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
NBER

and

John V. Pepper
Department of Economics
University of Virginia
Charlottesville, VA 22902

January 1998

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 the Federal Reserve Board, Northwestern University, and the University of Virginia. We have also benefitted from the comments of Jinyong Hahn and of Mark Manuszak.

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, 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. A particularly interesting special case of an MIV assumption is *monotone treatment selection* (MTS).

IV and MIV assumptions may be imposed alone or in combination with other assumptions. We study the identifying power of MIV assumptions in three informational settings: MIV alone; MIV combined with the classical linear response assumption; MIV combined with the *monotone treatment response* (MTR) assumption. We apply the results to the problem of inference on the returns to schooling. We analyze wage data reported by white male respondents to the National Longitudinal Survey of Youth (NLSY) and use the respondent's AFQT score as an MIV. We find that this MIV assumption has little identifying power when imposed alone. However combining the MIV assumption with the MTR and MTS assumptions yields fairly tight bounds on two distinct measures of the returns to schooling.

1. Introduction

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, 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, study their identifying power, and give an empirical application.

Whereas an IV assumption is a mean-independence condition holding that mean response is constant across the subpopulations of persons with different values of an observable covariate, an MIV assumption supposes that mean response varies weakly monotonically across these subpopulations. MIV assumptions may be imposed alone or in combination with other assumptions. We first characterize the identifying power of MIV assumptions alone, extending the Manski (1989, 1990, 1994) analysis of the identifying power of IV assumptions alone. We next study the identifying power of MIV assumptions in combination with two assumptions restricting the shapes of response functions -- the linear response assumption of classical econometric analysis and the monotone treatment response assumption of Manski (1997). We then give an empirical application to the problem of inference on the returns to schooling.

The findings that we report add to the emerging literature developing nonparametric bounds on treatment effects. Several contributors to this literature have examined the identifying power of IV assumptions and of variations on the IV theme. Manski (1990) reported sharp bounds on mean outcomes and average treatment effects under the IV assumption alone. Robins (1989) and Balke and Pearl (1997) have considered the statistical independence assumption that holds in classical randomized experiments; that is, response functions are statistically independent of assigned treatments, not just mean independent. Robins (1989) reported the same bound as Manski (1990), but Balke and Pearl (1997) later showed that this bound, although sharp under the IV assumption, sometimes is not sharp under the statistical independence assumption (see also Robins and Greenland, 1996).

Hotz, Mullins, and Sanders (1997) have studied *contaminated instrument* assumptions, which weaken IV assumptions in a different way than MIV assumptions do. They suppose that an IV assumption holds in a specified population, but the observable population is a probability mixture of this population and another one in which the IV assumption does not hold. Applying results in Horowitz and Manski (1995) to the resulting contaminated sampling problem, they derive sharp bounds on average treatment effects in the specified population.

This paper uses the same formal setup and notation as Manski (1997). We assume there is a probability space (J, Ω, P) of individuals. Each member j of population J has observable covariates $x_j \in X$ and a response function $y_j(\cdot)$: $T \to Y$ mapping the mutually exclusive and exhaustive *treatments* $t \in T$ into *outcomes* $y_j(t) \in Y$. Person 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. 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 IV and MIV assumptions. Let

$x = (w, v)$ and $X = W \times V$. Each value of (w, v) defines an observable subpopulation of persons. The IV (or mean-independence) 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 subpopulations $(w, v = u)$, $u \in V$. Thus,

<u>IV Assumption</u>: Covariate v is an instrumental variable 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 weaken the IV idea by replacing the equality in (1) by an inequality. Thus

<u>MIV Assumption</u>: Let V be an ordered set. Covariate v is a monotone instrumental variable if, for each $t \in T$, each value of w, and all $(u_1, u_2) \in (V \times V)$ such that $u_2 \geq u_1$,

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

Researchers contemplating application of IV and MIV assumptions should have a clear understanding of their content. With this in mind, Section 2 examines what these assumptions do and do not assert. The key is to integrate the concepts of treatments and covariates in the analysis of treatment response. We use the integrated framework to suggest MIV assumptions that might credibly be applied in analyses of the returns to schooling.

In Sections 3 through 5, we study the identifying power of MIV assumptions in three informational settings. In Section 3 we consider an MIV assumption alone and report sharp bounds on the conditional mean responses $E[y(t)|w, v = u]$, $u \in V$ and the marginal mean $E[y(t)|w]$. The MIV 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.

In Section 4, we combine an MIV assumption with the linear response assumption

(3) $$y_j(t) = \beta t + \epsilon_j,$$

where ϵ_j is an unobserved covariate. Classical econometric analysis of treatment response (see Hood and Koopmans, 1953) combines an IV assumption with assumption (3). The central finding is that assumptions (1) and (3) together identify the response parameter β, provided that z is not mean independent of v. We find here that when the IV assumption is weakened to an MIV assumption, the value of β is no longer identified but is bounded.

In Section 5, we combine an MIV assumption with the *monotone treatment response* (MTR) assumption studied in Manski (1997). This is

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

(4) $$t_2 \geq t_1 \Rightarrow y_j(t_2) \geq y_j(t_1).$$

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

In Section 6, we present an empirical case study applying the findings of Sections 3 through 5 to inference on the returns to schooling. Here we analyze wage data reported by white male respondents to the National Longitudinal Survey of Youth (NLSY) who are employed full time in 1994. We propose using the respondent's AFQT score as an MIV and find that this assumption has little identifying power when imposed alone. However substantively interesting findings emerge when we combine this MIV assumption with the MTR and MTS assumptions. We obtain an upper bound on the returns to schooling that is only slightly above the point estimates commonly reported in the literature.

In Section 7, we draw conclusions and call attention to two variations on the MIV theme. One might weaken the statistical independence assumption of classical randomized experiments to an assumption of weak stochastic dominance. *Fuzzy instrumental variables* weaken the mean independence of the standard IV assumption to some form of approximate mean independence.

To simplify the exposition in Sections 2 through 5, we henceforth leave implicit the conditioning on w maintained in the definitions of IVs and MIVs. Moreover, to keep the focus on identification, we treat identified, consistently estimable quantities as known. In the empirical analysis of Section 6, we explicitly condition on specified covariates w. There we discuss statistical considerations and use bootstrapped confidence intervals to measure the sampling precision of the estimates.

2. What Are IV And MIV Assumptions?

The concepts introduced in Section 1 suffice to define IV and MIV assumptions and to analyze their identifying power. Imbedding these concepts within a broader framework, however, helps to understand the meaning of these assumptions. Section 2.1 sets out this broader framework. Section 2.2 uses it to suggest MIV assumptions that might credibly be imposed when analyzing the returns to schooling.

2.1. Treatments and Covariates

The discussion of Section 1 suggests a sharp distinction between treatments and covariates. Treatments have been presented as quantities that may be manipulated autonomously, inducing variation in response. We have been careful to use separate symbols for the conjectural treatments $t \in T$ and for the actual treatment $z_j \in T$ realized by person j. Covariates have been presented only as realized quantities associated with the members of the population, with no mention of their manipulability. We have used $v_j \in V$ to denote the covariate value associated with person j. We have given no notation for conjectural values of this covariate.

