Identification of Local Treatment Effects Using a Proxy for an Instrument

Karim Chalak*†
Boston College

April 30, 2010

Abstract

The method of indirect least squares (ILS) using a proxy for a discrete instrument is shown to identify a weighted average of local treatment effects. The weights are nonnegative if and only if the proxy is *intensity preserving* for the instrument. A similar result holds for instrumental variables (IV) methods such as two stage least squares. Thus, one should carefully interpret estimates for causal effects obtained via ILS or IV using an error-laden proxy of an instrument, a proxy for an instrument with missing or imputed observations, or a binary proxy for a multivalued instrument. Favorably, the proxy need not satisfy all the assumptions required for the instrument. Specifically, an individual's proxy can depend on others' instrument and the proxy need not affect the treatment nor be exogenous. In special cases such as with binary instrument, ILS using any suitable proxy for an instrument identifies local average treatment effects.

Keywords: causality, compliance, indirect least squares, instrumental variables, local average treatment effect, measurement error, proxy, quadrant dependence, two stage least squares.

---

*Karim Chalak is Assistant Professor of Economics, Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467 (email: chalak@bc.edu).

†The author thanks the participants of the Boston College labor lunch seminar and the UC San Diego and Brown econometrics seminar, Gary Chamberlain, Stefan Hoderlein, Ivana Komunjer, Arthur Lewbel, and especially Guido Imbens and Halbert White for helpful comments and suggestions. All errors and omissions are the author's responsibility.

1 Introduction

This paper studies the scope of the method of indirect least squares (ILS) and instrumental variables (IV) methods for the identification of local average treatment effects (LATEs) of a cause or treatment D on an outcome of interest Y using a proxy W for an exogenous instrument Z. We study the case in which the instrument, proxy, and treatment are discrete; the outcome need not be discrete. Here an instrument Z satisfies definition 3 of Angrist, Imbens, and Rubin (1996). Three particular features of Z are that (1) it is an exogenous variable that (2) induces individuals' receipt of treatment in the same direction, and (3) drives the outcome only via the treatment. A proxy W for an instrument Z is a variable associated with Z, and possibly driven by it, that need not affect the treatment or outcome and thus need not be an instrument according to this definition. In particular, we assume that W is mean independent of the treatment and outcome conditional on the instrument. Examples of possible proxies include an error-laden measurement of an unobserved instrument, a measurement of an instrument with missing or imputed observations, or a binary coding for a multivalued instrument.

Building on the work of Imbens and Angrist (1994; thereafter IA), Angrist and Imbens (1995; thereafter AI), and Angrist, Imbens, and Rubin (1996, thereafter AIR), we study whether ILS and IV methods that use a proxy W for an instrument Z identify local average treatment effects. In doing so, we establish a relation between an ILS or IV estimand using the proxy W on one hand and ILS estimands using Z on the other. As we show, ILS methods that employ a suitable proxy for a multivalued instrument identify a weighted average of local average treatment effects based on the instrument. This is particularly troublesome if some of the weights are negative since in this case the sign of the identified weighted average local effect may be the opposite of that of some, or even all, of the LATEs. In effect, two researchers seeking to measure the same local average effect using different proxies for the same instrument may identify weighted average effects of different signs. We show that the weights are nonnegative if and only if the proxy is *intensity preserving* for the instrument, a specific form of dependence of the instrument on the proxy that we discuss. Furthermore, we show that a similar result holds for IV: IV methods using a function of a proxy for an instrument identify a weighted average of LATEs based on the instrument. We demonstrate that the weights are nonnegative if and only if the instrument and the function of the proxy *covary* with the proxy *concordantly*, a particular form of dependence among the instrument, proxy, and a function of the proxy that we introduce. Thus, researchers should cautiously interpret estimates for causal effects obtained via ILS or IV (e.g. TSLS) methods based on a proxy for an instrument including an error-laden measurement of an instrument, a proxy for an instrument with missing or imputed observations, and a binary proxy for a multivalued instrument.

The case of a binary instrument Z receives significant attention in the literature. For example, in the "compliance problem" studied in AIR, the treatment and the treatment assignment are binary and the latter may serve as an instrument. In this case, LATE is the average effect of the treatment for the subpopulation of "compliers" who always comply with the treatment assignment. We demonstrate that for binary Z, ILS or IV using any suitable proxy for the treatment assignment identifies LATE. Thus, in this case, identifying LATE via ILS is robust to certain types of measurement error and instrument misclassification. Also, the fact that ILS using any suitable proxy identifies LATE provides the foundation for testing the underlying assumptions.

Favorably, the proxy need not satisfy some of the assumptions imposed on the instrument. In particular, the proxy need not be individualistic: an individual's proxy may depend on others' instrument. Further, unlike the instrument, the proxy need not cause the treatment nor be exogenous but rather it suffices that it is associated with it in an appropriate sense. If economic theory suggests a proper unobserved instrument, then a researcher may make use of a suitable intensity preserving proxy to identify an informative weighted average of local effects. For example, a nonzero weighted average of local effects provides evidence against the hypothesis of no causal effect of the treatment on the outcome.

The results of this paper apply to a variety of situations in which researchers make use of a proxy for an instrument. To illustrate, consider an experiment to measure the effects of a training program on wage for a population of individuals $i = 1, ..., n$ in which financial aid status denoted by Z is randomized across individuals thereby affecting their enrollment cost if they attend the program. Denote joining the program by $D = 1$, otherwise $D = 0$. Last, Y denotes wage measured at a subsequent time. For $i = 1, ..., n$, the econometrician observes who joined the program and the individuals' wages, that is realizations of D and Y, but does not observe the financial aid status of individuals. Instead she only observes a proxy W for financial aid. The properties of the proxy may vary depending on the context. The proxy may denote an error-laden measurement of the instrument Z that incorrectly codes certain individual's financial aid status. The proxy may also contain missing observations for certain individuals' instrument. These observations could be coded as missing or otherwise replaced by some value which could possibly depend on others' observed instrument. Further, the proxy may code the instrument information coarsely. For example, Z may assign individuals zero financial aid, may wave his/her tuition, or may provide him/her with full financial aid covering tuition and living expenses, and W may only code whether individuals are assigned financial aid. In turn, this coding may also contain measurement error. We study these cases in what follows.

Examples in which researchers employ a proxy for an instrument are abundant in the literature. Consider for example the effects of military service denoted by a binary treatment D on an outcome Y denoting a measure of wage in Angrist (1990) and of civilian mortality in AIR and let Z denote the Vietnam draft lottery random sequence number drawn for each date of birth in a given year. Thus, for the year 1951 Z can take on 365 values. Individuals who were assigned a sequence number that falls below a subsequently announced ceiling (e.g. 125) were potentially drafted for military service. Angrist (1990) and AIR use a binary proxy W for the lottery number denoting whether the latter falls below the ceiling. The effect of the lottery number Z on military service need not be fully captured by the proxy W. For example, Angrist (1990, p. 314) notes that "many men with low numbers volunteered for the military to avoid being drafted and to improve their terms of service" and that "there was even a behavioral response to the lottery in enlistment rates for the 1953 cohort, although no one born in 1953 was drafted."

As another example, Angrist and Krueger (1991) study the returns to education. Here, the treatment D denotes completed years of education, the outcome Y is a measure of wage, and Z corresponds to date of birth. Angrist and Krueger (1991) argue that because of age at entry policy and compulsory schooling law, individuals with different birthdays may attain different levels of schooling. In particular, age of entry policy typically requires a student to be 6 years old by January of his/her first year at school. Further, a student is allowed to drop out of school only after he/she attains a legal age (e.g. 16 years old). Thus, individuals born earlier in the year may attain less schooling than those born later in the year because they enter school at an older age and drop out of school earlier. Angrist and Krueger (1991, p. 995) make use of a binary proxy W for date of birth, such that $W = 1$ if the date of birth is in the first quarter and 0 otherwise. Here too the proxy need not fully capture the effect of date of birth on education. For example, Angrist and Krueger (1991, p. 979-980) state that "the interaction of school-entry requirements and compulsory schooling laws compel students born in certain months to attend school longer than students born in other months." AI also apply ILS and TSLS using instruments denoting quarter of birth. As we demonstrate, in this and the previous example, under certain assumptions, ILS using the proxy W identifies a weighted average of local treatment effects based on Z with nonnegative weights.

The paper is organized as follows. Section 2 describes the data generating process and states the assumptions. We work simultaneously with a structural system and the Rubin Causal Model (Holland, 1986; AIR). Section 3 gives a general result which demonstrates that the ILS estimand using a proxy is a weighted average of ILS estimands using the instrument. Further, we introduce the notion of an *intensity preserving* proxy for an instrument and show that this property is necessary and sufficient for the weights to be nonnegative. Section 4 studies identification of causal effects via ILS using a proxy for an instrument. After stating a general result for multivalued variables, we examine special cases. First, we study the compliance problem with binary instrument, proxy, and treatment. We then study the case of a binary instrument and multivalued proxy. Last we study the case of multivalued instrument and binary proxy. Section 5 shows that IV methods using a proxy identify a weighted average of local treatment effects based on the instrument. We extend results in IA and AI to the general case of multivalued instrument, proxy, and treatment and provide a necessary and sufficient condition for the weights to be nonnegative. For this, we introduce the notions of *mean quadrant dependence* and of *concordance* specifying dependence among the instrument, proxy, and a function of the proxy. Section 6 relates the results here to other work on instrument proxy in linear structural systems and nonseparable structural systems with continuous variables and briefly discusses the implications of the results in this paper on testing the underlying assumptions and the hypothesis of no causal effects. Section 6 concludes. Mathematical proofs are gathered in the Appendix.

2 Data Generation

This section introduces the data generating process. In stating our assumptions, we work simultaneously with a structural system and the Rubin Causal Model (Holland, 1986) as in Vytlacil (2002). In what follows, we extend the assumptions in IA, AI, and AIR to accommodate a proxy W for the possibly unobserved instrument Z and to permit the instrument, proxy, and treatment to be multivalued.

Assumption 2.1 *Data Generating Structural System: (i) Let I denote a population of individuals indexed by $i = 1, ..., n$, $n \in \mathbb{N}^+ := \{1, 2, ...\}$. Let Z, W, D, and Y be random variables ranging over I and taking value[1] respectively in $S_Z := \{0, 1, ..., K\}$, $S_W := \{0, 1, ..., L\}$, $S_D := \{0, 1, ..., J\}$, $K, L, J \in \mathbb{N}^+$ finite, and $S_Y \subseteq \mathbb{R}$, and suppose that $E(D)$ and $E(Y)$ exist and are finite. (ii) Let a triangular structural system generate the random vectors of countable dimension U_W, U_D, and U_Y and the instrument Z ranging over I. The proxy w_i for the instrument, treatment d_i, and outcome y_i for individual i in population I are structurally generated as*

$$\begin{aligned} w_i &= p(\mathbf{z}, u_{W,i}), \\ d_i &= q(z_i, u_{D,i}), \text{ and} \\ y_i &= r(d_i, u_{Y,i}), \end{aligned}$$

where p, q, and r are unknown measurable functions mapping respectively to S_W, S_D, and S_Y and where $\mathbf{z} = (z_1, ..., z_n)'$. The realizations of W, D, and Y are observed, those of U_D, U_Y, U_W are not, and $\mathbf{z}$ may be unobserved.

