An Extended Class of Instrumental Variables for the Estimation of Causal Effects

Karim Chalak and Halbert White

July 31, 2006[1]

Abstract: This paper builds on the structural equations, treatment effect, and machine learning literatures to provide a causal framework that permits the identification and estimation of causal effects from observational studies. We begin by providing a causal interpretation for standard exogenous regressors and standard "valid" and "relevant" instrumental variables. We then build on this interpretation to characterize *extended instrumental variables* (EIV) methods, that is methods that make use of variables that need not be valid instruments in the standard sense, but that are nevertheless instrumental in the recovery of causal effects of interest. After examining special cases of *single* and *double* EIV methods, we provide necessary and sufficient conditions for the identification of causal effects by means of EIV and provide consistent and asymptotically normal estimators for the effects of interest.

JEL Classification Numbers: C10, C20, C30, C51.

[1] First Draft: March, 2005. The authors thank Kate Antonovics, Julian Betts, Graham Elliott, Marjorie Flavin, Clive Granger, Jinyong Hahn, Keisuke Hirano, Stephen Lauritzen, Mark Machina, Judea Pearl, Dimitris Politis, Ross Starr, Ruth Williams and the participants of the UCSD applied lunch seminar, the 4th annual Advances in Econometrics conference, the 2006 North American meeting of the Econometric Society, and the UCSD and UCLA Econometrics seminars. All errors and omissions are the authors' responsibility.

Karim Chalak, Department of Economics 0534, University of California, 9500 Gilman Drive, La Jolla, CA 92093-0534, kchalak@ucsd.edu, http://dss.ucsd.edu/~kchalak

Halbert White, Department of Economics 0508, University of California, 9500 Gilman Drive, La Jolla, CA 92093-0508, hwhite@ucsd.edu, http://weber.ucsd.edu/~mbacci/white

1. Introduction

Structural equations methods are predominant in economics for modeling, identifying, and estimating causal effects of interest. The early work of the Cowles Commission by Tinbergen, Frisch, Koopmans, Haavelmo, Marschak, Simon, Wold, Strotz, and others studied identification and estimation of causal effects (e.g. Haavelmo, 1943, 1944; Simon, 1953, 1954; Strotz and Wold, 1960; Fisher 1966). This literature also introduced notions of "endogeneity" and "exogeneity." Textbooks typically define these respectively as the correlation or lack thereof between a structural equation's observed explanatory variables and its unobserved "errors." With the introduction of these concepts, it became evident that standard methods of estimation such as least squares regression fail to provide a consistent estimator for the effect of interest in the endogeneous regressor case.

Reiersøl (1945) formalized the method of "instrumental variables" (IV), originally introduced by Philip Wright (1928) building on Sewall Wright's (1921, 1923) work on "path analysis", within the structural equations framework. Ever since, this method has played a central role in handling issues of endogeneity (e.g. Hausman and Taylor 1983; Heckman, 1997; Angrist and Krueger, 2001; Heckman, Urzua, and Vytlacil, 2005). In the familiar case of linear structural equations, "proper" IVs (variables that are "exogenous" or "valid," i.e., uncorrelated with the equation's error; and "relevant", i.e. correlated with the included explanatory variables) can deliver consistent estimates of effects.

Over the years, advances across a variety of disciplines have resulted in alternative approaches to identifying and estimating causal effects in the presence of endogeneity.

In particular, developments in labor economics (Roy 1951; Heckman and Robb, 1985; Hahn, 1998; Heckman, Ichimura, and Todd, 1998; Heckman, LaLonde, and Smith, 1999; Hirano, Imbens, and Ridder, 2003; Hirano and Imbens, 2004; Heckman and Vytlacil, 2005, etc.) have yielded a variety of methods, such as those based on matching and the propensity score, that permit this identification and estimation.

An extensive statistical literature on observational studies (e.g. Rubin, 1974; Rosenbaum, 2002) also emerged, building on the experimental design work of R.A. Fisher, Cox, Neymann, Kempthorne, and others. This "treatment effect" literature introduced the "potential outcome model" and notions of "ignorability" and "propensity score” for measuring causal effects (e.g., Rosenbaum and Rubin, 1983; Holland, 1986). Angrist, Imbens, and Rubin (1996) (AIR) relate this approach to IV methods.

Another line of research has emerged in the machine learning literature in the work of Pearl (1988, 1993a, 1993b, 1995, 2000), Spirtes, Glymour, and Scheines (1993), and Dawid (2002) among others. In particular, Pearl (1995) introduced two methods related to the labor economics and treatment effect literatures, the “back door” and “front door” methods. A distinctive feature of this literature is the use of directed acyclic graphs (DAGs) to represent causal relations and of graphical criteria to determine if particular causal effects are identifiable, with less attention to the estimation of these causal effects.

White (2006) and White and Chalak (2006a) (WC) propose the “settable system” framework as a means of unifying these various approaches. There, particular attention is paid to identifying and estimating causal effects in a setting closest to that of exogenous regressors. Here we broaden our focus and apply this framework to analyze identification and estimation of causal effects in the presence of endogenous regressors generally. Consistent with Dawid (1979, 2000), the methods that emerge, including all those above, require one or more independence or conditional independence relationships to hold between the observed and unobserved variables of the system.

Specifically, our contribution here is to provide a novel and detailed examination of the ways in which causal structures can yield observed variables other than the cause or treatment of interest that can play an *instrumental* role in identifying and estimating causal effects. We thus extend the standard concept of instrumental variables to accommodate variables that are not necessarily uncorrelated with unobserved causes of a response of interest but that are nevertheless instrumental in recovering causal effects.

Consider, for example, the following structural system, where X, Y, and Z are variables with observed realizations, U_x, U_y and U_z are unobserved causes of X, Y, and Z respectively, α_o is an unknown vector, and γ_o and δ_o are unknown scalars such that:

(1) $$X \overset{c}{=} U_x{}'\,\alpha_o$$

(2) $$Z \overset{c}{=} \gamma_o X + U_z$$

(3) $$Y \overset{c}{=} \delta_o Z + U_y.$$

We use "$\overset{c}{=}$" instead of the usual equality to emphasize the causal nature of the structural equations. As in Dawid (1979), "$\perp$" denotes independence between random variables and "$\not\perp$" denotes dependence. Here, we assume that $U_x \not\perp U_y$, $U_x \perp U_z$, and $U_y \perp U_z$.

Consider measuring the total causal effect of X on Y. Substituting *(2)* into *(3)* gives:

$$(3') \qquad Y \overset{c}{=} \beta_o X + \delta_o U_z + U_y,$$

where $\beta_o \equiv \gamma_o \delta_o$ is the total causal effect of X on Y.

Clearly, since $U_x \not\perp U_y$, the least squares estimator for β_o, say $\hat{\beta}$, is inconsistent, as X is endogenous. Further, Z is an invalid instrument, as it is correlated with the unobserved term in *(3′)*: from *(2)* we have Z correlated with U_z; also, since X causes Z from *(2)* and X and U_y are correlated, we have Z correlated with U_y.

This presents a situation where it might seem that the causal effect of X on Y cannot be consistently estimated. Nevertheless, results of Section 4.1.2 demonstrate that, under mild conditions, a consistent estimator for the total causal effect of X on Y is given by:

$$\tilde{\beta} = \{(\boldsymbol{X'X})^{-1}(\boldsymbol{X'Z})\} \times \{[\boldsymbol{Z'}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'})\boldsymbol{Z}]^{-1}[\boldsymbol{Z'}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'})\boldsymbol{Y}]\},$$

where $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ each denote $n \times 1$ vectors. As discussed below, this system permits use of Pearl's (1995) "front-door" method. The variable Z is instrumental in the recovery of the causal effect of X on Y even though it is an "invalid" instrument in the standard sense.

We can also view this example as one in which identification is achieved by combining exclusion and covariance restrictions as in Hausman and Taylor (1983). But Sections 4.1.1 and 4.2 provide further examples where Hausman and Taylor's necessary conditions fail, yet identification still holds. In fact, our results provide an extension of those of Hausman and Taylor in which identification is achieved not solely by exclusion and covariance restrictions, but may also rely on *conditional* covariance (or independence) restrictions involving not just unobservables but also observables.

A main goal of this paper is thus to show that by extending the standard notion of instrumental variables, we achieve a framework that not only incorporates classical identification results such as those of Hausman and Taylor, but that also enables identification and estimation of clearly defined causal effects in the endogenous regressor case for a wide array of situations in which standard methods fail. As it turns out, these methods extend readily to the general case in which structural equations are not separable between observables and unobservables. Classical methods are less tractable in this way.

Specifically, by focusing on *conditional exogeneity* rather than strict exogeneity, we obtain a class of *extended instrumental variables* (EIV) methods that permit identification and estimation of causal effects. These methods are characterized by alternative moment conditions and exclusion restrictions that parallel those in the standard case. This paper begins a systematic exploration of these methods and their interrelations.

In Section 2, we discuss the data generating structural equations systems of interest here. In Section 3, we use the framework of Section 2 to provide a fully explicit causal account of standard regression and IV methods, extending AIR and setting the stage for subsequent results. Section 3.1 examines the case of *exogenous regressors* (XR). In Section 3.2, we study causal identification via standard *exogenous instruments* (XI) Z. Appendix B describes how causal identification fails when standard IV methods fail.

Section 4 begins our study of EIV methods, where the use of *conditional* instruments Z, *conditioning* instruments W, or both together permits identification of the causal effect of endogenous X. Section 4.1 treats *single* EIV methods, which use either conditioning EIV or conditional EIV but not both to identify causal effects. Section 4.1.1 introduces the method of *conditionally exogenous regressors given conditioning instruments* (CXR|I), relating this to matching, Rosenbaum and Rubin's (1983) ignorability condition, Pearl's (1995) back door, and White's (2006) predictive proxies. Section 4.1.2 introduces the method of *conditionally exogenous instruments given regressors* (CXI|R). We relate these to standard IV and to Pearl's (1995) front door method.

In Section 4.2, we discuss *double* EIV methods where joint use of conditional and conditioning EIV identifies effects of interest. We introduce the methods of *conditionally exogenous instruments given conditioning instruments* (CXI|I), *conditionally exogenous instruments and regressors given conditioning instruments* (CXIR|I), and *conditionally exogenous instruments given regressors and conditioning instruments* (CXI|RI).

Section 5 states a "master theorem" that provides necessary and sufficient conditions for identification of causal effects via EIV methods. Section 6 discusses the use of causal matrices to characterize the cases where identification holds by EIV. We illustrate by showing that CXR|I and CXI|R exhaust the possibilities for identification using a single EIV. Section 7 states conditions ensuring consistency and asymptotic normality for the

EIV estimators considered here. Section 8 concludes, with final remarks and a discussion of directions for future research. Proofs of formal results are gathered into Appendix A.

2. Causal Data Generating Systems

Economists have long understood the distinction between predictive and causal inquiries and in particular the dictum that correlation need not imply causation. Goldberger (1991, p. 337) states, "the causal requirement that in regression the *x*'s have to be the variables that actually determine *y* does not appear in the specification of the [classical regression] model: nothing in the [classical regression] model requires that the *x*'s cause *y*." Thus, economists have been concerned with developing methods to measure causal effects beyond predictive linear regression (see, e.g., Angrist and Krueger, 1999; Heckman, LaLonde, and Smith, 1999; Heckman, 2000; and Hoover, 2001).

Here we employ a familiar structural equation system to represent a causal structure, S. In particular, we consider data generated as a special case of the recursive system

$$X_1 \overset{c}{=} r_1\,(X_0)$$

$$X_2 \overset{c}{=} r_2\,(X_1, X_0)$$

$$\vdots$$

$$X_J \overset{c}{=} r_J\,(X_{J\text{-}1}, \ldots, X_1, X_0),$$

where X_0 is a random vector, and for $j = 1, \ldots, J$, X_j is a random variable and r_j is an unknown scalar-valued response function.

We use the notation $\overset{c}{=}$ to emphasize that the structural equations of S are neither equations nor regressions. Instead, they represent directional "causal links" (Goldberger, 1972, p.979). In particular, the right hand side variables mechanistically determine the value of the corresponding left hand side variable, but the converse is not true. The structural equations are thus directional autonomous mechanisms describing how every variable in the system is generated (Haavelmo 1943, 1944; Strotz and Wold, 1960; Pearl, 2000; WC). Conceptually, these can be manipulated without modifying any of the other relations. This enables definition of causal effects by means of hypothetical interventions where X_j is set to some different value, X_j*. WC provides a rigorous formalization.

Observe that X_0 does not appear on the left hand side of any causal relation. If the system affords a complete description of the causal structure, then X_0 is not caused by any other variables of the system. Following WC, we call such variables *fundamental*.

Formally, we work with the following structures.

Assumption A.1(a): Data Generation: For $j = 1, \ldots, J$, let U_j be random vectors with unobserved realizations, and let the response functions r_j be unknown real-valued measurable functions such that observable random variables $X_1, \ldots, X_J$ are generated as:

$$X_1 \overset{c}{=} r_1\,(U_1)$$

$$X_2 \overset{c}{=} r_2\,(X_1,\, U_2)$$

$$\vdots$$

$$X_J \overset{c}{=} r_J\,(X_{J\text{-}1},\, \ldots,\, X_1,\, U_J). \qquad \blacksquare$$

The U_j's may have differing dimensions. We collect them together as $X_0 \equiv (U_1', \ldots, U_J')'$. The data are thus generated by the system $S \equiv (X_0, r_1, \ldots, r_J)$.

We formally view S as a settable system as defined by WC, so references here to notions of setting, response, cause, and effect are as formally defined there. Given A.1(a), the following working definition of causality (always relative to S) suffices. Specifically, X_j *does not cause* X_k when $j \geq k$ (including $k = 0$). For $j < k$, we say that X_j *does not cause* X_k if $r_k\,(X_{k\text{-}1}, \ldots, X_1, U_k)$ defines a function constant in X_j. Otherwise, we say that X_j *causes* X_k. This corresponds to "direct" or "immediate" causality as defined by Pearl (2000). We take U_k to be a cause of X_k, whereas U_j does not cause X_k for $j \neq k$.

In this structure, all variables have causal status. For conciseness, we do not introduce attributes, that is, non-causal response modifiers (see WC). A key role played by attributes is to introduce heterogeneity into the system, an essential aspect of economic reality (see, e.g., Heckman, 1997; Heckman, Urzua, and Vytlacil, 2005; and Heckman and Vytlacil, 2005). The structures we consider can be straightforwardly generalized to handle this, but we refrain from doing so here to maintain a sharp focus for our analysis.

The vector X_0 accommodates either the unobservability of known determinants of a given response, the researcher's ignorance of the full set of relevant causes, or both. In A.1(a) we do not specify X_0 as fundamental, so A.1(a) does not completely specify the causal structure. In particular, dependence among elements of X_0 may arise from causal relations among these elements; this then entails dependence between the observed and unobserved causes (endogeneity) in a given structural equation.

To describe further restrictions placed on X_0, we now write $U \equiv X_0$ and $X \equiv (X_1, ..., X_J)$ and let $V \equiv (V_1, ..., V_G) \equiv (X, U)$ denote the vector of all observed and unobserved variables in the system. We also use the fact that every settable system S has an associated "causal matrix," $C_S = [c_{gh}]$. This is an adjacency matrix in which every observed and unobserved variable has a corresponding row and a corresponding column. Thus C_S is a $G \times G$ matrix. An entry $c_{gh} = 1$ indicates that V_g is an immediate cause of V_h. An entry $c_{gh} = 0$ indicates that V_g does not immediately cause V_h. We impose the convention that a variable does not cause itself, so $c_{gg} = 0$ for $g = 1, ..., G$.

For example, when the unobservables are scalar, C_S has the form

$$C_S = \left|\begin{array}{c:c} C_{S_1} & C_{S_2} \\ \hdashline C_{S_3} & C_{S_4} \end{array}\right| = \begin{array}{c|ccc:ccc|} & X_1 & \cdots & X_J & U_1 & \cdots & U_J \\ X_1 & 0 & & & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots & \ddots & \vdots \\ X_J & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hdashline U_1 & 1 & 0 \cdots & 0 & 0 & & \\ \vdots & 0 & \ddots & \vdots & & \ddots & \\ \vdots & \vdots & & \vdots & & & \\ U_J & 0 & \cdots 0 & 1 & & & 0 \end{array}$$

The recursivity in A.1(a) makes C_{S_1} upper triangular with zeros along the diagonal. Blank entries in C_{S_1} indicate elements taking either the values 0 or 1, reflecting that X_j may or may not cause X_k when $j < k$. Assumption A.1(a) further specifies that none of the X's cause any U's. Thus C_{S_2} is a $J \times J$ zero matrix. We also have that U_k does not cause X_j for $j \neq k$. Thus, C_{S_3} is a $J \times J$ identity matrix. We permit U_k to cause U_j, leaving C_{S_4} unspecified, apart from the zero diagonal. Consequently, C_{S_1} and C_{S_4} determine C_S.

A causal matrix explicitly specifies all causal relationships, including those holding among the unobserved variables, but these are unspecified in A.1(a). In order to complete our specification of the causal systems of interest here, we impose

Assumption A.1(b): Acyclicality: S has the property that for each $h \leq G$ and each set of h distinct elements, say $\{g_1, ..., g_h\}$, of $\{1, ..., G\}$, we have $c_{g_1 g_2} \times c_{g_2 g_3} ... \times c_{g_h g_1} = 0$. ■

The acyclic structure of A.1(a, b) rules out mutual causality or cycles in S. Mutual causality occurs when V_g causes V_h and V_h causes V_g. Cycles occur when, for example, V_g causes V_h, V_h causes V_k, and V_k causes V_g. We impose this structure to keep our analysis focused; the general settable system framework does not require this.

