Identification and Estimation of Average Marginal Treatment Effects with a Bunching Design

Carolina Caetano, Gregorio Caetano, Eric Nielsen

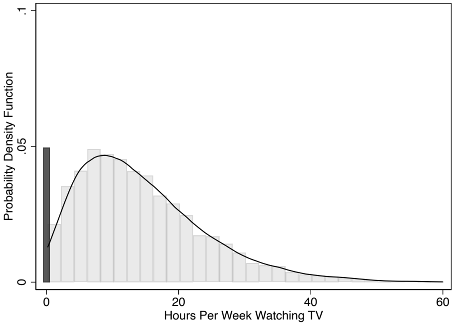
University of Georgia, The Federal Reserve Board of Governors

This paper

- We show that a causal effect can be nonparametrically identified using bunching.
- Causal effect: average marginal treatment effect among the marginal observations at the bunching point.
- Identification does NOT require:
  - Exclusion restrictions
  - Functional forms
  - Distributional assumptions
  - Special data structures
- Identification requires:
  - Continuity conditions
  - Local independence and monotonicity conditions

Bunching

Bunching: Concentration of observations at a point of an otherwise locally continuous distribution.



Hours per week children watch TV on CDS-PSID

Treatment variables with corner bunching

More common than you think!

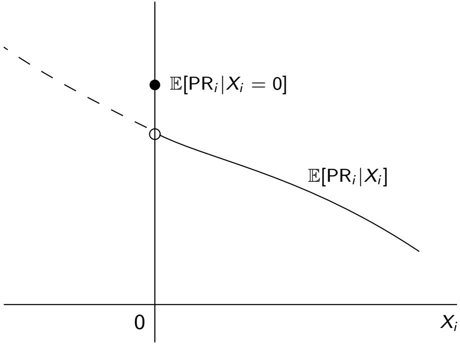
- Usually found when treatment is a choice: often there is bunching at zero.
  - Consumption goods: e.g. vitamin supplements, cigarettes, alcohol, and coffee.
  - Financial variables: e.g. credit card debt, credit access, bequests, and expenditure on ads.
  - Time use: e.g. exercising, working, doing homework, volunteering, and using social media.
- Artificially created restrictions: minimum schooling, minimum wage, minimum age to work, and minimum 401K contributions.
- Generally applicable whenever Caetano (2015)'s test can be applied at the boundary. Several examples in a variety of settings in economics, political science, and finance.

Main insight (1)

Distribution of confounders is discontinuous at the bunching point

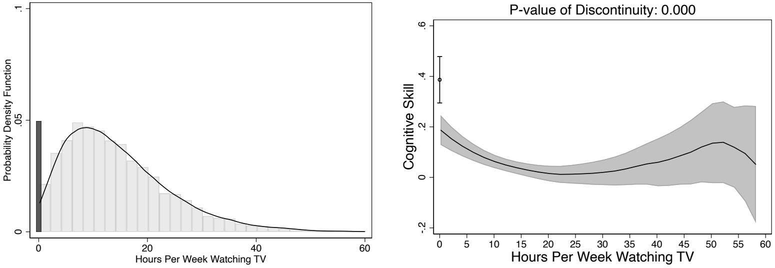
PR_i = how strict the parents are

X_i = hours watching TV per week



Main insight (2)

Discontinuity of the outcome at the bunching point reveals behavior of confounders



If effect of TV is continuous, discontinuity on skills is due to one reason: differences in confounders and/or their effects on the outcome among those at the bunching point and those near it.

Ordering at the bunching point

- Treatment variable X_i has bunching at $X_i = 0$.
- Suppose observations at the bunching point may be ordered by an unobservable $X_i^* \leq 0$. When $X_i > 0$, we write $X_i^* = X_i$, so

$$X_i = \max\{X_i^*, 0\}, \quad \mathbb{P}(X_i^* < 0) > 0$$

  - Helpful: X_i^* is a "**desired treatment**" under no constraint.
  - Helpful: X_i^* orders observations by **level of indifference** between $X_i = 0$ and $X_i = h$ for a small $h > 0$.
  - X_i^* **not a structural quantity**, only an index such that those with $X_i^* = -h$ are comparable with those with $X_i = h$.

Outcome and Treatment Effect

Potential outcomes model:

$$Y_i = Y_i(X_i)$$

Average Marginal Treatment Effect at $X_i = x$:

$$\beta(x) = \mathbb{E}[Y_i'(x)|X_i = x]$$

Fundamental problem of causality

$$Y_i = \underbrace{\mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*]}_{\text{treatment effect}} + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{\text{selection}} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\text{idiosyncratic error}}$$

Fundamental problem of causality

$$Y_i = \underbrace{\mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*]}_{\text{treatment effect}} + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{\text{selection}} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\text{idiosyncratic error}}$$

$$\frac{d}{dX_i}\mathbb{E}[Y_i|X_i = 0^+] := \underbrace{\mathbb{E}[Y_i'(0)|X_i^* = 0]}_{\beta(0)} + \underbrace{\frac{d}{dX_i^*}\mathbb{E}[Y_i(0)|X_i^* = 0]}_{\text{endogeneity bias}}$$

Assumption 1 (part 1)

1. **Distribution of $\mathbf{X_i^*}$:** for some $h > 0$,
   a. The support of the distribution of X_i is dense in $(0, h)$.
   b. X_i^* has a density for $x \in (-\infty, h]$
   c. $f_{X^*}(0) > 0$.
   d. $f_{X^*}(x)$ is right-continuous at $x = 0$

Continuous treatment and treatment effects conditional on X_i^*

2. **Treatment effects:** for all $x, x' \in [0, h)$, for some $h > 0$,
   a. $Y_i(x)$ is continuously differentiable a.s. conditional on $X_i^* = x'$
   b. For all $\epsilon > 0$, there exists $\delta > 0$ such that $x, x' \in [0, \delta) \implies \mathbb{E}[|Y_i'(x) - Y_i'(0)|\,|X_i^* = x'] < \epsilon$.
   c. $\mathbb{E}[|Y_i'(x)||X_i^* = x'] \leq C < \infty$.

