cemmap

centre for microdata methods and practice

The identification power of smoothness assumptions in models with counterfactual outcomes

Wooyoung Kim
Koohyun Kwon
Soonwoo Kwon
Sokbae Lee

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP17/14

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

THE IDENTIFICATION POWER OF SMOOTHNESS ASSUMPTIONS IN MODELS WITH COUNTERFACTUAL OUTCOMES

WOOYOUNG KIM[1], KOOHYUN KWON[2], SOONWOO KWON[2], AND SOKBAE LEE[2 3]

Abstract. In this paper, we investigate what can be learned about average counterfactual outcomes when it is assumed that treatment response functions are smooth. The smoothness conditions in this paper amount to assuming that the differences in average counterfactual outcomes are bounded under different treatments. We obtain a set of new partial identification results for the average treatment response by imposing smoothness conditions alone, by combining them with monotonicity assumptions, and by adding instrumental variables assumptions to treatment responses. We give a numerical illustration of our findings by reanalyzing the return to schooling example of Manski and Pepper (2000) and demonstrate how one can conduct sensitivity analysis by varying the degrees of smoothness assumption. In addition, we discuss how to carry out inference based on the existing literature using our identification results and illustrate its usefulness by applying one of our identification results to the Job Corps Study dataset. Our empirical results show that there is strong evidence of the gender and race gaps among the less educated population.

Keywords: Bounds, identification regions, instrumental variables, monotonicity, partial identification, sensitivity analysis, treatment responses, treatment selection.

JEL Classification: C14, C18, C21, C26.

[1]Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706-1393, USA.
[2]Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Republic of Korea.
[3]Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 7 Ridgmount Street, London, WC1E 7AE, UK.
E-mail addresses: wkim68@wisc.edu, doubleth@snu.ac.kr, kwon409@snu.ac.kr, sokbae@snu.ac.kr.
Date: 19 March 2014.
We would like to thank Andrew Chesher, Toru Kitagawa, Charles Manski, Adam Rosen and Kyungchul Song for helpful comments. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012-S1A3-A2033467) and by the European Research Council (ERC-2009-StG-240910-ROMETA).

1. Introduction

Partial identification has been increasingly popular in econometrics. For example, see monographs by Manski (2003, 2007), a recent review by Tamer (2010), and references therein. One important branch of this literature is concerned with bounding the distribution of the counterfactual outcomes or bounding the average treatment effects.[1] In this paper, we introduce a new set of "smoothness" assumptions for models with counterfactual outcomes. The smoothness conditions in this paper amount to assuming that the differences in average counterfactual outcomes are bounded under different treatments. The precise definition will be given later, but the basic idea is that the change in the average treatment effect cannot be too large if the change in the treatment is not large; hence it is called *smoothness* conditions.

To describe our setup, let $\Gamma \subset \mathbb{R}$ denote an ordered set that can be finite, countably infinite, or uncountable, and let $Y_i(t) : \Gamma \mapsto \mathbb{R}$ denote an individual-level, real-valued outcome function for treatment $t \in \Gamma$. Assume that we observe independent and identically distributed observations $\{(Y_i, Z_i) : i = 1, \ldots, n\}$, where $Z_i \in \Gamma$ is the actual treatment for individual i, and $Y_i \equiv Y_i(Z_i)$ is this individual's observed outcome. Let μ denote the probability distribution of Z_i, which may be discrete, continuous, or mixed.[2]

In this paper, we focus on the identification region of $g^*(t) \equiv E[Y_i(t)]$, namely the expected value of the counterfactual outcome $Y_i(t)$ for each $t \in \Gamma$. Define $g(t,s) \equiv E[Y_i(t)|Z_i = s]$ to be the expectation of $Y_i(t)$ conditional on the event that the realized treatment is s. With the empirical evidence alone, we can only identify $g(s,s)$. Let t_0 be the value of the treatment of interest. Suppose that $Y_i(t_0) \in [y_{\min}, y_{\max}]$, where $-\infty \le y_{\min} \le y_{\max} \le \infty$. Then, partial identification analysis for $g^*(t)$ starts from the well-known Manski's worst-case bound for the parameter (see, for example, Proposition 1.1 of Manski (2003)):

$$\begin{aligned} &E[Y_i|Z_i = t_0]P(Z_i = t_0) + y_{\min}P(Z_i \neq t_0) \\ &\le g^*(t_0) \\ &\le E[Y_i|Z_i = t_0]P(Z_i = t_0) + y_{\max}P(Z_i \neq t_0). \end{aligned}$$

[1] See, for example, Balke and Pearl (1997), Bhattacharya, Shaikh, and Vytlacil (2008, 2012), Blundell, Gosling, Ichimura, and Meghir (2007), Chesher (2005, 2010), Chiburis (2010), Fan and Park (2014), Fan, Sherman, and Shum (2014), Fan and Wu (2010), Heckman and Vytlacil (1999, 2005), Jun, Pinkse, and Xu (2011), Kitagawa (2009), Manski (1990, 1997, 2013), Manski and Pepper (2000, 2009, 2012), Shaikh and Vytlacil (2011) and Okumura and Usui (2013) among many others.

[2] Furthermore, we implicitly assume that all random variables, their functions, and all the events appearing in the paper are measurable.

This formulation of the identification region reveals that the identification power becomes weak when (i) the probability mass at $Z_i = t_0$ is small or (ii) $y_{max} - y_{min}$ is too large. Indeed, the identification region for $g^*(t_0)$ is $[y_{\min}, y_{\max}]$ if $P(Z_i = t_0) = 0$ and $(-\infty, \infty)$ if $y_{max} = \infty$ and $y_{min} = -\infty$.

The issue of small or zero probability mass occurs naturally when the treatment is evaluated on a continuous scale or on a discrete scale with many treatment options.[3] It is also easy to think of a situation where the difference between the upper and lower bounds of the outcome variable is too large. This motivates us to develop new identifying conditions under which one can obtain a meaningful identification region for $g^*(t)$ even in these circumstances.

The basic idea of this paper is to introduce smoothness conditions in $g(t, s)$ in both arguments t and s. They are called *smooth treatment response* and *smooth treatment selection* conditions. For the former, we assume that there exists a bound for the changes in the average treatment response with respect to the changes in the treatment; for the latter, we assume that the average counterfactual outcome cannot change too much if the treatment choice changes a little bit. We also combine the smoothness conditions with the other frequently imposed assumptions in the partial identification literature, namely the monotone treatment response (MTR), monotone instrument variable (MIV), and monotone treatment selection (MTS) assumptions, and derive the conditions under which the tighter bounds can be obtained. This is a natural step, since the philosophy of the partial identification is a "bottom-up" approach; starting from the no-assumption state, we add more and more assumptions if deemed justifiable. Here, we suggest that the smoothness assumptions can be utilized for the purpose of tightening the identification region.

Our smoothness conditions are inspired by Hausman and Newey (2013), who put the bounds on the partial derivative of the demand function with respect to the income in order to partially identify average consumer surplus. Moreover, Hall and Yatchew (2007, 2010) used information on the derivatives to recover the level of the function itself. They showed that faster convergence rate for nonparametric regression can be achieved in presence of the data on the derivatives. Our paper differs from the latter papers in that we only need information on the *bound* on the derivatives, leading to partial identification of the counterfactual outcomes. However, we share with Hausman and Newey (2013) and Hall

[3]This problem may arise under the extrapolation problem as well. For example, consider the case that the treatment t is years of schooling and we have individual level data, including years of schooling measured in integer values. We are interested in extending the current level of compulsory schooling, say some integer value $\underline{t}$, by only 6 months.Then in this case, the empirical probability mass at $Z_i = \underline{t} + 0.5$ is zero, implying that we have no prediction power for the new policy with the empirical evidence alone. See Figures 1.1 and 1.2 in Manski (2007, page 5) for a nice illustration.

and Yatchew (2007, 2010) the feature that information on derivatives or differences can help bound (or estimate) the mean outcomes of interest.

There is now a large body of literature on empirical applications of the various partial identification techniques. In particular, there exist many empirical papers using partial identification under monotonicity (MTR, MTS, and MIV).[4] However, most of them focus on binary or at most five treatments. An alternative approach to the continuous treatment case is to assume point identification using generalized propensity score matching (see, for example, Hirano and Imbens (2004) and Imai and Van Dyk (2004)).[5] We complement the existing literature by increasing the scope of applications of partial identification in the context of multiple/continuous treatments.

The remainder of the paper is organized as follows. In Section 2, we introduce new assumptions on treatment responses and obtain identification results. In addition, we provide discussions on bounding average treatment effects and other parameters. In Section 3, we show how to tighten the identification results obtained in Section 2 when an IV or an MIV exists. We find that the MTR-MTS bound of Manski and Pepper (2000) can be further tightened if we impose the smoothness conditions on the treatment response. In Section 4, we revisit the returns to schooling example of Manski and Pepper (2000) and demonstrate the usefulness of our smoothness conditions. In particular, we demonstrate how one can conduct sensitivity analysis by varying the degrees of smoothness assumption. In Section 5, we provide discussions on inference using the identification results. In Section 6, we present an actual application in the context of continuous treatment using the Job Corps Study dataset and find that the signs of gender and race gaps can be determined for some demographic groups under weak assumptions. Section 7 gives concluding remarks. In Appendix A, we consider restrictions on treatment selection and establish identification results. The basic idea in Appendix A is that the average counterfactual outcome cannot change too much if self-selected treatment choices are not too different. In Appendix B, we collect the proofs of all the identification results in the paper. Appendix C gives some additional theoretical results with respect to the average treatment effects and Appendix D provides some details about the empirical example.

As mentioned before, μ can be a general probability measure. Throughout the paper, we write the expectation of a function of Z as $E[\varphi(Z)] = \int \varphi(z)\mu(dz)$, where $\varphi(\cdot)$ is a given

[4]See De Haan (2011), Gerfin and Schellhorn (2006), Gonzalez (2006), Gundersen and Kreider (2008, 2009), Gundersen, Kreider, and Pepper (2012), Kreider and Hill (2009), Kreider and Pepper (2007), Kreider, Pepper, Gundersen, and Jolliffe (2012), Lee and Wilke (2009) and Pepper (2000) among others.

[5]For recent empirical applications of this method, see Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), Kluve, Schneider, Uhlendorff, and Zhao (2012), and others.

function. For example, if the distribution of Z is continuous, $E[\varphi(Z)] = \int \varphi(z)\mu(dz) = \int \varphi(z)p_\mu(z)dz$, where $p_\mu(\cdot)$ is the probability density function of Z. Alternatively, if the distribution of Z is discrete, $E[\varphi(Z)] = \int \varphi(z)\mu(dz) = \sum_j \varphi(z_j)p_\mu(z_j)$, where $p_\mu(\cdot)$ is now the probability mass function of Z. Other cases can be understood similarly. Finally, we let Roman letters such as $t, t', s, s' \in \Gamma$ denote generic arguments of $g(\cdot,\cdot)$ with different uses in different places.

2. Smooth Treatment Response

In this section, we introduce two assumptions on treatment responses: the one we call *smooth treatment response* (STR) and the other *smooth monotone treatment response* (SMTR). Both conditions are stated below in terms of the "local" behavior of $g(t, s)$ with respect to t.

