THE IDENTIFICATION POWER OF SMOOTHNESS ASSUMPTIONS IN MODELS WITH COUNTERFACTUAL OUTCOMES

WOOYOUNG KIM1, KOOHYUN KWON2, SOONWOO KWON3, AND SOKBAE LEE2 4

Abstract. In this paper, we investigate what can be learned about average counterfactual outcomes as well as average treatment effects when it is assumed that treatment response functions are smooth. We obtain a set of new partial identification results for both the average treatment response and the average treatment effect. In particular, we find that the monotone treatment response and monotone treatment selection bound of Manski and Pepper (2000) can be further tightened if we impose the smoothness conditions on the treatment response. Since it is unknown in practice whether the imposed smoothness restriction is met, it is desirable to conduct a sensitivity analysis with respect to the smoothness assumption. We demonstrate how one can carry out a sensitivity analysis for the average treatment effect by varying the degrees of smoothness assumption. We give a numerical illustration of our findings by reanalyzing the return to schooling example of Manski and Pepper (2000).

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

JEL Classification: C14, C18, C21, C26.

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

1Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706-1393, USA.
2Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Republic of Korea.
3Department of Economics, Yale University, PO Box 208268, New Haven, CT 06520-8268, USA.
4Centre 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, soonwoo.kwon@yale.edu, sokbae@snu.ac.kr.

Date: 23 October 2014.

We would like to thank Steve Bond, Andrew Chesher, Toru Kitagawa, Charles Manski, Adam Rosen and Kyungchul Song for helpful comments. This work was supported by the European Research Council (ERC-2009-StG-240910-ROMETA) and by the BK21Plus Program (Future-oriented innovative brain raising type, 21B20130000013) funded by the Ministry of Education and National Research Foundation of Korea.

In this paper, we build on Manski (1997) and Manski and Pepper (2000), and investigate what can be learned about average counterfactual outcomes as well as average treatment effects when it is assumed that treatment response functions are smooth. The smoothness conditions in this paper amount to assuming that there exists a bound for the changes in the average treatment response with respect to the changes in the treatment. 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 $\mu$ denote the probability distribution of $Z_i$, which may be discrete, continuous, or mixed.

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 $t \in \Gamma$. 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 $\mu$ denote the probability distribution of $Z_i$, which may be discrete, continuous, or mixed.

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 $-\infty \leq y_{\min} \leq y_{\max} \leq \infty$. Then, the partial identification analysis of $g^*(t)$ starts from the well-known Manski’s worst-case bound (see, for example, Proposition 1.1 of Manski (2003)):

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

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 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. This problem may arise under the extrapolation problem as well. It is also easy to think of a situation where the difference between the upper and lower bounds of the outcome variable is 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 paper is organized as follows. In Section 2, we introduce new assumptions on treatment responses and obtain corresponding identification results. In Section 3, we find that the monotone treatment response (MTR) and monotone treatment selection (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 a sensitivity analysis by varying the degrees of smoothness assumption. Section 5 gives concluding remarks. Appendix A contains the proofs of all the identification results in the paper.

The reader is referred to Kim, Kwon, Kwon, and Lee (2014), which is a working paper version of this paper, for related identification results when an instrumental variable or a monotone instrumental variable exists, discussions on inference methods, a real-data example, and remarks on the literature regarding empirical applications.

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 function and $\mu$ can be any probability measure as mentioned before. 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 $t, t', s, s', z \in \Gamma$ 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 (2014), 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, \( t \geq t' \Rightarrow Y_i(t) \geq Y_i(t') \) (2.1) for each individual \( i \). Our monotonicity assumption in the SMTR condition is in the same spirit as 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 (2014) used the bounds on the income effect to partially identify average consumer surplus. We follow Hausman and Newey (2014) 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

\[
-b \leq \frac{g(t,s) - g(t',s)}{t - t'} \leq b
\]

for all \( t \neq t' \) and for all \( s \).\(^2\) Regarding \( g(\cdot,s) \) as a function of only the first argument for each \( s \), the quotient in (2.2) 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 on the real line, this is equivalent to assuming that \( g(\cdot,s) \) is Lipschitz continuous with respect to the first argument uniformly in \( s \).

