Journal of Econometrics 178 (2014) 383–397

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Estimation and inference for distribution functions and quantile functions in treatment effect models[☆]

Stephen G. Donald[a], Yu-Chin Hsu[b,*]

[a] Department of Economics, University of Texas at Austin, United States

[b] Institute of Economics, Academia Sinica, Taiwan

ARTICLE INFO

Article history:
Received 16 June 2011
Received in revised form
4 October 2012
Accepted 21 March 2013
Available online 2 September 2013

JEL classification:
C01
C12
C21

Keywords:
Hypothesis testing
Stochastic dominance
Treatment effects
Propensity score

ABSTRACT

We propose inverse probability weighted estimators for the distribution functions of the potential outcomes under the unconfoundedness assumption and apply the inverse mapping to obtain the quantile functions. We show that these estimators converge weakly to zero mean Gaussian processes. A simulation method is proposed to approximate these limiting processes. Based on these results, we construct tests for stochastic dominance relations between the potential outcomes. Monte-Carlo simulations are conducted to examine the finite sample properties of our tests. We apply our test in an empirical example and find that a job training program had a positive effect on incomes.

1. Introduction

This paper is related to numerous papers in the treatment effects literature. Much of the literature to date has focused on examining the impact of a treatment by examining the effect of this treatment on certain features of the distributions of the treatment and control groups. Most papers focus on the mean treatment effects which are defined as the average treatment effect (ATE) and average treatment effect on the treated (ATT) (e.g., Rosenbaum and Rubin (1983, 1985), Heckman et al. (1997, 1998a,b), Hahn (1988), Rosenbaum (1987), Rubin and Thomas (1996) and Hirano et al. (2003, HIR hereafter)). More recently Firpo (2007a) has examined quantile treatment effects (QTE) and quantile treatment effects on the treated (QTT) which measure the difference between certain quantiles of the potential outcomes. Also, Firpo (2010) proposes estimators of the treatment effect on various measures of inequality. These include measures such as the coefficient of variation, interquartile range, Theil's index and the Gini coefficient. For recent reviews, please see Imbens (2004) and Imbens and Wooldridge (2009) among others.

In this paper we propose methods for estimating the entire distributions (CDF) of potential outcomes in a binary treatment effect model where the unconfoundedness assumption is satisfied. The estimation method is the inverse probability weighted (IPW) estimator of the CDF for potential outcomes and allows for the use of a nonparametric estimator of the propensity score. We also use these estimates and the inverse mapping to estimate the quantile processes associated with the distributions. All our estimators are treated as functions and our asymptotic analysis of the estimators shows that, under regularity assumptions, the estimators converge weakly to Gaussian stochastic processes at the usual parametric rate. Since it is well known that the covariance kernel of the associated Gaussian processes depend on the estimation error in the estimated propensity score, we propose simulation methods, based on the multiplier central limit theorem, for conducting inference that take this into account. We also propose estimators for the CDFs and quantile functions of the potential outcomes in the *treated* subpopulation and discuss simulation methods for conducting inference. To the best of our knowledge, this is the first paper to estimate the CDFs and the whole quantile functions of the potential outcomes under the unconfoundedness assumptions and to propose a method to approximate the limiting processes of the estimators. To demonstrate the usefulness of our results we propose Kolmogorov–Smirnov (KS) type tests for stochastic dominance relations between the distributions of potential outcomes. The importance of such tests has been

[☆] A previous version was circulated under the title "Testing for Stochastic Dominance in Treatment Effects".

* Correspondence to: 128 Academia Road, Section 2, Nankang, Taipei, 115, Taiwan. Tel.: +886 2 27822791x322; fax: +886 2 2785 3946.
E-mail address: ychsu@econ.sinica.edu.tw (Y.-C. Hsu).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.03.010

discussed by Imbens and Wooldridge (2009). Our paper also examines the properties of the proposed methods in finite samples by conducting a small scale simulation as well as applying the methods for the KS tests to examine the effect of a job training program in an empirical example using data from the National Supported Work Demonstration (NSW) job training program. In the empirical example we find that real earnings under job training stochastically dominates real earnings without job training and we find evidence against the opposite relationship. Thus our methods suggest that positive effects of job training are felt over the entire distribution which is a stronger conclusion than one would reach based simply on the mean.

This paper is related to some other recent papers that examine distributional effects in treatment effect models. Abadie (2002) constructs tests for stochastic dominance relationship between potential outcomes when the treatment assignment is endogenous and one has a binary instrument. Maier (2011) focuses on KS tests for the stochastic dominance relation in a treatment effect situation similar to ours but uses a different method for inference based on the bootstrap. Lee and Whang (2009), and Hsu (2011) consider the null hypothesis that the stochastic dominance relation between the potential outcomes holds conditional on the all possible values of the covariances. Shen (2011) proposes nonparametric and semiparametric tests for conditional stochastic dominance relation between the potential outcomes at a given covariate value. Their focus is different from ours, because we focus on the stochastic dominance relation between the unconditional CDFs. As noted above Firpo (2007a) examines quantile treatment effects for specific quantiles while Firpo (2010) examines specific measures of inequality. Our paper provides results for the entire CDF and quantile process and can be used in principle, along with functional delta methods, to construct any functional of either and conduct inference with KS tests for stochastic dominance being one empirically important example. Additionally our paper provides results for the treated subpopulation.

This paper is also related to numerous papers in the stochastic dominance literature. Papers that consider inference for stochastic dominance between two distributions include McFadden (1989), Anderson (1996), Davidson and Duclos (2000), Barrett and Donald (2003), Linton et al. (2005), Bennett (2009), Linton et al. (2010), Donald and Hsu (2010) and Donald et al. (2012) among others. Our paper considers the empirically important case where one comparing two distributions in the treatment effect context where selection to treatment depends on observables and does not rely on completely random assignment as would be required in order to use results of Barrett and Donald's (2003).[1]

The remainder of this paper is organized as follows. In Section 2, we discuss the treatment effect model. We also examine the identification of the CDFs of potential outcomes and propose estimators of the CDFs as well as the associated quantile functions. Section 3 presents regularity conditions and asymptotic results for the estimators of the CDFs and quantile functions. In Section 4, we introduce a simulation method to approximate the limiting stochastic processes of the estimated CDFs and quantile functions. We propose KS tests for the stochastic dominance relation between the distributions of potential outcomes and show the size and power properties of our tests in Section 5. We extend our results to the treated group in Section 6. Monte-Carlo simulation results are summarized in Section 7. Section 8 presents the empirical application, and Section 9 concludes. All mathematical proofs are contained in the Appendix.

[1] Linton et al. (2005, 2010) and Donald et al. (2012) allow for covariates in their model in a different way. In their framework, the variable of interest, say Y, is allowed to be a function of covariates. That is, Y can be defined as $Y = g(X, \theta)$, where X denotes a vector of covariates and $g(\cdot, \theta)$ is a real-valued function known up to the finite dimensional parameter θ. However, the dependence of the selection to treatment on covariates is not allowed in their model.

2. Estimating CDFs and quantile functions of the potential outcomes

2.1. Model framework

We consider the standard treatment effects framework. Let T be a dummy variable such that $T = 1$ if the individual receives treatment; otherwise, $T = 0$. Define $Y(1)$ as the potential outcome for the individual under treatment and $Y(0)$ as that without treatment. Let $F_0(\cdot) = P(Y(0) \le \cdot)$ and $F_1(\cdot) = P(Y(1) \le \cdot)$ denote the unconditional CDFs of $Y(0)$ and $Y(1)$ respectively. We observe $T, Y = T\cdot Y(1)+(1-T)\cdot Y(0)$, and X, a vector of covariates. We have a random sample of size N.

2.2. Identification of F_0 and F_1

In the treatment effect model, we never observe $Y(0)$ and $Y(1)$ simultaneously, so there is an identification problem due to the missing variable. We impose the unconfoundedness assumption (Rosenbaum and Rubin, 1983) to obtain identification.

Assumption 2.1 (*Unconfoundedness Assumption*). $(Y(0), Y(1)) \perp T|X$.

The unconfoundedness assumption requires that conditional on the observed individual characteristics, the treatment assignment is independent of the potential outcomes. Let $p(x) = P(T = 1|X = x)$ denote the propensity score function (Rosenbaum and Rubin, 1983, 1985). Under Assumption 2.1, $F_1(z)$ and $F_0(z)$ are identified by

$$F_0(z) = E\left[\frac{(1-T)\cdot 1(Y \le z)}{1-p(X)}\right], \qquad F_1(z) = E\left[\frac{T\cdot 1(Y \le z)}{p(X)}\right]. \tag{1}$$

The identification results in (1) are similar to those of $E[Y(1)]$ and $E[Y(0)]$ in HIR after we replace Y with indicator function $1(Y \le z)$.

2.3. Estimation of F_0 and F_1

Based on (1), the IPW estimators for $F_0(z)$ and $F_1(z)$ are:

$$\widehat{F}_0(z) = \sum_{i=1}^{N} \frac{(1-T_i)\cdot 1(Y_i \le z)}{1-\hat{p}(X_i)} \Bigg/ \sum_{i=1}^{N} \frac{1-T_i}{1-\hat{p}(X_i)}, \tag{2}$$

$$\widehat{F}_1(z) = \sum_{i=1}^{N} \frac{T_i\cdot 1(Y_i \le z)}{\hat{p}(X_i)} \Bigg/ \sum_{i=1}^{N} \frac{T_i}{\hat{p}(X_i)}, \tag{3}$$

where $\hat{p}(X_i)$ is a nonparametric estimator for $p(x)$. As in HIR, we use the Series Logit Estimator (SLE) to estimate $p(x)$ based on power series. Let $\lambda = (\lambda_1, \ldots, \lambda_r)' \in \mathbb{Z}_+^r$ be a r-dimensional vector of non-negative integers where $\mathbb{Z}_+$ denotes the set of non-negative integers, and define the norm for λ as $|\lambda| = \sum_{j=1}^{r} \lambda_j$. Let $\{\lambda(k)\}_{k=1}^{\infty}$ be a sequence including all distinct $\lambda \in \mathbb{Z}_+^r$ such that $|\lambda(k)|$ is non-decreasing in k and let $x^{\lambda} = \prod_{j=1}^{r} x_j^{\lambda_j}$. For any integer K, define $R^K(x) = (x^{\lambda(1)}, \ldots, x^{\lambda(K)})'$ as a vector of power functions. Let $\mathcal{L}(a) = \exp(a)/(1+\exp(a))$ be the logistic CDF. The SLE for $p(X_i)$ is defined as $\hat{p}(x) = \mathcal{L}\left(R^K(x)'\hat{\pi}_K\right)$ where

$$\hat{\pi}_K = \arg\max_{\pi_k} \frac{1}{N}\sum_{i=1}^{N}\Big(T_i \cdot \ln \mathcal{L}\left(R^K(X_i)'\pi_K\right) + (1-T_i)\cdot \ln\left(1-\mathcal{L}\left(R^K(X_i)'\pi_K\right)\right)\Big).$$

The asymptotic properties of $\hat{p}(x)$ are discussed in the Appendix of HIR.[2]

One can use other nonparametric estimators to estimate the propensity score function such as local polynomial estimators in Ichimura and Linton (2005), but one drawback of this approach is that the estimated propensity score is not necessarily bounded away from zero and one in finite samples and proper trimming is required. However, the propensity score function estimated by SLE is naturally bounded away from 0 and 1 and trimming is not required. One can also use the imputation estimator to estimate $\widehat{F}_0(z)$ and $\widehat{F}_1(z)$, e.g., Heckman et al. (1997, 1998b), Heckman et al. (1998b), and Hahn (1988). That is, $\widehat{F}_0(z)$ and $\widehat{F}_0(z)$ can be also estimated by

$$\widehat{F}_0(z) = \frac{1}{N}\sum_{i=1}^{N}\hat{F}_0(z|X_i), \qquad \widehat{F}_0(z) = \frac{1}{N}\sum_{i=1}^{N}\hat{F}_1(z|X_i)$$

where $\hat{F}_0(z|x)$ and $\hat{F}_1(z|x)$ are nonparametric estimators for $F_0(z|x)$ and $F_1(z|x)$ for all $x \in \mathcal{X}$. Under suitable assumptions, we expect that all the results discussed below still hold for the imputation estimators. However, a potential drawback of this method is that the estimated CDFs are not necessarily monotonically increasing.

2.4. Estimation of quantile functions

The quantile function $q(t)$ for $t \in [0, 1]$ of distribution function $F(z)$ is defined as

$$q(t) = \inf\{z : F(z) \geq t\}.$$

In the case where $F(z)$ is strictly increasing we have $q(F(z)) = z$ and we can write $q(t) = F^{-1}(t)$. The quantile functions of $Y(0)$ and $Y(1)$ are

$$q_0(t) = \inf\{z : F_0(z) \geq t\}, \qquad q_1(t) = \inf\{z : F_1(z) \geq t\}$$

for $t \in [0, 1]$, which are estimated by

$$\hat{q}_0(t) = \inf\{z : \widehat{F}_0(z) \geq t\}, \qquad \hat{q}_1(t) = \inf\{z : \widehat{F}_1(z) \geq t\}.$$

3. Asymptotic properties

3.1. Assumptions

In addition to the unconfoundedness assumption, we make the following assumptions which are similar to those in HIR.

Assumption 3.1 (*Distributions of $Y(0)$ and $Y(1)$*).

1. $Y(0)$ and $Y(1)$ have convex and compact supports $[z_{0\ell}, z_{0u}]$ and $[z_{1\ell}, z_{1u}]$. Let $\mathcal{Z} = [\min\{z_{0\ell}, z_{1\ell}\}, \max\{z_{0u}, z_{1u}\}]$ and without loss of generality, assume that $\mathcal{Z} = [0, \bar{z}]$ with $\bar{z} < \infty$.
2. F_0 and F_1 are continuous functions on $\mathcal{Z}$ with $F_0(0) = F_1(0) = 0$.

Assumption 3.2 (*Distribution of X*).

1. The support of the r-dimensional covariate X is a Cartesian product of compact intervals, $\mathcal{X} = \prod_{j=1}^{r}[x_{\ell j}, x_{uj}]$.
2. The density of X is bounded and bounded away from 0, on $\mathcal{X}$.

Let $F_0(z|x)$ and $F_1(z|x)$ be the conditional CDFs for $Y(0)$ and $Y(1)$ respectively.

Assumption 3.3 (*Conditional Distributions of $Y(0)$ and $Y(1)$*).

1. For any given $x \in \mathcal{X}$, $F_0(z|x)$ and $F_1(z|x)$ are continuous in $z \in \mathcal{Z}$.
2. For any given $z \in \mathcal{Z}$, $F_0(z|x)$ and $F_1(z|x)$ are continuously differentiable in $x \in \mathcal{X}$.

