Quantile Treatment Effect Estimators for Durations under the Conditional Independence Assumption

Michael Maier*
University of Mannheim

June 13, 2008

Abstract

This paper describes a method to estimate quantile treatment effects of a binary treatment variable on censored durations. The effects of interest are differences between quantiles of the counterfactual outcomes in both treatment states. Identification is based on the conditional independence assumption. Estimators for the whole population and for the subgroup of participants are proposed, consistency, asymptotic normality, and consistency of the variance estimators are shown. A simple transformation of the effects is given, which enables an easy interpretation of the results. A test procedure for several hypotheses on the whole quantile treatment effect process is described, that is, a testing method for hypotheses on treatment effects for all quantile levels simultaneously. Finally, pros and cons of the method compared to other approaches are discussed.

JEL-Classification: C14, C21, C24, C41.

Keywords: Conditional independence assumption, durations, econometric evaluation, Kolmogorov-Smirnov tests, quantile treatment effects.

---

*Contact address: University of Mannheim, Department of Economics, L 7, 3-5, 68131 Mannheim, Germany. Phone +49-621-181-3500, e-mail: michael.maier@uni-mannheim.de.
The author thanks Michael Lechner and Stefan Speckesser for helpful comments. Financial support by the German Research Foundation (DFG) through the research group *Heterogeneous Labor* at the University of Konstanz and the ZEW Mannheim is gratefully acknowledged.

1 Introduction

This paper describes a method to estimate quantile treatment effects of a binary treatment variable on censored durations. Using the framework of econometric evaluation methods, the effects of interest are the differences between quantiles of the counterfactual outcomes in both treatment states. Identification is based on the conditional independence assumption. The proposed model extends the model of Firpo (2007) by allowing censored outcomes.

Most approaches for analyzing durations are based on the concept of the hazard rate. Kiefer (1988) and van den Berg (2001) give economic examples and review numerous methods for estimating the hazard rates in a variety of different contexts. To derive causal effects in the presence of self-selected individuals, the general model of Abbring and van den Berg (2003a) examines the effect of a treatment at some point in time on an outcome duration. Abbring and van den Berg (2004) compare this approach to alternatives like basic (nondynamic) binary treatment effect models and panel data models. A method for implementation is proposed by Abbring and van den Berg (2003b), whereas Abbring (2003, 2006) give overviews and generalizations of the underlying basic dynamic approach. Abbring (2007) introduces a general framework where durations are defined by the crossing of a threshold by certain stochastic processes. In an instrumental variables setting, Abbring and van den Berg (2005) define treatment effects concerning hazard rates and consider identification and estimation in various situations with respect to time of treatment and compliance of the individuals to the instrumental variable.

The present paper uses a different approach to analyze treatment effects on durations. Instead of modelling hazard rates, the focus directly lies on durations. The influence of the treatment variable on the outcome duration is investigated by using the censored quantile regression model of Powell (1986), which adapts the quantile regression estimator of Koenker and Bassett (1978) to censored dependent variables. The applicability of quantile regressions for duration analyses is adressed by Koenker and Bilias (2001) and Koenker and Geling (2001) for the uncensored case. Empirical analyses of censored durations using quantile (or median) regressions are carried out by Horowitz and Neumann (1987) and Fitzenberger and Wilke (2005), for example.

The treatment effects may differ for each individual. Therefore, an evaluation approach based on the assumption of independence of the latent outcomes and the treatment conditional on the covariates is used. This is implemented by using the reweighting approach introduced in the econometric evaluation literature by Hirano, Imbens, and

Ridder (2003). In the following, the weighting scheme is applied to the censored quantile regression objective function of Powell (1986). This extends the quantile treatment effect model of Firpo (2007) by allowing censored outcomes. The propensity score necessary for this approach is estimated by the nonparametric series estimator of Horowitz and Mammen (2004), which avoids the strong differentiability assumption of the series estimator proposed by Hirano, Imbens, and Ridder (2003).

In contrast to the general timing-of-events approach (see Abbring and van den Berg (2003a), for example), the present approach does not make use of the exact time of the realization of a treatment. It considers whether a treatment occured in some time period and evaluates its impact on some subsequent outcome duration. In this respect, the present approach resembles to the multiple treatment effects framework, which considers how many treatments an individual choose in a certain time period and compares the outcomes to those of individuals which chose a different number of treatments, but without taking into account the exact position in time when treatments were taken (see Lechner and Miquel (2001)).

In the following, estimators for quantile treatment effects for the whole population and for the subgroup of participants are proposed. Section 2.1 describes the approach of Horowitz and Mammen (2004), which is used to estimate the propensity score. Section 2.2 describes the quantile treatment estimators and shows consistency and asymptotic normality. It also contains a simple transformation which makes a clear interpretation of the results possible. The transformation yields an expression of the relative change of the counterfactual durations. Variance estimators are defined and their consistency is shown in section 2.3. Quantile regression models enable in principle to estimate a continuum of quantile treatment effects. To test hypotheses for the whole quantile treatment effect process, a test procedure following Chernozhukov and Hansen (2006) and Chernozhukov and Fernandez-Val (2005) is described in section 3. Section 4 concludes. All proofs are given in the appendix.

2 Estimation Approach and Asymptotic Properties

2.1 Estimation of the Propensity Score

In this section, a short description of the series estimator of Horowitz and Mammen (2004) is given. Li and Racine (2007, sec. 15.3.3) provide a textbook account of this approach. Its first step is similar to the approach of Hirano, Imbens and Ridder (2003), but it avoids a strong differentiability assumption with respect to the true propensity score.

The treatment indicator D is is modelled as follows:

$$D = F(\mu + m_1(x_1) + \ldots + m_k(x_k)) + U. \quad (1)$$

Here, F is a known function (for example, the logit transformation, i.e., $F(x) = 1/(1+\exp(-x))$), x_j is the jth element of X, which is a k-dimensional random vector of (nonconstant) covariates, μ is an unknown constant, $m_j(x_j)$ is an unknown function of x_j, $j = 1,\ldots,k$, and U is an error term. Set $m(x) = \sum_{j=1}^{k} m_j(x_j)$. The unknown function $\mu + m(x)$ is approximated by a series of the elements of X. To this end, define

$$R_\kappa(X) = [1, r_1(x_1), \ldots, r_\kappa(x_1), r_1(x_2), \ldots, r_\kappa(x_2), \ldots, r_1(x_k), \ldots, r_\kappa(x_k)]',$$

where $\kappa \in \mathbb{N}$ and $r_l(x_j)$, $l = 1,\ldots,\kappa$, $j = 1,\ldots,k$, are elements of a series with $\kappa k + 1$ elements in total. The function $\mu + m(x)$ is approximated by a linear combination of the elements of $R_\kappa(x)$, i.e., by $R_\kappa(x)'\theta$, where $\theta \in \mathbb{R}^{\kappa k+1}$. The parameters θ of the approximation are determined by the following minimization problem:

$$\hat{\theta}_{n\kappa} = \arg\min_\theta S_{n\kappa} = \arg\min_\theta \frac{1}{n}\sum_{i=1}^{n}(D_i - F(R_\kappa(X_i)'\theta))^2. \quad (2)$$

As a result, preliminary estimates $\tilde{\mu} + \tilde{m}(X) = R_\kappa(X)'\hat{\theta}_{n\kappa}$ are obtained.

Until now, the procedure is basically identical to that of Hirano, Imbens, and Ridder (2003). Horowitz and Mammen (2004) propose to add a further step for the estimation of $m(x)$. To this end, define the following terms for $j = 0, 1$:

$$\begin{aligned} S'_{nj1}(x_1, \tilde{m}) &= -2\sum_{i=1}^{n}\left(D_i - F\left(\tilde{\mu} + \tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{X}_i)\right)\right) \\ &\quad \times F'\left(\tilde{\mu} + \tilde{m}_1(x_1) + \tilde{m}_{-1}(\tilde{X}_i)\right)(X_{1,i} - x_1)^j K_h(x_1 - X_{1,i}). \end{aligned}$$

Here, $\tilde{X} \equiv (x_2,\ldots,x_k)$, $\tilde{m}_{-1}(\tilde{X}_i) = \tilde{m}_2(x_{2,i}) + \ldots + \tilde{m}_k(x_{k,i})$, $x_{l,i}$ is observation i of element l of X. $K_h(u) = K(u/h)$ is a kernel function and $h > 0$ is a bandwidth. Further, define for $j = 0, 1, 2$:

$$S''_{nj1}(x_1, \tilde{m}) =$$

$$2\sum_{i=1}^{n} F\left(\tilde{\mu}+\tilde{m}_1(x_1)+\tilde{m}_{-1}(\tilde{X}_i)\right)^2 (X_{1,i}-x_1)^j K_h(x_1-X_{1,i})$$
$$-2\sum_{i=1}^{n}\left(D_i - F\left(\tilde{\mu}+\tilde{m}_1(x_1)+\tilde{m}_{-1}(\tilde{X}_i)\right)\right)$$
$$\times F''\left(\tilde{\mu}+\tilde{m}_1(x_1)+\tilde{m}_{-1}(\tilde{X}_i)\right)(X_{1,i}-x_1)^j K_h(x_1-X_{1,i}).$$

With these expressions, the second stage estimator of $m_1(x^1)$ can be defined by:

$$\hat{m}_1(x_1)=\tilde{m}_1(x_1)-\frac{S''_{n21}(x_1,\tilde{m})S'_{n01}(x_1,\tilde{m})-S''_{n11}(x_1,\tilde{m})S'_{n11}(x_1,\tilde{m})}{S''_{n01}(x_1,\tilde{m})S''_{n21}(x_1,\tilde{m})-S''_{n11}(x_1,\tilde{m})^2}.$$

$S'_{nj1}(x_1,\tilde{m})$ and $S''_{nj1}(x_1,\tilde{m})$ are formulated for the first nonconstant covariate. Similar expressions can also be defined for the second stage estimators of $m_l(x_l)$, $l=2,\ldots,k$. With estimates for $\hat{m}_1(\cdot),\ldots,\hat{m}_k(\cdot)$ at hand, a prediction of D (termed $\hat{p}(X)$ later on) can be obtained by $\hat{D}=F(\hat{\mu}+\hat{m}(X))$, where $\hat{\mu}=\tilde{\mu}$.

To derive the asymptotic properties of the estimator, Horowitz and Mammen (2004) state the following assumptions:

A1 The data $\{D_i,X_i\}_{i=1}^n$ are iid and the true model of the data is $E[D|X=x]=F(\mu+m(x))$.

A2 The support of X is (rescaled to be) $[-1,1]^k$, the distribution of X is absolutely continuous with respect to Lebesgue measure, the density of X is finite, bounded away from zero and twice differentiable. Define $U=D-F(\mu+m(X))$. The variance of U is bounded from below and above by c_V and C_V, respectively, and $E|U|^j \leqslant C_U^{j-2} j! E[U^2] < \infty$ for all $j \geqslant 2$ and a constant $C_U<\infty$.

A3 The unknown functions $m_j(\cdot)$, $j=1,\ldots,k$, are bounded in absolute value by a constant C_m and are twice differentiable. The known transformation F is bounded from above, its derivative is bounded from below and above. F is differentiable in a neighborhood of μ, where the size of the environment depends linearly on the number of covariates. The second derivative of F is Lipschitz continuous with constant C_{F3}.

A4 Define $Q_\kappa = E[F'(\mu+m(X))^2 R_\kappa(X)R_\kappa(X)']$. The absolute value of all elements of Q_κ is bounded from above, the smallest eigenvalue of Q_κ is bounded away from zero. The largest eigenvalue of $Q_\kappa^{-1}E[F'(\mu+m(X))^2\, Var(U|X)R_\kappa(X)R_\kappa(X)']Q_\kappa^{-1}$ is finite.

