Identification and Estimation of the Marginal Treatment Effect (MTE) without Instrumental Variable (IV)

Zhewen Pan[1]   Zhengxin Wang[1]   Junsen Zhang[2*]   Yahong Zhou[3]

[1] Zhejiang University of Finance & Economics

[2] Zhejiang University

[3] Shanghai University of Finance & Economics

June 29th, 2024 At AMES, Hangzhou

Contents

1. Introduction
2. Model
3. Identification
4. Estimation
5. Empirical application to Head Start
6. Conclusion

1. Introduction

- MTE (Heckman and Vytlacil, 1999, 2001, 2005) is a tool for describing, interpreting, and analyzing **heterogeneous** causal effects of a **nonrandom** treatment.
    - $\mathrm{MTE}(x, v) = E[Y_1 - Y_0 | X = x, V = v]$
- Existing methods of MTE rely heavily on IV.
    - In this paper, we attempt to **model**, **identify**, and **estimate** MTE without IV.
- Main value of our method
    - When IV is hard to find: consistently estimate heterogeneous causal effects
    - When IV is available but under question: conveniently test exclusion of IV
    - When IV is valid: check robustness to alternative identifying assumptions

1. Introduction: literature review

- Set identification for sample selection models: Honore and Hu (2020, 2023)
- Sensitivity analysis for exclusion restrictions: Conley et al (2012), Kippersluis (2018)
- Identification based on heteroscedasticity: Lewbel (2012, 2018)
- Extremal quantile regression for sample selection models: D'Haultfœuille et al (2018)
- Local irrelevance assumption in control function approach: D'Haultfœuille et al (2023)
- **Identification based on functional form**: Escanciano et al (2016)
  - linear outcome equations $Y_d = X'\beta_d + U_d$
  - nonlinear propensity score $\pi(x) = E[D|X = x]$
  - conditional mean independence $E[U_d|V, X] = E[U_d|V]$

1. Introduction: preliminaries

- Potential outcomes model
  - $Y = Y_0 + D(Y_1 - Y_0) = DY_1 + (1 - D)Y_0 = \begin{cases} Y_1 & \text{if } D = 1 \\ Y_0 & \text{if } D = 0 \end{cases}$
- Selection on observables $(Y_1, Y_0) \perp D \mid X$: PSM or IPW $\Rightarrow$ ATE $= E[Y_1 - Y_0]$
- Selection on unobservables $(Y_1, Y_0) \perp Z \mid X$: ivregress Y on $(D = Z)$ $X \Rightarrow$ LATE
  - Denote $D = ZD_1 + (1 - Z)D_0$ where Z is binary, then LATE $= E[Y_1 - Y_0 | D_1 = 1, D_0 = 0]$
  - Monotonicity assumption: $\Pr(D_1 \geq D_0) = 1$

1. Introduction: from LATE to MTE

- Selection model or generalized Roy model:
  - $D = 1\{\mu(X,Z) \geq U\}$ and $(Y_1, Y_0, U) \perp Z \mid X$
  - normalized to $D = 1\{F_{U|X}(\mu(X,Z)) \geq F_{U|X}(U)\} = 1\{\pi(X,Z) \geq V\}$, where
    - $\pi(x,z) = E[D|X = x, Z = z]$ is the propensity score
    - V is the normalized error term s.t. $V|X \sim \text{Uniform}(0,1)$ and $V \perp X$
- Vytlacil (2002) established **equivalence** of the selection model to the LATE model
- Marginal treatment effect is defined as $\text{MTE}(x,v) = E[Y_1 - Y_0|X = x, V = v]$
  - $\text{ATE}(x) = E[Y_1 - Y_0|X = x] = \int_0^1 \text{MTE}(x,v)dv$
  - $\text{LATE}(x) = \frac{1}{\pi_1 - \pi_0}\int_{\pi_0}^{\pi_1} \text{MTE}(x,v)dv$ where $\pi_1 = \pi(x,1)$ and $\pi_0 = \pi(x,0)$
- Identification of MTE: $\text{MTE}(x,v) = \frac{\partial E[Y|X=x,\pi(X,Z)=v]}{\partial v}$ (Z should be continuous)

Contents

1. Introduction
2. Model
3. Identification
4. Estimation
5. Empirical application to Head Start
6. Conclusion

2. Model

- Selection model without IV : $D = 1\{\mu(X) \geq U\}$
  - X is not necessarily stochastically independent of U
  - Separability or monotonicity is not required, e.g., U can depend functionally on X
  - Exclusion restriction is not required, namely, all X can appear in outcome equations

Example 1. *Consider a latent index rule for the treatment participation:*

$$D = 1\{m(X, \varepsilon) \geq 0\}, \tag{2}$$

where the observables X can be statistically correlated with the unobservables ε, and no restriction is imposed on the cross partials of the index function m. Without independence and additive separability, model (2) is known to be completely vacuous, imposing no restrictions on the observed or counterfactual outcomes (Heckman and Vytlacil, 2001). This general latent index rule can fit into the threshold crossing rule (1) by taking $\mu(X) = E[m(X, \varepsilon)|X]$ and $U = \mu(X) - m(X, \varepsilon)$.

2. Model: normalization

- Selection model without IV : $D = 1\{\mu(X) \geq U\}$
- Can be normalized to be : $D = 1\{\pi(X) \geq V\}$
- where $\pi(x) = E[D|X = x] = F_{U|X}(\mu(x)|x)$ is the propensity score
- and $V = F_{U|X}(U)$, satisfying $V|X \sim \text{Uniform}(0,1)$ and $V \perp X$
    - normalized error term
    - the unobservables projected onto the subspace orthogonal to that spanned by X
    - rank of U conditional on X
    - willingness to pay
    - resistence to treatment (cost) or distaste for treatment (preference)

2. Model: normalized error term

Example 2. *Suppose that X is a scalar and that*

$$\begin{pmatrix} U \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{UX} \\ \sigma_{UX} & \sigma_X^2 \end{pmatrix}\right).$$

By the property of bivariate normal distribution, we have $U\,|(X=x) \sim N\left(\mu_{U|X}(x), \sigma_{U|X}^2\right)$ *and* $F_{U|X}(u\,|x) = \Phi\left(\left[u - \mu_{U|X}(x)\right]/\sigma_{U|X}\right)$, *where* $\mu_{U|X}(x) = (\sigma_{UX}/\sigma_X^2)\,x$, $\sigma_{U|X}^2 = 1 - (\sigma_{UX}^2/\sigma_X^2)$, *and* $\Phi(\cdot)$ *denotes the standard normal CDF. Hence,*

$$V = F_{U|X}(U\,|X) = \Phi\left(\frac{U - \mu_{U|X}(X)}{\sigma_{U|X}}\right) \text{ and } V_x = \Phi\left(\frac{U - \mu_{U|X}(x)}{\sigma_{U|X}}\right).$$

It is straightforward that $V \perp\!\!\!\perp X$ *since* $F_{V|X}(v\,|x) = v$, *but that* $V_x \not\perp\!\!\!\perp X$ *since*

$$F_{V_x|X}(v\,|\tilde{x}) = \Phi\left(\frac{(\sigma_{UX}/\sigma_X^2)(x - \tilde{x})}{\sigma_{U|X}} + \Phi^{-1}(v)\right),$$

and that V_x *is not uniformly distributed since* $F_{V_x}(v) = \Phi\left(\mu_{U|X}(x) + \sigma_{U|X}\Phi^{-1}(v)\right)$.