Assumption 2.1 (Treatment Response Assumptions). *Assume one of the following conditions:*

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

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

Assumption 2.1, which is inspired by Manski (1997) and Hausman and Newey (2013), does not seem to be explored in the literature on models with counterfactual outcomes. Manski (1997) introduced the notion of monotone treatment response (MTR). That is,

$$t \geq t' \Rightarrow Y_i(t) \geq Y_i(t') \tag{2.1}$$

for each individual i. Our monotonicity assumption in SMTR is in the same spirit of Manski (1997), but slightly weaker than (2.1) since we focus on the identification region of the expected value $E[Y_i(t)]$.

What is different from Manski (1997) in this paper is that we have a bound on changes in $g(t, s)$ with respect to t. Hausman and Newey (2013, Theorem 7) used the bounds on the income effect to partially identify average consumer surplus. Their assumption has the following form for the income effect: (using their notation) there are constants b and B such that

$$b \leq \partial q(p, y, \eta)/\partial y \leq B, \tag{2.2}$$

where q is a demand function of price p, individual income y, and unobserved individual heterogeneity η. We follow Hausman and Newey (2013) to make Assumption 2.1, while allowing for the case that the treatment is not continuous.

The "smoothness" condition in both STR and SMTR conditions can be rewritten as

$$-b \le \frac{g(t,s) - g(t',s)}{t - t'} \le b \tag{2.3}$$

for all $t \neq t'$ and for all s.[6] Regarding $g(\cdot, s)$ as a function of only the first argument for each s, the quotient in (2.3) is called in general the difference quotient of $g(\cdot, s)$. Hence, part (i) of Assumption 2.1 amounts to assuming that $g(\cdot, s)$, as a function of the first argument, has bounded difference quotients uniformly in s. If $\Gamma = \mathbb{R}$, or an interval in the real line, this is equivalent to assuming that $g(\cdot, s)$ is Lipschitz continuous with respect to the first argument uniformly in s.

Note that the inequalities in (2.3) can be satisfied if

$$-b \le \frac{Y_i(t) - Y_i(t')}{t - t'} \le b \tag{2.4}$$

for all $t \neq t'$ and for each i. That is, assuming (2.4) amounts to bounding the individual-level treatment effect defined as $[Y_i(t) - Y_i(t')]/(t - t')$. Manski and Pepper (2009) considered the homogeneous-linear-response (HLR) assumption such that

$$Y_i(t) = \beta \times t + \delta_i,$$

where β is a slope parameter and δ_i is an unobserved random variable for each individual i. The STR condition is satisfied by the HLR assumption, as long as $\beta \le b$.

An alternative way of bounding the rate of change in the average counterfactual response is to impose further global restrictions in addition to monotonicity. Manski (1997, Section 4) added concavity to the basic assumption of monotonicity and showed formally that concavity has substantial identifying power. See also Okumura and Usui (2013) who combined concavity with the MTS assumption. Our approach imposes restrictions directly on the rate of change in its nature, whereas the combination of concavity and monotonicity, as in Manski (1997) and Okumura and Usui (2013), restricts the rate of change indirectly. Therefore, we view that two approaches are distinct as well as complementary.

[6]More generally, one may consider (2.3) with two different end points b_1 and b_2, similar to (2.2), as in Hausman and Newey (2013). Our STR and SMTR conditions are special cases of $(b_1, b_2) = (-b, b)$ and $(b_1, b_2) = (0, b)$, respectively. In this paper, we did not opt to introduce this more general setup since it introduces an additional tuning parameter.

In some applications, the derivative of a counterfactual outcome function is naturally bounded. For example, consider a production function for which the input is some raw material and the output is a processed product. When measured by the weight, the derivative cannot exceed 1. Another case is an inelastic downward sloping demand function where the treatment is price. In both cases, the STR and SMTR assumptions can be applied with $b = 1$. See also Hausman and Newey (2013) for how to set bounds (b and B in (2.2)) on the income effect for their empirical application on gasoline demand. There will be many other cases where we can set a plausible bound on the smoothness of the counterfactual outcome.

More generally speaking, we may interpret our identification analysis as a conditional one indexed by b. Furthermore, we may conduct sensitivity analysis by looking at different values of b. In Section 4.3, we provide an example of sensitivity analysis (for general discussions on sensitivity analysis, see Leamer (1985), Tamer (2010), and others).

We regard STR as a useful assumption that might complement the MTR assumption in a variety of applications. For example, consider the problem of bounding the return to schooling, as in Manski and Pepper (2000). In this example, $Y_i(t)$ is the counterfactual log hourly wage and the treatment t is years of schooling. Thus, the "smoothness" assumption here corresponds to setting the maximum value of the return to schooling.[7] In Section 4, we will come back to the example of Manski and Pepper (2000) to show some numerical illustration.

Before we give our first identification result, define $x^+ \equiv \max(x, 0)$ and $x^- \equiv \max(-x, 0)$ for any real number x. The following proposition provides sharp bounds for $g^*(t)$ under STR and SMTR, respectively.

Proposition 2.1. *Assume that the support of $Y_i(t)$ is unbounded. Then the following bounds are sharp:*

(i) *Under STR,* $E[Y_i] - bE[|Z_i - t|] \leq g^*(t) \leq E[Y_i] + bE[|Z_i - t|]$.
(ii) *Under SMTR,* $E[Y_i] - bE[(Z_i - t)^+] \leq g^*(t) \leq E[Y_i] + bE[(Z_i - t)^-]$.

Proposition 2.1 (i) states that under STR, the sharp bound is symmetric around $E[Y_i]$ and its width is $2bE[|Z_i - t|]$. Proposition 2.1 (ii) implies that under SMTR, the sharp bound is possibly asymmetric around $E[Y_i]$, and its width is now $bE[|Z_i - t|]$ since $|x| = x^+ + x^-$ for any real number x. Thus, adding the weak monotonicity to the STR condition shortens the width by half. In both cases, the strength of the identification power of the STR condition is determined by two factors: (i) the size of b and (ii) the distribution of the realized

[7]For instance, setting 0.2 as the bound on the return from one more year of schooling seems to be conservative enough.

treatment random variable Z_i. Also note that for either case, the width is minimized when the counterfactual treatment value is the median of Z_i.

We now focus on comparison between the SMTR condition and the original MTR assumption. First, if only the MTR condition in the equation (2.1) is assumed with unbounded $Y_i(t)$, then the identification region of $g^*(t)$ is unbounded (see Corollary M1.2 of Manski (1997)). Therefore, we have demonstrated that when the support of $Y_i(t)$ is unbounded but average changes in $Y_i(t)$ are bounded, we can obtain some informative identification results.

When the support of $Y_i(t)$ is bounded, the identification analysis is more complicated. For example, suppose that $Y_i(t) \leq y_{\max} < \infty$ for some known $y_{\max}$. Then it is straightforward to show that the SMTR upper bound for $g^*(t)$ is

$$g^*(t) \leq \int_{z<t} \min\left\{y_{\max}, (E[Y_i|Z_i = z] + b(t-z))\right\} \mu(dz) + E[Y_i|Z_i \geq t]P(Z_i \geq t). \tag{2.5}$$

Here, recall that μ denotes the probability distribution of Z_i, which may be discrete, continuous, or mixed. The upper bound (2.5) cannot be larger than the upper bound under the MTR assumption alone since the latter has the form (see again Corollary M1.2 of Manski (1997)):

$$g^*(t) \leq y_{\max}P(Z_i < t) + E[Y_i|Z_i \geq t]P(Z_i \geq t). \tag{2.6}$$

Note that the SMTR upper bound strictly improves the MTR upper bound if and only if the event such that $E[Y_i|Z_i] + b(t - Z_i) < y_{\max}$ 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 2.2. *Assume that the support of* $Y_i(t)$ *is* $[y_{\min}, y_{\max}]$, *where* $-\infty \leq y_{\min} \leq y_{\max} \leq \infty$. *Then we have:*

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

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

Note that given values of $(t, b, y_{\min}, y_{\max})$, one can test whether there is an strict improvement, since the sufficient and necessary conditions in Corollary 2.2 are expressed as population quantities that can be estimated consistently. To be more specific, suppose that we are interested in testing the null hypothesis of no strict improvement of the upper bound. That is, we consider testing

$$H_0 : \int_{z<t} 1\{U_{\text{SMTR}}(t,z) < 0\}\,\mu(dz) = 0 \text{ vs. } H_1 : \int_{z<t} 1\{U_{\text{SMTR}}(t,z) < 0\}\,\mu(dz) > 0. \tag{2.7}$$

Analogously, we can test for the improvement of the MTR-lower bound. It seems that there is no readily available method for testing (2.7) nonparametrically. This is an interesting topic for future research.

Remark 2.1. Taking minimum of the STR and MTR upper bounds generally does not lead to the SMTR upper bound. To see this, define

$$\begin{aligned}
I_1(t) &\equiv \int_{z<t} \min\{y_{\max}, E[Y_i|Z_i = z] + b|z-t|\}\mu(dz),\\
I_2(t) &\equiv \int_{z<t} y_{\max}\mu(dz),\\
I_3(t) &\equiv \int_{z\geq t} E[Y_i|Z_i = z]\mu(dz),\\
I_4(t) &\equiv \int_{z\geq t} E[Y_i|Z_i = z] + b|z-t|\mu(dz).
\end{aligned}$$

Then, the upper bounds for STR, MTR and SMTR with the knowledge of the support of $Y_i(t)$ are $I_1(t)+I_4(t)$, $I_2(t)+I_3(t)$, and $I_1(t)+I_3(t)$, respectively. Note that $I_1(t) \leq I_2(t)$ and $I_3(t) \leq I_4(t)$. Therefore, $\min\{I_1(t)+I_4(t), I_2(t)+I_3(t)\} > I_1(t)+I_3(t)$ if and only if $I_2(t) > I_1(t)$ and $I_4(t) > I_3(t)$, which easily holds if $y_{\max}$ is sufficiently large, and t is not the endpoint of the support of Z_i (for example, if μ is counting measure and t is the upper endpoint, $I_3(t) = I_4(t)$, which leads to the same STR upper bound as SMTR one). A similar argument can be applied to the case of the lower bound.

Remark 2.2. We may confine the STR and SMTR conditions to be only locally valid. This restriction is reasonable if we suspect that the underlying counterfactual response function exhibits non-smooth behavior in some region of the support. Making global assumptions may also result in an excessively large value of b, which may not lead to informative identification results. Let Γ_0 denote the subset of Γ where the STR and SMTR conditions locally hold. Then the identification results presented above (and those to be presented below) can be translated as those for $E[Y_i(t)|Z_i \in \Gamma_0]$ for $t \in \Gamma_0$.