A symmetric perspective on treatments and covariates emerges if, as in Manski (1997, Section 2.4), we enlarge the set of treatments from T to the Cartesian product set $T \times V$ and introduce a generalized response function $y^*_j(\cdot, \cdot): T \times V \to Y$ mapping elements of $T \times V$ into outcomes. Now, for each $(t, u) \in T \times V$, $y^*_j(t, u)$ is the outcome that person j would experience if she were to receive the conjectural treatment pair (t, u). The treatment pair realized by person j is (z_j, v_j) and her realized outcome is $y_j = y^*_j(z_j, v_j)$. The response function $y_j(\cdot): T \to Y$ introduced as a primitive in Section 1 is now a derived *sub-response function*, obtained by evaluating $y^*_j(\cdot, \cdot)$ with its second argument set at the realized treatment value v_j; that is,

$$(5) \quad y_j(\cdot) \equiv y^*_j(\cdot, v_j).$$

In this broadened framework, a covariate is synonymous with a realized treatment.

With this as background, observe that the familiar statement "covariate v does not affect response" has two distinct formal interpretations. One interpretation is that v is an IV and the other is that outcomes are constant under conjectural variations in v. The IV interpretation, rewritten here using the $y^*(\cdot, \cdot)$ notation, states that for each $t \in T$,

$$(1)' \quad E[y^*(t, v)|v = u'] = E[y^*(t, v)|v = u], \quad \forall\ u \in V,\ u' \in V.$$

The other interpretation is that for each $j \in J$ and $t \in T$,

$$(6) \quad y^*_j(t, u) = y^*_j(t, v_j), \qquad \forall\ u \in V.$$

The second interpretation, which has not yet been named, will be called a *constant treatment response* (CTR) assumption.

Similarly, the familiar statement: "response is monotone in v" has two interpretations. One is that v is an MIV and the other is that outcomes vary monotonically under conjectural variations in v. The MIV interpretation is that for each $t \in T$, (2') $u_2 \geq u_1 \Rightarrow E[y^*(t, v) | v = u_2] \geq E[y^*(t, v) | v = u_1]$.

The other is that for each $j \in J$ and $t \in T$,

(7) $u_2 \geq u_1 \Rightarrow y^*_j(t, u_2) \geq y^*_j(t, u_1)$.

The second interpretation is an MTR assumption as in (4), but here applied to u rather than to t.

Distinguishing appropriately between IV/MIV assumptions and CTR/MTR assumptions is critical to the informed analysis of treatment response. We cannot know how often empirical researchers, thinking loosely that "covariate v does not affect response," have imposed an IV assumption but really had a CTR assumption in mind. Introducing MIV assumptions here, we want to squelch from the start any confusion between MIV and MTR assumptions.

2.2. Application to The Returns to Schooling

Labor economists studying the returns to schooling commonly suppose that each individual j has a wage function $y_j(t)$, giving the wage that j would receive were she to obtain t years of schooling. Observing realized covariates, schooling, and wages in the population, labor economists often seek to learn the expected returns to the t^{th} year of schooling, namely $E[y(t)] - E[y(t-1)]$.

Researchers often use personal, family, and environmental attributes as instrumental variables for wages. 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), Griliches (1977), Garen (1984), Kenny et al. (1979), Lang and Ruud (1986), Osterbeek (1990), and Willis and Rosen (1979). Yet the validity of whatever IV assumption may be imposed seems inevitably to be questioned.

Some formal analysis suggests covariates that might plausibly serve as MIVs in analyses of the returns to schooling. We present a Lemma giving conditions sufficient for v to be an MIV. We then pose some covariates to which the lemma might be applied.

Lemma: Assume that person j's earning function has the form

(8) $y^*_j(t, u) = g(t, u, \alpha_j, \epsilon_j)$.

Here $t \in T$ is years of schooling and $u \in V$ is another treatment taking values in the ordered set V. The quantities (α_j, ϵ_j) are person j's realizations of the unobserved covariates (α, ϵ). Covariate α measures some form of ability and takes values in an ordered set A, while ϵ is unnamed and takes values in some space E. Assume that, for each $t \in T$ and each realization of ϵ, the sub-response function $g(t, \cdot, \cdot, \epsilon): V \times A \rightarrow R^1$ is weakly increasing in u and α. Assume that ϵ is statistically independent of (α, v). Assume that the distribution $P(\alpha|v)$ of ability conditional on v is weakly increasing in v; that is, $u_2 \geq u_1$ implies that $P(\alpha|v = u_2)$ weakly stochastically dominates $P(\alpha|v = u_1)$. Then v is an MIV. ■

Proof: Let $u_2 \geq u_1$. We need to show that $E[y^*(t, v)|v = u_2] \geq E[y^*(t, v)|v = u_1]$.

Let $u \in [u_1, u_2]$. The assumptions imposed imply that

$E[y^*(t, v)|v = u_2] - E[y^*(t, v)|v = u_1]$

$$= E[g(t, v, \alpha, \epsilon)|v = u_2] - E[g(t, v, \alpha, \epsilon)|v = u_1]$$

$$= \int\int g(t, u_2, \alpha, \epsilon)dP(\alpha|\epsilon, v = u_2)dP(\epsilon|v = u_2)$$
$$- \int\int g(t, u_1, \alpha, \epsilon)dP(\alpha|\epsilon, v = u_1)dP(\epsilon|v = u_1)$$

$$= \int [\int g(t, u_2, \alpha, \epsilon)dP(\alpha|v = u_2) - \int g(t, u_1, \alpha, \epsilon)dP(\alpha|v = u_1)]dP(\epsilon)$$

$$\geq \int [\int g(t, u, \alpha, \epsilon)dP(\alpha|v = u_2) - \int g(t, u, \alpha, \epsilon)dP(\alpha|v = u_1)]dP(\epsilon)$$

$$\geq 0.$$

The first equality applies (8). The second equality writes the expectations explicitly as integrals. The third equality applies the assumption that ϵ is statistically independent of (α, v). The fourth inequality applies the assumption that $g(t, \cdot, \alpha, \epsilon)$ is monotone in u. The fifth inequality applies the assumption that $g(t, u, \cdot, \epsilon)$ is monotone in α and that $P(\alpha|v = u_2)$ weakly dominates $P(\alpha|v = u_1)$.

Q.E.D.

Empirical studies of the returns to schooling have commonly maintained the assumptions of this lemma within parametric response models that also impose strong functional form assumptions on $g(\cdot, \cdot, \cdot, \cdot)$ and strong distributional assumptions on (α, ϵ). Among various classes of covariates that plausibly fit the conditions of the lemma, perhaps the most obvious are achievement measures -- standardized test scores, grade point averages, and the like. The lemma shows that sufficient conditions for an achievement measure to be a valid MIV are

(i) Holding fixed (t, α, ϵ), wages $g(t, \cdot, \alpha, \epsilon)$ are a weakly increasing function of measured achievement u.

(ii) The distribution of ability among persons with higher measured achievement weakly dominates the distribution of ability among persons with lower measured achievement.

These seem to us to be easily understood and plausible assumptions. Assumption (i) permits measured achievement to have either no direct impact on wages or a direct positive impact, as is implied by signaling models in which employers use measured achievement to predict ability. Assumption (ii) does not require that measured achievement be identical to ability, only that it be a weakly positive predictor of ability in the sense of weak stochastic dominance.

3. Identification Using an MIV Assumption Alone

The identifying power of an IV assumption alone, with no other assumptions imposed, has been studied in Manski (1990, 1994). 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 on these quantities.

Section 3.1 gives the general results. Section 3.2 applies these results to an important special case in which the form of the bounds simplifies considerably. This is the case of *monotone treatment selection*, which weakens the familiar assumption of exogenous treatment selection from an IV to an MIV. Section 3.3 discusses multi-dimensional IV and MIV assumptions.

