

Online Appendix for “Robustify and Tighten the Lee Bounds: A Sample Selection Model under Stochastic Monotonicity and Symmetry Assumptions”

YUTA OKAMOTO[†]

[†] *Graduate School of Economics, Kyoto University*
E-mail: okamoto.yuuta.57w@st.kyoto-u.ac.jp

Summary Section S1 provides omitted theoretical results (estimation/inference procedure without symmetry assumption, discussion on the sharpness of the bounds under symmetry, covariate adjustment, and proofs). Numerical simulation results are reported in Section S2. Section S3 is for additional results about empirical illustrations. Section S4 applies the stochastic monotonicity assumption to the setting of estimating bounds on the marginal treatment effect in the presence of sample selection.

S1. OMITTED RESULTS AND PROOFS

S1.1. Estimation and Inference without Symmetry

Assume a researcher knows that $\vartheta_L \geq \vartheta_F$ but hesitates to impose the symmetry. In such a case, one can estimate the lower bound based on

$$h(\beta) = \begin{pmatrix} (Y_i - \mu)S_i D_i \mathbf{1}\{Y_i \leq \beta^L\} \\ (\mathbf{1}\{Y_i > \beta^L\} - (1 - \vartheta_L q)) S_i D_i \\ (qS_i - \alpha) D_i \\ (S_i - \alpha)(1 - D_i) \\ (Y_i - \eta)S_i(1 - D_i) \end{pmatrix}$$

with an additional parameter μ and $\mu_0 = \mathbb{E}[Y_i | D_i = 1, S_i = 1, Y_i \leq y_{\vartheta_L q_0}^1]$, instead of $g(\beta)$. The estimator is given by $\widehat{\Delta}_1 = \widehat{\mu} - \widehat{\eta}$. $\widehat{\nabla}_1$ is defined similarly.

COROLLARY S.1.1. *Under the same assumptions in Corollary 4.1, $\sqrt{n}(\widehat{\Delta}_1 - \Delta) \rightarrow_d \mathcal{N}(0, \tilde{\Omega}_L + \Omega_C)$ and $\sqrt{n}(\widehat{\nabla}_1 - \nabla) \rightarrow_d \mathcal{N}(0, \tilde{\Omega}_U + \Omega_C)$, where*

$$\begin{aligned} \tilde{\Omega}_L &= \frac{\mathbb{V}[Y_i | D_i = 1, S_i = 1, Y_i \leq y_{\vartheta_L q_0}^1] + (y_{\vartheta_L q_0}^1 - \mu_L)^2 (1 - \vartheta_L q_0)}{\mathbb{P}[S_i = 1, D_i = 1] \vartheta_L q_0} + (y_{\vartheta_L q_0}^1 - \mu_L)^2 \Omega_Q, \\ \tilde{\Omega}_U &= \frac{\mathbb{V}[Y_i | D_i = 1, S_i = 1, Y_i \geq y_{1-\vartheta_L q_0}^1] + (y_{1-\vartheta_L q_0}^1 - \mu_U)^2 (1 - \vartheta_L q_0)}{\mathbb{P}[S_i = 1, D_i = 1] \vartheta_L q_0} + (y_{1-\vartheta_L q_0}^1 - \mu_U)^2 \Omega_Q, \\ \Omega_Q &= \frac{1 - \alpha_0/q_0}{\mathbb{P}[D_i = 1] \alpha_0/q_0} + \frac{1 - \alpha_0}{\alpha_0 \mathbb{P}[D_i = 0]}, \end{aligned}$$

$\mu_L = \mathbb{E}[Y_i | D_i = 1, S_i = 1, Y_i \leq y_{\vartheta_L q_0}^1]$, $\mu_U = \mathbb{E}[Y_i | D_i = 1, S_i = 1, Y_i \geq y_{1-\vartheta_L q_0}^1]$, and Ω_C is defined in Corollary 4.1.

REMARK S.1.1. The asymptotic variances with $\vartheta_L = 1$ coincide with those in Lee (2009, Proposition 3).

When a researcher does not know that $\vartheta_L \geq \vartheta_F$ and hesitates to impose the symmetry. We shall replace $\tilde{g}$ with

$$\tilde{h}(\gamma) = \begin{pmatrix} (Y_i - \mu^L) S_i D_i \mathbf{1}\{Y_i \leq \gamma^L\} \\ (Y_i - \mu^F) S_i D_i \mathbf{1}\{Y_i \leq \gamma^F\} \\ (\mathbf{1}\{Y_i > \gamma^L\} - (1 - \vartheta_L q)) S_i D_i \\ (\mathbf{1}\{Y_i > \gamma^F\} - (1 - (1 + q(1 - 1/\alpha)))) S_i D_i \\ (q S_i - \alpha) D_i \\ (S_i - \alpha)(1 - D_i) \\ (Y_i - \eta) S_i (1 - D_i) \end{pmatrix},$$

where μ^L, μ^F are additional parameters and

$$\widehat{\Delta}_2 = \max_{v \in \{L, F\}} (\hat{\mu}^v - \hat{\eta}).$$

Inference can be performed similarly to that we considered in the main text based on Chernozhukov et al. (2013).

S1.2. On Sharpness of the Bounds

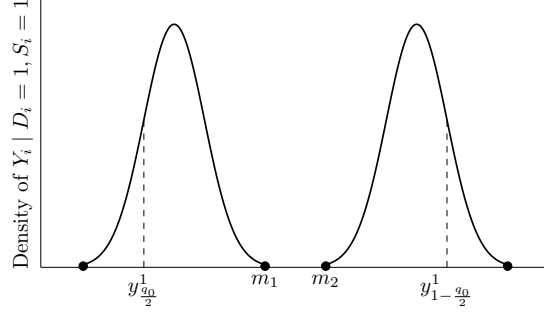
The bounds obtained in Theorem 3.1 in the main text may not be sharp under the stated assumptions, in the sense that there can possibly be an $m \in (\Delta^s, \nabla^s)$ such that there is *no* joint distribution of $(Y_{1i}^*, Y_{0i}^*, S_{1i}, S_{0i}, D_i)$ that is consistent with the assumptions, induce the distribution of observed data, and satisfy $\tau = m$. In other words, some regions or points in (Δ^s, ∇^s) may be inconsistent with the data and assumptions, and the resulting sharp bounds could be a set of subintervals of $[\Delta^s, \nabla^s]$. This can be observed in the following example:

EXAMPLE S.1.1. (NON-SHARP BOUNDS (1)) Assume the density of Y_i conditional on $D_i = 1, S_i = 1$ is given by Figure S1, wherein the support of the density is a set consisting of two disjoint intervals. Note that Assumption 3.2 in the main text holds. In this case, there is no joint distribution of $(Y_{1i}^*, Y_{0i}^*, S_{1i}, S_{0i}, D_i)$ that rationalises both $\text{Median}[Y_{1i}^* | S_{0i} = S_{1i} = 1] \in (m_1, m_2)$ and Assumption 3.1 (Mean-Median Coincidence) in the main text. Put differently, under this setup and Assumption 3.1, it holds that $\mathbb{E}[Y_{1i}^* | S_{0i} = S_{1i} = 1] \notin (m_1, m_2)$. This, in turn, implies that

$$\tau \notin (m_1 - \mathbb{E}[Y_i | D_i = 0, S_i = 1], m_2 - \mathbb{E}[Y_i | D_i = 0, S_i = 1]).$$

Hence, the bounds $[\Delta^s, \nabla^s]$ are not sharp in that the bounds include rejectable points.

