Quantitative Economics 5 (2014), 175–194 1759-7331/20140175

Concave-monotone treatment response and monotone treatment selection: With an application to the returns to schooling

Tsunao Okumura
Yokohama National University

Emiko Usui
Nagoya University and IZA

This paper identifies sharp bounds on the mean treatment response and average treatment effect under the assumptions of both the concave-monotone treatment response (concave-MTR) and the monotone treatment selection (MTS). We use our bounds and the U.S. National Longitudinal Survey of Youth 1979 to estimate mean returns to schooling. Our upper-bound estimates are substantially smaller than (i) estimates using only the concave-MTR assumption of Manski (1997), and (ii) estimates using only the MTR and MTS assumptions of Manski and Pepper (2000). Our upper-bound estimates fall in the range of the point estimates given in previous studies that assume linear wage functions.

Keywords. Nonparametric methods, partial identification, sharp bounds, treatment response, returns to schooling.

JEL classification. C14, J24.

1. Introduction

This paper examines the identifying power of the mean treatment response and average treatment effect when the concave-monotone treatment response (concave-MTR) assumption of Manski (1997) is combined with the monotone treatment selection (MTS) assumption of Manski and Pepper (2000). We are motivated by the fact that either assumption, taken alone, produces bounds that are too wide to have sufficient identifying power for many purposes. We then apply our bound analysis to estimate the returns to schooling and thus assess the tightness of our bound estimates.

Tsunao Okumura: okumura@ynu.ac.jp
Emiko Usui: usui@soec.nagoya-u.ac.jp

The authors thank the anonymous referees, Hidehiko Ichimura, Brendan Kline, Sokbae Lee, Adam Rosen, Elie Tamer, Edward Vytlacil, and especially Charles Manski for their helpful comments. The authors also thank participants at the annual meetings of the Econometric Society, the European Association of Labour Economists, the Midwest Econometrics Group, the Symposium on Econometric Theory and Applications (SETA), and the Western Economic Association, and those in seminars at Hitotsubashi University, Kyoto University, Northwestern University, and the University of Tokyo. This research is supported by JSPS Grant 21530167, the Kikawada Foundation, JCER, and TCER.

DOI: 10.3982/QE268

Manski (1997) studied sharp bounds on the mean treatment response and average treatment effect when the response functions are assumed to satisfy either the monotone treatment response (MTR) or concave-MTR. To enhance the identifying power on the bounds, Manski and Pepper (2000) combined the MTS assumption with the MTR assumption. They applied their bounds to estimate the returns to schooling. Their bound estimates are narrower than those of Manski (1997). However, they are still so large that they contain almost all of the point estimates of the returns to schooling in the existing empirical literature.

In this paper, we add the assumption of concavity to the assumptions of MTR and MTS. Concavity is a natural assumption because diminishing marginal returns are commonly assumed in economic analyses. We explore how the inclusion of this assumption tightens the sharp bounds on the mean treatment response and average treatment effect.

Using the 2000 wave of the U.S. National Longitudinal Survey of Youth 1979, we implement our bounds to estimate returns to schooling. Since our bounds use the min and max operations, the estimates of our bounds may have a finite-sample bias. To address this bias, we employ the three methods of bias correction proposed by Kreider and Pepper (2007), Haile and Tamer (2003), and Chernozhukov, Lee, and Rosen (2013). Our sharp upper-bound estimates of the returns to schooling are only between 14 and 28 percent of the estimates produced using only the concave-MTR assumption of Manski (1997), and are between 39 and 80 percent of the estimates produced using the MTR and MTS assumptions of Manski and Pepper (2000). Thus, our upper-bound estimates are substantially smaller than either of the estimates using only the concave-MTR assumption or the estimates using the MTR and MTS assumptions.

Our upper-bound estimates on college education fall in the range of the point estimates on returns to schooling reported in previous studies. Therefore, the concave-MTR and MTS assumptions have substantial identifying power. In previous studies, (i) the log-wage regression function has almost always been assumed to be linear in relation to years of schooling, and (ii) the point estimates would be biased unless the correlation between years of schooling and unobserved abilities were accounted for (i.e., by addressing the effect known as the ability bias). In contrast, the concave-MTR assumption allows for flexible and weakly concave-increasing wage functions, and the MTS assumption accommodates the ability bias in terms of the mean monotonicity of wages and schooling.

The bounds obtained using the concave-MTR and MTS assumptions have many applications in addition to the returns to schooling. A growing stream of literature uses the bounds obtained using the MTR and MTS assumptions of Manski and Pepper (2000) to estimate causal relationships between variables of interest. Some of these relationships can be estimated by using our bounds because they represent concave functions (e.g., diminishing marginal returns). For example, our bounds are applicable to the relationships studied in González (2005), Gerfin and Schellhorn (2006), Blundell, Gosling, Ichimura, and Meghir (2007), Kreider and Pepper (2007, 2008), Gundersen and Kreider (2009), Kreider and Hill (2009), de Haan (2011), Kang (2011), Gundersen, Kreider, and Pepper (2012), Kreider, Pepper, Gundersen, and Jolliffe (2012), and Huang, Maassen van den Brink, and Groot (2012).

In Section 2, we present our study of the sharp bounds on the mean treatment response and the average treatment effects under the concave-MTR and MTS assumptions. Section 3 applies the bounds to the estimation of returns to schooling. Section 4 presents our conclusions.

2. Concave-monotone treatment response and monotone treatment selection

2.1 *Background*

This section sets up basic concepts and notation, and summarizes the bounds derived by Manski (1997) and Manski and Pepper (2000). We employ the same setup as Manski (1997) and Manski and Pepper (2000). There is a probability space (J, Ω, P) of individuals. Each member j of population J has an individual-specific response function $y_j(\cdot): T \to Y$, mapping the mutually exclusive and exhaustive treatments $t \in T$ into outcomes $y_j(t) \in Y$. Each individual j has a realized treatment $z_j \in T$ and a realized outcome $y_j \equiv y_j(z_j)$, both of which are observable. The latent outcomes $y_j(t)$, $t \neq z_j$, are not observable. By combining the distribution of a random sample (z, y) with prior information, we intend to identify the mean treatment response $E[y(t)]$ and the average treatment effect $E[y(t_2)] - E[y(t_1)]$ for $t_1 < t_2$.[1]

Manski (1997) stated the MTR assumption as $y_j(t_1) \leq y_j(t_2)$ for each $j \in J$ and all $(t_1, t_2) \in T^2$, where T is an ordered set and $t_1 \leq t_2$. Under the MTR assumption, he showed the sharp bounds on $E[y(t)]$ to be

$$\begin{aligned} &\sum_{s\leq t} E[y|z=s]P(z=s) + y_0 P(z>t) \\ &\quad \leq E\big[y(t)\big] \leq \sum_{s\geq t} E[y|z=s]P(z=s) + y_1 P(z<t), \end{aligned} \tag{1}$$

where $[y_0, y_1]$ is the range of Y.

When $y_j(\cdot)$ satisfies the assumptions of concavity and MTR (i.e., the concave-MTR assumption), and when $T = [0, \delta]$ for some $\delta \in (0, \infty]$ and $Y = [0, \infty]$, Manski (1997) showed the sharp bounds on $E[y(t)]$ to be[2]

$$\begin{aligned} &\sum_{s\leq t} E[y|z=s]P(z=s) + E\left[\frac{y}{z}t \middle| z>t\right]P(z>t) \\ &\quad \leq E\big[y(t)\big] \leq \sum_{s\geq t} E[y|z=s]P(z=s) + E\left[\frac{y}{z}t \middle| z<t\right]P(z<t). \end{aligned} \tag{2}$$

[1] When there are covariates x, all results regarding the identification and estimation of $E[y(t)]$ and $E[y(t_2)] - E[y(t_1)]$ in this paper are applied to the identification and estimation of $E[y(t)|x]$ and $E[y(t_2)|x] - E[y(t_1)|x]$.