2. Model: definition of MTE

$$\text{MTE}(x, v) = E[Y_1 - Y_0 | X = x, V = v]$$

- Relationship between MTE and commonly-used causal parameters:

$$\text{ATE}(x) = E[Y_1 - Y_0 | X = x] = \int_0^1 \text{MTE}(x, v) dv$$

$$\text{ATT}(x) = E[Y_1 - Y_0 | X = x, D = 1] = \frac{1}{\pi(x)} \int_0^{\pi(x)} \text{MTE}(x, v) dv$$

$$\text{ATUT}(x) = E[Y_1 - Y_0 | X = x, D = 0] = \frac{1}{1 - \pi(x)} \int_{\pi(x)}^1 \text{MTE}(x, v) dv$$

$$\text{LATE}(x, v_0, v_1) = E[Y_1 - Y_0 | X = x, v_0 \leq V \leq v_1] = \frac{1}{v_1 - v_0} \int_{v_0}^{v_1} \text{MTE}(x, v) dv$$

Contents

1. Introduction
2. Model
3. Identification
4. Estimation
5. Empirical application to Head Start
6. Conclusion

3. Identification

- Selection model without IV : $D = 1\{\pi(X) \geq V\}$
- Identification by functional form (in a semiparametric version)
  - nonlinear $\pi(x)$
  - linear outcome equations $Y_d = X'\beta_d + U_d$
  - conditional mean independence $E[U_d|V,X] = E[U_d|V]$
  - $\pi(X) - X'\beta_d$ provides excluded variation, playing the role of a continuous IV
- Some notation before imposing the assumptions
  - $X = (X^C, X^D)$ where X^C is continuous and X^D is discrete
  - $\pi_0(x^C) = \pi(x^C, 0)$
  - Denote x_k, x_k^C, or x_k^D as the k-th element of x, x^C, or x^D

3. Identification: nonlinearity

Assumption NL (Non-Linearity). *Assume that π_0 satisfies the following NL1 when $\dim\left(X^C\right)=1$, or NL2 when $\dim\left(X^C\right)\geq 2$.*

– *NL1 $\left(\dim\left(X^C\right)=1\right)$: there exist two different constants x^C, $\tilde{x}^C$ in the support of X^C such that $\pi_0\left(x^C\right)=\pi_0\left(\tilde{x}^C\right)$.*

– *NL2 $\left(\dim\left(X^C\right)\geq 2\right)$: there exist two vectors x^C, $\tilde{x}^C$ in the support of X^C and two elements k, j of the set $\left\{1,2,\cdots,\dim\left(X^C\right)\right\}$ such that π_0 and $E\left[Y\left|X^C=x^C, X^D=0, D=d\right.\right]$, $d=0,1$, are differentiable at x^C and $\tilde{x}^C$, and that* (i) $\partial_k\pi_0\left(x^C\right)\neq 0$, (ii) $\partial_j\pi_0\left(x^C\right)\neq 0$, (iii) $\partial_k\pi_0\left(\tilde{x}^C\right)\neq 0$, (iv) $\partial_j\pi_0\left(\tilde{x}^C\right)\neq 0$, (v) $\partial_k\pi_0\left(x^C\right)/\partial_j\pi_0\left(x^C\right)\neq\partial_k\pi_0\left(\tilde{x}^C\right)/\partial_j\pi_0\left(\tilde{x}^C\right)$.

- Assumption NL2 will not hold if $\pi_0(x^C)=f(\gamma' x^C)$. Otherwise, it generally holds.

Example 3. *Consider the case of two continuous covariates. Suppose, for some smooth function f, $\pi_0\left(x^C\right)=f\left(\gamma_1 x_1^C+\gamma_2 x_2^C+\gamma_3 x_1^C x_2^C\right)$ or $\pi_0\left(x^C\right)=f\left(\gamma_1 x_1^C+\gamma_2 x_2^C+\gamma_3\left(x_1^C\right)^2\right)$. Then we have $\partial_1\pi_0\left(x^C\right)/\partial_2\pi_0\left(x^C\right)=\left(\gamma_1+\gamma_3 x_2^C\right)/\left(\gamma_2+\gamma_3 x_1^C\right)$ for the interaction case, or $\partial_1\pi_0\left(x^C\right)/\partial_2\pi_0\left(x^C\right)=\left(\gamma_1+2\gamma_3 x_1^C\right)/\gamma_2$ for the quadratic case. In both cases, Assumption NL2 (v) generally holds for x^C and $\tilde{x}^C$ satisfying $x_1^C\neq\tilde{x}_1^C$.*

3. Identification: linearity

Assumption L (Linearity). *Assume that* $E[Y_d|X=x]=\alpha_d+x'\beta_d$ *for some fixed* α_d *and* β_d, $d=0,1$.

Assumption CMI (Conditional Mean Independence). *Denote* $U_d=Y_d-X'\beta_d$, $d=0,1$. *Assume that* $E[U_d|V,X]=E[U_d|V]$ *with probability one for* $d=0,1$.

- Under Assumptions L and CMI, we have:

$$\Delta^{\text{MTE}}(x,v)=x'(\beta_1-\beta_0)+E[U_1-U_0|V=v]$$

$$E[Y|X=x,D=d]=x'\beta_d+g_d(\pi(x))$$

$$g_0(p)=E[U_0|V>p]=\frac{1}{1-p}\int_p^1 E[U_0|V=v]\,dv,$$

$$g_1(p)=E[U_1|V\le p]=\frac{1}{p}\int_0^p E[U_1|V=v]\,dv.$$

$$E[U_0|V=p]=g_0(p)-(1-p)g_0^{(1)}(p),$$

$$E[U_1|V=p]=g_1(p)+pg_1^{(1)}(p),$$

3. Identification: Assumption CMI

- $E[U_d|V,X] = E[U_d|V]$
  - Standard in the MTE literature
  - also referred as separability
- Note that by definition
  - $E[U_d|X] = \alpha_d = E[U_d]$
  - $V \perp X$
- Assumption CMI essentially requires the copula of (U_d, V) not depend on X
  - much weaker than $(U_d, U) \perp X$
  - does not rule out the marginal dependence of U_d or U on X

Example 4. *Suppose that X is a scalar and that*

$$\begin{pmatrix} U_d \\ U \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & \sigma_{dU} & 0 \\ \sigma_{dU} & 1 & \sigma_{UX} \\ 0 & \sigma_{UX} & \sigma_X^2 \end{pmatrix} \right).$$

Note that U is correlated with X in this setting. By the property of multivariate normal distribution, we have

$$\begin{pmatrix} U_d \\ U \end{pmatrix} \Big| (X = x) \sim N\left(\begin{pmatrix} 0 \\ \mu_{U|X}(x) \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & \sigma_{dU} \\ \sigma_{dU} & \sigma_{U|X}^2 \end{pmatrix} \right), \qquad (4)$$

where $\mu_{U|X}(x) = (\sigma_{UX}/\sigma_X^2)x$ and $\sigma_{U|X}^2 = 1 - (\sigma_{UX}^2/\sigma_X^2)$. Hence,

$$E[U_d | U = u, X = x] = \frac{\sigma_{dU}}{\sigma_{U|X}^2}(u - \mu_{U|X}(x)).$$

By Example 2, we have $V = \Phi([U - \mu_{U|X}(X)]/\sigma_{U|X})$, so that $U = \sigma_{U|X}\Phi^{-1}(V) + \mu_{U|X}(X)$. Consequently,

$$E[U_d | V = v, X = x] = E[U_d | U = \sigma_{U|X}\Phi^{-1}(v) + \mu_{U|X}(x), X = x] = \frac{\sigma_{dU}}{\sigma_{U|X}}\Phi^{-1}(v),$$

and Assumption CMI holds. More generally, to allow for the dependence of U_d on X as well, we can instead set

$$\begin{pmatrix} U_d \\ U \end{pmatrix} \Big| (X = x) \sim N\left(\begin{pmatrix} 0 \\ \mu_{U|X}(x) \end{pmatrix}, \begin{pmatrix} \sigma_d^2(x) & \sigma_{dU} \\ \sigma_{dU} & \sigma_{U|X}^2 \end{pmatrix} \right)$$

in place of (4), where $\sigma_d^2(x)$ is the conditional variance of U_d given $X = x$. Since $E[U_d | V = v, X = x]$ is irrelevant to the variance of U_d by the above analysis, Assumption CMI will still hold in the presence of such heteroscedastic U_d.

