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

Michael Maier*
ZEW Mannheim

February 14, 2007

Abstract

This note 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.

JEL-Classification: C14, C15, C21

Keywords: Econometric evaluation, conditional independence assumption, distributional treatment effect, Kolmogorov-Smirnov test, semiparametric efficiency, double robust estimation

*Contact address: Centre for European Economic Research, L 7,1, 68161 Mannheim, Germany. Phone +49-621-1235-362, Fax +49-621-1235-225, e-mail: maier@zew.de.

1 Introduction

This note 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), Imbens (2004), Tan (2006 a, b), or Heckman, LaLonde and Smith (1999)). They may be classified whether they are based on 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 (2007) 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. 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 (2005) 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). To derive bounds on differences between wage distributions, Blundell et al. (forthcoming) implement the restriction of positive selection into work by assuming stochastic dominance of the wage distribution of non-workers by that of workers. Recent contributions to the topic of stochastic dominance include Barrett and Donald (2003), Davidson and Duclos (2000), Horváth, Kokoszka and Zitikis (2006) and Linton, Maasoumi and Whang (2005), for example.

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. This approximation is plugged into the estimators of the latent distribution functions, which therefore are semiparametric estimators. 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 statistic is 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 2 describes the framework and the assumptions of the tests and the semiparametric estimators of the test statistics. Section 3 derives a semiparametric efficient test statistic and states its double robustness property. Resampling methods to obtain critical values of the test statistics are described in sections 2 and 3. All proofs are given in the appendix.

2 Tests for Distributional Treatment Effects

Consider the effect of a binary treatment D on an outcome Y. It is assumed that for every individual and for each value of the treatment D 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 note 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}, respectivly.

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:

$$F_{Y_1}(\gamma) = F_{Y_0}(\gamma) \;\; \forall \gamma \in \mathbb{R}, \tag{1}$$

$$F_{Y_1}(\gamma) \leqslant F_{Y_0}(\gamma) \;\; \forall \gamma \in \mathbb{R}, \tag{2}$$

$$\int_{-\infty}^{\gamma} F_{Y_1}(x)dx \leqslant \int_{-\infty}^{\gamma} F_{Y_0}(x)dx \;\; \forall \gamma \in \mathbb{R}. \tag{3}$$

(1) is the null hypothesis for equality of distributions, (2) and (3) are those for first and second order stochastic dominance, respectively.

To implement the tests, estimators of the latent outcome distributions $F_{Y_1}(\cdot)$ and $F_{Y_0}(\cdot)$ are constructed using results of Firpo (2005). 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). This 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 $F_A(a) = P(A \leqslant a) = E[1\{A \leqslant a\}]$, Firpo (2005) reweights the indicator function $1\{Y < \gamma\}$ to get an estimator of the distributions of the latent outcomes of treated and untreated. This approach was also suggestested 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 some covariates 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 every x in the support of X.
- A3 $E[1\{Y_j \leqslant \gamma\}|X]$ is continuously differentiable for all elements of the supports of X and Y ($j \in \{0,1\}$).
- A4 The propensity score $p(X) \equiv E[D|X]$ is continuously differentiable of order $s \geqslant 7r$ (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 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. A5 is a restiction of 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 (2005). Here, a slightly different representation is chosen, with n_1 and n_0 as the number of observations with $D=1$ and $D=0$, respectively:

$$\hat{F}_{Y_1}(\gamma) - \hat{F}_{Y_0}(\gamma) = \\ = \frac{1}{n_1}\sum_{i=1}^{n}\frac{D_i}{\hat{p}(X_i)}1\{Y_i \leqslant \gamma\} - \frac{1}{n_0}\sum_{i=1}^{n}\frac{1-D_i}{1-\hat{p}(X_i)}1\{Y_i \leqslant \gamma\}. \quad (4)$$

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:

$$\hat{F}_{Y_1|T=1}(\gamma) - \hat{F}_{Y_0|T=1}(\gamma) = \\ = \frac{1}{n_1}\sum_{i=1}^{n}\frac{D_i}{p}1\{Y_i \leqslant \gamma\} - \frac{1}{n_0}\sum_{i=1}^{n}\frac{1-D_i}{p}\frac{\hat{p}(X_i)}{1-\hat{p}(X_i)}1\{Y_i \leqslant \gamma\}. \quad (5)$$

The further exposure will mainly consider the effects for a randomly chosen individual (i.e., it will be based on estimator (4)). All results are also valid for estimator (5).