Assumption 1 (part 2)

Continuous selection on X_i^* at zero

3. **Selection:**
   a. As $x \downarrow 0$, $Y_i(0)|X_i^* = x \rightarrow_{\text{a.s.}} Y_i(0)|X_i^* = 0$.
   b. $\mathbb{E}[Y_i'(0)|X_i^* = x']$ is right-continuous at $x' = 0$.
   c. Either $\mathbb{E}[Y_i(0)|X_i^* = x]$ is equal to zero, or it is monotonic and differentiable for $x \leq 0$.
   d. $f_{\mathbb{E}[Y(0)|X^*]|X=0} \leq C < \infty$.

4. **Idiosyncratic error:** either
   a. Stronger condition: $Y_i(0) - \mathbb{E}[Y_i(0)|X_i^*] \perp\!\!\!\perp X_i^*|X_i = 0$.
   b. Weakest condition: define W_i a random variable derived from the deconvolution of $dF_{Y(0)-\mathbb{E}[Y(0)|X^*]|X=0^+}$ from $f_{Y(0)|X=0}$. Then $f_{W|X=0}(\mathbb{E}[Y_i(0)|X_i^* = 0]) = f_{\mathbb{E}[Y(0)|X^*]|X=0}(\mathbb{E}[Y_i(0)|X_i^* = 0])$ holds.

Assumption 1 (part 2)

Monotonic selection on X_i^* at zero

3. **Selection:**
   a. As $x \downarrow 0$, $Y_i(0)|X_i^* = x \rightarrow_{\text{a.s.}} Y_i(0)|X_i^* = 0$.
   b. $\mathbb{E}[Y_i'(0)|X_i^* = x']$ is right-continuous at $x' = 0$.
   c. Either $\mathbb{E}[Y_i(0)|X_i^* = x]$ is equal to zero, or it is monotonic and differentiable for $x \leq 0$].
   d. $f_{\mathbb{E}[Y(0)|X^*]|X=0} \leq C < \infty$.

4. **Idiosyncratic error:** either
   a. Stronger condition: $Y_i(0) - \mathbb{E}[Y_i(0)|X_i^*] \perp\!\!\!\perp X_i^*|X_i = 0$.
   b. Weakest condition: define W_i a random variable derived from the deconvolution of $dF_{Y(0)-\mathbb{E}[Y(0)|X^*]|X=0^+}$ from $f_{Y(0)|X=0}$. Then $f_{W|X=0}(\mathbb{E}[Y_i(0)|X_i^* = 0]) = f_{\mathbb{E}[Y(0)|X^*]|X=0}(\mathbb{E}[Y_i(0)|X_i^* = 0])$ holds.

Assumption 1 (part 2)

3. **Selection:**
   a. As $x \downarrow 0$, $Y_i(0)|X_i^* = x \rightarrow_{\text{a.s.}} Y_i(0)|X_i^* = 0$.
   b. $\mathbb{E}[Y_i'(0)|X_i^* = x']$ is right-continuous at $x' = 0$.
   c. Either $\mathbb{E}[Y_i(0)|X_i^* = x]$ is equal to zero, or it is monotonic and differentiable for $x \leq 0$.
   d. $f_{\mathbb{E}[Y(0)|X^*]|X=0} \leq C < \infty$.

Local independence of the idiosyncratic error and X_i^* at bunching point

4. **Idiosyncratic error:** either
   a. Stronger condition: $Y_i(0) - \mathbb{E}[Y_i(0)|X_i^*] \perp\!\!\!\perp X_i^*|X_i = 0$.
   b. Weakest condition: define W_i a random variable derived from the deconvolution of $dF_{Y(0)-\mathbb{E}[Y(0)|X^*]|X=0^+}$ from $f_{Y(0)|X=0}$. Then $f_{W|X=0}(\mathbb{E}[Y_i(0)|X_i^* = 0]) = f_{\mathbb{E}[Y(0)|X^*]|X=0}(\mathbb{E}[Y_i(0)|X_i^* = 0])$ holds.

How to identify $u'(0)$?

$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\varepsilon_i}$$

$$\text{endogeneity bias: } u'(0) = \frac{d}{dX_i^*}\mathbb{E}[Y_i(0)|X_i^* = 0]$$

$u(X_i^*)$ is a deterministic function of X_i^*.

How to identify $u'(0)$?

$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\varepsilon_i}$$

$$\text{endogeneity bias: } u'(0) = \frac{d}{dX_i^*}\mathbb{E}[Y_i(0)|X_i^* = 0]$$

$u(X_i^*)$ is a deterministic function of X_i^*. Change in Variables Theorem:

u invertible and differentiable $\implies$ $f_{u(X^*)}(u(x)) \cdot |u'(x)| = f_{X^*}(x)$

How to identify $u'(0)$?

