

**MONOTONE INSTRUMENTAL VARIABLES:
with an Application to the Returns to Schooling**

Charles F. Manski
Department of Economics and Institute for Policy Research
Northwestern University
Evanston, IL 60208

and

John V. Pepper
Department of Economics
University of Virginia
Charlottesville, VA 22903

Revised: April 1999

This research was supported in part by NSF Grant SBR-9722846. We have benefitted from the opportunity to present this work at the 1998 Winter Meeting of the Econometric Society and in seminars at Duke University, the Federal Reserve Board, Northwestern University, UCLA, the University of Bristol, the University of Mississippi, the University of Southampton, the University of Texas at Austin, and the University of Virginia. We have also benefitted from the comments of Jinyong Hahn, Mark Manuszak, and two anonymous reviewers.

MONOTONE INSTRUMENTAL VARIABLES: WITH AN APPLICATION TO THE RETURNS TO SCHOOLING

Charles F. Manski and John V. Pepper

Abstract

Econometric analyses of treatment response commonly use *instrumental variable* (IV) assumptions to identify treatment effects. Yet the credibility of IV assumptions is often a matter of considerable disagreement. There is therefore good reason to consider weaker but more credible assumptions. To this end, we introduce *monotone instrumental variable* (MIV) assumptions and the important special case of *monotone treatment selection* (MTS). We study the identifying power of MIV assumptions alone and combined with the assumption of *monotone treatment response* (MTR). We present an empirical application using the MTS and MTR assumptions to place upper bounds on the returns to schooling.

1. Introduction

For fifty years econometric analyses of treatment response have made extensive use of *instrumental variable* (IV) assumptions holding that mean response is constant across specified sub-populations of a population of interest.¹ Yet the credibility of mean independence conditions and other IV assumptions has often been a matter of considerable disagreement, with much debate about whether some covariate is or is not a "valid instrument" in an application of interest. There is therefore good reason to consider weaker but more credible assumptions. To this end, we introduce *monotone instrumental variable* (MIV) assumptions holding that mean response varies weakly monotonically across specified sub-populations. We study the identifying power of these MIV assumptions and give an empirical application. The findings reported here add to the literature developing nonparametric bounds on treatment effects.²

This paper uses the same formal setup as Manski (1997). There is a probability space (J, \mathbf{S}, P) of individuals. Each member j of population J has

¹ Whereas a variable v was originally called an *instrumental variable* if v has zero covariance with a residual ϵ , the modern usage of the term has broadened to embrace assumptions that specified functions of v and ϵ are orthogonal. Hence it is now necessary to specify the type of IV assumption one has in mind. Mean independence, quantile independence, and statistical independence assumptions (or the orthogonality conditions that these assumptions yield) have all been prominent in the literature. See Manski (1988) pp. 25-26 and Section 6.1 for discussion of the history and exposition of the variety of modern IV assumptions.

² Several contributors to this literature have examined the identifying power of IV assumptions in the absence of parametric restrictions on the form of response functions. Manski (1990, 1994) showed that mean independence assumptions imply sharp bounds on mean outcomes and average treatment effects. Robins (1989) and Balke and Pearl (1997) have considered the statistical independence assumption that holds in classical randomized experiments, where response functions are statistically independent of assigned treatments. Hotz, Mullins, and Sanders (1997) have studied *contaminated instrument* assumptions. These suppose that a mean-independence assumption holds in a population of interest, but the observed population is a probability mixture of the population of interest and one in which the assumption does not hold.

observable covariates $x_j \in X$ and a response function $y_j(\cdot): T \rightarrow Y$ mapping the mutually exclusive and exhaustive *treatments* $t \in T$ into *outcomes* $y_j(t) \in Y$. Person j has a realized treatment $z_j \in T$ and a realized outcome $y_j / y_j(z_j)$, both of which are observable. The latent outcomes $y_j(t)$, $t \neq z_j$ are not observable. An empirical researcher learns the distribution $P(x, z, y)$ of covariates, realized treatments, and realized outcomes by observing a random sample of the population. The researcher's problem is to combine this empirical evidence with assumptions in order to learn about the distribution $P[y(\cdot)]$ of response functions, or perhaps the conditional distributions $P[y(\cdot)|x]$.

With this background, we may formally define the MIV assumptions to be studied here. Let $x = (w, v)$ and $X = W \times V$. Each value of (w, v) defines an observable sub-population of persons. The familiar mean-independence form of IV assumption is that, for each $t \in T$ and each value of w , the mean value of $y(t)$ is the same in all of the sub-populations $(w, v = u)$, $u \in V$. Thus,

IV Assumption: Covariate v is an instrumental variable in the sense of mean-independence if, for each $t \in T$, each value of w , and all $(u, u') \in (V \times V)$,

$$(1) \quad E[y(t)|w, v = u'] = E[y(t)|w, v = u].$$

MIV assumptions replace the equality in (1) by an inequality, yielding a mean-monotonicity condition. Thus

MIV Assumption: Let V be an ordered set. Covariate v is a monotone instrumental variable in the sense of mean-monotonicity if, for each $t \in T$, each value of w , and all $(u_1, u_2) \in (V \times V)$ such that $u_2 \leq u_1$,

$$(2) \quad E[y(t)^*w, v = u_2] \geq E[y(t)^*w, v = u_1].$$

To illustrate how IV and MIV assumptions differ, let v measure a person's ability and consider how wage functions vary with realized schooling. To use v as an IV is to assume that persons with different measured ability have the same mean wage functions. To use v as an MIV is to assume that persons with higher measured ability have weakly higher mean wage functions than do those with lower measured ability. If the type of ability measured by v is valued in the labor market, it is reasonable to assume that v is an MIV but not that it is an IV.

