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

Karim Chalak and Halbert White
April 2006
First Draft: March, 2005

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.

DRAFT
DO NOT CITE WITHOUT PERMISSION

---

1 The authors thank Kate Antonovics, Julian Betts, Graham Elliott, Marjorie Flavin, Clive Granger, Keisuke Hirano, Stephen Lauritzen, Mark Machina, Dimitris Politis, Ross Starr, Ruth Williams and the participants of the UCSD Applied Lunch seminar, the 4th annual Advances in Econometrics conference, and the UCSD Econometrics seminar. 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

The structural equations framework is predominant in economics for the purposes of modeling, identifying, and estimating causal effects of interest. The early work of the Cowles Commission by Tinbergen, Frisch, Koopmans, Haavelmo, Marshack, Simon, Wold, Strotz, and others studied the 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.” For the most part, standard textbooks currently define these concepts to mean respectively the correlation or lack thereof between a structural equation’s observed explanatory variables and its unobserved variables. 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”, originally introduced by Philip Wright (1928) building on Sewall Wright’s (1921, 1923) work on “path analysis”, within the structural equations framework. Ever since, the method of instrumental variables has played a central role in econometrics to overcome the problem of endogeneity (e.g. Heckman, 1997; Angrist and Krueger, 2001; Heckman, Urzua, and Vytlacil, 2005.) The definition may vary somewhat depending on the context, but in the familiar context of a linear structural equations system, instrumental variables are defined as variables that are uncorrelated with the equation’s unobserved variables, i.e. “valid”, and correlated with the included explanatory variables, i.e “relevant.”

Recently, advances across a variety of disciplines have resulted in alternative frameworks and methods to identify and estimate causal effects from observational studies 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 put forward 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” as well as the notions of “ignorability” and “propensity score” for measuring causal effects (e.g., Rosenbaum and Rubin, 1983; Holland, 1986.) Angrist, Imbens, and Rubin (1996) have related this framework to the method of instrumental variables.

Another line of research into the identification of causal effects has emerged in the machine learning literature in the work of Pearl (1988, 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 the “front door” methods. A distinctive feature of this literature is the use of “directed acyclic graphs” (DAGs) to represent causal relationships and the use of graphical criteria to determine if particular causal effects are identifiable without particular attention to the estimation of these causal effects.

White (2006) and White and Chalak (2006a) propose the “settable system” framework as a means of unifying these distinct approaches to the study of causality. In those papers, particular attention is paid to the identification and estimation of causal effects in a setting most analogous to the classical case of exogenous regressors. Here we broaden our focus to apply this framework to analyze identification and estimation of causal effects in the presence of endogenous regressors generally. Consistent with Dawid (1969, 2000), we show that all of the methods that emerge, including those above, require one or more independence or conditional independence relationships to hold between observed variables and corresponding unobserved variables of the system under study.

Specifically, the contribution of this paper is to provide a novel and detailed examination of the ways in which causal structures can give rise to observed variables other than the cause or treatment of interest that can play an instrumental role in permitting the identification and estimation of causal effects of interest. In this sense, this paper extends the standard notion of instrumental variables to accommodate variables that are not necessarily uncorrelated with the unobserved causes of a response variable of interest but that can nevertheless be instrumental in permitting the recovery of useful estimates of causal effects of interest.

Consider, for example, the following simple structural equations 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, $\alpha_0$ is an unknown real vector, and $\gamma_0$ and $\delta_0$ are unknown real scalars such that:

\begin{align*}
(1) \quad X^c &= U_x^c \alpha_0 \\
(2) \quad Z^c &= \gamma_0 X + U_z \\
(3) \quad Y^c &= \delta_0 Z + U_y,
\end{align*}

where we use $\overset{c}{=}$ instead of the usual equality sign to emphasize the causal nature of the structural equations. Following Dawid (1979), we use $\perp$ to denote independence between two random variables or vectors and $\perp$ to denote otherwise. For this example, we assume that $U_x \perp U_y$, $U_x \perp U_z$, and $U_y \perp U_z$.

Consider measuring the total causal effect of $X$ on $Y$. Substituting structural equation (2) into structural equation (3) we get:

\begin{align*}
(3') \quad Y^c &= \beta_0 X + \delta_0 U_z + U_y,
\end{align*}
where $\beta_0 \equiv \gamma_0 \delta_0$ is the causal effect of $X$ on $Y$.

Clearly, since $U_x \perp U_y$, the least squares estimator for $\beta_0$, say $\hat{\beta}$, is inconsistent, as $X$ is endogenous in the standard sense. Further, $Z$ is an “invalid” instrumental variable, in the standard sense, as it is correlated with the unobserved term in $(3')$: from (2) we have that $Z$ is correlated with $U_z$. Also, since $X$ causes $Z$ from (2) and $X$ and $U_y$ are correlated, we have that $Z$ is correlated with $U_y$.

This system presents a situation where it seems 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:

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

where $X$, $Y$, and $Z$ each denote $n \times 1$ data vectors.