3. Identification: main result

Theorem 1. *If Assumptions L, NL, CMI, and S hold, then β_d and $g_d(p)$ at all p in the support of the propensity score P are identified for $d = 0, 1$.*

- Theorem 1 implies identification of MTE without IV :

$$\Delta^{\text{MTE}}(x, v) = x'(\beta_1 - \beta_0) + [g_1(v) - g_0(v)] + v g_1^{(1)}(v) + (1 - v) g_0^{(1)}(v)$$

- as well as other causal parameters :

$$\Delta^{\text{ATE}}(x) = x'(\beta_1 - \beta_0) + [g_1(1) - g_0(0)],$$

$$\Delta^{\text{TT}}(x) = x'(\beta_1 - \beta_0) + g_1(\pi(x)) + \frac{(1 - \pi(x)) g_0(\pi(x)) - g_0(0)}{\pi(x)},$$

$$\Delta^{\text{TUT}}(x) = x'(\beta_1 - \beta_0) + \frac{g_1(1) - \pi(x) g_1(\pi(x))}{1 - \pi(x)} - g_0(\pi(x)),$$

$$\Delta^{\text{LATE}}(x, v_1, v_2) = x'(\beta_1 - \beta_0) + \frac{v_2 g_1(v_2) - v_1 g_1(v_1) + (1 - v_2) g_0(v_2) - (1 - v_1) g_0(v_1)}{v_2 - v_1}.$$

3. Identification: sketch of proof

- The identification is grounded on $E[Y|X=x, D=d] = x'\beta_d + g_d\big(\pi(x)\big)$
  - Denote $m_d(x^C) = E[Y|X^C = x^C, X^D = 0, D = d] = x^{C\prime}\beta_d^C + g_d\big(\pi_0(x^C)\big)$
  - $m_d(x^C)$ and $\pi_0(x^C) = E[D|X^C = x^C, X^D = 0]$ are directly identified from the data
- $x_k^C \rightarrow \xi_k^C = x_k^C + \epsilon$ and $x_j^C \rightarrow \xi_j^C = x_j^C + \epsilon \cdot \left(-\frac{\partial_k \pi_0(x^C)}{\partial_j \pi_0(x^C)}\right)$, then $\pi_0(x^C)$ remains unchanged
  - intuition: the derivative of implicit function $\pi_0(x^C) = c$ is $\frac{\partial x_j^C}{\partial x_k^C} = -\frac{\partial_k \pi_0(x^C)}{\partial_j \pi_0(x^C)}$
  - $m_d(\xi^C) = x^{C\prime}\beta_d^C + g_d\big(\pi_0(x^C)\big) + \epsilon \cdot \beta_{d,k}^C - \epsilon \cdot \frac{\partial_k \pi_0(x^C)}{\partial_j \pi_0(x^C)} \cdot \beta_{d,j}^C$
  - $m_d(\xi^C) - m_d(x^C) = \epsilon \cdot \beta_{d,k}^C - \epsilon \cdot \frac{\partial_k \pi_0(x^C)}{\partial_j \pi_0(x^C)} \cdot \beta_{d,j}^C$
  - $m_d(\tilde{\xi}^C) - m_d(\tilde{x}^C) = \epsilon \cdot \beta_{d,k}^C - \epsilon \cdot \frac{\partial_k \pi_0(\tilde{x}^C)}{\partial_j \pi_0(\tilde{x}^C)} \cdot \beta_{d,j}^C$

Contents

1. Introduction
2. Model
3. Identification
4. Estimation
5. Empirical application to Head Start
6. Conclusion

4. Estimation: first stage

- First stage: estimation of propensity score $\pi(x) = E[D|X = x]$ and $P_i = \pi(X_i)$
- Recommendation: nonparametric estimation

$$\hat{\pi}(x) = \frac{\sum_{i=1}^{n} D_i \left[\prod_{l=1}^{\dim(X^C)} k_1\left(\left(X_{il}^C - x_l^C\right)/h_{1l}\right)\right] 1\left\{X_i^D = x^D\right\}}{\sum_{i=1}^{n} \left[\prod_{l=1}^{\dim(X^C)} k_1\left(\left(X_{il}^C - x_l^C\right)/h_{1l}\right)\right] 1\left\{X_i^D = x^D\right\}} \qquad \hat{P}_i = \hat{\pi}(X_i)$$

- Probit/Logit or semiparametric estimation are also allowed
  - $D = 1\{W'\gamma \geq U\}$ where W contains all covariates in X and their interactions or higher-order terms

4. Estimation: second stage

- Second stage: estimation of selection model $E[Y_i|X_i, D_i = d] = X_i{'}\beta_d + g_d(P_i)$
- Recommendation: semiparametric estimation

$$\hat{\beta}_d = \left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\hat{\omega}_{dij}\left(X_i - X_j\right)\left(X_i - X_j\right)'\right]^{-1}\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\hat{\omega}_{dij}\left(X_i - X_j\right)\left(Y_i - Y_j\right)\right]$$

where $\hat{\omega}_{dij} = 1\left\{D_i = D_j = d\right\} k_2\left(\frac{\hat{P}_i - \hat{P}_j}{h_2}\right)$

$$\begin{pmatrix} \hat{g}_d(p) \\ \hat{g}_d^{(1)}(p) \end{pmatrix} = \left[\sum_{i=1}^{n}\hat{w}_{dpi}\begin{pmatrix} 1 \\ \hat{P}_i - p \end{pmatrix}\begin{pmatrix} 1 \\ \hat{P}_i - p \end{pmatrix}'\right]^{-1}\left[\sum_{i=1}^{n}\hat{w}_{dpi}\begin{pmatrix} 1 \\ \hat{P}_i - p \end{pmatrix}\left(Y_i - X_i'\hat{\beta}_d\right)\right]$$

- Parametric estimation if we are willing to parameterize g_d or $E[U_d|V]$
  - polynomial or normal polynomial: $E[U_d|V = v] = \sum_{r=1}^{R}\rho_{dr}\Phi^{-r}(v)$