[2] When $y_j(\cdot)$ is concave-MTR but either T or Y does not have any known finite lower bounds, Manski (1997) showed that the sharp bounds on $E[y(t)]$ are represented by Equation (1).

Manski and Pepper (2000) introduced the assumption of MTS as $E[y(t)|z = t_1] \leq E[y(t)|z = t_2]$ for each $t \in T$ and all $(t_1, t_2) \in T^2$ such that $t_1 \leq t_2$. When schooling is a treatment, the MTS assumption asserts that people who select more schooling have weakly higher mean wage functions than those who select less schooling. Under the assumptions of both MTR and MTS, they showed the sharp bounds on $E[y(t)]$ to be

$$\begin{aligned} &\sum_{s \leq t} E[y|z = s]P(z = s) + E[y|z = t]P(z > t) \\ &\quad \leq E\bigl[y(t)\bigr] \leq \sum_{s \geq t} E[y|z = s]P(z = s) + E[y|z = t]P(z < t). \end{aligned} \tag{3}$$

2.2 *Sharp bounds on the mean treatment response*

This section demonstrates the sharp bounds on the mean treatment response ($E[y(t)]$) under both the concave-MTR and the MTS assumptions. We first illustrate the basic idea with an example, shown in Figure 1. We consider bounding the conditional mean of latent outcome $E[y(t)|z = s]$, which is point A when $t < s$. The MTS assumption implies that, for $u < s$, $E[y|z = u]$ (point F) $\leq E[y(u)|z = s]$ (point G). Furthermore, $E[y(\tau)|z = s]$ is concave-MTR in $\tau \in T$ and is $E[y|z = s]$ (point C) when $\tau = s$. Thus, when $u \leq t < s$, the value (point E) of the function that describes the straight line traversing $(u, E[y|z = u])$ (point F) and $(s, E[y|z = s])$ (point C), evaluated at t, is a lower bound on $E[y(t)|z = s]$ (point A). Given $(s, E[y|z = s])$ (point C), these lines are drawn for all realized points of

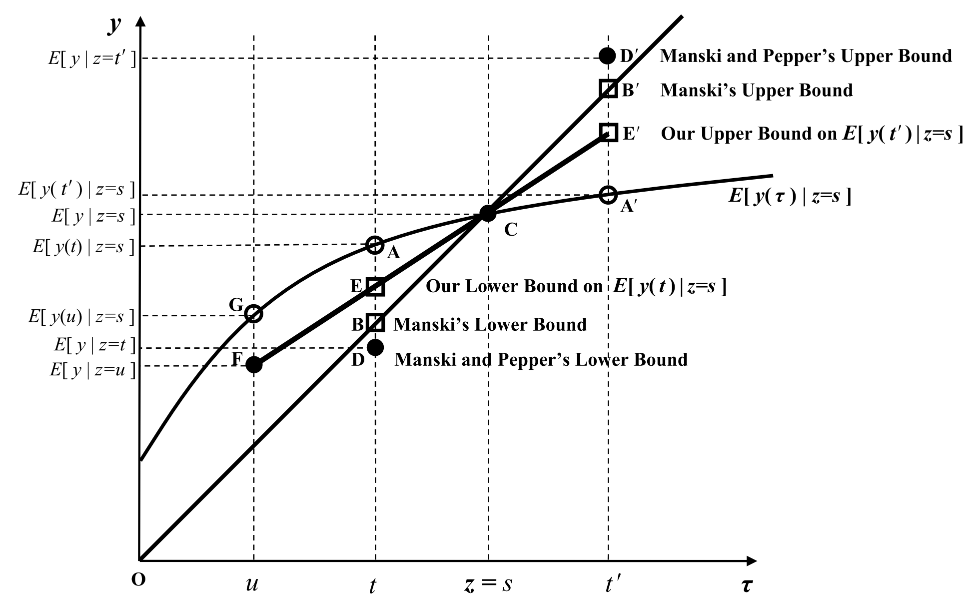

FIGURE 1. Sharp bounds on the mean treatment response. Filled circles indicate the realized treatments and the conditional means of realized outcomes. Open circles indicate the latent treatments and the conditional means of latent outcomes. Open squares indicate the bounds on the conditional-mean treatment responses.

$(u, E[y|z = u])$ for $u \leq t$ and for the origin (point O) when $Y = [0, \infty]$. The values of these functions evaluated at t (points B, D, and E) are all lower bounds on $E[y(t)|z = s]$ for $t < s$ (point A). These include the lower bound of Manski (1997) (i.e., point B on the line joining points C and O) and that of Manski and Pepper (2000) ($E[y|z = t]$, i.e., point D on the line joining points C and D). Our lower bound on $E[y(t)|z = s]$ for $t < s$ is the greatest among these lower bounds (i.e., point E is the greatest among points B, D, and E).

Similarly, when $t \geq s$ (denoted by t' in Figure 1), $E[y(t')|z = s]$ is point A'. For any $u < s \leq t'$, the value (point E') of the function that describes the straight line traversing $(u, E[y|z = u])$ (point F) and $(s, E[y|z = s])$ (point C), evaluated at t', is an upper bound on $E[y(t')|z = s]$ (point A'). Furthermore, the value (point B') of the function that describes the straight line traversing the origin (point O) and point C, evaluated at t', is an upper bound on $E[y(t')|z = s]$ (point A'), as is the value $E[y|z = t']$ (point D')—the latter because of the MTS assumption. Our upper bound (point E') on $E[y(t')|z = s]$ for $t' \geq s$ is the smallest among these upper bounds (points B', D', and E'). Note that point B' corresponds to the upper bound in Manski (1997), and point D' corresponds to the upper bound in Manski and Pepper (2000). Using the law of iterated expectations, our bounds on $E[y(t)]$ are narrower than or equal to those in Manski (1997) and Manski and Pepper (2000).

Proposition 1. *Let T be ordered and let Y be a closed subset of the extended real line. Assume that $y_j(\cdot)$, $j \in J$, satisfies the concave-MTR and the MTS assumptions. We then obtain the following three results.*

(a) *For $(t, s, \eta_1, \eta_2) \in T^4$,*

$$\begin{aligned}\sum_{s \leq t} E[y|z = s]P(z = s) + \sum_{s > t} \mathrm{LB}(s, t)P(z = s) \\ \leq E\bigl[y(t)\bigr] \leq \sum_{s \geq t} E[y|z = s]P(z = s) + \sum_{s < t} \mathrm{UB}(s, t)P(z = s),\end{aligned} \tag{4}$$

where

$$\mathrm{LB}(s, t) = \max_{\{(\eta_1, \eta_2)|\eta_1 \leq t < \eta_2 \leq s\}} \frac{\eta_2 - t}{\eta_2 - \eta_1} E[y|z = \eta_1] + \frac{t - \eta_1}{\eta_2 - \eta_1} E[y|z = \eta_2], \tag{5}$$

$$\mathrm{UB}(s, t) = \min_{\{(\eta_1, \eta_2)|s \leq \eta_2 \leq t \wedge \eta_1 < \eta_2\}} \frac{\eta_2 - t}{\eta_2 - \eta_1} E[y|z = \eta_1] + \frac{t - \eta_1}{\eta_2 - \eta_1} E[y|z = \eta_2]. \tag{6}$$

These bounds are sharp.

(b) *Furthermore,* (i) *let $T = [0, \delta]$ for some $\delta \in (0, \infty]$,* (ii) *let $Y = [0, \infty]$, and* (iii) *let $E[y|z = 0] = 0$ whenever $P(z = 0) = 0$. Then Equations* (4), (5), *and* (6) *hold. These bounds are sharp.*

(c) *In either case* (a) *or* (b), *the bounds represented by Equations* (4), (5), *and* (6) *are narrower than or equal to those using only the concave-MTR assumption of Manski* (1997), *as well as those using only the MTR and the MTS assumptions of Manski and Pepper* (2000).

We prove Proposition 1 in Appendix A (all appendices are available in a supplementary file on the journal website, http://qeconomics.org/supp/268/supplement.pdf). The bound $\mathrm{LB}(s,t)$, represented by Equation (5), divides the line segment joining $E[y|z=\eta_1]$ and $E[y|z=\eta_2]$ internally, whereas $\mathrm{UB}(s,t)$, represented by Equation (6), divides this line segment externally.