Notice that even in such a case, the endpoints of the bounds, Δ^s and ∇^s , cannot be rejected as the true value of τ . Therefore, the practical implication that “the true effect is greater than Δ^s and less than ∇^s , or equal to these values” remains intact even when the bounds are not sharp.

**Figure S1:** Example: Non-Sharp Bounds (1)

However, making an assumption that avoids the (perhaps somewhat extreme) cases like Example S.1.1 can ensure the sharpness of the bounds. For this aim, define

$$\mu_l := \int_{-\infty}^{y_{\frac{\vartheta q_0}{2}}^1} z \frac{f_y^1(z)}{\vartheta q_0/2} dz, \quad \mu_r := \int_{y_{1-\frac{\vartheta q_0}{2}}^1}^{\infty} z \frac{f_y^1(z)}{\vartheta q_0/2} dz,$$

and

$$\tilde{\mu}_l(m) := \int_{l(m)}^m z \frac{f_y^1(z)}{\vartheta q_0/2} dz, \quad \tilde{\mu}_r(m) := \int_m^{r(m)} z \frac{f_y^1(z)}{\vartheta q_0/2} dz,$$

where $l(m)$ and $r(m)$ satisfy

$$\int_{l(m)}^m f_y^1(z) dz = \int_m^{r(m)} f_y^1(z) dz = \frac{\vartheta q_0}{2}. \quad (\text{S.1})$$

Under Assumption S.1.1 below, there exists $l(m)$ and $r(m)$ satisfying (S.1). When $l(m)$ and $r(m)$ are not uniquely defined, one can arbitrarily define them so that (S.1) holds. Intuitively, μ_l and $\tilde{\mu}_l$ are the trimmed means using the data from the lower tail and the left near side of m . μ_r and $\tilde{\mu}_r$ are similar. Now, we consider the following assumption:

ASSUMPTION S.1.1. f_y^1 is continuous on its support. For any $m \in (y_{\frac{\vartheta q_0}{2}}^1, y_{1-\frac{\vartheta q_0}{2}}^1)$,

$$\frac{\mu_l + \tilde{\mu}_r(m)}{2} < m \leq \frac{\tilde{\mu}_l(m) + \mu_r}{2} \quad (\text{S.2})$$

and $f_y^1(m) > 0$.

Continuity and positivity are not very strong requirements. An important part will be (S.2), although it may seem less intuitive at first glance. Therefore, we provide an example where the assumption is violated.

EXAMPLE S.1.2. (NON-SHARP BOUNDS (2)) Assume the density of Y_i conditional on $D_i = 1, S_i = 1$ is given by Figure S2. Consider if $m - \mathbb{E}[Y_i | D_i = 0, S_i = 1]$ is attainable.

Firstly, observe that

$$\underbrace{\frac{\mu_l + \tilde{\mu}_r(m)}{2}}_{\text{hollow circle in Figure S2}} > m,$$

that is, (S.2) does not hold. Here, note that, in order for $\text{Median}[Y_{1i}^* | S_{0i} = S_{1i} = 1] = m$ to be satisfied, it should at least hold that $q_0/2$ values are below m and the other $q_0/2$ values are above m . Then we have $\mathbb{E}[Y_{1i}^* | S_{0i} = S_{1i} = 1] > (\mu_l + \tilde{\mu}_r(m))/2$. Hence, combining these two inequalities, this m is rejected as a candidate of $\mathbb{E}[Y_{1i}^* | S_{0i} = S_{1i} = 1]$, leading to $\tau \neq m - \mathbb{E}[Y_i | D_i = 0, S_i = 1]$. Therefore, the bounds are not sharp.

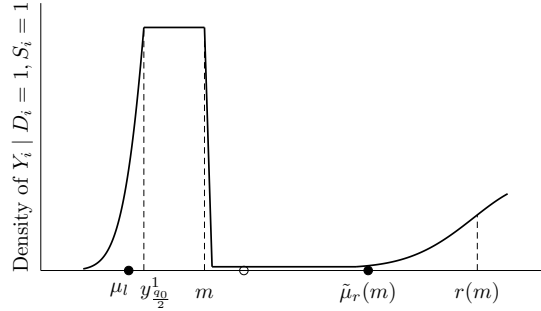


Figure S2: Example: Non-Sharp Bounds (2)

Intuitively speaking, (S.2) in Assumption S.1.1 requires that the conditional density f_y^1 is not “far separated” as in Figures S1 and S2. When f_y^1 is smoothly decaying unimodal density or is concentrated in a certain “middle” point, (S.2) would be satisfied. The density function of the Hindi test scores in our empirical example shows such a functional form (Figure 2 in the main text). Now, we give the next corollary. The proof is given in Section S1.4.

COROLLARY S.1.2. *Suppose the assumptions stated in Theorem 3.1 in the main text and Assumption S.1.1 hold. Then the bounds obtained in Theorem 3.1 (i.e., $[\Delta^s, \nabla^s]$) are sharp.*

S1.3. Covariates

We can use predetermined covariates W_i in the same manner as Lee (2009, p. 1086) to obtain possibly improved bounds. Instead of Assumption 2.1, assume

$$(Y_{1i}^*, Y_{0i}^*, S_{1i}, S_{0i}, W_i) \perp\!\!\!\perp D_i.$$

As for the stochastic monotonicity, in addition to Assumption 2.3-(i), the covariates should also satisfy

$$(W_i, Y_{1i}^*, Y_{0i}^*) \perp\!\!\!\perp S_{1i} | S_{0i} = 1,$$

which respects the stochastic nature of the decision flips. This is automatically satisfied under Lee's (2009) strict monotonicity. The mean-median coincidence is now

$$\text{Median}[Y_{1i}^*|S_{0i} = S_{1i} = 1, W_i] = \mathbb{E}[Y_{1i}^*|S_{0i} = S_{1i} = 1, W_i]. \quad (\text{S.3})$$

With such W_i , we first compute the same lower bound conditional on $W_i = w$, written by $\Delta(w)$ or $\Delta^s(w)$, and then we obtain (weakly) narrower bounds by computing

$$\begin{aligned} & \int \Delta(w) f_w(w|S_i = 1, D_i = 0) dw, \\ & \int \Delta^s(w) f_w(w|S_i = 1, D_i = 0) dw, \end{aligned}$$

where f_w is the conditional density of the covariates.

As a candidate of W_i , variables with predictive power for selection will be a good choice to narrow the bounds. Specifically, the variables that produce a difference in $q_0(w)$, defined as $q_0(w) = \mathbb{P}[S_i = 1|D_i = 0, W_i = w] / \mathbb{P}[S_i = 1|D_i = 1, W_i = w]$, would help tighten the bounds. When W_i is large-dimensional, discretisation of W_i via regression (Lee, 2009, Section 5.2) or clustering (Caetano et al., 2024, Remark 5.1) may be useful. The upper bounds can be treated similarly.