Section 2 studies the identifying power of MIV assumptions alone, not combined with other assumptions. We report sharp bounds on the conditional mean responses $E[y(t)^*w, v = u]$, $u \in V$ and the marginal mean $E[y(t)^*w]$. These bounds are informative if the outcome space Y is bounded and if the no-assumptions bounds of Manski (1989) are not monotone increasing in u . The MIV bounds take a particularly simple form in the case of *monotone treatment selection* (MTS), where the realized treatment z is itself an MIV.

Section 3 combines an MIV assumption with the *monotone treatment response* (MTR) assumption of Manski (1997). This is

MTR Assumption: Let T be an ordered set. For each $j \in J$,

$$(3) \quad t_2 \geq t_1 \implies Y_j(t_2) \geq Y_j(t_1).$$

The MIV and MTR assumptions make distinct contributions to identification. When imposed together, the two assumptions can have substantial identifying power. Combining the MTR and MTS assumptions yields a particularly interesting finding.

Whereas the MIV-MTR bounds are generally informative only when Y is a bounded outcome space, the MTS-MTR bounds are informative even if Y is unbounded.

Section 4 presents an empirical application using the MTS and MTR assumptions to draw conclusions about the returns to schooling. We analyze wage and schooling data in the National Longitudinal Survey of Youth (NLSY). In this application, the MTS assumption asserts that persons who realize higher years of schooling have weakly higher mean wage functions than do those who realize lower levels of schooling. The MTR assumption asserts that, *ceteris paribus*, wage rises as a function of conjectured years of schooling. We find that combining these assumptions yields informative upper bounds on the returns to schooling. In particular, we find that the average year-by-year return to college completion relative to high school completion is smaller than some of the point estimates reported recently by labor economists.

To simplify the exposition in Sections 2 and 3, we henceforth leave implicit the conditioning on w maintained in the definitions of MIVs. To keep the focus on identification, we treat identified quantities as known. In the empirical analysis of Section 4, we explicitly condition on specified covariates w and we discuss statistical considerations. Section 5 briefly calls attention to some variations on the MIV theme.

2. Identification Using an MIV Assumption Alone

We examine here the identifying power of an MIV assumption alone. We focus on the problem of inference on the conditional means $E[y(t)^*v = u]$, $u \in V$ and the marginal mean $E[y(t)]$. The findings are sharp bounds that weaken in obvious ways

the sharp IV bounds of Manski (1990, 1994). Section 2.1 gives the general results and Section 2.2 applies them to *monotone treatment selection*, which weakens the familiar assumption of exogenous treatment selection from an IV to an MIV.

2.1. General Case

The starting point for determination of the identifying power of MIV assumptions is the no-assumptions bound on $E[y(t)^*v]$ of Manski (1989). Let $[K_0, K_1]$ denote the range of Y . Let $u \in V$. Use the law of iterated expectations and the fact that $E[y(t)^*v = u, z = t] = E(y^*v = u, z = t)$ to write

$$(4) \quad E[y(t)^*v = u] = E(y^*v = u, z = t) \cdot P(z = t^*v = u) \\ + E[y(t)^*v = u, z \dots t] \cdot P(z \dots t^*v = u).$$

The sampling process identifies each of the quantities on the right side except for the censored mean $E[y(t)^*v = u, z \dots t]$, which may take any value in the interval $[K_0, K_1]$. This implies the sharp bound

$$(5) \quad E(y^*v = u, z = t) \cdot P(z = t^*v = u) + K_0 \cdot P(z \dots t^*v = u) \\ \# E[y(t)^*v = u] \# \\ E(y^*v = u, z = t) \cdot P(z = t^*v = u) + K_1 \cdot P(z \dots t^*v = u).$$

An MIV assumption implies the inequality restriction

$$(6) \quad u_1 \# u \# u_2 \quad Y \quad E[y(t)^*v = u_1] \# E[y(t)^*v = u] \# E[y(t)^*v = u_2].$$

Hence $E[y(t)^*v = u]$ is no smaller than the no-assumption lower bound on $E[y(t)^*v = u_1]$ and no larger than the no-assumption upper bound on $E[y(t)^*v = u_2]$. This holds for all $u_1 < u$ and all $u_2 \geq u$. There are no other restrictions on $E[y(t)^*v = u]$. Thus we have ^{3 4}

Proposition 1: Let the MIV Assumption (2) hold. Then for each $u \in V$,

$$(7) \quad \sup_{u_1 \# u} [E(y^*v = u_1, z = t) \cdot P(z = t^*v = u_1) + K_0 \cdot P(z \dots t^*v = u_1)] \\ \# E[y(t)^*v = u] \# \\ \inf_{u_2 \geq u} [E(y^*v = u_2, z = t) \cdot P(z = t^*v = u_2) + K_1 \cdot P(z \dots t^*v = u_2)].$$