\(^2\)More generally, one may consider (2.2) with two different end points \( b_1 \) and \( b_2 \), as in Hausman and Newey (2014). Our STR and SMTR conditions are special cases of \((b_1, b_2) = (-b, b)\) and \((b_1, b_2) = (0, b)\), respectively.
Note that the inequalities in (2.2) can be satisfied if
\[-b \leq \frac{Y_i(t) - Y_i(t')}{t - t'} \leq b\] (2.3)
for all \(t \neq t'\) and for each \(i\). Assuming (2.3) 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 \(\beta\) is a slope parameter and \(\delta_i\) is an unobserved random variable for each individual \(i\). The STR condition is satisfied by the HLR assumption, as long as \(\beta \leq b\).

Remark 2.1. 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) added concavity to the basic assumption of monotonicity and showed formally that concavity has substantial identifying power. See also Okumura and Usui (2014) 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 (2014), restricts the rate of change indirectly. Therefore, two approaches are distinct as well as complementary. □

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 (2014) for how to set bounds 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 a sensitivity analysis by looking at different values of \(b\). In Section 4 we provide an example of sensitivity analyses. See Leamer (1985), Tamer (2010), and others for general discussions on sensitivity analyses; see also Chen, Tamer, and Torgovitsky (2011) for a recent development on sensitivity analyses in semiparametric likelihood models in the context of partial identification.

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 the STR condition, the sharp bound is symmetric around $E[Y_i]$ and its width is $2bE[|Z_i - t|]$. Proposition 2.1 (ii) implies that under the SMTR condition, 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 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 the 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 $Y_i(t) \leq y_{\text{max}} < \infty$ for some known $y_{\text{max}}$. Then it is straightforward to show that the SMTR upper bound for $g^*(t)$ is

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

The upper bound (2.4) 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)):

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

(2.5)

Note that the SMTR upper bound strictly improves the MTR upper bound if and only if the event such that $E[Y_i|Z_i] + b(t - Z_i) < y_{\text{max}}$ has a strictly positive probability, conditional on $Z_i < t$. Analogous results can be established for 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 $\Gamma_0$ denote a closed subset of $\Gamma$ where the STR and SMTR conditions locally hold. Then the identification results presented above can be translated as those for $E[Y_i(t)|Z_i \in \Gamma_0]$ for $t \in \Gamma_0$. □

We now show that the STR or SMTR assumption alone does not provide a meaningful identification result for the average treatment effects.

**Proposition 2.2.** 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 $[0, b(t - t')]$.

Although the bound with the STR or SMTR condition alone is not attractive in terms of identifying the average treatment effects, our approach is useful to bound other parameters. To give such an example, suppose that $W_i$ is the gender of individual $i$. 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. 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, for example, Blundell, Gosling, Ichimura, and Meghir (2007) and Lee and Wilke (2009) for related results.

### 3. Adding the Smoothness Assumption to the MTR-MTS Bound

In this section, 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):

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

We examine the role of smoothness assumption for the MTR-MTS bound by replacing the MTR assumption with the SMTR condition. The following proposition gives the sharp bounds for the average counterfactual outcomes.
Proposition 3.1. Under the SMTR and MTS assumptions together, we have that
\[ E[Y_i(t)] \in [l_1(t), u_1(t)], \]
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.

It is useful to compare our bounds in Proposition 3.1 with the MTR-MTS bound of Manski and Pepper (2000):
\[ l_{MP}(t) \leq E[Y_i(t)] \leq u_{MP}(t), \]
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 see how the SMTR-MTS bound improves the MTR-MTS bound, note that the MTR-MTS constraint implies that \( E[Y_i|Z_i = t] \) is an increasing function of \( t \) (Manski and Pepper, 2000). Hence, the integrand for the second term of \( l_1(t) \) can be strictly larger than \( E[Y_i|Z_i = t] \) or the integrand for the first term of \( u_1(t) \) can be strictly smaller than \( E[Y_i|Z_i = t] \), provided that \( b \) is sufficiently small. However, when \( b \) is large enough, the SMTR-MTS bound reduces to the MTR-MTS bound of Manski and Pepper (2000). Thus, the SMTR-MTS bound can be made tighter than the MTR-MTS bound only if \( b \) is reasonably small, i.e. the treatment response is sufficiently smooth.

