# Nonparametric Tests of Conditional Treatment Effects

Sokbae Lee and Yoon-Jae Whang

University College London and Seoul National University

April 30, 2009

*Universidad Carlos III de Madrid*

## Outline

# 1. Introduction

## Motivation

- **Treatment Effects:** The causal effect of a binary (0 - 1) variable on an outcome of scientific or policy interest. E.g., Gov't subsidizing training program for disadvantaged workers, New drug,...
- **Treatment Heterogeneity**: Treatment effects vary across different subgroups and individuals. The usual average treatment effects for the entire population or for the treated may not provide a full picture of treatment effects.
- There has been relatively little research on testing for the presence of treatment effects that allows for the individual heterogeneity.
- **This paper**: *We provide a general nonparametric test for the treatment effects conditional on covariates. The hypotheses can be either one-sided or two-sided. We allow both conditional average (CATE) and conditional distributional treatment effects (CDTE)*.

## Basic Concepts

- Let

$$\begin{aligned} Y_i(0) &: \text{ potential outcome for individual } i \text{ without treatment} \\ Y_i(1) &: \text{ potential outcome for individual } i \text{ with treatment} \\ Z_i &: \text{ treatment intake indicator } (0 \text{ or } 1) \\ X_i &: \text{ vector of characteristics of individual } i \text{ (covariates)} \\ Y_i &= Y_i(1)Z_i + Y_i(0)(1 - Z_i) : \text{ realized outcome} \end{aligned}$$

$\Rightarrow$ For each $i$, we observe $(Y_i, X_i, Z_i)$.

- Under the unconfoundedness assumption $(Y(0), Y(1) \perp Z \mid X)$, we have

$$\begin{aligned} CATE(x) &\equiv E\left[Y(1) - Y(0) \mid X = x\right] \\ &= E\left[Y(1) \mid X = x, Z = 1\right] - E\left[Y(0) \mid X = x, Z = 0\right] \\ &= E\left[Y \mid X = x, Z = 1\right] - E\left[Y \mid X = x, Z = 0\right], \\ ATE &\equiv E\left[Y(1) - Y(0)\right] = E\left[CATE(X)\right]. \end{aligned}$$

## Hypotheses of Interest

- Let $(Y, X, Z) \in \mathcal{Y} \times \mathcal{X} \times \{0, 1\}$ be observed rv's, where $Z$ is a binary rv. and $X$ is a continuous rv. Let $\mathcal{W} = \mathcal{W}_y \times \mathcal{W}_x \subseteq \mathcal{Y} \times \mathcal{X}$.
- **One-sided Hypotheses**:

$$\begin{aligned} H_0 \quad &: \quad E[G(Y, y)|X = x, Z = 1] \leq E[G(Y, y)|X = x, Z = 0] \\ &\quad\ \text{for each } (y, x) \in \mathcal{W} \quad \textit{vs.} \end{aligned}$$

$$\begin{aligned} H_1 \quad &: \quad E[G(Y, y)|X = x, Z = 1] > E[G(Y, y)|X = x, Z = 0] \\ &\quad\ \text{for some } (y, x) \in \mathcal{W}. \end{aligned}$$

- Examples:
  - $G(Y, y) = -Y$ : testing *positive conditional average treatment effect* for each $x \in \mathcal{W}_x$. ( positive CATE)
  - $G(Y, y) = 1(Y \leq y)$: testing *conditional stochastic dominance* between treatment and control groups for each $x \in \mathcal{W}_x$. (positive CDTE)
  - $G(Y, y) = (y - Y)^{s-1} 1(Y \leq y)$: testing *higher-order* ($s$-th order) *conditional stochastic dominance*.

## Hypotheses of Interest (cont.)

- **Two-sided Hypotheses**:

$$
\begin{aligned}
H_0 \quad &: \quad E[G(Y,y)|X=x,Z=1] = E[G(Y,y)|X=x,Z=0] \\
&\quad \text{for each } (y,x) \in \mathcal{W} \qquad \textit{vs.}
\end{aligned}
$$

$$
\begin{aligned}
H_1 \quad &: \quad E[G(Y,y)|X=x,Z=1] \neq E[G(Y,y)|X=x,Z=0] \\
&\quad \text{for some } (y,x) \in \mathcal{W}.
\end{aligned}
$$

- Examples:
  - $G(Y,y) = -Y$ : testing *no conditional average treatment effect* for each $x \in \mathcal{W}_x$. (no CATE)
  - $G(Y,y) = 1(Y \leq y)$ : testing *equal conditional distributional effect* between treatment and control groups for each $x \in \mathcal{W}_x$. (no CDTE)

## Literature

- **Econometric Evaluation of Social Programs**: Huge. For a survey, see Abbring and Heckman (2007), Blundell and Costa Dias (2008), Heckman and Vytlacil (2007a, 2007b), Imbens (2004), Imbens and Wooldridge (2009), ...
- **Testing Stochastic Dominance:** Anderson (1996), Davidson and Duclos (1997, 2000), McFadden (1989), Klecan, McFadden, and McFadden (1991), Barrett and Donald (2003), Linton, Massoumi and Whang (2005), ...
- **Significance Testing and Conditional Independence Testing:** Fan and Li (1996), Linton and Gozalo (1997), Delgado and Gonzáz Manteiga (2001), Lavergne (2001), Angrist and Kuersteiner (2004, 2008), Su and White (2004, 2007, 2008), Song (2007), ...
- **Testing Treatment Effects**:
  - Crump, Hotz, Imbens, Mitnik (2008): Two-sided CATE.
  - Abadie (2002): Unconditional DTE

Literature

- **Giné, Mason and Zaitsev (2003)**: $L_1$-norm density estimator process. If $h \to 0$ and $n^{1/2}h \to \infty$, then

$$\sqrt{n}\left[\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|\, dx - E\int_{\mathbb{R}} |f_n(x) - Ef_n(x)|\, dx\right] \to N(0, \sigma^2),$$

  where

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

## Main Contributions of This Paper

- We develop a general class of nonparametric tests of *conditional* treatment effects.

| | *Null of Positive TE* | *Null of No TE* |
|---|---|---|
| *CATE* | ? | Crump et. al.(2008)<br>Tests for equality b/ regression functions |
| *CDTE* | ? | Tests for conditional independence |

## Main Contributions of This Paper

- We develop a general class of nonparametric tests of *conditional* treatment effects.

| | *Null of Positive TE* | *Null of No TE* |
|---|---|---|
| *CATE* | ? | Crump et. al.(2008)<br>Tests for equality b/ regression functions |
| *CDTE* | ? | Tests for conditional independence |

- We show that suitably studentized versions of our test statistics are asymptotically *standard normal* under the null hypotheses.

## Main Contributions of This Paper

- We develop a general class of nonparametric tests of *conditional* treatment effects.

| | *Null of Positive TE* | *Null of No TE* |
|---|---|---|
| *CATE* | ? | Crump et. al.(2008)<br>Tests for equality b/ regression functions |
| *CDTE* | ? | Tests for conditional independence |

- We show that suitably studentized versions of our test statistics are asymptotically *standard normal* under the null hypotheses.
- We show that our tests are *consistent* against general fixed alternatives and powerful against *some* $n^{-1/2}$ *local alternatives*.