In the absence of other information, this bound is sharp. \in

The MIV bound on the marginal mean $E[y(t)]$ is easily obtained from Proposition 1. Assume for simplicity that the set V is finite.⁵ Then we may use the law of iterated expectations to write

³ Whereas the MIV assumption (2) implies Proposition 1, the IV assumption (1) implies that $E[y(t)^*v = u]$ is constant across $u \in V$. Hence the common value of $E[y(t)^*v = u]$, $u \in V$ lies in the intersection of the bounds (5) across all the elements of V (Manski, 1990).

⁴ Proposition 1 also applies to *semi-monotone instrumental variable* assumptions, in which the set V is only semi-ordered rather than ordered. The inequality (2) holds as stated, it being understood that there may exist some pairs of covariate values that are not ordered.

⁵ If V is not finite, the result below continues to hold with the summation replaced by a Lebesgue integral, subject to measurability considerations.

$$(8) \quad E[y(t)] = \int_0^V P(v = u) \cdot E[y(t)^*v = u] \, du$$

Equation (7) shows that the MIV lower and upper bounds on $E[y(t)^*v = u]$ are weakly increasing in u . Hence the sharp joint lower (upper) bound on $\{E[y(t)^*v = u], u \in [0, V]\}$ is obtained by setting each of the quantities $E[y(t)^*v = u], u \in [0, V]$ at its lower (upper) bound as given in (7). Inserting these lower and upper bounds into (8) yields

Proposition 1, Corollary 1: Let the MIV Assumption (2) hold. Then

$$(9) \quad \int_0^V P(v = u) \left\{ \sup_{u_1 \neq u} [E(y^*v = u_1, z = t) \cdot P(z = t^*v = u_1) + K_0 \cdot P(z \dots t^*v = u_1)] \right\} \\ \# \quad E[y(t)] \quad \# \\ \int_0^V P(v = u) \left\{ \inf_{u_2 \leq u} [E(y^*v = u_2, z = t) \cdot P(z = t^*v = u_2) + K_1 \cdot P(z \dots t^*v = u_2)] \right\} \, du$$

In the absence of other information, this bound is sharp. \in

The MIV bounds in Proposition 1 and Corollary 1 necessarily are subsets of the corresponding no-assumptions bounds and supersets of the corresponding IV bounds. The MIV and no-assumptions bounds coincide if the no-assumptions lower and upper bounds on $E[y(t)^*v = u]$ weakly increase with u ; in such cases the MIV assumption has no identifying power. The MIV and IV bounds coincide if the no-assumptions lower and upper bounds on $E[y(t)^*v = u]$ weakly decrease with u ; in such cases, the MIV and IV assumptions have the same identifying power.

2.2. Monotone Treatment Selection

Certainly the most commonly applied IV assumption is *exogenous treatment selection* (ETS). Here the instrumental variable v is the realized treatment z . So the IV assumption (1) becomes

ETS Assumption: For each $t \in T$,

$$(10) \quad E[y(t)^*z = u'] = E[y(t)^*z = u], \quad \forall u \in T, u' \in T.$$

As is well known, this assumption implies that $E[y(t)] = E(y^*z = t)$.

Weakening equation (10) to an inequality yields the special MIV assumption that we call *monotone treatment selection* (MTS):

MTS Assumption: Let T be an ordered set. For each $t \in T$,

$$(11) \quad u_2 \leq u_1 \implies E[y(t)^*z = u_2] \leq E[y(t)^*z = u_1].$$

Applying Proposition 1 and Corollary 1 yield these sharp MTS bounds:

Proposition 1, Corollary 2: Let the MTS Assumption (11) hold. Then

$$(12) \quad \begin{aligned} u < t & \implies K_0 \leq E[y(t)^*z = u] \leq E(y^*z = t) \\ u = t & \implies E[y(t)^*z = u] = E(y^*z = t) \\ u > t & \implies E(y^*z = t) \leq E[y(t)^*z = u] \leq K_1. \end{aligned}$$

and

$$(13) \quad K_0 \cdot P(z < t) + E(y^* | z = t) \cdot P(z = t) \leq E[y(t)] \\ \leq K_1 \cdot P(z > t) + E(y^* | z = t) \cdot P(z = t).$$

In the absence of other information, these bounds are sharp. \square

To illustrate the ETS and MTS assumptions, consider again the variation of wages with schooling. The ETS assumption asserts that persons who select different levels of schooling have the same mean wage functions. The MTS assumption asserts that persons who select higher levels of schooling have weakly higher mean wage functions than do those who select lower levels of schooling. Many economic models of schooling choice and wage determination predict that persons with higher ability have higher mean wage functions and choose higher levels of schooling than do persons with lower ability. The MTS assumption is consistent with these models but the ETS assumption is not.