When the assumption of concavity is added to the MTR and MTS assumptions, the width of the bounds on the mean treatment response is narrowed by the quantity

$$\sum_{s>t}\bigl\{\mathrm{LB}(s,t)-E[y|z=t]\bigr\}P(z=s)+\sum_{s<t}\bigl\{E[y|z=t]-\mathrm{UB}(s,t)\bigr\}P(z=s). \tag{7}$$

The first term expresses the increase in the lower bound, whereas the second term expresses the decrease in the upper bound.

2.3 *Sharp bounds on the average treatment effects*

This section demonstrates the sharp bounds on the average treatment effects ($E[y(t_2)]-E[y(t_1)]$ for $t_1<t_2$). Figure 2 provides an example of the bounds given in Proposition 2. When $t_2=t'>s$, let $\mathrm{UB}(s,t_2)$ be the sharp upper bound on $E[y(t_2)|z=s]$ in Equation (6) (point E'') and let $\mathrm{AT}_1(t_1,s,t_2)$ be the value of the function that describes the line joining points O, F, E, C, I, and E'', evaluated at t_1. When $t_2=t\le s$, $E[y|z=s]$ (point H) is

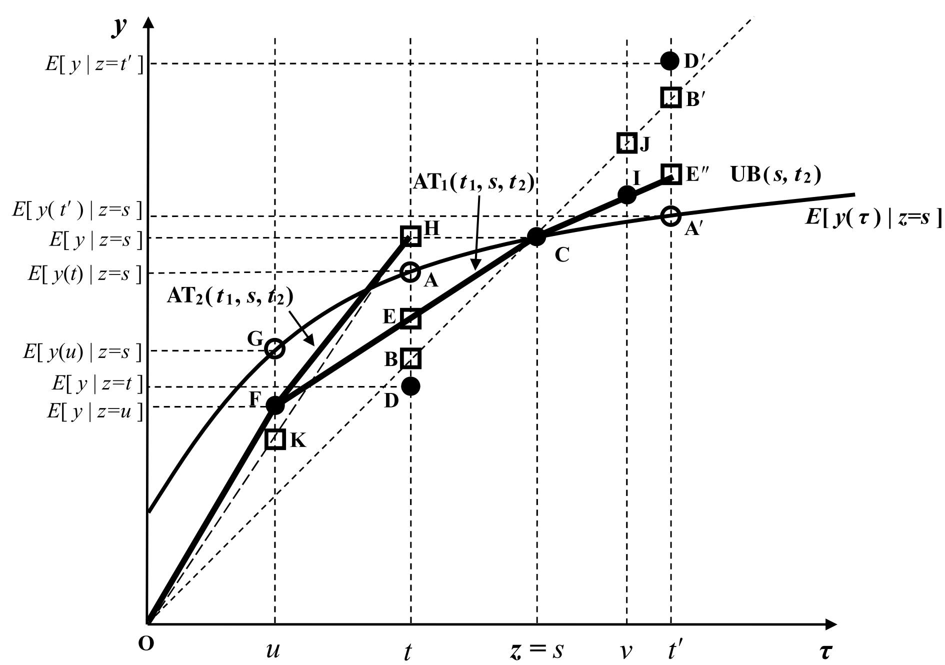

FIGURE 2. Sharp bounds on the average treatment effects. Filled circles indicate the realized treatments and the conditional means of realized outcomes. Open circles indicate the latent treatments and the conditional means of latent outcomes. Open squares indicate the bounds on the average treatment effects.

the sharp upper bound on $E[y(t_2)|z=s]$ in Equation (4). Let $\mathrm{AT}_2(t_1,s,t_2)$ be the value of the function that describes the line joining points O, F, and H, evaluated at t_1. Then our sharp upper bound on $E[y(t_2)|z=s]-E[y(t_1)|z=s]$ is $\mathrm{UB}(s,t_2)-\mathrm{AT}_1(t_1,s,t_2)$ for $t_2=t'>s$, and it is $E[y|z=s]-\mathrm{AT}_2(t_1,s,t_2)$ for $t_2=t\le s$. Using the law of iterated expectations, we obtain the sharp upper bound on the average treatment effect.

Proposition 2. *Let T be ordered and let Y be a closed subset of the extended real line. Assume that $y_j(\cdot)$, $j\in J$, satisfies the concave-MTR and MTS assumptions. We then obtain the following three results*:

(a) *For $(t_1,t_2,s,\eta_1,\eta_2)\in T^5$, where $t_1<t_2$,*

$$\begin{aligned}0&\le E\big[y(t_2)\big]-E\big[y(t_1)\big]\\&\le\sum_{s<t_2}\big[\mathrm{UB}(s,t_2)-\mathrm{AT}_1(t_1,s,t_2)\big]P(z=s)\\&\quad+\sum_{s\ge t_2}\big\{E[y|z=s]-\mathrm{AT}_2(t_1,s,t_2)\big\}P(z=s),\end{aligned}\tag{8}$$

where, for $s<t_2$,

$$\mathrm{AT}_1(t_1,s,t_2)=\begin{cases}\dfrac{t_1-s}{t_2-s}\mathrm{UB}(s,t_2)+\dfrac{t_2-t_1}{t_2-s}E[y|z=s], & \text{if } s\le t_1<t_2,\\ \mathrm{LB}(s,t_1), & \text{if } t_1<s<t_2,\end{cases}\tag{9}$$

and, for $t_2\le s$,

$$\mathrm{AT}_2(t_1,s,t_2)=\max_{\{(\eta_1,\eta_2)|\eta_1\le t_1<\eta_2\le t_2\}}\frac{\eta_2-t_1}{\eta_2-\eta_1}E[y|z=\eta_1]+\frac{t_1-\eta_1}{\eta_2-\eta_1}\mu(\eta_2),\tag{10}$$

where $\mu(\eta_2)=E[y|z=s]$ if $\eta_2=t_2$ and $E[y|z=\eta_2]$ if $\eta_2<t_2$. These bounds are sharp.

(b) *Furthermore,* (i) *let $T=[0,\delta]$ for some $\delta\in(0,\infty]$,* (ii) *let $Y=[0,\infty]$, and* (iii) *let $E[y|z=0]=0$ whenever $P(z=0)=0$. Then Equations* (8), (9), *and* (10) *hold. These bounds are sharp.*

(c) *In either case* (a) *or* (b), *the bounds represented by Equations* (8), (9), *and* (10) *are narrower than or equal to those using only the concave-MTR assumption of Manski* (1997), *as well as those using only the MTR and MTS assumptions of Manski and Pepper* (2000).

In Appendix B, we prove Proposition 2.

Proposition 2(a) and (b) show that our sharp upper bound on $E[y(t_2)]-E[y(t_1)]$ is attained when $\mathrm{AT}_1(\tau,s,t_2)$ is the mean response function of individuals whose realized treatment (s) is smaller than t_2 (i.e., $E[y(\tau)|z=s]$ for $s<t_2$) and when $\mathrm{AT}_2(\tau,s,t_2)$ is the mean response function of individuals whose realized treatment (s) is not smaller than t_2 (i.e., $E[y(\tau)|z=s]$ for $s\ge t_2$). The function $\mathrm{AT}_1(\tau,s,t_2)$ is a function in τ that describes the upper envelope of the points $(u,E[y|z=u])$ for all $u\le s$ and the point $(t_2,\mathrm{UB}(s,t_2))$ (i.e., a function in τ that describes the upper boundary of the convex hull for a set formed

by these points). The function $\mathrm{AT}_2(\tau, s, t_2)$ is a function in τ that describes the upper envelope of the points $(u, E[y|z=u])$ for all $u \leq t_2$ and the point $(t_2, E[y|z=s])$. Therefore, the curves of the functions $\mathrm{AT}_1(\tau, s, t_2)$ and $\mathrm{AT}_2(\tau, s, t_2)$ constitute the upper envelope (or upper boundary of the convex hull) of the conditional means of the realized outcomes.

