Three Essays on Semiparametric Econometric Evaluation Methods

Michael Maier

Dissertation

zur Erlangung des akademischen Grades
des Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)
am Fachbereich Wirtschaftswissenschaften
der Universität Konstanz

Vorgelegt von:
Michael Maier
Nietzschestr. 16
68165 Mannheim

1. Referent: Prof. Dr. Winfried Pohlmeier
2. Referent: Prof. Dr. Enno Mammen
Tag der mündlichen Prüfung: 22. April 2008

Konstanzer Online-Publikations-System (KOPS)
URL: http://www.ub.uni-konstanz.de/kops/volltexte/2008/5685/
URN: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-56856

Acknowledgements

I would like to thank Enno Mammen for supporting this thesis as second supervisor, for hints and comments on various parts of the thesis, and also for his lectures on empirical processes held at the University of Mannheim in winter term 2005/2006. I am also grateful to Michael Lechner for detailed comments on the three papers of the thesis. Helpful comments, hints and suggestions were also given by Alberto Abadie, Monika Kerekes, Anja Kuckulenz, François Laisney, Winfried Pohlmeier and Stefan Speckesser. The paper of section 3 also benefited from the comments of two anonymous journal referees.

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. I am also grateful to the Centre for European Economic Research for supporting the work on this thesis.

Finally, I would like to thank Monika Kerekes, Christiane Maaßen, Stefan Speckesser and Thomas Weber for support and encouragement during the last years.

Contents

Summary

This dissertation introduces three novel approaches for the econometric evaluation of heterogeneous treatment effects. The proposed methods consider the effects of a binary treatment on different characteristics of the outcome distribution.

Section 1 proposes an estimation method for various average treatment effects using weighted integrals of estimates of marginal treatment effects. Marginal treatment effects measure the effect of a treatment variable on the outcome for individuals who are indifferent between participating or not. By integration of suitable weighted marginal treatment effects, several average treatment effects can be calculated. The general assumptions of the method enable the use of various nonparametric estimators for the weights and the marginal treatment effects.

Section 2 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. Pros and cons of the method compared to other approaches are discussed.

Section 3 describes methods to test for distributional treatment effects under the conditional independence assumption. The differences between latent outcome distributions are judged by testing hypotheses of distributional equality and stochastic dominance. Furthermore, semiparametric efficient versions of the test statistics are given. The latter test statistics are double robust, i.e., they are consistent under misspecification of either the outcome equation or the propensity score. Consistent bootstrap procedures for deriving critical values of all tests are proposed.

Zusammenfassung

Die vorliegende Dissertation enthält drei neue Verfahren zur Evaluation heterogener Maßnahmeeffekte. Die vorgeschlagenen Methoden untersuchen die Wirkungen binärer Maßnahmen auf unterschiedliche Charakteristika der Verteilung der Ergebnisvariablen.

Abschnitt 1 schlägt eine Schätzmethode für mittlere Maßnahmeeffekte vor, die auf gewichteten Integralen von Schätzern marginaler Maßnahmeeffekte beruht. Marginale Maßnahmeeffekte messen die Wirkung einer Maßnahme auf die Ergebnisvariable von Individuen, die zwischen Teilnahme und Nichtteilnahme indifferent sind. Durch Integration geeignet gewichteter marginaler Maßnahmeeffekte können verschiedene mittlere Maßnahmeeffekte berechnet werden. Die allgemeinen Annahmen der Methode ermöglichen die Verwendung unterschiedlicher nichtparametrischer Schätzer der Gewichte und der marginalen Maßnahmeeffekte.

Abschnitt 2 beschreibt eine Methode zur Schätzung von Quantilsmaßnahmeeffekten binärer Maßnahmen für zensierte Verweildauern. Die untersuchten Effekte sind Differenzen von Quantilen der kontrafaktischen Ergebnisvariablen in beiden Zuständen der Maßnahmevariablen. Die Identifikation basiert auf der konditionalen Unabhängigkeitsannahme. Schätzer für die gesamte Population sowie für die Gruppe der Teilnehmer werden vorgeschlagen, Konsistenz, asymptotische Normalverteilung und Konsistenz der Varianzschätzer wird gezeigt. Eine einfache Umformung der Effekte wird hergeleitet, die eine einfache Interpretation der Ergebnisse ermöglicht. Ein Testverfahren für Hypothesen bezüglich des gesamten Prozesses der Quantilsmaßnahmeeffekte wird beschrieben. Die Vor- und Nachteile der Methode werden mit anderen Ansätzen verglichen.

Abschnitt 3 beschreibt Methoden zum Test von Verteilungsmaßnahmeeffekten unter der konditionalen Unabhängigkeitsannahme. Unterschiede zwischen den Verteilungen der latenten Ergebnisvariablen werden mit Hypothesentests hinsichtlich Gleichheit der Verteilungen und stochastischer Dominanz beurteilt. Darüber hinaus werden semiparametrisch effiziente Teststatistiken beschrieben. Diese Teststatistiken sind doppelt robust, d.h. sie bleiben konsistent, falls entweder die Teilnahmeneigung (*propensity score*) oder der konditionale Erwartungswert der Ergebnisvariablen fehlspezifiziert sind. Zur Herleitung der kritischen Werte der Tests werden konsistente Bootstrap-Verfahren beschrieben.

1 Asymptotic Properties of Average Treatment Effect Estimators based on Weighted Integrals of Marginal Treatment Effects

This paper proposes an estimation method for various average treatment effects using weighted integrals of estimates of marginal treatment effects. Marginal treatment effects measure the effect of a treatment variable on the outcome for individuals who are indifferent between participating or not. By integration of suitable weighted marginal treatment effects, several average treatment effects can be calculated. The general assumptions of the method enable the use of various nonparametric estimators for the weights and the marginal treatment effects.

1.1 Introduction

A large strand of current research in statistics and econometrics is concerned with estimation of the effects of a treatment variable on an outcome when the effects may differ for each observational unit. In this case, some kind of average effect is estimated. Following Rubins potential outcome approach, this average is defined as the mean difference of the outcomes for groups defined by the possible states of the treatment variable (so-called counterfactual outcomes). For different groups of individuals, different average effects may occur. This is especially the case if the choice of the treatment depends in some way on the difference of the counterfactual outcomes. For introductions and general reviews of econometric evaluation methods, see Heckman (2001), Heckman and Smith (1998), or Heckman, LaLonde, and Smith (1999), for example.

In this situation, the effect on individuals who are indifferent between choosing the treatment or not is of particular interest. For example, if a policy is considered which is intended to increase participation in the program, the so-called Marginal Treatment Effect (MTE) measures the impact on the group at the margin of participation. This view of the evaluation problem in the context of heterogeneous effects was basically proposed by Björklund and Moffitt (1987). In the context of econometric evaluation methods, several approaches are based (at least implicitly) on this framework. These approaches use exclusion restrictions and can be distingished whether the instrumental variable (IV) is binary (or countably finite) or continuous.

Imbens and Angrist (1994) discuss identification and estimation of heterogeneous treatment effects when a binary (or at least discrete) instrumental variable is available. They show what is identified by IV methods if effects

are heterogeneous and discuss interpretation of the so-called Local Average Treatment Effect (LATE). The MTE can be interpreted as limit version of the LATE. Further discussions of LATE models are contained in Angrist, Imbens, and Rubin (1996), Angrist (2004), and Heckman (1997). The LATE model can be extended to identify other characteristics of the outcomes than averages. Imbens and Rubin (1997) and Abadie (2002) use the LATE approach to identify distributions of latent outcomes and to test global hypotheses on equality of distributions and stochastic dominance.

Some other LATE models also incorporate covariate information. Tan (2006b) reviews the IV model and proposes regression and reweighting methods. Abadie, Angrist, and Imbens (2002) consider a quantile treatment effect model, while Abadie (2003) proposes semiparametric estimators for LATE analysis of possibly nonlinear models. Frölich (2007) describes a nonparametric approach to LATE estimation in the presence of covariates. Hong and Nekipelov (2007) discuss semiparametric efficiency of LATE models based on nonlinear moment restrictions. Among other things, they discuss efficiency of the quantile treatment effect estimators of Abadie, Angrist, and Imbens (2002).

When instrumental variables are continuous, the theoretical interpretation of LATE models and their practical computation is unclear. By using a latent index model for the treatment participation decision, Heckman and Vytlacil (1999, 2000, 2001a, 2005) relate the probability of treatment participation, MTEs, and the various average treatment effects to each other. In particular, they show that average treatment effects may be written as weighted integrals of MTEs. Furthermore, they propose to estimate MTEs by so-called Local Instrumental Variables (LIV) estimators. None of their papers, however, contains a derivation of the asymptotic properties of treatment effect estimators based on LIV estimators. In an application of this framework, Carneiro, Heckman, and Vytlacil (2005) use the bootstrap to compute confidence intervals. In the present paper, asymptotic properties of treatment effect estimators based on weighted integrals of MTEs are derived, which have not been described in the literature as yet.

A comprehensive review of the framework of Heckman and Vytlacil is given by Heckman, Urzua, and Vytlacil (2006). A similar idea was already used by Angrist, Graddy, and Imbens (2000). Theoretical applications of the framework involve Heckman and Vytlacil (2001b, 2007), who consider policy changes and the impact on individuals at the margin of participation in some detail, and Vytlacil and Yildiz (2007), who investigate identification of treatment effects in weakly separable models. Finally, the connection between LATE and MTE models is analyzed by Vytlacil (2002, 2006).

The remainder of this paper is organized as follows: section 1.2 gives

a review of the estimation approach based on MTEs. Section 1.3 shows consistency and $\sqrt{n}$–normality of the estimators based on weighted integrals of MTEs. Only general assumptions on the LIV estimators are stated; the results hold for a broad range of nonparametric estimators of the propensity score and of the MTE itself. Section 1.4 concludes.

1.2 Marginal Treatment Effects

Heckman and Vytlacil (1999, 2000, 2001a, 2005) show that econometric evaluation estimators like the average treatment effect or the average treatment effect on the treated can be expressed as weighted averages of the marginal treatment effect. In the following, a short description of this approach is given before some of its asymptotic properties are derived in the next section.

The basic model of Heckman and Vytlacil considers a binary treatment variable, which is assumed to be driven by a latent index $\mu_D(Z)$, i.e., $D = 1\{\mu_D(Z) - U_D \geq 0\}$, where Z is a set of instrumental variables and U_D is an error term which can be assumed without loss of generality to be uniformly distributed on the $[0, 1]$ interval (see Heckman and Vytlacil (2001a, p. 5/6)). If further covariates X are available, Z may contain some or all elements of X.

Various treatment effects are defined and used in the literature; for example, the average treatment effect and restrictions thereof to the subpopulations of treated ($D = 1$) or untreated individuals ($D = 0$):

$$\begin{aligned} \Delta^{ATE}(x) &= E[Y_1 - Y_0|X = x] \\ \Delta^{TT}(x) &= E[Y_1 - Y_0|X = x, D = 1] \\ \Delta^{TUT}(x) &= E[Y_1 - Y_0|X = x, D = 0]. \end{aligned}$$

Here, all effects are conditional on X. To be precise, note that all effects are averaged over the instruments Z, i.e., $E[Y_1 - Y_0|X] = E[E[Y_1 - Y_0|X, Z]|X]$.

Furthermore, the local average treatment effect of Imbens and Angrist (1994) is given by

$$\Delta^{LATE}(x, z, z') = E[Y_1 - Y_0|X = x, D(z) - D(z') = 1],$$

where $D(z)$ is the value of the treatment indicator for some value z of the instrument Z. The local average treatment effect therefore gives the treatment effect for the subgroup of individuals which change their treatment participation choice due to a change of the instrument. It can be rewritten as

$$\Delta^{LATE}(x, z, z') = \frac{E[Y|X = x, Z = z] - E[Y|X = x, Z = z']}{P(z) - P(z')},$$

where the propensity score $P(z)$ is defined as the expectation of the treatment choice indicator given the instruments, i.e., $P(z) \equiv E[D|Z = z]$ (see Imbens and Angrist (1994)).

For $P(z) \rightarrow P(z')$, $\Delta^{LATE}(x)$ corresponds to the marginal treatment effect, which is defined as

$$\Delta^{MTE}(x, u_D) = E[Y_1 - Y_0|X = x, U_D = u_D].$$

This expression can be interpreted as the average treatment effect of an individual which is indifferent between participation and nonparticipation given U_D. The importance of the marginal treatment effect lies in the fact that all other treatment effects can be written as weighted averages of the marginal treatment effect:

$$\Delta^{TE}(x) = \int_0^1 \omega^{TE}(x, u_D)\Delta^{MTE}(x, u_D)du_D,$$

where TE stands for any average treatment effect mentioned above. The average treatment effect is the unweighted integral of the marginal treatment effects, i.e., $\omega^{ATE}(x, u_D) \equiv 1 \; \forall \; u_D \in [0, 1]$ and $\forall \; x \in \text{supp}(X)$. The weights for the treatment effects of the subgroup of participants and nonparticipants are given by:

$$\begin{aligned} \omega^{TT}(x, u_D) &= \frac{1}{E[P(Z)|X = x]} \int_{u_D}^1 f(p|X = x)dp \\ \omega^{TUT}(x, u_D) &= \frac{1}{1 - E[P(Z)|X = x]} \int_0^{u_D} f(p|X = x)dp. \end{aligned}$$

$E[P(Z)|X = x]$ is the conditional probability for $D = 1$, and $f_{P(Z)}(p|\; X = x)$ is the density of the propensity score $P(Z)$. See Heckman and Vytlacil (2001a, p. 11-13) for a different interpretation of the weights above in terms of survival functions.

The intuition of this approach is that the average effect is composed by summing up a range of LATE estimators. The LATE is the effect of the subgroup of compliers; the approach of integrating the MTE over the whole $[0, 1]$–interval can be thought of as dividing the whole population in different groups of compliers and summing all these local effects. This idea was already stated by Angrist and Imbens (1995) and Imbens and Angrist (1994) in the context of LATE estimation for instuments with discrete (i.e., countably finite) support. In this situation, the LATE is given as weighted sum of local effects for pairs of values of the instrument. Note that the integral expression is no expectation of the MTE, as no probability measure appears. There is,

however, a link to a particular expectation of the MTEs; this will be used and detailed in the proof of Theorem 1 below. The weight for the Local Average Treatment Effect for the change of the instrument Z from z to z' can be expressed as

$$\omega^{LATE}(u_D) = \frac{1\{P(z') \leq u_D \leq P(z')\}}{P(z) - P(z')}.$$

For analyzing two policies which affect participation but not the marginal treatment effect, the policy relevant treatment effect may be of interest (see Heckman and Vytlacil (2001b, 2005) for a further discussion):

$$\Delta^{PRTE}(x) = \int_0^1 \omega^{PRTE}(x, u_D)\Delta^{MTE}(x, u_D)du_D.$$

Here, the weight is defined to be

$$\omega^{PRTE}(x, u_D) = \frac{F_{P(Z')}(u_D|X = x) - F_{P(Z)}(u_D|X = x)}{E[P(Z)|X = x] - E[P(Z')|X = x]}.$$

$P(Z')$ and $P(Z)$ are the propensity scores under the different policies. Finally, it should be mentioned that also ordinary least squares and (common) instrumental variables estimators may be written as averages of marginal treatment effects (see Heckman and Vytlacil (2005, table IB)).