2.1. **Bounds on Average Treatment Effects and Other Parameters.** All identification results obtained so far are concerned with the average outcomes of a particular treatment. In some cases, we are interested in average treatment effects, defined as $\Delta(t,t') \equiv$

$g^*(t) - g^*(t')$ for $t \neq t'$. Recall that the STR condition $-b \leq [g(t,s) - g(t',s)]/(t-t') \leq b$ implies $|\Delta(t,t')| \leq b(t-t')$ for $t \neq t'$. This shows that we are essentially bounding the size of the average treatment effect to make identification of $g^*(t)$ possible. This strategy results in weak identification power of the STR condition for the average treatment effect itself. For example, if we write the upper bound of $g^*(t) - g^*(t')$ as

$$(E[Y_i] + bE[|Z_i - t|]) - (E[Y_i] - bE[|Z_i - t'|]) = b\left(E[|Z_i - t|] + E[|Z_i - t'|]\right)$$

taking the upper and lower bounds of each counterfactual mean outcome, this bound is not sharp under the STR condition. Similar arguments hold for the SMTR condition.[8] However, the STR or SMTR condition can be useful to bound the average treatment effect when it is combined with other assumptions. For example, see Section 4.2 for this case.

Although the bound with the STR or SMTR condition alone is not attractive in terms of identifying average treatment effects, our approach is useful to bound other parameters. To give such an example, suppose that W_i is the gender. Then $E[Y_i(t)|W_i = \text{male}] - E[Y_i(t)|W_i = \text{female}]$ is the gender gap in the average counterfactual outcome. The upper bound of $E[Y_i(t)|W_i = \text{male}] - E[Y_i(t)|W_i = \text{female}]$ is the difference between the upper bound of $E[Y_i(t)|W_i = \text{male}]$ and the lower bound of $E[Y_i(t)|W_i = \text{female}]$. This bound is sharp if there is no cross restriction between males and females. The sharp lower bound is defined analogously.[9] We provide an empirical example of this type of bounds in Section 6 using the Job Corps Study dataset.

3. Adding Instrumental Variables Assumptions to Treatment Responses

In this section, we show how to tighten the identification results obtained in Section 2 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.[10]

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 $Y_i(t)$

[8]See Proposition C.1 and following discussions in Appendix C for details.

[9]Other examples of parameters of interest, which can be bounded sharply by the STR or SMTR condition, include trends of the average counterfactual outcome over time. See, e.g. Blundell, Gosling, Ichimura, and Meghir (2007) and Lee and Wilke (2009) for related results.

[10]Laffërs (2013) emphasized the importance of distinguishing conditional and unconditional versions of monotone treatment selection.

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 3.1 (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 (2.4) holds for each individual and CSMTR is satisfied when (2.1) and (2.4) hold for each individual. Hence, the conditional versions of STR and SMTR conditions can be motivated, as in Section 2. The IV and MIV conditions are well known in the literature. See, for example, Manski and Pepper (2000, 2009) among others.

Remark 3.1. A related identification assumption in the literature is “bounded instrumental variable” introduced in Manski and Pepper (2012, 2013). To express the assumption in our notation, 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, where $ATE_v \equiv E[Y(t_1)|V_i = v] - E[Y(t_0)|V_i = v]$. This condition is related with 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 3.1.

Proposition 3.1. *Assume that the support of* $Y_i(t)$ *is unbounded. Then the following bounds are sharp:*

(i) *Under CSTR and IV together,*

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

(ii) *Under CSMTR and IV together,*

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

(iii) *Under CSTR and MIV together,*

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

(iv) *Under CSMTR and MIV together,*

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

These results can be regarded as combinations of Proposition 2.1 and Proposition 1 of Manski and Pepper (2000). It follows immediately from Proposition 3.1 (iii) and (iv) that the identification region of $g^*(t)$ under the MIV assumption is given as below.

Corollary 3.2. *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,*

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

(ii) *Under CSMTR and MIV together,*

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

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 combination with MIV. Thus, in such cases, the MIV assumption has the same identifying power as the IV assumption.

3.1. **The SMTR and MTS Bounds.** In this subsection, we consider adding the smoothness assumption to the MTR-MTS bound of Manski and Pepper (2000). This bound is particularly useful because combining the MTR and MTS assumptions yields an informative bound even if Y_i is unbounded, as shown by Manski and Pepper (2000). Therefore, it is important to understand the role of smoothness assumption for the MTR-MTS bound.

Manski and Pepper (2000) introduced the following concept of monotone treatment selection (MTS):

$$s \geq s' \Rightarrow E[Y_i(t)|Z_i = s] \geq E[Y_i(t)|Z_i = s']. \tag{3.1}$$

Manski and Pepper (2000) pointed out that the MTS assumption is a special case of the MIV assumption when the instrumental variable V_i is the realized treatment Z_i. We examine the role of smoothness assumption for the MTR-MTS bound by replacing the MTR assumption with the SMTR condition.

Proposition 3.3. *Assume that the support of $Y_i(t)$ is unbounded. Then, under SMTR and MTS together, $E[Y_i(t)] \in [l_1(t), u_1(t)]$, where*

$$l_1(t) \equiv \int_{z<t} E[Y_i|Z_i = z]\mu(dz) + \int_{z \geq t} \sup_{s' \in [t,z]} (E[Y_i|Z_i = s'] + b(t - s'))\mu(dz),$$

$$u_1(t) \equiv \int_{z \leq t} \inf_{s' \in [z,t]} (E[Y_i|Z_i = s'] + b(t - s'))\mu(dz) + \int_{z>t} E[Y_i|Z_i = z]\mu(dz).$$

Moreover, this bound is sharp.

Recall that the MTR-MTS bound of Manski and Pepper (2000) has the form

$$l_{MP}(t) \leq E[Y_i(t)] \leq u_{MP}(t),$$

where

$$l_{MP}(t) \equiv E[Y_i|Z_i < t]P(Z_i < t) + E[Y_i|Z_i = t]P(Z_i \geq t),$$

$$u_{MP}(t) \equiv E[Y_i|Z_i > t]P(Z_i > t) + E[Y_i|Z_i = t]P(Z_i \leq t).$$

To characterize the case when the smoothness assumption improves the MTR-MTS bound, define

$$A_l(t) \equiv \{z \in \Gamma : \sup_{s' \in [t,z]} (E[Y_i|Z_i = s'] + b(t - s')) > E[Y_i|Z_i = t]\},$$
$$A_u(t) \equiv \{z \in \Gamma : \inf_{s' \in [z,t]} (E[Y_i|Z_i = s'] + b(t - s')) < E[Y_i|Z_i = t]\}.$$

Clearly, $A_l(t)$ excludes points such that $z \leq t$ and $A_u(t)$ excludes those such that $z \geq t$. The following proposition gives necessary and sufficient conditions for the strict improvement of the SMTR-MTS bound to the MTR-MTS bound.

Corollary 3.4. *The the SMTR-MTS bound is strictly tighter than the MTR-MTS bound if and only if $P(Z_i \in A_l(t)) > 0$ or $P(Z_i \in A_u(t)) > 0$. More specifically, we have*

(i) $l_1(t) > l_{MP}(t)$ *if and only if* $P(Z_i \in A_l(t)) > 0$.
(ii) $u_1(t) < u_{MP}(t)$ *if and only if* $P(Z_i \in A_u(t)) > 0$.

This corollary essentially means that SMTR-MTS bound can be made tighter than the MTR-MTS bound if b is reasonably small, i.e. treatment response is sufficiently smooth. In Section 4, using the empirical example of Manski and Pepper (2000), we will demonstrate the strict improvement of the MTR-MTS bound with a reasonable value of b. Furthermore, Section 4.3 presents sensitivity analysis that shows the SMTR-MTS bound improves the MTR-MTS bound for the average treatment effect for a range of different values of b.

4. Numerical Illustration: Manski and Pepper (2000) Revisited

In this section, we revisit the returns to schooling example of Manski and Pepper (2000) and illustrate the usefulness of our framework. In particular, we show that the SMTR-MTS bound becomes narrower than the MTR-MTS bound, which achieves the tightest bound in Manski and Pepper (2000), for a range of reasonable values of b. In this section, we will treat numerical results as if we knew population quantities, to focus on identification results and also to be comparable to Manski and Pepper (2000). In the next section, we will move to the issue of inference.

4.1. **Bounds on Average Counterfactual Outcomes.** In the example of Manski and Pepper (2000), t is years of schooling and $g^*(t)$ is the expectation of counterfactual log wages when the treatment is t years of schooling. To estimate the bounds developed in this paper and those in Manski and Pepper (2000), we only need to estimate $E[Y_i|Z_i = t]$, $P(Z_i = t)$, and the end points $[y_{\min}, y_{\max}]$ of the support of Y_i. Table I of Manski and Pepper (2000) gives information on the estimates of $E[Y_i|Z_i = t]$, $P(Z_i = t)$, which were obtained from the NLSY. According to the NBER working paper version of Manski and Pepper (1998), Manski and Pepper used $[y_{\min}, y_{\max}] = [1.4, 5.0]$.[11] We use the same values in our analysis of both the MTR bound and the MTS bound.

FIGURE 1. SMTR-STR-MTR comparison

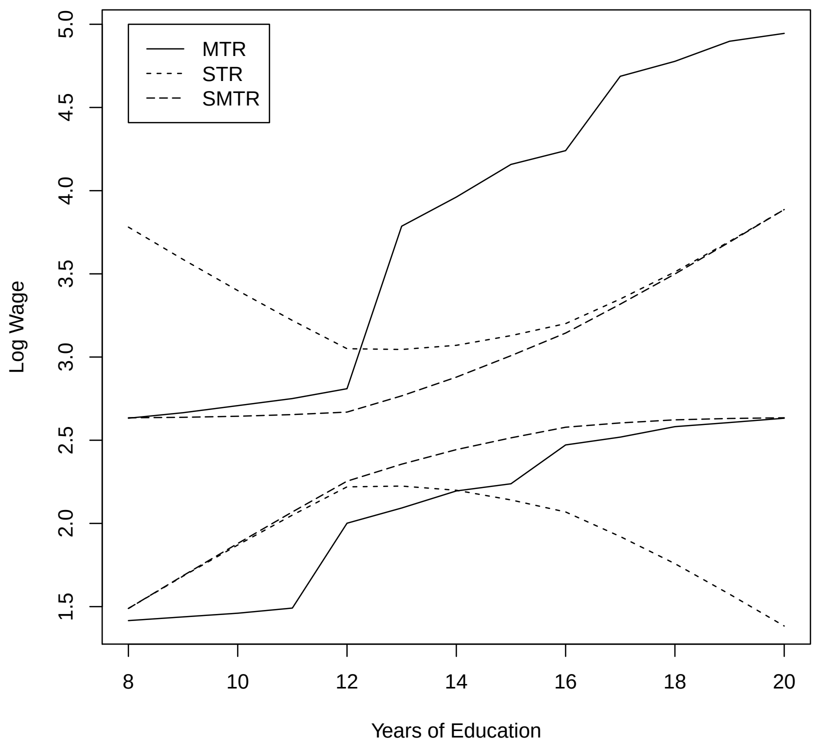


Figure 1 shows the SMTR, STR, and MTR bounds when the value of b is 0.2.[12] Roughly speaking, this corresponds to the maximum of 20 percentage points in the average return to one year of schooling. For US samples, OLS and IV estimates of the returns to education are typically less than 0.1 (see, for example, Table II of Card, 2001). Using local instrumental variables estimators with NLSY data, Carneiro, Heckman, and Vytlacil (2011) reported a baseline estimate of 0.0815 for the average treatment effect of one year of college. Their estimate varies between 0.0626 and 0.1409, across different samples and specifications (see

[11]See Manski and Pepper (1998) for details on their choice of $[y_{\min}, y_{\max}]$.