The fact that this convex hull is the smallest convex set that contains the conditional means of realized outcomes provides intuitive explanations of why the functions $\mathrm{AT}_1(\tau, s, t_2)$ and $\mathrm{AT}_2(\tau, s, t_2)$ satisfy the concave-MTR and MTS assumptions. First, because a convex set has a concave upper boundary, these functions are concave-MTR. Second, because the convex hull for a set formed by $(u, E[y|z=u])$ for all $u \leq s$ is included in the convex hull for a set formed by $(u, E[y|z=u])$ for all $u \leq s'$ and $s < s'$, and because $\mathrm{UB}(s, t_2)$ and $E[y|z=s]$ both weakly increase in s, it follows that $\mathrm{AT}_k(\tau, s, t_2) \leq \mathrm{AT}_k(\tau, s', t_2)$ for $k=1,2$ and for $s < s'$. That is, the functions $\mathrm{AT}_1(\tau, s, t_2)$ and $\mathrm{AT}_2(\tau, s, t_2)$ satisfy the MTS assumption. Figure 2 illustrates these intuitive explanations. Specifically, when $t_2 = t'$, the function $\mathrm{AT}_1(\tau, s, t_2)$ in Proposition 2(b) is a function that describes the line joining points O, F, E, C, I, and E''; that is, it is a function that describes the upper boundary of the convex hull for the set of points O, F, D, C, I, and E''. When $t_2 = t$, the function $\mathrm{AT}_2(\tau, s, t_2)$ is a function that describes the line joining points O, F, and H; that is, it is a function that describes the upper boundary of the convex hull for the set of points O, F, H, and D. These functions are both concave-MTR and MTS.

Proposition 2(c) shows that our sharp upper bound on the average treatment effect $(E[y(t_2)] - E[y(t_1)])$ is smaller than or equal to that of Manski (1997) and that of Manski and Pepper (2000). Specifically, when the assumption of concavity is added to the MTR and MTS assumptions of Manski and Pepper (2000), the upper bound on the average treatment effect is reduced by the quantity

$$\begin{aligned}
&\sum_{s<t_2}\bigl\{E[y|z=t_2] - \mathrm{UB}(s,t_2)\bigr\}P(z=s) + \sum_{s\leq t_1}\bigl\{\mathrm{AT}_1(t_1,s,t_2) - E[y|z=s]\bigr\}P(z=s) \\
&\quad + \sum_{t_1<s<t_2}\bigl\{\mathrm{AT}_1(t_1,s,t_2) - E[y|z=t_1]\bigr\}P(z=s) \\
&\quad + \sum_{s\geq t_2}\bigl\{\mathrm{AT}_2(t_1,s,t_2) - E[y|z=t_1]\bigr\}P(z=s).
\end{aligned} \tag{11}$$

However, when $T = [0, \delta]$ for some $\delta \in (0, \infty]$ and $Y = [0, \infty]$, and when the MTS assumption is added to the concave-MTR assumption of Manski (1997), the upper bound on the average treatment effect is reduced by the quantity

$$\begin{aligned}
&\sum_{s<t_2}\left\{E\left[\frac{y}{s}t_2\middle|z=s\right] - \mathrm{UB}(s,t_2)\right\}P(z=s) \\
&\quad + \sum_{s<t_2}\left\{\mathrm{AT}_1(t_1,s,t_2) - E\left[\frac{y}{s}t_1\middle|z=s\right]\right\}P(z=s) \\
&\quad + \sum_{s\geq t_2}\left\{\mathrm{AT}_2(t_1,s,t_2) - E\left[\frac{y}{t_2}t_1\middle|z=s\right]\right\}P(z=s).
\end{aligned} \tag{12}$$

To graphically explain Proposition 2(c) and Equations (11) and (12), we use Figure 2, which depicts the differences among the three sets of upper bounds on the average treatment effect that are provided in Manski (1997), Manski and Pepper (2000), and this paper. We define X_y as the y-coordinate of point X. The position relationships of t_1, t_2, and s $(= z)$ are classified into three cases. In the first case, in which $t_1 < s < t_2$ (represented by the case in which $t_2 = t'$ and $t_1 = t$ in Figure 2), our upper bound is $E''_y - E_y$, Manski and Pepper's upper bound is $D'_y - D_y$, and Manski's upper bound is $B'_y - B_y$. Therefore, the difference between our upper bound and that of Manski and Pepper is the sum of $D'_y - E''_y$ (the first term of Equation (11)) and $E_y - D_y$ (the third term of Equation (11)), whereas the difference between our upper bound and that of Manski is the sum of $B'_y - E''_y$ (the first term of Equation (12)) and $E_y - B_y$ (the second term of Equation (12)). In the second case, in which $s \le t_1 < t_2$ (represented by the case in which $t_2 = t'$ and $t_1 = v$ in Figure 2), our upper bound is $E''_y - I_y$, Manski and Pepper's upper bound is $D'_y - C_y$, and Manski's upper bound is $B'_y - J_y$. Therefore, the difference between our upper bound and that of Manski and Pepper is the sum of $D'_y - E''_y$ (the first term of Equation (11)) and $I_y - C_y$ (the second term of Equation (11)), whereas the difference between our upper bound and that of Manski is the sum of $B'_y - E''_y$ (the first term of Equation (12)) and $I_y - J_y$ (the second term of Equation (12)). In the third case, in which $t_1 < t_2 \le s$ (represented by the case in which $t_2 = t$ and $t_1 = u$ in Figure 2), our upper bound is $H_y - F_y$, Manski and Pepper's upper bound is $H_y - F_y$, and Manski's upper bound is $H_y - K_y$. Therefore, the difference between our upper bound and that of Manski and Pepper is zero (the fourth term in Equation (11)), whereas the difference between our upper bound and that of Manski is $F_y - K_y$ (the third term of Equation (12)).[3]

Manski (1995) and Manski and Pepper (2009) studied the identifying power of the homogeneous linear response (HLR) assumption and the exogenous treatment selection (ETS) assumption, which are imposed on ordinary least squares (OLS) regressions. The HLR assumption asserts that $y_j(t) = \beta t + \nu_j$, where ν_j is an unobserved covariate and β is a slope parameter that takes the same value for all j. The ETS assumption asserts that for $(t, t_1, t_2) \in T^3$, $E[y(t)|z = t_1] = E[y(t)|z = t_2]$. Under the HLR and ETS assumptions, Manski (1995) and Manski and Pepper (2009) showed that $E[y(t_2)] - E[y(t_1)] = E[y|z = t_2] - E[y|z = t_1] = \beta(t_2 - t_1)$. This quantity is not smaller than our upper bound for $t_1 \le s \le t_2$ (i.e., $\mathrm{UB}(s, t_2) - \mathrm{AT}_1(t_1, s, t_2)$), because $\mathrm{UB}(s, t_2) \le E[y|z = t_2]$ and $E[y|z = t_1] \le \mathrm{AT}_1(t_1, s, t_2)$. Figure 2 shows that when $t_2 = t'$ and $t_1 = t$, $E[y|z = t_2] - E[y|z = t_1]$ is equal to $D'_y - D_y$, whereas our upper bound is equal to $E''_y - E_y$. As will be shown in Section 3, our upper bound estimates on the returns to schooling are thus likely to be smaller than the estimates obtained through OLS regression.[4]

[3]For the bounds of this paper and Manski (1997), in the first case (in which $t_1 < s < t_2$), the upper bound on $E[y(t_2)|z = s] - E[y(t_1)|z = s]$ is equal to the difference between the upper bound on $E[y(t_2)|z = s]$ and the lower bound on $E[y(t_1)|z = s]$. However, both in the second case (in which $s \le t_1 < t_2$) and in the third case (in which $t_1 < t_2 \le s$), the upper bound on $E[y(t_2)|z = s] - E[y(t_1)|z = s]$ is less than or equal to the difference between the upper bound on $E[y(t_2)|z = s]$ and the lower bound on $E[y(t_1)|z = s]$ because the concavity of $y_j(t)$ is assumed. For the bounds of Manski and Pepper (2000), in all cases, the upper bound on $E[y(t_2)|z = s] - E[y(t_1)|z = s]$ is equal to the difference between the upper bound on $E[y(t_2)|z = s]$ and the lower bound on $E[y(t_1)|z = s]$.