3.1. General Case

The starting point for determination of the identifying power of both IV and MIV assumptions is the no-assumptions bound on $E[y(t)|v]$ reported in Manski (1989). Let $[K_0, K_1]$ denote the range of the outcome space 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

$$(9) \quad E[y(t)|v = u] = E(y|v = u, z = t)\cdot P(z = t|v = u) + E[y(t)|v = u, z \neq t]\cdot P(z \neq t|v = u).$$

The sampling process identifies each of the quantities on the right side except for the censored mean $E[y(t)|v = u, z \neq t]$. In the absence of assumptions, all that is known about this censored mean is that $K_0 \leq E[y(t)|v = u, z \neq t] \leq K_1$. This implies the sharp bound

$$(10) \quad E(y|v = u, z = t)\cdot P(z = t|v = u) + K_0\cdot P(z \neq t|v = u) \leq E[y(t)|v = u] \leq E(y|v = u, z = t)\cdot P(z = t|v = u) + K_1\cdot P(z \neq t|v = u).$$

This bound is informative if the treatment selection probability $P(z = t|v = u)$ is positive and if Y is a bounded outcome space, so K_0 and K_1 are finite.

Under the IV assumption, $E[y(t)|v = u]$ is constant across $u \in V$. It follows that the common value of $E[y(t)|v = u]$, $u \in V$ lies in the intersection of the bounds (10) across the elements of V. Any point in this intersection is feasible. Thus, for all $u \in V$, we obtain the common sharp bound

$$(11) \quad \sup_{u' \in V} [E(y|v = u', z = t) \cdot P(z = t|v = u') + K_0 \cdot P(z \neq t|v = u')]$$

$$\leq E[y(t)|v = u] \leq$$

$$\inf_{u' \in V} [E(y|v = u', z = t) \cdot P(z = t|v = u') + K_1 \cdot P(z \neq t |v = u')].$$

This is also the sharp bound on the marginal mean $E[y(t)]$. See Manski (1990, 1994) for further discussion.

The IV bound (11) is necessarily a subset of the no-assumptions bound (10). It is a proper subset for some $u \in V$ if and only if the no-assumptions bounds for $u \in V$ do not all coincide. This is a *rank condition* in the spirit of, but formally distinct from, the familiar rank condition associated with classical econometric analysis of treatment response (see Section 4).

Now consider an MIV assumption. In this case, $E[y(t)|v = u]$ need not be constant across $u \in V$ but we do have the inequality restriction

$$(12) \quad u_1 \leq u \leq u_2 \Rightarrow E[y(t)|v = u_1] \leq E[y(t)|v = u] \leq E[y(t)|v = u_2].$$

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

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

$$(13) \qquad \sup_{u_1 \le u} [E(y)|v = u_1, z = t) \cdot P(z = t|v = u_1) + K_0 \cdot P(z \ne t|v = u_1)]$$

$$\le \quad E[y(t)|v = u] \quad \le$$

$$\inf_{u_2 \ge u} [E(y|v = u_2, z = t) \cdot P(z = t|v = u_2) + K_1 \cdot P(z \ne t|v = u_2)].$$

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

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. Then we may use the law of iterated expectations to write

$$(14) \quad E[y(t)] \quad = \quad \sum_{u \in V} P(v = u) \cdot E[y(t)|v = u]$$

Equation (13) 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 $[E[y(t)|v = u], u \in V\}$ is obtained by setting each of the quantities $E[y(t)|v = u]$, $u \in V$ at its lower (upper) bound as given in (13). Inserting these lower and upper bounds into (14) yields

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

$$(15)\quad \sum_{u \in V} P(v = u)\{ \sup_{u_1 \le u} [E(y|v = u_1,\ z = t)\cdot P(z = t|v = u_1) + K_0\cdot P(z \ne t|v = u_1)]\}$$

$$\le \quad E[y(t)] \quad \le$$

$$\sum_{u \in V} P(v = u)\{ \inf_{u_2 \ge u} [E(y|v = u_2,\ z = t)\cdot P(z = t|v = u_2) + K_1\cdot P(z \ne t|v = u_2)]\}.$$

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

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

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 bounds 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 bounds 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 assumption has the same identifying power as does the IV assumption.

Figures 1 through 4 illustrate how the identifying power of the IV and MIV assumptions depend on the way the no-assumptions bound on $E[y(t)|v = u]$ varies with u. In Figure 1 the no-assumptions bound is [0.3, 0.7] for all values of u, so neither the MIV nor the IV assumption has identifying power. In each of Figures 2 through 4, the no-assumptions bound varies sufficiently with u that their intersection contains only one point, namely 0.5. Hence the IV assumption reveals in each case that $E[y(t)|v = u] = 0.5$ for all u.

The identifying power of the MIV assumption varies considerably from Figure 2 to Figure 4. In Figure 2, the no-assumptions bound weakly increases with u, so the MIV assumption has no identifying power. In Figure 3, the no-assumption bound varies non-monotonically with u, so the identifying power of the MIV assumption varies with u. In particular, the MIV assumption has no identifying power at $v = 1$ and at $v = 10$, but substantially narrows the no-assumption bound at $v = 5$. In Figure 4, the no-assumptions bound weakly decreases with u, so the MIV assumption reveals that $E[y(t)|v = u] = 0.5$ for all u.

3.2. Monotone Treatment Selection

Certainly the most prominent IV assumption in the literature is that of *exogenous treatment selection* (ETS), in which the instrumental variable v is the realized treatment z. Then the IV assumption (1) becomes

<u>ETS Assumption</u>: For each $t \in T$,

$$(16) \quad E[y(t)|z = u'] = E[y(t)|z = u], \qquad \forall\, u \in T,\ u' \in T.$$

It is well known that the IV bound (11) reduces to an equality in this case. Observe that $P(z = t|v = u') = P(z = t|z = u') = 1$ if $u' = t$ and equals zero otherwise. Hence (11) becomes

$$(17) \quad E(y|z = t) \le E[y(t)|z = u] \le E(y|z = t).$$

Thus we obtain the equality

$$(18)\quad E[y(t)|z = u] = E(y|z = t).$$

Let us now weaken equation (16) to an inequality. This yields a new assumption which we call *monotone treatment selection* (MTS):

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

$$(19)\quad u_2 \geq u_1 \Rightarrow E[y(t)|z = u_2] \geq E[y(t)|z = u_1].$$

Applying Proposition 1 and Corollary 1 now yield these simple sharp bounds on $E[y(t)|z = u]$ and $E[y(t)]$:

Proposition 1, Corollary 2: Let the MTS Assumption (19) hold. Then

$$(20)\quad \begin{aligned} u < t &\Rightarrow K_0 \leq E[y(t)|z = u] \leq E(y|z = t) \\ u = t &\Rightarrow E[y(t)|z = u] = E(y|z = t) \\ u > t &\Rightarrow E(y|z = t) \leq E[y(t)|z = u] \leq K_1. \end{aligned}$$

and

$$(21)\quad K_0 \cdot P(z < t) + E(y|z = t) \cdot P(z \geq t) \leq E[y(t)]$$
$$\leq K_1 \cdot P(z > t) + E(y|z = t) \cdot P(z \leq t).$$

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

To illustrate the ETS and MTS assumptions, consider the returns to 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.

Figure 5 shows visually the identifying power of the ETS and MTS assumptions. The figure displays no-assumption bounds on $E[y(t)|z = u]$ when $y(t)$ is a binary outcome variable. The no-assumption bound is informative only when $z = t$, in which case the conditional expectation is revealed to equal 0.5. The intersection of the no-assumptions bounds across different values of z also equals 0.5. Thus the ETS assumption implies that $E[y(t)|z = u] = 0.5$ for all u. The no-assumption bound varies non-monotonically with z, so the identifying power of the MIV assumption varies across the subpopulations receiving different treatments. When $z < t$, the MIV lower bound is uninformative and the upper bound equals 0.5. When $z > t$, the MIV upper bound is uninformative and the lower bound equals 0.5.

