Supplementary Material

Supplement to "The identification power of smoothness assumptions in models with counterfactual outcomes"

(*Quantitative Economics*, Vol. 9, No. 2, July 2018, 617–642)

Wooyoung Kim
Department of Economics, University of Wisconsin-Madison

Koohyun Kwon
Department of Economics, Yale University

Soonwoo Kwon
Department of Economics, Yale University

Sokbae Lee
Centre for Microdata Methods and Practice, Institute for Fiscal Studies and Department of Economics, Columbia University

Appendix A: Omitted proofs

In this section, we provide lemmas that establish the sharpness of the bounds given in Propositions 3.1, 4.1, and 4.2, and any proofs that were omitted from the main text of the paper.

Lemma A.1. *The bounds given in Proposition* 3.1 *are sharp.*

Proof. *Part* (i). Consider a data generating process (DGP) subject to (s.t.) $E[Y_i(t)|Z_i = s] = E[Y_i|Z_i = s] + b|s - t|\ \forall t, s \in \Gamma$. This ensures $E[Y_i(t)]$ attains the upper bound. It remains to show this DGP satisfies STR, which is followed by

$$\begin{aligned}\left|E\left[Y_i(t_1)|Z_i = s\right] - E\left[Y_i(t_2)|Z_i = s\right]\right| &= b\big||s - t_1| - |s - t_2|\big| \\ &\le b|t_1 - t_2|.\end{aligned}$$

On the other hand, the DGP $E[Y_i(t)|Z_i = s] = E[Y_i|Z_i = s] - b|s - t|\ \forall t, s \in \Gamma$ attains the lower bound, and the convex combinations between the two DGPs yield all the values between the lower and upper bounds. It can also be shown that they obey STR.

Wooyoung Kim: wkim68@wisc.edu
Koohyun Kwon: koohyun.kwon@yale.edu
Soonwoo Kwon: soonwoo.kwon@yale.edu
Sokbae Lee: sl3841@columbia.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE545

Part (ii). Consider a DGP s.t. $E[Y_i(t)|Z_i = s] = E[Y_i|Z_i = s] + b(t - s)$ when $s \leq t$ and $E[Y_i(t)|Z_i = s] = E[Y_i|Z_i = s]$ when $s > t$. This ensures $E[Y_i(t)]$ attains the upper bound. To show that this DGP satisfies SMTR, note that for any t_1 and t_2 satisfying $t_1 > t_2$, we have

$$E\big[Y_i(t_1)|Z_i = s\big] - E\big[Y_i(t_2)|Z_i = s\big] = \begin{cases} b(t_1 - t_2), & \text{if } t_1 > t_2 \geq s, \\ 0, & \text{if } t_2 < t_1 < s, \\ b(t_1 - s), & \text{if } t_2 < s \leq t_1. \end{cases}$$

This implies that SMTR holds since $(t_1 - s) \leq (t_1 - t_2)$ when $t_2 < s \leq t_1$. The lower bound can be attained similarly, and furthermore, as in part (i), the convex combinations between the two polar DGPs yield all the values between the lower and upper bounds. □

Proof of Proposition 3.2. To verify the sharpness of the STR upper bound, consider a HLR-type DGP s.t. $Y_i(t) = \beta \times t + \delta_i$ with $E\delta_i = 0$ and $\beta = b$ satisfies STR, as mentioned in the main text, and this DGP yields $\Delta(t, t') = b(t - t')$. Likewise, the sharp lower bound is $-b(t - t')$ (take $\beta = -b$). Now the convex combinations between the two DGPs yield all the values between the lower and upper bounds. Identical arguments yield that the sharp SMTR upper and lower bounds for the average treatment effect are $b(t - t')$ and 0, respectively. □

Proof of Proposition 3.3. The proof is omitted since it is basically the same as that of Proposition 3.1. □

Lemma A.2. *The bounds given in Proposition* 4.1 *are sharp.*

Proof. We start by taking $g(s, s)$ to be a monotone increasing function of s since the MTR-MTS assumption requires such property to hold. To verify the sharpness of the upper bound, consider the following DGP:

$$g(t, s) = \begin{cases} \inf_{s' \in [s,t]} \big(g(s', s') + b(t - s')\big), & s \leq t, \\ g(s, s), & s > t \end{cases}$$

for all s, t.

First, we will check whether SMTR holds, that is, $0 \leq g(t_2, s) - g(t_1, s) \leq b(t_2 - t_1)$ for $t_1 < t_2$. There are three cases: (1) $t_1 < t_2 \leq s$, (2) $t_1 \leq s < t_2$, and (3) $s < t_1 < t_2$. For the first case, note that

$$\begin{aligned} g(t_2, s) - g(t_1, s) &= g(s, s) - g(s, s) \\ &= 0, \end{aligned}$$

so SMTR holds. For the second case,

$$g(t_2, s) - g(t_1, s) = \inf_{s' \in [s,t_2]} \big(g(s', s') + b(t_2 - s')\big) - g(s, s)$$

$$
\begin{aligned}
&\geq \inf_{s' \in [s,t_2]} g(s',s') - g(s,s) \\
&= 0
\end{aligned}
$$

since $g(s,s)$ is an increasing function with respect to s. Therefore, MTR holds. Moreover,

$$
\begin{aligned}
g(t_2,s) - g(t_1,s) &= \inf_{s' \in [s,t_2]} \big(g(s',s') + b(t_2 - s') \big) - g(s,s) \\
&\leq g(s,s) - b(t_2 - s) - g(s,s) \\
&= b(t_2 - s) \\
&\leq b(t_2 - t_1)
\end{aligned}
$$

since $t_1 \leq s$ for the second case. This verifies the smoothness condition, so SMTR holds. For the third case,

$$
g(t_2,s) - g(t_1,s) = \inf_{s' \in [s,t_2]} \big(g(s',s') + b(t_2 - s') \big) - \inf_{s' \in [s,t_1]} \big(g(s',s') + b(t_1 - s') \big).
$$

Note that

$$
\begin{aligned}
&\inf_{s' \in [s,t_2]} \big(g(s',s') + b(t_2 - s') \big) \\
&\quad = \min\Big\{ \inf_{s' \in [s,t_1]} \big(g(s',s') + b(t_2 - s') \big), \inf_{s' \in [t_1,t_2]} \big(g(s',s') + b(t_2 - s') \big) \Big\}.
\end{aligned}
$$

For the first term inside the minimum operator,

$$
\begin{aligned}
\inf_{s' \in [s,t_1]} \big(g(s',s') + b(t_2 - s') \big) &= \inf_{s' \in [s,t_1]} \big(g(s',s') + b(t_1 - s') + b(t_2 - t_1) \big) \\
&\geq \inf_{s' \in [s,t_1]} \big(g(s',s') + b(t_1 - s') \big).
\end{aligned}
$$

Moreover, for the second term inside the minimum operator, we can see that

$$
\begin{aligned}
\inf_{s' \in [t_1,t_2]} \big(g(s',s') + b(t_2 - s') \big) &\geq \inf_{s' \in [t_1,t_2]} g(s',s') \\
&= g(t_1,t_1) \\
&\geq \inf_{s' \in [s,t_1]} \big(g(s',s') + b(t_1 - s') \big).
\end{aligned}
$$

Therefore, MTR holds. For the case of SMTR, let $A \equiv \inf_{s' \in [s,t_1]} (g(s',s') + b(t_1 - s'))$ and $B \equiv \inf_{s' \in [t_1,t_2]} (g(s',s') + b(t_1 - s'))$. Then

$$
\begin{aligned}
g(t_2,s) - g(t_1,s) &= \inf_{s' \in [s,t_2]} \big(g(s',s') + b(t_2 - s') \big) - \inf_{s' \in [s,t_1]} \big(g(s',s') + b(t_1 - s') \big) \\
&= \inf_{s' \in [s,t_2]} \big(g(s',s') + b(t_1 - s') \big) + b(t_2 - t_1) - \inf_{s' \in [s,t_1]} \big(g(s',s') + b(t_1 - s') \big) \\
&= \min\{A, B\} + b(t_2 - t_1) - A.
\end{aligned}
$$

Since $\min\{A, B\} \leq A$, we can see that SMTR holds.

Next, we need to check MTS, that is, $g(t,s_2)-g(t,s_1)\geq 0$ for $s_1<s_2$. There are again three cases to consider: (1) $s_1<s_2\leq t$, (2) $s_1\leq t<s_2$, and (3) $t<s_1<s_2$. For the first case,

$$\begin{aligned} g(t,s_2)-g(t,s_1) &= \inf_{s'\in[s_2,t]}\big(g(s',s')+b(t-s')\big)-\inf_{s'\in[s_1,t]}\big(g(s',s')+b(t-s')\big)\\ &\geq 0, \end{aligned}$$

since $[s_2,t]\subset[s_1,t]$. For the second case,

$$\begin{aligned} g(t,s_2)-g(t,s_1) &= g(s_2,s_2)-\inf_{s'\in[s_1,t]}\big(g(s',s')+b(t-s')\big)\\ &\geq g(s_2,s_2)-g(t,t)\\ &\geq 0. \end{aligned}$$

For the third,

$$\begin{aligned} g(t,s_2)-g(t,s_1) &= g(s_2,s_2)-g(s_1,s_1)\\ &\geq 0. \end{aligned}$$

Therefore, MTS holds for this DGP.

Likewise, the DGP that achieves the lower bound is

$$g(t,s)=\begin{cases} g(s,s), & s\leq t,\\ \sup_{s'\in[t,s]}\big(g(s',s')-b(s'-t)\big), & s>t. \end{cases}$$

We omit the proof that it satisfies SMTR-MTS, since it is analogous to the previous one, and the convex combination of the previous DGP and this one yields all the values between the upper and the lower bounds. □

LEMMA A.3. *The bounds given in Proposition* 4.2 *are sharp.*

PROOF. The DGP that achieves the sharp bound is constructed by the following manner. First, fix t_1 and t_2. Then define

$$f_2(s)=\begin{cases} f_I(s,t_2), & s\leq t_2,\\ g(s,s), & s>t_2, \end{cases}$$

$$\tilde{f}_1(s)=\begin{cases} g(s,s), & s\leq t_1,\\ f_S(s,t_1), & s>t_1, \end{cases}$$

and

$$f_1(s)=\max\big\{\tilde{f}_1(s),f_2(s)-b(t_2-t_1)\big\}.$$

Consider the following DGP:

$$g(t,s)=\begin{cases} f_1(s), & t \leq t_1, \\ f_1(s)+\left(\dfrac{f_2(s)-f_1(s)}{t_2-t_1}\right)(t-t_1), & t_1 < t \leq t_2, \\ f_2(s), & t > t_2. \end{cases}$$

We can easily check that this form of DGP leads to the sharp upper bound for $E[Y(t_2)-Y(t_1)]$.