## Main Contributions of This Paper

- We develop a general class of nonparametric tests of *conditional* treatment effects.

| | *Null of Positive TE* | *Null of No TE* |
|---|---|---|
| *CATE* | ? | Crump et. al.(2008)<br>Tests for equality b/ regression functions |
| *CDTE* | ? | Tests for conditional independence |

- We show that suitably studentized versions of our test statistics are asymptotically *standard normal* under the null hypotheses.
- We show that our tests are *consistent* against general fixed alternatives and powerful against *some* $n^{-1/2}$ *local alternatives*.
- In the case of the one-sided hypothesis, we propose *a more powerful test* by estimating the "contact set" on which the inequality restriction is binding.

# 2. Test Statistics and Assumptions

# Test Statistics

- Define

$$\tau_0(y,x) = E[G(Y,y)|X=x, Z=1] - E[G(Y,y)|X=x, Z=0].$$

- **The null hypotheses of interest**:

$$\begin{aligned} H_0 &: \tau_0(y,x) \le 0 \text{ for each } (y,x) \in \mathcal{W}, \\ H_0^D &: \tau_0(y,x) = 0 \text{ for each } (y,x) \in \mathcal{W}. \end{aligned}$$

- **Test Statistics**:

$$\begin{aligned} \hat{T} &= \int\int \sqrt{n} \max\{\hat{\tau}(y,x), 0\} w(y,x) dy dx, \\ \hat{D} &= \int\int \sqrt{n} \left|\hat{\tau}(y,x)\right| w(y,x) dy dx, \end{aligned}$$

where $w(y,x)$ is a weight function with support $\mathcal{W}$.

## Definitions

- Let

$$
\begin{aligned}
f(x) &: \text{ density of } X \\
p_j(x) &= \Pr(Z=j|X=x)f(x) \text{ for } j=0,1.
\end{aligned}
$$

- **Estimator of** $\tau_0(y,x)$ :

$$
\begin{aligned}
\hat{\tau}(y,x) &= \hat{E}[G(Y,y)|X=x,Z=1]-\hat{E}[G(Y,y)|X=x,Z=0] \\
&= n^{-1}\sum_{i=1}^{n} G(Y_i,y)\hat{\phi}(x,Z_i)K_h(x-X_i),
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{p}_j(x) &= n^{-1}\sum_{i=1}^{n} 1(Z_i=j)K_h(x-X_i), \\
\hat{\phi}(x,z) &= \frac{1(z=1)}{\hat{p}_1(x)}-\frac{1(z=0)}{\hat{p}_0(x)}, \\
K_h(\cdot) &= K(\cdot/h)/h.
\end{aligned}
$$

## Definitions (cont.)

$$K_*(t) = \int K(\xi) K(\xi + t) \, d\xi.$$

$$\mu_1(y, y', x) := \sum_{j \in \{0,1\}} \frac{E[G(Y, y)G(Y, y')|X = x, Z = j]}{p_j(x)},$$

$$\mu_2(y, y', x) := \sum_{j \in \{0,1\}} \frac{E[G(Y, y)|X = x, Z = j] \, E[G(Y, y')|X = x, Z = j]}{p_j(x)}.$$

$$\begin{aligned} \rho_1(y, y', x, t) &= \left\{\mu_1(y, y', x) - \mu_2(y, y', x)\right\} K_*(t), \\ \rho_2(y, x) &= \left\{\mu_1(y, y, x) - \mu_2(y, y, x)\right\} K_*(0), \\ \rho(y, y', x, t) &= \frac{\rho_1(y, y', x, t)}{\sqrt{\rho_2(y, x)\rho_2(y', x)}}. \end{aligned}$$

Assumption 3.1

1. The distribution of $X$ is absolutely continuous with respect to Lebesgue measure and the probability density function $f$ of $X$ is continuously differentiable;
2. The distribution of $Y$ is absolutely continuous with respect to Lebesgue measure;
3. $w(\cdot, \cdot)$ is a continuous function with support $\mathcal{W} = \mathcal{W}_y \times \mathcal{W}_x$, where $\mathcal{W}_y$ is a subset of $\mathcal{Y}$ (possibly, the entire $\mathcal{Y}$) and $\mathcal{W}_x$ is a strict compact subset of $\mathcal{X}$;
4. $p_1(\cdot)$ and $p_0(\cdot)$ are bounded away from zero on $\mathcal{W}_x$;
5. $K$ is a second-order kernel function with support $[-1/2, 1/2]$, symmetric, integrates to 1 and is twice continuously differentiable;

Assumption 3.1

6. As functions of $x$, $E[G(Y,y)|X=x,Z=j]$, $f(x)$, $p_j(x)$ for $j=0,1$ are twice continuously differentiable for each $y$ with uniformly bounded derivatives;
7. $\sup_{(y,x)\in\mathcal{W}} E\left[|G(Y,y)|^3 \,|X=x,Z=j\right] < \infty$ for $j=0,1$;
8. $\{G(\cdot,y) : y \in \mathcal{W}_y\}$ is a VC class of functions with an envelope function $\mathbf{M}$ satisfying $E[\mathbf{M}^2(Y)|X=x] < \infty$;
9. The bandwidth satisfies $h = c_1 n^{-\delta}$ for some positive constant $c_1$ with $1/4 < \delta < 1/3$.

# 3. Asymptotic Null Distributions

## One-Sided Test

- **Result:** If $\tau_0(y,x) = 0 \;\forall\; (y,x) \in \mathcal{W}$ (i.e,. LFC), then

$$\frac{\hat{T} - a_n}{\sigma_0} \xrightarrow{d} N(0,1),$$

where

$$a_n = h^{-1/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{\rho_2(y,x)} w(y,x) dy dx \cdot E \max\{\mathbb{Z}_1, 0\},$$

$$\begin{aligned} \sigma_0^2 \;\; = \;\; & \int_{-1}^{1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Cov \Big[ \max\{ \sqrt{1 - \rho^2(y,y',x,t)} \mathbb{Z}_1 \\ & + \rho(y,y',x,t) \mathbb{Z}_2, 0\}, \max\{\mathbb{Z}_2, 0\} \Big] \\ & \times \sqrt{\rho_2(y,x)\rho_2(y',x)} w(y,x) dy dy' dx dt. \end{aligned}$$

## Two-Sided Test

- **Result**: Under $H_0^D$,

$$\frac{\hat{D} - a_{n,D}}{\sigma_{0,D}} \xrightarrow{d} N(0,1),$$

where

$$a_{n,D} = h^{-1/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{\rho_2(y,x)} w(y,x) dy dx \cdot E\left|\mathbb{Z}_1\right|,$$

$$\begin{aligned} \sigma_{0,D}^2 &= \int_{-1}^{1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Cov\left[\left|\sqrt{1-\rho^2(y,y',x,t)}\mathbb{Z}_1 \right.\right. \\ &\quad \left.\left. +\rho(y,y',x,t)\mathbb{Z}_2\right|, \left|\mathbb{Z}_2\right|\right] \\ &\quad \times \sqrt{\rho_2(y,x)\rho_2(y',x)} w(y,x) dy dy' dx dt. \end{aligned}$$