3.3. Multi-dimensional IV and MIV Assumptions

It is straightforward to combine multiple scalar IV and MIV assumptions. One simply takes the intersection of the bounds obtained under each assumption imposed. An illustration will be given in the empirical analysis of Section 6, where we use AFQT score as an MIV and also assume monotone treatment selection.

We caution the reader that combining multiple scalar assumptions is not the same as imposing one multi-dimensional assumption. To make the point it suffices to consider two scalar covariates, say $v_a \in V_a$ and $v_b \in V_b$, where V_a and V_b are both subsets of the real line. One might assume that v_a and v_b are each scalar IVs, or one might assume that the pair (v_a, v_b) is a two-dimensional IV. The former assumption states that for each $t \in T$,

$$E[y(t)|v_a = u'] = E[y(t)|v_a = u], \quad \text{all } (u, u') \in V_a \times V_a$$
$$E[y(t)|v_b = u'] = E[y(t)|v_b = u], \quad \text{all } (u, u') \in V_b \times V_b .$$

The latter assumption states that

$$E[y(t)|(v_a, v_b) = (u_a', u_b')] = E[y(t)|(v_a, v_b) = (u_a, u_b)],$$
$$\text{all } [(u_a, u_b), (u_a', u_b')] \in (V_a \times V_b) \times (V_a \times V_b).$$

The latter assumption implies the former one. Hence the IV bound (11) formed using (v_a, v_b) as a two-dimensional IV is necessarily a subset of the intersection of the bounds formed using v_a and v_b as two scalar IVs. A direct proof of this can be based on the reasoning in Manski (1994, note 4, p. 167).

Now consider v_a and v_b as MIVs. One might assume that v_a and v_b are each scalar MIVs; that is,

$$E[y(t)|v_a = u_2] \geq E[y(t)|v_a = u_1], \quad \text{all } (u_1, u_2) \in V_a \times V_a \quad \text{s.t. } u_2 \geq u_1$$
$$E[y(t)|v_b = u_2] \geq E[y(t)|v_b = u_1], \quad \text{all } (u_1, u_2) \in V_b \times V_b \quad \text{s.t. } u_2 \geq u_1.$$

Alternatively, one might assume that the pair (v_a, v_b) is a two-dimensional *semi-monotone instrumental variable* (SMIV).

We define an SMIV in the same manner as an MIV except that the set V is assumed only to be semi-ordered rather than ordered. The inequality (2) holds as stated, it being understood that there may exist some pairs of (u_1, u_2) values that are not ordered. In the present case, $u_1 = (u_{a1}, u_{b1})$, $u_2 = (u_{a2}, u_{b2})$, and we define $u_2 \geq u_1$ if and only if $u_{a2} \geq u_{a1}$ and $u_{b2} \geq u_{b1}$. So (v_a, v_b) is an SMIV if

$$E[y(t)|(v_a, v_b) = (u_{a2}, u_{b2})] \geq E[y(t)|(v_a, v_b) = (u_{a1}, u_{b1})],$$

$$\text{all } [(u_{a1}, u_{b1}), (u_{a2}, u_{b2})] \in (V_a \times V_b) \times (V_a \times V_b) \text{ s.t. } u_{a2} \geq u_{a1} \text{ and } u_{b2} \geq u_{b1}.$$

This SMIV assumption does not imply the earlier pair of MIVs, nor vice versa.

Proposition 1 holds as stated for SMIVs, the sup and inf operations being taken over the pairs of (u_1, u) and (u, u_2) values that are ordered. With some increase in notational burden, the analysis in the remainder of this paper can be extended to SMIVs as well, much in the manner that Manski (1997) extends the analysis of monotone treatment response to semi-monotone treatment response.

4. Identification Using MIV and Linear Response Assumptions

The central finding of the classical econometric literature on treatment response is that the response parameter β of the assumed linear response function (3) is identified given an IV assumption (1) and the rank condition that z is not mean independent of v. We present a simple proof taken from Manski (1995, page 152). We then weaken the IV assumption to an MIV assumption and show that this renders β unidentified but bounded.

Let $u_1 \in V$ and $u_2 \in V$ be any two points on the support of the distribution of covariate v. Assumptions (1) and (3) imply that

$$(22)\quad E(\epsilon|v = u_2) = E(\epsilon|v = u_1).$$

Assumption (3) implies that $\epsilon_j = y_j - \beta z_j$ for each person j. Hence

$$(23)\quad E(y - \beta z|v = u_2) = E(y - \beta z|v = u_1).$$

Solving (23) for β yields

$$(24)\quad \beta = \frac{E(y|v = u_2) - E(y|v = u_1)}{E(z|v = u_2) - E(z|v = u_1)},$$

provided that the dominator is non-zero. The rank condition is that this denominator should be non-zero for some pair (u_1, u_2) or, equivalently, that z should not be mean independent of v. Each of the quantities on the right side of (24) is identified and can be nonparametrically estimated. Hence assumptions (1), (3), and the rank condition identify β and make it estimable.

Analyses of treatment response often seek to learn average treatment effects of the form $E[y(t_2)] - E[y(t_1)]$ for specified $t_1 \in T$ and $t_2 \in T$. When the linear response model (3) is assumed, the average treatment effect is $\beta(t_2 - t_1)$. Hence identification of β is equivalent to identification of the average treatment effect.

Now replace the IV assumption with MIV assumption (2). Let V be an ordered set and let $u_2 > u_1$ be two points on the support of v. Assumptions (2) and (3) imply that

$$(25) \quad E(\varepsilon | v = u_2) \geq E(\varepsilon | v = u_1).$$

Assumption (3) still implies that $\varepsilon_j = y_j - \beta z_j$ for each person j. Hence

$$(26) \quad E(y - \beta z | v = u_2) \geq E(y - \beta z | v = u_1).$$

Solving for ß yields the inequality

$$(27a) \quad \beta \leq \frac{E(y|v = u_2) - E(y|v = u_1)}{E(z|v = u_2) - E(z|v = u_1)} \quad \text{if} \quad E(z|v = u_2) - E(z|v = u_1) > 0$$

$$(27b) \quad \beta \geq \frac{E(y|v = u_2) - E(y|v = u_1)}{E(z|v = u_2) - E(z|v = u_1)} \quad \text{if} \quad E(z|v = u_2) - E(z|v = u_1) < 0.$$

This proves

<u>Proposition 2</u>: Let the MIV Assumption (2) and the linear response model (3) hold. Then ß lies in the intersection of the inequalities (27) over $(u_1, u_2) \in V \times V$ such that $u_2 > u_1$. In the absence of other information, this bound is sharp. ■

Proposition 2 yields an informative bound on β if and only if z is not mean independent of v. Thus the rank condition here is the same as when an IV assumption is combined with the linear response model. The bound in Proposition

2 typically does not identify β. The sign of ß may or may not be identified. Inspection of (27) shows that sgn(ß) is identified as negative if there exists a $u_2 > u_1$ such that $E(y|v = u_2) - E(y|v = u_1) < 0$ and $E(z|v = u_2) - E(z|v = u_1) > 0$. Sgn(ß) is identified as positive if there exists a $u_2 > u_1$ such that $E(y|v = u_2) - E(y|v = u_1) < 0$ and $E(z|v = u_2) - E(z|v = u_1) < 0$. Sgn(ß) is not identified if $E(y|v = u_2) - E(y|v = u_1) \geq 0$ for all $u_2 > u_1$.

