_i^M))^2}{G_0^4(T_i^M)} \sum_{i=1}^n \left(\int_0^{Z_i^t u/\sqrt{n}} (I\{\varepsilon_i^{(\tau)} \le s\} - I\{\varepsilon_i^{(\tau)} \le 0\}) ds\right)^2.$$

The first factor on the right hand side of (4.7) is of order $O_P(1)$ because of the process convergence of the Kaplan-Meier estimator, assumption (A4) and (A5). A direct calculation involving a Taylor expansion yields

$$E\left[\left(\int_0^{Z_i^t u/\sqrt{n}} (I\{\varepsilon_i^{(\tau)} \le s\} - I\{\varepsilon_i^{(\tau)} \le 0\}) ds\right)^2\right] = O\left(\frac{1}{n^{3/2}}\right)$$

13

uniformly with respect to τ. A similar argument as in the derivation of (4.3) and (4.7) shows $P_2(\tau, \hat{G}) = o_P(1)$ uniformly with respect to τ. Combining the estimates (4.3), (4.4), (4.5) with (4.1) yields the stochastic expansion

$$f_n(u, \tau, \hat{G}) - f_n(0, \tau, \hat{G}) = u^t \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i^M Z_i}{G_0(Y_i^M)} (I\{\varepsilon_i^{(\tau)} \le 0\} - \tau) + \int_0^M H_\tau^{(n)}(s) dM_n(s) \right)$$

(4.8)
$$+ \frac{1}{2} u^t \Sigma(\tau) u + o_P(1)$$

for every u, and the approximation holds uniformly with respect to τ.

Now we approximate the second summand on the right hand side of (4.8). Note that the functions $s \mapsto H_\tau(s), H_\tau^{(n)}(s)$ are left continuous and adapted to the family of σ-algebras $\{\mathcal{F}_s : 0 \le s\}$, where

$$\mathcal{F}_s = \sigma\{I\{Y_i \le t, \delta_i = 1\}, I\{Y_i \le t, \delta_i = 0\} : 0 \le t \le s, \ i = 1, \dots, n\}.$$

Thus H_τ and $H_\tau - H_\tau^{(n)}$ are predictable processes [which follows from Lemma 1.4.1 of Fleming and Harrington (1991)]. The classical Glivenko-Cantelli theorem and the uniform almost sure consistency of the Kaplan-Meier estimator on the interval $[0, M]$ give

$$\sum_{i=1}^{n} \int_0^M \left(\frac{1}{\sqrt{n}} H_\tau - H_\tau^{(n)} \right)(s) \left(\frac{1}{\sqrt{n}} H_\tau - H_\tau^{(n)} \right)^t (s) I\{Y_i \ge s\} d\Lambda_C(s) \xrightarrow{a.s.} 0.$$

Now Theorem 5.1.1 of Fleming and Harrington (1991) yields

(4.9)
$$\int_0^M H_\tau^{(n)}(s) dM_n(s) = \frac{1}{\sqrt{n}} \int_0^M H_\tau(s) dM_n(s) + o_P(1)$$

for each τ. Next we show that this statement holds uniformly with respect to $\tau \in [\tau_L, \tau_U]$ which will follow from Theorem 1.5.7 of van der Vaart and Wellner (1996) once we establish asymptotic uniform equicontinuity of the processes $\int_0^M H_\tau^{(n)}(s) dM_n(s)$ and $n^{-1/2} \int_0^M H_\tau(s) dM_n(s)$. First we use integration by parts to obtain

$$\int_0^M H_\tau^{(n)}(s) dM_n(s) = M_n(M) H_\tau^{(n)}(M) - \int_0^M M_n(s) dH_\tau^{(n)}(s)$$

(4.10)
$$= M_n(M) H_\tau^{(n)}(M) - \int_0^M M_n(s) h^{(1)}(s) dh_\tau^{(2)}(s) - \int_0^M M_n(s) h_\tau^{(2)}(s) dh^{(1)}(s),$$

where the functions $h^{(1)}$ and $h^{(2)}$ are defined by

$$h^{(1)}(s) = \frac{1}{\sqrt{n}} \frac{\hat{G}(s-)}{G_0(s)} \frac{n}{\sum_{k=1}^{n} I\{Y_k \ge s\}},$$
$$h_\tau^{(2)}(s) = \mathrm{E}\left[Z_1(\tau - I\{\varepsilon_1^{(\tau)} \le 0\}) I\{T_1^M \ge s\} \right].$$