We now show that the SMTR-MTS condition holds for this DGP. For SMTR, it is necessary to verify that $0 \leq g(t',s)-g(t,s) \leq b(t'-t)$, and there are six cases to consider: (1) $t < t' \leq t_1 < t_2$, (2) $t \leq t_1 < t' \leq t_2$, (3) $t_1 < t < t' \leq t_2$, (4) $t_1 < t \leq t_2 < t'$, (5) $t \leq t_1 < t_2 < t'$, and (6) $t_2 < t < t'$. Checking each of these six cases shows that:

(1) $g(t',s)-g(t,s)=f_1(s)-f_1(s)=0$.

(2) first note that

$$\left(\frac{f_2(s)-f_1(s)}{t_2-t_1}\right) \leq b \tag{A.1}$$

by the definition of $f_1(s)$. Moreover, $f_2(s)-f_1(s) \geq 0$ by the definitions of $f_S(s,t_1)$ and $f_I(s,t_2)$. Then $g(t',s)-g(t,s)=(\frac{f_2(s)-f_1(s)}{t_2-t_1})(t'-t_1) \geq 0$ and $(\frac{f_2(s)-f_1(s)}{t_2-t_1})(t'-t_1) \leq b(t'-t_1) \leq b(t'-t)$ since $t \leq t_1$.

(3) $g(t',s)-g(t,s)=(\frac{f_2(s)-f_1(s)}{t_2-t_1})(t'-t)$ which is greater than 0 and smaller than $b(t'-t)$ by (A.1).

(4) $g(t',s)-g(t,s)=(\frac{t_2-t}{t_2-t_1})(f_2(s)-f_1(s)) \geq 0$ and $(\frac{t_2-t}{t_2-t_1})(f_2(s)-f_1(s)) \leq b(t_2-t) \leq b(t'-t)$ since $t' > t_2$.

(5) $g(t',s)-g(t,s)=f_2(s)-f_1(s)$. The discussion regarding the second case shows this is greater than 0 and smaller than $b(t_2-t_1)$.

(6) $g(t',s)-g(t,s)=f_2(s)-f_2(s)=0$.

Next, for the MTS, we need to show that $g(t,s_2)-g(t,s_1)$ for $s_1 < s_2$. There are three cases to consider: (1) $t \leq t_1$, (2) $t_1 < t \leq t_2$, and (3) $t_2 < t$. For the first case,

$$\begin{aligned} g(t,s_2)-g(t,s_1) &= f_1(s_2)-f_1(s_1) \\ &= \max\{\tilde{f}_1(s_2), f_2(s_2)-b(t_2-t_1)\} \\ &\quad - \max\{\tilde{f}_1(s_1), f_2(s_1)-b(t_2-t_1)\}. \end{aligned}$$

In order to show this to be greater than 0, first note that $a_2 \geq a_1$ and $b_2 \geq b_1$ implies $\max\{a_2,b_2\} \geq \max\{a_1,b_1\}$. Therefore, we only need to show (i) $\tilde{f}_1(s_2) \geq \tilde{f}_1(s_1)$ and (ii) $f_2(s_2) \geq f_2(s_1)$. We have already verified inequality (ii) in the proof of Proposition 4.1. Also, note that inequality (i) is analogous to inequality (ii). Therefore, case (1) has been dealt with. Now for the second case,

$$\begin{aligned} g(t,s_2) - g(t,s_1) \\ &= f_1(s_2) + \left(\frac{f_2(s_2) - f_1(s_2)}{t_2 - t_1}\right)(t - t_1) - f_1(s_1) - \left(\frac{f_2(s_1) - f_1(s_1)}{t_2 - t_1}\right)(t - t_1) \\ &= \left(\frac{t_2 - t}{t_2 - t_1}\right)\bigl(f_1(s_2) - f_1(s_1)\bigr) + \left(\frac{t - t_1}{t_2 - t_1}\right)\bigl(f_2(s_2) - f_2(s_1)\bigr), \end{aligned}$$

which is greater than 0 when $f_1(s_2) \geq f_1(s_1)$ and $f_2(s_2) \geq f_2(s_1)$, and we have already shown that the last two inequalities hold. For the last case, $g(t,s_2) - g(t,s_1) = f_2(s_2) - f_2(s_1)$, which is greater than 0 as mentioned before. Therefore, MTS holds. □

Proof of Propositions 4.3 and 4.4. The detailed proofs of Propositions 4.3 and 4.4 are omitted since they are basically analogous to those of Propositions 4.1 and 4.2. One notable modification that is necessary to prove the sharpness of Proposition 4.2 is to define $g(t,s) = f_1(s) + (\frac{f_2(s)-f_1(s)}{\omega(t_2-t_1)})(\omega(t-t_1))$ if $t_1 < t \leq t_2$. Note that the additional requirement $\omega(t_2 - t_1) > 0$ is used in this step. □

Appendix B: Binary treatment

In this section, we consider the binary treatment case to better understand the role of the smoothness assumption. Suppose throughout this section there are only two treatment levels: $t_1 = 1$ and $t_0 = 0$.

Remark B.1 (Binary Treatment—Proposition 3.1). Proposition 3.1(ii) can be written as

$$\begin{aligned} E[Y_i] - bP(Z_i = 1) \leq g^*(0) \leq E[Y_i], \\ E[Y_i] \leq g^*(1) \leq E[Y_i] + bP(Z_i = 0). \end{aligned}$$

Specifically, a strict improvement from the MTR bounds occurs when $y_{\max} > g(0,0) + b$ (for $g^*(1)$) or $g(1,1) - b > y_{\min}$ (for $g^*(0)$). This shows how b is restricting the possible values for the unobserved outcomes, especially when $y_{\max}$ is too large or $y_{\min}$ is too small.

Remark B.2 (Binary Treatment—Proposition 4.1). Proposition 4.1 can be written as

$$\begin{aligned} l_1(0) &= g(0,0)P(Z_i = 0) + \sup\bigl\{g(0,0), g(1,1) - b\bigr\}P(Z_i = 1), \\ u_1(0) &= E[Y_i], \\ l_1(1) &= E[Y_i], \\ u_1(1) &= \inf\bigl\{g(0,0) + b, g(1,1)\bigr\}P(Z_i = 0) + g(1,1)P(Z_i = 1). \end{aligned}$$

Specifically, the values inside the supremum and infimum operators show that improvement from the MTR-MTS bounds occurs when $g(1,1) - g(0,0) > b$. Note that the improved bounds become equivalent to the case without monotone treatment selection.

Remark B.3 (Binary Treatment—Proposition 4.2). For the binary treatment case, $0 \leq \Delta(0,1) \leq b$ (trivial upper bound) when $g(1,1) - g(0,0) > b$, and $0 \leq \Delta(0,1) \leq g(1,1) - g(0,0)$ (MTR-MTS upper bound) when $g(1,1) - g(0,0) \leq b$. Therefore, imposing the smoothness assumption is not useful for bounding the average treatment effect when the treatment is binary.

Appendix C: Adding instrumental variables assumptions to treatment responses

In this section, we show how to tighten the identification results obtained in Section 3 when an instrumental variable exists. In particular, we follow Manski and Pepper (2000) and study the identifying power of instrumental variable (IV) and monotone instrumental variable (MIV) assumptions, as they are combined with conditional versions of STR and SMTR conditions.[1]

Assume from now on that we observe independent and identically distributed observations $\{(Y_i, Z_i, V_i) : i = 1, \ldots, n\}$, where $V_i \in \mathcal{V} \subset \mathbb{R}$ is a real-valued instrumental variable for individual i. Define $g(t,s,v) \equiv E[Y_i(t)|Z_i = s, V_i = v]$ to be the expectation of $Y_i(t)$ conditional on $Z_i = s$ and $V_i = v$. We now state the STR and SMTR assumptions conditional on $V_i = v$ (hence called CSTR and CSMTR) as well as the IV and MIV assumptions of Manski and Pepper (2000).

Assumption C.1 (Treatment Response and Instrumental Variable Assumptions). *Consider the following assumptions*:

(i) (Condition CSTR) *There exists a constant* $b > 0$ *such that* $|g(t,s,v) - g(t',s,v)| \leq b|t - t'|\ \forall (t,t',s,v) \in (\Gamma \times \Gamma \times \Gamma \times \mathcal{V})$.

(ii) (Condition CSMTR) *The CSTR condition in part* (i) *holds with a constant* $b > 0$. *In addition,* $g(t,s,v) \geq g(t',s,v)\ \forall (t,t',s,v) \in (\Gamma \times \Gamma \times \Gamma \times \mathcal{V})$ *satisfying* $t \geq t'$.

(iii) (Condition IV) $E[Y_i(t)|V_i = v] = E[Y_i(t)|V_i = v']$ *for all* $(v, v', t) \in (\mathcal{V} \times \mathcal{V} \times \Gamma)$.

(iv) (Condition MIV) *If* $v \geq v'$, *then* $E[Y_i(t)|V_i = v] \geq E[Y_i(t)|V_i = v']$ *for all* $(v, v', t) \in (\mathcal{V} \times \mathcal{V} \times \Gamma)$.

Condition CSTR is met if (3.3) holds for each individual and CSMTR is satisfied when (3.1) and (3.3) hold for each individual. Hence, the conditional versions of STR and SMTR conditions can be motivated, as in Section 3. The IV and MIV conditions are well known in the literature; see, for example, Manski and Pepper (2000, 2009) among others.

