Department of Economics
Seminar Series

Yu-Chin Hsu
University of Texas at Austin

"Testing for Stochastic Dominance in Treatment Effects"

Monday,
January 25, 2010
10:00 am
N201/202 Memorial Union

Testing for Stochastic Dominance in Treatment Effects

Yu-Chin Hsu

Department of Economics
University of Texas at Austin

Job Market Paper

First version: January 31, 2009
This version: November 20, 2009

∗ Acknowledgement: I thank Jason Abrevaya, Richard Chiburis, Chung-Ming Kuan, Robert Lieli and Maxwell Stinchcombe for their useful comments. I especially thank Stephen Donald for his guidance. All errors and omissions are my own responsibility.

Abstract

In this paper, we propose Kolmogorov-Smirnov type tests for stochastic dominance relations between the potential outcomes of a binary treatment under the unconfoundedness assumption. Our stochastic dominance tests compare every point of the cumulative distribution functions (CDF), so they can fully utilize all information in the distributions. For first order stochastic dominance, the test statistic is defined as the supremum of the difference of two inverse-probability-weighting estimators for the CDFs of the potential outcomes. The critical values are approximated based on a simulation method. We show that our test has good size properties and is consistent in the sense that it can detect any violation of the null hypothesis asymptotically. First order stochastic dominance tests in the treated subpopulation, and higher order stochastic dominance tests in the whole population and among the treated are shown to share the same properties. The tests are applied to evaluate the effect of a job training program on incomes, and we find that job training has a positive effect on real earnings. Finally, we extend our tests to cases in which the unconfoundedness assumption does not hold.

JEL classification: C01, C12, C21

Keywords: Hypothesis testing, stochastic dominance, treatment effects, propensity score.

1 Introduction

Economists are often interested in the effect of a binary treatment or policy on some outcome variable. Many studies focus on estimating the average treatment effect (ATE) and average treatment effect on the treated (ATT) when the treatment assignment is unconfounded (e.g., Rosenbaum and Rubin (1983, 1985), Heckman, Ichimura, and Todd (1997, 1998), Heckman, Ichimura, Smith and Todd (1998), Hahn (1998), Rosenbaum (1987), Rubin, and Thomas (1996), Hirano, Imbens, and Ridder (2003, HIR)).

While ATE and ATT describe the mean effect of a treatment, they do not fully characterize its distributional impact. Nevertheless, it is often important to obtain this information. For example, one might be interested in the effects of a job training program on the lower or upper tails of the distributions of the potential outcomes. Therefore, the overall quantile treatment effects (QTE) and quantile treatment effects on the treated (QTT) have been proposed to evaluate heterogeneous impacts on different points of the potential outcome distributions (e.g., Firpo (2007a, 2007b)). The distributional impact of a treatment can also be captured by taking the difference of inequality measures of the distributions such as coefficient of variation, interquartile range, Theil index and Gini coefficient. Firpo (2008) defines these distribution impacts as the inequality treatment effects (ITE) and the inequality treatment effects on the treated (ITT).

QTE, QTT, ITE and ITT do not fully utilize the information in the distributions, since the comparison is made based on a certain aspect of the distributions. As a result, examining the stochastic dominance relations between the potential outcomes is a more desirable approach, since it compares every point of the cumulative distribution functions (CDF). The importance of analyzing various forms of stochastic relations between income distributions has been emphasized in many inequality and poverty studies (e.g., McFadden (1989), Anderson (1996), Davidson and Duclos (2000), and Barrett and Donald (2003)). The stochastic dominance relations and the social welfare orderings are closely related. For example, if CDF H_1 first order stochastic dominates CDF H_2, then for any social welfare function based on a monotonically increasing utility function, the social welfare of H_1 is greater than or equal to that of H_2. Also, if H_1 second order stochastic dominates H_2, then for any social welfare function based on a monotonically increasing and concave utilitarian function, the social welfare of H_1 is greater than or equal to that of H_2.

The aim of this paper is to propose Kolmogorov-Smirnov tests for the stochastic dominance relations between the potential outcome distributions. Under an unconfoundedness assumption, we propose inverse propensity score weighting (IPW) estimators for the CDFs of the potential outcomes based on identification results similar to those in HIR. A test statistic for first order stochastic dominance is defined as the supremum of the difference of two estimated CDFs with proper scaling. The critical value (p-value) is approximated by a Monte-Carlo based method that takes into account the estimation effect of the propensity score. We prove that our test can control the size under the desired level asymptotically and is consistent. Tests for first order stochastic dominance on the treated, and tests for higher order stochastic dominance on the whole group and on the treated are also discussed and are shown to share the same properties.

Part of our contribution is to extend Barrett and Donald's (2003) tests to allow for covariates in the model. For Barrett and Donald's (2003) tests to work, it is required that the treatment is completely random. Random assignment implies the unconfoundedness assumption for any given set of covariates, so our tests are also valid if treatment is randomly assigned but covariates are still used. On the other hand, if the random assignment assumption is violated, then their tests may give wrong conclusions, but our tests will still be valid given that the unconfoundedness assumption holds for a given set of covariates. In this sense, our test is more robust than theirs.

We examine the finite-sample properties of our tests by conducting small-scale simulations, and the results confirm our theoretical findings. We apply our tests to the data from National Supported Work Demonstration job training program. We evaluate the effect of the job training program on real earnings. We find that real earnings under job training stochastically dominates real earnings without job training and we find evidence against the reverse relations. We conclude that there is a positive effect of job training on earnings and moreover that using any social welfare function and not just the mean would lead to this conclusion.

Without the unconfoundedness assumption, treatment assignment is correlated with the potential outcomes even conditional on the covariates. As a result, the difference between the CDFs of the potential outcomes is not identified. However, if a valid binary instrument is available, the difference between the CDFs can be identified in the subpopulation of compliers. Hence, we extend our tests to test the stochastic dominance relations among compliers. These are referred as the local stochastic dominance relations.

The remainder of this paper is organized as follows. In Section 2, we formulate the hypothesis for the first order stochastic dominance relation between the potential outcomes. We discuss the identification and estimation of the CDFs of the the potential outcomes and present our test statistic and decision rule for testing first order stochastic dominance. Section 3 presents the asymptotic results for the estimators of the CDFs and the test statistic. In Section 4, we introduce simulation methods for the critical value (p-value) of the test and prove the size and power properties of our test. Section 5 extends the results to first order stochastic dominance tests among the treated. Section 6 extends the first order stochastic dominance tests to higher order stochastic dominance tests. Monte-Carlo simulation results are summarized in Section 7, and the empirical application is presented in Section 8. In Section 9, we extend our results to cases in which the unconfoundedness assumption does not hold. Section 10 concludes, and all mathematical proofs are deferred to the Appendix.

2 Hypothesis and Test Statistic for First Order Stochastic Dominance

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. It is known that H_1 first order stochastically dominates (SD1) H_2 if and only if (iff) the welfare of H_1 will be greater than or equal to that of H_2 for any welfare function based on a monotonically increasing utility function, i.e., $W_U(H_1) \geq W_U(H_2)$ for all $U(z)$ such that $U'(z) \geq 0$. For example, $U(z) = z$ is monotonically increasing and the social welfare function with $U(z) = z$ gives the means of the distributions. Consequently, if H_1 SD1 H_2, we have $\int zdH_1(z) \geq \int zdH_2(z)$. This implies that if the potential outcome under the treatment SD1 that without treatment, then the ATE is non-negative. First order stochastic dominance can also be defined in terms of the underlying distributions. That is, H_1 SD1 H_2 iff $H_1(z) \leq H_2(z)$ 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., $U'(z) \geq 0$ and $U''(z) \leq 0$. Similarly, H_1 SD2 H_2 iff $\int_0^z H_1(t)dt \leq \int_0^z H_2(t)dt$ for all z. Furthermore, it is true that first order stochastic dominance implies second order stochastic dominance, but not vice versa.[1] In this section, we focus on test for first order stochastic dominance and we will discuss higher order stochastic dominance later.

2.1 Hypothesis Formulation

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 $F_0(\cdot)$ and $F_1(\cdot)$ 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.

We want to know if $Y(1)$ SD1 $Y(0)$, i.e. $F_1(z) \leq F_0(z)$ for all $z \in \mathcal{Z}$, where $\mathcal{Z}$ is the common support of $Y(0)$ and $Y(1)$. Hence, we formulate our hypotheses as follows:

$$
\begin{aligned}
&H_0 : F_1(z) \leq F_0(z) \quad \text{for all } z \in \mathcal{Z}; \\
&H_1 : F_1(z) > F_0(z) \quad \text{for some } z \in \mathcal{Z}. \qquad (1)
\end{aligned}
$$

2.2 Identification and Estimation of $F_1(z)$ and $F_0(z)$

The main issue in the treatment effects literature is that we never observe both $Y(0)$ and $Y(1)$, so there is an identification problem. This problem can be avoided by imposing the unconfoundedness assumption (Rubin (1978), Rosenbaum and Rubin (1983), Hirano, Imbens and Ridder (2003), Firpo (2007a)) which is equivalent to the selection-on-observables assumption (Barnow, Cain and Goldberger (1980)). The unconfoundedness assumption requires that conditional on the observed individual characteristics, the treatment assignment is independent of the potential outcomes.

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

Let $p(x) = P(T = 1|X = x)$ denote the propensity score (Rosenbaum and Rubin (1983, 1985)), the probability of getting treatment for an individual with covariate x. Under

[1] 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 $(-1)^k U^{(k)}(z) \leq 0$ for $k = 1, \ldots, j$. It is also true that SDj implies SD$(j+1)$, but not vice versa.

Assumption 2.1, $F_1(z)$ is identified by

$$\begin{aligned}
& E\left[\frac{T\cdot 1(Y\leq z)}{p(X)}\right] \\
=& E\left[E\left[\frac{T\cdot 1(Y\leq z)}{p(X)}\Big|X\right]\right] \\
=& E\left[\frac{1}{p(X)}E\left[E\big[T\cdot 1(Y\leq z)\Big|T=1,\ X\big]\right]\cdot P\big(T=1|X\big)\right] \\
=& E\left[E\big[1(Y(1)\leq z)\big|T=1,\ X\big]\right] \\
=& E\left[E[1(Y(1)\leq z)\big|X]\right] \\
=& F_1(z). \qquad (2)
\end{aligned}$$

The first equality follows from the law of iterated expectations (LIE). By expanding the conditional expectation, we get the second equality. The third one holds since $P\big(T=1|X\big)=p(X)$ and $Y=Y(1)$ conditional on $T=1$. The fourth one follows from the unconfoundedness assumption. We get the last equality by applying LIE again. Similarly, $F_0(z)$ is identified by

$$F_0(z)=E\left[\frac{(1-T)\cdot 1(Y\leq z)}{1-p(X)}\right]. \qquad (3)$$

Based on (2) and (3), the IPW estimators for $F_0(z)$ and $F_1(z)$ are as follows:

$$\widehat{F}_0(z)=\frac{1}{N}\sum_{i=1}^{N}\frac{(1-T_i)\cdot 1(Y_i\leq z)}{1-\hat{p}(X_i)}, \qquad \widehat{F}_1(z)=\frac{1}{N}\sum_{i=1}^{N}\frac{T_i\cdot 1(Y_i\leq z)}{\hat{p}(X_i)}, \qquad (4)$$

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 $\lambda=(\lambda_1,...,\lambda_r)'\in\mathbb{Z}_+^r$ 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 $R^K(x)=(x^{\lambda(1)},...,x^{\lambda(K)})'$ 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 $\hat{p}(x)=\mathcal{L}\big(R^K(x)'\hat{\pi}_K\big)$ where

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

The asymptotic properties of $\hat{p}(x)$ are discussed in Appendix A of HIR.

2.3 Test Statistic and Decision Rule

Given $\widehat{F}_0(z)$ and $\widehat{F}_1(z)$, we define the statistic as

$$\widehat{S}_N = \sqrt{N} \sup_{z \in \mathcal{Z}} \big(\widehat{F}_1(z) - \widehat{F}_0(z)\big). \tag{5}$$

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 be 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 $\widehat{S}_N = \sqrt{N} \max\{\widehat{F}_1(z) - \widehat{F}_0(z) | z = 0, \bar{z}, Y_1, ..., Y_N\}$, which only involves taking maximum over a finite number of values.

3 Asymptotic Properties

3.1 Assumptions

In addition to the unconfoundedness assumption, we make the following assumptions which are similar to those in HIR. The first assumption summarizes the properties of the CDFs of $Y(0)$ and $Y(1)$.

Assumption 3.1 (Distributions of $Y(0)$ and $Y(1)$):

1. *$Y(0)$ and $Y(1)$ have a common compact support $\mathcal{Z} = [0, \bar{z}]$ with $\bar{z} < \infty$.*
2. *F_0 and F_1 are continuous functions on $\mathcal{Z}$ with $F_0(0) = F_1(0) = 0$.*

We impose conditions on the distribution of X and the conditional CDFs of $Y(0)$ and $Y(1)$.

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 $x \in \mathcal{X}$.*

The following assumption requires the smoothness of the propensity score function.

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: $0 < \underline{p} \leq p(x) \leq \overline{p} < 1$.*

The last assumption restricts the growth rate of the number of approximating functions to be included in the series approximation to the propensity score function.

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

Remarks on the Assumptions:

1. Assumption 3.1 requires that $Y(0)$ and $Y(1)$ to be continuous random variables on $\mathcal{Z}$. However, our theoretical results presented later are still valid when the probabilities of $Y(0)$ and $Y(1)$ at 0 are strictly positive. That is, we can modify our proofs to allow for the cases where $F_0(0) > 0$ or $F_1(0) > 0$ or both. As in our empirical examples, we have $F_0(0) > 0$ and $F_1(0) > 0$.

2. 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.

3. As in HIR, Assumption 3.4 ensures the existence of a ν satisfying the conditions in Assumption 3.5.

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

Lemma 3.6 *Suppose Assumption 2.1 and 3.1-3.5 hold. Then*

$$\sqrt{N}(\widehat{F}_0(z) - F_0(z)) \Rightarrow \Psi_0(z), \quad \sqrt{N}(\widehat{F}_1(z) - F_1(z)) \Rightarrow \Psi_1(z),$$

where $\Rightarrow$ denotes weak convergence, and $\Psi_0(z)$ and $\Psi_1(z)$ are zero-mean Gaussian processes with covariance functions $\Omega_0(z_1, z_2) = E[\Psi_0(z_1)\Psi_0(z_2)]$ and $\Omega_1(z_1, z_2) = E[\Psi_1(z_1)\Psi_1(z_2)]$ such that for $z_1 \leq z_2$

$$\Omega_0(z_1, z_2) = E\left[\frac{F_0(z_1|X_i)}{1 - p(X_i)}\right] - F_0(z_1)F_0(z_2) - E\left[\frac{p(X_i)F_0(z_1|X_i)F_0(z_2|X_i)}{1 - p(X_i)}\right],$$
$$\Omega_1(z_1, z_2) = E\left[\frac{F_1(z_1|X_i)}{p(X_i)}\right] - F_1(z_1)F_1(z_2) - E\left[\frac{(1 - p(X_i))F_1(z_1|X_i)F_1(z_2|X_i)}{p(X_i)}\right].$$

To show Lemma 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:

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

where $W \equiv \{Y, T, X\}$,

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

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 both Donsker classes.[2] Then by Donsker's Theorem or functional central limit theorem, we

[2] Let P denote a probability measure. A class $\mathcal{F}$ of measure functions $f : X \to \mathbb{R}$ is said to be a P-Donsker set if the general empirical processes converge weakly to a mean zero Gaussian processes indexed by $\mathcal{F}$ with covariance $E[f(X)g(X)] - E[f(X)]E[g(X)]$. For example, let X_i's be i.i.d. random variables with CDF $F(t)$ and $\mathcal{F} = \{1(x \leq t)|t \in \mathbb{R}\}$. It is known that $\mathcal{F} = \{1(x \leq t)|t \in \mathbb{R}\}$ is Donsker for any probability measure. It implies that the empirical distribution function $\mathbb{F}_N(t)$

$$\sqrt{N}(\mathbb{F}_N(t) - F(t)) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\big(1(X_i \leq t) - F(t)\big) \Rightarrow \mathbb{F}(t),$$

where $\mathbb{F}(t)$ is a mean zero Gaussian processes with $Cov(\mathbb{F}(t), \mathbb{F}(s)) = F(\min\{s, t\}) - F(s)F(t)$. For more details, please refer to Kosorok (2008), and van der Vaart and Wellner (1996).

have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\big(\psi_0(W_i,z)-F_0(z)\big)\Rightarrow\Psi_0(z),\qquad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\big(\psi_1(W_i,z)-F_1(z)\big)\Rightarrow\Psi_1(z),$$

and Lemma 3.6 follows.

