Quantitative Economics 9 (2018), 617–642 1759-7331/20180617

The identification power of smoothness assumptions in models with counterfactual outcomes

Wooyoung Kim
Department of Economics, University of Wisconsin-Madison

Koohyun Kwon
Department of Economics, Yale University

Soonwoo Kwon
Department of Economics, Yale University

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

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 illustrate our findings by reanalyzing the return to schooling example of Manski and Pepper (2000) and also by measuring the effect of the length of job training on the labor market outcomes.

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

JEL classification. C14, C18, C21, C26.

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

We would like to thank the Editor, Rosa Matzkin, three anonymous referees, Steve Bond, Andrew Chesher, Toru Kitagawa, Charles Manski, Adam Rosen, Kyungchul Song, and Elie Tamer for helpful comments. This work is in part supported by the European Research Council (ERC-2014-CoG-646917-ROMIA) and by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A2A01014041).

 https://doi.org/10.3982/QE545

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; see, for example, 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), Jun, Pinkse, and Xu (2011), Kitagawa (2009), Lee (2009), Manski (1990, 1997, 2013), Manski and Pepper (2000, 2009, forthcoming), Okumura and Usui (2014), and Shaikh and Vytlacil (2011) among many others.

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 they are called *smoothness* conditions.[1]

To describe our setup, let $\Gamma \subset \mathbb{R}$ denote the treatment space that can be finite, countably infinite, or uncountable, and let $Y_i(t)$ denote a random variable that gives the individual-level potential outcome for treatment $t \in \Gamma$. Hence, $\{Y_i(t) : t \in \Gamma\}$ is a real-valued stochastic process that we would like to learn about.

Assume that we observe independent and identically distributed observations $\{(Y_i, Z_i) : i = 1, \dots, n\}$, where Z_i is the actual treatment for individual i that takes values in a subset of Γ, and $Y_i \equiv Y_i(Z_i)$ is this individual's observed outcome.[2] Let μ denote the probability distribution of Z_i, which may be discrete, continuous, or mixed.[3]

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 the partial identification analysis of $g^*(t)$ starts from the well-known Manski's worst-case bound (see, e.g., Proposition 1.1 of Manski (2003)):

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

[1]We argue that there are three motivations to smoothness assumptions: first, economic theory may deliver smoothness; second, smoothness can be used for sensitivity analysis; third, smoothness conditions are typically assumed for nonparametric estimation and inference. We would like to thank one of anonymous referees for pointing out the third motivation.

[2]It is possible that the support of Z_i is a strict subset of Γ.

[3]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 $y_{\max}$ and $y_{\min}$ 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 describe two empirical examples with which we will illustrate the usefulness of our approach. In Section 3, we introduce new assumptions on treatment responses and obtain corresponding identification results. In particular, we find that the monotone treatment response (MTR) bound of Manski (1997) can be tightened if we impose the smoothness conditions on the treatment response. In Section 4, we show that adding the smoothness to the treatment response improves the monotone treatment response and monotone treatment selection (MTR-MTS) bound of Manski and Pepper (2000). In Section 5, we revisit the return to schooling example of Manski and Pepper (2000), and in Section 6, we use data from the National Job Corps Study and show how to bound the effect of the length of job training on the labor market outcomes. Moreover, in Sections 5 and 6, we demonstrate how one can conduct a sensitivity analysis by varying the degrees of smoothness assumption. Section 7 gives concluding remarks. The online supplement consists of six Appendices and is available in a supplementary file on the journal website, http://qeconomics.org/supp/545/supplement.pdf. Appendix A contains all proofs omitted from the main text, including the proofs for the sharpness results. In Appendix B, we focus on the binary treatment case to better understand the role of the smoothness assumption. In Appendix C, we show how to tighten the identification results obtained in Section 3 when an instrumental variable exists. In Appendix D, we show how to use the smoothness assumption with respect to treatment selection. Appendix E provides discussions on statistical inference. Finally, Appendix F gives additional empirical results that are omitted from the main text. Replication files are available in a supplementary file on the journal website, http://qeconomics.org/supp/545/code_and_data.zip.

Notation

Throughout the paper, we write the expectation of a function of Z_i as $E[\varphi(Z_i)] = \int \varphi(z)\mu(dz)$, where $\varphi(\cdot)$ is a given function and μ can be any probability measure as mentioned before. For example, if the distribution of Z_i is continuous, $E[\varphi(Z_i)] = \int \varphi(z)\mu(dz) = \int \varphi(z) p_\mu(z)\,dz$, where $p_\mu(\cdot)$ is the probability density function of Z_i. Alternatively, if the distribution of Z_i is discrete, $E[\varphi(Z_i)] = \int \varphi(z)\mu(dz) = \sum_j \varphi(z_j) p_\mu(z_j)$, where $p_\mu(\cdot)$ is now the probability mass function of Z_i. 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. Let $x^+ \equiv \max(x, 0)$ and $x^- \equiv \max(-x, 0)$ for any real number x.

2. Motivating empirical applications

2.1 *Return to schooling*

As our first empirical example, we revisit the return to schooling example of Manski and Pepper (2000). They use data from the NLSY79 (National Longitudinal Survey of Youth 1979) and restrict the sample to be 1257 white males who were full-time year-round workers in 1994. In this example, the treatment t is years of schooling, $Y_i(t)$ is the logarithm of counterfactual hourly wages for t years of schooling, and Y_i and Z_i are observed log hourly wages in 1994 and years of schooling. Manski and Pepper (2000) argue that the monotone treatment response (MTR) and monotone treatment selection (MTS) assumptions are plausible in the case of return to schooling and obtain the upper bound of the average differences between two different years of schooling.[4] We will demonstrate that the MTR-MTS bound of Manski and Pepper (2000) can be tightened further if we add smoothness conditions.

2.2 *Effects of job corps*

Our second empirical example is based on the National Job Corps[5] Study (NJCS), which contain data for applicants for Job Corps between November 1994 and February 1996. This study is based on the random assignment of eligible applicants who were divided into treatment, control, and nonresearch groups.

We use the survey analysis sample, which includes 11,313 youths who completed a 48-month interview, belonging to treatment (6828) and control (4485) groups. Among these, there are 774 individuals with missing values for the wages or the training duration, and we are excluding them from our analysis. We should also note that there are noncompliers in the treatment group (2111) and the control group (198). Excluding observations with missing vales and noncompliers, we end up with 8306 observations (4207 treatments and 4099 controls).

For a detailed description of the Job Corps study methodology, the dataset and previous research, see Schochet, Burghardt, and Glazerman (2001), Schochet, Burghardt, and McConnell (2008), Lee (2009), Flores-Lagunes, Gonzalez, and Neumann (2010), and Flores, Flores-Lagunes, Gonzalez, and Neumann (2012) among others. As pointed out by Flores et al. (2012), the length of the training program varies among individuals since each individual designs his/her own training program with the help of JC counselors. Therefore, the NJCS provides a possibility of measuring the effect of the length of job training on the labor market outcomes.