Acyclicality also ensures the system possesses at least one fundamental variable (Bang-Jensen and Gutin, 2001, prop. 1.4.2). Given A.1(a), none of the X_j's ($j > 0$) can be fundamental, so it must be that at least one element of X_0 is. We denote the vector of fundamental variables U_0. This may contain some or all of the elements of the U_j's. The causal matrix column for a fundamental variable is a vector of zeroes.

A convenient device for representing causal relations is the causal graph, often used below. For each causal matrix C_S there is a causal graph G_S. This is a variant of the graphs used in Wright's path analysis (Wright, 1921, 1923) and that are lately revived in the machine learning literature as "semi-markovian directed acyclic graphs" (DAGs). See, for example, Pearl (1988, 1993a, 1993b, 2000) and Spirtes, Glymour, and Scheines (1993). There, the unobserved variables are typically not explicitly represented. In contrast, we explicitly represent these due to their central role in econometrics.

The graph G_S consists of a set of nodes (vertices), one for each element of $\mathbb{V}$, and a set of arrows $\mathbb{A}$, corresponding to ordered pairs of distinct vertices. An arrow a_{gh} denotes that V_g directly causes V_h. Solid arrows will denote causal relationships between observables. A dashed arrow from U_j to U_k denotes that U_j causes U_k. A dashed arrow also denotes that U_j causes an observable. A simple dashed line between U_j and U_k indicates that (i) U_j causes U_k; (ii) U_k causes U_j; or (iii) an unobserved cause (e.g., U_0) (directly or indirectly) causes both U_j and U_k (we omit depicting the unobserved common cause). The convention that variables do not cause themselves corresponds to the absence of self-directed arrows in G_S. The absence of dashed lines or arrows between U_j and U_k implies that $U_j \perp U_k$, as discussed further below.

We next impose some significant simplifying structure:

Assumption A.2: Linearity and Separability: For $j = 1, \ldots, J$, assume that r_j is linear and separable so that the data generating structural equations system S is given by:

$$X_1 \overset{c}{=} U_1{}' \alpha_1$$

$$X_2 \overset{c}{=} \beta_{2,1}X_1 + U_2'\alpha_2$$

$$\vdots$$

$$X_J \overset{c}{=} \beta_{J,J\text{-}1}X_{J\text{-}1} + \ldots + \beta_{J,1}X_1 + U_J'\alpha_J,$$

where, for $j = 1, \ldots, J$, $E(U_j) = 0$, α_j is an unknown real vector conforming to U_j, and for $j = 2, \ldots, J$, $\beta_{j,1}, \ldots, \beta_{j,j\text{-}1}$ are unknown real scalars. We put $\beta_j \equiv (\beta_{j,j\text{-}1}, \ldots, \beta_{j,1})'$. ■

The compelling motivation for imposing this strong structure is to provide clear insights in a familiar context, permitting us to make our main points without being distracted by further complications. Nevertheless, our key results extend to the modern non-separable setting where much milder conditions replace A.2 (see, e.g., Matzkin, 2003, 2004, and 2005; Imbens and Newey, 2003; WC). We take this up elsewhere.

Throughout, our interest attaches to identifying and estimating the *average total causal effect* of an observable[2] X_j on another observable X_k. Thus, we focus on the full effect of X_j on a response X_k, channeled via all routes in the system, averaged over the unobserved causes U_k of X_k. Given A.2, this quantity is a constant.

Below, we require certain independence or conditional independence conditions to hold between variables. Following Dawid (1979), we write $X \perp U \mid W$ to denote that X is independent of U conditional on W. Just as independence $X \perp U$ entails $f_{U|X}(u|\, x) = f_U(u)$ (using an obvious shorthand notation for (conditional) density functions), conditional independence $X \perp U \mid W$ entails $f_{U|X,W}(u \mid x, w) = f_{U|W}(u \mid w)$. Given A.2, these assumptions are stronger than will be necessary to obtain identification results. Weaker conditions suffice, such as conditional mean independence ($E(U \mid X, W) = E(U \mid W)$) or conditional non-correlation ($E(X\,U \mid W) = E(X \mid W)\, E(U \mid W)$). Nevertheless, we work primarily with (conditional) independence, first for convenience and second because such conditions are required for identification of causal effects generally, such as for the non-separable case or when interest attaches to causal effects other than average effects, such as effects on the quantiles of the response (see WC).

Conditional independence implies conditional mean independence and conditional non-correlation. As a convenience to accommodate the stronger than necessary

---

2 As we implicitly rely on WC's settable system framework, it is more appropriate to refer to causal relationships as holding between *settable variables*, as defined there. Our present usage is intended as a convenient shorthand.

assumptions adopted here, when we speak of conditional (unconditional) dependence, we will understand this to result from conditional (unconditional) correlation, also implying conditional (unconditional) mean dependence.

3. Causal Identification with Exogenous Regressors or Instruments

We first employ the framework of Section 2 to provide a fully explicit causal interpretation of standard regression and IV methods. This covers some familiar ground. Nevertheless, we discover important features previously unrecognized. This also sets the stage for subsequent developments.

Goldberger (1991, p. 337) notes that there is no necessary causal structure embodied in standard regression. The same is true for standard treatments of IV. Although causal relationships were clearly of concern to the Cowles Commission pioneers, an explicit causal focus has disappeared from much of the subsequent literature. For example, White (2001) treats IV estimation extensively, but without any reference to causal structure. The properties of the estimators studied are driven solely by stochastic properties of the variables involved, and in particular certain key moment conditions.

Exceptions to this agnosticism are provided by recent articles of Angrist, Imbens, and Rubin (1996) (AIR), Heckman (1997), and Heckman, Urzua, and Vytlacil, (2005), for example. A main goal of AIR is explicitly to provide a causal account of IV. Here we provide a causal account of IV complementary to and extending AIR, designed also to accord with the philosophy literature, which requires causal foundations to drive causal conclusions, as expressed in Cartwright's (1989) dictum "no causes in, no causes out."

To proceed, we elaborate our notation. Let Y now denote the scalar response of interest, let $X \equiv [X_1, \ldots, X_k]'$ denote the observed causes of interest, and let $Z \equiv [Z_1, \ldots, Z_\ell]'$ denote variables potentially instrumental to identifying causal effects of interest, all generated as in A.1 and A.2. (The X_j's of A.1 and A.2 are now the elements of X, Y, and Z.) We denote by U_y, $U_x \equiv [U_{x_1}', \ldots, U_{x_k}']'$, and $U_z \equiv [U_{z_1}', \ldots, U_{z_\ell}']'$ the unobserved causes corresponding to the responses, causes, and instruments, respectively. As above, the fundamental unobservables are U_0. We let $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ denote $n\times k$, $n\times 1$, and $n\times \ell$ matrices containing n identically distributed observations on X, Y, and Z respectively.

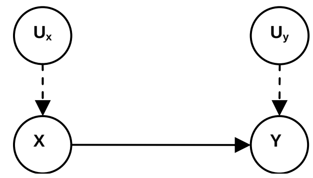
3.1 Exogenous Regressors

We first consider exogenous regressors, for which linear regression identifies the effect of X on Y. Consider structural equations system S_1 and its corresponding causal graph G_1:

(1) $X \overset{c}{=} \alpha_x U_x$

(2) $Y \overset{c}{=} X'\beta_o + U_y$,

where $U_x \perp U_y$.



Graph 1 (G_1)
Exogenous Regressors (XR)

In *(1)*, α_x is a matrix of unknown coefficients that map unobserved causes U_x to observed causes X. The coefficients β_o have causal meaning by virtue of *(2)*.

In S_1, X and Y do not share a common cause, as reflected by the condition

(XR) *Exogenous Regressors*: $X \perp U_y$.

Together, A.1 and XR ensure the key moment condition

$$E(XU_y) = 0. \qquad (M1)$$

From *(2)*, $U_y = Y - X'\beta_o$ (an equality, not a causal link). Substituting this into $M1$ gives

$$E(XY) - E(XX')\,\beta_o = 0.$$

This condition *structurally identifies* causal coefficients β_o by relating them solely to moments of observable variables. When *stochastic identification* also holds, that is, $E(XX')$ is non-singular, β_o is *fully identified* as

$$\beta_o = [E(XX')]^{-1}[E(XY)].$$

Formally, we have:

Proposition 3.1.1 Suppose A.1 and A.2 hold such that: (i) $Y \overset{c}{=} X'\beta_o + U_y$, where X is $k \times 1$, $k > 0$, β_o is an unknown finite $k \times 1$ vector, and $E(XX')$ and $E(XY)$ exist and are finite. Suppose further that (ii) $E(XX')$ is non-singular; and (iii) XR: $X \perp U_y$.

Then β_o, the average total causal effect of X on Y, is fully identified as:

$$\beta_o = [E(XX')]^{-1}[E(XY)]. \quad \blacksquare$$

The plug-in estimator for β_o is the usual OLS estimator for a linear regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$, $\hat{\beta}_n^{XR} \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y})$. Section 7 gives straightforward conditions ensuring consistency and asymptotic normality for $\hat{\beta}_n^{XR}$ and the other estimators we discuss.

Reichenbach's (1956) principle of common cause, applicable here, states that two random variables can exhibit correlation only if one causes the other or if they share a common cause. Here X and Y are correlated. We assume that Y does not cause X. We also exclude the possibility that X and Y share a common cause. The association between X and Y can thus only be explained as the effect of X on Y.

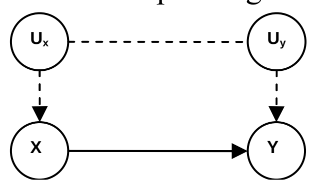
When control over X is possible, XR can be ensured by randomization. In observational studies, where control is absent, it is often hard to argue for XR. We now examine the failure of XR from a causal standpoint.

Specifically, consider the structural system S_2 and its corresponding causal graph G_2:

(1) $X \overset{c}{=} \alpha_x U_x$

(2) $Y \overset{c}{=} X'\beta_o + U_y$

where $U_x \perp U_y$.



Graph 2 (G_2)
Endogenous Regressors

In S_2, XR does not hold since $X \perp U_y$. When this results from $E(XU_y) \neq 0$, then regression no longer structurally identifies β_o, as unobservables appear in the moment equation $E(XY) = E(XX')\,\beta_o + E(XU_y)$. In G_2, the association between X and Y could be due to the joint response of X and Y to U_y, to U_x, or to U_0, an unobserved common cause of U_y and U_x. Following standard parlance, we call failure of XR *regressor endogeneity* and call X *endogenous regressors*. We also call this *confoundedness* of causes. In S_2, either U_y, U_x, or U_0 is a *confounding variable*. Thus, under A.1, an endogenous regressor is one sharing an unobserved common cause with the response. Simultaneity is absent from this system and is thus not responsible for the endogeneity.

3.2 Exogenous Instruments

When regressors are endogenous, XR is not available to identify β_o. But it is well known that identification can be achieved using a vector of "proper" instrumental variables, Z. A standard definition is that Z is proper if it is "valid," i.e. $E(ZU_y) = 0$, and "relevant," i.e. $E(XZ) \neq 0$ (e.g., Hamilton, 1994 p.238; Hayashi, 2000 p.191; Wooldridge, 2002 p.83-84).

P.G. Wright (1928) first used instrumental variables, which he called "curve shifters," to identify supply and demand elasticities (see Morgan, 1990; Angrist and Krueger, 2001;

Stock and Trebbi, 2003). In describing these, P.G. Wright states: “Such additional factors may be factors which (A) *affect* demand conditions without affecting cost conditions or (B) *affect* cost conditions without affecting demand conditions” (P.G. Wright, 1928 p.312; our italics). The use of the term “affect” suggests that Wright was thinking about causality, not just correlation. The first thoughts on instrumental variables thus appear to have been driven by causal reasoning and not by statistical or algebraic study.

As we discuss next, standard instrumental variables methods fall into one of two causally meaningful subcategories. In both cases, we refer to these standard instruments as *exogenous instruments* (XI) and refer to this as the XI method.

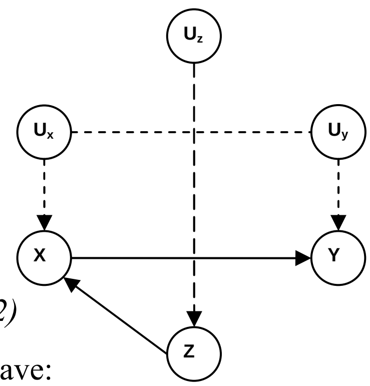
3.2.1 Observed Exogenous Instruments

Consider the following structural equations system S_3 and its associated causal graph G_3:

(1) $Z \overset{c}{=} \alpha_z U_z$

(2) $X \overset{c}{=} \gamma_x Z + \alpha_x U_x$

(3) $Y \overset{c}{=} X'\beta_o + U_y$



Graph 3 (G_3)
Observed Exogenous Instruments (OXI)

where γ_x is a $k \times k$ matrix (so $\ell = k$), $U_x \perp\!\!\!\!\!\!\not\;\; U_y$, $U_x \perp U_z$, and $U_y \perp U_z$. Substituting structural equation *(2)* into structural equation *(3)* and setting $\pi_o \equiv \gamma_x{}'\beta_o$, we have:

(3′) $Y \overset{c}{=} Z'\pi_o + U_x{}'\beta_o + U_y$.

In S_3, X is endogenous. Nevertheless, structural identification of β_o is ensured by:

(XI) *Exogenous Instruments*: $Z \perp U_y$.

Together with A.1 and A.2, this implies

$$E(ZU_y) = 0. \qquad (M2)$$

Using *(3)* then gives the structural identifying condition $E(ZY) - E(ZX')\beta_o = 0$. Identification is complete given stochastic identification, that is, that $E(ZX')$ is non-singular. This directly embodies the standard rank and order conditions. We have

Proposition 3.2.1 Suppose A.1 and A.2 hold such that: (i) $Y \overset{c}{=} X'\beta_o + U_y$, $X \overset{c}{=} \gamma_x Z + \alpha_x U_x$ (with $\ell = k$), and $E(ZX')$ and $E(ZY)$ exist and are finite. Suppose further that (ii) $E(ZX')$ is non-singular; and (iii) XI: $Z \perp U_y$.

Then β_o, the average total causal effect of X on Y, is fully identified as:

$$\beta_o = [E(ZX')]^{-1}[E(ZY)]. \quad \blacksquare$$

The familiar result that the plug-in estimator $\hat{\beta}_n^{XI} \equiv (\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{Y})$ is consistent and asymptotically normal for β_o then holds under mild conditions, provided in Section 7.

In S_3, Z satisfies the following three causal properties that accord with XI and that make Z instrumental for identifying β_o when X and Y are confounded:

(CP:OXI): **Causal Properties of Observed Exogenous Instruments** (i) Z directly causes X, and the effect of Z on X is identified via XR; (ii) Z indirectly causes Y, and the effect of Z on Y is identified via XR; (iii) Z causes Y only via X. ■

As Z is observable, we call Z *observed exogenous instruments* (OXI).

These properties justify the indirect least squares (ILS) interpretation of instrumental variables (Haavelmo, 1943, 1944). Specifically, in S_3, since $Z \perp U_x$ and given $E(XZ')$ and $E(ZZ')$ finite with $E(ZZ')$ non-singular, Proposition 3.1.1 identifies the effect of Z on X as $\gamma_x = E(XZ')[E(ZZ')]^{-1}$. Similarly, since $Z \perp U_x$ and $Z \perp U_y$ and given $E(ZY)$ finite, Proposition 3.1.1 identifies the effect of Z on Y as $\pi_o = [E(ZZ')]^{-1}E(ZY)$. By CP:OXI (iii), Z causes Y and only via X. The effect of X on Y, β_o, is thus the “ratio” of the effect of Z on Y to that of Z on X, so that $\beta_o = \gamma_x'^{-1}\,\pi_o$ for γ_x non-singular. That is,

$$\beta_o = \gamma_x'^{-1}\,\pi_o = \{[E(ZZ')]^{-1}E(ZX')\}^{-1}\{[E(ZZ')]^{-1}E(ZY)\} = [E(ZX')]^{-1}E(ZY).$$

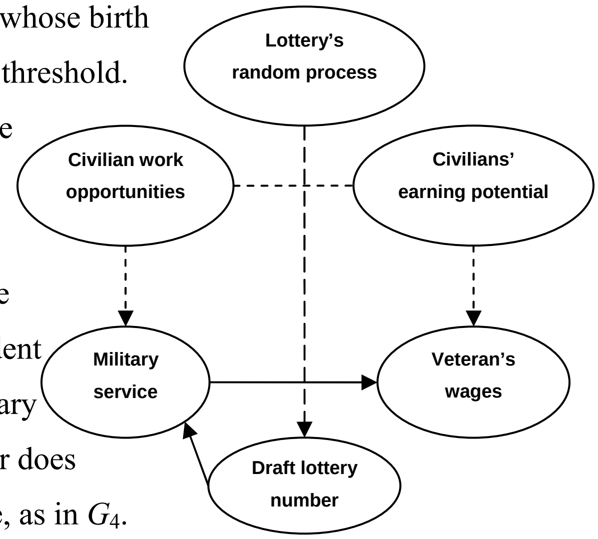
This can be consistently estimated by indirect least squares, that is, as the “ratio” of the two consistent OLS estimators of the effect of Z on Y and the effect of Z on X:

$$\{(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})\}^{-1}\{(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{Y})\} = (\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{Y}) = \hat{\beta}_n^{XI}.$$

This classical account of IV estimation bears explicit statement for two reasons. First, it makes fully explicit all the causal components; second, it provides a “base case” against which further developments, provided below, can be compared.