3. Identification Combining MIV and MTR Assumptions

We examine here the identifying power of an MIV assumption of the form (2) combined with a monotone treatment response assumption of the form (3). Section 3.1 motivates MTR assumptions and explains how they differ from MIV assumptions. Section 3.2 gives the general findings. Section 3.3 focuses on the important special case of monotone treatment selection and response. Section 3.4 obtains bounds on average treatment effects.

3.1. MTR Assumptions

Classical econometric analysis of treatment response (Hood and Koopmans, 1953) combines an IV assumption of form (1) with the linear response assumption

$$(14) \quad y_j(t) = \beta t + \epsilon_j,$$

where ϵ_j is an unobserved covariate. The central finding is that assumptions (1) and (14) together identify the response parameter β , provided that z is not mean independent of v .⁶ For many years, empirical researchers have applied linear response models even though these models are not grounded in economic theory or other substantive reasoning. The literature has not provided compelling, or even suggestive, arguments in support of the hypothesis that response varies linearly with treatment and that all persons have the same response parameter.

Much of the empirical research that has applied linear response models could more plausibly apply monotone treatment response assumptions of the form (3) stating that, *ceteris paribus*, response varies monotonically with treatment. Consumer theory suggests that, *ceteris paribus*, the demand for a product weakly decreases as a function of the product's price. The theory of production suggests that, *ceteris paribus*, the output of a product weakly increases as a function of each input into the production process. Human capital theory suggests that, *ceteris paribus*, the wage that a worker earns weakly increases as a function of the worker's years of schooling. In these and other settings, MTR

⁶ It can be shown that if assumption (1) is weakened to an MIV assumption of form (2), the value of β is no longer identified but is bounded. The proof is available from the authors.

assumptions have a reasonably firm foundation.

The MIV assumption (2) and MTR assumption (3) are distinct in form and they have distinct implications for the conditional means $E[y(t)^*v = u]$. Whereas the MIV assumption implies the sharp bound given in Proposition 1, the MTR assumption implies this sharp bound (Manski, 1997, Corollary M1.2):

$$(15) \quad E(y^*v = u, t \leq z) \cdot P(t \leq z^*v = u) + K_0 \cdot P(t < z^*v = u) \leq E[y(t)^*v = u] \\ \leq E(y^*v = u, t \geq z) \cdot P(t \geq z^*v = u) + K_1 \cdot P(t > z^*v = u).$$

It is important to understand how the MTS and MTR assumptions differ from one another. Consider the variation of wages with schooling. It is common to hear the verbal assertion that "wages increase with schooling." The MTS and MTR assumptions interpret this statement in different ways. The MTS interpretation is that persons who select higher levels of schooling have weakly higher mean wage functions than do those who select lower levels of schooling; that is, $u_2 \geq u_1 \implies E[y(t)^*z = u_2] \geq E[y(t)^*z = u_1]$ for each $t \in [0, T]$. The MTR interpretation is that each person's wage function is weakly increasing in conjectured years of schooling; that is, for each $j \in [0, J]$, $t_2 \geq t_1 \implies y_j(t_2) \geq y_j(t_1)$. Although the MTS and MTR interpretations of the statement "wages increase with schooling" are distinct, they are not mutually exclusive.⁷

⁷ As discussed earlier, the MTS assumption is consistent with economic models of schooling choice and wage determination which predict that persons with higher ability have higher mean wage functions and choose higher levels of schooling than do persons with lower ability. The MTR assumption is consistent with economic models of the production of human capital through schooling.

3.2. General Findings

It is straightforward to combine an MTR assumption with an MIV assumption. We simply repeat the derivation of Section 2.1, with the MTR bound (15) replacing the no-assumptions bound (5). Let v be an MIV. Then we have the inequality (6). Hence $E[y(t)^*v = u]$ is no smaller than the MTR lower bound on $E[y(t)^*v = u_1]$ and no larger than the MTR upper bound on $E[y(t)^*v = u_2]$. This holds for all $u_1 < u$ and all $u_2 > u$. There are no other restrictions on $E[y(t)^*v = u]$. Thus

Proposition 2: Let the MIV and MTR Assumptions (2) and (3) hold. Then

$$(16) \quad \sup_{u_1 \# u} [E(y^*v = u_1, t \# z) \cdot P(t \# z^*v = u_1) + K_0 \cdot P(t < z^*v = u_1)]$$

$$\# E[y(t)^*v = u] \#$$

$$\inf_{u_2 \# u} [E(y^*v = u_2, t \# z) \cdot P(t \# z^*v = u_2) + K_1 \cdot P(t > z^*v = u_2)]$$

for each $u \in V$. In the absence of other information, this bound is sharp. \in

The MIV-MTR bound on the marginal mean $E[y(t)]$ is obtained from (16). Recall the application of the law of iterated expectations given in (8). Proposition 2 shows that the MIV-MTR lower and upper bounds on $E[y(t)^*v = u]$ are weakly increasing in u . Hence the sharp joint lower (upper) bound on $\{E[y(t)^*v = u], u \in V\}$ is obtained by setting each of the quantities $E[y(t)^*v = u], u \in V$ at its lower (upper) bound in (16). Inserting these lower and upper bounds into (8) yields