This structural equation system permits the use of Pearl’s (1995) “front-door” method, providing an example of a situation where a variable $Z$ is instrumental in the recovery of the causal effect of $X$ on $Y$ even though it is an “invalid” instrument as currently defined in the literature. The challenge here is not that $\beta_0$ is unidentified, but that it is not identified by exogenous regressors or standard exogenous instruments. Nor is this example the only possibility of this sort.

In a nutshell, the standard method of instrumental variables is not the whole story. There are other extended instrumental variables (EIV) methods that we can employ to identify and estimate causal effects of interest in the endogenous regressor case. These methods are characterized by alternative moment conditions and exclusion restrictions that parallel those in the standard instrumental variable case. A main goal of this paper is to begin a systematic exploration of these methods and their interrelations.

This paper is organized as follows. In Section 2, we state our assumptions and discuss the data generating structural equations systems of interest to us here. In Section 3, we employ the framework of Section 2 to provide a fully explicit causal interpretation of standard regression and IV methods, extending previous work and setting the stage for subsequent developments. Section 3.1 examines the case of exogenous regressors (XR) where regressors $X$ act as their own instruments to identify the causal effects of interest. In Section 3.2, we study causal identification via standard exogenous instruments (XI) $Z$. There, we examine the standard “validity” and “relevancy” conditions, and we employ our framework to relax conditions presented in previous studies, such as that of Angrist, Imbens, and Rubin (1996), supporting a causal interpretation of standard instrumental variables methods. Section 3.3 shows how causal identification breaks down in situations where standard IV methods fail.

Section 4 begins our study of EIV methods, where the use of conditional extended instrumental variables $Z$, conditioning extended instrumental variables $W$, or both together permit the identification of the causal effect of a potentially endogenous $X$ on the
response of interest $Y$. In Section 4.1 we discuss single EIV methods, that is, methods that use either conditioning EIV or conditional EIV but not both to identify causal effects. Section 4.1.1 defines and discusses conditioning instruments and the method of conditionally exogenous regressors given conditioning instruments (CXRI), relating these to the method of matching, Rosenbaum and Rubin’s (1983) ignorability condition, Pearl’s (1995) back door method, and White’s (2006) predictive proxies. In Section 4.1.2, we examine conditional instruments and the method of conditionally exogenous instruments given regressors (CXIR). We relate these to the standard method of instrumental variables and to Pearl’s (1995) front door method.

In Section 4.2, we discuss special cases of double EIV methods where the joint use of conditional and conditioning EIV is needed for the identification of effects of interest. In particular, we discuss the methods of conditionally exogenous instruments given conditioning instruments (CXII), conditionally exogenous instruments and regressors given conditioning instruments (CXRIII), and conditionally exogenous instruments given regressors and conditioning instruments (CXIIIR).

Section 5 provides a “master theorem” that contains the various EIV identification results as special cases or delivers them as immediate corollaries, stating both necessary and sufficient conditions for identification of causal effects.

Section 6 discusses how causal matrices can be used to characterize the cases where the identification of causal effects via EIV methods obtains. We illustrate by showing that our single EIV methods exhaust the possibilities for identification of causal effects using a single EIV. Section 7 presents straightforward conditions that ensure the consistency and asymptotic normality of the extended instrumental variables estimators considered here. Section 8 concludes, with final remarks and a discussion of directions for future research. Proofs of formal results are gathered into the Mathematical Appendix.

2. Causal Data Generating Systems

Economists and econometricians interested in measuring causal effects 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 that “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 the linear regressions that they perceive as convenient carriers of predictive relationships in the form of conditional correlations between variables (see, e.g., Angrist and Krueger, 1999; Heckman, LaLonde, and Smith, 1999; Heckman, 2000; and Hoover, 2001.)

We employ a familiar structural equations system to represent a causal structure, $S$. In particular, we consider data generated as a special case of the recursive system
\[
\begin{align*}
X_1^c &= r_1 (X_0) \\
X_2^c &= r_2 (X_1, X_0) \\
&\quad \cdots \\
X_J^c &= r_J (X_{J-1}, \ldots, X_1, X_0),
\end{align*}
\]

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.

In writing this system, we use the notation \(\cdot^c\) to emphasize that the \(J\) structural equations appearing in the system \(S\) are neither equations nor regressions. Instead, they represent “causal links” (Goldberger, 1972, p.979) that embody directionality from cause to effect. In particular, the right hand side variables of every structural equation determine the value of the corresponding left hand side variable but the converse is not necessarily true. The structural equations are thus directional “autonomous” mechanisms describing how every variable in the model is generated (Haavelmo 1943, 1944; Strotz and Wold, 1960; Pearl, 2000; White and Chalak, 2006a.) Conceptually, these mechanisms can be independently manipulated without necessarily modifying any of the other generating structural equations in the system. The autonomy of these equations is a vital requirement, for it enables the researcher to evaluate causal relationships by means of hypothetical interventions where \(X_j\) is set to some different value, \(X_j^*\). Strotz and Wold (1960) describe such a manipulation as “wiping out” the structural equation that generates \(X_j\), thus suspending the mechanism by which \(X_j\) was naturally generated, and setting \(X_j\) to \(X_j^*\), whenever it appears as a right hand side variable in the other structural equations of the system, thus enabling the manipulation to manifest its effect on the rest of the variables in the system. White and Chalak (2006a) provide a rigorous formalization of this notion.