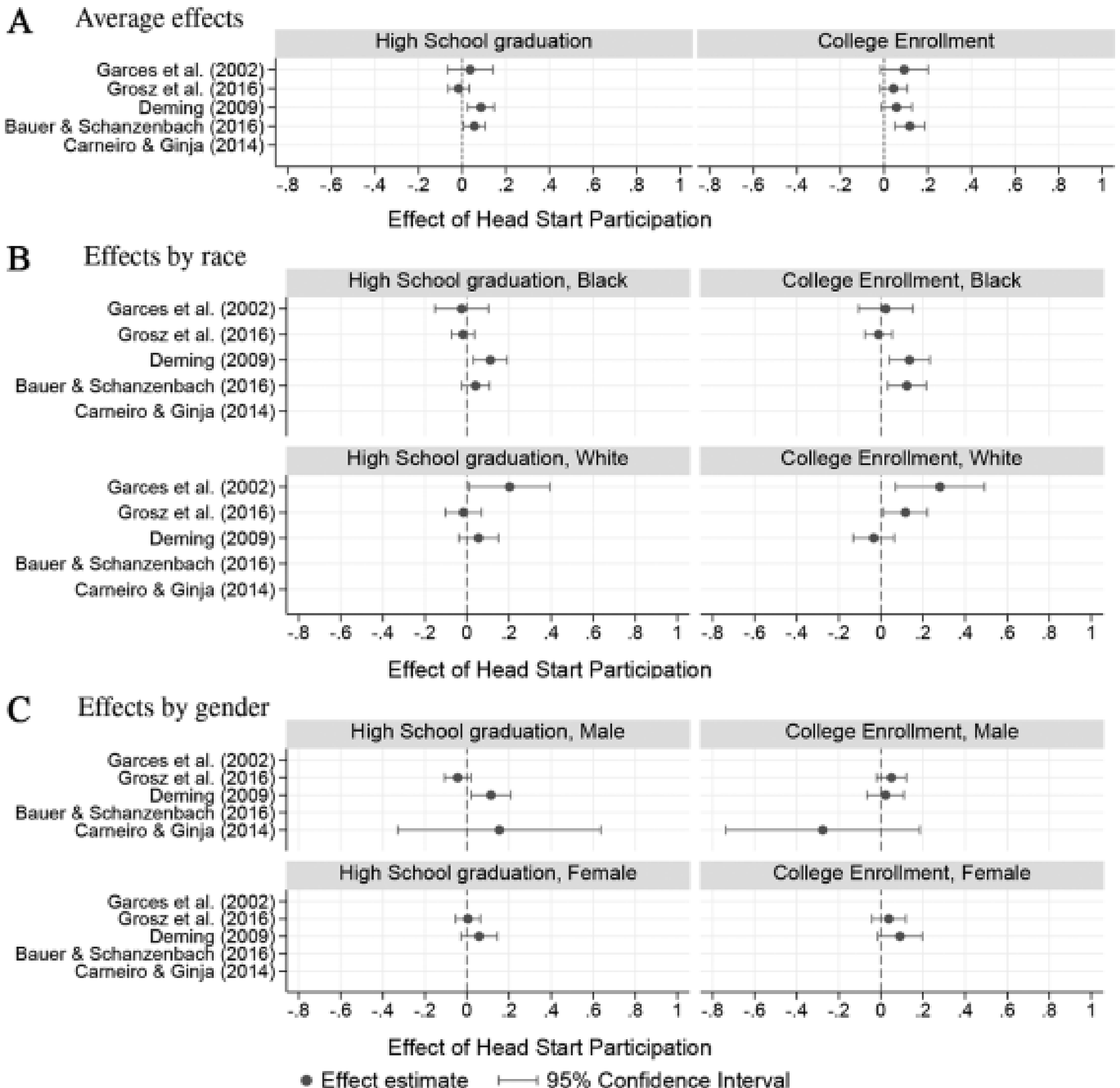
Contents

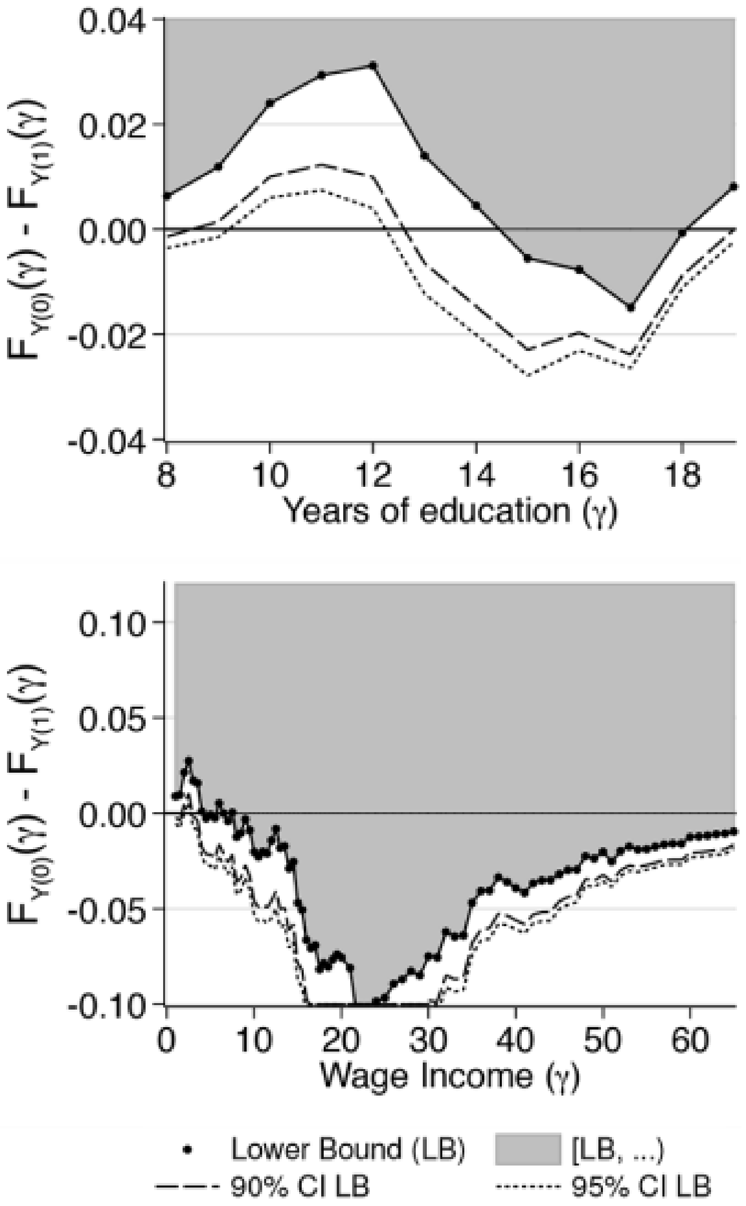
1. Introduction
2. Model
3. Identification
4. Estimation
5. Empirical application to Head Start
6. Conclusion

5. Application

- **Head Start** is a major federally funded preschool program in the US
  - targeted at children from low-income (below the poverty line) families
  - serving more than 1 million children at a cost of ＄10 billion in 2019
- Many studies show short-term positive effects on cognitive outcomes
- However, results on longer-term effects of Head Start are far from united
  - relatively more results on crime and health outcomes
  - less agreement regarding educational attainment and earnings
- De Haan and Leuven (2020, JoLE) attempts to fill this gap
  - National Longitudinal Study of Youth (NLSY) 1979
  - distributional treatment effects
  - partial identification without needing IV
- We revisit long-term effects of Head Start using De Haan and Leuven’s dataset.