REMARK S.1.2. (PARTIAL MEAN-MEDIAN COINCIDENCE) *Let $W_i \in \{0, 1\}$ for simplicity. In some cases, the conditional mean-median coincidence (S.3) seems to hold only for those with $W_i = 1$. In such a case, we can combine the two bounds obtained in Theorems 2.1 and 3.1. Namely, we first compute $\Delta(0)$ and $\Delta^s(1)$ and then obtain the lower bound as follows:*

$$\Delta(0)\mathbb{P}[W_i = 0|S_i = 1, D_i = 0] + \Delta^s(1)\mathbb{P}[W_i = 1|S_i = 1, D_i = 0].$$

S1.4. Proofs

Proof of Corollary 4.1: Assuming the parameter space for q be $[0, 1/\vartheta_L]$, the consistency, $\hat{\beta} \rightarrow_p \beta_0$, holds under the boundedness assumption of the support of Y_i and $\mathbb{P}[S_i = 1|D_i = 0] > 0$. This follows from a similar argument to Proposition 2 of Lee (2009), utilising Theorem 2.6 from Newey and McFadden (1994). Define $g_0(\beta) = \mathbb{E}[g(\beta)]$ and $\hat{g}_n(\beta) = n^{-1} \sum_{i=1}^n g(\beta)$. Let G_β denote the derivative of g_0 at $\beta = \beta_0$ and Σ_β be the asymptotic variance of $\hat{g}_n(\beta_0)$. By Newey and McFadden (1994, Theorem 7.2) and Andrews (1994, Theorems 1–3), under the stated assumptions, we have the asymptotic normality of $\hat{\beta}$ with the asymptotic variance $G_\beta^{-1} \Sigma_\beta (G'_\beta)^{-1}$. Here, we can compute

$$G_\beta = \begin{pmatrix} -f_y^1 \left(y_{\frac{\vartheta_L q_0}{2}}^1 \right) \mathbb{E}[S_i D_i] & \frac{\vartheta_L}{2} \mathbb{E}[S_i D_i] & 0 & 0 \\ 0 & \mathbb{E}[S_i D_i] & -\mathbb{P}[D_i = 1] & 0 \\ 0 & 0 & -\mathbb{P}[D_i = 0] & 0 \\ 0 & 0 & 0 & -\mathbb{E}[S_i(1 - D_i)] \end{pmatrix},$$

and $\Sigma_\beta = \text{diag}(\Sigma_1, \Sigma_3, \Sigma_4, \Sigma_5)$, where these elements are defined in the proof of Lemma 4.1 below. A straightforward calculation yields that

$$G_\beta^{-1} = \left\{ f_y^1 \left(y_{\frac{\vartheta_L q_0}{2}}^1 \right) \right\}^{-1} \frac{1}{\mathbb{P}[S_i = 1, D_i = 1]} \begin{pmatrix} -1 & \frac{\vartheta_L}{2} & -\frac{\vartheta_L}{2} \frac{\mathbb{P}[D_i = 1]}{\mathbb{P}[D_i = 0]} & 0 \\ 0 & C_1 & C_2 & 0 \\ 0 & 0 & C_3 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + M_\eta^{-1},$$

where C_j are some constants and M_η^{-1} is a 4×4 matrix where the bottom right element is $-1/\mathbb{E}[S_i(1 - D_i)]$ and all others are 0. Then, the asymptotic variance of $\hat{\beta}^L - \hat{\eta}$ is given by

$$\left\{ f_y^1 \left(y_{\frac{\vartheta_L q_0}{2}}^1 \right) \right\}^2 \frac{1}{\mathbb{P}[S_i = 1, D_i = 1]^2} \left\{ \Sigma_1 + \frac{\vartheta_L^2}{4} \Sigma_3 + \frac{\vartheta_L^2}{4} \frac{\mathbb{P}[D_i = 1]^2}{\mathbb{P}[D_i = 0]^2} \Sigma_4 \right\} + \Omega_C.$$

The asymptotic variance of the upper bound is similar. $\square$

Proof of Lemma 4.1: The asymptotic variance is given by $\Omega_{\gamma, L} = G_{\gamma, L}^{-1} \Sigma_\gamma (G'_{\gamma, L})^{-1}$; $G_{\gamma, L}$ is given by

$$\begin{pmatrix} -f_y^1(\gamma_0^L) \mathbb{E}[S_i D_i] & 0 & \frac{\vartheta_L}{2} \mathbb{E}[S_i D_i] & 0 & 0 \\ 0 & -f_y^1(\gamma_0^F) \mathbb{E}[S_i D_i] & \frac{1 - 1/\alpha_0}{2} \mathbb{E}[S_i D_i] & \frac{q_0}{2\alpha_0^2} \mathbb{E}[S_i D_i] & 0 \\ 0 & 0 & \mathbb{E}[S_i D_i] & -\mathbb{E}[D_i] & 0 \\ 0 & 0 & 0 & -\mathbb{E}[1 - D_i] & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + M_\eta;$$

M_η is a 5×5 matrix where the bottom right element is $-\mathbb{E}[S_i(1 - D_i)]$ and all others are 0; and $\Sigma_\gamma = \text{diag}(\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4, \Sigma_5)$, where

$$\Sigma_1 = \frac{\vartheta_L q_0}{2} \left(1 - \frac{\vartheta_L q_0}{2} \right) \mathbb{P}[S_i = 1, D_i = 1], \quad \Sigma_2 = \frac{\vartheta_F q_0}{2} \left(1 - \frac{\vartheta_F q_0}{2} \right) \mathbb{P}[S_i = 1, D_i = 1],$$

$$\Sigma_3 = \alpha_0 (q_0 - \alpha_0) \mathbb{P}[D_i = 1], \quad \Sigma_4 = \alpha_0 (1 - \alpha_0) \mathbb{P}[D_i = 0],$$

and $\Sigma_5 = \mathbb{V}[Y_i | S_i = 1, D_i = 0] \mathbb{E}[S_i(1 - D_i)]$. The proof is similar to that of Corollary 4.1 above. $\square$

Proof of Corollary S.1.2: Under Assumptions 2.1, 2.3, 3.1, 3.2, and S.1.1, the endpoints of the bounds are not rejected. Hence, it suffices to show that $m - \mathbb{E}[Y_i | D_i = 0, S_i = 1]$ is attainable for any $m \in (y_{\vartheta q_0/2}^1, y_{1-\vartheta q_0/2}^1)$. Fix m . Recall the probabilities and conditional probabilities defined in the proof of Theorem 3.1 (in the main text), except for π_{11}^* .