The work of Angrist (1990) provides an example of OXI in which all causal elements are clear. Angrist is interested in the effect of Vietnam War military service on a veteran’s post-war civilian wage.Vietnam War service and civilian wage could be confounded by variables such as individual ability or education, as either might affect both joining the military and civilian wages. Since military service is thus potentially endogenous, Angrist employs the Vietnam draft lottery number as an observed exogenous instrument in our parlance. This number was randomly assigned based on birth date; it dictated military service for individuals whose birth date corresponds to a lottery number below a certain threshold. Those with a lottery number above the threshold were not required to serve.

Angrist's use of lottery number as an instrument satisfies CP:OXI. Implicitly, Angrist assumes that the randomness of the lottery number makes it independent of unobserved factors affecting an individual's military service or civilian wages, and that the lottery number does not affect veteran's wages except via military service, as in G_4. If the data are indeed generated in this way, the lottery number is a legitimate OXI.



Graph 4 (G_4)
OXI for the Effect of Military Service on Veteran's Wage

A main goal of AIR is to provide an explicit causal account of the IV method. For this, AIR employ the "potential outcome" framework. We now compare the present approach with that of AIR. We maintain AIR's notation except that we use X_i and $\boldsymbol{X}$ instead of their D_i and $\boldsymbol{D}$ to denote receipt of treatment. AIR let $i = 1, \ldots, n$ denote individuals in the population of interest and list the following sufficient assumptions for the IV estimator to have a "causal interpretation," namely that of "an average causal effect for a subgroup of units, the compliers":

AIR Conditions (a) Stable Unit Treatment Value Assumption (SUTVA): if $Z_i = Z_i'$ then $X_i(\boldsymbol{Z}) = X_i(\boldsymbol{Z}')$; if $Z_i = Z_i'$ and $X_i = X_i'$ then $Y_i(\boldsymbol{Z}, \boldsymbol{X}) = Y_i(\boldsymbol{Z}', \boldsymbol{X}')$; (b) The treatment assignment Z_i is random (or, more generally, "ignorable"); (c) Exclusion restriction: $\boldsymbol{Y}(\boldsymbol{Z}, \boldsymbol{X}) = \boldsymbol{Y}(\boldsymbol{Z}', \boldsymbol{X})$ for all $\boldsymbol{Z}$ and $\boldsymbol{Z}'$ and for all $\boldsymbol{X}$; (d) Nonzero average causal effect of Z on X; (e) Monotonicity: $X_i(1) \geq X_i(0)$ for all $i = 1, \ldots, n$. ■

If we let the variables X, Y, Z, U_x, U_y, and U_z in S_3 pertain to a given individual in the population, the OXI case satisfies AIR's assumptions. Assumption (a) is satisfied by A.1, as the left- and right-hand side variables in every equation in S_3 pertain only to a given individual. Assumption (b) and the OXI case share the same statistical implications, as $U_x \perp U_z$ and $U_y \perp U_z$. Assumption (c) states that any effect of Z on Y must be via X; this is ensured by S_3 *(2)* and *(3)*. Assumption (d) requires Z to have an effect on the treatment X, which is ensured by S_3 *(2)* with γ_x non-singular. Assumption (e) requires a monotonic causal relation between Z and X in that the direction of the effect of the assignment to treatment on the actual treatment is the same for all individuals. This is implicitly ensured in S_3, as γ_x is the same for each individual, given our assumed absence of heterogeneity.

3.2.2 Proxies for Unobserved Exogenous Instruments

In satisfying CP:OXI, S_3 imposes the strong requirements that $U_x \perp U_z$ and $U_y \perp U_z$, so that Z is random. But random instruments are hard to justify in economics generally, as it is largely an observational science. Further, neither the standard relevance and validity conditions nor Proposition 3.2.1 require Z to be random or ignorable. In particular, Proposition 3.2.1, as is standard (e.g. Heckman, 1996), does not require $Z \perp U_x$: the effect of Z on X, usually estimated from a "first stage" regression, need not be identified.

We now relax AIR's conditions for the causal interpretation of IV by giving a causal account that does not require ignorability for Z, but that only relies on the standard relevance and validity conditions. For this, consider structural equations system S_5 and its associated causal graph G_5:

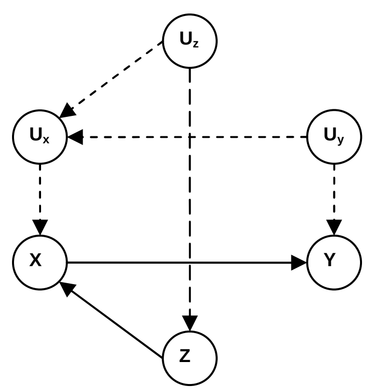
$$(1)\ Z \overset{c}{=} \alpha_z U_z$$

$$(2)\ X \overset{c}{=} \gamma_x Z + \alpha_x U_x$$

$$(3)\ Y \overset{c}{=} X' \beta_o + U_y$$

where γ_x is a $k \times k$ matrix (so $\ell = k$), $U_x \not\perp U_y$, $U_x \not\perp U_z$ and $U_y \perp U_z$. Substituting *(2)* into *(3)* gives

$$(3')\ Y \overset{c}{=} Z' \pi_o + U_x' \alpha_x' \beta_o + U_y.$$



Graph 5 (G_5)
Proxies for Unobserved Exogenous Instruments (PXI)

where $\pi_o \equiv \gamma_x' \beta_o$. Here $U_x \not\perp U_z$, whereas in S_3 we have $U_x \perp U_z$.

As before, XR cannot identify β_o. However, XI is satisfied, as $Z \perp U_y$. As we also have $Z \not\perp X$, Z is a proper instrument. Thus, Proposition 3.2.1 applies, identifying β_o as $[E(ZX')]^{-1} E(ZY)$. Nevertheless, S_5 differs fundamentally from S_3, in that the causal properties making Z in S_5 instrumental for identifying β_o are satisfied not by Z but instead by unobservable causes U_z. If these were observable, they could act as proper instruments. In their absence, the observable Z acts as a proxy for the unobservable U_z. Accordingly, we call Z in S_5 *proxies for (unobserved) exogenous instruments* (PXI) to distinguish this case from OXI. Causal properties for this case are:

(CP:PXI) Causal Properties for Proxies for Unobserved Exogenous Instruments

(i) U_z indirectly causes X, and the full effect of U_z on X could be identified via XR had U_z been observed; (ii) U_z indirectly causes Y, and the full effect of U_z on Y could be identified via XR had U_z been observed; (iii) U_z causes Y only via X; (iv) if Z causes Y, it does so only via X. ■

Conditions (i), (ii) and (iii) of CP:PXI are essentially identical to their analogues of CP:OXI with U_z replacing Z. Note that in (i) and (ii) we refer to the *full* effect of U_z on X and Y respectively. In (i), this includes not only the effect of U_z on X through Z, but also through U_x, and similarly for the effect on Y in (ii) (see *(3′)*). Condition (iv) is analogous to the exclusion restriction (iii) of CP:OXI, but here we do not require that Z cause Y.

In the PXI case, the ILS account of IV holds with U_z replacing Z, but the unobservability of U_z prohibits empirical application of this version of ILS. Moreover, in S_5 the effects of Z on Y and of Z on X are *not* identified as they for OXI. For example, observe that Z is endogenous in the reduced form *(3')*. The classical ILS account now *fails* for instruments Z. Fortunately, however, Z's role as a proxy for U_z enables it to structurally identify β_o, given the causal structure of S_5.

Specifically, this structure ensures that Z and X as well as Z and Y are confounded by the same variables, U_z. When U_z renders the OLS estimator of π_o inconsistent, it also renders the estimator of γ_x inconsistent in just the right way to leave the ratio of these confounded effects informative for the effect of interest.

To demonstrate, suppose that $E(XZ')$, $E(ZZ')$, $E(ZY)$, and $E(U_xZ')$ are finite and that needed inverses exist. The effect of Z on X, γ_x, is not identified as $E(XZ')\ [E(ZZ')]^{-1}$ from *(2)*, as $Z \not\perp U_x$. Instead, $\gamma_x = E(XZ')[E(ZZ')]^{-1} - \alpha_x\, E(U_x\, Z')[E(ZZ')]^{-1}$. Similarly, as $Z \not\perp U_x$, the effect of Z on Y, π_o, is not identified as $[E(ZZ')]^{-1}E(ZY)$ from *(3′)*. Instead, $\pi_o = [E(ZZ')]^{-1}E(ZY) - [E(ZZ')]^{-1}E(ZU_x')\ \alpha_x'\ \beta_o$. Yet β_o is identified from *(3)*:

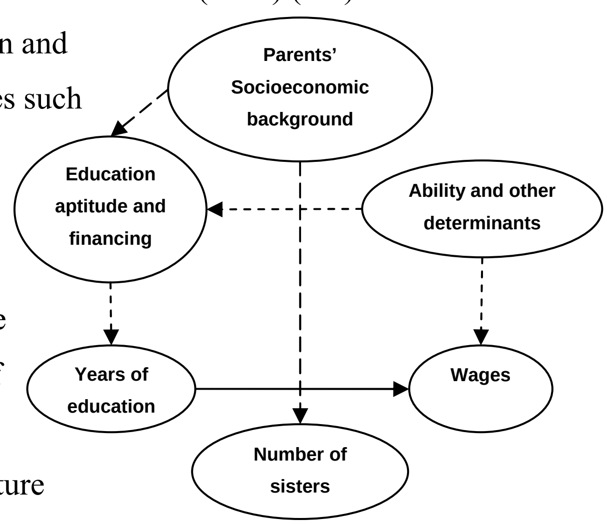
$$
\begin{aligned}
E(ZX')^{-1}E(ZY) &= \{[E(ZZ')]^{-1}E(ZX')\}^{-1}\,[E(ZZ')]^{-1}E(ZY) \\
&= \{\,\gamma_x' + [E(ZZ')]^{-1}E(ZU_{\mathbf{x}}')\,\alpha_x'\}^{-1}\,[\pi_o + [E(ZZ')]^{-1}E(ZU_{\mathbf{x}}')\,\alpha_x'\,\beta_o] \\
&= \{\,\gamma_x' + [E(ZZ')]^{-1}E(ZU_{\mathbf{x}}')\,\alpha_x'\}^{-1}\,\{\gamma_x' + [E(ZZ')]^{-1}E(ZU_{\mathbf{x}}')\,\alpha_x'\}\beta_o \\
&= \beta_o.
\end{aligned}
$$

The above expression even permits $\gamma_x = 0$, so that Z need not cause X, whereas γ_x had to be invertible in S_3 to support ILS. As long as Z and X share a common unobserved cause (U_z), they possess the correlation required to identify the effect of interest. We provide an example below. When $\gamma_x = 0$, Z is a "pure predictive proxy" for U_z, the true causal variable instrumental to identifying the effect of X on Y. The PXI case thus relies on causally meaningful instruments (Z) that satisfy the relevance and validity conditions but do not satisfy AIR's conditions. In particular, PXI provides a causal account of IV that removes two of AIR's assumptions, namely, ignorability of Z (assumption (b)) and nonzero average causal effect of Z on X (assumption (d)).

In the PXI case, a function of two inconsistent estimators, the OLS estimators of γ_x' and π_o from *(2)* and *(3')*, is itself consistent for the effect of interest, β_o. Thus, identification strategies that advocate recovery of causal effects as functions only of *identifiable* effects (Pearl 2000, p.153-154) miss recovering certain identifiable effects.

A number of applied papers in economics that use standard IV to estimate the effect on Y of a potentially endogenous X implicitly employ PXI to justify the validity and relevance of their instruments. Consider, for example, measuring the effect of a person's years of education on their future wages, as in Butcher and Case (1994) (BC).

As BC note, an individual's years of education and wages can be confounded by unobserved variables such as the individual's ability, making the "years of education" variable potentially endogenous. To overcome this problem, BC employ the number of sisters in a family as an instrument. They argue that daughters in families with a larger number of sisters tend to have lower levels of education and that this association is unlikely to be related to future



Graph 6 (G_6)
PXI for the effect of Education of wages

wages by means other than educational attainment. In terms of our framework, BC exploit the correlation between the number of sisters and a daughter's level of education without requiring that one causes the other. For example, suppose that parents' socioeconomic background and capacity to help finance daughters' education generates the correlation between the number of sisters and education level (see G_6). If the data are indeed generated as in G_6, then number of sisters is a legitimate proxy for the unobserved exogenous instrument "parents' socioeconomic background."

Both OXI and PXI employ instruments Z satisfying the standard validity and relevance conditions. We call such Z "proper standard instruments." Equally important is that Z provides or proxies for a source of variation preceding the cause of interest X and affecting Y only via X, if at all. We thus also call such Z *pre-cause instrumental variables*. We also call any Z satisfying XI an *unconditional* instrumental variable, to distinguish it from the *conditional* instrumental variables discussed below.

3.3 Failures of Identification

Our causal framework accounts for not only the successes of standard IV but also its failures. So as not to disrupt the flow of our main discussion, Appendix B contains a detailed causal description of how structural identification of β_o via XI fails in the standard "irrelevant instrument," "invalid instrument," and "under-identified" cases.

4 Extended Instruments

We now investigate situations in which vectors Z or W are not valid instruments in the standard sense, as they are correlated with U_y, but are nevertheless instrumental in identifying the effect of X on Y. We call these *extended* instrumental variables (EIV). In particular, we introduce *conditional* and *conditioning* EIV.

4.1 Single EIV Methods

We first treat the case in which a single EIV can identify the causal effect of interest.

4.1.1 Conditioning Instruments

The treatment effect literature has introduced two central methods to treat the problem of confoundedness: randomization and matching (e.g., Rubin, 1974; Rosenbaum, 2002). Randomization, introduced by R.A. Fisher (1935, ch.2), randomly assigns units to treatment and control groups. If feasible, this ensures the absence of confounding variables for the cause of interest, as in S_1 and S_3. But randomization is rare in observational studies.

In non-randomized studies, matching observations from the treatment and control groups that share common causes or attributes provides a way forward. By conditioning on the information in the confounding variables, one can interpret the remaining conditional association between the putative cause and its response as the causal effect of the first on the second. Developments along these lines include “selection on observables” (Barnow, Cain, and Goldberger, 1980; Heckman and Robb, 1985), ignorability and the “propensity score” (Rubin, 1974; Rosenbaum and Rubin, 1983), the “back-door” method (Pearl, 1995), and “predictive proxies” (White, 2006; WC). In labor economics, matching methods are well established and have been applied in the contexts of the distribution of earnings, policy evaluation, and the returns to education and training programs (Roy, 1951; Heckman and Robb, 1985; Heckman, Ichimura, and Todd, 1998).

We now study causal structures in which matching can be effected by the use of *conditioning* instruments W that act as proxies for unobserved confounding variables. We write $W \equiv [W_1, \ldots, W_m]'$, $U_w \equiv [U_{w_1}{}', \ldots, U_{w_m}{}']'$, consistent with A.1, and let $\boldsymbol{W}$ denote an $n \times m$ matrix of identically distributed observations on W.

Specifically, consider S_2, where X is endogenous because $U_x \perp U_y$. Suppose that this dependence arises because U_x and U_y have a common cause. To gain insight we start with the extreme case where we actually observe the confounding variables W that determine U_x and U_y. This violates our convention that observables do not cause unobservables (A.1(a)), so this is only a temporary expedient that we will remove shortly. To proceed, consider structural equations system S_{7a} and its associated causal graph G_{7a}:

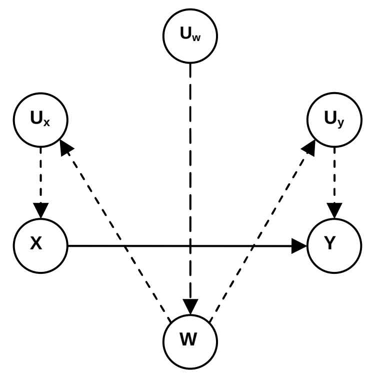
(1) $W \stackrel{c}{=} \alpha_w U_w$

(2) $U_{x_1} \stackrel{c}{=} \gamma_{x_1} W$

(3) $U_{y_1} \stackrel{c}{=} \gamma_{y_1} W$

(4) $X \stackrel{c}{=} \alpha_{x_1} U_{x_1} + \alpha_{x_2} U_{x_2}$

(5) $Y \stackrel{c}{=} X' \beta_o + U_{y_1} + U_{y_2}$



Graph 7*a* (G_{7a})
Conditioning Instruments

so that $U_x \perp U_y$, $U_x \perp U_w$, and $U_y \perp U_w$, where $U_x \equiv (U_{x_1}', U_{x_2}')'$ and $U_y \equiv (U_{y_1}, U_{y_2})'$, with $U_w \perp U_{x_2}$, $U_w \perp U_{y_2}$, and $U_{x_2} \perp U_{y_2}$. Regressor endogeneity arises from correlation between U_{x_1} and U_{y_1} resulting from the common cause W. The unobservable causes U_{x_2} and U_{y_2} provide independent sources of variation[3] ensuring that X is not entirely determined by W and that Y is not entirely determined by X and W.

In S_{7a}, once we condition on W, we are guaranteed that the remaining association between X and Y can be interpreted only as the causal effect of X on Y. The key conditional independence condition obvious in S_{7a} that parallels XR and XI above is:

(CXR|I) *Conditionally Exogenous Regressors given Conditioning Instruments:*

$$X \perp U_y \mid W.$$

When this condition holds for some vector W generally, we call W *conditioning instruments* to emphasize their role in ensuring this conditional exogeneity.

We emphasize that the role of S_{7a} is merely to motivate CXR|I; by no means is S_{7a} a necessary structure for CXR|I. As we discuss shortly, CXR|I can also hold for properly chosen W even when the true confounding variables for X and Y cannot be observed.