3.3 Asymptotic Properties of the Test Statistic

We define $\overline{S}=\sup_{z\in\mathcal{Z}}\Psi(z)$, where $\Psi(z)=\Psi_1(z)-\Psi_0(z)$. Following from Lemma 3.6, $\Psi(z)$ is asymptotically equivalent to the limit of $\sqrt{N}\big((\widehat{F}_1(z)-\widehat{F}_0(z))-(F_1(z)-F_0(z))\big)$ and by the continuous mapping theorem (CMT), $\sup_{z\in\mathcal{Z}}\sqrt{N}\big((\widehat{F}_1(z)-\widehat{F}_0(z))-(F_1(z)-F_0(z))\big)\xrightarrow{D}\overline{S}$. Thus, we have the following result concerning the asymptotic properties of the test statistic $\widehat{S}_N$.

Proposition 3.7 *Suppose that Assumption 2.1 and 3.1-3.5 hold and c is a positive finite number, then:*

1. *if H_0 is true,* $\limsup P(\textit{reject } H_0)=\limsup P(\widehat{S}_N>c)\leq P(\overline{S}>c)\equiv\bar{\alpha}(c)$.
2. *if H_0 is false,* $\lim_{N\to\infty}P(\textit{reject } H_0)=1$.

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 $z\in\mathcal{Z}$. Hence, the null distribution depends both on the $F_0(z)$ and $F_1(z)$. 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

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

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

$$\widehat{S}_N=\sqrt{N}\sup_{z\in\mathcal{Z}}(\widehat{F}_1(z)-\widehat{F}_0(z))\leq\sqrt{N}\sup_{z\in\mathcal{Z}}\big((\widehat{F}_1(z)-\widehat{F}_0(z))-(F_1(z)-F_0(z))\big).$$

The statistic is never greater than $\sqrt{N}\big((\widehat{F}_1(z)-\widehat{F}_0(z))-(F_1(z)-F_0(z))\big)$ 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 $z \in \mathcal{Z}$ and $\overline{S}$ 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 the sense that the type I error of the test can be strictly smaller than the pre-determined level, and the first part of Proposition 3.7 confirms this. The most important implication of this result is that if we can find a c such that $\bar{\alpha}(c)$ is equal to some desired significance level (e.g. 0.01 or 0.05), then we can control the size of our test to be no greater than $\bar{\alpha}(c)$ asymptotically.

On the other hand, if the alternative hypothesis is true, then there exists z^* such that $F_1(z^*) - F_0(z^*) > 0$. In this case, $\widehat{F}_1(z^*) - \widehat{F}_0(z^*)$ will converge to infinity. Given the fact that $\widehat{S}_N \geq \widehat{F}_1(z^*) - \widehat{F}_0(z^*)$, $\widehat{S}_N$ converges to infinity, i.e., $\lim_{N\to\infty} P(\widehat{S}_N > c) = 1$ for any $c > 0$. Therefore, the second result follows and it implies that our test is consistent in the sense that it can detect any violation of the null hypothesis in the limit.

4 Simulated Critical Value, and Size and Power Properties

As noted by McFadden (1989) and Barrett and Donald (2003), the main difficulty with Kolmogorov-Smirnov tests for stochastic dominance is in constructing an appropriate critical value (p-value) for conducting the tests since the limiting distribution of the test statistics under the LFC depends on the underlying distributions, $F_0(z)$ and $F_1(z)$. Our case is more complicated, since the limiting distribution of the test statistic depends not only on $F_0(z)$ and $F_1(z)$, but also on $p(x)$, $F_0(z|x)$ and $F_1(z|x)$. As in McFadden (1989) and Barrett and Donald (2003), we propose and justify a Monte-Carlo method to estimate the critical value.

Before introducing the Monte-Carlo method to conduct the test, 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.

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

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

$$\widetilde{F}_1(z|x) = \left(\sum_{i=1}^{N} \frac{1(Y_i \leq z)T_i}{\hat{p}(X_i)} R^K(X_i)\right)' \left(\sum_{i=1}^{N} R^K(X_i) R^K(X_i)'\right)^{-1} R^K(x). \tag{6}$$

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 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 $\widetilde{F}_1(z|x)$.[3]

Without loss of generality (WLOG), we assume 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 $Y_{(i)} \le z < Y_{(i+1)}$

$$\begin{aligned}\widehat{F}_0(z|x) =& \widehat{F}_0(Y_{(i-1)}|x) \cdot 1\big(0 \le \widetilde{F}_0(Y_{(i)}|x) \le \widehat{F}_0(Y_{(i-1)}|x)\big) \\ &+ \widetilde{F}_0(Y_{(i)}|x) \cdot 1\big(\widehat{F}_0(Y_{(i-1)}|x) < \widetilde{F}_0(Y_{(i)}|x) \le 1\big) + 1\big(\widetilde{F}_0(Y_{(i)}|x) > 1\big).\end{aligned}$$

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 $\widetilde{F}_0(z|x) > 1$. 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.

Lemma 4.1 *Suppose Assumption 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*

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

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 $x \in \mathcal{X}$. Similar arguments apply to $\widehat{F}_1(z|x)$.

Let U_1, U_2,... be independent standard normals on a probability space $(\Omega_u, \mathcal{F}_u, P_u)$ independent of the sequence $\mathcal{W} = \{W_1, W_2, ...\}$ which is on $(\Omega_w, \mathcal{F}_w, P_w)$. For all $z \in \mathcal{Z}$

[3] 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.

define the simulated stochastic process as

$$\Psi_u(z) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i\Big\{\frac{T_i\cdot 1(Y_i\le z)}{\hat{p}(X_i)} - \frac{(1-T_i)\cdot 1(Y_i\le z)}{1-\hat{p}(X_i)} - (\widehat{F}_1(z)-\widehat{F}_0(z))$$
$$- (T-\hat{p}(X_i))\Big(\frac{\widehat{F}_1(z|X_i)}{\hat{p}(X_i)} + \frac{\widehat{F}_0(z|X_i)}{1-\hat{p}(X_i)}\Big)\Big\}. \tag{7}$$

Lemma 4.2 *Suppose Assumption 2.1 and 3.1-3.5 hold. Then $\Psi_u(z) \Rightarrow \Psi(z)$ given $\mathcal{W}$ in probability which is denoted by $\Psi_u(z) \overset{P_{\mathcal{W}}}{\Rightarrow} \Psi(z)$.*

Lemma 4.2 indicates that we can approximate $\Psi(z)$ by the simulated stochastic process $\Psi_u(z)$, since $\Psi_u(z)$ weakly converges to $\Psi(z)$ given $\mathcal{W}$ in probability. To show this, we first show that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i\Big\{\frac{T_i\cdot 1(Y_i\le z)}{p(X_i)} - \frac{(1-T_i)\cdot 1(Y_i\le z)}{1-p(X_i)} - (F_1(z)-F_0(z))$$
$$- (T-p(X_i))\Big(\frac{F_1(z|X_i)}{p(X_i)} + \frac{F_0(z|X_i)}{1-p(X_i)}\Big)\Big\} \overset{P_{\mathcal{W}}}{\Rightarrow} \Psi(z),$$

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.

Given $\Psi_u(z)$, we can approximate the null distribution under LFC $\overline{S}$ by $\overline{S}_u = \sup_z \Psi_u(z)$. Since the supremum operation is a continuous function, then by CMT we can show that $\overline{S}_u$ converges in distribution to $\overline{S}$ given $\mathcal{W}$ in probability. 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

$$\hat{c} = \sup\{q|P_u(\overline{S}_u \le q) \le 1-\alpha_0\}. \tag{8}$$

Let $c_0 = \sup\{q|P(\overline{S} \le q) \le 1-\alpha_0\}$, the $(1-\alpha_0)$-th quantile of $\overline{S}$, which satisfies $\bar{\alpha}(c_0) = \alpha_0$. We have that if $\alpha_0 < 1/2$, then $\hat{c}$ converges in probability to c_0. The following lemma summarizes these results.

Lemma 4.3 *Suppose Assumption 2.1 and 3.1-3.5 hold, and $\alpha_0 < 1/2$, then*

1. $\overline{S}_u \xrightarrow{D} \overline{S}$ *given $\mathcal{W}$ in probability.*

2. $\hat{c} \xrightarrow{p} c_0$.

The size and power properties of our test 4.3 are summarized in the following theorem which follows from Lemma 4.3.

Theorem 4.4 *Suppose Assumption 2.1 and 3.1-3.5 hold and $\alpha_0 < 1/2$. If we reject the H_0 when $\widehat{S}_N > \hat{c}$, then:*

1. *if H_0 is true,* $\limsup P(\textit{reject } H_0) = \limsup P(\widehat{S}_N > \hat{c}) \leq \alpha_0$.

2. *if H_0 is false,* $\lim_{N\to\infty} P(\textit{reject } H_0) = 1$.

Theorem 4.4 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.

Remarks:

1. 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 α_0, $\widehat{S}_N > \hat{c}$ iff $\hat{p}(\widehat{S}_N) < \alpha_0$. Therefore, the critical value method and the p-value method are equivalent.[4]

2. To calculate $\hat{c}$ practically, we need to approximate the distribution of $\overline{S}_u$. To be more specific, let $\{U_i^r\}_{i=1}^N$ be the r-th sample of U_i for $r = 1, ..., R$, where R is the number of repetitions in the Monte-Carlo simulation. The rth realization of the statistic is

$$\overline{S}_u^r = \sup_{z\in\mathcal{Z}} \frac{1}{\sqrt{N}} \sum_{i=1}^N U_i^r \Big\{ \frac{T_i \cdot 1(Y_i \leq z)}{\hat{p}(X_i)} - \frac{(1-T_i)\cdot 1(Y_i \leq z)}{1-\hat{p}(X_i)}$$
$$-(\widehat{F}_1(z) - \widehat{F}_0(z)) - (T - p(X_i))\Big(\frac{\widehat{F}_1(z|X_i)}{p(X_i)} + \frac{\widehat{F}_0(z|X_i)}{1-p(X_i)}\Big)\Big\}.$$

   Then the critical value $\hat{c}$ can be approximated by the $(1-\alpha_0)$-th quantile of the empirical distribution of $\{\overline{S}_u^r\}_{r=1}^R$. Additionally, the $\hat{p}(\widehat{S}_N)$ can be approximated by $\sum_{r=1}^R 1(\overline{S}_u^r > \widehat{S}_N)/R$.

   Similar to $\widehat{S}_N$, the calculation of $\overline{S}_u^r$ involves taking the supremum over a compact set. However, $\Psi_u(z)$ is a step function, so we can can exactly calculate the value

[4] It is easier to show the properties of our test based on the critical value method. However, in practice, it is easier to make inference based on the p-value method especially when we pick several different significance levels. Note that the p-value remains the same for any given significance level α_0, but the critical values for different significance levels varies. Consequently, the p-value method requires less computation and our simulations and empirical studies are implemented based on the p-value method.

of $\overline{S}_u$ by taking maximum over a finite number of values. Given the computation time is proportional to NR, so when N and R are too large, it can be very time consuming to do it. Hence, as in Barrett and Donald (2003), one can also use a grid to approximate the supremum. By increasing the number of values of a grid, the approximation can be made as accurate as one wants. For example, let $0 = z_0 < z_1 \cdots < z_K = \bar{z}$ be a grid of values on $[0, \bar{z}]$ which can be a subset of Y_i's, where K is the number of subintervals. That is, the r-th realization of the statistic can be approximated by

$$\overline{S}_u^r = \max_{k=0,\ldots,K} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i^r \Big\{ \frac{T_i \cdot 1(Y_i \le z_k)}{\hat{p}(X_i)} - \frac{(1-T_i) \cdot 1(Y_i \le z_k)}{1-\hat{p}(X_i)}$$
$$-(\widehat{F}_1(z_k) - \widehat{F}_0(z_k)) - (T - p(X_i)) \Big(\frac{\widehat{F}_1(z_k|X_i)}{p(X_i)} + \frac{\widehat{F}_0(z_k|X_i)}{1-p(X_i)} \Big) \Big\}.$$

3. Note that our test without any covariate is equivalent to Barrett and Donald's (2003) test with the p-values generated by Equation (6) which is labeled as 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 is no covariate in our test, the only regressor in the SLE is the constant term and we have $\hat{p} = N_1/N$, the proportion of the treated individuals. 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,

$$\widehat{S}_N = \sqrt{N} \sup_{z \in \mathcal{Z}} (\widehat{F}_1(z) - \widehat{F}_0(z)) = \sqrt{\frac{N^2}{N_1 N_0}} \left(\sqrt{\frac{N_1 N_0}{N}} \sup_{z \in \mathcal{Z}} (\widehat{F}_1(z) - \widehat{F}_0(z)) \right),$$

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

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

where the term in the biggest 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 $\sqrt{\frac{N^2}{N_1 N_0}}$. The test statistic and critical value are all rescaled by the same positive number $\sqrt{\frac{N^2}{N_1 N_0}}$, so the conclusion of both tests will be the same too. As a result, our test without any covariate is equivalent to Barrett and Donald's (2003) test with p-values generated by Equation (6).

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, $F_0(z) = E[1(Y \leq z)|D = 0]$ and $F_1(z) = E[1(Y \leq z)|D = 1]$, so their test based on the treated sample and the untreated people works. Furthermore, the random assignment assumption also implies that 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 $F(z|T = 0) = E[1(Y \leq z)|T = 0]$ and $F(z|T = 1) = E[1(Y \leq z)|T = 1]$ are not equal to $F_0(z)$ and $F_1(z)$ in general. Consequently, their test may give wrong conclusions, but our test is still valid. In this sense, our test is more robust than theirs.

5 Tests for First Order Stochastic Dominance among the Treated

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 subpopulation such as the treated. This section discusses first order stochastic dominance test among the treated.