Define

$$\pi_{1|1}^* := \begin{cases} (1 - \alpha_1) \frac{1}{\vartheta q_0} \mathbb{P} [\tilde{Y}_{1i}^* \leq y | \tilde{S}_{1i} = 1] & \text{if } y \leq y_{\frac{\vartheta q_0}{2}}^1 \\ (1 - \alpha_1) \frac{1}{2} + \alpha_1 \frac{1}{\vartheta q_0} \mathbb{P} [l(m) \leq \tilde{Y}_{1i}^* \leq y] & \text{if } y_{\frac{\vartheta q_0}{2}}^1 < y \leq m \\ \frac{1}{2} + \alpha_2 \frac{1}{\vartheta q_0} \mathbb{P} [m \leq \tilde{Y}_{1i}^* \leq y] & \text{if } m < y \leq r(m) \\ \frac{1}{2} + \frac{\alpha_2}{2} + (1 - \alpha_2) \frac{1}{\vartheta q_0} \mathbb{P} [y_{1 - \frac{\vartheta q_0}{2}}^1 \leq \tilde{Y}_{1i}^* \leq y] & \text{if } r(m) < y \end{cases},$$

where $\alpha_1 \in (0, 1]$ and $\alpha_2 \in [0, 1]$. Then, it holds that $\text{Median}[\tilde{Y}_{1i}^* | \tilde{S}_{0i} = \tilde{S}_{1i} = 1] = m$ under the latter half of Assumption S.1.1. Plus, by Assumption S.1.1,

$$\begin{cases} \tilde{\mu}_r(m) - m \in [m - \tilde{\mu}_l(m), m - \mu_l] & \text{if } m - \tilde{\mu}_l(m) \leq \tilde{\mu}_r(m) - m \\ m - \tilde{\mu}_l(m) \in (\tilde{\mu}_r(m) - m, \mu_r - m] & \text{if } m - \tilde{\mu}_l(m) > \tilde{\mu}_r(m) - m \end{cases},$$

and hence, noting that

$$\mathbb{E} [\tilde{Y}_{1i}^* | \tilde{S}_{0i} = \tilde{S}_{1i} = 1] = \frac{1}{2} \{ (1 - \alpha_1) \mu_l + \alpha_1 \tilde{\mu}_l(m) \} + \frac{1}{2} \{ \alpha_2 \tilde{\mu}_r(m) + (1 - \alpha_2) \mu_r \},$$

we can obtain $\mathbb{E} [\tilde{Y}_{1i}^* | \tilde{S}_{0i} = \tilde{S}_{1i} = 1] = m$ with an appropriate choice of α_1, α_2 . Then, by a similar argument to that in the proof of Theorem 3.1, we can show that $m - \mathbb{E} [Y_i | D_i = 0, S_i = 1]$ is attainable. $\square$

S2. NUMERICAL EXPERIMENT

We report simulation results to assess the performance of the proposed bounds and CIs under the symmetry assumption. We construct the data generating process that mimics the Mindspark intervention data used in the empirical illustration. We assume $\mathbb{P} [S_{1i} = 1 | S_{0i} = 1] = 1$, $\mathbb{P} [S_i = 1 | D_i = 1] = 0.89$, $\mathbb{P} [S_i = 1 | D_i = 0] = 0.85$, and

$$\begin{cases} f_{Y_0^*}(y | S_{1i} = 1, S_{0i} = 1) = \phi(y; 0, 1), \\ f_{Y_1^*}(y | S_{1i} = 1, S_{0i} = 1) = \phi(y; 1/6, 1), \\ f_{Y_1^*}(y | S_{1i} = 1, S_{0i} = 0) = \phi(y; 1/6, 4), \end{cases}$$

where $\phi(\cdot; \mu, \sigma^2)$ is the density function of the normal distribution $\mathcal{N}(\mu, \sigma^2)$; $f_{Y_0^*}(y | S_{1i} = 0, S_{0i} = 0)$ and $f_{Y_1^*}(y | S_{1i} = 0, S_{0i} = 0)$ are not specified. Then, the average treatment effect for always-takers is $\tau = 1/6 \simeq 0.17$, Lee's bounds are $[\Delta, \nabla] = [0.06, 0.28]$, and the bounds under mean-meadian coincidence are $[\Delta^s, \nabla^s] = [0.11, 0.22]$.

We randomly generate samples of size $n = 500, 1000$, and 3000 (half of them are assigned to the treatment group) with 500 iterations and compute empirical bounds, average length and empirical coverage of Imbens and Manski's 95% CIs. The results, summarised in Table S1, indicate that the point estimates are precise and the average length of CIs are tightened, while the empirical coverage is over 95%, which is consistent with Imbens and Manski's (2004) asymptotic theory. Further, the probability that the CIs do not contain zero has more than doubled.

Table S1: Simulation Results

	LB	UB	Length of CIs	Coverage of CIs (%)	% of $0 \notin$ CIs
$n = 500$					
$[\Delta, \nabla]$	0.06 (0.12)	0.29 (0.12)	0.64	99.8	9.2
$[\Delta^s, \nabla^s]$	0.11 (0.11)	0.24 (0.12)	0.55	99.8	19.8
$n = 1000$					
$[\Delta, \nabla]$	0.05 (0.09)	0.27 (0.09)	0.50	100	14.0
$[\Delta^s, \nabla^s]$	0.11 (0.08)	0.23 (0.09)	0.41	99.2	33.0
$n = 3000$					
$[\Delta, \nabla]$	0.06 (0.05)	0.27 (0.05)	0.37	100	31.0
$[\Delta^s, \nabla^s]$	0.11 (0.05)	0.22 (0.05)	0.28	99.6	71.6

Note: In the first column (LB), the average of the estimated lower bounds are shown with standard errors in the parenthesis. The second column (UB) is for upper bounds. The third and fourth columns show the average length and empirical coverage of Imbens and Manski’s CIs. In the last column, the percentage that the bounds do not include zero is reported.

S3. EMPIRICAL ILLUSTRATIONS: ADDITIONAL RESULTS

S3.1. Mindspark Intervention

In Table S2, we report the covariate-adjusted bounds. Following Muralidharan et al. (2019, Table A8), we use the location where the Mindspark intervention was held (Mindspark centres) as a covariate.

In most cases, the introduction of the symmetry assumption tightens the bounds compared to the standard Lee bounds, and incorporating covariates further narrows these bounds. When assuming strict monotonicity, the bounds under symmetry become sufficiently narrow to derive policy implications. However, the bounds under stochastic monotonicity are widened again. Regarding the CIs, relaxing the strict monotonicity assumption results in $0 \in$ CIs only for Hindi scores; thus, the qualitative implications remain similar to those presented in the main text.

S3.2. Log Wage Distributions of American Workers

We present density estimates for the log wages of American workers. The dataset is sourced from Honoré and Hu (2020). Following Honoré and Hu (2020), our analysis is confined to samples from Arizona, California, New Mexico, and Texas, spanning the years 2003 to 2016, with a specific focus on non-Hispanic whites and Mexican-Americans. For additional details, refer to Honoré and Hu (2020, p. 1027). The estimated densities are illustrated in Figures S3a–S3d, wherein the sample mean and median are also reported. The Epanechnikov kernel and Sheather and Jones’s (1991) bandwidth selector are used.

