MORE ON MONOTONE INSTRUMENTAL VARIABLES

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

and

John V. Pepper
Department of Economics
University of Virginia

Revised: August 2008

Abstract

Econometric analyses of treatment response often use *instrumental variable* (IV) assumptions to identify treatment effects. The traditional IV assumption holds that mean response is constant across the sub-populations of persons with different values of an observed covariate. Manski and Pepper (2000) introduced *monotone instrumental variable* (MIV) assumptions, which replace equalities with weak inequalities. This paper presents further analysis of the MIV idea. We use an explicit response model to enhance understanding of the content of MIV and traditional IV assumptions. We study the identifying power of MIV assumptions when combined with the *homogeneous linear response* assumption maintained in many studies of treatment response. We also consider estimation of MIV bounds, with particular attention to finite-sample bias.

This paper was prepared for the tenth anniversary issue of the *Econometric Journal*. Our research on monotone instrumental variables (MIVs) was first circulated in 1998, the year that the journal began publication. We are grateful for this opportunity to report further findings on MIVs and, in doing so, to mark the tenth anniversary of both the journal and the subject. We have benefitted from the comments of a referee. Manski's research was supported in part by NSF Grant SES-0549544.

1. Introduction

Econometric analyses of treatment response often use *instrumental variable* (IV) assumptions to identify treatment effects. The traditional IV assumption holds that mean response is constant across the sub-populations of persons with different values of an observed covariate. The credibility of this assumption is often a matter of considerable disagreement, as evidenced by frequent debates about whether some covariate is a "valid instrument." There is therefore good reason to consider weaker but more credible assumptions.

To this end, Manski and Pepper (2000) introduced *monotone instrumental variable* (MIV) assumptions. An MIV assumption weakens the traditional IV assumption by replacing its equality of mean response across sub-populations with a weak inequality. We studied the identifying power of MIV assumptions when imposed alone and when combined with the assumption of *monotone treatment response* (MTR). We reported an empirical application using MIV and MTR assumptions to bound the wage returns to schooling.

This paper presents further analysis of the MIV idea. We draw in part on previously unpublished research that appeared in our original working paper on MIVs (Manski and Pepper, 1998) but that was not included in our 2000 *Econometrica* article. We also discuss statistical issues associated with estimation of MIV bounds.

As prelude, Section 2 sets up basic concepts and notation, summarizes the main analytical findings of Manski and Pepper (2000), and describes subsequent findings. Section 3 uses an explicit response model to enhance understanding of the content of MIV and traditional IV assumptions. 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 and other studies of production. Section 3 is a revised version of Manski and Pepper (1998, Section 2).

Section 4 studies the identifying power of MIV assumptions when combined with the *homogeneous linear response* (HLR) assumption maintained in many studies of treatment response. We think that the HLR assumption is rarely credible; hence, we do not endorse its use in practice. However, widespread application of the assumption makes it important that researchers understand its implications. It has been common to combine the HLR assumption with a traditional IV assumption to achieve point-identification of treatment effects. We show that combining the HLR assumption with an MIV assumption yields a bound on treatment effects. Section 4 is a revised and extended version of Manski and Pepper (1998, Section 4).

Section 5 considers estimation of the bounds reported in Sections 2 and 4. An important statistical concern, noted but not analyzed in Manski and Pepper (2000), is that analog estimates of the bounds have finite-sample bias that make them tend to be narrower than the true bounds. We explain this bias, give Monte Carlo evidence on its magnitude, and describe the bias-correction procedure of Kreider and Pepper (2007). We also show how the so-called *weak-instruments* problem that arises in analog estimation under HLR and IV assumptions manifests itself when HLR and MIV assumptions are combined.

2. Background

2.1. Concepts and Notation

We use the same formal setup as Manski and Pepper (2000). 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 $y_j(\cdot)$: $T \rightarrow Y$ mapping the mutually exclusive and exhaustive treatments $t \in T$ into real-valued outcomes $y_j(t) \in Y$. The outcome space Y has greatest lower bound $K_0 \equiv \inf Y$ and least upper bound $K_1 \equiv \sup Y$.

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 define an MIV assumption. 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 $w \in W$, the mean value of $y(t)$ is the same in all of the sub-populations $(w, v = u)$, $u \in V$. Thus,

<u>IV Assumption</u>: Covariate v is an instrumental variable in the sense of mean-independence if, for each $t \in T$, each $w \in W$, and all $(u_1, u_2) \in V \times V$,

$$(1) \qquad E[y(t)|w, v = u_2] = E[y(t)|w, v = u_1].$$

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

<u>MIV Assumption</u>: 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 $w \in W$, and all $(u_1, u_2) \in V \times V$ such that $u_2 \geq u_1$,

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

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

<u>ETS Assumption</u>: For each $t \in T$, each $w \in W$, and all $(u_1, u_2) \in T \times T$,

$$(3) \qquad E[y(t)|w, z = u_2] = E[y(t)|w, z = u_1].$$

Weakening equation (3) to an inequality yields the special MIV assumption that we call *monotone treatment selection* (MTS):

<u>MTS Assumption</u>: Let T be an ordered set. For each $t \in T$, each $w \in W$, and all $(u_1, u_2) \in T \times T$ such that $u_2 \geq u_1$,

$$(4) \qquad u_2 \geq u_1 \Rightarrow E[y(t)|w, z = u_2] \geq E[y(t)|w, z = u_1].$$

The MTS assumption should not be confused with the *monotone treatment response* (MTR) assumption of Manski (1997). This is

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

$$(5) \qquad t_2 \geq t_1 \Rightarrow y_j(t_2) \geq y_j(t_1).$$

2.2. Findings

This section summarizes the main analytical findings of Manski and Pepper (2000). To simplify the exposition, we henceforth leave implicit the conditioning on w maintained in the definitions of MIVs.