Proposition 2, Corollary 1: Let the MIV-MTR Assumptions (2) and (3) hold. Then

$$(17) \quad \int_0^V P(v = u) \left\{ \sup_{u_1 \neq u} [E(y^*v = u_1, t \leq z) \cdot P(t \leq z^*v = u_1) + K_0 \cdot P(t < z^*v = u_1)] \right\} \\ \# E[y(t)] \#$$

$$\int_0^V P(v = u) \left\{ \inf_{u_2 \leq u} [E(y^*v = u_2, t \neq z) \cdot P(t \neq z^*v = u_2) + K_1 \cdot P(t > z^*v = u_2)] \right\}.$$

In the absence of other information, this bound is sharp. \in

3.3. Monotone Treatment Selection and Response

In general, the MIV-MTR bounds on $E[y(t)^*v = u]$ and $E[y(t)]$ are informative only if the outcome space Y is bounded. Yet there is an important special case in which these bounds are informative even if Y is unbounded. This is the case of monotone treatment selection, in which $v = z$. Application of Proposition 2 yields this MTS-MTR bound on $E[y(t)^*z = u]$:

$$(18) \quad u < t \quad Y \quad \sup_{u_1 \neq u} E(y^*z = u_1) \# E[y(t)^*z = u] \# \inf_{u_2 \leq t} E(y^*z = u_2)$$

$$u = t \quad Y \quad \sup_{u_1 \neq t} E(y^*z = u_1) \# E[y(t)^*z = u] \# \inf_{u_2 \leq t} E(y^*z = u_2)$$

$$u > t \quad Y \quad \sup_{u_1 \neq t} E(y^*z = u_1) \# E[y(t)^*z = u] \# \inf_{u_2 \leq u} E(y^*z = u_2).$$

It follows from the MTS and MTR assumptions that ⁸

$$(19) \quad u' \# u \quad Y \quad E(y^*z = u') = E[y(u')^*z = u'] \# E[y(u)^*z = u'] \\ \# E[y(u)^*z = u] = E(y^*z = u).$$

Combining (18) and (19) yields these MTS-MTR bounds, which are informative even if Y is unbounded:

Proposition 2, Corollary 2: Let the MTS-MTR Assumptions (11) and (3) hold. Then

$$(20) \quad u < t \quad Y \quad E(y^*z = u) \# E[y(t)^*z = u] \# E(y^*z = t) \\ u = t \quad Y \quad E[y(t)^*z = u] = E(y^*z = t) \\ u > t \quad Y \quad E(y^*z = t) \# E[y(t)^*z = u] \# E(y^*z = u)$$

and

$$(21) \quad \int_{u < t} E(y^*z = u) \cdot P(z = u) + E(y^*z = t) \cdot P(z \leq t) \# E[y(t)] \\ \# \int_{u > t} E(y^*z = u) \cdot P(z = u) + E(y^*z = t) \cdot P(z \geq t).$$

In the absence of other information, these bounds are sharp. ϵ

⁸ Equation (19) suggests a test of the joint MTS-MTR hypothesis. Under this hypothesis, $E(y^*z = u)$ must be a weakly increasing function of u . Hence we should reject the hypothesis if $E(y^*z = u)$ is not weakly increasing in u . This test is a weakened version of the stochastic dominance test proposed in Manski (1997, p.1327) for testing the joint hypothesis that treatment response is monotone and that z is statistically independent of $y(\cdot)$.

3.4. Bounds on Average Treatment Effects

Propositions 1 and 2 give sharp bounds on mean outcomes under specified treatments. Let s and t be two such treatments, with $s < t$. Often the objects of interest are average treatment effects of the form

$$(22) \quad \tau(s, t) = E[y(t)] - E[y(s)].$$

If only an MIV assumption is imposed, the sharp lower (upper) bound on $\tau(s, t)$ is obtained by subtracting the lower (upper) bound on $E[y(t)]$ from the upper (lower) bound on $E[y(s)]$. If MIV and MTR assumptions are imposed, a bound on $\tau(s, t)$ may be obtained in the same manner but this bound is not sharp. The general case is complex to analyze, but the MTS-MTR case is straightforward.

Let $u \in [0, T]$. Under the MTS-MTR assumption, equation (20) gives sharp bounds on each of $E[y(t) | z = u]$ and $E[y(s) | z = u]$. Inspection of these bounds shows that it is jointly feasible for $E[y(t) | z = u]$ to be at its upper bound and $E[y(s) | z = u]$ to be at its lower bound. Thus the MTS-MTR sharp upper bound on $\tau(s, t)$ is the upper bound on $E[y(t)]$ minus the lower bound on $E[y(s)]$, namely

$$(23) \quad \tau(s, t) \leq \int_{u > t} E(y^* | z = u) \cdot P(z = u) + E(y^* | z = t) \cdot P(z \in [s, t]) \\ - \int_{u < s} E(y^* | z = u) \cdot P(z = u) - E(y^* | z = s) \cdot P(z \in [s, t]) \\ = \int_{u < s} [E(y^* | z = t) - E(y^* | z = u)] \cdot P(z = u) + [E(y^* | z = t) - E(y^* | z = s)] \cdot P(z \in [s, t]) \\ + \int_{u > t} [E(y^* | z = u) - E(y^* | z = s)] \cdot P(z = u).$$

Observe that, by (19), the right side of (23) is non-negative and no smaller than $E(y^*z = t) - E(y^*z = s)$, which is the value of $\gamma(s, t)$ under the ETS assumption.