Table S2: Empirical Bounds with Covariate

Bounds under Monotonicity ($\vartheta = 1$)				
	Math		Hindi	
	(1)	(2)	(3)	(4)
Symmetry		Yes		Yes
Bounds	[0.316, 0.485]	[0.357, 0.410]	[0.143, 0.286]	[0.127, 0.175]
CI	(0.168, 0.649)	(0.220, 0.547)	(−0.006, 0.435)	(0.002, 0.300)
Bounds under Stochastic Monotonicity ($\vartheta = 0.95$)				
	Math		Hindi	
	(1)	(2)	(3)	(4)
Symmetry		Yes		Yes
Bounds	[0.248, 0.558]	[0.323, 0.434]	[0.071, 0.358]	[0.096, 0.217]
CI	(0.116, 0.686)	(0.193, 0.566)	(−0.053, 0.480)	(−0.021, 0.338)

The estimated densities in Figures S3a–S3c seem almost symmetric. The density in Figure S3d looks asymmetric, while the sample mean and median are mostly the same.

S4. MARGINAL TREATMENT EFFECTS IN THE PRESENCE OF SAMPLE SELECTION

The stochastic monotonicity assumption will be applicable to many settings that are analysed under the monotonicity assumption. In this section, as a simple but important application, we consider the identification of the marginal treatment effect (MTE) in the presence of sample selection, which is analysed by Bartalotti et al. (2023a). In their work, the monotone sample selection as in Assumption S.4.2 below plays an important role in obtaining useful bounds; they wrote, “the identified set can be substantially tightened by imposing that the sample selection mechanism is monotone in the treatment.”

Again, in the case of MTE, the monotonicity may not always hold, as discussed in the main text. Therefore, empirical researchers can only obtain bounds that are too large to be informative when they are concerned about the monotonicity assumption and abandon the bounds based on this assumption. Stochastic monotonicity will be an effective alternative in such a case.

S4.1. Model, Assumptions, and Identification

Following Bartalotti et al. (2023a), we consider the next model:

$$D_i = \mathbf{1}\{V_i \leq P(Z_i)\}, \quad (\text{S.4a})$$

$$S_i = S_{1i}D_i + S_{0i}(1 - D_i), \quad (\text{S.4b})$$

$$Y_i = S_i \cdot \{Y_{1i}^*D_i + Y_{0i}^*(1 - D_i)\}, \quad (\text{S.4c})$$

$$(Y_i, S_i, D_i, Z_i) \text{ is observed}, \quad (\text{S.4d})$$

where all the characters are defined in the same manner as before, except for Z_i and V_i . Z_i is a vector of continuous instrumental variables, and V_i is a latent variable. As in Bartalotti et al. (2023a), our target is the following MTE function for the always-takers:

$$\mu(v) = \mathbb{E}[Y_{1i}^* - Y_{0i}^* | V = v, S_{0i} = 1, S_{1i} = 1].$$

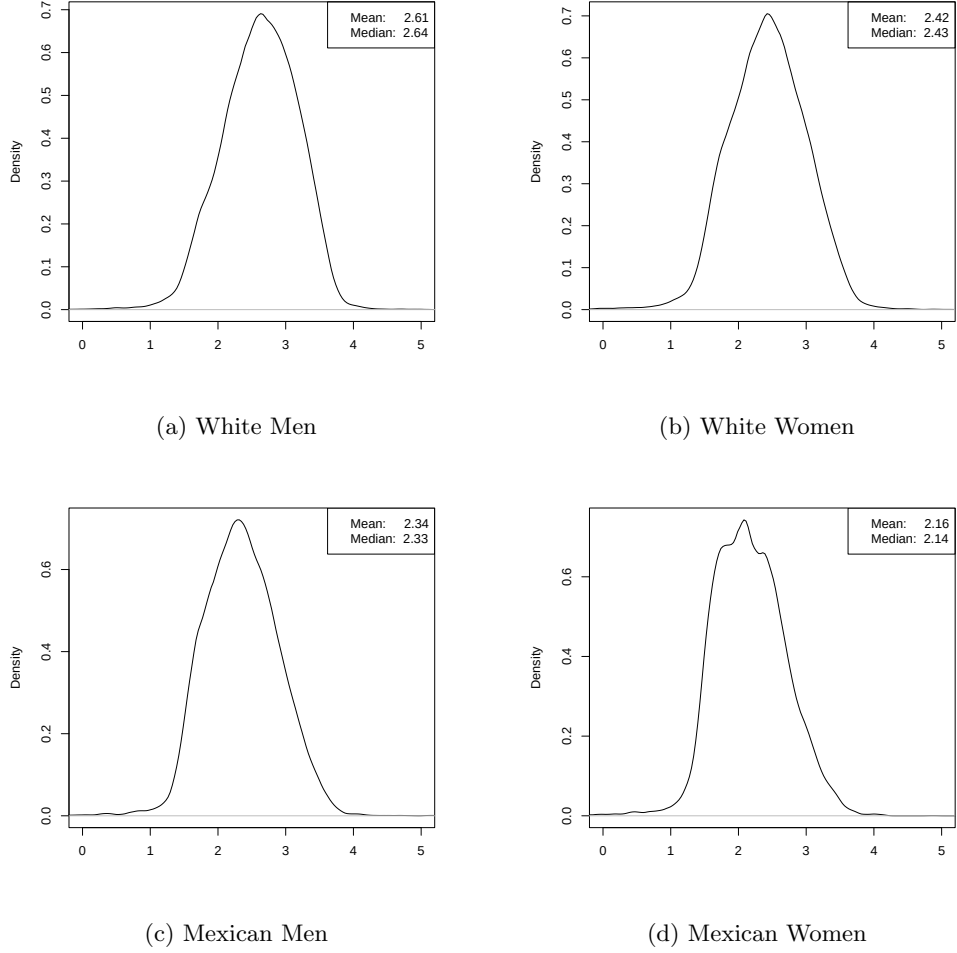


Figure S3: Log Wage Distributions

Bartalotti et al. (2023a) shows the identification result under the next set of assumptions and Assumption S.4.2, i.e., the monotonicity assumption.

ASSUMPTION S.4.1. (BARTALOTTI ET AL., 2023A)

- (i) The instruments Z_i are independent of $(Y_{1i}^*, Y_{0i}^*, S_{1i}, S_{0i}, V_i)$.
- (ii) $P(z)$ is a nontrivial function of z and the random variable $P(Z_i)$ is absolutely continuous with support given by an interval $\mathcal{P} := [p_L, p_U] \subseteq [0, 1]$.
- (iii) $0 < \mathbb{P}[D_i = 1] < 1$ and $\mathbb{P}[S_{0i} = 1, S_{1i} = 1 | V = v] > 0$ for any $v \in \mathcal{P}$.
- (iv) $\mathbb{E}[|Y_{di}^*| | S_{0i} = 1, S_{1i} = 1, V = v] < \infty$ for any $v \in \mathcal{P}$ and $d \in \{0, 1\}$.
- (v) The unconditional distribution of V_i is uniform over $[0, 1]$.

ASSUMPTION S.4.2. (MONOTONICITY) $\mathbb{P}[S_{1i} = 1 | S_{0i} = 1] = 1$.