MIV Assumptions

Consider an MIV assumption alone, not combined with other assumptions. Proposition 1 gives sharp bounds on the conditional mean responses $E[y(t)|v = u], u \in V$ and the marginal mean $E[y(t)]$. These bounds are informative if the outcome space Y has finite range $[K_0, K_1]$. The MIV bounds have particularly simple forms in the case of monotone treatment selection.

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

$$(6)\quad \sup_{u_1 \leq u} E(y|v = u_1, z = t)\cdot P(z = t|v = u_1) + K_0\cdot P(z \neq t|v = u_1)$$

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

$$\inf_{u_2 \geq u} E(y|v = u_2, z = t)\cdot P(z = t|v = u_2) + K_1\cdot P(z \neq t|v = u_2).$$

It follows that

$$(7)\quad \sum_{u \in V} P(v = u)\{ \sup_{u_1 \leq u} E(y|v = u_1, z = t)\cdot P(z = t|v = u_1) + K_0\cdot P(z \neq t|v = u_1)\}$$

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

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

Let the MTS Assumption (4) hold. Then bound (6) reduces to

$$(8)\quad u < t \Rightarrow K_0 \leq E[y(t)|z = u] \leq E(y|z = t)$$

$$u = t \Rightarrow E[y(t)|z = u] = E(y|z = t)$$

$$u > t \;\Rightarrow\; E(y|z = t) \;\le\; E[y(t)|z = u] \;\le\; K_1.$$

It follows that

$$(9) \qquad K_0 \cdot P(z < t) + E(y|z = t) \cdot P(z \ge t) \;\le\; E[y(t)] \;\le\; K_1 \cdot P(z > t) + E(y|z = t) \cdot P(z \le t).$$

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

The basic finding in Proposition 1 is inequality (6), from which the other findings are derived. The lower bound in (6) is the supremum of the no-assumptions bounds on $E[y(t)|v = u_1]$ over all $u_1 \le u$. The upper bound is the infimum of the no-assumptions bounds on $E[y(t)|v = u_2]$ over $u_2 \ge u$. Hence, the MIV bound on $E[y(t)|v = u]$ is a subset of the no-assumption bound on $E[y(t)|v = u]$. The MIV assumption has no identifying power if the no-assumptions lower and upper bounds on $E[y(t)|v = u]$ weakly increase with u. Otherwise, it has identifying power in the formal sense that at least some MIV bounds are proper subsets of the corresponding no-assumptions bounds.

Bound (6) is a superset of the Manski (1990) IV bound on $E[y(t)|v = u]$, which is the intersection of the no-assumptions bounds on $E[y(t)|v = u_1]$ over all $u_1 \in V$. The MIV and IV bounds coincide if the no-assumptions bounds on $E[y(t)|v = u]$ weakly decrease with u. In this case, the MIV and IV assumptions have the same identifying power.

MIV and MTR Assumptions

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 informative only when Y has finite range, the MTS-MTR bounds are informative even if Y is unbounded. Proposition 2 gives these results.

Proposition 2: Let the MIV and MTR Assumptions (2) and (5) hold. Then for each $u \in V$,

$$(10) \quad \sup_{u_1 \le u} E(y|v = u_1, t \ge z)\cdot P(t \ge z|v = u_1) + K_0\cdot P(t < z|v = u_1)$$

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

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

It follows that

$$(11) \quad \sum_{u \in V} P(v = u) \; \{ \sup_{u_1 \le u} E(y|v = u_1, t \ge z)\cdot P(t \ge z|v = u_1) + K_0\cdot P(t < z|v = u_1)\}$$

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

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

Let the MTS and MTR Assumptions (4) and (5) hold. Then bound (10) reduces to

$$(12) \quad \begin{aligned} u < t \; &\Rightarrow \; E(y|z = u) \; \le \; E[y(t)|z = u] \; \le \; E(y|z = t) \\ u = t \; &\Rightarrow \; E[y(t)|z = u] \; = \; E(y|z = t) \\ u > t \; &\Rightarrow \; E(y|z = t) \; \le \; E[y(t)|z = u] \; \le \; E(y|z = u). \end{aligned}$$

It follows that

$$(13) \quad \sum_{u<t} E(y|z=u)\cdot P(z=u) + E(y|z=t)\cdot P(z \geq t) \;\leq\; E[y(t)]$$

$$\leq \sum_{u>t} E(y|z=u)\cdot P(z=u) + E(y|z=t)\cdot P(z \leq t).$$

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

2.3. Subsequent Findings

The analytical findings described above have been applied in several empirical studies of treatment response. As mentioned earlier, we reported MIV bounds on the wage returns to schooling. Subsequently, Gonzáles (2005) has studied the wage returns to English proficiency. Gerfin and Schellhorn (2006) have studied the effect of deductibles in health care insurance on doctor visits. Kreider and Hill (2008) have studied the effect of health insurance coverage on health care utilization.