Remark

In the case of CATE, in which $G(Y,y) = -Y$, we have

$$\rho(y, y', x, t) = \frac{\int K(\xi) K(\xi + t) d\xi}{\int K^2(\xi) d\xi} \equiv \rho(t),$$

so that the expressions for $a_n$ and $\sigma_0^2$ are simplified to:

$$\begin{aligned} a_n &= h^{-1/2} \int_{-\infty}^{\infty} \sqrt{\rho_2(x)} w(x) dx \cdot \frac{1}{\sqrt{2\pi}}, \\ \sigma_0^2 &= \int_{-1}^{1} Cov \left[ \max\{\sqrt{1-\rho(t)}\mathbb{Z}_1 + \rho(t)\mathbb{Z}_2, 0\}, \max\{\mathbb{Z}_2, 0\} \right] dt \\ &\quad \times \int_{-\infty}^{\infty} \rho_2(x) w(x) dx, \text{ where} \\ \rho_2(x) &= \kappa_2^2 \sum_{j\in\{0,1\}} \frac{E\left[Y^2|X=x, Z=j\right] - \left(E\left[Y|X=x, Z=j\right]\right)^2}{p_j(x)}. \end{aligned}$$

and likewise for the two sided test.

## Feasible Tests

- **Standardized Test Statistics**:

$$\hat{S} = \frac{\hat{T} - \hat{a}_n}{\hat{\sigma}},$$
$$\hat{S}_D = \frac{\hat{D} - \hat{a}_{n,D}}{\hat{\sigma}_D}.$$

- **Decision Rules**:

$$\text{Reject } H_0 \text{ if } \hat{S} > z_{1-\alpha},$$
$$\text{Reject } H_0^D \text{ if } \hat{S}_D > z_{1-\alpha}$$

at the nominal significance level $\alpha$, where $z_\alpha$ is the $\alpha$ quantile of the standard normal distribution for $0 < \alpha < 1$.

## Asymptotic Size Properties

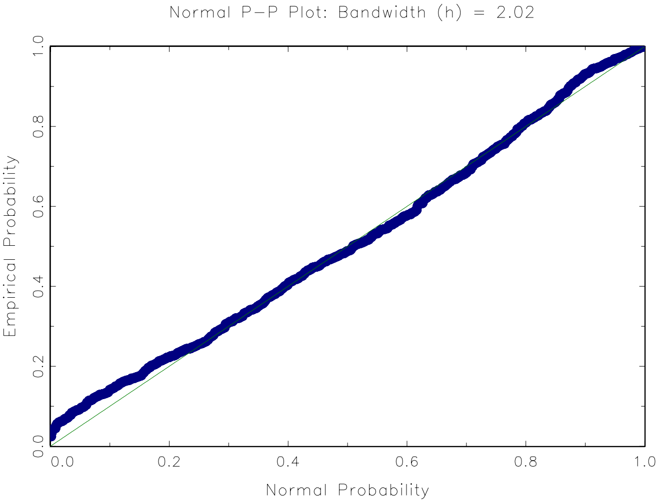
- **Theorem 1 :** *Let Assumption 3.1 hold. Then, (a) under the null hypothesis* $H_0$,

$$\lim_{n\to\infty} \Pr\left(\hat{S} > z_{1-\alpha}\right) \leq \alpha,$$

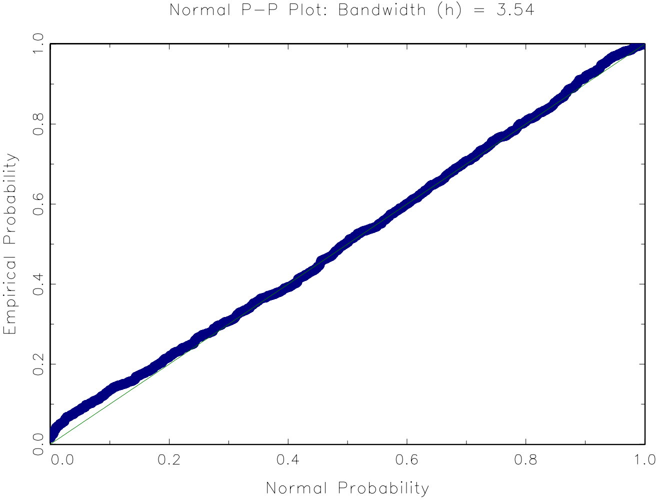
*with equality when* $\tau_0(y,x) = 0$ *for each* $(y,x) \in \mathcal{W}$ *and (b) under the null hypothesis* $H_0^D$,

$$\lim_{n\to\infty} \Pr\left(\hat{S}_D > z_{1-\alpha}\right) = \alpha.$$

# One-Sided Test (Normal P-P Plot)



# Two-Sided Test (Normal P-P Plot)



## Asymptotic Theory: Outline

Although the tests are easy to implement, the asymptotic theory for the tests involves several lengthy steps.

1. The asymptotic approximation of $\hat{T}$ by $T_n$, using the uniform approximation of $\hat{\tau}(y, x)$ by $\tau_n(y, x)$. (cf.) Ghosal, Sen, and van der Vaart (2000).
2. Get the asymptotic distribution of $T_N$, a *Poissonized* version $T_n$. (cf.) Giné, Mason and Zaitsev (2003).
3. D*e-Poissonize* $T_N$ to derive the asymptotic normality of $T_n$ and hence $\hat{T}$.

Asymptotic Theory

- **First-order Approximation:** Under Assumption 3.1,

$$\hat{T} = T_n^* + o_p(1),$$

where

$$T_n^* = \int\int \sqrt{n} \max\{\tau_0(y,x) + [\tau_n(y,x) - E\tau_n(y,x)], 0\} w(y,x) dy dx,$$

$$\begin{aligned}
\tau_n(y,x) &= \frac{1}{n}\sum_{i=1}^{n} \varphi(W_i, y, x) \\
&= \frac{1}{n}\sum_{i=1}^{n} [G(Y_i, y)\phi(x, Z_i) - \chi(Z_i, y, x)] K_h(x - X_i), \\
\chi(z,y,x) &= \chi_1(y,x)1(z=1) - \chi_0(y,x)1(z=0), \\
\chi_j &= \frac{E[G(Y,y)|X=x, Z=j]}{p_j(x)}.
\end{aligned}$$

# 4. Asymptotic Power Properties

Consistency

- **Fixed Alternative Hypotheses**:

$$
\begin{aligned}
H_1 &: \int\int \max\{\tau_0(y,x),0\}\, w(y,x)dydx > 0 \\
H_1^D &: \int\int |\tau_0(y,x)|\, w(y,x)dydx > 0
\end{aligned}
$$