A5 The elements of the series r_k are symmetric orthonormal functions, i.e., $\int_{-1}^{1} r_k(v)dv = 0$, $\int_{-1}^{1} r_j(v)r_k(v)dv = \delta_{j,k}$, where $\delta_{j,k} = 1\{j = k\}$. Assume $\zeta_\kappa = \sup_{x\in\text{supp}(X)} \|R_\kappa(x)\| > c_\kappa$ for sufficiently large κ and $\zeta_\kappa = O(\sqrt{\kappa})$ for $\kappa \to \infty$. For $C_\theta < \infty$ and $\theta_{\kappa 0} \in [-C_\theta, C_\theta]^{k\kappa+1}$ it holds that $\sup_{x\in\text{supp}(X)} |\mu + m(x) - R_\kappa(x)'\theta_{\kappa 0}| = O(\kappa^{-2})$ for $\kappa \to \infty$. Finally, $\theta_\kappa \in \text{int}[-C_\theta, C_\theta]^{k\kappa+1}$.

A6 New elements are included in the series with rate $\kappa = C_\kappa n^{4/15+\nu}$ for $0 < C_\kappa < \infty$ and $0 < \nu < 1/30$. For the bandwidth h it holds that $h = C_h n^{-1/5}$ for $0 < C_h < \infty$.

A7 The kernel function is bounded, continuous, symmetric and integrates to one.

Assumption A1 is standard. Assumption A2 imposes no restrictions on covariates with support larger than $[-1, 1]$, as for any (finite) $x \in \mathbb{R}$, a value obtained by the monotone transformation $T : \mathbb{R} \to [-1, 1]$, defined as

$$T(x) \quad = \quad 2\frac{x - \min_i\{x_i\}}{\max_i\{x_i\} - \min_i\{x_i\}} - 1,$$

does not alter the estimation problem and fulfills the assumption. Discrete covariates are more difficult to handle; if the sample is large enough, the model can be estimated separately for all cells of combinations of the values of the discrete covariates (for a review of adaptions of nonparametric (kernel-based) methods to discrete variables, see sections 3 and 4 of Li and Racine (2007)). Assumption A3 requires the true functions only to be two times differentiable. Note that this is a considerably weaker assumption than that of Hirano, Imbens, and Ridder (2003), who require a number of derivatives of the true propensity score which is a multiple of the dimension of the covariates. Assumption A4 ensures regular properties of the covariance matrix of the first-stage estimates. A5 concerns the approximation error of the series estimator and its convergence properties, A6 the rate of inclusion of new elements in the series and the rate with which the bandwidth h converges to zero. A7 contains standard assumptions on the kernel function. For more discussion of the technical assumptions, see Horowitz and Mammen (2004).

Under these assumptions, Horowitz and Mammen (2004) derive the asymptotic properties of the estimator. They show consistency and asymptotic normality and derive the speed of convergence of the estimator. They also discuss bandwidth choice for the estimator. Their proposed procedure for this purpose is not directly suited for the framework of the present paper, since optimal estimation of the propensity score does not necessarily imply optimal estimation of the quantile treatment effects with respect to some mean square error criterion. For an analysis of this problem in the context of mean treatment effects, see Frölich (2005), Ichimura and Linton (2005), or Imbens, Newey, and Ridder (2005).

2.2 Quantile Treatment Effect Estimators for Censored Durations

Let T be a duration and D a binary indicator of treatment participation. Assume that for both treatment states latent values T_1 and T_0 exist. Define $q_\tau(T)$ to be the τ-quantile of T and $q_{j,\tau} = q_\tau(T_j)$, $j \in \{0, 1\}$, as the τ-quantiles of the latent outcomes. The interest of the present paper lies in estimating the quantile treatment effect Δ_τ of D on T, which is given by

$$\Delta_\tau = q_\tau(T_1) - q_\tau(T_0). \tag{3}$$

Note that this is not the τ-quantile of the difference $T_1 - T_0$, as long as one is not willing to assume rank invariance, i.e., that all individuals take the same ranks in both treatment states. See for example Chernozhukov and Hansen (2005) or Firpo (2007) for a discussion of rank invariance or, synonymously, rank preservation.

Eq. (3) is the quantile treatment effect for an average individual of the population. Another effect of interest is the treatment effect on the treated, which is the effect for an arbitrary individual of the subpopulation of participants. Denote this treatment effect by

$$\Delta_{\tau|D=1} = q_{\tau|D=1}(T_1) - q_{\tau|D=1}(T_0). \tag{4}$$

Here, $q_{\tau|D=1}(T)$ is the τ-quantile of T for the subgroup of participants.

Firpo (2007) shows identification of the quantiles of T_1 and T_0 under the conditional independence assumption by using the reweighting approach of Hirano, Imbens, and Ridder (2003). Without censoring of T, Firpos estimation approach would be directly applicable by using $\log T$ as outcome variable. The log-transformation is necessary for the estimation procedure to handle outcomes which cannot become negative. If T is censored, which is usually the case in duration analysis, this approach would lead to inconsistent estimates. Therefore, the present paper combines the reweighting approach with results of Powell (1986) for censored quantile regressions.

Let $\{\tilde{T}_i, C_i, D_i, X_i\}_{i=1}^n$ be a random sample. $\tilde{T}_i$ is the observed (possibly censored) outcome, which is equal to T_i if $\tilde{T}_i \leqslant C_i$ and is equal to C_i otherwise, where C_i is the censoring time. Note that the censoring time has to be known also for uncensored observations.

This requirement is necessary for application of the approach of Powell (1986). In the case of censoring due to the end of the survey, this information is clearly available. This situation is frequently encountered when analyzing administrative data sets of active labour market programs, for example. D_i is the treatment participation indicator, and X_i is a k-dimensional vector of covariates.

Extending the weighted quantile regression estimators of Firpo (2007) by the objective function of censored quantile regressions, the following estimators are suggested:

$$\begin{aligned}\hat{\Delta}_\tau &= \arg\min_{q_{1,\tau}} \frac{1}{n}\sum_{i=1}^{n} \frac{D_i}{\hat{p}(X_i)}\rho_\tau(\tilde{T}_i - \min\{q_{1,\tau}, C_i\}) \\ &\quad - \arg\min_{q_{0,\tau}} \frac{1}{n}\sum_{i=1}^{n} \frac{1-D_i}{1-\hat{p}(X_i)}\rho_\tau(\tilde{T}_i - \min\{q_{0,\tau}, C_i\}),\end{aligned}$$

$$\begin{aligned}\hat{\Delta}_{\tau|D=1} &= \arg\min_{q_{1,\tau|D=1}} \frac{1}{n}\sum_{i=1}^{n} \frac{D_i}{\bar{p}}\rho_\tau(\tilde{T}_i - \min\{q_{1,\tau|D=1}, C_i\}) \\ &\quad - \arg\min_{q_{0,\tau|D=1}} \frac{1}{n}\sum_{i=1}^{n} \frac{1-D_i}{\bar{p}}\frac{\hat{p}(X_i)}{1-\hat{p}(X_i)}\rho_\tau(\tilde{T}_i - \min\{q_{0,\tau|D=1}, C_i\}),\end{aligned}$$

where $\hat{p}(\cdot)$ is a nonparametric estimate of the propensity score $E[D|X]$, $\bar{p}$ is an estimator of the unconditional expectation of D, and $\rho_\tau(u) \equiv (\tau - 1\{u < 0\})u$, $\tau \in (0,1)$, is the usual quantile regression check function (see Koenker and Bassett (1978) or Koenker (2005)). The identification results of Firpo (2007) carry over to the present adaption, as noninformative censoring is assumed. Estimation of the propensity score is proposed to be carried out by the nonparametric series estimator of Horowitz and Mammen (2004), which was described in section 2.1. Asymptotic properties of the estimators are given by the following theorem.

Theorem 1: *In addition to assumptions A1 - A7, assume the following:*

A8 Conditional independence of D and the latent outcomes T_1 and T_0: $D \perp\!\!\!\perp (T_1, T_0)|X$.

A9 The propensity score $p(X) \equiv E[D|X]$ is bounded away from zero and one.

A10 The latent quantiles $q_\tau(T_j)$ and $q_{\tau|D=1}(T_j)$, $j \in \{0,1\}$, take values in a compact set $\mathcal{Q} \subset \mathbb{R}$.

Then the following holds:

1. $\hat{\Delta}_\tau - \Delta_\tau = o_p(1)$ *and* $\hat{\Delta}_{\tau|D=1} - \Delta_{\tau|D=1} = o_p(1)$.

2. $\sqrt{n}(\hat{\Delta}_\tau - \Delta_\tau) \xrightarrow{D} \mathcal{N}(0, \Omega_\tau)$ *and* $\sqrt{n}(\hat{\Delta}_{\tau|D=1} - \Delta_{\tau|D=1}) \xrightarrow{D} \mathcal{N}(0, \Omega_{\tau|D=1})$.

Discussion of the variance expressions and their estimation is deferred to section 2.3. For identification of $\Delta_{\tau|D=1}$, assumptions A8 and A9 can be weakened to $D \perp\!\!\!\perp T_0|X$ and $p(X) < 1$ (see Imbens (2004), for example).

Usually interest lies in the effect of D on the duration T as such. To circumvent the need for an estimation procedure which handles the nonnegativity of the outcome, however, the logarithm of T is used for estimation. The quantile treatment effects are therefore the differences of the quantiles of the logarithms of the considered durations, which makes direct interpretation difficult. A simple transformation gives an expression which is more easily to interpret:

$$\begin{aligned}\exp(\Delta_\tau) &= \exp(q_\tau(\ln T_1) - q_\tau(\ln T_0)) \\ &= \exp(\ln(q_\tau(T_1)) - \ln(q_\tau(T_0))) \\ &= \exp\left(\ln\left(\frac{q_\tau(T_1)}{q_\tau(T_0)}\right)\right) \\ &= \frac{q_\tau(T_1)}{q_\tau(T_0)}.\end{aligned}$$

The second equality follows by the invariance of quantiles with respect to monotone (rank-preserving) transformations (see for example Koenker and Geling (2001, sec. 2.3)). Therefore, $\exp(\hat{\Delta}_\tau)$ is an estimate of the relative increase or decline of the τ-quantile of the duration due to the treatment. For example, $\exp(\hat{\Delta}_\tau) = .9$ means that the treatment decreases the τ-quantile of the outcome duration by ten percent. Note that this is a causal effect only under the assumption of rank invariance. The distribution of $\exp(\hat{\Delta}_\tau)$ follows directly by the continuous mapping theorem and by the Delta method:

$$\begin{aligned}\sqrt{n}(\exp(\hat{\Delta}_\tau) - \exp(\Delta_\tau)) &\xrightarrow{D} \mathcal{N}(0, (\partial \exp(\Delta_\tau)/\partial \Delta_\tau)^2 \Omega_\tau) \\ &= \mathcal{N}(0, \exp(\Delta_\tau)^2 \Omega_\tau).\end{aligned}$$

2.3 Variance Estimation

To derive asymptotic variances of the estimators, a number of terms are defined. First, following Chen, Linton, and Van Keilegom (2003), consider the following representations of the objective functions for Δ_τ and $\Delta_{\tau|D=1}$:

$$M_\tau(q_\tau, p) = E\left[\frac{D}{p(X)} 1\{q_{1,\tau} < C\}(\tau - 1\{\tilde{T} \leqslant q_{1,\tau}\})\right.$$

$$-\frac{1-D}{1-p(X)}1\{q_{0,\tau} < C\}(\tau - 1\{\tilde{T} \leqslant q_{0,\tau}\})\Bigg]$$

$$M_{\tau|D=1}(q_{\tau|D=1}, p)$$
$$= E\left[\frac{D}{\check{p}}1\{q_{1,\tau|D=1} < C\}(\tau - 1\{\tilde{T} \leqslant q_{1,\tau|D=1}\})\right.$$
$$\left.-\frac{1-D}{\check{p}}\frac{p(X)}{1-p(X)}1\{q_{0,\tau|D=1} < C\}(\tau - 1\{\tilde{T} \leqslant q_{0,\tau|D=1}\})\right].$$

Here, $\check{p}$ denotes $E[D]$. The objectives depend on the finite dimensional parameters $q_\tau = (q_{1,\tau}, q_{0,\tau})'$ or $q_{\tau|D=1} = (q_{1,\tau|D=1}, q_{0,\tau|D=1}, \check{p})'$ and the infinite dimensional parameter $p(X)$, which is partly abbreviated by p in the following.