Manski and Pepper (2000) proposed MIVs specifically to weaken the traditional mean-independence IV assumption in analysis of treatment response. However, our broad idea was to enhance the credibility of empirical research by replacing traditional distributional equality assumptions with weak inequalities. Other manifestations of this idea have developed subsequently.

The findings of our 2000 article apply immediately to inference on the population mean of an outcome from a random sample with missing outcome data. Let T = {0, 1}, with t = 1 indicating that the outcome is observable and t = 0 that is it unobservable. Let $y_j(1)$ be the outcome of interest. Let $z_j = 1$ if $y_j(1)$ is observable and $z_j = 0$ otherwise. With these definitions, Proposition 1 gives the sharp bound on E[y(1)] when a covariate v satisfies MIV assumption (2). This application of the MIV idea is developed in

Manski (2003, Section 2.5), with the notational simplification that y(1) here is called y there.

Blundell *et al.* (2007) apply the missing-outcome version of the MIV assumption to inference on wages. The outcome of interest is wage, which is observable when a person works but not otherwise. Blundell *et al.* perform inference when a traditional statistical-independence form of IV assumption is replaced by a weak stochastic dominance assumption. To accomplish this, they apply Proposition 1 to the distribution function $P[y(1) \le r] = E\{1[y(t) \le r]\}$, all $r \in R^1$. This work partially addresses the Manski and Pepper (2000) call for research on MIVs that replaces the statistical-independence assumption of classical randomized experiments with weak stochastic dominance. Blundell *et al.* also contribute new analytical findings on identification of quantiles and inter-quantile differences.

MIV assumptions may also be applied when a researcher faces a data problem other than missing outcomes. Kreider and Pepper (2007, 2008) consider inference on a conditional mean when the data problem is partial misreporting of either the conditioning variable or the outcome variable. They maintain Assumption (2) for a given covariate v and obtain a version of Proposition 1 in which the lower (upper) bound in inequality (6) is replaced by the supremum (infimum) of several v-specific lower (upper) bounds that hold in the misreporting context.

Yet another related line of work begins with traditional parametric econometric models in which the parameters of interest solve a set of moment equations, and replaces these equations with a set of weak moment inequalities. Although replacement of parametric moment equations with inequalities is very much within the broad MIV theme, the approaches used to study partial identification of parametric models are quite different from and more complex than the transparent nonparametric analysis of Propositions 1 and 2. See Manski and Tamer (2002), Chernozhukov, Hong, and Tamer (2007), and Rosen (2006).

3. What Are IV And MIV Assumptions?

The concepts introduced in Section 2 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 3.1 sets out this broader framework. Section 3.2 uses it to develop a lemma suggesting MIV assumptions that might credibly be imposed when analyzing the returns to schooling. Section 3.3 discusses other applications of the lemma to production analysis.

3.1. Treatments and Covariates

The discussion of Section 2 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 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 to denote a covariate value associated with person j. We have given no notation for conjectural values of covariates.

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 $T \times S$ and introduce a generalized response function $y^*_j(\cdot, \cdot)$: $T \times S \rightarrow Y$ mapping elements of $T \times S$ into outcomes. Now, for each $(t, s) \in T \times S$, $y^*_j(t, s)$ is the outcome that person j would experience if she were to receive the conjectural treatment pair (t, s). The treatment pair realized by person j is (z_j, ζ_j) and her realized outcome is $y_j = y^*_j(z_j, \zeta_j)$, where $\zeta_j \in S$ is the realized value of s. The response function $y_j(\cdot)$: $T \rightarrow Y$ introduced as a primitive in Section 2 is now a derived *sub-response function*, obtained by evaluating $y^*_j(\cdot, \cdot)$ with its second argument set at the realized treatment value ζ_j. That is,

(14) $\quad y_j(\cdot) \equiv y^*_j(\cdot, \zeta_j)$.

In this broadened framework, covariate ζ_j is the realized value of treatment s.

With this as background, observe that the familiar statement "variable v does not affect response" has two distinct formal interpretations. One interpretation is that v is an IV, as defined in (1). The other is that outcomes are constant under conjectural variations in v. Under the latter interpretation, S = V and

(15) $\quad y^*_j(t, u) = y^*_j(t, v_j) = y_j(t)$

for all $j \in J$ and $(t, u) \in T \times V$. Assumption (15) has not yet been named. We call it 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, as defined in (2). The other is that outcomes vary monotonically under conjectural variations in v. Under the latter interpretation, S = V and

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

for all $j \in J$ and $t \in T$. This is an MTR assumption.

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 "variable 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.

3.2. Researcher-Measured Ability and 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 mean wage E[y(t)] that would occur if a randomly drawn member of the population were to receive t years of schooling. Researchers often use personal, family, and environmental attributes as instrumental variables for years of schooling. Yet the validity of whatever IV assumption may be imposed seems inevitably to be questioned.

Some formal analysis using the concepts of Section 3.1 suggests that school grades, test scores, and other measures of ability or achievement that are commonly observed by researchers are plausible MIVs for inference on the returns to schooling. For succinctness, we refer below to measures of ability rather than to measures of ability or achievement.