In contrast, it is not jointly feasible for $E[y(t)^*z = u]$ to be at its lower bound and $E[y(s)^*z = u]$ to be at its upper bound. Placing $E[y(t)^*z = u]$ and $E[y(s)^*z = u]$ at these limit points yields a non-positive value for $\gamma(s, t)$. Under the MTR assumption, however, the lower bound on $\gamma(s, t)$ must be no less than zero (see Manski, 1997, and Pepper, 1997). Our application of the MTS-MTR assumption in the next section uses (23) as the upper bound on $\gamma(s, t)$ and zero as the lower bound.

4. Empirical Analysis of The Returns to Schooling

4.1. Maintained Assumptions

Labor economists studying schooling as a treatment commonly suppose that each individual j has a log(wage) function $y_j(t)$, giving the log(wage) that j would receive were he to obtain t years of schooling. In theories of schooling and labor supply, $y_j(\cdot)$ is interpreted as person j 's production function for human capital. Observing realized covariates, schooling, and wages, labor economists often seek to learn features of the distribution of these production functions in a sub-population of interest. In particular, many studies report estimates of the expected returns to completing t years of schooling relative to s years for $s < t$, namely $\gamma(s, t^*w)$, where the covariates w define the sub-population of

interest.⁹

To infer the returns to schooling, some researchers assume that treatment selection is exogenous. Others use family and environmental attributes as instrumental variables.¹⁰ Yet the validity of whatever IV assumption is imposed seems inevitably to be questioned. Researchers also often assume, without compelling reason, that $\log(\text{wage})$ varies linearly with conjectured schooling.

The empirical analysis presented here assumes only that treatment selection and response are monotone. As discussed in Section 3.1, the MTS and MTR assumptions are consistent with conventional theories of human capital accumulation. Even if these assumptions do not warrant unquestioned acceptance, they certainly merit serious consideration.¹¹

4.2. Data

We use data from the NLSY. In its base year of 1979, the NLSY interviewed

⁹ The literature on the returns to schooling, as other applied literatures on treatment response, exhibits varying definitions of the sub-population of interest. The definition matters substantively whenever there is concern that treatment response may vary from person to person. Considerations in defining the sub-population of interest are discussed in many sources, including Heckman and Robb (1985), Björkland and Moffitt (1987), and Manski (1996). The theoretical analysis in the present paper applies to any sub-population of known composition; that is, to a sub-population defined by conditioning on observable covariates w . Our analysis does not apply to sub-populations of unknown composition, such as those defined by Imbens and Angrist (1994) and by Heckman and Vytlacil (1999) in their studies of local average and local IV treatment effects.

¹⁰ See, for example, Angrist and Kreuger (1991), Blakemore and Low (1984), Blackburn and Neumark (1993, 1995), Butcher and Case (1994), Card (1993, 1994), Frazis (1993), Garen (1984), Kenny et al. (1979), Lang and Ruud (1986), Murphy and Welch (1989), Osterbeek (1990), and Willis and Rosen (1979).

¹¹ A potential counter to the MTS assumption is suggested by Card (1994), who speculates that ability and taste for schooling may be negatively associated.

12,686 persons who were between the ages of 14 to 22 at that time. Nearly half of the respondents were randomly sampled, the remaining being selected to over-represent certain demographic groups (see Center for Human Resource Research, 1995). We restrict attention to the 1,257 randomly sampled white males who, in 1994, reported that they were full-time year-round workers with positive wages. We exclude the self-employed. Thus our empirical analysis concerns the sub-population of persons who, in the notation introduced in Section 1 but since left implicit, have the shared observable covariates

(24) $w =$ (white males, full-time year-round workers in 1994, not self-employed).

We observe each respondent's 1994 hourly wage and realized years of schooling. In our notation, z is realized years of schooling, the response variable $y_j(t)$ is the $\log(\text{wage})$ that person j would experience if he were to have t years of schooling, and $y_j = y_j(z_j)$ is the observed hourly $\log(\text{wage})$.

4.3. Statistical Considerations

The bounds developed in Propositions 1 and 2 are continuous functions of various nonparametrically estimable conditional probabilities and mean responses. In our application, we need only estimate the MTS-MTR upper bound on $\beta(s, t)$ given in (23). Thus we need only estimate the probabilities $P(z)$ of realizing z years of schooling and the expectations $E(y^*z)$ of $\log(\text{wage})$ conditional on schooling. For each value of z , we use the empirical distribution of schooling to estimate $P(z)$ and the sample average $\log(\text{wage})$ of respondents with z years of schooling to estimate $E(y^*z)$. So estimation of the MTS-MTR upper bound is a

simple matter.