Observe that the vector \(X_0\) does not appear on the left hand side of any causal relation in this system. If the system provides a complete description of the causal relations of the structure, then \(X_0\) is not caused by any of the other variables of the system. Following White and Chalak (2006a), we refer to such variables as fundamental.

Our first formal assumption modifies the notation above somewhat to accommodate the structures of interest to us here.

**Assumption A.1(a): Data Generating Structural Equations System:** 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:

\[
\begin{align*}
X_1^c &= r_1 (U_1) \\
X_2^c &= r_2 (X_1, U_2) \\
&\quad \cdots \\
X_J^c &= r_J (X_{J-1}, \ldots, X_1, X_0),
\end{align*}
\]
The \( U_j \)'s may have differing dimensions. We collect them together into the random vector \( X_0 \equiv (U_1', \ldots, U_J')' \). The data are thus generated by the system \( S \equiv (X_0, r_1, \ldots, r_J) \) associated with the above collection of \( J \) structural equations.

In A.1(a) we do not specify that \( X_0 \) is fundamental, implying that A.1(a) does not provide a complete specification of the causal structure. This provides flexibility in handling dependence, and in particular causal relationships, that may hold among the \( U_j \)'s. For the moment, we defer specifying these.

For notational convenience, we write \( U \equiv X_0 \) and \( X \equiv (X_1, \ldots, X_J) \), and we let \( V \equiv (X, U) \equiv (V_1, \ldots, V_G) \) denote the vector of all observed and unobserved variables in the system. To maintain a tight focus for our analysis, we study the identification and estimation of causal effects in the structural equation system given in A.1(a) for the rest of this paper, leaving the task of suggesting and testing for causal models for other work (in progress).

We formally view the structure of A.1(a) as a settable system as defined by White and Chalak (2006a), so that references here to notions of setting, cause, and effect are as formally defined there. In particular, we view settings as the means by which manipulations of the mechanism that naturally generates a variable can be effected (Holland, 1986; Spirtes, Glymour, and Scheines, 1993; Pearl, 2000; and White and Chalak, 2006a.) Given A.1(a), the following working definition of causality suffices.

Specifically, \( X_j \) does not cause \( X_k \) when \( j \geq k \) (including \( k = 0 \)), whereas \( X_j \) may cause \( X_k \) when \( j < k \). For \( j < k \), we say that \( X_j \) does not cause \( X_k \) relative to \( S \) if \( r_k (X_{k-1}, \ldots, X_1, U_k) \) defines a function constant in \( X_j \). In this case, we say that the response function \( r_k \) depends trivially on \( X_j \). Otherwise, we say that \( X_j \) causes \( X_k \) relative to \( S \). This notion corresponds to “direct” or “immediate” causality as defined by Pearl (2000.) As we have written this system, we view \( U_k \) as a cause of \( X_k \), whereas \( U_j \) does not cause \( X_k \) for \( j \neq k \).

The object of interest in this paper is the average total causal effect of an observable\(^2\) \( X_j \) on another observable \( X_k \). Thus, we are interested in the full effect of \( X_j \) on the dependent variable \( X_k \), channeled via all routes in the system and averaged over the unobserved causes \( U_k \) of \( X_k \). This is in contrast to the direct or immediate average effect on a variable where the effects due to intermediate causes are not accounted for. White and Chalak (2006a) discuss the identification of causal effects more generally, including covariate-conditioned effects on other features of the distribution of the response such as the variance and the quantiles.

---

\(^2\) As we implicitly rely on White and Chalak’s (2006a) 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 to be a convenient shorthand.
The unobserved causes \( U \) are included in the structural equations to accommodate either the unobservability of known variables key to determining the target variable, the researcher’s ignorance of the full causal mechanism that generates the target, or both. We permit dependence among the elements of \( U \) to accommodate dependence between the observed and unobserved variables of the same structural equation, resulting in endogeneity. This dependence may arise from causal relations among the elements of \( U \).

In the structure provided by A.1(a), all variables have a causal status. For simplicity, we have not introduced attributes, that is, non-causal response modifiers (see White and Chalak, 2006a). A key role played by attributes is to introduce heterogeneity into the structural 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 heterogeneity, but we refrain from doing so here in order to maintain a sharp focus for our analysis.

For what follows we make use of 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 of system \( S \) 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 \( g = 1, \ldots, G \).