Lemma: Let person j's earning function be

$$(17) \qquad y_j^*(t, s) = g(t, s, \epsilon_j).$$

Here $t \in T$ is years of schooling and $s \in S$ is an ordered measure of ability that is observable by employers. The quantity ϵ_j is person j's realization of another variable that takes values in a space E and that is observable by employers. Assume that, for each $(t, \epsilon) \in T \times \mathrm{E}$, the sub-response function $g(t, \cdot, \epsilon)$ is weakly increasing on S. Thus, wage increases with employer-measured ability. Let $v \in V$ be an ordered measure of ability that is observable by a researcher.

Let ζ be the realized value of employer-measured ability. Assume that $P(\zeta | v)$ is weakly increasing in v in the sense that $P(\zeta | v = u_2)$ weakly stochastically dominates $P(\zeta | v = u_1)$ when $u_2 \geq u_1$. Assume that ϵ is statistically independent of (ζ, v). Then v is an MIV. ■

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

$$
\begin{aligned}
E[y(t) | v = u_2] - E[y(t) | v = u_1] &= E[y^*(t, \zeta) | v = u_2] - E[y^*(t, \zeta) | v = u_1] \\
&= E[g(t, \zeta, \epsilon) | v = u_2] - E[g(t, \zeta, \epsilon) | v = u_1] \\
&= \int g(t, \zeta, \epsilon) dP(\zeta, \epsilon | v = u_2) - \int g(t, \zeta, \epsilon) dP(\zeta, \epsilon | v = u_1) \\
&= \int [\int g(t, \zeta, \epsilon) dP(\zeta | v = u_2) - \int g(t, \zeta, \epsilon) dP(\zeta | v = u_1)] dP(\epsilon) \\
&\geq 0.
\end{aligned}
$$

The first and second equalities apply (14) and (17). The third equality writes the expectations explicitly as integrals. The fourth equality applies the assumption that ϵ is statistically independent of (ζ, v). The final inequality applies the assumptions that $g(t, \cdot, \epsilon)$ is monotone and that $P(\zeta | v = u_2)$ weakly dominates $P(\zeta | v = u_1)$.

Q. E. D.

The lemma does not require that employers or researchers measure ability accurately. Nor does it require that employers and researchers observe the same measure of ability. It only requires that (i) wage increases with employer-measured ability and (ii) researcher-measured ability be a weakly positive predictor of employer-measured ability, in the sense of stochastic dominance. These are understandable and plausible assumptions. Indeed, we think the assumption that $P(\zeta | v)$ weakly increases in v in the sense of stochastic dominance appropriately formalizes what many empirical researchers have in mind when they state that some observed variable v is a "proxy" for an unobserved variable ζ.

Although researcher-measured ability provides a credible MIV for inference on the returns to schooling, we found in our own application that this MIV has little identifying power. Analyzing data from the National Longitudinal Study of Youth, Manski and Pepper (1998) reported no-assumptions bounds on the returns to schooling as well as bounds that use a respondent's score on the Armed Forces Qualifying Test (AFQT) as an MIV. We found that the no-assumptions bounds computed for different AFQT scores are nearly monotone increasing with the score. As a consequence, the MIV bounds were only slightly narrower than the no-assumptions bounds.

3.3. Other Applications to Production Analysis

For concreteness, we used the returns to schooling to motivate the lemma developed in Section 3.2. We observe here that the lemma has other applications to production analysis.

In abstraction, let equation (17) give the production function for agent j, who might be a firm producing a commodity, a person investing in human capital, or another entity. Let t and s be two conjectural factors of production whose realized values (z_j, ζ_j) are either chosen by the agent or predetermined. Let ϵ_j be a production shock of some kind. For example, in agricultural production, $g(t, s, \epsilon)$ might be crop output per acre, which varies with seed input t, planting effort s, and weather quality ϵ following planting.

Suppose that, for each agent j, a researcher observes the realized output y_j and input z_j. The researcher does not observe input ζ_j or the shock ϵ_j, but he does observe a "proxy" v_j for ζ_j. For example, in the agricultural setting, the researcher might observe crop output y_j, seed input z_j, and labor hours v_j allocated to planting. However, he might not observe cultivation effort and weather quality.

Suppose the researcher finds it credible to assume that $P(\zeta | v)$ weakly increases in v in the sense of stochastic dominance and that ϵ is statistically independent of (ζ, v). For example, these assumptions are credible in the agricultural setting described above. Then the lemma of Section 3.2 shows that v is an MIV.

4. Combining MIV and HLR Assumptions

Classical econometric analysis of treatment response, as codified in the literature on linear simultaneous equations (Hood and Koopmans, 1953), supposes that the outcome space Y is the entire real line and combines an IV assumption of form (1) with the homogeneous-linear-response assumption

$$(18) \qquad y_j(t) = \beta t + \delta_j.$$

Here treatments are real-valued, δ_j is an unobserved covariate, and β is a slope parameter that takes the same value for all members of the population. The central finding is that assumptions (1) and (18) point-identify the response parameter β, provided that z is not mean independent of v. For example, numerous studies of the returns to schooling report estimates of β, interpreted as the common return that all members of the population experience per year of schooling.

Empirical researchers have long used the HLR assumption even though it is 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. Manski (1997) and Manski and Pepper (2000) argue that much of the research that has used the HLR assumption could more plausibly use the MTR assumption. 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 years of schooling. In these and other settings, MTR assumptions have a reasonably firm foundation.