Pointwise consistency of $\hat{F}_{Y_1}(\cdot)$ and $\hat{F}_{Y_0}(\cdot)$ 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 (1) - (3) 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:

$$T_n^{eq} = \sqrt{\frac{n_1 n_0}{n}} \sup_{\gamma \in \mathbb{R}} \left| \hat{F}_{Y_1}(\gamma) - \hat{F}_{Y_0}(\gamma) \right|, \tag{6}$$

$$T_n^{fsd} = \sqrt{\frac{n_1 n_0}{n}} \sup_{\gamma \in \mathbb{R}} \left(\hat{F}_{Y_1}(\gamma) - \hat{F}_{Y_0}(\gamma) \right), \tag{7}$$

$$T_n^{ssd} = \sqrt{\frac{n_1 n_0}{n}} \sup_{\gamma \in \mathbb{R}} \int_{-\infty}^{\gamma} \left(\hat{F}_{Y_1}(x) - \hat{F}_{Y_0}(x) \right) dx. \tag{8}$$

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 (6) - (8) (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. Proposition 1 notes that the procedure just described shares these features.

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

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 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. These concepts 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 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

$$Z_n(\gamma, \alpha) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Pi^{\perp}(h(\gamma, \alpha), \alpha)(X_i). \tag{9}$$

Here, $h(\gamma, \alpha)$ is an element of the tangent space of the test statistic, which may depend on the nuisance parameter α. 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 $\mu(\gamma|j,X) \equiv E[1\{Y \leqslant \gamma\}|D=j,X]$. 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:

$$\begin{aligned}
&\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma) = \\
&= \frac{1}{n_1}\sum_{i=1}^{n}\left(\frac{D_i}{\hat{p}(X_i)}1\{Y_i \leqslant \gamma\} - \frac{D_i - \hat{p}(X_i)}{\hat{p}(X_i)}\hat{\mu}(\gamma|1,X)\right) \\
&\quad -\frac{1}{n_0}\sum_{i=1}^{n}\left(\frac{1-D_i}{1-\hat{p}(X_i)}1\{Y_i \leqslant \gamma\} - \frac{D_i - \hat{p}(X_i)}{1-\hat{p}(X_i)}\hat{\mu}(\gamma|0,X)\right). \qquad (10)
\end{aligned}$$

Comparing (10) with (4), 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:

$$\begin{aligned}
&\hat{F}^{\text{eff}}_{Y_1|T=1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0|T=1}(\gamma) \\
&= \frac{1}{n_1}\sum_{i=1}^{n}\left(\frac{D_i}{p}(1\{Y_i \leqslant \gamma\} - \hat{\mu}(\gamma|1,X))\right) \\
&\quad -\frac{1}{n_0}\sum_{i=1}^{n}\left(\frac{1-D_i}{p}\frac{\hat{p}(X_i)}{1-\hat{p}(X_i)}(1\{Y_i \leqslant \gamma\} - \hat{\mu}(\gamma|0,X))\right) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{p}(X_i)}{p}(\hat{\mu}(\gamma|1,X) - \hat{\mu}(\gamma|0,X)). \qquad (11)
\end{aligned}$$

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

$$\Pi^{\perp}(h(\gamma,\hat{\alpha}),\hat{\alpha}) = \hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma). \qquad (12)$$

An estimator of the score process (9) is equal to (12) 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 (6) - (8) 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 proposition.

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

$$Z_n(\gamma, \hat{\alpha}) = \sqrt{n}\left(\hat{F}^{\mathit{eff}}_{Y_1}(\gamma) - \hat{F}^{\mathit{eff}}_{Y_0}(\gamma)\right). \tag{13}$$

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

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 eq. (6) - (8).
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.
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 proposition.