Let $F_0(z|\, T = 1) = E[1(Y(0) \leq z)|\, T = 1]$ and $F_1(z|\, T = 1) = E[1(Y(1) \leq z)|\, T = 1]$ denote the conditional CDFs of the potential outcomes of the treated. As in (1), we formulate the related hypothesis as follows:

$$\begin{aligned} &H_{0,treated} : F_1(z|\, T = 1) \leq F_0(z|\, T = 1) \quad \text{for all } z \in \mathcal{Z}; \\ &H_{1,treated} : F_1(z|\, T = 1) > F_0(z|\, T = 1) \quad \text{for some } z \in \mathcal{Z}. \end{aligned} \tag{9}$$

Define $G_0(x)$ and $G_1(x)$ as

$$G_0(x) = E\left[\frac{p(X)(1-T)\cdot 1(Y \leq z)}{1-p(X)}\right], \quad G_1(x) = E\left[T \cdot 1(Y \leq z)\right]. \tag{10}$$

Note that

$$
\begin{aligned}
G_1(x) &= E\left[T \cdot 1(Y \le z)\right] \\
&= E\big[p(X)E[1(Y(1) \le z)|X]\big] \\
&= E\big[E[T|X]E[1(Y(1) \le z)|X]\big] \\
&= E\big[E[T \cdot 1(Y(1) \le z)|X]\big] \\
&= E[T \cdot 1(Y(1) \le z)] \\
&= E[1(Y(1) \le z)|T = 1]P(T = 1) \\
&= F_1(z|T = 1)E[p(X)].
\end{aligned}
$$

The first equality follows from the same argument as (2) without $p(X)$ and since $p(X) = E[T|X]$, the second equality follows. Because of the unconfoundedness assumption and LIE, we have the third and fourth equalities. By expanding the expectation, we have the fifth equality. The last equality holds, since $P(T = 1) = E[E[T|X]] = E[p(X)]$. Similarly,

$$
G_0(z) = F_0(z|T = 1)E[p(X)]. \tag{11}
$$

Since $E[p(X)] > 0$, then $F_1(z|T = 1) \le F_0(z|T = 1)$ if and only if $G_1(z) \le G_0(z)$ for all $z \in \mathcal{Z}$. As a result, the hypotheses in (9) are equivalent to

$$
\begin{aligned}
&H_{0,G} : G_1(z) \le G_0(z) \quad \text{for all } z \in \mathcal{Z}; \\
&H_{1,G} : G_1(z) > G_0(z) \quad \text{for some } z \in \mathcal{Z}.
\end{aligned} \tag{12}
$$

As a result, we can make the inference on the treated by comparing $G_0(z)$ and $G_1(z)$. By the definition of $G_0(z)$ and $G_1(z)$, we estimate them by

$$
\widehat{G}_0(z) = \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{p}(X_i)(1 - T_i) \cdot 1(Y_i \le z)}{1 - \hat{p}(X_i)}, \qquad \widehat{G}_1(z) = \frac{1}{N}\sum_{i=1}^{N} T_i \cdot 1(Y_i \le z). \tag{13}
$$

The statistic is defined as

$$
\widehat{S}_{N,G} = \sqrt{N} \sup_{z \in \mathcal{Z}} \big(\widehat{G}_1(z) - \widehat{G}_0(z)\big). \tag{14}
$$

Define the simulated stochastic process as

$$
\begin{aligned}
\Psi_{u,G}(z) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_i \Big\{ & T_i\big(1(Y_i \le z) - \widehat{F}_1(z|X_i)\big) \\
& - \frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)}(1 - T_i)\big(1(Y_i \le z) - \widehat{F}_0(z|X_i)\big) \\
& + T_i\big(\widehat{F}_1(z|X_i) - \widehat{F}_0(z|X_i)\big) - \big(\widehat{G}_1(z) - \widehat{G}_0(z)\big)\Big\},
\end{aligned} \tag{15}
$$

where U_i are i.i.d. standard normals. Define the simulated null process $\overline{S}_{u,G} = \sup_{z \in \mathcal{Z}} \Psi_{u,G}(z)$. Given the significance level α_0, the simulated critical value, $\hat{c}_G$, is defined as the $(1-\alpha_0)$-th quantile of $\overline{S}_{u,G}$.

We first modify Assumption 3.1 for the treated case.

Assumption 5.1 (Conditional Distributions of $Y(0)$ and $Y(1)$ on the Treated):

1. *Conditional on the treated, $Y(0)$ and $Y(1)$ have a common compact support $\mathcal{Z} = [0, \bar{z}]$ where $\bar{z} < \infty$.*
2. *$F_1(z|T=1)$ and $F_0(z|T=1)$ are continuous functions on $\mathcal{Z}$ with $F_1(0|T=1) = F_0(0|T=1) = 0$.*

The second part of Assumption 5.1 implies that $G_1(z)$ and $G_0(z)$ are continuous functions with $G_1(z) = G_0(z) = 0$.[5] The following theorem analogous to Theorem 4.4 summarizes the size and power properties of the test on the treated.

Theorem 5.2 *Suppose Assumption 2.1, 3.2-3.5 and 5.1 hold, and $\alpha_0 < 1/2$. If we reject the $H_{0,treated}$ (or $H_{0,G}$) when $\widehat{S}_{N,G} > \hat{c}_G$, then:*

1. *if $H_{0,treated}$ is true,* $\limsup P(\textit{reject } H_{0,treated}) \leq \alpha_0$.
2. *if $H_{0,treated}$ is false,* $\lim_{N\to\infty} P(\textit{reject } H_{0,treated}) = 1$.

6 Tests for Higher Order Stochastic Dominance

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

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

[5] As in Assumption 3.1, we can allow for the cases where $F_1(0|T=1) > 0$ or $F_0(0|T=1) > 0$ or both.

Accordingly, F_1 SDj F_0 corresponds to $\mathcal{I}_j(z; F_1) \leq \mathcal{I}_j(z; F_0)$ for all z. Hence, the hypothesis for F_1 SDj F_0 is the following:

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

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), we can show that

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

which is a piecewise polynomial and so 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 compute simply.[6]

Following the same argument in Section 3.3, 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 null distribution under the LFC is $\overline{S}^j = \sup_{z \in \mathcal{Z}} \mathcal{I}_j(z; \Psi)$. As a result, 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 following theorem summarizes the size and power properties of the test of higher order stochastic dominance.

Theorem 6.1 *Suppose Assumption 2.1 and 3.1-3.5 hold, and* $\alpha_0 < 1/2$. *If we reject the* H_0^j *when* $\widehat{S}_N^j > \hat{c}^j$, *then*

1. *if* H_0^j *is true,* $\limsup P(\textit{reject } H_0^j) \leq \alpha_0$.
2. *if* H_0^j *is false,* $\lim_{N \to \infty} P(\textit{reject } H_0^j) = 1$.

[6] In the simulations and the empirical studies, we use the left Reimann 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 $0 = z_0 < z_1 < ... < z_k = \bar{z}$ and we approximate $\mathcal{I}_2(z_j; \widehat{F}_1 - \widehat{F}_0)$ and $\widehat{S}_N^2$ respectively by

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

Similarly, we can test the higher order stochastic dominance relations among the treated. As in Section 5, we formulate the hypotheses as follows:

$$H^j_{0,treated} : \mathcal{I}_j(z; F_1(\cdot|\ T = 1)) \le \mathcal{I}_j(z; F_0(\cdot|\ T = 1)) \ \text{ for all } z \in \mathcal{Z};$$
$$H^j_{1,treated} : \mathcal{I}_j(z; F_1(\cdot|\ T = 1)) > \mathcal{I}_j(z; F_0(\cdot|\ T = 1)) \ \text{ for some } z \in \mathcal{Z},$$

which are equivalent to the following:

$$H^j_{0,G} : \mathcal{I}_j(z; G_1(\cdot)) \le \mathcal{I}_j(z; G_0(\cdot)) \ \text{ for all } z \in \mathcal{Z};$$
$$H^j_{1,G} : \mathcal{I}_j(z; G_1(\cdot)) > \mathcal{I}_j(z; G_0(\cdot)) \ \text{ for some } z \in \mathcal{Z}.$$

Define the statistic as

$$\widehat{S}^j_{N,G} = \sqrt{N} \sup_{z\in\mathcal{Z}} \mathcal{I}_j(z; \widehat{G}_1(\cdot) - \widehat{G}_0(\cdot)), \tag{16}$$

and $\overline{S}^j_{u,G} = \sup_{z\in\mathcal{Z}} \mathcal{I}_j(z; \Psi_{u,G})$. Given the significance level α_0, define the simulated critical value, $\hat{c}^j_G$, as the $(1-\alpha_0)$-th quantile of $\overline{S}^j_{u,G}$. We have the following theorem.

Theorem 6.2 *Suppose Assumption 2.1, 3.2-3.5 and 5.1 hold, and $\alpha_0 \le 1/2$. If we reject the $H^j_{0,treated}$ (or $H^j_{0,G}$) when $\widehat{S}^j_{N,G} > \hat{c}^j_G$, then:*

1. *if $H^j_{0,treated}$ is true, $\limsup P(\text{reject } H^j_{0,treated}) \le \alpha_0$.*
2. *if $H^j_{0,treated}$ is false, $\lim_{N\to\infty} P(\text{reject } H^j_{0,treated}) = 1$.*

7 Monte-Carlo Studies

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:*

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

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. The unconditional CDFs of $Y(0)$ and $Y(1)$ are identical, so $Y(1)$ SD1 and SD2 $Y(0)$. We can show that $F_1(z|T=1) \leq F_0(z|T=1)$ for all z, which implies that $Y(1)$ conditional on the treated SD1 and SD2 $Y(0)$ conditional on the treated. We plot the corresponding CDFs and integrated CDFs in Figure 1.

Example 7.2 *Let the DGP be:*

$$\begin{aligned}
&X = U_x, \quad X_0 = 0.5 + 0.3X, \qquad T = 1(U_t < X_0),\\
&Y(0) = 1(U_{y_0} \leq X_0)\frac{U_{y_0}^2}{X_0} + 1(U_{y_0} > 1 - X_0)\Big(1 - \frac{(1-U_{y_0})^2}{1-X_0}\Big),\\
&Y(1) = U_{y_1},\\
&Y = TY(1) + (1-T)Y(0),
\end{aligned}$$

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

In Example 7.2, we have $F_1(z|X) > F_0(z|X)$ for all $z > X_0$ which implies that $Y(1)$ SD1 $Y(0)$ neither unconditionally nor conditional on the treated. However, we have $\int_0^z F_1(t|X)dt < \int_0^z F_0(t|X)dt$ for all $z > 0$ and it follows that $Y(1)$ SD2 $Y(0)$ unconditionally and conditional on the treated. Figure 2 presents the corresponding CDFs and integrated CDFs.

Example 7.3 *Let the DGP be:*

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

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

In Example 7.3, $Y(1)$ neither SD1 nor SD2 $Y(0)$ conditional on X, since $F_1(z|X) > F_0(z|X)$ for all $z \leq X$ and $F_1(z|X) = F_0(z|X)$ for all $z > X$. It follows that $Y(1)$ neither SD1 nor SD2 $Y(0)$ unconditionally and conditional on the treated group. The plots for the corresponding CDFs and integrated CDFs can be found in Figure 3.

In all examples, we set the sample sizes $N = 200$. The rejection rate is calculated based on 1,000 simulations, and for each simulation the p-value is approximated by 1,000 repetitions. We use all of the different values of Y_i as the gridpoints. The propensity score function is estimated by SLE with the power series: 1, X and X^2. The significance level is 5%. Let F_j denote the CDF of potential outcome $Y(j)$ and F_j^t the conditional CDF of potential outcome $Y(j)$ of the treated for $j = 0$ and 1. Table 1 summarizes the rejection rates for different cases.

All the rejection rates corresponding to Example 7.1 are around the 5% level. The rejection rates in Example 7.2 are much smaller than 5% for both second order stochastic dominance tests. These support the size properties of our tests. In Example 7.2, the null hypotheses are wrong for the SD1 cases and the rejection rates are greater than 5% (14% and 12% respectively). All the null hypotheses are violated in Example 7.3 and we find that the rejection rates are between 26% and 45%. The rejections rates of Example 7.3 are much higher than Example 7.2 due to the larger difference between the two CDFs in Example 7.3.

We redo the simulations with a larger sample size, $N = 400$. The propensity score function is estimated by SLE with the power series: 1, X, X^2 and X^3. The rejection rates corresponding to correct null hypotheses are still around or less than 5% level. When the null hypotheses are wrong, the corresponding rejection rates are higher as the sample size increases. These results again support the size and power properties of our methods and also illustrate the consistency of our tests.

Furthermore, the nominal sizes for the treated cases in Example 7.1 and the SD2 cases in Example 7.2 are strictly smaller than 5% due to the LFC of our tests. On the other hand, he nominal sizes for the unconditional cases in Example 7.1 are close to 5%, since we have $F_1(z) = F_0(z)$ for all z which is the LFC of our tests.

We compare the performances between our tests and Barrett and Donald's (2003) KS2 method, which is equivalent to our test without any covariate. When there is no covariate in our model, then the only regressor in SLE is the constant term. Let the DGPs be the same as previous examples except that T is generated by $T = 1(U_t < 0.5)$. That is, T is randomly assigned. When there is no covariate, the constant term is the only regressor when we implement the SLE. The other set we consider contains X and we use the power series of X up to order two as the regressors in the SLE estimation. The simulation results are summarized in Table 2. We find that for all cases, the rejection rates under two different specifications are not significantly different. We conclude that, at least in this case, it is not harmful to include some covariates in the tests when T is randomly assigned.

We use the following examples to demonstrate that Barrett and Donald's (2003) tests may lead to wrong conclusion when the random assignment assumption is violated.

Example 7.4 *Let the DGP be:*

$$\begin{aligned}
&X = 0.2 + 0.6U_x, \qquad T = 1(U_t < 1 - X), \\
&Y(0) = U_{y_0}, \\
&Y(1) = \begin{cases} \sqrt{XU_{y_1}}, & \text{if } X \geq 0.5 \text{ and } U_{y_1} \leq X, \\ (1-x) - \sqrt{(1-x)^2 - (1-x)U_{y_1}}, & \text{if } X < 0.5 \text{ and } U_{y_1} \leq 1 - X, \\ U_{y_1}, & \text{othetwise}, \end{cases} \\
&Y = TY(1) + (1-T)Y(0),
\end{aligned}$$

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

The CDFs are presented in Figure 4. In this example, $F_1(z) = F_0(z)$ for all $z \in \mathcal{Z}$ that implies F_1 first order and second order stochastically dominates F_0. However, $F(z|T = 1) > F(z|T = 0)$ for $z \in (0, 0.8)$ that implies that $F(z|T = 1)$ neither first order nor second order stochastically dominates $F(z|T = 0)$.

Example 7.5 *Let GDP be the same as that in Example 7.4 except that $X = 0.25+0.4U_x$ and $T = 1(U_t < 0.1 + 0.8U_x^2)$.*

The CDFs are presented in Figure 5. In this example, $F_1(z) > F_0(z)$ for all $z \in (0, 0.65)$, then F_1 neither order and second order stochastically dominates F_0. However, $F(z|T = 1) \leq F(z|T = 0)$ for $z \in \mathcal{Z}$ and $F(z|T = 1)$ both first order and second order stochastically dominates $F(z|T = 0)$.

Table 3 summarizes the simulation results of Example 7.4 and 7.5. Regarding Example 7.4, the null hypotheses are correct. The rejection rates of our tests which use the covariate X to estimate the propensity score are around 5%, However, those of Barrett and Donald's (2003) which assume the treatment is completely random and do not include any covariate are around 15% and they are much higher than the significance level 5%. On the other hand, concerning Example 7.5 where the null hypotheses are not true, the rejections of our tests are around 17%, but those of Barrett and Donald's (2003) are lower than 5%. We redo the simulations with sample size $N = 400$. Regarding Example 7.4, our tests still control the size well (5.7% and 4.6% respectively) and the rejection rates of Barrett and Donald's (2003) increase by 2-3% . With respect to Example 7.5, the rejection rates of our tests increase by 5-6% and those of Barrett and Donald's (2003) are still lower than 5%. Hence, if the treatment is unconfounded conditional on the covariates, then Barrett and Donald's (2003) tests may give wrong conclusions and our tests are more robust than theirs.

8 Empirical Illustration

We apply our tests to the data from National Supported Work Demonstration (NSW) job training program. These data sets are first analyzed by LaLonde (1986) and subsequently by Heckman and Hotz (1989), Dehejia and Wahba (1999), Smith and Todd (2001), Firpo (2007b) and Abadie and Imbens (2007). The NSW data is well described in 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 nonexperimental comparison group with 2490 individuals from the PSID. Summary statistics for the two data sets are given in Table 4. We plot the empirical CDFs for the treated and untreated groups for both data sets in Figure 6.

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, since the treatment is randomly assigned in this subset. Our tests are implemented under three different estimations of the propensity score. In the first one, 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 tests are summarized in Table 5. The estimated CDFs and integrated CDFs under different specifications are summarized in Figure 7.