$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\varepsilon_i}$$

$$\text{endogeneity bias: } u'(0) = \frac{d}{dX_i^*}\mathbb{E}[Y_i(0)|X_i^* = 0]$$

$u(X_i^*)$ is a deterministic function of X_i^*. Change in Variables Theorem:

u invertible and differentiable $\implies f_{u(X^*)}(u(x)) \cdot |u'(x)| = f_{X^*}(x)$

If $f_{u(X^*)}(u(x)) \neq 0$,

$$u'(x) = \text{sgn}(u'(x)) \cdot \frac{f_{X^*}(x)}{f_{u(X^*)}(u(x))}$$

How to identify $u'(0)$?

$$u'(x) = \text{sgn}(u'(x)) \cdot \frac{f_{X^*}(x)}{f_{u(X^*)}(u(x))}$$

How to identify $u'(0)$?

$$u'(x) = \text{sgn}(u'(x)) \cdot \frac{f_{X^*|X=0}(x)}{f_{u(X^*)|X=0}(u(x))}$$

We can identify if $u(x)$ is increasing or decreasing.

We can identify $f_{X^*}(x)$ for $x > 0$.

We can identify $f_{u(X^*)|X^* \leq 0}(u(x))$ for $u(x) \leq u(0)$ if u is increasing,
for $u(x) \geq u(0)$ if u is decreasing.

How to identify $u'(0)$?

$$u'(x) = \text{sgn}(u'(x)) \cdot \frac{f_{X^*|X=0}(x)}{f_{u(X^*)|X=0}(u(x))}$$

We can identify if $u(x)$ is increasing or decreasing.

We can identify $f_{X^*}(x)$ for $x > 0$.

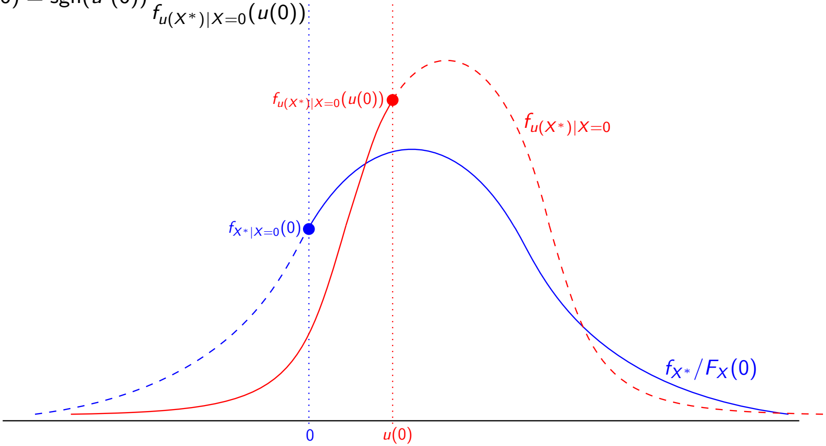
We can identify $f_{u(X^*)|X^* \leq 0}(u(x))$ for $u(x) \leq u(0)$ if u is increasing,

for $u(x) \geq u(0)$ if u is decreasing.

How to identify $u'(0)$?

Suppose u increasing.

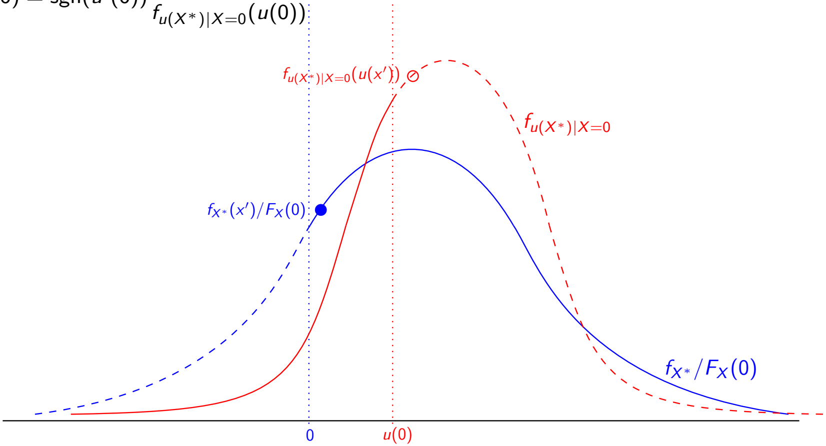
$$u'(0) = \mathrm{sgn}(u'(0)) \frac{f_{X^*|X=0}(0)}{f_{u(X^*)|X=0}(u(0))}$$



How to identify $u'(0)$?

Suppose u increasing.

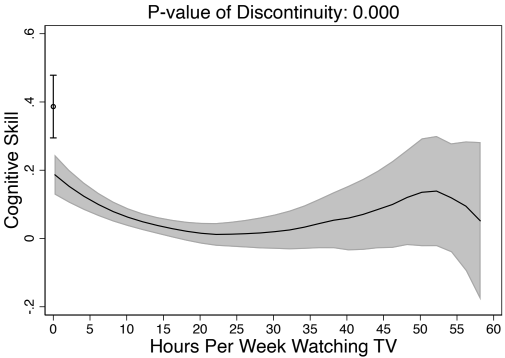
$$u'(0) = \operatorname{sgn}(u'(0)) \frac{f_{X^*|X=0}(0)}{f_{u(X^*)|X=0}(u(0))}$$



How to identify $sgn(u'(0))$?