Proposition 2 yields a particularly simple conclusion in the case of an MTS assumption. Let $v = z$. Then $E(z|v = u_2) - E(z|v = u_1) = u_2 - u_1$. Hence the rank condition necessarily holds and Proposition 2 yields this upper bound on ß:

<u>Proposition 2, Corollary 1</u>: Let the MTS Assumption (19) and the linear response model (3) hold. Then

$$(28) \qquad ß \leq \inf_{(u_2,\, u_1):\, u_2 > u_1} \frac{E(y|z = u_2) - E(y|z = u_1)}{u_2 - u_1}.$$

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

Equation (24) showed that under the ETS assumption, which is the boundary case of MTS, $ß = [E(y|z = u_2) - E(y|z = u_1)]/(u_2 - u_1)$, all $(u_2, u_1): u_2 > u_1$. Corollary 1 reveals that estimates of ß obtained under the ETS assumption are biased upward if the MTS Assumption holds but the ETS Assumption does not.

5. Identification Using MTR and MIV Assumptions

For many years, empirical researchers have applied linear response models of the form (3) 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 (4), stating that response varies monotonically with treatment. The theory of the firm suggests that supply functions slope upward. Consumer theory suggests that demand functions slope downward. Human capital theory suggests that wages increase with years of schooling. In these and other settings, MTR assumptions have a reasonably firm foundation.

The identifying power of an MTR assumption alone, with no other assumptions imposed, has been studied in Manski (1997). Let $t \in T$, let v denote an observed covariate, and let (4) hold. Corollary M1.2 of Manski (1997) gives this sharp bound on the conditional mean response $E[y(t)|v = u]$:

$$(29) \quad E(y|v = u,\ t \geq z)\cdot P(t \geq z|v = u) + K_0\cdot P(t < z|v = u) \quad \leq \quad E[y(t)|v = u]$$
$$\leq \quad E(y|v = u,\ t \leq z)\cdot P(t \leq z|v = u) + K_1\cdot P(t > z|v = u).$$

It is straightforward to combine the MTR assumption with the assumption that v is an IV or an MIV. We simply repeat the derivation of Section 3.1, with the MTR bound (29) replacing the no-assumptions bound (10).

Let v be an IV. Then $E[y(t)|v = u]$ 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 (29) across the elements of V. Any point in this intersection is feasible. Thus, for all $u \in V$, we obtain the common sharp bound

$$(30) \quad \sup_{u' \in V} [E(y|v = u', t \geq z) \cdot P(t \geq z|v = u') + K_0 \cdot P(t < z|v = u')]$$

$$\leq E[y(t)|v = u] \leq$$

$$\inf_{u' \in V} [E(y|v = u', t \leq z) \cdot P(t \leq z|v = u') + K_1 \cdot P(t > z|v = u')].$$

This is also the sharp bound on the marginal mean $E[y(t)]$.

Now let v be an MIV. In this case, $E[y(t)|v = u]$ need not be constant across $u \in V$ but we do have the inequality restriction given earlier in (12), namely

$$u_1 \leq u \leq u_2 \quad \Rightarrow \quad E[y(t)|v = u_1] \leq E[y(t)|v = u] \leq E[y(t)|v = u_2].$$

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 we have

Proposition 3: Let the MIV and MTR Assumptions (2) and (4) hold. Then for each $u \in V$,

$$(31) \quad \sup_{u_1 \le u} [E(y|v = u_1, t \ge z) \cdot P(t \ge z|v = u_1) + K_0 \cdot P(t < z|v = u_1)]$$

$$\le \quad E[y(t)|v = u] \quad \le$$

$$\inf_{u_2 \ge u} [E(y|v = u_2, t \le z) \cdot P(t \le z|v = u_2) + K_1 \cdot P(t > z|v = u_2)].$$

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

The MTR-MIV bound on the marginal mean $E[y(t)]$ is obtained from (31). Recall the application of the law of iterated expectations given in (14), namely

$$E[y(t)] \quad = \quad \sum_{u \in V} P(v = u) \cdot E[y(t)|v = u].$$

Proposition 3 shows that the MTR-MIV lower and upper bounds on $E[y(t)|v = u]$ are weakly increasing in u. Hence the sharp joint lower (upper) bound on $\{E[y(t)|v = u], u \in V\}$ is obtained by setting each of the quantities $E[y(t)|v = u]$, $u \in V$ at its lower (upper) bound as given in (31). Inserting these lower and upper bounds into (14) yields

Proposition 3, Corollary 1: Let the MIV and MTR Assumptions (2) and (4) hold. Then

$$(32)\quad \sum_{u \in V} P(v = u)\ \{ \sup_{u_1 \le u} [E(y|v = u_1,\ t \ge z)\cdot P(t \ge z|v = u_1) + K_0\cdot P(t < z|v = u_1)]\}$$

$$\le\ E[y(t)]\ \le$$

$$\sum_{u \in V} P(v = u)\ \{ \inf_{u_2 \ge u} [E(y|v = u_2,\ t \le z)\cdot P(t \le z|v = u_2) + K_1\cdot P(t > z|v = u_2)]\}.$$

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

In general, the MTR-MIV 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 3 yields this MTR-MTS bound on $E[y(t)|z = u]$:

$$(33)\quad u < t \Rightarrow \sup_{u_1 \le u} E(y|z = u_1) \le E[y(t)|z = u] \le \inf_{u_2 \ge t} E(y|z = u_2)$$

$$u = t \Rightarrow \sup_{u_1 \le t} E(y|z = u_1) \le E[y(t)|z = u] \le \inf_{u_2 \ge t} E(y|z = u_2)$$

$$u > t \Rightarrow \sup_{u_1 \le t} E(y|z = u_1) \le E[y(t)|z = u] \le \inf_{u_2 \ge u} E(y|z = u_2).$$

It follows from the MTR and MTS assumptions that

$$(34) \quad u' \le u \Rightarrow E(y|z = u') = E[y(u')|z = u'] \le E[y(u)|z = u'] \le E[y(u)|z = u] = E(y|z = u).$$

Combining (33) and (34) yields these quite simple MTR-MTS bounds, which are informative even if Y is unbounded:

<u>Proposition 3, Corollary 2</u>: Let the MTR and MTS Assumptions (4) and (19) hold. Then

$$(35) \quad \begin{aligned} u < t &\Rightarrow E(y|z = u) \le E[y(t)|z = u] \le E(y|z = t) \\ u = t &\Rightarrow E[y(t)|z = u] = E(y|z = t) \\ u > t &\Rightarrow E(y|z = t) \le E[y(t)|z = u] \le E(y|z = u) \end{aligned}$$

and

$$(36) \quad \sum_{u < t} E(y|z = u)\cdot P(z = u) + E(y|z = t)\cdot P(z \ge t) \le E[y(t)] \le \sum_{u > t} E(y|z = u)\cdot P(z = u) + E(y|z = t)\cdot P(z \le t).$$

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

Equation (34) suggests a simple test of the joint MTR-MTS hypothesis. If the MTR and the MTS assumptions both hold, then $E(y|z = u)$ must be a weakly increasing function of u. Hence we should reject the MTR-MTS 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 $y(\cdot)$.

6. Empirical Analysis of The Returns to Schooling

6.1. Specification and Data

We now apply the propositions developed in Sections 3 through 5 to the problem of inference on the returns to schooling, discussed earlier in Section 2.2. We have suggested a number of assumptions that are consistent with standard theories of human capital accumulation, productivity, and wages. In Section 2.2 we suggested that achievement measures are credible MIVs. In Section 3.2 we suggested that realized years of schooling is itself a credible MIV, so the MTS Assumption (19) holds. In Section 5 we suggested that wages increase with conjectured years of schooling, so the MTR Assumption (4) holds. In this section we use data from the 1979 cohort of the National Longitudinal Survey of Youth (NLSY) to determine what empirical findings about the returns to schooling emerge when these weak assumptions are imposed. We also report findings when two strong assumptions are imposed, the linear response assumption (3) and the ETS assumption (16).