CXR|I delivers structural identification of β_o, as it implies the key moment condition

$$E(XU_y \mid W) = E(X \mid W) \times E(U_y \mid W). \qquad (M3)$$

To see how this condition structurally identifies β_o, rewrite *(M3)* as

$$E([X - E(X \mid W)]\, U_y \mid W) = 0,$$

replace $E(X|W)$ with its regression representation $E(XW')[E(WW')]^{-1}W$, and take expectations on both sides above to get

$$E(\,[X - E(XW')[E(WW')]^{-1}W\,]\, U_y) = 0.$$

This and structural equation *(5)* imply

[3] In subsequent structural equations systems of Section 4, we may drop explicit reference to components of vectors of unobserved causes for notational convenience, keeping in mind that these vectors are not entirely determined by other unobserved causes and thus that they include independent sources of variation, such as U_{x_2} and U_{y_2} in S_{7a} (and S_9 below), necessary for stochastic identification.

$$E([X - E(XW')[E(WW')]^{-1}W]\,[Y - X'\,\beta_o]\,) = 0,$$

so that β_o is structurally identified as

$$\{E(XX') - E(XW')[E(WW')]^{-1}\,E(WX')\}\,\beta_o = E(XY) - E(XW')[E(WW')]^{-1}E(WY).$$

Note that this derivation relies only on $Y \overset{c}{=} X'\beta_o + U_{y_1} + U_{y_2}$, the linear regression representation of $E(X \mid W)$, and CXR|I. The specific structure of S_{7a} is not required.

When stochastic identification holds, i.e., $E(XX') - E(XW')[E(WW')]^{-1}\,E(WX')$ is non-singular, β_o, the average total causal effect of X on Y, is identified as:

$$\beta_o = \{E(XX') - E(XW')[E(WW')]^{-1}E(WX')\}^{-1} \times \{E(XY) - E(XW')[E(WW')]^{-1}E(WY)\}.$$

Under mild conditions, a consistent, asymptotically normal plug-in estimator for β_o is

$$\hat{\beta}_n^{CXR|I} \equiv \{\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}')\,\boldsymbol{X}\}^{-1}\{\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}')\,\boldsymbol{Y}\}.$$

Even though W plays an instrumental role in identifying β_o, there is no requirement that W be exogenous. For example, in S_{7a} W is clearly endogenous, as $W \not\perp U_y$. Conditioning instruments are thus not standard instruments, motivating their description as extended instrumental variables (EIV). We call $\hat{\beta}_n^{CXR|I}$ an EIV estimator.

Inspecting $\hat{\beta}_n^{CXR|I}$, we see that it is a standard IV estimator using as *derived* standard instruments estimated residuals of the regression of X on W, $X - E(XW')[E(WW')]^{-1}W$. Nevertheless, we do not place these "residual instruments" on an equal footing with W, as it is W that carries the causal information enabling recovery of the effect of X on Y. Of even greater significance is that, as WC show, when A.2 is relaxed to permit non-separable structures, these residual instruments no longer appear, whereas W (the "predictive proxies") continue to play their instrumental role.

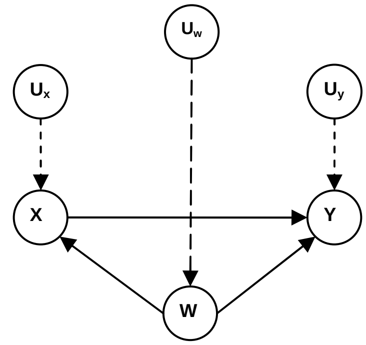
$\hat{\beta}_n^{CXR|I}$ is also the Frisch-Waugh (1933) partial regression estimator, obtained by regressing $\boldsymbol{Y}$ on the residuals $(\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}')\boldsymbol{X}$ from a regression of $\boldsymbol{X}$ on $\boldsymbol{W}$. This is equivalently the coefficient estimator associated with $\boldsymbol{X}$ from a linear regression of $\boldsymbol{Y}$ on both $\boldsymbol{X}$ and $\boldsymbol{W}$. This latter regression emerges naturally from S_{7a}, after performing the substitutions required to enforce our convention that observables do not cause unobservables. Substituting *(2)* into *(4)* and *(3)* into *(5)* in S_{7a} gives the structure S_{7b}:

$$(1)\ W \overset{c}{=} \alpha_w U_w$$

$$(2)\ X \overset{c}{=} \gamma_x W + \alpha_x U_x$$

$$(3)\ Y \overset{c}{=} X'\beta_o + W'\gamma_o + U_y$$

with $U_w \perp U_x$, $U_w \perp U_y$, and $U_x \perp U_y$.



Graph 7*b* (G_{7b})
Conditioning Instruments

In writing S_{7b}, we adjust the notation in the natural way. With the given independence conditions, Proposition 3.1.1 applies, as X and W jointly satisfy XR. In S_{7b}, both β_o, the *full* causal effect of X on Y, and γ_o, the *direct* causal effect of W on Y, are identified. The *full* causal effect of W on Y is $\gamma_o + \gamma_x'\beta_o$, identified from a regression of Y on W only.

As noted above, S_{7a} (S_{7b}) is not necessary for CXR|I. Structures satisfying Pearl's (1995; 2000, pp. 79-81) "back-door" criterion, in which an observable (here W) mediates a link between X and Y, also ensure CXR|I. In Pearl's framework, W is either the vector of common causes (G_{7a}, G_{7b}), or a response to the unobserved common cause and a cause of either Y or X (G_{8a}, G_{8b}). In G_{8a} and G_{8b}, CXR|I holds because an unobserved confounding common cause of X and Y causes Y via W (G_{8a}) or X via W (G_{8b}). In each case, W acts as an observable proxy for the unobserved common cause.

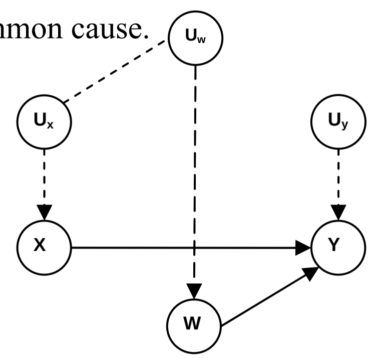
Specifically, let S_{8a} be given by

$$(1)\ W \overset{c}{=} \alpha_w U_w$$

$$(2)\ X \overset{c}{=} \alpha_x U_x$$

$$(3)\ Y \overset{c}{=} X'\beta_o + W'\gamma_o + U_y$$

with $U_w \not\perp U_x$, $U_w \perp U_y$, and $U_x \perp U_y$.



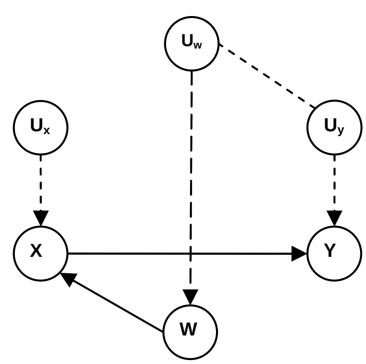
Graph 8*a* (G_{8a})
Conditioning Instruments

Similarly, let S_{8b} be given by

$$(1)\ W \overset{c}{=} \alpha_w U_w$$

$$(2)\ X \overset{c}{=} \gamma_x W + \alpha_x U_x$$

$$(3)\ Y \overset{c}{=} X'\beta_o + U_y$$



Graph 8*b* (G_{8b})
Conditioning Instruments

with $U_w \perp U_x$, $U_w \not\perp U_y$, and $U_x \perp U_y$.

There are several noteworthy features to these structures. In S_{8a}, the causal direction between U_x and U_w in G_{8a} is unspecified, so S_{8a} corresponds to three possible back door structures. In each, X and W jointly satisfy XR in *(3)*, so Proposition 3.1.1 holds. Here, W is a structurally relevant exogenous variable correlated with X, so omitting W from the identifying regression leads to the classical "omitted variable bias."

Now consider S_{8b}. For concreteness, suppose U_y causes U_w. Now both X and W are endogenous, as $X \not\perp U_y$ and $W \not\perp U_y$. Yet β_o is structurally identified by CXR|I. Given stochastic identification, β_o is fully identified from a regression of Y on both X and W. According to the textbooks, this regression should yield nonsense, as it contains not only endogenous regressors X, but also *structurally irrelevant* and *endogenous* regressors W. (W "is structurally irrelevant," as it does not appear in S_{8b} *(3)*.) Nevertheless, this regression identifies causally meaningful coefficients β_o.

What about the remaining regression coefficients, those associated with W? In the context of S_{8b}, these have no causal interpretation. Instead they have only a predictive interpretation, as discussed in detail by White (2006) and WC. Thus, some regression coefficients have causal meaning (those associated with X), but others do not (those associated with W). In other words, not all the regression coefficients need have signs and magnitudes that make causal sense, constituting an instance of what Heckman (2006) has termed "Marschak's maxim": we may identify certain economically meaningful components of a structure (β_o) without having to identify the entire structure.

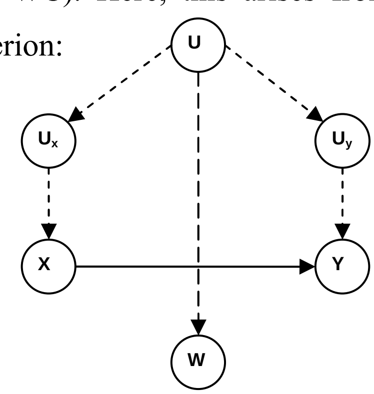
Nor does Pearl's back door method exhaust the possibilities for CXR|I. Another possibility is that of "predictive proxies" (White, 2006; WC). Here, this arises from structures such as S_9, which violates Pearl's back-door criterion:

(1) $W \stackrel{c}{=} \alpha_{w_1} U_{w_1} + \alpha_{w_2} U_{w_2}$

(2) $U_{x_1} \stackrel{c}{=} \gamma_{x_1} U_{w_1}$

(3) $X \stackrel{c}{=} \alpha_{x_1} U_{x_1} + \alpha_{x_2} U_{x_2}$

(4) $U_{y_1} \stackrel{c}{=} \gamma_{y_1} U_{w_1}$



Graph 9 (G_9)
Conditioning Instruments

(5) $Y \overset{c}{=} X'\beta_o + U_{y_1} + U_{y_2}$

where $U_{w_1} \perp U_{w_2}$, $U_{w_2} \perp U_{x_2}$, $U_{w_2} \perp U_{y_2}$, and $U_{x_2} \perp U_{y_2}$, with $U_w \equiv (U_{w_1}', U_{w_2}')'$, $U_x \equiv (U_{x_1}', U_{x_2}')$, and $U_y \equiv (U_{y_1}', U_{y_2})'$, so that $W \not\perp U_y$, and $X \not\perp U_y$.

In S_9, U_{w_1} is an unobserved common cause for X and Y; the predictive proxy W is a measurement error-laden version of U_{w_1}. Again, both X and W are endogenous; however, Proposition 4.4 of WC applies to establish CXR|I. The key to this is W's ability to predict U_w (hence X) sufficiently well that U_y contains no additional information useful in predicting X. As in S_{8b}, β_o is identified from a regression containing endogenous X and structurally irrelevant endogenous W. Our comments about S_{8b} fully apply to S_9.

Given its role as an observable proxy for unobserved common causes of X and Y, we call W a vector of *common cause instruments*.

CXR|I enables matching, in the language of the treatment effects literature. Letting Y_x denote the value Y would take had X been set to x (the "potential outcome"), it follows that $Y \overset{c}{=} X'\beta_o + U_y$ and CXR|I imply the key "ignorability" or "unconfoundedness" condition $Y_x \perp X \mid W$ of Rosenbaum and Rubin (1983) (White, 2006, proposition 3.2).

Although CXR|I and the methods of Section 4.2 permit structural identification, the order condition necessary for identification in Hausman and Taylor (1983) fails in these cases. In particular, in S_9 the number of unconstrained coefficients in *(5)* exceeds the number of "predetermined" (uncorrelated with U_y) variables for (*5*) ($k > 0$). Thus, the limited information order condition of Hausman and Taylor (1983, proposition 4) fails. Similarly, since S_9 does not impose any covariance restrictions ($U_x \not\perp U_y$, $U_x \not\perp U_z$, and $U_y \not\perp U_z$), Hausman and Taylor (1983, proposition 6) does not apply, and their sufficient condition for identification in the full information context (proposition 9) is not satisfied. Instead, a restriction on the *conditional* covariance of the unobserved causes, specifically $U_x \perp U_y \mid U_w$, ensures CXR|I here, ensuring the structural identification of β_o in S_9.

We conclude this section with a formal identification result under CXR|I.

Proposition 4.1.1 Suppose A.1 and A.2 hold such that: (i) $Y \overset{c}{=} X'\beta_o + U_y$, and $E(XX')$ and $E(XY)$ exist and are finite. Suppose further that (ii) there exists a random vector W

such that $E(XW')$, $E(WW')$, and $E(WY)$ exist and are finite; $E(WW')$ is non-singular and $E(X\,|W) = E(XW')[E(WW')]^{-1}W$; (iii) $E(XX') - E(XW')[E(WW')]^{-1}E(WX')$ is non-singular; and (iv) CXR|I: $X \perp U_y \mid W$.

Then β_o, the average total causal effect of X on Y, is fully identified as:

$$\beta_o = \{E(XX') - E(XW')[E(WW')]^{-1}E(WX')\}^{-1}\ \{E(XY) - E(XW')[E(WW')]^{-1}E(WY)\}. \ \blacksquare$$

Note that in contrast to XI, CXR|I does not require $\ell = k$.

WC present further substantial analysis for identification of average and other causal effects using predictive proxies for the general nonlinear and non-separable case (where A.2 is removed). White and Chalak (2006b) discuss related parametric and nonparametric estimation methods and provide several tests for CXR|I.

Because of the straightforward framework provided by CXR|I for identifying causal effects (in particular, because there are no necessary exclusion restrictions involved) we do not provide a list of causal properties for CXR|I parallel to CP:OXI or CP:PXI. Nevertheless, we conjecture that in S_9 (and G_9) CXR|I implies (possibly with some mild additional conditions) that X cannot cause W (see *(1)* in S_9). In particular, observe that if X causes W in S_9 then conditioning on a common effect W of X and U_w generally renders X and U_w conditionally dependent given W. Since U_w causes U_y, this could lead CXR|I to fail. We leave a formal treatment of this conjecture for other work.

4.1.2 Conditional Instruments

We now examine how a single vector of *conditional instruments* Z can identify the causal effect of endogenous X on Y as the product of the effects of X on Z and that of Z on Y. We call this EIV class *intermediate cause* instrumental variables, as these variables mediate the effects of X on Y. To illustrate, consider system S_{10} and causal graph G_{10}:

(1) $X \overset{c}{=} \alpha_x U_x$

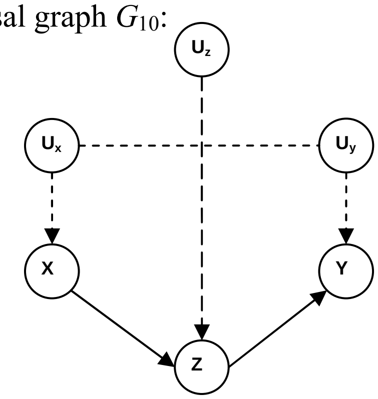
(2) $Z \overset{c}{=} \gamma_z X + \alpha_z U_z$

(3) $Y \overset{c}{=} Z'\delta_o + U_y$

where $U_x \perp U_y$, $U_x \perp U_z$, and $U_y \perp U_z$.

Substituting *(2)* into *(3)* with $\beta_o \equiv \gamma_z'\delta_o$ gives:

(3′) $Y \overset{c}{=} X'\beta_o + U_z'\alpha_z'\delta_o + U_y$.



Graph 10 (G_{10})
Conditional Instruments

This is the structure described in the introduction. Clearly X and Z are endogenous in *(3′)*, so neither XR nor XI can identify β_o. Nor do we have CXR|I in *(3′)*, as $X \not\perp U_y \mid Z$. Nevertheless, β_o is structurally identified as a result of

(CXI|R) *Conditionally Exogenous Instruments given Regressors:* $Z \perp U_y \mid X$.

For our linear separable system, CXI|R implies the key moment condition

$$E(ZU_y \mid X) = E(Z \mid X) \times E(U_y \mid X). \qquad (M4)$$

Parallel to our analysis of CXR|I, it follows from this moment condition that

$$E(\,[Z - E(ZX')[E(XX')]^{-1}X\,]\,U_y\,) = 0.$$

Thus, CXR|I with regressors Z and conditioning instruments X identifies δ_o in *(3)* as

$$\{E(ZZ') - E(ZX')[E(XX')]^{-1}E(XZ')\}\ \delta_o = E(ZY) - E(ZX')[E(XX')]^{-1}E(XY).$$

Full identification holds given non-singularity of $\{E(ZZ') - E(ZX')[E(XX')]^{-1}E(XZ')\}$.

If γ_z can also be identified, then identification of β_o follows, as $\beta_o \equiv \gamma_z'\delta_o$. In S_{10}, γ_z is structurally identified from *(2)* by XR, as $X \perp U_z$. If γ_z is stochastically identified ($E(XX')$ is non-singular), Proposition 3.1.1 gives $\gamma_z' = [E(XX')]^{-1}E(XZ')$. Thus we have

Proposition 4.1.2 Suppose A.1 and A.2 hold such that: (i) $Z \overset{c}{=} \gamma_z X + \alpha_z U_z$, $Y \overset{c}{=} Z'\delta_o + U_y$, where $E(XX')$, $E(XZ')$, $E(ZZ')$, $E(ZY)$, and $E(XY)$ exist and are finite. Suppose further that (ii) (a) $E(XX')$ is non-singular and (b) $\{E(ZZ') - E(ZX')[E(XX')]^{-1}E(XZ')\}$ is non-singular; and (iii) (a) XR: $X \perp U_z$ and (b) CXI|R: $Z \perp U_y \mid X$.