Assumption 3.4 (*Propensity Score*). For all $x \in \mathcal{X}$, the propensity score $p(x)$ satisfies the following conditions:

1. $p(x)$ is continuously differentiable of order $s \geq 7r$, where r is the dimension of $\mathcal{X}$.
2. $p(x)$ is bounded away from zero and one: $0 < \underline{p} \leq p(x) \leq \overline{p} < 1$.

Assumption 3.5 (*Series Estimator*). The SLE of $p(x)$ uses a power series with $K = a \cdot N^{\nu}$ for some $a > 0$ and $r/4(s-r) < \nu < 1/9$.

Remarks on the assumptions:

1. Assumption 3.1 requires that $Y(0)$ and $Y(1)$ have compact supports. This is not restrictive in that the theory regarding the CDF estimators remains the same when $Y(0)$ and $Y(1)$ have support on the whole real line. However, to estimate the whole quantile functions at the parametric rate, it is required that the density functions of $Y(0)$ and $Y(1)$ are bounded away from 0 on their support and this rules out the unbounded support case.[3]
2. Assumption 3.1 requires that $Y(0)$ and $Y(1)$ are continuous random variables on $\mathcal{Z}$. However, our results extend easily to cases where the probabilities of $Y(0)$ and $Y(1)$ at 0 are strictly positive, as in our empirical examples where in fact we have $F_0(0) > 0$ and $F_1(0) > 0$.[4]
3. Assumption 3.2 requires that all of the covariates are continuous. However, at the expense of additional notation, we can deal with the case where X has both continuous and discrete components. For example and for simplicity, let X_1 be a binary variable taking value on $\{0, 1\}$. Let the power series in the SLE be $R^K(x) = (1(x_1 = 0)\tilde{R}^K(x_{-1})', 1(x_1 = 1)\tilde{R}^K(x_{-1}))'$, where $\tilde{R}^K(x_{-1})$ is a vector of power functions in $x_{-1} \equiv (x_2, \ldots, x_r)$. We can generalize this approach to deal with discrete variables in the model.[5]
4. As in HIR, Assumption 3.4 ensures the existence of a ν satisfying the conditions in Assumption 3.5. Note that our theory holds for any choices of a and ν such that $K = a \cdot N^{\nu}$ that satisfies Assumption 3.5, and how to pick K in finite sample is still an open question. However, we also note that the simulation results in this paper are not very sensitive to the choice of K.
5. Also, as noted by Khan and Tamer (2010), the assumption that the propensity score is bounded away from zero and one plays an important role in determining the convergence rate of inverse probability weighted estimators.

3.2. Asymptotic properties of $\widehat{F}_0(z)$ and $\widehat{F}_1(z)$

Define $\widehat{\mathbf{F}} = (\widehat{F}_0, \widehat{F}_1)'$, $\mathbf{F} = (F_0, F_1)'$ and $\zeta = (z_1, z_2)' \in \mathcal{Z} \times \mathcal{Z}$.

Theorem 3.6. *Suppose Assumptions 2.1 and 3.1–3.5 hold. Then*

$$\sqrt{N}(\widehat{\mathbf{F}}(\cdot) - \mathbf{F}(\cdot)) \Rightarrow \Psi(\cdot),$$

where $\Rightarrow$ denotes weak convergence, $\Psi(\cdot)$ is a two dimensional mean zero Gaussian process with covariance functions $\Omega(\zeta_1, \zeta_2) = E[\psi(W, \zeta_1)\psi(W, \zeta_2)']$ where $W \equiv \{Y, T, X\}$ and $\psi(W, \zeta) = (\psi_0(W, z_1), \psi_1(W, z_2))'$ with

[2] All the asymptotic results of this paper will not change if one uses other distribution functions to construct the series estimators for the propensity score and estimate F_1 and F_0 accordingly. For example, one can use the standard normal distribution to construct the Series Probit Estimator. We use SLE because the theory for SLE is established by HIR.

[3] However, we can still estimate the quantile functions $q_0(t)$ and $q_1(t)$ for $t \in [\epsilon, 1-\epsilon]$ for any $0 < \epsilon < 1/2$ on which the density functions of $Y(0)$ and $Y(1)$ are bounded away from 0.

[4] In this case, we are not able to estimate the whole quantile functions of the potential outcomes.

[5] This method is equivalent to sample splitting. That is, we split the sample into two groups based on $X_1 = 0$ and $X_1 = 0$. Then we can estimate $\widehat{F}_0(z|X_1 = 1)$ and $\widehat{F}_0(z|X_1 = 0)$ because the unconfoundedness assumption holds in the subgroups. Next, $\widehat{F}_0(z)$ is estimated by $\hat{p}_{x_1}\widehat{F}_0(z|X_1 = 1) + (1 - \hat{p}_{x_1})\widehat{F}_0(z|X_1 = 0)$, where $\hat{p}_{x_1} = \sum_{i=1}^{N} 1(X_{1i} = 1)/N$.

$$\psi_0(W,z) = \frac{(1-T)\cdot 1(Y\le z)}{1-p(X)} + \frac{F_0(z|X)}{1-p(X)}(T-p(X)) - F_0(z),$$
$$\psi_1(W,z) = \frac{T\cdot 1(Y\le z)}{p(X)} - \frac{F_1(z|X)}{p(X)}(T-p(X)) - F_1(z),$$

and the convergence is in $\ell^\infty(\mathcal{Z})\times\ell^\infty(\mathcal{Z})$.[6,7]

To show Theorem 3.6, we first show that $\sqrt{N}(\widehat{F}_0(z)-F_0(z))$ and $\sqrt{N}(\widehat{F}_1(z)-F_1(z))$ are asymptotically equivalent to the following linear expressions respectively:

$$\sup_{z\in\mathcal{Z}}\left|\sqrt{N}(\widehat{F}_0(z)-F_0(z)) - \frac{1}{\sqrt{N}}\sum_{i=1}^N \psi_0(W_i,z)\right| = o_p(1),$$
$$\sup_{z\in\mathcal{Z}}\left|\sqrt{N}(\widehat{F}_1(z)-F_1(z)) - \frac{1}{\sqrt{N}}\sum_{i=1}^N \psi_1(W_i,z)\right| = o_p(1).$$

Second, we show that $\mathcal{K}_0 = \{\psi_0(W,z)|z\in\mathcal{Z}\}$ and $\mathcal{K}_1 = \{\psi_1(W,z)|z\in\mathcal{Z}\}$ are Donsker classes. Note that the Cartesian product of two Donsker classes of functions is also a Donsker class as in van der Vaart (2000, p. 270). Hence, by Donsker's Theorem or the functional central limit theorem, Theorem 3.6 follows.

3.3. Asymptotic properties of $\hat{q}_0$ and $\hat{q}_1$

We introduce some additional assumptions. Let $f_0(z)$ and $f_1(z)$ be the probability density functions of $Y(0)$ and $Y(1)$.

Assumption 3.7. $f_0(z)$ and $f_1(z)$ are continuous and bounded away from 0 on $[z_{0\ell}, z_{0u}]$ and $[z_{1\ell}, z_{1u}]$ respectively. Furthermore, $f_0(z)$ and $f_1(z)$ are continuously differentiable of order 2.

Assumption 3.7 implies that $F_0(z)$ and $F_1(z)$ are strictly increasing on $[z_{0\ell}, z_{0u}]$ and $[z_{1\ell}, z_{1u}]$, and the quantiles functions, $q_0(t)$ and $q_1(t)$, are well-defined on $[0, 1]$. Define $\widehat{\mathbf{q}} = (\hat{q}_0, \hat{q}_1)'$, $\mathbf{q} = (q_0, q_1)'$ and $\tau = (t_1, t_2)' \in [0, 1]\times[0, 1]$.

Theorem 3.8. *Suppose Assumptions 2.1, 3.1–3.5 and 3.7 hold. Then*

$$\sqrt{N}(\widehat{\mathbf{q}}(\cdot)-\mathbf{q}(\cdot)) \Rightarrow \mathcal{Q}(\cdot),$$

where $\mathcal{Q}(\tau) = (\mathcal{Q}_0(t_1), \mathcal{Q}_1(t_2))$ *is a two dimensional mean zero Gaussian process such that*

$$\mathcal{Q}_0(t_1) \equiv -\frac{\Psi_0(q_0(t_1))}{f_0(q_0(t_1))}, \qquad \mathcal{Q}_1(t_2) \equiv -\frac{\Psi_1(q_1(t_2))}{f_1(q_1(t_2))}.$$

The convergence is in $\ell^\infty([0,1])\times\ell^\infty([0,1])$.

Given that the quantile map is Hadamard differentiable, Theorem 3.8 follows from the functional delta method. Theorem 3.8 extends Theorem 1 of Firpo (2007a), which focuses on the asymptotics of the pointwise quantiles to the asymptotics of the quantile processes.

4. Simulating the limiting processes

4.1. Estimators for $F_0(z|x)$ and $F_1(z|x)$

Before introducing the simulation-based method, we construct estimators for $F_0(z|x)$ and $F_1(z|x)$ which are bounded between 0 and 1, monotonically increasing in z for any given x, and converge in probability to $F_0(z|x)$ and $F_1(z|x)$ uniformly in both arguments z and x. These estimators play important roles in the simulation-based method.

Let $\widetilde{F}_0(z|x)$ and $\widetilde{F}_1(z|x)$ be the series estimators for $F_0(z|x)$ and $F_1(z|x)$:

$$\widetilde{F}_0(z|x) = \left(\sum_{i=1}^N \frac{1(Y_i\le z)(1-T_i)}{1-\hat{p}(X_i)}R^K(X_i)\right)' \times\left(\sum_{i=1}^N R^K(X_i)R^K(X_i)'\right)^{-1} R^K(x),$$
$$\widetilde{F}_1(z|x) = \left(\sum_{i=1}^N \frac{1(Y_i\le z)T_i}{\hat{p}(X_i)}R^K(X_i)\right)' \times\left(\sum_{i=1}^N R^K(X_i)R^K(X_i)'\right)^{-1} R^K(x), \tag{4}$$

where $R^K(x)$ is the same power series used in SLE estimator. For any given x, $\widetilde{F}_0(z|x)$ and $\widetilde{F}_1(z|x)$ are step functions in z with jumps at Y_i's. It is true that $\widetilde{F}_0(z|x)$ and $\widetilde{F}_1(z|x)$ converge in probability to $F_0(z|x)$ and $F_1(z|x)$ uniformly in both arguments z and x, but they are not necessarily bounded between 0 and 1 and are not necessarily monotonically increasing in z for any given x. However, we can construct estimators for $F_0(z|x)$ and $F_1(z|x)$ satisfying all requirements based on $\widetilde{F}_0(z|x)$ and $\widetilde{F}_1(z|x)$.[8]

Without loss of generality, we assume that there are no ties between Y_i's and we add $Y_{(0)} = 0$ and $Y_{(N+1)} = \bar{z}$. Let $Y_{(i)}$ denote the i-th smallest element among the Y_i's so that we have $0 = Y_{(0)} < Y_{(1)} < \cdots < Y_{(N)} < Y_{(N+1)} = \bar{z}$. We define $\widehat{F}_0(z|x)$ by induction. Define $\widehat{F}_0(z|x) = \widetilde{F}_0(z|x) = 0$ for $Y_{(0)} \le z < Y_{(1)}$ and $\widehat{F}_0(Y_{(N+1)}|x) = 1$. Suppose $\widehat{F}_0(z|x) = 0$ is defined for $Y_{(0)} \le z < Y_{(i)}$, then we define for $Y_{(i)} \le z < Y_{(i+1)}$

$$\widehat{F}_0(z|x) = \widehat{F}_0(Y_{(i-1)}|x)\cdot 1\left(0\le \widetilde{F}_0(Y_{(i)}|x)\le \widehat{F}_0(Y_{(i-1)}|x)\right) + \widetilde{F}_0(Y_{(i)}|x)\cdot 1\left(\widehat{F}_0(Y_{(i-1)}|x) < \widetilde{F}_0(Y_{(i)}|x)\le 1\right) + 1\left(\widetilde{F}_0(Y_{(i)}|x) > 1\right).$$

The idea is that if $\widetilde{F}_0(z|x)$ jumps down at $Y_{(i)}$, then we set $\widehat{F}_0(z|x) = \widehat{F}_0(Y_{(i-1)}|x)$ for $Y_{(i)} \le z < Y_{(i+1)}$. At the same time, we trim $\widetilde{F}_0(z|x)$ between 0 and 1 by defining $\widehat{F}_0(z|x) = 0$ when $\widetilde{F}_0(z|x) < 0$ and defining $\widehat{F}_0(z|x) = 1$ when $\widetilde{F}_0(z|x) \ge 1$. And $\widehat{F}_1(z|x) = 0$ is defined in the same way. The properties of $\widehat{F}_0(z|x)$ and $\widehat{F}_1(z|x)$ are summarized in the following lemma.[9]

Lemma 4.1. *Suppose Assumptions 2.1 and 3.1–3.5 hold. Then for any given x,* $\widehat{F}_0(z|x)$ *and* $\widehat{F}_1(z|x)$ *are bounded between 0 and 1 and monotonically increasing in z, and*

$$\sup_{z\in\mathcal{Z},x\in\mathcal{X}}|\widehat{F}_0(z|x)-F_0(z|x)| = o_p(1),$$
$$\sup_{z\in\mathcal{Z},x\in\mathcal{X}}|\widehat{F}_1(z|x)-F_1(z|x)| = o_p(1).$$

It is easy to check that $\widehat{F}_0(z|x)$ and $\widehat{F}_1(z|x)$ are bounded between 0 and 1, and monotonically increasing in z. The last part of Lemma 4.1 follows from the facts that $\sup_{z\in\mathcal{Z},x\in\mathcal{X}}|\widetilde{F}_0(z|x) - F_0(z|x)| = o_p(1)$ and $\sup_{z\in\mathcal{Z}}|\widehat{F}_0(z|x)-F_0(z|x)| \le \sup_{z\in\mathcal{Z}}|\widetilde{F}_0(z|x)-F_0(z|x)|$ for all $x\in\mathcal{X}$. Note that the compactness of $\mathcal{X}$ is needed to obtain the uniform result. Similar arguments apply to $\widehat{F}_1(z|x)$.

---

6 The weak convergence is in the sense of Definition 1.3.3 of van der Vaart and Wellner (1996).

7 $\ell^\infty(\mathcal{Z})$ denotes the set of all bounded real function on $\mathcal{Z}$.

8 Instead of the series estimators, one can also use kernel estimators to estimate $F_0(z|x)$ and $F_1(z|x)$ so that the monotonicity and boundedness between 0 and 1 hold directly. Our main results still hold if the kernel estimators have the properties in Lemma 4.1.