[4]For $t_2 < s$, it is indeterminate which is larger: our upper bound (i.e., $E[y|z = s] - \mathrm{AT}_2(t_1, s, t_2)$) or the average treatment effect under the HLR and ETS assumptions (i.e., $E[y|z = t_2] - E[y|z = t_1]$) because

When the ETS assumption does not appear to hold in empirical applications, the instrumental variable (IV) assumption is often used to identify the mean treatment response and average treatment effect. Manski and Pepper (2000, 2009) studied the identifying power of the IV assumption, which asserts the mean independence of outcomes and instrumental variables. They also introduced the monotone instrumental variable (MIV) assumption, which weakens the IV assumption to the mean monotonicity of outcomes and instrumental variables, and they studied the identifying power of this assumption. The MTS assumption is the MIV assumption when the instrumental variable is the realized treatment. The bounds on the mean treatment response represented by Equations (4), (5), and (6), and those on the average treatment effect represented by Equations (8), (9), and (10) can be further narrowed when the concave-MTR and MTS assumptions are combined with either the IV or the MIV assumption. In Appendix C, we provide the bounds on the mean treatment response and average treatment effect by combining either the IV or MIV assumption with the concave-MTR and MTS assumptions.

2.4 *Sharp bounds on D-outcomes*

In this section, using Proposition 1, we derive the sharp bounds on the parameter that respects stochastic dominance.[5] We impose two assumptions on the distribution of outcomes $F_{y(t)}(r) := P(y(t) \leq r)$ for $r \in Y$.

First, we make the following assumption: for $(t_1, t_2) \in T^2$, where $t_1 \leq t_2$, and $r \in Y$,

$$F_{y(t)}(r|z = t_1) \geq F_{y(t)}(r|z = t_2). \tag{13}$$

The assumption represented by Equation (13) is from Blundell et al. (2007). It asserts that a higher value of the realized treatment $z = t_2$ leads to a distribution of outcomes that first-order stochastically dominates the distribution of outcomes with a lower value of the realized treatment $z = t_1$. Second, we make the assumption expressed by the following two equations: for $(t_1, t_2, t) \in T^3$, where $t_1 \leq t_2$, $r \in Y$, and $\alpha \in [0, 1]$,

$$F_{y(t_1)}(r|z = t) \geq F_{y(t_2)}(r|z = t), \tag{14}$$

$$F_{y(\alpha t_1 + (1-\alpha)t_2)}(r|z = t) \leq \alpha F_{y(t_1)}(r|z = t) + (1 - \alpha) F_{y(t_2)}(r|z = t). \tag{15}$$

The assumption represented by Equations (14) and (15) is from Boes (2010); it asserts that the conditional distribution of outcomes is convex and decreasing in relation to treatments. For example, when schooling is a treatment, Equation (13) asserts that even

$E[y|z = s] \geq E[y|z = t_2]$ and $\mathrm{AT}_2(t_1, s, t_2) \geq E[y|z = t_1]$. Figure 2 shows that when $t_2 = t$ and $t_1 = u$, our upper bound is $H_y - F_y$, whereas $E[y|z = t_2] - E[y|z = t_1]$ is $D_y - F_y$; thus, the former is larger than the latter.

[5] Manski (1997) defined the parameter $D[F_v(r)]$ that respects stochastic dominance as the parameter that satisfies $D[F_{v_1}(r)] \geq D[F_{v_2}(r)]$ whenever $F_{v_1}(r) \leq F_{v_2}(r)$, where $F_v(r) := P(v \leq r)$ for a real random variable v and a real number r. This parameter includes not only means but also quantiles and upper-tail probabilities. A growing strand of literature on quantile regression shows the advantages of quantiles over the mean, such as their robustness to outliers and their capacity to measure the tendency and dispersion of the treatment effects. Therefore, bounding the parameter that respects stochastic dominance makes possible useful analysis for many purposes.

if people who actually select a high level of schooling had selected a low level of schooling, these people would be more likely to earn high wages than people who actually do select a low level of schooling. Equations (14) and (15) assert that the probability that people are in a high-income group is concave-increasing in relation to their schooling levels.

Equation (13) is equivalent to the condition that the indicator function $1(y(t) > r)$ satisfies the MTS assumption. Equations (14) and (15), taken together, are equivalent to the condition that $E[1(y(t) > r)|z]$ satisfies the concave-MTR assumption. Therefore, under the assumptions of Equations (13), (14), and (15), we obtain the sharp bounds on the parameter $D[F_{y(t)}(r)]$ that respects stochastic dominance.

Proposition 3. *Let T be ordered. Assume that the conditional probability $F_{y(t)}(r|z)$ satisfies Equations* (13), (14), *and* (15). *We then obtain, for $(t, s, \eta_1, \eta_2) \in T^4$ and $r \in Y$,*

$$\begin{aligned} &D\Bigl[\sum_{s\le t} F_y(r|z=s)P(z=s) + \sum_{s>t} \mathrm{LBP}(s,t)P(z=s)\Bigr] \\ &\quad\le D\bigl[F_{y(t)}(r)\bigr] \le D\Bigl[\sum_{s\ge t} F_y(r|z=s)P(z=s) + \sum_{s<t} \mathrm{UBP}(s,t)P(z=s)\Bigr], \end{aligned} \tag{16}$$

where

$$\mathrm{LBP}(s,t) = \min_{\{(\eta_1,\eta_2)|\eta_1\le t<\eta_2\le s\}} \frac{\eta_2 - t}{\eta_2 - \eta_1} F_y(r|z=\eta_1) + \frac{t-\eta_1}{\eta_2-\eta_1} F_y(r|z=\eta_2), \tag{17}$$

$$\begin{aligned} \mathrm{UBP}(s,t) = \max\Bigl\{&0, \\ &\max_{\{(\eta_1,\eta_2)|s\le\eta_2\le t\wedge\eta_1<\eta_2\}} \frac{\eta_2 - t}{\eta_2 - \eta_1} F_y(r|z=\eta_1) + \frac{t-\eta_1}{\eta_2-\eta_1} F_y(r|z=\eta_2)\Bigr\}. \end{aligned} \tag{18}$$

These bounds are sharp.

In Appendix D, we prove Proposition 3.

3. Estimation of returns to schooling

3.1 *Data*

We use the 2000 wave of the U.S. National Longitudinal Survey of Youth 1979 (NLSY79), which is representative of the U.S. noninstitutionalized civilian population who were between the ages of 14 and 22 in 1979. Like Manski and Pepper (2000), who used the 1994 wave of the NLSY79, we use a random sample of white men who reported that they were full-time, year-round workers and not self-employed. Their hourly rate of pay and realized years of schooling were observed. Thus, the realized outcome y_j is the logarithm of the observed hourly wage.[6] The sample size is 1280. The percentage of the respondents

[6]We excluded four individuals whose hourly wages are less than one dollar. Thus, the realized outcomes y_j are positive.

Table 1. Mean log hourly wages and distribution of schooling.

z	Years of Schooling	$E[y\|z]$	$P(z)$	Sample Size
1	<12	2.516	0.078	100
2	12	2.727	0.423	542
3	13–15	2.980	0.184	235
4	16	3.251	0.180	231
5	>16	3.336	0.134	172
Total			1	1280

Note: The function $E[y|z]$ is estimated by local constant kernel regression using a quartic kernel and a rule-of-thumb bandwidth presented in Fan and Gijbels (1996).

with 12 years of schooling is 42.3 and the percentage of those with 16 years of schooling is 18.1, but the percentages of those with years of schooling other than 12 and 16 are small.