For example, when Assumption A.1(a) holds and the unobservables are scalar, \( C_S \) has the following form:

\[
C_S = \begin{bmatrix}
C_{S_1} & C_{S_2} \\
C_{S_3} & C_{S_4}
\end{bmatrix}
= \begin{bmatrix}
X_1 & \ldots & X_J & U_1 & \ldots & U_J \\
X_1 & 0 & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
X_J & 0 & \ldots & 0 & 0 & 0 \\
U_1 & 1 & 0 & \ldots & 0 & \ldots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
U_J & 0 & \ldots & 0 & 1 & 0
\end{bmatrix}
\]

The triangularity of the system assumed in A.1(a) ensures that \( C_{S_1} \) is upper triangular with zeros along the diagonal. Blank entries in \( C_S \) above indicate elements that can take either the values 0 or 1, reflecting the fact 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 can cause any of the \( U \)’s. Thus \( C_{S_2} \) is the \( J \times J \) zero matrix. We also have that \( U_k \) does not cause \( X_j \) for \( j \neq k \). As a result, \( C_{S_3} \) is the \( J \times J \) identity matrix. We do not rule out the possibility that \( U_k \) can cause \( U_j \). This leaves the elements of \( C_{S_4} \) unspecified, apart from the zero diagonal. Consequently, \( C_{S_1} \) and \( C_{S_4} \) determine \( C_S \).
We note that there corresponds a unique structural equations system $S$ given by A.1(a) for a given causal matrix $C_S$ but that the converse is not true. This is so because the causal matrix explicitly specifies all causal relationships, including those holding among the unobserved variables, but these are unspecified in the structural equations system $S$ of A.1(a). It follows that different causal matrices can potentially generate the same statistical dependence relationships among the unobserved variables in $S$. We discuss this further below.

In addition to the recursive structure among the observed variables imposed in A.1(a), we assume the following:

**Assumption A.1(b): Acyclicality:** For each $h \leq G$ and each set of $h$ distinct elements, say $\{g_1, \ldots, g_h\}$, of $\{1, \ldots, G\}$, we have:

$$c_{g_1g_2} \times c_{g_2g_3} \cdots \times c_{g_hg_1} = 0.$$ 

The recursive 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 particular recursive structure for simplicity; the general settable system framework does not require this.

The acyclicality imposed in A.1(b) ensures that there exists at least one fundamental variable among the variables of the system, a consequence of proposition 1.4.2 of Bang-Jensen and Gutin (2001.) 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 $U$ is. For ease of reference, we denote the vector of fundamental variables $U_0$. This may contain some or all of the elements of the $U_j$’s. We could alternatively specify an additional vector of unobserved fundamental variables $U_0$ to which the other $U_j$’s may be causally related. We forgo this, however, to avoid elaborating our structure beyond what is strictly necessary to deliver our desired results. Note that whenever a variable is fundamental, its corresponding causal matrix column contains a vector of zeroes.

A convenient device for representing causal relations in simple situations that we employ repeatedly below is the “directed causal graph.” For each causal matrix $C_S$, there is a corresponding causal graph $G_S$. These are variants 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). In that literature, the unobserved components of the system are typically not explicitly represented. In contrast, we explicitly represent these due to the central role that they play in econometrics.

The graph $G_S$ consists of a set of nodes (also referred to as vertices), one for each element of $V$, and a set of arrows $A$, corresponding to ordered pairs of distinct vertices. An arrow $a_{gh}$ denotes that variable $V_g$ is an immediate cause of $V_h$. We use solid arrows to denote direct causal relationships between variables with observed realizations. Thus, a solid
arrow from $X_j$ to $X_k$ denotes that $X_j$ is an immediate cause of $X_k$. For variables with unobserved realizations, we use a dashed arrow from $U_j$ to $U_k$ to denote that $U_j$ causes $U_k$. We also use a dashed arrow to denote that $U_j$ is a cause of an observed variable $X_j$. As a convenient shorthand, we use a dashed line connecting $U_j$ and $U_k$ to indicate that either $U_j$ is an immediate cause of $U_k$ or that $U_k$ is an immediate cause of $U_j$, or to indicate that there is an unobserved cause (e.g., $U_0$) that causes both $U_j$ and $U_k$. (In the latter case we omit depiction of the unobserved common cause.) The lack of dashed arrows or dashed lines between the unobserved variables $U_j$ and $U_k$ indicates that $U_j \perp U_k$ as discussed further below. The convention that variables do not cause themselves corresponds to the absence of self-directed arrows in the causal graph $G_S$.

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:

$$
\begin{align*}
X_1^c &= U_1' \alpha_1 \\
X_2^c &= \beta_{2,1}X_1 + U_2' \alpha_2 \\
\vdots \\
X_J^c &= \beta_{J,J-1}X_{J-1} + \ldots + \beta_{J,1}X_1 + U_J' \alpha_J,
\end{align*}
$$

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

The compelling motivation for imposing the strong structure of linearity and separability is to provide clear insight into the nature of EIV methods in a simple and familiar context, permitting us to make our main points without being distracted by further complications that otherwise arise. Nevertheless, the key insights of this paper extend to the modern “nonparametric” (more accurately, non-separable) setting in which A.2 is replaced by much milder conditions (see, e.g., Matzkin, 2003, 2004, and 2005; Imbens and Newey, 2003; White and Chalak, 2006a). We take this up elsewhere.