[12]For the SMTR and STR bounds, we calculated unbounded version of them, following Proposition 2.1.

Table 6 of Carneiro, Heckman, and Vytlacil, 2011). In view of these estimates, we regard our choice of b as a plausible upper bound.

The STR bound alone or the MTR bound alone gives a relatively wide bound; however, the SMTR bound seems much tighter, especially in the middle of the distribution of Z_i. Note that the SMTR bound is narrower than the envelope of the STR and MTR bounds, as discussed in Remark 2.1. Figure 1 demonstrates that there could be a substantial shrinkage of the identification region if one combines the smoothness condition with the monotonicity assumption.

Table 1. Comparing SMTR-MTS bound and MTR-MTS bound

	SMTR-MTS		MTR-MTS	
t	lower	upper	lower	upper
8	2.247	2.635	2.247	2.635
9	2.299	2.636	2.299	2.636
10	2.195	2.632	2.195	2.632
11	2.339	2.640	2.339	2.640
12	2.477	2.651	2.477	2.651
13	2.560	2.730	2.560	2.730
14	2.552	2.719	2.552	2.719
15	2.571	2.754	2.571	2.754
16	2.627	2.875	2.627	2.875
17	**2.618**	2.792	2.615	2.792
18	2.636	**2.973**	2.636	3.002
19	2.636	3.004	2.636	3.004
20	2.635	2.933	2.635	2.933

Table 1 shows the SMTR-MTS bound given in Proposition 3.3 when $b = 0.2$. It also shows Manski and Pepper (2000)'s MTR-MTS bound in the last two columns. The improved bounds are marked as boldface. The lower bound for $t = 17$ and the upper bound for $t = 18$ are improved with the addition of the smoothness assumption with $b = 0.2$.

4.2. **Bounds on Average Treatment Effects.** We now consider the average treatment effect:

$$\triangle(s,t) \equiv E[Y_i(t)] - E[Y_i(s)]. \tag{4.1}$$

As discussed in Section 2.1, one may obtain the upper bound of $\triangle(s,t)$ by taking the difference between the upper bound of $E[Y_i(t)]$ and the lower bound of $E[Y_i(s)]$. On one hand, this type of the upper bound based on our SMTR-MTS assumption is not sharp[13];

[13]See Proposition C.2 in Appendix C for a formal proof.

on the other hand, the same type of the upper bound using the MTR-MTS assumption is sharp, as discussed in Manski and Pepper (2000). However, it is possible that our upper bound of $\triangle(s,t)$ can be strictly smaller than that of Manski and Pepper (2000). For example, in Table 1, consider $\triangle(16,18)$, which corresponds roughly to the return to two years of postgraduate education. Note that the MTR-MTS upper bound of $\triangle(16,18)$ is 0.375, whereas the SMTR-MTS upper bound is 0.346 with $b = 0.2$. As a result, the upper bound decreases by 2.9 percentage points.

Remark 4.1. Recall that the SMTR bound alone imposes the upper bound of $0.4 = 2b$ for $\triangle(16,18)$. It can be seen that the SMTR bound alone is improved upon by combining it with the MTS condition, resulting in a better bound than the MTR-MTS bound. Therefore, although the SMTR condition alone does not provide useful information on the average treatment effect, which was discussed in Section 2.1, the upper bound of $\triangle(s,t)$ based on the SMTR-MTS assumption can provide a tighter bound for the average treatment effect .

FIGURE 2. The upper bound for $\Delta(16,18)$

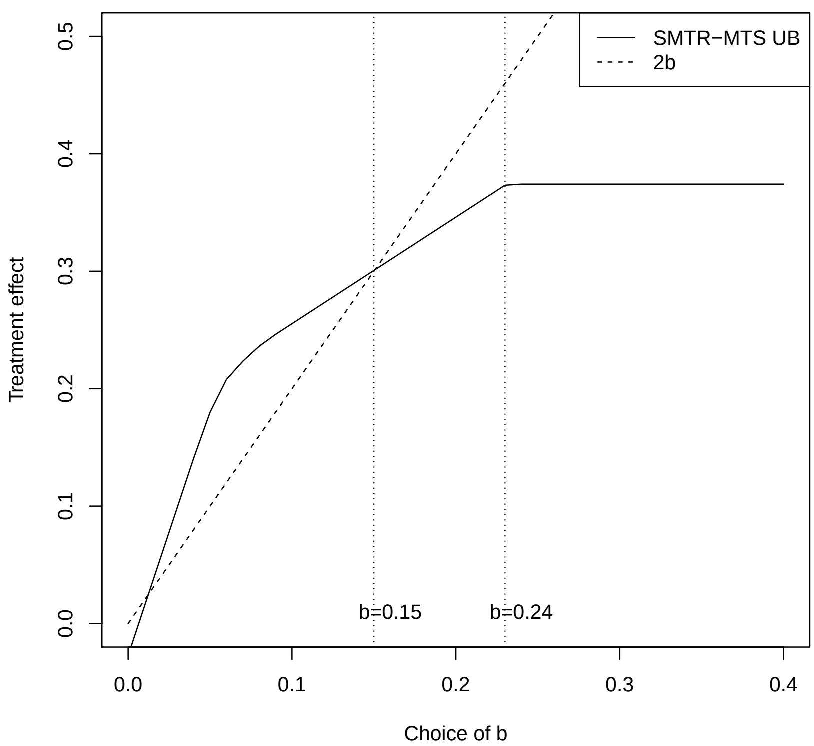
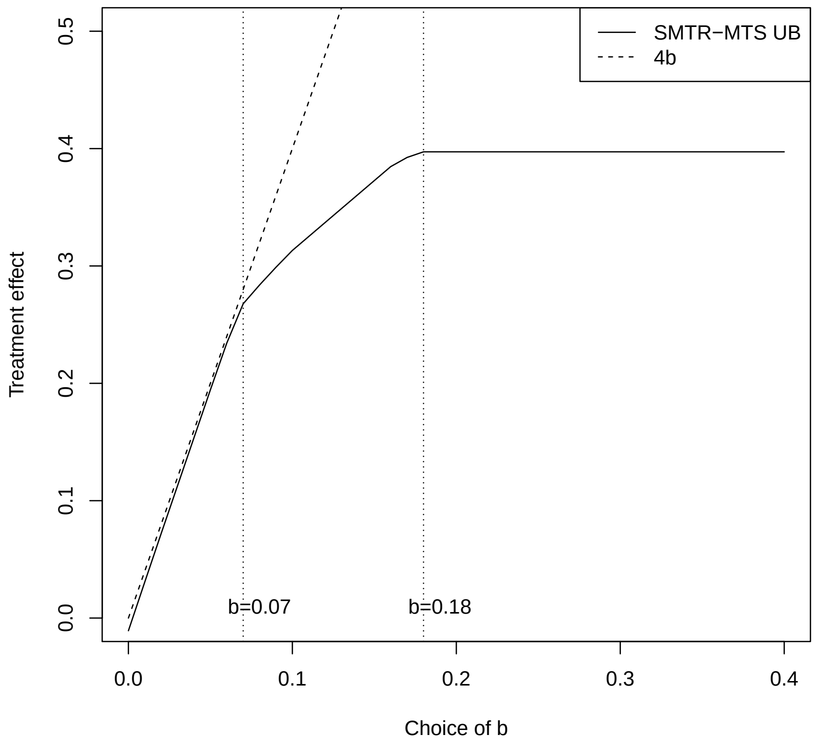


Figure 3. The upper bound for $\Delta(12, 16)$



4.3. **Sensitivity Analysis.** The results up to now are based on the choice of $b = 0.2$, which seems reasonable but nonetheless arbitrary. We now present the results of sensitivity analysis by considering all possible values of b. Figure 2 shows how different choices of b affect the identification region of the avearage treatment effect. The solid line is the upper bound for $\Delta(16, 18)$ obtained by subtracting the lower bound of $E[Y_i(16)]$ from the upper bound of $E[Y_i(18)]$ under the SMTR-MTS assumption. Note that choosing b from (0.15,0.24) gives both meaningful (smaller than $2b$ - denoted by the dashed line) and improved (smaller than the MTR-MTS upper bound, whose value is marked by the constant solid line when $b > 0.24$) upper bound for the average treatment effect. That is, for the region of (0.15,0.24), the smoothness assumption provides useful information to tighten the bound for $\Delta(16, 18)$. Hence, we may call such region the *effective region* of b for identification of $\Delta(16, 18)$. Obtaining the effective region of b amounts to conducting sensitivity analysis in this example. By looking at all possible values of b, we can see how the identification region of the average treatment effect changes. This approach gives a more complete picture of partial identification analysis than the approach with a fixed choice of b.

In general, the effective region of b will be dependent on the average treatment effect of interest. Figure 3 presents the analogous result when our parameter of interest is $\Delta(12, 16)$, and the effective region of b, which turns out to be (0.07,0.18).

5. 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.

5.1. **Inference Using Proposition 2.1.**

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

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

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

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

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

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

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

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:

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

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

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

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

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

One may develop alternative methods for inference, noting that the bounds in Proposition 2.1 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), Romano and Shaikh (2010) and Rosen (2008) among others.

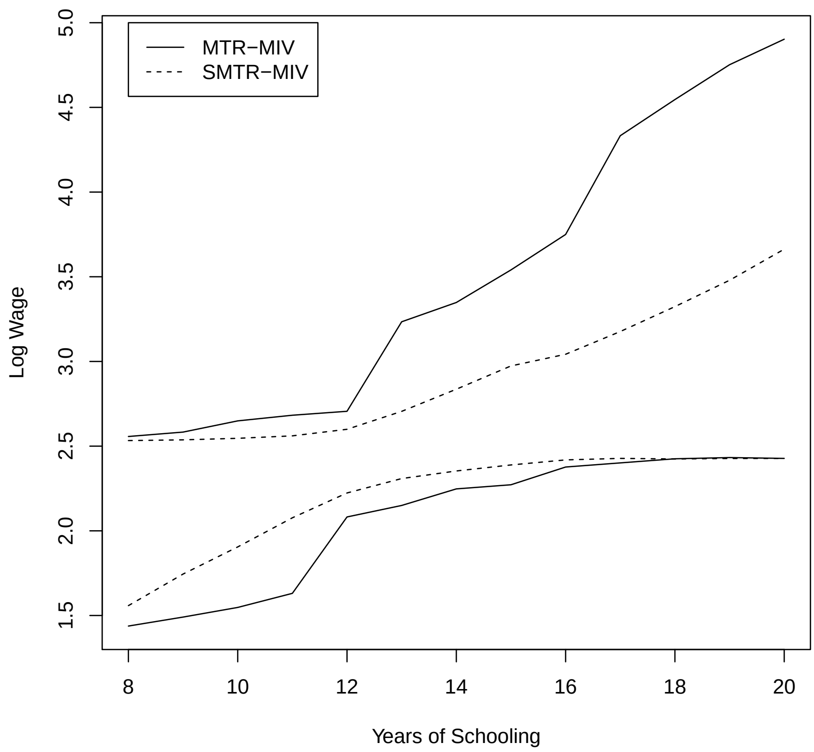
5.2. **Inference Using Proposition 3.1.** The identification region obtained in each case of Proposition 3.1 corresponds to the form of intersection bounds considered in Chernozhukov, Lee, and Rosen (2013). Therefore, a pointwise confidence interval for $g^*(t)$ can be obtained, following the inference method developed in Chernozhukov, Lee, and Rosen (2013) directly. Alternatively, one can use inference methods developed for conditional moment inequalities, such as Andrews and Shi (2013), Armstrong (2011a,b), Armstrong and Chan (2013), Chetverikov (2011) and Lee, Song, and Whang (2013a,b) among others. Among these methods, Lee, Song, and Whang (2013a) 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.[14] Here, $Y_i(t)$ is the counterfactual log wage given that the individual received t years of schooling, and V_i is

[14]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.

the Armed Forces Qualifying Test (AFQT) score which is used as MIV; i.e. those with higher AFQT scores will earn more wages on average. We used $b = 0.2$ for the smoothness parameter. Figure 4 shows the pointwise 95% confidence intervals for $E[Y_i(t)|V_i = 0]$ using both the MTR-MIV and SMTR-MIV bounds.[15] 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 (2013).[16] 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.