14

Note that $M_n(s)$ is a sum of centered independent identically distributed random variables and $n^{-1/2}M_n(\cdot)$ can be viewed as empirical process indexed by the class of functions $\{(\delta, y) \mapsto (1 - \delta)I\{y \leq s\} - \int_0^s I\{y \geq x\}d\Lambda_C(x) : s \in [0, M]\}$. This class is Donsker because the functions $y \mapsto I\{y \leq s\}, y \mapsto \int_0^s I\{y \geq x\}d\Lambda_C(x)$ are bounded and monotone. This observation and the continuous mapping theorem yield

$$\frac{1}{\sqrt{n}} \sup_{s \in [0,M]} |M_n(s)| = O_P(1).$$

Further note that for n sufficiently large a simple calculation shows that

$$\|H_{\tau_1}^{(n)}(s) - H_{\tau_2}^{(n)}(s)\| \leq \frac{O_P(1)}{\sqrt{n}}|\tau_1 - \tau_2|,$$

$$\|h_{\tau_1}^{(2)}(s) - h_{\tau_2}^{(2)}(s)\| \leq 2C_Z|\tau_1 - \tau_2|,$$

where the term $O_P(1)$ does not depend on τ_1, τ_2. The theorem of dominated convergence implies

$$\frac{\mathrm{d}h_\tau^{(2)}(s)}{\mathrm{d}s} = \mathrm{E}\left[Z_1(I\{\log s \leq Z_1^t\beta_0(\tau)\} - \tau)f(s|Z_1)\right],$$

which yields

$$\left\|\int_0^M dh_{\tau_1}^{(2)}(s) - \int_0^M dh_{\tau_2}^{(2)}(s)\right\| \leq 2C_Z|\tau_1 - \tau_2|.$$

The mapping $\sqrt{n}h^{(1)}$ is, on the interval $[0, M]$, uniformly bounded with probability tending to one and can be shown to be of bounded variation because it is a product of uniformly bounded [again with probability tending to one], monotone functions.

The last statements together with (4.10) imply

$$\left\|\int_0^M H_{\tau_1}^{(n)}(s)dM_n(s) - \int_0^M H_{\tau_2}^{(n)}(s)dM_n(s)\right\| \leq |\tau_1 - \tau_2|O_P(1),$$

where the $O_P(1)$ does not depend on τ_1, τ_2, which implies asymptotic equicontinuity in probability. Similarily but with less effort (note that $H_\tau(s)$ has a derivative with respect to s which is Lipschitz continuous with respect to τ) we also obtain asymptotic equicontinuity in probability of the process $1/\sqrt{n}\int_0^M H_\tau(s)dM_n(s)$. This finally gives that the statement (4.9) holds uniformly with respect to τ. Observing the definition of W_n in (2.9) this implies

(4.11) $$f_n(u, \tau, \hat{G}) - f_n(0, \tau, \hat{G}) = u^t W_n(\tau) + \frac{1}{2}u^t\Sigma(\tau)u + o_P(1)$$

uniformly with respect to τ.

Note that the sequence $(W_n(\cdot))_n$ is clearly a sequence of bounded stochastic processes (where the bound depends on n). Now we show stochastic boundedness (for $n \to \infty$) of the quantity

15

$\sup_{\tau \in [\tau_L, \tau_U]} \|W_n(\tau)\|$. For each τ the random variable $W_n(\tau)$ is a sum of centered independent identically distributed random variables with finite variance and scaled by $n^{-1/2}$. It thus converges weakly to a normal random variable. The class of functions

$$\{(\delta, z, t) \mapsto \delta z / G_0(t)(I\{\log t \leq \beta^t z\} - \tau) : \ \tau \in [\tau_L, \tau_U], \ \beta \in B\}$$

is Donsker, which can be shown by similar arguments as given above. This implies tightness of the process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i^M Z_i}{G_0(Y_i^M)} (I\{\varepsilon_i^{(\tau)} \leq 0\} - \tau).$$