We find that the testing results are robust to the estimations 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 the tests 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 two different estimations of the propensity score. One is proposed by Dehejia and Wahba (1999) in which in addition to those regressors in Wooldridge's (2002) specification, it 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 6. We plot the estimated G_i's and integrated G_i's under different specifications in Figure 8.

The testing results are robust to how the propensity score is estimated except for one case. In both specifications, we accept that conditional on the treated, the outcome under job training first-order and second-order stochastically dominates that without job training. For the first order stochastic dominance of the outcome without job training against that under job training, we reject the null hypothesis at the 1% significance level under Dehejia and Wahba's (1999) specification, but only at the 10% significance level under Firpo's (2007b). On the other hand, we have stronger evidence against second order stochastic dominance, since the p-values are both equal to 0. Hence, we reject second order stochastic dominance at the 1% significance level under both specifications.

To sum up, the empirical results suggest that the real earning under job training SD1 that without job training which implies the SD2 relations. Equivalently, the social welfare 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. On the other hand, we have strong evidence against that F_0 (F_0^t) SD2 F_1 (F_1^t), but weaker evidence against F_0 (F_0^t) SD1 F_1 (F_1^t).

9 Tests for Stochastic Dominance without Unconfoundedness

In previous sections, the stochastic dominance tests are constructed under the unconfoundedness assumption. In this section, we extend our results to cases in which the treatment is not unconfounded. Suppose a binary instrument V is available that takes values on $\{0,1\}$. Let the dummy variable $T(v)$ denote the potential treatment status given $V=v$, but we only observe the treatment status indicator $T = VT(1)+(1-V)T(0)$. Let $Y(v,t)$ denote the potential outcome of the individual that we would observe if $V=v$ and $T=t$. The following assumption is the same as the Assumption 2.1 of Abadie (2003).

Assumption 9.1

1. Independence of the instrument: Conditional on X, V is independent of $Y(0,0)$, $Y(0,1)$, $Y(1,0)$, $Y(1,1)$, $T(1)$ and $T(0)$.
2. Exclusion of the instrument: $P(Y(1,v)=Y(0,v)|X)=1$ for $v\in\{0,1\}$.
3. First stage: $0<P(V=1|X)<1$ and $P(T(1)=1|X)>P(T(0)=1|X)$.
4. Monotonicity: $P(T(1)\geq T(0)|X)=1$.

The exclusion of the instrument assumption indicates that the switch the of instrument does not change the potential outcomes other than through T and this allows us to define $Y(0)=Y(0,0)=Y(1,0)$ and $Y(1)=Y(0,1)=Y(1,1)$. We observe $Y=TY(1)+(1-T)Y(0)$. The first stage assumption implies that V and T are correlated conditional on X. These two assumptions guarantee that V is a valid instrument for T. Under Assumption 9.1, we can identify the average treatment response for the subpopulation called compilers,[7] e.g. Abadie (2002, 2003), Abadie, Angrist and Imbens (2002), Angrist, Imbens and Rubin (1996), Frölich (2007), Frölich and Melly (2008), and Hong and Nekipelov (2008). Let $\mathcal{C}$ denote the group of compilers, and $F_0(z|\mathcal{C})$ and $F_1(z|\mathcal{C})$ the CDFs of $Y(0)$ and $Y(1)$ conditional on compilers respectively. We are interested in if $F_1(z|\mathcal{C})$ first order stochastically dominates $F_0(z|\mathcal{C})$ and refer this relation as the local first order stochastic dominance of $Y(1)$ to $Y(0)$. Hence, the hypotheses are

[7] Those individuals with $D(1)>D(0)$, or equivalently $T(1)=1$ and $T(0)=0$, are compilers. Always-takers are defined by $T(0)=T(1)=1$ and never-takers by $T(0)=T(1)=0$. Finally, the defiers are those with $T(1)<T(0)$. Under the monotonicity assumption, there will be no defiers in our model.

defined as follows:

$$\begin{aligned}
&H_0^c : F_1(z|\mathcal{C}) - F_0(z|\mathcal{C}) \leq 0 \;\text{ for all } z \in \mathcal{Z};\\
&H_1^c : F_1(z|\mathcal{C}) - F_0(z|\mathcal{C}) > 0 \;\text{ for some } z \in \mathcal{Z}.
\end{aligned} \tag{17}$$

Note that

$$\begin{aligned}
&T = VT(1) + (1-V)T(0),\\
&1 - T = V(1-T(1)) + (1-V)(1-T(0)),\\
&Y = TY(1) + (1-T)Y(0).
\end{aligned}$$

These indicate that

$$\begin{aligned}
Y &= TY(1) + (1-T)Y(0)\\
&= (VT(1) + (1-V)T(0))Y(1) + (V(1-T(1)) + (1-V)(1-T(0)))Y(0)\\
&= V(T(1)Y(1) + (1-T(1))Y(0)) + (1-V)(T(0)Y(1) + (1-T(0))Y(0))\\
&= VM(1) + (1-V)M(0) = M(V) = M,
\end{aligned}$$

where we define $M(v) = T(v)Y(1) + (1-T(v))Y(0)$ for $v = 0$ and 1. Define $F_0^m(z) = E[1(M(0) \leq z)]$ and $F_1^m(y) = E[1(M(1) \leq z)]$ which are the CDFs of $M(0)$ and $M(1)$. Therefore,

$$\begin{aligned}
F_1^m(z) - F_0^m(z) &= E[1(M(1) \leq z) - 1(M(0) \leq z)]\\
&= E[(T(1) - T(0))(1(Y(1) \leq z)) - 1(Y(0) \leq z))]\\
&= E[1(Y(1) \leq z) - 1(Y(0) \leq z))|T(1) - T(0) = 1]P[T(1) - T(0) = 1]\\
&= \big(F_1(z|\mathcal{C}) - F_0(z|\mathcal{C})\big)P[T(1) - T(0) = 1].
\end{aligned} \tag{18}$$

The second equality follows from that for $v = 0$ or 1,

$$\begin{aligned}
1(M(v) \leq z) &= 1\big([T(v)Y(1) + (1-T(v))Y(0)] \leq z\big)\\
&= T(v) \cdot 1(Y(1) \leq z) + (1-T(v)) \cdot 1(Y(0) \leq z).
\end{aligned}$$

The third equality holds because $T(1) - T(0)$ takes values on 0 and 1, and $T(1) > T(0)$ iff $T(1) - T(0) = 1$. Hence, given $P[T(1) - T(0) = 1] > 0$ and according to (18), we have $F_1(z|\mathcal{C}) - F_0(z|\mathcal{C}) \leq 0$ iff $M_1(z) - M_0(z) \leq 0$. Therefore, the hypotheses in (17) are equivalent to:

$$\begin{aligned}
&H_0^m : F_1^m(z) - F_0^m(z) \leq 0 \;\text{ for all } z \in \mathcal{Z};\\
&H_1^m : F_1^m(z) - F_0^m(z) > 0 \;\text{ for some } z \in \mathcal{Z}.
\end{aligned} \tag{19}$$

As in Abadie (2002), determining the local first-order stochastic dominance relation between $F_0(y|\mathcal{C})$ and $F_1(y|\mathcal{C})$ is equivalent to determining that between $F_0^w(y)$ and $F_0^w(y)$. Note that the independence of the instrument assumption implies that $(M(0), M(1)) \perp V \,|X$. Therefore, if we treat V as the primary treatment and $M(0)$ and $M(1)$ as the potential outcomes, then the primary treatment is unconfounded. Hence, we can use the same technique developed in Section 3 and 4 to test (17) by testing (19). Consequently, we also extend Abadie's (2002) test to allows for the presence of covariates.

Recent studies also focus on the treatment effect conditional on the treated compliers, e.g. Hong and Nekipelov (2008) and Donald, Hsu and Lieli (2009). As a result, it is interesting to determine the local first order stochastic dominance between the potential outcomes among the treated. We also discuss how to use extend previous technique to this case.

Let $F_0(z|\mathcal{C}, T = 1)$ and $F_1(z|\mathcal{C}, T = 1)$ denote the conditional CDFs of the potential outcomes of the treated compliers. The hypotheses of our interest are

$$\begin{aligned}
&H^c_{0,treated} : F_1(z|\mathcal{C}, T = 1) - F_0(z|\mathcal{C}, T = 1) \le 0 \ \text{ for all } z \in \mathcal{Z};\\
&H^c_{1,treated} : F_1(z|\mathcal{C}, T = 1) - F_0(z|\mathcal{C}, T = 1) > 0 \ \text{ for some } z \in \mathcal{Z}.
\end{aligned} \tag{20}$$

Note that conditional on the compliers, we have $V = T$ and it follows that $F_0(z|\mathcal{C}, T = 1) = F_0(z|\mathcal{C}, V = 1)$ and $F_1(z|\mathcal{C}, T = 1) = F_1(z|\mathcal{C}, V = 1)$. Let $F_0^m(z|V = 1) = E[1(M(0) \le z)|V = 1]$ and $F_1^m(z|V = 1) = E[1(M(1) \le z)|V = 1]$. Therefore,

$$\begin{aligned}
&F_1^m(z|V = 1) - F_0^m(z|V = 1)\\
=&E[1(M(1) \le z) - 1(M(0) \le z)|V = 1]\\
=&E[(T(1) - T(0))(1(Y(1) \le z)) - 1(Y(0) \le z))|V = 1]\\
=&E[1(Y(1) \le z) - 1(Y(0) \le z))|T(1) - T(0) = 1, V = 1]P[T(1) - T(0) = 1|V = 1]\\
=&\big(F_1(z|\mathcal{C}, V = 1) - F_0(z|\mathcal{C}, V = 1)\big)P[T(1) - T(0) = 1|V = 1].\\
=&\big(F_1(z|\mathcal{C}, T = 1) - F_0(z|\mathcal{C}, T = 1)\big)P[T(1) - T(0) = 1|V = 1].
\end{aligned}$$

Since $P[T(1) - T(0) = 1|V = 1]$ is strictly positive, the hypotheses in (20) are equivalent to:

$$\begin{aligned}
&H^m_{0,treated} : F_1^m(z|V = 1) - F_0^m(z|V = 1) \le 0 \ \text{ for all } z \in \mathcal{Z};\\
&H^m_{0,treated} : F_1^m(z|V = 1) - F_0^m(z|V = 1) > 0 \ \text{ for some } z \in \mathcal{Z}.
\end{aligned} \tag{21}$$

As a result, we can use the same technique developed in Section 5 to test (20) by testing (21). Similarly, we can use the same technique in Section 6 to test the local higher order stochastic dominance between the potential outcomes in the whole population and among the treated.

10 Conclusion

In this paper, we propose Kolmogorov-Smirnov-type tests for stochastic dominance relations between the potential outcomes of a binary treatment under the unconfoundedness assumption to evaluate its distributional impacts. Although measures such as QTE, QTT, ITE and ITT are capable of evaluating the distributional impacts of a treatment, they do not fully utilize all information in the distributions. Our stochastic dominance tests compare every point of the CDFs, so they do not have this drawback. For first order stochastic dominance, the test statistic is defined as the supremum of the difference of two inverse probability weighting estimators for the CDFs of the potential outcomes. The critical values (p-values) are approximated based on a simulation method. We show that our test has good size properties and is consistent in the sense that it can detect any violation of the null hypotheses asymptotically. First order stochastic dominance tests in the treated subpopulation, and higher order stochastic dominance tests in the whole population and among the treated are shown to share the same properties. Simulation results support our theoretical findings.

We apply our tests to the National Supported Work Demonstration data and find that job training has a positive effect on real earnings. That is, real earnings under job training stochastically dominates real earnings without job training, and we also find evidence against the reverse relations. Moreover, using any social welfare function and not just the mean would lead to this conclusion.

Finally, we discuss how to apply the technique we have developed to cases in which the unconfoundedness assumption does not hold. Furthermore, one may extend our results to cases in which the comparison is made for any compact subinterval common to both supports. This is useful for poverty comparisons in which one focuses on the welfare for the poor (e.g., Davidson and Duclos (2000)). In addition, one can extend our tests to other situations where one is interested in comparing curves and testing for dominance relations. For example, testing for Lorenz dominance relations is another possible direction.

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 A1: *Suppose Assumption 2.1-3.5 hold. Then*

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

Proof of Lemma A1: The proof is similar to that in the Addendum Proof of Hirano, Imbens and Ridder's (2003). By replacing the Y_i with $1(Y_i \le z)$ in the addendum proof in Hirano, Imbens and Ridder's (2003), we can find a uniform bound for the difference between $\sqrt{N}(\widehat{F}_0(z)-F_0(z))$ and $\sum_{i=1}^{N}\big(\psi_0(W_i,z)-F_0(z)\big)/\sqrt{N}$ over $z \in \mathcal{Z}$. The argument for $\sqrt{N}(\widehat{F}_1(z)-F_1(z))$ is the same. These complete the proof of Lemma A1. $\square$

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. The follow lemma shows that $\mathcal{K}_0$ and $\mathcal{K}_1$ are both $\mathcal{P}$-Donsker.

Lemma A2: *$\mathcal{K}_0$ and $\mathcal{K}_1$ are $\mathcal{P}$-Donsker.*

Proof of Lemma A2: Let $\mathcal{Q}_1 = \{1(Y \le z)] \mid z \in \mathcal{Z}\}$ and it is well-known that $\mathcal{Q}_1$ is a $\mathcal{P}$-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 ϵ-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 bounbed[8] which implies that $\mathcal{Q}_2$ is $\mathcal{P}$-Donsker.

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

Define $g_1(W) = (1-T)/(1-p(X))$ and $g_1(W)$ is a uniformly bounded and measurable function. By (v) of Corollary 9.32 of Kosorok (2008), $g_1 \cdot \mathcal{Q}_1$ is $\mathcal{P}$-Donsker. Similarly, $g_2 \cdot \mathcal{Q}_2$ is $\mathcal{P}$-Donsker where $g_2(W) = (T - p(X))/(1 - p(X))$. Hence, $\mathcal{K}_0 = \{g_1 \cdot 1(Y \leq z) + g_2 \cdot F_0(z|X)] \mid z \in \mathcal{Z}\}$ is $\mathcal{P}$-Donsker follows from (i) of Corollary 9.32, (i) of Theorem 9.30 and the fact that $\mathcal{K}_0$ is a subset of $g_1 \cdot \mathcal{Q}_1 + g_2 \cdot \mathcal{Q}_2$. Similarly, $\mathcal{K}_1$ is also $\mathcal{P}$-Donsker. $\square$

Proof of Lemma 3.6: Let $\mathbb{P}_N$ denote the empirical measure of a sample of W_i and $\mathcal{P}$ be the common distribution of the W_i. Note that $\mathcal{P}\psi_0(W,z) = E[\psi_0(W,z)] = F_0(z)$ for all $z \in \mathcal{Z}$. Lemma A2 and the Donsker's Theorem impliy that $\Psi_0(z)$, the limit of $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \big(\psi_0(W_i,z) - F_0(z)\big)$, is a zero-mean Gaussian process with covariance function $\Omega_0(z_1,z_2) = E[(\psi_0(W_i,z_1) - F_0(z_1))(\psi_0(W_i,z_2) - F_0(z_2))]$. Hence, $\sqrt{N}(\widehat{F}_0(z) - F_0(z)) \Rightarrow \Psi_0(z)$. Similarly, $\sqrt{N}(\widehat{F}_1(z) - F_1(z)) \Rightarrow \Psi_1(z)$. $\square$

Proof of Proposition 3.7: Define $\overline{S}_N \equiv \sqrt{N} \sup_{z\in\mathcal{Z}} \big(\widehat{F}_1(z) - \widehat{F}_0(z) - (F_1(z) - F_0(z))\big)$. By the Proposition 3.6, we have $\sqrt{N}\big(\widehat{F}_1(z) - \widehat{F}_0(z) - (F_1(z) - F_0(z))\big) \Rightarrow \Psi_1(z) - \Psi_0(0)$ and by CMT, $\overline{S}_N \xrightarrow{D} \overline{S}$.