Part (i) of Assumption 2.1 introduces the random variables. Part (ii) imposes structure on the data generating process. We let $i = 1, \dots, n$ denote individuals in a population I. The random variable Z takes value in S_Z and denotes, under Assumption 2.1(ii), the instrument. The vector $\mathbf{z} = (z_1, ..., z_n)'$ collects the realizations of Z for all individuals. Typically, we assume $\mathbf{z}$ is unobserved. In some cases, $\mathbf{z}$ may be observed and the econometrician nevertheless employs a suitable proxy W for the instrument Z as in the examples discussed in the Introduction. The proxy W takes value in S_W, the treatment D takes value in S_D, and we denote by Y taking value in S_Y the outcome of interest. We let Z, W, and

[1] Throughout, the support of a random variable X, supp(X):= S_X, denotes the smallest set S such that $P[X \in S] = 1$.

D be multivalued discrete random variables and Y be any random variable. For $z \in S_Z$, we define the *potential treatment* $D(z)$ as

$$D(z) := q(z, U_D),$$

and similarly, for $d \in S_D$, we define the *potential outcome* as

$$Y(d) := r(d, U_Y)$$

The random vectors U_W, U_D, and U_Y reflect the heterogeneity of individuals in population I. For example, for the same value of the instrument z, $d_i(z) := q(z, u_{D,i})$ may differ from $d_j(z) := q(z, u_{D,j})$ and therefore individuals i and j may receive different treatments.

The triangular structure above defines an inherent ordering of variables, implicit in the potential outcome notation, in which predecessors influence successors but not the opposite (see Chalak and White, 2008). In particular, the random vectors U_W, U_D, and U_Y precede the instrument Z. In turn, U_W, U_D, U_Y, and Z precede the proxy W, all of which precede the treatment D, and the outcome Y succeeds all the system's variables.

Two assumptions are implicit in Assumption 2.1(ii). Following Manski (2010), we refer to the first one as *individualistic treatment receipt and treatment response*. In particular, the potential treatment $d_i(z)$ for an individual i, depends only on his/her variables but does not depend on others' treatment assignment or proxies. Similarly, the potential outcome $y_i(d)$ for an individual i depends only on his/her variables. Observe that Assumption 2.1(ii) permits the *potential proxy* for an individual i $w_i(\mathbf{z}') := p(\mathbf{z}', u_{W,i})$ to depend on the vector of treatment assignment $\mathbf{z}'$ for all individuals. Strengthening Assumption 2.1(ii) to further restrict the potential proxy to be individualistic ($p(\mathbf{z}', u_{W,i}) = p(\mathbf{z}'', u_{W,i})$ for all $\mathbf{z}'$ and $\mathbf{z}''$ such that $z'_i = z''_i$ and all i) gives the "stable unit treatment value assumption" (SUTVA) also known as "no interference between units" (See e.g. Rubin, 1986; AIR).

The second assumption implicit in Assumption 2.1(ii) is an *exclusion restriction* which imposes structure on the impact of the instrument and proxy on the treatment and outcome.

In particular, Assumption 2.1(ii) states that the receipt of treatment is affected by the instrument but is not affected by the proxy. This would be violated if for example an individual who is assigned to treatment decides to not take the treatment because the proxy incorrectly codes him as assigned to the control group. On the other hand, this assumption is plausible if individuals believe that the treatment assignment is coded without error. Similarly, Assumption 2.1(ii) states that the response is affected by the treatment but is not otherwise affected by the instrument or its proxy. This would be violated if for example the instrument affects the outcome via a channel other than the treatment.

A notion of the causal effects on D and Y of the intervention $z \to z'$ to Z, defined as two points $z, z' \in S_Z$ (see White and Chalak, 2009), for the i^{th} individual can be given by:

$$\begin{aligned} \text{Causal effect of } z &\to z' \text{ on } d_i : d_i(z') - d_i(z), \text{ and} \\ \text{Causal effect of } z &\to z' \text{ on } y_i : y_i(d_i(z')) - y_i(d_i(z)). \end{aligned}$$

The "fundamental problem of causal inference" (Holland, 1986) is that we do not observe the counterfactuals. For example, we may observe $d_i(z)$ or $d_i(z')$ but not both ($z \neq z'$). We thus focus attention on certain average causal effects over the population I. Together with Assumption 2.1, our next assumption permits us to equate average causal effects with differences of conditional average responses. In what follows, $\perp_m$ denotes mean independence[2].

Assumption 2.2 *Mean Ignorability of the Instrument:*

$$D(z) \perp_m Z \quad \textit{and} \quad Y(D(z)) \perp_m Z \quad \textit{for all } z \in S_Z.$$

For controlled or natural experiments, Assumption 2.2 is ensured by randomization of Z. Also, observe that exogeneity of the instrument, $Z \perp (U_D, U_Y)$ where $\perp$ denotes independence as in Dawid (1979), is sufficient for Assumption 2.2 to hold by Dawid (1979,

[2]We write $X_1 \perp_m X_2 | X_3$ if $E(X_1|X_2, X_3) = E(X_1|X_3)$ provided these means exist and are finite.

lemma 4.2). It follows that under Assumptions 2.1, and 2.2, the average causal effect on D of the intervention $z' \to z$ to Z, $E[D(z') - D(z)]$ is identified with $E(D|Z = z') - E(D|Z = z)$. Similarly, the average causal effect on Y of the intervention $z' \to z$ to Z, $E[Y(D(z')) - Y(D(z))]$ is identified with $E(Y|Z = z') - E(Y|Z = z)$ (see e.g. Rubin, 1978).

The next two assumptions impose structure on the effect of the instrument on the treatment as in IA, AI, and AIR.

Assumption 2.3 *Nonzero Causal Effect of Z on D: There exists $z, z' \in S_Z$, $z < z'$, such that* $\Pr[D(z) \neq D(z')] > 0$.

Assumption 2.3 ensures that there is a non-negligible set of individuals for whom the instrument affects the treatment, that is $d_i(z) \neq d_i(z')$. The next assumption imposes structure on the way in which the instrument affects the treatment.

Assumption 2.4 *Weak Monotonicity: There exists $z, z' \in S_Z$, $z < z'$, such that* $\Pr(D(z) \leq D(z')) = 1$.

Assumption 2.4 requires that $q(z, U_D)$ *is locally monotonic with probability 1.* IA, AI, and AIR discuss the plausibility of this assumption and the consequences of its failure. For $J = K = 1$, Assumption 2.4 essentially rules out from the population "defiers" who systematically undertake the opposite treatment than that assigned to them (see AIR). When Assumption 2.4 holds for all $k - 1, k \in S_Z$, we order the instrument values such that $D(0) \leq D(1) \leq ... \leq D(K)$ with probability 1. Given Assumption 2.2 this implies $E(D|Z = 0) \leq E(D|Z = 1) \leq ... \leq E(D|Z = K)$. This imposes a particular from of dependence of the treatment on the instrument, namely that $E(D|Z = l)$ is monotone in l for all $l \in S_Z$. In linear structural systems, the instrument Z is required to be "relevant," that is correlated with D. Here a specific form of relevance is imposed. We discuss this further in Section 4. If Assumption 2.3 also holds for all $k - 1, k \in S_Z$, we show that we then obtain $E(D|Z = 0) < E(D|Z = 1) < ... < E(D|Z = K)$. For example, when $J = 1$, Assumption 2.4 ensures that with probability 1, $D(z') - D(z)$ is either 0 or 1. Assumption 2.3 further gives that that $\Pr[D(z') - D(z) = 1] \neq 0$ so that the average causal effect of Z on D is nonzero, $E[D(z') - D(z)] \neq 0$. Assumption 2.2 then gives that $E[D|Z = z'] - E[D|Z = z] \neq 0$.

The last assumption concerns the proxy W.

Assumption 2.5 *Conditional Mean Independence of the Proxy from the Treatment and Outcome Given the Instrument:*

$$D \perp_m W|Z \quad \text{and} \quad Y \perp_m W|Z.$$

Assumption 2.5 ensures that the proxy W is irrelevant for predicting the average treatment and outcome given the instrument information Z. Assumption 2.1 ensures that the treatment and outcome are not affected by the proxy thereby eliminating a possibility through which Assumption 2.5 may be violated. Assumption 2.5 may also be violated if the proxy and the treatment or outcome are jointly driven by factors other than the instrument, or if the proxy is contaminated based on the treatment or outcome. Assumption 2.5 holds, when $(U_D, U_Y) \perp W|Z$ so that conditional on Z, W does not predict the heterogeneity in treatment receipt and response. When the proxy is individualistic, $W = q(Z, U_W)$, Assumption 2.5 is satisfied provided $(Z, D, Y) \perp U_W$ which can hold for example when U_W is measurement error occurring at random in recording the instrument, when U_W is a binary indicator for missing at random instrument values which the proxy codes by $L = K + 1$ so that $W = ZU_W + L(1 - U_W)$, or trivially when U_W is constant and W is a binary proxy for a multivalued instrument.

The next Proposition shows that exogeneity of the instrument and proxy $(Z, W) \perp (U_D, U_Y)$ is sufficient for Assumptions 2.2 and 2.5 to hold.

Proposition 2.1 *Suppose that Assumption 2.1 hold. If $(Z, W) \perp (U_D, U_Y)$ then Assumptions 2.2 and 2.5 hold.*

2.1 Comparison of Instrument and Proxy

It is useful to compare the assumptions imposed on the instrument Z and the proxy W. (1) In contrast to the instrument Z, W need not be determined individualistically, it could be a function of the instrument of all individuals. For example, the econometrician may impute missing instrument observations provided Assumption 2.5 is satisfied. (2) We assume that an individual's treatment, and through it outcome, may be affected by his/her instrument but not by his/her proxy. (3) We do not require that the proxy is exogenous. In particular, we do not assume that $(U_D, U_Y) \perp W$ or that $D(z) \perp_m W$ and $Y(D(z)) \perp_m W$ for all $z \in S_Z$. If we have that $(U_D, U_Y) \perp W|Z$ then $D(z) \perp_m W|Z$ and $Y(D(z)) \perp_m W|Z$. Given Assumption 2.2, this implies that $D(z) \perp_m W$ and $Y(D(z)) \perp_m W$. (4) In sharp contrast to Assumption 2.3 for the instrument, the proxy need not cause the treatment. Thus, the proxy W need not satisfy all the assumptions imposed on the instrument Z.