- **Theorem 2 :** *Let Assumption 3.1 hold. Then, (a) under the alternative hypothesis* $H_1$*,*

$$\lim_{n\to\infty} \Pr\left(\hat{S} > z_{1-\alpha}\right) = 1$$

*and (b) under the alternative hypothesis* $H_1^D$*,*

$$\lim_{n\to\infty} \Pr\left(\hat{S}_D > z_{1-\alpha}\right) = 1.$$

- **Local Alternatives**:

$$
\begin{aligned}
H_a &: \tau_0(y,x) = n^{-1/2}\delta(y,x), \\
H_a^D &: \tau_0(y,x) = n^{-1/2}\delta_D(y,x),
\end{aligned}
$$

where $\delta(\cdot,\cdot)$ is non-negative on $\mathcal{W}$ with $\int\int \delta(y,x)w(y,x)dydx > 0$ and $\int\int |\delta_D(y,x)|\, w(y,x)dydx > 0.$

- **Theorem 3:** *Let Assumption 3.1 hold. Then, (a) under the alternative hypothesis $H_a$ ,*

$$\lim_{n\to\infty} \Pr\left(\hat{S} > z_{1-\alpha}\right) > \alpha$$

*and (b) under the alternative hypothesis $H_a^D$,*

$$\lim_{n\to\infty} \Pr\left(\hat{S}_D > z_{1-\alpha}\right) \geq \alpha.$$

# Remarks

- Under $H_a$, we show that

$$\hat{S} = \frac{\hat{T} - \tilde{a}_n}{\sigma_0} + \frac{\tilde{a}_n - a_n}{\sigma_0} + o_p(1),$$

$$\frac{\hat{T} - \tilde{a}_n}{\sigma_0} \xrightarrow{d} N(0,1), \text{ where}$$

$$\tilde{a}_n = \int\int E \max\left\{\delta(y,x) + h^{-1/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\} w(y,x)dydx.$$

## Remarks

- Under $H_a$, we show that

$$\hat{S} = \frac{\hat{T} - \tilde{a}_n}{\sigma_0} + \frac{\tilde{a}_n - a_n}{\sigma_0} + o_p(1),$$

$$\frac{\hat{T} - \tilde{a}_n}{\sigma_0} \xrightarrow{d} N(0,1), \text{ where}$$

$$\tilde{a}_n = \int\int E \max\left\{\delta(y,x) + h^{-1/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\} w(y,x)dydx.$$

- Since $a \geq \max\{a+b,0\} - \max\{b,0\} \geq a1(b \geq 0)$ for $a \geq 0$,

$$\begin{aligned}\tilde{a}_n - a_n &= \int\int E\left[\max\left\{\delta(y,x) + h^{-1/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\}\right. \\ &\quad \left. - \max\left\{h^{-1/2}\sqrt{\rho_2(y,x)}\mathbb{Z}, 0\right\}\right] w(y,x)dydx \\ &\geq \frac{1}{2}\int\int \delta(y,x)w(y,x)dydx > 0.\end{aligned}$$

$\Rightarrow \hat{S}$ has asymptotically non-trivial local power against $H_a$.

# 5. Extension: A More Powerful One-Sided Test

## Approximation under the Inequality Constraint

- Let
$$C = \{(y, x) \in \mathcal{W} : \tau_0(y, x) = 0\}.$$
- **Result (Pointwise)**: Suppose $\int \int_C w(y, x) dy dx > 0$. *Then, under* $H_0 : \tau_0(y, x) \leq 0 \ \forall \ (y, x) \in \mathcal{W}$,
$$\hat{T} = T_n(C) + o_p(1), \text{ where}$$
$$T_n(C) = \int \int_C \sqrt{n} \max\{[\tau_n(y, x) - E\tau_n(y, x)], 0\} w(y, x) dy dx.$$
- Why?: Recall that
$$\hat{T} = T_n^* + o_p(1),$$
$$T_n^* = \int \int \sqrt{n} \max\{\tau_0(y, x) + [\tau_n(y, x) - E\tau_n(y, x)], 0\} w(y, x) dy dx.$$

## Test Statistic

- **Contact Set Estimator**:

$$\hat{C} = \{(y,x) \in \mathcal{W} : |\hat{\tau}(y,x)| \leq \eta_n\}.$$

- **Test Statistic**:

$$\hat{S}_C = \frac{\hat{T} - \hat{a}_n(\hat{C})}{\hat{\sigma}(\hat{C})}.$$

- **Decision Rule**:

$$\text{Reject } H_0 \text{ if } \hat{S}^* > z_{1-\alpha} \ ,$$

where

$$\hat{S}^* = \begin{cases} \hat{S}_C & \text{if } \int\int_{\hat{C}} w(y,x)dydx > 0 \\ \hat{S} & \text{if } \int\int_{\hat{C}} w(y,x)dydx = 0 \end{cases} .$$

Assumption 4.1

1. $C = C_1 \times C_0$, where $C_1 \subset \mathcal{W}_y$ and $C_0 \subset \mathcal{W}_x$ are Borel sets.
2. Whenever the Lebesgue measure $\lambda(C)$ of $C$ is strictly positive, the boundary of $C$ satisfies

$$0 < \lim_{t \to 0^+} \inf \frac{h^*(t)}{t^\gamma} \leq \lim_{t \to 0^+} \sup \frac{h^*(t)}{t^\gamma} = c,$$

   for $\gamma \geq 1$ and some finite constant $c$, where $h^*(t) = \lambda(\{(y,x) : 0 < |\tau_0(y,x)| \leq t\})$.
3. The tuning parameter $\eta_n$ satisfies $\eta_n = c_2 n^{-\alpha}$ for some constant $c_2$ with $2\delta/(1+\gamma) < \alpha < (1-\delta)/2$.

## Size and Global Power

- **Theorem 4.1.** *Suppose that Assumptions 3.1 and 4.1 hold. Then, (a) under the null hypothesis* $H_0$,

$$\lim_{n\to\infty} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) \leq \alpha,$$

*with equality when* $\int\int_C w(y,x)dydx > 0$ *and (b) under the alternative hypothesis* $H_1$,

$$\lim_{n\to\infty} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) = 1.$$

# Local Power

- **Local Alternatives:**

$$H_a^* : \tau_0(y,x) = \mu(y,x) + n^{-1/2}\delta(y,x),$$

where $\mu \leq 0$ on $\mathcal{W}$, $\mu = 0$ on $C_a$ and $\int\int_{C_a} \delta(y,x)w(y,x)dydx > 0$.