For empirical analyses, an estimator of the Marginal Treatment Effect is needed. Heckman and Vytlacil (2005) propose the Local Instrumental Variables estimator:

$$\Delta^{LIV}(x, p) = \frac{\partial E[Y|X = x, P(Z) = p]}{\partial p}.$$

That $\Delta^{LIV}(x, p)$ is an estimator of $\Delta^{MTE}(x, u_D)$ can be seen as follows (see Heckman (2005, footnote 89, p. 80)):

$$\begin{aligned}
\frac{\partial E[Y|X, P(Z)]}{\partial P(Z)} &= \frac{\partial}{\partial P(Z)} E[DY_1 + (1 - D)Y_0|X, P(Z)] \\
&= \frac{\partial}{\partial P(Z)} E[E[DY_1 + (1 - D)Y_0|X, D, P(Z)]|X, P(Z)] \\
&= \frac{\partial}{\partial P(Z)} E[P(Z)E[Y_1|X, D = 1, P(Z)] \\
&\qquad +(1 - P(Z))E[Y_0|X, D = 0, P(Z)]|X, P(Z)] \\
&= E[E[Y_1|X, D = 1, P(Z)] - E[Y_0|X, D = 0, P(Z)]|X, P(Z)] \\
&= E[Y_1 - Y_0|X, P(Z)].
\end{aligned}$$

This derivation is based on the law of iterated expectations. This gives pointwise estimates of $\Delta^{MTE}(x, u_D)$ for each value of $u_D \in [0, 1]$. Together with estimates of the weights $\omega^{TE}(x, u_D)$, treatment effects like $\Delta^{ATE}(x)$ or $\Delta^{TT}(x)$ can be estimated. By averaging over the empirical distribution of X, unconditional effects can be obtained.

1.3 Average Treatment Effects

As described in section 1.2, average treatment effects can be expressed by

$$\Delta^{TE} = \iint \omega^{TE}(x, p)\Delta^{MTE}(x, p)f_X(x)\ dp\ dx,$$

where TE stands for any average treatment effect of section 1.2. The expressions of the various treatment effects vary only by the weight $\omega^{TE}(x, p)$.

To implement these estimators, estimates of $\omega^{TE}(x, p)$ and $\Delta^{MTE}\ (x, p)$ are needed. To estimate these, the propensity score needs to be estimated in turn. To estimate the latter, a range of different methods is available. To prevent misspecification, a nonparametric approach is appropriate. Hirano, Imbens, and Ridder (2003) and Horowitz and Mammen (2004) propose nonparametric series estimators for binary dependent variables. The latter approach is preferable, as it circumvents a strong differentiability assumption of the former. Yu, Park, and Mammen (forthcoming) propose an estimator for generalized additive models based on the smooth backfitting algorithm of Mammen, Linton, and Nielsen (1999). For the latter approach as well as for the model of Horowitz and Mammen (2004), see also Li and Racine (2007, sec. 9.1.4 and 15.3.3).

To estimate the MTEs, estimates of the derivative of the expectation of the outcome variable given the propensity score and possibly some covariates are needed. An obvious candidate for obtaining these estimates are local quadratic regressions (see Fan and Gijbels (1996), Li and Racine (2007, sec. 2.4 and 2.5), Ruppert and Wand (1994), or Wand and Jones (1995), for example), as the derivative of the outcome with respect to the regressor is a direct estimation result, i.e., no numerical differentiation by some differencing scheme is needed. The parameter of interest is the coefficient of the local linear term, but to lower bias, an additional term should be included in the approximation (see Li and Racine (2007, pp. 87/88)), as the difference between the order of the parameter of interest and the overall order of the estimated polynomial should be odd.

A complication arises by the fact that the propensity score is unknown and therefore needs to be estimated. By using this term as a conditioning variable, a generated regressor problem arises, as the additional bias and

variability due to estimation has to be accounted for when deriving bias and variance of the nonparametric estimator of the derivative of the conditional expectation of the outcome. This problem is well known in the context of variance estimation of two step least squares models. For a review of this problem in a parametric framework, see Pagan (1984). Contributions to the generated regressor problem in nonparametric models involve Ahn (1997), Rilstone (1996), Li and Wooldridge (2002), and Sperlich (2007). Heckman, Ichimura, and Todd (1998) derive asymptotic properties of local polynomial regressions with generated regressors under general conditions for the estimated functions and the generated variables.

The weighting functions can be estimated by similar methods. In what follows, it is assumed that nonparametric estimators for the MTEs and the weighting functions of the various treatment effects are available. Assumptions which the estimates have to fulfil are stated below. Furthermore, only unconditional estimators are considered in the following.

Now, define the following estimator:

$$\hat{\Delta}^{TE} = \sum_{\{k\}} \omega^I(k)\hat{\omega}^{TE}(k)\hat{\Delta}^{MTE}(k),$$

where $\omega^I(k)$ is a deterministic weight for approximating an integral with the property that

$$\int f(x)dx - \sum_{\{k\}} \omega^I(k)f(k) = o(1).$$

The set $\{k\}$ is a suitable grid over the domain of integration. In the present case (i.e., with domain equal to [0,1]), a very simple (and somewhat crude) algorithm could use $\{k\} \equiv \{(j-1)/(K-1)|j = 1, \ldots, K\}$ and $\omega^I(k) \equiv K^{-1}$ for some number of grid points K. For details on numerical integration and further algorithms, see any textbook on basic numerical mathematics, for example Press et al. (2007, sec. 4) or Quarteroni, Sacco, and Saleri (2007, sec. 9).

Consider the following assumptions, which follow closely those of Frölich (2007):

A1 The iid sample is given by $\{Y_i, D_i, Z_i\}_{i=1}^n$.

A2 $f_{P(Z)}(p) > 0 \; \forall \; p \in [0, 1]$.

A3 $E[Y^2] < \infty$.

A4 The estimators of the MTE and the weights are asymptotically linear, i.e.,

$$\begin{aligned}\hat{\Delta}^{MTE}(p) - \Delta^{MTE}(p) &= \frac{1}{n}\sum_{i=1}^{n} \psi_1(Y_i, P(Z_i), p) + b_1(p) + R_1(p) \\ \hat{\omega}(p) - \omega(p) &= \frac{1}{n}\sum_{i=1}^{n} \psi_2(P(Z_i), p) + b_2(p) + R_2(p).\end{aligned}$$

where the various terms have the following properties ($l = 1, 2$):

(a) $E[\psi_l(Y_i, P(Z_i), P)|P = p] = 0$,
(b) $E[\psi_l(Y_i, P(Z_i), P(Z_j))^2] = o(n)$,
(c) $n^{-1}\sum_{i=1}^{n} b_l(P(Z_i)) = o_P(n^{-1/2})$,
(d) $n^{-1}\sum_{i=1}^{n} R_l(P(Z_i)) = o_P(n^{-1/2})$.

Assumption A1 is standard. A2 states that observations are available for every possible value of the propensity score (i.e., for each value of the interval $[0, 1]$). This is a strong assumption. The application and the dataset have to be suitable to generate estimates of the propensity score which cover the whole $[0, 1]$ interval. If the estimated values of the propensity score cover only a subset of $[0, 1]$, it is still possible to identify bounds of the treatment effects (see Heckman and Vytlacil (2001a, sec. 5.1)). A3 is a standard regularity assumption. The properties of the estimators of the MTEs and the weights stated in assumption A4 are fulfilled by a range of nonparametric estimators (see, for example, Frölich (2007) or Heckman, Ichimura, and Todd (1998)). For a general description of asymptotic linearity, see Newey and McFadden (1994, sec. 3). If it can be shown that a given estimator is asymptotic linear, consistency follows by the mean zero property of the influence function ψ as well as by the convergence to zero of the averaged bias and residual terms b_l and R_l. The asymptotic variance of the estimator is equal to $E[\psi^2]$. Here and in the proof below it is assumed that the bias terms already contain the terms resulting from the use of generated regressors; these will not be marked explicitely (that is, to be more precise, $\hat{P}(Z)$ should be written instead of $P(Z)$).

The asymptotic properties of the class of estimators defined in this section are given by the following theorem.

Theorem 1: *Under assumptions A1 – A4, the following holds for the average treatment effect estimators $\hat{\Delta}^{TE}$ described above:*

1. $\hat{\Delta}^{TE} - \Delta^{TE} = o_p(1)$

2. $\sqrt{n}\left(\hat{\Delta}^{TE} - \Delta^{TE}\right) \xrightarrow{D} \mathcal{N}(0, V)$.

The proof is given in appendix 1.5. A concrete expression of the variance V depends in part on the estimators used and is not detailed in the present paper.

1.4 Conclusion

This paper proposes a method for estimating various average treatment effects based on weighted integrals of marginal treatment effects. By using a simple integral approximation, a convenient estimator is obtained. Several estimates are needed in the first step; by using general assumptions for the proposed estimator, a range of nonparametric intermediate estimators are applicable. For the future, the semiparametric efficiency of the estimator remains to be considered and expressions for efficient influence functions have to be derived.

1.5 Proof of the Theorem

Proof of Theorem 1: The proof uses some results for U-statistics. Lee (1990), Koroljuk and Borovskich (1994), Serfling (1980, sec. 5), Shorack (2000, sec. 17.1), and van der Vaart (1998, sec. 12) provide introductions to U-statistics. Applications to nonparametric econometrics are given by Frölich (2007), Heckman, Ichimura, and Todd (1998), Powell, Stock, and Stoker (1989), and Zheng (1996), for example. To prove the theorem, it will be shown that $\hat{\Delta}^{TE}$ can be approximated by a suitable U-statistic. By the properties of this U-statistic, those of $\hat{\Delta}^{TE}$ stated in Theorem 1 will follow.

First, note that the treatment effect Δ^{TE} can be rewritten as

$$\int_0^1 \omega^{TE}(p)\Delta^{MTE}(p)dp \;=\; \int_0^1 \frac{\omega^{TE}(p)\Delta^{MTE}(p)}{f_{P(Z)}(p)} f_{P(Z)}(p)dp.$$

This equivalence has an intuitive interpretation. The treatment effect estimator on the lhs weights the integrand (i.e., $\omega^{TE}(p)\Delta^{MTE}(p)$) for all values of the domain of integration (which is equal to $[0, 1]$, the codomain of the propensity score) equally. In contrast, an usual expectation weights $\omega^{TE}(p)\Delta^{MTE}(p)$ by $f_{P(Z)}(p)$, the density of the propensity score; as long as the propensity score is not uniformly distributed (which is very unlikely in practice), the expectation of the weighted MTE will differ from the weighted intergrals of the definition of the treatment effects. If $\omega^{TE}(p)\Delta^{MTE}(p)$ is divided by $f_{P(Z)}(p)$, however, the expectation thereof is equal to the definition of the treatment

effects. This procedure downweights areas of the $[0,1]$-interval where the density of the propensity score is large. Consider the following estimator of the treatment effect:

$$\hat{\Delta}^{TE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\omega}^{TE}(P(Z_i))\hat{\Delta}^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}.$$

Here, $\hat{f}_{P(Z)}(p)$ is some nonparametric estimator of the density of the propensity score (for example, a kernel density estimator). Of course, for this expression to be valid for the present purpose, it is necessary that $P(Z)$ covers the whole $[0,1]$-interval, which is assured by assumption A2. In principle, this expression could be used as an estimator of the average treatment effects Δ^{TE}. This would require, however, an estimate of the density of the propensity score $P(Z)$, and, for a nonparametric estimator of it, the choice of one more smoothing parameter. Furthermore, the estimated density might be rather small (it is assumed merely to be positive), which could cause instability problems. Therefore, and for simplicity, it is suggested to base estimation directly on the integral approximation formula. For deriving the asymptotic properties, the usual expectation expression above is more useful, as one gets rid of the summation over $\{k\}$ and one obtains a standard U-statistic.