In its 1979 base year, the NLSY interviewed 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 respondents 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 of the returns to schooling concerns the subpopulation of persons who, in the notation introduced in Section 1 but since left implicit, have the shared observable covariates

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

We observe each respondent's 1994 hourly wage and years of schooling. We also observe the respondent's Armed Forces Qualification Test (AFQT) score obtained when the AFQT was administered during the 1979 and 1980 interviews. In terms of our notation, z is the observed years of schooling and v is the AFQT score, a covariate that measures student achievement. The response variable $y_j(t)$ is the 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 wage.

We examine two distinct features of the returns to schooling. First, following the tradition in the literature, we use log(wage) to measure outcomes and report findings on the expected returns to an additional year of schooling, namely

$$(38)\quad \Delta_1(t) = E\{\log[y(t)]\} - E\{\log[(y(t-1)]\}.$$

Second we examine the effect of an additional year of schooling on the probability that wage exceeds ten dollars per hour; that is,

$$(39)\quad \Delta_2(t) = P[y(t) > 10] - P[y(t-1) > 10]$$

$$= E\{1[y(t) > 10]\} - E\{1[y(t-1) > 10]\}.$$

We report findings for the values $t = 12$, 15, and 16; that is, on the returns to the 12^{th}, 15^{th}, and 16^{th} years of schooling.

The indicator functions in (39) take the values 0 and 1, so we set $K_0 = 0$ and $K_1 = 1$ when applying Propositions 1 and 3 to the problem of inference on $\Delta_2(t)$. The log functions in (38) have unbounded range, so we must set $K_0 = -\infty$ and $K_1 = \infty$ for inference on $\Delta_1(t)$ unless we assume that the distribution of wages has bounded support. With this in mind, we assume that the lowest possible wage is \$4.25 per hour, which was the official minimum wage in 1994. We assume that the highest possible wage is \$150 per hour, which exceeds the highest wage reported by any NLSY respondent in 1994, namely \$138 per hour. Thus we set $K_0 = \log(4.25) = 1.4$ and $K_1 = \log(150) = 5.0$ for inference on $\Delta_1(t)$. A reader who does not accept these support assumptions should interpret us as reporting findings on a trimmed mean of log(wage) rather than the mean itself in those cases where the bound depends on the values of K_0 and K_1. This caveat does not apply to the bounds derived under the linear response model nor to those that combine the MTR and MTS assumptions. As shown in Sections 4 and 5, these bounds do not depend on the values of K_0 and K_1.

Whereas Proposition 2 directly yields sharp bounds on average treatment effects, Propositions 1 and 3 give sharp bounds on the mean outcomes $E\{\log[y(t)]\}$, $E\{\log[y(t-1)]\}$, $P[y(t) > 10]$, and $P[y(t-1) > 10]$. Our objects of interest are the average treatment effects $\Delta_1(t)$ and $\Delta_2(t)$, which take differences of these quantities. It suffices to consider $\Delta_1(t)$, as the same considerations apply to $\Delta_2(t)$.

In the case of Proposition 1, the sharp lower (upper) bound on $\Delta_1(t)$ is obtained by subtracting the lower (upper) bound on $E\{\log[y(t)]\}$ from the upper (lower) bound on $E\{\log[y(t-1)]\}$. In the case of Proposition 3, we may obtain bounds on $\Delta_1(t)$ in the same manner but these bounds are not necessarily sharp. In particular we also know that, under the MTR assumption alone, the lower bound on $\Delta_1(t)$ must be no less than zero (see Manski, 1997, and Pepper, 1997). The bound that we report uses this information as well as the information yielded by Proposition 3.

6.2. Statistical Considerations

The bounds developed in Propositions 1 through 3 and their corollaries are continuous functions of various nonparametrically estimable conditional expectations. Hence any consistent nonparametric regression method may be used to obtain consistent estimates of the bounds. Treating the covariates v (AFQT score) and z (realized years of schooling) as continuous conditioning variables, we use kernel methods to estimate the relevant conditional expectations, much as in Manski, Sandefur, McLanahan, and Powers (1992) and in Pepper (1997). In particular, we use a standard normal kernel with the bandwidths for v and z set, following some experimentation, to the values 10 and 0.8 respectively. We take the suprema and infima required by Propositions 1 through 3 only over the integer values of v and z that are realized in the NLSY data.

Bootstrapping provides an heuristically appealing and computationally tractable way to form confidence intervals for our estimates of bounds on the returns to schooling. The bootstrapped sampling distribution of an estimate 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 (37) equals the empirical distribution of these variables in the sample of 1,257 randomly sampled NLSY respondents.

Beneath each [lower bound, upper bound] estimate, we display the interval

(40) (A, B) ≡

(0.05 quantile of bootstrapped sampling distribution of lower bound estimate, 0.95 quantile of bootstrapped sampling distribution of upper bound estimate).

Let [L, U] denote a [lower bound, upper bound] estimate. By construction, $L \le U$ in all samples. Under the bootstrap assumption that the population and empirical distributions coincide, $Prob(A \le L) \ge 0.95$ and $Prob(U \le B) \ge 0.95$. Now consider the event $(A \le L \cap U \le B)$.

$$(41) \quad Prob(A \le L \cap U \le B) = Prob(A \le L) + Prob(U \le B) - Prob(A \le L \cup U \le B)$$
$$\ge 0.95 + 0.95 - 1 = 0.90.$$

Hence the interval (A, B) gives a conservative bootstrapped 90 percent confidence interval for the bound estimate.

We caution the reader that the available asymptotic theory for bootstrapped sampling distributions does not appear to immediately cover our bound estimates, which are functions of various kernel estimates. It would be useful to extend the asymptotic theory of the bootstrap to cover such estimates. It would also be useful to analyze the finite sample bias of our bound estimates. Finite sample bias may be a particular concern in those cases where the bound estimate is obtained by taking the suprema or infima of several kernel estimates. We do not attempt to resolve these statistical questions in the present paper, which is primarily concerned with identification.

We also caution the reader not to confuse the statistical questions that arise in bound estimation with those examined in the large literature on the sampling distributions of point estimates obtained under the classical econometric model combining an IV assumption with the linear response assumption (3). Consider, in particular, the matter of *weak instruments* examined by Nelson and Startz (1990), Bound, Jaeger, and Baker (1995), Staiger and Stock (1997), and others. The classical econometric model is formally identified given a weak instrument. The problem is that the sampling distribution of the usual estimate may be poorly behaved.

In our setting, an instrumental variable may be weak in two different senses. One, which is not a concern in the classical literature, is that the IV may have little identifying power, in that the IV bound improves little on the no-assumptions bound. The other sense is that the sampling distribution of an estimate of the IV bound may be poorly behaved. It is an open question whether classical IV point estimates and our nonparametric IV (or MIV) bound estimates have poorly behaved sampling distributions in similar situations.

6.3. Findings

Tables 1 and 2 report our findings on $\Delta_1(t)$ and $\Delta_2(t)$ respectively, for t = 12, 15, 16. We shall focus our discussion on the findings in Table 1 for $\Delta_1(12)$, the mean log(wage) return to the 12^{th} year of schooling. The patterns for the other estimates are similar.