9 Note that an alternative to this method of monotonization is the rearrangement method of Chernozhukov et al. (2010).

4.2. Simulating $\Psi(\zeta)$

Let $U_1, U_2, \ldots$ be i.i.d. random variables with mean equal to 0 and variance 1 and be independent of the sequence $\mathcal{W} \equiv W^\infty = \{W_1, W_2, \ldots\}$. For all $\zeta \in \mathcal{Z} \times \mathcal{Z}$ define the simulated stochastic processes as $\Psi^u(\zeta) = (\Psi_0^u(z_1), \Psi_1^u(z_2))'$ with

$$\begin{aligned}\Psi_0^u(z_1) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N U_i\left(\frac{(1-T_i)\cdot 1(Y_i \le z_1)}{1-\hat{p}(X_i)} - \widehat{F}_0(z_1)\right.\\ &\quad \left. + (T-\hat{p}(X_i))\frac{\widehat{F}_0(z_1|X_i)}{1-\hat{p}(X_i)}\right)\\ \Psi_1^u(z_2) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N U_i\left(\frac{T_i\cdot 1(Y_i \le z_2)}{\hat{p}(X_i)}\right.\\ &\quad \left. - \widehat{F}_1(z_2) - (T_i-\hat{p}(X_i))\frac{\widehat{F}_1(z_2|X_i)}{\hat{p}(X_i)}\right). \end{aligned} \tag{5}$$

Lemma 4.2. *Suppose Assumptions 2.1 and 3.1–3.5 hold. Then $\Psi^u(\cdot) \Rightarrow \Psi(\cdot)$ conditional on sample path $\mathcal{W}$ with probability approaching 1 (w.p.a. 1) which is denoted by $\Psi^u(\cdot) \overset{p}{\Rightarrow} \Psi(\cdot)$.*[10]

Lemma 4.2 shows that we can approximate $\Psi(\cdot)$ by the simulated stochastic process $\Psi^u(\cdot)$. To show this, we first show that $N^{-1/2}\sum_{i=1}^N U_i\psi(W_i,\cdot) \overset{p}{\Rightarrow} \Psi(\cdot)$ by the conditional multiplier central limit theorem (Theorem 2.9.6 of van der Vaart and Wellner (1996)). Second, we show that the effect of estimation errors of $\widehat{F}_0(z)$, $\widehat{F}_1(z)$, $\hat{p}(X_i)$, $\widehat{F}_0(z|X_i)$ and $\widehat{F}_1(z|X_i)$ will disappear in the limit. Monotonicity of $\widehat{F}_0(z|X_i)$ and $\widehat{F}_1(z|X_i)$ plays an important role in the second part of the proof.

4.3. Simulating $\mathcal{Q}(\tau)$

We first introduce uniformly consistent IPW kernel estimators for $f_0(z)$ and $f_1(z)$ in $\mathcal{Z}$ which plays an important role when showing that the estimation error of the proposed simulated process is asymptotically negligible.

Let h denote the bandwidth, which depends on sample size N, and $K(u)$ a kernel function. Define $\tilde{f}_0(z)$ and $\tilde{f}_1(z)$ as

$$\tilde{f}_0(z) = \frac{1}{Nh}\sum_{i=1}^N \frac{1-T_i}{1-\hat{p}(X_i)}K\left(\frac{Y_i - z}{h}\right),$$

$$\tilde{f}_1(z) = \frac{1}{Nh}\sum_{i=1}^N \frac{T_i}{\hat{p}(X_i)}K\left(\frac{Y_i - z}{h}\right).$$

The estimators for $f_0(z)$ and $f_1(z)$ are defined as

$$\hat{f}_0(z) = \begin{cases}\tilde{f}_0(z_{0\ell}+h) & \text{if } z \in [z_{0\ell}, z_{0\ell}+h],\\ \tilde{f}_0(z) & \text{if } z \in [z_{0\ell}+h, z_{0u}-h],\\ \tilde{f}_0(\bar{z}-h) & \text{if } z \in (z_{0u}-h, z_{0u}],\end{cases}$$

$$\hat{f}_1(z) = \begin{cases}\tilde{f}_1(z_{1\ell}+h) & \text{if } z \in [z_{1\ell}, z_{1\ell}+h],\\ \tilde{f}_1(z) & \text{if } z \in [z_{1\ell}+h, z_{1u}-h],\\ \tilde{f}_1(\bar{z}-h) & \text{if } z \in (z_{1u}-h, z_{1u}].\end{cases}$$

It is well known that the density estimator $\tilde{f}_0(z)$ is in general inconsistent around the boundary points $z_{0\ell}$ and z_{0u}. So is $\tilde{f}_1(z)$. Therefore, we modify $\tilde{f}_0(z)$ and $\tilde{f}_1(z)$ around the boundary points to obtain uniformly consistent estimators for $f_0(z)$ and $f_1(z)$. This method is also used in Donald et al. (2012). We make the following assumptions on $K(u)$ and h.

Assumption 4.3. Suppose $K(u)$ is non-negative and has support $[-1, 1]$. $K(u)$ is symmetric around 0 and is continuously differentiable of order 1. The bandwidth h satisfies $h \to 0$, $Nh^4 \to \infty$ and $Nh/\log N \to \infty$ when $N \to \infty$.

Lemma 4.4. *Suppose Assumptions 2.1, 3.1–3.5, 3.7 and 4.3 hold. Then $\sup_{z\in[z_{0\ell},z_{0u}]}|\hat{f}_0(z) - f_0(z)| = o_p(1)$ and $\sup_{z\in[z_{1\ell},z_{1u}]}|\hat{f}_1(z) - f_1(z)| = o_p(1)$.*

We define the simulated processes as $\mathcal{Q}^u(\tau) = (\mathcal{Q}_0^u(t_1), \mathcal{Q}_1^u(t_2))'$ with

$$\mathcal{Q}_0^u(\tau) = -\frac{\Psi_0^u(\hat{q}_0(t_1))}{\hat{f}_0(\hat{q}_0(t_1))}, \qquad \mathcal{Q}_1^u(t_2) = -\frac{\Psi_1^u(\hat{q}_1(t_2))}{\hat{f}_1(\hat{q}_1(t_2))}.$$

Theorem 4.5. *Suppose all conditions in Lemma 4.4 hold. Then $\mathcal{Q}^u(\cdot) \overset{p}{\Rightarrow} \mathcal{Q}(\cdot)$.*

Theorem 4.5 shows that $\mathcal{Q}^u(\tau)$ can approximate $\mathcal{Q}(\tau)$ well. The results regarding quantile functions allow us to test for the Lorenz dominance relations between the potential outcomes and to test for the quantile average treatment effects over a continuum of quantile indexes.

5. Tests for stochastic dominance

To demonstrate the usefulness of our results, we construct tests for the stochastic dominance relations between the distributions of the potential outcomes.

Determining stochastic dominance relations is important in social welfare comparisons. Let $W_U(H)$ denote a social welfare function of the form $W_U(H) = \int U(z)dH(z)$ where H is the distribution of income and U is any utility function. Let H_1 and H_2 be two CDFs. H_1 first order stochastically dominates (SD1) H_2 iff the welfare of H_1 will be greater than or equal to that of H_2 for any welfare function based on any monotonically increasing utility function, i.e., $W_U(H_1) \ge W_U(H_2)$ for all $U(z)$ such that $U'(z) \ge 0$. First order stochastic dominance can also be defined in terms of the underlying distributions. That is, H_1 SD1 H_2 iff $H_1(z) \le H_2(z)$ for all $z \in \mathbb{R}$.

First order stochastic dominance requires that the welfare function gives the same ordering over the class of monotonically increasing functions. This requirement may be too strong, so it is of interest to consider the weaker notion of higher order stochastic dominance. The importance of higher order stochastic dominance is discussed by McFadden (1989), Anderson (1996), Davidson and Duclos (2000) and Barrett and Donald (2003). Second order stochastic dominance (SD2) of H_1 over H_2 requires that the welfare function gives the same ordering for any monotonically increasing and concave function U, i.e., $U'(z) \ge 0$ and $U''(z) \le 0$. Similarly, H_1 SD2 H_2 iff $\int_0^z H_1(t)dt \le \int_0^z H_2(t)dt$ for all z. Furthermore, it is true that first order stochastic dominance implies second order stochastic dominance, but not vice versa.[11] We discuss the test for first order stochastic dominance in detail and briefly present the extension to higher order stochastic dominance cases.

---

[10] The conditional weak convergence is in the sense of Section 2.9 of van der Vaart and Wellner (1996). To be more specific, $\Psi_n^u \overset{p}{\Rightarrow} \Psi$ in the separable metric space $(\mathbb{D}, d)$ if and only if $\sup_{f\in BL_1}|E_u f(\Psi_n^u) - Ef(\Psi)| \xrightarrow{p} 0$ and $E_u f(\Psi_n^u)^* - E_u f(\Psi_n^u)_* \xrightarrow{p} 0$, where the subscript u in E_u indicates conditional expectation over the weights U_i's given the remaining data, BL_1 is the space of functions $f : \mathbb{D} \to R$ with Lipschitz norm bounded by 1, and $f(\Psi_n^u)^*$ and $f(\Psi_n^u)_*$ denote measurable majorants and minorants with respect to the joint data including the U_i's. We use the notation $\Psi_n^u \overset{a.s.}{\Rightarrow} \Psi$ to mean the same thing except with all $\xrightarrow{p}$'s replaced by $\xrightarrow{a.s.}$'s. Note that by Lemma 1.9.2(ii) of van der Vaart and Wellner (1996) and given that $\Psi(\cdot)$ is Borel measurable and separable, it is true that $\Psi_N^u \overset{p}{\Rightarrow} \Psi$ if and only if every subsequence k_N of N has a further subsequence ℓ_N of k_N such that $\Psi_{\ell_N}^u \overset{a.s.}{\Rightarrow} \Psi$.

[11] The stochastic dominance relation can be extended to any order. For example, j-th order stochastic dominance (SDj) of F_1 over F_0 requires that the welfare function gives the same ordering for any U such that $(-1)^k U^{(k)}(z) \le 0$ for $k = 1, \ldots, j$. It is also true that SDj implies $SD(j+1)$, but not vice versa.

5.1. Hypothesis formulation

We formulate the hypotheses regarding $Y(1)$ SD1 $Y(0)$ as follows:

$$H_0 : F_1(z) \leq F_0(z) \quad \text{for all } z \in \mathcal{Z};$$
$$H_1 : F_1(z) > F_0(z) \quad \text{for some } z \in \mathcal{Z}. \tag{6}$$

5.2. Test statistic and decision rule

The KS statistic is

$$\widehat{S}_N = \sqrt{N} \sup_{z\in\mathcal{Z}} \left(\widehat{F}_1(z) - \widehat{F}_0(z)\right). \tag{7}$$

Given a critical value c which will be defined later, the decision rule of the test is:

Reject H_0 if $\widehat{S}_N > c$.

Note that the calculation of $\widehat{S}_N$ involves taking the supremum over a compact set that might seem difficult in practice. However, given $\widehat{F}_0(z)$ and $\widehat{F}_1(z)$ are both step functions, we have $\widehat{F}_1(z)-\widehat{F}_0(z)$ is a step function in z with jumps at Y_i for all $i = 1, \ldots, N$. As a result, we can exactly calculate $\widehat{S}_N$, since $\widehat{S}_N = \sqrt{N}\max\{\widehat{F}_1(z) - \widehat{F}_0(z)|z = 0, \bar{z}, Y_1, \ldots, Y_N\}$, which only involves taking maximum over a finite number of values.

5.3. Least favorable configuration

When the null hypothesis involves an inequality, the set of points satisfying the null hypothesis is usually not a singleton. For example, if we fix $F_1(z)$, there will be infinitely many $F_0(z)$ satisfying $F_1(z) \leq F_0(z)$ for all $z \in \mathcal{Z}$. The typical way to resolve this is to apply the least favorable configuration (LFC) to find a point in the null hypothesis least favorable to the alternative hypothesis.

To obtain the LFC for our test, note that for any z

$$\begin{aligned}\widehat{F}_1(z) - \widehat{F}_0(z) &= (\widehat{F}_1(z) - \widehat{F}_0(z)) - (F_1(z) - F_0(z)) \\ &\quad + (F_1(z) - F_0(z)) \\ &\leq (\widehat{F}_1(z) - \widehat{F}_0(z)) - (F_1(z) - F_0(z)).\end{aligned}$$

The second inequality holds because under the null hypothesis $F_1(z) - F_0(z) \leq 0$ for all z. It follows that

$$\begin{aligned}\widehat{S}_N &= \sqrt{N}\sup_{z\in\mathcal{Z}}(\widehat{F}_1(z) - \widehat{F}_0(z)) \\ &\leq \sqrt{N}\sup_{z\in\mathcal{Z}}\left((\widehat{F}_1(z) - \widehat{F}_0(z)) - (F_1(z) - F_0(z))\right).\end{aligned}$$

The statistic is never greater than $\sqrt{N}((\widehat{F}_1(z) - \widehat{F}_0(z)) - (F_1(z) - F_0(z)))$ no matter what F_1 and F_0 are, and the equality holds if and only if $F_1(z) - F_0(z) = 0$ for all z. This suggests that the LFC is the point where $F_1(z) = F_0(z)$ for all $z \in \mathcal{Z}$. We define $\overline{S} \equiv \sup_{z\in\mathcal{Z}}(\Psi_1(z) - \Psi_0(z))$ which is the limiting null distribution under the LFC. If a null hypothesis is rejected under the LFC, then it will be rejected regardless of which point we use in the null hypothesis to construct the null distribution. Hence, a test based on the LFC is usually conservative in that the type I error of the test can be strictly smaller than the pre-determined level. We construct our tests based on the LFC.[12]

5.4. Simulated critical value, and size and power properties

To obtain the critical value, we first approximate $\overline{S}$, limit of the KS test statistic under the LFC, by $\overline{S}_u = \sup_z(\Psi_1^u(z) - \Psi_0^u(z))$. Because the supremum operator is a continuous function, then by the continuous mapping theorem we can show that $\overline{S}_u$ converges in distribution to $\overline{S}$ conditional on $\mathcal{W}$ w.p.a. 1. As a result, given the significance level α_0, the simulated critical value $\hat{c}$ is defined as the $(1-\alpha_0)$-th quantile of $\overline{S}_u$ such that

$$\hat{c} = \sup\{q|P_u(\overline{S}_u \leq q) \leq 1 - \alpha_0\}.$$

Let $c_0 = \sup\{q|P(\overline{S} \leq q) \leq 1 - \alpha_0\}$, the $(1-\alpha_0)$-th quantile of $\overline{S}$. If $\alpha_0 < 1/2$, $\hat{c}$ converges in probability to c_0.

Theorem 5.1. *Suppose Assumptions 2.1 and 3.1–3.5 hold and $\alpha_0 < 1/2$. If we reject the H_0 in (6) when $\widehat{S}_N > \hat{c}$, then:*

1. *if H_0 is true,* $\limsup P(\text{reject } H_0) = \limsup P(\widehat{S}_N > \hat{c}) \leq \alpha_0$ *where the equality holds when $F_0(z) = F_1(z)$ for all $z \in \mathcal{Z}$.*
2. *if H_0 is false,* $\lim_{N\to\infty} P(\text{reject } H_0) = 1$.