The derivatives of $M_\tau(q_\tau, p)$ and $M_{\tau|D=1}(q_{\tau|D=1}, p)$ with respect to the finite dimensional parameters are given by $\Gamma_{1,\tau}(q, p^*)$ and $\Gamma_{1,\tau|D=1}$ (q, p^*), respectively, those with respect to the infinite dimensional parameter in direction $p-p^*$ by $\Gamma_{2,\tau}(q,p^*)[p-p^*]$ and $\Gamma_{2,\tau|D=1}(q,p^*)[p-p^*]$:

$$\Gamma_{1,\tau}(q,p^*) = \begin{pmatrix} -E\left[\frac{D}{p^*(X)}(f_C(q_{1,\tau})(\tau - F_{\tilde{T}|X}(q_{1,\tau})) \right. \\ \left. +(1-F_C(q_{1,\tau}))f_{\tilde{T}|X}(q_{1,\tau}))\right] \\ E\left[\frac{1-D}{1-p^*(X)}(f_C(q_{0,\tau})(\tau - F_{\tilde{T}|X}(q_{0,\tau})) \right. \\ \left. +(1-F_C(q_{0,\tau}))f_{\tilde{T}|X}(q_{0,\tau}))\right] \end{pmatrix}$$

$$\Gamma_{1,\tau|D=1}(q,p^*)$$
$$= \begin{pmatrix} -E\left[\frac{D}{\check{p}}(f_C(q_{1,\tau|D=1})(\tau - F_{\tilde{T}|X}(q_{1,\tau|D=1})) \right. \\ \left. +(1-F_C(q_{1,\tau|D=1}))f_{\tilde{T}|X}(q_{1,\tau|D=1}))\right] \\ E\left[\frac{1-D}{\check{p}}\frac{p^*(X)}{1-p^*(X)}(f_C(q_{0,\tau|D=1})(\tau - F_{\tilde{T}|X}(q_{0,\tau|D=1})) \right. \\ \left. +(1-F_C(q_{0,\tau|D=1}))f_{\tilde{T}|X}(q_{0,\tau|D=1}))\right] \\ -E\left[\frac{D}{\check{p}^2}(1-F_C(q_{1,\tau|D=1}))(\tau - F_{\tilde{T}|X}(q_{1,\tau|D=1})) \right. \\ \left. -\frac{1-D}{\check{p}^2}\frac{p^*(X)}{1-p^*(X)}(1-F_C(q_{0,\tau|D=1}))(\tau - F_{\tilde{T}|X}(q_{0,\tau|D=1}))\right] \end{pmatrix}$$

$$\Gamma_{2,\tau}(q,p^*)[p-p^*]$$
$$= -E\left[\frac{D(p(X)-p^*(X))}{(p^*(X))^2}(1-F_C(q_{1,\tau}))(\tau - F_{\tilde{T}|X}(q_{1,\tau}))\right.$$
$$\left.-\frac{(1-D)(p(X)-p^*(X))}{(1-p^*(X))^2}(1-F_C(q_{0,\tau}))(\tau - F_{\tilde{T}|X}(q_{0,\tau}))\right]$$

$$\Gamma_{2,\tau|D=1}(q,p^*)[p-p^*]$$

$$
\begin{aligned}
= \quad & -E\left[\frac{1-D}{\check{p}}\frac{p(X)-p^*(X)}{(1-p^*(X))^2}(1-F_C(q_{0,\tau|D=1}))\right. \\
& \left.\times(\tau - F_{\tilde{T}|X}(q_{0,\tau|D=1}))\right].
\end{aligned}
$$

Here, q^*, p^* and $\check{p}$ are the true values of q, p, and $E[D]$, respectively. F_C and f_C are the cumulative distribution function and the density of the censoring time, $F_{\tilde{T}|X}$ and $f_{\tilde{T}|X}$ are those of the outcome duration. For a note on the derivation of the above expressions, see part 2.2 of the proof of Theorem 1.

Consider the following expression:

$$
M(q^*, p^*) + \Gamma_2(q^*, p^*)[\hat{p} - p^*].
$$

Condition 2.6 of Theorem 2 of Chen, Linton, and Van Keilegom (2003) assumes the existence of an matrix V for which it holds that

$$
\sqrt{n}(M(q^*, p^*) + \Gamma_2(q^*, p^*)[\hat{p} - p^*]) \xrightarrow{D} \mathcal{N}(0, V).
$$

Of course, separate expressions of V exist for both estimators $\hat{\Delta}_\tau$ and $\hat{\Delta}_{\tau|D=1}$. Define

$$
\begin{aligned}
\xi_{\tau,i} \quad = \quad & \frac{D}{p^*(X)} 1\{q^*_{1,\tau} < C\}(\tau - 1\{\tilde{T} \leqslant q^*_{1,\tau}\}) \\
& -\frac{1-D}{1-p^*(X)} 1\{q^*_{0,\tau} < C\}(\tau - 1\{\tilde{T} \leqslant q^*_{0,\tau}\}) \\
& -E\left[\frac{D(p(X)-p^*(X))}{(p^*(X))^2}(1-F_C(q^*_{1,\tau}))(\tau - F_{\tilde{T}|X}(q^*_{1,\tau}))\right. \\
& \left.\quad + \frac{(1-D)(\hat{p}(X)-p^*(X))}{(1-p^*(X))^2}(1-F_C(q^*_{0,\tau}))(\tau - F_{\tilde{T}|X}(q^*_{0,\tau}))\right].
\end{aligned}
$$

The variance V_τ is then given by $Var(\xi_{\tau,i}) = E[\xi^2_{\tau,i}] - (E[\xi_{\tau,i}])^2$. $V_{\tau|D=1}$ is equal to $Var(\xi_{\tau|D=1,i}) = E[\xi^2_{\tau|D=1,i}] - (E[\xi_{\tau|D=1,i}])^2$, where $\xi_{\tau|D=1,i}$ is given by:

$$
\begin{aligned}
\xi_{\tau|D=1,i} \quad = \quad & \frac{D}{\check{p}^*} 1\{q^*_{1,\tau|D=1} < C\}(\tau - 1\{\tilde{T} \leqslant q^*_{1,\tau|D=1}\}) \\
& -\frac{1-D}{\check{p}^*}\frac{p^*(X)}{1-p^*(X)} 1\{q^*_{0,\tau|D=1} < C\}(\tau - 1\{\tilde{T} \leqslant q^*_{0,\tau|D=1}\}) \\
& -E\left[\frac{1-D}{\check{p}^*}\frac{p(X)-p^*(X)}{(1-p^*(X))^2}(1-F_C(q^*_{0,\tau|D=1}))(\tau - F_{\tilde{T}|X}(q^*_{0,\tau|D=1}))\right].
\end{aligned}
$$

Then, following Theorem 2 of Chen, Linton, and Van Keilegom (2003), the variances Ω_τ and $\Omega_{\tau|D=1}$ of Δ_τ and $\Delta_{\tau|D=1}$ are given by:

$$
\Omega_\tau \quad = \quad \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \Gamma'_{1,\tau} W V_\tau W \Gamma_{1,\tau} \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1},
$$

$$\Omega_{\tau|D=1} = \left(\Gamma'_{1,\tau|D=1} W \Gamma_{1,\tau|D=1}\right)^{-1} \Gamma'_{1,\tau|D=1} W V_{\tau|D=1} W \times \Gamma_{1,\tau|D=1} \left(\Gamma'_{1,\tau|D=1} W \Gamma_{1,\tau|D=1}\right)^{-1},$$

where W is some symmetric positive definite matrix.

To derive estimators of the variances Ω_τ and $\Omega_{\tau|D=1}$, estimators of all components are needed. Define

$$\begin{aligned}
\hat{\xi}_{\tau,i} &= \frac{D_i}{\hat{p}(X_i)} 1\{\hat{q}_{1,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant \hat{q}_{1,\tau}\}) \\
&- \frac{1-D_i}{1-\hat{p}(X_i)} 1\{\hat{q}_{0,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant \hat{q}_{0,\tau}\}) \\
&- \hat{E}\left[\frac{D_i(\hat{p}(X_i) - \hat{E}[\hat{p}(X_i)])}{\hat{p}(X_i)^2}(1 - \hat{F}_C(\hat{q}_{1,\tau}))(\tau - \hat{F}_{\tilde{T}_i|X_i}(\hat{q}_{1,\tau})) \right. \\
&\left. + \frac{(1-D_i)(\hat{p}(X_i) - \hat{E}[\hat{p}(X_i)])}{(1-\hat{p}(X_i))^2}(1 - \hat{F}_C(\hat{q}_{0,\tau}))(\tau - \hat{F}_{\tilde{T}_i|X_i}(\hat{q}_{0,\tau}))\right]
\end{aligned}$$

and

$$\begin{aligned}
\hat{\xi}_{\tau,i|D=1} &= \frac{D_i}{\bar{p}} 1\{\hat{q}_{1,\tau|D=1} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant \hat{q}_{1,\tau|D=1}\}) \\
&- \frac{1-D_i}{\bar{p}} \frac{\hat{p}(X_i)}{1-\hat{p}(X_i)} 1\{\hat{q}_{0,\tau|D=1} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant \hat{q}_{0,\tau|D=1}\}) \\
&- \hat{E}\left[\frac{1-D_i}{\bar{p}} \frac{\hat{p}(X_i) - \hat{E}[\hat{p}(X_i)]}{(1-\hat{p}(X_i))^2}(1 - \hat{F}_C(\hat{q}_{0,\tau|D=1}))(\tau - \hat{F}_{\tilde{T}|X}(\hat{q}_{0,\tau|D=1}))\right].
\end{aligned}$$

Using the mean as an estimator of the unconditional expectation, the variances V_τ and $V_{\tau|D=1}$ can be estimated by

$$\begin{aligned}
\hat{V}_\tau &= \frac{1}{n}\sum_{i=1}^n \hat{\xi}^2_{\tau,i} - \left(\frac{1}{n}\sum_{i=1}^n \hat{\xi}_{\tau,i}\right)^2 \\
\hat{V}_{\tau|D=1} &= \frac{1}{n}\sum_{i=1}^n \hat{\xi}^2_{\tau,i|D=1} - \left(\frac{1}{n}\sum_{i=1}^n \hat{\xi}_{\tau,i|D=1}\right)^2.
\end{aligned}$$

For estimation of $\xi_{\tau,i}$ and $\xi_{\tau,i|D=1}$, replace expectations by means, estimate for example $E[p(X)]$ by $n^{-1}\sum_{i=1}^n \hat{p}(X_i)$ and $\bar{p}$ by $n^{-1}\sum_{i=1}^n D_i$. The (unconditional) cumulative distribution function of the censoring times may be estimated simply by $\hat{F}_C(q) = n^{-1}\sum_{i=1}^n 1\{C_i < q\}$. The conditional cumulative distribution function of $\tilde{T}$ at some point t given X may be estimated by some regression of $1\{\tilde{T} < t\}$ on X. To circumvent misspecification of the functional form, a nonparametric approach is preferable. For example, the model used for estimating the propensity score can also be used here. Note that only two different regressions are necessary for each variance estimate, i.e. regressions of $1\{\tilde{T} < \hat{q}_{1,\tau}\}$ and $1\{\tilde{T} < \hat{q}_{0,\tau}\}$ (or $1\{\tilde{T} < \hat{q}_{1,\tau|D=1}\}$ and $1\{\tilde{T} < \hat{q}_{0,\tau|D=1}\}$) on X.