In such finite samples, the estimates of the bounds on the mean treatment response and the average treatment effect may be biased. To alleviate the finite-sample bias problem, we estimate these bounds for five broad schooling groups: (i) less than 12 years of schooling (high-school dropouts, the realized treatment $z = 1$), (ii) 12 years of schooling (high-school graduates, $z = 2$), (iii) 13–15 years of schooling (some college, $z = 3$), (iv) 16 years of schooling (college graduates, $z = 4$), and (v) 16–20 years of schooling (more than 4 years of college, $z = 5$). Table 1 shows the local constant kernel estimates of $E[y|z]$, the empirical probability $P(z)$, and the sample size for schooling group z.[7] The sample size of each schooling group is not smaller than 100. Estimates of $E[y|z]$ increase with z and thus are consistent with the MTR and MTS assumptions.[8]

3.2 *Estimation results*

In this section, we estimate (i) the bounds on the mean treatment response $E[y(t)]$ represented by Equations (4), (5), and (6), and (ii) the bounds on the average treatment effect $E[y(t_2)] - E[y(t_1)]$ represented by Equations (8), (9), and (10). Because $y_j(t)$ is the logarithm of the hourly rate of pay that a person j would obtain if he were to have t schooling level, the average treatment effect, $E[y(t_2)] - E[y(t_1)]$ for $t_1 < t_2$, is the expected return to completing a t_2 schooling level relative to a t_1 schooling level.

As noted by Manski and Pepper (2000, 2009), Haile and Tamer (2003), Kreider and Pepper (2007), and Chernozhukov, Lee, and Rosen (2013), because the minima and maxima of the functions of the estimates of $E[y|z]$ in Equations (5), (6), (9), and (10) have a finite-sample bias, estimates of the bounds on $E[y(t)]$ represented by Equations (4), (5), and (6) and those on $E[y(t_2)] - E[y(t_1)]$ represented by Equations (8), (9), and (10) are tighter than the population bounds. Therefore, we estimate the bounds using the following three methods of bias correction: (i) a method proposed by Kreider and Pepper

[7] The local constant kernel estimates of $E[y|z]$ use the quartic kernel and the rule-of-thumb bandwidth presented in Fan and Gijbels (1996).

[8] If $E[y|z]$ is not weakly increasing in z, then the MTR and MTS assumptions are rejected.

TABLE 2. Lower and upper bounds on $E[y(t)]$: concave-MTR and MTS assumptions.

		Lower Bounds on $E[y(t)]$					
t	Schooling	Estimate (1)	KP Estimate (2)	HT Estimate (3)	HT 0.05 CI (4)	CLR Estimate (5)	CLR 0.05 CI (6)
1	<12	2.516	2.514	2.516	2.371	2.516	2.446
2	12	2.725	2.721	2.719	2.673	2.716	2.687
3	13–15	2.845	2.842	2.841	2.799	2.838	2.816
4	16	2.922	2.922	2.921	2.869	2.919	2.895
5	>16	2.933	2.934	2.933	2.884	2.933	2.911
		Upper Bounds on $E[y(t)]$					
t	Schooling	Estimate (1)	KP Estimate (2)	HT Estimate (3)	HT 0.95 CI (4)	CLR Estimate (5)	CLR 0.95 CI (6)
1	<12	2.933	2.934	2.933	2.981	2.934	2.956
2	12	2.950	2.951	2.950	2.997	2.950	2.973
3	13–15	3.056	3.066	3.062	3.124	3.071	3.113
4	16	3.205	3.229	3.216	3.295	3.253	3.303
5	>16	3.336	3.375	3.338	3.429	3.384	3.440

Note: See text for a description of the estimator. *KP estimate* and *HT estimate* are abbreviations for the bias-corrected estimates in Kreider and Pepper (2007) and Haile and Tamer (2003), respectively, while *CLR estimate* is an abbreviation for the median unbiased estimate in Chernozhukov, Lee, and Rosen (2013). The CI denotes confidence interval.

(2007) and Manski and Pepper (2009) (abbreviated here as the KP method), (ii) a method proposed by Haile and Tamer (2003) (the HT method), and (iii) a method proposed by Chernozhukov, Lee, and Rosen (2013) (the CLR method). The KP and HT estimators are computationally simpler but have less formal justification of their asymptotic properties than do the CLR estimators. In Appendix E, we provide an implementation guide for applying these three methods to our bounds.

The upper and lower panels of Table 2 report the estimates of the lower and upper bounds, respectively, on the mean treatment response $E[y(t)]$. Column 1 of Table 2 reports the bound estimates that are not bias-corrected. Columns 2, 3, and 5 report the bound estimates that are bias-corrected by the KP, HT, and CLR methods, respectively. Columns 4 and 6 report the confidence intervals produced by the HT and CLR methods, respectively.[9] The bias-corrected bound estimates take similar values among the KP, HT, and CLR estimates; the CLR bound estimates are the widest. The CLR confidence intervals are narrower than the HT confidence intervals except for the 0.95 confidence intervals for $t = 4$ and 5. For comparison, Table 3 reports the bound estimates on $E[y(t)]$ using only the concave-MTR assumption of Manski (1997), and the bound estimates using only the MTR and MTS assumptions of Manski and Pepper (2000). The estimates of our bounds, which are bias-corrected by the KP, HT, and CLR methods, are narrower

[9]Table 2 reports the estimates and confidence intervals of the bounds obtained in Proposition 1(a). These bounds provide the same estimation results as those obtained in Proposition 1(b), except for the HT estimates of the upper bounds, which have negligibly small differences.

Table 3. Lower and upper bounds on $E[y(t)]$.

		Manski's (1997) Bounds				Manski and Pepper's (2000) Bounds			
		Lower Bounds on $E[y(t)]$		Upper Bounds on $E[y(t)]$		Lower Bounds on $E[y(t)]$		Upper Bounds on $E[y(t)]$	
t	Schooling	Estimate (1)	0.05 CI (2)	Estimate (3)	0.95 CI (4)	Estimate (5)	0.05 CI (6)	Estimate (7)	0.95 CI (8)
1	<12	1.193	1.146	2.933	2.981	2.516	2.371	2.933	2.981
2	12	2.189	2.129	3.130	3.201	2.711	2.647	2.950	2.997
3	13–15	2.607	2.561	3.904	4.022	2.837	2.778	3.077	3.150
4	16	2.844	2.798	4.860	5.040	2.922	2.872	3.263	3.354
5	>16	2.933	2.884	5.963	6.200	2.933	2.884	3.336	3.474

Table 4. Upper bounds on returns to schooling: concave-MTR and MTS assumptions.

		Upper Bounds on $E[y(t_2)] - E[y(t_1)]$					
t_1	t_2	Estimate (1)	KP Estimate (2)	HT Estimate (3)	HT 0.95 CI (4)	CLR Estimate (5)	CLR 0.95 CI (6)
1	2	0.434	0.437	0.434	0.588	0.434	0.508
2	3	0.270	0.281	0.277	0.349	0.292	0.335
3	4	0.230	0.245	0.240	0.280	0.261	0.288
4	5	0.161	0.175	0.163	0.218	0.194	0.228
2	4	0.459	0.486	0.475	0.564	0.514	0.573
Average effect		0.115	0.122	0.119	0.141	0.129	0.143

Note: See the text for a description of the estimator. *KP estimate* and *HT estimate* are abbreviations for the bias-corrected estimates in Kreider and Pepper (2007) and Haile and Tamer (2003), respectively, while *CLR estimate* is an abbreviation for the median unbiased estimate in Chernozhukov, Lee, and Rosen (2013). The CI denotes confidence interval.

than those of Manski (1997); they are also narrower than or equal to those of Manski and Pepper (2000), except for $t = 5$.