5. Application: a review





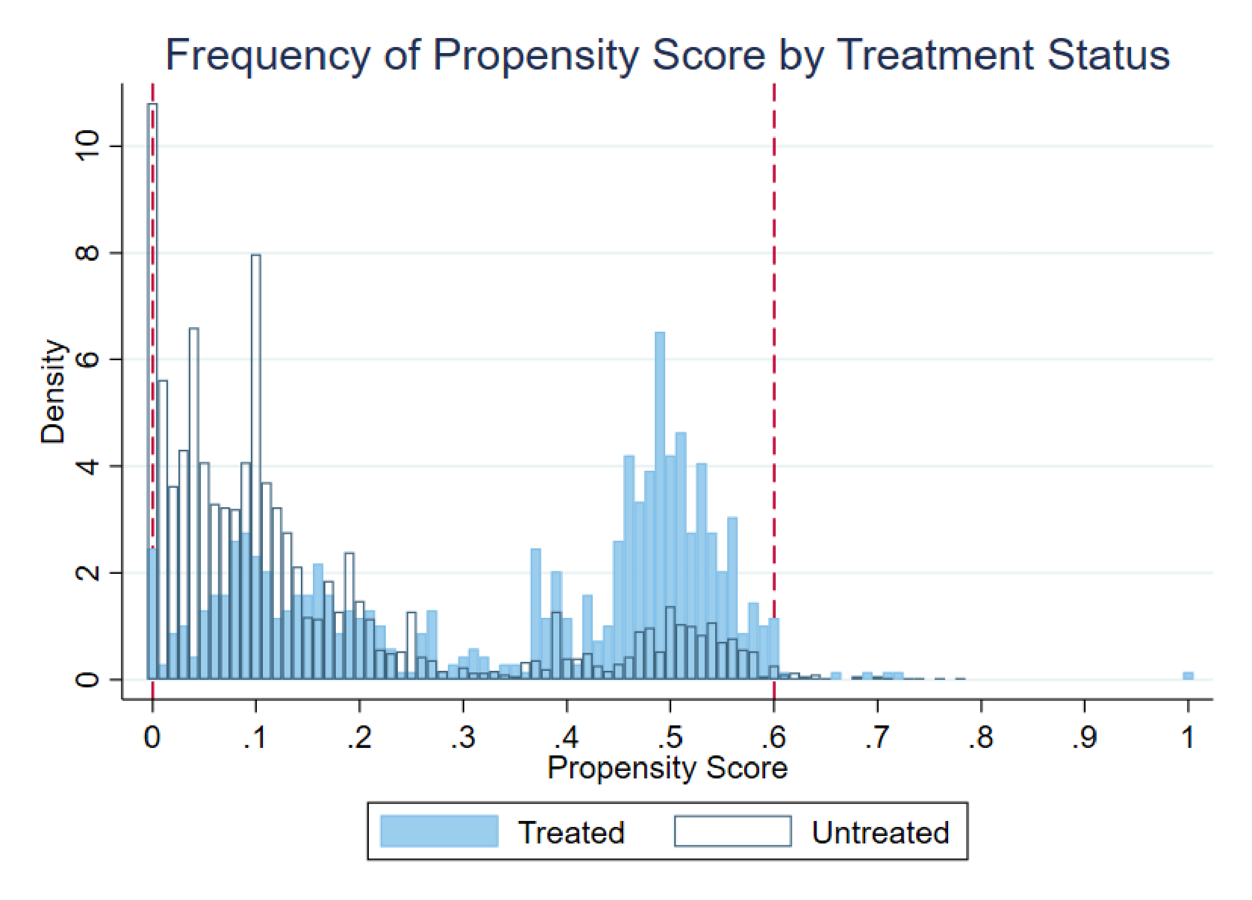
5. Application: data

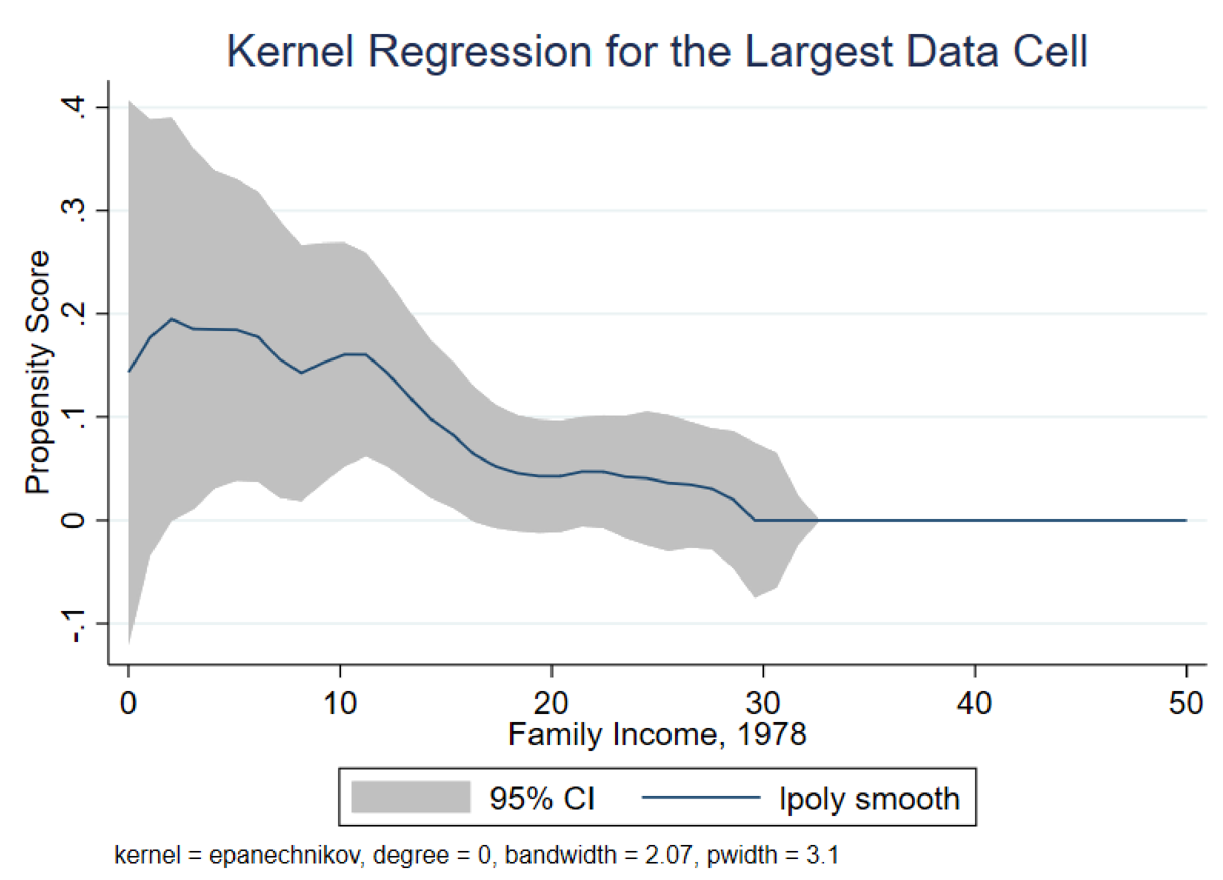
Table 1
Descriptive Statistics

		Head Start		Race		
	All	Yes	No	White	Black	Hispanic
Head Start	.23			.08	.49	.21
Age	32.1	32.0	32.1	32.1	32.1	32.0
Female	.50	.52	.50	.49	.51	.51
Race:						
White	.49	.16	.59			
Black	.31	.66	.21			
Hispanic	.20	.17	.20			
Parental education:						
Less than high school	.21	.26	.19	.10	.19	.50
Some high school	.15	.22	.13	.11	.25	.11
High school	.40	.38	.41	.47	.40	.24
College, 1–3 years	.12	.07	.13	.14	.09	.08
College, ≥4 years	.12	.07	.14	.18	.06	.06
Family income 1978	16,303	11,603	17,759	21,096	10,946	13,077
Years of education	12.8	12.6	12.8	13.1	12.6	12.1
Wage income	22,633	19,637	23,456	25,226	19,057	20,790
N	4,876	1,132	3,744	2,404	1,518	954

NOTE.—Sample sizes for wage income are 3,781 (all), 815 (Head Start yes), 2,966 (Head Start no), 1,985 (white), 1,060 (black), and 736 (Hispanic).