FIGURE 4. SMTR-MTR comparison when combined with MIV



5.3. **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 3.2 differ from the intersection bounds; they are rather averages of intersection bounds.

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

[16]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, Kim, Lee, and Rosen (2013).

There does not seem to 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 3.3 seem 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. It is an open question how to carry out inference in general non-smooth setups, including our bounds as special cases.

6. An Empirical Application

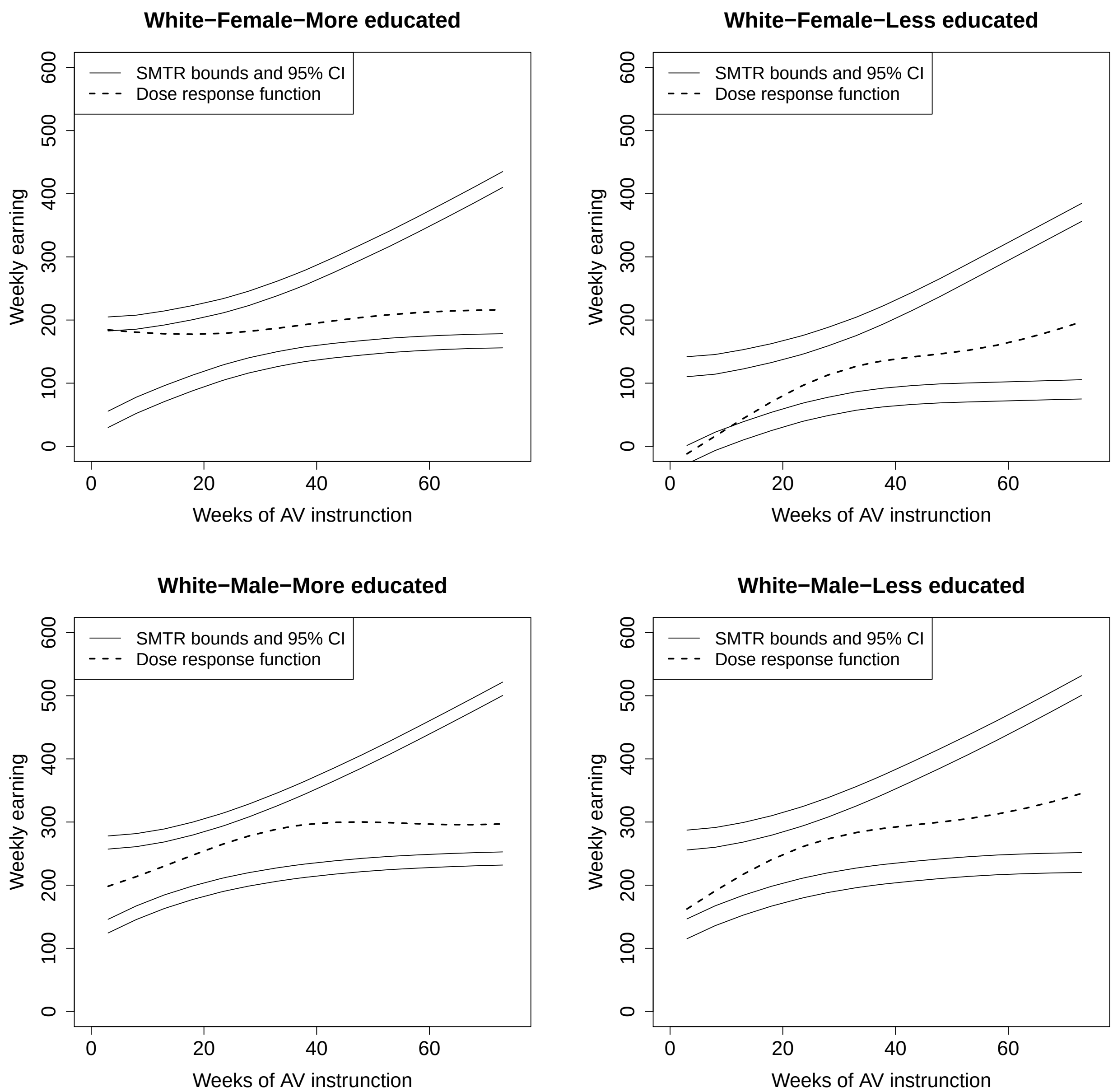
In this section, we use data from the National Job Corps Study (NJCS) and investigate what can be learned by our partial identification approach about the relationship between the length of job training and the labor market outcome. In particular, we focus on the SMTR bounds, present an empirical application of our inference method described in Section 5.1, and illustrate the usefulness of our approach by bounding counterfactual gender and race gaps.

To describe the setting, the treatment variable Z_i is the length of exposure to the academic and vocational instruction (henceforth "AV instruction") measured in weeks, and the outcome variable $Y_i(t)$ is the weekly earnings of an individual one year after the individual left the training program, given that the individual received t weeks of AV instruction during the program. For a detailed description of this dataset and previous research, refer to the literature that analysed the same dataset, such as Schochet, Burghardt, and Glazerman (2001), Schochet, Burghardt, and McConnell (2008), Flores-Lagunes, Gonzalez, and Neumann (2010), and most recently Flores, Flores-Lagunes, Gonzalez, and Neumann (2012).

Our smoothness assumption seems particularly relevant in this example. If only the MTR assumption is imposed, the resulting bound depends on the support of $Y_i(t)$, which seems difficult to know since the outcome variable is the weekly earnings of an individual. Furthermore, since the treatment changes continuously, it would be difficult to come up with an informative bound with only the MTR assumption.

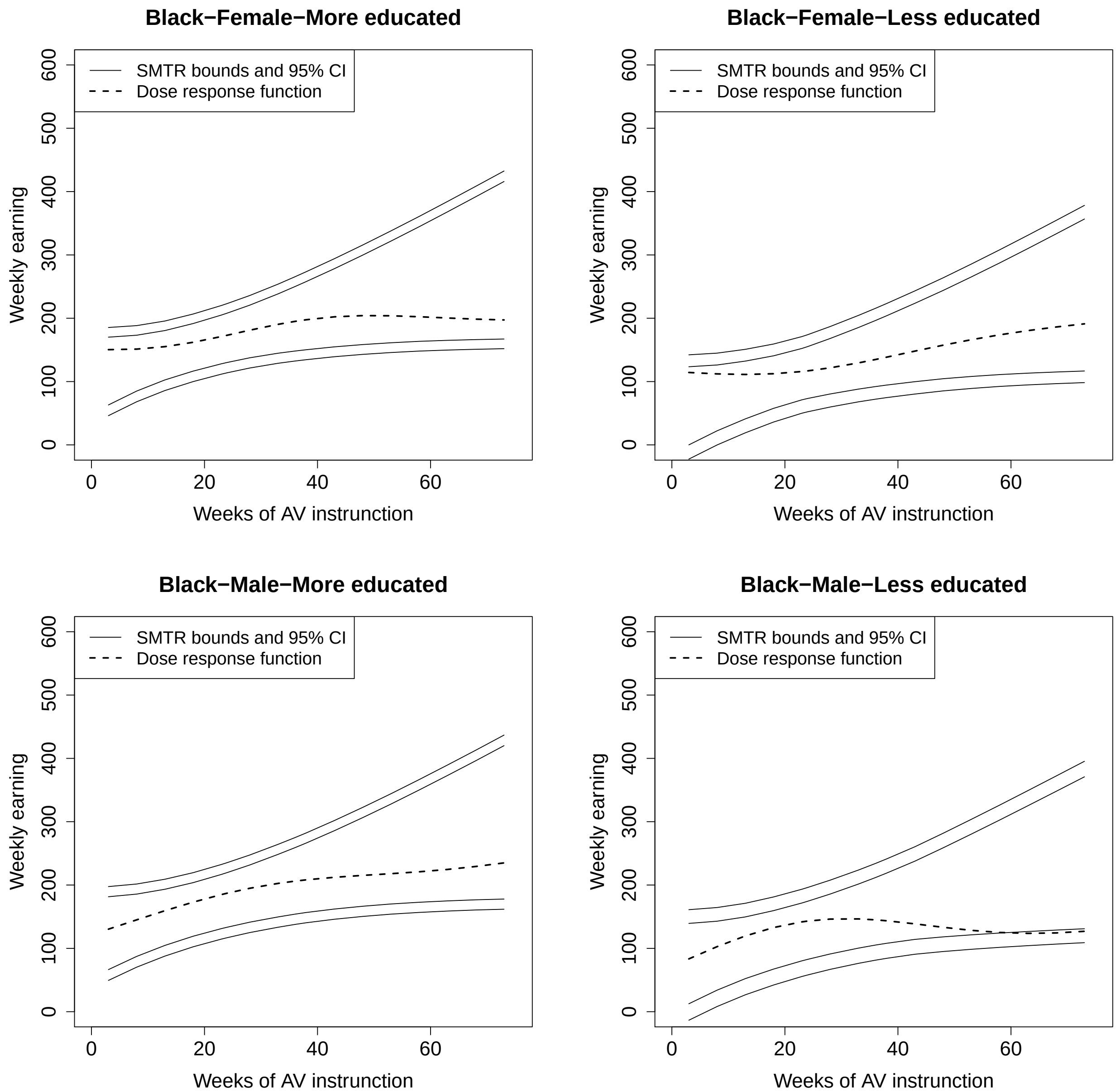
6.1. **Bounds on the Average Responses of Earnings to Job Training.** We are interested in obtaining the upper and lower bounds of $E[Y_i(t)]$ using smoothness assumptions along with monotonicity. Here we impose the SMTR assumption, meaning that taking more instructions at least does not hurt the labor outcome on average. The smoothness parameter we set here is $b = 5$; i.e. we assume that one more week of AV instruction cannot increase the average weekly earnings by more than 5 dollars. We chose $b = 5$ since it is larger than the largest estimate of average derivatives reported in Flores, Flores-Lagunes, Gonzalez, and Neumann (2012, the second panel of Table 3).

FIGURE 5. Potential earning for the white population



*“More educated” means more than or equal to 10 years of education. Otherwise individuals are classified as “Less educated.” Weekly earnings are measured in dollars.