Remark C.1. A related identification assumption in the literature is "bounded instrumental variable" introduced in Manski and Pepper (2013). To express this assumption in

[1] Lafférs (2013) emphasize the importance of distinguishing conditional and unconditional versions of monotone treatment selection.

our notation, let $ATE_v \equiv E[Y(t_2)|V_i = v] - E[Y(t_1)|V_i = v]$ for some fixed $t_1 \neq t_2$. Then V_i is called a bounded instrumental variable if

$$|ATE_{v_1} - ATE_{v_2}| \leq \Delta$$

for some $\Delta > 0$ for all v_1 and v_2. This condition is related to our CSTR assumption in the following sense: $|ATE_{v_1} - ATE_{v_2}|$ is less than or equal to $|ATE_{v_1}| + |ATE_{v_2}|$ by triangular inequality, and each $|ATE_{v_1}|$ and $|ATE_{v_2}|$ is bounded due to the CSTR assumption.

The following proposition gives identification results under several possible combinations of the conditions in Assumption C.1.

Proposition C.1. *Assume that the support of $Y_i(t)$ is unbounded. Then the following bounds are sharp*:

(i) *Under CSTR and IV together,*

$$\sup_{v\in\mathcal{V}}\big\{E[Y_i|V_i = v] - bE\big[|Z_i - t||V_i = v\big]\big\} \leq g^*(t) \leq \inf_{v\in\mathcal{V}}\big\{E[Y_i|V_i = v] + bE\big[|Z_i - t||V_i = v\big]\big\}.$$

(ii) *Under CSMTR and IV together,*

$$\sup_{v\in\mathcal{V}}\big\{E[Y_i|V_i = v] - bE\big[(Z_i - t)^+|V_i = v\big]\big\} \leq g^*(t) \leq \inf_{v\in\mathcal{V}}\big\{E[Y_i|V_i = v] + bE\big[(Z_i - t)^-|V_i = v\big]\big\}.$$

(iii) *Under CSTR and MIV together,*

$$\begin{aligned}
&\sup_{v_1\in\mathcal{V}:v_1\leq v}\big\{E[Y_i|V_i = v_1] - bE\big[|Z_i - t||V_i = v_1\big]\big\}\\
&\leq E\big[Y_i(t)|V_i = v\big]\\
&\leq \inf_{v_2\in\mathcal{V}:v_2\geq v}\big\{E[Y_i|V_i = v_2] + bE\big[|Z_i - t||V_i = v_2\big]\big\}.
\end{aligned}$$

(iv) *Under CSMTR and MIV together,*

$$\begin{aligned}
&\sup_{v_1\in\mathcal{V}:v_1\leq v}\big\{E[Y_i|V_i = v_1] - bE\big[(Z_i - t)^+|V_i = v_1\big]\big\}\\
&\leq E\big[Y_i(t)|V_i = v\big]\\
&\leq \inf_{v_2\in\mathcal{V}:v_2\geq v}\big\{E[Y_i|V_i = v_2] + bE\big[(Z_i - t)^-|V_i = v_2\big]\big\}.
\end{aligned}$$

These results can be regarded as combinations of Proposition 3.1 and Proposition 1 of Manski and Pepper (2000). It follows immediately from Proposition C.1(iii) and (iv) that the identification region of $g^*(t)$ under the MIV assumption is given as below.

COROLLARY C.2. *Assume that the support of $Y_i(t)$ is unbounded. Let F_V denote the probability measure of V_i. Then the following bounds are sharp*:

(i) *Under CSTR and MIV together,*

$$\begin{aligned}
&\int \sup_{v_1 \in \mathcal{V}: v_1 \leq v} \big\{E[Y_i|V_i = v_1] - bE\big[|Z_i - t||V_i = v_1\big]\big\} F_V(dv) \\
&\leq g^*(t) \\
&\leq \int \inf_{v_2 \in \mathcal{V}: v_2 \geq v} \big\{E[Y_i|V_i = v_2] + bE\big[|Z_i - t||V_i = v_2\big]\big\} F_V(dv).
\end{aligned}$$

(ii) *Under CSMTR and MIV together,*

$$\begin{aligned}
&\int \sup_{v_1 \in \mathcal{V}: v_1 \leq v} \big\{E[Y_i|V_i = v_1] - bE\big[(Z_i - t)^+|V_i = v_1\big]\big\} F_V(dv) \\
&\leq g^*(t) \\
&\leq \int \inf_{v_2 \in \mathcal{V}: v_2 \geq v} \big\{E[Y_i|V_i = v_2] + bE\big[(Z_i - t)^-|V_i = v_2\big]\big\} F_V(dv).
\end{aligned}$$

As in the only-MIV bound of Manski and Pepper (2000), the CSTR-MIV (CSMTR-MIV) bound coincides with the only-CSTR (only-CSMTR) bound if the CSTR (CSMTR) lower and upper bounds for $E[Y_i(t)|V_i = v]$ are weakly increasing in v. Hence, in this case, the MIV assumption has no identifying power. Likewise, if these bounds are weakly decreasing in v, then combining IV with CSTR or CSMTR yields the same result as MIV with CSTR or CSMTR. Thus, in such cases, the MIV assumption has the same identifying power as the IV assumption.

PROOF OF PROPOSITION C.1 AND COROLLARY C.2. *Part* (i). Note that under CSTR, Proposition 3.1(i) leads to

$$E[Y_i|V_i = u] - bE\big[|Z_i - t||V_i = u\big] \leq E\big[Y_i(t)|V_i = u\big] \leq E[Y_i|V_i = u] + bE\big[|Z_i - t||V_i = u\big]$$

which holds for any $u \in \mathcal{V}$. Due to Assumption C.1 (iii), $E[Y_i(t)|V_i = u]$ is no larger than the CSTR upper bound on $E[Y_i(t)|V_i = u']$, and no smaller than the CSTR lower bound, for any $u' \in \mathcal{V}$. There are no other restriction on $E[Y_i(t)|V_i = u]$, so the bound is sharp. *Part* (ii) can be proved in a similar way. For *Part* (iii), we again use the CSTR bound for $E[Y_i(t)|V_i = u]$ presented above. Due to Assumption MIV, $E[Y_i(t)|V_i = u]$ is no smaller than the CSTR lower bound on $E[Y_i(t)|V_i = u_1]$, and no larger than the CSTR upper bound on $E[Y_i(t)|V_i = u_2]$, for any $u_1 \leq u \leq u_2$. There is no other restriction on $E[Y_i(t)|V_i = u]$, so the bound is sharp. *Part* (iv) can be proved in the similar way.

Corollary C.2 can be proved by observing that

$$g^*(t) = \int E\big[Y_i(t)|V_i = v\big] F_V(dv).$$

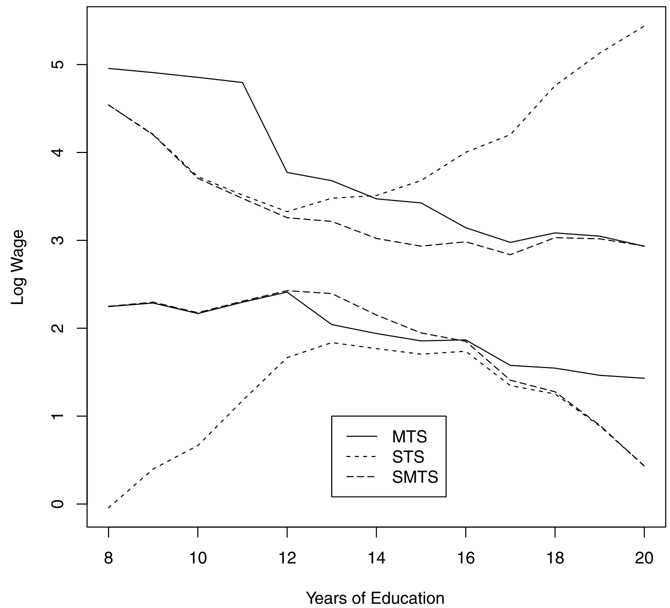

FIGURE A-1. SMTS-STS-MTS comparison.

Setting $E[Y_i(t)|V_i = v]$ at its lower (upper) bound given in Proposition C.1(iii) and (iv) yields the result. □

APPENDIX D: SMOOTH TREATMENT SELECTION

In this section, we introduce the condition that $E[Y_i(t)|Z_i = s]$ is a "smooth" function of s. As in Section 3, we focus on two assumptions on treatment selection: the one we call *smooth treatment selection* (STS) and the other *smooth monotone treatment selection* (SMTS). Both conditions are now stated below in terms of the "local" behavior of $g(t, s)$ with respect to s.

ASSUMPTION D.1 (Treatment Selection Assumptions). *Assume one of the following conditions*:

(i) (Condition STS) *There exists a constant* $a > 0$ *such that* $|g(t, s) - g(t, s')| \leq a|s - s'|\ \forall t, s, s' \in \Gamma$.

(ii) (Condition SMTS) *The STS condition in part* (i) *holds with a constant* $a > 0$. *In addition,* $g(t, s) \geq g(t, s')\ \forall t, s, s' \in \Gamma$ *satisfying* $s \geq s'$.

Note that in both STS and SMTS conditions, we have a bound on changes in $g(t, s)$ with respect to s. The "smoothness" condition in Assumption D.1 can be rewritten as

$$-a \leq \frac{g(t, s) - g\bigl(t, s'\bigr)}{s - s'} \leq a \tag{D.1}$$

for all $s \neq s'$ and t, which is equivalent with $g(t, \cdot)$ having uniformly bounded difference quotients when viewed as a function of only the second argument for each t. The condition in (D.1) assumes that the average outcome cannot change "too much" as the selection of the treatment varies. If we think about plausibility in the context of bounding the return to schooling, the STS assumption is consistent with the economic models that predict that persons who select similar levels of schooling have similar levels of ability on average.

Remark D.1. Assumption D.1 does not seem to be imposed in the literature before; the most related discussions we can find in the literature are from Manski (2003) and Manski and Pepper (forthcoming). Using our notation, equation (9.21) in Section 9.4 of Manski (2003, p. 149) states that

$$\left|E\left[Y_i(t)|V_i=v\right]-E\left[Y_i(t)|V_i=v'\right]\right| \leq C \quad \forall\left(v, v', t\right) \in (\mathcal{V} \times \mathcal{V} \times \Gamma),$$

where $C > 0$ is a specified constant. Manski (2003) motivated this restriction as a form of "approximate" mean independence of instruments but just mentioned it without developing any identification result. Manski and Pepper (forthcoming) considered assumptions of bounded variations; using our notation, a simplified version of their assumption is that for any (t, d, w) and (t', d', w'),

$$C_L \leq E_d\left[Y_i(t)|W_i=w\right]-E_{d'}\left[Y_i\left(t'\right)|W_i=w'\right] \leq C_U,$$

where d and d' refer to possibly different time periods, W_i is a vector of covariates, and C_L and C_U are constants chosen by the researcher. Our STS and SMTS assumptions are distinct and have different motivations since we emphasize the nature of continuity or smoothness of the treatment selection.