Theorem 5.1 shows that the size of our test will never be greater than the pre-specified significance level asymptotically and that our test is consistent in the sense that it will reject the null hypothesis with probability approaching 1 when the null hypothesis is false. The first part holds with inequality instead of equality because our test is constructed based on the LFC. In fact, when $F_1(z) \leq F_0(z)$ for all $z \in \mathcal{Z}$ and $F_1(z) < F_0(z)$ for some $z \in \mathcal{Z}$, then the asymptotic size will be strictly smaller than α_0.

Next, we show that the probability of rejection is larger than α_0 under a sequence of local alternatives that converge at rate $N^{-1/2}$ to the least favorable case. Let $F_{1,N}(z) = E[F_1(z|X) + \delta_1(z,X)/\sqrt{N}]$ and $F_{0,N}(z) = E[F_0(z|X) + \delta_0(z,X)/\sqrt{N}]$ such that $E[F_1(z|X)] = E[F_0(z|X)]$ for all $z \in \mathcal{Z}$, and $\delta(z) \equiv E[\delta_1(z|X) - \delta_0(z|X)] \geq 0$ for all z and the strict inequality holds for some z.[13] Define the local alternatives as

$$H_{1,N} : F_{1,N}(z) - F_{0,N}(z) = \frac{\delta(z)}{\sqrt{N}}. \tag{8}$$

Theorem 5.2. *Suppose Assumptions 2.1 and 3.1–3.5 hold for each N and $\alpha_0 < 1/2$. Under the local alternatives (8),* $\lim_{N\to\infty} P(\text{reject } H_0) \geq \alpha_0$.

Theorem 5.2 shows the local power of our test is greater than or equal to α_0 under local alternatives that converge to the least favorable case and the deviation, $\delta(z)$, is non-negative for all $z \in \mathcal{Z}$.

Remarks. 1. Note that all of our results introduced so far crucially depend on the validity of the unconfoundedness assumption. If the unconfoundedness assumption is not correct, our results have to be interpreted more carefully. To the best of our knowledge, the only direct test for this assumption is proposed by Donald et al. (2011), although their test only applies to cases where there is a valid binary instrument variable for the treatment indicator satisfying the one-sided non-compliance assumption.[14,15] There are indirect methods to assess the unconfoundedness assumption, (e.g., Rosenbaum (1987), Heckman et al. (1997), and Heckman and Hotz (1989)) that test whether a treatment effect that is known to be zero is actually zero. However, these tests are indirect in that the rejections of these tests are interpreted as "weakening" the unconfoundedness assumption. Please refer to Imbens and Wooldridge (2009, Section 5.11) for a review on these indirect tests.

12 To keep this paper more focused, we do not employ the recentering method or generalized moment selection method as in Donald and Hsu (2010), Andrews and Shi (2013) and Linton et al. (2010) to construct more powerful tests without resorting to LFC, but the extension of our tests to this case is straightforward.

13 For each N, we assume that $F_1(z|x) + \delta_1(z,x)/\sqrt{N}$ and $F_0(z|x) + \delta_0(z,x)/\sqrt{N}$ satisfy all the regularity conditions introduced in Section 3.

14 For a direct test, we mean that a rejection of the null hypothesis is interpreted as a failure of the unconfoundedness assumption.

15 Please refer to Donald et al. (2011) for more details regarding their tests and the definitions of the validity of an IV and the one-sided non-compliance assumption.

2. So far, we construct our test based on the critical value method; however, one can also implement our test based on the p-value method. Define the p-value of our test as $\hat{p}(\widehat{S}_N) = P(\overline{S}_u > \widehat{S}_N)$. For $\alpha_0 < 1/2$, we reject H_0 when $\hat{p}(\widehat{S}_N) < \alpha_0$. It is true that for a given $\alpha_0 < 1/2$, $\widehat{S}_N > \hat{c}$ iff $\hat{p}(\widehat{S}_N) < \alpha_0$. Therefore, the critical value method and the p-value method are equivalent.
3. Note that our test without any covariate is equivalent to Barrett and Donald's (2003) test with the p-values generated by their Eq. (6) which is labeled as the KS2 method in their simulations. To see this, define $N_1 = \sum_i T_i$ and $N_0 = \sum_i(1 - T_i)$ which are the numbers of the treated individuals and the untreated ones respectively. When there are no covariates, the only regressor in the SLE is the constant term and we have $\hat{p} = N_1/N$, the proportion of treated individuals. Also, we have $\sum_i T_i/\hat{p} = N$ and $\sum_i(1 - T_i)/(1 - \hat{p}) = N$. Therefore,

$$\widehat{F}_1(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i \cdot 1(Y_i \le z)}{\hat{p}} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i \cdot 1(Y_i \le z)}{\frac{N_1}{N}} = \frac{1}{N_1}\sum_{\{i:T_i=1\}} 1(Y_i \le z),$$

which is the empirical CDF based on the treated individuals. Similarly, $\widehat{F}_0(z)$ is the empirical CDF based on the untreated individuals. As a result,

$$\widehat{S}_N = \sqrt{N}\sup_{z\in\mathcal{Z}}(\widehat{F}_1(z) - \widehat{F}_0(z)) = \sqrt{\frac{N^2}{N_1N_0}}\left(\sqrt{\frac{N_1N_0}{N}}\sup_{z\in\mathcal{Z}}(\widehat{F}_1(z) - \widehat{F}_0(z))\right),$$

where the term in the largest parentheses is the test statistic of Barrett and Donald's (2003). Similarly, we can rewrite $\Psi_u(z)$ as

$$\begin{aligned}\Psi_u(z) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}U_i\left\{\frac{T_i \cdot (1(Y_i \le z) - \widehat{F}_1(z))}{\hat{p}(X_i)} - \frac{(1 - T_i)\cdot 1(Y_i \le z) - \widehat{F}_0(z)}{1 - \hat{p}(X_i)}\right\}\\ &= \sqrt{\frac{N^2}{N_1N_0}}\left(\sqrt{\frac{N_1N_0}{N}}\left(\frac{1}{N_1}\times\sum_{\{i:T_i=1\}}U_i\{1(Y_i \le z) - \widehat{F}_1(z)\} - \frac{1}{N_0}\sum_{\{i:T_i=0\}}U_i\{1(Y_i \le z) - \widehat{F}_0(z)\}\right)\right),\end{aligned}$$

where the term in the largest parentheses is the simulated process of Barrett and Donald's (2003). These imply that the simulated null distribution and the critical value of our test are those of Barrett and Donald's (2003) with a rescale factor $\sqrt{\frac{N^2}{N_1N_0}}$. Therefore, the tests will be identical in this special case. For Barrett and Donald's (2003) test to work, it is required that the treatment is randomly assigned, i.e. $(Y(0), Y(1), X) \perp T$. In this case, $F_0(z) = E[1(Y \le z)|D = 0] = F(z|T = 0)$ and $F_1(z) = E[1(Y \le z)|D = 1] = F(z|T = 1)$, so their test based on the treated sample and the untreated sample works. Furthermore, the random assignment assumption also implies that $(Y(0), Y(1)) \perp T|X_1$ for any $X_1 \subseteq X$, so we can implement our tests based on any subset of X. However, if the treatment assignment is not random, but only unconfounded conditional on some covariates, then $F(z|T = 0) = E[1(Y \le z)|T = 0]$ and $F(z|T = 1) = E[1(Y \le z)|T = 1]$ are not equal to $F_0(z)$ and $F_1(z)$ in general and their test will be invalid whereas our test will be valid. In this sense, our test is more robust.
4. The simulation we propose is different from the bootstrap procedure in Maier (2011). One difference is that our procedure does not require repeated estimation of the various nonparametric components including the propensity score based on the SLE. It is an open question as to which is better in practice.

5.5. Tests for higher order stochastic dominance

Let $\mathcal{I}_j(.; G)$ be the functional that integrates the function G to order $j - 1$ so that

$$\begin{aligned}\mathcal{I}_1(z; G) &= G(z),\\ \mathcal{I}_2(z; G) &= \int_0^z G(t)dt = \int_0^z \mathcal{I}_1(t; G)dt,\\ &\vdots\\ \mathcal{I}_j(z; G) &= \int_0^z \mathcal{I}_{j-1}(t; G)dt.\end{aligned}$$

Accordingly, F_1 *SDj* F_0 corresponds to $\mathcal{I}_j(z; F_1) \le \mathcal{I}_j(z; F_0)$ for all z. Hence, the hypotheses for F_1 *SDj* F_0 are:

$$H_0^j : \mathcal{I}_j(z; F_1) \le \mathcal{I}_j(z; F_0) \quad \text{for all } z \in \mathcal{Z};$$
$$H_1^j : \mathcal{I}_j(z; F_1) > \mathcal{I}_j(z; F_0) \quad \text{for some } z \in \mathcal{Z}.$$

Define the statistic as

$$\widehat{S}_N^j = \sqrt{N}\sup_{z\in\mathcal{Z}}\mathcal{I}_j(z; \widehat{F}_1 - \widehat{F}_0).$$

As in Davidson and Duclos (2000) and Barrett and Donald (2003), one can show that

$$\mathcal{I}_j(z; \widehat{F}_1) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(j-1)!}\frac{T_i}{\hat{p}(X_i)}1(Y_i \le z)(z - X_i)^{j-1},$$

which is a piecewise polynomial as is $\mathcal{I}_j(z; \widehat{F}_0)$. As a result, $\mathcal{I}_j(z; \widehat{F}_1 - \widehat{F}_0) = \mathcal{I}_j(z; \widehat{F}_1) - \mathcal{I}_j(z; \widehat{F}_0)$ is also a piecewise polynomial and the supremum of it can be computed simply.[16]

The LFC in this case is the point where $\mathcal{I}_j(z; F_1) = \mathcal{I}_j(z; F_0)$ for all z. Note that $\mathcal{I}_j(z; F_1) = \mathcal{I}_j(z; F_0)$ for all z if and only if $F_1(z) = F_0(z)$ for all Z. Hence, the limiting random variable under the LFC is $\overline{S}^j = \sup_{z\in\mathcal{Z}}\mathcal{I}_j(z; \Psi)$. The simulated critical value $\hat{c}^j$ with significance level α_0 is defined as the $(1 - \alpha_0)$-th quantile of $\overline{S}_u^j = \sup_{z\in\mathcal{Z}}\mathcal{I}_j(z; \Psi_u)$. The size and power properties of the test of higher order stochastic dominance and the proof are similar to the first order stochastic dominance case, and we omit the statements.[17]

6. Estimation among the treated group

In general, researchers are interested not only in a relation between the potential outcomes across the whole population, but also between the potential outcomes in certain subpopulations

16 In the empirical implementation, we use the left Riemann sum to approximate the $\mathcal{I}_2(z; \widehat{F}_1 - \widehat{F}_0)$ and $\widehat{S}_N^2$ for a given set of gridpoints. Let $0 = z_0 < z_1 < \cdots < z_k = \bar{z}$ and we approximate $\mathcal{I}_2(z_j; \widehat{F}_1 - \widehat{F}_0)$ and $\widehat{S}_N^2$ respectively by

$$\mathcal{I}_2(z_j; \widehat{F}_1 - \widehat{F}_0) = \sum_{\ell=0}^{j-1}\left(\widehat{F}_1(z_\ell) - \widehat{F}_0(z_\ell)\right)(z_{\ell+1} - z_\ell),$$
$$\widehat{S}_N^2 = \max_{j=1,\ldots,k}\mathcal{I}_2(z_j; \widehat{F}_1 - \widehat{F}_0).$$

17 The proof can be done by applying continuous mapping theorem because the integral operator $\mathcal{I}_j$ is a continuous function from $\ell^\infty(\mathcal{Z})$ to $\ell^\infty(\mathcal{Z})$.

such as the treated. In this section we extend our estimation results and the simulation method to the treated subpopulation.

Let $F_0^t(z) = E[1(Y(0) \le z)|T = 1]$ and $F_1^t(z) = E[1(Y(1) \le z)|T = 1]$ denote the conditional CDFs of the potential outcomes of the treated. Similar to (1), $F_0^t(z)$ and $F_1^t(z)$ are identified by

$$F_0^t(z) = E\left[\frac{p(X)(1-T)\cdot 1(Y \le z)}{1-p(X)}\right] \Big/ E\left[\frac{p(X)(1-T)}{1-p(X)}\right],$$
$$F_1^t(z) = E\left[T \cdot 1(Y \le z)\right]/E[T]. \tag{9}$$

Based on (9), we estimate $F_0^t(z)$ and $F_1^t(z)$ by

$$\widehat{F}_0^t(z) = \frac{1}{N}\sum_{i=1}^N \frac{\hat{p}(X_i)(1-T_i)\cdot 1(Y_i \le z)}{1-\hat{p}(X_i)}/\hat{p}_0,$$
$$\widehat{F}_1^t(z) = \frac{1}{N}\sum_{i=1}^N T_i \cdot 1(Y_i \le z)/\hat{p}_1,$$

where

$$\hat{p}_0 = \frac{1}{N}\sum_{i=1}^N \frac{\hat{p}(X_i)(1-T_i)}{1-\hat{p}(X_i)}, \qquad \hat{p}_1 = \frac{1}{N}\sum_{i=1}^N T_i.$$

Also, we estimate the quantile functions on the treated by

$$\hat{q}_0^t(t) = \inf\{z : \widehat{F}_0^t(z) \ge t\}, \qquad \hat{q}_1^t(t) = \inf\{z : \widehat{F}_1^t(z) \ge t\}.$$

Let $f_0^t(z)$ and $f_1^t(z)$ be the probability density functions of $Y(0)$ and $Y(1)$ conditional on the treated.

We modify Assumption 3.1 for the treated case.

Assumption 6.1 (*Conditional Distributions of $Y(0)$ and $Y(1)$ on the Treated*)**.**

1. $F_0^t(z)$ and $F_1^t(z)$ have supports $[z_{0\ell}, z_{0u}]$ and $[z_{1\ell}, z_{1u}]$.
2. $F_0^t(z)$ and $F_1^t(z)$ are continuous functions on $[z_{0\ell}, z_{0u}]$ and $[z_{1\ell}, z_{1u}]$ with $F_0^t(0) = F_1^t(0) = 0$.
3. $f_0^t(z)$ and $f_1^t(z)$ are continuous and bounded away from 0 on $[z_{0\ell}, z_{0u}]$ and $[z_{1\ell}, z_{1u}]$. Furthermore, $f_0^t(z)$ and $f_1^t(z)$ are continuously differentiable of order 2.