Rewrite now the difference between the approximation of the treatment effect estimator and its true value as follows:

$$\begin{aligned}
&\hat{\Delta}^{TE} - \Delta^{TE} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\omega}^{TE}(P(Z_i))\hat{\Delta}^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))} - E\left[\frac{\omega^{TE}(P(Z))\Delta^{MTE}(P(Z))}{f_{P(Z)}(P(Z))}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\omega}^{TE}(P(Z_i))\hat{\Delta}^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))} - \frac{\hat{\omega}^{TE}(P(Z_i))\Delta^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\right. \\
&\quad + \frac{\hat{\omega}^{TE}(P(Z_i))\Delta^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))} - \frac{\omega^{TE}(P(Z_i))\Delta^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))} \\
&\quad \left. + \frac{\omega^{TE}(P(Z_i))\Delta^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\right) - E\left[\frac{\omega^{TE}(P(Z))\Delta^{MTE}(P(Z))}{f_{P(Z)}(P(Z))}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\omega}^{TE}(P(Z_i))\left(\hat{\Delta}^{MTE}(P(Z_i)) - \Delta^{MTE}(P(Z_i))\right)}{\hat{f}_{P(Z)}(P(Z_i))}\right. \\
&\quad \left. \times\frac{\omega^{TE}(P(Z_i))}{\omega^{TE}(P(Z_i))} + \frac{\left(\hat{\omega}^{TE}(P(Z_i)) - \omega^{TE}(P(Z_i))\right)\Delta^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\right.
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\omega^{TE}(P(Z_i))\Delta^{MTE}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\Bigg)\frac{f_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\\
&-E\left[\frac{\omega^{TE}(P(Z))\Delta^{MTE}(P(Z))}{f_{P(Z)}(P(Z))}\right]\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\Bigg(\frac{\omega^{TE}(P(Z_i))\left(\hat{\Delta}^{MTE}(P(Z_i))-\Delta^{MTE}(P(Z_i))\right)}{f_{P(Z)}(P(Z_i))}\\
&\times\frac{\hat{\omega}^{TE}(P(Z_i))}{\omega^{TE}(P(Z_i))}+\frac{\left(\hat{\omega}^{TE}(P(Z_i))-\omega^{TE}(P(Z_i))\right)\Delta^{MTE}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\\
&+\frac{\omega^{TE}(P(Z_i))\Delta^{MTE}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\Bigg)\frac{f_{P(Z)}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\\
&-E\left[\frac{\omega^{TE}(P(Z))\Delta^{MTE}(P(Z))}{f_{P(Z)}(P(Z))}\right].
\end{aligned}
$$

These derivations and the following are very similar to those of Frölich (2007, sec. A.3). Consider first the term $f_{P(Z)}(P(Z_i))/\hat{f}_{P(Z)}(P(Z_i))$:

$$
\begin{aligned}
&\frac{f_{P(Z)}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\\
&\quad=\quad\frac{f_{P(Z)}(P(Z_i))-\hat{f}_{P(Z)}(P(Z_i))+\hat{f}_{P(Z)}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\\
&\quad=\quad\frac{f_{P(Z)}(P(Z_i))-\hat{f}_{P(Z)}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}+1\\
&\quad=\quad1-\frac{\hat{f}_{P(Z)}(P(Z_i))-f_{P(Z)}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\\
&\quad=\quad1-\frac{\hat{f}_{P(Z)}(P(Z_i))-f_{P(Z)}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\frac{f_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\\
&\quad=\quad1-\frac{\hat{f}_{P(Z)}(P(Z_i))-f_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\frac{f_{P(Z)}(P(Z_i))}{\hat{f}_{P(Z)}(P(Z_i))}\\
&\quad=\quad1-\frac{\hat{f}_{P(Z)}(P(Z_i))-f_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\left(\frac{\hat{f}_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\right)^{-1}\\
&\quad=\quad1-\frac{\hat{f}_{P(Z)}(P(Z_i))-f_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))}
\end{aligned}
$$

$$
\begin{aligned}
&\times \left(\frac{\hat{f}_{P(Z)}(P(Z_i)) - f_{P(Z)}(P(Z_i)) + f_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))} \right)^{-1} \\
= \; & 1 - \frac{\hat{f}_{P(Z)}(P(Z_i)) - f_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))} \\
&\times \left(1 + \frac{\hat{f}_{P(Z)}(P(Z_i)) - f_{P(Z)}(P(Z_i))}{f_{P(Z)}(P(Z_i))} \right)^{-1} \\
= \; & 1 - o_P(1)(1 + o_P(1))^{-1} \\
= \; & 1 - o_P(1)O_P(1) \\
= \; & 1 - o_P(1).
\end{aligned}
$$

For the term $\hat{\omega}^{TE}(P(Z_i))/\omega^{TE}(P(Z_i))$ it holds that:

$$
\begin{aligned}
\frac{\hat{\omega}^{TE}(P(Z_i))}{\omega^{TE}(P(Z_i))} &= \frac{\hat{\omega}^{TE}(P(Z_i)) - \omega^{TE}(P(Z_i)) + \omega^{TE}(P(Z_i))}{\omega^{TE}(P(Z_i))} \\
&= 1 + \frac{\hat{\omega}^{TE}(P(Z_i)) - \omega^{TE}(P(Z_i))}{\omega^{TE}(P(Z_i))} \\
&= 1 + o_P(1)
\end{aligned}
$$

The convergence of $\hat{f}_{P(Z)}(p)$ and $\hat{\omega}^{TE}(p)$ may be slower than that of parametric estimators; cross-product terms will be of higher order and can be neglected.

Using these results, $\hat{\Delta}^{TE} - \Delta^{TE}$ is asymptotically equivalent to:

$$
\begin{aligned}
&\hat{\Delta}^{TE} - \Delta^{TE} \\
&\simeq \frac{1}{n} \sum_{i=1}^{n} \frac{\omega^{TE}(P(Z_i)) \left(\hat{\Delta}^{MTE}(P(Z_i)) - \Delta^{MTE}(P(Z_i)) \right)}{f_{P(Z)}(P(Z_i))} \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \frac{\left(\hat{\omega}^{TE}(P(Z_i)) - \omega^{TE}(P(Z_i)) \right) \Delta^{MTE}(P(Z_i))}{f_{P(Z)}(P(Z_i))} \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \frac{\omega^{TE}(P(Z_i)) \Delta^{MTE}(P(Z_i))}{f_{P(Z)}(P(Z_i))} \\
&\quad - E\left[\frac{\omega^{TE}(P(Z)) \Delta^{MTE}(P(Z))}{f_{P(Z)}(P(Z))} \right].
\end{aligned}
$$

Here, $a \simeq b$ means that a equals b asymptotically up to higher order terms. The difference of the last two lines converges to zero by a law of large numbers. To analyze the first line, insert the influence function and the bias and

residual terms of $\hat{\Delta}^{MTE}(P(Z_i))$:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{\omega^{TE}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\left(\hat{\Delta}^{MTE}(P(Z_i))-\Delta^{MTE}(P(Z_i))\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\omega^{TE}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\\
&\quad\times\left(\frac{1}{n}\sum_{j=1}^{n}\psi_1(Y_j,P(Z_j),P(Z_i))+b_1(P(Z_i))+R_1(P(Z_i))\right)\\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{\omega^{TE}(P(Z_i))}{f_{P(Z)}(P(Z_i))}\psi_1(Y_j,P(Z_j),P(Z_i))\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\frac{\omega^{TE}(P(Z_i))}{f_{P(Z)}(P(Z_i))}(b_1(P(Z_i))+R_1(P(Z_i)))\\
&= \frac{1}{n}\sum_{i=1}^{n}E[\psi_1(Y_1,P(Z_1),P(Z))|Y_1=Y_i,P(Z_1)=P(Z_i)]\\
&\quad+o_P(n^{-1/2}).
\end{aligned}$$

For the projection result of the last equation to be applicable, $E[\|\psi_1\|]$ needs to be finite (see Serfling (1980, p. 188) or Powell, Stock, and Stoker (1989, Lemma 3.1), which follows by assumption A4(b). In the summation above, all observations with $i = j$ are dropped. As pointed out by Powell, Stock, and Stoker (1989, p. 1409, footnote 6), this does not alter the convergence properties of estimator, as long as the omitted terms satisfy the condition just stated. More formally, note that without this slight adaption, the estimator is a V-statistic (see Serfling (1980, sec. 5.1.2, 5.7.3, and also 6.3.2) or Shorack (2000, p. 457), for example). As V-statistics are asymptotically equivalent to U-statistics (Lemma 5.7.3 of Serfling (1980, p. 206)), the substitution of the V-statistic leads asymptotically to the same result.

An analogous result can be shown for the term containing $\hat{\omega}^{TE}(p) - \omega^{TE}(p)$. Therefore, $\hat{\Delta}^{TE}$ is asymptotically linear with influence function $\xi_i \equiv \tilde{\psi}_1(Y_i, P(Z_i), P(Z)) + \tilde{\psi}_2(Y_i, P(Z_i), P(Z))$, where

$$\begin{aligned}
\tilde{\psi}_1(Y_i,P(Z_i),P(Z)) &= \frac{\omega^{TE}(P(Z_i))}{f_{P(Z)}(P(Z_i))}E\left[\psi_1(Y_i,P(Z_i),P(Z))|Y_i,X_i\right]\\
\tilde{\psi}_2(Y_i,P(Z_i),P(Z)) &= \frac{\Delta^{TE}(P(Z_i))}{f_{P(Z)}(P(Z_i))}E\left[\psi_2(Y_i,P(Z_i),P(Z))|Y_i,X_i\right].
\end{aligned}$$

The variance is given by $V(\hat{\Delta}^{TE}) = E[\xi^2]$. From this, Theorem 1 follows. □

1.6 References

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

——— (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263.

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

Ahn, H. (1997): "Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors," *Econometric Theory*, 13, 3–31.

Angrist, J. (2004): "Treatment Effect Heterogeneity in Theory and Practice," *Economic Journal*, 114, C52–C83.

Angrist, J., K. Graddy, and G. Imbens (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67, 499–527.

Angrist, J., and G. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442.

Angrist, J., G. Imbens, and D. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–472, with discussion.

Björklund, A., and R. Moffitt (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69, 42–49.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2005): "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education," Mimeo, University of Chicago.

Fan, J., and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Frölich, M. (2007): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 139, 35–75.

Heckman, J. J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441–462, with addendum (vol. 33, 247) and discussion (vol. 34, 823–837).

——— (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 109, 673–748.

——— (2005): "The Scientific Model of Causality," *Sociological Methodology*, 35, 1–162, with discussion.

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Markets Programs," in *Handbook of Labor Economics*, ed. by A. Ashenfelter, and D. Card, vol. 3, pp. 1864–2097. Elsevier, Amsterdam.

Heckman, J. J., and J. Smith (1998): "Evaluating the Welfare State," in *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, ed. by S. Strøm, pp. 241–318. Cambridge University Press, Cambridge.

Heckman, J. J., S. Urzua, and E. J. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88, 389–432.

Heckman, J. J., and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734.

——— (2000): "The Relationship Between Treatment Parameters Within a Latent Variable Framework," *Economics Letters*, 66, 33–39.

——— (2001a): "Local Instrumental Variables," in *Nonlinear Statistical Modelling*, ed. by C. Hsiao, K. Morimune, and J. J. Powell, pp. 1–46. Cambridge University Press, Cambridge.

——— (2001b): "Policy-Relevant Treatment Effects," *American Economic Review, Papers and Proceedings*, 91, 107–111.

——— (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73, 669–738.

——— (2007): "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," mimeo, Columbia University.

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

Hong, H., and D. Nekipelov (2007): "Semiparametric Efficiency in Nonlinear LATE Models," mimeo, Stanford University.

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

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Imbens, G. W., and D. B. Rubin (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555–574.

Koroljuk, V. S., and Y. V. Borovskich (1994): *Theory of U-Statistics*. Kluwer, Dordrecht.

Lee, A. J. (1990): *U-Statistics: Theory and Practice*. Marcel Dekker, New York.

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

Li, Q., and J. M. Wooldridge (2002): "Semiparametric Estimation of Partially Linear Models for Dependent Data with Generated Regressors," *Econometric Theory*, 18, 625–645.

Mammen, E., O. Linton, and J. P. Nielsen (1999): "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm under Weak Conditions," *Annals of Statistics*, 27, 1443–1490.

Newey, W. K., and D. L. McFadden (1995): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2111–2245. Elsevier, Amsterdam.

Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2007): *Numerical Recipes*. Cambridge University Press, Cambridge, 3rd edn.

Quarteroni, A., R. Sacco, and F. Saleri (2007): *Numerical Mathematics*. Springer, Berlin, 2nd edn.

Rilstone, P. (1996): "Nonparametric Estimation of Models with Generated Regressors," *International Economic Review*, 37, 299–313.

Ruppert, D., and M. P. Wand (1994): "Multivariate Locally Weighted Least Squares Regression," *Annals of Statistics*, 22, 1346–1370.

Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shorack, G. R. (2000): *Probability for Statisticians*. Springer, New York.

Sperlich, S. (2007): "A Note on Nonparametric Estimation with Constructed Variables and Generated Regressors," mimeo, University of Göttingen.

Tan, Z. (2006): "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 101, 1607–1618.

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

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.

——— (2006): "Ordered Discrete-Choice Selection Models and Local Average treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results," *Review of Economics and Statistics*, 88, 578–581.

Vytlacil, E., and N. Yildiz (2007): "Dummy Endogenous Variables in Weakly Separable Models," *Econometrica*, 75, 757–779.

Wand, M. P., and M. C. Jones (1995): *Kernel Smoothing*. Chapman & Hall, London.

Yu, K., B. U. Park, and E. Mammen (forthcoming): "Smooth Backfitting in Generalized Additive Models," *Annals of Statistics*.

Zheng, J. X. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–289.

2 Quantile Treatment Effect Estimators for Durations under the Conditional Independence Assumption

This paper describes a method to estimate quantile treatment effects of a binary treatment variable on right 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. Pros and cons of the method compared to other approaches are discussed.

2.1 Introduction

This paper describes a method to estimate quantile treatment effects of a binary treatment variable on right 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 (2007a) by allowing censored outcomes.

Most approaches for analyzing durations are based on the concept of the hazard rate. Fewer approaches consider directly the durations as such. In the following, the literature of both approaches is reviewed, before describing the model of the present paper. Kiefer (1988) and van den Berg (2001) give economic examples of durations and review numerous methods for estimating the hazard rates in a variety of different contexts. Heckman and Singer (1984) consider economic examples of duration models, the influence of unobservable factors and nonparametric identification. Theoretical accounts involve Ridder (1990), who proposes a generalization which encompasses different duration models, and discusses its identification, and Hahn (1994), who analyzes semiparametric efficiency of the mixed proportional hazard model. Identification of a special class of semiparametric duration models is given by Chesher (2002). Ham and LaLonde (1996) discuss self-selection for duration models and analyze an experimental data set. 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 account to alternatives like basic (nondynamic) binary treatment effect models and panel data approaches. 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 very general framework where durations are defined by the crossing of a threshold by certain stochastic processes. Heckman and Navarro (2007) analyze identification of dynamic treatment effects in the context of dynamic discrete choice models. Cunha, Heckman, and Navarro (2007) consider ordered choice models with stochastic thresholds and show the relationship to mixed proportional hazard models. Approaches using instrumental variables are derived by Bijwaard (2001) for Ridders (1990) generalized accelerating failure time model and by Bijwaard and Ridder (2005) for the mixed proportional hazard model. Abbring and van den Berg (2005) define treatment effects concerning hazard rates of survival probabilities and consider identification and estimation in various situations with respect to time of treatment and compliance of the individuals to the instrumental variables.

The present paper uses a different approach to analyze treatment effects on durations. Instead of modelling hazard rates, the focus lies directly on the 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 actual computation of censored quantile regression estimators is quite challenging; see Buchinsky and Hahn (1998), Chernozhukov and Hong (2002), Fitzenberger (1997), Fitzenberger and Winker (2007), or Portnoy (2003), among others. Other LAD-type models with censored data are examined by Bang and Tsiatis (2002) and Huang, Ma, and Xie (2005).

In the present paper, treatment effects are identified by assuming independence of the latent outcomes and the treatment conditional on the covariates. 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 (2007a) by allowing censored outcomes. The propen-

sity score 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 sophisticated and 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 take 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. For an extensive description of this evaluation framework, 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 2.2.1 describes the approach of Horowitz and Mammen (2004), which is used to estimate the propensity score. Section 2.2.2 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. Consistency of the variance estimators is shown in section 2.2.3. 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 2.3. Section 2.4 concludes. All proofs are given in appendix 2.5.