3 ILS Using A Proxy For An Instrument

To state a result relating the ILS estimand using a proxy to those using an instrument, it is convenient to introduce the following notation. For all $k, k' \in S_Z$, $k < k'$, such that $E(D|Z = k) \neq E(D|Z = k')$, define $\alpha_{k,k'}$ the ILS estimand using the instrument Z evaluated at k and k':

$$\alpha_{k,k'} := \frac{E(Y|Z = k') - E(Y|Z = k)}{E(D|Z = k') - E(D|Z = k)}.$$

Similarly, for all $l, l' \in S_W$, $l < l'$, such that $E(D|W = l) \neq E(D|W = l')$, define $\beta_{l,l'}$, the ILS estimand using the proxy W evaluated at l and l':

$$\beta_{l,l'} := \frac{E(Y|W = l') - E(Y|W = l)}{E(D|W = l') - E(D|W = l)}.$$

This section demonstrates that $\beta_{l,l'}$ can be represented by a weighted average of $\alpha_{k-1,k}$, $k \in S_Z|\{0\}$. The results of this section do not require the structure in Assumption 2.1(ii) as they apply to any discrete random variables Z, W, D, and any random variable Y.

3.1 Intensity Preserving Proxy

We begin by providing a definition for a property that a proxy for an instrument may have.

Definition 3.1 *Intensity Preserving Proxy: Let Z and W be as in Assumption 2.1(i) and put $\Lambda_{k,l} := \Pr(Z \leq k|W = l)$, $k \in S_Z$, $l \in S_W$. For $l, l' \in S_W$, $l < l'$, we say that that the proxy W at (l, l') is positively (respectively negatively) intensity preserving for the instrument Z if $\Lambda_{k,l} \geq \Lambda_{k,l'}$ (respectively $\Lambda_{k,l} \leq \Lambda_{k,l'}$) for all $k \in S_Z$. We say that W is positively (respectively negatively) intensity preserving for Z if for all $l, l' \in S_W$, such that $l < l'$, W at (l, l') is positively (respectively negatively) intensity preserving for Z. Last, we say that W (at (l, l')) is intensity preserving for Z if W (at (l, l')) is either positively or negatively intensity preserving for Z.*

The notion that W is positively (respectively negatively) intensity preserving for Z corresponds to the definition in Lehmann (1966) of Z being either "positively (respectively "negatively") regression dependent" on W. Intensity preservation is thus a restriction on the probability distribution of (Z, W). If a proxy W is positively intensity preserving for Z then knowledge of W being small increases the probability of Z being small. Intensity preservation at (l, l') is the corresponding local notion.

Often researchers use a deterministic function of the instrument $W = f(Z)$ as a proxy. Then a sufficient but not necessary condition for W to be intensity preserving for Z is that f is monotonic.

Proposition 3.1 *Let Z and W be as in Assumption 2.1(i) and for $f : S_Z \to S_W$ let $W = f(Z)$. Suppose that f is an increasing (respectively decreasing) monotonic function. Then W is positively (respectively negatively) intensity preserving for Z. Further, monotonicity of f is not necessary for W to be intensity preserving for Z.*

3.2 ILS Using A Proxy

The next result establishes a relation between $\beta_{l,l'}$ and $\alpha_{k-1,k}$, $k \in S_Z|\{0\}$.

Theorem 3.2 *ILS Using a Proxy: Suppose that Assumption 2.1(i) and 2.5 hold. Let* $\lambda_{k,l} := \Pr(Z = k|W = l)$, $k \in S_Z$, $l \in S_W$*, and suppose that* $E(D|Z = k-1) \neq E(D|Z = k)$, $k = 1, ..., K$ *and* $\sum_{k=0}^{K}(\lambda_{k,l'} - \lambda_{k,l})E(D|Z = k) \neq 0$*. Then* $\beta_{l,l'}$, $l, l' \in S_W$, $l < l'$*, exists, is finite, and is equal to a weighted average of* $\alpha_{k-1,k}$*:*

$$\beta_{l,l'} = \sum_{k=1}^{K} \nu_k^{l,l'} \alpha_{k-1,k},$$

with weights

$$\nu_k^{l,l'} := \frac{[E(D|Z = k) - E(D|Z = k-1)][\sum_{m=k}^{K}(\lambda_{m,l'} - \lambda_{m,l})]}{\sum_{p=1}^{K}[E(D|Z = p) - E(D|Z = p-1)][\sum_{m=p}^{K}(\lambda_{m,l'} - \lambda_{m,l})]},$$

and $\sum_{k=1}^{K} \nu_k^{l,l'} = 1$*. Further, order the elements of* S_Z *such that* $E(D|Z = k-1) < E(D|Z = k)$*, then* $0 \leq \nu_1^{l,l'}, ..., \upsilon_K^{l,l'} \leq 1$ *if and only if* W *at* (l, l') *is intensity preserving for* Z*.*

Theorem 3.2 demonstrates the consequences under Assumption 2.5 on the ILS estimand of substituting a proxy W for an instrument Z. Theorem 3.2 applies to any discrete random variables Z, W, D, and any random variable Y. In particular, Z need not be a valid instrument here. The result shows that $\beta_{l,l'}$, the ILS estimand using the proxy W evaluated at l and l', is a weighted average of the ILS estimands formed using the instrument Z and evaluated at adjacent points of S_Z. The condition $\sum_{k=0}^{K}(\lambda_{k,l'} - \lambda_{k,l})E(D|Z = k) \neq 0$ is clearly violated if $Z \perp W$ since then $\lambda_{k,l'} = \lambda_{k,l}$ for all $k \in S_Z$ and $l, l' \in S_W$. While the weights sum up to 1, they could be negative. It follows that $\beta_{l,l'}$ could be negative or equal to 0 even when $\alpha_{k-1,k} > 0$ for all $k \in S_Z \backslash \{0\}$. When the instrument is suitably relevant so that $E(D|Z = k-1) < E(D|Z = k)$ for $k = 1, ..., K$, a necessary and sufficient condition for the weights to be positive is that W at (l, l') is intensity preserving for Z.

4 Identification of Local Causal Effects via ILS using a Proxy

We employ the results of Section 3 to study the identification of local causal effects via ILS using a proxy for an instrument.

Proposition 4.1 *Identification via ILS Using a Proxy: (a) Suppose that Assumptions 2.1, and 2.2 hold, and that Assumptions 2.3 and 2.4 hold for $k, k' \in S_Z$ $(k < k')$. Then $\alpha_{k,k'}$ exists, is finite, and is equal to the weighted conditional average causal effect $\alpha^*_{k,k'}$:*

$$\alpha_{k,k'} = \sum_{j=1}^{J} \omega_j E[Y(j) - Y(j-1)|D(k) < j \leq D(k')] := \alpha^*_{k,k'},$$

with

$$\omega_j := \frac{\Pr[D(k) < j \leq D(k')]}{\sum_{j=1}^{J} \Pr[D(k) < j \leq D(k')]},$$

and where $0 \leq \omega_j \leq 1$ and $\sum_{j=1}^{J} \omega_j = 1$.

(b) Suppose further that Assumptions 2.3 and 2.4 hold for all $k-1$ and k, $k \in S_Z|\{0\}$, that Assumption 2.5 holds, and that $\sum_{k=0}^{K} (\lambda_{k,l'} - \lambda_{k,l}) E(D|Z = k) \neq 0$. Then $\beta_{l,l'}$ exists, is finite, and is identified with a weighted average of $\alpha^*_{k-1,k}$:

$$\beta_{l,l'} = \sum_{k=1}^{K} \nu_k^{l,l'} \alpha^*_{k-1,k},$$

where $\sum_{k=1}^{K} \nu_k^{l,l'} = 1$. Further, $0 \leq \nu_1^{l,l'}, ..., \upsilon_K^{l,l'} \leq 1$ if and only if W at (l, l') is intensity preserving for Z.

Proposition 4.1(a) directly extends theorem 1 of AI with binary instrument to accommodate a generally multivalued instrument. As AI demonstrate, ILS using an observed instrument identifies a weighted average of local average treatments effects for subpopulations of individuals whose treatment receipt is induced by the instrument. AI refer to $\alpha^*_{0,1}$ in the case of binary observed instrument as the "average causal response." When Z is unobserved $\alpha_{k,k'}$ is not directly estimable. Part (b) of Proposition 4.1 demonstrates the consequences of using a proxy for an instrument on the identification of local effects via ILS. Under the stated assumptions, the ILS estimand $\beta_{l,l'}$ is a weighted average of "average causal responses" evaluated at adjacent points of the support of the instrument. It follows that $\beta_{l,l'}$ is a weighted average of local average treatments effects for subpopulations of individuals whose treatment receipt is induced by the instrument. Since the weights could be negative, $\beta_{l,l'}$ may be negative even when all the local average effects are positive. Assumptions 2.2 and 2.3 -2.4 for all $k-1, k \in S_Z$, ensure that $E(D|Z=0) < E(D|Z=1) < ... < E(D|Z=K)$. Thus, a necessary and sufficient condition for the weights to be nonnegative is that W at (l, l') is intensity preserving for Z. This holds for example when the proxy is a monotonic function of the instrument as shown in Proposition 3.1. Next, we examine the consequences of Theorem 4.1 for special cases of interest.

4.1 The Compliance Problem With Proxy for Unobserved Treatment Assignment

Consider the case in which $J = K = L = 1$ so that the treatment assignment (instrument), proxy, and treatment can each take on only two values. Suppose that realizations of the instrument are observed without error. Proposition 4.1(a) gives that under the stated assumptions the ILS estimand $\alpha_{0,1}$ exists, is finite, and is identified with the local average treatment effect (LATE) $\alpha^*_{0,1}$:

$$\alpha_{0,1} := \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)} = E[Y(1) - Y(0)|D(1) - D(0) = 1] := \alpha^*_{0,1}.$$

As discussed in AIR, in this case, $\alpha_{0,1}$ is identified with $\alpha^*_{0,1}$, the average treatment effect for the subpopulation of "compliers," the individuals who always comply with the treatment assignment.

Suppose now that the econometrician only observes realizations of the proxy W for the instrument Z. Then under Assumptions 2.1 -2.4 condition $\sum_{k=0}^{1}(\lambda_{k,1} - \lambda_{k,0})E(D|Z=k) \neq 0$ reduces to $Z \not\perp W$ or $(\lambda_{1,1} - \lambda_{1,0}) \neq 0$, and given Assumption 2.5 ILS using the proxy W identifies the LATE, $\alpha^*_{0,1}$:

$$\beta_{0,1} := \frac{E(Y|W=1) - E(Y|W=0)}{E(D|W=1) - E(D|W=0)} = E[Y(1) - Y(0)|D(1) - D(0) = 1] := \alpha^*_{0,1}.$$

Indeed multiplying the numerator and denominator in the expression for $\alpha_{0,1}$ by $(\lambda_{1,1} - \lambda_{1,0})$ gives the expression for $\beta_{0,1}$ (see also Hernán and Robins, 2006, theorem 5). The result demonstrates the robustness of ILS methods to certain types of measurement error in the instrument for the purpose of identification of LATE. Examples include the case of "error at random" in recording the treatment assignment. Another example is the case of imputed "missing at random" assignment values.