The following theorem similar to Theorem 3.6 summarizes the asymptotics of the estimators. Define $\widehat{\mathbf{F}}^t = (\widehat{F}_0^t, \widehat{F}_1^t)'$, $\mathbf{F}^t = (F_0^t, F_1^t)'$, $\widehat{\mathbf{q}}^t = (\hat{q}_0^t, \hat{q}_1^t)'$, and $\mathbf{q} = (q_0^t, q_1^t)'$.

Theorem 6.2. *Suppose Assumptions 2.1, 3.2–3.5 and 6.1 hold. Then*

$$\sqrt{N}(\widehat{\mathbf{F}}^t(\cdot) - \mathbf{F}^t(\cdot)) \Rightarrow \Psi^t(\cdot), \qquad \sqrt{N}(\widehat{\mathbf{q}}^t(\cdot) - \mathbf{q}^t(\cdot)) \Rightarrow \mathcal{Q}^t(\cdot),$$

where $\Psi^t(\cdot)$ is a two dimensional mean zero Gaussian process with covariance functions $\Omega^t(\zeta_1,\zeta_2) = E[\psi^t(W,\zeta_1)\psi^t(W,\zeta_2)']$ where $\psi(W,\zeta) = (\psi_0^t(W,z_1), \psi_1^t(W,z_2))'$ with

$$\psi_0^t(W,z) = \frac{1}{E[T]}\left(\frac{p(X)(1-T)\left(1(Y \le z) - F_0(z|X)\right)}{1-p(X)} + T\left(F_0(z|X) - F_0^t(z)\right)\right),$$
$$\psi_1^t(W,z) = \frac{1}{E[T]}\left(T \cdot (1(Y \le z) - F_1^t(z))\right),$$

and $\mathcal{Q}^t(\tau) = (\mathcal{Q}_0^t(t_1), \mathcal{Q}_1^t(t_2))$ is a two dimensional mean zero Gaussian process such that

$$\mathcal{Q}_0^t(t_1) \equiv -\frac{\Psi_0^t(q_0^t(t_1))}{f_0^t(q_0^t(t_1))}, \qquad \mathcal{Q}_1^t(t_2) \equiv -\frac{\Psi_1^t(q_1^t(t_2))}{f_1^t(q_1^t(t_2))}.$$

To describe the simulation method here we first define $\tilde{f}_0^t(z)$ and $\tilde{f}_1^t(z)$ as

$$\tilde{f}_0(z) = \frac{1}{Nh}\sum_{i=1}^N \frac{\hat{p}(X_i)(1-T_i)}{1-\hat{p}(X_i)} K\left(\frac{Y_i - z}{h}\right)/\hat{p}_0,$$
$$\tilde{f}_1(z) = \frac{1}{Nh}\sum_{i=1}^N T_i K\left(\frac{Y_i - z}{h}\right)/\hat{p}_1.$$

The estimators for $f_0^t(z)$ and $f_1^t(z)$ are defined as

$$\hat{f}_0^t(z) = \begin{cases} \tilde{f}_0^t(z_{0\ell}+h) & \text{if } z \in [z_{0\ell}, z_{0\ell}+h], \\ \tilde{f}_0^t(z) & \text{if } z \in [z_{0\ell}+h, z_{0u}-h], \\ \tilde{f}_0^t(\bar{z}-h) & \text{if } z \in (z_{0u}-h, z_{0u}], \end{cases}$$
$$\hat{f}_1^t(z) = \begin{cases} \tilde{f}_1^t(z_{1\ell}+h) & \text{if } z \in [z_{1\ell}, z_{0\ell}+h], \\ \tilde{f}_1^t(z) & \text{if } z \in [z_{1\ell}+h, z_{1u}-h], \\ \tilde{f}_1^t(\bar{z}-h) & \text{if } z \in (z_{1u}-h, z_{1u}]. \end{cases}$$

The simulated stochastic processes are defined as

$$\Psi_{0,t}^u(z) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{U_i}{\hat{p}_0}\left(\frac{\hat{p}(X_i)(1-T_i)(1(Y_i \le z) - \widehat{F}_0(z|X_i))}{1-\hat{p}(X_i)} + T_i\left(\widehat{F}_0(z|X_i) - \widehat{F}_0^t(z)\right)\right),$$
$$\Psi_{1,t}^u(z) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{U_i}{\hat{p}_1}\left(T_i \cdot 1(Y_i \le z) - \widehat{F}_1^t(z)\right),$$
$$\mathcal{Q}_{0,t}^u(t) = -\frac{\Psi_{0,t}^u(\hat{q}_0^t(t))}{\hat{f}_0^t(\hat{q}_0^t(t))}, \qquad \mathcal{Q}_{1,t}^u(t) = -\frac{\Psi_{1,t}^u(\hat{q}_1^t(t))}{\hat{f}_1^t(\hat{q}_1^t(t))}.$$

Similar to Theorem 4.2 and Theorem 4.5, we can show that $\Psi^t(\cdot)$ and $\mathcal{Q}^t(\cdot)$ can be approximated by $\Psi_t^u(\cdot) = (\Psi_{0,t}^u(\cdot), \Psi_{1,t}^u(\cdot))$ and $\mathcal{Q}_t^u(\cdot) = (\mathcal{Q}_{0,t}^u(\cdot), \mathcal{Q}_{1,t}^u(\cdot))$. Extending the stochastic dominance tests to the treated cases is also straightforward.

7. Simulations

In this section we conduct small-scale Monte Carlo studies to illustrate the power and size properties of our tests.

Example 7.1. Let the data generating process (DGP) be:

$$X = 0.3 + 0.4U_x, \qquad T = 1(U_t < X),$$
$$Y(0) = 1(U_{y_0} \le X)\frac{U_{y_0}^2}{X} + 1(U_{y_0} > X)U_{y_0},$$
$$Y(1) = 1(U_{y_1} \le 1-X)\frac{U_{y_1}^2}{1-X} + 1(U_{y_1} > 1-X)U_{y_1},$$
$$Y = TY(1) + (1-T)Y(0),$$

where U_x, U_t, U_{y_0} and U_{y_1} are independent uniform distributions over [0, 1].

In Example 7.1, T is independent of $(Y(0), Y(1))$ conditional on X, i.e. T is unconfounded. This example is used to examine the size properties of the test. The unconditional CDFs of $Y(0)$ and $Y(1)$ are identical, so $Y(1)$ SD1 $Y(0)$. In this case we are interested in whether the simulation gives a test with approximately nominal size (5%). We can also show that $F_1(z|T = 1) \le F_0(z|T = 1)$ for all z, which implies that $Y(1)$ conditional on the treated SD1 $Y(0)$ conditional on the treated. In this case, we expect test to have size less than nominal since it is based on the LFC.

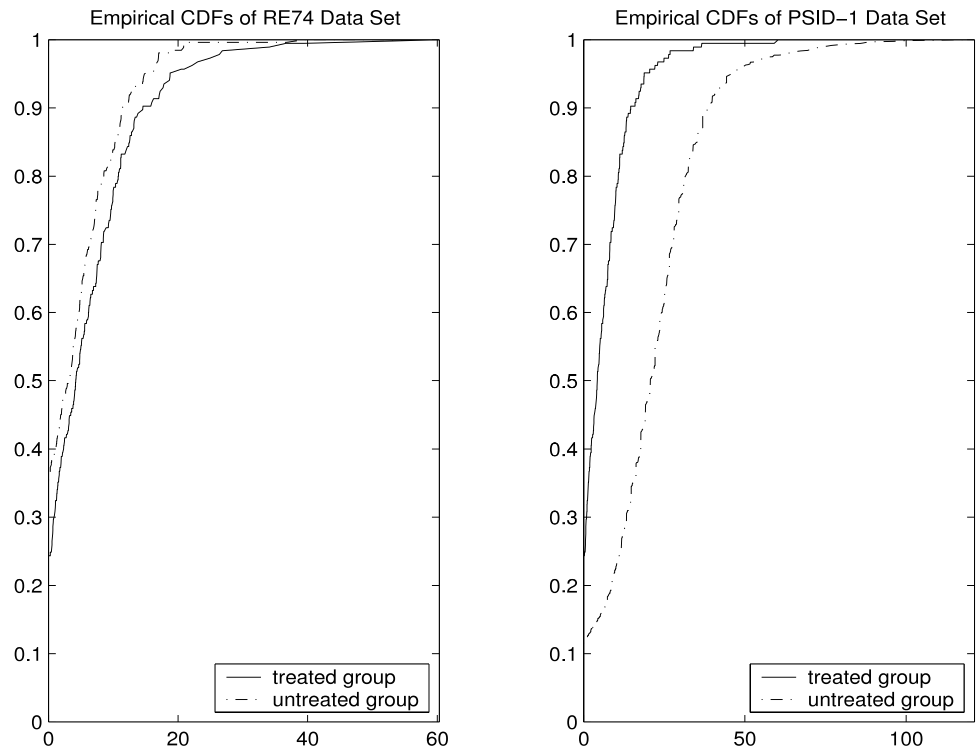

Fig. 1. Empirical CDFs of RE74 and PSID-1 data sets.

Example 7.2. Let the DGP be:

$$X = 0.3 + 0.4U_x, \qquad T = 1(U_t < X),$$
$$Y(0) = U_{y_0},$$
$$Y(1) = 1(U_{y_1} \le X)\frac{U_{y_1}^2}{X} + 1(U_{y_1} > X)U_{y_1},$$
$$Y = TY(1) + (1 - T)Y(0),$$

where U_x, U_t, U_{y_0} and U_{y_1} are independent uniform distributions over [0, 1].

In Example 7.2, $Y(1)$ does not SD1 $Y(0)$ conditional on X, because $F_1(z|X) > F_0(z|X)$ for all $0 < z \le X$ and $F_1(z|X) = F_0(z|X)$ for all $1 > z > X$. It follows that $Y(1)$ does not SD1 $Y(0)$ unconditionally and conditional on the treated group because $F_1(z) > F_0(z)$ and $F_1^t(z) > F_0^t(z)$ for all $0 < z \le X$. We use this example to examine the power properties of our tests.

We first set the sample sizes $N = 200$. The U_i's used to simulate the Gaussian processes, are iid standard normal. The rejection rate is calculated based on 1000 simulations, and for each simulation the p-value is approximated by 1000 repetitions. We use all of the different values of Y_i as the gridpoints to compute the test statistics as well as the suprema of the simulated stochastic processes. The propensity score function is estimated by SLE with the power series: $1, X$ and X^2. The significance level is 5%. Table 1 summarizes the rejection rates for different cases.

The rejection rates corresponding to Example 7.1 are 5.8% and 4.1% indicating that the procedure has size properties consistent with the theory even in a very small sample. The null hypotheses are violated in Example 7.2 and we find that the rejection rates are between 43.7% and 45.1% indicating that the procedure is able to detect violations of the null hypothesis. When we redo the simulations with sample size $N = 400$, we estimate the propensity score function using: 1, X, X^2 and X^3. The rejection rates under the null and alternative hypotheses suggest if anything that size is closer to nominal in the LFC case, is slightly more conservative for the null in the treated case and that power is increased in the larger sample size. These results, though limited, suggest that even in very small samples the procedures have good properties.

Table 1
Rejection rates.

N		F_1 SD1 F_0	F_1^t SD1 F_0^t
200	Example 7.1	0.058	0.041
	Example 7.2	0.437	0.451
400	Example 7.1	0.055	0.039
	Example 7.2	0.781	0.816

8. Empirical illustration

We apply our test to the data from National Supported Work Demonstration (NSW) job training program. These data sets were first analyzed by LaLonde (1986) and subsequently by Heckman and Hotz (1989), Dehejia and Wahba (1999), Imbens (2003), Smith and Todd (2001, 2005), Firpo (2007b), Abadie and Imbens (2011) and Wooldridge (2002). For a detailed description of the NSW data, please refer to LaLonde (1986).

The data sets we use correspond to the sub-samples termed "RE74 subset" and "PSID-1" in Dehejia and Wahba (1999). The treatment variable T is equal to 1 if the individual participates in the job training. RE74 subset contains an experimental sample from a randomized evaluation of the NSW program in which 185 individuals receive the treatment and 260 do not. PSID-1 contains the experimental participants in the RE74 subset and a non-experimental comparison group with 2490 individuals from the PSID. We plot the empirical CDFs for the treated and untreated groups for both data sets in Fig. 1.

We are interested in the distributional effect of the job training program on earnings in 1978. We only apply the tests for the whole group to RE74 subset, because the treatment is randomly assigned in this subset, which implies the CDFs for the whole group are the same as the CDFs for the treated group. Our tests are implemented for three different estimates of the propensity score. In the first one,

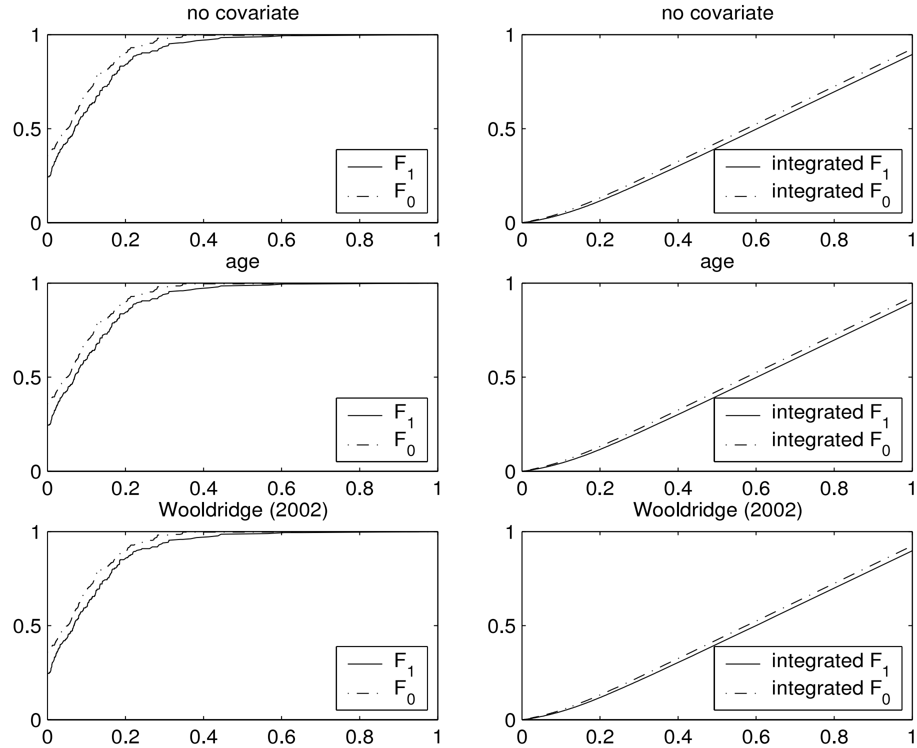

Fig. 2. Estimated CDFs and integrated CDFs for RE74 under different specifications.