- **Theorem 4.2**: *Suppose that Assumptions 3.1, 4.1 (iii) and 4.2 hold. Then, under the alternative hypothesis $H_a^*$, we have (a)*

$$\lim_{n\to\infty} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) > \alpha$$

*and (b)*

$$\lim_{n\to\infty} \Pr\left(\hat{S}^* > z_{1-\alpha}\right) > \lim_{n\to\infty} \Pr\left(\hat{S} > z_{1-\alpha}\right),$$

*whenever*

$$\int\int_{\mathbb{R}^2} w(y,x)dydx > \int\int_{C_a} w(y,x)dydx.$$

## Assumption 4.2

1. $C_a = C_{1a} \times C_{0a}$, where $C_{1a} \subset \mathcal{W}_y$ and $C_{0a} \subset \mathcal{W}_x$ are Borel sets.
2. $\int\int_{C_a} w(y,x)dydx > 0$, where $C_a = \{(y,x) \in \mathcal{W} : \mu(y,x) = 0\}$.
3. $\sup_{(y,x)\in\mathcal{W}} \mu(y,x) \leq 0$ .
4. $\delta(\cdot,\cdot)$ is a non-negative function on $\mathcal{W}$ with $\int\int_{C_a} \delta(y,x)w(y,x)dydx > 0$ and $\sup_{(y,x)\in\mathcal{W}} \delta(y,x) < \infty$.
5. The boundary of $C_a$ satisfies

$$0 < \lim_{t\to 0^+} \inf \frac{h^{**}(t)}{t^\gamma} \leq \lim_{t\to 0^+} \sup \frac{h^{**}(t)}{t^\gamma} = c,$$

for $\gamma \geq 1$ and some finite constant $c$, where $h^{**}(t) = \lambda(\{(y,x) : 0 < |\mu(y,x)| \leq t\})$.

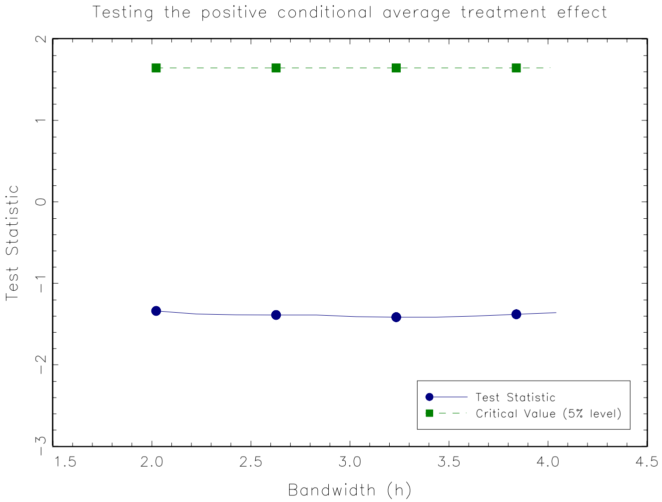
# 5. An Empirical Example and Monte Carlo Experiments

## Setup

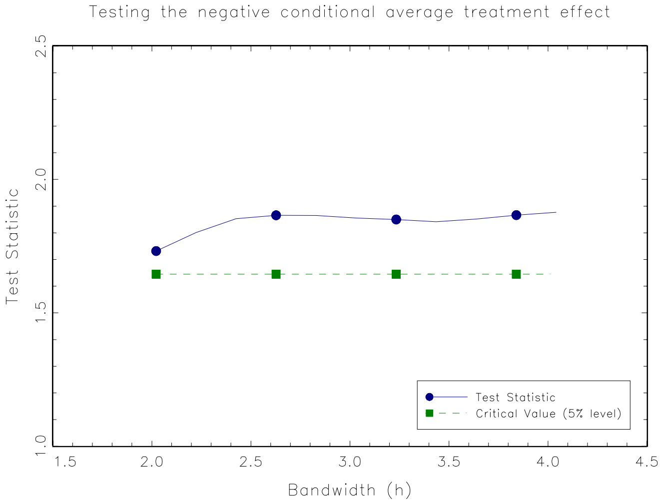
- **Data**: The experimental data from *National Supported Work (NSW) Demonstration* program.
  - *NSW*: A randomized, temporary employment program in the U.S. in the mid-1970s designed to help disadvantaged workers.
  - *Empirical Analysis*: LaLonde (1986), Dehejia and Wahba (1999, 2002) and Smith and Todd (2005), etc.
  - *Observations*: Treatment group (297) + Control group (425) = Total (722)
- **Variables**:
  - $Y$ : `RE78` (earnings in 1978) or `RE78-RE75` (changes in earnings between 1978 (post-intervention year) and 1975 (pre-intervention year))
  - $X$ : Age
  - $Z = 1$ (Treatment group), $Z = 0$ (Control group)

# Empirical Results (RE78)

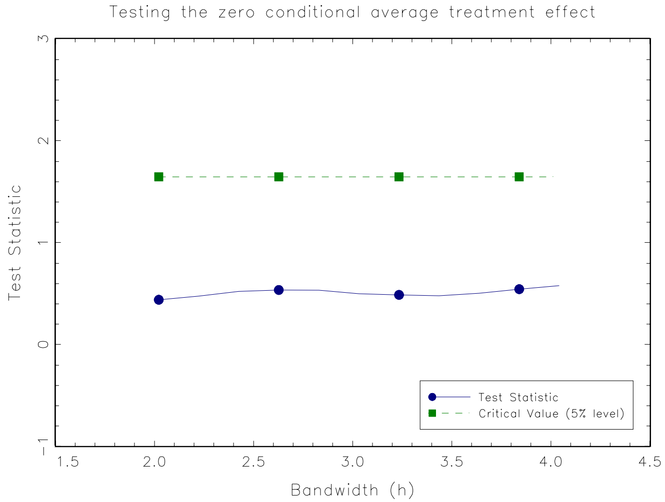
# Empirical Results (RE78)



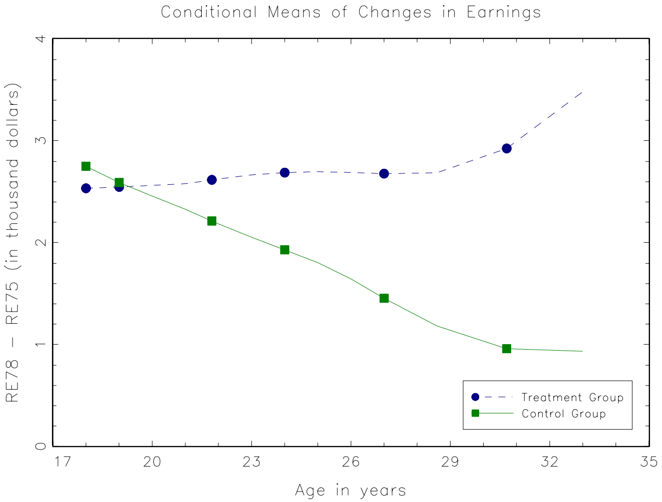
# Empirical Results (RE78)