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

The next proposition 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 may show 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, labor market experience, for example. 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 estimators seem to be less variable when estimates of the propensity score near zero or one occur.

Proposition 4: *The semiparametric efficient estimators* $\hat{F}^{eff}_{Y_j}(\gamma)$ *and* $\hat{F}^{eff}_{Y_j|T=1}(\gamma)$ $(j \in \{0,1\})$ *remain consistent under misspecification of either the propensity score or the outcome equation.*

4 Conclusion

This note 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.

One point has to be kept in mind when interpreting estimates of distributional or quantile treatment effects, however. Contrary to average treatment effects, distributional or quantile treatment effects identify causal effects only under the assumption of rank invariance. In the context of this note, this means that individuals take the same rank in the outcome distribution in both treatment states. This is a strong assumption (see the discussion in Chernozhukov and Hansen (2005) and Firpo (2007)). Without this assumption, the effect can be interpreted as difference of marginal distributions in both treatment states.

A Proofs

Proof of Proposition 1: It will be shown that the derivations of Abadie (2002) are also valid for the procedure described in section 2. The first step is to show that the estimator of the difference of the latent outcome distribution functions converge to some Gaussian process. To this end, it is shown that $\sqrt{n_1}(\hat{F}_{Y_1} - F_{Y_1}) \Rightarrow \mathbb{G}_{F_{Y_1}}$ and $\sqrt{n_0}(\hat{F}_{Y_0} - F_{Y_0}) \Rightarrow \mathbb{G}_{F_{Y_0}}$. 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 (2000), 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 $(0,1)$ are Donsker classes (see examples 2.5.4 (p. 129) and 2.6.21 (p. 149) of van der Vaart and Wellner (2000)). The set $\{1/f | f \in \mathcal{F},\ \mathcal{F}$ 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 (2000)). These results, together with a permanence property of Donsker classes (example 2.10.8 (p. 192) of van der Vaart and Wellner (2000)) 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

$$\begin{aligned} &\sqrt{\frac{n_1 n_0}{n}} \left(\hat{F}_{Y_1} - \hat{F}_{Y_0} \right) \\ &= \sqrt{\frac{n_1 n_0}{n}} \left(\hat{F}_{Y_1} - F_{Y_1} - \left(\hat{F}_{Y_0} - F_{Y_0} \right) + (F_{Y_1} - F_{Y_0}) \right). \end{aligned}$$

Under the null hypothesis $F_{Y_1} = F_{Y_0}$ $(\equiv F)$, and by $n_1/n \to \pi \in (0,1)$ (which can be deduced from the common support assumption A4), weak convergence of the following expression results:

$$\sqrt{\frac{n_0}{n}} \sqrt{n_1} \left(\hat{F}_{Y_1} - F_{Y_1} \right) - \sqrt{\frac{n_1}{n}} \sqrt{n_0} \left(\hat{F}_{Y_0} - F_{Y_0} \right) \Rightarrow \sqrt{1-\pi}\,\mathbb{G}_F - \sqrt{\pi}\,\mathbb{G}'_F,$$

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.

Following again Abadie (2002), convergence of the bootstrap test statistics to some Gaussian process will be shown. By Theorem 3.7.6 (p. 365) of van der Vaart and Wellner (2000), this follows by the Donsker property of the estimators. Abadie (2002) 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. Therefore, the derivations of Abadie (2002) apply also to the tests of this note. In addition, consistency against any fixed alternative follows.