Table 2
Stochastic dominance in RE74 data set.

Specification		F_1 SD1 F_0	F_1 SD2 F_0	F_0 SD1 F_1	F_0 SD2 F_1
No covariate	Statistic	0	0	2.787	0.626
	p-value	1.000	1.000	0.019	0.004
Age	Statistic	0	0	2.824	0.596
	p-value	1.000	1.000	0.019	0.005
Wooldridge	Statistic	0	0	2.768	0.551
	p-value	1.000	1.000	0.021	0.009

Table 3
Stochastic dominance in PSID-1 data set.

Specification		F_1^t SD1 F_0^t	F_1^t SD2 F_0^t	F_0^t SD1 F_1^t	F_0^t SD2 F_1^t
Wooldridge	Statistic	1.595	0.000	7.938	0.507
	p-value	0.789	1.000	0.152	0.164
Dehejia and Wahba	Statistic	0.033	0.000	2.430	1.710
	p-value	0.986	1.000	0.068	0.000
Firpo	Statistic	1.887	0.000	8.069	0.511
	p-value	0.774	1.000	0.143	0.163

we do not include any covariates. In the second one, we use age as the only covariate in the conditioning set, and a constant, age and the squared age are the regressors in the SLE. In the last one, we use the specification in Example 18.2 of Wooldridge (2002), which includes a constant, age, the squared age, real earnings in 1974 and 1975, a binary high school degree indicator, marital status, and dummy variables for black and Hispanic in the SLE. Since affine transformations preserve the stochastic dominance relations, we apply an affine transformation on the outcome variables so that the minimum is 0 and maximum is 1. We use all different values of transformed outcome variables as the gridpoints. The p-values for various tests are approximated by 10,000 repetitions. Test statistics and p-values regarding various specifications are summarized in Table 2. The estimated CDFs and integrated CDFs under different specifications are summarized in Fig. 2.

We find that the testing results are robust to the different estimates of the propensity score. We accept that the 1978 real earnings under job training first-order and second-order stochastically dominates that without job training since all of the p-values are equal to 1. Regarding stochastic dominance of the outcome without job training against that under job training, we reject first-order stochastic dominance at the 5% significance level and reject second-order stochastic dominance at the 1% significance level.

We apply our test on the treated to PSID-1 as Firpo (2007b) suggests, since the non-experimental comparison group is essentially different from the treated group. We conduct our tests according to three different estimates of the propensity score. One is Wooldridge's (2002), which is used in previous empirical example. Another is proposed by Dehejia and Wahba (1999) in which in addition to those regressors in Wooldridge's (2002), it also contains education, squared education, squared real earnings in 1974 and 1975, and the interaction term between the dummy for black and the dummy for unemployed in 1974. The other one is proposed by Firpo (2007b), which is different from Dehejia and Wahba's (1999) in the interaction terms. The interaction terms Firpo (2007b) uses are marital status with real earnings in 1974 and marital status with the dummy for unemployed in 1974. We present the test results in Table 3. We plot the estimated F_i^t's and integrated F_i^t's under different specifications in Fig. 3.

The testing results depend on how the propensity score is estimated. However, in all specifications, we accept that conditional on the treated, the outcome under job training SD1 and SD2 that without job training. On the other hand, under Wooldridge (2002) and Firpo's (2007b) specifications, we accept that conditional on the treated, the outcome without job training SD1 and SD2 that under job training. However, under Dehejia and Wahba's (1999) specification, we reject the first order stochastic dominance of the outcome without job training against that under job training at the 10% significance level and reject second order stochastic dominance at the 1% significance level.

To sum up, the empirical results suggest that the distribution of real earnings under job training SD1 and SD2 the distribution of real earnings without job training. Equivalently, the social welfare

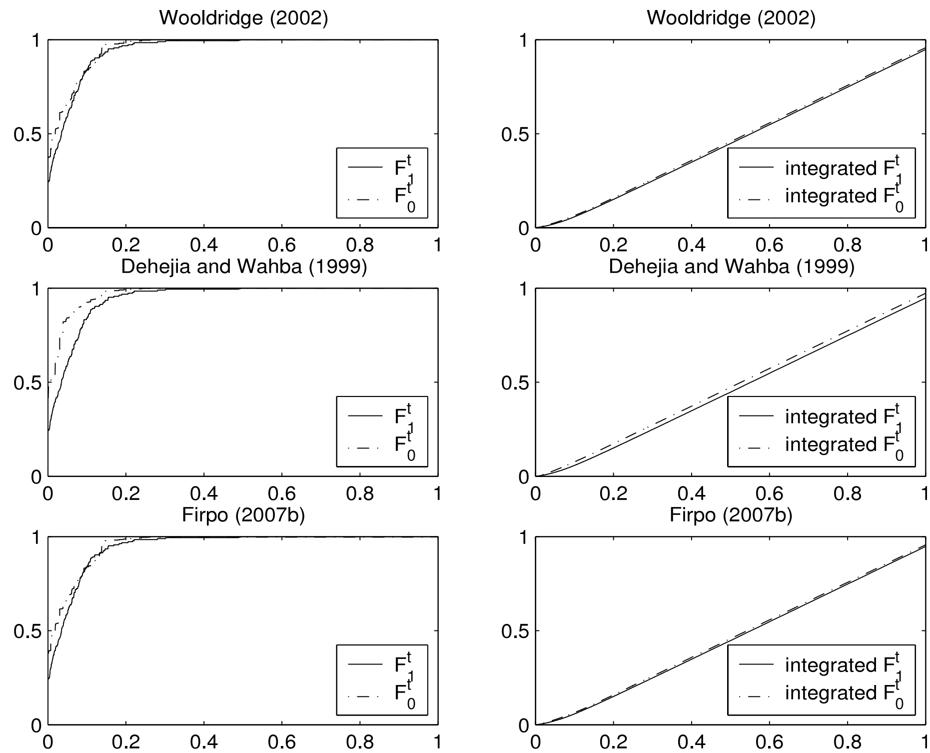

Fig. 3. Estimated F_i^t's and integrated F_i^t's for PSID-1 under different specifications.

in F_1 (F_1^t) is at least as large as that in F_0 (F_0^t) for any non-decreasing function U. Also, we have strong evidence against the hypothesis that F_0 SD2 F_1, but weaker evidence against F_0 SD1 F_1. The testing results for F_0^t SD1 F_1^t and F_0^t SD2 F_1^t depend on the estimation of the propensity score function. As a result, a unified criterion to determine the power series we should use in the SLE estimation is a very important direction for future studies.

9. Conclusion

This paper proposes IPW estimators for the distribution functions of the potential outcomes of a binary treatment under unconfoundedness assumption. The proposed estimators weakly converge to a two dimensional mean zero Gaussian process. A simulation-based method taking into account the estimation effect of the propensity score to approximate the limiting Gaussian processes is introduced. These results can be used to conduct inference on various functionals of the distributions of potential outcomes. In the paper we show how the results can be used to derive properties of estimates of the entire quantile function obtained by inversion. We use the functional delta method to derive the asymptotic properties for the quantile process. We also suggest a simulation method to approximate the limiting quantile process. The usefulness of our results is further demonstrated by constructing KS tests for stochastic dominance relations between the distributions of potential outcomes. The results are also extended to the situation where one is interested in the distributions of potential outcomes in the treated population as is useful for examining treatment effects on the treated subpopulation. A small scale simulation suggests that the proposed methods for testing stochastic dominance have good finite sample properties. Finally our results are applied to test stochastic dominance relations between potential outcomes in the National Supported Work Demonstration data and suggest that job training has a positive effect on real earnings in the stronger sense of stochastic dominance.

Acknowledgments

We thank Han Hong, an associate editor, two referees, Alberto Abadie, Jason Abrevaya, Richard Chiburis, Stefan Hoderlein, Chung-Ming Kuan, Lung-Fei Lee, Robert P. Lieli, Blaise Melly, Maxwell Stinchcombe, and seminar participants at The Ohio State University, University of Missouri at Columbia, University of Wisconsin at Milwaukee, and 2012 North American Winter Meeting of the Econometric Society for helpful comments. All errors and omissions are our own responsibility.

Appendix

We show several useful lemmas before we proceed. Let Δ be a generic constant which varies in different cases. All limits are taken as $N \to \infty$.

Lemma A.1. *Suppose Assumptions 2.1 and 3.1–3.5 hold. Then*

$$\sup_{z\in\mathcal{Z}}\left|\sqrt{N}(\widehat{F}_0(z)-F_0(z))-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\psi_0(W_i,z)-F_0(z))\right| = o_p(1),$$

$$\sup_{z\in\mathcal{Z}}\left|\sqrt{N}(\widehat{F}_1(z)-F_1(z))-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\psi_1(W_i,z)-F_0(z))\right| = o_p(1).$$

Proof of Lemma A.1. First, we replace the Y_i with $1(Y_i \le z)$ in the addendum of Hirano et al. (2003) and we can show that

$$\sup_{z\in\mathcal{Z}}\left|\frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N}\frac{T_i\cdot 1(Y_i\le z)}{\hat{p}(X_i)}-F_1(z)\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\psi_1(W_i,z)-F_1(z))\right| = o_p(1).$$

Also, by replacing Y_i's with 1's in the Proof of Theorem 1 in HIR, we can show

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{\hat{p}(X_i)}-1\right)=o_p(1),$$

which implies that

$$\sqrt{N}\left(N/\sum_{i=1}^{N}\frac{T_i}{\hat{p}(X_i)}-1\right)=o_p(1).$$

Given this, we have

$$\begin{aligned}
&\sup_{z\in\mathcal{Z}}\left|\sqrt{N}\widehat{F}_1(z)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_i\cdot 1(Y_i\le z)}{\hat{p}(X_i)}\right|\\
&\quad\le\sqrt{N}\left|\left(N/\sum_{i=1}^{N}\frac{T_i}{\hat{p}(X_i)}-1\right)\right|\cdot\sup_{z\in\mathcal{Z}}\left|\frac{1}{N}\sum_{i=1}^{N}\frac{T_i\cdot 1(Y_i\le z)}{\hat{p}(X_i)}\right|\\
&\quad=\sqrt{N}\left|\left(N/\sum_{i=1}^{N}\frac{T_i}{\hat{p}(X_i)}-1\right)\right|\cdot\left|\frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{\hat{p}(X_i)}\right|\\
&\quad=o_p(1)\cdot O_p(1)=o_p(1).
\end{aligned}$$

By the triangle inequality, we have

$$\begin{aligned}
&\sup_{z\in\mathcal{Z}}\left|\sqrt{N}(\widehat{F}_1(z)-F_1(z))-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\psi_1(W_i,z)-F_0(z))\right|\\
&\quad=o_p(1).
\end{aligned}$$

The argument for $\sqrt{N}(\widehat{F}_1(z)-F_1(z))$ is the same. □

Let $\mathcal{P}$ be the common distribution of the W_i. Recall that $\mathcal{K}_0=\{\psi_0(W,z)|z\in\mathcal{Z}\}\mathcal{K}_1=\{\psi_1(W,z)|z\in\mathcal{Z}\}$ are collections of measurable functions from W to $\mathbb{R}$. indexed by z.

Lemma A.2. *$\mathcal{K}_0$ and $\mathcal{K}_1$ are $\mathcal{P}$-Donsker.*

Proof of Lemma A.2. Let $\mathcal{Q}_1=\{1(Y\le z)\mid z\in\mathcal{Z}\}$ and it is well-known that $\mathcal{Q}_1$ is $\mathcal{P}$-Donsker. Let $\mathcal{Q}_2=\{F_0(z|X)\mid z\in\mathcal{Z}\}$ and we claim that $\mathcal{Q}_2$ is $\mathcal{P}$-Donsker by Theorem 2.3 of Kosorok (2008). Given that $F_0(z)$ is continuous on a compact set $\mathcal{Z}$ with $F_0(0)=0$, then for any $\epsilon>0$, we can find a finite collection of $0=z_0<z_1<\cdots<z_k=\bar{z}$ so that $F_0(z_j)-F_0(z_{j-1})\le\epsilon^2$ for all $1\le j\le k$ and this can be done in such a way that $k\le 1+\epsilon^{-2}$. Consider the brackets $\{(\ell_j,u_j),\,1\le j\le k\}$ with $\ell_j=F_0(z_{j-1}|X)$ and $u_j=F_0(z_j|X)$ and we find that each $F_0(z|X)$ is in at least one of these brackets. Also, we have $\|u_j-\ell_j\|_{\mathcal{P},2}\le[\|u_j-\ell_j\|_{\mathcal{P},1}]^{1/2}=[F_0(z_j)-F_0(z_{j-1})]^{1/2}\le\epsilon$, where $\|g\|_{P,r}=[\int|g(w)|^r dP(w)]^{1/r}$ for $1\le r<\infty$. Hence, the minimum number of $L_2\epsilon$-brackets to cover $\mathcal{Q}_2$ is bounded by $1+\epsilon^{-2}$. As a result, the bracketing integral, $J_{[]}(\infty,\mathcal{Q}_2,L_2(\mathcal{P}))$, is bounded[18] which implies that $\mathcal{Q}_2$ is $\mathcal{P}$-Donsker.

Let $g_1(W)=(1-T)/(1-p(X))$ and $g_2(W)=(T-p(X))/(1-p(X))$. Note that $g_1(W)$ and $g_2(W)$ are uniformly bounded and measurable functions. By Examples 2.10.7 and 2.10.10 of van der Vaart and Wellner (1996), $\mathcal{K}_0$ is $\mathcal{P}$-Donsker. Similar arguments apply to $\mathcal{K}_1$. □

Proof of Theorem 3.6. Note that by van der Vaart (2000, p. 270), the Cartesian product of two Donsker classes of functions is also a Donsker class. Hence, Lemma A.2 and the Donsker's Theorem imply Theorem 3.6.

Proof of Theorem 3.8. The result follows when we apply the functional delta method to the quantile map, e.g., Section 3.9 of van der Vaart and Wellner (1996). □

[18] The definition of $J_{[]}(\infty,\mathcal{Q}_2,L_2(\mathcal{P}))$ can be found in p. 17 of Kosorok (2008).