The obvious striking finding is how the widths of the estimated bounds vary with the assumptions imposed. First we should report some qualitatively meaningful but quantitatively uninteresting findings. The no-assumptions bounds show that the data alone reveal almost nothing about the returns to schooling as measured by the difference in mean log(wage) between 11 and 12 years of schooling. All we can say based on the data alone is that the return to the 12^{th} year of schooling is in the interval [-3.052, 2.497]. The assumption that AFQT score is an MIV has very little identifying power, the bound being [-2.841, 2.320]. The MTS assumption, namely that the persons who select more years of schooling have higher mean log(wage) has somewhat more power, the bound now being [-2.384, 1.362]. The MTR assumption, namely that wages increase weakly with conjectured schooling, yields a yet narrower bound of [0, 1.333], but this range of possible values of $\Delta_1(12)$ is still very wide relative to the conventional wisdom about the magnitude of the returns to schooling.

Quantitatively interesting findings begin to emerge when the MTR and MTS assumptions are combined. The MTR-MTS bound on the mean log(wage) effect of the 12^{th} year of schooling is [0.0, 0.199]. In Section 5 we observed that the MTR-MTS assumption is a testable hypothesis, which should be rejected if $E(y|z = u)$ is not weakly increasing in u. Our nonparametric estimates of mean log(wage) do increase in realized schooling, the estimates being

$E(y|z = 9) = 2.269$ $E(y|z = 10) = 2.339$ $E(y|z = 11) = 2.470$

$E(y|z = 12) = 2.505$ $E(y|z = 13) = 2.554$ $E(y|z = 14) = 2.644$

$E(y|z = 15) = 2.773$ $E(y|z = 16) = 2.848$ $E(y|z = 17) = 2.866$

$E(y|z = 18) = 2.946$.

Hence the MTR-MTS assumption is consistent with the empirical evidence.

Under the joint MIV-MTS-MTR assumption, the bound on the mean log(wage) effect of the 12^{th} year of schooling is narrowed further to [0, 0.126]. Thus, by combining three weak assumptions, we are able to conclude -- subject to considerations of finite-sample precision and bias -- that the 12^{th} year of schooling increases mean log(wage) by 0.126 at most. This is a substantively interesting upper bound on the returns to schooling.

We do not obtain a substantively interesting lower bound on the returns to schooling. The lower bound under the joint MIV-MTS-MTR assumption is zero, which is implied algebraically by the MTR Assumption alone. Proposition 3, Corollary 2 shows that combining the MTR Assumption and the MTS Assumption cannot improve on the MTR lower bound of zero. MIV assumptions other than the MTS one can, in principle, yield positive lower bounds on treatment effects. In our application, however, the AFQT variable turns out not to have this kind of identifying power.

The conservative ninety percent bootstrapped confidence interval for the MIV-MTR-MTS bound estimate is (0, 0.148). This interval is just 17 percent wider than the bound estimate itself. Inspection of Tables 1 and 2 shows that the confidence intervals displayed below the bound estimates are generally not much wider than the estimates themselves. Thus identification appears to be the dominant problem in our efforts to infer the returns to schooling. Subject to our earlier caveat about the absence of a demonstrated theoretical foundation for our applications of the bootstrap, we find that the sampling precision of the estimates is no more than a second-order concern.

Much of the literature on the returns to schooling assumes that treatment selection is exogenous and that the log(wage) response function is linear in years of conjectured schooling with a common slope parameter across persons, as in (3). Some labor economists combine the linear response model with instrumental variable assumptions or other identifying restrictions. The estimates of $\Delta_1(t)$ reviewed by Card (1994) typically lie between 0.07 and 0.09. More recently, Blackburn and Neumark (1995),using NLSY data and assuming that family background characteristics are IVs, estimate the return to an additional year of schooling to be 0.08.

The ETS assumption alone suffices to identify the returns to schooling. Our estimate of $\Delta_2(12)$ under this assumption is 0.035. When the ETS assumption is combined with the linear response assumption, the estimate is 0.077. When the linear response model is combined with the assumption that AFQT score is an MIV, we find that an additional year of schooling at most increases mean log(wage) by 0.068. When we combine the linear response model with the MTS assumption, we find that an additional year of schooling at most increases mean log(wage) by 0.018.

Our estimated upper bound of 0.018 when the linear response model and the MTS assumption are combined lies well below the point estimates of the returns to schooling reported in the literature reviewed by Card (1994). The reader must draw his or her own conclusions from the inconsistency of our finding with those in the literature. Under the hypothesis that the linear response model is correct, the inconsistency implies either that the MTS assumption is incorrect or that the IV identifying assumptions made in the literature are incorrect. However it may be that the linear response model is incorrect. If so, the inconsistency of the findings carries no implications for the validity of the MTS assumption or of the IV assumptions in the literature.

7. Conclusion

This paper has introduced the general idea of a monotone instrumental variable and the important special case of monotone treatment selection.

Propositions 1 through 3 have characterized the identifying power of an MIV assumption when imposed alone, when combined with the linear response model, and when combined with the assumption of monotone treatment response. Consideration of the problem of inference on the returns to schooling has demonstrated that MIV assumptions may be credible in situations where IV assumptions are controversial. Our empirical analysis of the returns to schooling has shown that MIV assumptions are easy to apply and has given a sense of their identifying power in practice.

It is easy to think of variations on the MIV theme that warrant study and that may prove useful in empirical research. One direction for future work would be to combine the MIV idea with the idea of contaminated instruments introduced by Hotz, Mullins, and Sanders (1997). They assume that the observed population is a probability mixture of two subpopulations, one that satisfies an IV assumption and another that does not. A *contaminated MIV* assumption would hold if the first subpopulation satisfies an MIV assumption instead.

A second direction for future work would be to begin from the statistical independence assumption that holds in classical randomized experiments, namely

Statistical Independence Assumption: For each value of w and all $(u, u') \in (V \times V)$,

(42) $P[y(\cdot)|w, v = u'] = P[y(\cdot)|w, v = u]$.

Weakening the equality in (42) to an inequality might be interpreted as asserting a weak stochastic dominance assumption, namely

Weak Stochastic Dominance Assumption: Let V be an ordered set. For each $t \in T$, each value of w, all $(u_1, u_2) \in (V \times V)$ such that $u_2 \geq u_1$, and all $c \in R$,

$$(43)\quad P[y(t) > c|w, v = u_2] \geq P[y(t) > c|w, v = u_1].$$

Whereas statistical independence implies the IV assumption, weak stochastic dominance implies the MIV assumption. We presently have only a limited understanding of the identifying power of the statistical independence assumption. Considering the case of two treatments and a binary outcome, Balke and Pearl (1997) show that the sharp bound under statistical independence solves a certain linear programming problem. Analysis of the identifying power of the weak stochastic dominance assumption may pose a challenging task.

Another variation on the MIV theme would be to weaken the mean independence of the standard IV assumption to some form of approximate mean independence. One way to formalize this notion would be to assert a *fuzzy instrumental variable* (FIV) assumption of the form

FIV Assumption: Covariate v is a fuzzy instrumental variable if, for each $t \in T$, each value of w, all $(u, u') \in (V \times V)$, and a specified $C \geq 0$,

$$(44)\quad \|E[y(t)|w, v = u'] - E[y(t)|w, v = u]\| \leq C.$$

The constant C, which is specified by the researcher, gives the maximal variation of the conditional expectation on the covariate space V and so controls the degree of *fuzziness* that the researcher wishes to permit. Setting $C = 0$ yields the IV Assumption. Setting $C = \infty$ imposes no restriction at all.