Under the null hypothesis, $F_1(z) \leq F_0(z)$ for all $z \in \mathcal{Z}$ and $\widehat{S}_N = \sqrt{N} \sup_{z\in\mathcal{Z}} \big(\widehat{F}_1(z) - \widehat{F}_0(z)\big) \leq \overline{S}_N$. Hence, $P(\widehat{S}_N \leq c) \geq P(\overline{S}_N \geq c)$. By taking lim inf on both sides, we have

$$\liminf P(\widehat{S}_N \leq c) \geq \liminf P(\overline{S}_N \geq c) = P(\overline{S} \leq c),$$

and the last equality follows from the fact that $\overline{S}_N \xrightarrow{D} \overline{S}$. Finally,

$$\begin{aligned}\limsup P(\widehat{S}_N > c) &= \limsup 1 - P(\overline{S}_N \leq c) = 1 - \liminf P(\overline{S}_N \leq c)\\ &\leq 1 - P(\overline{S} \leq c) = P(\overline{S} > c)\end{aligned}$$

which shows the first part.

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

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

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. This shows the second part. $\square$

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

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

The usual bound for series estimators applies:

$$\sup_{z\in\mathcal{Z},x\in\mathcal{X}} \left|\Phi_K(z)'\xi_K^{-1}R^K(X_i) - F_1(z|x)\right| \le \Delta_1\zeta(K)O_p(\sqrt{\frac{\zeta(K)}{N}}) + \Delta_2 K^{\frac{-s'}{r}}$$

where s' is the number of continuous derivatives of $F_1(z|x)$ and $\zeta(K) = \sup_{x\in\mathcal{X}}\|R^K(x)\|$. Under the assumptions, we have $\sup_{z\in\mathcal{Z},x\in\mathcal{X}}\left|\Phi_K(z)'\xi_K^{-1}R^K(X_i) - F_1(z|x)\right| = o_p(1)$. Also,

$$\begin{aligned}
&\sup_{z\in\mathcal{Z}}\|\widehat{\Phi}_K(z) - \Phi_K(z)\| = \sup_{z\in\mathcal{Z}}\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{p}(X_i) - p(X_i)}{p(X_i)\hat{p}(X_i)}1(Y_i \le z)T_iR^K(X_i)\right\| \\
&\le \sup_{z\in\mathcal{Z}}\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{p(X_i)\hat{p}(X_i)}\right|\cdot|\hat{p}(X_i) - p(X_i)|\cdot|1(Y_i \le z)T_i|\cdot\|R^K(X_i)\| \\
&\le \Delta\sup_{x\in\mathcal{X}}|\hat{p}(x) - p(x)|\sup_{x\in\mathcal{X}}\|R^K(x)\| = \Delta_1\zeta^2(K)O_p\Big(\sqrt{\frac{\zeta(K)}{N}}\Big) + \Delta_2\zeta(K)K^{\frac{-s}{r}}.
\end{aligned}$$

Hence, under the assumptions,

$$\begin{aligned}
&\sup_{z\in\mathcal{Z},x\in\mathcal{X}}|\widehat{\Phi}_K(z)'\xi_K^{-1}R^K(X_i) - \Phi_K(z)'\xi_K^{-1}R^K(X_i)| \\
&\le \Delta_1\zeta^3(K)O_p\Big(\sqrt{\frac{\zeta(K)}{N}}\Big) + \Delta_2\zeta^2(K)K^{\frac{-s}{r}} = o_p(1).
\end{aligned}$$

Therefore,

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

WLOG we assume that $\widetilde{F}_1(z|x)$ is bounded between 0 and 1.[9] 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 $F_1(z|x) - \widetilde{F}_1(z|x) > F_1(z|x) - \widehat{F}_1(z|x) > 0$.

[9] 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 $\widetilde{F}_1(z|x) > 1$.

If $\widehat{F}_1(z|x) \leq F_1(z|x)$, 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

$$|\widehat{F}_1(z|x) - F_1(z|x)| \leq \max\{|\widetilde{F}_1(y_{h-1}|x) - F_1(y_{h-1}|x)|, |\widetilde{F}_1(z|x) - F_1(z|x)|\}.$$

As a result,

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

Then by induction, we can show that

$$\sup_{z\in\mathcal{Z}} |\widehat{F}_1(z|x) - F_1(z|x)| \leq \sup_{z\in\mathcal{Z}} |\widetilde{F}_1(z|x) - F_1(z|x)|$$

It follows that

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

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

Proof of Lemma 4.2:

We rewrite $\Psi_u(z)$ as

$$\begin{aligned}
\Psi_u(z) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i \Big\{ &\frac{T_i \cdot 1(Y_i \leq z)}{p(X_i)} - \frac{(1-T_i)\cdot 1(Y_i \leq z)}{1-p(X_i)} - (F_1(z) - F_0(z)) \\
&- (T - p(X_i))\Big(\frac{F_1(z|X_i)}{p(X_i)} + \frac{F_0(z|X_i)}{1-p(X_i)}\Big)\Big\} \\
&+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i \Big\{\frac{T_i \cdot 1(Y_i \leq z)}{\hat{p}(X_i)} - \frac{T_i \cdot 1(Y_i \leq z)}{p(X_i)}\Big\} \\
&- \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i \Big\{\frac{(1-T_i)\cdot 1(Y_i \leq z)}{1-\hat{p}(X_i)} - \frac{(1-T_i)\cdot 1(Y_i \leq z)}{1-p(X_i)}\Big\} \\
&- \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i \Big\{\big(p(X_i) - \hat{p}(X_i)\big)\Big(\frac{\widehat{F}_1(z|X_i)}{\hat{p}(X_i)} + \frac{\widehat{F}_0(z|X_i)}{1-\hat{p}(X_i)}\Big)\Big\} \\
&- \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i \Big\{\big(T_i - p(X_i)\big)\Big[\Big(\frac{\widehat{F}_1(z|X_i)}{\hat{p}(X_i)} + \frac{\widehat{F}_0(z|X_i)}{1-\hat{p}(X_i)}\Big) \\
&\qquad - \Big(\frac{F_1(z|X_i)}{p(X_i)} + \frac{F_0(z|X_i)}{1-p(X_i)}\Big)\Big]\Big\} \\
&- \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i \Big\{(\widehat{F}_1(z) - \widehat{F}_0(z)) - (F_1(z) - F_0(z))\Big\}
\end{aligned} \tag{22}$$

We first claim that

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

For any sequence $\mathcal{W}$, define that

$$f_{Ni}(U_i,z|\mathcal{W})=\frac{U_i}{\sqrt{N}}\left\{\frac{T_i\cdot 1(Y_i\le z)}{\hat{p}(X_i)}-\frac{T_i\cdot 1(Y_i\le z)}{p(X_i)}\right\},$$

$$\chi_N(z|\mathcal{W})=\sum_{i=1}^{N}\left\{f_{Ni}(U_i,z|\mathcal{W})\right\}$$

There exists $M>0$ such that $F_{N,i}=\frac{M}{\sqrt{N}}|U_i|$ is the envelope of $f_{Ni}(U_i,z|\mathcal{W})$ since T_i, $1(Y_i\le z)$, $\hat{p}(X_i)$ and $p(X_i)$ are all bounded. First, we show (i)-(v) of Theorem 10.6 (functional central limit theorem) of Pollard (1990) with probability 1. By the definition of $F_{N,i}$, it is true that (iii) and (iv) hold in probability. For $z_1\le z_2$,

$$\begin{aligned}E_u[\chi_N(z_1|\mathcal{W})\chi_N(z_2|\mathcal{W})]&=\frac{1}{N}\sum_{i=1}^{N}E_u\left[U_i^2T_i\cdot 1(Y_i\le z_1)\left\{\frac{1}{\hat{p}(X_i)}-\frac{1}{p(X_i)}\right\}^2\right]\\&=\frac{1}{N}\sum_{i=1}^{N}T_i\cdot 1(Y_i\le z_1)\left\{\frac{1}{\hat{p}(X_i)}-\frac{1}{p(X_i)}\right\}^2\\&=\frac{1}{N}\sum_{i=1}^{N}\frac{T_i\cdot 1(Y_i\le z_1)}{\hat{p}^2(X_i)p^2(X_i)}\left\{p(X_i)-\hat{p}(X_i)\right\}^2\\&\le\Delta\sup_x|\hat{p}(x)-p(x)|^2\to 0,\quad\text{given }\mathcal{W}\text{ in probability.}\end{aligned}$$

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 since $\hat{p}(X)$ and $p(x)$ are both bounded away from 0 and $\sup_x|\hat{p}(x)-p(x)|^2\xrightarrow{p}0$. As a result, (ii) holds. Define

$$\rho_N(z_1,z_2)=\left(\sum_{i=1}^{N}E_u\big[f_{Ni}(U_i,z_1|\mathcal{W})-f_{Ni}(U_i,z_2|\mathcal{W})\big]\right)^{\frac{1}{2}}.$$

Note that for any $z_1<z_2$,

$$\begin{aligned}\rho_N(z_1,z_2)^2&=\sum_{i=1}^{N}E_u\big[f_{Ni}(U_i,z_1|\mathcal{W})-f_{Ni}(U_i,z_2|\mathcal{W})\big]^2\\&=\frac{1}{N}\sum_{i=1}^{N}E_u\left[U_i^2\cdot T_i\cdot 1(z_1<Y_i\le z_2)\left\{\frac{1}{\hat{p}(X_i)}-\frac{1}{p(X_i)}\right\}^2\right]\\&\le\Delta\sup_x|\hat{p}(x)-p(x)|^2\to 0,\quad\text{given }\mathcal{W}\text{ in probability.}\end{aligned}$$

Since, $\rho_N(z_1, z_2)$ converges uniformly to 0 over z_1 and z_2 given $\mathcal{W}$ in probability, it implies that (v) holds given $\mathcal{W}$ in probability

Given $\mathcal{W}$ and $\omega_u \in \Omega_u$, let

$$F_{Ni}(\omega_u) = \frac{M}{\sqrt{N}}|U_i(\omega_u)|, \quad \overrightarrow{F_N}(\omega_u) = \big(F_{N1}(\omega_u), ..., F_{NN}(\omega_u)\big),$$
$$f_{Ni}(U_i(\omega_u), z|\mathcal{W}) = \left[\frac{U_i(\omega_u)T_i}{\sqrt{n}}\left(\frac{1}{\hat{p}(X_i)} - \frac{1}{p(X_i)}\right)\right] 1(Y_i \le z),$$
$$\overrightarrow{f_N}(z, \omega_u|\mathcal{W}) = \big(f_{N1}(U_1(\omega_u), z|\mathcal{W}), ..., f_{NN}(U_N(\omega_u), z|\mathcal{W})\big),$$
$$\mathcal{F}_{N\omega_u} = \{\overrightarrow{f_N}(z, \omega_u|\mathcal{W}) : \ z \in [0, \bar{z}]\}$$

For any two vectors $a = (a_1, ..., a_N)$ and $b = (b_1, ..., b_N)$ in $\mathbb{R}^N$, define the the pointwise product $\odot$ as $a \odot b = (a_1 b_1, ..., a_N b_N)$. Define $\Theta_N = (\theta_1, ..., \theta_N)$ be a vector of non-negative weights and $\Theta_N \odot \mathcal{F}_{N\omega_u} = \{\Theta_N \odot \overrightarrow{f_N}(z, \omega_u|\mathcal{W}) : \ z \in [0, \bar{z}]\}$. The packing number $D(\epsilon, T_0)$ for a subset of T_0 of a metric space with metric d is defined as the largest m for which there exist points $t_1, ..., t_m$ in T_0 with $d(t_i, t_j) > \epsilon$ for $i \neq j$. We use the ℓ_1 norm on $\mathbb{R}^N$ which is defined as $|(v_1, ..., v_N)|_1 = \sum_{i=1}^N |v_i|$. Since $D(\epsilon|\Theta_N \odot \overrightarrow{F_N}|_1, \Theta_N \odot \mathcal{F}_{N\omega_u}) = D(\epsilon|h\Theta_N \odot \overrightarrow{F_N}|_1, h\Theta_N \odot \mathcal{F}_{N\omega_u})$ for all $h > 0$, WLOG we can rescale Θ_N such that $|\Theta_N \odot \overrightarrow{F_N}|_1 = 1$. Define

$$\begin{aligned} d(z_1, z_2) &= \big|\overrightarrow{f_N}(z_1, \omega_u|\mathcal{W}) - \overrightarrow{f_N}(z_2, \omega_u|\mathcal{W})\big|_1 \\ &= \sum_{i=1}^N \big|f_{Ni}(U_i(\omega_u), z_1|\mathcal{W}) - f_{Ni}(U_i(\omega_u), z_2|\mathcal{W})\big|. \end{aligned}$$

Since $f_{Ni}(U_i(\omega_u), z|\mathcal{W})$ is monotonic in z for each i, then it follows that for any $z_1 \le z_2 \le z_3$,

$$\begin{aligned} d(z_3, z_1) &= \sum_{i=1}^N \big|f_{Ni}(U_i(\omega_u), z_3|\mathcal{W}) - f_{Ni}(U_i(\omega_u), z_1|\mathcal{W})\big| \\ &= \sum_{i=1}^N \big|f_{Ni}(U_i(\omega_u), z_3|\mathcal{W}) - f_{Ni}(U_i(\omega_u), z_2|\mathcal{W}) \\ &\qquad\qquad + f_{Ni}(U_i(\omega_u), z_2|\mathcal{W}) - f_{Ni}(U_i(\omega_u), z_1|\mathcal{W})\big| \\ &= \sum_{i=1}^N \big|f_{Ni}(U_i(\omega_u), z_3|\mathcal{W}) - f_{Ni}(U_i(\omega_u), z_2|\mathcal{W})\big| \\ &\qquad\qquad + \big|f_{Ni}(U_i(\omega_u), z_2|\mathcal{W}) - f_{Ni}(U_i(\omega_u), z_1|\mathcal{W})\big| \\ &= d(z_3, z_2) + d(z_2, z_1). \end{aligned}$$

We claim that $D(\epsilon, \Theta_N \odot \mathcal{F}_{N\omega_u}) \leq \frac{1}{\epsilon} + 1$. Suppose not, then there exists $0 \leq z_1 < z_2 < \ldots < z_k \leq \bar{z}$ with $k \geq \frac{1}{\epsilon} + 1$ and $d(z_i, z_j) > \epsilon$ for all $i \neq j$ which implies that

$$\begin{aligned} d(0, \bar{z}) =& d(0, z_1) + d(z_1, z_2) + \cdots + d(z_k, \bar{z}) \\ & \geq d(z_1, z_2) + \cdots + d(z_{k-1}, z_k) \geq k\epsilon > (k-1)\epsilon \geq \frac{1}{\epsilon}\epsilon = 1. \end{aligned}$$

But this contradicts to the fact that $d(0, \bar{z}) \leq |\Theta_N \odot \overrightarrow{F_N}|_1 = 1$. Hence, for all sequences W_1, W_2, ..., we have $D(\epsilon, \Theta_N \odot \mathcal{F}_{N,\omega_u}) \leq \frac{1}{\epsilon} + 1 \equiv \beta(\epsilon)$ for all $\omega_u \in \Omega_u$ for any Θ_N and $\int_0^1 \sqrt{\beta(\epsilon)} d\epsilon < \infty$. That is, (i) holds for all W.