The above notwithstanding, it is important that researcher who want to maintain the HLR assumption should understand its implications when combined with other assumptions. This section studies identification when an MIV assumption of form (2) is combined with assumption (18).

4.1. The Classical Result

As background, we first give a proof of the classical result that originally appeared in Manski (1995, page 152), with further exposition in Manski (2007, Chapter 8). This proof is much simpler than those offered in traditional treatments of linear simultaneous equations. Moreover, it extends easily when we replace the IV assumption with an MIV assumption in the next section.

Let $u_1 \in V$ and $u_2 \in V$ be any two points on the support of the distribution of covariate v. Given (18), assumption (1) states that

$$(19) \qquad E(\delta \mid v = u_2) = E(\delta \mid v = u_1).$$

Given (18), $\delta_j = y_j - \beta z_j$ for each person j. Hence,

$$(20) \qquad E(y - \beta z \mid v = u_2) = E(y - \beta z \mid v = u_1).$$

Solving (20) for β yields

$$(21) \qquad \beta = \frac{E(y \mid v = u_2) - E(y \mid v = u_1)}{E(z \mid v = u_2) - E(z \mid v = u_1)},$$

provided that the dominator is non-zero. The requirement that the denominator be non-zero is called the *rank condition* in the classical literature.

Each quantity on the right-hand side of (21) is point-identified. Hence, assumptions (1), (18), and the rank condition point-identify β. If V contains multiple (u_1, u_2) pairs that satisfy the rank condition, then there exist correspondingly multiple versions of equation (21). The parameter β must equal the right-hand-side of all such equations. If the right-hand side of all versions of (21) have the same value, β is said to be *over-identified*. If versions of (21) differ in value, either the IV or the HLR assumption is incorrect.

Equation (21) takes a particularly simple form in the case of an ETS assumption. Let v = z. Then $E(z|v = u_2) - E(z|v = u_1) = u_2 - u_1$. Hence, the rank condition holds and (21) reduces to

(22) $$\beta = [E(y|z = u_2) - E(y|z = u_1)]/(u_2 - u_1).$$

4.2. Weakening an IV to an MIV

Now replace IV assumption (1) with MIV assumption (2). Let V be an ordered set and let $u_2 > u_1$ be any two points on the support of v. Given (18), assumption (2) states that

(23) $$E(\delta|v = u_2) \geq E(\delta|v = u_1).$$

Recall that $\delta_j = y_j - \beta z_j$ for each person j. Hence,

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

Solving for β yields the inequality

$$(25a) \qquad \beta \leq \frac{E(y|v = u_2) - E(y|v = u_1)}{E(z|v = u_2) - E(z|v = u_1)} \qquad \text{if } E(z|v = u_2) - E(z|v = u_1) > 0,$$

$$(25b) \qquad \beta \geq \frac{E(y|v = u_2) - E(y|v = u_1)}{E(z|v = u_2) - E(z|v = u_1)} \qquad \text{if } E(z|v = u_2) - E(z|v = u_1) < 0.$$

This proves

Proposition 3: Let MIV Assumption (2) and HLR assumption (18) hold. Then β lies in the intersection of the inequalities (25) 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 3 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 HLR assumption. It may turn out that no value of β satisfies all of the inequalities in (25). If so, either assumption HLR or MIV is incorrect.

In general, the bound in Proposition 3 does not point-identify β. However, the sign of β may be identified. Inspection of (25) shows that $\text{sgn}(\beta)$ 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 $E(z|v = u_2) - E(z|v = u_1) > 0$. $\text{Sgn}(\beta)$ 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 $E(z|v = u_2) - E(z|v = u_1) < 0$. $\text{Sgn}(\beta)$ is not identified if $E(y|v = u_2) - E(y|v = u_1) \geq 0$ for all $u_2 > u_1$.

Inequalities (25) take a particularly simple form in the case of an MTS assumption. Let $v = z$. Then $E(z|v = u_2) - E(z|v = u_1) = u_2 - u_1 > 0$, so only (25a) is applicable. These upper bounds on β reduce to

$$(26) \qquad \beta \leq \inf_{(u_2,\, u_1):\; u_2 > u_1} \frac{E(y \mid z = u_2) - E(y \mid z = u_1)}{u_2 - u_1}.$$

Thus, the MTS and HLR assumptions together imply an upper bound on β but no lower bound.

4.3. Restricted Outcome Spaces

The above analysis has supposed that the outcome space Y is the entire real line. Empirical researchers often apply the HLR assumption in settings where Y is a restricted part of the real line, perhaps a discrete set or a bounded interval. In such cases, the structure of Y implies constraints on β that must hold even in the absence of an IV or MIV assumption. These constraints, which typically are ignored in empirical studies, generally have no substantive interpretation and often have the nonsensical implication that β must equal zero. These difficulties are regularly overlooked in practice, so we call attention to them here, in the hope that empirical researchers will henceforth be more judicious in their applications of the HLR assumption.