It remains to derive estimators for $\Gamma_{1,\tau}(q,p)$ and $\Gamma_{1,\tau|D=1}(q,p)$. This can be done analogously to the estimators of V_τ and $V_{\tau|D=1}$; for example, an estimator of $\Gamma_{1,\tau}(q,p)$ is given by

$$\hat{\Gamma}_{1,\tau}(\hat{q},\hat{p}) = \begin{pmatrix} -\hat{E}\left[\frac{D_i}{\hat{p}(X_i)}(\hat{f}_C(\hat{q}_{1,\tau})(\tau - \hat{F}_{\tilde{T}|X}(\hat{q}_{1,\tau})) \right. \\ \left. \qquad +(1-\hat{F}_C(\hat{q}_{1,\tau}))\hat{f}_{\tilde{T}|X}(\hat{q}_{1,\tau}))\right] \\ -\hat{E}\left[\frac{1-D_i}{1-\hat{p}(X_i)}(\hat{f}_C(\hat{q}_{0,\tau})(\tau - \hat{F}_{\tilde{T}|X}(\hat{q}_{0,\tau})) \right. \\ \left. \qquad +(1-\hat{F}_C(\hat{q}_{0,\tau}))\hat{f}_{\tilde{T}|X}(\hat{q}_{0,\tau}))\right] \end{pmatrix}.$$

The densities of the censoring times and of the outcome given X can be estimated by standard kernel estimators (for estimation of conditional densities, see sec. 5 of Li and Racine (2007)).

Given all these estimators, the estimated variances of Δ_τ and $\Delta_{\tau|D=1}$ are given by:

$$\begin{aligned} \hat{\Omega}_\tau &= \left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} \hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1}, \\ \hat{\Omega}_{\tau|D=1} &= \left(\hat{\Gamma}'_{1,\tau|D=1} W \hat{\Gamma}_{1,\tau|D=1}\right)^{-1} \hat{\Gamma}'_{1,\tau|D=1} W \hat{V}_{\tau|D=1} W \hat{\Gamma}_{1,\tau|D=1} \\ &\quad \times \left(\hat{\Gamma}'_{1,\tau|D=1} W \hat{\Gamma}_{1,\tau|D=1}\right)^{-1}, \end{aligned}$$

Consistency of the variance estimators is addressed by the following theorem.

Theorem 2: $\hat{\Omega}_\tau - \Omega_\tau = o_P(1)$ *and* $\hat{\Omega}_{\tau|D=1} - \Omega_{\tau|D=1} = o_P(1)$.

3 Hypothesis Tests for the Quantile Treatment Effect Process

To judge on all quantile treatment effects simultaneously, hypotheses on the quantile treatment effect process can be tested. A number of hypotheses are of interest:

1. Zero effects: $H_0 : \Delta_\tau = 0 \ \forall \ \tau$ vs. $H_A : \exists \tau_0$ such that $\Delta_{\tau_0} \neq 0$.
2. Constant effects: $H_0 : \Delta_\tau = \Delta_{\tau_0} \ \forall \ \tau$ and some τ_0 vs. $H_A : \exists \tau_0, \tau_1$ such that $\Delta_{\tau_0} \neq \Delta_{\tau_1}$.
3. Only positive effects: $H_0 : \Delta_\tau \geqslant 0 \ \forall \ \tau$ vs. $H_A : \exists \tau_0$ such that $\Delta_{\tau_0} < 0$.

4. No difference to conventional censored quantile regression: $H_0 : \Delta_\tau = \beta_\tau \; \forall \; \tau$ vs. $H_A : \exists \tau_0$ such that $\Delta_{\tau_0} \neq \beta_{\tau_0}$, where β_τ is the probability limit of the estimated parameter of D from an ususal (unweighted) censored quantile regression of $\tilde{T}$ on D and X.
5. No difference between treatment effects for the whole population and for the treated: $H_0 : \Delta_\tau = \Delta_{\tau|D=1} \; \forall \; \tau$ vs. $H_A : \exists \tau_0$ such that $\Delta_{\tau_0} \neq \Delta_{\tau_0|D=1}$.
6. Larger effects for the treated as for the whole population: $H_0 : \Delta_{\tau|D=1} \geqslant \Delta_\tau \; \forall \; \tau$ vs. $H_A : \exists \tau_0$ such that $\Delta_{\tau_0|D=1} < \Delta_{\tau_0}$.

The first four hypotheses were proposed by Chernozhukov and Hansen (2006) and Chernozhukov and Fernandez-Val (2005), the last two emerge in the context of the present model. Of course, hypotheses 1 to 4 are also applicable to $\Delta_{\tau|D=1}$. The last two hypotheses can be used to consider whether the individuals of the population which benefit most from the treatment are indeed selected to participate. If H_0 of hypothesis 6 is rejected, this is a hint that the selection is not based solely on the extend of the expected treatment effect, or that the expectations were incorrect. For a further discussion of the relationship between overall treatment effects and treatment effects for the treated, see for example Smith (2000).

All null hypotheses can be stated by the following expression:

$$\upsilon_n(\tau) \equiv \Delta_\tau - a(\tau) = 0 \; \forall \; \tau, \tag{5}$$

where the scalar $a(\tau)$ might be unknown. In this case, it will be substituted by an estimate $\hat{a}(\tau)$. Viewed as a function of $\tau \in (0, 1)$, eq. (5) is called the quantile treatment effect process. Eq. (5) is a simplified version of the expression of Chernozhukov and Hansen (2006). Their estimation approach yields a vector of estimated coeffiecients, for which hypotheses can be stated by more general expressions. As the approach of the present paper only involves scalar estimates (i.e., the scalar treatment effects), a simplification is possible.

Hypotheses 1-6 may be fitted in this framework as follows: For testing the hypothesis of a zero effect, let $a(\tau) \equiv 0$. To test for constant effects, set $a(\tau) = \Delta_{\tau_0}$ for all τ and for some arbitrary τ_0. The dominance hypotheses may not be fitted directly in this framework, consider here for hypothesis 3 directly the smallest value of Δ_τ, or, for technical reasons of the definition of the test statistic, the equivalent expression $\max\{-\Delta_\tau, 0\}$. Hypothesis 6 can be based on $\max\{-(\Delta_{\tau|D=1} - \Delta_\tau), 0\}$. To test whether the treatment effects differ from conventional censored quantile regression estimates, let $a(\tau) \equiv \hat{\beta}_\tau$, where $\hat{\beta}_\tau$ is the estimate of D from a usual censored quantile regression. Hypothesis 5 may be tested simply by setting $a(\tau) = \Delta_{\tau|D=1}$.

To test the hypotheses, Kolmogorov–Smirnov tests are used (for a description of these tests, see sec. 19.3 of van der Vaart (1998), for example). Kolmogorov–Smirnov tests can be used for hypotheses which are defined not for a countably finite set of values (for example, for some coefficients like in the standard F-test), but for a continuum of values like in the present case, where the hypotheses consider quantile treatment effects for all $\tau \in (0, 1) \subset \mathbb{R}$. The Kolmogorov–Smirnov test statistic is the scaled supremum of the test process $\upsilon_n(\tau)$:

$$S_n = \sqrt{n} \sup_{\tau \in (0,1)} \|\upsilon_n(\tau)\|.$$

To test the dominance hypothesis 3, use directly

$$S_n = \sqrt{n} \sup_{\tau \in (0,1)} \max\{-\Delta_\tau, 0\}$$

and, for hypothesis 6,

$$S_n = \sqrt{n} \sup_{\tau \in (0,1)} \max\{-(\Delta_{\tau|D=1} - \Delta_\tau), 0\}.$$

The distribution of the test statistics under the null hypothesis is unknown (see van der Vaart (1998, sec. 19.3)). Therefore, Chernozhukov and Hansen (2006) and Chernozhukov and Fernandez-Val (2005) propose a resampling procedure (see also Abadie (2002) for a similar problem). The critical value of the test statistic may be obtained by the following algorithm:

1. Calculate the test statistics S_n of the various hypotheses using the original dataset.
2. Draw a random sample with replacement. Compute the bootstrap test statistics $\tilde{S}_{n,j}$, where j denotes the number of the resample. Repeat this step B times.
3. Compute the p-value as

$$\frac{1}{B} \sum_{j=1}^{B} 1 \left\{ \tilde{S}_{n,j} > S_n \right\}.$$

The bootstrapped critical value $\tilde{c}$ is the value of $\tilde{S}_n$ which is exceeded by $(1-\alpha)B$ realizations of $\tilde{S}_n$, where α is the significance level of the test.

Asymptotic properties of the test procedure are proven in the following theorem. This corresponds to Theorem 4 of Chernozhukov and Hansen (2006). As the optimization of the estimator of the present paper does not involve time consuming grid searches, no score approximations as in Chernozhukov and Hansen (2006) are used. Furthermore, no subsampling scheme is used, as the nonparametric estimation of the propensity score should be based on as much observations as possible (i.e., the bootstrap samples contain n observations).

Theorem 3: *If $\sqrt{n}(\hat{\Delta}_{(\cdot)} - \Delta_{(\cdot)}) \Rightarrow b(\cdot)$ and $\sqrt{n}(\hat{a}(\cdot) - a(\cdot)) \Rightarrow d(\cdot)$ jointly in ℓ^∞, where $b(\cdot)$ and $d(\cdot)$ are mean zero Gaussian processes with possibly different laws under the null and the alternative hypothesis, then the following holds:*

1. *If $\Delta_\tau - a(\tau) = 0 \; \forall \; \tau$, then $S_n \xrightarrow{D} S \equiv f(\upsilon(\cdot))$, where $\upsilon(\cdot) = b(\cdot) - d(\cdot)$ and $f(\cdot)$ is the Kolmogorov-Smirnov test statistic.*

Let $\tilde{c}$ be the bootstrapped critical value. Then:

2. *If furthermore υ has a nondegenerate covariance kernel, and for $\alpha < 1/2$, $P(S_n > \tilde{c}_{1-\alpha}) \rightarrow \alpha = P(f(\upsilon(\cdot)) > c_{1-\alpha})$, where for the critical value $c_{1-\alpha}$ it holds that $P(f(\upsilon(\cdot)) > c_{1-\alpha}) = \alpha$.*

3. *If $\Delta_\tau - a(\tau) \neq 0$ for some τ, then $S_n \xrightarrow{D} \infty$ and $P(S_n > \tilde{c}_{1-\alpha}) \rightarrow 1$.*

The assumption that $\alpha < 1/2$ is not restrictive, as the significanc level of a test usually is choosen no larger than 10 per cent (i.e., α is usually less than or equal 1/10). The assumption of a nondegenerate covariance kernel means that for two arbitrary scalar values τ_0 and τ_1, the joint (two-dimensional) distribution of $\upsilon(\tau_0)$ and $\upsilon(\tau_1)$ is nondegenerate.

4 Discussion

From a theoretical point of view, distributions of durations and hazard rates or survival functions are equivalent. Efron and Johnstone (1990) and Ritov and Wellner (1988) show that a one-to-one mapping between the set of distributions of durations and the set of hazard functions exist. Therefore, the two representations contain the same information.

For practical analyses, the ease of interpretation and the appropriateness for the question at hand differ for the two approaches. The hazard rate is a function of time, whereas the quantile treatment effects of durations may be well subsumed by a finite set of scalars. This might be advantageous for the presentation and the communication of the results.