2.2 Estimation Approach and Asymptotic Properties

2.2.1 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. It is similar to the approach of Hirano, Imbens and Ridder (2003), but avoids a strong differentiability assumption with respect to the true propensity score.

The treatment indicator D is is modelled as follows:

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

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:

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

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

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 $j = 0, 1$:

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

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 $j = 0, 1, 2$:

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

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 Y can be obtained by $\hat{Y} = F(\hat{\mu} + \hat{m}(X))$, 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 $U = Y - F(\mu + m(X))$. The variance of U is bounded from below and above by c_V and C_V, respectively, and $E|U|^j \leq C_U^{j-2} j! E[U^2] < \infty$ for all $j \geq 2$ 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) $x \in \mathbb{R}$, 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 average treatment effects, see Frölich (2005), Ichimura and Linton (2005), or Imbens, Newey, and Ridder (2005).

2.2.2 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. For later use, define also $q_{j,\tau} = q_\tau(T_j)$, $j \in \{0, 1\}$. 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 (2007a) for a discussion of rank invariance or, synonymously, rank preservation.

Eq. (3) is the 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 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 (2007a) 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. Note that identification in the present framework follows by Lemma 1 and Corollary 1 of Firpo (2007a) by substituting the moment conditions of the uncensored quantile estimator by those of censored quantile estimator and by assuming noninformative censoring.

Let $\{\tilde{T}_i, C_i, D_i, X_i\}_{i=1}^n$ be a random sample. $\tilde{T}_i$ is the observed (possibly right censored) outcome, which is equal to T_i if $\tilde{T}_i \leq C_i$ and is equal to C_i otherwise, where C_i is the censoring time. Note that the censoring time has also to be known for uncensored observations. This requirement is necesarry 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 (2007a) by the objective function of censored quantile regressions, the following estimators are suggested:

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

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)). Estimation of the propensity score is carried out by the nonparametric series estimator of Horowitz and Mammen (2004), which was described in section 2.2.1. 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* $\sqrt{n}(\hat{\Delta}_{\tau|D=1} - \Delta_{\tau|D=1}) \xrightarrow{D} \mathcal{N}(0, \Omega_{\tau|D=1})$.

The proof is given in appendix 2.5. The discussion of the variance expressions and their estimation is deferred to section 2.2.3. 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 non-negativity of the outcome, 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:

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

2.2.3 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 $\Delta_{\tau|D=1}$:

$$\begin{aligned}
M_\tau(q_\tau, p) &= E\left[\frac{D}{p(X)} 1\{q_{1,\tau} < C\}(\tau - 1\{\tilde{T} \leq q_{1,\tau}\})\right. \\
&\quad \left. -\frac{1-D}{1-p(X)} 1\{q_{0,\tau} < C\}(\tau - 1\{\tilde{T} \leq q_{0,\tau}\})\right]
\end{aligned}$$

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

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

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

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

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

[1]For a review of derivatives in a general setting, see Appendix A5 of Bickel et al. (1998).

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

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

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

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:

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

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:

$$\begin{aligned}
\Omega_\tau &= \left(\Gamma_{1,\tau}'W\Gamma_{1,\tau}\right)^{-1}\Gamma_{1,\tau}'WV_\tau W\Gamma_{1,\tau}\left(\Gamma_{1,\tau}'W\Gamma_{1,\tau}\right)^{-1}, \\
\Omega_{\tau|D=1} &= \left(\Gamma_{1,\tau|D=1}'W\Gamma_{1,\tau|D=1}\right)^{-1}\Gamma_{1,\tau|D=1}'WV_{\tau|D=1}W \\
&\quad \times\Gamma_{1,\tau|D=1}\left(\Gamma_{1,\tau|D=1}'W\Gamma_{1,\tau|D=1}\right)^{-1},
\end{aligned}$$

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

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

and

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

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

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

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

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

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

The proof is given in appendix 2.5.

2.3 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: $H_0 : \Delta_\tau \geq 0 \; \forall \; \tau$ 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: $H_0 : \Delta_{\tau|D=1} \geq \Delta_\tau \; \forall \; \tau$ 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 average 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 $a(\tau)$ is the value of Δ_τ under the null hypothesis. For example, in hypotheses 1 and 3, $a(\tau) = 0$ for all $\tau \in (0, 1)$. For hypothesis 4, $a(\tau)$ is equal to the conventional censored quantile regression parameters. For implemening the test statistic, $a(\tau)$ will be replaced by an estimate $\hat{a}(\tau)$. Viewed as a function of $\tau \in (0, 1)$, eq. (5) is called 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 coefficients, 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 (again, see sec. 19.3 of van der Vaart (1998)). 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* $\upsilon(\cdot) = b(\cdot) - d(\cdot)$

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

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

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. ℓ^∞ denotes the set of bounded functions (here, on the domain of $b(\cdot)$ and $d(\cdot)$). The proof is contained in appendix 2.5.

2.4 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).

2.5 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

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

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

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

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 $\inf_{\|q_\tau - q_\tau^*\| > \delta} \|M_n(q_\tau, p^*)\| \geq \epsilon(\delta) > 0$, 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′ $\sup_{q_\tau \in \mathcal{Q} \times \mathcal{Q}, \|p - p^*\| \leq \delta_n} \|M_n(q_\tau, p) - M(q_\tau, p)\| = o_P(1)$, 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)[2], 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 $\mathcal{Q}_\delta \equiv \{q_\tau \in \mathcal{Q} \times \mathcal{Q} | \; \|q_\tau - q_\tau^*\| \leq \delta\}$ and $\mathcal{P}_\delta \equiv \{p \in \mathcal{P} | \; \|p - p^*\| \leq \delta\}$ with

[2]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.

$\delta > 0$. By Theorem 2 of Chen, Linton, and Van Keilegom (2003), asymptotic normality follows by the following conditions[3]:

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

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

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

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

The derivative Γ_1 is therefore given as

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

The derivative of the objective of $\Delta_{\tau|D=1}$ is given in section 2.2.3. Assuming sufficiently smooth distributions of the censoring and

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

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. $\|M(q_\tau, p) - M(q_\tau, p^*) - \Gamma_2(q_\tau, p^*)(p - p^*)\| \leq k\|p - p^*\|^2$ for $k \geq 0$.
Rewrite the derivative as[4]

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

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

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 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:

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

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 $\delta_n = o_P(1)$,

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

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.

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

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 $r \geq 2$, i.e.,

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

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

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

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

Condition 3.2 will be fulfilled, if $E[|1\{q' < C\} - 1\{q < C\}|]$, $E[|1\{\tilde{T} \leq q\} - 1\{\tilde{T} \leq q'\}|]$, 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

$$
\begin{aligned}
& q - \delta < q < q + \delta \\
& \quad \Rightarrow \;\; 1\{q - \delta \; < C\} \geq 1\{q < C\} \geq 1\{q + \delta < C\},
\end{aligned}
$$

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

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

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

where the last line follows by the fact that $1\{q - \delta \leq C\} \geq 1\{q + \delta \leq C\}$. 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 $q + \delta$:

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

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:

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

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 $E[|1\{\tilde{T} \leq q\} - 1\{\tilde{T} \leq q'\}|]$. 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

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

The 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

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

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 2.2.3). 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 2.2.3. □

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 $o_P(1)$.

$$\begin{aligned}
\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} \\
&\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

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

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

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

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:

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

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:

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

The first two lines are bounded by

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

To bound the second difference, consider:

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

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

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

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

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

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

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. □

2.6 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.

Bang, H., and A. A. Tsiatis (2002): "Median Regression with Censored Cost Data," *Biometrics*, 58, 643–649.

Bickel, P. J., C. Klaassen, Y. Ritov, and J. Wellner (1998): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.

Bijwaard, G. E. (2001): "Instrumental Variable Estimation for Duration Data: A Reappraisal of the Illinois Reemployment Bonus Experiment," Econometric Institute Report EI 2002-39, Erasmus University Rotterdam.

Bijwaard, G. E., and G. Ridder (2005): "Correcting for Selective Compliance in a Re-Employment Bonus Experiment," *Journal of Econometrics*, 125, 77–111.

Buchinsky, M., and J. Hahn (1998): "An Alternative Model for the Censored Quantile Regression Estimnator," *Econometrica*, 66, 653–671.

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 I. Fernandez-Val (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhya*, 67, 253–276.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

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

Chernozhukov, V., and H. Hong (2002): "Three-Step Censored Quantile Regression and Extramarital Affairs," *Journal of the American Statistical Association*, 97, 872–882.

Chesher, A. (2002): "Semiparametric Identification in Duration Models," Cemmap Working Paper 20/02.

Cunha, F., J. J. Heckman, and S. Navarro (2007): "The Identification and Economic Content of Ordered Choice Models with Stochastic Tresholds," NBER Technical Working Paper 340.

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

Fitzenberger, B. (1997): "A Guide to Censored Quantile Regressions," in *Handbook of Statistics*, ed. by G. S. Maddala, and C. R. Rao, vol. 15, pp. 405–437. Elsevier, Amsterdam.

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

Fitzenberger, B., and P. Winker (2007): "Improving the Computation of Censored Quantile Regressions," *Computational Statistics & Data Analysis*, 52, 88–108.

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

Hahn, J. (1994): "The Efficiency Bound of the Mixed Proportional Hazard Model," *Review of Economic Studies*, 61, 607–629.

Ham, J. C., and R. J. LaLonde (1996): "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, 64, 175–205.

Heckman, J. J., and S. Navarro (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136, 341–396.

Heckman, J. J., and B. Singer (1984): "Econometric Duration Analysis," *Journal of Econometrics*, 24, 63–132, Correction: Vol. 27 (1985), 137-138.

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.

Huang, J., S. Ma, and H. Xie (2005): "Least Absolute Deviations Estimation for the Accelerated Failure Time Model," Technical Report No. 350, Department of Statistics and Actuarial Science, University of Iowa.

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.

Portnoy, S. (2003): "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012, Correction by T. Neocleous, K. Vanden Branden and S. Portnoy, vol. 101 (2006), 860-861.

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.

Ridder, G. (1990): "The Non-Parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57, 167–182.

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. North Holland, 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.

3 Double Robust Semiparametric Efficient Tests for Distributional Treatment Effects under the Conditional Independence Assumption

This paper describes methods to test for distributional treatment effects under the conditional independence assumption. The differences between latent outcome distributions are judged by testing hypotheses of distributional equality and stochastic dominance. Furthermore, semiparametric efficient versions of the test statistics are given. The latter test statistics are double robust, i.e., they are consistent under misspecification of either the outcome equation or the propensity score. Consistent bootstrap procedures for deriving critical values of all tests are proposed.

3.1 Introduction

This paper presents methods to test for significance of distributional treatment effects under the conditional independence assumption. Using the framework of econometric evaluation methods, the cumulative distribution functions of latent outcomes for treated and untreated individuals are compared. To judge on the significance of the impacts of a treatment on various parts of the outcome distribution, tests for equality and stochastic dominance are used.

To evaluate effects of treatments on the whole distribution of outcomes, several quantile and distributional treatment effect models were proposed in the literature. The approaches use the econometric evaluation framework and consider some difference between functionals of latent outcome distributions (for general overviews of econometric evaluation methods see Angrist (2004), Heckman, LaLonde, and Smith (1999), Imbens (2004), or Tan (2006 a, b)). They may be classified whether they are based on the horizontal or vertical difference of the distribution functions.

Following Doksum (1974), quantile treatment effects are defined as differences between quantiles of the latent outcome distributions for treated and untreated individuals (i.e., as horizontal difference). Such models were proposed by Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2005, 2006) using exclusion restrictions, and by Firpo (2007a) under the conditional independence assumption. Athey and Imbens (2006) derive quantile treatment effects for a difference-in-differences model. To summarize quantile treatment effects, Chernozhukov and Hansen (2005) and Chernozhukov and Fernandez-Val (2005) present formal procedures to test hypotheses on a set of quantile treatment effects. Within the conditional independence

framework, Bitler, Gelbach, and Hoynes (2006) apply the basic Kolmogorov-Smirnov testing approach of Abadie (2002) to test for any significant effects within a set of estimates for different quantiles.

Distributional treatment effect models examine the vertical distance between distributions, i.e. the difference between the distributions of treated and untreated individuals at a given point in the support of the outcome variable. Using exclusion restrictions, Abadie (2002) derives procedures to test for equality and stochastic dominance of the distributions of compliers, i.e., individuals who change their participation decision due to a change of the binary instrument (for the underlying concept of local average treatment effects see Imbens and Angrist (1994) or Angrist, Imbens, and Rubin (1996)), extending the work of Imbens and Rubin (1997). Using the conditional independence assumption, Imbens (2004) suggests estimators for cumulative distribution functions of latent outcomes based on the reweighting method of Hirano, Imbens, and Ridder (2003). Following this approach, Firpo (2007b) proposes estimators for functionals of distributions like variance or interquartile range.

Distributional treatment effect models are useful especially for examining stochastic dominance hypotheses. For definitions and connections to economic theory see McFadden (1989), for example. An empirical application is the analysis of Blundell et al. (2007). To derive bounds on differences between wage distributions, the restriction of positive selection into work is implemented by assuming stochastic dominance of the wage distribution of non-workers by that of workers. Methodological contributions involve Barrett and Donald (2003), who derive consistent tests for stochastic dominance of various orders, and Horváth, Kokoszka, and Zitikis (2006), who present a weighted test statistic to prevent degenerate limit distributions when testing for stochastic dominance of order three or higher. Davidson and Duclos (2000) consider tests for various measures of poverty and inequality and the corresponding limit distributions. Linton, Maasoumi, and Whang (2005) provide some generalizations by considering comparisons between more than two states of the quantity of interest and by admitting dependent observations.

The methods of this note are based on those of Abadie (2002), but differ in several aspects. First, identification relies on a different assumption. As opposed to the instrumental variable approach of Abadie (2002), the methods proposed here use the conditional independence assumption. Second, a semiparametric efficient test statistic is derived. Using the reweighting approach of Hirano, Imbens, and Ridder (2003), the propensity score is modelled by a nonparametric series estimator, which is plugged in the estimators of the latent distribution functions. In parlance of the theory of

semiparametric efficiency, the propensity score is an infinite dimensional nuisance parameter. For testing semiparametric hypotheses, Bickel, Ritov, and Stoker (2006) suggest efficient tests which concentrate on deviations of the test statistics that are not influenced by the nuisance parameter. Finally, it is shown that the semiparametric efficient test statistics are double robust in the sense of Scharfstein, Rotnitzky, and Robins (1999). In the context of this note, this means that the estimator of the test statistic is consistent whenever the propensity score or the outcome equation is correctly specified. Put differently, either the propensity score or the outcome equation may be misspecified without leading to inconsistent estimates of the test statistic.