The source of the constraints is the fact that the outcomes generated by the HLR assumption must logically be elements of Y. That is, β must be such that $\beta t + \delta_j \in Y$ for all $t \in T$ and $j \in J$. Recall that $\delta_j = y_j - \beta z_j$ for each person j. Hence, β must satisfy the constraints

$$(27) \qquad \beta(t - z_j) + y_j \in Y, \quad \forall\ t \in T \text{ and } j \in J.$$

These constraints are unnatural when the outcome space is discrete. If Y and T are both discrete, at most a discrete set of β values can satisfy (26). If Y is discrete and T is a continuum, the only parameter value that satisfies (26) is $\beta = 0$. Hence, the HLR assumption is not sensible when the outcome space is discrete.

The constraints imply a set of inequalities on β when the outcome space is a bounded interval. Let $Y = [K_0, K_1]$. Then (27) is the set of inequalities

(28) $\quad K_0 \leq \beta(t - z_j) + y_j \leq K_1, \quad \forall\ t \in T \text{ and } j \in J.$

Manipulation of (28) yields these inequalities on β:

(29a) $\quad (K_0 - y_j)/(t_1 - z_j) \leq \beta \leq (K_1 - y_j)/(t_1 - z_j), \quad \forall\ j \in J,$

(29b) $\quad (K_1 - y_j)/(t_0 - z_j) \leq \beta \leq (K_0 - y_j)/(t_0 - z_j), \quad \forall\ j \in J,$

where $t_0 \equiv \inf T$ and $t_1 \equiv \sup T$. These inequalities generally lack substantive interpretation. Indeed, the only value satisfying (29a) is $\beta = 0$ if the population contains a member j for whom $(y_j = K_0, z_j < t_1)$ and a member k for whom $(y_k = K_1, z_k < t_1)$. Similarly, the only value satisfying (29b) is $\beta = 0$ if the population contains a member j for whom $(y_j = K_0, z_j > t_0)$ and a k for whom $(y_{kj} = K_1, z_k > t_0)$. Thus, the HLR assumption generally is not sensible when Y is a bounded interval.

5. Estimation of the Bounds

Although identification usually is the dominant inferential problem in analysis of treatment response, finite-sample statistical inference can be a serious concern as well. It is easy to show that analog estimates of the bounds in Propositions 1 through 3 are consistent. Each bound is a continuous function of point-identified conditional means. When V is finite, these conditional means are consistently estimated by the corresponding sample averages. When V is a continuum, nonparametric regression methods may be used.

The bounds under the MTS assumption in (9) and (13) have simple explicit forms, and the sampling distributions of analog estimates are correspondingly simple. However, the sup and inf operations in inequalities (6) and (10) significantly complicate the bounds under other MIV assumptions, rendering it difficult to analyze the sampling behavior of analog estimates. Moreover, the methods for forming asymptotically valid confidence sets for partially identified parameters developed by Horowitz and Manski (2000), Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Rosen (2006), and others appear not to apply.

An important statistical concern, noted but not analyzed in Manski and Pepper (2000), is that analog estimates of bounds (6) and (10) have finite-sample bias that make the estimates tend to be narrower than the true bounds. Section 5.1 explains this bias, gives Monte Carlo evidence on its magnitude, and describes an heuristically motivated bias-correction method. Section 5.2 considers analog estimation of the bound of Proposition 3. Although the present discussion focuses on estimation of MIV bounds, with obvious modifications it applies as well to estimation of the IV bounds of Manski (1990)

5.1. Estimation of the Bounds of Propositions 1 and 2

The finite-sample bias of analog estimates is most transparent when V is a finite set. We restrict attention to this case and focus on the lower bound in (6). Analysis of the upper bound in (6) and the bounds in (10) is analogous.

To begin, observe that the lower bound in (6) may be rewritten as

$$(30) \qquad \max_{u_1 \leq u} \; E\{y \cdot 1[z = t] + K_0 \cdot 1[z \neq t] \,|\, v = u_1\}.$$

Let a random sample of size N be drawn. Let N(u) be the sub-sample size with (v = u). Consider inference conditional on the vector $N(V) \equiv [N(u), u \in V]$ of sub-sample sizes. The analog estimate of (30) is

$$(31) \qquad \max_{u_1 \le u} E_{N(V)}\{y \cdot 1[z = t] + K_0 \cdot 1[z \ne t] \mid v = u_1\},$$

where $E_{N(V)}$ denotes the empirical mean. Suppose that all components of N(V) are positive, so the estimate exists. Each term in (31) is an unbiased estimate of the corresponding term in (30). It follows by Jensen's inequality that

$$(32) \qquad E^* \{ \max_{u_1 \le u} E_{N(V)}\{y \cdot 1[z = t] + K_0 \cdot 1[z \ne t] \mid v = u_1\}\} \ge \max_{u_1 \le u} E\{y \cdot 1[z = t] + K_0 \cdot 1[z \ne t] \mid v = u_1\},$$

where E^* denotes the expected value of the estimate across repeated samples of sizes N(V). The inequality in (32) is strict in the ordinary case where the distributions $P\{y \cdot 1[z = t] + K_0 \cdot 1[z \ne t] \mid v = u_1\}$, $u_1 \le u$ are non-degenerate and have overlapping supports.

The above shows that the analog estimate of the lower bound is biased upwards for each vector N(V) such that the estimate exists. Similar analysis shows that the estimate of the upper bound is biased downward. Thus, the mean estimate of the bound always is a subset of the true bound and ordinarily is a proper subset.