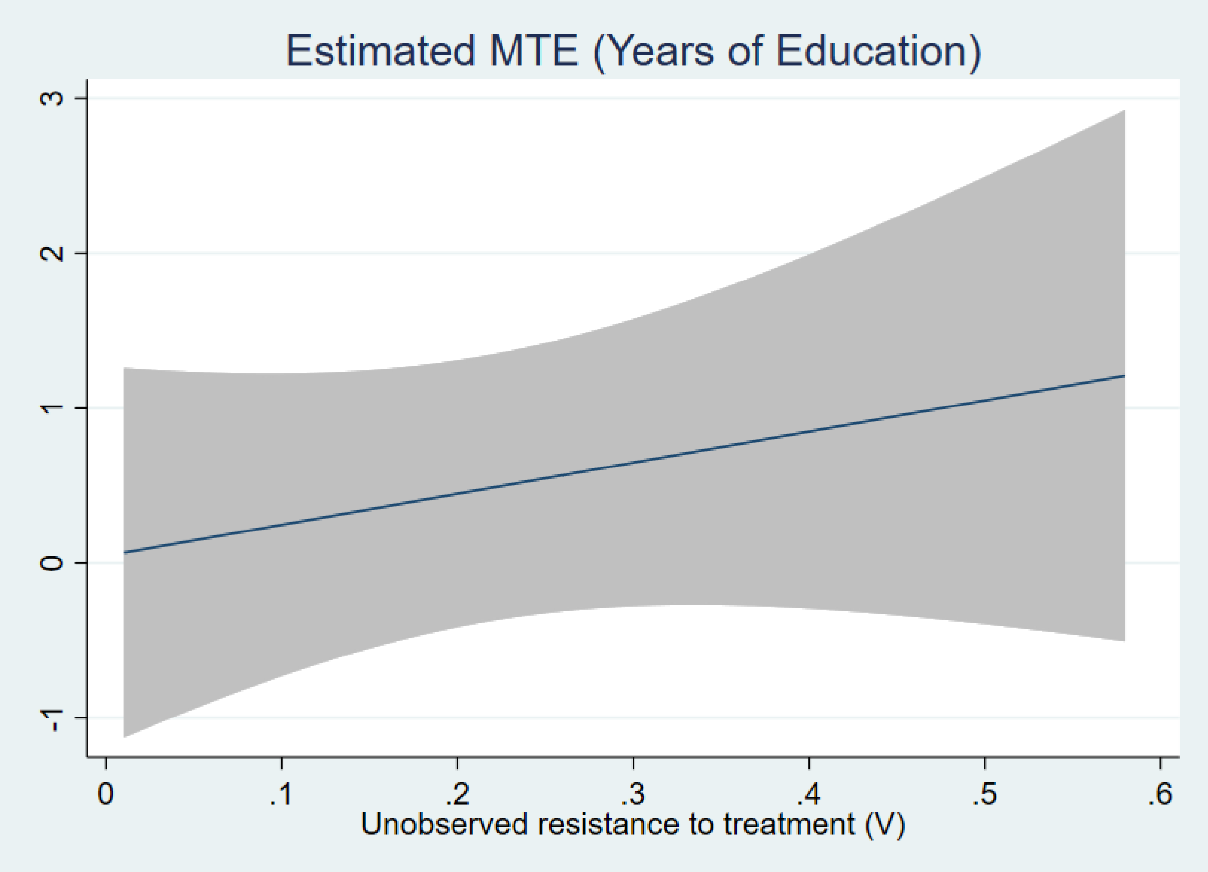
5. Application: first stage

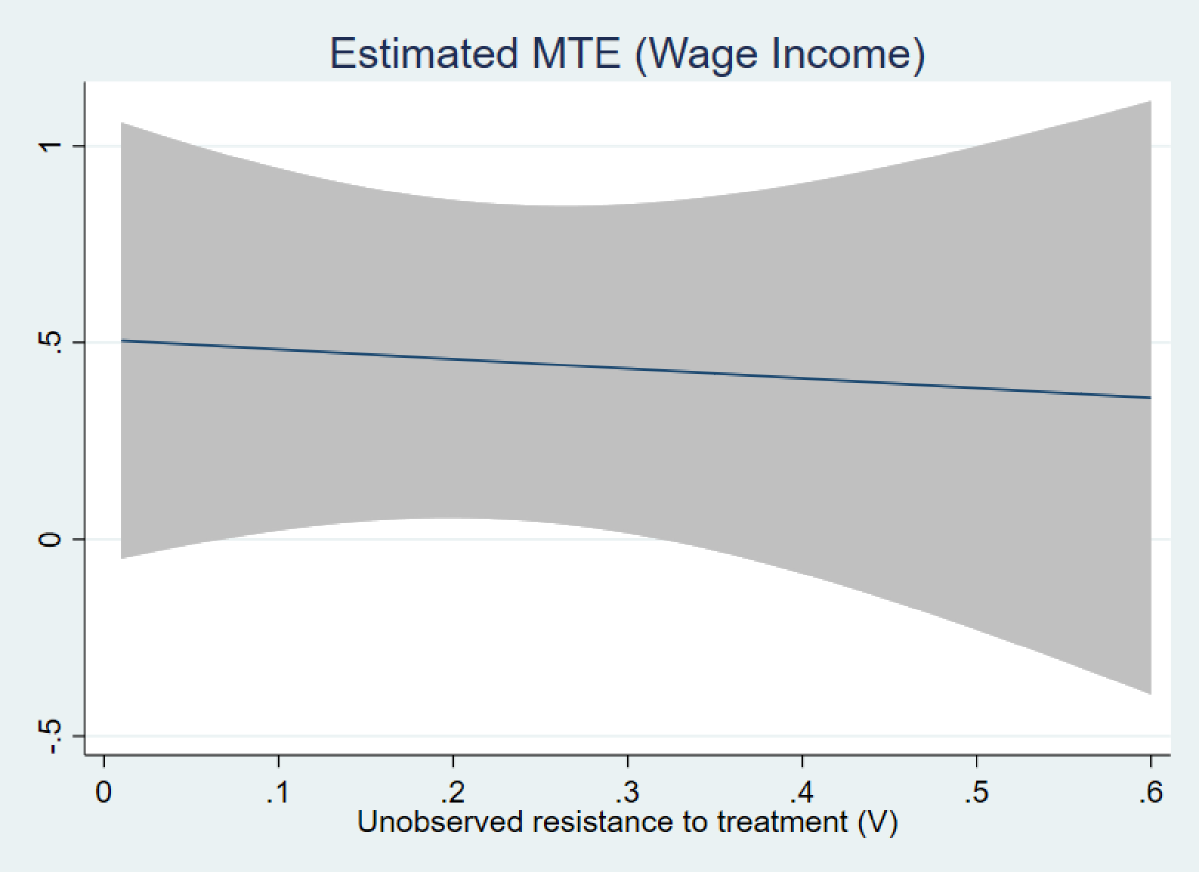




5. Application: parametric second stage

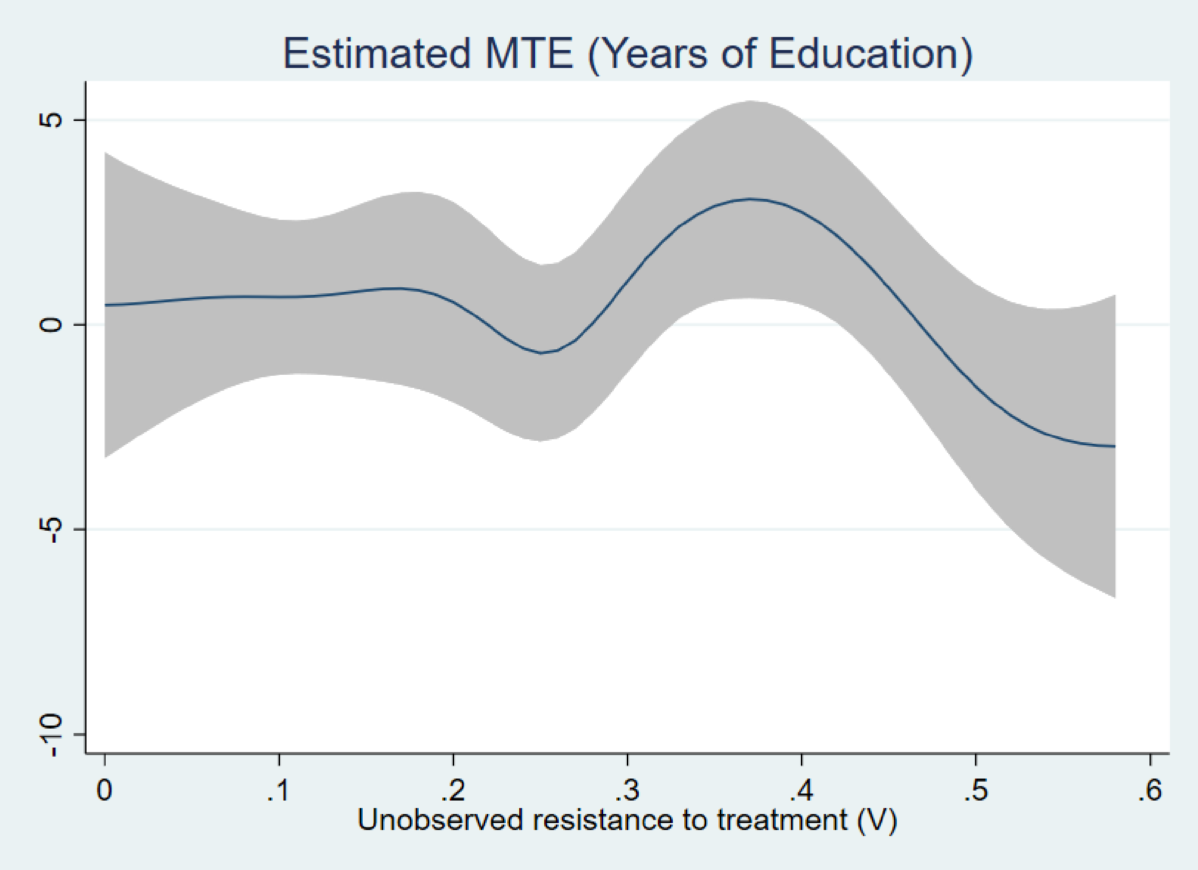
- Linear specification: $E[U_d|V = v] = \theta_{d0} + \theta_{d1}v$
- Bootstrapped confidence interval with 1000 replications