To derive the sharp bounds on the average treatment effect \( \Delta(t_1, t_2) \equiv E[Y_i(t_2)] - E[Y_i(t_1)] \), define
\[ f_S(s, t_1) \equiv \sup_{s' \in [t_1, s]} \{ E[Y_i|Z_i = s'] - b(s' - t_1) \} \quad \text{for } t_1 \leq s, \]
\[ f_I(s, t_2) \equiv \inf_{s' \in [s, t_2]} \{ E[Y_i|Z_i = s'] + b(t_2 - s') \} \quad \text{for } t_2 \geq s. \]

The following proposition gives the main result of the paper.
Proposition 3.2. Suppose $t_2 > t_1$. Then, under the SMTR and MTS assumptions together, $\Delta(t_1, t_2) \in [0, u_2(t_1, t_2)]$, where

$$u_2(t_1, t_2) \equiv \int_{s < t_1} \min \{f_I(s, t_2) - g(s, s), b(t_2 - t_1)\} \mu(ds) + \int_{t_1 \leq s \leq t_2} \min \{f_I(s, t_2) - f_S(s, t_1), b(t_2 - t_1)\} \mu(ds) + \int_{s > t_2} \min \{g(s, s) - f_S(s, t_1), b(t_2 - t_1)\} \mu(ds).$$

Moreover, this bound is sharp.

Remark 3.1. When $b$ is large enough, the upper bound $u_2(t_1, t_2)$ reduces to the sharp bound of the MTR-MTS bound. Specifically, if $b$ is large enough that $b(t_2 - t_1)$ is not binding in any of the minimum operators in $u_2(t_1, t_2)$, we have that

$$u_2(t_1, t_2) = u_{MP}(t_2) - l_{MP}(t_1).$$

On the other hand, when $b$ is sufficiently small, $u_2(t_1, t_2) = b(t_2 - t_1)$. For intermediate values of $b$ between these two extremes, we have a meaningful upper bound that is strictly tighter than the MTR-MTS bound without the smoothness assumption. To emphasize these intermediate values, we define the effective region of $b$ to be the range of $b$ which gives the smaller upper bound for the average treatment effect than the MTR-MTS bound and also gives the smaller upper bound than $b(t_2 - t_1)$. □

Since the value of $b$ is unknown in practice, it is natural to present the identification result obtained in Proposition 3.2 for all possible values of $b$, thereby identifying the effective region of $b$. This exercise can be viewed as a sensitivity analysis with respect to the degrees of the smoothness assumption. In Section 4, using the empirical example of Manski and Pepper (2000), we present a sensitivity analysis that shows the SMTR-MTS bound improves the MTR-MTS bound for a wide range of $b$. 

In this section, we revisit the return 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 $b$.

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 hourly 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 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. In the NBER working paper version of Manski and Pepper (2000), Manski and Pepper (1998) used $[y_{\min}, y_{\max}] = [1.4, 5.0] \approx [\ln(4.25), \ln(150)]$, where $4.25$ per hour is the official minimum wage in 1994 and $150$ per hour exceeds the sample maximum ($138$ per hour) in 1994. We use the same values in our analysis of the MTR bound but did not use them for the STR and SMTR bounds, following Proposition 2.1.

Figure 1 shows the SMTR, STR, and MTR bounds when the value of $b$ is 0.2. 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 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. Figure 1 demonstrates that there could be a substantial shrinkage of the identification region if one combines the smoothness condition with the monotonicity assumption.
4.2. Bounds on Average Treatment Effects and Sensitivity Analysis. We now consider the average treatment effect $\Delta(t_1, t_2) \equiv E[Y_i(t_2)] - E[Y_i(t_1)]$ under the SMTR-MTS assumptions. Recall that MTR-MTS constraint requires $g(t, t)$ to be weakly increasing in $t$. However, the estimated function of $g(t, t)$ reported in Manski and Pepper (2000) is not an increasing function of $t$, possibly due to random sampling errors. Following Chernozhukov, Fernandez-Val, and Galichon (2009), we sort the estimates of $g(t, t)$ in an increasing order and rearrange them to construct monotonized estimates of $g(t, t)$.