Proof of Lemma 4.1. By definition of $\widehat{F}_1(z|x)$, it not hard to show that it is monotonically increasing. Then we claim $\sup_{z\in\mathcal{Z},x\in\mathcal{X}}|\widetilde{F}_1(z|x)-F_1(z|x)|=o_p(1)$. For a matrix A, let $\|A\|$ denote the matrix norm of A such that $\|A\|=\sqrt{tr(A'A)}$. Define

$$\Phi_K(z)=\frac{1}{N}\sum_{i=1}^{N}\frac{1(Y_i\le z)T_i}{p(X_i)}R^K(X_i),$$

$$\widehat{\Phi}_K(z)=\frac{1}{N}\sum_{i=1}^{N}\frac{1(Y_i\le z)T_i}{\hat{p}(X_i)}R^K(X_i),$$

$$\xi_K=\frac{1}{N}\sum_{i=1}^{N}R^K(X_i)R^K(X_i)'.$$

Because $p(X)$ is bounded away from 0, the conditional variance of $1(Y_i\le z)T_i/p(X_i)$ conditional on $X=x$ is bounded. Hence, the bound in Newey (1997) for series estimators applies:

$$\begin{aligned}
&\sup_{z\in\mathcal{Z},x\in\mathcal{X}}\left|\Phi_K(z)_K'^{-1}R^K(X_i)-F_1(z|x)\right|\\
&\quad\le\Delta_1\zeta(K)O_p\left(\sqrt{\frac{\zeta(K)}{N}}\right)+\Delta_2K^{\frac{-s'}{r}}
\end{aligned}$$

where s' is the number of continuous derivatives of $F_1(z|x)$ and $\zeta(K)=\sup_{x\in\mathcal{X}}\|R^K(x)\|$. Note that the compactness of $\mathcal{X}$ is needed to obtain the uniform bound over $\mathcal{X}$. By similar argument of HIR,

$$\begin{aligned}
&\sup_{z\in\mathcal{Z},x\in\mathcal{X}}|\widetilde{F}_1(z|x)-F_1(z|x)|\\
&\quad\le\sup_{z\in\mathcal{Z},x\in\mathcal{X}}|\widehat{\Phi}_K(z)_K'^{-1}R^K(X_i)-\Phi_K(z)_K'^{-1}R^K(X_i)|\\
&\qquad+\sup_{z\in\mathcal{Z},x\in\mathcal{X}}|\Phi_K(z)_K'^{-1}R^K(X_i)-F_1(z|x)|=o_p(1).
\end{aligned}$$

Without loss of generality, we assume that $\widetilde{F}_1(z|x)$ is bounded between 0 and 1.[19] For a given x, Suppose Y_h is the first point at which $\widetilde{F}_1(z|x)$ jumps down. Then for $y_h\le z<y_{h+1}$, $\widehat{F}_1(z|x)=\widetilde{F}_1(y_{h-1}|x)>\widetilde{F}_1(y_h|x)=\widetilde{F}_1(z|x)$ and for $y_{h-1}\le z<y_h$, $\widehat{F}_1(z|x)=\widetilde{F}_1(y_{h-1}|x)$. For $y_h\le z<y_{h+1}$, if $\widehat{F}_1(z|x)\le F_1(z|x)$, then we have $F_1(z|x)-\widehat{F}_1(z|x)>F_1(z|x)-\widetilde{F}_1(z|x)>0$. If $\widehat{F}_1(z|x)>F_1(z|x)$, then we have $\widetilde{F}_1(y_{h-1}|x)-F_1(y_{h-1}|x)-\widehat{F}_1(z|x)-F_1(z|x)>0$. These imply that

$$\begin{aligned}
|\widehat{F}_1(z|x)-F_1(z|x)|\le\max\{&|\widetilde{F}_1(y_{h-1}|x)\\
&-F_1(y_{h-1}|x)|,|\widetilde{F}_1(z|x)-F_1(z|x)|\}.
\end{aligned}$$

As a result,

$$\sup_{0\le z<y_{h+1}}|\widehat{F}_1(z|x)-F_1(z|x)|\le\sup_{0\le z<y_{h+1}}|\widetilde{F}_1(z|x)-F_1(z|x)|.$$

Then by induction, we can show that

$$\sup_{z\in\mathcal{Z}}|\widehat{F}_1(z|x)-F_1(z|x)|\le\sup_{z\in\mathcal{Z}}|\widetilde{F}_1(z|x)-F_1(z|x)|=o_p(1).$$

Also, $\sup_{z\in\mathcal{Z},x\in X}|\widehat{F}_0(z|x)-F_0(z|x)|=o_p(1)$. □

Proof of Lemma 4.2. We rewrite $\Psi_1^u(z)$ as

$$\begin{aligned}
\Psi_1^u(z)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}U_i\Bigg(&\frac{T_i\cdot 1(Y_i\le z)}{p(X_i)}\\
&-F_1(z)-(T-p(X_i))\frac{F_1(z|X_i)}{p(X_i)}\Bigg)
\end{aligned}\qquad(10)$$

[19] If $\widetilde{F}_1(z|x)<0$, then $0>0-F_1(z|x)>\widetilde{F}_1(z|x)-F_1(z|x)$ or $|0-F_1(z|x)|<|\widetilde{F}_1(z|x)-F_1(z|x)|$. If we trim $\widetilde{F}_1(z|x)$ at 0, then the deviation from $F_1(z|x)$ is smaller. Similar argument applies when $\widetilde{F}_1(z|x)>1$.

$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i \left(\frac{T_i \cdot 1(Y_i \le z)}{\hat{p}(X_i)} - \frac{T_i \cdot 1(Y_i \le z)}{p(X_i)} \right) \quad (11)$$

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i \left((p(X_i) - \hat{p}(X_i)) \frac{\widehat{F}_1(z|X_i)}{\hat{p}(X_i)} \right) \quad (12)$$

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i \left((T_i - p(X_i)) \left(\frac{\widehat{F}_1(z|X_i)}{\hat{p}(X_i)} - \frac{F_1(z|X_i)}{p(X_i)} \right) \right) \quad (13)$$

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i \left(\widehat{F}_1(z) - F_1(z)\right). \quad (14)$$

We first show that the process in (11) weakly converges to a zero process conditional on the sample path $\mathcal{W}$ w.p.a. 1, i.e.,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i \left(\frac{T_i \cdot 1(Y_i \le z)}{\hat{p}(X_i)} - \frac{T_i \cdot 1(Y_i \le z)}{p(X_i)} \right) \overset{p}{\Rightarrow} 0. \quad (15)$$

Note that the proof for (12) and (13) is similar and we omit it. According to Footnote 10 in page 13, we need to show that for any subsequence k_N of N, there exists a further subsequence ℓ_N of k_N such that

$$\frac{1}{\sqrt{\ell_N}} \sum_{i=1}^{\ell_N} U_i \left(\frac{T_i \cdot 1(Y_i \le z)}{\hat{p}_{\ell_N}(X_i)} - \frac{T_i \cdot 1(Y_i \le z)}{p(X_i)} \right) \overset{a.s.}{\Rightarrow} 0. \quad (16)$$

Note that for any subsequence k_N of N, there exists a further subsequence ℓ_N of k_N such that $\sup_{x\in\mathcal{X}} |\hat{p}_{\ell_N}(x) - p(x)| \overset{a.s.}{\to} 0$ where $\hat{p}_{\ell_N}(x)$ denotes the estimator at ℓ_N. Define

$$\mathcal{W}_\ell \equiv \left\{ \omega \in \mathcal{W} : \sup_{x\in\mathcal{X}} |\hat{p}_{\ell_N}(x)(\omega) - p(x)| \to 0 \right\}, \quad (17)$$

where $\hat{p}_{\ell_N}(x)(\omega)$ denotes the realization at ω. It is true that $P(\mathcal{W}_\ell) = 1$. Note that for any $\delta > 0$ and for any $\omega \in \mathcal{W}_\ell$, we have $\sup_{x\in\mathcal{X}} |\hat{p}_{\ell_N}(x)(w) - p(x)| \le \delta$ eventually. When we pick $\delta = 1/2 \min\{\underline{p}, 1 - \bar{p}\}$, it is obvious that for any $\omega \in \mathcal{W}_\ell$, $0 < \underline{p} - \delta \le \hat{p}_{\ell_N}(x)(w) \le \bar{p} + \delta < 1$, and $1/\hat{p}_{\ell_N}(x)(w)$ and $1/(1-\hat{p}_{\ell_N}(x)(w))$ are bounded by M for some M large enough and for all ℓ_N large enough.

For any $\omega \in \mathcal{W}_\ell$, define

$$f_{\ell_N i}(U_i, z|\omega) = \frac{U_i}{\sqrt{\ell_N}} \left(\frac{T_i \cdot 1(Y_i \le z)}{\hat{p}_{\ell_N}(X_i)(\omega)} - \frac{T_i \cdot 1(Y_i \le z)}{p(X_i)} \right),$$

$$\chi_{\ell_N}(z|\omega) = \sum_{i=1}^{\ell_N} f_{\ell_N i}(U_i, z|\omega).$$

We first show that the triangular array $\{f_{\ell_N i}(U_i, z|\omega)\}$ satisfies (i)–(v) of Theorem 10.6 of Pollard (1990). Note that we have conditioned on the sample path ω, so the randomness is from the U_i's which is independent of the sample path ω. Define $F_{\ell_N,i} = M|U_i|/\sqrt{\ell_N}$ is the envelope of $f_{\ell_N i}(U_i, z|\omega)$ since T_i and $1(Y_i \le z)$ are bounded, and $p(X_i)$ and $\hat{p}_{\ell_N}(X_i)(\omega)$ are bounded away from 0 and 1. By the definition of $F_{\ell_N,i}$, (iii) and (iv) hold. For $z_1 \le z_2$,

$$\begin{aligned}
&E_u[\chi_{\ell_N}(z_1|\omega)\chi_{\ell_N}(z_2|\omega)] \\
&= \frac{1}{\ell_N} \sum_{i=1}^{\ell_N} E_u \left[U_i^2 T_i \cdot 1(Y_i \le z_1) \left(\frac{1}{\hat{p}_{\ell_N}(X_i)(\omega)} - \frac{1}{p(X_i)} \right)^2 \right] \\
&= \frac{1}{\ell_N} \sum_{i=1}^{\ell_N} T_i \cdot 1(Y_i \le z_1) \left(\frac{1}{\hat{p}_{\ell_N}(X_i)(\omega)} - \frac{1}{p(X_i)} \right)^2 \\
&= \frac{1}{\ell_N} \sum_{i=1}^{\ell_N} \frac{T_i \cdot 1(Y_i \le z_1)}{\hat{p}^2_{\ell_N}(X_i)(\omega) p^2(X_i)} \left(p(X_i) - \hat{p}_{\ell_N}(X_i)(\omega)\right)^2 \\
&\le \Delta \sup_x |\hat{p}_{\ell_N}(x)(\omega) - p(x)|^2 \to 0
\end{aligned}$$

where the convergence follows from the fact that $\omega \in \mathcal{W}_\ell$.

The first equality follows from the fact that U_i's are mutually independent with mean 0, $T_i^2 = T_i$ and $1(Y_i \le z_1)1(Y_i \le z_2) = 1(Y_i \le z_1)$. The second one follows from that $E[U_i^2] = 1$ and the last inequality holds because $\hat{p}_{\ell_N}(x)(\omega)$ and $p(x)$ are bounded away from 0 when ℓ_N is large enough and $\sup_x |\hat{p}_{\ell_N}(x)(\omega) - p(x)| \to 0$. As a result, (ii) holds. Define

$$\rho_{\ell_N}(z_1, z_2) = \left(\sum_{i=1}^{\ell_N} E_u \left[f_{\ell_N i}(U_i, z_1|\omega) - f_{\ell_N i}(U_i, z_2|\omega) \right] \right)^{\frac{1}{2}}.$$

For any $z_1 < z_2$,

$$\begin{aligned}
\rho_{\ell_N}(z_1, z_2)^2 &= \sum_{i=1}^{\ell_N} E_u \left[f_{\ell_N i}(U_i, z_1|\omega) - f_{Ni}(U_i, z_2|\omega) \right]^2 \\
&= \frac{1}{\ell_N} \sum_{i=1}^{\ell_N} E_u \left[U_i^2 \cdot T_i \cdot 1(z_1 < Y_i \le z_2) \left(\frac{1}{\hat{p}_{\ell_N}(X_i)(\omega)} - \frac{1}{p(X_i)} \right)^2 \right] \\
&\le \Delta \sup_x |\hat{p}_{\ell_N}(x)(\omega) - p(x)|^2 \to 0.
\end{aligned}$$

$\rho_{\ell_N}(z_1, z_2)$ converges uniformly to 0 over z_1 and z_2 and (v) holds.

Finally, $\{f_{\ell_N i}(U_i, z|\omega)\}$ is manageable with respect to $F_{\ell_N,i} = M|U_i|/\sqrt{\ell_N}$ because $1(Y_i \le z_1)$ forms a Vapnik–Chervonenkis class of functions and (i) holds.

We have shown that for all $\omega \in \mathcal{W}_\ell$, the triangular array $\{f_{\ell_N i}(U_i, z|\omega)\}$ satisfies (i)–(v) of Theorem 10.6 of Pollard (1990) and this implies that $\omega \in \mathcal{W}_\ell$

$$\frac{1}{\sqrt{\ell_N}} \sum_{i=1}^{\ell_N} U_i \left(\frac{T_i \cdot 1(Y_i \le z)}{\hat{p}_{\ell_N}(X_i)(\omega)} - \frac{T_i \cdot 1(Y_i \le z)}{p(X_i)} \right) \Rightarrow 0. \quad (18)$$

Therefore, (16) holds and (15) follows.[20]

By the same proof of Proposition 2 of Barrett and Donald (2003), the process in (14) converges to a zero process. Hence,

$$\begin{aligned}
\Psi_1^u(z) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i &\left(\frac{T_i \cdot 1(Y_i \le z)}{p(X_i)} - F_1(z) \right. \\
&\left. - (T - p(X_i)) \frac{F_1(z|X_i)}{p(X_i)} \right) \overset{p}{\Rightarrow} 0.
\end{aligned}$$

Finally, Corollary 2.9.3 of van der Vaart and Wellner (1996) implies that

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i \left(\frac{T_i \cdot 1(Y_i \le z)}{p(X_i)} - F_1(z) - (T - p(X_i)) \frac{F_1(z|X_i)}{p(X_i)} \right) \\
&\overset{p}{\Rightarrow} \Psi_1(z),
\end{aligned}$$

and $\Psi_1^u(z) \overset{p}{\Rightarrow} \Psi_1(z)$ follows. By similar argument, we can show that $\Psi_0^u(z) \overset{p}{\Rightarrow} \Psi_0(z)$ and Lemma 4.2 holds. □

Proof of Lemma 4.4. Without loss of generality, we assume $[z_{1\ell}, z_{1u}] = \mathcal{Z}$. Define

$$\ddot{f}_1(z) = \frac{1}{Nh} \sum_{i=1}^{N} \frac{T_i}{p(X_i)} K\left(\frac{Y_i - z}{h} \right)$$

and $Z_N = [h, \bar{z} - h]$. By similar argument in Lemma 2 of Donald et al. (2012) and Masry (1996), we can show that $\sup_{z\in Z_N} |\ddot{f}_1(z) - f_1(z)| = o_p(1)$. On the other hand,