Since (i)-(v) hold given $\mathcal{W}$ in probability, then $\chi_N(z|\mathcal{W})$ weakly converges to a tight Gaussian process given W in probability with a covariance function equal to the limit of $E_u[\chi_N(z_1)\chi_N(z_2)]$ which is equal to 0 for all $z_1 \leq z_2$. These imply that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N U_i \Big\{ \frac{T_i \cdot 1(Y_i \leq z)}{\hat{p}(X_i)} - \frac{T_i \cdot 1(Y_i \leq z)}{p(X_i)} \Big\} \overset{\mathbb{P}_w}{\Rightarrow} 0.$$

Similarly,

$$\begin{aligned} &\frac{1}{\sqrt{N}} \sum_{i=1}^N U_i \Big\{ \frac{(1-T_i) \cdot 1(Y_i \leq z)}{1 - \hat{p}(X_i)} - \frac{(1-T_i) \cdot 1(Y_i \leq z)}{1 - p(X_i)} \Big\} \overset{\mathbb{P}_w}{\Rightarrow} 0, \\ &\frac{1}{\sqrt{N}} \sum_{i=1}^N U_i \Big\{ \big(p(X_i) - \hat{p}(X_i)\big) \Big(\frac{\widehat{F}_1(z|X_i)}{\hat{p}(X_i)} + \frac{\widehat{F}_0(z|X_i)}{1 - \hat{p}(X_i)} \Big) \Big\} \overset{\mathbb{P}_w}{\Rightarrow} 0. \end{aligned}$$

Define

$$\begin{aligned} g_{Ni}(U_i, z|\mathcal{W}) &= \frac{U_i}{\sqrt{N}} \Big\{ \big(T_i - p(X_i)\big) \Big(\frac{\widehat{F}_1(z|X_i)}{\hat{p}(X_i)} + \frac{\widehat{F}_0(z|X_i)}{1 - \hat{p}(X_i)} \Big) \Big\}, \\ \varphi_N(z|\mathcal{W}) &= \sum_{i=1}^N g_{Ni}(U_i, z|\mathcal{W}), \end{aligned}$$

and pick M large enough so that $F_{N_i} = \frac{M}{\sqrt{N}}|U_i|$ is the envelope of $g_{Ni}(U_i, z|\mathcal{W})$. It follows that (iii) and (iv) hold in probability. Since $g_{Ni}(U_i, z|\mathcal{W})$ is monotonic in z for each i, we can show that (i) holds by the previous arguments. For any $z_1 \in \mathcal{Z}$

$$\begin{aligned} E_u[\varphi_N^2(z_1|\mathcal{W})] &= \frac{1}{N} \sum_{i=1}^N E_u \Big[U_i^2 \Big\{ \big(T_i - p(X_i)\big) \Big(\frac{\widehat{F}_1(z_1|X_i)}{\hat{p}(X_i)} + \frac{\widehat{F}_0(z_1|X_i)}{1 - \hat{p}(X_i)} \Big)^2 \Big] \\ &= \frac{1}{N} \sum_{i=1}^N \big(T_i - p(X_i)\big)^2 \Big(\frac{F_1(z_1|X_i)}{p(X_i)} + \frac{F_0(z_1|X_i)}{1 - p(X_i)} \Big)^2 \\ &\quad + \frac{1}{N} \sum_{i=1}^N \big(T_i - p(X_i)\big)^2 \Big\{ \Big(\frac{\widehat{F}_1(z_1|X_i)}{\hat{p}(X_i)} + \frac{\widehat{F}_0(z_1|X_i)}{1 - \hat{p}(X_i)} \Big)^2 - \Big(\frac{F_1(z_1|X_i)}{p(X_i)} + \frac{F_0(z_1|X_i)}{1 - p(X_i)} \Big)^2 \Big\} \end{aligned}$$

Since $(T_i - p(X_i))^2 \big((F_1(z_1|X_i)/p(X_i)) + (F_0(z_1|X_i)/1 - p(X_i))\big)^2$ is bounded, then by LLN's, the first term after the second equality converges in probability to the unconditional mean $E[(T_i - p(X_i))^2 \big((F_1(z_1|X_i)/p(X_i)) + (F_0(z_1|X_i))/1 - p(X_i))\big)^2]$. The second term converges to 0 in probability, since

$$
\begin{aligned}
&\Big|\frac{1}{N}\sum_{i=1}^{N}(T_i - p(X_i))^2\Big\{\Big(\frac{\widehat{F}_1(z_1|X_i)}{\hat{p}(X_i)} + \frac{\widehat{F}_0(z_1|X_i)}{1-\hat{p}(X_i)}\Big)^2 - \Big(\frac{F_1(z_1|X_i)}{p(X_i)} + \frac{F_0(z_1|X_i)}{1-p(X_i)}\Big)^2\Big\}\Big| \\
\leq&\Delta \sup_{z,x}\Big|\Big(\frac{\widehat{F}_1(z|x)}{\hat{p}(x)} - \frac{\widehat{F}_0(z|x)}{1-\hat{p}(x)}\Big)^2 - \Big(\frac{F_1(z|X_i)}{p(x)} + \frac{F_0(z|x)}{1-p(x)}\Big)\Big| \xrightarrow{p} 0.
\end{aligned}
$$

The last line follow from the fact that $\widehat{F}_0(z|x)$, $\widehat{F}_1(z|x)$ and $\hat{p}(x)$ converge uniformly to $F_0(z|x)$, $F_1(z|x)$ and $p(x)$ respectively. It implies that

$$
E_u[\varphi_N^2(z_1|\mathcal{W})] \to E\Big[(T_i - p(X_i))^2\Big(\frac{F_1(z_1|X_i)}{p(X_i)} + \frac{F_0(z_1|X_i)}{1-p(X_i)}\Big)^2\Big], \quad \text{given } \mathcal{W} \text{ in probability.}
$$

Following the same argument, we have for any $z_1 \leq z_2$ and given $\mathcal{W}$ in probability

$$
\begin{aligned}
&E_u[\varphi_N(z_1|\mathcal{W})\varphi_N(z_2|\mathcal{W})] \\
\to&E\Big[(T_i - p(X_i))^2\Big(\frac{F_1(z_1|X_i)}{p(X_i)} + \frac{F_0(z_1|X_i)}{1-p(X_i)}\Big)\Big(\frac{F_1(z_2|X_i)}{p(X_i)} + \frac{F_0(z_2|X_i)}{1-p(X_i)}\Big)\Big] \equiv \Omega_\varphi(z_1, z_2).
\end{aligned}
$$

Hence, (ii) holds. For any z_1 and z_2 define

$$
\begin{aligned}
\varrho_N(z_1, z_2) &= \left(\sum_{i=1}^{N} E_u\big[g_{Ni}(U_i, z_1|\mathcal{W}) - g_{Ni}(U_i, z_2|\mathcal{W})\big]^2\right)^{\frac{1}{2}}, \\
\varrho(z_1, z_2) &= \left(E\Big[(T_i - p(X_i))^2\Big(\frac{F_1(z_2|X_i) - F_1(z_1|X_i)}{p(X_i)} + \frac{F_0(z_2|X_i) - F_0(z_1|X_i)}{1-p(X_i)}\Big)\Big]\right)^{\frac{1}{2}}.
\end{aligned}
$$

Applying the same argument for (ii) and by uniform law of large number (Lemma 2.4 of Newey and McFadden (1994)), we show that given $\mathcal{W}$ in probability

$$
\sup_{z_1, z_2} |\varrho_N^2(z_1, z_2) - \varrho^2(z_1, z_2)| \to 0.
$$

As a result, this is sufficient to show (v) holds. It follows that $\varphi_N(z|\mathcal{W})$ weakly converges to a tight Gaussian process given $\mathcal{W}$ in probability with covariance function $\Omega_\varphi(z_1, z_2)$. Also, define

$$
\begin{aligned}
q_{Ni}(U_i, z|\mathcal{W}) &= \frac{U_i}{\sqrt{N}}\Big\{(T_i - p(X_i))\Big(\frac{F_1(z|X_i)}{p(X_i)} + \frac{F_0(z|X_i)}{1-p(X_i)}\Big)\Big\}, \\
\eta_N(z|\mathcal{W}) &= \sum_{i=1}^{N} q_{Ni}(U_i, z|\mathcal{W}).
\end{aligned}
$$

Similarly, it is true that given $\mathcal{W}$ in probability $\eta_N(z|\mathcal{W})$ weakly converges to a tight Gaussian process with the same covariance function as $\varphi_N(z|\mathcal{W})$. Note that the fifth term in (22) is equal to $\varphi_N(z|\mathcal{W}) - \eta_N(z|\mathcal{W})$ and the limit of it is also a tight Gaussian process given $\mathcal{W}$ in probability. It is easy to show that limit of the covariance function, $E[(\varphi_N(z_1|\mathcal{W}) - \eta_N(z_2|\mathcal{W}))(\varphi_N(z_2|\mathcal{W}) - \eta_N(z_2|\mathcal{W}))]$, is equal to 0 for all z_1 and z_2. Therefore, the fifth term in (22) weakly converges to 0 given $\mathcal{W}$ in probability.

For the last term in (22),

$$
\begin{aligned}
& P\left(\sup_{z\in\mathcal{Z}}\left|\left(\widehat{F}_1(z) - \widehat{F}_0(z)) - (F_1(z) - F_0(z)\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i\right| > \epsilon\right) \\
\leq & \frac{E\left(\sup_{z\in\mathcal{Z}}\left|(\widehat{F}_1(z) - \widehat{F}_0(z)) - (F_1(z) - F_0(z))\right|\right)^2}{\epsilon^2} \to 0.
\end{aligned}
$$

The first inequality follows from the Chebyshev's inequality and $E(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i)^2 = 1$. The last line holds because

$$
E\left(\sup_{z\in\mathcal{Z}}\left|(\widehat{F}_1(z) - \widehat{F}_0(z)) - (F_1(z) - F_0(z))\right|\right)^2 \to 0
$$

by dominated convergence theorem.

Accordingly,

$$
\Psi_u(z) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i\left\{\frac{T_i\cdot 1(Y_i\leq z)}{p(X_i)} - \frac{(1-T_i)\cdot 1(Y_i\leq z)}{1-p(X_i)} - (F_1(z) - F_0(z))\right\} \overset{\mathbb{P}_w}{\Rightarrow} 0.
$$

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

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i\left\{\frac{T_i\cdot 1(Y_i\leq z)}{p(X_i)} - \frac{(1-T_i)\cdot 1(Y_i\leq z)}{1-p(X_i)} - (F_1(z) - F_0(z))\right\} \overset{\mathbb{P}_w}{\Rightarrow} \Psi_(z),
$$

and Lemma 4.2 follows. $\square$

Proof of Lemma 4.3: By CMT, we have

$$
\overline{S}_u \xrightarrow{D} \overline{S} \quad \text{given } \mathcal{W} \text{ in probability.}
$$

Note that $P(\overline{S}\leq 0)\leq 1/2$ 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, $\hat{c}$ the $(1-\alpha_0)$th quantile of $\overline{S}_u$, converges in probability to $c_0 > 0$, the $(1-\alpha_0)$th quantile of $\overline{S}$. $\square$

Proof of Theorem 4.4: Under the null,

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

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 Proposition 3.7 and the definition of c_0. This shows the first part of Theorem 4.4.

The second part follows from the fact that $\widehat{S}_N \xrightarrow{p} \infty$ when the null hypothesis is wrong, but $\hat{c} \xrightarrow{p} c_0 < \infty$. $\square$

Proof of Theorem 5.2: It is not hard to show the following:

$$\sup_{z\in\mathcal{Z}}\left|\sqrt{N}\big(\widehat{G}_1(z)-\widehat{G}_0(z)-(G_1(z)-G_0(z))\big)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_t(W_i,z)\right|=o_p(1),$$

where

$$\begin{aligned}\psi_t(W,z)=&T(1(Y\le z)-F_1(z|X))-\frac{p(X)}{1-p(X)}(1-T)(1(Y\le z)-F_0(z|X))\\&+T\big(F_1(z|X)-F_0(z|X)\big)-(G_1(z)-G_0(z)).\end{aligned}$$

Given this result, the rest of the proof is similar to that of the unconditional case and we omit it. $\square$

Proof of Theorem 6.1: Theorem 4.4 shows the case for $j=1$. We show the case for $j=2$ and the proof for general cases is similar.

The integral operator $\mathcal{I}_2$ is a continuous function from $\ell^\infty(\mathcal{Z})$ to $\ell^\infty(\mathcal{Z})$, the set of all uniformly bounded real functions on $\mathcal{Z}$. By CMT, we have

$$\begin{aligned}&\mathcal{I}_2(z;\widehat{F}_1-\widehat{F}_0-F_1+F_0)\Rightarrow\mathcal{I}_2(z;\Psi),\\&\mathcal{I}_2(z;\Psi_u)\Rightarrow\mathcal{I}_2(z;\Psi)\quad\text{given }\mathcal{W}\text{ in probability.}\end{aligned}$$

Define $\overline{S}^j=\sup_{z\in\mathcal{Z}}\mathcal{I}_2(z;\Psi)$. By the same proof of Theorem 4.4, we can show that when H_0^j is true, $\limsup P(\text{reject } H_0^j)=\limsup P(\widehat{S}_N^j>\hat{c}^j)\le P(\overline{S}^j>c_0^j)=\alpha_0$. This shows the first part.

When the null hypothesis is wrong, the second part follows from the fact that $\widehat{S}_N^j\xrightarrow{p}\infty$ and $\hat{c}^j$ is bounded in probability. $\square$

Proof of Theorem 6.2: The proof is similar to that of Theorem 6.1 and we omit it. $\square$

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

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

Anderson, G. (1996): "Nonparametric Tests for Stochastic Dominance," *Econometrica*, **64**, 1183–1193.

Angrist, J., G. Imbens and D. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of American Statistical Association*, **91**, 444–455.

Barnow, B., G. Cain and A. Goldberger (1980): "Issues in the Analysis of Selectivity Bias" in *Evaluation Studies*, Vol. 5, ed. by E. Stromsorfer and G. Farkas. San Francisco: Sage.

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

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

Dehejia, R. and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, **94**, 1053–1062.

Donald, S. G. Y.-C. Hsu, and R. P. Lieli (2009): "Efficient Nonparametric IV Estimation of Local Average Treatment Effects Using the Estimated Propensity Score," 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.*

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

Frolich, M. and B. Melly (2008): "Unconditional Quantile Treatment Effects under Endogeneity," IZA Discussion Paper, 3288.

Heckman, J. and J. Hotz (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., H. Ichimura and P. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, **64**, 605–654.

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

Heckman, J., H. Ichimura, J. Smith and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, **66**, 1017–1098.

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

Hong, H., and D. Nekipelov (2008): "Semiparametric Efficiency in Nonlinear LATE Models," Working Paper.

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.

McFadden, D. (1989): Testing for Stochastic Dominance, in Studies in Economics of Uncertainty: In Honor of Josef Hadar, ed. by T. B. Fombay and T. K. Seo, New York, Berlin, London and Tokyo: Springer.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Volume 4, ed. R. F. Engle and D. McFadden. Amsterdam: North Holland, 2111–2245.

Pollard, D. (1990): Empirical Processes: Theory and Application, CBMS Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

Rosenbaum, P. (1987): "Model-Based Direct Adjustment," *Journal of American Statistical Association*, **82**, 387–394.

Rosenbaum, P. and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, **70**, 41–55.

Rosenbaum, P. and D. Rubin (1985): "Reducing Bias in Observational Studies Using Subclassfication on the Propensity Score," *Journal of American Statistical Association*, **79**, 516–524.

Rubin D. (1978): "Bayesian Inference for Causal Effects: the Role of Randomization," *Annals of Statistics*, **6**, 34–58.

Rubin, D. and N. Thomas (1996): "Matching Using Estimated Propensity Scores: Relating Theory to Practice," *Biometrics*, **52**, 249–264.

Smith, J. A. and P. E. Todd (2001): "Reconciling Con.icting Evidence on the Performance of Propensity-Score Matching Methods," *American Economic Review*, Papers and Proceedings, **91**, 112–118.