We emphasize that the proxy need not satisfy all the assumptions required for the instrument. In particular, the proxy need not be individualistic: an individual's proxy could depend on other individuals' assignment. Also, importantly, the proxy need not cause the treatment nor be exogenous. Instead, it suffices that $D \perp_m W|Z$, $Y \perp_m W|Z$, and that $Z \not\perp W$. This significantly relaxes AIR's conditions to identify LATE: any suitable proxy for the instrument can be used to identify the LATE.

4.2 Multivalued Proxy for Binary Instrument

It is also of interest to consider situations in which the instrument is binary and a proxy for it is multivalued so that $K = 1 < L$. For example, the proxy may code instrument observations that are missing at random. To illustrate, the instrument Z may be a binary treatment assignment, and the proxy W may take on three values $S_W = \{0, 1, 2\}$ corresponding to being assigned to the control group ($W = 0$), to the treatment group ($W = 1$), or to missing assignment information ($W = 2$). Suppose further that the treatment is binary, $S_D = \{0, 1\}$. In this case, Proposition 4.1 gives that, under the stated assumptions, the ILS estimand $\beta_{l,l'}$ evaluated at *any* $l, l' \in S_W$ is identified with the LATE $\alpha^*_{0,1}$:

$$\beta_{l,l'} := \frac{E(Y|W = l') - E(Y|W = l)}{E(D|W = l') - E(D|W = l)} = E[Y(1) - Y(0)|D(1) - D(0) = 1] := \alpha^*_{0,1}.$$

Under Assumptions 2.1-2.4, the condition $\sum_{k=0}^{1}(\lambda_{k,l'} - \lambda_{k,l})E(D|Z = k) \neq 0$ reduces to $(\lambda_{1,l'} - \lambda_{1,l}) \neq 0$. Interestingly, $\beta_{l,l'}$ could be evaluated at any $l, l' \in S_W$ including $l' = 2$ denoting a missing instrument observation. For example, $\beta_{1,2}$ is identified with $\alpha^*_{0,1}$ under Assumptions 2.1-2.5 provided that $(\lambda_{1,2} - \lambda_{1,1}) \neq 0$ which can hold for example if $\lambda_{1,1} = 1$

so that recorded instrument observations are correct and $\lambda_{1,2} \neq 1$ so that the missing instrument observations are not solely from the treatment group. This result demonstrates that identifying LATE via ILS is robust to certain types of instrument misclassification.

More generally, the treatment can take on $J+1$ values such as when it has multiple intensities. Examples include the number of courses an individual completes at a training program or the number of years of education an individual receives. Proposition 4.1 then gives that under our assumptions $\beta_{l,l'}$ evaluated at *any* $l, l' \in S_W$ is identified with the average causal response $\alpha^*_{0,1}$, a weighted average of LATEs:

$$\beta_{l,l'} := \frac{E(Y|W=l') - E(Y|W=l)}{E(D|W=l') - E(D|W=l)} = \sum_{j=1}^{J} \omega_j E[Y(j) - Y(j-1)|D(0) < j \leq D(1)] := \alpha^*_{0,1}.$$

4.3 Binary Proxy for Multivalued Instrument

Often econometricians employ a binary proxy for a multivalued instrument so that $K > 1 = L$. For example, the proxy may code information less finely than the instrument. Proposition 4.1 demonstrates that ILS using W identifies a weighted average of local treatment effects based on Z. For example, let $K = 2$ so that the instrument can take on three values. Suppose further that the proxy and treatment are binary $L = J = 1$. Then

$$\beta_{0,1} := \frac{E(Y|W=1) - E(Y|W=0)}{E(D|W=1) - E(D|W=0)} = \nu_1^{0,1}\alpha^*_{0,1} + \nu_2^{0,1}\alpha^*_{1,2}$$
$$= \nu_1^{0,1}E[Y(1) - Y(0)|D(0) < D(1)] + \nu_2^{0,1}E[Y(1) - Y(0)|D(1) < D(2)].$$

with

$$\nu_1^{0,1} = \frac{[E(D|Z=1) - E(D|Z=0)](\lambda_{1,1} - \lambda_{1,0} + \lambda_{2,1} - \lambda_{2,0})]}{\sum_{p=1}^{2}[E(D|Z=p) - E(D|Z=p-1)][\sum_{m=p}^{2}(\lambda_{m,1} - \lambda_{m,0})]},$$
$$\nu_2^{0,1} = \frac{[E(D|Z=2) - E(D|Z=1)][(\lambda_{2,1} - \lambda_{2,0})]}{\sum_{p=1}^{2}[E(D|Z=p) - E(D|Z=p-1)][\sum_{m=p}^{2}(\lambda_{m,1} - \lambda_{m,0})]}.$$

To illustrate, consider the training program example discussed in the Introduction. Suppose that the unobserved instrument Z denotes whether individuals are offered zero financial aid ($Z = 0$), partial financial aid in the form of waived tuition ($Z = 1$), or full financial aid covering tuition and living expenses ($Z = 2$). Suppose that the proxy records only whether individuals are offered financial aid ($W = 1$) or not ($W = 0$). Suppose that the proxy is coded without error so that $\Lambda_{0,0} = \lambda_{0,0} = 1$ and $\Lambda_{0,1} = \lambda_{0,1} = 0$. Since $\Lambda_{0,0} > \Lambda_{0,1}$ and $\Lambda_{1,0} = \lambda_{0,0} + \lambda_{1,0} = 1 > \Lambda_{1,1}$, we have that the proxy is intensity preserving for the instrument and therefore the weights are positive and given by

$$\nu_1^{0,1} = \frac{E(D|Z=1) - E(D|Z=0)}{[E(D|Z=1) - E(D|Z=0)] + [E(D|Z=2) - E(D|Z=1)]\lambda_{2,1}},$$

$$\nu_2^{0,1} = \frac{[E(D|Z=2) - E(D|Z=1)]\lambda_{2,1}}{[E(D|Z=1) - E(D|Z=0)] + [E(D|Z=2) - E(D|Z=1)]\lambda_{2,1}}.$$

Suppose instead that the proxy only records whether individuals' tuition is waived. In this case, $\lambda_{1,1} = 1$ and $\lambda_{1,0} = 0$. We then have $\Lambda_{0,0} > \Lambda_{0,1} = \lambda_{0,1} = 0$ and $\Lambda_{1,0} = \lambda_{0,0} + \lambda_{1,0} = \lambda_{0,0} < \lambda_{0,1} + \lambda_{1,1} = \Lambda_{1,1} = 1$. Since the proxy is *intensity disrupting* for the instrument, one of the weights must be negative. Indeed, the weights are given by

$$\nu_1^{0,1} = \frac{[E(D|Z=1) - E(D|Z=0)](1 - \lambda_{2,0})]}{[E(D|Z=1) - E(D|Z=0)](1 - \lambda_{2,0})] - [E(D|Z=2) - E(D|Z=1)]\lambda_{2,0}},$$

$$\nu_2^{0,1} = \frac{[E(D|Z=2) - E(D|Z=1)](-\lambda_{2,0})}{[E(D|Z=1) - E(D|Z=0)](1 - \lambda_{2,0})] - [E(D|Z=2) - E(D|Z=1)]\lambda_{2,0}},$$

and $\nu_2^{0,1}$ is negative. Now suppose that $E[D(1) - D(0)] = 2E[D(2) - D(1)] > 0$ and that $\lambda_{2,0} = \frac{1}{3}$ then $\nu_1^{0,1} = \frac{4}{3}$ and $\nu_2^{0,1} = -\frac{1}{3}$. It follows that $\beta_{0,1}$ is negative whenever $0 < \alpha^*_{0,1} < \frac{1}{4}\alpha^*_{1,2}$ in which case the sign of $\beta_{0,1}$ is the opposite of that of the two LATEs $\alpha^*_{0,1}$ and $\alpha^*_{1,2}$.

These two examples are special cases in which the proxy is a deterministic function of the instrument. This need not hold in general. For instance, error in recording the instrument is sufficient for one of the weights in the expression for $\beta_{0,1}$ to be negative. To illustrate, consider the first financial aid proxy but now suppose that $\lambda_{0,0} = 0.9$, $\lambda_{1,0} = 0.025$, $\lambda_{0,1} = 0.05$, and $\lambda_{1,1} = 0.925$. Then $\Lambda_{0,0} = 0.9 > 0.05 = \Lambda_{0,1}$ and $\Lambda_{1,0} = 0.9 + 0.025 <$

$0.05 + 0.925 = \Lambda_{1,1}$. It follows that one of the weights is negative and therefore that the sign of $\beta_{0,1}$ may be the opposite of that of the two LATEs in the expression for $\beta_{0,1}$.

As discussed in the Introduction, Angrist (1990), AIR, and Angrist and Krueger (1991) employ a binary proxy that is a monotonic function of a multivalued instrument. Thus by Proposition 3.1 these proxies are intensity preserving for their respective instruments. It follows that under the assumptions of Proposition 4.1, ILS using these proxies identifies a weighted average of the LATEs evaluated at adjacent instrument values with nonnegative weights.

The results of this section demonstrate the need to interpret $\beta_{0,1}$ carefully since it is a weighted average of LATEs. Further, even when the signs of the LATEs in the expression for $\beta_{0,1}$ coincide, $\beta_{0,1}$ may have the opposite sign if one or more of the weights are negative. This possibility arises if and only if the proxy is not intensity preserving for the instrument.

5 Identification of Local Causal Effects via IV using a Proxy For An Instrument

Often researchers employ IV methods such as TSLS for the purpose of identifying average causal effects. Consider a function $g(Z)$ of the instrument where $g : S_Z \to S_g$ is a function mapping S_Z to $S_g := \{0, ..., M\}$, $M \in \mathbb{N}^+$ finite. Then the IV estimand β_Z^g is given by:

$$\beta_Z^g := \frac{Cov(Y, g(Z))}{Cov(D, g(Z))} = \frac{E\{Y[g(Z) - E(g(Z))]\}}{E\{D[g(Z) - E(g(Z))]\}}.$$

The particular choice $g(Z) = E(D|Z)$ corresponds to the method of two stage least squares whose estimand we denote by β_Z^{TSLS}. Theorem 2 of IA demonstrates that for binary treatment with observed instrument Z and $g(\cdot)$ monotonic, β_Z^g is identified with a weighted average of LATEs $\alpha_{k-1,k}^*$, $k = 1, ..., K$ with nonnegative weights. Theorem 2 of AI generalizes this result to cover multivalued treatment in the case of TSLS where $g(Z) = E(D|Z)$.