Then $\beta_o \equiv \gamma_z'\delta_o$, the average total causal effect of X on Y, is identified as:

$$\beta_o = [E(XX')]^{-1}E(XZ') \times$$
$$\{E(ZZ') - E(ZX')[E(XX')]^{-1}E(XZ')\}^{-1} \times \{E(ZY) - E(ZX')[E(XX')]^{-1}E(XY)\}\ \blacksquare$$

In contrast to XI but like CXR|I, CXI|R does not require $\ell = k$.

A consistent and asymptotically normal plug-in estimator, treated in Section 7, is:

$$\hat{\beta}_n^{CXI|R} \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Z}) \times [\boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Z}]^{-1}[\boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y}]$$
$$\equiv \hat{\gamma}_n^{XR\,\prime}\ \hat{\delta}_n^{CXR|I}.$$

Although CXI|R uses a single vector of EIVs Z to identify β_o, both Z and X play dual roles. The EIVs Z play the dual role of a response for X and a cause for Y. The regressors

X are exogenous regressors with respect to U_z in *(2)* and conditioning instruments with respect to U_y in *(3)*. We reflect these latter roles in our notation $\hat{\gamma}_n^{XR}$ and $\hat{\delta}_n^{CXR|I}$ above.

Analogous to XI, we can state a succinct set of causal properties required to ensure structural identification using conditional instruments Z:

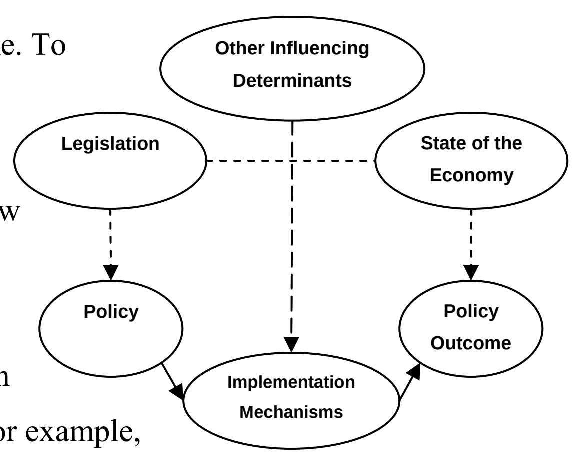
(CP:CXI|R): **Causal Properties of Conditionally Exogenous Instruments Given Regressors** (i) The effect of X on Z is identified via XR; (ii) The effect of Z on Y is identified via CXR|I with conditioning instruments X; (iii) If X causes Y, it does so only via Z. ■

As is readily verified, S_{10} satisfies CP:CXI|R.

The CXI|R method corresponds to the "front-door" method introduced by Pearl (1995, 2000). Whereas the treatment effect literature applies CXR|I (back door) to identify the effect of interest in the presence of confounding by conditioning on a covariate (W) that is *not* affected by the treatment, the CXI|R (front-door) method makes use of a variable that *is* affected by the treatment (indeed, that mediates it) to structurally identify the causal effect on Y of the treatment X.

The structure of S_{10} can also be analyzed via Hausman and Taylor (1983). As $U_z \perp U_y$ and γ_z is estimable from a regression on *(2)*, proposition 6 of Hausman and Taylor (1983) applies to ensure that the residuals from a regression based on *(2)* in S_{10} can play the role of a standard instrument for Z in *(3)* and thus yield the identification of δ_o.

The CXI|R method can play a particularly useful role in measuring policy effects, as illustrated in G_{11}. Consider evaluating the outcome of a policy that we think is endogenous since it is determined by legislation that is correlated with the state of the economy, which also determines the policy outcome. To illustrate, we might be interested in evaluating the effect on students' performance in public schools, as measured by their standardized test scores, of new legislation for education reform (see, for example, Gordon and Vegas, 2005) but suspect that the new education law is endogenous, as it is correlated with unobserved causes of the students' performance. For example,



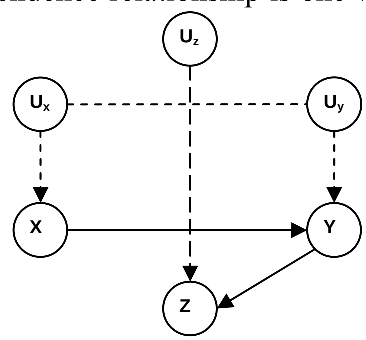
Graph 11 (G_{11})
Policy Evaluation by means of CXI|R

suppose the legislation passed due to the poor state of the economy, which itself is a cause of the students' unsatisfactory performance. Under these circumstances, one can recover the effect of the legislation using as instruments intermediate causes affected by the new policy that in turn affect student performance. These EIVs should be implementation mechanisms that are responses only to the new policy and are not otherwise caused by the unobserved common confounding causes of the policy and the response of interest. In this example, potential intermediate cause instruments could be funding per student, number of teachers per school, educational attainment of teachers, class size, and so forth.

4.1.3 Other Potential Single Extended Instruments

So far, we have examined four single EIV methods for ensuring identification of the effect of X on Y: XR: $X \perp U_y$; XI: $Z \perp U_y$; CXR|I: $X \perp U_y \mid W$; and CXI|R: $Z \perp U_y \mid X$. The remaining possibilities for (conditional) independence from U_y are associated with Y. Now $Y \perp U_y$, $Y \perp U_y \mid X$, and $Y \perp U_y \mid W$ as, by definition, U_y is an immediate cause of Y. Similarly, $X \perp U_y \mid Y$, as conditioning on a common response generally renders the possibly independent X and U_y necessarily dependent. The final possibility to consider is whether identification holds when Z is conditionally independent of U_y given Y.

A causal structure generating this conditional independence relationship is one where Z is a *post-response* instrument, as in S_{12}:



Graph 12 (G_{12})

(1) $X \overset{c}{=} \alpha_x U_x$

(2) $Y \overset{c}{=} X'\beta_o + U_y$

(3) $Z \overset{c}{=} \gamma_z Y + \alpha_z U_z$

where $U_x \perp U_y$, $U_x \perp U_z$, and $U_y \perp U_z$.

Substituting structural equation *(2)* into structural equation *(3)* with $\delta_o \equiv \beta_o \gamma_z$ we get:

(3′) $Z \overset{c}{=} X'\delta_o + \gamma_z U_y + \alpha_z U_z$.

This might look promising, as one may consider the possibility of identifying the effect on Y of the endogenous X as the ratio of the effect of X on Z and that of Y on Z, analogous to ILS. Unfortunately, the condition $Z \perp U_y \mid Y$ is not sufficient to identify β_o.

From this condition and algebra analogous to that above, we obtain the moment equations

$$\{E(ZY') - E(ZY')[E(YY')]^{-1}E(YY')\} - \{E(ZX') - E(ZY')[E(YY')]^{-1}E(YX')\}\, \beta_o = 0.$$

But the first term above is always zero, so even if $\{E(ZX') - E(ZY')[E(YY')]^{-1} E(YX')\}$ is non-singular, the solution for β_o is also always zero. The problem is that the effect of X on Z is not identified, as X and Z are confounded by the unobserved common cause of X and Y. Proposition 7 of Hausman and Taylor (1983) also rules out using the estimated residuals from a regression on *(3)* as standard instruments to identify β_o, as $Y \perp U_z$ in S_{12}.

This exhausts the possibilities for structural identification via a single vector of EIVs.

4.2 Double Extended Instrumental Variables Methods

Economic theory can suggest causal structures that permit identification of causal effects by jointly using conditional instruments Z and conditioning instruments W. We now examine the corresponding EIV methods.

For this, let Y be the response of interest, let the elements of the $k_1 \times 1$, ..., $k_p \times 1$ random vectors X_1, …, X_p be the causes of interest, and let the elements of the $\ell_1 \times 1$, …, $\ell_q \times 1$ random vectors Z_1, ..., Z_q and the $m_1 \times 1$, …, $m_s \times 1$ random vectors W_1, ..., W_s be extended instrumental variables, all with observed realizations as specified in A.1 and A.2. The corresponding unobserved causes are U_y and the elements of vectors U_{x_1}, ..., U_{x_p}, U_{z_1}, ..., U_{z_q}, and U_{w_1}, ..., U_{w_s}. Put $X \equiv [X_1', ..., X_p']'$, $Z \equiv [Z_1', ..., Z_q']'$ and $W \equiv [W_1', ..., W_s']'$, where X is $k \times 1$ with $k \equiv k_1 + ... + k_p$, Z is $\ell \times 1$ with $\ell \equiv \ell_1 + ... + \ell_q$, and W is $m \times 1$ with $m \equiv m_1 + ... + m_s$. Similarly, put $U_x \equiv [U_{x_1}', ..., U_{x_p}']'$, $U_z \equiv [U_{z_1}', ..., U_{z_q}']'$, and $U_w \equiv [U_{w_1}', ..., U_{w_s}']'$. Boldface symbols denote vectors and matrices of observations of X, Y, Z, and W, as above.

4.2.1 Conditional and Conditioning Instruments: CXI|I

4.2.1.a OCXI|I

Our first case is that of *observed conditionally exogenous instruments given conditioning instruments* (OCXI|I). To illustrate, consider structural system S_{13a} with associated causal graph G_{13a}, where S_{13a} is given by:

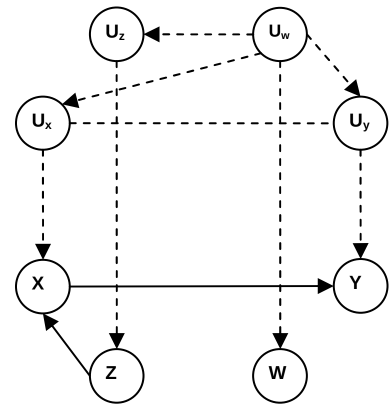
(1) $W \overset{c}{=} \alpha_w U_w$

(2) $Z \overset{c}{=} \alpha_z U_z$

(3) $X \overset{c}{=} \gamma_x Z + \alpha_x U_x$

(4) $Y \overset{c}{=} X'\beta_o + U_y$,

where $U_x \perp U_y$, $U_x \perp U_z$, $U_x \perp U_w$, $U_y \perp U_z$, $U_y \perp U_w$, and $U_z \perp U_w$. Substituting structural equation *(3)* into *(4)* and setting $\pi_o \equiv \gamma_x{}' \beta_o$ gives



Graph 13a (G_{13a})
Observed Conditionally Exogenous Instruments
Given Conditioning Instruments (OCXI|I)

(4′) $Y \overset{c}{=} Z'\pi_o + U_x{}'\alpha_x{}'\beta_o + U_y$

The key conditional independence relationship that holds in S_{13a} when W is a sufficiently good predictor for U_w (hence Z) is:

(CXI|I) *Conditionally Exogenous Instruments given Conditioning Instruments*: $Z \perp U_y \mid W$.

Given A.2, the key moment condition resulting from CXI|I is:

$$E(ZU_y \mid W) = E(Z \mid W) \times E(U_y \mid W). \qquad (M5)$$

Algebra similar to that for CXR|I delivers the structural identification of β_o under CXI|I.

Proposition 4.2.1 Suppose A.1 and A.2 hold such that: (i) $Y \overset{c}{=} X'\beta_o + U_y$. Suppose further that (ii) there exist random vectors W and Z such that and that $\ell = k$; $E(ZY)$, $E(ZW')$, $E(WW')$, $E(WY)$, and $E(ZX')$ exist and are finite; $E(WW')$ is non-singular and $E(Z \mid W) = E(ZW')[E(WW')]^{-1}W$; (iii) $E(ZX') - E(ZW')[E(WW')]^{-1} E(WX')$ is non-singular; and (iv) CXI|I: $Z \perp U_y \mid W$.

Then, β_o, the average total causal effect of X on Y, is fully identified as

$$\beta_o = \{E(ZX') - E(ZW')[E(WW')]^{-1}E(WX')\}^{-1} \times \{E(ZY) - E(ZW')[E(WW')]^{-1}E(WY)\}. \ \blacksquare$$

Note that $\ell = k$, analogous to XI. No previous method can identify β_o in this case, as none of the other admissible conditional independence relationships hold in S_{13a}.

The plug-in CXI|I estimator, covered by the results of Section 7, is

$$\hat{\beta}_n^{CXI|I} \equiv [\boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}')\boldsymbol{X}]^{-1}[\boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}')\boldsymbol{Y}].$$

This is standard IV with residual instruments $Z - E(ZW')E(WW')^{-1} W$, but as discussed above for CXR|I, the key role in identifying β_o is played by W and Z, not these residuals.

In S_{13a}, Z satisfies the following causal properties that parallel CP:OXI.

(CP:OCXI|I): **Causal Properties of Observed Conditionally Exogenous Instruments given Conditioning Instruments** (i) Z directly causes X, and the effect of Z on X is identified via CXR|I with conditioning instruments W; (ii) Z indirectly causes Y, and the effect of Z on Y is identified via CXR|I with conditioning instruments W; (iii) Z causes Y only via X. ■

As Z is observed, we call this case *observed conditionally exogenous instruments given conditioning instruments* (OCXI|I). Similar to the method of XI, the effect of X on Y is identified here as the "ratio" of the identified effects of Z on X and that of Z on Y.

4.2.1.b PCXI|I

As for XI and CXR|I, the true underlying cause need not be observed; a suitable proxy suffices. In fact, this feature applies to all EIV methods we discuss (see Theorem 5.1 below). We illustrate this in S_{13b} and associated causal graph G_{13b}, where S_{13b} is given by:

(1) $W \stackrel{c}{=} \alpha_w U_w$

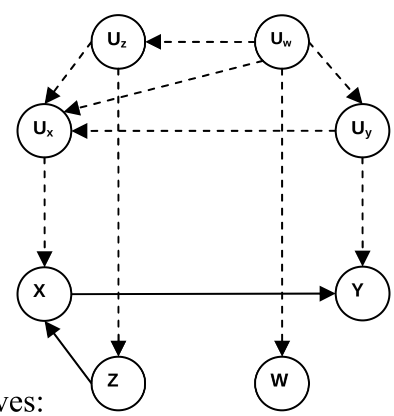
(2) $Z \stackrel{c}{=} \alpha_z U_z$

(3) $X \stackrel{c}{=} \gamma_x Z + \alpha_x U_x$

(4) $Y \stackrel{c}{=} X'\beta_o + U_y$,

where $U_x \perp U_y$, $U_x \perp U_z$, $U_x \perp U_w$, $U_y \perp U_z$, $U_y \perp U_w$, and $U_z \perp U_w$. Substituting (3) into (4) and setting $\pi_o \equiv \gamma_x{}' \beta_o$ gives:

(4′) $Y \stackrel{c}{=} Z'\pi_o + U_x{}'\alpha_x{}'\beta_o + U_y$



Graph 13b (G_{13b})
Proxy for Unobserved Conditionally Exogenous Instruments Given Conditioning Instruments (PCXI|I)

Here CXI|I holds when W is a sufficiently good predictor for U_w; Proposition 4.2.1 then applies to fully identify β_o, as for OCXI|I. Here, however, the effects of Z on X and of Z on Y are no longer identified, so CP:OCXI|I fails. Instead, U_z plays the key role, and Z acts as a proxy for U_z. Parallel to PXI, we call Z *proxies for (unobserved) conditionally exogenous instruments given conditioning instruments* (PCXI|I). The parallel causal properties permitting identification of β_o are:

(CP:PCXI|I) Causal Properties for Proxies for Unobserved Conditionally Exogenous Instruments given Conditioning Instruments (i) U_z indirectly causes X, and the full effect of U_z on X could be identified via CXR|I with conditioning instruments W had U_z been observed; (ii) U_z indirectly causes Y, and the full effect of U_z on Y could be identified via CXR|I with conditioning instruments W had U_z been observed; (iii) U_z causes Y only via X; (iv) if Z causes Y, it does so only via X . ■

CP:PCXI|I parallels CP:PXI, but identification in (i) and (ii) is via CXR|I with conditioning instruments W, not XR. Our comments about PXI fully apply here, in that the analog of ILS fails. Nevertheless, the ratio of two inconsistent CXR|I estimators remains informative for β_o. As in the PXI case, Z is not required to cause X in S_{13b}. When $\gamma_x = 0$, Z acts as a "pure predictive proxy" for U_z.

4.2.2 Conditional and Conditioning Instruments: CXIR|I

When conditioning instruments W render only a subvector X_2 of $X \equiv [X_1', X_2']'$ conditionally exogenous, the previous methods cannot structurally identify $\beta_o \equiv [\beta_1', \beta_2']'$. Nevertheless, identification obtains given conditional instruments Z for X_1 that are conditionally exogenous given W. To illustrate, let S_{14} be given by:

(1) $W \stackrel{c}{=} \alpha_w U_w$

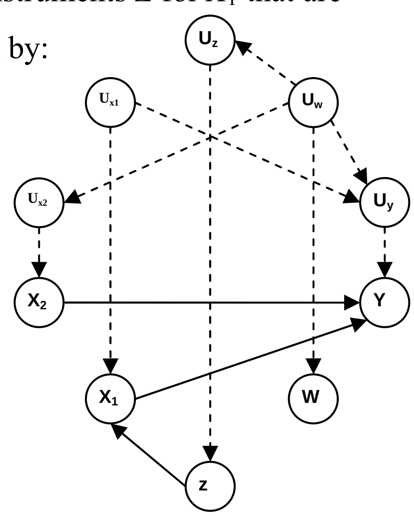
(2) $Z \stackrel{c}{=} \alpha_z U_z$

(3) $X_1 \stackrel{c}{=} \gamma_{x_1} Z + \alpha_{x_1} U_{x_1}$

(4) $X_2 \stackrel{c}{=} \alpha_{x_2} U_{x_2}$

(5) $Y \stackrel{c}{=} X_1'\beta_1 + X_2'\beta_2 + U_y$

where $U_{x_1} \perp U_{x_2}$, $U_{x_1} \not\perp U_y$, $U_{x_2} \not\perp U_y$, $U_{x_1} \perp U_z$, $U_{x_2} \not\perp U_z$, $U_{x_1} \perp U_w$, $U_{x_2} \not\perp U_w$, $U_y \not\perp U_z$, $U_y \not\perp U_w$, $U_z \not\perp U_w$.