The following proposition provides sharp bounds for $g^*(t)$ under these two assumptions.

Proposition D.1. *Assume that the support of $Y_i(t)$ is unbounded. Then the following bounds are sharp*:

(i) *Under STS*, $E[Y_i|Z_i=t]-aE[|Z_i-t|] \leq g^*(t) \leq E[Y_i|Z_i=t]+aE[|Z_i-t|]$.

(ii) *Under SMTS*, $E[Y_i|Z_i=t]-aE[(Z_i-t)^-] \leq g^*(t) \leq E[Y_i|Z_i=t]+aE[(Z_i-t)^+]$.

Proof of Proposition D.1. *Part* (i). We only prove the case of the upper bound. The proof for the lower bound is analogous. Under STS,

$$\begin{aligned}
E\left[Y_i(t)\right] &= \int E\left[Y_i(t)|Z_i=z\right]\mu(dz) \\
&\leq \int \left(E[Y_i|Z_i=t]+a|z-t|\right)\mu(dz) \\
&= E[Y_i|Z_i=t]+aE\left[|Z_i-t|\right].
\end{aligned}$$

For the sharpness, consider a DGP s.t. $E[Y_i(t)|Z_i = s] = E[Y_i(t)|Z_i = t] + a|s - t|\ \forall t, s \in \Gamma$. This ensures that $E[Y_i(t)]$ attains the upper bound. We can also show that this DGP satisfies STS:

$$\begin{aligned} E\big[Y_i(t)|Z_i = s_1\big] &= E\big[Y_i(t)|Z_i = t\big] + a|s_1 - t|, \\ E\big[Y_i(t)|Z_i = s_2\big] &= E\big[Y_i(t)|Z_i = t\big] + a|s_2 - t| \\ \Rightarrow \quad \big|E\big[Y_i(t)|Z_i = s_1\big] - E\big[Y_i(t)|Z_i = s_2\big]\big| &= a\big\||s_1 - t| - |s_2 - t|\big\| \le a|s_1 - s_2|. \end{aligned}$$

On the other hand, the DGP s.t. $E[Y_i(t)|Z_i = s] = E[Y_i(t)|Z_i = t] - a|s - t|\ \forall t, s \in \Gamma$, attains the lower bound, and the convex combinations between the two DGPs yield all the values between the lower and upper bounds. Also, they all obey STS.

Part (ii). We only consider the upper bound. Again, the proof for the lower bound is analogous. Under SMTS,

$$\begin{aligned} E\big[Y_i(t)\big] &= \int_{z \le t} E\big[Y_i(t)|Z_i = z\big]\mu(dz) + \int_{z>t} E\big[Y_i(t)|Z_i = z\big]\mu(dz) \\ &\le \int_{z \le t} \big(E[Y_i|Z_i = t] + 0\big)\mu(dz) + \int_{z>t} \big(E[Y_i|Z_i = t] + a(z - t)\big)\mu(dz) \\ &= E[Y|Z_i = t] + aE\big[(Z_i - t)^+\big]. \end{aligned}$$

For the sharpness, consider a DGP s.t. $E[Y_i(t)|Z_i = s] = E[Y_i(t)|Z_i = t]$ when $s \le t$ and $E[Y_i(t)|Z_i = s] = E[Y_i(t)|Z_i = t] + a(s - t)$ when $s > t$. This ensures that $E[Y_i(t)]$ attains the upper bound. To show that this DGP satisfies SMTS, note that for any s_1 and s_2 satisfying $s_1 > s_2$, we have

$$E\big[Y_i(t)|Z_i = s_1\big] - E\big[Y_i(t)|Z_i = s_2\big] = \begin{cases} 0, & \text{if } s_2 < s_1 \le t, \\ a(s_1 - s_2), & \text{if } s_1 > s_2 > t, \\ a(s_1 - t), & \text{if } s_1 > t \ge s_2. \end{cases}$$

This implies that SMTS holds since $(s_1 - t) \le (s_1 - s_2)$ when $s_1 > t \ge s_2$. Then the rest can be proved as in part (ii) of the proof of Proposition 3.1. □

The interpretation of Proposition D.1 is similar to that of Proposition 3.1. Proposition D.1(i) states that under STS, the sharp bound is symmetric around $E[Y_i|Z_i = t]$ and its width is $2aE[|Z_i - t|]$. Proposition D.1(ii) implies that under SMTS, the sharp bound is asymmetric around $E[Y_i|Z_i = t]$ and also its width is just $aE[|Z_i - t|]$ (the half of the width under STS). Again, the width is minimized when the counterfactual treatment value is the median of Z_i.

As in the case of the treatment response assumptions, the identification region of $g^*(t)$ is unbounded when only the MTS condition in the equation (4.1) is assumed (see Proposition 1, Corollary 2 of Manski and Pepper (2000)). This implies that the STS assumption can provide additional information for identification when the support of $Y_i(t)$ is unbounded.

When the support of $Y_i(t)$ is bounded, that is, $Y_i(t) \le y_{\max} < \infty$ for some known $y_{\max}$, we can show that the upper bound for $g^*(t)$ isp

$$g^*(t) \le \int_{z>t} \min\{y_{\max}, \big(E[Y_i|Z_i=t] + a(z-t)\big)\}\mu(dz) + E[Y_i|Z_i=t]P(Z_i \le t). \tag{D.2}$$

The upper bound (D.2) cannot be larger than the upper bound under the MTS assumption alone since the latter has the form (see again Proposition 1, Corollary 2 of Manski and Pepper (2000)):

$$g^*(t) \le y_{\max} P(Z_i > t) + E\big[Y_i(t)|Z_i = t\big]P(Z_i \le t). \tag{D.3}$$

Similar to the discussion under SMTR, note that the SMTS upper bound strictly improves the MTS upper bound if and only if the event such that $E[Y_i|Z_i=t] + a(Z_i - t) < y_{\max}$ has a strictly positive probability, conditional on $Z_i > t$. Analogous results can be established for the lower bound, and we summarize our findings below.

Corollary D.2. *Assume that the support of $Y_i(t)$ is $[y_{\min}, y_{\max}]$, where $-\infty \le y_{\min} \le y_{\max} \le \infty$. Then we have*:

(i) *The upper bound of the SMTS bound is strictly smaller than that of the MTS bound if and only if $\int_{z>t} 1\{U_{\text{SMTS}}(t,z) < 0\}\mu(dz) > 0$, where $U_{\text{SMTS}}(t,z) \equiv E[Y_i|Z_i=t] + a(z-t) - y_{\max}$.*

(ii) *The lower bound of the SMTS bound is strictly larger than that of the MTS bound if and only if $\int_{z<t} 1\{L_{\text{SMTS}}(t,z) > 0\}\mu(dz) > 0$, where $L_{\text{SMTS}}(t,z) \equiv E[Y_i|Z_i=t] - a(t-z) - y_{\min}$.*

The proof of Corollary D.2 is omitted since it is straightforward. As in the SMTR case, one can test whether there is an strict improvement using Corollary D.2.

D.1 *The STR and STS bounds*

If we combine STR with STS, we obtain the following result.

Proposition D.3. *Assume that the support of $Y_i(t)$ is unbounded. Then, under STR and STS together, $E[Y_i(t)] \in [l_2(t), u_2(t)]$, where*

$$l_2(t) \equiv \int \max\{E[Y_i|Z_i=t] - a|z-t|, E[Y_i|Z_i=z] - b|z-t|\}\mu(dz),$$

$$u_2(t) \equiv \int \min\{E[Y_i|Z_i=t] + a|z-t|, E[Y_i|Z_i=z] + b|z-t|\}\mu(dz).$$

Moreover, this bound is sharp.

In contrast to the case where only one of STS and STR holds, the length of the identification region is generally not minimized at the median of Z_i. To make a comparison with the STR bound in Proposition 3.1 and the STS bound in Proposition D.1, we present the special case such that $a = b$ as the following corollary.

COROLLARY D.4. *Suppose $a = b = \bar{k}$, where $\bar{k}$ denotes the common value. Define $A(t)$ as the event such that $E[Y_i|Z_i = t] \le E[Y_i|Z_i]$ for each t. Assume that the support of $Y_i(t)$ is unbounded. Then, under STR and STS together, $E[Y_i(t)] \in [l_3(t), u_3(t)]$, where*

$$l_3(t) \equiv E[Y_i|Z_i = t]P\big(A(t)^c\big) + E\big[Y_i|A(t)\big]P\big(A(t)\big) - \bar{k}E\big[|Z_i - t|\big],$$
$$u_3(t) \equiv E[Y_i|Z_i = t]P\big(A(t)\big) + E\big[Y_i|A(t)^c\big]P\big(A(t)^c\big) + \bar{k}E\big[|Z_i - t|\big].$$

As a polar case, suppose that $E[Y_i|Z_i = t] \le E[Y_i|Z_i]$ holds with probability one. Then the lower and upper bounds of the STR-STS bound reduce to

$$l_3(t) = E[Y_i] - \bar{k}E\big[|Z_i - t|\big] \quad \text{and} \quad u_3(t) = E[Y_i|Z_i = t] + \bar{k}E\big[|Z_i - t|\big],$$

respectively. Thus, in this case, as long as $E[Y_i|Z_i = t] < E[Y_i]$, we can conclude that the upper bound of the STR-STS bound is strictly smaller than that of the STR bound in Proposition 3.1 and that the lower bound of the STR-STS bound is strictly larger than that of the STS bound in Proposition 3.1.