The next theorem extends the results in IA and AI in three directions. First, we give a necessary and sufficient condition for the weights in the expression for an IV estimand β_W^g as a weighted average of ILS estimands $\beta_{l-1,l}$, $l = 1, ..., L$, to be nonnegative. Second, we show that under Assumption 2.5, the IV estimand β^g_W using a proxy W is a weighted average of ILS estimands using the instrument Z, $\alpha_{k-1,k}$, $k = 1, ..., K$ and we give a necessary and sufficient condition for the weights to be nonnegative. These two results hold for any random variables as defined in Assumption 2.1(i). Last, we show that under Assumptions 2.1-2.5, the IV estimand β^g_W using the proxy W is identified with a weighted average of LATEs $\alpha^*_{k-1,k}$, $k = 1, ..., K$.

5.1 Mean Quadrant Dependence and Concordance

The following definition of *mean quadrant dependence* is useful in stating the next theorem. Although mean quadrant dependence and intensity preservation may hold for any random variables, we state the definitions for discrete random variables to avoid introducing new notation.

Definition 5.1 *Mean Quadrant Dependence: Let W be as in Assumption 2.1(i) and let V be a random variable that takes value in S_g with $E(V)$ finite. We say that V is positively mean quadrant dependent on W if $E(V|W \leq l) \leq E(V)$ for all $l \in S_W$ and that V is negatively mean quadrant dependent on W if $E(V|W \leq l) \geq E(V)$ for all $l \in S_W$. We say that V is mean quadrant dependent on W if V is either positively or negatively mean quadrant dependent on W*

V is positively (respectively negatively) mean quadrant dependent on W if the expected value of V given that $W \leq l$ for any $l \in S_W$ is not larger (respectively not smaller) than the unconditional expected value of V. Noting that[3] for all $l \in S_W$,

$$\sum_{m=1}^{M} \Pr(V \geq m|W \leq l) = E(V|W \leq l),$$

[3] We have $\sum_{m=1}^{M} \Pr(V \geq m|W \leq l) = \sum_{m=1}^{M} \sum_{p=m}^{M} \Pr(V = p|W \leq l) = \sum_{p=1}^{M} \sum_{m=1}^{p} \Pr(V = p|W \leq l) = \sum_{p=1}^{M} p \Pr(V = p|W \leq l) = E(V|W \leq l)$.

a sufficient condition for V to be mean quadrant dependent on W is that (W, V) are either "positively quadrant dependent" so that $\Pr(V \geq m|W \leq l) \leq \Pr(V \geq m)$ for all $l \in S_W$ and $m \in S_g$ or "negatively quadrant dependent" so that the previous condition holds with the sign reversed (Lehmann, 1966). In turn, a sufficient condition for this is that either W is intensity preserving for V (so that $\Pr(V \geq m|W = l)$ is monotone in l for all $m \in S_g$) or conversely that V is intensity preserving for W (Lehmann, 1966, lemma 4).

Using argument similar to lemma 4 in Lehmann (1966), the next proposition gives two increasingly stronger conditions than mean quadrant dependence.

Proposition 5.1 *Let W and V be as in Definition 5.1. Consider the following statements:*

$$\begin{aligned}
(a) \quad E(V|W &= l) \text{ is non-decreasing in } l \in S_W, \\
(b) \quad E(V|W &\leq l) \leq E(V|W \leq l') \text{ for all } l, l' \in S_W,\ l < l', \text{ and} \\
(c) \quad E(V|W &\leq l) \leq E(V) \text{ for all } l \in S_W.
\end{aligned}$$

We have (a)$\Rightarrow$(b)$\Rightarrow$(c).

As discussed in Section 2, Assumptions 2.1, 2.2, and 2.4 for all $k-1, k,\ k \in S_Z$, imply that $E(D|Z = l)$ is monotone in l for all $l \in S_Z$, a form of dependence between D and Z analogous to condition (a).

Definition 5.1 is a restriction on the joint distribution of (W, V). We are interested in the special case in which $V = g(W)$. The next proposition demonstrates that monotonicity of g is sufficient but not necessary for mean quadrant dependence.

Proposition 5.2 *Let W be as in Assumption 2.1(i) and let $g : S_W \to S_g$ be an increasing (respectively decreasing) monotonic function such that $E(g(W))$ exists and is finite. Then $g(W)$ is positively (respectively negatively) mean quadrant dependent on W. Further, monotonicity of g is not necessary for $g(W)$ to be mean quadrant dependent on W.*

As $g(W)$ is a function of a proxy for an instrument, we are concerned with the joint probability distribution of $(Z, W, g(W))$ and particularly with the following notion of dependence among the instrument Z, proxy W, and $g(W)$.

Definition 5.2 *Concordance: Let Z and W be as in Assumption 2.1(i) and let $g : S_W \to S_g$ with $E(g(W))$ finite. We say that the instrument Z and a function $g(W)$ of the proxy covary with W concordantly if either $[E(g(W)) - E(g(W)|W \leq l-1)][\Lambda_{k,l-1} - \Lambda_{k,l}]$ is nonnegative for all $(l,k) \in \{1, ..., L\} \times S_Z$ or nonpositive for all $(l,k) \in \{1, ..., L\} \times S_Z$.*

Clearly, a sufficient condition for Z and $g(W)$ to covary with W concordantly is that W is intensity preserving for Z and that $g(W)$ is mean quadrant dependent on W.

5.2 Identification of Local Causal Effects via IV using a Proxy

Part (a) of the next theorem demonstrates that the IV estimand using a function $g(W)$ of a suitable proxy W for Z is a weighted average of ILS estimands using the instrument Z and provides necessary and sufficient conditions for the weights to be nonnegative. Part (b) further demonstrates that under our assumptions, IV identifies a weighted average of local treatment effects.

Theorem 5.3 *Identification via IV Using a Proxy: (a) Suppose that the hypothesis of Theorem 3.2 holds for all $l-1$ and l, $l \in S_W|\{0\}$, that $E(g(W))$ exists and is finite, and that for $\pi_q := \Pr(W = q)$, $q \in S_W$, $\sum_{q=0}^{L} \pi_q E(D|W = q)[g(q) - E(g(W))] \neq 0$. Then the IV estimand β_W^g exists, is finite, and is equal to a weighted average of $\alpha_{k-1,k}$:*

$$\beta_W^g = \sum_{l=1}^{L} \mu_l^g \beta_{l-1,l} = \sum_{l=1}^{L} \sum_{k=1}^{K} \phi_{l,k}^g \alpha_{k-1,k},$$

where

$$\mu_l^g = [E(D|W = l) - E(D|W = l-1)] \frac{\sum_{q=l}^{L} \pi_q [g(q) - E(g(W))]}{\sum_{q=0}^{L} \pi_q E(D|W = q)[g(q) - E(g(W))]},$$

and

$$\phi^g_{l,k} := \mu^g_l . \nu^{l-1,l}_k,$$

with $\sum_{l=1}^{L} \mu^g_l = \sum_{k=1}^{K} \nu^{l-1,l}_k = \sum_{l=1}^{L}\sum_{k=1}^{K} \phi^g_{l,k} = 1$. *Further, order the elements of* S_W *such that* $E(D|W = l-1) < E(D|W = l)$, *then* $0 \leq \mu^g_1, ..., \mu^g_L \leq 1$ *if and only if* $g(W)$ *is mean quadrant dependent on* W. *If also the elements of* S_Z *are ordered as in Theorem 3.2, then* $0 \leq \phi^g_{1,1}, \phi^g_{1,2}, ..., \phi^g_{2,1}, ...\phi^g_{L,K} \leq 1$ *if and only if the instrument* Z *and* $g(W)$ *covary with the proxy* W *concordantly.*

(b) Suppose that Assumptions 2.1, 2.2, and 2.5 hold and that Assumptions 2.3 and 2.4 hold for all $k-1$ *and* k, $k \in S_Z|\{0\}$. *Suppose further that* $\sum_{k=0}^{K}(\lambda_{k,l} - \lambda_{k,l-1})E(D|Z = k) \neq 0$ *for all* $l \in \{1, ..., L\}$ *and* $\sum_{m=0}^{L} \pi_m E(D|W = m)[g(m) - E(g(W))] \neq 0$. *Then the IV estimand* β^g_W *exists, is finite, and is identified with a weighted average of* $\alpha^*_{k-1,k}$*:*

$$\beta^g_W := \sum_{l=1}^{L}\sum_{k=1}^{K} \phi_{l,k}\alpha^*_{k-1,k}.$$

If the elements of S_W *are ordered as in (a) then* $0 \leq \phi^g_{1,1}, \phi^g_{1,2}, ..., \phi^g_{2,1}, ...\phi^g_{L,K} \leq 1$ *if and only if the instrument* Z *and* $g(W)$ *covary with the proxy* W *concordantly.*

Observe that for $L = 1$ and any function g, β^g_W is equal to $\beta_{0,1}$. Further, the concordance condition in Theorem 5.3(a) for the weights $\phi^g_{1,k}$ to be nonnegative restricts the sign of $[E(g(W)) - E(g(W)|W \leq 0)[\Lambda_{k,0} - \Lambda_{k,1}]$ or equivalently the sign of $\Lambda_{k,0} - \Lambda_{k,1}$, to be the same for all $k \in S_Z$. In this case, the results in Theorem 5.3 reduce to the results in Proposition 4.1.

Theorem 2 of IA provides a sufficient condition for $\mu^g_1, ..., \mu^g_L$ to be nonnegative, namely that $E(D|W = l-1) < E(D|W = l)$ for all $l \in S_W|\{0\}$ and that $g(W)$ is monotone in W. Indeed, Proposition 5.2 shows that monotonicity of g is sufficient for $g(W)$ to be mean quadrant dependent on W and thus for $\mu^g_1, ..., \mu^g_L$ to be nonnegative given the ordering of the elements of S_W. Also, theorem 2 of AI considers the special case of TSLS in which $g(W) = E(D|W)$. In this case, ordering the elements of S_W such that $E(D|W =$

$l-1) < E(D|W=l)$ gives that $g(W)$ is monotonic in W and therefore that the weights $\mu_1^{TSLS}, ..., \mu_L^{TSLS}$ are nonnegative.

Similar to the discussion in Section 4, because the IV estimand is a weighted average of LATE with potentially negative weights, one should carefully interpret estimates for causal effects obtained via IV using a function of an instrument or of a proxy for it.

6 Further Comments

6.1 Nonseparable Structural System with Discrete Variables

For linear structural equations, IV using a measurement that is "valid," that is uncorrelated with the unobserved causes of the response Y, and "relevant," that is correlated with the causes of interest D, identifies the causal effect of D on Y. As pointed in Heckman (1996, p. 460), this measurement need not be uncorrelated with the unobserved drivers of D. In particular, it need not be exogenous in the reduced form equation. Accordingly, Chalak and White (2009) distinguish between observed exogenous instruments (OXI) and proxies for unobserved exogenous instruments (PXI). As they show, for linear structural systems, IV methods using a valid and relevant proxy for an unobserved exogenous instrument identifies the causal effect of the treatment on the response.