Asymptotically-valid sampling confidence intervals for the bounds may be computed using the delta method or bootstrap approaches. We apply the percentile bootstrap method. To be precise, the bootstrap sampling distribution of an estimate of the MTS-MTR upper bound (23) is its sampling distribution under the assumption that the unknown population distribution of (AFQT score, realized years of schooling, realized wages) among persons with the covariates w specified in (24) equals the empirical distribution of these variables in the sample of 1,257 randomly sampled NLSY respondents. Next to each upper bound estimate we report the 0.95 quantile of its bootstrap sampling distribution.¹²

4.4. Findings

Table 1 gives the estimates of $E(y^*z)$ and $P(z)$ that we use to estimate the MTS-MTR bounds. The table shows that 41 percent of the NLSY respondents have 12 years of schooling and 19 percent have 16 years, but the support of the schooling distribution stretches from 8 years to 20 years. Hence we are able to report findings on $\gamma(s, t)$ for $t = 9$ through 20 and $8 \leq s < t$.

In Section 3.3 we observed that the MTS-MTR assumption is a testable

¹² We caution readers that some applications may not be as straightforward from a statistical perspective. In general, applications of Proposition 1 and 2 requiring taking infs and sups of collections of nonparametric regression estimates. The consistency of the resulting bounds estimates is easy to establish, but the sampling distributions of these estimates is not yet well understood. A particular concern is that lower (upper) bound estimates, being sups (infs) of collections of nonparametric regression estimates, may have non-negligible positive (negative) finite-sample biases. This concern does not arise in the special case of MTS-MTR bounds, which do not require taking sups and infs. We do not attempt to resolve the open statistical questions associated with general applications of MIV bounds in this paper, which is primarily concerned with identification.

hypothesis, which should be rejected if $E(y^*z = u)$ is not weakly increasing in u . The estimate of $E(y^*z)$ in Table 1 for the most part does increase with z , but there are occasional dips. Computing a uniform 95 percent confidence band for the estimate of $E(y^*z)$, we have found that the band contains everywhere monotone functions. Hence we proceed on the basis that the MTS-MTR assumption is consistent with the empirical evidence.

Table 1: NLSY Empirical Mean log(wages) and Distribution of Years of Schooling

<u>z</u>	<u>E(y z)</u>	<u>P(z)</u>	<u>Sample Size</u>
8	2.249	0.014	18
9	2.302	0.018	22
10	2.195	0.018	23
11	2.346	0.025	32
12	2.496	0.413	519
13	2.658	0.074	93
14	2.639	0.083	104
15	2.693	0.035	44
16	2.870	0.189	238
17	2.775	0.038	48
18	3.006	0.051	64
19	3.009	0.020	25
20	2.936	0.021	27
Total		1	1257

Table 2 reports the estimates and 0.95 bootstrap quantiles of the MTS-MTR upper bounds on $\beta(t-1, t)$, $t = 9, \dots, 20$ followed by the upper bound on $\beta(12, 16)$, which compares high school completion with college completion.¹³ To provide context for these upper bound estimates, let us review some of the point estimates of $\beta(t-1, t)$ reported in the recent empirical literature on the returns to schooling. Most of the point estimates cited in the survey by Card

¹³ Point estimates of these treatment effects under the assumption of exogenous treatment selection may be obtained directly from the first column of Table 1. Under the ETS assumption, $\beta(s, t) = E(y^*z = t) - E(y^*z = s)$.

(1994) are between 0.07 and 0.09. Card (1993), using a linear response model and assuming that proximity to college is an IV, reports a point estimate of 0.132. Ashenfelter and Krueger (1994), assuming that treatment selection is exogenous within their sample of twins, report various estimates and conclude that (p. 1171): "our best estimate is that increased schooling increases averages wage rates by about 12-16 percent per year completed."

Table 2: MTS-MTR Upper Bounds on Returns to Schooling

<u>s</u>	<u>t</u>	Upper Bound on $\beta(s, t)$	
		<u>Estimate</u>	<u>0.95 Bootstrap Quantile</u>
8	9	0.390	0.531
9	10	0.334	0.408
10	11	0.445	0.525
11	12	0.313	0.416
12	13	0.253	0.307
13	14	0.159	0.226
14	15	0.202	0.288
15	16	0.304	0.369
16	17	0.165	0.256
17	18	0.386	0.485
18	19	0.368	0.539
19	20	0.296	0.486
--	--	-----	-----
12	16	0.397	0.450

None of the estimates of upper bounds on $\beta(t - 1, t)$ in Table 2 lie below the point estimates reported in the literature. The smallest of the upper bound estimates are 0.159 for $\beta(13, 14)$ and 0.165 for $\beta(16, 17)$. These are about equal to the largest of the point estimates known to us, namely those in Ashenfelter and Krueger (1994). It might therefore appear that the MTS-MTR assumption does not, in this application, have sufficient identifying power to affect current thinking about the magnitude of the returns to schooling.

A different conclusion emerges with consideration of the estimate of the

upper bound on $\delta(12, 16)$. We estimate that completion of a four-year college yields at most an increase of 0.397 in mean $\log(\text{wage})$ relative to completion of high school. This implies that the average value of the four year-by-year treatment effects $\delta(12, 13)$, $\delta(13, 14)$, $\delta(14, 15)$, and $\delta(15, 16)$ is at most 0.099, which is well below the point estimates of Card (1993) and Ashenfelter and Krueger (1994). This conclusion continues in force if, acting conservatively, one uses the 0.95 bootstrap quantile of 0.450 to estimate the upper bound on $\delta(12, 16)$. Then the implied upper bound on the average value of the year-by-year treatment effects is 0.113. Thus we find that, under the MTS-MTR assumption, the returns to college-level schooling are smaller than some of the point estimates reported recently.