Instead of imposing this exact monotonicity assumption (Assumption S.4.2), we con-

sider the identification under a slightly modified version of the stochastic monotonicity assumption in the main text:

ASSUMPTION S.4.3. (STOCHASTIC MONOTONICITY FOR MTE)

- (i) $\mathbb{P}[S_{1i} = 1 | S_{0i} = 1, V_i = v] \geq \vartheta_L$ with known $\vartheta_L \in (0, 1]$ for any $v \in \mathcal{P}$.
- (ii) (Y_{1i}^*, Y_{0i}^*) is independent of S_{1i} conditional on $S_{0i} = 1, V = v$.

The intuition behind this assumption is similar to before. Under this version of monotonicity, we can again obtain bounds as follows:

THEOREM S.4.1. Assume the model (S.4). Let Y_{0i}^* and Y_{1i}^* be continuous random variables. Under Assumptions S.4.1 and S.4.3, for any $v \in \mathcal{P}$, the upper and lower (uniformly) sharp bounds, $\nabla_\mu(v)$ and $\Delta_\mu(v)$, for $\mu(v)$ are given by

$$\begin{aligned}\nabla_\mu(v) &= \mathbb{E}[\tilde{Y}_{1i} | D_i = 1, S_i = 1, P(Z_i) = v, \tilde{Y}_{1i} \geq y_{1-s(v)}^1] - \mathbb{E}[\tilde{Y}_{0i} | D_i = 0, S_i = 1, P(Z_i) = v], \\ \Delta_\mu(v) &= \mathbb{E}[\tilde{Y}_{1i} | D_i = 1, S_i = 1, P(Z_i) = v, \tilde{Y}_{1i} \leq y_{s(v)}^1] - \mathbb{E}[\tilde{Y}_{0i} | D_i = 0, S_i = 1, P(Z_i) = v],\end{aligned}$$

where $s(v) = \max\{\vartheta_L \eta(v), \tilde{\alpha}(v)\}$. The conditional distribution of $\tilde{Y}_{di}$ (given $D_i = d$, $S_i = 1$, and $P(Z_i) = v$), written by $\tilde{F}_d$, is given by

$$\tilde{F}_d(y) = \left(\frac{\partial \mathbb{P}[Y \leq y, S_i = 1, D_i = d | P(Z_i) = v]}{\partial v} \right) \bigg/ \left(\frac{\partial \mathbb{P}[S_i = 1, D_i = d | P(Z_i) = v]}{\partial v} \right),$$

and $y_r^1 = \tilde{F}_1^{-1}(r)$. The functions η and $\tilde{\alpha}$ are given by

$$\eta(v) = - \left(\frac{\partial \mathbb{P}[S_i = 1, D_i = 0 | P(Z_i) = v]}{\partial v} \right) \bigg/ \left(\frac{\partial \mathbb{P}[S_i = 1, D_i = 1 | P(Z_i) = v]}{\partial v} \right),$$

and

$$\begin{aligned}\tilde{\alpha}(v) &= \max \left\{ - \frac{\partial \mathbb{P}[S_i = 1, D_i = 0 | P(Z_i) = v]}{\partial v} + \frac{\partial \mathbb{P}[S_i = 1, D_i = 1 | P(Z_i) = v]}{\partial v} - 1, 0 \right\} \\ &\quad \bigg/ \left(\frac{\partial \mathbb{P}[S_i = 1, D_i = 1 | P(Z_i) = v]}{\partial v} \right).\end{aligned}$$

Proof of Theorem S.4.1: We first summarize for convenience important identification results that hold under Assumption S.4.1 derived in Bartalotti et al. (2023a, Section 3.1). Let $\mathcal{Y} \subseteq \mathbb{R}$ be the support of Y_{0i}^* and Y_{1i}^* . For any $v \in \mathcal{P}$ and any Borel set $\mathcal{A} \subseteq \mathcal{Y}$,

$$\mathbb{P}[Y_{1i}^* \in \mathcal{A}, S_{1i} = 1 | V_i = v] = \frac{\partial \mathbb{P}[Y_i \in \mathcal{A}, S_i = 1, D_i = 1 | P(Z_i) = v]}{\partial v}, \quad (\text{S.5})$$

$$\mathbb{P}[Y_{0i}^* \in \mathcal{A}, S_{0i} = 1 | V_i = v] = - \frac{\partial \mathbb{P}[Y_i \in \mathcal{A}, S_i = 1, D_i = 0 | P(Z_i) = v]}{\partial v}. \quad (\text{S.6})$$

Also, we immediately have that

$$\mathbb{P}[S_{1i} = 1 | V_i = v] = \frac{\partial \mathbb{P}[S_i = 1, D_i = 1 | P(Z_i) = v]}{\partial v}, \quad (\text{S.7})$$

$$\mathbb{P}[S_{0i} = 1 | V_i = v] = - \frac{\partial \mathbb{P}[S_i = 1, D_i = 0 | P(Z_i) = v]}{\partial v}. \quad (\text{S.8})$$

Given the above relations, we can show that under Assumptions S.4.1 and S.4.3, it

holds that

$$\mathbb{E}[Y_{0i}^*|V_i = v, S_{0i} = 1, S_{1i} = 1] = \mathbb{E}[\tilde{Y}_{0i}|S_i = 1, D_i = 0, P(Z_i) = v].$$

Hence, we will consider the bounds for $\mathbb{E}[Y_{1i}^*|V_i = v, S_{0i} = 1, S_{1i} = 1]$.

Firstly, by Bartalotti et al. (2023a, equation (3.7)), it holds that

$$\begin{aligned} \mathbb{P}[Y_{1i}^* \in \mathcal{A}|S_{1i} = 1, V_i = v] &= \alpha(v) \mathbb{P}[Y_{1i}^* \in \mathcal{A}|S_{0i} = 1, S_{1i} = 1, V_i = v] \\ &\quad + (1 - \alpha(v)) \mathbb{P}[Y_{1i}^* \in \mathcal{A}|S_{0i} = 0, S_{1i} = 1, V_i = v], \end{aligned}$$

for any Borel set $\mathcal{A} \subseteq \mathcal{Y}$, where

$$\alpha(v) = \frac{\mathbb{P}[S_{1i} = 1, S_{0i} = 1|V = v]}{\mathbb{P}[S_{1i} = 1|V = v]}.$$

By Lemma 1 of Bartalotti et al. (2023a), the Fréchet inequalities produce a sharp lower bound $\alpha(v) \geq \tilde{\alpha}(v)$ under Assumption S.4.1. Also, Assumption S.4.3 and (S.7)-(S.8) imply that $\alpha(v) \geq \vartheta_L \eta(v)$. Therefore, for any v , $\alpha(v) \geq s(v) := \max\{\vartheta_L \eta(v), \tilde{\alpha}(v)\}$. Thus, Proposition 4 of Horowitz and Manski (1995) implies that the bounds $[\Delta_\mu(v), \nabla_\mu(v)]$ are valid.