Schennach, White, and Chalak (2009) study the method of ILS in a general nonseparable structural equations system with continuous cause, response, and instrument. They demonstrate that ILS using an observed exogenous instrument identifies a weighted average marginal effect with some of the weights possibly negative. They show that in special cases, such as when the treatment is separably determined, this weighted average effect reduces to an instrument-conditioned average marginal effect. Further, they show that the availability of two or more suitable proxies for an unobserved exogenous instrument permits estimating this weighted average effect.

This paper studies a general nonseparable structural system with discrete instrument, proxy, cause, and a general response. We demonstrate that ILS and IV methods using a proxy for an instrument generally identify a weighted average of local causal effects based on the instrument. This analysis thus complements the results in Chalak and White (2009) and Schennach, White, and Chalak (2009). It is of interest to study the consequences in the continuous case (see e.g. Hernán and Robins, 2006, theorem 6) of the dependence properties between the proxy and the instrument introduced here.

6.2 Implications for Testing

The results of this paper provide a way to test two hypotheses of interest. First, the availability of a proxy for an unobserved instrument permits testing the hypothesis of no causal effect of the treatment. In particular, under this hypothesis the weighted average effect identified by ILS using a proxy should be zero. This hypothesis could also be tested via IV using a function of a proxy. Second, the availability of one multivalued proxy or of two binary proxies for a binary instrument permits identifying the same LATE via ILS using the multivalued instrument evaluated at any two adjacent points of its support or using any of the two binary proxies. This provides the foundation for tests of the underlying assumptions. We leave developing these tests for future work.

7 Conclusion

We study the scope of the method of indirect least squares (ILS) and of instrumental variables (IV) methods for the identification of local effects of an endogenous cause D on an outcome of interest Y using a proxy W for a possibly unobserved instrument Z. We study the case in which the instrument, proxy, and treatment are discrete; the outcome need not be discrete. ILS using a suitable proxy W for an instrument Z identifies a weighted average of local treatment effects based on the instrument Z. The weights are nonnegative if and only if the proxy is intensity preserving for the instrument. Similarly, IV using a function of a proxy for an instrument identifies a weighted average of local treatment effects based on the instrument and the weights are nonnegative if and only if the instrument Z and $g(W)$ covary with the proxy W concordantly. Thus, researchers should carefully interpret estimates for causal effects obtained via ILS or IV, such as the method of two stage least squares, using an error-laden proxy for an instrument, a proxy for an instrument with missing or imputed observations, or a binary proxy for a multivalued instrument for example. Positively, the proxy need not satisfy all the assumptions required for the instrument. In particular, the proxy need not be individualistic, it can depend on other individuals' instrument. Importantly, unlike the instrument, the proxy need not cause the treatment nor be exogenous. This is particularly useful because in special circumstances such as when the instrument is binary, ILS using any suitable proxy identifies local average treatment effects. Last, it is of interest to study the consequences in the continuous case of the dependence between a proxy and an instrument studied here. Also, it is of interest to develop tests for the underlying assumptions and for the hypothesis of no causal effect based on the availability of one or more proxies for an instrument. We leave this for future research.

Mathematical Appendix

Proof of Proposition 2.1: We refer to lemmas 4.1, 4.2, and 4.3 of Dawid (1979) in what follows. To show that $(Z, W) \perp (U_D, U_Y)$ is sufficient for Assumption 2.2 to hold, lemma 4.2(i) gives $(q(z, U_D), r(q(z, U_D), U_Y)) \perp Z$ and thus that $(D(z), Y(D(z))) \perp Z$ which implies Assumption 2.2. To show that $(Z, W) \perp (U_D, U_Y)$ implies Assumption 2.5, we make use of the converse of lemma 4.3 which, as stated in Dawid (1979, p. 5), holds. This gives $(U_D, U_Y) \perp W|Z$. Applying lemma 4.1 gives $(Z, U_D, U_Y) \perp (Z, W)|Z$. Then lemma 4.2(i) gives $(q(Z, U_D), r(q(Z, U_D), U_Y)) \perp W|Z$ or $(D, Y) \perp W|Z$ which implies Assumption 2.5. ■

Proof of Proposition 3.1: We first prove sufficiency. Suppose that f is an increasing monotonic function. We have

$$\Lambda_{k,l} := \Pr(Z \leq k|W = l) = \begin{cases} 0 \text{ if } f(k) < l \\ c \in (0,1] \text{ if } f(k) = l \\ 1 \text{ if } f(k) > l \end{cases}$$

since for $f(k) = l$, there may exist $k_1, ... < k_h$, $h \leq K$, such that $f(k_1) = ... = f(k_h) = l$.

Then, for arbitrary $l < l'$,

$$\Lambda_{k,l} - \Lambda_{k,l'} = \begin{cases} 0 \text{ if } f(k) < l < l' \\ c \in (0,1] \text{ if } l \leq f(k) < l' \\ 1 - c',\ c' \in (0,1],\ \text{if } l < l' \leq f(k) \end{cases}$$

and therefore for all $l, l' \in S_W$, $l < l'$, we have $\Lambda_{k,l} \geq \Lambda_{k,l'}$ for all $k \in S_Z$.

Similar arguments shows that if f is a decreasing monotonic function then for all $l, l' \in S_W$, $l < l'$, we have $\Lambda_{k,l} \leq \Lambda_{k,l'}$ for all $k \in S_Z$.

We provide an example to show that the monotonicity of f is not necessary for W to be intensity preserving for Z. Suppose that $K = 3$, and that Z is uniformly distributed so that $P(Z = k) = \frac{1}{4}$ for $k = 0, 1, 2, 3$. Suppose that

$$W = \begin{cases} 0 \text{ if } Z = 0, 2 \\ 1 \text{ if } Z = 1, 3 \end{cases}.$$

We have

$$\Lambda_{k,l} := \Pr(Z \leq k|W = l) = \sum_{p=0}^{k} \Pr(Z = p|W = l),$$

and by Bayes' theorem

$$\Pr(Z = p|W = l) = \frac{\Pr(W = l|Z = p)\Pr(Z = p)}{\sum_{k=0}^{K} \Pr(W = l|Z = k)\Pr(Z = k)}.$$

Thus $\Lambda_{0,0} = \frac{1}{2} > 0 = \Lambda_{0,1}$, $\Lambda_{1,0} = \frac{1}{2} = \Lambda_{1,1}$, $\Lambda_{2,0} = 1 > \frac{1}{2} = \Lambda_{2,1}$, and $\Lambda_{3,0} = 1 = \Lambda_{3,1}$ and therefore W is intensity preserving for Z. ■

Proof of Theorem 3.2: Assumption 2.1(i) ensures that $E(Y)$ exist and is finite and thus $E(Y|Z = k)$ and $E(Y|W = l)$ exit and are finite for all $k \in S_Z$ and $l \in S_W$. By

Assumption 2.5

$$
\begin{aligned}
E(Y|W &= l) = \sum_{m=0}^{K} E(Y|W=l, Z=m) \Pr(Z=m|W=l) \\
&= \sum_{m=0}^{K} E(Y|Z=m)\lambda_{m,l}.
\end{aligned}
$$

We can write

$$
\begin{aligned}
E(Y|W &= l') - E(Y|W=l) \\
&= \sum_{m=0}^{K} (\lambda_{m,l'} - \lambda_{m,l}) E(Y|Z=m) \\
&= (\lambda_{0,l'} - \lambda_{0,l}) E(Y|Z=0) + \sum_{m=1}^{K} (\lambda_{m,l'} - \lambda_{m,l}) E(Y|Z=m) \\
&= ((1 - \sum_{m=1}^{K} \lambda_{m,l'}) - (1 - \sum_{m=1}^{K} \lambda_{m,l})) E(Y|Z=0) + \sum_{m=1}^{K} (\lambda_{m,l'} - \lambda_{m,l}) E(Y|Z=m) \\
&= \sum_{m=1}^{K} (\lambda_{m,l'} - \lambda_{m,l})[E(Y|Z=m) - E(Y|Z=0)].
\end{aligned}
$$

Substituting for

$$
E(Y|Z=m) - E(Y|Z=0) = \sum_{k=1}^{m} [E(Y|Z=k) - E(Y|Z=k-1)],
$$

we obtain

$$
\begin{aligned}
E(Y|W=l') - E(Y|W=l) &= \sum_{m=1}^{K} (\lambda_{m,l'} - \lambda_{m,l}) \sum_{k=1}^{m} [E(Y|Z=k) - E(Y|Z=k-1)] \\
&= \sum_{k=1}^{K} [E(Y|Z=k) - E(Y|Z=k-1)][\sum_{m=k}^{K} (\lambda_{m,l'} - \lambda_{m,l})].
\end{aligned}
$$

Similarly, by Assumption 2.1(i), $E(D)$ exists, is finite and thus $E(D|Z=k)$ and $E(D|W=l)$ exist and are finite for all $k \in S_Z$ and $l \in S_W$. By Assumption 2.5, a similar derivation as above gives

$$
\begin{aligned}
E(D|W &= l') - E(D|W=l) = \sum_{k=0}^{K} (\lambda_{k,l} - \lambda_{k,l'}) E(D|Z=k) \\
&= \sum_{k=1}^{K} [E(D|Z=k) - E(D|Z=k-1)][\sum_{m=k}^{K} (\lambda_{m,l'} - \lambda_{m,l})].
\end{aligned}
$$

Given that $E(D|Z=k-1) \neq E(D|Z=k)$, $k=1,...,K$, we have

$$E(Y|W=l')-E(Y|W=l) = \sum_{k=1}^{K} \alpha_{k-1,k}[E(D|Z=k)-E(D|Z=k-1)][\sum_{m=k}^{K}(\lambda_{m,l'}-\lambda_{m,l})].$$

Since $\sum_{k=0}^{K}(\lambda_{k,l'}-\lambda_{k,l})E(D|Z=k) \neq 0$, we have that $\beta_{l,l'}$ exists, is finite, and is given by

$$\begin{aligned}
\beta_{l,l'} \quad : \quad &= \frac{E(Y|W=l')-E(Y|W=l)}{E(D|W=l')-E(D|W=l)} \\
&= \frac{\sum_{k=1}^{K}\alpha_{k-1,k}[E(D|Z=k)-E(D|Z=k-1)][\sum_{m=k}^{K}(\lambda_{m,l'}-\lambda_{m,l})]}{\sum_{p=1}^{K}[E(D|Z=p)-E(D|Z=p-1)][\sum_{m=p}^{K}(\lambda_{m,l'}-\lambda_{m,l})]} \\
&= \sum_{k=1}^{K}\nu_k^{l,l'}\alpha_{k-1,k}.
\end{aligned}$$

It is immediate that the weights are such that $\sum_{k=1}^{K}\nu_k^{l,l'} = 1$. All weights $\nu_k^{l,l'}$, $k = 1,...,K$, are nonnegative if there does not exist $\nu_k^{l,l'}, \nu_{k'}^{l,l'}$, $k, k' \in \{1,...,K\}$ with numerators of opposite sign. If $E(D|Z=k)-E(D|Z=k-1) > 0$, this is equivalent to either $\sum_{m=k}^{K}\lambda_{m,l'} \geq \sum_{m=k}^{K}\lambda_{m,l}$ for all $k \in \{1,...,K\}$ or $\sum_{m=k}^{K}\lambda_{m,l'} \leq \sum_{m=k}^{K}\lambda_{m,l}$ for all $k \in \{1,...,K\}$. Noting that $\sum_{m=k}^{K}\lambda_{m,l} = 1-\Lambda_{k-1,l}$, for $k=1,...,K$ and $l \in S_W$ and that $\Lambda_{K,l} = 1$ for all $l \in S_W$ completes the proof. ■