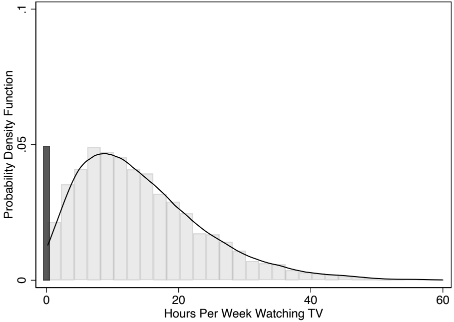
$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + Y_i - \mathbb{E}[Y_i|X_i^*]$$

$$\Delta := \mathbb{E}[Y_i|X_i = 0^+] - \mathbb{E}[Y_i|X_i = 0] = u(0) - \mathbb{E}[u(X_i^*)|X_i^* \leq 0]$$



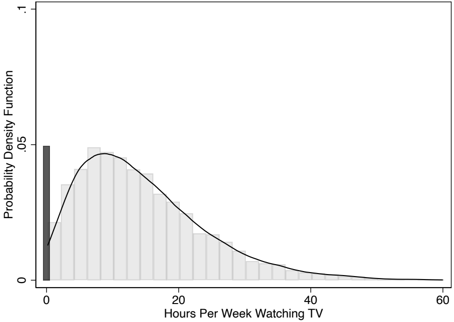
How to identify $f_{X^*|X_i=0}(0)$?

$$f_{X^*|X=0}(0) = \frac{f_{X^*}(0)}{F_X(0)}$$



How to identify $f_{X^*|X_i=0}(0)$?

$$f_{X^*|X=0}(0) = \frac{f_{X^*}(0)}{F_X(0)} = \frac{\lim_{x\downarrow 0} f_X(x)}{F_X(0)} =: \frac{f_X(0^+)}{F_X(0)}$$



How to identify $f_{u(X^*)|X=0}(u(0))$?

$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\varepsilon_i}$$

Convolution: for $X_i = 0$, $Y_i = u(X_i^*) + \varepsilon_i$

How to identify $f_{u(X^*)|X=0}(u(0))$?

$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\varepsilon_i}$$

Convolution: for $X_i = 0$, $Y_i = u(X_i^*) + \varepsilon_i$, so

$$f_{Y|X=0}(y) = \int f_{u(X^*)|X=0}(y-e)dF_{\varepsilon|X=0}(e)$$

How to identify $f_{u(X^*)|X=0}(u(0))$?

$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\varepsilon_i}$$

Convolution: for $X_i = 0$, $Y_i = u(X_i^*) + \varepsilon_i$, so

$$f_{Y|X=0}(y) = \int f_{u(X^*)|X=0}(y-e)dF_{\varepsilon|X=0}(e)$$

By the continuity of the distribution of $\varepsilon|X = x$, we can identify

$$F_{\varepsilon|X=0}(e) =: F_{Y-\mathbb{E}[Y|X=0^+]|X=0^+}(e)$$

How to identify $f_{u(X^*)|X=0}(u(0))$?

$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\varepsilon_i}$$

Convolution: for $X_i = 0$, $Y_i = u(X_i^*) + \varepsilon_i$, so

$$f_{Y|X=0}(y) = \int f_{u(X^*)|X=0}(y-e)dF_{\varepsilon|X=0}(e)$$

By the continuity of the distribution of $\varepsilon|X=x$, we can identify

$$F_{\varepsilon|X=0}(e) =: F_{Y-\mathbb{E}[Y|X=0^+]|X=0^+}(e)$$

By the Fourier representation:

$$f_{u(X^*)|X=0}(u(0)) = \frac{1}{2\pi}\int \frac{\mathbb{E}[e^{\mathbf{i}\xi Y_i}|X_i=0]}{\mathbb{E}[e^{\mathbf{i}\xi(Y_i-u(0))}|X_i=0^+]} e^{-\mathbf{i}\xi u(0)} d\xi,$$

where $\mathbf{i} = \sqrt{-1}$.

How to identify $f_{u(X^*)|X=0}(u(0))$?

$$Y_i = \mathbb{E}[Y_i(X_i) - Y_i(0)|X_i^*] + \underbrace{\mathbb{E}[Y_i(0)|X_i^*]}_{u(X_i^*)} + \underbrace{Y_i - \mathbb{E}[Y_i|X_i^*]}_{\varepsilon_i}$$

Convolution: for $X_i = 0$, $Y_i = u(X_i^*) + \varepsilon_i$, so

$$f_{Y|X=0}(y) = \int f_{u(X^*)|X=0}(y-e)dF_{\varepsilon|X=0}(e)$$

By the continuity of the distribution of $\varepsilon|X = x$, we can identify

$$F_{\varepsilon|X=0}(e) =: F_{Y-\mathbb{E}[Y|X=0^+]|X=0^+}(e)$$

By the Fourier representation:

$$f_{u(X^*)|X=0}(u(0)) = \frac{1}{2\pi}\int \frac{\mathbb{E}[e^{\mathbf{i}\xi Y_i}|X_i = 0]}{\mathbb{E}[e^{\mathbf{i}\xi(Y_i - \mathbb{E}[Y_i|X_i=0^+])}|X_i = 0^+]} e^{-\mathbf{i}\xi\mathbb{E}[Y_i|X_i=0^+]}d\xi,$$

where $\mathbf{i} = \sqrt{-1}$.