Table 4 reports the estimates of the upper bounds on the average treatment effect, $E[y(t_2)] - E[y(t_1)]$, namely, the estimates of the upper bounds on the returns to schooling.[10] The layout of the table is the same as for Table 2. The bias-corrected upper-bound estimates take similar values among the KP, HT, and CLR estimates; the CLR estimates are the largest and the HT estimates are the smallest except for $(t_1, t_2) = (1, 2)$. The CLR confidence intervals are smaller for $(t_1, t_2) = (1, 2)$ and $(2, 3)$, and are larger for $(t_1, t_2) = (3, 4)$ and $(4, 5)$, compared to the HT confidence intervals. For comparison, Table 5 reports (i) the estimates of the upper bounds on the returns to schooling using

[10]Table 4 reports the estimates and confidence intervals of the upper bounds obtained in Proposition 2(a). These bounds provide the same estimation results as those obtained in Proposition 2(b), except for the HT estimates of the upper bounds, which have negligibly small differences. All sharp lower bounds on the returns to schooling of this paper, Manski (1997), and Manski and Pepper (2000) are always zero (see Proposition 2).

TABLE 5. Upper bounds on returns to schooling.

		Upper Bounds on $E[y(t_2)] - E[y(t_1)]$			
		Manski's (1997) Bounds		Manski and Pepper's (2000) Bounds	
t_1	t_2	Estimate (1)	0.95 CI (2)	Estimate (3)	0.95 CI (4)
1	2	1.565	1.601	0.434	0.588
2	3	1.301	1.341	0.366	0.461
3	4	1.215	1.260	0.426	0.542
4	5	1.193	1.240	0.414	0.563
2	4	2.430	2.520	0.552	0.665
Average effect		0.608	0.630	0.138	0.166

the concave-MTR assumption of Manski (1997), and (ii) those using the MTR and MTS assumptions of Manski and Pepper (2000). Our upper-bound estimates are substantially smaller than those of Manski (1997) and Manski and Pepper (2000). Specifically, our upper-bound estimates, which are bias-corrected by the KP, HT, and CLR methods, are only between 14 and 28 percent as large as those of Manski (1997), and they are only between 39 and 80 percent as large as those of Manski and Pepper (2000) except for the returns for $(t_1, t_2) = (1, 2)$.[11]

The last row of Table 4 shows the estimates of our upper bounds on the average yearly returns from completing 4 years of college relative to completing high school (i.e., $\{E[y(4)] - E[y(2)]\}/4$). The bias-uncorrected estimate is 0.115, the KP estimate is 0.122, the HT estimate is 0.119, and the CLR estimate is 0.129. In Section 2.3, we demonstrated that under the HLR and ETS assumptions, the average yearly returns from completing 4 years of college relative to completing high school is $\{E[y|z = 4] - E[y|z = 2]\}/4$. The point estimate of this quantity is 0.131, which is obtained using Table 1. The OLS estimate using our sample with 12 and 16 years of schooling is 0.131 (with a standard error of 0.009). Card (1999, 2001) showed that point estimates on the returns to schooling in previous studies that use linear regressions and U.S. data are between 0.052 and 0.132. Therefore, our upper-bound estimates of the average yearly returns from completing 4 years of college relative to completing high school are informative in the sense that they are lower than the point estimates that use our sample and also lower than some of the point estimates reported in previous studies. In comparison, the last row of Table 5 shows that the upper-bound estimate of the average yearly returns from completing 4 years of college relative to completing high school, computed using the concave-MTR assumption of Manski (1997), is 0.608 and the estimate using the MTR and MTS assumptions of Manski and Pepper (2000) is 0.138. These upper-bound estimates are, therefore, larger than the point estimates that use our sample and all of the point estimates reported in previous studies. We thus conclude that it is the combination of the concave-MTR and the MTS assumptions that has substantial identifying power for the

[11]The bounds of Proposition 2(a) and Manski and Pepper (2000) are identical when $(t_1, t_2) = (1, 2)$ (see Equation (11)).

returns to college-level schooling. This empirical result supports the prediction made in Section 2.3: our upper bound on the average treatment effect may be smaller than the average treatment effect point-identified by the HLR and ETS assumptions.

In the remainder of this section, we discuss the economic interpretation and implications of the assumptions that have been used to estimate the returns to schooling by the linear regressions, Manski and Pepper (2000), Manski (1997), and this paper. Most of the previous empirical studies on returns to schooling are based on the HLR assumption; the returns to schooling are, therefore, estimated by linearly regressing log wages on years of schooling. The HLR assumption asserts that the individuals' log-wage functions are all linear in years of schooling and have the same slope parameter across the population. Among these empirical studies, those that use the OLS technique are also based on the ETS assumption. Schooling may, however, be correlated with the error terms because of unobserved abilities (the ability bias) and because of measurement error in relation to schooling (the attenuation bias); for these reasons, the validity of the ETS assumption is often questioned. To correct for these biases, institutional, personal, and/or family attributes are used as the instrumental variables (IVs) for schooling.[12]

The MTR and MTS assumptions, which Manski and Pepper (2000) imposed to estimate the returns to schooling, are distinct from the HLR and ETS assumptions. The MTR assumption simply asserts that log wage increases weakly in relation to schooling and is, therefore, consistent with conventional theories of human capital accumulation. The MTS assumption asserts the mean monotonicity of wages and schooling. Therefore, the MTS assumption accommodates the ability bias in light of the large stream of literature on returns to schooling that assumes that people with higher ability select more schooling and have higher mean wage functions than those with lower ability. However, the MTS assumption is restrictive in light of the Willis and Rosen (1979) model and the Roy model, which represent another large stream of literature on returns to schooling. Specifically, these latter two models assume that individuals differ in their skill endowment and select schooling levels at which their own skills are most rewarded (i.e., for $(t, s) \in T^2$, $E[y(t)|z = s] \leq E[y|z = s]$). Therefore, in the Willis–Rosen and Roy models, it is possible that if individuals who select a high level of schooling had selected a low level of schooling, they would be poorer than those who actually do select a low level of schooling (i.e., for $t_1 < t_2$, $E[y(t_1)|z = t_2] < E[y|z = t_1]$). The MTS assumption rules out this possibility.[13] Moreover, the MTS assumption cannot address the attenuation bias caused by measurement errors in the years of schooling.

[12]In the case in which the HLR assumption does not hold, Lang (1993), Imbens and Angrist (1994), Card (1995, 1999, 2001), and Carneiro, Heckman, and Vytlacil (2011) argued that the estimated returns to schooling using the IV technique identify a local average treatment effect (LATE) and thus do not express the average treatment effect of the returns to schooling. Specifically, when costs of schooling are used as the instrumental variables, these estimates could be higher than the population average return because they identify the return for credit-constrained individuals, who are induced to go to college by changes in the instrumental variables (the effect known as the discount rate bias).

[13]By using the local instrumental-variable technique, Carneiro, Heckman, and Vytlacil (2011) and Heckman, Eisenhauer, and Vytlacil (2011) estimated the returns to schooling in the Willis–Rosen and Roy models.

In this paper, we add the assumption of concavity, which is used by Manski (1997), to the MTR and MTS assumptions. The concave-MTR assumption asserts that the log-wage function is weakly concave-increasing in relation to schooling. The conventional theories of human capital accumulation assume that production functions for human capital have diminishing marginal returns to schooling (e.g., Card (1999, 2001)). Thus, the concave-MTR assumption is still consistent with conventional theories of human capital accumulation. However, if there are wage premiums for obtaining higher credentials such as a high-school diploma or college degree (known as the sheepskin effect), this assumption may not be justified.

4. Conclusion

We identify sharp bounds on the mean treatment response and average treatment effect under both the concave-MTR and the MTS assumptions. We then estimate the bounds on the returns to schooling by utilizing our bounds and the NLSY79 data. Estimates obtained using our bounds are substantially tighter than either estimates using only the concave-MTR assumption of Manski (1997) or estimates using only the MTR and MTS assumptions of Manski and Pepper (2000). Moreover, our upper-bound estimates fall in the range of point estimates reported in the previous literature.