Summing up, the quantile treatment effect estimators for durations proposed in this paper are a simple way to evaluate heterogeneous effects of a treatment on censored durations. They only consider whether or not a treatment was chosen, and do not use the exact point in time of treatment realization as timing-of-events approaches do. They yield, however, results which are easy to interpret and to communicate.

The semiparametric efficiency of the estimators remains to be considered. A possible extension would be to derive results for random instead of fixed censoring (see Honoré, Khan, and Powell (2002), for example).

A Proofs

Proof of Theorem 1: First, consistency of the estimators is shown. Due to the nonsmooth character of the objective function and the use of an infinite dimensional parameter, consistency of $\hat{\Delta}_\tau$ is shown by checking the conditions of Theorem 1 of Chen, Linton, and Van Keilegom (2003). Define $Z_i \equiv (C_i, D_i, X_i)$ and let $q_\tau = (q_{1,\tau}, q_{0,\tau})'$ ($\in \mathcal{Q} \times \mathcal{Q}$) denote the vector of finite dimensional parameters, and define also

$$\begin{aligned} M_n(q_\tau, p) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i}{p(X_i)}1\{q_{1,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant q_{1,\tau}\})\right. \\ &\qquad \left. - \frac{1-D_i}{1-p(X_i)}1\{q_{0,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant q_{0,\tau}\})\right) \\ &\equiv \frac{1}{n}\sum_{i=1}^{n} m(Z_i, q_\tau, p) \\ M(q_\tau, p) &\equiv E[m(Z_i, q_\tau, p)]. \end{aligned}$$

Chen, Linton, and Van Keilegom (2003) show that their Theorem 1 is implied by the followig conditions:

1.1 $\|M_n(\hat{q}_\tau, \hat{p})\| \leqslant \inf_{q_\tau \in \mathcal{Q}} \|M_n(q_\tau, \hat{p})\| + o_p(1)$.

This condition follows for given $\hat{p}$ directly by Theorem 1 of Powell (1986).

1.2 $\forall \delta > 0 \; \exists \; \epsilon(\delta) > 0$ such that $\inf_{\|q_\tau - q^*_\tau\| > \delta} \|M_n(q_\tau, p^*)\| \geqslant \epsilon(\delta) > 0$, where q^*_τ and p^* denote the true values of q_τ and p.

Again, for given p^*, this follows by Theorem 1 of Powell (1986).

1.3 $M(q_\tau, p)$ is continuous in p at p^* uniformly for all $q_\tau \in \mathcal{Q} \times \mathcal{Q}$.
This condition follows directly by the fact that p enters $M_n(q_\tau, p)$ as multiplicative factors $1/p$ and $1/(1-p)$.

1.4 $\|p - p^*\| = o_P(1)$.
Consistency of the estimator of the propensity score follows by Theorem 2 of Horowitz and Mammen (2004).

1.5′ $\sup_{q_\tau \in \mathcal{Q}\times\mathcal{Q}, \|p-p^*\| \leqslant \delta_n} \|M_n(q_\tau, p) - M(q_\tau, p)\| = o_P(1)$, where $\delta_n = o(1)$.
This condition will be fulfilled if $\{m(Z_i, q_\tau, p) | q_\tau \in \mathcal{Q}\times\mathcal{Q}, p \in \mathcal{P}\}$ is a Glivenko-Cantelli class, where $\mathcal{P}$ is the set of infinite dimensional parameters, i.e., the set of propensity scores. By the preservation result for Glivenko-Cantelli classes in Theorem 3 of van der Vaart and Wellner (2000)[1], it suffices to show that $p(X)$ and the censored quantile regression objective function form Glivenko-Cantelli classes, because both terms are linked continuously.

The propensity score belongs to the set of monotone functions which is a Glivenko-Cantelli class by Theorem 2.4.1 in connection with Theorem 2.7.5 of van der Vaart and Wellner (1996). The objective function of the censored quantile regression may be rewritten as product of indicator functions, which form classes with finite covering numbers, see Example 19.6 of van der Vaart (1998) or Example 2.4.2 of van der Vaart and Wellner (1996). The desired result follows by Theorem 3 of van der Vaart and Wellner (2000).

This shows consistency of $\hat{\Delta}_\tau$. Consistency of $\hat{\Delta}_{\tau|D=1}$ follows similarly.

Next, the asymptotic distributions of $\hat{\Delta}_\tau$ and $\hat{\Delta}_{\tau|D=1}$ are considered. Let $\mathcal{Q}_\delta \equiv \{q_\tau \in \mathcal{Q}\times\mathcal{Q} | \; \|q_\tau - q_\tau^*\| \leqslant \delta\}$ and $\mathcal{P}_\delta \equiv \{p \in \mathcal{P} | \; \|p - p^*\| \leqslant \delta\}$ with $\delta > 0$. By Theorem 2 of Chen, Linton, and Van Keilegom (2003), asymptotic normality follows by the following conditions[2]:

2.1 $\|M_n(\hat{q}_\tau, \hat{p})\| = \inf_{q_\tau \in \mathcal{Q}_\delta} \|M_n(q_\tau, \hat{p})\| + o_P(n^{-1/2})$.
This condition follows directly by results of Powell (1984, 1986).

2.2 i. Let

$$\Gamma_1 \equiv \Gamma_1(q_\tau, p^*) = \frac{\partial M(q_\tau, p^*)}{\partial q_\tau} = \frac{\partial E[m(Z_i, q_\tau, p^*)]}{\partial q_\tau}.$$

[1] Theorem 3 of van der Vaart and Wellner (2000) states that if $\mathcal{F}_1, \ldots, \mathcal{F}_k$ are Glivenko-Cantelli classes and $\varphi(f_1, \ldots, f_k)$, $f_1 \in \mathcal{F}_1, \ldots, f_k \in \mathcal{F}_k$, is a continuous function from $\mathbb{R}^k \to \mathbb{R}$, then the set $\{\varphi(f_1, \ldots, f_k) | f_1 \in \mathcal{F}_1, \ldots, f_k \in \mathcal{F}_k\}$ is also a Glivenko-Cantelli class.

[2] Chen, Linton, and Van Keilegom (2002) give an extensive discussion of an example, which shows how to verify the conditions of their approach.

Assume that $\Gamma_1(q_\tau, p^*)$ exists for $q_\tau \in \mathcal{Q}_\delta$ and is continuous at q_τ^*.
To show that these assumptions hold in the present application, express $E[m(Z_i, q_\tau, p^*)]$ as

$$\begin{aligned}
&E[m(Z_i, q_\tau, p^*)] \\
&= E\left[\frac{D_i}{p^*(X_i)} 1\{q_{1,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant q_{1,\tau}\}) \right. \\
&\qquad \left. - \frac{1-D_i}{1-p^*(X_i)} 1\{q_{0,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant q_{0,\tau}\})\right] \\
&= E\left[E\left[\frac{D_i}{p^*(X_i)} 1\{q_{1,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant q_{1,\tau}\}) \right.\right. \\
&\qquad \left.\left. - \frac{1-D_i}{1-p^*(X_i)} 1\{q_{0,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant q_{0,\tau}\}) \right| Z\right]\right] \\
&= E\left[\frac{D_i}{p^*(X_i)} (1 - F_C(q_{1,\tau}))(\tau - F_{\tilde{T}|X}(q_{1,\tau})) \right. \\
&\qquad \left. - \frac{1-D_i}{1-p^*(X_i)} (1 - F_C(q_{0,\tau}))(\tau - F_{\tilde{T}|X}(q_{0,\tau}))\right].
\end{aligned}$$

The derivative Γ_1 is therefore given as

$$\frac{\partial M(q_\tau, p^*)}{\partial q_\tau} = \begin{pmatrix} -E\left[\frac{D_i}{p^*(X_i)}(f_C(q_{1,\tau})(\tau - F_{\tilde{T}|X}(q_{1,\tau})) \right. \\ \left. +(1 - F_C(q_{1,\tau}))f_{\tilde{T}|X}(q_{1,\tau}))\right] \\ E\left[\frac{1-D_i}{1-p^*(X_i)}(f_C(q_{0,\tau})(\tau - F_{\tilde{T}|X}(q_{0,\tau})) \right. \\ \left. +(1 - F_C(q_{0,\tau}))f_{\tilde{T}|X}(q_{0,\tau}))\right] \end{pmatrix}.$$

The derivative of the objective of $\Delta_{\tau|D=1}$ is given in section 2.3. Assuming sufficiently smooth distributions of the censoring and outcome variables, the condition is shown to hold in the present application.

ii. Γ_1 is assumed to be of full column rank.
This is obvious, as the co-domain of $M(q_\tau, p)$ is one-dimensional.

2.3 Define the functional derivative of $M(q_\tau, p)$ with respect to $p \in \mathcal{P}$ as

$$\Gamma_2(q_\tau, p)(\tilde{p} - p) = \lim_{\theta \to 0} \frac{M(q_\tau, p + \theta(\tilde{p} - p)) - M(q_\tau, p)}{\theta}.$$

For all $q_\tau \in \mathcal{Q}_\delta$, it is assumed that the limit in all directions $(\tilde{p} - p) \in \mathcal{P}$ exists. For all $(q_\tau, p) \in \mathcal{Q}_\delta \times \mathcal{P}_\delta$, it holds furthermore that:

i. $\|M(q_\tau, p) - M(q_\tau, p^*) - \Gamma_2(q_\tau, p^*)(p - p^*)\| \leqslant k\|p - p^*\|^2$ for $k \geqslant 0$.

Rewrite the derivative as[3]

$$\begin{aligned}
&\Gamma_2(q_\tau, p^*)(p - p^*) \\
&= \lim_{\theta \to 0} \frac{E[m(Z_i, q_\tau, p^* + \theta(p - p^*)) - m(Z_i, q_\tau, p^*)]}{\theta} \\
&= \lim_{\theta \to 0} \frac{1}{\theta} E\Bigg[\left(\frac{D_i}{p^*(X_i) + \theta(p(X_i) - p^*(X_i))} - \frac{D_i}{p^*(X_i)}\right) \\
&\quad \times (1 - F_C(q_{1,\tau}))(\tau - F_{Y|X}(q_{1,\tau})) \\
&\quad - \left(\frac{1 - D_i}{1 - (p^*(X_i) + \theta(p(X_i) - p^*(X_i)))} - \frac{1 - D_i}{1 - p^*(X_i)}\right) \\
&\quad \times (1 - F_C(q_{0,\tau}))(\tau - F_{Y|X}(q_{0,\tau})) \\
&= E\Bigg[\frac{\partial}{\partial\theta}\Bigg(\frac{D_i}{p^*(X_i) + \theta(p(X_i) - p^*(X_i))} \\
&\qquad \times (1 - F_C(q_{1,\tau}))(\tau - F_{Y|X}(q_{1,\tau})) \\
&\qquad - \frac{1 - D_i}{1 - (p^*(X_i) + \theta(p(X_i) - p^*(X_i)))} \\
&\qquad \times (1 - F_C(q_{0,\tau}))(\tau - F_{Y|X}(q_{0,\tau}))\Bigg)\Bigg|_{\theta=0}\Bigg] \\
&= -E\Bigg[\frac{D_i(p(X_i) - p^*(X_i))}{(p^*(X_i))^2}(1 - F_C(q_{1,\tau}))(\tau - F_{Y|X}(q_{1,\tau})) \\
&\qquad - \frac{(1 - D_i)(p(X_i) - p^*(X_i))}{(1 - p^*(X_i))^2} \\
&\qquad \times (1 - F_C(q_{0,\tau}))(\tau - F_{Y|X}(q_{0,\tau}))\Bigg]
\end{aligned}$$