Figure 6. Potential earning for the black population



*"More educated" means more than or equal to 10 years of education. Otherwise individuals are classified as "Less educated." Weekly earnings are measured in dollars.

To compare the estimates from our method with the existing ones, we computed the estimates of $E[Y_i(t)]$ using the generalized propensity score (GPS) under the uncomfoundedness assumption (the estimate is often called "average dose response function"). For theoretical background, see Hirano and Imbens (2004) and Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), the latter of which applies the GPS matching method to the Job Corps dataset. For simplicity we used the parametric specification in estimating the GPS and the average dose response function; the semiparametric methods were used in Flores, Flores-Lagunes, Gonzalez, and Neumann (2012) instead. For actual implementation of the parametric method, refer to Bia and Mattei (2008).

Figures 5 and 6 show the upper and lower bounds of $E[Y_i(t)]$ for different subsamples divided according to the gender, race (white or black), and education (at least 10 years of education or less - corresponding to finishing at least the half of high school education), along with 95% confidence intervals obtained by the method described in Section 5.1. The sample was restricted to those with data on variables of interest - the outcome, treatment, demographic variables used to divide the sample, as well as other conditioning variables necessary to calculate the GPS.[17] The estimates using GPS are also drawn together. We can observe that the latter estimates usually lie between the confidence intervals of SMTR bounds. Moreover, the monotonicity assumption we made seems generally consistent with the results obtained by the GPS approach, except the case of the less educated black male population.

6.2. **Bounds on the Counterfactual Gender and Race Gaps.** In this subsection, we illustrate the usefulness of the SMTR bound by looking at the implied counterfactual gender and race gaps. As discussed in Section 2.1, the following corollary gives the sharp SMTR bound of the average group differences such as the gender or race gap, as long as there is no cross restriction between two groups.

Corollary 6.1. *Assume that the support of $Y_i(t)$ is unbounded. Let $W_i \in \{0,1\}$ denote a binary indicator that splits the population into two exclusive groups. Suppose that there is no cross restriction between two groups. Then under SMTR, the following bound is sharp:*

$$\begin{aligned}
&\{E[Y_i|W_i=1]-E[Y_i|W_i=0]\} - b\left\{E[(Z_i-t)^+|W_i=1]+E[(Z_i-t)^-|W_i=0]\right\}\\
&\le E[Y_i(t)|W_i=1]-E[Y_i(t)|W_i=0]\\
&\le \{E[Y_i|W_i=1]-E[Y_i|W_i=0]\} + b\left\{E[(Z_i-t)^-|W_i=1]+E[(Z_i-t)^+|W_i=0]\right\}.
\end{aligned}$$

Corollary 6.1 implies that the upper bound (UB) of the gender (or race) gap is defined as the difference between the upper SMTR bound of male (white) and the lower SMTR bound of female (black). The lower bound (LB) is defined similarly. Tables 2 and 3 show

[17]Choice of the conditioning variables followed the specification of Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), but we excluded some of them. See Appendix D for details.

corresponding estimates of bounds for $E[Y_i(t)|\text{Male}]-E[Y_i(t)|\text{Female}]$ and $E[Y_i(t)|\text{White}]-E[Y_i(t)|\text{Black}]$, respectively. These bound estimates are obtained in different subsamples, at $t=4$ (one month), 24 (six months), and 36 (nine months), which approximately correspond to 25%, 50%, and 75% quantiles of the outcome variable. The estimated bound is marked as boldface when the sign of the gap can be determined.

Table 2. Estimation of gender outcome gaps (Male - Female)

	White/High Edu		White/Low Edu		Black/High Edu		Black/Low Edu	
	LB	UB	LB	UB	LB	UB	LB	UB
1 month	-18.35	183.16	**53.09**	**237.66**	-85.90	100.70	-92.21	120.68
6 months	-0.25	165.06	**64.86**	**225.88**	-74.68	89.48	-72.59	101.05
9 months	-17.35	182.16	**44.18**	**246.57**	-95.97	110.77	-91.18	119.64

*"High Edu" means more than or equal to 10 years of education. Otherwise individuals are classified as "Low Edu." The unit is dollars per week.

Table 3. Estimation of race outcome gaps (White - Black)

	Male/High Edu		Male/Low Edu		Female/High Edu		Female/Low Edu	
	LB	UB	LB	UB	LB	UB	LB	UB
1 month	-18.53	173.69	**24.42**	**226.03**	-95.37	100.51	-103.84	92.01
6 months	-7.08	162.23	**37.75**	**212.71**	-77.51	82.65	-85.76	73.93
9 months	-27.72	182.87	**19.07**	**231.39**	-95.26	100.41	-106.36	94.54

*"High Edu" means more than or equal to 10 years of education. Otherwise individuals are classified as "Low Edu." The unit is dollars per week.

The estimation result shows that there is strong evidence of the gender gap among the white, less educated population; the minimum gap varies from 44 dollars to 64 dollars a week. There is also evidence of the race gap among less educated males, the minimum gap varying from 19 dollars to 37 dollars a week.

7. Concluding Remarks

In this paper, we have investigated the identification power of smoothness assumptions in the context of partial identification of average counterfactual outcomes. We have obtained a set of new identification results for the average treatment response by imposing smoothness conditions alone, by combining them with monotonicity assumptions, and by adding instrumental variables assumptions to treatment responses. Our result can also be used to tighten the average treatment effect when we combine the smoothness assumption with instrumental variables assumptions. We have demonstrated the usefulness of our approach by reanalyzing the return to schooling example of Manski and Pepper (2000) and also by applying it to the Job Corps Study dataset.

Since information on the upper bound of the average treatment effect is useful for conducting our identification analysis, our approach may be suitable when a policymaker tries to predict the average counterfactual outcome of a new policy, when some average treatment effect estimates are available from previous studies (or a lower-cost pilot study using randomized experiments). Also, our result may be useful when a policymaker makes contingent predictions for the average counterfactual outcome of a new policy, depending on various scenarios of the effectiveness of the treatments. The latter corresponds to the sensitivity analysis approach.

It might be important to extend our analysis to the identification of the entire distribution of counterfactual responses, not just average outcomes. It would also be useful to develop new inference tools for our identification results that cannot be covered by the existing literature. These are interesting topics for future research.

Appendix A. 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 2, 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 A.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 A.1 can be rewritten as

$$-a \leq \frac{g(t, s) - g(t, s')}{s - s'} \leq a \qquad \text{(A.1)}$$

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 (A.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 A.1. Assumption A.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 (2012). Using our notation, equation (9.21) in Section 9.4 of Manski (2003, page 149) states that:

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

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 (2012) considered assumptions of bounded variations: using our notation, for any (t, d, w) and (t', d', w'),

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

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

The interpretation of Proposition A.1 is similar to that of Proposition 2.1. Proposition A.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 A.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 (3.1) 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, i.e. $Y_i(t) \leq y_{\max} < \infty$ for some known $y_{\max}$, we can show that the upper bound for $g^*(t)$ is

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

The upper bound (A.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)):

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

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

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

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

As in the SMTR case, one can test whether there is an strict improvement using Corollary A.2. Moreover, a similar argument as Remark 2.1 can be made to deduce that taking minimum (maximum) of STS and MTS upper (lower) bounds does not give SMTS upper (lower) bound.

A.1. **The STR and STS Bounds.** If we combine STR with STS, we obtain the following result.

Proposition A.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 2.1 and the STS bound in Proposition A.1, we present the special case such that $a = b$ as the following corollary.

Corollary A.4. *Suppose $a = b = \bar{k}$, where $\bar{k}$ denotes the common value. Define $A(t)$ as the event such that $E[Y_i|Z_i = t] \leq E[Y_i|Z_i]$ 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*

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

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

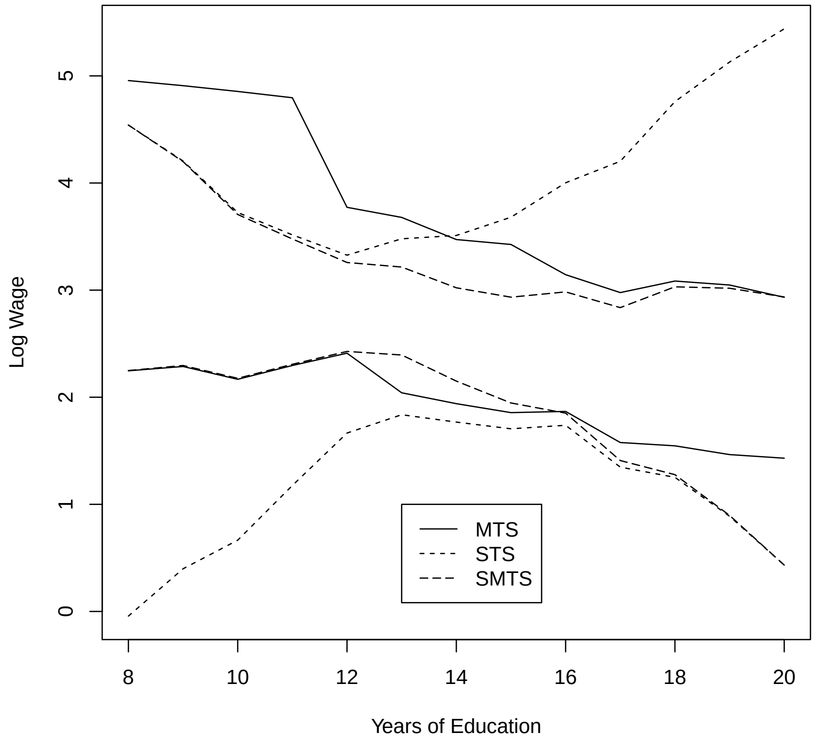
$$l_3(t) = E[Y_i] - \bar{k}E[|Z_i - t|] \text{ and } u_3(t) = E[Y_i|Z_i = t] + \bar{k}E[|Z_i - t|],$$

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 2.1 and that the lower bound of the STR-STS bound is strictly larger than that of the STS bound in Proposition 2.1.

A.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 4 and illustrate the usefulness of STR and SMTS assumptions. Figure 7 shows the SMTS, STS, and MTS bounds when the value of a is 0.4. Here, SMTS and STS bounds are calculated as unbounded version. 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 1. Again Figure 7 shows that there could be a substantial advantage if one combines the smoothness condition with the monotonicity assumption.

A.3. **Inference Using Proposition A.1.** The bounds given in Proposition A.1 are similar to those given in Proposition 2.1. The important difference is that the bounds are around the conditional expectation $E[Y_i|Z_i = t]$, not the overall mean $E[Y_i]$. To reflect this

FIGURE 7. SMTS-STS-MTS comparison



difference, define

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

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

where $\widehat{E[Y_i|Z_i = t]}$ is a local linear estimator of $E[Y_i|Z_i = t]$.[18] Let $\widehat{s}^2(t)$ denote a consistent estimator of the asymptotic variance of $\sqrt{nh}\left(\widehat{E[Y_i|Z_i = t]} - E[Y_i|Z_i = t]\right)$, where h is a bandwidth used in local linear estimation. Since $\widehat{E[Y_i|Z_i = t]}$ cannot be estimated by a rate of $n^{-1/2}$, while $\frac{1}{n}\sum_{i=1}^{n}|Z_i - t|$ can be estimated by $n^{-1/2}$, we set

$$\widetilde{\sigma}_\ell^2 = \widetilde{\sigma}_u^2 \equiv \widehat{s}^2(t),$$

[18]We can allow for other nonparametric estimators, provided that a pointwise normal approximation is readily available.