PROOF OF PROPOSITION D.3 AND COROLLARY D.4. The bounds in Proposition D.3 naturally follow from Propositions 3.1 and D.1. For the sharpness, note that there are two cases to consider; STR and STS can hold with (1) $a \ge b$ or (2) $a < b$. For case (1), consider a DGP s.t. $E[Y_i|Z_i = z] = c \; \forall z$, where c indicates some constant. Furthermore, suppose $g(t,s) = g(s,s) + b|t-s| = c + b|t-s|$. Note that in case (1), $\int \min\{E[Y_i|Z_i = t] + a|z - t|, E[Y_i|Z_i = z] + b|z-t|\}\mu(dz) = \int \min\{c + a|z-t|, c + b|z-t|\}\mu(dz) = \int c + b|z-t|\mu(dz)$. Moreover, note that $E[Y_i(t)] = \int E[Y_i(t)|Z_i = z]\mu(dz) = \int c + b|t-z|\mu(dz)$. Therefore, the upper bound is sharp in this case. Likewise, if we change the DGP into $g(t,s) = g(s,s) - b|t-s| = c - b|t-s|$, we can show that the lower bound is also sharp. Finally, the DGP s.t. $g(t,s) = g(s,s) + k|t-s| = c + k|t-s|$, $s \in (-b, b)$ generates different values for $E[Y_i(t)]$ which are between the upper and the lower bound.

It remains to show these DGPs satisfy STR and STS. However, this can be easily checked since these DGPs have the same form as in the DGPs appearing in the proofs for Propositions 3.1 and D.1. For case (2), replacing b with a leads to the analogous argument which completes the proof.

For the upper bound in Corollary D.4, note that

$$\begin{aligned} g^*(t) &\le \int \min\big(E[Y_i|Z_i = t] + \bar{k}|z-t|, E[Y_i|Z_i = z] + \bar{k}|z-t|\big)\mu(dz) \\ &= \int \big[\min\big(E[Y_i|Z_i = t], E[Y_i|Z_i = z]\big) + \bar{k}|z-t|\big]\mu(dz) \\ &= \int \min\big(E[Y_i|Z_i = t], E[Y_i|Z_i = z]\big)\mu(dz) + \bar{k}E|Z_i - t| \\ &= \int_{A(t)} E[Y_i|Z_i = t]\mu(dz) + \int_{A(t)^c} E[Y_i|Z_i = z]\mu(dz) + \bar{k}E|Z_i - t|. \end{aligned}$$

The result for the lower bound can be shown similarly. □

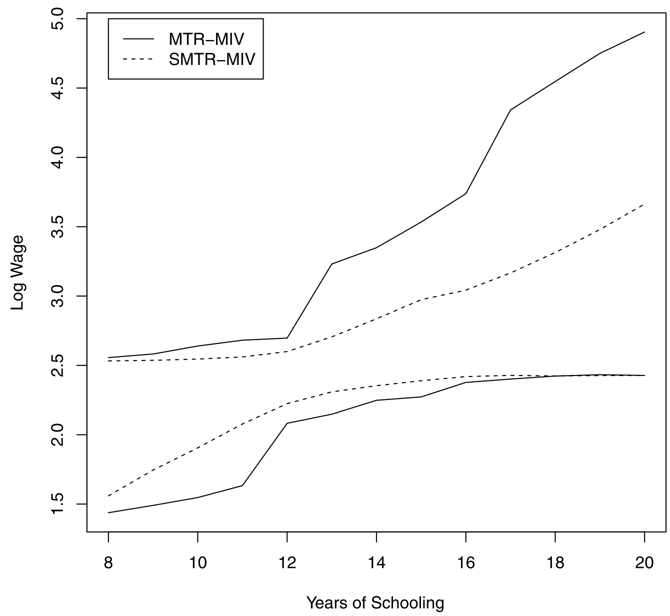

FIGURE A-2. SMTR-MTR comparison when combined with MIV.

D.2 *Numerical illustration: Manski and Pepper (2000) revisited*

In this subsection, we go back to the returns to schooling example of Manski and Pepper (2000) in Section 5 and illustrate the usefulness of STR and SMTS assumptions. Figure A-1 shows the SMTS, STS, and MTS bounds when the value of a is 0.4. Here, SMTS and STS bounds are calculated under the assumption that log wages are unbounded. It seems more difficult to come up with a reasonable value of a in this example. We set $a = 2b = 0.4$ to have a relatively large value for a. The estimation results are similar to those in Figure 2. Again Figure A-1 shows that there could be a substantial advantage if one combines the smoothness condition with the monotonicity assumption.

APPENDIX E: INFERENCE

In this section, we provide discussions on inference using the identification results obtained in the paper and give directions for further research by mentioning open questions in inference methods.

E.1 *Inference using Proposition 3.1*

We first describe how to carry out inference under STR, following Imbens and Manski (2004) and Stoye (2009). First, define

$$\widehat{\theta}_\ell \equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i - b|Z_i - t|\big],$$

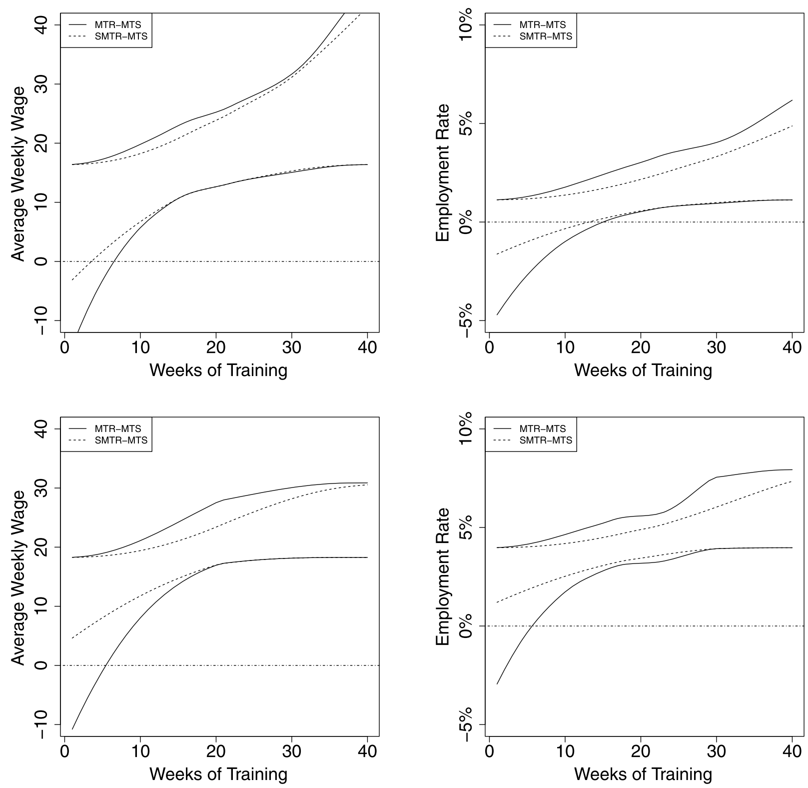

FIGURE A-3. Comparison between the MTR-MTS and SMTR-MTS bounds. *Notes*: Each figure compares the MTR-MTS bounds with the SMTR-MTS bounds. The top panels show results for males and the bottom panels for females, respectively. In each row, the left and right figures correspond to weekly earnings and employment rates, respectively.

$$\widehat{\theta}_u \equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i + b|Z_i - t|\big],$$

$$\widehat{\sigma}_\ell^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i - b|Z_i - t|\big]^2 - \widehat{\theta}_\ell^2,$$

$$\widehat{\sigma}_u^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i + b|Z_i - t|\big]^2 - \widehat{\theta}_u^2,$$

and $\widehat{\Delta} \equiv 2b\,n^{-1}\sum_{i=1}^{n}|Z_i - t|$. For each t, let

$$\mathrm{CI}_\alpha^{\mathrm{STR}}(t) \equiv \left[\widehat{\theta}_\ell - \frac{c_\alpha \widehat{\sigma}_\ell}{\sqrt{n}}, \widehat{\theta}_u + \frac{c_\alpha \widehat{\sigma}_u}{\sqrt{n}}\right],$$

where c_α solves

$$\Phi\left(c_\alpha + \frac{\sqrt{n}\widehat{\Delta}}{\max\{\widehat{\sigma}_\ell, \widehat{\sigma}_u\}}\right) - \Phi(-c_\alpha) = 1 - \alpha.$$

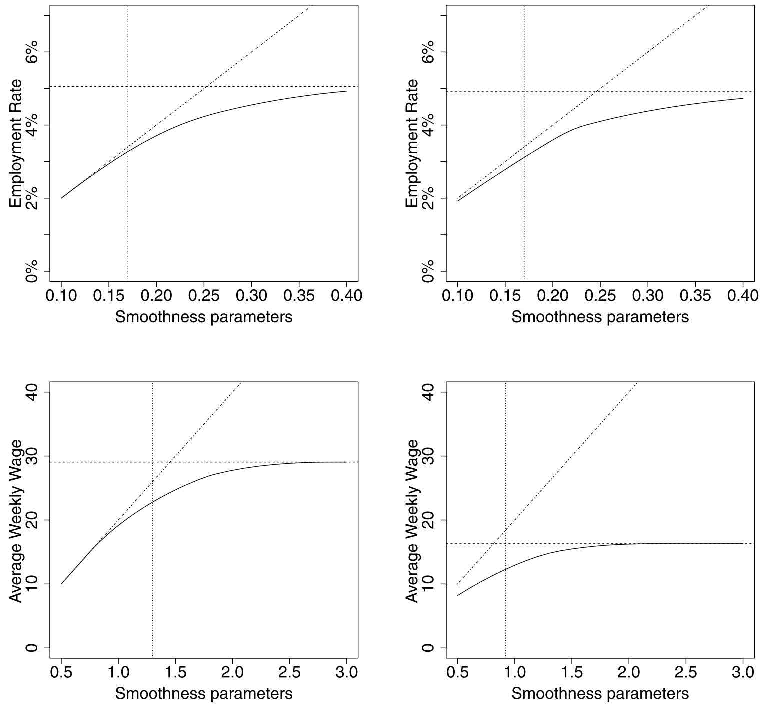

FIGURE A-4. Sensitivity analysis on the upper bound of $\Delta(16, 36)$. *Notes*: The top and bottom panels show sensitivity analysis results on $\Delta(16, 36)$ for the employment probability in percentage and weekly earnings, respectively. In each row, the left and right panels show the results for male and females, respectively. The dash horizontal lines show the MTR-MTS bounds and the dash-dot upward sloping lines represent $20b$. The dotted vertical lines correspond to the baseline values of b used in Table 1.