In what follows, we specify that certain independence or conditional independence conditions hold between the variables of interest. 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 $f_{UX}(u \mid x) = f_U(u)$ (using the obvious shorthand notation for density or conditional density functions), conditional independence $X \perp U \mid W$ entails $f_{UXW}(u \mid x, w) = f_{UW}(u \mid w)$. Given the linear separable structure assumed in A.2, these assumptions are stronger than is strictly necessary to obtain identification results. Weaker conditions can suffice given A.2, such as suitable conditional mean independence ($E(U \mid X, W) = E(U \mid W)$) or conditional non-correlation ($E(X U \mid W) = E(X \mid W) E(U \mid W)$) assumptions. Nevertheless, we work primarily with independence or conditional independence, first for simplicity
and second because these conditions are required for identification of causal effects in general, such as for the non-separable case or when interest attaches to causal effects on aspects of the distribution of the response other than average effects, such as effects on the quantiles or distribution of the response (see White and Chalak, 2006a.)

Observe that conditional independence implies conditional mean independence and conditional non-correlation. As a convenient convention to accommodate the stronger than necessary independence or conditional independence assumptions adopted here, when we speak of dependence or conditional dependence, we may understand this to result from unconditional or conditional correlation, also implying unconditional or conditional 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. In placing the standard methods in this context, we will cover some very familiar ground. Nevertheless, by doing so we discover certain aspects of these familiar cases that have previously been overlooked, permitting us to extend previous work and to set the stage for subsequent developments.

Just as Goldberger (1991, p. 337) notes that there is no necessary causal structure embodied in the variables appearing in the standard regression model, there is also no necessary causal structure embodied in the standard treatments of instrumental variables. Although causal relationships were clearly of concern to the Cowles Commission pioneers, an explicit causal focus has disappeared from much of the subsequent literature on instrumental variables methods. For example, White (2001) treats instrumental variables estimation extensively, but nowhere is there any reference to causal structure. The statistical 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 about causal structure in the instrumental variables context are provided by the 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 the method of instrumental variables. Here we provide a causal account of instrumental variables methods complementary to and extending that of AIR. Our account is designed to accord with the philosophy literature on causality, which requires causal inputs in order to derive causal conclusions, as expressed in Cartwright’s (1989) dictum “no causes in, no causes out.” After all, why would we rely on statistical methods that allege to have identified causal effects without providing a causal account to support that claim?

To facilitate our analysis, we further elaborate our notation. We now let $Y$ denote the scalar response of interest, let the $k$ random variables $X_1, \ldots, X_k$ denote the observed causes of interest, and let the $\ell$ random variables $Z_1, \ldots, Z_\ell$ denote variables potentially instrumental to identifying the causal effects of interest, all with observed realizations specified in a manner consistent with A.1 and A.2. We put $X \equiv [X_1, \ldots, X_k]'$ and $Z \equiv [Z_1, \ldots, Z_\ell]'$. 
... $Z_l$. (The $X_i$'s of A.1 and A.2 now correspond to the elements of $X$, $Y$, and $Z$; this new notation helps us to keep track of the various roles played by the different variables of the system.) We denote by $U_y$, $U_{x_1}$, ..., $U_{x_k}$, and $U_{z_1}$, ..., $U_{z_l}$ the unobserved causes corresponding to the responses, causes, and instruments, respectively, and we write $U_x = [U_{x_1}', ..., U_{x_k}']'$, and $U_z = [U_{z_1}', ..., U_{z_l}']'$. We continue to denote the fundamental unobservables as $U_0$. We also let $X$, $Y$, and $Z$ denote $n \times k$, $n \times 1$, and $n \times \ell$ matrices containing $n$ identically distributed random observations on $X$, $Y$, and $Z$ respectively.

3.1 Exogenous Regressors

The first case we consider is that of exogenous regressors. Under our assumptions, this is the familiar case where simple regression identifies the effect of $X$ on $Y$. Consider structural equations system $S_1$ and its corresponding causal graph $G_1$:

\begin{align*}
(1) \hspace{1cm} X^c &= \alpha_x U_x \\
(2) \hspace{1cm} Y^c &= X' \beta_o + U_y,
\end{align*}

where $U_x \perp U_y$.

In (1) above, $\alpha_x$ is a matrix of unknown coefficients that maps the unobserved causes $U_x$ to the observed causes $X$. The coefficients $\beta_o$ have causal meaning by virtue of (2).

A main feature of $S_1$ is that $X$ and $Y$ do not share a common cause. This structure ensures the following unconditional independence relationship:

$$ (XR) \hspace{1cm} X \perp U_y $$

In conformity with standard nomenclature, we refer to $X$ as \textit{exogenous regressors}. Together, A.1 and XR ensure the key moment condition

$$ E(XU_y) = 0. \hspace{1cm} (M1) $$

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

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

This condition \textit{structurally identifies} the causal coefficients $\beta_o$ by relating them solely to moments of observable variables. When \textit{stochastic identification} also holds, that is, $E(XX')$ is non-singular, we have \textit{full identification} of $\beta_o$. In this case, we have
$$\beta_o = [E(XX')]^{-1}[E(XY)].$$

We formalize this as follows:

**Proposition 3.1.1** Suppose A.1 and A.2 hold such that: (i) \(Y^c = X'\beta_o + U_y\), where \(X\) is \(k \times 1\), \(k > 0\), \(\beta_o\) 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: \(X \perp U_y\) holds.