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. W. and J. A. Wellner (1996): Weak Convergence and Empirical Processes: With Application to Statistics, New York: Springer-Verlag.

Wooldridge, J. M. (2002): Econometric Analysis of Cross Section and Panel Data, Cambridge, MA: MIT Press.

Table 1: Rejection rates.

	$N = 200$			
	F_1 SD1 F_0	F_1 SD2 F_0	F_1^t SD1 F_0^t	F_1^t SD2 F_0^t
Ex. 7.1	0.058	0.053	0.041	0.038
Ex. 7.2	0.141	0.002	0.117	0.004
Ex. 7.3	0.437	0.266	0.451	0.289
	$N = 400$			
Ex. 7.1	0.055	0.057	0.039	0.042
Ex. 7.2	0.321	0.002	0.273	0.001
Ex. 7.3	0.781	0.466	0.816	0.490

Table 2: T is Randomly Assigned.

	No covariate	
	F_1 SD1 F_0	F_1 SD2 F_0
Ex. 7.1	0.043	0.049
Ex. 7.2	0.146	0.004
Ex. 7.3	0.452	0.274
	X	
	F_1 SD1 F_0	F_1 SD2 F_0
Ex. 7.1	0.048	0.051
Ex. 7.2	0.146	0.005
Ex. 7.3	0.471	0.279

Table 3: T is not Randomly Assigned.

No Covariate, $N = 200$		
	F_1 SD1 F_0	F_1 SD2 F_0
Ex. 7.4	0.162	0.154
Ex. 7.5	0.043	0.033

X, $N = 200$		
	F_1 SD1 F_0	F_1 SD2 F_0
Ex. 7.4	0.047	0.047
Ex. 7.5	0.168	0.165

No Covariate, $N = 400$		
	F_1 SD1 F_0	F_1 SD2 F_0
Ex. 7.4	0.188	0.188
Ex. 7.5	0.037	0.011

X, $N = 400$		
	F_1 SD1 F_0	F_1 SD2 F_0
Ex. 7.4	0.057	0.046
Ex. 7.5	0.232	0.212

Table 4: Summary Statistics.

	RE74 Treated	RE74 Control	PSID-1 Control
Sample size	185	260	2490
Earnings in 1978	6.349	4.555	21.554
Earnings in 1974	2.057	2.107	19.429
Earnings in 1975	1.532	1.267	19.063
Age	25.81	25.05	34.85
Education	10.35	10.09	12.11
Black	0.84	0.83	0.25
Hispanic	0.059	0.1	0.032
No degree	0.71	0.83	0.31
Married	0.19	015	0.87
Unemployed in 1974	0.708	0.75	0
Unemployed in 1975	0.6	0.685	0
Maximum earning in 1978	60.3	39.7	121.2

Note: Earnings in calender year = real earnings in calendar year in thousand dollars; Age = age in years; Education = number of years of schooling; Black = 1 if black, 0 otherwise; Hispanic = 1 if hispanic, 0 otherwise; No degree = 1 if Education $<$ 12, 0 otherwise; Married = 1 if married, 0 otherwise; Unemployed in calendar year =1 if unemployed for all calendar year.

Table 5: Stochastic Dominance in RE74 Data Set.

	No covariate			
	F_1 SD1 F_0	F_1 SD2 F_0	F_0 SD1 F_1	F_0 SD2 F_1
statistic	-0.002	0	2.788	0.627
p-value	1.000	1.000	0.018	0.003
	Age			
	F_1 SD1 F_0	F_1 SD2 F_0	F_0 SD1 F_1	F_0 SD2 F_1
statistic	-0.004	0	2.825	0.600
p-value	1.000	1.000	0.018	0.004
	Wooldridge's specification			
	F_1 SD1 F_0	F_1 SD2 F_0	F_0 SD1 F_1	F_0 SD2 F_1
statistic	0.096	0	2.790	0.638
p-value	1.000	1.000	0.017	0.002

Table 6: Stochastic Dominance in PSID-1 Data Set.

	Dehejia and Wahba's 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
statistic	-1.047	0.000	2.430	1.710
p-value	1.000	1.000	0.004	0.000
	Firpo's 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
statistic	-0.083	0.000	0.714	0.273
p-value	1.000	1.000	0.066	0.000

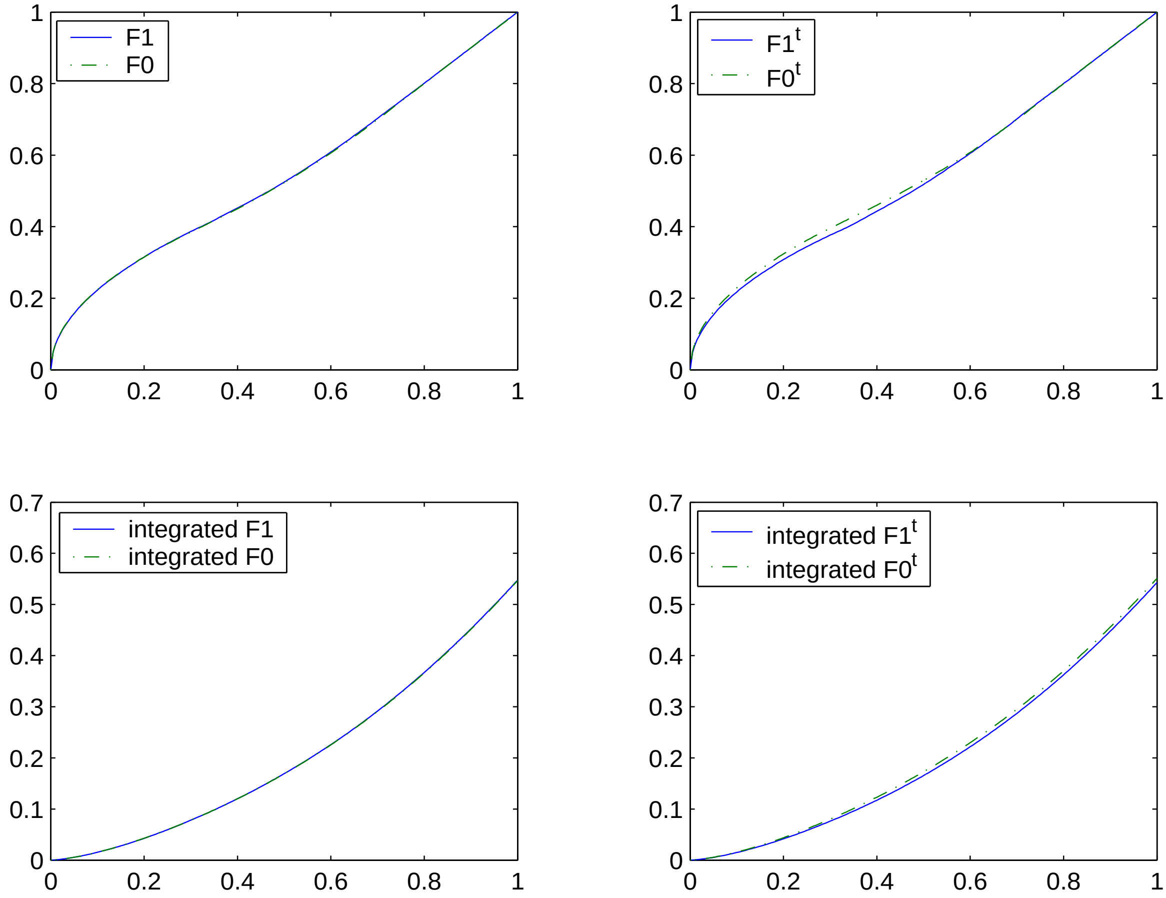

Figure 1: The CDFs and the integrated CDFs of Example 7.1.

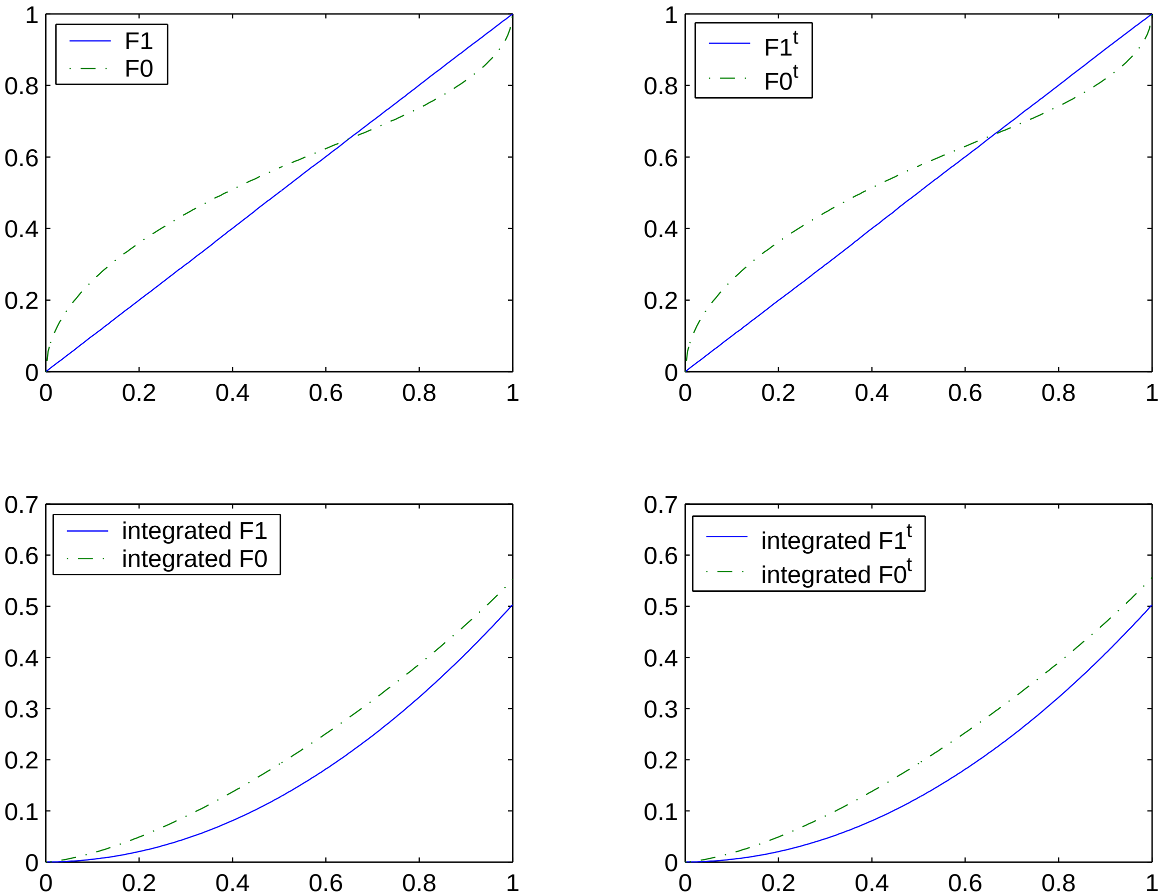

Figure 2: The CDFs and the integrated CDFs of Example 7.2.

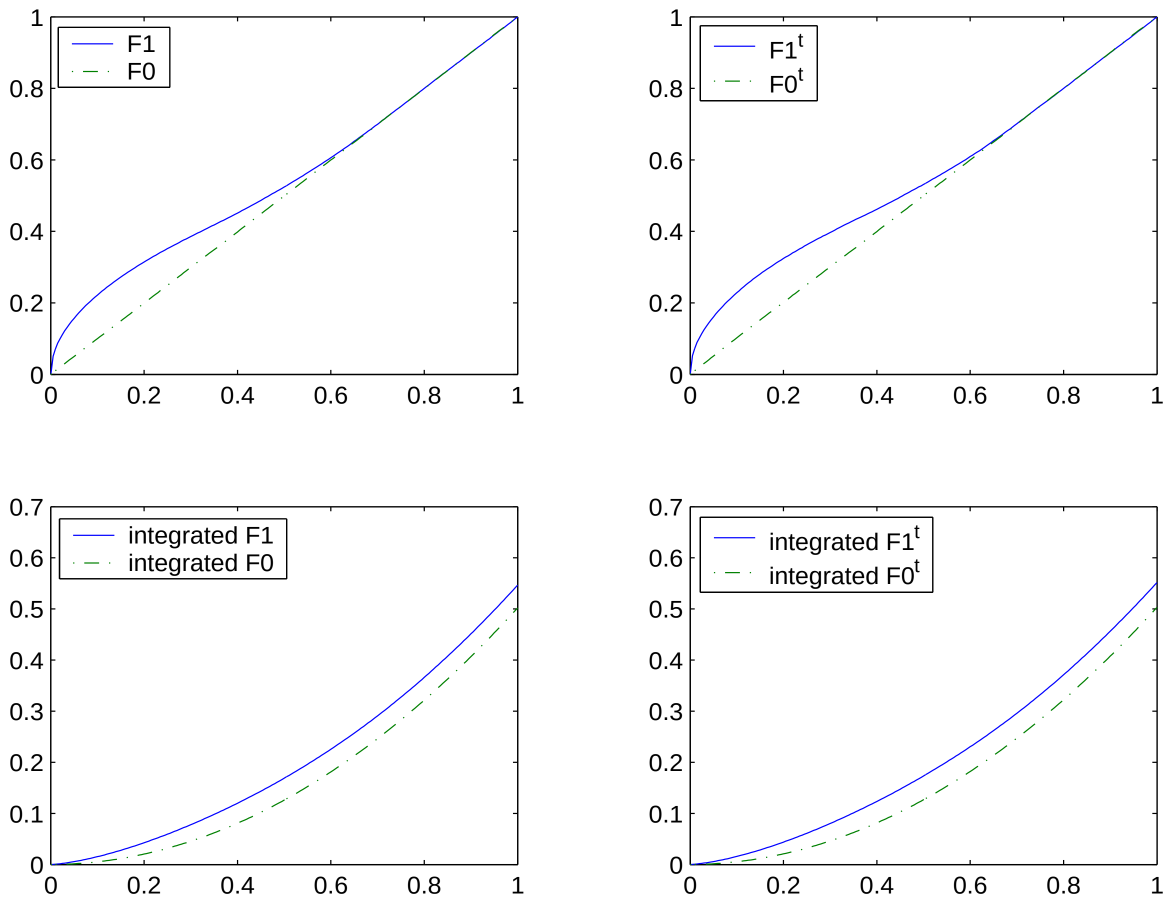

Figure 3: The CDFs and the integrated CDFs of Example 7.3.