Graph 14 (G_{14})
Conditionally Exogenous Instruments and Regressors Given Conditioning Instruments (CXIR|I)

The key conditional independence relationship that holds in S_{14} is:

(CXIR|I) *Conditionally Exogenous Instruments and Regressors given Conditioning Instruments:* $(Z, X_2) \perp U_y \mid W$.

CXIR|I is the special case of CXI|I in which X_2 plays the role of a conditionally exogenous instrument for itself. We call $\tilde{Z} = [Z', X_2']'$ *conditionally exogenous instruments and regressors given conditioning instruments*. The key moment condition is

$$E(\tilde{Z}\, U_y \mid W) = E(\tilde{Z} \mid W) \times E(U_y \mid W). \qquad (M6)$$

The corresponding identification result (in which $\ell = k_1$ is necessary) is

Proposition 4.2.2 Suppose A.1 and A.2 hold such that: (i) $Y \overset{c}{=} X_1'\beta_1 + X_2'\beta_2 + U_y$, with $X \equiv [X_1', X_2']'$ and $\beta_o \equiv [\beta_1', \beta_2']'$. Suppose further that (ii) there exist random vectors W and Z such that $\ell = k_1$ and that with $\tilde{Z} = [Z', X_2']'$, $E(\tilde{Z} X')$, $E(\tilde{Z} W')$, $E(WW')$, $E(WX')$, $E(\tilde{Z} Y)$, and $E(WY)$ exist and are finite; $E(WW')$ is non-singular and $E(\tilde{Z} \mid W) = E(\tilde{Z} W')[E(WW')]^{-1}W$; (iii) $E(\tilde{Z} X') - E(\tilde{Z} W')[E(WW')]^{-1} E(WX')$ is non-singular; and (iv) CXIR|I: $(Z, X_2) \perp U_y \mid W$.

Then β_o, the average total causal effect of X on Y, is fully identified as

$$\beta_o = \{E(\tilde{Z} X') - E(\tilde{Z} W')[E(WW')]^{-1}E(WX')\}^{-1}\{E(\tilde{Z} Y) - E(\tilde{Z} W')[E(WW')]^{-1}E(WY)\}$$

■

The CXIR|I plug-in estimator treated in Section 7 is

$$\hat{\beta}_n^{CXIR|I} \equiv [\,\tilde{\boldsymbol{Z}}' (\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}')\boldsymbol{X}]^{-1}[\,\tilde{\boldsymbol{Z}}' (\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}')\boldsymbol{Y}].$$

4.2.3 Conditional and Conditioning Instruments: CXI|RI

A generalization of CXI|R occurs when CXI|R fails but conditioning instruments W and regressors X jointly render extended instruments Z conditionally exogenous, as in S_{15}:

(1) $W \overset{c}{=} \alpha_w U_w$

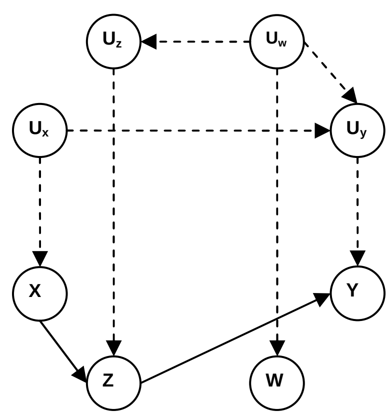
(2) $X \overset{c}{=} \alpha_x U_x$

(3) $Z \overset{c}{=} \gamma_z X + \alpha_z U_z$

(4) $Y \overset{c}{=} Z'\delta_o + U_y$

where $U_x \perp U_y$, $U_x \perp U_z$, $U_x \perp U_w$, $U_y \perp U_z$, $U_y \perp U_w$, and $U_z \perp U_w$. Substituting equation *(3)* into equation *(4)* with $\beta_o \equiv \gamma_z'\delta_o$ gives:

(4') $Y \overset{c}{=} X'\beta_o + U_z'\alpha_z'\delta_o + U_y$.



Graph 15 (G_{15})
Conditionally Exogenous Instruments given Regressors and Conditioning Instruments (CXI|RI)

The key conditional independence relationship that holds in S_{15} is:

(CXI|RI) *Conditionally Exogenous Instruments given Regressors and Conditioning Instruments:* $Z \perp U_y \mid (X, W)$.

We call Z *conditionally exogenous instruments given regressors and conditioning instruments*. The key moment condition resulting from CXR|RI is:

$$E(ZU_y \mid \tilde{W}) = E(Z \mid \tilde{W}) \times E(U_y \mid \tilde{W}), \tag{M7}$$

where $\tilde{W} = [X', W']'$. Like CXI|R, CXI|RI identifies β_o as the product of the effects of X on Z and of Z on Y. The causal properties of CXI|RI parallel those of CP:CXI|R.

Proposition 4.2.3 Suppose A.1 and A.2 hold such that: (i) $Z \overset{c}{=} \gamma_z X + \alpha_z U_z$, $Y \overset{c}{=} Z'\delta_o + U_y$, where $E(XX')$, $E(XZ')$, $E(ZZ')$, and $E(ZY)$ exist and are finite. Suppose further that (ii) there exists a random vector W such that with $\tilde{W} = [X', W']'$, $E(Z\tilde{W}')$, $E(\tilde{W}\tilde{W}')$, and $E(\tilde{W}Y)$ exist and are finite; $E(\tilde{W}\tilde{W}')$ is non-singular and $E(Z \mid \tilde{W}) = E(Z\tilde{W}')[E(\tilde{W}\tilde{W}')]^{-1}\tilde{W}$; (iii) (a) $E(XX')$ and (b) $\{E(ZZ') - E(Z\tilde{W}')[E(\tilde{W}\tilde{W}')]^{-1} E(\tilde{W}Z')\}$ are non-singular; and (iv) (a) XR: $X \perp U_z$ and (b) CXI|RI: $Z \perp U_y \mid \tilde{W}$.

Then $\beta_o \equiv \gamma_z'\delta_o$, the average total causal effect of X on Y, is fully identified as:

$$\beta_o = [E(XX')]^{-1}E(XZ') \times$$
$$\{E(ZZ') - E(Z\tilde{W}')[E(\tilde{W}\tilde{W}')]^{-1}E(\tilde{W}Z')\}^{-1}\{E(ZY) - E(Z\tilde{W}')[E(\tilde{W}\tilde{W}')]^{-1}E(\tilde{W}Y)\} \quad \blacksquare$$

A key feature of S_{15} is that $X \perp U_z$. It should now be clear that this can be relaxed to a conditional independence relationship, such as CXR|I: $X \perp U_z \mid W_1$, with W_1 a suitable vector of conditioning instruments. Pearl (1995, 2000) provides graphical criteria for structural identification to obtain in such a manner via his "front door" method.

The plug-in estimator $\hat{\beta}^{CXI|RI}$ treated in Section 7 is

$$\hat{\beta}_n^{CXI|RI} \equiv [(\boldsymbol{X}'\boldsymbol{X})]^{-1}(\boldsymbol{X}'\boldsymbol{Z}) \times [\boldsymbol{Z}'(\boldsymbol{I} - \tilde{\boldsymbol{W}}(\tilde{\boldsymbol{W}}'\tilde{\boldsymbol{W}})^{-1}\tilde{\boldsymbol{W}}')\boldsymbol{Z}]^{-1}[\boldsymbol{Z}'(\boldsymbol{I} - \tilde{\boldsymbol{W}}(\tilde{\boldsymbol{W}}'\tilde{\boldsymbol{W}})^{-1}\tilde{\boldsymbol{W}}')\boldsymbol{Y}]$$
$$\equiv \hat{\gamma}_n^{XR\,\prime}\,\hat{\delta}_n^{CXR|I}.$$

4.2.4 Further Comments on Double Extended Instrumental Variables

As in the single EIV case, conditions of the form $Z \perp U_y \mid (Y, W)$ do not permit structural identification of β_o. The demonstration is entirely parallel to that of Section 4.1.3.

The double EIV methods CXI|I, CXIR|I, and CXI|RI, together with the single EIV methods of Sections 3.1, 3.2, and 4.1 thus provide a basis for all EIV methods discussed so far. In fact XR, XI, CXR|I, CXI|I, and CXRI|I constitute an exhaustive set of "primitive" methods, since other EIV methods, such as CXI|R and CXI|RI, identify causal effects as functions of effects identified by use of one or more of these primitives.

5. A Master Theorem for EIV Identification

We now summarize our previous results by stating a "master theorem" that provides not just sufficient conditions for identification, but necessary and sufficient conditions.

Theorem 5.1 Suppose A.1 and A.2 hold for a structural system S such that: (i) $Y \overset{c}{=} X'\beta_o + U_y$, where X is $k \times 1$, $k > 0$, and β_o is finite and $k \times 1$. Suppose further that (ii) Z ($\ell \times 1$, $\ell \geq 0$) and W ($m \times 1$, $m \geq 0$) are random vectors determined by S, and let $\tilde{Z}$ and $\tilde{W}$ be $k \times 1$ and $\tilde{m} \times 1$ vectors respectively such that $[\tilde{Z}', \tilde{W}']' = \boldsymbol{A}\,[X', Z', W']'$, for a given $(k + \tilde{m}) \times (k + \ell + m)$ matrix $\boldsymbol{A}$, and that $E(\tilde{Z}\,Y)$, $E(\tilde{Z}\,X')$, $E(\,E(\tilde{Z}\,|\,\tilde{W})\;Y\,)$, and $E(\,E(\tilde{Z}\,|\,\tilde{W})\,X')$ exist and are finite. Then

(a) $E\{[\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]U_y\}$ exists and is finite.

(b) Stochastic identification holds, that is, there exists a unique β^* such that

$$E(\,[\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]\,X')\,\beta^* = E([\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]Y) - E\{[\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]U_y\}$$

if and only if $E(\,[\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]\,X')$ is non-singular.

(c) Structural identification holds, that is β_o satisfies

$$E(\,[\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]\,X')\,\beta_o - E([\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]\;Y) = 0$$

if and only if $E\{[\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]U_y\} = 0$.

(d) The average causal effect β_o is fully identified as

$$\beta_o = \{\,E(\,[\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]\,X')\}^{-1} \times \{E(\,[\tilde{Z} - E(\tilde{Z}\,|\,\tilde{W})]\;Y\,)\}$$

if and only if stochastic and structural identification jointly hold. ■

For the sake of generality, we now do not impose linear structure on $E(\tilde{Z}\,|\,\tilde{W})$. The previous sections employ this for concreteness. When stochastic identification holds but not structural identification, then we have

$$\beta^* = \beta_o + \{ E([\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] X') \}^{-1} E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] U_y\}.$$

With linearity for $E(\tilde{Z} \mid \tilde{W})$, β^* is the probability limit of the plug-in EIV estimator

$$\hat{\beta}_n^{EIV} \equiv [\tilde{\boldsymbol{Z}}' (\boldsymbol{I} - \tilde{\boldsymbol{W}} (\tilde{\boldsymbol{W}}' \tilde{\boldsymbol{W}})^{-1} \tilde{\boldsymbol{W}}') \boldsymbol{X}]^{-1} [\tilde{\boldsymbol{Z}}' (\boldsymbol{I} - \tilde{\boldsymbol{W}} (\tilde{\boldsymbol{W}}' \tilde{\boldsymbol{W}})^{-1} \tilde{\boldsymbol{W}}') \boldsymbol{Y}].$$

Thus, $\hat{\beta}_n^{EIV}$ converges to the true average causal effect, β_o, plus a "causal discrepancy,"

$$\delta^* \equiv \{E([\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] X') \}^{-1} E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] U_y\}.$$

If structural but not stochastic identification holds, then the estimating equations

$$[\tilde{\boldsymbol{Z}}' (\boldsymbol{I} - \tilde{\boldsymbol{W}} (\tilde{\boldsymbol{W}}' \tilde{\boldsymbol{W}})^{-1} \tilde{\boldsymbol{W}}') \boldsymbol{X}] \beta - [\tilde{\boldsymbol{Z}}' (\boldsymbol{I} - \tilde{\boldsymbol{W}} (\tilde{\boldsymbol{W}}' \tilde{\boldsymbol{W}})^{-1} \tilde{\boldsymbol{W}}') \boldsymbol{Y}] = 0$$

define a set of solutions converging stochastically to a set that contains β_o, but there is insufficient information to identify which element of the set is the true causal effect.

Theorem 5.1 contains XR, CXR|I, XI, CXI|I, and CXRI|I as special cases. For these, an exclusion restriction acts to ensure that when Z is present, Z causes Y only via X. It follows easily that conditional independence ($\tilde{Z} \perp U_y \mid \tilde{W}$), conditional mean independence ($E(U_y \mid \tilde{Z}, \tilde{W}) = E(U_y \mid \tilde{W})$), and conditional non-correlation ($E(\tilde{Z} U_y \mid \tilde{W}) = E(\tilde{Z} \mid \tilde{W}) E(U_y \mid \tilde{W})$) each imply the necessary structural identification condition $E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] U_y\} = 0$. The derived standard instruments $\tilde{Z} - E(\tilde{Z} \mid \tilde{W})$ satisfy this moment condition. We also call these "residual instruments," analogous to those of Hausman and Taylor (1983).

The next result extends Theorem 5.1 to cover cases such as CXI|R and CXI|RI, where causal effects are identified as a function of effects identifiable as in Theorem 5.1.

Corollary 5.2 Suppose A.1 and A.2 hold for a structural system S such that $Y \overset{c}{=} X'\beta_o + U_y$, where X is $k \times 1$, $k > 0$, and β_o is finite and $k \times 1$. For $H > 0$, let $\theta_1, \ldots, \theta_H$, be real-valued vectors of structural coefficients of S, and let $b(\cdot)$ be a known measurable real vector-valued function such that $\beta_o = b(\theta_1, \ldots, \theta_H)$. If $\theta_1, \ldots, \theta_H$ are each fully identified as in Theorem 5.1, then, β_o is fully identified as $b(\theta_1, \ldots, \theta_H)$. ■

6. Characterization of Structural Identification via Causal Matrices: Single EIV

Causal matrices effectively characterize the causal structures in which the identification of given causal effects of interest obtains. In particular, Chalak and White (2006) give a procedure to generate *conditional independence matrices* from causal matrices. These characterize the conditional independence relationships holding among the variables of a given system S, conditioning on any subset of system variables. (The empty set yields the *independence matrix*.) Thus, by inspecting the conditional independence matrices one can determine whether the necessary exogeneity or conditional exogeneity relationships hold for structural identification of given causal effects.

Every causal matrix C_S also has an associated *path matrix* P_S. The (k, l) entry of P_S, p_{kl}, is 1 if there is a (V_k, V_l)-path in G_S and is 0 otherwise. Formally, $P_S = p(C_S)$ where:

$$
\begin{aligned}
p_{kl} &= 1 && \text{if there exists } h > 0 \text{ and a set } \{g_1, \ldots, g_h\} \text{ with elements} \\
& && \text{in } \{1, \ldots, G\} \text{ such that } c_{kg_1} \times \ldots \times c_{g_h l} = 1; \\
p_{kl} &= c_{kl} && \text{otherwise.}
\end{aligned}
$$

Thus, P_S summarizes all direct and indirect causal relationships between the variables of S. The path matrix, together with its corresponding causal matrix, expresses concisely the exclusion restrictions necessary for the identification of causal effects.

By examining the causal, path, and conditional independence matrices, one can check whether structural identification of given causal effects obtains. We illustrate this for the case of a single EIV Z and single cause and response variables X and Y. For the single EIV case under A.1 and A.2, the causal matrix has the form

$$
C_S \;=\; \left|\begin{array}{c:c} C_{S_1} & C_{S_2} \\ \hdashline C_{S_3} & C_{S_4} \end{array}\right| \;=\;
\begin{array}{c|ccc:ccc|}
 & X & Y & Z & U_x & U_y & U_z \\
\hline
X & 0 & & & 0 & 0 & 0 \\
Y & 0 & 0 & & 0 & 0 & 0 \\
Z & & & 0 & 0 & 0 & 0 \\
\hdashline
U_x & 1 & 0 & 0 & 0 & & \\
U_y & 0 & 1 & 0 & & 0 & \\
U_z & 0 & 0 & 1 & & & 0
\end{array}
$$

The entries in the off-diagonal blocks follow by our conventions, as do the diagonal elements. We have $c_{21} = 0$ by acyclicality and because the effect of interest is that of X on Y. Further, acyclicality imposes on C_{S_1} three constraints of the form $c_{jk} \times c_{kj} = 0$ and two constraints of the form $c_{jk} \times c_{kl} \times c_{lj} = 0$ for $j, k, l = 1, 2, 3$ and on C_{S_4} three constraints of the form $c_{jk} \times c_{kj} = 0$ and two constraints of the form $c_{jk} \times c_{kl} \times c_{lj} = 0$ for $j, k, l = 4, 5, 6$.