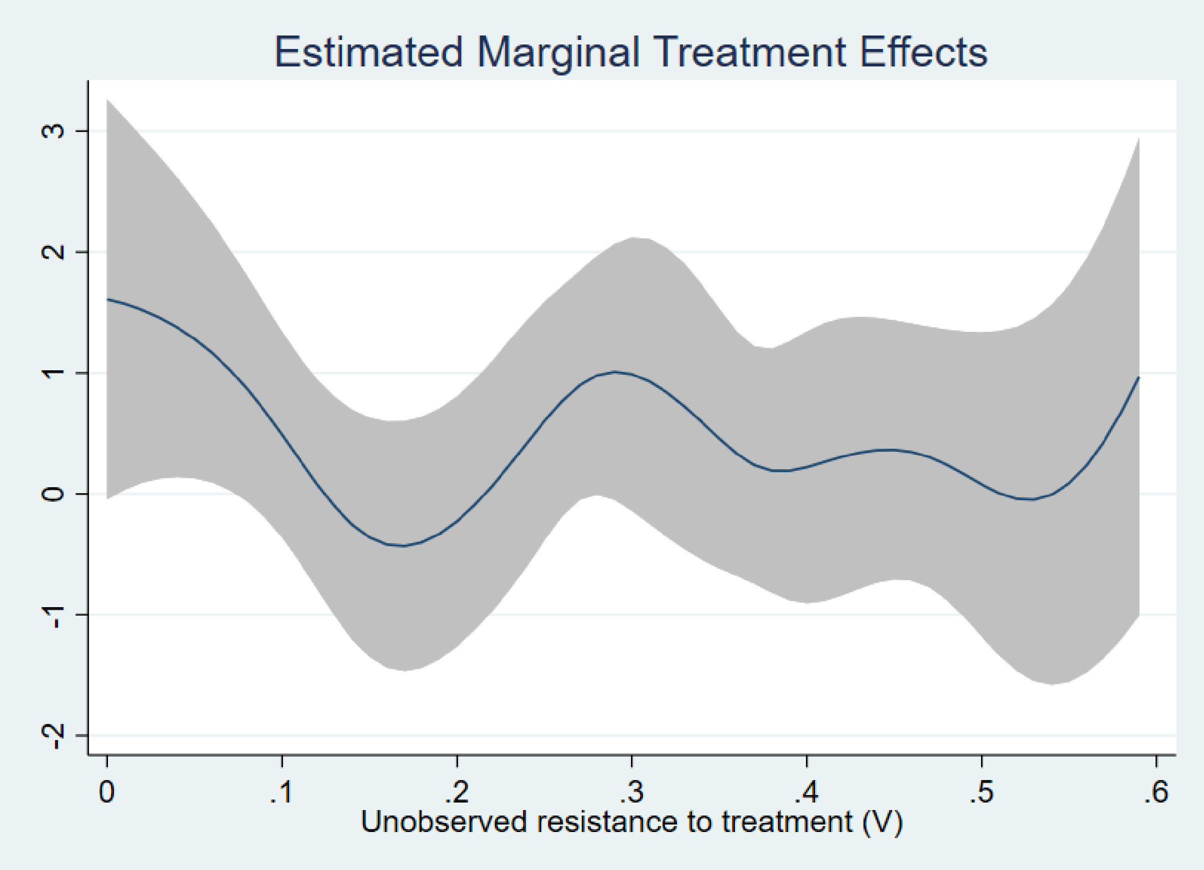




5. Application: semiparametric second stage

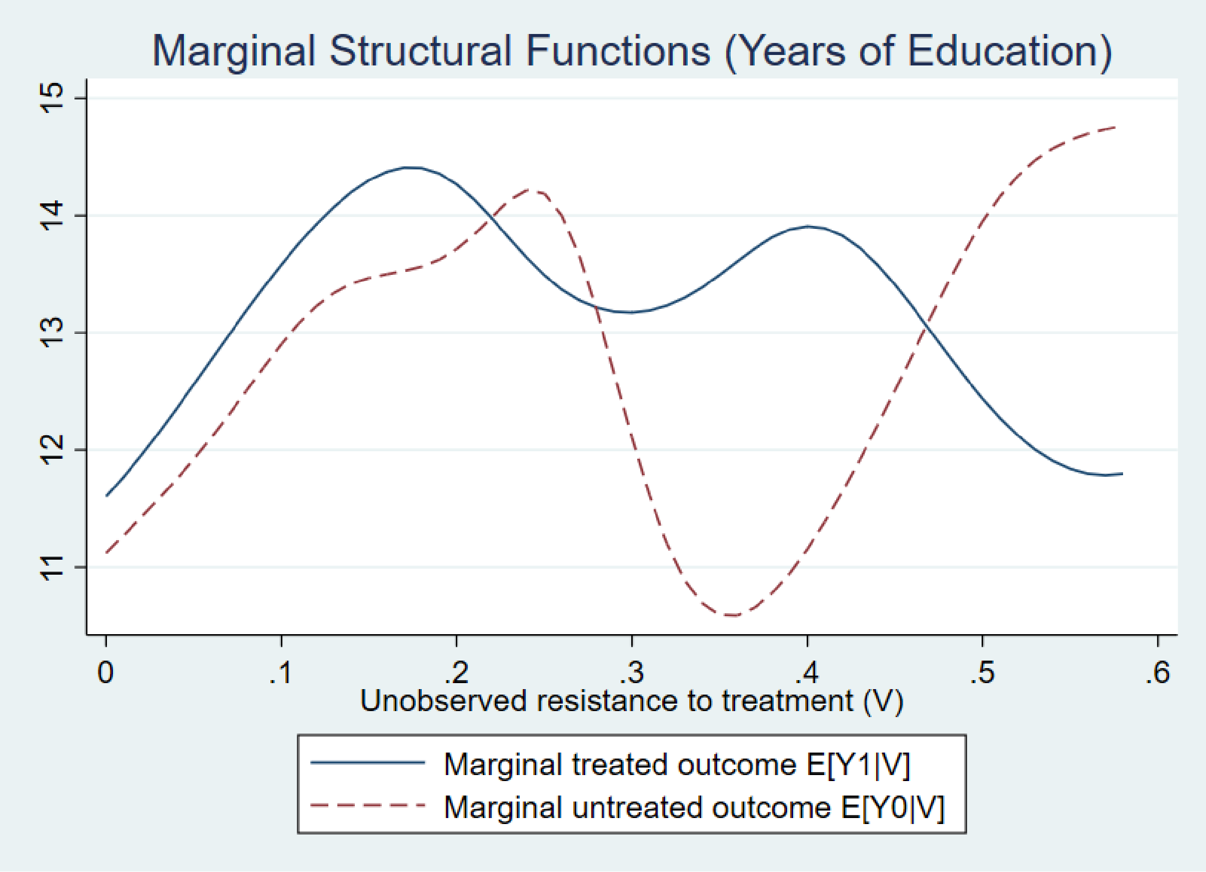
- $E[U_d|V = v]$ is nonparametrically specified
- Bootstrapped confidence interval with 1000 replications