To describe the setting, let the treatment variable Z_i be the length of exposure to the academic and vocational instruction (henceforth "AV instruction") in weeks[6] for individual i, and let the counterfactual outcome $Y_i(t)$ be the weekly earnings (in dollars) or the binary employment status (1 if employed; 0 otherwise) of individual i 48 months after

[4]The MTR assumption is given in (3.1) and the MTS assumption is presented in (4.1).

[5]Job Corps, established in 1964, is a nationwide job-training program administered by the U.S. Department of Labor that provides vocational education to young people ages 16 through 24.

[6]It is calculated by dividing the total training hours by 40, and hence, can be treated as continuous.

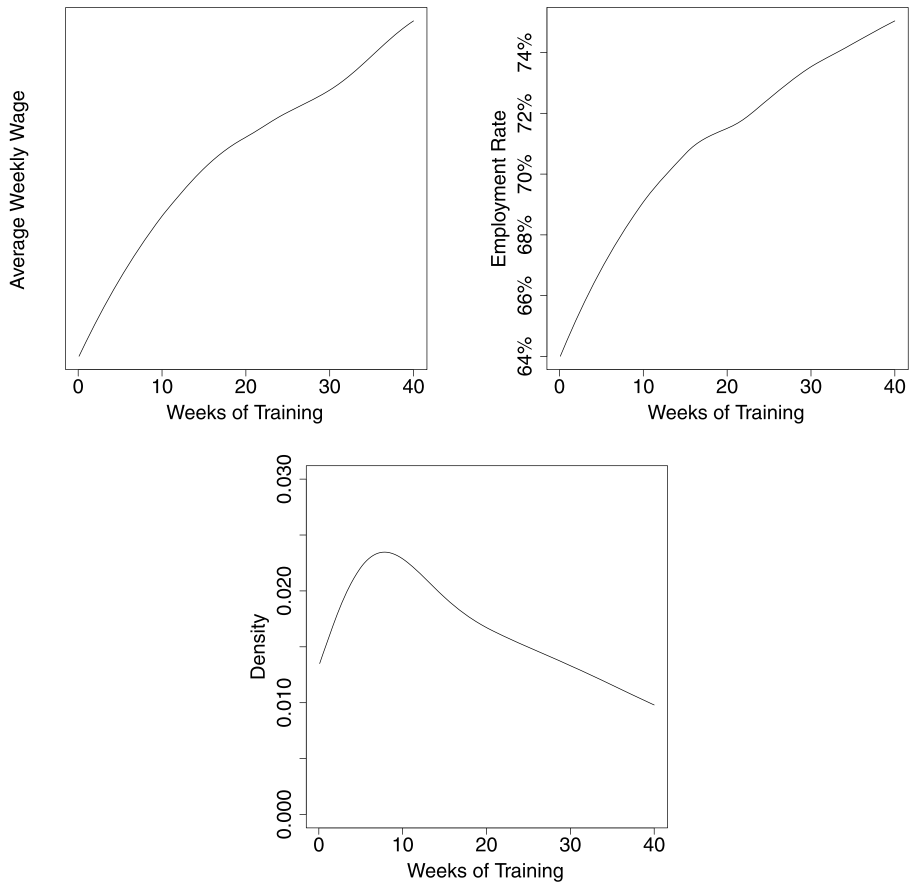

FIGURE 1. Wages, employment status and training durations in the treatment group. *Notes*: The top left and right panels of the figure show nonparametric estimates of the conditional means of weekly earnings and employment status given weeks of training, respectively, and the bottom panel depicts the kernel density estimates of training durations.

random assignment, given that the individual received t weeks of AV instruction during the program.[7] The outcomes after 4 years of random assignment are the usual outcomes of interest in the literature.

Figure 1 and Table A-2 in Online Appendix F provide descriptive statistics for the data used in the paper. Throughout this example, weights were used in all calculations to adjust for the sample and survey designs. It can be seen in Table A-2 that the mean earnings for treatment and control groups are 218.81 and 202.24 dollars, respectively. The percentages employed for these groups are 71.26 and 68.46, respectively. These figures replicate those reported in the NJCS (the corresponding figures are \$217.5, \$199.4, 71.1%, and 68.7% in Table A1 of Schochet, Burghardt, and McConnell (2008)), indicating that our estimation sample does not seem to be biased in any particular way. For both males

[7]Precisely, we define the weekly earnings to be the average of weekly wages in the 16th quarter after random assignment and the employment status to be one if the weekly wage is positive and 0 otherwise.

and females, the Job Corps program increased the average earnings and employment rates.

The average enrollment to the program is about 8 months but the realized treatment varies substantially in the data. The maximum is greater than 200 weeks but the 90% sample quantile is just 43 weeks. In what follows, we will restrict the treatment space Γ to be $[0, 40]$ for identification analysis. This is mainly because we will utilize monotonicity assumptions, which may not hold for excessively large values of training weeks. For example, the MTR condition assumes that $Y_i(t)$ is a nondecreasing function of t; it may not be plausible to assume that the earnings will still increase after a long period of job training. We will make further comments in Remark 3.2.

In Figure 1, the top left and right panels of the figure show nonparametric estimates of the conditional means of weekly earnings and employment status given weeks of training, respectively, and the bottom panel depicts the kernel density estimates of training durations.[8] In each figure, the horizontal axis is limited to 40 weeks, as we restricted the treatment space Γ to be $[0, 40]$. Note that the conditional means of both outcomes are increasing in the realized training weeks. Since the MTR and MTS conditions together imply that $E[Y_i|Z_i = z]$ is a nondecreasing function of z (Manski and Pepper (2000)), it seems plausible to make the MTR and MTS assumptions in this example.

Thanks to the random assignment, the NJCS data provide credible estimates of the counterfactual outcomes without the Job Corps program. Using this feature of the dataset, our parameters of interest will be two different types of the average treatment effects: first, $E[Y_i(t)] - E[Y_i|i \in \text{control group}]$, which measures the impact of a treatment t relatively to the control group; second, $E[Y_i(t_1)] - E[Y_i(t_2)]$ for different t_1 and t_2, which measures the average changes of the counterfactual outcomes by varying t within the treatment group. Both will provide useful information about the effects of the length of job training.

As mentioned in the Introduction, the difficulty of identification arises from the fact that we only observe $Y_i \equiv Y_i(Z_i)$ in the treatment group, not all the counterfactual outcomes. Flores et al. (2012) estimate semiparametrically "dose-response functions" under the uncomfoundedness assumption, employing generalized propensity score introduced by Hirano and Imbens (2004). In this paper, we revisit their analysis but adopt a partial identification approach.

3. 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. Recall that $g(t, s) \equiv E[Y_i(t)|Z_i = s]$.