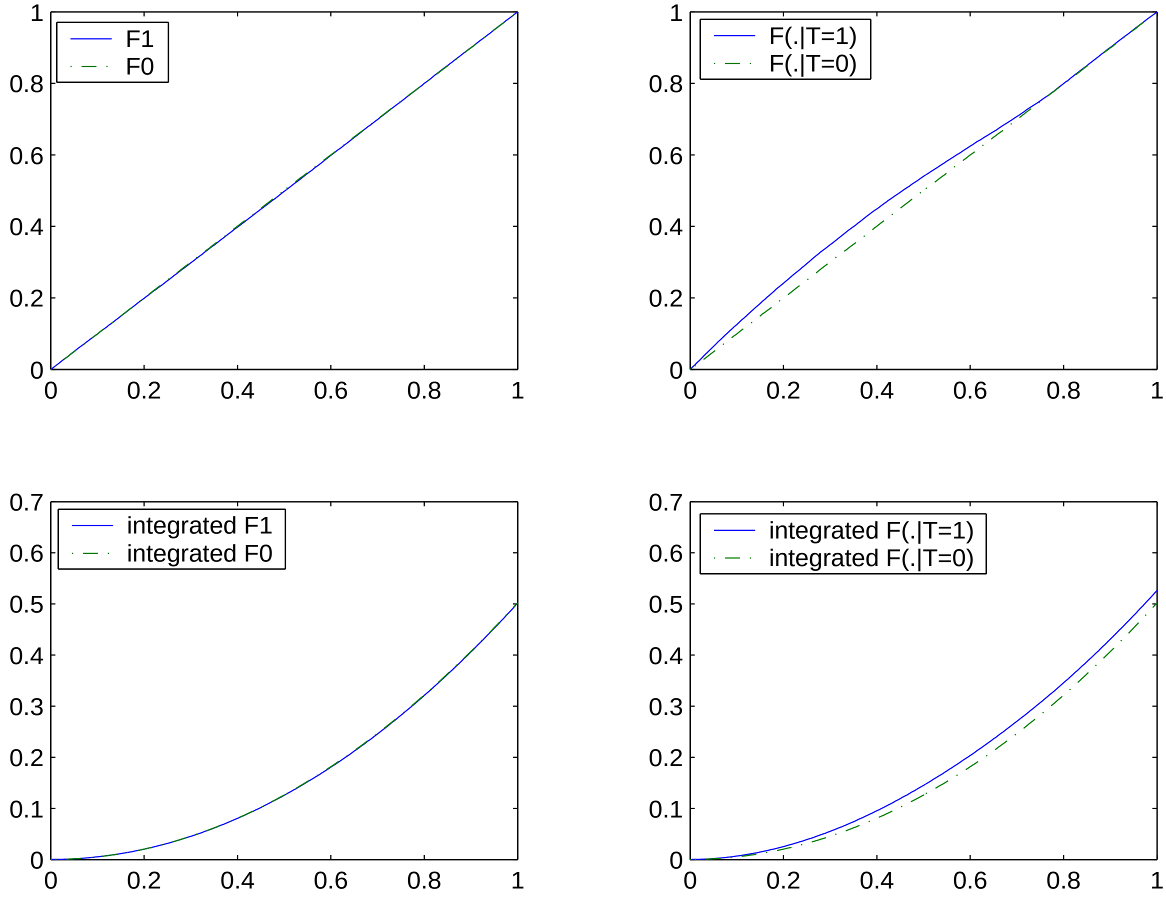

Figure 4: The CDFs and the integrated CDFs of Example 7.4.

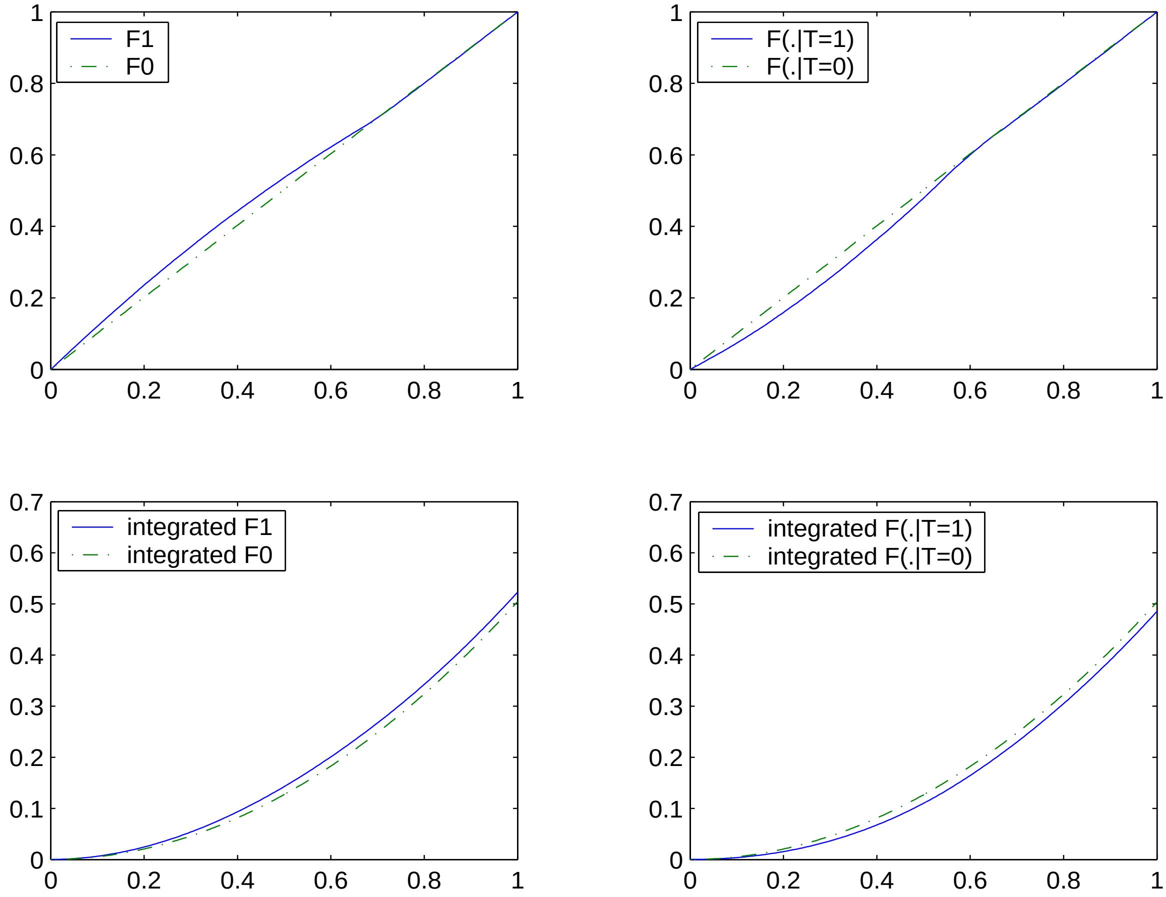

Figure 5: The CDFs and the integrated CDFs of Example 7.5.

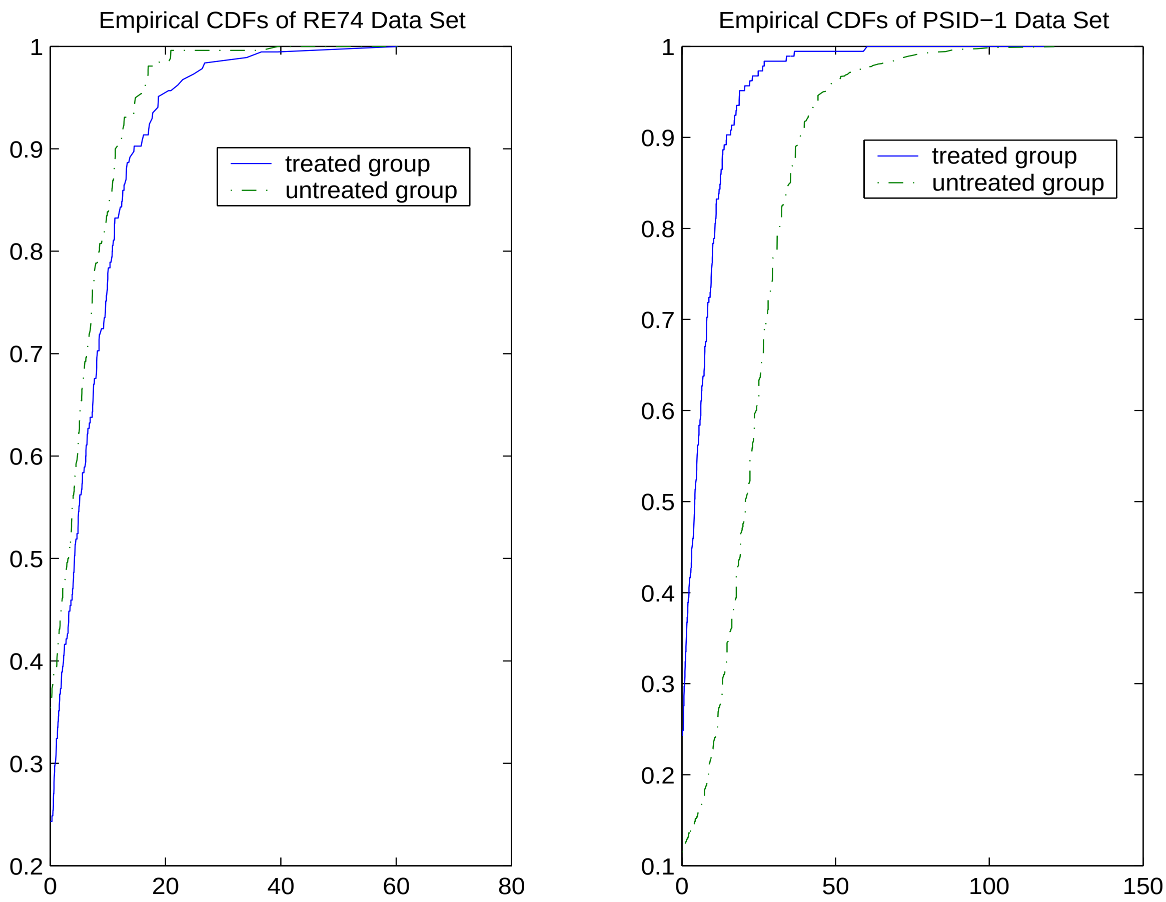

Figure 6: Empirical CDFs of RE74 and PSID-1 data sets.

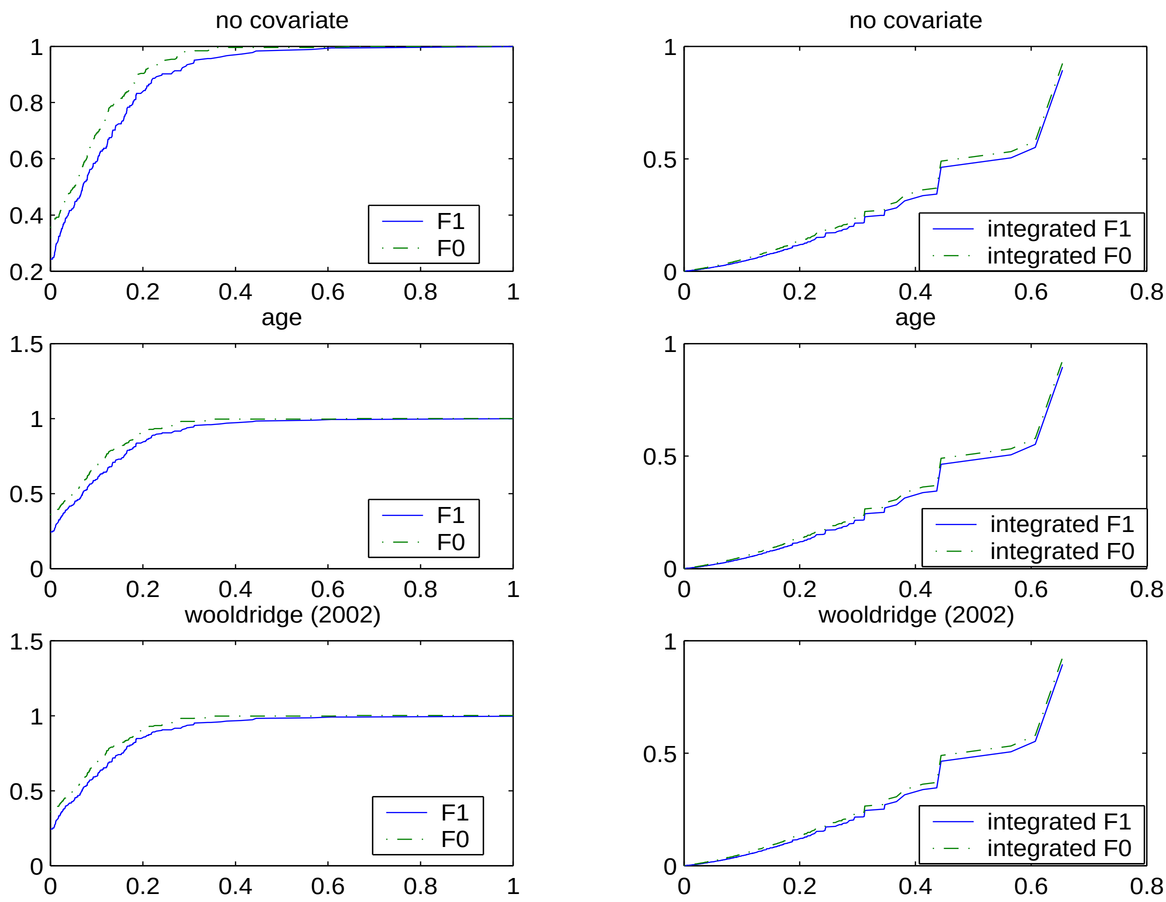

Figure 7: Estimated CDFs and Integrated CDFs for RE74 under Different Specifications.

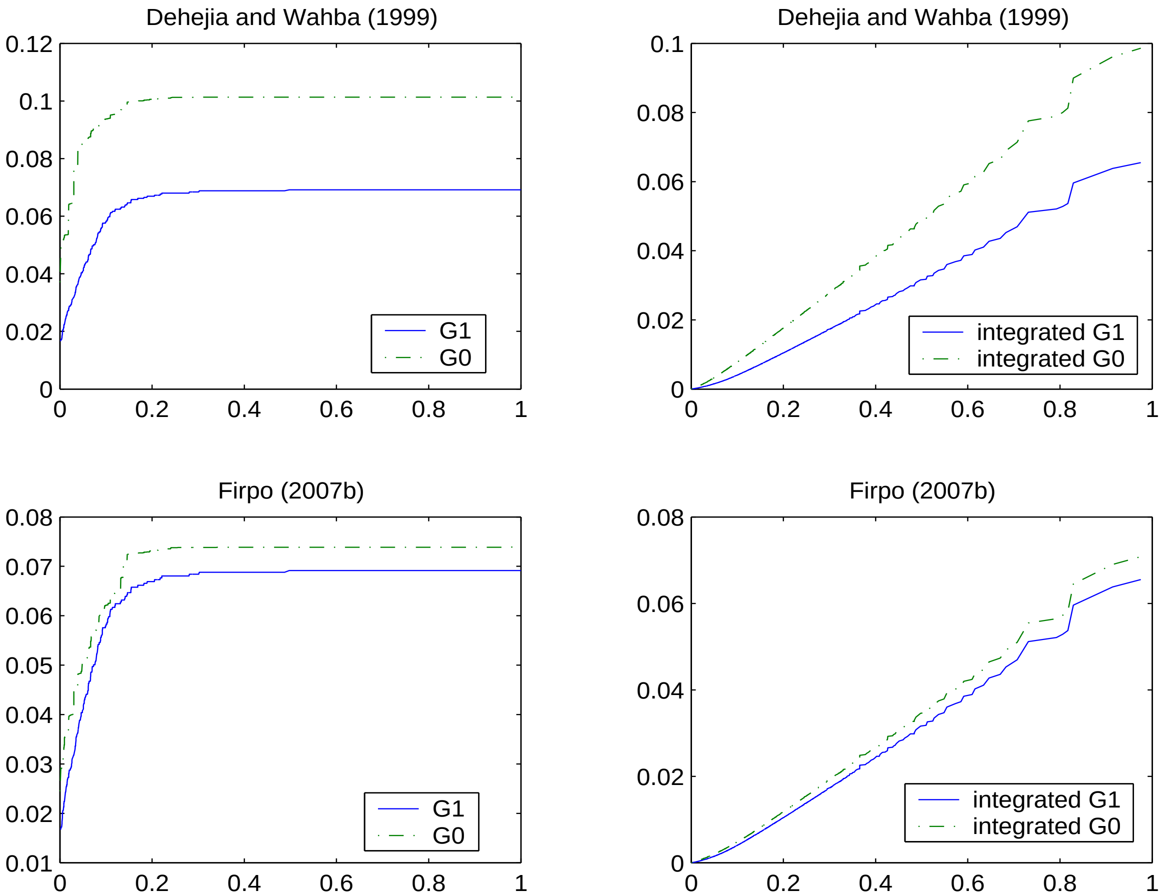

Figure 8: Estimated G_i's and Integrated G_i's for PSID-1 under Different Specifications.