Then \(\beta_o\), the average total causal effect of \(X\) on \(Y\), is fully identified as:

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

Thus, Proposition 3.1.1 identifies the causal coefficient \(\beta_o\) with the statistical association between \(X\) and \(Y\), \([E(XX')]^{-1}[E(XY)]\). We refer the use of exogenous regressors to identify the causal effect \(\beta_o\) in this way as the XR method.

The plug-in estimator for \(\beta_o\) is the familiar OLS estimator for a simple linear regression of \(Y\) on \(X\), \(\hat{\beta}_o^{XR} = (X'X)^{-1}(X'Y)\). In Section 7, we state straightforward conditions ensuring that this estimator and the others we introduce are consistent and asymptotically normal for the causal effect \(\beta_o\).

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 we have that \(X\) and \(Y\) are correlated. We know that \(Y\) does not cause \(X\). The XR condition rules out the possibility that \(X\) and \(Y\) share a common cause. Given that \(Y = X'\beta_o + U_y\), it follows that the association between \(X\) and \(Y\) can only be explained as the effect of \(X\) on \(Y\).

When control over \(X\) is possible, XR can be ensured by randomization, for example. Nevertheless, in observational studies, where control is not possible, it is often hard to argue that XR holds. We now examine the case where XR fails from a causal standpoint.

Specifically, consider the structural equations system \(S_2\) and its corresponding causal graph \(G_2\):

1. \(X^c = \alpha_x U_x\)
2. \(Y^c = X'\beta_o + U_y\)

where \(U_x \perp U_y\).

\(\begin{array}{c}
U_x \\
\downarrow \\
X \\
\downarrow \\
Y \\
\downarrow \\
U_y
\end{array}\)

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 $\beta_o$ is no longer structurally identified. Instead, we have $E(XY) = E(XX') \beta_o + E(XU_y)$, in which unknown moments involving unobservables appear. We thus have the familiar result that regression fails to structurally identify $\beta_o$. In particular, the OLS estimator from a linear regression of $Y$ on $X$ is inconsistent for $\beta_o$.

In $G_2$, we cannot necessarily explain the association between $X$ and $Y$ as due to $X$ causing $Y$, as this could equally be due to the joint response of $X$ and $Y$ to $U_y$, to $U_x$, or to an unobserved common cause of $U_y$ and $U_x$, or to an unobserved common cause of $U_y$ and $U_x$. In accord with standard parlance, we refer to the failure of XR as regressor endogeneity and refer to $X$ as endogenous regressors when $X \perp U_y$. We also refer to failure of XR as confoundedness of causes. In $S_2$, either $U_y$ or $U_x$ (or $U_0$) is a confounding variable for $X$ and $Y$. Thus, under A.1, an endogenous regressor is one that shares at least one unobserved common cause with the response variable. Observe that simultaneity is absent from this system and is therefore not responsible for the endogeneity.

### 3.2 Exogenous Instruments

The presence of endogenous regressors means that $\beta_o$ cannot be identified by the method of exogenous regressors. Nevertheless, identification is possible when one has available a vector of “proper” instrumental variables, $Z$. The standard textbook definition is that variables $Z$ are “proper” instrumental variables if they are “valid,” i.e. uncorrelated with the “error term,” and “relevant,” i.e. correlated with the endogenous regressors (e.g., Hamilton, 1994 p.238; Hayashi, 2000 p.191; Wooldridge, 2002 p.83-84):

- (i) $Z$ is “valid” if and only if $\text{Corr}(Z, U_y) = 0$
- (ii) $Z$ is “relevant” if and only if $\text{Corr}(X, Z) \neq 0$

where $\text{Corr} (\cdot, \cdot)$ denotes correlation.

P.G. Wright (1928) first used instrumental variables, which he referred to as “curve shifters,” to identify supply and demand elasticities (see Morgan, 1990; Angrist and Krueger, 2001; Stock and Trebbi, 2003.) In describing these variables, 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 and not only about statistical correlation. The first thoughts on instrumental variables appear, as expected, 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 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 $G_3$:

\begin{align*}
(1) \ Y &= \alpha_z U_z \\
(2) \ X &= \gamma_x Z + \alpha_x U_x \\
(3) \ Y &= X' \beta_o + U_y
\end{align*}

where $\gamma_x$ is a $k \times k$ matrix (so $\ell = k$), $U_x \perp U_y$, $U_x \perp U_z$, 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:

\begin{align*}
(3') \ Y &= Z' \pi_o + U_x' \beta_o + U_y.
\end{align*}

In $S_3$, $X$ is endogenous since XR does not hold. Nevertheless, structural identification of the effect of $X$ on $Y$ is ensured by:

\[(XI) \quad \text{Exogenous Instruments: } Z \perp U_y\]

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

\[E(ZU_y) = 0. \quad (M2)\]

Using (3) then gives the structural identifying condition $E(ZY) - E(ZX') \beta_o = 0$.

Identification is complete provided stochastic identification holds; for this, we now require that $E(ZX')$ is non-singular. This directly embodies the standard rank and order conditions. Parallel to Proposition 3.1.1, we have

**Proposition 3.2.1** Suppose A.1 and A.2 hold such that: (i) $Y = X' \beta_o + U_y$, $X = \gamma_x Z + \alpha_x U_x$ (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: $Z \perp U_y$ hold.