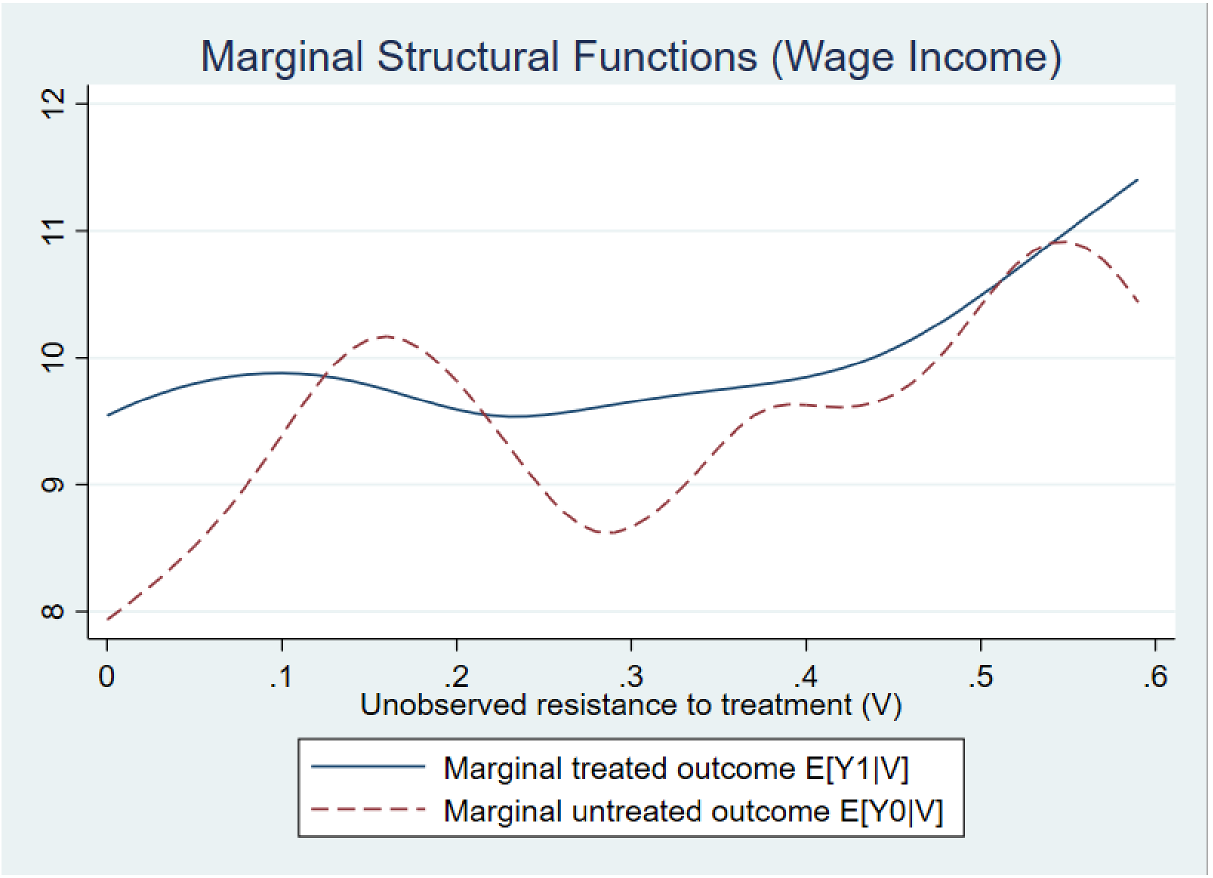




5. Application: counterfactual

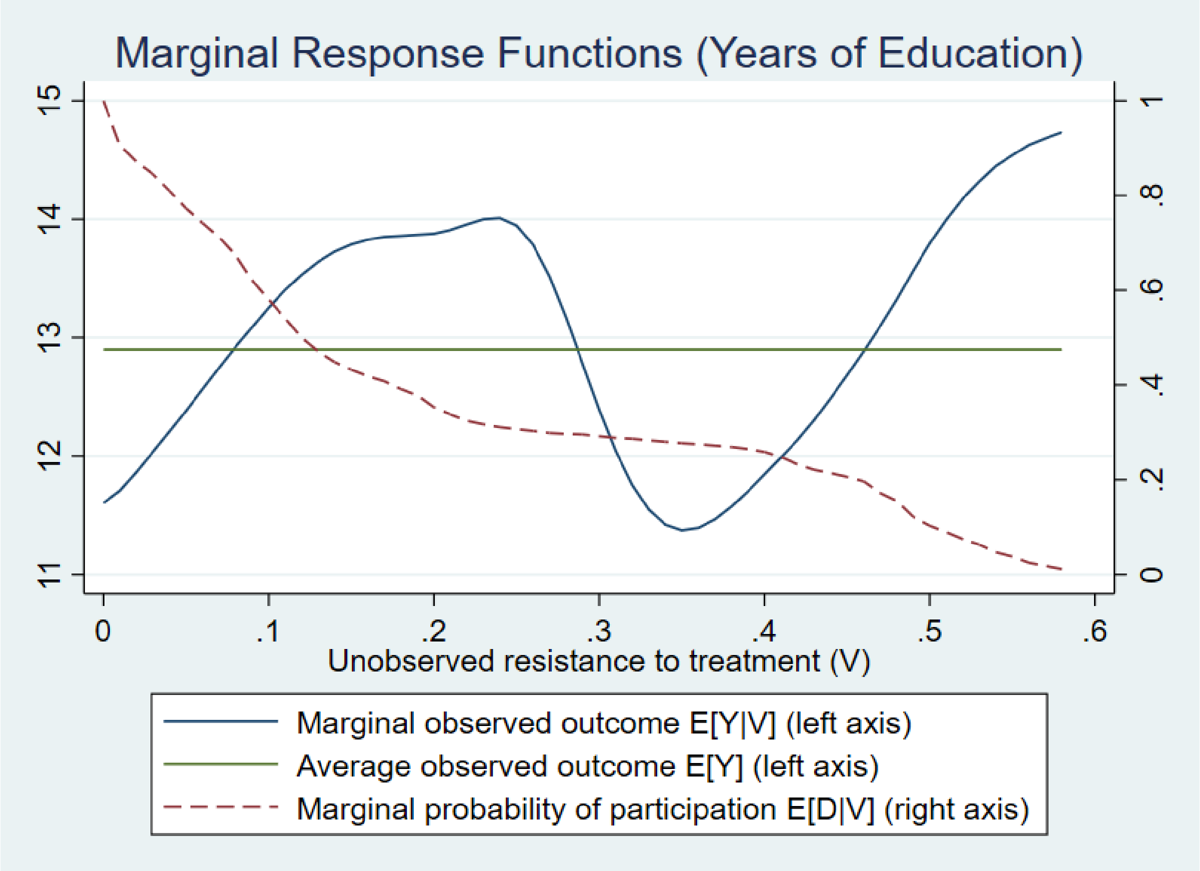
- $E[Y_d|X = x, V = v] = x'\beta_d + E[U_d|V = v]$
- $E[Y_d|V = v] = (EX)'\beta_d + E[U_d|V = v]$

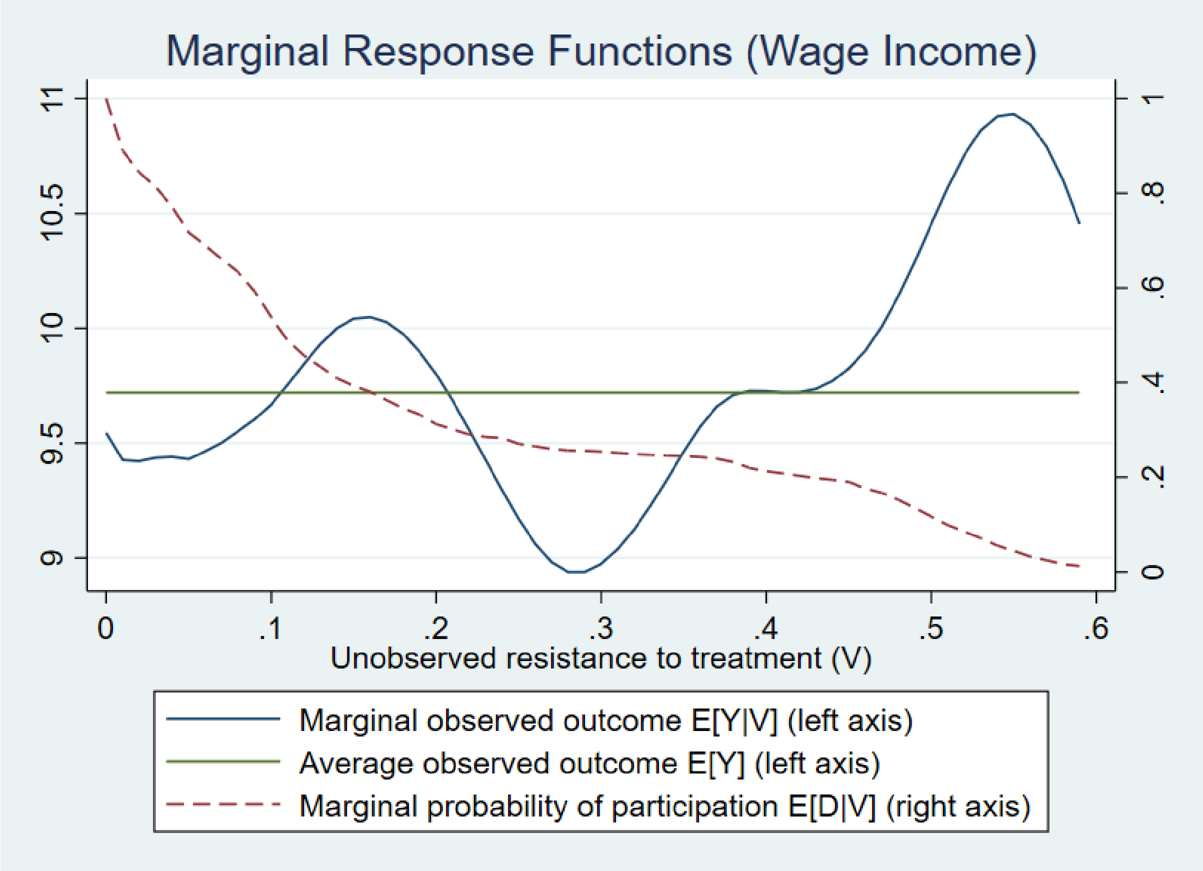




5. Application: interpretation

- $E[Y|V = v] = E[Y_0|V = v] + E[Y_1 - Y_0|D = 1, V = v] \cdot E[D|V = v]$





Contents

1. Introduction
2. Model
3. Identification
4. Estimation
5. Empirical application to Head Start
6. Conclusion

6. Conclusion

- We propose an IV-free model for MTE that nests most IV models for MTE
- We give a set of sufficient conditions that guarantees identification of MTE without IV
  - based on the method of identification by functional form
- We provides an empirical application to illustrate the usefulness of our method

Thank you for your listening

My e-mail: panzhew@zufe.edu.cn