In the following, section 3.2 describes the framework and the assumptions of the tests and the semiparametric estimators of the test statistics. Section 3.3 derives semiparametric efficient test statistics and states their double robustness property. Resampling methods to obtain critical values of the test statistics are described in sections 3.2 and 3.3. All proofs are given in appendix 3.5.

3.2 Tests for Distributional Treatment Effects

Consider the effect of a binary treatment D on a nonnegative outcome Y. It is assumed that for every individual and for each value of the treatment D (nonnegative) latent outcomes Y_1 and Y_0 exist. The observable outcome is therefore given by $Y = DY_1 + (1-D)Y_0$. Contrary to most of the econometric evaluation literature, this paper considers the effect on the whole outcome distribution. The cumulative distribution functions of the latent outcomes Y_1 and Y_0 are denoted by F_{Y_1} and F_{Y_0}, respectively.

The distributional treatment effect of D on Y at a point γ of the support of Y is defined to be $F_{Y_1}(\gamma) - F_{Y_0}(\gamma)$. To judge on distributional inequality, Kolmogorov-Smirnov tests are used to compare the distribution functions of Y_1 and Y_0 for all values of the outcome. Following Abadie (2002), the following null hypotheses are tested:

$$F_{Y_1}(\gamma) = F_{Y_0}(\gamma) \ \ \forall \gamma \in \mathbb{R}_+, \tag{9}$$

$$F_{Y_1}(\gamma) \leq F_{Y_0}(\gamma) \ \ \forall \gamma \in \mathbb{R}_+, \tag{10}$$

$$\int_{-\infty}^{\gamma} F_{Y_1}(x)dx \leq \int_{-\infty}^{\gamma} F_{Y_0}(x)dx \ \ \forall \gamma \in \mathbb{R}_+. \tag{11}$$

$\mathbb{R}_+$ is defined as $\{x \in \mathbb{R} | x > 0\}$. (9) is the null hypothesis for equality of distributions, (10) and (11) are those for first and second order stochastic dominance, respectively. If the distribution F_{Y_1} first-order stochastically dominates F_{Y_0}, the distribution of Y_1 is (equal or) spread more widely than

that of Y_0. For example, in the context of the evaluation of active labor market programs this means that the latent outcome in the treatment case is less concentrated than in the nontreatment case. This concept, however, is of little use if none of the investigated distributions first-order dominates the other. Furthermore, not all requirements of social welfare analysis are met by first order stochastic dominance. To circumvent some of these problems, the concept of second order stochastic dominance can be used. For more details on the measurement of inequality, see Cowell (2000), for example.

To implement the tests, estimators of the latent outcome distributions $F_{Y_1}(\gamma)$ and $F_{Y_0}(\gamma)$ are constructed using results of Firpo (2007b). This approach applies the reweighting method of Hirano, Imbens, and Ridder (2003), which yields estimators for $E[Y_1]$ and $E[Y_0]$ by reweighting the observable outcome Y by suitable functions of the conditional probability of receiving the treatment (propensity score). The underlying approach traces back to Horvitz and Thompson (1952). In the context of general missing data models it is used by Robins, Rotnitzky, and Zhao (1994), for example; see also Tsiatis (2006). Hirano, Imbens, and Ridder (2003) estimate the propensity score by a nonparametric series estimator (see their Appendix A for details). Using the fact that $F_A(a) = P(A \leq a) = E[1\{A \leq a\}]$, Firpo (2007b) reweights the indicator function $1\{Y \leq \gamma\}$ to get an estimator of the distribution functions of the latent outcomes of the treated and untreated. This approach was also suggested by Imbens (2004). Using the weights, the unobservable (latent) distribution functions F_{Y_1} and F_{Y_0} can be expressed by functions of the observable variables Y, D, and X.

Let $\{Y_i, D_i, X_i\}_{i=1}^n$ be a (large) sample. The assumptions of the following analysis are based on those of Hirano, Imbens, and Ridder (2003):

A1 The latent outcomes are independent of the treatment conditional on the covariates: $(Y_1, Y_0) \perp\!\!\!\perp D|X$.

A2 The support of X is the cartesian product of compact intervals and the density of X is bounded from above and away from zero, and is continuously differentiable for all x in the support of X.

A3 $E[1\{Y_j \leq \gamma\}|X]$ is continuously differentiable for all elements of the support of X ($j \in \{0, 1\}$).

A4 The propensity score $p(X) \equiv E[D|X]$ is continuously differentiable of order $s \geq 7r$ (where r is the dimension of X) and is bounded away from zero and one.

A5 The series logit estimator of $p(X)$ uses a power series with $K = n^\nu$ elements, where $1/(4(s/r - 1)) < \nu < 1/9$.

Assumption A1 is the usual conditional independence assumption. A2 excludes discrete explanatory variables; this situation can be coped with by conditioning on cells of combinations of the discrete variables, if the sample is large enough. A3 is needed to show pointwise consistency of the estimators. The assumption of a high number of derivatives (A4) of the nonparametrically estimated function is usually made in the literature of series estimators (see Newey (1994, 1995, 1997), for example), though it is a strong assumption. It is used by Hirano, Imbens, and Ridder (2003) to ensure the existence of a ν in assumption A5, which restricts in turn the rate of inclusion of new elements in the series to assure consistency of the estimators.

Under these assumptions and given an estimate $\hat{p}(X)$ of the propensity score, an estimator for the difference $F_{Y_1}(\gamma) - F_{Y_0}(\gamma)$ at a given point γ can be obtained using a proposal of Firpo (2007b):

$$\begin{aligned}\hat{F}_{Y_1}(\gamma) - \hat{F}_{Y_0}(\gamma) &= \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{D_i}{\hat{p}(X_i)}1\{Y_i \leq \gamma\} - \frac{1}{n}\sum_{i=1}^{n}\frac{1-D_i}{1-\hat{p}(X_i)}1\{Y_i \leq \gamma\}. \qquad (12)\end{aligned}$$

The latent distribution functions for the subgroup of treated individuals may be estimated using different weights, with $p \equiv n^{-1}\sum_{i=1}^{n} D_i$ as the unconditional mean of D:

$$\begin{aligned}\hat{F}_{Y_1|D=1}(\gamma) - \hat{F}_{Y_0|D=1}(\gamma) &= \\ = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i}{p}1\{Y_i \leq \gamma\} - \frac{1}{n}\sum_{i=1}^{n}\frac{1-D_i}{p}\frac{\hat{p}(X_i)}{1-\hat{p}(X_i)}1\{Y_i \leq \gamma\}. \qquad (13)\end{aligned}$$

The further exposure will mainly consider the effects for a randomly chosen individual (i.e., it will be based on estimator (12)). All results are also valid for estimator (13).

Pointwise consistency of $\hat{F}_{Y_1}(\gamma)$ and $\hat{F}_{Y_0}(\gamma)$ follows directly by adapting results for the mean case of Hirano, Imbens, and Ridder (2003). With estimators of F_{Y_1} and F_{Y_0} at hand, test statistics for hypotheses (9) - (11) can be defined. These statistics resemble those of Abadie (2002). The two groups to be compared are devided by the value of D, which is a difference to the original tests of Abadie (2002), who divides the groups with respect to the value of the binary instrumental variable. The Kolmogorov-Smirnov test statistics for the hypotheses of equality of distributions and first and second order stochastic dominance are given by:

$$T_n^{eq} = \sqrt{\frac{n}{2}}\sup_{\gamma\in\mathbb{R}_+}\left|\hat{F}_{Y_1}(\gamma) - \hat{F}_{Y_0}(\gamma)\right|, \qquad (14)$$

$$T_n^{fsd} = \sqrt{\frac{n}{2}} \sup_{\gamma \in \mathbb{R}_+} \left(\hat{F}_{Y_1}(\gamma) - \hat{F}_{Y_0}(\gamma) \right), \tag{15}$$

$$T_n^{ssd} = \sqrt{\frac{n}{2}} \sup_{\gamma \in \mathbb{R}_+} \int_{-\infty}^{\gamma} \left(\hat{F}_{Y_1}(x) - \hat{F}_{Y_0}(x) \right) dx. \tag{16}$$

See the proof of Theorem 1 for a note on the factor $\sqrt{1/2}$.

The distributions of the test statistics depend on the unknown true probability distribution (see sec. 19.3 of van der Vaart (1998) or Romano (1988)). Therefore, following again Abadie (2002), Algorithm 1 proposes a bootstrap procedure to approximate the distribution of the test statistics.

Algorithm 1:

1 The test statistics given by (14) - (16) (henceforth abbreviated by $T_n^{\cdot}$) are computed using the original sample. Set n_1 equal to the number of individuals with $D = 1$ and let $n_0 = n - n_1$.

2 A resample with replacement is drawn. The first n_1 observations are classified as treated (i.e., D is set equal to one), the remaining n_0 as untreated ($D = 0$). Using the resample, test statistics $\tilde{T}_n^{\cdot}$ are computed. Note that the propensity score is re-estimated for each subsample.

3 To approximate the distribution of the test statistics, the second step is repeated B times.

4 The p-values of the tests are calculated as the fraction of test statistics $\tilde{T}_{n,j}^{\cdot}$ of the subsamples which exceed the original value $T_n^{\cdot}$:

$$\text{p-value of } T_n^{\cdot} = \frac{1}{B} \sum_{j=1}^{B} 1\{\tilde{T}_{n,j}^{\cdot} > T_n^{\cdot}\}.$$

Abadie (2002) derives asymptotic properties of the bootstrap procedure. Theorem 1 notes that the procedure just described shares these features.

Theorem 1: *The test procedure described by Algorithm 1 has correct size, is consistent against fixed alternatives and has power against contiguous alternatives.*

The proof is given in appendix 3.5. The main step is showing that the test statistics converge uniformly to a Brownian bridge. The properties of Theorem 1 then follow essentially by the same reasoning as in the proof of Proposition 2.2 of Abadie (2002).

3.3 Semiparametric Efficient Tests

Due to the semiparametric nature of the estimators used for the construction of the test statistics, an extension of the approach of section 3.2 is appropriate. For semiparametric goodness of fit tests, Bickel, Ritov, and Stoker (2006) suggest the use of a semiparametric efficient formulation of the test statistic as the base for the testing procedure. As the propensity score is approximated nonparametrically, it is an infinite-dimensional nuisance parameter for the estimation of the distribution functions of Y_1 and Y_0 at a given point γ. The test statistic is influenced not only by differences between F_{Y_1} and F_{Y_0}, but also by deviations induced by the estimated propensity score. Therefore, to concentrate on differences of the distribution functions, the test statistic is adjusted for influences of the propensity score. Technically, only movements of the test statistic in the tangent space associated with differences between the distribution functions are considered without paying attention to movements in the nuisance parameter tangent space. That is, the test statistic is projected on the orthogonal complement of the nuisance parameter tangent space, which corresponds to using the difference of the test statistic and its projection on the nuisance parameter tangent space. The underlying concepts of semiparametric efficiency are discussed in detail in Bickel et al. (1998), Groeneboom and Wellner (1992), Newey (1990), Tsiatis (2006) and van der Vaart (1998, 2002), for example.

The basic building block of the approach of Bickel, Ritov and Stoker (2006) is what they call the (test) score process, which is $\sqrt{n}$ times an estimator of the projection $\Pi^{\perp}(h(\cdot))$ of the test score on the orthogonal complement of the nuisance parameter tangent space. The score process is given by

$$Z_n(\gamma, \alpha) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Pi^{\perp}(h(\gamma, \alpha), \alpha)(X_i). \tag{17}$$

Here, $h(\gamma, \alpha)$ is some test statistic, which may depend on the nuisance parameter α. In the present application $h(\gamma, \alpha)$ is the difference of the latent cumulative distribution functions at some point γ, and the nuissance parameter α is the propensity score. Given an estimate $\hat{\alpha}$, $Z_n(\gamma, \hat{\alpha})$ can be used to implement an test. Proposition 2 below shows that $Z_n(\gamma, \hat{\alpha})$ converges weakly to $Z(\gamma, \alpha_0)$.

This general testing approach can be adapted to the question of this note. First, an expression of the efficient test statistic is needed. Define $\mu(\gamma|j, X) \equiv E[1\{Y \leq \gamma\}|D = j, X]$. Hahn (1998), Hirano, Imbens, and Ridder (2003) and Tsiatis (2006, sec. 13.5) derive semiparametric efficient estimators for average treatment effects, which can be directly adapted to

the present framework:

$$\begin{aligned}
&\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i}{\hat{p}(X_i)}1\{Y_i \leq \gamma\} - \frac{D_i - \hat{p}(X_i)}{\hat{p}(X_i)}\hat{\mu}(\gamma|1, X_i)\right) \\
&\quad - \left(\frac{1 - D_i}{1 - \hat{p}(X_i)}1\{Y_i \leq \gamma\} - \frac{D_i - \hat{p}(X_i)}{1 - \hat{p}(X_i)}\hat{\mu}(\gamma|0, X_i)\right)\Bigg). \qquad (18)
\end{aligned}$$

Comparing (18) with (12), one notes additional terms. These are the projections of the estimators of F_{Y_1} and F_{Y_0} on the nuisance parameter tangent space. The corresponding expression for the subgroup of treated individuals is:

$$\begin{aligned}
\hat{F}^{\text{eff}}_{Y_1|D=1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0|D=1}(\gamma) &= \frac{1}{n}\sum_{i=1}^{n}\Bigg(\frac{D_i}{p}(1\{Y_i \leq \gamma\} - \hat{\mu}(\gamma|1, X_i)) \\
&\quad - \frac{1 - D_i}{p}\frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)}(1\{Y_i \leq \gamma\} - \hat{\mu}(\gamma|0, X_i)) \\
&\quad + \frac{\hat{p}(X_i)}{p}(\hat{\mu}(\gamma|1, X_i) - \hat{\mu}(\gamma|0, X_i))\Bigg). \qquad (19)
\end{aligned}$$

Therefore, for the specific testing problem of this note, the (estimated) projection on the orthogonal complement of the nuisance parameter space is given by

$$\Pi^{\perp}(h(\gamma, \hat{\alpha}), \hat{\alpha}) = \hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma). \qquad (20)$$

An estimator of the test score process (17) is equal to (20) multiplied by $\sqrt{n}$. Here, $h(\gamma, \hat{\alpha})$ is the difference between $\hat{F}_{Y_1}(\cdot)$ and $\hat{F}_{Y_1}(\cdot)$, γ corresponds to a value in the support of Y, and α is the nuisance parameter, i.e., the propensity score. The semiparametric efficient test statistics are given by applying the transformations of (14) - (16) to the score process, i.e., $Z_n(\gamma, \hat{\alpha})$ replaces the simple difference $\hat{F}_{Y_1} - \hat{F}_{Y_0}$. Following Hirano, Imbens, and Ridder (2003, p. 1172), the conditional expectation $\mu(j, X)$ can be estimated by a series approximation like that used for the propensity score. The weak convergence of the test statistic to its true value is stated in the following theorem.