5. Some Variations on the Theme

This paper has introduced the general idea of a monotone instrumental variable and the important special case of monotone treatment selection. It is easy to think of variations on the MIV theme that warrant study. One would be to combine the MIV idea with the idea of contaminated instruments introduced by Hotz, Mullins, and Sanders (1997). Another would be to begin from the statistical independence assumption that holds in classical randomized experiments and weaken it to a stochastic dominance assumption. Yet another variation on the MIV theme would be to weaken mean independence to some form of approximate mean independence.

References

- Angrist, J. and A. Krueger (1991), "Does Compulsory Schooling Affect Schooling and Earnings?" Quarterly Journal of Economics 106, 979-1014.
- Ashenfelter, O. and A. Krueger (1994), "Estimates of the Economic Returns to Schooling from a New Sample of Twins," American Economic Review, 84, 1157-1173.
- Balke, A. and J. Pearl(1997), "Bounds on Treatment Effects from Studies With Imperfect Compliance," Journal of the American Statistical Association, 92, 1171-1177.
- Björklund, A. and R. Moffitt (1987), "Estimation of Wage Gains and Welfare Gains in Self-Selection Models," Review of Economics and Statistics,69, 42-49.
- Blackburn, M.L. and D. Neumark (1993), "Omitted Ability Bias and the Increase in the Return to Schooling," Journal of Labor Economics 11, 521-44.
- Blackburn, M.L. and D. Neumark (1995), "Are OLS Estimates of the Return to Schooling Biased Downward? Another Look," Review of Economics and Statistics 77, 217-30.
- Blakemore, A. and S. Low (1984), "The High-School Dropout Decision and Its Wage Consequences," Economics of Education Review 3, 111-119.
- Butcher, K. and A. Case (1994), "The Effect of Sibling Sex Composition on Women's Education and Earnings," Quarterly Journal of Economics 109, 531-564
- Card, D. (1993), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," Working Paper 4483, National Bureau of Economic Research.
- Card, D. (1994), "Earnings, Schooling, and Ability Revisited," NBER Working Paper 4832, National Bureau of Economic Research, Cambridge, Mass.
- Center for Human Resource Research(1995), "NLS Handbook 1995. The National Longitudinal Surveys of Labor Market Experience, "Columbus:Ohio State University.
- Frazis, H. (1993), "Selection Bias and the Returns to Education," Journal of Human Resources 28, 538-54.
- Garen, J. (1984), "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," Econometrica 52, 1199-1218.
- Heckman, J. and R. Robb (1985), "Alternative Methods for Evaluating the Impact of Interventions." in J. Heckman and B. Singer (editors), Longitudinal Analysis of Labor Market Data,New York: Cambridge University Press.
- Heckman, J. and E. Vytlacil (1999), "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," Department of Economics, University of Chicago.

Hood, W. and T. Koopmans (editors) (1953), Studies in Econometric Method, New York: Wiley.

Hotz, J., C. Mullins, and S. Sanders (1997), "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing," Review of Economic Studies 64, 575-603.

Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," Econometrica, 62, 467-476.

Kenny, L., L. Lee, G.S. Maddala, and R. Troust (1979), "Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data," International Economic Review 20, 775-89.

Lang, K. and P. Ruud (1986), "Returns to Schooling, Implicit Discount Rates and Black-White Wage Differentials," Review of Economics and Statistics 68, 41-47.

Manski, C. (1988), Analog Estimation Methods in Econometrics, London: Chapman & Hall.

Manski, C. (1989), "Anatomy of the Selection Problem", Journal of Human Resources, 24, 343-360.

Manski, C. (1990), "Nonparametric Bounds on Treatment Effects," American Economic Review Papers and Proceedings, 80, 319-323.

Manski, C. (1994), "The Selection Problem," in C. Sims (editor) Advances in Econometrics: Sixth World Congress, Cambridge, UK: Cambridge University Press.

Manski, C. (1996), "Learning about Treatment Effects from Experiments with Random Assignment of Treatments," Journal of Human Resources 31, 707-733.

Manski, C. (1997), "Monotone Treatment Response," Econometrica, 65, 1311-1334.

Murphy, K. and F. Welch (1989), "Wage Premiums for College Graduates: Recent Growth and Possible Explanations," Educational Researcher, 18(4), 17-26.

Oosterbeek, H. (1990), "Education and Earnings in the Netherlands: An Empirical Analysis," European Economic Review 34, 1353-75.

Pepper, J. (1997), "The Intergenerational Transmission of Welfare Receipt," Department of Economics, University of Virginia.

Robins, J. (1989), "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in Sechrest, L., H. Freeman, and A. Mulley. eds. Health Service Research Methodology: A Focus on AIDS, NCHSR, U.S. Public Health Service.

Willis, R. and S. Rosen (1979), "Education and Self-Selection," Journal of Political Economy 87, S7-S37.