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

$$\mathrm{CI}_{\alpha}^{\mathrm{STS}}(t) \equiv \left[\widetilde{\theta}_{\ell} - \frac{c_{\alpha}\widetilde{\sigma}_{\ell}}{\sqrt{nh}}, \widetilde{\theta}_{u} + \frac{c_{\alpha}\widetilde{\sigma}_{u}}{\sqrt{nh}} \right],$$

where c_{α} solves

$$\Phi\left(c_{\alpha} + \frac{\sqrt{nh}\widetilde{\Delta}}{\widehat{s}(t)} \right) - \Phi\left(-c_{\alpha}\right) = 1 - \alpha.$$

Again, since $\widetilde{\theta}_{\ell} \leq \widetilde{\theta}_{u}$ by construction, we can modify Lemma 3 and Proposition 1 of Stoye (2009) and rely on undersmoothing to obtain the asymptotic uniform validity of the proposed confidence interval. Analogously, we can obtain a confidence interval for $g^{*}(t)$ under SMTS by redefining $\widetilde{\theta}_{\ell}$, $\widetilde{\theta}_{u}$ and $\widetilde{\Delta}$ as the following:

$$\widetilde{\theta}_{\ell} \equiv \widehat{E[Y_i|Z_i = t]} - a\frac{1}{n}\sum_{i=1}^{n}(Z_i - t)^{-},$$

$$\widetilde{\theta}_{u} \equiv \widehat{E[Y_i|Z_i = t]} + a\frac{1}{n}\sum_{i=1}^{n}(Z_i - t)^{+},$$

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

The STR-STS bound in Proposition A.3 is expressed as the expectation of the maximum between two conditional expectations. Its form in Corollary A.4 is simpler but still involves several terms, including $E[Y_i|Z_i = t]$. For these bounds, there does not seem to exist a suitable inference method yet in the literature. It is an interesting future research topic to develop inference methods for such bounds.

Appendix B. Proofs

In this section, we give the proofs of propositions and corollaries in the paper. The proofs of Corollaries 2.2, 6.1 and A.2 are omitted since they are straightforward.

Proof of Proposition 2.1. **Part (i).** Under STR, we have

$$\int (E[Y_i|Z_i = z] - b|z - t|)\mu(dz) \leq \int E[Y_i(t)|Z_i = z]\mu(dz) \leq \int (E[Y_i|Z_i = z] + b|z - t|)\mu(dz),$$

equivalently,

$$E[Y_i] - bE[|Z_i - t|] \leq \int E[Y_i(t)|Z_i = z]\mu(dz) \leq E[Y_i] + bE[|Z_i - t|].$$

Hence, we obtained the desired bound since $g^{*}(t) = E[Y_i(t)] = \int E[Y_i(t)|Z_i = z]\mu(dz)$.

For the sharpness, consider a DGP 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:

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

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.

Part (ii). We only prove the case for the upper bound. The proof for the lower bound is similar. Under SMTR,

$$\begin{aligned}E[Y_i(t)] &= \int_{z \leq t} E[Y_i(t)|Z_i = z]\mu(dz) + \int_{z > t} E[Y_i(t)|Z_i = z]\mu(dz) \\ &\leq \int_{z \leq t} (E[Y_i|Z_i = z] + b(t - z))\mu(dz) + \int_{z > t} (E[Y_i|Z_i = z] + 0)\mu(dz) \\ &= E[Y_i] + bE[(Z_i - t)^-].\end{aligned}$$

For the sharpness, 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

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

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 3.1 and Corollary 3.2. **Part (i).** Note that under CSTR, Proposition 2.1(i) leads to

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

which holds for any $u \in \mathcal{V}$. Due to Assumption IV, $E[Y_i(t)|V_i = u]$ 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 3.2 can be proved by observing that

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

Setting $E[Y_i(t)|V_i = v]$ at its lower (upper) bound given in Proposition 3.1 (iii) and (iv) yields the result. □

Proof of Proposition 3.3. Suppose $s < t$. Note by the SMTR condition that $g(s,s) \leq g(t,s) \leq g(s,s) + b(t-s)$. Then, for all $s' \in [s,t]$, we have $g(t,s) \leq g(s',s') + b(t-s')$ by MTS and thus $g(t,s) \leq \inf_{s' \in [s,t]}(g(s',s') + b(t-s'))$. Thus, we obtain $g(s,s) \leq g(t,s) \leq \inf_{s' \in [s,t]}(g(s',s')+b(t-s'))$ for all $s < t$. In a similar manner, we obtain $\sup_{s' \in [t,s]}(g(s',s')+b(t-s')) \leq g(t,s) \leq g(s,s)$ for all $s > t$. Hence, it follows that

$$\begin{aligned} s < t &\Rightarrow g(s,s) \leq g(t,s) \leq \inf_{s' \in [s,t]}(g(s',s') + b(t-s')). \\ s = t &\Rightarrow g(t,s) = g(t,t) \\ s > t &\Rightarrow \sup_{s' \in [t,s]}(g(s',s') + b(t-s')) \leq g(t,s) \leq g(s,s). \end{aligned}$$

The lower and upper bounds follow immediately by integrating out s.

For the sharpness, consider a DGP that satisfies

$$g(t,s) = \begin{cases} g(s,s) + b(t-s) & \text{if } s \leq t \\ g(s,s) & \text{if } s > t \end{cases} \tag{B.1}$$

and $g(t,s) \leq g(t,s')$ $\forall s \leq s', \forall t$. Note that this implies that the DGP in (B.1) satisfies $g(s',s') \geq g(s,s) + b(s'-s)$ $\forall s \leq s' \leq t$. Then, if $s < t$, it follows that

$$\inf_{s' \in [s,t]}(g(s',s') + b(t-s')) = g(s,s) + b(t-s) = g(t,s),$$

where the last equality follows from (B.1). Hence, we have

$$\begin{aligned} s < t &\Rightarrow g(t,s) = \inf_{s' \in [s,t]}(g(s',s') + b(t-s')) \\ s = t &\Rightarrow g(t,s) = g(t,t) \\ s > t &\Rightarrow g(t,s) = g(s,s). \end{aligned}$$

This ensures that $E[Y_i(t)]$ attains the upper bound. For the lower bound, consider a DGP that satisfies

$$g(t,s)=\begin{cases} g(s,s) & \text{if } s<t \\ g(s,s)+b(t-s) & \text{if } s\geq t \end{cases}$$

and $g(t,s) \leq g(t,s') \ \forall s \leq s', \forall t$. An analogous argument shows that this DGP indeed attains the lower bound. The convex combinations between the two DGPs yield all the values between the lower and upper bounds. It is obvious that such DGPs satisfy SMTR and MTS. The latter is directly imposed. To see the former, consider the DGP we considered for the upper bound, and observe that we have, for any t_1 and t_2 satisfying $t_1 > t_2$,

$$g(t_1,s)-g(t_2,s)=\begin{cases} b(t_1-t_2) & \text{if } t_1\geq t_2\geq s \\ b(t_1-s) & \text{if } t_1\geq s>t_2 \\ 0 & \text{if } s>t_1\geq t_2, \end{cases}$$

which implies that the DGP satisfies SMTR. □

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

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

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:

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

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,

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

For the sharpness, consider a DGP s.t. $E[Y_i(t)|Z_i = s] = E[Y_i(t)|Z_i = t]$ when $s \leq t$ 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

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

This implies that SMTS holds since $(s_1 - t) \leq (s_1 - s_2)$ when $s_1 > t \geq s_2$. Then the rest can be proved as in part (ii) of the proof of Proposition 2.1. □

Proof of Proposition A.3 and Corollary A.4. The bounds in Proposition A.3 naturally follow from Propositions 2.1 and A.1. For the sharpness, note that there are two cases to consider; STR and STS can hold with 1) $a \geq b$ 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 the propositions 2.1 and A.1. For case 2), replacing b with a leads to the analogous argument which completes the proof.

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

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

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

Appendix C. Additional Theoretical Results

In this section, we provide some additional theoretical results with respect to the average treatment effects.

Proposition C.1. *Consider the average treatment effect,* $\Delta(t, t') \equiv g^*(t) - g^*(t')$ *with* $t > t'$. *Under STR, the sharp bound for* $\Delta(t, t')$ *is* $[-b(t - t'), b(t - t')]$. *Under SMTR, the sharp bound for* $\Delta(t, t')$ *is* $[0, b(t - t')]$.

Proof of Proposition C.1. 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. □

Note that subtracting the lower bound from the upper bound using Proposition 2.1 yields the bound of $\Delta(t, t')$ in the following form:

$$[-b\left(E[|Z_i - t|] + E[|Z_i - t'|]\right), b\left(E[|Z_i - t|] + E[|Z_i - t'|]\right)]. \tag{C.1}$$

Let $t > t'$. For simplicity, suppose that Z_i is a continuous random variable. Now consider the upper bound in (C.1):

$$
\begin{aligned}
& b\left\{E[|Z_i - t|] + E[|Z_i - t'|]\right\} \\
&= b\left\{\int_{z<t}(t-z)\mu(dz) + \int_{z>t}(z-t)\mu(dz) + \int_{z<t'}(t'-z)\mu(dz) + \int_{z>t'}(z-t')\mu(dz)\right\} \\
&= b\left\{\int_{z<t'}(t+t'-2z)\mu(dz) + \int_{z>t}(2z-t-t')\mu(dz) + \int_{t'<z<t}(t-t')\mu(dz)\right\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
b\left\{E[|Z_i - t|] + E[|Z_i - t'|]\right\} - b(t-t') &= 2b\left\{\int_{z<t'}(t'-z)\mu(dz) + \int_{z>t}(z-t)\mu(dz)\right\} \\
&> 0,
\end{aligned}
$$

provided that t or t' is in the interior of the support of Z_i. Therefore, subtracting the lower bound from the upper bound can never achieve the sharp upper bound for the average treatment effect under the STR assumption. Similar arguments can be made for the STR lower bound as well as the SMTR lower and upper bounds.