Theorem 2: *The estimator of the score process is given by*

$$Z_n(\gamma, \hat{\alpha}) = \sqrt{n}\left(\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma)\right). \qquad (21)$$

Under assumptions A1 - A5, $Z_n(\gamma, \hat{\alpha})$ converges weakly to $Z(\gamma, \alpha_0)$, which is a Gaussian process with mean zero.

The proof is given in appendix 3.5 and consists in checking that the assumptions of Theorem 3.2 of Bickel, Ritov, and Stoker (2006) hold in the present application.

The critical values of the test can be determined by a bootstrap procedure.

Algorithm 2:

1 Derive the test statistics for the original sample by estimating $Z_n(\gamma, \hat{\alpha})$ over a grid of values γ in the support of Y and apply the transformations $T_n^{\cdot}$ described by equations (14) - (16).

2 A resample with replacement is drawn. The first n_1 observations are viewed as treated (i.e., $D = 1$), the remaining n_0 as untreated ($D = 0$). Following Bickel, Ritov, and Stoker (2006, sec. 3.3), the transformations $T_n^{\cdot}$ are applied to the centered projections:

$$\tilde{Z}_n(\gamma, \hat{\alpha}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\Pi^{\perp}(h(\gamma, \hat{\tilde{\alpha}}), \hat{\tilde{\alpha}})(\tilde{X}_i) - \Pi^{\perp}(h(\gamma, \hat{\alpha}), \hat{\alpha})(X_i) \right).$$

Here, $\hat{\tilde{\alpha}}$ is an estimate of the propensity score based on the bootstrap sample. The resulting test statistics are denoted by $\tilde{T}_n^{\cdot}$.

3 To approximate the distribution of the test statistics, the second step is repeated B times.

4 The p-values of the tests are calculated as the fraction of test statistics of the resamples which exceed the original value:

$$\text{p-value of } T_n^{\cdot} = \frac{1}{B} \sum_{j=1}^{B} 1\{\tilde{T}_{n,j}^{\cdot} > T_n^{\cdot}\}.$$

The applicability of the bootstrap procedure of Algorithm 2 is shown by the following theorem.

Theorem 3: *The bootstrap test statistic of Algorithm 2 converges weakly to the true test statistic.*

The proof of this statement applies Theorem 2.1 of Bickel and Ren (2001) and is contained in appendix 3.5.

The next theorem shows double robustness of the efficient test statistics. This concept was introduced by Scharfstein, Rotnitzky, and Robins (1999). Other descriptions of this approach are Bang and Robins (2005), Neugebauer and van der Laan (2005), Robins and Rotnitzky (2001), Tsiatis (2006) and

van der Laan and Robins (2003). Here, this property means that the estimates of the test statistic are consistent under the presence of inconsistent estimates of either the propensity score or the outcome equation.

An example shows the usefulness of double robust estimators in empirical evaluation studies. Let D be an indicator for further training, and Y be some income measure. Assume selection in D depends on observable characteristics only, such as age, eduction and labor market experience. This might well be the case in some programs of active labor market policy. Beside these factors, Y may depend furthermore on some unobservable factors like motivation. Normally, this is a serious problem for methods relying on the conditional independence assumption. If panel data are available, a possible solution is to apply some kind of differencing scheme to remove time-invariant unobservable factors (see Heckman et al. (1998) for such an approach). A double robust estimator like that of the present paper would yield consistent results in the presence of unobserved influence factors, as long as the propensity score is correctly specified. A further potential advantage is stated by Tsiatis (2006, p. 150), who notes that double robust reweighting estimators seem to be less variable when estimates of the propensity score near zero or one occur.

Theorem 4: *The semiparametric efficient estimators* $\hat{F}^{eff}_{Y_j}(\gamma)$ *and* $\hat{F}^{eff}_{Y_j|D=1}(\gamma)$ $(j \in \{0,1\})$ *remain consistent if either the estimator of the propensity score or of that the outcome equation is inconsistent.*

The proof is given in section 3.5. It follows a derivation of Tsiatis (2006, sec. 13.5) for a mean treatment effect model.

3.4 Conclusion

This paper describes testing methods for distributional treatment effects under the conditional independence assumption. Hypotheses for equality of distributions and stochastic dominance are considered. An extension describes a semiparametric efficient formulation of the test statistic, for which the property of double robustness is shown to hold. Bootstrap procedures for obtaining critical values of these tests are presented.

3.5 Proofs

Proof of Theorem 1: It will be shown that the derivations of Abadie (2002) are basically also valid for the procedure described in section 3.2. The first step is to show that the estimator of the difference of the latent outcome distribution functions converge to a Gaussian process. To this end, it is

shown that $\sqrt{n}(\hat{F}_{Y_1} - F_{Y_1}) \Rightarrow \mathbb{G}_{F_{Y_1}}$ and $\sqrt{n}(\hat{F}_{Y_0} - F_{Y_0}) \Rightarrow \mathbb{G}_{F_{Y_0}}$, where '$\Rightarrow$' denotes weak functional convergence, i.e., uniform convergence of a function to a stochastic process (here, to Brownian bridges). This can be deduced from results of empirical process theory, which is described in Andrews (1994a,b), Pollard (1984, 1990), van de Geer (1990, 2000), van der Vaart (1998, 2002) and van der Vaart and Wellner (1996), for example. Weak convergence will follow if the estimators of F_{Y_1} and F_{Y_1} form Donsker classes.

Sets of indicator functions and monotone mappings from $\mathbb{R}$ to $(0, 1)$ in general are Donsker classes (see example 2.6.21 (p. 149) of van der Vaart and Wellner (1996)). The set $\{1/f | f \in \mathcal{F}, \mathcal{F} \text{ is a Donsker class}\}$ is also a Donsker class if every $f > 0$ (example 2.10.9 (p. 192) of van der Vaart and Wellner (1996)). These results, together with a permanence property of Donsker classes (example 2.10.8 (p. 192) of van der Vaart and Wellner (1996)) and assumption A4 show the Donsker property and hence the weak convergence of the estimators.

The scaled difference of the estimators can be rewritten as

$$\sqrt{\frac{n}{2}}\left(\hat{F}_{Y_1} - \hat{F}_{Y_0}\right) = \sqrt{\frac{n}{2}}\left(\hat{F}_{Y_1} - F_{Y_1} - \left(\hat{F}_{Y_0} - F_{Y_0}\right) + (F_{Y_1} - F_{Y_0})\right).$$

Under the null hypothesis $F_{Y_1} = F_{Y_0}$ ($\equiv F$), weak convergence of the following expression results:

$$\sqrt{\frac{1}{2}}\sqrt{n}\left(\hat{F}_{Y_1} - F\right) - \sqrt{\frac{1}{2}}\sqrt{n}\left(\hat{F}_{Y_0} - F\right) \Rightarrow \sqrt{\frac{1}{2}}\mathbb{G}_F - \sqrt{\frac{1}{2}}\mathbb{G}'_F,$$

where $\mathbb{G}_F$ and $\mathbb{G}'_F$ are independent Brownian bridges with law F. The difference of the right hand side converges also to a Brownian bridge, which shows convergence of the scaled difference of the distribution functions to a Gaussian process under the null hypothesis. The factor $\sqrt{1/2}$ is necessary for the convergence of the difference of two Brownian bridges (see the results for transformations of Brownian motion in Shorack (2000, p. 303)).

To proceed, convergence of the bootstrap test statistics to some Gaussian process will be shown. By Theorem 23.7 of van der Vaart (1998), the bootstrap empirical distribution functions converge in probability to the empirical distribution function. Since the estimators of the latent cdfs are Hadamard differentiable with respect to the empirical distribution function (which is simply multiplied by some other function), it follows by the Delta method for the bootstrap (Theorem 23.9 of van der Vaart (1998)), that the bootstrap test statistics converge to the same limit as the original test statistics (see also example 23.11 of van der Vaart (1998)).

The rest of the proof follows closely Abadie (2002). He shows continuity of all $T_n^{\cdot}$, which follows by the fact that these transformations are norms

or can be bounded by norms. Furthermore, the $T_n^{\cdot}$ are convex continuous functionals. Correct asymptotic size follows by results for such functionals. In addition, consistency against any fixed alternative follows.

Power against contiguous alternatives can also be shown analogous to Abadies (2002) derivations. Rewrite the test statistic under the null hypothesis as a general empirical process (in the present case, the functions $f \in \mathcal{F}$ are empirical distribution functions):

$$\begin{aligned}
&\sqrt{\frac{n}{2}}\left(\hat{F}_{Y_1} - \hat{F}_{Y_0}\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{D_i}{\hat{p}(X_i)} - \frac{1-D_i}{1-\hat{p}(X_i)}\right)(f(Y_i) - P(f(Y_i)) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{D_i}{E[D|X_i]} - \frac{1-D_i}{1-E[D|X_i]}\right)(f(Y_i) - Pf(Y_i)) + o_P(1),
\end{aligned}$$

where $Pf(Y) = \int f(y)dP_Y(y)$ and P is the probability measure of Y under the null hypothesis. Consider now local alternatives P_n with the following property:

$$\int\left(\sqrt{n}\left(dP_n^{1/2} - dP^{1/2}\right) - \frac{1}{2}xdP^{1/2}\right)^2 \to 0,$$

where x is a measurable function (see section 3.10.1 of van der Vaart and Wellner (1996)). With this setting, and by the Donsker property of $\mathcal{F}$, contiguity of the test procedure follows essentially by the same reasoning as in the proof of Proposition 2.2 of Abadie (2002) or in van der Vaart and Wellner (1996, sec. 3.10.1).

Similar derivations hold for the tests of the subpopulation of treated individuals. Therefore, Algorithm 1 shares all asymptotic properties of the method of Abadie (2002). □

Proof of Theorem 2: In the following, conditions M0, M1, M3, N1 and N2 of Bickel, Ritov, and Stoker (2006) are shown to hold for the score process of this note, which proves the applicability of their Theorem 3.2. From this, weak convergence of the test score process follows. Like the proof of Theorem 1, the following derivations rely primarily on empirical process methods.

M0 It has to be shown that $\{\Pi^{\perp}(h_\gamma(\cdot,\alpha_0),\alpha_0)|\gamma \in \mathbb{R}_+\}$ is an universal Donsker class. That is, the class is a P-Donsker class for every probability measure P on the sample space (see van der Vaart and Wellner (1996, p. 82)). In the present application, this set is given by:

$$\left\{\Pi^{\perp}(h_\gamma(\cdot,\alpha_0),\alpha_0)\middle|\,\gamma \in \mathbb{R}_+\right\} = \left\{F_{Y_1}^{\text{eff}}(\gamma) - F_{Y_0}^{\text{eff}}(\gamma)\middle|\,\gamma \in \mathbb{R}_+\right\}$$

$$= \left\{ \frac{D}{p(X)} 1\{Y \leq \gamma\} - \frac{D - p(X)}{p(X)} \mu(\gamma|1, X) - \left(\frac{1-D}{1-p(X)} 1\{Y \leq \gamma\} - \frac{D - p(X)}{1-p(X)} \mu(\gamma|0, X) \right) \middle| \gamma \in \mathbb{R}_+ \right\}.$$

It is well known that the set of empirical distribution functions is a Donsker class (van der Vaart (1998, sec. 19.2)). The propensity score belongs to the set of bounded functions from $\mathbb{R} \to (0, 1)$. Uniformity in P follows by the facts that classes of monotone functions are Donsker classes for all probability measures P, and that $\{1/f | f \in \mathcal{F}, \mathcal{F}$ is a Donsker class$\}$ is also a Donsker class (Examples 2.6.21 (p. 149) and 2.10.9 (p. 192) of van der Vaart and Wellner (1996)).

M1 $\|(P_n - P_0)(\Pi(h_\gamma(\cdot, \hat{\alpha}), \hat{\alpha}) - \Pi(h_\gamma(\cdot, \hat{\alpha}), \alpha_0))\|_\infty = o_P(n^{-1/2})$, where $\|f\|_\infty \equiv \sup_x |f(x)|$ (for $x \in \text{supp}(X)$) is the supremum norm. In the context of the present application, this condition is equivalent to

$$\left\| (P_n - P_0) \left(\frac{D - \hat{p}(X)}{\hat{p}(X)} \hat{\mu}(\gamma|1, X) + \frac{D - \hat{p}(X)}{1 - \hat{p}(X)} \hat{\mu}(\gamma|0, X) \right) - \left(\frac{D - p(X)}{p(X)} \mu(\gamma|1, X) + \frac{D - p(X)}{1 - p(X)} \mu(\gamma|0, X) \right) \right\|_\infty$$

being $o_P(n^{-1/2})$. This condition is fulfilled if the projection of the test statistic on the nuissance parameter tangent space, i.e.,

$$\Pi(h_\gamma(\cdot, \hat{\alpha}), \hat{\alpha}) = \frac{D - \hat{p}(X)}{\hat{p}(X)} \hat{\mu}(\gamma|1, X) + \frac{D - \hat{p}(X)}{1 - \hat{p}(X)} \hat{\mu}(\gamma|0, X),$$

is stochastic equicontinuous. As all functions are bounded, their entropy integral and the envelope function are finite. Therefore, convergence follows by empirical process arguments (see van de Geer (2000, sec. 6), for example).

M3 $\|(P_n - P_0)(h_\gamma(\cdot, \hat{\alpha}) - h_\gamma(\cdot, \alpha_0))\|_\infty = o_P(n^{-1/2})$, which follows if $h_\gamma(\cdot, \hat{\alpha})$ is stochastic equicontinuous. This function is equal to the difference of the (simple) estimators of the counterfactual distributions, i.e., $h_\gamma(\cdot, \hat{\alpha}) = \hat{F}_{Y_1}(\gamma) - \hat{F}_{Y_0}(\gamma)$. These functions are bounded with finite entropy, which shows the validity of condition M3 for the present application.