Theorem: Identification of tre

If Assumption 1 holds, then the $\beta(0) = \mathbb{E}[Y_i'(0)|X_i^* = 0]$ is identifiable as

$$\beta(0) = \lim_{x \downarrow 0} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - \text{sgn}(u'(0)) \cdot \frac{f_{X^*|X=0}(0)}{f_{u(X^*)|X=0}(u(0))},$$

where

$$\Delta = \mathbb{E}[Y_i|X_i = 0^+] - \mathbb{E}[Y_i|X_i = 0]$$

$$f_{u(X^*)|X=0}(u(0)) = \frac{1}{2\pi}\int \frac{\mathbb{E}[e^{\mathrm{i}\xi Y_i}|X_i = 0]}{\mathbb{E}[e^{\mathrm{i}\xi(Y_i - \mathbb{E}[Y_i|X_i=0^+])}|X_i = 0^+]} e^{-\mathrm{i}\xi\mathbb{E}[Y_i|X_i=0^+]} d\xi$$

Theorem: Identification of tre

If Assumption 1 holds, then the $\beta(0) = \mathbb{E}[Y_i'(0)|X_i^* = 0]$ is identifiable as

$$\beta(0) = \lim_{x \downarrow 0} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - \text{sgn}(\Delta) \cdot \frac{f_{X^*|X=0}(0)}{f_{u(X^*)|X=0}(u(0))},$$

where

$$\Delta = \mathbb{E}[Y_i|X_i = 0^+] - \mathbb{E}[Y_i|X_i = 0]$$

$$f_{u(X^*)|X=0}(u(0)) = \frac{1}{2\pi}\int \frac{\mathbb{E}[e^{i\xi Y_i}|X_i = 0]}{\mathbb{E}[e^{i\xi(Y_i - \mathbb{E}[Y_i|X_i=0^+])}|X_i = 0^+]} e^{-i\xi\mathbb{E}[Y_i|X_i=0^+]} d\xi$$

Theorem: Identification of tre

If Assumption 1 holds, then the $\beta(0) = \mathbb{E}[Y_i'(0)|X_i^* = 0]$ is identifiable as

$$\beta(0) = \lim_{x \downarrow 0} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - \text{sgn}(\Delta) \cdot \frac{f_{X^*|X=0}(0)}{f_{u(X^*)|X=0}(u(0))},$$

where

$$\Delta = \mathbb{E}[Y_i|X_i = 0^+] - \mathbb{E}[Y_i|X_i = 0]$$

$$f_{u(X^*)|X=0}(u(0)) = \frac{1}{2\pi}\int \frac{\mathbb{E}[e^{i\xi Y_i}|X_i = 0]}{\mathbb{E}[e^{i\xi(Y_i - \mathbb{E}[Y_i|X_i=0^+])}|X_i = 0^+]} e^{-i\xi\mathbb{E}[Y_i|X_i=0^+]} d\xi$$

Theorem: Identification of tre

If Assumption 1 holds, then the $\beta(0) = \mathbb{E}[Y_i'(0)|X_i^* = 0]$ is identifiable as

$$\beta(0) = \lim_{x \downarrow 0} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - \operatorname{sgn}(\Delta) \cdot \frac{f_X(0^+)/F_X(0)}{f_{u(X^*)|X=0}(u(0))},$$

where

$$\Delta = \mathbb{E}[Y_i|X_i = 0^+] - \mathbb{E}[Y_i|X_i = 0]$$

$$f_{u(X^*)|X=0}(u(0)) = \frac{1}{2\pi}\int \frac{\mathbb{E}[e^{i\xi Y_i}|X_i = 0]}{\mathbb{E}[e^{i\xi(Y_i - \mathbb{E}[Y_i|X_i=0^+])}|X_i = 0^+]} e^{-i\xi\mathbb{E}[Y_i|X_i=0^+]} d\xi$$

Theorem: Identification of tre

If Assumption 1 holds, then the $\beta(0) = \mathbb{E}[Y_i'(0)|X_i^* = 0]$ is identifiable as

$$\beta(0) = \lim_{x \downarrow 0} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - \text{sgn}(\Delta) \cdot \frac{f_X(0^+)/F_X(0)}{f_{u(X^*)|X=0}(u(0))},$$

where

$$\Delta = \mathbb{E}[Y_i|X_i = 0^+] - \mathbb{E}[Y_i|X_i = 0]$$

$$f_{u(X^*)|X=0}(u(0)) = \frac{1}{2\pi}\int \frac{\mathbb{E}[e^{i\xi Y_i}|X_i = 0]}{\mathbb{E}[e^{i\xi(Y_i - \mathbb{E}[Y_i|X_i=0^+])}|X_i = 0^+]} e^{-i\xi\mathbb{E}[Y_i|X_i=0^+]} d\xi$$

What else is on the paper?

- If $u'(x) = g(x;\theta)$ is invertible on θ, then all treatment effects are identified. Example: $\mathbb{E}[Y_i(0)|X_i^*] = \delta_0 + \delta X_i^*$.

What else is on the paper?

- If $u'(x) = g(x;\theta)$ is invertible on θ, then all treatment effects are identified. Example: $\mathbb{E}[Y_i(0)|X_i^*] = \delta_0 + \delta X_i^*$.
- Conditioning on controls, Z_i, weakens assumptions:
  1. Monotonicity of $\mathbb{E}[Y_i(0)|X_i^*, Z_i]$ may switch signs.
  2. $Y_i - \mathbb{E}[Y_i|X_i^*, Z_i] \perp\!\!\!\perp X_i^*|X_i = 0, Z_i$.
  3. Identification of treatment effects elsewhere may be done by assuming, e.g. $\mathbb{E}[Y_i(0)|X_i^*, Z_i] = \delta_0(Z_i) + \delta(Z_i)X_i^*$.

What else is on the paper?

- If $u'(x) = g(x;\theta)$ is invertible on θ, then all treatment effects are identified. Example: $\mathbb{E}[Y_i(0)|X_i^*] = \delta_0 + \delta X_i^*$.
- Conditioning on controls, Z_i, weakens assumptions:
  1. Monotonicity of $\mathbb{E}[Y_i(0)|X_i^*, Z_i]$ may switch signs.
  2. $Y_i - \mathbb{E}[Y_i|X_i^*, Z_i] \perp\!\!\!\perp X_i^*|X_i = 0, Z_i$.
  3. Identification of treatment effects elsewhere may be done by assuming, e.g. $\mathbb{E}[Y_i(0)|X_i^*, Z_i] = \delta_0(Z_i) + \delta(Z_i)X_i^*$.
- Estimators, and estimators in the control case: methods for large vector of mixed discrete/continuous controls.