This shows the claimed existence of the derivative in all directions under the assumptions of the theorem. To show condition 2.3.i, consider a Taylor expansion of $M(q_\tau, p)$ around $p^*(X_i)$:

$$\begin{aligned}
M(q_\tau, p) &= M(q_\tau, p^*) + \Gamma_2(q_\tau, p^*)(p(X_i) - p^*(X_i)) \\
&\quad + M^{(2)}(q_\tau, p^*)(p(X_i) - p^*(X_i))^2 + \tilde{R}(X_i),
\end{aligned}$$

where $M^{(2)}(q_\tau, p^*)$ is the second derivative of $M(q_\tau, p)$ with respect to $p(X_i)$ and $\tilde{R}(X_i)$ is the remainder of the expansion. Note that $\|M(q_\tau, p) - M(q_\tau, p^*) - \Gamma_2(q_\tau, p^*)(p - p^*)\|$ is equal to the quadratic term and the remainder. An inspection of $\Gamma_2(q, p^*)$ shows that its derivative (i.e., the second derivative of $M(q, p)$ with respect to $p(X_i)$ at $p^*(X_i)$) is

[3]See the example in Chen, Linton, and Van Keilegom (2002) for similar calculations.

bounded. Therefore, $\|M(q,p)-M(q,p^*)-\Gamma_2(q,p^*)(p-p^*)\|$ is bounded by $K\|p-p^*\|^2$ for a suitable constant K (see the proof of Proposition 1 of Chen, Linton, and Van Keilegom (2002) for the same line of argument for a different estimator.) That shows that condition 2.3.i holds for the present estimator. It can be shown similarly that this condition also holds for $\Delta_{\tau|D=1}$.

ii. $\|\Gamma_2(q_\tau,p^*)(p-p^*)-\Gamma_2(q_\tau^*,p^*)(p-p^*)\| = o(1)$.
Rewrite this condition as follows:

$$\begin{aligned}
&\|\Gamma_2(q,p^*)(p-p^*)-\Gamma_2(q^*,p^*)(p-p^*)\| \\
= &\left\| E\left[\frac{D_i(p(X_i)-p^*(X_i))}{(p^*(X_i))^2}\right.\right. \\
&\quad\times((1-F_C(q_{1,\tau}^*))(\tau-F_{Y|X}(q_{1,\tau}^*)) \\
&\quad-((1-F_C(q_{1,\tau}))(\tau-F_{Y|X}(q_{1,\tau}))) \\
&\quad+\frac{(1-D_i)(p(X_i)-p^*(X_i))}{(1-p^*(X_i))^2} \\
&\quad\times((1-F_C(q_{0,\tau}^*))(\tau-F_{Y|X}(q_{0,\tau}^*)) \\
&\quad\left.\left.-(1-F_C(q_{0,\tau}))(\tau-F_{Y|X}(q_{0,\tau})))\right]\right\|.
\end{aligned}$$

As $(1-F_C(q^*))(\tau-F_{Y|X}(q^*))-(1-F_C(q))(\tau-F_{Y|X}(q))$ is bounded and $p-p^*$ is $o_P(1)$, condition 2.3.ii holds.

2.4 $\hat{p}\in\mathcal{P}$ with probability one for $n\to\infty$ and $\|\hat{p}-p^*\| = o_P(n^{-1/4})$.
The first part follows by results of Horowitz and Mammen (2004) and assumption A9, the second solely by results of Horowitz and Mammen (2004).

2.5′ For $\delta_n = o_P(1)$,

$$\sup_{\|q-q^*\|\leqslant\delta_n,\|p-p^*\|\leqslant\delta_n} \|M_n(q,p)-M(q,p)-M_n(q^*,p^*)\| = o_P(n^{-1/2}).$$

Assume $m(z,q_\tau,p) = m_c(z,q_\tau,p)+m_{lc}(z,q_\tau,p)$. Theorem 3 of Chen, Linton, and Van Keilegom (2003) shows that condition 2.5′ is implied by their conditions 3.1 – 3.3 on $m_c(z,q_\tau,p)$ and $m_{lc}(z,q_\tau,p)$:

3.1 $m_c(z,q_\tau,p)$ is Hölder continuous with respect to q_τ and p, i.e.

$$|m_c(z,q_\tau,p)-m_c(z,q_\tau',p')| \leqslant b_j(z)\left(\|q_\tau-q_\tau'\|^{s_1}+\|p-p'\|^{s_2}\right).$$

As $m_c(z,q_\tau,p)=0$ for all $q_\tau\in\mathcal{Q}\times\mathcal{Q}$ and $p\in\mathcal{P}$ in the present application, this condition is trivially satisfied.

3.2 $m_{lc}(z, q_\tau, p)$ is locally uniformly $L_r(P)$ continuous with respect to q_τ and p for $r \geqslant 2$, i.e.,

$$E\left[\sup_{\|q_\tau - q'_\tau\| \leqslant \delta, \|p-p'\| \leqslant \delta} |m_{lc}(Z, q'_\tau, p') - m_{lc}(Z, q_\tau, p)|^r\right]^{1/r} \leqslant K\delta^{s_2},$$

where in the bounds of conditions 3.1 and 3.2 $b(z)$ is a measurable function such that $E[b(Z)]^r < \infty$, $s_1, s_2 \in (0,1]$, and $\delta = o(1)$.

In the following only the first term of $m_{lc}(z, q_\tau, p)$ will be considered, i.e., the term concerning $q_\tau(T_1)$. It can be shown by similar arguments that the condition holds for the quantile estimator of T_0, too. As both objective functions depend multiplicatively on D and on $1 - D$, respectively, the cross product term emerging by multiplying out the squared difference vanishes, as $D(1 - D) = 0$. Therefore, a separate analysis is possible. Abbreviate $q_\tau(T_1)$ in the following by q.

Rewrite the squared difference $m_{lc}(Z, q', p') - m_{lc}(Z, q, p)$ of the first term of $m_{lc}(Z, q, p)$ by adding and subtracting terms as

$$\begin{aligned}
&|m_{lc}(Z, q', p') - m_{lc}(Z, q, p)|^2 = \\
&= \left|\frac{D}{p(X)'}1\{q' < C\}(\tau - 1\{\tilde{T} \leqslant q'\}) \right.\\
&\qquad \left. - \frac{D}{p(X)}1\{q < C\}(\tau - 1\{\tilde{T} \leqslant q\})\right|^2 \\
&= \left|\frac{D}{p'(X)}1\{q' < C\}(\tau - 1\{\tilde{T} \leqslant q'\}) \right.\\
&\qquad - \frac{D}{p'(X)}1\{q < C\}(\tau - 1\{\tilde{T} \leqslant q'\}) \\
&\qquad + \frac{D}{p'(X)}1\{q < C\}(\tau - 1\{\tilde{T} \leqslant q'\}) \\
&\qquad - \frac{D}{p'(X)}1\{q < C\}(\tau - 1\{\tilde{T} \leqslant q\}) \\
&\qquad + \frac{D}{p'(X)}1\{q < C\}(\tau - 1\{\tilde{T} \leqslant q\}) \\
&\qquad \left. - \frac{D}{p(X)}1\{q < C\}(\tau - 1\{\tilde{T} \leqslant q\})\right|^2 \\
&= \left|\frac{D}{p'(X)}1\{\tilde{T} \leqslant q'\}(1\{q' < C\} - 1\{q < C\}) \right.\\
&\qquad + \frac{D}{p'(X)}1\{q < C\}(1\{\tilde{T} \leqslant q\}) - 1\{\tilde{T} \leqslant q'\})
\end{aligned}$$

$$
\begin{aligned}
& + \frac{D(p(X)-p'(X))}{p'(X)p(X)} 1\{q < C\}(\tau - 1\{\tilde{T} \leqslant q\}) \Big|^2 \\
\leqslant\ & D|(K_1(1\{q' < C\} - 1\{q < C\}) \\
& +K_2(1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}) + K_3(p(X) - p'(X))|^2 \\
\leqslant\ & K_4 D(|1\{q' < C\} - 1\{q < C\}| \\
& +|1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}| + |p(X) - p'(X)|)^2 \\
=\ & K_4 D(|1\{q' < C\} - 1\{q < C\}|^2 \\
& +|1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}|^2 + |p(X) - p'(X)|^2 \\
& +2|1\{q' < C\} - 1\{q < C\}||1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}| \\
& +2|1\{q' < C\} - 1\{q < C\}||p(X) - p'(X)| \\
& +2|1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}||p(X) - p'(X)|) \\
\leqslant\ & K_4 D(|1\{q' < C\} - 1\{q < C\}|^2 \\
& +|1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}|^2 + |p(X) - p'(X)|^2) \\
& +K_5 D(|1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}| \\
& +|1\{q' < C\} - 1\{q < C\}| + |p(X) - p'(X)|)
\end{aligned}
$$

Condition 3.2 will be fulfilled, if $E[|1\{q' < C\} - 1\{q < C\}|]$, $E[|1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}|]$, and $E[|p(X) - p'(X)|]$ are bounded by $K\delta$. This will be shown similar to examples 1 and 2 of Chen, Linton, and Van Keilegom (2003). First, note that

$$
\begin{aligned}
& q - \delta < q < q + \delta \\
& \quad \Rightarrow \quad 1\{q - \delta < C\} \geqslant 1\{q < C\} \geqslant 1\{q + \delta < C\}, \\
& q - \delta < q' < q + \delta \\
& \quad \Rightarrow \quad 1\{q - \delta < C\} \geqslant 1\{q' < C\} \geqslant 1\{q + \delta < C\}.
\end{aligned}
$$

The second line follows by $\|q - q'\| \leqslant \delta \leqslant 1$, which implies $q - \delta \leqslant q' \leqslant q + \delta$. Furthermore, as $1\{q' < C\} \leqslant 1\{q - \delta < C\}$ and $1\{q < C\} \geqslant 1\{q + \delta < C\} \Leftrightarrow -1\{q < C\} \leqslant -1\{q + \delta < C\}$, it follows that

$$
\begin{aligned}
|1\{q' < C\} - 1\{q < C\}| &\leqslant |1\{q - \delta < C\} - 1\{q + \delta < C\}| \\
&= 1\{q - \delta < C\} - 1\{q + \delta < C\},
\end{aligned}
$$

where the last line follows by the fact that $1\{q - \delta \leqslant C\} \geqslant 1\{q + \delta \leqslant C\}$. As this expression is equal to one or zero, the square can be dropped. The expectation of this expression is equal to the probability that C lies between $q - \delta$ and $q + \delta$:

$$
E[1\{q - \delta < C\} - 1\{q + \delta < C\}]
$$

$$\begin{aligned} &= 1 - F_C(q-\delta) - (1 - F_C(q+\delta)) \\ &= F_C(q+\delta) - F_C(q-\delta) \\ &= Pr(q-\delta < C < q+\delta). \end{aligned}$$

This expression is bounded by $K\delta$ if the distribution of C is Lipschitz continuous.

With these derivations, $L_r(P)$ continuity of $m_{lc}(Z,q,p)$ follows immediately. Taking the derivations above into account, the continuity condition for $E[|1\{q' < C\} - 1\{q < C\}|]$ reads as:

$$\begin{aligned} & E\left[\sup_{\|q-q'\|\leqslant\delta} |1\{q' < C\} - 1\{q < C\}|\right] \\ \leqslant\ & E\left[1\{q-\delta < C\} - 1\{q+\delta < C\}\right] \\ =\ & Pr(q-\delta < C < q+\delta) \\ =\ & F_C(q+\delta) - F_C(q-\delta). \end{aligned}$$

This expression is bounded by $K\delta$ for some $K > 0$, if the cumulative distribution function of C is assumed to be Lipschitz continuous. By the law of iterated expectations, similar arguments show $L_r(P)$ continuity of $E[|1\{\tilde{T} \leqslant q\} - 1\{\tilde{T} \leqslant q'\}|]$. The condition for $E[|p(X) - p'(X)|]$ follows directly. For the subpopulation of treated individuals, analoguous derivations are valid. Therefore, condition 3.2 of Theorem 3 of Chen, Linton, and Van Keilegom (2003) holds for the present application.