[8]The nonparametric conditional mean estimates are obtained by the `loess` command in R with default options and the density is estimated by the kernel density estimator using the Gaussian kernel with the Silverman's rule of thumb bandwidth. The bounds reported in Section 6 are computed via numerical integration using these nonparametric estimates.

Assumption 3.1 (Treatment Response Assumptions). *Assume one of the following conditions*:

(i) (Condition STR) *There exists a known 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 known constant* $b > 0$. *In addition,* $g(t,s) \geq g(t',s)\ \forall t, t', s \in \Gamma$ *satisfying* $t \geq t'$.

Assumption 3.1, which is inspired by Manski (1997) and Hausman and Newey (2016), 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,

$$t \geq t' \quad \Rightarrow \quad Y_i(t) \geq Y_i(t') \tag{3.1}$$

for each individual i. Our monotonicity assumption in the SMTR condition is in the same spirit as Manski (1997), but slightly weaker than (3.1) 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 (2016) used the bounds on the income effect to partially identify average consumer surplus. We follow Hausman and Newey (2016) to make Assumption 3.1, while allowing for the case that the treatment is not continuous.

The "smoothness" condition in the STR condition can be rewritten as

$$-b \leq \frac{g(t,s) - g(t',s)}{t - t'} \leq b \tag{3.2}$$

for all $t \neq t'$ and for all s.[9] Regarding $g(\cdot,s)$ as a function of only the first argument for each s, the quotient in (3.2) is called in general the difference quotient of $g(\cdot,s)$. Hence, part (i) of Assumption 3.1 amounts to assuming that $g(\cdot,s)$, as a function of the first argument, has bounded difference quotients uniformly in s. This is equivalent to assuming that $g(\cdot,s)$ is Lipschitz continuous with respect to the first argument uniformly in s.[10]

Note that the inequalities in (3.2) can be satisfied if

$$-b \leq \frac{Y_i(t) - Y_i(t')}{t - t'} \leq b \tag{3.3}$$

for all $t \neq t'$ and for each i. Assuming (3.3) amounts to bounding the individual-level treatment effect defined as $[Y_i(t) - Y_i(t')]/(t - t')$. Manski and Pepper (2009) considered

[9]More generally, one may consider (3.2) with two different end points b_1 and b_2, as in Hausman and Newey (2016). Our STR and SMTR conditions are special cases of $(b_1, b_2) = (-b, b)$ and $(b_1, b_2) = (0, b)$, respectively.

[10]In order to achieve identification, it is assumed that that $E[Y_i(t)|Z_i = s]$ is smooth in t for all s. Strictly speaking, assuming smoothness on $E[Y_i(t)|Z_i = s]$ is different from assuming smoothness on the bounding function. The main nonparametric component in estimation of the bounding function is $E[Y_i|Z_i = s]$. That is, it is necessary to assume smoothness assumption on $g(t,t)$ for the purpose of estimation; however, we need to have smoothness on $g(t,s)$ that is beyond the case that $t = s$.

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 $\beta \leq b$.

Remark 3.1. 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 (2016) 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.

In other applications, we perceive that choosing b inherently involves some subjective belief about the maximum size of treatment effects, and our identification result is obtained conditional on that belief. One possible route to choose b formally is to rely on Bayesian inference using presamples or information from prior elicitation. Using existing experimental results or from previous research, one may obtain a posterior distribution regarding b and use a high quantile of the posterior distribution as a possible value of b. If there is no related information available, the prior on b will not be updated and we need to rely on our purely subjective belief on the value of b. We now explain how we choose b in our examples.

Example (Return to Schooling). We use a simple theoretical model of endogenous schooling in Card (2001) to motivate how we choose b in the return to schooling example. In Card (2001, equation (1)), an optimal schooling choice Z_i for individual i satisfies the following condition: in our notation,

$$f_i'(Z_i)/f_i(Z_i) = d_i(Z_i), \tag{3.4}$$

where the left-hand side $f_i'(Z_i)/f_i(Z_i)$ of (3.4) is the individual-specific marginal return to schooling and the right-hand side $d_i(Z_i)$ is the individual-specific marginal costs.

The form of the marginal return to schooling involves a first-order derivative of $f_i(\cdot)$, which is the individual-level earnings function of schooling. Thus, the existence of the marginal return implicitly assumes that the log earning function $t \mapsto \log f_i(t)$ is differentiable, thereby implying that the counterfactual outcome $Y_i(t) \equiv \log f_i(t)$ is a smooth function of t. A simple specification of $f_i'(t)/f_i(t)$ in Card (2001, equation (2)) is

$$f_i'(t)/f_i(t) = \beta_i - k_1 t, \tag{3.5}$$

where β_i is an individual-specific random variable and $k_1 \geq 0$ is a constant. Equation (3.5) implies that

$$Y_i(t) \equiv \log f_i(t) = \alpha_i + \beta_i t - \frac{1}{2} k_1 t^2, \tag{3.6}$$

where α_i is an individual-specific random intercept term. Note that the MTR condition is satisfied in (3.6) if $\beta_i - k_1 t \geq 0$ for each $t \in \Gamma$ and for each individual i. To motivate the SMTR condition, suppose that $t_1 \geq t_2$ and write

$$E\big[Y_i(t_1)|Z_i = s\big] - E\big[Y_i(t_2)|Z_i = s\big] = (t_1 - t_2)\left\{E[\beta_i|Z_i = s] - \frac{1}{2} k_1 (t_1 + t_2)\right\}.$$

Thus, the SMTR condition is satisfied if there exists a known constant $b > 0$ such that

$$0 \leq E[\beta_i|Z_i = s] - \frac{1}{2} k_1 (t_1 + t_2) \leq b$$

for every $s, t_1, t_2 \in \Gamma$. Note that as long as $E[\beta_i|Z_i = s]$ is bounded, the simple structural model in (3.5) and (3.6) implies that the mapping $t \mapsto E[Y_i(t)|Z_i = s]$ is Lipschitz continuous for each s with a universal constant b. In short, the theoretical argument ensures that there exists such a constant b but it does not deliver the known value for b; it is an empirical question how to choose b.

We now link our choice of b to the previous studies. Chamberlain and Imbens (2003) considered a simple model of earnings and schooling and used a nonparametric approach to Bayesian inference. In their model, potential log earnings follow a linear model with a slope parameter γ (in their notation), which is common to all individuals. They used Angrist and Krueger (1991)'s data to obtain the posterior distribution on γ, which is analogous to $E[Y_i(t) - Y_i(t-1)]$ in our model, assuming that this quantity is constant for all t. They found that the 97.5th percentile of the posterior distribution of γ turned out to be 0.132 and that a normal approximation with mean 0.089 and standard deviation 0.021 would provide a good approximation to their posterior distribution.