Table 1: Effect of Years of Schooling on Mean Log(Wage)

Prior Information	Years of Schooling		
	t = 12 years	t = 15 years	t = 16 years
Nonparametric Analysis			
I. No MIV			
No Assumptions	[-3.052, 2.497] (-3.082 , 2.551)	[-3.312, 3.376] (-3.344, 3.399)	[-3.210, 3.099] (-3.242, 3.137)
MTS	[-2.384, 1.362] (-2.418 , 1.418)	[-1.563, 1.515] (-1.593, 1.562)	[-1.590, 1.211] (-1.623, 1.250)
MTR	[0, 1.333] (0, 1.362)	[0, 2.114] (0, 2.147)	[0, 2.151] (0, 2.181)
MTR-MTS	[0, 0.199] (0, 0.222)	[0, 0.255] (0, 0.299)	[0, 0.256] (0, 0.296)
ETS	0.035 (0.024 , 0.047)	0.129 (0.091 , 0.169)	0.075 (0.043, 0.102)
II. AFQT Score as MIV			
No Other Assumptions	[-2.841, 2.320] (-2.877 , 2.370)	[-3.247, 3.340] (-3.297, 3.365)	[-3.206, 2.762] (-3.236, 2.832)
MTS	[-2.371, 1.217] (-2.398, 1.273)	[-1.494, 1.393] (-1.535, 1.436)	[-1.501, 1.117] (-1.543, 1.162)
MTR	[0, 1.010] (0, 1.059)	[0, 1.153] (0, 1.237)	[0, 1.250] (0, 1.334)
MTR-MTS	[0 , 0.126] (0 , 0.148)	[0, 0.162] (0, 0.197)	[0, 0.167] (0, 0.206)
Linear Response Model			
MIV	[$-\infty$, 0.068] ($-\infty$, 0.080)	[$-\infty$, 0.068] ($-\infty$, 0.080)	[$-\infty$, 0.068] ($-\infty$, 0.080)
MTS	[$-\infty$, 0.018] ($-\infty$, 0.038)	[$-\infty$, 0.018] ($-\infty$, 0.038)	[$-\infty$, 0.018] ($-\infty$, 0.038)
ETS	0.077 (0.069, 0.085)	0.077 (0.069, 0.085)	0.077 (0.069, 0.085)

Note: Bound estimates are in brackets. Conservative boostrapped ninety percent confidence intervals for the bound estimates are in parentheses.

Table 2: Effect of Years of Schooling on Prob(Wage > $10 per Hour)

Prior Information	Years of Schooling		
	t = 12 years	t = 15 years	t = 16 years
Nonparametric Analysis			
I. No MIV			
No Assumptions	[-0.699, 0.862] (-0.722, 0.877)	[-0.952, 0.931] (-0.959, 0.941)	[-0.828, 0.948] (-0.844, 0.956)
MTS	[-0.321 , 0.216] (-0.343, 0.236)	[-0.554, 0.555] (-0.575, 0.579)	[-0.611, 0.597] (-0.631, 0.623)
MTR	[0, 0.763] (0, 0.783)	[0, 0.508] (0, 0.529)	[0, 0.483] (0, 0.506)
MTR-MTS	[0, 0.117] (0, 0.138)	[0 , 0.129] (0, 0.156)	[0, 0.132] (0, 0.160)
ETS	0.032 (0.015, 0.047)	0.071 (0.040, 0.101)	0.037 (0.018, 0.061)
II. AFQT Score as MIV			
No Other Assumptions	[-0.555, 0.821] (-0.582, 0.841)	[-0.948, 0.923] (-0.954, 0.932)	[-0.840, 0.928] (-0.856, 0.941)
MTS	[-0.292, 0.167] (-0.322, 0.185)	[-0.575, 0.556] (-0.606, 0.585)	[-0.631, 0.615] (-0.662, 0.642)
MTR	[0, 0.648] (0, 0.689)	[0, 0.295] (0, 0.334)	[0, 0.294] (0, 0.336)
MTR-MTS	[0, 0.071] (0., 0.092)	[0, 0.087] (0, 0.113)	[0, 0.106] (0, 0.130)

Note: Bound estimates are in brackets. Conservative boostrapped ninety percent confidence intervals for the bound estimates are in parentheses.

References

Angrist, J. and A. Krueger (1991), "Does Compulsory Schooling Affect Schooling and Earnings?" Quarterly Journal of Economics 106, 979-1014.

Balke, A. and J. Pearl (1997), "Bounds on Treatment Effects from Studies With Imperfect Compliance," Journal of the American Statistical Association, 92, 1171-1177.

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.

Bound, J., D. Jaeger, and R. Baker (1995), "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," Journal of the American Statistical Association, 90, 443-450.

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," NBER Working Paper 4483, National Bureau of Economic Research, Cambridge, Mass.

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: The 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.

Griliches, Z. (1977), "Estimating the Returns to Schooling: Some Econometric Problems," Econometrica 45, 1-22.

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

Horowitz, J. and C. Manski (1995), "Identification and Robustness with Contaminated and Corrupted Data," Econometrica, 63, 281-302.

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.

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. (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. (1995), Identification Problems in the Social Sciences, Cambridge, Mass.: Harvard University Press.

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

Manski, C., G. Sandefur, S. McLanahan, and D. Powers (1992), "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation," Journal of the American Statistical Association, 87, 25-37.

Nelson, C. and R. Startz (1990), "The Distribution of the Instrumental Variable Estimator and its *t*-ratio When the Instrument is a Poor One," Journal of Business, 63, 5125-5140.

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.

Robins, J. and S. Greenland (1996), "Comment on Angrist, Imbens, and Rubin's 'Identification of Causal Effects Using Instrumental Variables'," Journal of the American Statistical Association, 91, 456-458.

Staiger, D. and J. Stock (1997), "Instrumental Variables Regression with Weak Instruments," Econometrica, 65, 557-586.

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

Figure 1

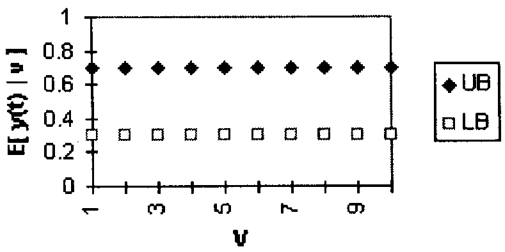

Figure 2

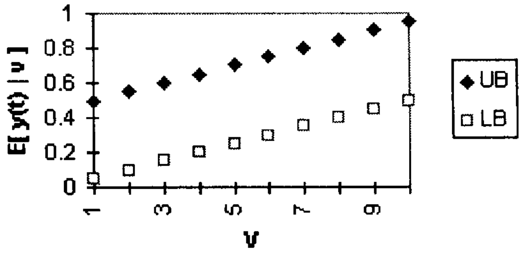

Figure 3

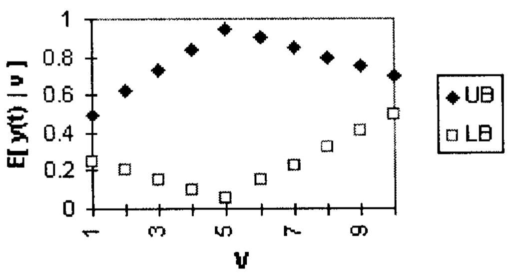

Figure 4

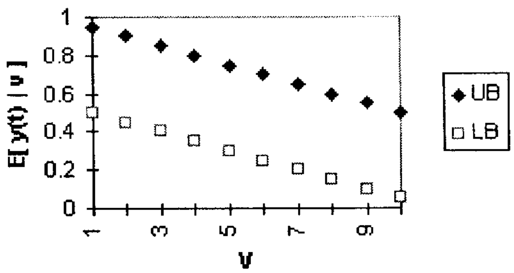

Figure 5