3.3 $\mathcal{Q}$ is a compact subset of $\mathbb{R}$ and $\mathcal{P}$ has a finite entropy integral.

Compactness of $\mathcal{Q}$ is assumed, the latter condition follows by the fact that the propensity score is a bounded monotone function (see example 2.6.21 of van der Vaart and Wellner (1996, p. 149)).

2.6 $\sqrt{n}(M_n(q^*,p^*) + \Gamma_2(q^*,p^*)(\hat{p}-p^*)) \xrightarrow{D} \mathcal{N}(0,V)$ for a finite matrix V.

For Δ_τ, $M_n(q^*,p^*) + \Gamma_2(q^*,p^*)(\hat{p}-p^*)$ is given by

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i}{p^*(X_i)}1\{q^*_{1,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant q^*_{1,\tau}\})\right. \\ & \qquad \left. - \frac{1-D_i}{1-p^*(X_i)}1\{q^*_{0,\tau} < C_i\}(\tau - 1\{\tilde{T}_i \leqslant q^*_{0,\tau}\})\right) \\ & - E\left[\frac{D(\hat{p}(X) - p^*(X))}{(p^*(X))^2}(1 - F_C(q^*_{1,\tau}))(\tau - F_{\tilde{T}|X}(q^*_{1,\tau}))\right. \end{aligned}$$

$$+\frac{(1-D)(\hat{p}(X)-p^*(X))}{(1-p^*(X))^2}(1-F_C(q^*_{0,\tau}))(\tau-F_{\tilde{T}|X}(q^*_{0,\tau}))\Bigg].$$

The difference of the estimator $\hat{p}(X)$ and the true propensity score $p^*(X)$ can be rewritten by adding and subtracting as

$$(\hat{p}(X)-E[\hat{p}(X)])+(E[\hat{p}(X)]-p^*(X)).$$

The second term is the bias of $\hat{p}(X)$. Following a similar argument in Example 1 of Chen, Linton, and Van Keilegom (2003), and using results of Horowitz and Mammen (2004), it follows that $(E[\hat{p}(X)]-p^*(X))$ is equal to a bounded function times a term of order $o_P(n^{-1/2})$. Therefore, $M_n(q^*,p^*)+\Gamma_2(q^*,p^*)(\hat{p}-p^*)$ can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\Bigg(\frac{D_i}{p^*(X_i)}1\{q^*_{1,\tau}<C_i\}(\tau-1\{\tilde{T}_i\leqslant q^*_{1,\tau}\})\\
&\qquad-\frac{1-D_i}{1-p^*(X_i)}1\{q^*_{0,\tau}<C_i\}(\tau-1\{\tilde{T}_i\leqslant q^*_{0,\tau}\})\Bigg)\\
&-E\Bigg[\frac{D(\hat{p}(X)-E[\hat{p}(X)])}{(p^*(X))^2}(1-F_C(q^*_{1,\tau}))(\tau-F_{\tilde{T}|X}(q^*_{1,\tau}))\\
&\qquad+\frac{(1-D)(\hat{p}(X)-E[\hat{p}(X)]}{(1-p^*(X))^2}(1-F_C(q^*_{0,\tau}))(\tau-F_{\tilde{T}|X}(q^*_{0,\tau}))\Bigg]\\
&\qquad+o_P(n^{-1/2})\\
\equiv&\quad\frac{1}{n}\sum_{i=1}^{n}\xi_{\tau,i}+o_P(n^{-1/2}).
\end{aligned}$$

An inspection of these terms shows zero mean and boundedness of $\xi_{\tau,i}$. Therefore, the expression converges in distribution to $\mathcal{N}(0,V)$, where $V=Var(\xi_{\tau,i})$. For $\Delta_{\tau|D=1}$, the condition is also satisfied (for the expression in this case, see section 2.3). This shows that condition 2.6 holds for the application of the present paper.

Under these conditions asymptotic normality follows by Theorem 2 of Chen, Linton, and Van Keilegom (2003). The asymptotic variances are given in section 2.3. □

Proof of Theorem 2: The theorem is proved by adding and subtracting some terms to $\hat{\Omega}_\tau-\Omega_\tau$ and showing that the resulting diffences are $o_P(1)$.

$$\hat{\Omega}_\tau-\Omega_\tau \;=\; \left(\hat{\Gamma}'_{1,\tau}W\hat{\Gamma}_{1,\tau}\right)^{-1}\hat{\Gamma}'_{1,\tau}W\hat{V}_\tau W\hat{\Gamma}_{1,\tau}\left(\hat{\Gamma}'_{1,\tau}W\hat{\Gamma}_{1,\tau}\right)^{-1}$$

$$
\begin{aligned}
&\quad - \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \Gamma'_{1,\tau} W V_\tau W \Gamma_{1,\tau} \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \\
&= \left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} \hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} \\
&\quad - \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} \\
&\quad + \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} \\
&\quad - \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \\
&\quad + \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \\
&\quad - \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \Gamma'_{1,\tau} W V_\tau W \Gamma_{1,\tau} \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \\
&= \left(\left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} - \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1}\right) \\
&\qquad \times \hat{\Gamma}'_{1,\tau} W \hat{V}_{1,\tau} W \hat{\Gamma}_{1,\tau} \left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} \qquad (6) \\
&\quad + \left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} \hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \\
&\qquad \times \left(\left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} - \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1}\right) \qquad (7) \\
&\quad + \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1} \left(\hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau}\right. \\
&\qquad \left. - \Gamma'_{1,\tau} W V_\tau W \Gamma_{1,\tau}\right) \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1}. \qquad (8)
\end{aligned}
$$

Consider eqs. (6) and (7) first. For these terms to be stochastically negligible, it has to be shown that the following difference is $o_P(1)$:

$$
\left(\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau}\right)^{-1} - \left(\Gamma'_{1,\tau} W \Gamma_{1,\tau}\right)^{-1}.
$$

As it follows from $A_n \xrightarrow{P} A$ that $A_n^{-1} \xrightarrow{P} A^{-1}$ (see Davidson (1994, p. 287), the difference in eqs. (6) and (7) is $o_P(1)$ if

$$
\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau} - \Gamma'_{1,\tau} W \Gamma_{1,\tau}
$$

is $o_P(1)$. By adding and subtracting, this is equivalent to

$$
\begin{aligned}
&\hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau} - \Gamma'_{1,\tau} W \Gamma_{1,\tau} \\
&\quad = \hat{\Gamma}'_{1,\tau} W \hat{\Gamma}_{1,\tau} - \Gamma'_{1,\tau} W \hat{\Gamma}_{1,\tau} + \Gamma'_{1,\tau} W \hat{\Gamma}_{1,\tau} - \Gamma'_{1,\tau} W \Gamma_{1,\tau} \\
&\quad = \left(\hat{\Gamma}_{1,\tau} - \Gamma_{1,\tau}\right)' W \hat{\Gamma}_{1,\tau} + \Gamma'_{1,\tau} W \left(\hat{\Gamma}_{1,\tau} - \Gamma_{1,\tau}\right).
\end{aligned}
$$

The difference $\hat{\Gamma}_{1,\tau} - \Gamma_{1,\tau}$ is given by:

$$\hat{\Gamma}_{1,\tau} - \Gamma_{1,\tau} = \begin{pmatrix} -\hat{E}\left[\frac{D_i}{\hat{p}(X_i)}(\hat{f}_C(\hat{q}_{1,\tau})(\tau - \hat{F}_{\tilde{T}|X}(\hat{q}_{1,\tau})) \right. \\ \left. +(1-\hat{F}_C(\hat{q}_{1,\tau}))\hat{f}_{\tilde{T}|X}(\hat{q}_{1,\tau}))\right] \\ -\hat{E}\left[\frac{1-D_i}{1-\hat{p}(X_i)}(\hat{f}_C(\hat{q}_{0,\tau})(\tau - \hat{F}_{\tilde{T}|X}(\hat{q}_{0,\tau})) \right. \\ \left. +(1-\hat{F}_C(\hat{q}_{0,\tau}))\hat{f}_{\tilde{T}|X}(\hat{q}_{0,\tau}))\right] \end{pmatrix} - \begin{pmatrix} -E\left[\frac{D}{p^*(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde{T}|X}(q^*_{1,\tau})) \right. \\ \left. +(1-F_C(q^*_{1,\tau}))f_{\tilde{T}|X}(q^*_{1,\tau}))\right] \\ -E\left[\frac{1-D}{1-p^*(X)}(f_C(q^*_{0,\tau})(\tau - F_{\tilde{T}|X}(q^*_{0,\tau})) \right. \\ \left. +(1-F_C(q^*_{0,\tau}))f_{\tilde{T}|X}(q^*_{0,\tau}))\right] \end{pmatrix}.$$

Only the first element of this vector will be considered in the following, as the second can be bounded similarly. Again, the difference is rewritten by adding and subtracting a number of terms:

$$\begin{aligned} & E\left[\frac{D}{p^*(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde{T}|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde{T}|X}(q^*_{1,\tau}))\right] \\ & -\hat{E}\left[\frac{D_i}{\hat{p}(X_i)}(\hat{f}_C(\hat{q}_{1,\tau})(\tau - \hat{F}_{\tilde{T}|X}(\hat{q}_{1,\tau})) + (1-\hat{F}_C(\hat{q}_{1,\tau}))\hat{f}_{\tilde{T}|X}(\hat{q}_{1,\tau}))\right] \\ = \; & E\left[\frac{D}{p^*(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde{T}|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde{T}|X}(q^*_{1,\tau}))\right] \\ & -\hat{E}\left[\frac{D}{p^*(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde{T}|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde{T}|X}(q^*_{1,\tau}))\right] \\ & +\hat{E}\left[\frac{D}{p^*(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde{T}|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde{T}|X}(q^*_{1,\tau}))\right] \\ & -\hat{E}\left[\frac{D_i}{\hat{p}(X_i)}(\hat{f}_C(\hat{q}_{1,\tau})(\tau - \hat{F}_{\tilde{T}|X}(\hat{q}_{1,\tau})) + (1-\hat{F}_C(\hat{q}_{1,\tau}))\hat{f}_{\tilde{T}|X}(\hat{q}_{1,\tau}))\right]. \end{aligned}$$

The first two equations of this expression are asymptotically equal to zero in probability by convergence of the sample mean to the expectation, i.e., because $E[A] = \hat{E}[A] + o_P(1)$. The difference of the last two will be stochastically negligible if the terms within the brackets converge to each other, which again can be shown by an add-and-subtract strategy:

$$\begin{aligned} & \frac{D}{p^*(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde{T}|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde{T}|X}(q^*_{1,\tau})) \\ & -\frac{D_i}{\hat{p}(X_i)}(\hat{f}_C(\hat{q}_{1,\tau})(\tau - \hat{F}_{\tilde{T}|X}(\hat{q}_{1,\tau})) + (1-\hat{F}_C(\hat{q}_{1,\tau}))\hat{f}_{\tilde{T}|X}(\hat{q}_{1,\tau})) \\ = \; & \frac{D}{p^*(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde{T}|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde{T}|X}(q^*_{1,\tau})) \end{aligned}$$

$$
\begin{aligned}
&-\frac{D}{\hat p(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde T|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde T|X}(q^*_{1,\tau}))\\
&+\frac{D}{\hat p(X)}(f_C(q^*_{1,\tau})(\tau - F_{\tilde T|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde T|X}(q^*_{1,\tau}))\\
&-\frac{D_i}{\hat p(X_i)}(\hat f_C(\hat q_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau})) + (1-\hat F_C(\hat q_{1,\tau}))\hat f_{\tilde T|X}(\hat q_{1,\tau})).
\end{aligned}
$$