Moreover, the process $1/\sqrt{n} \int_0^M H_\tau(s) dM_n(s)$ is asymptotically tight by the arguments given in the last paragraph. Thus $W_n(\cdot)$ converges weakly to a Gaussian process in the space $l^\infty([\tau_L, \tau_U])^p$. Again the continuous mapping theorem implies

(4.12)
$$\sup_{\tau \in [\tau_L, \tau_U]} \|W_n(\tau)\| = O_P(1).$$

Because $f_n(u, \tau, \hat{G}) - f_n(0, \tau, \hat{G})$ is convex in u for each τ and bounded in τ for each u and $\Sigma(\tau)$ is a positive definite matrix with uniformly bounded eigenvalues on $[\tau_L, \tau_U]$ the first assertion of Theorem 2.3 follows from Theorem 2 of Kato (2009). The second assertion follows from the first one, the arguments in the last paragraph for establishing weak convergence and tightness, and by a tedious but straight forward covariance calculation. $\qquad\square$

4.2   Proof of Theorem 2.5

Because of the first assertion of Theorem 2.3 we have

$$\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau)) = \Sigma(\tau)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i^M}{G_0(T_i^M)} Z_i(I\{\varepsilon_i^{(\tau)} \leq 0\} - \tau) + \frac{1}{\sqrt{n}} \int_0^t H_\tau(s) dM_n(s) \right)$$
$$+ o_P(1),$$

uniformly with respect to $\tau \in [\tau_L, \tau_U]$ and the term $o_P(1)$ is understood with respect to the probability space generated by (Z, δ, T). Using the results of Appendix 3 of Zhou (2006) and similar arguments as in the proof of Theorem 2.3 yields the approximation

$$\sqrt{n}(\beta^*(\tau) - \beta_0(\tau)) = \Sigma(\tau)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi_i \delta_i^M}{G_0(T_i^M)} Z_i(I\{\varepsilon_i^{(\tau)} \leq 0\} - \tau) + \frac{1}{\sqrt{n}} \int_0^t H_\tau(s) dM_n^{(\xi)}(s) \right)$$
$$+ o_P(1),$$

where $M_n^{(\xi)}(s) = N_\xi(s) - \int_0^s \sum_{i=1}^n \xi_i I\{Y_i \geq s\}d\Lambda_C(s)$ uniformly with respect to τ and for almost all sequences $(\xi_i)_i$, and the remainder term $o_P(1)$ is understood with respect to the measure generated by $(T_1, \delta_1, Z_1), \ldots, (T_n, \delta_n, Z_n)$ holds. Thus the representation

$$\sqrt{n}(\beta^*(\tau) - \hat{\beta}(\tau)) = \Sigma(\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n (\xi_i - 1)X_i(\tau) + r_n(\tau),$$

where the random variables $X_i(\tau)$ are given by

$$X_i(\tau) = \frac{\delta_i^M Z_i}{G_0(T_i^M)}(I\{\varepsilon_i^{(\tau)} \leq 0\} - \tau) - \int_0^t H_\tau(s)d\left((1 - \delta_i)I\{Y_i \leq s\} - \int_0^s I\{Y_i \geq s\}d\Lambda_C(s)\right).$$

and $\sup_{\tau \in [\tau_L, \tau_U]} \|r_n(\tau)\| = o_P(1)$ conditionally on (Z_i, δ_i, Y_i) along almost all sequences (ξ_i). For $h \in BL_1$ we directly obtain

$$\left|\mathrm{E}_\xi\left[h(\sqrt{n}(\beta^*(\cdot) - \hat{\beta}(\cdot)))\right] - \mathrm{E}\left[h(-\Sigma(\cdot)^{-1}W(\cdot))\right]\right|$$

$$\leq \left|\mathrm{E}_\xi\left[h\left(-\Sigma(\cdot)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi_i - 1)X_i(\cdot)\right)\right] - \mathrm{E}\left[h(-\Sigma(\cdot)^{-1}W(\cdot))\right]\right| + o_P(1).$$

The proof of Theorem 2.3 and Theorem 2.9.6 of van der Vaart and Wellner (1996) now yield the assertion of Theorem 2.5. □