We take $b = 0.2$ as the conservative baseline value. If we use the normal approximation as suggested by Chamberlain and Imbens (2003), the posterior probability that $\gamma > 0.2$ is almost zero. Roughly speaking, this corresponds to the maximum of 20 percentage points in the average return to 1 year of schooling. For US samples, OLS and IV estimates of the returns to education are typically less than 0.1 (see, e.g., Card (2001, Table II)). 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 1 year of college. Their estimate varies between 0.0626 and 0.1409, across different samples and specifications (see Carneiro, Heckman, and Vytlacil (2011, Table 6)). In view of these estimates, we regard our choice of b as a plausible upper bound.

Example (Effects of Job Corps). We consider two outcomes: weekly earnings and employment rates. Economic theory suggests that the job training program can increase both: the latter can be achieved by providing career counseling and encouraging individuals to enter labor force. The former may increase due to the increase in labor supply and/or because of the increase in human capital; see, for example, Heckman, Lalonde, and Smith (1999) for overview of evaluations of labor market programs.

In view of the standard economic models, it is reasonable to assume that total earnings and employment rates are, on average, nondecreasing smooth functions of job training intensity. For example, in modeling a worker's optimal choice of job training duration, marginal increases in expected total earnings or those in the probability of being employed may be equated to marginal increases in job training costs. This effectively implies that under the rational expectation assumption, counterfactual average outcomes are smooth functions of job training intensity. Hence, the SMTR assumption is reasonable in this example.

As in the previous example, choosing the value of b is an empirical question; and it is helpful to have the average treatment estimates between treatment and control groups in the NJCS data. The estimated impact on average weekly earnings and percentage employed 16 quarters after random assignment are 25.2 dollars and 3.3%, respectively (see the last column of Table A1 of Schochet, Burghardt, and McConnell (2008)). The average duration of job training in the NJCS data is 30 weeks (see Table A-2 in Online Appendix F), implying that the impact estimates are translated into the increases of 84 cents and 0.11% per week of training. These provide starting points for the choice of b.

When the outcome is the employment status, we take $b = 0.17\%$ by simply increasing the impact estimate of the NJCS by 50%. For the earnings, we consider two channels of increases, as we have described above. First, we calibrate the wage effect by relying on the results of Lee (2009), who obtains the upper bound of the wage effect of the Job Corps program under relative weak assumptions. His point estimates of the upper bound of the average treatment for *log wages* are 0.093 in Table 4 of Lee (2009) and 0.0899 in Table 5 of Lee (2009). The latter is tighter than the former because of the use of covariates. To be on the conservative side, we take the value of 0.12, which is greater than either of the upper ends of the 95% confidence intervals reported in Table 4 of Lee (2009), to compute the upper bound of the average return to one week of additional training. Then the value of 0.12 amounts to the average increase of 0.4% per week ($= 100 \times 0.12/30$). Recall that the average weekly earnings for the control group (including zeros) are about 200.2 dollars (see Table A1 of Schochet, Burghardt, and McConnell (2008)). We take 80.1 cents (0.4% increment of 200.2 dollars) as the upper bound of the wage effect. Second, we consider the labor supply effect. Since the impact estimate of an additional week of training on the probability of employment is 0.0011, its effect on earnings is 22 cents. Again to be on the conservative side, by increasing the impact estimate by 50%, we take 33 cents to compute the upper bound of the labor supply effect. Adding these two effects together and rounding it up, our baseline value of b for weekly earnings is $b = 1.14$ (1 dollar and 14 cents). When we look at males and females separately, we recalibrate b for weekly earnings using their respective earnings for the control group. These are $b = 1.30$ for males and $b = 0.92$ for females, respectively.

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.[11] In Section 5, 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, recall that $x^+ \equiv \max(x, 0)$ 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 3.1. *Assume that the support of $Y_i(t)$ is unbounded. Then the following bounds are sharp*:

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

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

Proof. (i) Under STR, we have

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

equivalently,

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

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

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

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

[11] There is a sensitivity parameter that is related to but different from our sensitivity parameter b in the literature on partial identification. To study identification with contaminated and corrupted data, Horowitz and Manski (1995) considered an ex ante known upper bound on the probability of a data error.

The sharpness of such bounds for both parts (i) and (ii) follows from Lemma A.1. □

Proposition 3.1(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 3.1(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 (3.1) 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 $Y_i(t) \le y_{\max} < \infty$ for some known $y_{\max}$. Then we can show that the SMTR upper bound for $g^*(t)$ is

$$\begin{aligned} g^*(t) \le & \int_{z<t} \min\{y_{\max}, \big(E[Y_i|Z_i = z] + b(t - z)\big)\}\mu(dz) \\ & + E[Y_i|Z_i \ge t]P(Z_i \ge t). \end{aligned} \tag{3.7}$$

The upper bound (3.7) 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)):

$$g^*(t) \le y_{\max}P(Z_i < t) + E[Y_i|Z_i \ge t]P(Z_i \ge t). \tag{3.8}$$

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.

Remark 3.2. 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 nonsmooth 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 denote a closed subset of the support of Z_i and assume that the STR and SMTR conditions locally hold on Γ. Then the identification results presented above can be translated as those for $E[Y_i(t)|Z_i \in \Gamma]$ for $t \in \Gamma$. As we mentioned in Section 2.2, we will restriction Γ to be $[0, 40]$ when we analyze the effect of the length of job training, so that all conditional expectations and probabilities in that example are conditional on $Z_i \in [0, 40]$.

The following proposition asserts that the STR or SMTR assumption alone does not provide a meaningful identification result for the average treatment effect $\Delta(t, t') \equiv g^*(t) - g^*(t')$ for $t > t'$.

PROPOSITION 3.2. *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 et al. (2007) and Lee and Wilke (2009) for related results.

3.1 *Generalizations of the smoothness condition: Modulus of continuity*

Our objective is to bound the magnitude of $|g(t, s) - g(t', s)|$ relative to the magnitude of $|t - t'|$ in some systematic way, and imposing Lipschitz continuity on the function $g(\cdot, s)$ (uniformly in s) is likely one of the simplest such structures we can think of. A more general form will be rather assuming $|g(t, s) - g(t', s)| \leq \omega(|t - t'|)$, where $\omega : [0, \infty) \to [0, \infty)$ with $\lim_{x\to 0} \omega(x) = 0$. Such a function ω is called the modulus of continuity. It quantifies the uniform continuity of functions, and a function $f(\cdot)$ admits a modulus of continuity if and only if $f(\cdot)$ is uniformly continuous.

ASSUMPTION 3.2 (Treatment Response Assumptions Under the Modulus of Continuity). *Assume one of the following conditions*:

(i) (Condition STR-MoC) *There exists a known function* $\omega(\cdot)$ *such that*

$$\left|g(t, s) - g\left(t', s\right)\right| \leq \omega\left(\left|t - t'\right|\right) \quad \forall t, t', s \in \Gamma,$$

where $\omega : [0, \infty) \to [0, \infty)$ *with* $\lim_{x\to 0} \omega(x) = \omega(0) = 0$.