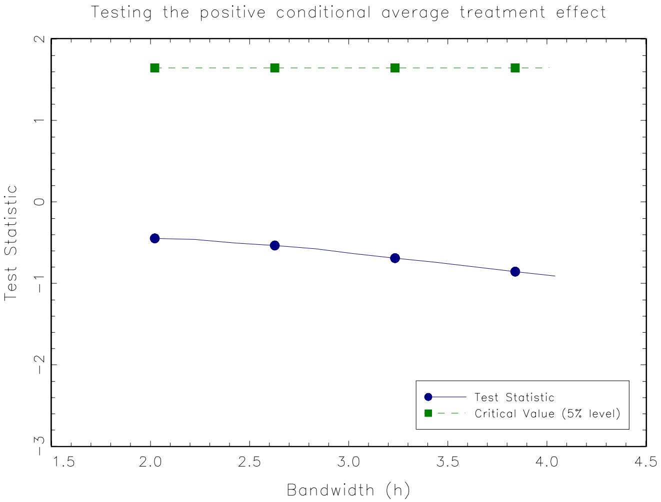
# Empirical Results (RE78)



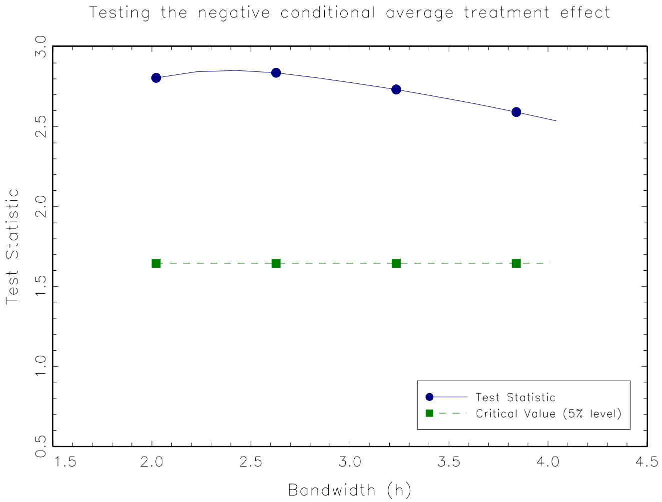
# Empirical Results (RE78-RE75)



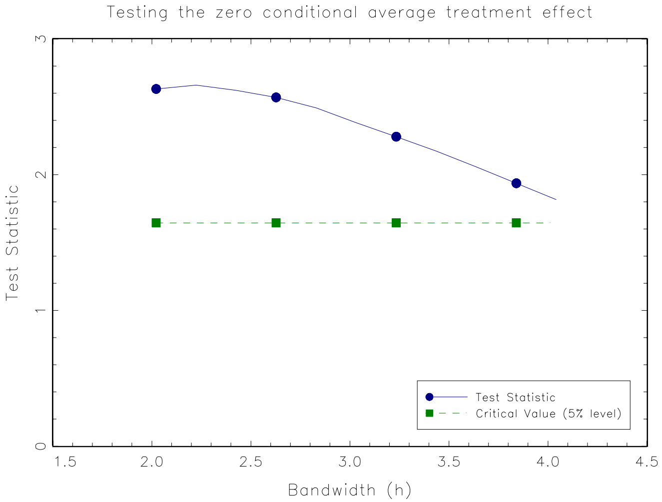
# Empirical Results (RE78-RE75)



# Empirical Results (RE78-RE75)



# Empirical Results (RE78-RE75)

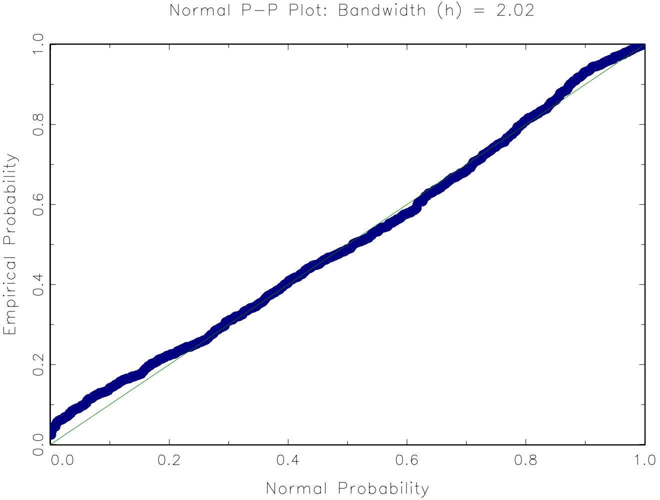


## Monte Carlo Experiments

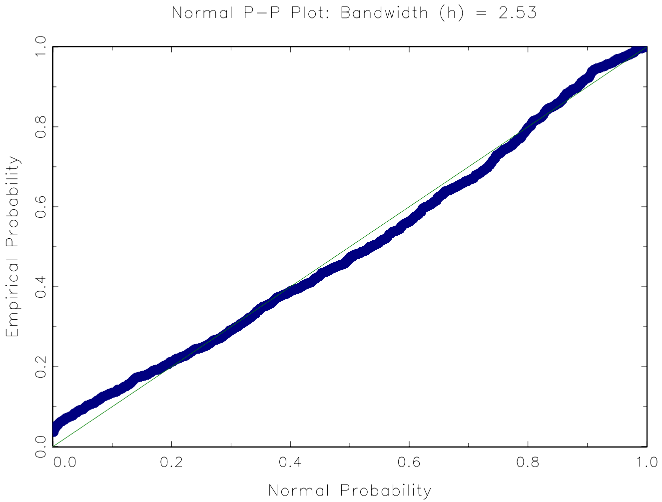
- We use the NSW data to get simulation data, in order to evaluate the finite sample performance of our tests in more practical situations.
- **DGP1**: 1,000 repeated samples are generated randomly with replacement from the NSW data, with the restriction that $(Y, X)$ and $Z$ are generated independently. $\Rightarrow$ LFC
- **DGP2**: 1,000 repeated samples are generated randomly with replacement from the NSW data, with the joint distribution of $(Y, X, Z)$ being left intact. $\Rightarrow$ Interior of the Null & Power
- Other Choices:

$$\begin{aligned} K(u) &= \frac{3}{2}\left(1-(2u)^2\right) 1\left\{|u| \leq \frac{1}{2}\right\} \\ h &= C_h \hat{s}_X n^{-2/7} \\ n &= 722 \text{ or } 1,444. \end{aligned}$$