4.3 Proof of Theorem 3.1

By definition we have

$$\sqrt{n}(\hat{\beta}^L(\tau) - \beta_0(\tau)) = \mathrm{argmin}_u V_n(u, \tau),$$

where

$$V_n(u, \tau) = f_n(u, \tau, \hat{G}) - f_n(0, \tau, \hat{G}) + \lambda \sum_{j=2}^p \left(\left|\beta_{0,j}(\tau) + \frac{u_j}{\sqrt{n}}\right| - |\beta_{0,j}(\tau)|\right).$$

Equation (4.11) and results from Knight and Fu (2000) directly yield

$$V_n(u, \cdot) \xrightarrow{w} u^t W(\cdot) + \frac{1}{2}u^t\Sigma(\cdot)u + \lambda_0 \sum_{j=2}^q u_j\mathrm{sgn}(\beta_{0,j}(\cdot)) + \lambda_0 \sum_{j=q+1}^p |u_j| = V(u, \cdot)$$

in $C([\tau_L, \tau_U])$. The mappings $u \mapsto V(u, \tau), V_n(u, \tau)$ are strictly convex for every τ and the functions $\tau \mapsto V(u, \tau), V_n(u, \tau)$ are continuous for every u. Moreover $\tau \mapsto \mathrm{argmin}_u V_n(u, \tau)$ is clearly bounded. For a large norm of u the mapping $u \mapsto V(u, \tau)$ is dominated by the quadratic term $u^t\Sigma(\tau)u$ and (A2) yields that $u_\infty(\cdot) = \mathrm{argmin}_u V(u, \cdot) \in l^\infty([\tau_L, \tau_U])^p$. We will show at the

17

end of this proof that the function u_∞ is continuous. Therefore the assertion of Theorem 3.1 follows from Theorem 1 of Kato (2009).

In order to prove that u_∞ is continuous for $\tau \in [\tau_L, \tau_U]$ we assume the contrary. In this case there exists a sequence $\tau_n \to \tau$ and a constant $\delta_1 > 0$ such that $\|u_\infty(\tau_n) - u_\infty(\tau)\| > \delta_1$ for all n. The continuity and strict convexity of $V(\cdot, \tau)$ yield the existence of a constant $\delta_2 > 0$ such that

$$(4.13) \qquad\qquad V(u_\infty(\tau_n), \tau) - V(u_\infty(\tau), \tau) > \delta_2 \qquad \text{for all } n.$$

Boundedness of the set $\{u_\infty(\tau_n) : \ n \in \mathbb{N}\}$ and the continuity of $V(u, \cdot)$ imply that for all $\varepsilon > 0$ we can choose n such that the inequalities

$$|V(u_\infty(\tau), \tau) - V(u_\infty(\tau), \tau_n)| < \varepsilon, |V(u_\infty(\tau_n), \tau_n) - V(u_\infty(\tau_n), \tau)| < \varepsilon$$

hold. These imply

$$V(u_\infty(\tau), \tau) > V(u_\infty(\tau), \tau_n) - \varepsilon > V(u_\infty(\tau_n), \tau_n) - \varepsilon > V(u_\infty(\tau_n), \tau) - 2\varepsilon,$$

which is a contradiction to (4.13) if ε is chosen sufficiently small. This complete the proof of Theorem (3.1). □

4.4 Proof of Theorem 3.3

As in the proof of Theorem 3.1 we obtain

$$\sqrt{n}(\hat{\beta}^{AL}(\tau) - \beta_0(\tau)) = \text{argmin}_u \tilde{V}_n(u, \tau),$$

where the function $\tilde{V}_n$ is defined by

$$\tilde{V}_n(u, \tau) = f_n(u, \tau, \hat{G}) - f_n(0, \tau, \hat{G}) + \lambda \sum_{j=2}^{p} \frac{|\beta_{0,j}(\tau) + u_j/\sqrt{n}| - |\beta_{0,j}(\tau)|}{|\bar{\beta}_j(\tau)|}.$$

Assumption (A7) directly yields

$$\lambda \sup_{\tau \in [\tau_j, \tau_u]} \frac{|\beta_{0,j}(\tau) + u_j/\sqrt{n}| - |\beta_{0,j}(\tau)|}{|\bar{\beta}_j(\tau)|} \xrightarrow{P} 0$$

for all $j \leq q$ and

$$\lambda \inf_{\tau \in [\tau_L, \tau_U]} \frac{|\beta_{0,j}(\tau) + u_j/\sqrt{n}| - |\beta_{0,j}(\tau)|}{|\bar{\beta}_j(\tau)|} \xrightarrow{P} \infty$$