Table I displays all possible acyclic causal structures that can relate X, Y, and a single

EIV *Z*. As illustrated there, under A1 C_{S_1} admits 9 possible values that we label in relation to *Z*. These include the pre-cause, intermediate cause, and post-response instrument cases. In addition, entries (1, 1), (1, 2), (2, 1), and (2, 2) of Table I depict common cause instrument cases, valid when appropriate causal relationships hold among the unobserved variables. Structures not obeying the exclusion restrictions for identification appear in the second column of the second, third, and fourth rows. We refer to cases in the first row as the *non-causal* case in entry (1, 1), the *joint cause* case in entry (1, 2), and the *joint response* case in entry (1, 3).

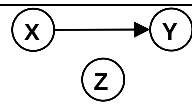
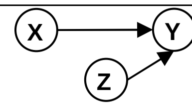
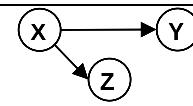
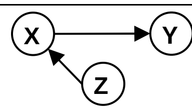
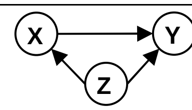
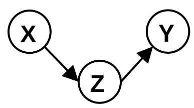
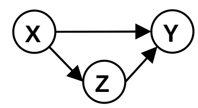
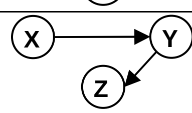
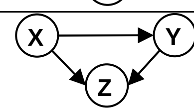
Non Causal, Joint Cause, and Joint Response			
Pre-Cause			
Intermediate Cause			
Post-Response			

Table I
Acyclic Causal Relationships for C_{S_1} in the
Single Extended Instrumental Variable Case.

Inspection of C_{S_4} reveals that for every entry of Table I, there are 25 possible acyclic structures that relate the unobserved variables. Thus, C_S can represent 225 (25×9) potential acyclic causal structures in the single EIV case. The analysis simplifies by restricting attention to the presence or absence of statistical independence among the unobservables, as is standard practice in the literature. The 25 possible acyclic structures reduce to 8 possible sets of independence/dependence relationships among U_x, U_y, and U_z. We thus have 72 (8×9) possible structural equations systems.

The cases discussed in Sections 3.1, 3.2, and 4.1 are the only ones for which the structural identification of the effect of *X* on *Y* holds in the single EIV case. Specifically, the values of C_S that have corresponding conditional independence matrices indicating that at least one exogeneity or conditional exogeneity relationship holds, together with corresponding path matrices indicating the needed exclusion restrictions, exhaustively characterize the acyclic causal structures admitting structural identification. These are precisely the cases presented in Sections 3.1, 3.2, and 4.1.

For example, the second columns of the pre-cause and intermediate-cause categories in Table I violate the exclusion restrictions that Z causes Y only via X in the first case and that X causes Y only via Z in the second case. Hence, identification is not possible in these cases, even when appropriate exogeneity or conditional exogeneity conditions hold.

7. Asymptotic Properties of EIV Estimators

With linearity assumed for $E(\tilde{Z}\,|\,\tilde{W})$, plug-in EIV estimators for causal coefficients identified by Theorem 5.1 have the form

$$\hat{\beta}_n^{EIV} \equiv [\,\tilde{\boldsymbol{Z}}'\,(\boldsymbol{I} - \tilde{\boldsymbol{W}}\,(\tilde{\boldsymbol{W}}'\,\tilde{\boldsymbol{W}}\,)^{-1}\,\tilde{\boldsymbol{W}}'\,)\boldsymbol{X}]^{-1}[\,\tilde{\boldsymbol{Z}}'\,(\boldsymbol{I} - \tilde{\boldsymbol{W}}\,(\tilde{\boldsymbol{W}}'\,\tilde{\boldsymbol{W}}\,)^{-1}\,\tilde{\boldsymbol{W}}'\,)\boldsymbol{Y}]\,.$$

Standard arguments easily yield an asymptotic normality result for this estimator.

Theorem 7.1 Suppose the conditions of Theorem 5.1 ensuring the identification of β_o hold with $E(\tilde{Z}\,|\,\tilde{W}) = E(\tilde{Z}\,\tilde{W}')[E(\tilde{W}\,\tilde{W}')]^{-1}\,\tilde{W}$, where $E(\tilde{Z}\,\tilde{W}')$ and $E(\tilde{W}\,\tilde{W}')$ exist and are finite, and that $E(\tilde{W}\,\tilde{W}')$ is nonsingular. Suppose further that

(i) $\tilde{\boldsymbol{Z}}'\,(\boldsymbol{I} - \tilde{\boldsymbol{W}}\,(\tilde{\boldsymbol{W}}'\,\tilde{\boldsymbol{W}}\,)^{-1}\,\tilde{\boldsymbol{W}}'\,)\boldsymbol{X}\,/\,n \xrightarrow{p} Q \equiv E(\tilde{Z}\,X') - E(\tilde{Z}\,\tilde{W}')[E(\tilde{W}\,\tilde{W}')]^{-1}E(\tilde{W}\,X')$;

(ii) $n^{-1/2}\sum_{i=1}^{n}\,[\tilde{Z}_i - E(\tilde{Z}_i\,|\,\tilde{W}_i)]U_{y,i} \xrightarrow{d} N(0, V)$, where V is finite and positive definite.

Then $n^{1/2}\,(\hat{\beta}_n^{EIV} - \beta_o) \xrightarrow{d} N(0, Q^{-1}V\,Q'^{\,-1})$. ■

Plug-in EIV estimators for average causal effects identified by Corollary 5.2 are given by

$$\hat{\beta}_n^{EIV} \equiv b(\hat{\theta}_n^{EIV}),$$

where $\hat{\theta}_n^{EIV} = (\hat{\theta}_{1,n}^{EIV}{}',\ \ldots,\ \hat{\theta}_{H,n}^{EIV}{}')'$ is a vector of plug-in EIV estimators of the form covered by Theorem 7.1. To state a formal result, let

$$\hat{\theta}_{h,n}^{EIV} \equiv [\,\tilde{\boldsymbol{Z}}_{\boldsymbol{h}}'\,(\boldsymbol{I} - \tilde{\boldsymbol{W}}_{\boldsymbol{h}}(\tilde{\boldsymbol{W}}_{\boldsymbol{h}}'\,\tilde{\boldsymbol{W}}_{\boldsymbol{h}})^{-1}\tilde{\boldsymbol{W}}_{\boldsymbol{h}}'\,)\,\boldsymbol{X}_{\boldsymbol{h}}]^{-1}[\,\tilde{\boldsymbol{Z}}_{\boldsymbol{h}}'\,(\boldsymbol{I} - \tilde{\boldsymbol{W}}_{\boldsymbol{h}}\,(\tilde{\boldsymbol{W}}_{\boldsymbol{h}}'\,\tilde{\boldsymbol{W}}_{\boldsymbol{h}})^{-1}\tilde{\boldsymbol{W}}_{\boldsymbol{h}}'\,)\boldsymbol{Y}_{\boldsymbol{h}}\,] \quad h = 1, \ldots, H,$$

$$\zeta_{h,i} \equiv [\,\tilde{Z}_{h,i} - E(\tilde{Z}_{h,i}\,|\,\tilde{W}_{h,i})]\;U_{y_h,i} \qquad i = 1, \ldots, n;\ \ h = 1, \ldots, H,$$

and put $\zeta_i \equiv (\zeta_{1,i}{}', \ldots, \zeta_{H,i}{}')'$.

Theorem 7.2 Suppose the conditions of Corollary 5.2 hold with $\theta_o \equiv (\theta_1', \ldots, \theta_H')'$, with $E(\tilde{Z}_h \mid \tilde{W}_h) = E(\tilde{Z}_h \tilde{W}_h')[E(\tilde{W}_h \tilde{W}_h')]^{-1} \tilde{W}_h$, where $E(\tilde{Z}_h \tilde{W}_h')$ and $E(\tilde{W}_h \tilde{W}_h')$ exist and are finite, and that $E(\tilde{W}_h \tilde{W}_h')$ is nonsingular, $h = 1, \ldots, H$. Suppose further that

(i) $\tilde{\boldsymbol{Z}}_{\boldsymbol{h}}' (\boldsymbol{I} - \tilde{\boldsymbol{W}}_{\boldsymbol{h}} (\tilde{\boldsymbol{W}}_{\boldsymbol{h}}' \tilde{\boldsymbol{W}}_{\boldsymbol{h}})^{-1} \tilde{\boldsymbol{W}}_{\boldsymbol{h}}') \boldsymbol{X}_{\boldsymbol{h}} / n \xrightarrow{p} Q_h \equiv E(\tilde{Z}_h X_h') - E(\tilde{Z}_h \tilde{W}_h')[E(\tilde{W}_h \tilde{W}_h')]^{-1} E(\tilde{W}_h X_h')$, $h = 1, \ldots, H$;

(ii) $n^{-1/2} \sum_{i=1}^{n} \zeta_i \xrightarrow{d} N(0, V)$, where V is finite and positive definite.

Then $n^{1/2} (\hat{\theta}_n^{EIV} - \theta_o) \xrightarrow{d} N(0, Q^{-1} V Q'^{-1})$, where $Q = \text{diag}\,(Q_1, \ldots, Q_H)$.

Suppose further that b is continuously differentiable at θ_o such that $\nabla b(\theta_o)$ (the gradient of b at θ_o) has full column rank. Then with $\hat{\beta}_n^{EIV} \equiv b(\hat{\theta}_n^{EIV})$ and $\beta_o \equiv b(\theta_o)$,

$$n^{1/2} (\hat{\beta}_n^{EIV} - \beta_o) \xrightarrow{d} N(0, \nabla b(\theta_o)' Q^{-1} V Q'^{-1} \nabla b(\theta_o)). \qquad \blacksquare$$

White (2001, ch. 3, 5) gives straightforward primitive conditions ensuring hypotheses (i) (law of large numbers) and (ii) (central limit theorem) of Theorems 7.1 and 7.2.

These plug-in estimators are straightforward to compute, and their asymptotic covariance matrices can be robustly estimated in the usual way under mild conditions (e.g., as in White, 2001, ch. 6). Nevertheless, they are not necessarily asymptotically efficient. Efficiency arises from optimally choosing the extended instruments in a manner similar to the way in which optimal instruments are chosen in the standard IV framework. In particular, GLS-like corrections for conditional heteroskedasticity are involved in obtaining the optimal instruments for EIV. We leave this analysis to other work.

8. Conclusion

Building on the structural equations, treatment effects, and machine learning literatures, we utilize the settable system framework of White (2006) and White and Chalak (2006a) to present an explicit and rigorous framework that permits the identification and estimation of causal effects with the aid of *extended instrumental variables* (EIV). EIV methods use variables that are not valid instruments in the traditional sense, but that emerge from a given causal structure to enable the recovery of causal effects of interest. We analyze *single* and *double* extended instrumental variables methods.

In the single EIV case, we demonstrate how the use of a single vector of *unconditional*, *conditional*, or *conditioning* EIVs permits identification of causal effects of potentially endogenous causes on a response of interest. Specifically, we analyze the cases of *exogenous regressors* (XR), *exogenous instruments* (XI), *conditionally exogenous regressors given conditioning instruments* (CXR|I), and *conditionally exogenous instruments given regressors* (CXI|R). For XI, we provide a causal account for two subcategories: *observed exogenous instruments* (OXI) and *proxies for unobserved exogenous instruments* (PXI), thereby extending work of Angrist, Imbens and Rubin (1996). We also causally explain the failure of XI in the standard irrelevant instrument, invalid instrument, and under-identified cases.

In the *double* EIV case, we show how joint use of conditional and conditioning EIVs permits identification of causal effects. We analyze the cases of *conditionally exogenous instruments given conditioning instruments* (CXI|I), *conditionally exogenous instruments and regressors given conditioning instruments* (CXIR|I), and *conditionally exogenous instruments given regressors and conditioning instruments* (CXI|RI). For CXR|I and double EIV methods, we show how identification results from restrictions on certain conditional covariances, extending results of Hausman and Taylor (1983). We state a master theorem giving necessary and sufficient conditions for the identification of causal effects via EIV methods and provide straightforward high-level conditions ensuring consistency and asymptotic normality for EIV plug-in estimators.

By using *causal*, *path*, and *conditional independence matrices* one can characterize the cases where structural identification holds. We illustrate this in the single EIV case, demonstrating that the XR, XI, CXR|I, and CXI|R methods exhaust the single EIV methods capable of structurally identifying causal effects. Chalak and White (2006) give procedures for generating conditional independence matrices from causal matrices and establishing identification results for EIV methods more generally.

Here, we consider identification of causal effects given causal structures specified *a priori*. Chalak and White (2006) give methods for generating the class of causal matrices that agree with a collection of given (observed) conditional independence matrices. This yields methods for suggesting or ruling out potential causal structures. There we propose methods for causal inference based on those results and our present identification results.

Future work will analyze asymptotic efficiency for EIV in the linear separable case. In other work (White and Chalak, 2006a, 2006b), we analyze nonparametric identification and estimation of general causal effects, relaxing A.2 to the non-separable case. White (2006) and White and Chalak (2006b) give tests of conditional exogeneity. Other planned work extends these and studies new tests for use with EIV methods.

Throughout this paper, we have provided examples of the use of EIV methods relevant to the labor economics and policy evaluation literatures. Our hope is that these methods will prove broadly helpful in empirical applications focused on modeling, understanding, and measuring causal effects of interest. Our methods also offer a possible alternative to handling the consequences of weak instruments: when standard instruments are weak, there may be extended instruments that are either less weak or not at all weak for identifying effects of interest. This is another interesting avenue for further research.

Appendix A: Mathematical Proofs

Proof of Proposition 3.1.1 From (iii), $E(XU_y) = 0$. From (i), $U_y = Y - X'\beta_o$. Substituting this into $E(XU_y) = 0$ gives $E(XY) - E(XX')\beta_o = 0$. From (ii), $E(XX')$ is non-singular. Thus β_o is fully identified as $\beta_o = [E(XX')]^{-1}\,[E(XY)]$ ■

Proof of Proposition 3.2.1 Analogous to 3.1.1, *mutatis mutandis*. ■

Proof of Proposition 4.1.1 From (iii), $E(XU_y \mid W) = E(X \mid W)\, E(U_y \mid W)$. Equivalently,

$$E([X - E(X \mid W)]\, U_y \mid W) = 0.$$

From (ii), $E(WW')$ is non-singular and $E(X \mid W) = E(XW')[E(WW')]^{-1} W$, so

$$E(\,[X - E(XW')[E(WW')]^{-1}W]\, U_y \mid W) = 0,$$

By the law of iterated expectations

$$E(\,[X - E(XW')[E(WW')]^{-1}W\,]\, U_y) = 0.$$

From (i), $U_y = Y - X'\beta_o$. Substituting this gives:

$$E([X - E(XW')[E(WW')]^{-1}W]\,[Y - X'\beta_o]\,) = 0$$

or

$$\{E(XX') - E(XW')[E(WW')]^{-1}\, E(WX')\}\;\beta_o = E(XY) - E(XW')[E(WW')]^{-1}E(WY).$$

By (iii), $\{E(XX') - E(XW')[E(WW')]^{-1} E(WX')\}$ is non-singular. Thus β_o is fully identified as

$$\beta_o = \{E(XX') - E(XW')[E(WW')]^{-1} E(WX')\}^{-1} \{E(XY) - E(XW')[E(WW')]^{-1}E(WY)\} \blacksquare$$

Proof of Proposition 4.1.2 From (iii)(a), $E(XU_z) = 0$. From (i), $\alpha_z U_z = Z - \gamma_z X$, and from (ii)(a) $E(XX')$ is non-singular. Proposition 3.1.1 thus ensures that γ_z' is fully identified as $\gamma_z' = [E(XX')]^{-1}E(XZ')$. Similarly, from (iii)(b), $E(ZU_y \mid X) = E(Z \mid X) \times E(U_y \mid X)$; from (i), $U_y = Y - Z'\delta_o$; and from (ii)(b), $\{E(ZZ') - E(ZX')[E(XX')]^{-1} E(XZ')\}$ is non-singular. Since we also have $E(Z \mid X) = E(ZX')[E(XX')]^{-1}X$, δ_o is fully identified by Proposition 4.1.1 as $\delta_o = \{E(ZZ') - E(ZX')[E(XX')]^{-1}E(XZ')\}^{-1} \times \{E(ZY) - E(ZX')[E(XX')]^{-1}E(XY)\}$. Since $\beta_o \equiv \gamma_z'\delta_o$, β_o is thus fully identified as:

$$\beta_o = [E(XX')]^{-1}E(XZ') \times \{E(ZZ') - E(ZX')[E(XX')]^{-1}E(XZ')\}^{-1} \times \{E(ZY) - E(ZX')[E(XX')]^{-1}E(XY)\} \blacksquare$$

Proof of Proposition 4.2.1 Analogous to 4.1.1, *mutatis mutandis*. ■

Proof of Proposition 4.2.2 Analogous to 4.2.1, replacing Z with $\tilde{Z}$. ■

Proof of Proposition 4.2.3 Analogous to 4.1.2, *mutatis mutandis*. ■

Proof of Theorem 5.1: (a) From (i) and (ii),

$$E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})]U_y\}$$

$$= E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})]\,(Y - X'\beta_o)\}$$

$$= E(\tilde{Z}\,Y) - E(E(\tilde{Z} \mid \tilde{W})\,Y) - E\,(\tilde{Z}X')\,\beta_o + E(E(\tilde{Z} \mid \tilde{W})X')\,\beta_o$$

Since β_o is finite and $E(\tilde{Z}\,Y)$, $E(E(\tilde{Z} \mid \tilde{W})\,Y)$, $E\,(\tilde{Z}X')$, and $E(E(\tilde{Z} \mid \tilde{W})X')$ exist and are finite, it follows that $E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})]U_y\}$ exists and is finite.

(b) Consider the system of equations

$$E\{\,[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})]\,X'\}\,\beta = E\{\,[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})Y\,\} - E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})]U_y\}.$$

It is a standard result of linear algebra that this system admits a unique solution β^* if and only if $E\{\,[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})]\,X'\}$ is non-singular.