N1 (i) $\hat{\alpha}$ ist consistent with respect to the Hellinger (pseudo-) metric, i.e.,

$$H^2(P_{\hat{\alpha}}, P_{\alpha_0}) = \frac{1}{2} \int \left(\sqrt{\frac{dP_{\hat{\alpha}}}{d\mu}} - \sqrt{\frac{dP_{\alpha_0}}{d\mu}} \right)^2 d\mu \to 0.$$

Let $f_{\alpha_0} = dP_{\alpha_0}/d\mu$ and $f_{\hat{\alpha}} = dP_{\hat{\alpha}}/d\mu$. Lemma 1.1 of van de Geer (1993) shows the following relationship:

$$\int_{f_0>0} \left(\sqrt{\frac{f_{\hat{\alpha}}}{f_{\alpha_0}}} - 1 \right) d(P_n - P_0) \geq H^2(f_{\hat{\alpha}}, f_{\alpha_0}). \tag{22}$$

Define the set $\mathcal{G} = \{(\sqrt{f_\theta/f_0} - 1)1\{f_0 > 0\} | \theta \in \Theta\}$. Consistency of $\hat{\alpha}$ with respect to the Hellinger metric results by (22), if the following holds:

$$\sup_{g \in \mathcal{G}} \left| \int g d(P_N - P_0) \right| \to 0 \text{ a.s.} \tag{23}$$

Validity of (23) can be shown by empirical process methods; in the present context, the conditions of Theorem 3.1 of van de Geer (1993) will be shown to hold.

To this end, consider the following densities:

$$\begin{aligned} f_{\alpha_0}(y, d, x) &= (f_1(y|x)p(x))^d (f_0(y|x)(1 - p(x)))^{1-d} f(x), \\ f_{\hat{\alpha}}(y, d, x) &= (f_1(y|x)\hat{p}(x))^d (f_0(y|x)(1 - \hat{p}(x)))^{1-d} f(x). \end{aligned}$$

f_{α_0} is the true density of the data, $f_{\hat{\alpha}}$ is the density given an estimate of the propensity score. $f_j(y|x)$ is the marginal conditional density of Y_j of the joint density $f(y_1, y_0|x)$ of the latent outcomes Y_1 and Y_0. Define $\mathcal{G}_0 = \{\sqrt{f_\theta} \, | \theta \in \Theta\}$ and let $\mathcal{H}(\delta, \mathcal{G}_0, \| \cdot \|_{P_n,\infty})$ denote the δ-entropy (i.e., the logarithm of the δ-covering number) of $\mathcal{G}_0$. If the following three conditions hold, $H^2(\hat{f}_{\hat{\alpha}}, f_{\alpha_0}) \to 0$ a.s. by Theorem 3.1 of van de Geer (1993).

(a) $\mathcal{G}_0$ is uniformly bounded,
(b) $\int \sqrt{f_0} d\mu < \infty$,
(c) $n^{-1}\mathcal{H}(\delta, \mathcal{G}_0, \| \cdot \|_{P_n,\infty}) \to 0$.

Conditions (a) and (b) are satisfied by suitable assumptions on the densities (see assumption A2). The finiteness of entropy of condition (c) follows by the boundedness of the densities of $\mathcal{G}_0$, which follows by assumption A2. Therefore, N1 (i) is shown to hold in the present context.

(ii) For $s(\alpha) = \sqrt{dP_\alpha/d\mu}$, $\hat{s} = s(\hat{\alpha})$ and $s_0 = s(\alpha_0)$ it holds that

$$\left\| \frac{(\hat{s} - s_0)^2}{\hat{s}} \right\|_{2,\mu}^2 = o_P(n^{-1}),$$

where $\|f\|_{p,\mu} \equiv (\int |f|^p d\mu)^{1/p}$ is the L_p norm. The norm condition can be rewritten as

$$\begin{aligned}
\left\| \frac{(\hat{s}-s_0)^2}{\hat{s}} \right\|_{2,\mu}^2 &= \int \frac{(\hat{s}-s_0)^4}{\hat{s}^2}\, d\mu \;=\; o_P(n^{-1}) \\
\Leftrightarrow \quad n \int \frac{(\hat{s}-s_0)^4}{\hat{s}^2}\, d\mu &\;=\; o_P(1) \\
\Leftrightarrow \quad \int \left(\sqrt{n}(\hat{s}-s_0)\right)^2 (\hat{s}-s_0)^2 \frac{1}{\hat{s}^2}\, d\mu &\;=\; o_P(1). \qquad (24)
\end{aligned}$$

Assume that $s(\alpha)$ is differentiable in quadratic mean. By an intermediary result of the proof of Theorem 7.2 of van der Vaart (1998, p. 94), $(\sqrt{n}(\hat{s}-s_0))^2$ converges in quadratic mean to $1/4 h^T I_\alpha h < \infty$, where I_α is the Fisher information matrix and h is a vector. The third factor of (24) is also bounded, the second converges to zero. Therefore, condition N1 (ii) is fulfilled in the present application.

(iii) With Π_μ as projection on the tangent space of $s(\alpha)$ for $s(\alpha_0)$, it holds that

$$\|\hat{s} - s_0 - \Pi_\mu(\hat{s}-s_0)\|_{2,\mu} = o_P(n^{-1/2}).$$

The condition may be rewritten as:

$$\begin{aligned}
&\|\hat{s} - s_0 - \Pi_\mu(\hat{s}-s_0)\|_{2,\mu}^2 \\
&= \int (\hat{s} - s_0 - \Pi_\mu(\hat{s}-s_0) d\mu)^2\, d\mu \;=\; o_P(n^{-1}) \\
&\Leftrightarrow \int \left(\frac{\hat{s}-s_0}{n^{-1/2}} - \frac{\Pi_\mu(\hat{s}-s_0)}{n^{-1/2}} \right)^2 d\mu \;=\; o_P(1)
\end{aligned}$$

As $\Pi_\mu(\hat{s}-s_0)$ is an element of the tangent space at s_0, the condition reads as

$$\int \left(\frac{\hat{s}-s_0}{n^{-1/2}} - \frac{\dot{s}}{n^{-1/2}} \right)^2 d\mu,$$

which is $o_P(1)$ by mean square differentiability of $s(\alpha)$. This way of reasoning was used by Bickel, Ritov and Stoker (2006) in their example 4.1.

N2 (i) $\sup_{\gamma,\alpha} \|h_\gamma(\cdot,\alpha)\|_\infty < \infty$ and (ii) $\sup_{\gamma,\alpha} \|\Pi(h_\gamma(\cdot,\alpha))\|_\infty < \infty$. This follows directly from the definitions and assumptions, as both functions are bounded for all γ and α. □

Proof of Theorem 3: To show convergence of the bootstrap test statistic to a Brownian bridge process, a result of Bickel and Ren (2001) is used. This result shows the convergence of a test statistic to a Brownian bridge process under the assumption that the estimated process (for example, some distribution function) converges weakly to its true value, and that the transformation applied to this process is suitably smooth. To this end, view the efficient test score as an operator S mapping the Banach space $\mathcal{F}_0$ of cumulative distribution functions to some other Banach space (here, $\mathbb{R}$). Assumptions A1 and A2 of Theorem 2.1 of Bickel and Ren (2001) are:

(a) The efficient test score (i.e., the semiparametrically efficient influence function of the test statistic) is Hadamard differentiable, i.e.,

$$\sup_G \left\| \frac{S(F+\delta G) - S(F)}{\delta} - \dot{S}(\delta G) \right\| = o(1)$$

for $\delta \to 0$ and with G as any element of any compact subset of $\mathcal{F}_0$. A definition of Hadamard differentiability and relationships to other differentiability concepts for normed spaces are given by Bickel et al. (1998, Appendix A.5), Fernholz (1983), Gill (1989), van der Vaart (1998, sec. 20), or van der Vaart and Wellner (1996, sec. 3.9)). This concept is frequently used in non- and semiparametric statistics, see for example Andersen et al. (1993, sec. II.8) or Gill (1994) for applications in survival analysis, and van der Vaart (1991), who considers the semiparametric efficiency of functionals of efficient estimators.

(b) $\sqrt{n}(\hat{F}_n - F) \Rightarrow \mathbb{G}$, where $\mathbb{G}$ is a Brownian bridge. That is, the estimator of the functional converges weakly to its true value.

The efficient test statistic contains the cumulative distribution function linearly and via a conditional expectation. The linear term is clearly differentiable, the second is by the dominated convergence theorem. The latter theorem is applicable as the cumulative distribution function is bounded and therefore a finite envelope function exists. For the dominated convergence theorem and differentiability under the integral, see Elstrodt (2004, pp. 145) or Pollard (2002, p. 32). Condition (b) was shown to hold in the proof of Theorem 1. □

Proof of Theorem 4: Following Tsiatis (2006, sec. 13.5), the mechanics of double robust estimators are shown first for $\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma)$. For the efficient estimator (18) of $F_{Y_1} - F_{Y_0}$ it holds that

$$\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma) \xrightarrow{p} E\left[\frac{D}{p(X)} 1\{Y \leq \gamma\} - \frac{D - p(X)}{p(X)} \mu(\gamma|1, X) \right.$$

$$- \frac{1-D}{1-p(X)} 1\{Y \leq \gamma\} + \frac{D-p(X)}{1-p(X)} \mu(\gamma|0,X) \Bigg].$$

By the fact that $DY = DY_1$ and $(1-D)Y = (1-D)Y_0$, this can be rewritten as

$$\begin{aligned}
&\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma) = \\
&\quad E\Bigg[1\{Y_1 \leq \gamma\} + \frac{D-p(X)}{p(X)}(1\{Y_1 \leq \gamma\} - \mu(\gamma|1,X)) \\
&\qquad -1\{Y_0 \leq \gamma\} + \frac{D-p(X)}{1-p(X)}(1\{Y_0 \leq \gamma\} - \mu(\gamma|0,X))\Bigg] + o_P(1).
\end{aligned}$$

By the law of the iterated expectation it follows that the second terms of each line are zero whenever the propensity score or the outcome equation is correctly specified, but not necessarily both. That is, inconsistent estimates of either the propensity score or the outcome equation do not lead to inconsistent estimates of the distribution function of the latent outcome, which shows double robustness of the efficient estimator of $F_{Y_1} - F_{Y_0}$.

Double robustness of the estimator for the subgroup of treated individuals can be shown using results of Hirano, Imbens, and Ridder (2003) on weighted average treatment effects. The estimator $\hat{F}^{\text{eff}}_{Y_1|D=1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0|D=1}(\gamma)$ is equal to the expression for $\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma)$ given above times $p(X)/p$, which converges in probability to $E[p(X)]/\ E[p]$. As $p(X)$ is a conditional expectation, $E[p(X)]$ converges to $E[p]$ by the law of iterated expectations, and double robustness of $\hat{F}^{\text{eff}}_{Y_1|D=1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0|D=1}(\gamma)$ follows by the derivations above for $\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma)$. □

3.6 References

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

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993): *Statistical Models Based on Counting Processes*. Springer, New York.

Andrews, D. W. K. (1994a): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62, 43–72.

——— (1994b): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2247–2294. Elsevier, Amsterdam.

Angrist, J. (2004): "Treatment Effect Heterogeneity in Theory and Practice," *Economic Journal*, 114, C52–C83.

Angrist, J., G. Imbens, and D. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–472, with discussion.

Athey, S., and G. W. Imbens (2006): "Identification and Inference in Nonlinear Difference-in-Difference Models," *Econometrica*, 74, 431–497.

Bang, H., and J. M. Robins (2005): "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61, 962–973.

Barrett, G. F., and S. G. Donald (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104.

Bickel, P. J., C. Klaassen, Y. Ritov, and J. Wellner (1998): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.

Bickel, P. J., and J.-J. Ren (2001): "The Bootstrap in Hypothesis Testing," in *State of the Art in Probability and Statistics. Festschrift for Willem R. van Zwet*, ed. by M. de Gunst, C. Klaasen, and A. van der Vaart, vol. 36 of *IMS Lecture Notes-Monograph Series*, pp. 91–112. Institute of Mathematical Statistics, Beachwood, Ohio.

Bickel, P. J., Y. Ritov, and T. M. Stoker (2006): "Tailor-Made Tests for Goodness of Fit to Semiparametric Hypotheses," *Annals of Statistics*, 34, 721–741.

Bitler, M. P., J. B. Gelbach, and H. H. Hoynes (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 69, 1127–1160.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75, 323–363.

Chernozhukov, V., and I. Fernandez-Val (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhya*, 67, 253–276.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

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

Cowell, F. A. (2000): "Measurement of Inequality," in *Handbook of Income Distribution*, ed. by A. B. Akinson, and F. Bourguignon, pp. 87–166. Elsevier, Amsterdam.

Davidson, R., and J.-Y. Duclos (2000): "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, 68, 1435–1464.

Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *Annals of Statistics*, 2, 267–277.

Elstrodt, J. (2004): *Maß- und Integrationstheorie*. Springer, Berlin, 4th edn.

Fernholz, L. T. (1983): *von Mises Calculus for Statistical Functionals*, vol. 19 of *Lecture Notes in Statistics*. Springer, New York.

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

——— (2007b): "Inequality Treatment Effects," mimeo, Pontifícia Universidade Católica do Rio de Janeiro.

Gill, R. D. (1989): "Non- and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 1)," *Scandinavian Journal of Statistics*, 16, 97–128.

——— (1994): "Lectures on Survival Analysis," in *Lectures on Probability Theory*, ed. by P. Bernard, vol. 1581 of *Lecture Notes in Mathematics*, pp. 115–241. Springer, Berlin.

Groeneboom, P., and J. A. Wellner (1992): *Information Bounds and Nonparametric Maximum Likelihood Estimation*, vol. 19 of *DMV-Seminar*. Birkhäuser, Basel.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing selection bias using experimental data," *Econometrica*, 66, 1017–1098.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Markets Programs," in *Handbook of Labor Economics*, ed. by A. Ashenfelter, and D. Card, vol. 3, pp. 1864–2097. Elsevier, Amsterdam.

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

Horváth, L., P. Kokoszka, and R. Zitikis (2006): "Testing for Stochastic Dominance Using the Weighted McFadden-type Statistic," *Journal of Econometrics*, 133, 191–205.

Horvitz, D. G., and D. J. Thompson (1952): "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685.

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

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Imbens, G. W., and D. B. Rubin (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555–574.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies*, 72, 735–765.

McFadden, D. (1989): "Testing for Stochastic Dominance," in *Studies in the Economics of Uncertainty*, ed. by T. B. Fomby, and T. K. Seo, pp. 113–134. Springer, New York.

Neugebauer, R., and M. van der Laan (2005): "Why Prefer Double Robust Estimators in Causal Inference?," *Journal of Statistical Planning and Inference*, 129, 405–426.

Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.

——— (1994): "Series Estimation of Regression Functionals," *Econometric Theory*, 10, 1–28.