Proposition C.2. *Consider the treatment effect, $g^*(t) - g^*(t')$. The bound for the average treatment effect implied by Proposition 3.3, $[l_1(t) - u_1(t'), u_1(t) - l_1(t')]$, is not sharp.*

Proof of Proposition C.2. We show that $g^*(t) - g^*(t')$ can never obtain the implied upper bound $u_1(t) - l_1(t')$ for any $t > t'$ under any DGP satisfying the conditions of Proposition 3.3. Suppose there exists some t and t' with $t > t'$ such that $g^*(t) - g^*(t') = u_1(t) - l_1(t')$. Investigating the proof of Proposition 5.3, we can see that this is possible only when $g(\cdot,\cdot)$ satisfies

$$
g(t,s) = \begin{cases} g(s,s) + b(t-s) & \text{if } s \le t \\ g(s,s) & \text{if } s > t \end{cases}
$$

and

$$
g(t',s) = \begin{cases} g(s,s) & \text{if } s < t' \\ g(s,s) + b(t'-s) & \text{if } s \ge t' \end{cases}
$$

almost surely. That is, $P(A) = 0$, where $A \equiv \{s \in \Gamma : s$ does not satisfy the two conditions above$\}$. Then, however, we have

$$g(t,s) - g(t',s) = \begin{cases} b(s-t') & \text{if } s \geq t > t' \\ b(t-t') & \text{if } t > s > t' \\ b(t-s) & \text{if } t > t' \geq s, \end{cases}$$

and thus $|g(t,s) - g(t',s)| > b|t-t'|$ for s such that $s > t$ or $s < t'$ on A^c. Hence, provided that $P(B) > 0$, where $B \equiv \{s \in \Gamma : s > t \text{ or } s < t'\}$, $g(\cdot,\cdot)$ does not satisfy SMTR. A similar argument can also be made to show that the lower bound cannot be obtained. □

Appendix D. Variables Employed in the Estimation of The GPS

In choosing the variables for the estimation of the generalized propensity score in the section 6.1, we tried to follow the specification of Flores, Flores-Lagunes, Gonzalez, and Neumann (2012). However, we had to omit some of the variables employed by the authors, due to (i) the small sample size when we delete observations with missing values for the variables we employ[19] and (ii) data availability issue (the variables related to the geographic information of the participants are not publicly available).

The demographic variables employed here are nonresidential slot indicator, age (as well as the quadratic and cubic terms), and indicators for living in primary metropolitan statistical area or just metropolitan statistical area, having ever been arrested, having a child, being married, being the head of household, living with two parents and using English as native language. Since the gender and race variables were used in dividing the sample, they are not listed here. The health variables include indicators for good/fair/poor health, smoking, drinking and smoking marijuana.

The variables related to education and work experience of the participants include the highest grade measured in years of education (as well as the quadratic term), average weekly earnings at the time of baseline interview and the indicator for having high school diploma, having general educational development diploma, having vocational diploma, having attended or training program in the last year and having ever worked before. Lastly, the variables related to Job Corps (JC) enrolment include the estimated probability of not enrolling in JC, the estimated probability of staying in JC for 30/90/180/270 days, and the indicators for being worried about JC, having heard about JC from parents, having known what center a participant wished to attend, and having known desired job training.

[19]Since we divide our sample into subsamples finer than Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), the small sample problem is more serious.

The variables omitted from the original specification used by Flores, Flores-Lagunes, Gonzalez, and Neumann (2012) due to the small sample issue include the indicators for the type of training a participant wanted from the program (11 categories), expectations from the program (7 categories) and reasons for joining JC (7 categories). The variables excluded due to the data availability issues are Local Unemployment Rates (LURs), LUR at time of JC exit, LUR of 16 to 35 year olds of an individuals race, JC center attended (109 centers), and the State of residence (48 states), all of them necessitating the private geographic information of the participants.

Moreover, interactions with female, white, black, and Hispanic indicators were not used since the bounds are estimated for each subsample of the gender/race category.

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.

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.

Andrews, D. W. K., and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Armstrong, T. B. (2011a): "Asymptotically Exact Inference in Conditional Moment Inequality Models," Working Paper, Stanford University.

——— (2011b): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," Working Paper, Stanford University.

Armstrong, T. B., and H. P. Chan (2013): "Multiscale Adaptive Inference on Conditional Moment Inequalities," Discussion Paper 1885, Cowles Foundation.

Balke, A., and J. Pearl (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92(439), 1171–1176.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2008): "Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization," *American Economic Review*, 98(2), 351–56.

——— (2012): "Treatment effect bounds: An application to Swan-Ganz catheterization," *Journal of Econometrics*, 168(2), 223–243.

Bia, M., and A. Mattei (2008): "A Stata package for the estimation of the dose-response function through adjustment for the generalized propensity score," *The Stata Journal*, 8(3), 354–373.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Unemployment Composition Using Bounds," *Econometrica*, 75(2), 323–363.

Bugni, F. (2010): "Bootstrap Inference for Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

Canay, I. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156(2), 408–425.

Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–81.

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.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Chernozhukov, V., W. Kim, S. Lee, and A. Rosen (2013): "Implementing intersection bounds in Stata," CeMMAP working paper CWP38/13.

Chernozhukov, V., S. Lee, and A. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Chesher, A. (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73(5), 1525–1550.

——— (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78(2), 575–601.

Chetverikov, D. (2011): "Adaptive rate optimal test of conditional moment inequalities," working paper, MIT.

Chiburis, R. (2010): "Semiparametric bounds on treatment effects," *Journal of Econometrics*, 159(2), 267–275.

De Haan, M. (2011): "The Effect of parents schooling on childs schooling: A nonparametric bounds analysis," *Journal of Labor Economics*, 29(4), 859–892.

Fan, Y., and S. Park (2014): "Nonparametric inference for counterfactual means: Bias-correction, confidence sets, and weak IV," *Journal of Econometrics*, 178(Part 1), 45–56.

Fan, Y., R. Sherman, and M. Shum (2014): "Identifying Treatment Effects under Data Combination," *Econometrica*, forthcoming.

Fan, Y., and J. Wu (2010): "Partial Identification of the Distribution of Treatment Effects in Switching Regime Models and its Confidence Sets," *Review of Economic Studies*, 77(3), 1002–1041.

Flores, C. A., A. Flores-Lagunes, A. Gonzalez, and T. C. Neumann (2012): "Estimating the effects of length of exposure to instruction in a training program: the case of job corps," *Review of Economics and Statistics*, 94(1), 153–171.

Flores-Lagunes, A., A. Gonzalez, and T. C. Neumann (2010): "Learning but not Earning? The Impact of Job Corps Training on Hispanic Youth," *Economic Inquiry*, 48(3), 651–667.

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.

Gerfin, M., and M. Schellhorn (2006): "Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits - Swiss evidence," *Health Economics*, 15(9), 1011–1020.

Gonzalez, L. (2006): "Nonparametric bounds on the returns to language skills," *Journal of Applied Econometrics*, 20(6), 771–795.

Gundersen, C., and B. Kreider (2008): "Food Stamps and Food Insecurity What Can Be Learned in the Presence of Nonclassical Measurement Error?," *Journal of Human Resources*, 43(2), 352–382.

——— (2009): "Bounding the effects of food insecurity on childrens health outcomes," *Journal of Health Economics*, 28(5), 971–983.

Gundersen, C., B. Kreider, and J. Pepper (2012): "The impact of the National School Lunch Program on child health: A nonparametric bounds analysis," *Journal of Econometrics*, 166(1), 79–91.

Hall, P., and A. Yatchew (2007): "Nonparametric estimation when data on derivatives are available," *Annals of Statistics*, 35(1), 300–323.

——— (2010): "Nonparametric least squares estimation in derivative families," *Journal of Econometrics*, 157(2), 362 – 374.

Hausman, J., and W. Newey (2013): "Individual heterogeneity and average welfare," CeMMAP working paper CWP34/13.

Heckman, J. J., and E. Vytlacil (1999): "Local instrumental variables and latent variable models for identifying and bounding treatment effects," *Proceedings of the National Academy of Sciences*, 96(8), 4730–4734.

Heckman, J. J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation1," *Econometrica*, 73(3), 669–738.

Hirano, K., and G. W. Imbens (2004): "The propensity score with continuous treatments," in *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, ed. by A. Gelman, and X.-L. Meng. Wiley InterScience, West Sussex, England.

Imai, K., and D. A. Van Dyk (2004): "Causal inference with general treatment regimes," *Journal of the American Statistical Association*, 99(467), 854–866.

Imbens, G., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Jun, S. J., J. Pinkse, and H. Xu (2011): "Tighter bounds in triangular systems," *Journal of Econometrics*, 161(2), 122–128.

Kitagawa, T. (2009): "Identification region of the potential outcome distributions under instrument independence," CeMMAP working paper CWP30/09.

Kluve, J., H. Schneider, A. Uhlendorff, and Z. Zhao (2012): "Evaluating continuous training programmes by using the generalized propensity score," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(2), 587–617.

Kreider, B., and S. C. Hill (2009): "Partially identifying treatment effects with an application to covering the uninsured," *Journal of Human Resources*, 44(2), 409–449.

Kreider, B., and J. V. Pepper (2007): "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," *Journal of the American Statistical Association*, 102(478), 432–441.

Kreider, B., J. V. Pepper, C. Gundersen, and D. Jolliffe (2012): "Identifying the effects of SNAP (Food Stamps) on child health outcomes when participation is endogenous and misreported," *Journal of the American Statistical Association*, 107(499), 958–975.

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

Leamer, E. E. (1985): "Sensitivity Analyses Would Help," *American Economic Review*, 75(3), 308–313.

Lee, S., K. Song, and Y.-J. Whang (2013a): "Testing a General Class of Functional Inequalities," arXiv working paper, arXiv:1311.1595.

——— (2013b): "Testing Functional Inequalities," *Journal of Econometrics*, 172(1), 14–32.

Lee, S., and R. Wilke (2009): "Reform of Unemployment Compensation in Germany: A Nonparametric Bounds Analysis Using Register Data," *Journal of Business and Economic Statistics*, 27(2), 193–205.

Manski, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80(2), 319–323.

——— (1997): "Monotone Treatment Response," *Econometrica*, 65(6), 1311–1334.

——— (2003): *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

——— (2007): *Identification for Prediction and Decision*. Harvard University Press, New York.

——— (2013): "Identification of treatment response with social interactions," *Econometrics Journal*, 16(1), S1–S23.

Manski, C. F., and J. V. Pepper (1998): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," NBER Technical Working Paper No. 224.

——— (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

——— (2009): "More on Monotone Instrumental Variables," *Econometrics Journal*, 12(s1), S200–S216.

——— (2012): "Partial identification of treatment response with data on repeated cross sections," Working Paper. Northwestern University and University of Virginia.

——— (2013): "Deterrence and the Death Penalty: Partial Identification Analysis Using Repeated Cross Sections," *Journal of Quantitative Criminology*, 29(1), 123–141.

Okumura, T., and E. Usui (2013): "Concave-Monotone Treatment Response and Monotone Treatment Selection: With an Application to the Returns to Schooling," *Quantitative Economics*, forthcoming.

Pepper, J. V. (2000): "The intergenerational transmission of welfare receipt: A nonparametric bounds analysis," *Review of Economics and Statistics*, 82(3), 472–488.

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.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78(1), 169–211.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117.

Schochet, P. Z., J. Burghardt, and S. Glazerman (2001): *National Job Corps Study: The Impacts of Job Corps on Participants Employment and Related Outcomes*. Mathematica Policy Research, Princeton, NJ.

Schochet, P. Z., J. Burghardt, and S. McConnell (2008): "Does job corps work? Impact findings from the National Job Corps Study," *American Economic Review*, 98(5), 1864–1886.

Shaikh, A. M., and E. J. Vytlacil (2011): "Partial Identification in Triangular Systems of Equations with Binary Dependent Variables," *Econometrica*, 79(3), 949–955.

Stoye, J. (2009): "More on Confidence Regions for Partially Identified Parameters," *Econometrica*, 77(4), 1299–1315.

Tamer, E. (2010): "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167–195.