(ii) (Condition SMTR-MoC) *The STR-MoC condition in part* (i) *holds with a known function* $\omega(\cdot)$. *In addition,* $g(t, s) \geq g(t', s)$ $\forall t, t', s \in \Gamma$ *satisfying* $t \geq t'$.

Note that $w(x) = b|x|$ corresponds to our original Lipschitz continuity and $w(x) = b|x|^a$ for a nonnegative constant a to Hölder continuity. The bounds on $g^*(t)$ can be generalized straightforwardly. Proposition 3.1 can be modified as follows.

Proposition 3.3. *Assume that the support of $Y_i(t)$ is unbounded. Then the following bounds are sharp*:

(i) *Under STR-MoC*, $E[Y_i] - E[\omega(|Z_i - t|)] \le g^*(t) \le E[Y_i] + E[\omega(|Z_i - t|)]$.

(ii) *Under SMTR-MoC*, $E[Y_i] - E[\omega((Z_i - t)^+)] \le g^*(t) \le E[Y_i] + E[\omega((Z_i - t)^-)]$.

3.2 *Statistical inference*

Although the main theme of this paper is on identification, it is important to discuss the corresponding inference problem. In Online Appendix E, 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.

4. 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):

$$s \ge s' \quad \Rightarrow \quad E\big[Y_i(t)|Z_i = s\big] \ge E\big[Y_i(t)|Z_i = s'\big]. \tag{4.1}$$

As emphasized in Manski and Pepper (2000), the MTS assumption is consistent with standard economic models of schooling and wages that assume that individuals with higher ability have higher counterfactual wages and choose higher levels of schooling than do those with lower ability. Just as schooling, it is plausible to assume that individuals with higher ability stay longer in the training program than do those with lower ability (up to 40 weeks, as we restrict the treatment space Γ to be $[0, 40]$). Hence, the MTS assumption is also reasonable in our second example in view of standard economic theories of human capital accumulation. Thus, we will explore the identification power of the MTS assumption in both examples.

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 4.1. *Under the SMTR and MTS assumptions together, we have that* $E[Y_i(t)] \in [l_1(t), u_1(t)]$, *where*

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

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

Moreover, this bound is sharp.

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

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

$$s = t \quad \Rightarrow \quad g(t,s) = g(t,t), \quad \text{and}$$

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

The lower and upper bounds follow immediately by integrating out s. The sharpness of such bounds follows by Lemma A.2. □

It is useful to compare our bounds in Proposition 4.1 with the MTR-MTS bound of Manski and Pepper (2000):

$$l_{\text{MP}}(t) \le E\big[Y_i(t)\big] \le u_{\text{MP}}(t),$$

where

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

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

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, that is, 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

$$f_S(s,t_1) \equiv \sup_{s'\in[t_1,s]}\big\{E\big[Y_i|Z_i = s'\big] - b\big(s'-t_1\big)\big\} \quad \text{for } t_1 \le s,$$

$$f_I(s,t_2) \equiv \inf_{s'\in[s,t_2]}\big\{E\big[Y_i|Z_i = s'\big] + b\big(t_2-s'\big)\big\} \quad \text{for } t_2 \ge s.$$

The following proposition gives the upper bound for the average treatment effect.

Proposition 4.2. *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*

$$u_2(t_1,t_2) \equiv \int_{s<t_1} \min\big\{f_I(s,t_2) - g(s,s), b(t_2-t_1)\big\}\mu(ds)$$

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

Moreover, this bound is sharp.

Proof. Note that

$$g^*(t_2) - g^*(t_1) = \int_{s<t_1} g(t_2, s) - g(t_1, s)\mu(ds)$$
$$+ \int_{t_1 \le s \le t_2} g(t_2, s) - g(t_1, s)\mu(ds)$$
$$+ \int_{s>t_2} g(t_2, s) - g(t_1, s)\mu(ds).$$

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 lower bound for $g(t_1, s)$ from the upper bound for $g(t_2, s)$, which is obtained in the proof for the Proposition 4.1. Therefore, we get $\int_{s<t_1} g(t_2, s) - g(t_1, s)\mu(ds) \le \int_{s<t_1} \min\{f_I(s, t_2) - g(s, s), b(t_2 - t_1)\}\mu(ds)$. The last two terms in the proposition can be derived in a similar manner. Sharpness follows by Lemma A.3. □

Proposition 4.2 shows that the smoothness assumption can help bound the average treatment effect when it is combined with the MTS assumption. While the sharp upper bound under the SMTR assumption is $b(t_2 - t_1)$ as shown in Proposition 3.2, the minimum operators in $u_2(t_1, t_2)$ show the bound can now be nontrivial, that is, less than $b(t_2 - t_1)$, for the case of SMTR-MTS, and this occurs when at least one of the three terms in the minimum operators is less than $b(t_2 - t_1)$. This also shows that the smaller the smoothness parameter, the smaller the gain from assuming the MTS condition.

Remark 4.1. 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

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

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 $b(t_2 - t_1)$.

Remark 4.2 (Ternary Treatment). Suppose there are three treatment levels: t_2, t_1, and t_0 with $t_2 > t_1 > t_0$. If $g(t_1, t_1) - g(t_0, t_0) > b(t_1 - t_0)$, then $u_2(t_0, t_1) = b(t_1 - t_0)$, the trivial upper bound. Moreover, when $g(t_2, t_2) - g(t_0, t_0) \le b(t_1 - t_0)$, $u_2(t_0, t_1) = u_{\text{MP}}(t_1) - l_{\text{MP}}(t_0)$.

However, for the intermediate case where $g(t_1, t_1) - g(t_0, t_0) < b(t_1 - t_0) < g(t_2, t_2) - g(t_0, t_0)$, $u_2(t_0, t_1)$ is smaller than both $b(t_1 - t_0)$ and the MTR-MTS upper bound. The cases for the other similar parameters of interest are analogous. This shows that as the number of possible treatment options increases, the room for improvement by imposing the smoothness assumption increases as well, relative to the original MTR-MTS bound. The case of binary treatment is presented in Online Appendix B.

4.1 *Results under the modulus of continuity*

We return to the case of modulus of continuity in Section 3.1 and obtain generalized versions of Propositions 4.1 and 4.2.

PROPOSITION 4.3. *Under the SMTR-MoC and MTS assumptions together, we have that* $E[Y_i(t)] \in [l_1(t), u_1(t)]$, *where*

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

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

Moreover, this bound is sharp.