Further research could pursue several interesting directions. First, by using the growing literature on inference for partially identified models, one could provide tests for the concave-MTR and MTS assumptions, such as a test for the null hypothesis that the bounds obtained using these assumptions do not cross.[14] Second, using Proposition 3, one could derive the sharp bounds on the quantile treatment response and quantile treatment effect under the concave-MTR and MTS assumptions. These bounds could be used to tighten (i) the bounds obtained by Giustinelli (2011) on the quantile treatment effect under the MTR and MTS assumptions, and (ii) the bounds, obtained using the quantile treatment response in Okumura (2011), on shift variables within a nonparametric simultaneous equations model. Third, using the bounds obtained in Appendix C, one could estimate the upper bounds on returns to schooling under a combination of the IV or MIV assumptions with the concave-MTR and MTS assumptions; one could then compare these bound estimates with the point estimates obtained using linear IV models.

Fourth, because conventional production theory is consistent with the concave-MTR and MTS assumptions, our bounds could be applied not only to the literature introduced in Section 1, but also to estimating the production functions of firms. Production theory often asserts that the output of a product increases with input. This assertion has dual interpretations. The first interpretation is that a production function weakly increases and marginal product weakly decreases with input, which follows from the concave-MTR assumption for this example. The second interpretation is that firms that

[14]The literature on inference for partially identified models includes Manski and Tamer (2002), Imbens and Manski (2004), Blundell et al. (2007), Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Rosen (2008), Stoye (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano and Shaikh (2010), and Andrews and Shi (2013).

select greater levels of output have weakly greater average production functions than those that select smaller levels of output, which follows from the MTS assumption for this example. Therefore, under the concave-MTR and MTS assumptions about the production process, the bound approach in this paper could be applied to reveal the average production function (i.e., $E[y(t)]$) and the average increase in a firm's production as input increases (i.e., $E[y(t_2)] - E[y(t_1)]$).

References

Andrews, D. W. K. and X. Shi (2013), "Inference based on conditional moment inequalities." *Econometrica*, 81 (2), 609–666. [191]

Andrews, D. W. K. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78 (1), 119–157. [191]

Beresteanu, A. and F. Molinari (2008), "Asymptotic properties for a class of partially identified models." *Econometrica*, 76 (4), 763–814. [191]

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007), "Changes in the distribution of male and female wages accounting for employment composition using bounds." *Econometrica*, 75 (2), 323–363. [176, 184, 191]

Boes, S. (2010), "Convex treatment response and treatment selection." Working paper, Socioeconomic Institute, University of Zurich. [184]

Bugni, F. A. (2010), "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set." *Econometrica*, 78 (2), 735–753. [191]

Canay, I. A. (2010), "EL inference for partially identified models: Large deviations optimality and bootstrap validity." *Journal of Econometrics*, 156 (2), 408–425. [191]

Card, D. (1995), "Earnings, schooling, and ability revisited." In *Research in Labor Economics*, Vol. 14 (S. W. Polachek, ed.), 23–48, JAI Press, Greenwich, Connecticut. [190]

Card, D. (1999), "The causal effect of education on earnings." In *Handbook of Labor Economics*, Vol. 3A (O. Ashenfelter and D. Card, eds.), 1801–1863, Chapter 30, North-Holland, Amsterdam. [189, 190, 191]

Card, D. (2001), "Estimating the return to schooling: Progress on some persistent econometric problems." *Econometrica*, 69 (5), 1127–1160. [189, 190, 191]

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011), "Estimating marginal returns to education." *American Economic Review*, 101 (6), 2754–2781. [190]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75 (5), 1243–1284. [191]

Chernozhukov, V., S. Lee, and A. M. Rosen (2013), "Intersection bounds: Estimation and inference." *Econometrica*, 81 (2), 667–737. [176, 186, 187, 188]

de Haan, M. (2011), "The effect of parents' schooling on child's schooling: A nonparametric bounds analysis." *Journal of Labor Economics*, 29 (4), 859–892. [176]

Fan, J. and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London. [186]

Gerfin, M. and M. Schellhorn (2006), "Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits—Swiss evidence." *Health Economics*, 15 (9), 1011–1020. [176]

Giustinelli, P. (2011), "Non-parametric bounds on quantiles under monotonicity assumptions: With an application to the Italian education returns." *Journal of Applied Econometrics*, 26 (5), 783–824. [191]

González, L. (2005), "Nonparametric bounds on the returns to language skills." *Journal of Applied Econometrics*, 20 (6), 771–795. [176]

Gundersen, C. and B. Kreider (2009), "Bounding the effects of food insecurity on children's health outcomes." *Journal of Health Economics*, 28 (5), 971–983. [176]

Gundersen, C., B. Kreider, and J. Pepper (2012), "The impact of the National School Lunch Program on child health: A nonparametric bounds analysis." *Journal of Econometrics*, 166 (1), 79–91. [177]

Haile, P. A. and E. Tamer (2003), "Inference with an incomplete model of English auctions." *Journal of Political Economy*, 111 (1), 1–51. [176, 186, 187, 188]

Heckman, J. J., P. Eisenhauer, and E. Vytlacil (2011), "Generalized Roy model and cost–benefit analysis of social programs." Working paper, Department of Economics, University of Chicago. [190]

Huang, J., H. Maassen van den Brink, and W. Groot (2012), "Does education promote social capital? Evidence from IV analysis and nonparametric-bound analysis." *Empirical Economics*, 42 (3), 1011–1034. [177]

Imbens, G. W. and J. D. Angrist (1994), "Identification and estimation of local average treatment effects." *Econometrica*, 62 (2), 467–475. [190]

Imbens, G. W. and C. F. Manski (2004), "Confidence intervals for partially identified parameters." *Econometrica*, 72 (6), 1845–1857. [191]

Kang, C. (2011), "Family size and educational investments in children: Evidence from private tutoring expenditures in South Korea." *Oxford Bulletin of Economics and Statistics*, 73 (1), 59–78. [176]

Kreider, B. and S. C. Hill (2009), "Partially identifying treatment effects with an application to covering the uninsured." *Journal of Human Resources*, 44 (2), 409–449. [176]

Kreider, B. and J. V. Pepper (2007), "Disability and employment: Reevaluating the evidence in light of reporting errors." *Journal of the American Statistical Association*, 102 (478), 432–441. [176, 186, 187, 188]

Kreider, B. and J. Pepper (2008), "Inferring disability status from corrupt data." *Journal of Applied Econometrics*, 23 (3), 329–349. [176]

Kreider, B., J. V. Pepper, C. Gundersen, and D. Jolliffe (2012), "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported." *Journal of the American Statistical Association*, 107 (499), 958–975. [177]

Lang, K. (1993), "Ability bias, discount rate bias, and the return to education." Working paper, Department of Economics, Boston University. [190]

Manski, C. F. (1995), *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge, Massachusetts. [183]

Manski, C. F. (1997), "Monotone treatment response." *Econometrica*, 65 (6), 1311–1334. [175, 176, 177, 179, 181, 182, 183, 184, 187, 188, 189, 190, 191]

Manski, C. F. and J. V. Pepper (2000), "Monotone instrumental variables: With an application to the returns to schooling." *Econometrica*, 68 (4), 997–1010. [175, 176, 177, 178, 179, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191]

Manski, C. F. and J. V. Pepper (2009), "More on monotone instrumental variables." *Econometrics Journal*, 12 (S1), S200–S216. [183, 184, 186, 187]

Manski, C. F. and E. Tamer (2002), "Inference on regressions with interval data on a regressor or outcome." *Econometrica*, 70 (2), 519–546. [191]

Okumura, T. (2011), "Nonparametric estimation of labor supply and demand factors." *Journal of Business & Economic Statistics*, 29 (1), 174–185. [191]

Romano, J. P. and A. M. Shaikh (2010), "Inference for the identified set in partially identified econometric models." *Econometrica*, 78 (1), 169–211. [191]

Rosen, A. M. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities." *Journal of Econometrics*, 146 (1), 107–117. [191]

Stoye, J. (2009), "More on confidence intervals for partially identified parameters." *Econometrica*, 77 (4), 1299–1315. [191]

Willis, R. J. and S. Rosen (1979), "Education and self-selection." *Journal of Political Economy*, 87 (5), S7–S36. [190]

Submitted April, 2012. Final version accepted June, 2013.