Since $\widehat{\theta}_\ell \leq \widehat{\theta}_u$ by construction, Lemma 3 and Proposition 1 of Stoye (2009) imply that $g^*(t) \in \mathrm{CI}_\alpha^{\mathrm{STR}}(t)$ with probability $1-\alpha$ uniformly as $n \to \infty$, provided that the data generating process satisfies mild regularity conditions given in Assumption 1(i) and (ii) of Stoye (2009). Note that the confidence interval in $\mathrm{CI}_\alpha^{\mathrm{STR}}(t)$ is pointwise in t. It would require more complicated approximations than simple normal approximations in Imbens and Manski (2004) and Stoye (2009) to obtain a uniform confidence band for $g^*(t)$.

Analogously, we can obtain a confidence interval for $g^*(t)$ under SMTR by redefining $\widehat{\theta}_\ell$, $\widehat{\theta}_u$, $\widehat{\sigma}_\ell^2$, $\widehat{\sigma}_u^2$, and $\widehat{\Delta}$ as the following:

$$\widehat{\theta}_\ell \equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i - b(Z_i - t)^+\big],$$

$$\widehat{\theta}_u \equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i + b(Z_i - t)^-\big],$$

$$\widehat{\sigma}_\ell^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i - b(Z_i - t)^+\big]^2 - \widehat{\theta}_\ell^2,$$

$$\widehat{\sigma}_u^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i + b(Z_i - t)^-\big]^2 - \widehat{\theta}_u^2,$$

$$\widehat{\Delta} \equiv b\frac{1}{n}\sum_{i=1}^{n}|Z_i - t|.$$

One may develop alternative methods for inference, noting that the bounds in Proposition 3.1 can be expressed as unconditional moment inequality restrictions. Existing inference methods include Andrews and Barwick (2012), Andrews and Guggenberger (2009), Andrews and Soares (2010), Beresteanu and Molinari (2008), Bugni (2010), Canay (2010), Chernozhukov, Hong, and Tamer (2007), Galichon and Henry (2009), Romano and Shaikh (2008, 2010), and Rosen (2008) among others.

E.2 *Inference using Proposition C.1*

The identification region obtained in each case of Proposition C.1 corresponds to the form of intersection bounds considered in Chernozhukov, Lee, and Rosen (2013). Therefore, a pointwise confidence interval for $g^*(t)$ can be obtained, following the inference method developed in Chernozhukov, Lee, and Rosen (2013) directly. Alternatively, one can use inference methods developed for conditional moment inequalities,

TABLE A-1. The upper bounds for $\Delta(12, 16)$ and $\Delta(16, 18)$.

Smoothness	Parameter of Interest	
b	$\Delta(12, 16)$	$\Delta(16, 18)$
MTR-MTS	0.316	0.338
SMTR-MTS		
0.02	0.080	0.040
0.04	**0.155**	0.080
0.06	**0.219**	0.120
0.08	**0.262**	**0.159**
0.10	**0.284**	**0.189**
0.12	**0.302**	**0.217**
0.14	**0.314**	**0.243**
0.16	0.316	**0.264**
0.18	0.316	**0.284**
0.20	0.316	**0.303**
0.22	0.316	**0.323**
0.24	0.316	**0.326**
0.26	0.316	**0.329**
0.28	0.316	**0.332**
0.30	0.316	**0.334**
0.32	0.316	**0.336**
0.34	0.316	**0.337**
0.36	0.316	0.338

Note: The bold font corresponds to the case when the upper bound for $\Delta(t_1, t_2)$ is strictly less than the MTR-MTS bound and also strictly less than $b(t_2 - t_1)$.

such as Andrews and Shi (2013), Armstrong (2014, 2015), Armstrong and Chan (2016), Chetverikov (2018), and Lee, Song, and Whang, Lee, Song, and Whang (2017, 2013) among others. Among these methods, Lee, Song, and Whang (2017) can be used to obtain a uniform confidence band for $g^*(t)$.

As an illustration of the inference method in this section, we compare the MTR-MIV bound with the SMTR-MIV bound by revisiting the return to education example using the data from the National Longitudinal Survey of Youth of 1979.[2] Here, $Y_i(t)$ is the counterfactual log wage given that the individual received t years of schooling, and V_i is the Armed Forces Qualifying Test (AFQT) score which is used as MIV; that is, those with higher AFQT scores will earn more wages on average. We used $b = 0.2$ for the smoothness parameter. Figure A-2 shows the pointwise 95% confidence intervals for $E[Y_i(t)|V_i = 0]$ using both the MTR-MIV and SMTR-MIV bounds.[3] Each confidence interval was obtained by the inference method of Chernozhukov, Lee, and Rosen (2013) using a STATA command in Chernozhukov, Kim, Lee, and Rosen (2015).[4] We can observe that imposing the smoothness assumption substantially tightens the original MTR-MIV confidence interval, in particular the upper confidence interval at more than 12 years of schooling.

[2]In particular, we use the same data extract as Carneiro and Lee (2009). See also Carneiro, Heckman, and Vytlacil (2011) for the dataset and recent advances in estimating returns to schooling.

[3]The variable V_i was normalized so that it has mean zero and variance one in the NLSY population.

[4]In particular, the series estimator with cubic B-splines was employed. The pointwise confidence intervals were obtained by inverting a test, which is implementable by the `clr3bound` command in Chernozhukov et al. (2015).

Table A-2. Summary statistics on the job corps data.

Group	N	Mean	St. Dev.	Min	Max	Median
Observed outcome: Weekly earnings 48 months after random assignment (in dollars)						
Treatment	4207	218.81	219.38	0.00	1890.49	202.10
Control	4099	202.24	207.44	0.00	2388.1	187.66
Treatment (male)	2376	249.94	236.87	0.00	1890.49	245.35
Control (male)	2524	225.83	216.57	0.00	2345.77	217.06
Treatment (female)	1831	178.43	186.84	0.00	1884.65	150.13
Control (female)	1575	164.45	185.82	0.00	2388.10	130.42
Observed outcome: Whether employed (in percentage)						
Treatment	4207	71.26	45.26	0	100	100
Control	4099	68.46	46.47	0	100	100
Treatment (male)	2376	73.06	44.37	0	100	100
Control (male)	2524	70.64	45.55	0	100	100
Treatment (female)	1831	68.92	46.29	0	100	100
Control (female)	1575	64.95	47.73	0	100	100
Realized treatment: Training weeks						
Treatment	4207	30.14	27.42	0.06	204.01	22.50
Treatment (male)	2376	29.64	27.45	0.06	204.01	21.95
Treatment (female)	1831	30.79	27.37	0.10	192.80	23.31

Note: This table gives summary statistics of the data extract used in this paper. All figures were calculated using design weights.

E.3 *Open questions in inference*

The existing literature does not provide inference methods for all the bounds we developed in this paper. First, the bounds given in Corollary C.2 differ from the intersection bounds; they are rather averages of intersection bounds. To our knowledge, there does not exist a suitable inference method yet in the literature. It is an interesting future research topic to develop inference methods for such bounds, including bounds given in equations (9) and (17) of Manski and Pepper (2000).

The SMTR-MTS bounds in Proposition 4.1 is difficult to deal with. Note that the SMTR-MTS bounds can be estimated consistently by plugging in suitable sample analogs; however, they are not sufficiently smooth functionals of the underlying population distribution. Santos (2012), Fang and Santos (2015), and Chernozhukov, Newey, and Santos (2015) have developed general inference methods for nonparametric functionals in models with partial identification; however, it is an open question how to carry out inference for our bounds by extending their results or by developing new tools of inference.

TABLE A-3. Empirical results by assumptions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Lower Bound				Upper Bound			
Training Duration (in Weeks)	MTR	SMTR	MTR + MTS	SMTR + MTS	MTR	SMTR	MTR + MTS	SMTR + MTS
Panel (a)—Males								
Outcome: employment status (percentage employed)								
4	−63.61	−1.20	−3.13	−1.15	4.66	1.16	1.23	1.16
16	−35.87	0.21	0.14	0.26	16.47	1.79	2.54	1.79
36	−3.73	1.10	1.09	1.10	27.83	4.29	5.20	4.22
Outcome: weekly earnings in US dollars (including zero earnings)								
4	−204.51	−1.40	−5.59	0.54	193.49	16.62	16.96	16.62
16	−111.34	9.38	11.26	11.26	848.16	21.43	23.55	21.43
36	−0.87	16.21	16.15	16.21	1560.02	40.61	40.32	38.00
Panel (b)—Females								
Outcome: employment status (percentage employed)								
4	−59.09	1.52	−0.89	1.69	7.33	4.00	4.09	4.00
16	−33.96	2.99	2.94	3.16	19.75	4.56	5.34	4.56
36	−0.73	3.95	3.97	3.97	33.50	7.01	7.85	6.85
Outcome: weekly earnings in US dollars (including zero earnings)								
4	−147.26	4.99	−3.17	7.24	173.89	18.41	18.69	18.41
16	−82.87	12.97	14.50	15.22	817.33	21.47	24.87	21.47
36	5.84	18.14	18.26	18.26	1614.05	34.70	30.79	29.99

Note: The table shows the lower and upper bounds of the average treatment effect $E[Y_i(t)] - E[Y_i | i \in \text{control group}]$, where the length of enrollment to the program (t) is 4, 16, and 36 weeks.

Table A-4. Sensitivity analysis on the upper bound of $\Delta(16, 36)$.

Y = Employment Probability		Y = Weekly Earning	
Panel (a)—Male			
MTR-MTS	5.06	MTR-MTS	29.06
SMTR-MTS		SMTR-MTS	
$b = 0.1$	2.00	$b = 0.5$	10.00
0.15	**2.94**	1	**19.15**
0.2	**3.71**	1.5	**24.71**
0.25	**4.23**	2	**27.77**
0.3	**4.55**	2.5	**28.86**
0.35	**4.78**	3	29.06
0.4	**4.93**		
Effective region: (0.10, 0.40)		Effective region: (0.7, 2.8)	
Panel (b)—Female			
MTR-MTS	4.91	MTR-MTS	16.29
SMTR-MTS		SMTR-MTS	
$b = 0.1$	**1.92**	$b = 0.5$	**8.20**
0.15	**2.78**	1	**12.92**
0.2	**3.59**	1.5	**15.47**
0.25	**4.10**	2	**16.23**
0.3	**4.38**	2.5	16.29
0.35	**4.59**	3	16.29
0.4	**4.73**		
Effective region: (0.10, 0.40)		Effective region: (0.5, 2.2)	

Note: The employment probability is in percentage. The bold font corresponds to the case when the upper bound for $\Delta(16, 36)$ is strictly less than the MTR-MTS bound and also strictly less than $20b$.