Define

$$f_{S,\omega}(s, t_1) \equiv \sup_{s'\in[t_1,s]} \{E[Y_i|Z_i = s'] - \omega(s' - t_1)\} \quad \text{for } t_1 \leq s,$$

$$f_{I,\omega}(s, t_2) \equiv \inf_{s'\in[s,t_2]} \{E[Y_i|Z_i = s'] + \omega(t_2 - s')\} \quad \text{for } t_2 \geq s.$$

PROPOSITION 4.4. *Suppose* $t_2 > t_1$, *and assume SMTR-MoC and MTS with* $\omega(t_2 - t_1) > 0$. *Then,* $\Delta(t_1, t_2) \in [0, u_2(t_1, t_2)]$, *where*

$$\begin{aligned} u_2(t_1, t_2) \equiv &\int_{s<t_1} \min\{f_{I,\omega}(s, t_2) - g(s, s), \omega(t_2 - t_1)\}\mu(ds) \\ &+ \int_{t_1\leq s\leq t_2} \min\{f_{I,\omega}(s, t_2) - f_{S,\omega}(s, t_1), \omega(t_2 - t_1)\}\mu(ds) \\ &+ \int_{s>t_2} \min\{g(s, s) - f_{S,\omega}(s, t_1), \omega(t_2 - t_1)\}\mu(ds). \end{aligned}$$

Moreover, this bound is sharp.

5. RETURN TO SCHOOLING: MANSKI AND PEPPER (2000) REVISITED

In this section, we return to the example in Section 2.1 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.

5.1 *Bounds on average counterfactual outcomes*

Recall that in this example, 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 do not use them for the STR and SMTR bounds, following Proposition 3.1.

Figure 2 shows the SMTR, STR, and MTR bounds when the value of b is 0.2. We have explained this choice of b in Section 3. 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 2 demonstrates that there could be a substantial shrinkage of the identification region if one combines the smoothness condition with the monotonicity assumption.

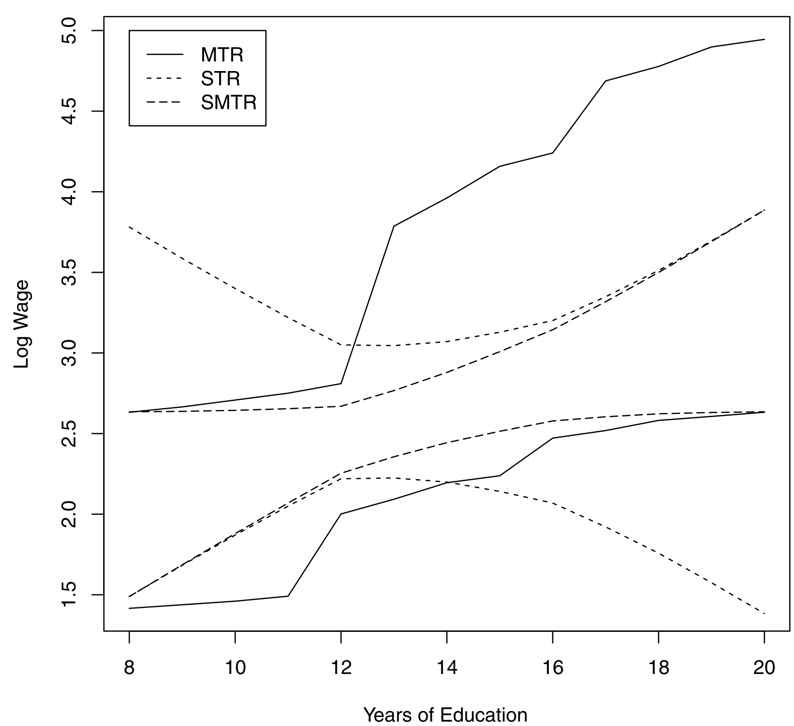

Figure 2. SMTR-STR-MTR comparison.

5.2 *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 calculate the sharp upper bounds on the average treatment effect for all possible values of b. Figure 3 reports how different choices of b affect the identification region of the average treatment effect. In the figure, the solid lines are the sharp upper bounds for $\Delta(s, t)$ under the SMTR-MTS assumption. As b increases, the upper bound becomes flat, approaching the MTR-MTS upper bound.[12]

Recall that in Remark 4.1, 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 $b(t_2 - t_1)$. 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.[13]

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

[12]In our empirical exercise, the MTR-MTS upper bound is tighter than the one reported in Manski and Pepper (2000) because we use the rearranged estimates of $g(t, t)$, whereas Manski and Pepper (2000) use the unconstrained estimates.

[13]See Table A-1 in Online Appendix F for the concrete numbers for the upper bounds.

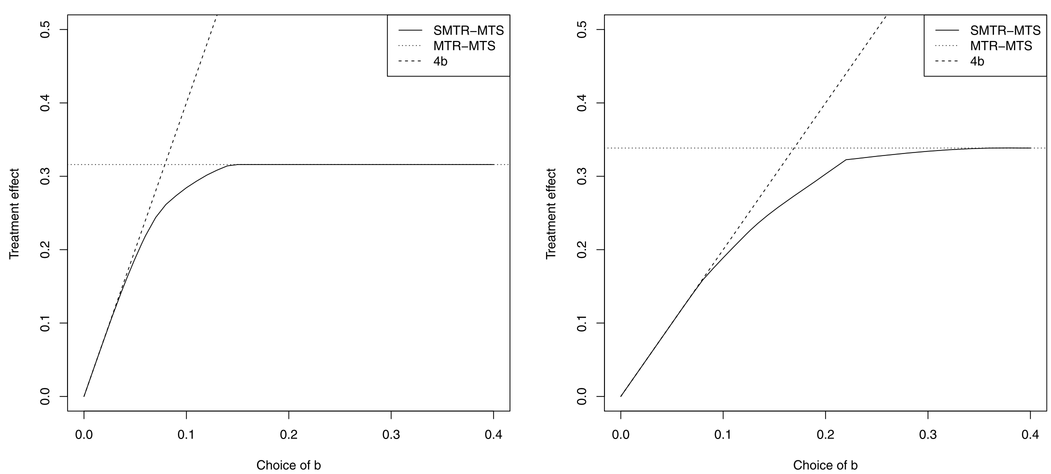

FIGURE 3. The upper bound for $\Delta(12, 16)$ and $\Delta(16, 18)$.

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 $Z_i = 12$. In fact, there is no improvement in the upper bound for the return to college when b is our baseline value ($b = 0.2$) and only minimal gains when $b = 0.12$ or 0.14.

6. Effects of the length of training on labor market outcomes

In this section, we use the example described in Section 2.2 and show how to vary identifying assumptions to see the efficacy of smoothness conditions. This example is particularly appealing to rely on smoothness conditions since the treatment variable is continuously distributed. In Section 6.1, we first present empirical results for average counterfactual outcomes in comparison with the control group. In this subsection, we fix b at the baseline values that are given in Section 3. Then we move to Section 6.2 in which we focus on average treatment effects within the treatment group and carry out sensitivity analyses with respect to b.