[20] We thank a referee for pointing out that if U_i's are standard normals, we can shorten the proof significantly by applying the Dudley's entropy integral bound because conditioning on $\mathcal{W}$, $\chi_N(z|\mathcal{W})$ is a Gaussian process. We do not use this approach because U_i's are allowed to be i.i.d. random variables with mean equal to 0 and variance 1.

$$\begin{aligned}
&\sup_{z\in Z_N}|\hat{f}_1(z)-\ddot{f}_1(z)| \\
&= \sup_{z\in Z_N}\left|\frac{1}{Nh}\sum_{i=1}^{N}\left(\frac{T_i}{\hat{p}(X_i)}-\frac{T_i}{p(X_i)}\right)K\left(\frac{Y_i-z}{h}\right)\right| \\
&\le \Delta\sup_{x\in\mathcal{X}}|\hat{p}(x)-p(x)|\sup_{z\in Z_N}\frac{1}{Nh}\sum_{i=1}^{N}\left|K\left(\frac{Y_i-z}{h}\right)\right| \\
&= \Delta\cdot o_p(1)\cdot O_p(1)=o_p(1).
\end{aligned}$$

The second line follows from the fact that $p(x)$ and $\hat{p}(x)$ are both bounded above and bounded away from 0. It can be shown by similar argument for $\ddot{f}_1(z)$ that $\sum_{i=1}^{N}|K((Y_i-z)/h)|/(Nh)$ is a uniformly consistent estimator for $f(z)$, the probability density function of Y, on Z_N and this implies that it is $O_p(1)$. Given $\sup_{x\in\mathcal{X}}|\hat{p}(x)-p(x)|=o_p(1)$, the last line follows. Hence, by the triangle inequality

$$\begin{aligned}
\sup_{z\in Z_N}|\hat{f}_1(z)-f_1(z)| &\le \sup_{z\in Z_N}|\hat{f}_1(z)-\ddot{f}_1(z)| \\
&\quad+\sup_{z\in Z_N}|\ddot{f}_1(z)-f_1(z)|=o_p(1).
\end{aligned}$$

On the other hand, for all $z\in[0,h]$ and for some $M>0$,

$$\begin{aligned}
|\hat{f}_1(z)-f_1(z)| &= |\hat{f}_1(h)-f_1(z)| \\
&\le |\hat{f}_1(h)-f_1(h)|+|f_1(h)-f_1(z)| \\
&\le |\hat{f}_1(h)-f_1(h)|+\Delta h.
\end{aligned}$$

The first equality holds by the definition of $\hat{f}_1(z)$ for $z\in[0,h]$. The second inequality holds by the triangle inequality. The second term of the last line holds by the fact that $f_1(z)$ is continuously differentiable of order 2 on $\mathcal{Z}$. This is sufficient to show that $\sup_{z\in[0,h]}|\hat{f}_1(z)-f_1(z)|=o_p(1)$ and $\sup_{z\in[\bar{z}-h,\bar{z}]}|\hat{f}_1(z)-f_1(z)|=o_p(1)$. These imply that $\sup_{z\in\mathcal{Z}}|\hat{f}_1(z)-f_1(z)|=o_p(1)$. The result regarding $\hat{f}_0(z)$ can be shown similarly. □

Proof of Theorem 4.5. First, it is straightforward to show that

$$-\frac{\Psi_0^u(q_0(\tau))}{f_0(q_0(\tau))}\overset{P_{\mathcal{W}}}{\Rightarrow}\mathcal{Q}_0(\tau).$$

Note that

$$\begin{aligned}
&\sup_{\tau\in[0,1]}|\hat{f}_0(\hat{q}_0(\tau))-f_0(q_0(\tau))| \\
&\le \sup_{\tau\in[0,1]}|\hat{f}_0(\hat{q}_0(\tau))-f_0(\hat{q}_0(\tau))|+\sup_{\tau\in[0,1]}|f_0(\hat{q}_0(\tau))-f_0(q_0(\tau))| \\
&\le \sup_{z\in\mathcal{Z}}|\hat{f}_0(z)-f_0(z)|+\Delta\sup_{\tau\in[0,1]}|\hat{q}_0(\tau)-q_0(\tau)|=o_p(1).
\end{aligned}$$

The first inequality follows from the triangle inequality. The second inequality follows from the fact that $\hat{q}_0(\tau)\in\mathcal{Z}$ for all τ and $f_0(z)$ is continuously differentiable of order 2 on $\mathcal{Z}$. The result follows from the fact that both $\hat{f}_0(z)$ and $\hat{q}_0(\tau)$ are uniformly consistent estimators. Conditioning on the sample path $\mathcal{W}$ w.p.a. 1, it is true that $\sup_{\tau\in[0,1]}|\Psi_0^u(\hat{q}_0(\tau))-\Psi_0^u(q_0(\tau))|=o_p(1)$ by the equicontinuity of $\Psi_0^u(z)$ and the uniform consistency of $\hat{q}_0(\tau)$. As a result,

$$\begin{aligned}
&\sup_{\tau\in[0,1]}\left|\frac{\Psi_0^u(\hat{q}_0(\tau))}{\hat{f}_0(\hat{q}_0(\tau))}-\frac{\Psi_0^u(q_0(\tau))}{f_0(q_0(\tau))}\right| \\
&\le \sup_{\tau\in[0,1]}\left|\frac{\Psi_0^u(\hat{q}_0(\tau))}{\hat{f}_0(\hat{q}_0(\tau))}-\frac{\Psi_0^u(\hat{q}_0(\tau))}{f_0(q_0(\tau))}\right| \\
&\quad+\sup_{\tau\in[0,1]}\left|\frac{\Psi_0^u(\hat{q}_0(\tau))}{f_0(q_0(\tau))}-\frac{\Psi_0^u(q_0(\tau))}{f_0(q_0(\tau))}\right| \\
&\le \Delta\sup_{\tau\in[0,1]}|\hat{f}_0(\hat{q}_0(\tau))-f_0(q_0(\tau))|\sup_{\tau\in[0,1]}|\Psi_0^u(\hat{q}_0(\tau))| \\
&\quad+\Delta\sup_{\tau\in[0,1]}|\Psi_0^u(\hat{q}_0(\tau))-\Psi_0^u(q_0(\tau))| \\
&= \Delta\cdot o_p(1)\cdot O_p(1)+\Delta\cdot o_p(1)=o_p(1).
\end{aligned}$$

This is sufficient to show that

$$\mathcal{Q}_0^u(\tau)\equiv-\frac{\Psi_0^u(\hat{q}_0(\tau))}{\hat{f}_0(\hat{q}_0(\tau))}\overset{P_{\mathcal{W}}}{\Rightarrow}\mathcal{Q}_0(\tau).$$

The result regarding $\mathcal{Q}_1^u(\tau)^u$ can be shown similarly. □

Proof of Theorem 5.1. By continuous mapping theorem, we have $\overline{S}_u\xrightarrow{D}\overline{S}$ given $\mathcal{W}$ in probability. Note that $P(\overline{S}\le 0)\le 1/2$ and the distribution function of $\overline{S}$ is continuous by Tsirel'son (1975). It follows that if $\alpha_0<1/2$, then $c_0>0$. Therefore, $\hat{c}$ converges in probability to $c_0>0$.

Under the null,

$$\limsup P(\widehat{S}_N>\hat{c})\le\limsup P(\overline{S}_N>\hat{c})=P(\overline{S}>c_0)=\alpha_0.$$

The first equality follows from the fact that $\overline{S}_N\xrightarrow{D}\overline{S}$, $\hat{c}\xrightarrow{p}c_0$ and the CDF of $\overline{S}$ is continuous at c_0. The second equality follows from the fact that $\overline{S}_N\xrightarrow{D}\overline{S}$ and the definition of c_0. This shows the first part of Theorem 5.1.

Suppose $F_1(z^*)>F_0(z^*)$ for some $z^*\in\mathcal{Z}$. Then

$$\begin{aligned}
\widehat{S}_N &= \sqrt{N}\sup_{z\in\mathcal{Z}}\left(\widehat{F}_1(z)-\widehat{F}_0(z)\right)\ge\sqrt{N}\left(\widehat{F}_1(z^*)-\widehat{F}_0(z^*)\right) \\
&= \sqrt{N}\left(\widehat{F}_1(z^*)-\widehat{F}_0(z^*)-(F_1(z^*)-F_0(z^*))\right) \\
&\quad+\sqrt{N}\left(F_1(z^*)-F_0(z^*)\right)\xrightarrow{p}\infty.
\end{aligned}$$

The last line holds since the first term in the third line is asymptotically normal which is bounded in probability and the second term diverges to infinity. The second part follows from the fact that $\widehat{S}_N\xrightarrow{p}\infty$ when the null hypothesis is wrong, but $\hat{c}\xrightarrow{p}c_0<\infty$. □

Proof of Theorem 5.2. By the same proof of Theorem 3.6, we have $\overline{S}_N\xrightarrow{D}\sup_{z\in\mathcal{Z}}(\Psi_1(z)-\Psi_0(z)+\delta(z))$ and by the same proof of Theorem 5.1, the critical value will converge to c_0, the $(1-\alpha_0)$-th quantile of $\sup_{z\in\mathcal{Z}}(\Psi_1(z)-\Psi_0(z))$. Given that $\delta(z)>0$ for some z, this implies that $\sup_{z\in\mathcal{Z}}(\Psi_1(z)-\Psi_0(z)+\delta(z))$ first order stochastically dominates $\sup_{z\in\mathcal{Z}}(\Psi_1(z)-\Psi_0(z))$. This implies that $\lim_{N\to\infty}P(\overline{S}_N>\hat{c})=P(\sup_{z\in\mathcal{Z}}(\Psi_1(z)-\Psi_0(z)+\delta(z))>c_0)\ge\alpha_0$. □

Proof of Theorem 6.2. The proofs are similar to those for Theorems 3.6 and 3.8 and we omit it.

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., Imbens, G., 2011. Bias corrected matching estimators for average treatment effects. Journal of Business and Economic Statistics 29, 1–11.
Anderson, G., 1996. Nonparametric tests for stochastic dominance. Econometrica 64, 1183–1193.
Andrews, D.W.K., Shi, X., 2013. Inference based on conditional moment inequalities. Econometrica 81, 609–666.
Barrett, G.F., Donald, S.G., 2003. Consistent tests for stochastic dominance. Econometrica 71, 71–104.
Bennett, C.J., 2009. Consistent integral-type tests for stochastic dominance. Working paper.
Chernozhukov, V., Fernandez-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78, 559–575.
Davidson, R.S., Duclos, J.-Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica 68, 1435–1464.
Dehejia, R., Wahba, S., 1999. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. Journal of the American Statistical Association 94, 1053–1062.
Donald, S.G., Hsu, Y.-C., 2010. Improving the power of tests of stochastic dominance. Working paper.
Donald, S.G., Hsu, Y.-C., Barrett, G.F., 2012. Incorporating covariates in the measurement of welfare and inequality: methods and applications. The Econometrics Journal 15, C1–C30.
Donald, S.G., Hsu, Y.-C., Lieli, R.P., 2011. Testing the unconfoundedness assumption via inverse probability weighted estimators of (L)ATT. Working paper.

Firpo, S., 2007a. Efficient semiparametric estimation of quantile treatment effects. Econometrica 75, 259–276.
Firpo, S., 2007b. Sumplement to efficient semiparametric estimation of quantile treatment effects. Econometrica.
Firpo, S., 2010. Identification and estimation of distributional impacts of interventions using changes in inequality measures. Working paper.
Hahn, J., 1988. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica 66, 315–331.
Heckman, J., Hotz, J., 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. Journal of the American Statistical Association 84, 862–874.
Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998a. Characterizing selection bias using experimental data. Econometrica 66, 1017–1098.
Heckman, J., Ichimura, H., Todd, P., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training program. Review of Economic Studies 64, 605–654.
Heckman, J., Ichimura, H., Todd, P., 1998b. Matching as an econometric evaluations estimator. Review of Economic Studies 65, 261–294.
Hirano, K., Imbens, G., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161–1189.
Hsu, Y.-C., 2011. Consistent tests for conditional treatment effects. Working paper.
Ichimura, H., Linton, O., 2005. Asymptotic expansions for some semiparametric program evaluation estimators. In: Andrews, D.W.K., Stock, J. (Eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press.
Imbens, G.W., 2003. Sensitivity to exogeneity assumptions in program evaluation. American Economic Review, Papers and Proceedings 93, 126–132.
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., Wooldridge, J.W., 2009. Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47, 5–86.
Khan, S., Tamer, E., 2010. Irregular identification, support conditions, and inverse weight estimation. Econometrica 6, 2021–2042.
Kosorok, M.R., 2008. Introduction to Empirical Processes and Semiparametric Inference. Springer Science+Business Media.
LaLonde, R., 1986. Evaluating the econometric evaluations of training programs with experimental data. American Economic Review 76, 604–620.
Lee, S., Whang, J.-Y., 2009. Nonparametric tests of conditional treatment effects. Working paper.
Linton, O., Maasoumi, E., Whang, Y.-J., 2005. Consistent testing for stochastic dominance under general sampling schemes. Review of Economic Studies 72, 735–765.
Linton, O., Song, K., Whang, Y.-J., 2010. An improved bootstrap test of stochastic dominance. Journal of Econometrics 154, 186–202.
Maier, M., 2011. Tests for distributional treatment effects under unconfoundedness. Economics Letters 110, 49–51.
Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. Journal of Time Series Analysis 17, 571–599.
McFadden, D., 1989. Testing for stochastic dominance. In: Fombay, T.B., Seo, T.K. (Eds.), Studies in Economics of Uncertainty: in Honor of Josef Hadar. Springer, New York, Berlin, London and Tokyo.
Newey, W., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.
Pollard, D., 1990. Empirical Processes: Theory and Application. In: CBMS Conference Series in Probability and Statistics, vol. 2. Institute of Mathematical Statistics, Hayward, CA.
Rosenbaum, P., 1987. Model-based direct adjustment. Journal of the American Statistical Association 82, 387–394.
Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.
Rosenbaum, P., Rubin, D., 1985. Reducing bias in observational studies using subclassfication on the propensity score. Journal of the American Statistical Association 79, 516–524.
Rubin, D., Thomas, N., 1996. Matching using estimated propensity scores: relating theory to practice. Biometrics 52, 249–264.
Shen, S., 2011. Tests for distributional partial effects. Working paper.
Smith, J.A., Todd, P.E., 2001. Reconciling conflicting evidence on the performance of propensity-score matching methods. American Economic Review, Papers and Proceedings 91, 112–118.
Smith, J.A., Todd, P.E., 2005. Does matching address LaLonde's critique of nonexperimental estimators. Journal of Econometrics 125, 305–353.
Tsirel'son, V.S., 1975. The density of the distribution of the maximum of a Gaussian process. Theory of Porbability and its Application 16, 847–856.
van der Vaart, A., 2000. Asymptotic Statistics. Cambridge University Press.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes: With Application to Statistics. Springer-Verlag, New York.
Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.