Appendix F: Additional empirical results

In this section, we provide supplementary empirical results that are omitted from the main text. Table A-1 gives the upper bounds for $\Delta(12, 16)$ and $\Delta(16, 18)$ shown in Figure 3. Table A-2 presents summary statistics for Job Corps data. Extra empirical results using Job Corps data are grouped into the following two subsections.

F.1 *Effects of Job Corps for subgroups*

In this subsection, we revisit the example of Section 6 and repeat the analysis for the subgroups defined by gender. If we look at Table A-3, we can see that there are some differences between males and females. It seems that the average treatment effect for the employment status is larger for females but that for the weekly earnings is higher for males especially at $t = 36$. The difference between the MTR-MTS and SMTR-MTS assumptions is largest in the lower bound for female earnings at week four; adding the smoothness tightens the bound substantially (by more than 10 dollars), so that the negative lower MTR-MTS bound (-3.17) becomes the positive SMTR-MTS bound (7.24). Fig-

TABLE A-5. Sensitivity Analysis (Y = Employment Probability).

Panel (a)—All individuals

	$E[Y_i(t)] - E[Y_i \mid i \in$ control group] at:					
	$t = 4$		$t = 16$		$t = 36$	
	LB	UB	LB	UB	LB	UB
MTR-MTS	−1.98	2.37	1.31	3.65	2.25	6.24
SMTR-MTS						
$b = 0.1$	**0.89**	**2.28**	**1.72**	**2.65**	2.25	**4.12**
0.15	**0.20**	**2.29**	**1.46**	**2.84**	2.25	**5.04**
0.2	**−0.36**	**2.30**	1.31	**3.03**	2.25	**5.82**
0.25	**−0.77**	**2.31**	1.31	**3.22**	2.25	**6.17**
0.3	**−1.17**	**2.32**	1.31	**3.41**	2.25	6.24
0.35	**−1.46**	**2.33**	1.31	**3.55**	2.25	6.24
0.4	**−1.70**	**2.34**	1.31	**3.63**	2.25	6.24

Panel (b)—Male

	$E[Y_i(t)] - E[Y_i \mid i \in$ control group] at:					
	$t = 4$		$t = 16$		$t = 36$	
	LB	UB	LB	UB	LB	UB
MTR-MTS	−3.13	1.23	0.14	2.54	1.09	5.20
SMTR-MTS						
b =0.1	**−0.24**	**1.14**	**0.59**	**1.51**	**1.11**	**2.99**
0.15	**−0.90**	**1.15**	**0.34**	**1.71**	**1.10**	**3.89**
0.2	**−1.49**	**1.16**	**0.17**	**1.90**	**1.10**	**4.64**
0.25	**−1.94**	**1.17**	0.14	**2.10**	1.09	**5.11**
0.3	**−2.33**	**1.18**	0.14	**2.29**	1.09	5.20
0.35	**−2.65**	**1.19**	0.14	**2.45**	1.09	5.20
0.4	**−2.88**	**1.20**	0.14	**2.53**	1.09	5.20

Panel (c)—Female

	$E[Y_i(t)] - E[Y_i \mid i \in$ control group] at:					
	$t = 4$		$t = 16$		$t = 36$	
	LB	UB	LB	UB	LB	UB
MTR-MTS	−0.89	4.09	2.94	5.34	3.97	7.85
SMTR-MTS						
$b = 0.1$	**2.58**	**3.99**	**3.45**	**4.32**	3.97	**5.72**
0.15	**1.94**	**4.00**	**3.24**	**4.49**	3.97	**6.54**
0.2	**1.32**	**4.00**	**3.05**	**4.67**	3.97	**7.30**
0.25	**0.78**	**4.01**	**2.95**	**4.84**	3.97	**7.71**
0.3	**0.38**	**4.02**	2.94	**5.01**	3.97	**7.84**
0.35	**0.06**	**4.03**	2.94	**5.15**	3.97	7.85
0.4	**−0.23**	**4.04**	2.94	**5.26**	3.97	7.85

Note: The employment probability is in percentage. The bold font corresponds to the case when the SMTR-MTS lower bound (LB) or upper bound (UB) strictly tightens the equivalent of the MTR-MTS bound.

Table A-6. Sensitivity Analysis ($Y =$ Weekly Earning).

Panel (a)—All individuals

	$E[Y_i(t)] - E[Y_i \mid i \in$ control group] at:					
	$t = 4$		$t = 16$		$t = 36$	
	LB	UB	LB	UB	LB	UB
MTR-MTS	−0.73	18.12	13.31	23.67	17.56	36.03
SMTR-MTS						
$b = 0.5$	**10.79**	**17.78**	**14.97**	**19.61**	**17.63**	**26.95**
1	**4.97**	**17.87**	**13.32**	**21.53**	**17.57**	**34.78**
1.5	**1.25**	**17.97**	13.31	**23.28**	17.56	36.03
2	**−0.60**	**18.06**	13.31	23.67	17.56	36.03
2.5	−0.73	18.12	13.31	23.67	17.56	36.03
3	−0.73	18.12	13.31	23.67	17.56	36.03

Panel (b)—Male

	$E[Y_i(t)] - E[Y_i \mid i \in$ control group] at:					
	$t = 4$		$t = 16$		$t = 36$	
	LB	UB	LB	UB	LB	UB
MTR-MTS	−5.59	16.96	11.26	23.55	16.16	40.32
SMTR-MTS						
$b = 0.5$	**9.54**	**16.47**	**13.69**	**18.32**	**16.31**	**25.70**
1	**3.19**	**16.57**	**11.49**	**20.27**	**16.25**	**34.54**
1.5	**−0.98**	**16.66**	11.26	**22.18**	**16.19**	**39.35**
2	**−3.78**	**16.76**	11.26	**23.40**	16.16	40.32
2.5	**−5.36**	**16.86**	11.26	23.55	16.16	40.32
3	−5.59	**16.95**	11.26	23.55	16.16	40.32

Panel (c)—Female

	$E[Y_i(t)] - E[Y_i \mid i \in$ control group] at:					
	$t = 4$		$t = 16$		$t = 36$	
	LB	UB	LB	UB	LB	UB
MTR-MTS	−3.17	18.69	14.50	24.87	18.26	30.79
SMTR-MTS						
$b = 0.5$	**11.79**	**18.34**	**16.13**	**20.01**	18.26	**26.08**
1	**6.38**	**18.42**	**15.05**	**21.75**	18.26	**30.36**
1.5	**1.52**	**18.51**	14.50	**23.49**	18.26	30.79
2	**−1.78**	**18.59**	14.50	**24.78**	18.26	30.79
2.5	**−3.16**	**18.67**	14.50	24.87	18.26	30.79
3	−3.17	18.69	14.50	24.87	18.26	30.79

Note: The bold font corresponds to the case when the SMTR-MTS lower bound (LB) or upper bound (UB) strictly tightens the equivalent of the MTR-MTS bound.

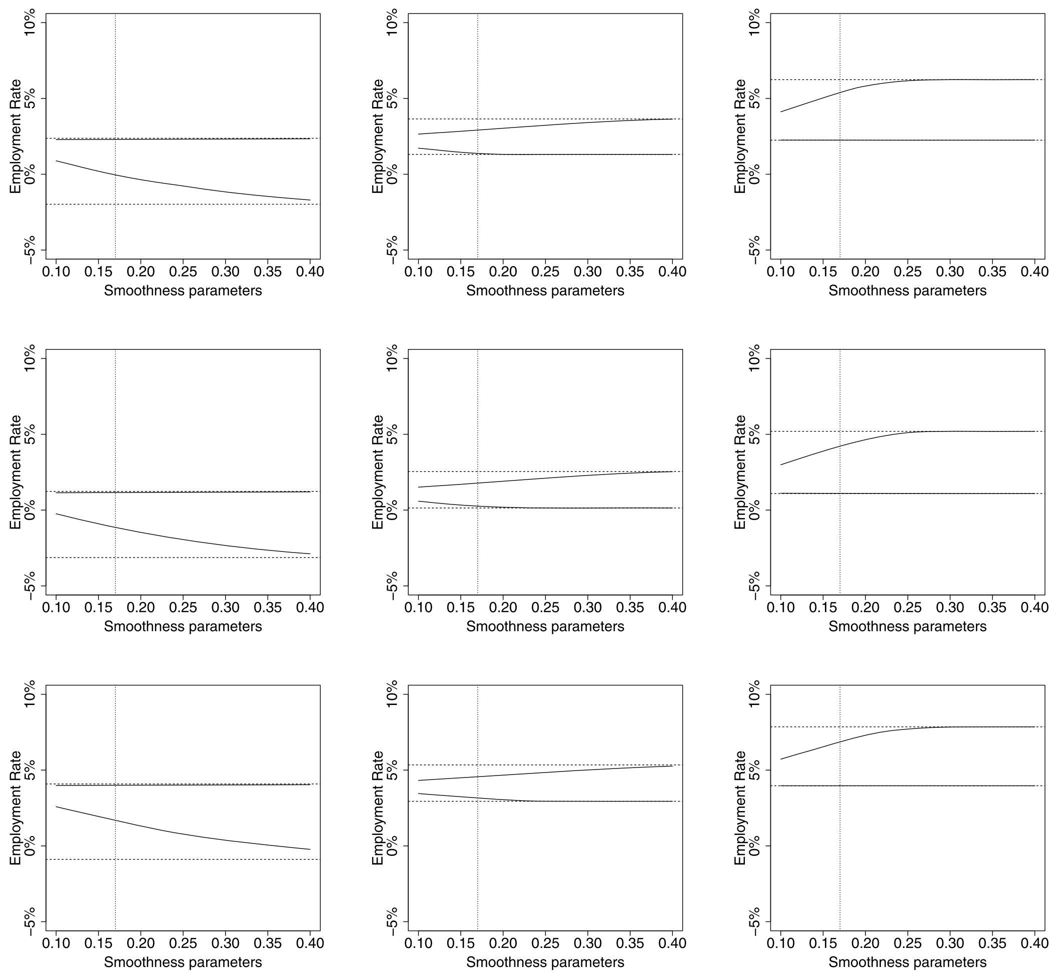

FIGURE A-5. Sensitivity Analysis ($Y =$ Employment Probability). *Notes*: The top, middle and bottom panels show the SMTR-MTS bounds as functions of b for all individuals, male and females, respectively. The average treatment effects at $t = 4$, 16, 36 weeks are shown from left to right, respectively. The dash horizontal lines depict the MTR-MTS bounds. The dotted vertical lines correspond to the baseline values of b used in Table 1.

ure A-3 gives a graphical summary similar to Figure 4. Table A-4 and Figure A-4 reports subgroup-specific sensitivity analysis results analogous to Table 2 and Figure 5.