What else is on the paper?

- If $u'(x) = g(x;\theta)$ is invertible on θ, then all treatment effects are identified. Example: $\mathbb{E}[Y_i(0)|X_i^*] = \delta_0 + \delta X_i^*$.
- Conditioning on controls, Z_i, weakens assumptions:
  1. Monotonicity of $\mathbb{E}[Y_i(0)|X_i^*, Z_i]$ may switch signs.
  2. $Y_i - \mathbb{E}[Y_i|X_i^*, Z_i] \perp\!\!\!\perp X_i^*|X_i = 0, Z_i$.
  3. Identification of treatment effects elsewhere may be done by assuming, e.g. $\mathbb{E}[Y_i(0)|X_i^*, Z_i] = \delta_0(Z_i) + \delta(Z_i)X_i^*$.
- Estimators, and estimators in the control case: methods for large vector of mixed discrete/continuous controls.
- Application to the effects of TV watching on children's cognitive and cognitive skills. Results are positive on cognitive skills and slightly more negative on non-cognitive skills.

Conclusion

- We show that bunching may be used for nonparametric identification of a treatment effect in the presence of endogeneity.
- Method identifies the average marginal treatment effect among the marginal observations at the bunching point.
- At the bunching point, requirements are: (1) ordering according to X_i^*, (2) monotonicity of $\mathbb{E}[Y_i(0)|X_i^*]$, (3) independence of $Y_i(0) - \mathbb{E}[Y_i(0)|X_i^*]$ and X_i^*.
- Above (near) the bunching point, requirement are (1) the right-continuity of treatment effects, (2) right continuity of $\mathbb{E}[Y_i'(0)|X_i^*]$ and distribution of $Y_i(0)|X_i^*$.
- Identification uses the change-in-variables theorem to derive $\frac{d}{dX_i^*}$ from the comparison between the density of X_i^* and the density of $\mathbb{E}[Y_i(0)|X_i^*]$ at $X_i* = 0$.

Average Marginal Treatment Effects elsewhere

$$\text{For } x > 0, \ \frac{d}{dx}\mathbb{E}[Y_i|X = x] = \beta(x) + u'(x)$$

- Extend requirements on differentiability to $x > 0$.
- In this case, identification of all marginal effects depends exclusively on assumptions on $u'(X_i^*) = \frac{d}{dx}\mathbb{E}[Y_i(0)|X_i^*]$:

Average Marginal Treatment Effects elsewhere

$$\text{For } x > 0,\ \frac{d}{dx}\mathbb{E}[Y_i|X = x] = \beta(x) + u'(x)$$

- Extend requirements on differentiability to $x > 0$.
- In this case, identification of all marginal effects depends exclusively on assumptions on $u'(X_i^*) = \frac{d}{dx}\mathbb{E}[Y_i(0)|X_i^*]$:

Assumption 2

Suppose that $u'(x) = g(x;\theta)$, where g is known, but the scalar θ is not. Suppose also that $g(0;\theta)$ is invertible in θ.

Average Marginal Treatment Effects elsewhere

$$\text{For } x > 0,\ \frac{d}{dx}\mathbb{E}[Y_i|X = x] = \beta(x) + u'(x)$$

- Extend requirements on differentiability to $x > 0$.
- In this case, identification of all marginal effects depends exclusively on assumptions on $u'(X_i^*) = \frac{d}{dx}\mathbb{E}[Y_i(0)|X_i^*]$:

Assumption 2

Suppose that $u'(x) = g(x;\theta)$, where g is known, but the scalar θ is not. Suppose also that $g(0;\theta)$ is invertible in θ.

If Assumptions 1 and 2 hold, then

$$u'(x) = g(x;\theta) \implies \theta = g^{-1}(0; u'(0))$$

and so,

$$\beta(x) = \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - g(x; g^{-1}(0; u'(0))).$$

Identification of $\beta(0)$ including controls

- $\mathcal{Z} = \{z \in \text{supp}(Z) \text{ s. t. } \mathbb{P}(X_i = 0|Z_i) > 0\}$

If Assumption 1 conditional on controls holds on $\mathcal{Z}$ with probability one, then $\beta(0)$ is identifiable.

$$\beta(0) = \mathbb{E}[\beta(0, Z_i)|X_i = 0] = \mathbb{E}[\beta(0, Z_i)|X_i = 0, Z_i \in \mathcal{Z}]\mathbb{P}(Z_i \in \mathcal{Z}|X_i = 0).$$

We provide estimators for cases with discrete, continuous and mixed large dimensional controls.

Estimation

$$\hat{\beta}(0) = d_X \hat{\mathbb{E}}[Y_i|X_i = 0^+] - \mathrm{sgn}(\hat{\Delta}) \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

where

$$d_X \hat{\mathbb{E}}[Y_i|X_i = 0^+] = \lim_{x \downarrow 0} \frac{d}{dx} \hat{\mathbb{E}}[Y_i|X_i = x]$$

$$\hat{\Delta} = \hat{\mathbb{E}}[Y_i|X_i = 0^+] - \hat{\mathbb{E}}[Y_i|X_i = 0]$$