Monte-Carlo Evidence

To obtain a sense of the magnitude of the bias, we have performed a Monte Carlo experiment. Ceteris paribus, the bias is most serious when the no-assumption bound is the same for all values of v, implying that the MIV assumption has no identifying power. We consider such a setting. In particular,

a) v has a multinomial distribution with M equal-probability mass points {1/M, 2/M, ... , 1}.

b) $T = \{0, 1\}$, $z_j = 1[v_j + \epsilon_j > 0]$, and ϵ is distributed $N(0, 1)$.

c) $y_j = \min\{\max[-1.96, \eta_j], 1.96\}$, and η is distributed $N(0, \sigma^2)$. Thus, y is censored normal.

d) The random variables η, v, and ϵ are mutually statistically independent.

In this setting,

$$(33) \quad E\{y\cdot 1[z = t] + K_0\cdot 1[z \neq t] \mid v = u_1\} = -1.96\cdot P(z \neq t \mid v = u_1).$$

Let t = 1 and u = 1. Then (30) reduces to

$$(34) \quad -1.96\cdot \min_{u_1 \leq 1} P(z = 0 \mid v = u_1) = -1.96\cdot P(z = 0 \mid v = 1) \cong -0.31$$

for all values of M and σ^2.

Fix values of N, M, and σ^2. To measure the bias of the analog estimate, we draw 1000 random samples of size N from the distribution of $\{\eta, v, \epsilon\}$, and compute the analog estimate of the MIV lower bound for each sample. The bias is then measured as the difference between the average of the 1000 estimates and the true lower bound -0.31.

Table 1 displays the bias for $N \in \{100, 500, 1000)$, $M \in \{4, 8\}$, and $\sigma^2 \in \{1, 4, 25)$. Qualitatively, the upward bias increases with M and σ^2 and decreases with N. These findings are sensible. The difference between the left and right-hand sides of (32) increases with the dispersion of the estimates. Dispersion increases with σ^2 and M, while it decreases with N. Observe that the mean number of observations of y per value of v is N/M. Hence, for each value of v, the sample size for estimation of $E\{y\cdot 1[z = t] + K_0\cdot 1[z \neq t] \mid v\}$ tends to increase with N and decrease with M.

Quantitatively, the bias is enormous when (N = 100, M = 8) for all values of σ^2. Indeed, the mean estimate of the bound is an empty interval when $\sigma^2 \in \{4, 25\}$. The bias is negligible when (N = 1000, M =

4) for all values of σ^2. Small to moderate biases occur at other values of (N, M, σ^2).

Table 1: Bias of the Analog Estimate of the MIV Lower Bound

	M = 4			M = 8		
N	$\sigma^2 = 1$	$\sigma^2 = 4$	$\sigma^2 = 25$	$\sigma^2 = 1$	$\sigma^2 = 4$	$\sigma^2 = 25$
100	0.09	0.15	0.19	0.31	0.43	0.53
500	0.01	0.03	0.04	0.08	0.12	0.15
1000	0.01	0.01	0.02	0.04	0.07	0.09

Bias-Corrected Estimates

To counter the bias of the analog estimate, it is natural to seek bias-corrected methods. Krieder and Pepper(2007) proposed a bootstrap bias-corrected estimator and applied it to their misreporting problem. The idea is to estimate the bias using the bootstrap distribution, and then adjust the analog estimate accordingly. For a random sample of size N, let T_N be the analog estimate of the lower bound in (31) and let $E^b(T_N)$ be the mean of this estimate using the bootstrap distribution. The bias is then estimated to equal $E^b(T_N)-T_N$, and the proposed bias-corrected estimator is $2T_N - E^b(T_N)$. Analyzing a partial identification problem that is substantively different but has similar mathematical structure, Haile and Tamer (2003) used a smoothing function to reduce the variability of analog estimators across different values of an index. While both correction methods seem reasonable and tractable, neither has a firm theoretical foundation as of now.

Evidence on the efficacy of these corrections in finite samples can be obtained from Monte Carlo experiments. We have assessed the Krieder and Pepper (2007) estimator using the simulation design described earlier. To do this, we conducted further simulations to assess the bootstrap distribution. For each of the 1000 random samples of size N, we drew 1000 random pseudo samples of size N, and used these pseudo samples to compute the mean $E^b(T_N)$of the bootstrap distribution. The bias of the Krieder-Pepper estimator was then measured as the difference between the average of the 1000 estimates and the true lower bound −0.31.

Table 2 displays the bias of the proposed estimator. Relative to the analog estimator in (31), substantial reductions in bias are realized in cases where N = 100 or M =8. For example, when (N=100, M=4, $\sigma^2 = 25$), the bias falls from 0.19 to 0.05. Overall, the biases are negligible (less than or equal to 0.05) except in the extreme case when (N=100, M=8). The biases are moderate in this case, but considerably smaller than the analogous biases of the analog estimator.

Table 2: Bias of the Krieder-Pepper Estimate of the MIV Lower Bound

	M = 4			M = 8		
N	$\sigma^2 = 1$	$\sigma^2 = 4$	$\sigma^2 = 25$	$\sigma^2 = 1$	$\sigma^2 = 4$	$\sigma^2 = 25$
100	0.01	0.03	0.05	0.12	0.16	0.21
500	-0.01	-0.01	-0.01	0.02	0.03	0.04
1000	-0.01	-0.00	-0.01	0.00	0.01	0.02

5.2. Estimation of the Bound of Proposition 3

This final section considers estimation of the bound (25) obtained by combining an MIV assumption with the HLR assumption. We first examine the special case in which the MIV is the realized treatment. Here inequality (26) gave the resulting sharp upper bound on β. The argument of Section 5.1 shows that the analog estimate of this bound is downward biased.