F.2 *Additional sensitivity analyses for effects of Job Corps*

In this subsection, we report sensitivity analyses by varying the values of b for the average treatment effects reported in Table 1. Table A-5 compares the MTR-MTS bounds and the SMTR-MTS bounds for different b's, when the outcome variable is the employment rates. Table A-6 does the same for weekly earnings. Figures A-5 and A-6 present

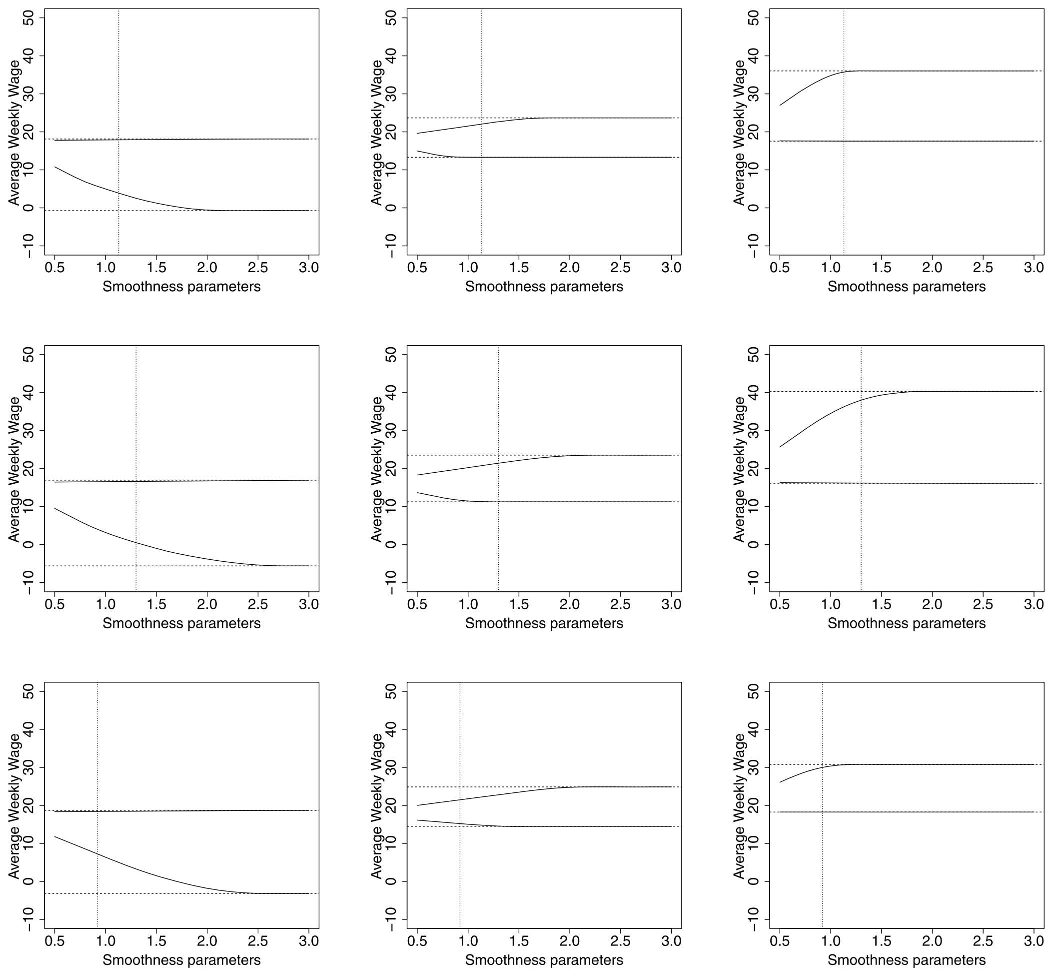

FIGURE A-6. Sensitivity analysis (Y = Weekly Earning). *Notes*: The top, middle, and bottom panels show the SMTR-MTS bounds as functions of b for all individuals, male and females, respectively. The average treatment effects at $t = 4$, 16, 36 weeks are shown from left to right, respectively. The dash horizontal lines depict the MTR-MTS bounds. The dotted vertical lines correspond to the baseline values of b used in Table 1.

the graphical representation of the sensitivity analysis results. The improvement is most noticeable at $t = 4$ and also for subsamples of males and females.

REFERENCES

Andrews, D. W. K. and P. J. Barwick (2012), "Inference for parameters defined by moment inequalities: A recommended moment selection procedure." *Econometrica*, 80 (6), 2805–2826. [18]

Andrews, D. W. K. and P. Guggenberger (2009), "Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities." *Econometric Theory*, 25, 669–709. [18]

Andrews, D. W. K. and X. Shi (2013), "Inference based on conditional moment inequalities." *Econometrica*, 81 (2), 609–666. [19]

Andrews, D. W. K. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78 (1), 119–157. [18]

Armstrong, T. B. (2014), "Weighted KS statistics for inference on conditional moment inequalities." *Journal of Econometrics*, 181 (2), 92–116. [19]

Armstrong, T. B. (2015), "Asymptotically exact inference in conditional moment inequality models." *Journal of Econometrics*, 186 (1), 51–65. [19]

Armstrong, T. B. and H. P. Chan (2016), "Multiscale adaptive inference on conditional moment inequalities." *Journal of Econometrics*, 194 (1), 24–43. [19]

Beresteanu, A. and F. Molinari (2008), "Asymptotic properties for a class of partially identified models." *Econometrica*, 76 (4), 763–814. [18]

Bugni, F. (2010), "Bootstrap inference for partially identified models defined by moment inequalities: Coverage of the identified set." *Econometrica*, 78 (2), 735–753. [18]

Canay, I. (2010), "EL inference for partially identified models: Large deviations optimality and bootstrap validity." *Journal of Econometrics*, 156 (2), 408–425. [18]

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011), "Estimating marginal returns to education." *American Economic Review*, 101 (6), 2754–2781. [19]

Carneiro, P. and S. Lee (2009), "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality." *Journal of Econometrics*, 149, 191–208. [19]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75 (5), 1243–1284. [18]

Chernozhukov, V., W. Kim, S. Lee, and A. Rosen (2015), "Implementing intersection bounds in Stata." *Stata Journal*, 15 (1), 21–44. [19]

Chernozhukov, V., S. Lee, and A. Rosen (2013), "Intersection bounds: Estimation and inference." *Econometrica*, 81 (2), 667–737. [18, 19]

Chernozhukov, V., W. Newey, and A. Santos (2015), "Constrained conditional moment restriction models." CeMMAP Working Paper CWP59/15. [20]

Chetverikov, D. (2018), "Adaptive Tests of Conditional Moment Inequalities." *Econometric Theory*, 34, 186–227. [19]

Fang, Z. and A. Santos (2015), "Inference on directionally differentiable functions." ArXiv working paper, available at http://arxiv.org/abs/1404.3763. [20]

Galichon, A. and M. Henry (2009), "A test of non-identifying restrictions and confidence regions for partially identified parameters." *Journal of Econometrics*, 159 (2), 186–196. [18]

Imbens, G. and C. F. Manski (2004), "Confidence intervals for partially identified parameters." *Econometrica*, 72 (6), 1845–1857. [15, 17]

Lafférs, L. (2013), "A note on bounding average treatment effects." *Economics Letters*, 120 (3), 424–428. [7]

Lee, S., K. Song, and Y.-J. Whang (2013), "Testing functional inequalities." *Journal of Econometrics*, 172 (1), 14–32. [19]

Lee, S., K. Song, and Y.-J. Whang (forthcoming), "Testing a general class of functional inequalities." *Econometric Theory.* [19]

Manski, C. F. (2003), *Partial Identification of Probability Distributions*. Springer-Verlag, New York. [11]

Manski, C. F. and J. V. Pepper (2000), "Monotone instrumental variables: With an application to the returns to schooling." *Econometrica*, 68 (4), 997–1010. [7, 8, 9, 12, 13, 15, 20]

Manski, C. F. and J. V. Pepper (2009), "More on monotone instrumental variables." *Econometrics Journal*, 12 (s1), S200–S216. [7]

Manski, C. F. and J. V. Pepper (2013), "Deterrence and the death penalty: Partial identification analysis using repeated cross sections." *Journal of Quantitative Criminology*, 29 (1), 123–141. [7]

Manski, C. F. and J. V. Pepper (forthcoming), "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions." *Review of Economics and Statistics.* [11]

Romano, J. P. and A. M. Shaikh (2008), "Inference for identifiable parameters in partially identified econometric models." *Journal of Statistical Planning and Inference*, 138 (9), 2786–2807. [18]

Romano, J. P. and A. M. Shaikh (2010), "Inference for the identified set in partially identified econometric models." *Econometrica*, 78 (1), 169–211. [18]

Rosen, A. M. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities." *Journal of Econometrics*, 146, 107–117. [18]

Santos, A. (2012), "Inference in nonparametric instrumental variables with partial identification." *Econometrica*, 80 (1), 213–275. [20]

Stoye, J. (2009), "More on confidence regions for partially identified parameters." *Econometrica*, 77 (4), 1299–1315. [15, 17]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 13 February, 2015; final version accepted 26 July, 2017; available online 22 September, 2017.