Estimation

$$\hat{\beta}(0) = d_X \hat{\mathbb{E}}[Y_i | X_i = 0^+] - \text{sgn}(\hat{\Delta}) \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

where

$$d_X \hat{\mathbb{E}}[Y_i | X_i = 0^+] = \lim_{x \downarrow 0} \frac{d}{dx} \hat{\mathbb{E}}[Y_i | X_i = x]$$

$$\hat{\Delta} = \hat{\mathbb{E}}[Y_i | X_i = 0^+] - \hat{\mathbb{E}}[Y_i | X_i = 0]$$

- $\hat{F}_X(0) = \frac{1}{n} \sum_{i=1}^{n} 1(X_i = 0)$

Estimation

$$\hat{\beta}(0) = d_X \hat{\mathbb{E}}[Y_i|X_i = 0^+] - \text{sgn}(\hat{\Delta}) \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

where

$$d_X \hat{\mathbb{E}}[Y_i|X_i = 0^+] = \lim_{x \downarrow 0} \frac{d}{dx} \hat{\mathbb{E}}[Y_i|X_i = x]$$

$$\hat{\Delta} = \hat{\mathbb{E}}[Y_i|X_i = 0^+] - \hat{\mathbb{E}}[Y_i|X_i = 0]$$

- $\hat{F}_X(0) = \frac{1}{n} \sum_{i=1}^{n} 1(X_i = 0)$
- $\hat{\mathbb{E}}[Y_i|X_i = 0] = \hat{F}_X(0)^{-1} \frac{1}{n} \sum_{i=1}^{n} Y_i \cdot 1(X_i = 0)$

Estimation

$$\hat{\beta}(0) = d_X\hat{\mathbb{E}}[Y_i|X_i = 0^+] - \text{sgn}(\hat{\Delta}) \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

where

$$d_X\hat{\mathbb{E}}[Y_i|X_i = 0^+] = \lim_{x \downarrow 0} \frac{d}{dx}\hat{\mathbb{E}}[Y_i|X_i = x]$$

$$\hat{\Delta} = \hat{\mathbb{E}}[Y_i|X_i = 0^+] - \hat{\mathbb{E}}[Y_i|X_i = 0]$$

- $\hat{F}_X(0) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i = 0)$
- $\hat{\mathbb{E}}[Y_i|X_i = 0] = \hat{F}_X(0)^{-1}\frac{1}{n}\sum_{i=1}^{n} Y_i \cdot 1(X_i = 0)$
- $\hat{\mathbb{E}}[Y_i|X_i = 0^+]$ is the intercept, and $d_X\hat{\mathbb{E}}[Y_i|X_i = 0^+]$ is the slope of a local linear regression of Y_i onto X_i at $X_i = 0$ using only observations such that $X_i > 0$.

Estimation

$$\hat{\beta}(0) = d_X \hat{\mathbb{E}}[Y_i | X_i = 0^+] - \frac{\hat{\Delta}}{|\hat{\Delta}|} \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

- $\hat{f}_X(0)_+$ uses Pinkse and Schurter (2021)'s estimator.

$$\hat{f}_X(0)_+ = \frac{1}{D} \frac{1}{nh} \sum_{i=1}^{n} k\left(X_i/h\right),$$

$$D = \frac{3}{2} \cdot \frac{2 + \hat{L}'_X(0)^2 h^2 - e^{\hat{L}'_X(0)h}(2 - 2\hat{L}'_X(0)h)}{\hat{L}'_X(0)^3 h^3}$$

$$\hat{L}'_X(0) = \frac{\frac{1}{nh}\sum_{i=1}^{n}(1 - 2X_i/h)1(0 \leq X_i \leq h)}{\frac{1}{nh}\sum_{i=1}^{n} X_i(1 - X_i/h)1(0 \leq X_i \leq h)},$$

Estimation

$$\hat{\beta}(0) = d_X \hat{\mathbb{E}}[Y_i | X_i = 0^+] - \frac{\hat{\Delta}}{|\hat{\Delta}|} \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

- $\hat{f}_{Y|X=0}(Y_i)$ is a classic deconvolution estimator.

Estimation

$$\hat{\beta}(0) = d_X \hat{\mathbb{E}}[Y_i|X_i = 0^+] - \frac{\hat{\Delta}}{|\hat{\Delta}|} \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

- $\hat{f}_{Y|X=0}(Y_i)$ is a classic deconvolution estimator.

  1. Program the function $\hat{\phi}(\xi) = \mathbf{e}_1(\mathbf{x}'\mathbf{k}\mathbf{x})^{-1}\mathbf{x}'\mathbf{k}A(\xi)$
     - $A(\xi) := (e^{\mathrm{i}\xi(Y_1 - \hat{\mathbb{E}}[Y_i|X=0^+]}, \dots, e^{\mathrm{i}\xi(Y_n - \hat{\mathbb{E}}[Y_i|X=0^+]})'$
     - $\mathbf{x}$ is the matrix with with rows $(1, X_i)'$
     - $\mathbf{k} = Diag\{k_3(X_i/h_3)1(X_i > 0)\}_{i=1}^{n}$
     - $\mathbf{e}_1 = (1, 0)'$

Estimation

$$\hat{\beta}(0) = d_X \hat{\mathbb{E}}[Y_i|X_i = 0^+] - \frac{\hat{\Delta}}{|\hat{\Delta}|} \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

- $\hat{f}_{Y|X=0}(Y_i)$ is a classic deconvolution estimator.