Power against contiguous alternatives can also be shown analogous to Abadies (2002) derivations. Let Q be the distribution of D conditional on X. Define $M = QP$. Rewrite the difference of cdfs as

$$\begin{aligned}
D_n(f) &= \sqrt{\frac{n_0 n_1}{n}} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} f(Y_i) - \frac{1}{n_0} \sum_{i=n_1+1}^{n} f(Y_i) \right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\sqrt{\frac{n_0}{n_1}} D_i - \sqrt{\frac{n_1}{n_0}} (1 - D_i) \right) (f(Y_i) - Pf(Y_i)) \\
&\to \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\sqrt{\frac{1-\pi}{\pi}} D_i - \sqrt{\frac{\pi}{1-\pi}} (1 - D_i) \right) (f(Y_i) - Pf(Y_i)).
\end{aligned}$$

Let $M_n = QP_n$ be local alternatives and P_n conditional probability measures, where $P_n = P_{d,n}$ for $D = d$. Assume that P_n converges in mean square for $d = 0, 1$, i.e.:

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

From this, convergence of M_n follows (with $x = D_i x_1(Y_i) + (1 - D_i) x_0(Y_i)$):

$$\int \left(\sqrt{n}(dM_n^{1/2} - dM^{1/2}) - \frac{1}{2} x dM^{1/2} \right)^2 \to 0$$

M_n^n and M^n are contiguous and the sequence of log likelihood ratios has a linear expansion. $D_n(f)$ and $\log \frac{M_n^n}{M^n}$ converge under M^n in distribution to a multivariate normal variable. Further, $D_n(f)$ converges under M_n^n in distribution to a multivariate normal variable. By the Donsker property of $\{f\}$, it can be shown that D_n converges weakly to a Gaussian process plus a mean zero random variable. From this, power against contiguous alternatives follows. Therefore, Algorithm 1 of this note shares the asymptotic properties of the method of Abadie (2002). □

Proof of Proposition 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. Like the proof of Proposition 1, the following derivations rely primarily on empirical process methods.

M0 It has to be shown that $\{\Pi^{\perp}(h_\gamma(\cdot,\alpha_0),\alpha_0)|\gamma\in\Gamma\}$ 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 (2000, p. 82)). In the present application, this set is given by:

$$\begin{aligned}\left\{\Pi^{\perp}(h_\gamma(\cdot,\alpha_0),\alpha_0)\middle|\,\gamma\in\Gamma\right\} &= \left\{F^{\text{eff}}_{Y_1}(\gamma)-F^{\text{eff}}_{Y_0}(\gamma)\middle|\,\gamma\in\Gamma\right\}\\ = \Bigg\{&\frac{D}{p(X)}1\{Y\leqslant\gamma\}-\frac{D-p(X)}{p(X)}\mu(\gamma|1,X)-\\ &\left(\frac{1-D}{1-p(X)}1\{Y\leqslant\gamma\}-\frac{D-p(X)}{1-p(X)}\mu(\gamma|0,X)\right)\Bigg|\,\gamma\in\Gamma\Bigg\}.\end{aligned}$$

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 (2000)).

M1 $\|(P_n-P_0)(\Pi(h_\gamma(\cdot,\hat\alpha),\hat\alpha)-\Pi(h_\gamma(\cdot,\hat\alpha),\alpha_0))\|_\infty=o_P(n^{-1/2})$. In the context of the present application, this is equivalent to

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

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 are 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.,

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

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:

$$\int_{f_0>0}\left(\sqrt{\frac{f_{\hat{\alpha}}}{f_{\alpha_0}}} - 1\right) d(P_n - P_0) \geqslant H^2(f_{\hat{\alpha}}, f_{\alpha_0}). \tag{14}$$

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 (14), if the following holds:

$$\sup_{g\in\mathcal{G}}\left|\int gd(P_N - P_0)\right| \rightarrow 0 \text{ a.s.} \tag{15}$$

Validity of (15) 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 (logarithm of the δ-covering number) of $\mathcal{G}_0$. If the following three conditions hold, $H^2(\hat{f}_{\hat{\alpha}}, f_{\alpha_0}) \rightarrow 0$ a.s. by Theorem 3.1 of van de Geer (1993).

(a) $\mathcal{G}_0$ is uniformly bounded,

(b) $\sqrt{f_0} < \infty$,

(c) $n^{-1}\mathcal{H}(\delta, \mathcal{G}_0, \|\cdot\|_{P_n,\infty}) \rightarrow 0$.