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

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

The familiar result that the plug-in estimator $\hat{\beta}_n^{XI} \equiv (Z'X)^{-1}(Z'Y)$ is consistent and asymptotically normal for $\beta_o$ 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 $\beta_o$ 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

(ii) $Z$ indirectly causes $Y$, and the effect of $Z$ on $Y$ is identified

(iii) $Z$ causes $Y$ only via $X$

We refer to $Z$ as *observed exogenous instruments* (OXI), as it is the observable vector $Z$ that satisfies the specified causal properties.

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 assuming that $E(XZ')$ and $E(ZZ')$ exist and are finite with $E(ZZ')$ non-singular, Proposition 3.1.1 establishes that $\gamma_x$, the effect of $Z$ on $X$, is identified as

$$\gamma_x = \frac{E(XZ')}{E(ZZ')}.$$

Under mild assumptions, the OLS estimator from a regression on structural equation (2), $$(Z'Z)^{-1}(Z'X),$$

is consistent for $\gamma_x$. Similarly, since $Z \perp U_X$ and $Z \perp U_Y$ and assuming that $E(ZY)$ exists and is finite, Proposition 3.1.1 establishes that $\pi_o$, the effect of $Z$ on $Y$, is identified as

$$\pi_o = \frac{E(ZY)}{E(ZZ')}.$$

By CP:OXI (iii) the only way in which $Z$ can affect $Y$ is via $X$. The effect of $X$ on $Y$, $\beta_o$, thus equals 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 $\gamma_x$ non-singular. $\beta_o$ is then identified as:

$$\beta_o = \gamma_x^{-1} \pi_o = \frac{E(ZZ')}{E(ZY)} = \frac{E(XZ')}{E(ZY)}.$$

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

$$\frac{E(ZX')}{E(ZY)} = \frac{E(ZY)}{E(ZZ')} = \beta_o.$$ 

This is essentially the classical account of IV estimation of the coefficients of one of a system of structural equations. It bears explicit statement, however, for two reasons. First, it makes fully explicit all the causal components; second, it provides a “base case” against which variations on this account, provided below, can be compared.

The work of Angrist (1990) provides an example of the use of OXI in which all the causal elements are clear. Angrist is interested in measuring the effect of serving in the military during the Vietnam War on the civilian wage of a veteran after the war. Serving in the Vietnam War and the veteran’s civilian wage could possibly be confounded by variables such as the individual’s ability and education level, since these variables might jointly affect whether an individual joins the military and his/her civilian wages. Since serving in the military could thus potentially be an endogenous regressor, Angrist employs the Vietnam draft lottery number as an observed exogenous instrument in our
terminology. The lottery number was randomly assigned to individuals based on their date of birth and dictated that individuals whose date of birth corresponds to a low number (a one below a certain threshold) have to serve in the army, whereas those whose date of birth corresponds to a high number do not.

We now show that Angrist’s use of the draft lottery number as an instrument satisfies CP:OXI. Implicitly, Angrist assumes that the randomness of the draft lottery number makes it statistically independent of unobserved factors that affect whether an individual joins the military or his/her civilian wages, and that the draft lottery number does not affect the veteran’s wages except via serving in the army (see Graph 4.) If the data are indeed generated as in $G_4$, then the draft lottery number is a proper observed exogenous instrument.

A main goal of AIR is to provide an explicit causal account of the operation of the method of instrumental variables. To this end, AIR employ the “potential outcome” framework. We now compare the present approach with that of AIR. We maintain AIR’s notation with the only change being our use of $X_i$ and $X$ instead of their $D_i$ and $D$ to denote the receipt of treatment. AIR let $i = 1, \ldots, n$ denote individuals in the population of interest and assume that the assignment $Z_i$ to a binary treatment is “ignorable” but that the receipt of the treatment $X_i$ is “non-ignorable.” AIR 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”:

(a) Single Unit Treatment Value Assumption (SUTVA):
\[
\text{if } Z_i = Z_i' \text{ then } X_i(Z) = X_i(Z');
\]
\[
\text{if } Z_i = Z_i' \text{ and } X_i = X_i' \text{ then } Y_i(Z, X) = Y_i(Z', X')
\]

(b) The treatment assignment $Z_i$ is random

(c) Exclusion restriction: $Y(Z, X) = Y(Z', X)$ for all $Z$ and $Z'$ and for all $X$

(d) Nonzero average causal effect of $Z$ on $X$

(e) Monotonicity: $X_i(1) \geq X_i(0)$ for all $i = 1, \ldots, n$. 

Graph 4 ($G_4$)
OXI for the Effect of Serving in the Army on Veteran’s Wage
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. In that case, assumption (a) is satisfied by A.1, so that the left- and right-hand side variables in every structural equation in $S_3$ pertain only to a given individual. Assumption (b) in AIR (or more weakly the ignorability of the assignment of $Z$) 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 go through $X$; this is ensured by structural equations (2) and (3) in $S_3$. Assumption (d) states that $Z$ has an effect on the treatment $X$, which is ensured by structural equation (2) in $S_3$ with $\gamma_x$ non-singular. Finally, assumption (e) assumes that the causal relationship between $Z$ and $X$ is monotonic in the sense 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 the structural equations of $S_3$ as $\gamma_x$ is the same for each individual, given our assumed absence of heterogeneity.