6.1 *Bounds on average counterfactual outcomes in comparison with the control group*

In this subsection, the parameter of interest is the average treatment effect $E[Y_i(t)] - E[Y_i|i \in \text{control group}]$, where t is the length of enrollment to the program. In other words, for each t, we bound the average counterfactual outcomes in comparison with the average observed outcomes for the control group in the NJCS.

Table 1 reports the estimated bounds of the average treatment effects at $t = 4, 16, 36$ weeks, under different assumptions.[14] Columns (1)–(4) show the lower bounds, whereas

[14] All bounds are computed using nonparametric estimates, as explained in footnote 8, while rearranging the nonparametric estimates of $E[Y_i|Z_i]$ when they are nonmonotone (this happens with the estimates for females on some small range of Z_i between 30 and 40 weeks).

Table 1. Bounds under different assumptions.

Training Duration (in Weeks)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Lower Bound				Upper Bound			
	MTR	SMTR	MTR + MTS	SMTR + MTS	MTR	SMTR	MTR + MTS	SMTR + MTS
Outcome: employment status (percentage employed)								
4	−61.55	−0.08	−1.98	−0.05	5.81	2.30	2.37	2.30
16	−34.45	1.34	1.31	1.37	18.07	2.92	3.65	2.92
36	−2.44	2.24	2.25	2.25	30.18	5.41	6.24	5.39
Outcome: weekly earnings in US dollars (including zero earnings)								
4	−180.20	1.97	−0.74	3.79	190.96	17.90	18.12	17.90
16	−97.36	11.49	13.31	13.31	852.17	22.06	23.67	22.06
36	2.68	17.55	17.56	17.56	1586.21	38.80	36.03	35.82

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

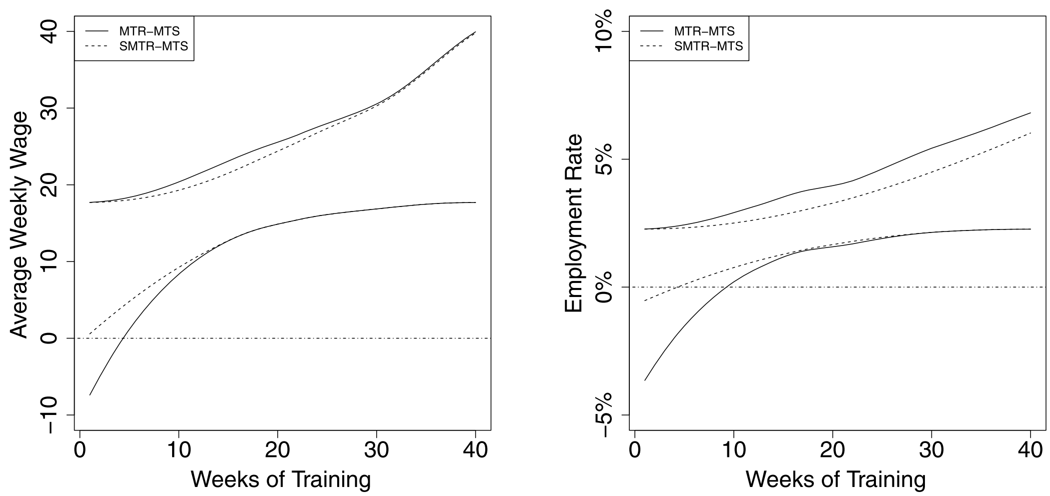

FIGURE 4. Comparison between the MTR-MTS and SMTR-MTS bounds.

columns (5)–(8) the upper bounds. Results for male and female subgroups are given in Table A-3 in Online Appendix F.

We first look at the bounds for the employment rates. On one hand, the MTR bounds are very wide, although the support of the employment status is bounded. On the other hand, the SMTR bounds are much tighter; for example, the SMTR bounds at 16 and 36 weeks are $[1.45\%, 2.84\%]$ and $[2.25\%, 5.04\%]$, respectively. Under the MTR and MTS assumptions together, the bounds are much tighter and comparable to those from the SMTR bounds. Also, it can be seen that for the employment status, the MTS assumption tightens the SMTR bound only marginally.

Moving to weekly earnings, we can see that the MTR bounds are worse for this outcome since the support of earnings is much wider (we took the sample minimum and maximum values for the end points of the support). As in the employment rates, the SMTR or MTR-MTS bounds are much narrower. The tightest bounds are from the SMTR-MTS assumption. Looking at the lower SMTR-MTS bounds, it is clear that the average treatment effect is increasing from 4 dollars to 18 dollars as t gets larger.

To see the overall effect of the smoothness assumption, Figure 4 compares the MTR-MTS bounds with the SMTR-MTS bounds. In Figure 4, the left and right figures correspond to weekly earnings and employment rates, and the horizontal axis is weeks of training (t). The smoothness assumption improves the lower bounds quite substantially when t is relatively small.

6.2 *Bounds on average treatment effects within the treatment group*

We now consider the average treatment effect $\Delta(t_1, t_2) \equiv E[Y_i(t_2)] - E[Y_i(t_1)]$ within the treatment group. Proposition 4.2 gives the upper bound for $\Delta(t_1, t_2)$ under the SMTR-MTS assumptions and allows us to carry out a sensitivity analysis. As an illustration, in this subsection, we focus on $\Delta(16, 36)$ that measures incremental improvements in the

Table 2. Sensitivity analysis on the upper bound of $\Delta(16, 36)$.

Y = Employment Probability		Y = Weekly Earning	
MTR-MTS	4.93	MTR-MTS	22.72
SMTR-MTS		SMTR-MTS	
$b = 0.1$	2.00	$b = 0.5$	10.00
0.15	**2.99**	1	**18.05**
0.2	**3.76**	1.5	**21.35**
0.25	**4.16**	2	**22.58**
0.3	**4.46**	2.5	22.72
0.35	**4.67**	3	22.72
0.4	**4.81**		
Effective region: (0.11, 0.40)		Effective region: (0.7, 2.2)	

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

average outcomes out of 5 more months of training, after enrolling in the program for 4 months already. This upper bound can be useful in a cost-benefit analysis; the upper bound limits the range of the benefits of additional training and can be compared to the costs of 5 extra months of training for the individuals who have been in the program for 4 months.

Table 2 shows the upper bounds of both the MTR-MTR and SMTR-MTS bounds for $\Delta(16, 36)$, while Figure 5 presents the graphical representation of the sensitivity analysis results. We see from the left panel of the table that the increase in employment for all individuals from the SMTR-MTS is at most 4.81% (with $b = 0.4$), one percentage point lower than the MTR-MTS bound. Moreover, for separate analysis of males and females, we can observe that there are strict improvements, by adding smoothness condition (up to $b = 0.4$), and we find similar reductions for the weekly earnings over a wide range of b. See Table A-4 in Online Appendix F for details.