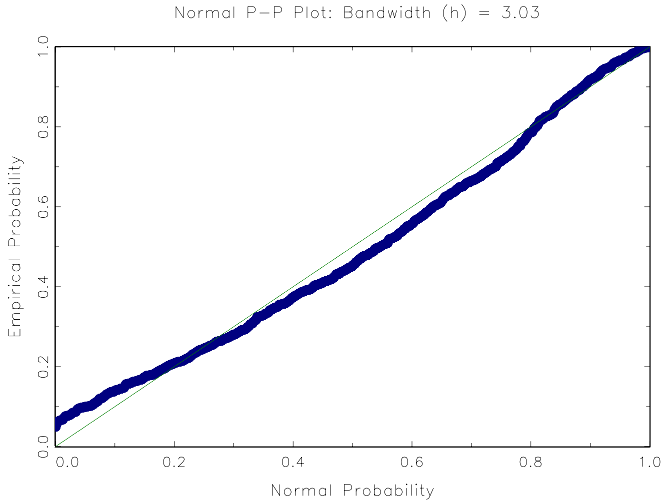
# One-Sided Test



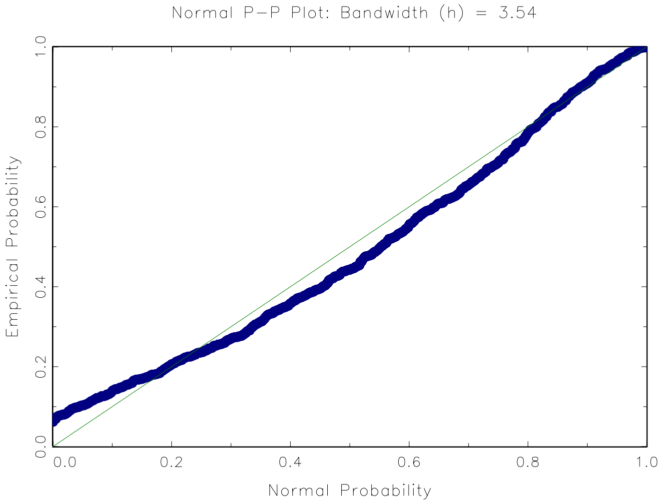
# One-Sided Test



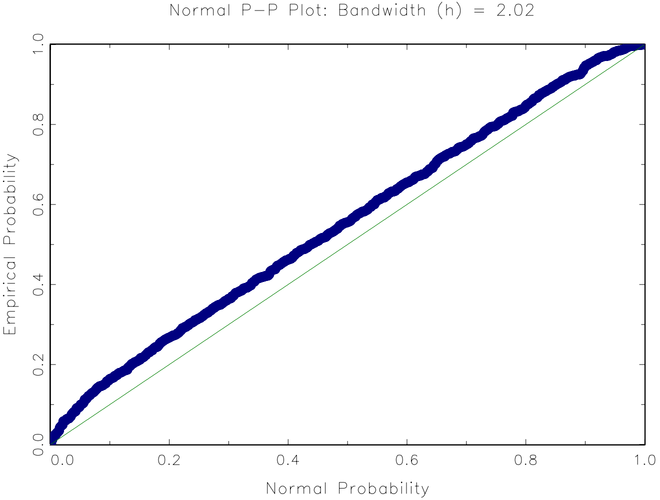
# One-Sided Test



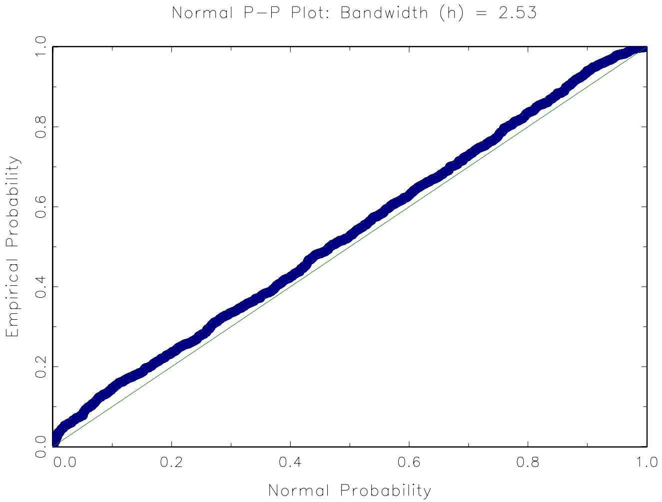
# One-Sided Test



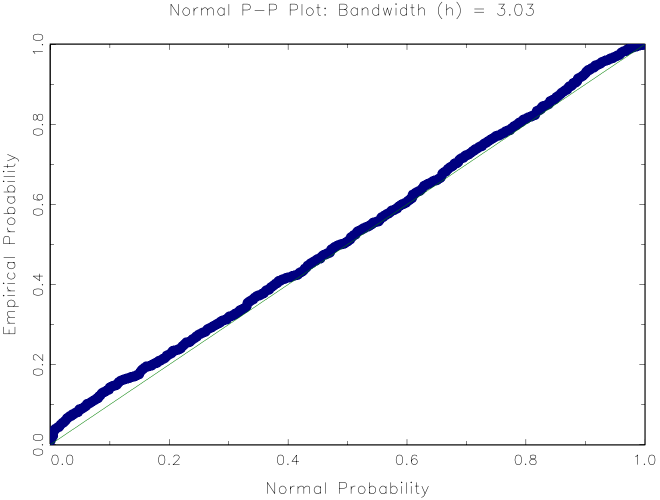
# Two-Sided Test



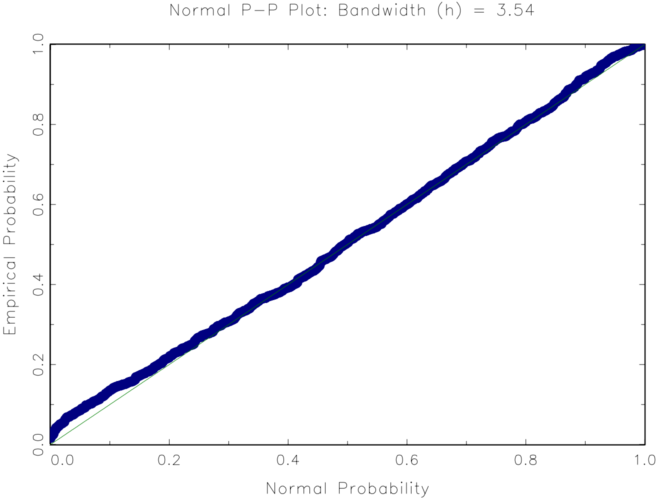
# Two-Sided Test



# Two-Sided Test



# Two-Sided Test



Rejection Probabilities (Our Test, One-Sided)

| Sample Size $(n)$ | Bandwidth $(h)$ | DGP1: $H_0$ is true (least favorable case) Nominal Probabilities 0.10 | 0.05 | 0.01 | DGP2: Mimicking the NSW data Nominal Probabilities 0.10 | 0.05 | 0.01 |
|---|---|---|---|---|---|---|---|
| $H_0$: the CATE is positive for each $x$ (One-Sided Test) | | | | | | | |
| 722 | 2.021 | 0.127 | 0.075 | 0.031 | 0.012 | 0.006 | 0.001 |
| | 2.526 | 0.119 | 0.075 | 0.030 | 0.009 | 0.004 | 0.001 |
| | 3.031 | 0.120 | 0.071 | 0.030 | 0.005 | 0.002 | 0.001 |
| | 3.537 | 0.117 | 0.071 | 0.027 | 0.003 | 0.002 | 0.001 |
| 1444 | 1.658 | 0.134 | 0.074 | 0.025 | 0.032 | 0.010 | 0.002 |
| | 2.072 | 0.119 | 0.066 | 0.024 | 0.005 | 0.002 | 0.002 |
| | 2.487 | 0.122 | 0.067 | 0.022 | 0.003 | 0.002 | 0.001 |
| | 2.901 | 0.126 | 0.064 | 0.021 | 0.003 | 0.001 | 0.000 |