By assumption A2, conditions (a) and (b) are satisfied. 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

$$\begin{aligned} \left\| \frac{(\hat{s} - s_0)^2}{\hat{s}} \right\|_{2,\mu}^2 &= \int \frac{(\hat{s} - s_0)^4}{\hat{s}^2} \, d\mu = o_P(n^{-1}) \\ \Leftrightarrow \ n \int \frac{(\hat{s} - s_0)^4}{\hat{s}^2} \, d\mu &= o_P(1) \\ \Leftrightarrow \ \int \left(\sqrt{n}(\hat{s} - s_0)\right)^2 (\hat{s} - s_0)^2 \frac{1}{\hat{s}^2} \, d\mu &= o_P(1). \quad (16) \end{aligned}$$

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 $1/4h^T I_\alpha h < \infty$, where I_α is the Fisher information matrix and h is a vector. The first and third factors of (16) are 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 $s(\alpha)$.

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 Proposition 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 some 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 score of the test 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 score (i.e., the semiparametrically efficient influence function) of the test 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 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), van der Vaart and Wellner (2000, sec. 3.9) or any monograph of functional analysis (for example, Werner (2005, sec. III.5)). 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, van der Vaart (1991), who considers the semiparametric efficiency of functionals of efficient estimators, and Freitag and Munk (2005), who consider goodness of fit tests for the k-sample problem with respect to structural relationships between the samples.

(b) $\sqrt{n}(\hat{F}_n - F) \Rightarrow Z$, where Z 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 an conditional expectation. The first relationship is clearly differentiable, the second by the dominated convergence theorem. The latter holds as the cdf is bounded and therefore a finite envelope function exists. For the dominated convergence theorem and differentiability under the integral see Elstrodt (2004, pp. 147) or Pollard (2002, p. 32). Condition (b) was already shown to hold in the proof of Proposition 1. □

Proof of Proposition 4: Following Tsiatis (2006, sec. 13.5), the mechanics of double robust estimators are shown first for $\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma)$.

The efficient estimator (10) of $F_{Y_1} - F_{Y_0}$ converges in probability to

$$\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma) \rightarrow E\left[\frac{D}{p(X)}1\{Y \leqslant \gamma\} - \frac{D-p(X)}{p(X)}\mu(\gamma|1,X) \right. \\ \left. - \frac{1-D}{1-p(X)}1\{Y \leqslant \gamma\} + \frac{D-p(X)}{1-p(X)}\mu(\gamma|0,X)\right].$$

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

$$\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma) = \\ E\left[1\{Y_1 \leqslant \gamma\} + \frac{D-p(X)}{p(X)}(1\{Y_1 \leqslant \gamma\} - \mu(\gamma|1,X)) \right. \\ \left. -1\{Y_0 \leqslant \gamma\} + \frac{D-p(X)}{1-p(X)}(1\{Y_0 \leqslant \gamma\} - \mu(\gamma|0,X))\right] + o_P(1).$$

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 more general weighted average treatment effects. The estimator $\hat{F}^{\text{eff}}_{Y_1|T=1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0|T=1}(\gamma)$ is equal to the expression for $\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma)$ 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 $\hat{F}^{\text{eff}}_{Y_1|T=1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0|T=1}(\gamma)$ follows by the derivations above for $\hat{F}^{\text{eff}}_{Y_1}(\gamma) - \hat{F}^{\text{eff}}_{Y_0}(\gamma)$. □

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. Klassen, 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 (forthcoming): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*.

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.

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. (2005): "Inequality Treatment Effects," mimeo, University of British Columbia.

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

Freitag, G., and A. Munk (2005): "On Hadamard Differentiability in k-Sample Semiparametric Models – with Applications to the Assessment of Structural Relationships," *Journal of Multivariate Analysis*, 94, 123–158.

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.

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, vol. 1781 of *Lecture Notes in Mathematics*, pp. 331–457. Springer, Berlin.

van der Vaart, A., and J. Wellner (2000): *Weak Convergence and Empirical Processes*. Springer, New York.

Werner, D. (2005): *Funktionalanalysis*. Springer, Berlin, 5th edn.