18

for all $j \geq q+1$. Together with the arguments given in the proof of Theorem 2.3 this implies the weak epiconvergence (compare e.g. Geyer (1996))

$$\tilde{V}_n(u, \cdot) \xrightarrow{w} \begin{cases} \frac{1}{2}(u^{(1)})^t \Sigma^{(11)}(\cdot) u^{(1)} + W^{(1)}(\cdot)^t u^{(1)}, & \text{if } u^{(2)}(\cdot) = 0_{p-q} \\ \infty & \text{otherwise} \end{cases},$$

where $u = ((u^{(1)})^t, (u^{(2)})^t)^t$, $u^{(1)} \in \mathbb{R}^q$ and $u^{(2)} \in \mathbb{R}^{p-q}$. Therefore we obtain that the probability of the set $A = \{\sup_{\tau \in [\tau_L, \tau_U], q+1 \leq j \leq p} |\hat{\beta}_j^{AL}(\tau)| = 0\}$ converges to 1, which proves (3.5). On the set A the weak convergence of $\operatorname{argmin}_{u^{(1)}} \tilde{V}_n((u^{(1)}, 0_{p-q}), \cdot)$ follows as in the proof of Theorem 2.3. Combined with $P(A) \to 1$ the first assertion (3.4) of Theorem 3.3 follows. □

References

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

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

Chen, K. and Lo, S.-H. (1997). On the rate of uniform convergence of the product-limit estimator: Strong and weak laws. *The Annals of Statistics*, 25(3):pp. 1050–1087.

Csörgö, S. and Horváth, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 62:411–426.

Efron, B. (1967). The two-sample problem with censored data. In Le Cam, L. and Neyman, J., editors, *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics IV*, pages 831–853. Prentice-Hall, New York.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348–1360.

Fitzenberger, B. (1997). A guide to censored quantile regressions. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics: Robust Inference*, volume 15, pages 405–437. North-Holland, Amsterdam.

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

Geyer, C. J. (1996). On the asymptotics of convex stochastic optimization. *Unpublished Manuscript.*

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

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis*, 100:1816–1829.

Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *Annals of Statistics*, 26:755–770.

Knight, K. and Fu, W. (2000). Asymptotics for Lasso-type estimators. *Annals of Statistics*, 28:1356–1378.

Koenker, R. (2005). Quantile regression. In *Economic Society Monographs*. Cambridge University Press.

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

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

Leng, C. and Tong, X. (2012). A quantile regression estimator for censored data. *Bernoulli, in press.*

Louis, T. (1981). Nonparametric analysis of an accelerated failure time model. *Biometrika*, 68:381–390.

Miller, R. (1976). Least squares regression with censored data. *Biometrika*, 63:449–464.

Newey, W. K. and Powell, J. L. (1990). Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory*, 6:295–317.

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

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

Portnoy, S. and Lin, G. (2010). Asymptotics for censored regression quantiles. *Journal of Nonparametric Statistics*, 22:115–130.

Powell, J. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25:303–325.

Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics*, 32:143–155.

Shows, J. H., Lu, W., and Zhang, H. H. (2010). Sparse estimation and inference for censored median regression. *Journal of Statistical Planning and Inference*, 140(7):1903–1917.

Tang, Y., Wang, H., He, X., and Zhu, Z. (2012). An informative subset-based estimator for censored quantile regression. *TEST*, pages 1–21. 10.1007/s11749-011-0266-y.

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

Wang, H. and Wang, L. (2009). Locally weighted censored quamtile regression. *Journal of the American Statistical Association*, 2104:1117–1128.

Wang, H. J., Zhou, J., and Li, Y. (2012). Variable selection for censored quantile regression. *Statistica Sinica, in press*.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19:801–817.

Ying, Z., Jung, S., and Wei, L. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association*, 90:178–184.

Yu, K., Lu, Z., and Stauder, J. (2003). Quantile regression: applications and current research areas. *The Statistician*, 53:331–350.

Zhou, L. (2006). A simple censored median regression estimator. *Statistica Sinica*, 16:1043–1058.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Annals of Statistics*, 36:1108–1126.