——— (1995): "Convergence Rates for Series Estimatiors," in *Advances in Econometrics and Quantitative Economics*, ed. by G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, pp. 254–275. Blackwell, Oxford.

——— (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79, 147–168.

Pollard, D. (1984): *Convergence of Stochastic Processes*. Springer, New York.

——— (1990): *Empirical Processes: Theory and Applications*, vol. 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, California.

——— (2002): *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge.

Robins, J. M., and A. Rotnitzky (2001): "Comments on 'Inference for Semiparametric Models: Some Questions and an Answer'," *Statistica Sinica*, 11, 920–936.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994): "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89, 846–866.

Romano, J. P. (1988): "A Bootstrap Revival of Some Nonparametric Distance Tests," *Journal of the American Statistical Association*, 83, 698–708.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins (1999): "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models," *Journal of the American Statistical Association*, 94, 1096–1146, with discussion.

Shorack, G. R. (2000): *Probability for Statisticians*. Springer, New York.

Tan, Z. (2006a): "A Distributional Approach for Causal Inference Using Propensity Scores," *Journal of the American Statistical Association*, 101, 1619–1637.

——— (2006b): "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 101, 1607–1618.

Tsiatis, A. A. (2006): *Semiparametric Theory and Missing Data*. Springer, New York.

van de Geer, S. (1990): "Estimating a Regression Function," *Annals of Statistics*, 18, 907–924.

——— (1993): "Hellinger-Consistency of Certain Nonparametric Maximum Likelihood Estimators," *Annals of Statistics*, 21, 14–44.

——— (2000): *Empirical Processes in M-estimation*. Cambridge University Press, Cambridge.

van der Laan, M. J., and J. M. Robins (2003): *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York.

van der Vaart, A. (1991): "Efficiency and Hadamard Differentiability," *Scandinavian Journal of Statistics*, 18, 63–75.

——— (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

——— (2002): "Semiparametric Statistics," in *Lectures on Probability Theory and Statistics*, ed. by P. Bernard, no. 1781 in Lecture Notes in Mathematics, pp. 331–457. Springer, Berlin.

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

Complete References

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

——— (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263.

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

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.

Ahn, H. (1997): "Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors," *Econometric Theory*, 13, 3–31.

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993): *Statistical Models Based on Counting Processes*. Springer, New York.

Andrews, D. W. K. (1994a): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62, 43–72.

——— (1994b): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2247–2294. Elsevier, Amsterdam.

Angrist, J. (2004): "Treatment Effect Heterogeneity in Theory and Practice," *Economic Journal*, 114, C52–C83.

Angrist, J., K. Graddy, and G. Imbens (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67, 499–527.

Angrist, J., and G. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442.

Angrist, J., G. Imbens, and D. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–472, with discussion.

Athey, S., and G. W. Imbens (2006): "Identification and Inference in Nonlinear Difference-in-Difference Models," *Econometrica*, 74, 431–497.

Bang, H., and J. M. Robins (2005): "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61, 962–973.

Bang, H., and A. A. Tsiatis (2002): "Median Regression with Censored Cost Data," *Biometrics*, 58, 643–649.

Barrett, G. F., and S. G. Donald (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71–104.

Bickel, P. J., C. Klaassen, Y. Ritov, and J. Wellner (1998): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.

Bickel, P. J., and J.-J. Ren (2001): "The Bootstrap in Hypothesis Testing," in *State of the Art in Probability and Statistics. Festschrift for Willem R. van Zwet*, ed. by M. de Gunst, C. Klaasen, and A. van der Vaart, vol. 36 of *IMS Lecture Notes-Monograph Series*, pp. 91–112. Institute of Mathematical Statistics, Beachwood, Ohio.

Bickel, P. J., Y. Ritov, and T. M. Stoker (2006): "Tailor-Made Tests for Goodness of Fit to Semiparametric Hypotheses," *Annals of Statistics*, 34, 721–741.

Bijwaard, G. E. (2001): "Instrumental Variable Estimation for Duration Data: A Reappraisal of the Illinois Reemployment Bonus Experiment," Econometric Institute Report EI 2002-39, Erasmus University Rotterdam.

Bijwaard, G. E., and G. Ridder (2005): "Correcting for Selective Compliance in a Re-Employment Bonus Experiment," *Journal of Econometrics*, 125, 77–111.

Bitler, M. P., J. B. Gelbach, and H. H. Hoynes (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 69, 1127–1160.

Björklund, A., and R. Moffitt (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69, 42–49.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75, 323–363.

Buchinsky, M., and J. Hahn (1998): "An Alternative Model for the Censored Quantile Regression Estimnator," *Econometrica*, 66, 653–671.

Carneiro, P., J. J. Heckman, and E. Vytlacil (2005): "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education," Mimeo, University of Chicago.

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.

Chernozhukov, V., and H. Hong (2002): "Three-Step Censored Quantile Regression and Extramarital Affairs," *Journal of the American Statistical Association*, 97, 872–882.

Chesher, A. (2002): "Semiparametric Identification in Duration Models," Cemmap Working Paper 20/02.

Cowell, F. A. (2000): "Measurement of Inequality," in *Handbook of Income Distribution*, ed. by A. B. Akinson, and F. Bourguignon, pp. 87–166. Elsevier, Amsterdam.

Cunha, F., J. J. Heckman, and S. Navarro (2007): "The Identification and Economic Content of Ordered Choice Models with Stochastic Tresholds," NBER Technical Working Paper 340.

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

Davidson, R., and J.-Y. Duclos (2000): "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, 68, 1435–1464.

Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *Annals of Statistics*, 2, 267–277.

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

Elstrodt, J. (2004): *Maß- und Integrationstheorie*. Springer, Berlin, 4th edn.

Fan, J., and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Fernholz, L. T. (1983): *von Mises Calculus for Statistical Functionals*, vol. 19 of *Lecture Notes in Statistics*. Springer, New York.

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

——— (2007b): "Inequality Treatment Effects," mimeo, Pontifícia Universidade Católica do Rio de Janeiro.

Fitzenberger, B. (1997): "A Guide to Censored Quantile Regressions," in *Handbook of Statistics*, ed. by G. S. Maddala, and C. R. Rao, vol. 15, pp. 405–437. Elsevier, Amsterdam.

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

Fitzenberger, B., and P. Winker (2007): "Improving the Computation of Censored Quantile Regressions," *Computational Statistics & Data Analysis*, 52, 88–108.

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

——— (2007): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 139, 35–75.

Gill, R. D. (1989): "Non- and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 1)," *Scandinavian Journal of Statistics*, 16, 97–128.

——— (1994): "Lectures on Survival Analysis," in *Lectures on Probability Theory*, ed. by P. Bernard, vol. 1581 of *Lecture Notes in Mathematics*, pp. 115–241. Springer, Berlin.

Groeneboom, P., and J. A. Wellner (1992): *Information Bounds and Nonparametric Maximum Likelihood Estimation*, vol. 19 of *DMV-Seminar*. Birkhäuser, Basel.

Hahn, J. (1994): "The Efficiency Bound of the Mixed Proportional Hazard Model," *Review of Economic Studies*, 61, 607–629.

——— (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331.

Ham, J. C., and R. J. LaLonde (1996): "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, 64, 175–205.

Heckman, J. J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441–462, with addendum (vol. 33, 247) and discussion (vol. 34, 823–837).

——— (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 109, 673–748.

——— (2005): "The Scientific Model of Causality," *Sociological Methodology*, 35, 1–162, with discussion.

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing selection bias using experimental data," *Econometrica*, 66, 1017–1098.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Markets Programs," in *Handbook of Labor Economics*, ed. by A. Ashenfelter, and D. Card, vol. 3, pp. 1864–2097. Elsevier, Amsterdam.

Heckman, J. J., and S. Navarro (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136, 341–396.

Heckman, J. J., and B. Singer (1984): "Econometric Duration Analysis," *Journal of Econometrics*, 24, 63–132, Correction: Vol. 27 (1985), 137-138.

Heckman, J. J., and J. Smith (1998): "Evaluating the Welfare State," in *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, ed. by S. Strøm, pp. 241–318. Cambridge University Press, Cambridge.

Heckman, J. J., S. Urzua, and E. J. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88, 389–432.

Heckman, J. J., and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734.

——— (2000): "The Relationship Between Treatment Parameters Within a Latent Variable Framework," *Economics Letters*, 66, 33–39.

——— (2001a): "Local Instrumental Variables," in *Nonlinear Statistical Modelling*, ed. by C. Hsiao, K. Morimune, and J. J. Powell, pp. 1–46. Cambridge University Press, Cambridge.

——— (2001b): "Policy-Relevant Treatment Effects," *American Economic Review, Papers and Proceedings*, 91, 107–111.

——— (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73, 669–738.

——— (2007): "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," mimeo, Columbia University.

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

Hong, H., and D. Nekipelov (2007): "Semiparametric Efficiency in Nonlinear LATE Models," mimeo, Stanford University.

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.

Horváth, L., P. Kokoszka, and R. Zitikis (2006): "Testing for Stochastic Dominance Using the Weighted McFadden-type Statistic," *Journal of Econometrics*, 133, 191–205.

Horvitz, D. G., and D. J. Thompson (1952): "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685.

Huang, J., S. Ma, and H. Xie (2005): "Least Absolute Deviations Estimation for the Accelerated Failure Time Model," Technical Report No. 350, Department of Statistics and Actuarial Science, University of Iowa.

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.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Imbens, G. W., and D. B. Rubin (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555–574.

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.

Koroljuk, V. S., and Y. V. Borovskich (1994): *Theory of U-Statistics*. Kluwer, Dordrecht.

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.

Lee, A. J. (1990): *U-Statistics: Theory and Practice*. Marcel Dekker, New York.

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

Li, Q., and J. M. Wooldridge (2002): "Semiparametric Estimation of Partially Linear Models for Dependent Data with Generated Regressors," *Econometric Theory*, 18, 625–645.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies*, 72, 735–765.

Mammen, E., O. Linton, and J. P. Nielsen (1999): "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm under Weak Conditions," *Annals of Statistics*, 27, 1443–1490.

McFadden, D. (1989): "Testing for Stochastic Dominance," in *Studies in the Economics of Uncertainty*, ed. by T. B. Fomby, and T. K. Seo, pp. 113–134. Springer, New York.

Neugebauer, R., and M. van der Laan (2005): "Why Prefer Double Robust Estimators in Causal Inference?," *Journal of Statistical Planning and Inference*, 129, 405–426.

Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.

——— (1994): "Series Estimation of Regression Functionals," *Econometric Theory*, 10, 1–28.

——— (1995): "Convergence Rates for Series Estimatiors," in *Advances in Econometrics and Quantitative Economics*, ed. by G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, pp. 254–275. Blackwell, Oxford.

——— (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79, 147–168.

Newey, W. K., and D. L. McFadden (1995): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2111–2245. Elsevier, Amsterdam.

Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247.

Pollard, D. (1984): *Convergence of Stochastic Processes*. Springer, New York.

——— (1990): *Empirical Processes: Theory and Applications*, vol. 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics, Hayward, California.

——— (2002): *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge.

Portnoy, S. (2003): "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012, Correction by T. Neocleous, K. Vanden Branden and S. Portnoy, vol. 101 (2006), 860-861.

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.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2007): *Numerical Recipes*. Cambridge University Press, Cambridge, 3rd edn.

Quarteroni, A., R. Sacco, and F. Saleri (2007): *Numerical Mathematics*. Springer, Berlin, 2nd edn.

Ridder, G. (1990): "The Non-Parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57, 167–182.

Rilstone, P. (1996): "Nonparametric Estimation of Models with Generated Regressors," *International Economic Review*, 37, 299–313.

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

Robins, J. M., and A. Rotnitzky (2001): "Comments on 'Inference for Semiparametric Models: Some Questions and an Answer'," *Statistica Sinica*, 11, 920–936.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994): "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89, 846–866.

Romano, J. P. (1988): "A Bootstrap Revival of Some Nonparametric Distance Tests," *Journal of the American Statistical Association*, 83, 698–708.

Ruppert, D., and M. P. Wand (1994): "Multivariate Locally Weighted Least Squares Regression," *Annals of Statistics*, 22, 1346–1370.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins (1999): "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models," *Journal of the American Statistical Association*, 94, 1096–1146, with discussion.

Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shorack, G. R. (2000): *Probability for Statisticians*. Springer, New York.

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

Sperlich, S. (2007): "A Note on Nonparametric Estimation with Constructed Variables and Generated Regressors," mimeo, University of Göttingen.

Tan, Z. (2006a): "A Distributional Approach for Causal Inference Using Propensity Scores," *Journal of the American Statistical Association*, 101, 1619–1637.

——— (2006b): "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 101, 1607–1618.

Tsiatis, A. A. (2006): *Semiparametric Theory and Missing Data*. Springer, New York.

van de Geer, S. (1990): "Estimating a Regression Function," *Annals of Statistics*, 18, 907–924.

——— (1993): "Hellinger-Consistency of Certain Nonparametric Maximum Likelihood Estimators," *Annals of Statistics*, 21, 14–44.

——— (2000): *Empirical Processes in M-estimation*. Cambridge University Press, Cambridge.

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 Laan, M. J., and J. M. Robins (2003): *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York.

van der Vaart, A. (1991): "Efficiency and Hadamard Differentiability," *Scandinavian Journal of Statistics*, 18, 63–75.

——— (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

——— (2002): "Semiparametric Statistics," in *Lectures on Probability Theory and Statistics*, ed. by P. Bernard, no. 1781 in Lecture Notes in Mathematics, pp. 331–457. Springer, Berlin.

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.

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.

——— (2006): "Ordered Discrete-Choice Selection Models and Local Average treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results," *Review of Economics and Statistics*, 88, 578–581.

Vytlacil, E., and N. Yildiz (2007): "Dummy Endogenous Variables in Weakly Separable Models," *Econometrica*, 75, 757–779.

Wand, M. P., and M. C. Jones (1995): *Kernel Smoothing*. Chapman & Hall, London.

Yu, K., B. U. Park, and E. Mammen (forthcoming): "Smooth Backfitting in Generalized Additive Models," *Annals of Statistics*.

Zheng, J. X. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–289.

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, insbesondere Promotionsberater, waren an der inhaltlich materiellen Erstellung dieser Arbeit nicht beteiligt. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Mannheim, 15.11.2007

(Michael Maier)

Abgrenzung

Ich erkläre hiermit, dass ich die Abschnitte 1, 2 und 3 der vorliegenden Arbeit ohne Koautoren geschrieben habe.

Mannheim, 15.11.2007

(Michael Maier)