(c) The result follows immediately from (a).

(d) If stochastic and structural identification hold, we have that $E\{ [\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] X' \}$ is non-singular and

$$E\{ [\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] X' \} \beta_o = E\{ [\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] Y\}.$$

It follows that β_o is then fully identified as

$$\beta_o = [E\{ [\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] X' \}]^{-1} E([\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] Y).$$

To establish the converse, suppose that either stochastic or structural identification fails. If stochastic identification fails, then the inverse of $E\{ [\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] X' \}$ does not exist, so β_o cannot have the form given above. If structural identification fails, then $E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})]U_y\}$ is not zero. By (a), β_o satisfies

$$E\{[\tilde{Z} - E(\tilde{Z} \mid \tilde{W})]U_y\} = E([\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] Y) - [E\{ [\tilde{Z} - E(\tilde{Z} \mid \tilde{W})] X' \}] \beta_o.$$

But this is incompatible with the expression above, and the result follows. ■

Proof of Corollary 5.2 Immediate.■

Proof of Theorem 7.1 The proof follows that of theorem 4.26 of White (2001). ■

Proof of Theorem 7.2 The proof of the first result follows that of theorem 4.26 of White (2001). The second result follows from theorem 4.39(i) of White (2001). ■

Appendix B: Failures of Identification from a Causal Perspective

Here we examine how structural identification of β_o via the XI method fails in the standard "irrelevant instrument," "invalid instrument," and "under-identified" cases.

B.1 Irrelevant Exogenous Instruments

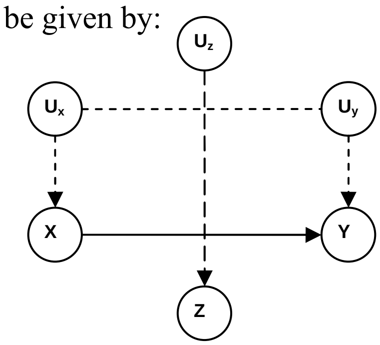
Proper instruments Z must be both valid and relevant to ensure structural identification. System S_{16} and its causal graph G_{16} depict the irrelevant XI case and demonstrate how an irrelevant XI satisfies neither CP:OXI nor CP:PXI. Let S_{16} be given by:

(1) $Z \overset{c}{=} \alpha_z U_z$

(2) $X \overset{c}{=} \alpha_x U_x$

(3) $Y \overset{c}{=} X'\beta_o + U_y$

where $U_x \perp U_y$, $U_x \perp U_z$ and $U_y \perp U_z$.



Graph 16(G_{16})
Irrelevant Exogenous Instruments

Although Z is valid and satisfies XI, it fails to identify β_o, because the effect of X on Y cannot be represented as the ratio of the effect of Z (resp. U_z) on Y and the effect of Z (resp. U_z) on X – both these effects are zero. In S_{16}, neither CP:OXI(i) nor CP:PXI(i) hold, since neither Z nor U_z cause X, justifying the label "irrelevant exogenous variables." When $\ell = k$ (as assumed here), the presence of irrelevant exogenous variables causes stochastic identification (condition (ii)) to fail in Proposition 3.2.1.

B.2 Endogenous Instruments

We next examine the failure of XI, condition (iii) of Proposition 3.2.1. In this case $Z \not\perp U_y$; such Z are endogenous. This can occur in several ways.

First, a potential instrument Z can be both irrelevant and endogenous. An example is a Z such that Z doesn't cause X and $U_z \perp U_x$, but both U_x and U_z cause U_y. Turning to relevant instruments, consider the system S_{17}:

$$(1)\ Z \overset{c}{=} \alpha_z U_z$$

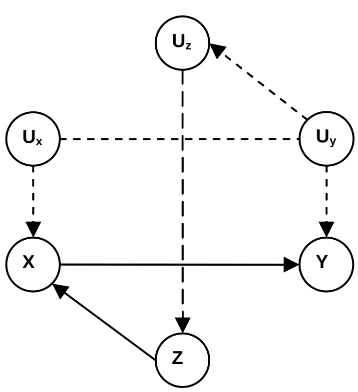
$$(2)\ X \overset{c}{=} \gamma_x Z + \alpha_x U_x$$

$$(3)\ Y \overset{c}{=} X'\beta_o + U_y$$

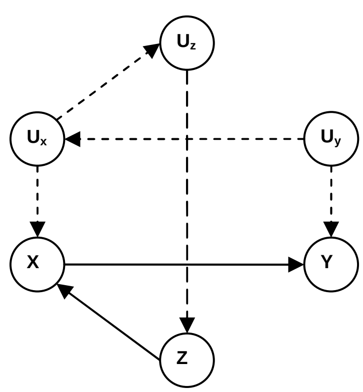
where $U_x \not\perp U_y$, $U_x \not\perp U_z$, and $U_y \not\perp U_z$. Substituting *(2)* into *(3)* with $\pi_o \equiv \gamma_x{}'\,\beta_o$ gives

$$(3')\ Y \overset{c}{=} Z'\pi_o + U_x{}'\alpha_x{}'\,\beta_o + U_y.$$

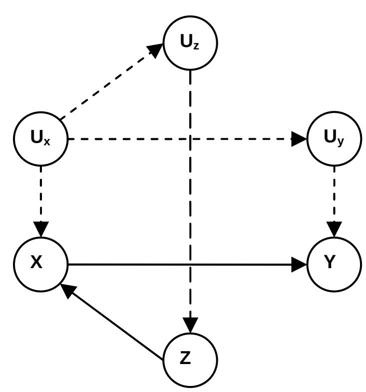
Because $U_y \not\perp U_z$, XI fails for Z. This can occur in several ways. For example, correlation between U_z and U_y can arise because either U_y causes U_z (G_{17a}, G_{17b}) or U_x causes both U_z and U_y (G_{17c}). CP:OXI(ii) fails, as Z and Y are confounded; and CP:PXI(ii) fails, as U_z and Y are confounded. CP:PXI(i) also fails, as U_z and X are also confounded.



Graph 17*a* (G_{17a})
Endogenous Instruments

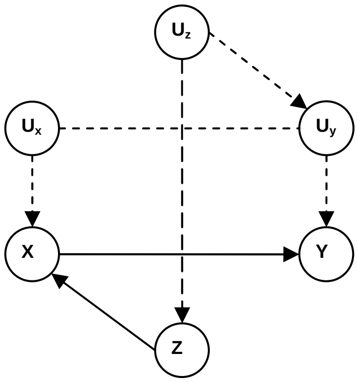


Graph 17*b* (G_{17b})
Endogenous Instruments

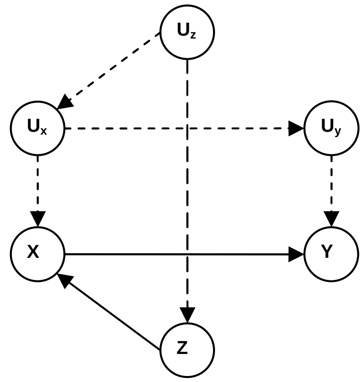


Graph 17*c* (G_{17c})
Endogenous Instruments

Alternatively, Z is endogenous when U_z affects U_y via a channel other than X. As we assume Z can't cause U_y, we need consider only the case where U_z causes Y via an intermediate cause other than X (G_{17d} and G_{17e}). Now CP:OXI(ii) fails as Z and Y are confounded. CP:PXI(iii) fails as U_z causes Y via an intermediate cause other than X; the effect of X on Y is thus no longer the ratio of the effect of U_z on Y and that of U_z on X.



Graph 17*d* (G_{17d})
Endogenous Instruments



Graph 17*e* (G_{17e})
Endogenous Instruments

Proposition 3.2.1 fails because structural identification fails. From *(3)* we have

$$E(ZY) = E(ZX') \beta_o + E(ZU_y),$$

but $E(ZU_y)$ does not vanish..

B.3 Under-Identified Exogenous Instruments

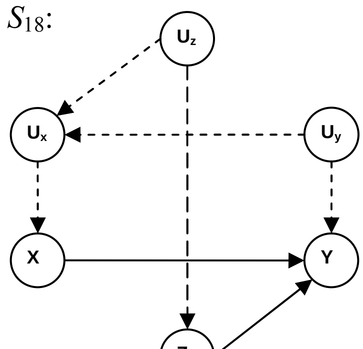
Finally, consider what happens when instruments Z are valid and relevant, but condition (i) of Proposition 3.2.1 fails. Specifically, consider the system S_{18}:

(1) $Z \stackrel{c}{=} \alpha_z U_z$

(2) $X \stackrel{c}{=} \alpha_x U_x$

(3) $Y \stackrel{c}{=} X' \beta_o + Z' \gamma_o + U_y$

where $U_x \not\perp U_y$, $U_x \not\perp U_z$, and $U_y \perp U_z$.



Graph 18 (G_{18})
Under-identified XI

In this case, the regressors X are endogenous, as $U_x \not\perp U_y$, but we have that Z is relevant since $X \not\perp Z$, and valid since $Z \perp U_y$. It follows from *(3)*, however, that

$$[E(ZX')]^{-1} E(ZY) = [E(ZX')]^{-1} E[Z(X'\beta_o + Z'\gamma_o + U_y)]$$

$$= \beta_o + [E(ZX')]^{-1}E(ZZ')\, \gamma_o,$$

Once again, structural identification of β_o fails, this time due to the presence of the unknown (non-zero) γ_o. The problem is that Z determines Y directly, and not solely via X. This violates CP:OXI(iii) and CP:PXI(iv).

Viewed in this way, the lack of structural identification appears as a form of "omitted variables bias," resulting from the failure to include Z in the instrumental variables regression. But one cannot resolve this problem by including Z, as then one is attempting to identify both β_o and γ_o, and there are not enough proper instruments for this. This is the classical "under-identified" case in which there are more right-hand side variables than valid instruments. Condition (ii) of Proposition 3.2.1 fails for the IV regression that includes both X and Z as regressors and uses only Z as instruments.

References

Angrist, J. (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from the Social Security Administrative Records," *American Economic Review*, 80, 313-336.

Angrist, J., G. Imbens, and D. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables" (with Discussion), *Journal of the American Statistical Association*, 91(434), 444-455.

Angrist, J. and A. Krueger (1999), "Empirical Strategies in Labor Economics," in *The Handbook of Labor Economics*, Vol. 3A, O. Ashenfelter and D. Card (eds.), Amsterdam: Elsevier Science.

Angrist, J. and A. Krueger (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *The Journal of Economic Perspectives*, 15(4), 69-85.

Bang-Jensen, J. and G. Gutin (2001). *Digraphs: Theory, Algorithms and Applications*. London: Springer Verlag.

Barnow, B., G. Cain, and A. Goldberger (1980), "Issues in the Analysis of Selectivity Bias," in E. Stromsdorfer and G. Farkas (eds.), *Evaluation Studies*, Vol. 5, San Francisco: Sage, 43-59.

Butcher, K. and A. Case (1994), “The Effects of Sibling Sex Composition on Women’s Education and Earnings,” *The Quarterly Journal of Economics*, 109, 531-563.

Chalak, K. and H. White (2006), “Independence and Conditional Independence in Causal Systems,” UCSD Department of Economics Discussion Paper.

Cartwright, N. (1989). *Nature’s Capacities and their Measurement*. Oxford: Clarendon.

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

Dawid, A.P. (2000), “Causal Inference without Counterfactuals” (with Discussion), *Journal of the American Statistical Association*, 95, 407-448.

Dawid, A.P. (2002), “Influence Diagrams for Causal Modeling and Inference,” *International Statistical Review*, 70, 161-189.

Fisher, F. (1966). *The Identification Problem in Econometrics*. New York: McGraw-Hill.

Fisher, R.A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.

Frisch, R. and F. Waugh (1933), “Partial Regressions as Compared with Individual Trends,” *Econometrica*, 1(4), 939-953.

Goldberger, A. (1972), “Structural Equation Methods in the Social Sciences,” *Econometrica*, 40(6), 979-1001.

Goldberger, A. (1991). *A Course in Econometrics*. Cambridge: Harvard University Press.

Gordon, N. and E. Vegas (2005), “Educational Finance Equalization, Spending, Teacher Quality and Student Outcomes: The Case of Brazil’s FUNDEF,” in E. Vegas (ed.), *Incentives to Improve Teaching: Lessons from Latin America*. Washington, DC: The World Bank, 2005.

Haavelmo, T. (1943), “The Statistical Implications of a System of Simultaneous Equations,” *Econometrica*, 11(1), 1-12.

Haavelmo, T. (1944), “The Probability Approach in Econometrics,” *Econometrica*, 12 (Supplement), iii-vi and 1-115.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hahn, J. (1998), “On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effect,” *Econometrica*, 66(2), 315-331.

Hausman, J. A. and W. E. Taylor (1983), “Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation,” *Econometrica*, 51(5), 1527-1550.

Hayashi, F. (2000). *Econometrics*. Princeton: Princeton University Press.

Heckman, J. (1996), “Comment on ‘Identification of Causal Effects Using Instrumental Variables’ by Angrist, J., G. Imbens, and D. Rubin,” *Journal of the American Statistical Association*, 91, 459-462.

Heckman, J. (1997), “Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations,” *Journal of Human Resources*, 32, 441-462.

Heckman, J. (2000), “Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective,” *Quarterly Journal of Economics*, February 2000, 115, 45-97.

Heckman, J. (2006), “The Scientific Model of Causality,” *Sociological Methodology* (forthcoming).

Heckman, J. and R. Robb (1985), “Alternative Methods for Evaluating the Impact of Interventions,” in J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*. Cambridge: Cambridge University Press, 146-245.

Heckman, J., H. Ichimura, and P. Todd (1998), “Matching as an Econometric Evaluation Estimator,” *The Review of Economic Studies*, 65, 261-294.

Heckman, J., R. LaLonde, and J. Smith (1999), “The Economics and Econometrics of Active Labor Market Programs,” *Handbook of Labor Economics*, Vol. 3A, O. Ashenfelter and D. Card (eds.), Amsterdam: North Holland, 1865-2097.

Heckman, J., S. Urzua, and E. Vytlacil (2005), “Understanding Instrumental Variables in Models with Essential Heterogeneity,” *Review of Economics and Statistics*, (forthcoming).

Heckman, J. and E. Vytlacil (2005), “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, 73(3), 669-738.

Hirano, K. and G. Imbens (2004), "The Propensity Score with Continuous Treatments," in A. Gelman and X.-L. Meng (eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. New York: Wiley.

Hirano, K., G. Imbens, and G. Ridder (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71(4), 1161-1189.

Holland, P.W. (1986), "Statistics and Causal inference" (with Discussion), *Journal of the American Statistical Association*, 81, 945-970.

Hoover, K.D. (2001). *Causality in Macroeconomics*. Cambridge University Press.

Imbens, G.W. and W. K. Newey (2003), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," manuscript.

Matzkin, R. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339-1375.

Matzkin, R. (2004), "Unobservable Instruments," Northwestern University Department of Economics Working Paper.

Matzkin, R. (2005), "Identification of Nonparametric Simultaneous Equations," Northwestern University Department of Economics Working Paper.

Morgan, M.S. (1990). *The History of Econometric Ideas*. Cambridge: Cambridge University Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufman.

Pearl, J. (1993a), "Aspects of Graphical Methods Connected with Causality," in *Proceedings of the 49th Session of the International Statistical Institute*, pp. 391-401.

Pearl, J. (1993b), "Comment: Graphical Models, Causality, and Intervention," *Statistical Science*, 8, 266-269.

Pearl, J. (1995), "Causal Diagrams for Empirical Research" (with Discussion), *Biometrika*, 82, 669-710.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Reiersøl, O. (1945), "Confluence Analysis by Means of Instrumental Sets of Variables," *Akiv för Matematik*, *Astronomi och Fysik*, 32a, 1-119.

Roy, A. D. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers* (New Series), 3, 135-146.

Rosenbaum, P. R. (2002). *Observational Studies*. 2nd edition, Berlin: Springer-Verlag.

Rosenbaum, P. R. and D. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

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.

Simon, H. (1953), "Causal Ordering and Identifiability," In *Studies in Econometric Method*, W. C. Hood and T. C. Koopmans (eds.), Cowles Commission Monograph no. 14. New York: Wiley, pp. 49–74.

Simon, H. (1954), "Spurious Correlation: A Causal Interpretation," *Journal of the American Statistical Association,* 49, 467-479.

Spirtes, P., C. Glymour, and R. Scheines (1993). *Causation, Prediction and Search*. Berlin: Springer-Verlag.

Stock, J. and Francesco Trebbi (2003), "Who Invented Instrumental Variable Regression?" *Journal of Economic Perspectives*, 17, 177-194.

Strotz, R. and H. Wold (1960), "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis," *Econometrica*, 28(2), 417-427.

White, H. (2001). *Asymptotic Theory for Econometricians*. New York: Academic Press.

White, H. (2006), "Time Series Estimation of the Effects of Natural Experiments," *Journal of Econometrics* (in press).

White, H. and K. Chalak (2006a), "A Unified Framework for Defining and Identifying Causal Effects," UCSD Dept. of Economics Discussion Paper.

White, H. and K. Chalak (2006b), "Parametric and Nonparametric Estimation of Covariate-Conditioned Average Effects," UCSD Dept. of Economics Discussion Paper.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Wright, P. G. (1928). *The Tariff on Animal and Vegetable Oil*. New York: Macmillan.

Wright, S. (1921), "Correlation and Causation," *Journal of Agricultural Research*, 20, 557-85.

Wright, S. (1923), "The Theory of Path Coefficients: A Reply to Niles' Criticism," *Genetics*, 8, 239-255.