The first two lines are bounded by

$$
K\left(\frac{D}{p^*(X)} - \frac{D}{\hat p(X)}\right) = K\frac{D(\hat p(X) - p^*(X))}{\hat p(X)} = o_P(1).
$$

To bound the second difference, consider:

$$
\begin{aligned}
&(f_C(q^*_{1,\tau})(\tau - F_{\tilde T|X}(q^*_{1,\tau})) + (1-F_C(q^*_{1,\tau}))f_{\tilde T|X}(q^*_{1,\tau}))\\
&-(\hat f_C(\hat q_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau})) + (1-\hat F_C(\hat q_{1,\tau}))\hat f_{\tilde T|X}(\hat q_{1,\tau}))\\
&\quad = \quad (f_C(q^*_{1,\tau})(\tau - F_{\tilde T|X}(q^*_{1,\tau})) - (\hat f_C(\hat q_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau}))\\
&\qquad\quad +(1-F_C(q^*_{1,\tau}))f_{\tilde T|X}(q^*_{1,\tau})) - (1-\hat F_C(\hat q_{1,\tau}))\hat f_{\tilde T|X}(\hat q_{1,\tau})).
\end{aligned}
$$

Only the first line will be considered in the following, as the second can be bounded similarly:

$$
\begin{aligned}
&f_C(q^*_{1,\tau})(\tau - F_{\tilde T|X}(q^*_{1,\tau})) - \hat f_C(\hat q_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau}))\\
&\quad = \quad f_C(q^*_{1,\tau})(\tau - F_{\tilde T|X}(q^*_{1,\tau})) - f_C(q^*_{1,\tau})(\tau - F_{\tilde T|X}(\hat q_{1,\tau}))\\
&\qquad\quad +f_C(q^*_{1,\tau})(\tau - F_{\tilde T|X}(\hat q_{1,\tau})) - f_C(q^*_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau}))\\
&\qquad\quad +f_C(q^*_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau})) - f_C(\hat q_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau}))\\
&\qquad\quad +f_C(\hat q_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau})) - \hat f_C(\hat q_{1,\tau})(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau}))\\
&\quad = \quad f_C(q^*_{1,\tau})(F_{\tilde T|X}(\hat q_{1,\tau}) - F_{\tilde T|X}(q^*_{1,\tau}))\\
&\qquad\quad +f_C(q^*_{1,\tau})(\hat F_{\tilde T|X}(\hat q_{1,\tau}) - F_{\tilde T|X}(\hat q_{1,\tau}))\\
&\qquad\quad +(f_C(q^*_{1,\tau}) - f_C(\hat q_{1,\tau}))(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau}))\\
&\qquad\quad +(f_C(\hat q_{1,\tau}) - \hat f_C(\hat q_{1,\tau}))(\tau - \hat F_{\tilde T|X}(\hat q_{1,\tau})).
\end{aligned}
$$

The first line converges to zero by consistency of $\hat q_{1,\tau}$ and if the distribution of $\tilde T$ is sufficiently smooth, the second by consistency of the estimator of the cumulative distribution function, the third by consistency of $\hat q_{1,\tau}$ and smoothness of the density of the censoring time, and the last by consistency of the kernel density estimators.

Finally, consider eq. (8). This will be asymptotically zero if the middle term is $o_P(1)$. By the same strategy as above, this part can be rewritten as

$$
\hat\Gamma'_{1,\tau}W\hat V_\tau W\hat\Gamma_{1,\tau} - \Gamma'_{1,\tau}WV_\tau W\Gamma_{1,\tau}
$$

$$
\begin{aligned}
&= \hat{\Gamma}'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} - \Gamma'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \\
&\quad + \Gamma'_{1,\tau} W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} - \Gamma'_{1,\tau} W V_\tau W \hat{\Gamma}_{1,\tau} \\
&\quad + \Gamma'_{1,\tau} W V_\tau W \hat{\Gamma}_{1,\tau} - \Gamma'_{1,\tau} W V_\tau W \Gamma_{1,\tau} \\
&= (\hat{\Gamma}_{1,\tau} - \Gamma_{1,\tau})' W \hat{V}_\tau W \hat{\Gamma}_{1,\tau} \\
&\quad + \Gamma'_{1,\tau} W (\hat{V}_\tau - V_\tau) W \hat{\Gamma}_{1,\tau} \\
&\quad + \Gamma'_{1,\tau} W V_\tau W (\hat{\Gamma}_{1,\tau} - \Gamma_{1,\tau}).
\end{aligned}
$$

Convergence of $\hat{\Gamma}_{1,\tau} - \Gamma_{1,\tau}$ to zero was shown just above. Rewrite the difference in the middle equation by inserting the definitions as

$$
\begin{aligned}
\hat{V}_\tau - V_\tau &= \frac{1}{n}\sum_{i=1}^n \xi_i^2 - \left(\frac{1}{n}\sum_{i=1}^n \xi_i\right)^2 - \left(E[\xi^2] - E[\xi]^2\right) \\
&= \frac{1}{n}\sum_{i=1}^n \xi_i^2 - E[\xi^2] - \left(\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right)^2 - E[\xi]^2\right).
\end{aligned}
$$

As $\xi_i^2 < \infty$, the first difference is $o_P(1)$ by a law of large numbers. The second difference converges also stochastically to zero by a law of large numbers together with the fact that for $A_n \xrightarrow{P} A$, it also holds that $g(A_n) \xrightarrow{P} g(A)$ for a measurable function $g(\cdot)$ which is continuous at the limit of the argument (see Theorem 18.10 of Davidson (1994, p. 286)). This shows consistency of the variance estimator of Δ_τ. Similarly, consistency of $\Omega_{\tau|D=1}$ can be shown, which completes the proof of Theorem 2. □

Proof of Theorem 3: The first claim of Theorem 3 follows, if the conditions of Theorem 4 of Chernozhukov and Hansen (2006) are met. These conditions state that $\sqrt{n}(\hat{\Delta}_{(\cdot)} - \Delta_{(\cdot)}) \Rightarrow b(\cdot)$ and $\sqrt{n}(\hat{a}(\cdot) - a(\cdot)) \Rightarrow d(\cdot)$, where $b(\cdot)$ and $d(\cdot)$ are mean zero Gaussian processes. This follows if $\Delta_{(\cdot)}$ and $a(\cdot)$ belong to Donsker classes.

Δ_τ consists of the reciprocal of the propensity score and a term with an indicator function. The first term form a Donsker class by Examples 2.6.21 (p. 149) and 2.10.9 (p. 192) of van der Vaart and Wellner (1996). Similarly, the second term is a Donsker class. By Theorem 2.10.6 of van der Vaart and Wellner (1996, p. 192), the product of both terms is also a Donsker class. This shows the first part of the condition of Theorem 4 of Chernozhukov and Hansen (2006) holds in the present application. The second part (i.e., convergence of $a(\cdot)$) does not differ from Chernozhukov and Hansen (2006). For the case $a(\cdot) = \Delta_{(\cdot)|D=1}$, the result follows by the Donsker property of $\Delta_{\tau|D=1}$. Therefore, the first claim of Theorem 3 is shown and convergence of S_n to $f(\upsilon(\cdot))$ holds.

Now, convergence of the bootstrap test statistics is shown. By the Donsker property of the test statistic, the bootstapped test statistic converges to the true test statistic (See van der Vaart (1998, Theorem 23.7, p. 333; see also van der Vaart and Wellner (1996, sec. 3.6) and Kosorok (2006, sec. 10). Now, the claims of the theorem can be shown as in the proof of Theorem 4 of Chernozhukov and Hansen (2006). Therefore, Theorem 3 is proven. □

References

ABADIE, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292.

ABBRING, J. H. (2003): "Dynamic Econometric Program Evaluation," IZA Discussion Paper No. 804.

——— (2006): "The Event-History Approach to Program Evaluation," Tinbergen Institute Discussion Paper 2006-057/3.

——— (2007): "Mixed Hitting–Time Models," Cemmap Working Paper CWP 15/07.

ABBRING, J. H., AND G. VAN DEN BERG (2003a): "The Nonparametric Identification of Treatment Effects in Duration Models," *Econometrica*, 71, 1491–1517.

——— (2003b): "A Simple Procedure for the Evaluation of Treatment Effects on Duration Variables," IFAU Working Paper 2003:19.

——— (2004): "Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data," *Empirical Economics*, 29, 5–20.

——— (2005): "Social Experiments and Instrumental Variables with Duration Outcomes," Tinbergen Institute Discussion Paper 2005-047/3.

CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2002): "Estimation of Semiparametric Models when the Criterion Function is not Smooth," Discussion Paper 0204, Institut de Statistique, Université Catholique de Louvain.

——— (2003): "Estimation of Semiparametric Models when the Criterion Function is not Smooth," *Econometrica*, 71, 1591–1608.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

CHERNOZHUKOV, V., AND I. FERNANDEZ-VAL (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhya*, 67, 253–276.

Chernozhukov, V., and C. Hansen (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491–525.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press, Oxford.

Efron, B., and I. M. Johnstone (1990): "Fisher's Information in Terms of the Hazard Rate," *Annals of Statistics*, 18, 38–62.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276.

Fitzenberger, B., and R. A. Wilke (2005): "Using Quantile Regression for Duration Analysis," ZEW Discussion Paper 05-65, ZEW Mannheim.

Frölich, M. (2005): "Matching Estimators and Optimal Bandwidth Choice," *Statistics and Computing*, 15, 197–215.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

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

Horowitz, J. H., and E. Mammen (2004): "Nonparametric Estimation of an Additive Model with a Link Function," *Annals of Statistics*, 32, 2412–2443.

Horowitz, J. H., and G. R. Neumann (1987): "Semiparametric Estimation of Employment Duration Models," *Econometric Reviews*, 6, 5–84 and 257–270, with discussion.

Ichimura, H., and O. Linton (2005): "Asymptotic Expansions for Some Semiparametric Program Evaluation Estimators," in *Identification and Inference for Econometric Models*, ed. by D. W. K. Andrews, and J. H. Stock, pp. 149–170. Cambridge University Press, Cambridge.

Imbens, G., W. Newey, and G. Ridder (2005): "Mean-square-error Calculations for Average Treatment Effects," IEPR Working Paper 05.34.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4–29.

Kiefer, N. M. (1988): "Economic Duration Data and Hazard Functions," *Journal of Economic Literature*, 26, 646–679.

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

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

Koenker, R., and Y. Bilias (2001): "Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments," *Empirical Economics*, 26, 199–220.

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

Kosorok, M. R. (2006): "Introduction to Empirical Processes and Semiparametric Inference," draft, University of North Carolina, Chapel Hill.

Lechner, M., and R. Miquel (2001): "A Potential Outcome Approach to Dynamic Programme Evaluation - Part I: Identification," mimeo, University of St. Gallen.

Li, Q., and J. S. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton.

Powell, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

——— (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

Ritov, Y., and J. A. Wellner (1988): "Censoring, Martingales and the Cox Model," *Contemporary Mathematics*, 80, 191–219.

Smith, J. (2000): "A Critical Survey of Empirical Methods for Evaluating Active Labor Market Programs," *Swiss Journal of Economics and Statistics*, 136, 247–268.

van den Berg, G. (2001): "Duration Models: Specification, Identification and Multiple Durations," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5, pp. 3381–3460. Elsevier, Amsterdam.

van der Vaart, A. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

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

——— (2000): "Preservation Theorems for Glivenko-Cantelli and Uniform Glivenko-Cantelli Classes," in *High Dimensional Probability II*, ed. by E. Giné, D. M. Mason, and J. A. Wellner, pp. 115–133. Birkhäuser, Boston.