Proof of Proposition 4.1: (a) The proof directly extends the proof of theorem 1 of Angrist and Imbens (1995) for binary Z to the general multivalued case, $K \geq 1$. Let $I(A)$ denote the indicator function of the event A and define $\rho_{k,j} := I(D(k) \geq j)$ for $k \in S_Z$ and $j \in S_D \cup \{J+1\}$, then $\rho_{k,0} = 1$ and $\rho_{k,J+1} = 0$ for all $k \in S_Z$. Let $Z^k = I(Z=k)$. Given Assumption 2.1, we write

$$Y = \sum_{k=0}^{K} Z^k Y(D(k)).$$

and

$$Y(D(k)) = \sum_{j=0}^{J}(\rho_{k,j}-\rho_{k,j+1})Y(j).$$

Given Assumption 2.1(i), $E(Y|Z=k)$ exist and is finite for all $k \in S_Z$. Using the above expression for Y, we can write:

$$\begin{aligned} E(Y|Z &= k') - E(Y|Z=k) \\ &= E[\sum_{k=0}^{K} Z^k Y(D(k))|Z=k'] - E[\sum_{k=0}^{K} Z^k Y(D(k))|Z=k] \\ &= E[Y(D(k'))|Z=k'] - E[Y(D(k))|Z=k]. \end{aligned}$$

Under Assumption 2.2, we have $Y(D(k)) \perp_m Z$ for all $k \in S_Z$ and thus

$$E[Y(D(k'))|Z=k'] - E[Y(D(k))|Z=k] = E[Y(D(k')) - Y(D(k))].$$

Using the expression for $Y(D(k))$ we have

$$\begin{aligned} & E[Y(D(k')) - Y(D(k))] \\ = \;& E[\sum_{j=0}^{J} (\rho_{k',j} - \rho_{k',j+1} - \rho_{k,j} + \rho_{k,j+1}) Y(j)] \\ = \;& E[(\rho_{k',0} - \rho_{k,0}) Y(0) + \sum_{j=1}^{J} (\rho_{k',j} - \rho_{k,j})(Y(j) - Y(j-1)) - (\rho_{k',J+1} - \rho_{k,J+1}) Y(J)] \\ = \;& E[\sum_{j=1}^{J} (\rho_{k',j} - \rho_{k,j})(Y(j) - Y(j-1))], \end{aligned}$$

where the last equation follows because $\rho_{k,0} = 1$ and $\rho_{k,J+1} = 0$ for all $k \in S_Z$.

Assumption 2.4 ensures that for $k, k' \in S_Z, k < k'$, $\rho_{k',j} - \rho_{k,j}$ equals 0 or 1 with probability 1 for all $j \in S_D$ and we can write:

$$\begin{aligned} E(Y|Z &= k') - E(Y|Z=k) = \sum_{j=1}^{J} E(Y(j) - Y(j-1)|\rho_{k',j} - \rho_{k,j} = 1) \Pr(\rho_{k',j} - \rho_{k,j} = 1) \\ &= \sum_{j=1}^{J} E[Y(j) - Y(j-1)|D(k) < j \leq D(k')] \Pr[D(k) < j \leq D(k')]. \end{aligned}$$

Similarly, under Assumption 2.1, we write

$$\begin{aligned} D &= \sum_{k=0}^{K} Z^k D(k), \quad \text{and} \\ D(k) &= \sum_{j=0}^{J} (\rho_{k,j} - \rho_{k,j+1}) j. \end{aligned}$$

Given Assumption 2.1(i), $E(D|Z=k)$ exist and is finite for all $k \in S_Z$. Using the expression for D, we write:

$$\begin{aligned} E(D|Z &= k') - E(D|Z=k) \\ &= E(D(k')|Z=k') - E(D(k)|Z=k). \end{aligned}$$

Under Assumption 2.2, we have $D(k) \perp_m Z$ for $k \in S_Z$. Then

$$E(D(k')|Z=k') - E(D(k)|Z=k) = E(D(k') - D(k)).$$

Substituting for the expression for $D(k)$ gives

$$\begin{aligned} & E(D(k') - D(k)) \\ = \; & E[\sum_{j=0}^{J}(\rho_{k',j} - \rho_{k',j+1} - \rho_{k,j} + \rho_{k,j+1})j] \\ = \; & E[\sum_{j=1}^{J}(\rho_{k',j} - \rho_{k,j}) - (\rho_{k',J+1} - \rho_{k,J+1})J] \\ = \; & E[\sum_{j=1}^{J}(\rho_{k',j} - \rho_{k,j})], \end{aligned}$$

where the last equality follows since $\rho_{k,J+1} = 0$ for all $k \in S_Z$. Assumption 2.4 gives that

$$E[\sum_{j=1}^{J}(\rho_{k',j} - \rho_{k,j})] = \sum_{j=1}^{J} \Pr[D(k) < j \leq D(k')],$$

and Assumption 2.3 then ensures that $\sum_{j=1}^{J} \Pr[D(k) < j \leq D(k')] > 0$. Taking the ratio of $E(Y|Z=k') - E(Y|Z=k)$ to $E(D|Z=k') - E(D|Z=k)$, it follows that $\alpha_{k,k'}$ exists, is finite, and is identified with the weighted conditional average causal effect $\alpha^*_{k,k'}$. It is immediate that $0 \leq \omega_j \leq 1$ and $\sum_{j=1}^{J} \omega_j = 1$.

(b) Given (a) and that Assumptions 2.3 and 2.4 hold for all $k, k-1$ for $k \in S_Z \backslash \{0\}$, we have that $\alpha_{k-1,k}$ exists, is finite, and is identified with $\alpha^*_{k-1,k}$, for all $k \in S_Z \backslash \{0\}$. Since $\sum_{k=0}^{K}(\lambda_{k,l'} - \lambda_{k,l})E(Y|Z=k) \neq 0$, and since Assumptions 2.1, 2.2, and 2.3-2.4 for $k-1, k$,

$k \in S_Z$ give $E(D|Z = k-1) < E(D|Z = k)$, the result obtains immediately from Theorem 3.2. ■

Proof of Proposition 5.1: Picking $l' = L$ gives that $(b) \Rightarrow (c)$. We show that $(a) \Rightarrow (b)$. For any $l, l' \in S_W$, $l < l'$, we have

$$E(V|W \leq l) = \frac{\sum_{m=0}^{M} m \Pr(V = m, W \leq l)}{\Pr(W \leq l)} \text{ and}$$
$$E(V|W \leq l') = \frac{\sum_{m=0}^{M} m \Pr(V = m, W \leq l) + \sum_{m=0}^{M} m \Pr(V = m, l < W \leq l')}{\Pr(W \leq l) + \Pr(l < W \leq l')}.$$

To show that $E(V|W \leq l) \leq E(V|W \leq l')$ it suffices to show that

$$\Pr(l < W \leq l') \sum_{m=0}^{M} m \Pr(V = m, W \leq l) \leq \Pr(W \leq l) \sum_{m=0}^{M} m \Pr(V = m, l < W \leq l').$$

Note that

$$\begin{aligned}
\sum_{m=0}^{M} m \Pr(V &= m, l < W \leq l') = \sum_{m=0}^{M} \sum_{q=l+1}^{l'} m \Pr(V = m, W = q) \\
&= \sum_{m=0}^{M} \sum_{q=l+1}^{l'} m \Pr(V = m|W = q) \Pr(W = q) \\
&= \sum_{q=l+1}^{l'} E(V|W = q) \Pr(W = q),
\end{aligned}$$

and similarly

$$\sum_{m=0}^{M} m \Pr(V = m, W \leq l) = \sum_{p=0}^{l} E(V|W = p) \Pr(W = p).$$

By (a) $E(V|W = p)$ is non-decreasing in p. Thus

$$\begin{aligned}
\Pr(l &< W \leq l') \sum_{m=0}^{M} m \Pr(V = m, W \leq l) = \Pr(l < W \leq l') \sum_{p=0}^{l} E(V|W = p) \Pr(W = p) \\
&\leq \Pr(l < W \leq l') E(V|W = l) \Pr(W \leq l) \\
&\leq \Pr(W \leq l) \sum_{q=l+1}^{l'} E(V|W = q) \Pr(W = q) = \Pr(W \leq l) \sum_{m=0}^{M} m \Pr(V = m, l < W \leq l').
\end{aligned}$$

which completes the proof. ■

Proof of Proposition 5.2: We have

$$\begin{aligned}
E(g(W)|W &\leq l) - E(g(W)) = \sum_{q=0}^{l} \frac{\pi_q}{\sum_{p=0}^{l} \pi_p} g(q) - \sum_{m=0}^{L} \pi_m g(m) \\
&= \sum_{q=0}^{l} (\frac{1}{\sum_{p=0}^{l} \pi_p} - 1) \pi_q g(q) - \sum_{m=l+1}^{L} \pi_m g(m) \\
&= \sum_{q=0}^{l} (\frac{\sum_{m=l+1}^{L} \pi_m}{\sum_{p=0}^{l} \pi_p}) \pi_q g(q) - \sum_{m=l+1}^{L} \pi_m g(m).
\end{aligned}$$

We also have

$$\sum_{q=0}^{l} (\frac{\sum_{m=l+1}^{L} \pi_m}{\sum_{p=0}^{l} \pi_p}) \pi_q = \sum_{q=0}^{l} (\frac{\pi_q}{\sum_{p=0}^{l} \pi_p}) \sum_{m=l+1}^{L} \pi_m = \sum_{m=l+1}^{L} \pi_m.$$

Now suppose that g is an increasing monotone function so that $g(l) \leq g(l')$ for $l < l'$. Let $\underline{g} = \min\{g(l+1), ...g(L)\}$ and $\bar{g} = \max\{g(0), ...g(l)\}$. Then $\underline{g} \geq \bar{g}$ and

$$\sum_{q=0}^{l} (\frac{\sum_{m=l+1}^{L} \pi_m}{\sum_{p=0}^{l} \pi_p}) \pi_q g(q) \leq \bar{g} \sum_{q=0}^{l} (\frac{\sum_{m=l+1}^{L} \pi_m}{\sum_{p=0}^{l} \pi_p}) \pi_q \leq \underline{g} \sum_{m=l+1}^{L} \pi_m \leq \sum_{m=l+1}^{L} \pi_m g(m).$$