Using the modified estimates on $g(t, t)$, we calculated the sharp upper bounds on the average treatment effect for all possible values of $b$. Table 1 and Figures 2 and 3 report how different choices of $b$ affect the identification region of the average treatment effect. In the figures, the solid line is the sharp upper bound for $\Delta(s, t)$ under the SMTR-MTS assumption. As $b$ increases, the upper bound becomes flat, approaching the MTR-MTS upper bound.\footnote{In our empirical exercise, the MTR-MTS upper bound is tighter than the one reported in Manski and Pepper (2000) because we used the rearranged estimates of $g(t, t)$, whereas Manski and Pepper (2000) used the unconstrained estimates.}
Table 1. The upper bounds for $\Delta(12, 16)$ and $\Delta(16, 18)$

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

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

Recall that in Remark 3.1, we have defined the effective region of $b$ to be the range of $b$ which strictly improves the upper bound for the average treatment effect under the MTR-MTS assumption but also gives the smaller upper bound than $b(t_2 - t_1)$. The upper bounds corresponding to the effective region of $b$ are marked as boldface in Table 1. For $\Delta(12, 16)$ and $\Delta(16, 18)$, the effective regions turn out to be $[0.04, 0.14]$ and $[0.08, 0.34]$, respectively. Obtaining the effective region of $b$ amounts to conducting a 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 $b$. We can see that the upper bound for $\Delta(16, 18)$ is improved for more values of $b$. This is not surprising since the smoothness assumption can be more useful when $P(Z_i = t)$ is small and there are fewer observations with $Z_i = 18$ than those with $Z_i = 12$. 
5. 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 as well as the average treatment effect by imposing smoothness conditions alone and by combining them with monotonicity assumptions. We have demonstrated the usefulness of our approach by reanalyzing the return to schooling example of Manski and Pepper (2000).

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 results may be useful when a policymaker makes contingent predictions for both the average counterfactual outcome and the average treatment effect of a new policy, depending on various scenarios of the effectiveness of the treatments. The latter corresponds to the sensitivity analysis approach.
The existing literature does not provide inference methods for the SMTR-MTS bounds we developed in this paper. 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. It might be important to extend our analysis to the identification of the entire distribution of counterfactual responses and also to the identification of quantile treatment effects, not just average outcomes or average treatment effects. These are interesting topics for future research.

APPENDIX A. PROOFS

In this section, we give the proofs of propositions in the paper.

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

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

The proof of Part (ii) follows a similar argument.
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:

\[ |E[Y_i(t_1)|Z_i = s] - E[Y_i(t_2)|Z_i = s]| = b|s - t_1| - |s - t_2| \]
\[ \leq b|t_1 - t_2|. \]

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,

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

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.

\[ \square \]

*Proof of Proposition 2.2.* To verify the sharpness of the STR upper bound, consider a HLR-type DGP s.t. \( Y_i(t) = \beta \times t + \delta_i \) with \( E\delta_i = 0 \) and \( \beta = b \) satisfies STR, as mentioned
in the main text, and this DGP yields $\Delta(t, t') = b(t - t')$. Likewise, the sharp lower bound is $-b(t - t')$ (take $\beta = -b$). Now the convex combinations between the two DGPs yield all the values between the lower and upper bounds. Identical arguments yield that the sharp SMTR upper and lower bounds for the average treatment effect are $b(t - t')$ and 0, respectively. □

**Proof of Proposition 3.1.** 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

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

$s = t \Rightarrow g(t, s) = g(t, t)$

$s > t \Rightarrow \sup_{s' \in [t, s]} (g(s', s') + b(t - s')) \leq g(t, s) \leq g(s, s)$.

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

For the sharpness, we will first assume that $g(s, s)$ is a monotone increasing function of $s$ since the MTR-MTS assumption requires $g(s, s)$ to be monotone increasing. Then, consider the following DGP:

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

for all $s, t$.