Rejection Probabilities (Our Test, One-Sided)

| Sample Size ($n$) | Bandwidth ($h$) | DGP1: $H_0$ is true (least favorable case) Nominal Probabilities 0.10 | 0.05 | 0.01 | DGP2: Mimicking the NSW data Nominal Probabilities 0.10 | 0.05 | 0.01 |
|---|---|---|---|---|---|---|---|
| $H_0$: the CATE is negative for each $x$ (One-Sided Test) | | | | | | | |
| 722 | 2.021 | 0.126 | 0.073 | 0.019 | 0.861 | 0.767 | 0.559 |
| | 2.526 | 0.115 | 0.076 | 0.020 | 0.834 | 0.746 | 0.535 |
| | 3.031 | 0.112 | 0.066 | 0.019 | 0.803 | 0.719 | 0.509 |
| | 3.537 | 0.112 | 0.064 | 0.019 | 0.780 | 0.684 | 0.477 |
| 1444 | 1.658 | 0.110 | 0.068 | 0.021 | 0.983 | 0.976 | 0.914 |
| | 2.072 | 0.098 | 0.061 | 0.018 | 0.993 | 0.984 | 0.910 |
| | 2.487 | 0.102 | 0.065 | 0.020 | 0.987 | 0.970 | 0.892 |
| | 2.901 | 0.100 | 0.065 | 0.022 | 0.981 | 0.960 | 0.859 |

Rejection Probabilities (Our Test, Two-Sided)

| Sample Size ($n$) | Bandwidth ($h$) | DGP1: $H_0$ is true (least favorable case) Nominal Probabilities 0.10 | 0.05 | 0.01 | DGP2: Mimicking the NSW data Nominal Probabilities 0.10 | 0.05 | 0.01 |
|---|---|---|---|---|---|---|---|
| $H_0$: the CATE is zero for each $x$ (Two-Sided Test) | | | | | | | |
| 722 | 2.021 | 0.149 | 0.095 | 0.033 | 0.801 | 0.693 | 0.460 |
| | 2.526 | 0.134 | 0.087 | 0.029 | 0.703 | 0.581 | 0.365 |
| | 3.031 | 0.121 | 0.077 | 0.024 | 0.613 | 0.497 | 0.306 |
| | 3.537 | 0.114 | 0.071 | 0.018 | 0.547 | 0.447 | 0.261 |
| 1444 | 1.658 | 0.154 | 0.077 | 0.028 | 0.980 | 0.977 | 0.944 |
| | 2.072 | 0.123 | 0.071 | 0.020 | 0.990 | 0.968 | 0.877 |
| | 2.487 | 0.116 | 0.067 | 0.021 | 0.969 | 0.940 | 0.808 |
| | 2.901 | 0.115 | 0.064 | 0.019 | 0.946 | 0.888 | 0.722 |

## Rejection Probabilities (Crump et. al., Two-Sided # 1)

| Sample Size ($n$) | Order of Power Series ($K-1$) | DGP1: $H_0$ is true (least favorable case) Nominal Probabilities 0.10 | 0.05 | 0.01 | DGP2: Mimicking the NSW data Nominal Probabilities 0.10 | 0.05 | 0.01 |
|---|---|---|---|---|---|---|---|
| $H_0$: the CATE is zero for each $x$ (Test Statistic $T$) | | | | | | | |
| 722 | 1 | 0.099 | 0.072 | 0.028 | 0.512 | 0.432 | 0.293 |
| | 2 | 0.108 | 0.068 | 0.022 | 0.452 | 0.368 | 0.236 |
| | 3 | 0.064 | 0.045 | 0.023 | 0.163 | 0.127 | 0.069 |
| | 4 | 0.063 | 0.039 | 0.018 | 0.136 | 0.096 | 0.056 |
| | 5 | 0.065 | 0.040 | 0.020 | 0.459 | 0.403 | 0.301 |
| 1444 | 1 | 0.114 | 0.080 | 0.042 | 0.798 | 0.741 | 0.626 |
| | 2 | 0.114 | 0.075 | 0.040 | 0.740 | 0.653 | 0.516 |
| | 3 | 0.071 | 0.051 | 0.026 | 0.218 | 0.173 | 0.100 |
| | 4 | 0.049 | 0.032 | 0.016 | 0.169 | 0.120 | 0.051 |
| | 5 | 0.051 | 0.028 | 0.014 | 0.725 | 0.671 | 0.573 |

## Rejection Probabilities (Crump et. al., Two-Sided #2)

| Sample Size ($n$) | Order of Power Series ($K-1$) | DGP1: $H_0$ is true (least favorable case) Nominal Probabilities 0.10 | 0.05 | 0.01 | DGP2: Mimicking the NSW data Nominal Probabilities 0.10 | 0.05 | 0.01 |
|---|---|---|---|---|---|---|---|
| $H_0$: the CATE is zero for each $x$ (Test Statistic $Q$) | | | | | | | |
| 722 | 1 | 0.098 | 0.042 | 0.004 | 0.504 | 0.362 | 0.140 |
| | 2 | 0.100 | 0.045 | 0.007 | 0.443 | 0.295 | 0.113 |
| | 3 | 0.059 | 0.034 | 0.013 | 0.153 | 0.097 | 0.033 |
| | 4 | 0.059 | 0.030 | 0.010 | 0.131 | 0.074 | 0.029 |
| | 5 | 0.061 | 0.031 | 0.011 | 0.453 | 0.355 | 0.231 |
| 1444 | 1 | 0.109 | 0.059 | 0.013 | 0.794 | 0.686 | 0.409 |
| | 2 | 0.105 | 0.055 | 0.012 | 0.725 | 0.587 | 0.336 |
| | 3 | 0.065 | 0.036 | 0.010 | 0.211 | 0.138 | 0.049 |
| | 4 | 0.045 | 0.021 | 0.004 | 0.166 | 0.086 | 0.024 |
| | 5 | 0.046 | 0.020 | 0.004 | 0.714 | 0.626 | 0.462 |

# 7. Conclusions

## Conclusions

- We develop a general nonparametric test for the treatment effects conditional on covariates, explicitly allowing for the individual heterogeneity.
- The hypotheses can be either one-sided (positive effect) or two-sided (no effect) and we allow both *conditional average (CATE)* or *conditional distributional treatment effects (CDTE)*.
- The test is easy to implement, using the *standard normal* critical values. The test is consistent and powerful against some $n^{-1/2}$ local alternatives. Under certain contexts, we may improve the power of the one-sided test by estimating the "contact set".
- The poissonization technique seems to be widely applicable to many contexts in econometrics. E.g., Anderson, Linton and Whang (2009), "Nonparametric Estimation of a Polarization Measure" : $\theta = \int \min \{f(x), g(x)\} dx$.
- Future works: Uniformity issues, Bootstrap critical values, Bandwidth choices, More simulations and applications,...
- In near future, the paper will be posted as Cemmap working paper at `http://www.cemmap.ac.uk/`.