1. Program the function $\hat{\phi}(\xi) = \mathbf{e}_1(\mathbf{x}'\mathbf{k}\mathbf{x})^{-1}\mathbf{x}'\mathbf{k}A(\xi)$
   - $A(\xi) := (e^{i\xi(Y_1 - \hat{\mathbb{E}}[Y_i|X=0^+]}, \ldots, e^{i\xi(Y_n - \hat{\mathbb{E}}[Y_i|X=0^+]})'$
   - $\mathbf{x}$ is the matrix with with rows $(1, X_i)'$
   - $\mathbf{k} = Diag\{k_3(X_i/h_3)1(X_i > 0)\}_{i=1}^n$
   - $\mathbf{e}_1 = (1, 0)'$
2. Calculate for Y_i such that $X_i = 0$,
   $$g(Y_i) = \frac{1}{\hat{F}_X(0) \cdot 2\pi} \int e^{i\xi(Y_i - \hat{\mathbb{E}}[Y_i|X_i=0^+])} \frac{\phi_K(h_4\xi)}{A(\xi)} d\xi$$
   - $\phi_{k_4}(h_4\xi) = \int k_4(\nu) e^{ih_4\xi\nu} d\nu$ is the Fourier transform of the kernel k_4.

Estimation

$$\hat{\beta}(0) = d_X \hat{\mathbb{E}}[Y_i|X_i = 0^+] - \frac{\hat{\Delta}}{|\hat{\Delta}|} \cdot \frac{\hat{f}_X(0)_+}{\hat{F}_X(0) \cdot \hat{f}_{u(X^*)|X=0}(u(0))},$$

- $\hat{f}_{Y|X=0}(Y_i)$ is a classic deconvolution estimator.

1. Program the function $\hat{\phi}(\xi) = \mathbf{e}_1(\mathbf{x}'\mathbf{k}\mathbf{x})^{-1}\mathbf{x}'\mathbf{k}A(\xi)$
   - $A(\xi) := (e^{i\xi(Y_1 - \hat{\mathbb{E}}[Y_i|X=0^+]}, \ldots, e^{i\xi(Y_n - \hat{\mathbb{E}}[Y_i|X=0^+]})'$
   - $\mathbf{x}$ is the matrix with with rows $(1, X_i)'$
   - $\mathbf{k} = Diag\{k_3(X_i/h_3)1(X_i > 0)\}_{i=1}^n$
   - $\mathbf{e}_1 = (1, 0)'$
2. Calculate for Y_i such that $X_i = 0$,
   $$g(Y_i) = \frac{1}{\hat{F}_X(0) \cdot 2\pi} \int e^{i\xi(Y_i - \hat{\mathbb{E}}[Y_i|X_i=0^+])} \frac{\phi_K(h_4\xi)}{A(\xi)} d\xi$$
   - $\phi_{k_4}(h_4\xi) = \int k_4(\nu)e^{ih_4\xi\nu}d\nu$ is the Fourier transform of the kernel k_4.
3. $$\hat{f}_{u(X^*)|X=0}(u(0)) = \frac{1}{nh_4}\sum_{i=1}^n g(Y_i)1(X_i = 0)$$

Application

- Panel Study of Income Dynamics Child Development Supplement (PSID-CDS)
- Waves 1997, 2002 and 2007.
- All children 5-18 years of age with complete time diaries and non-missing skill measures.
- $N = 4,396$.

Outcome Variables (standardized)	All	Grades K-5	Grades 6-8	Grades 9-12
Cognitive Skill	0.07	-0.35	0.30	0.56
	(0.93)	(0.93)	(0.76)	(0.72)
Non-cognitive Skill	0.01	-0.02	-0.03	0.11
	(0.95)	(0.95)	(0.98)	(0.92)
Treatment Variable (hours per day)				
Watch TV	1.99	1.92	2.12	1.97
	(1.46)	(1.30)	(1.51)	(1.67)
Watch TV = 0	0.05	0.03	0.04	0.09
	(0.22)	(0.17)	(0.20)	(0.28)
Covariates				
Child is Male	0.51	0.53	0.46	0.50
	(0.50)	(0.50)	(0.50)	(0.50)
Child is White	0.48	0.49	0.47	0.47
	(0.50)	(0.50)	(0.50)	(0.50)
Child is Black	0.40	0.39	0.41	0.41
	(0.49)	(0.49)	(0.49)	(0.49)
Child is Hispanic	0.08	0.08	0.07	0.08
	(0.26)	(0.27)	(0.26)	(0.27)
Child is Another Race	0.05	0.05	0.05	0.04
	(0.21)	(0.21)	(0.21)	(0.20)
Child is in Grade PreK-5	0.47	1.00	0.00	0.00
	(0.50)	(0.00)	(0.00)	(0.00)
Child is in Grade 6-8	0.27	0.00	1.00	0.00
	(0.44)	(0.00)	(0.00)	(0.00)
Child is in Grade 9-12	0.27	0.00	0.00	1.00
	(0.44)	(0.00)	(0.00)	(0.00)
Household Income (in $1,000s)	73.62	66.63	72.79	86.77
	(82.21)	(67.65)	(70.87)	(109.95)
Age (years)	11.29	8.10	12.38	15.82
	(3.64)	(2.13)	(0.95)	(1.19)
Observations	4,396	2,060	1,167	1,169

Application

X_i=Watching TV (hours per week)

Y_i= cognitive and non-cognitive skills (standard deviation units)

	Bandwidth (in hours per week)		
Cognitive Skill	$h = 5$	$h = 7$	$h = 10$
$\beta(0)$	0.178	0.259	0.338
	(0.122)	(0.113)	(0.106)
$u'(0)$	-0.232	-0.295	-0.357
	(0.118)	(0.105)	(0.097)
Non-Cognitive Skill	$h = 5$	$h = 7$	$h = 10$
$\beta(0)$	-0.215	-0.277	-0.331
	(0.103)	(0.098)	(0.092)
$u'(0)$	0.212	0.269	0.326
	(0.099)	(0.087)	(0.085)