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

$$g(t_2, s) - g(t_1, s) = g(s, s) - g(s, s) = 0,$$

so SMTR holds. For the second case,

$$g(t_2, s) - g(t_1, s) = \inf_{s' \in [s, t_2]} (g(s', s') + b(t_2 - s')) - g(s, s) \geq \inf_{s' \in [s, t_2]} g(s', s') - g(s, s) = 0.$$
since \( g(s, s) \) is an increasing function with respect to \( s \). Therefore, MTR holds. Moreover,
\[
g(t_2, s) - g(t_1, s) = \inf_{s' \in [s, t_2]} (g(s', s') + b(t_2 - s')) - g(s, s)
\]
\[
\leq g(s, s) - b(t_2 - s) - g(s, s)
\]
\[
= b(t_2 - s)
\]
\[
\leq b(t_2 - t_1)
\]
since \( t_1 \leq s \) for the second case. This verifies the smoothness condition, so SMTR holds.
For the third case
\[
g(t_2, s) - g(t_1, s) = \inf_{s' \in [s, t_2]} (g(s', s') + b(t_2 - s')) - \inf_{s' \in [s, t_1]} (g(s', s') + b(t_1 - s')).
\]
Note that
\[
\inf_{s' \in [s, t_2]} (g(s', s') + b(t_2 - s')) = \min \left\{ \inf_{s' \in [s, t_1]} (g(s', s') + b(t_2 - s')), \inf_{s' \in [t_1, t_2]} (g(s', s') + b(t_2 - s')) \right\}.
\]
For the first term inside the minimum operator,
\[
\inf_{s' \in [s, t_1]} (g(s', s') + b(t_2 - s')) = \inf_{s' \in [s', t_1]} (g(s', s') + b(t_1 - s') + b(t_2 - t_1))
\]
\[
\geq \inf_{s' \in [s', t_1]} (g(s', s') + b(t_1 - s')).
\]
Moreover, for the second term inside the minimum operator, we can see that
\[
\inf_{s' \in [t_1, t_2]} (g(s', s') + b(t_2 - s')) \geq \inf_{s' \in [t_1, t_2]} g(s', s')
\]
\[
= g(t_1, t_1)
\]
\[
\geq \inf_{s' \in [s, t_1]} (g(s', s') + b(t_1 - s')).
\]
Therefore, MTR holds. For the case of SMTR, let \( A \equiv \inf_{s' \in [s, t_1]} (g(s', s') + b(t_1 - s')) \) and \( B \equiv \inf_{s' \in [t_1, t_2]} (g(s', s') + b(t_1 - s')) \). Then,
\[
g(t_2, s) - g(t_1, s) = \inf_{s' \in [s, t_2]} (g(s', s') + b(t_2 - s')) - \inf_{s' \in [s, t_1]} (g(s', s') + b(t_1 - s'))
\]
\[
= \inf_{s' \in [s, t_2]} (g(s', s') + b(t_1 - s')) + b(t_2 - t_1) - \inf_{s' \in [s, t_1]} (g(s', s') + b(t_1 - s'))
\]
\[
= \min \{ A, B \} + b(t_2 - t_1) - A.
\]
Since \( \min \{ A, B \} \leq A \), we can see that SMTR holds.
Next, we need to check MTS, i.e. \( g(t, s_2) - g(t, s_1) \geq 0 \) for \( s_1 < s_2 \). There are again three cases to consider: 1) \( s_1 < s_2 \leq t \), 2) \( s_1 \leq t < s_2 \), 3) \( t < s_1 < s_2 \). For the first case,
\[
g(t, s_2) - g(t, s_1) = \inf_{s' \in [s_2, t]} (g(s', s') + b(t - s')) - \inf_{s' \in [s_1, t]} (g(s', s') + b(t - s')) \geq 0,
\]
since \([s_2, t] \subset [s_1, t]\). For the second case,
\[
g(t, s_2) - g(t, s_1) = g(s_2, s_2) - \inf_{s' \in [s_1, t]} (g(s', s') + b(t - s')) \geq g(s_2, s_2) - g(t, t) \geq 0.
\]
For the third,
\[
g(t, s_2) - g(t, s_1) = g(s_2, s_2) - g(s_1, s_1) \geq 0.
\]
Therefore, MTS holds for this DGP. Likewise, the DGP that achieves the lower bound is
\[
g(t, s) = \begin{cases} 
g(s, s) & s \leq t \\
\sup_{s' \in [t, s]} (g(s', s') - b(s' - t)) & s > t
\end{cases}
\]
We omit the proof that it satisfies SMTR-MTS, since it is analogous to the previous one, and the convex combination of the previous DGP and this one yields all the values between the upper and the lower bounds.

\[\square\]

**Proof of Proposition 3.2.** Note that
\[
g^*(t_2) - g^*(t_1) = \int_{s < t_1} g(t_2, s) - g(t_1, s) \mu(ds) + \int_{t_1 \leq s \leq t_2} g(t_2, s) - g(t_1, s) \mu(ds) + \int_{s > t_2} g(t_2, s) - g(t_1, s) \mu(ds).
\]
For \( s < t_1 \), \( g(t_2, s) - g(t_1, s) \) is less than \( b(t_2 - t_1) \) due to the smoothness condition. Moreover, it is also less than \( f_I(s, t_2) - g(s, s) \), which is clear if we subtract the upper bound for \( g(t_2, s) \) from the lower bound for \( g(t_1, s) \), which are obtained in the proof for the Proposition 3.1. Therefore, we get
\[
\int_{s < t_1} g(t_2, s) - g(t_1, s) \mu(ds) \leq \int_{s < t_1} f_I(s, t_2) - g(s, s) \mu(ds).
\]
\[ \int_{s<t_1} \min \{ f(s,t_2) - g(s,s), b(t_2 - t_1) \} \mu(ds). \] The discussion regarding the last two terms in the proposition can be proceeded by the similar manner.