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 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) and also by applying it to the Job Corps Study dataset.

Our identification analysis can be useful for policymakers. Suppose that some average treatment effect estimates are available from previous studies (for instance the National Job Corps Study as in our empirical example, or a lower-cost pilot study using randomized experiments). Then our approach may be suitable when a policymaker tries to predict the average counterfactual outcome of a new policy. Also, our results may be

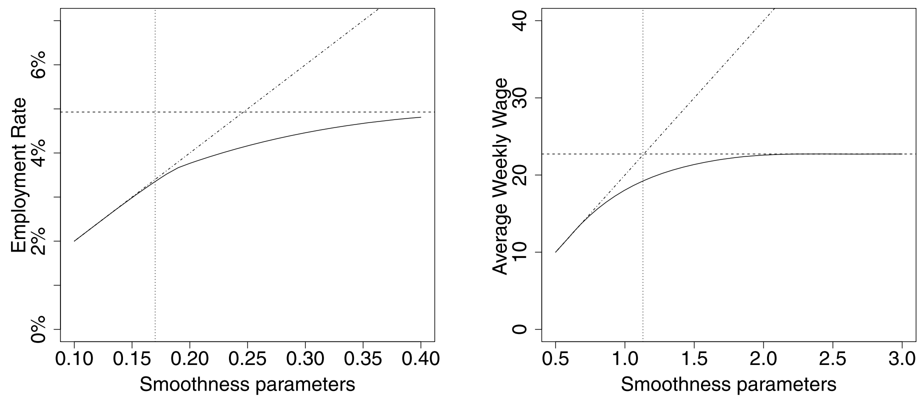

Figure 5. Sensitivity analysis on the upper bound of $\Delta(16, 36)$. *Notes*: The dash horizontal lines show the MTR-MTS bounds and the dash-dot upward sloping lines represent $20b$. The dotted vertical lines correspond to the baseline values of b used in Table 1.

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.

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.

References

Angrist, J. D. and A. B. Krueger (1991), "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106 (4), 979–1014. [625]

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), 351–356. [618]

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2012), "Treatment effect bounds: An application to Swan–Ganz catheterization." *Journal of Econometrics*, 168 (2), 223–243. [618]

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. [618, 629]

Card, D. (2001), "Estimating the return to schooling: Progress on some persistent econometric problems." *Econometrica*, 69 (5), 1127–1160. [624, 625]

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011), "Estimating marginal returns to education." *American Economic Review*, 101 (6), 2754–2781. [625, 626]

Chamberlain, G. and G. W. Imbens (2003), "Nonparametric applications of Bayesian inference." *Journal of Business & Economic Statistics*, 21 (1), 12–18. [625]

Chen, X., E. Tamer, and A. Torgovitsky (2011), "Sensitivity analysis in semiparametric likelihood models." Cowles Foundation Discussion Paper No. 1836. [627]

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2009), "Improving point and interval estimators of monotone functions by rearrangement." *Biometrika*, 86 (3), 559–575. [635]

Chesher, A. (2005), "Nonparametric identification under discrete variation." *Econometrica*, 73 (5), 1525–1550. [618]

Chesher, A. (2010), "Instrumental variable models for discrete outcomes." *Econometrica*, 78 (2), 575–601. [618]

Chiburis, R. (2010), "Semiparametric bounds on treatment effects." *Journal of Econometrics*, 159 (2), 267–275. [618]

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. [618]

Fan, Y., R. Sherman, and M. Shum (2014), "Identifying treatment effects under data combination." *Econometrica*, 82 (2), 811–822. [618]

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. [618]

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. [620, 622]

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. [620]

Hausman, J. A. and W. K. Newey (2016), "Individual heterogeneity and average welfare." *Econometrica*, 84 (3), 1225–1248. [623, 624]

Heckman, J. J., R. J. Lalonde, and J. A. Smith (1999), "The economics and econometrics of active labor market programs." In *Handbook of Labor Economics, Part A, Chapter 31*, Vol. 3 (O. C. Ashenfelter and D. Card, eds.), 1865–2097, Elsevier. [626]

Hirano, K. and G. W. Imbens (2004), "The propensity score with continuous treatments." In *Applied Bayesian Modeling and Causal Inference From Incomplete-Data Perspectives* (A. Gelman and X.-L. Meng, eds.). Wiley InterScience, West Sussex, England. [622]

Horowitz, J. L. and C. F. Manski (1995), "Identification and robustness with contaminated and corrupted data." *Econometrica*, 63 (2), 281–302. [627]

Jun, S. J., J. Pinkse, and H. Xu (2011), "Tighter bounds in triangular systems." *Journal of Econometrics*, 161 (2), 122–128. [618]

Kitagawa, T. (2009), "Identification region of the potential outcome distributions under instrument independence." CeMMAP working paper CWP30/09. [618]

Leamer, E. E. (1985), "Sensitivity analyses would help." *American Economic Review*, 75 (3), 308–313. [627]

Lee, D. S. (2009), "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *Review of Economic Studies*, 76 (3), 1071–1102. [618, 620, 626]

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. [629]

Manski, C. F. (1990), "Nonparametric bounds on treatment effects." *American Economic Review*, 80 (2), 319–323. [618]

Manski, C. F. (1997), "Monotone treatment response." *Econometrica*, 65 (6), 1311–1334. [618, 619, 623, 624, 628]

Manski, C. F. (2003), *Partial Identification of Probability Distributions*. Springer-Verlag, New York. [618]

Manski, C. F. (2007), *Identification for Prediction and Decision*. Harvard University Press, New York. [618]

Manski, C. F. (2013), "Identification of treatment response with social interactions." *Econometrics Journal*, 16 (1), S1–S23. [618]

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. [634]

Manski, C. F. and J. V. Pepper (2000), "Monotone instrumental variables: With an application to the returns to schooling." *Econometrica*, 68 (4), 997–1010. [617, 618, 619, 620, 622, 630, 631, 633, 634, 635, 638]

Manski, C. F. and J. V. Pepper (2009), "More on monotone instrumental variables." *Econometrics Journal*, 12 (s1), S200–S216. [618, 623]

Manski, C. F. and J. V. Pepper (forthcoming), "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions." *Review of Economics and Statistics.* [618]

Okumura, T. and E. Usui (2014), "Concave-monotone treatment response and monotone treatment selection: With an application to the returns to schooling." *Quantitative Economics*, 5 (1), 175–194. [618, 624]

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. [620]

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. [620, 621, 626]

Shaikh, A. M. and E. J. Vytlacil (2011), "Partial identification in triangular systems of equations with binary dependent variables." *Econometrica*, 79 (3), 949–955. [618]

Tamer, E. (2010), "Partial identification in econometrics." *Annual Review of Economics*, 2, 167–195. [618, 627]

Co-editor Rosa L. Matzkin handled this manuscript.

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