The uniform sharpness can be shown in a similar way to Proposition 1 of Bartalotti et al. (2023a,b). Define $s_N(v) = \max\{\vartheta_L \eta_N(v), \tilde{\alpha}_N(v)\}$ where

$$\begin{aligned} \eta_N(v) &= - \left(\frac{\partial \mathbb{P}[S_i = 1, D_i = 0|P(Z_i) = v]}{\partial v} \right), \\ \tilde{\alpha}_N(v) &= \max \left\{ - \frac{\partial \mathbb{P}[S_i = 1, D_i = 0|P(Z_i) = v]}{\partial v} + \frac{\partial \mathbb{P}[S_i = 1, D_i = 1|P(Z_i) = v]}{\partial v} - 1, 0 \right\}, \end{aligned}$$

i.e., the “N”umerator of $\eta(v), \tilde{\alpha}(v)$. Below, we show that the lower bound Δ_μ is achievable. The upper bound is similar. Following Bartalotti et al. (2023b, p. 7), define $\tilde{V} \sim \text{Uniform}[0, 1]$, and

$$k(v) := \begin{cases} s_N(v) & \text{if } v \in \mathcal{P} \\ \epsilon_0 & \text{if } v \in [0, p_L) \\ \epsilon_1 & \text{if } v \in (p_U, 1] \end{cases},$$

where $\epsilon_0, \epsilon_1 \in [0, 1]$ are some appropriate constants. Define

$$\begin{aligned} \mathbb{P}[\tilde{S}_{1i} = s_1|\tilde{V}_i = v] &:= \begin{cases} \frac{\partial \mathbb{P}[S_i = s_1, D_i = 1|P(Z_i) = v]}{\partial v} & \text{if } v \in \mathcal{P} \\ \frac{\mathbb{P}[S_i = s_1, D_i = 1|P(Z_i) = p_L]}{p_L} & \text{if } v \in [0, p_L) \\ k(v)\mathbf{1}\{s_1 = 1\} + (1 - k(v))\mathbf{1}\{s_1 = 0\} & \text{if } v \in (p_U, 1] \end{cases} \\ \mathbb{P}[\tilde{S}_{0i} = s_0|\tilde{V}_i = v] &:= \begin{cases} - \frac{\partial \mathbb{P}[S_i = s_0, D_i = 0|P(Z_i) = v]}{\partial v} & \text{if } v \in \mathcal{P} \\ k(v)\mathbf{1}\{s_0 = 1\} + (1 - k(v))\mathbf{1}\{s_0 = 0\} & \text{if } v \in [0, p_L) \\ \frac{\mathbb{P}[S_i = s_0, D_i = 0|P(Z_i) = p_U]}{1 - p_U} & \text{if } v \in (p_U, 1] \end{cases} \end{aligned}$$

where $s_1, s_0 \in \{0, 1\}$. Then, defining $\pi_{11} := \mathbb{P}[\tilde{S}_{0i} = 1, \tilde{S}_{1i} = 1 | \tilde{V}_i = v] := k(v)$, we have

$$\begin{aligned}\pi_{01} &:= \mathbb{P}[\tilde{S}_{0i} = 0, \tilde{S}_{1i} = 1 | \tilde{V}_i = v] = \mathbb{P}[\tilde{S}_{1i} = 1 | \tilde{V}_i = v] - k(v), \\ \pi_{10} &:= \mathbb{P}[\tilde{S}_{0i} = 1, \tilde{S}_{1i} = 0 | \tilde{V}_i = v] = \mathbb{P}[\tilde{S}_{0i} = 1 | \tilde{V}_i = v] - k(v), \\ \pi_{00} &:= \mathbb{P}[\tilde{S}_{0i} = 0, \tilde{S}_{1i} = 0 | \tilde{V}_i = v] = 1 - (\pi_{11} + \pi_{01} + \pi_{10}),\end{aligned}$$

these are in $[0, 1]$ by Assumption S.4.3 and construction. We can see that Assumption S.4.3-(i) is satisfied. Define

$$\begin{aligned}\mathbb{P}[\tilde{Y}_{1i}^* \leq y_1, \tilde{S}_{1i} = s_1 | \tilde{V}_i = v] \\ &:= \begin{cases} \frac{\partial \mathbb{P}[Y_i \leq y_1, S_i = s_1, D_i = 1 | P(Z_i) = v]}{\partial v} & \text{if } v \in \mathcal{P} \\ \frac{\mathbb{P}[Y_i \leq y_1, S_i = s_1, D_i = 1 | P(Z_i) = p_L]}{p_L} & \text{if } v \in [0, p_L) \\ \frac{\{k(v)\mathbf{1}\{s_1 = 1\} + (1 - k(v))\mathbf{1}\{s_1 = 0\}\} \mathbb{P}[Y_i \leq y_1]}{p_U} & \text{if } v \in (p_U, 1] \end{cases} \\ \mathbb{P}[\tilde{Y}_{0i}^* \leq y_0, \tilde{S}_{0i} = s_0 | \tilde{V}_i = v] \\ &:= \begin{cases} -\frac{\partial \mathbb{P}[Y_i \leq y_0, S_i = s_0, D_i = 0 | P(Z_i) = v]}{\partial v} & \text{if } v \in \mathcal{P} \\ \frac{\{k(v)\mathbf{1}\{s_0 = 1\} + (1 - k(v))\mathbf{1}\{s_0 = 0\}\} \mathbb{P}[Y_i \leq y_0]}{1 - p_U} & \text{if } v \in [0, p_L), \\ \frac{\mathbb{P}[Y_i \leq y_0, S_i = s_0, D_i = 0 | P(Z_i) = p_U]}{1 - p_U} & \text{if } v \in (p_U, 1] \end{cases}\end{aligned}$$

which are proper distribution functions by (S.5), (S.6). Hence, letting $\tilde{Y}_{0i} \sim F_{\tilde{Y}_{0i}^* | \tilde{S}_{0i}=1, \tilde{V}=v}$ and $\tilde{Y}_{1i} \sim F_{\tilde{Y}_{1i}^* | \tilde{S}_{1i}=1, \tilde{V}=v}$, we can define

$$\begin{aligned}\mathbb{P}[\tilde{Y}_{1i}^* \leq y_1 | \tilde{S}_{0i} = 1, \tilde{S}_{1i} = s_1, \tilde{V} = v] &:= \mathbb{P}\left[\tilde{Y}_{1i} \leq y_1 | \tilde{Y}_{1i} \leq F_{\tilde{Y}_{1i}^*}^{-1}\left(\frac{\pi_{11}}{\pi_{11} + \pi_{01}}\right)\right], \\ \mathbb{P}[\tilde{Y}_{1i}^* \leq y_1 | \tilde{S}_{0i} = 0, \tilde{S}_{1i} = s_1, \tilde{V} = v] &:= \mathbb{P}\left[\tilde{Y}_{1i} \leq y_1 | \tilde{Y}_{1i} > F_{\tilde{Y}_{1i}^*}^{-1}\left(\frac{\pi_{11}}{\pi_{11} + \pi_{01}}\right)\right], \\ \mathbb{P}[\tilde{Y}_{0i}^* \leq y_0 | \tilde{S}_{0i} = s_0, \tilde{S}_{1i} = s_1, \tilde{V} = v] &:= \mathbb{P}[\tilde{Y}_{0i} \leq y_0],\end{aligned}$$