The DGP that achieves the sharp bound is constructed by the following manner. Note that we again assume \( g(s,s) \) is increasing in \( s \). First, fix \( t_1 \) and \( t_2 \). Then, define

\[
f_2(s) = \begin{cases} f_1(s,t_2) & s \leq t_2 \\ g(s,s) & s > t_2, \end{cases}
\]

\[
n_1(s) = \begin{cases} g(s,s) & s \leq t_1 \\ n_S(s,t_1) & s > t_1, \end{cases}
\]

\[
f_1(s) = \max \left\{ n_1(s), f_2(s) - b(t_2 - t_1) \right\}.
\]

Our DGP is:

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

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

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

1) \( g(t',s) - g(t,s) = f_1(s) - f_1(s) = 0 \).
2) first note that

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

by the definition of \( f_1(s) \). Moreover, also note that \( f_2(s) - f_1(s) \geq 0 \) by the definitions of \( f_S(s,t_1) \) and \( f_1(s,t_2) \). Then, \( g(t',s) - g(t,s) = \left( \frac{f_2(s) - f_1(s)}{t_2 - t_1} \right) (t' - t_1) \geq 0 \) and \( \left( \frac{f_2(s) - f_1(s)}{t_2 - t_1} \right) (t' - t_1) \leq b(t' - t_1) \leq b(t' - t) \) since \( t \leq t_1 \).
3) \( g(t',s) - g(t,s) = \left( \frac{f_2(s) - f_1(s)}{t_2 - t_1} \right) (t' - t) \) which is greater than 0 and smaller than \( b(t' - t) \) by (A.1).
4) \( g(t',s) - g(t,s) = \left( \frac{t_2 - t}{t_2 - t_1} \right) (f_2(s) - f_1(s)) \geq 0 \) and \( \left( \frac{t_2 - t}{t_2 - t_1} \right) (f_2(s) - f_1(s)) \leq b(t_2 - t) \leq b(t' - t) \) since \( t' > t_2 \).
5) \( g(t',s) - g(t,s) = f_2(s) - f_1(s) \). The discussion regarding the second case shows this is greater than 0 and smaller than \( b(t_2 - t_1) \).
6) \( g(t', s) - g(t, s) = f_2(s) - f_2(s) = 0. \)

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

\[
g(t, s_2) - g(t, s_1) = f_1(s_2) - f_1(s_1) = \max \left\{ \tilde{f}_1(s_2), f_2(s_2) - b(t_2 - t_1) \right\} - \max \left\{ \tilde{f}_1(s_1), f_2(s_1) - b(t_2 - t_1) \right\}.
\]

In order to show this to be greater than 0, first note that \( a_2 \geq a_1 \) and \( b_2 \geq b_1 \) implies \( \max \{a_2, b_2\} \geq \max \{a_1, b_1\} \). Therefore we only need to show (i) \( \tilde{f}_1(s_2) \geq \tilde{f}_1(s_1) \) and (ii) \( f_2(s_2) \geq f_2(s_1) \). We have already verified inequality (ii) in the proof of Proposition 3.1. Also, note that inequality (i) is analogous to inequality (ii). Therefore, case 1) has been dealt with. Now for the second case,

\[
g(t, s_2) - g(t, s_1) = f_1(s_2) + \left( \frac{f_2(s_2) - f_1(s_2)}{t_2 - t_1} \right)(t - t_1) - f_1(s_1) - \left( \frac{f_2(s_1) - f_1(s_1)}{t_2 - t_1} \right)(t - t_1)
\]

\[
= \left( \frac{t_2 - t}{t_2 - t_1} \right)(f_1(s_2) - f_1(s_1)) + \left( \frac{t - t_1}{t_2 - t_1} \right)(f_2(s_2) - f_2(s_1)),
\]

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

References