As earlier, consider inference conditional on N(V). The analog estimate of (26) is

$$(35) \qquad \min_{(u_2, u_1):\, u_2 > u_1} \frac{E_{N(V)}(y \mid z = u_2) - E_{N(V)}(y \mid z = u_1)}{u_2 - u_1}.$$

Let all components of N(V) be positive, so the estimate exists. Each term in (35) is an unbiased estimate of the corresponding term in (26). It follows by Jensen's inequality that

$$(36)\quad E^{*}\left\{ \min_{(u_2,\, u_1):\, u_2 > u_1} \frac{E_{N(V)}(y \mid z = u_2) - E_{N(V)}(y \mid z = u_1)}{u_2 - u_1} \right\} \leq \min_{(u_2,\, u_1):\, u_2 > u_1} \frac{E(y \mid z = u_2) - E(y \mid z = u_1)}{u_2 - u_1}.$$

Thus, the estimate of the upper bound is biased downward for each vector N(V) such that the estimate exists. Hence, the estimate is biased downward conditional on existence.

Now consider the general form of Proposition 3, where β lies in the intersection of the inequalities (25). The analog estimate of the upper bound is

$$(37a)\quad \min_{(u_2,\, u_1) \in W_U} \frac{E_{N(V)}(y \mid v = u_2) - E_{N(V)}(y \mid v = u_1)}{E_{N(V)}(z \mid v = u_2) - E_{N(V)}(z \mid v = u_1)}$$

where $W_U \equiv \{(u_2, u_1)\colon u_2 > u_1 \text{ and } E_{N(V)}(z \mid v = u_2) - E_{N(V)}(z \mid v = u_1) > 0\}$. Similarly, the estimate of the lower bound is

$$(37b)\quad \max_{(u_2,\, u_1) \in W_L} \frac{E_{N(V)}(y \mid v = u_2) - E_{N(V)}(y \mid v = u_1)}{E_{N(V)}(z \mid v = u_2) - E_{N(V)}(z \mid v = u_1)},$$

where $W_L \equiv \{(u_2, u_1)\colon u_2 > u_1 \text{ and } E_{N(V)}(z \mid v = u_2) - E_{N(V)}(z \mid v = u_1) < 0\}$.

The structure of this estimate is complex. In particular, the estimate is highly sensitive to small variations in $E_{N(V)}(z \mid v = u_2) - E_{N(V)}(z \mid v = u_1)$ when this quantity is near zero, with discontinuity at zero. Realizations of $E_{N(V)}(z \mid v = u_2) - E_{N(V)}(z \mid v = u_1)$ that are near zero tend to occur frequently if the population mean difference $E(z \mid v = u_2) - E(z \mid v = u_1)$ is near zero and/or the dispersion of its estimate is large. Hence, estimate (37) has subtle sampling behavior in such cases. This is the MIV manifestation of the so-called *weak instruments* problem that has drawn much attention in the literature on analog estimation under the HLR and IV assumptions. See, for example, Nelson and Startz (1990), Bound, Jaeger, and Baker (1995), and Staiger and Stock (1997).

Observe that this weak-instruments problem does not occur in analog estimation of the bounds of Propositions 1 and 2. In those cases, the estimate always varies continuously as a function of multiple sample averages. Even when an MIV has no identifying power at all, the bounds of Propositions 1 and 2 exist and their analog estimates are consistent.

References

Beresteanu, A. and F. Molinari (2008), "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763-814.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007), "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75, 323-363.

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.

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243-1284.

Gerfin, M. and M. Schellhorn (2006), "Nonparametric Bounds on the Effect of Deductibles in Health Care Insurance on Doctor Visits—Swiss Evidence," *Health Economics*, 15, 1011-1020.

González, L. (2005), "Nonparametric Bounds on the Returns to Language Skills," *Journal of Applied Econometrics*, 20, 771-795.

Haile, P. and E. Tamer (2003), "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.

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

Horowitz, J. and C. Manski (2000),"Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77–84.

Kreider, B. and S. Hill, (2008). "Partially Identifying Treatment Effects with an Application to Covering the Uninsured," *Journal of Human Resources*, forthcoming.

Kreider, B. and J. Pepper (2007), "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," *Journal of the American Statistical Association*, 102, 432-441.

Kreider, B. and J. Pepper. (2008). "Inferring Disability in Corrupt Data," *Journal of Applied Econometrics,* 23, 329-349.

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

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. (2003), *Partial Identification of Probability Distributions*, New York: Springer-Verlag.

Manski, C. (2007), *Identification for Prediction and Decision*, Cambridge, Mass.: Harvard University Press.

Manski, C. and J. Pepper (1998), "Monotone Instrumental Variables: with an Application to the Returns to Schooling," National Bureau of Economic Research Technical Working Paper t0224.

Manski, C. and J. Pepper (2000), "Monotone Instrumental Variables: with an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010.

Manski, C. and E. Tamer (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome,." *Econometrica*, 70,: 519–546.

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.

Rosen, A. (2006), "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," Department of Economics, University College London.

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