Suppose instead that g is an decreasing monotone function so that $g(l) \geq g(l')$ for $l < l'$. Let $\underline{g} = \min\{g(0), ...g(l)\}$ and $\bar{g} = \max\{g(l+1), ...g(L)\}$. Then $\underline{g} \geq \bar{g}$ and

$$\sum_{q=0}^{l} (\frac{\sum_{m=l+1}^{L} \pi_m}{\sum_{p=0}^{l} \pi_p}) \pi_q g(q) \geq \underline{g} \sum_{q=0}^{l} (\frac{\sum_{m=l+1}^{L} \pi_m}{\sum_{p=0}^{l} \pi_p}) \pi_q \geq \bar{g} \sum_{m=l+1}^{L} \pi_m \geq \sum_{m=l+1}^{L} \pi_m g(m).$$

We provide an example to show that monotonicity of g is not necessary for $g(W)$ to be mean quadrant dependent on W. Let $K = 3$, and W be uniformly distributed so that $P(W = l) = \frac{1}{4}$ for $l = 0, 1, 2, 3$. Suppose that

$$g(W) = \left\{ \begin{array}{l} 0 \text{ if } W = 0, 2 \\ 1 \text{ if } W = 1, 3 \end{array} \right. .$$

Then $E(g(W)) = \frac{1}{2}$ and $E(g(W)|W \leq 0) = 0$, $E(g(W)|W \leq 1) = \frac{1}{2}$, $E(g(W)|W \leq 2) = \frac{1}{3}$, and $E(g(W)|W \leq 3) = \frac{1}{2}$. Thus, $E(g(W)|W \leq l) \leq E(g(W))$ for all $l \in S_W$. ■

Proof of Theorem 5.3: (a) From Theorem 3.2 we have that for all l and $l-1$, $l \in S_W|\{0\}$, $\beta_{l-1,l}$ exists, is finite and is given by $\beta_{l-1,l} = \sum_{k=1}^{K} \nu_k^{l,l'} \alpha_{k-1,k}$. Arguments similar to those in theorem 2 of IA and theorem 2 of AI give that $\beta_W^g = \sum_{l=1}^{L} \mu_l^g \beta_{l-1,l}$. We state these for completeness. We have

$$\begin{aligned} E\{Y[g(W) - E(g(W))]\} &= E\{E(Y|W)[g(W) - E(g(W))]\} \\ &= \sum_{q=0}^{L} \pi_q E(Y|W=q)[g(q) - E(g(W))], \end{aligned}$$

exists and is finite since $E(Y|W=l)$, $g(l)$, and $E(g(W))$ exist and are finite for all $l \in S_W$. Since

$$\sum_{q=0}^{L} \pi_q [g(q) - E(g(W))] = 0,$$

we can rewrite the above expression as

$$\begin{aligned} \sum_{q=0}^{L} \pi_q [E(Y|W &= q) - E(Y|W=0)][g(q) - E(g(W))] \\ &= \sum_{q=1}^{L} \pi_q [E(Y|W=q) - E(Y|W=0)][g(q) - E(g(W))] \end{aligned}$$

Substituting for

$$E(Y|W=q) - E(Y|W=0) = \sum_{l=1}^{q} E(Y|W=l) - E(Y|W=l-1),$$

gives

$$\begin{aligned} \sum_{q=1}^{L} \pi_q [\sum_{l=1}^{q} E(Y|W &= l) - E(Y|W=l-1)][g(q) - E(g(W))] \\ &= \sum_{l=1}^{L} E(Y|W=l) - E(Y|W=l-1) \sum_{q=l}^{L} \pi_q [g(q) - E(g(W))] \\ &= \sum_{l=1}^{L} \beta_{l-1,l} [E(D|W=l) - E(D|W=l-1) \sum_{q=l}^{L} \pi_q [g(q) - E(g(W))], \end{aligned}$$

given that $E(D|W=l) - E(D|W=l-1)$ for all l and $l-1$, $l \in S_W|\{0\}$.

A similar derivation shows that

$$\begin{aligned} E\{D[g(Z) - E(g(Z))]\} &= \sum_{q=0}^{L} \pi_q E(D|W=q)[g(q) - E(g(W))] \\ &= \sum_{l=1}^{L} [E(D|W=l) - E(D|W=l-1)] \sum_{q=l}^{L} \pi_q [g(q) - E(g(W))], \end{aligned}$$

exists and is finite since $E(D|W=l)$, $g(l)$, and $E(g(W))$ exist and are finite for all $l \in S_W$. Dividing the expression for $E\{Y[g(Z)-E(g(Z))]\}$ with the expression for $E\{D[g(Z)-E(g(Z))]\}$ which we assume is nonzero gives that $\beta_W^g = \sum_{l=1}^{L} \mu_l^g \beta_{l-1,l}$ where $\sum_{l=1}^{L} \mu_l^g = 1$ is obvious.

The weights μ_l^g, $l = 1, ..., L$, are nonnegative if there does not exist $\mu_l^g, \mu_{l'}^g$, $l, l' \in \{1, ..., L\}$ with numerators of opposite sign. With $E(D|W=l) - E(D|W=l-1) > 0$ for all $l \in \{1, ..., L\}$, this is equivalent to $\sum_{q=l}^{L} \pi_q[g(q) - E(g(W))]$ not having opposite signs for any $l, l' \in \{1, ..., L\}$. But

$$\begin{aligned}
&\sum_{q=l}^{L} \pi_q[g(q) - E(g(W))] \\
&= [E(g(W) - \sum_{q=0}^{l-1} \pi_q g(q)] - E(g(W)(1 - \sum_{q=0}^{l-1} \pi_q) \\
&= E(g(W)) \sum_{q=0}^{l-1} \pi_q - \sum_{q=0}^{l-1} \pi_q g(q).
\end{aligned}$$

Noting that $E(g(W)|W \leq l-1) = \sum_{q=0}^{l-1} \frac{\pi_q}{\sum_{p=0}^{l-1} \pi_p} g(q)$ and that $E(g(W)|W \leq L) = E(g(W))$ gives that $0 \leq \mu_1^g, ..., \mu_L^g \leq 1$ if and only if $E(g(W)|W \leq l) \leq E(g(W))$ for all $l \in S_W$ or $E(g(W)|W \leq l) \geq E(g(W))$ for all $l \in S_W$.

The weights $\phi_{l,k}^g$, $(l,k), (l',k') \in \{1, ..., L\} \times \{1, ..., K\}$ are nonnegative if there does not exist $\phi_{l,k}^g, \phi_{l',k'}^g$, $(l,k), (l',k') \in \{1, ..., L\} \times \{1, ..., K\}$, with numerators of opposite signs. With $E(D|W=l) - E(D|W=l-1) > 0$ for all $l-1, l \in S_W$ and $E(D|Z=k) - E(D|Z=k-1) > 0$ for all $k-1, k \in S_Z$, this is equivalent to $\sum_{q=l}^{L} \pi_q[g(q) - E(g(W))][\sum_{m=k}^{K}(\lambda_{m,l} - \lambda_{m,l-1})]$ not having opposite signs for any $(l,k), (l',k') \in \{1, ..., L\} \times \{1, ..., K\}$. Using the expression for $\sum_{q=l}^{L} \pi_q[g(q) - E(g(W))]$ from above and that $\Lambda_{k-1,l} = 1 - \sum_{m=k}^{K} \lambda_{m,l}$ we have

$$\begin{aligned}
&\sum_{q=l}^{L} \pi_q[g(q) - E(g(W))][\sum_{m=k}^{K}(\lambda_{m,l} - \lambda_{m,l-1})] \\
&= [E(g(W)) \sum_{q=0}^{l-1} \pi_q - \sum_{q=0}^{l-1} \pi_q g(q)][\Lambda_{k-1,l-1} - \Lambda_{k-1,l}].
\end{aligned}$$

Dividing by $\sum_{p=0}^{l-1} \pi_p > 0$ preserves the sign of this expression and gives

$$\begin{aligned} & [E(g(W)) - \sum_{q=0}^{l-1} \frac{\pi_q}{\sum_{p=0}^{l-1} \pi_p} g(q)][\Lambda_{k-1,l-1} - \Lambda_{k-1,l}] \\ = \; & [E(g(W)) - E(g(W)|W \leq l-1)][\Lambda_{k-1,l-1} - \Lambda_{k-1,l}]. \end{aligned}$$

Since $\Lambda_{K,l-1} = \Lambda_{K,l} = 1$ for all $l \in \{1, ..., L\}$, the weights are nonnegative if and only if $[E(g(W)) - E(g(W)|W \leq l-1)][\Lambda_{k,l-1} - \Lambda_{k,l}]$ is either nonnegative for all $(l,k) \in \{1, ..., L\} \times S_Z$ or nonpositive for all $(l,k) \in \{1, ..., L\} \times S_Z$.

(b) Under Assumptions 2.1, 2.2, and 2.5, and Assumptions 2.3-2.4 for all $k-1$ and k, $k \in S_Z|\{0\}$, and given $\sum_{k=0}^{K}(\lambda_{k,l'} - \lambda_{k,l})E(D|Z = k) \neq 0$, Proposition 4.1 gives that $\beta_{l-1,l}$ exists, is finite, and is identified as $\beta_{l-1,l} = \sum_{k=1}^{K} \nu_k^{l-1,l} \alpha_{k-1,k}^*$ for all $l \in \{1, ..., L\}$. Assumptions 2.1, 2.2, and 2.3-2.4 for all $k-1, k$, $k \in S_Z$, give $E(D|Z = k-1) < E(D|Z = k)$, the result then follows from part (a). ■

References:

Angrist, J. (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 80, 313-335.

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

Angrist, J. and G. Imbens (1995), "Two–Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431-442.

Angrist, J., G. Imbens, and D. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables," (with discussion) *Journal of the American Statistical Association*, 91, 444-472.

Chalak, K. and H. White (2008), "Independence and Conditional Independence in Structural Systems," Boston College Department of Economics Working paper.

Chalak, K. and H. White (2009), "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," Boston College Department of Economics Working paper.

Dawid, A.P. (1979), "Conditional Independence in Statistical Theory," (with Discussion) *Journal of the Royal Statistical Society*, Series B, 41, 1-31.

Heckman, J. (1996), "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association*, 91, 459-462.

Hernán, M. and J. Robins (2006), "Instruments for causal inference: An epidemiologist's dream?," *Epidemiology*, 17, 360-372.

Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 61, 467-476.

Lehmann, E. (1966), "Some Concepts of Dependence," *The Annals of Mathematical Statistics*, 37, 1137-1153.

Manski, C. (2010), "Identification of Treatment Response with Social Interactions," Northwestern University Department of Economics working paper.

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688-701.

Rubin, D. (1986), "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers,"*Journal of the American Statistical Association*, 81, 961-962.

Schennach, S. M., H. White, and K. Chalak (2009) "Local Indirect Least Squares and Average Marginal Effects in Nonseparable Structural Systems," Boston College Department of Economics Working paper.

Vytlacil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331-341.

White, H. and K. Chalak (2009), "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning," *Journal of Machine learning Research*, 10, 1759-1799.