for $\forall s_1, s_0 \in \{0, 1\}$. Then, Assumption S.4.3-(ii) is satisfied. Define $\tilde{D}_i = \mathbf{1}\{\tilde{V}_i \leq P(Z_i)\}$, then we can see that

$$\begin{aligned}\mathbb{P}[\tilde{Y}_i \leq y, \tilde{S}_i = 1, \tilde{D}_i = 1 | Z_i = z] &= \mathbb{P}[\tilde{Y}_{1i}^* \leq y, \tilde{S}_{1i} = 1, \tilde{V}_i \leq P(z)] \\ &= \int_0^{P(z)} \mathbb{P}[\tilde{Y}_{1i}^* \leq y, \tilde{S}_{1i} = 1 | \tilde{V}_i = v] dv \\ &= \mathbb{P}[Y_i \leq y, S_i = 1, D_i = 1 | P(Z_i) = P(z)] \\ &= \mathbb{P}[Y_i \leq y, S_i = 1, D_i = 1 | Z_i = z],\end{aligned}$$

where the third equality uses the above definition of the integrand; for the second equality, see Bartalotti et al. (2023a, p. 570). A similar argument holds for $\mathbb{P}[\tilde{Y}_i \leq y, \tilde{S}_i = 1, \tilde{D}_i = 0 | Z_i = z]$. Hence, the defined set of variables with tildes induces the joint distribution of

the observed data. Finally, we can obtain that, for $v \in \mathcal{P}$,

$$\Delta_\mu(v) = \mathbb{E} \left[\tilde{Y}_{1i}^* - \tilde{Y}_{0i}^* | \tilde{S}_{0i} = 1, \tilde{S}_{1i} = 1, \tilde{V} = v \right],$$

that is, the lower bound is attainable. $\square$

REMARK S.4.1. *The uniform sharpness means that for any function $f : \mathcal{P} \rightarrow \mathbb{R}$ such that $f \in [\Delta_\mu, \nabla_\mu]$, we can construct random variables that are consistent with the model and assumptions, imply the joint distribution of observables is the same as the observed distribution, and produce f as the MTE function. See Bartalotti et al. (2023a, Remark 3).*

Obviously, (i) Assumption S.4.1+Assumption S.4.2 $\implies$ (ii) Assumption S.4.1+Assumption S.4.3 $\implies$ (iii) Assumption S.4.1. Bartalotti et al. (2023a) derive identification under (i) and (iii), and find that the bounds under (iii) are too wide to be informative (Bartalotti et al., 2023b, p. 29). Hence, the bounds under (ii), which are obtained in Theorem S.4.1, will be much more informative than the bounds under (iii) but are more robust than (i).

EXAMPLE S.4.1. (NUMERICAL EXAMPLE) *We here compute the MTE bounds using the same data generating process (DGP) as in Bartalotti et al. (2023b, p. 29). Consider the following DGP, which is adopted from Bartalotti et al. (2023b, p. 29):*

$$D = \mathbf{1}\{V \leq \Phi(Z)\}, S = \mathbf{1}\{U_S \leq 0.1 + 0.4D\}, \\ Y^* = Y_1^*D + Y_0^*(1 - D), Y = Y^*S,$$

where $V = \Phi(\epsilon_V)$, $U_S = (\epsilon_V + \epsilon_S)/\sqrt{2}$, $(\epsilon_V, \epsilon_S, Z, \xi) \sim \mathcal{N}(0, I)$, I is the identity matrix, Φ is the distribution function of $\mathcal{N}(0, 1)$, and

$$Y_0^* = \mathbf{1}\{\xi \geq 0\} \cdot \epsilon_V + \mathbf{1}\{\xi < 0\} \cdot (-\epsilon_V), \\ Y_1^* = \mathbf{1}\{\xi \geq 0\} \cdot 5\epsilon_V + \mathbf{1}\{\xi < 0\} \cdot (-\epsilon_V).$$

The bounds under (i) Assumption S.4.1+Assumption S.4.2, (ii) Assumption S.4.1+Assumption S.4.3, and (iii) Assumption S.4.1 are shown in Figure S4; the dashed red lines are bounds under (i), the solid green lines are bounds under (ii) with $\vartheta_L = 0.8$, the dot-dashed blue lines are bounds under (iii), and the black solid line in the centre is the true $\mu(v)$. Note that we set $\vartheta_L = 0.8$ to obtain visually clear bounds, and we use numerical integration to compute bounds. Note that all sets of assumptions are valid in this example. The solid green line is the bound under the stochastic monotonicity (i.e., (ii) with $\vartheta_L = 0.8$). The bounds are, of course, wider than the bounds under (i), while the bounds are much more informative than those under (iii). Therefore, Assumption S.4.3 will be useful when researchers lack confidence in the validity of the monotonicity assumption.

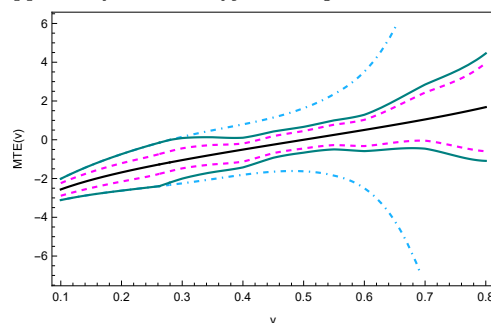


Figure S4: Bounds for $\mu(v)$ under Several Sets of Assumptions

REFERENCES

- Andrews, D. W. (1994). Empirical Process Methods in Econometrics. In *Handbook of Econometrics*, Volume 4, Chapter 37, pp. 2247–2294. Elsevier.
- Bartalotti, O., D. Kédagni, and V. Possebom (2023a). Identifying Marginal Treatment Effects in the Presence of Sample Selection. *Journal of Econometrics* 234(2), 565–584.
- Bartalotti, O., D. Kédagni, and V. Possebom (2023b). Supporting Information (Online Appendix): Identifying Marginal Treatment Effects in the Presence of Sample Selection. *Journal of Econometrics*.
- Caetano, C., G. Caetano, and E. Nielsen (2024). Correcting for Endogeneity in Models with Bunching. *Journal of Business & Economic Statistics* 42(3), 851–863.
- Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection Bounds: Estimation and Inference. *Econometrica* 81(2), 667–737.
- Honoré, B. E. and L. Hu (2020). Selection Without Exclusion. *Econometrica* 88(3), 1007–1029.
- Horowitz, J. L. and C. F. Manski (1995). Identification and Robustness with Contaminated and Corrupted Data. *Econometrica* 63(2), 281–302.
- Imbens, G. W. and C. F. Manski (2004). Confidence Intervals for Partially Identified Parameters. *Econometrica* 72(6), 1845–1857.
- Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies* 76(3), 1071–1102.
- Muralidharan, K., A. Singh, and A. J. Ganimian (2019). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. *American Economic Review* 109(4), 1426–60.
- Newey, W. K. and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing. In *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Elsevier.
- Sheather, S. J. and M. C. Jones (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 53(3), 683–690.