### 3.2.2 Proxies for Unobserved Exogenous Instruments

In satisfying CP:OXI, $S_3$ provides a causal account of the standard method of instrumental variables, but it imposes the strong requirements that $U_x \perp U_z$ and $U_y \perp U_z$, or that $Z$ is random (or ignorable) in AIR’s language. Nevertheless, random instruments are infrequent and hard to argue for in economics generally, as it is largely an observational science. Further, neither the standard relevancy and validity conditions nor Proposition 3.2.1 necessarily require $Z$ to be ignorable or random. It suffices that $Z$ is correlated with $X$ and uncorrelated with $U_y$. In particular, Proposition 3.2.1, as is standard in the econometrics literature (e.g. Heckman, 1996), does not require $Z \perp U_z$: the effect of $Z$ on $X$, usually estimated from a “first stage” regression, need not be identified.

In this section, we present a causal account of the method of instrumental variables when $Z$ is relevant and valid but is not random or ignorable. We thus present a causal explanation supporting the standard relevancy and validity moment conditions without having to impose further assumptions. In so doing, we relax AIR’s conditions for the standard method of instrumental variables to consistently estimate a causal effect.

For this, consider the structural equations system $S_5$ and its associated causal graph $G_5$:

\[
\begin{align*}
(1) & \quad Z^c = \alpha_x U_z \\
(2) & \quad X^c = \gamma_x Z + \alpha_x U_x \\
(3) & \quad Y^c = X^c \beta_o + U_y
\end{align*}
\]

where $\gamma_x$ is a $k \times k$ matrix (so $\ell = k$), $U_x \perp U_y$, $U_x \perp U_z$ and $U_y \perp U_z$.

Substituting structural equation (2) into structural equation (3) with $\pi_o \equiv \gamma_x \beta_o$ we have:

Graph 5 ($G_5$)  
Proxies for Unobserved Exogenous Instruments (PXI)
\[(3') \ Y = Z' \pi_0 + U_x' \alpha_0 \beta_o + U_y\]

Note that here \(U_x \perp U_z\), where in \(S_3\) we have \(U_x \perp U_z\).

Since XR does not hold, Proposition 3.1.1 need not hold, and the usual OLS estimator is generally inconsistent for \(\beta_o\). However, XI is satisfied as \(Z \perp U_y\). Since we further have that \(Z \perp X\), \(Z\) is a relevant and valid standard instrumental variable. As a result, Proposition 3.2.1 applies to \(S_5\), so \(\beta_o\) is identified as \(\beta_o = \left[ E(ZX') \right]^{-1} E(ZY)\).

Clearly, \(Z\) in \(S_5\) satisfies XI. Nevertheless, \(S_5\) differs fundamentally from \(S_3\), in that the causal properties making \(Z\) in \(S_5\) instrumental for identifying \(\beta_o\) are satisfied not by \(Z\) but instead by the unobservable causes \(U_z\). If these were observable, they could act as standard instruments. In their absence, the observable vector \(Z\) turns out to act as a proxy for the unobservables \(U_z\). Accordingly, we refer to \(Z\) in \(S_5\) as *proxies for (unobserved)* exogenous instruments (PXI) to distinguish this case from OXI. The 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 the unobservable \(U_z\) appearing instead of the observable \(Z\). 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 its effect 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 \(Z\) causes \(Y\).

In the PXI case, the effect of \(X\) on \(Y\) can be represented as the “ratio” of the full effect of \(U_z\) on \(Y\) to the full effect of \(U_z\) on \(X\); however, the unobservability of \(U_z\) prohibits a direct computation. Moreover, in \(S_5\) the effects of \(Z\) on \(Y\) and of \(Z\) on \(X\) are not identified as they are in the OXI case. The classical account of instrumental variables as indirect least squares does not work here. It thus might seem that the identification of the causal effect of interest is precluded in this case. Fortunately, however, \(Z\) plays the role of a proxy for \(U_z\) that, given the causal structure of \(S_5\), enables it to identify \(\beta_o\), the effect of \(X\) on \(Y\).

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 regression estimator of the effect of \(Z\) on \(X\) inconsistent, it simultaneously and systematically renders the estimator of the effect of \(Z\)
on Y 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 \( E(U,Z')\) exist and are finite and that the needed inverses exist. The effect of \( Z \) on \( X \), \( \gamma \), is not identified as \( E(XZ')\) \( [E(ZZ')]^{-1} \) from (2) since \( Z \perp U_x \). Instead, we have

\[
\gamma = E(XZ')\{E(ZZ')\}^{-1} - \alpha X \{E(XZ')\}^{-1} E(ZZ')\}^{-1}.
\]

Similarly, since \( Z \perp U_z \), the effect of \( Z \) on \( Y \), \( \pi_o \), is not identified as \( [E(ZZ')]^{-1} E(ZY)\) from (3'). Instead we have

\[
\pi_o = [E(ZZ')]^{-1} E(ZY) - [E(ZZ')]^{-1} E(ZU_x') \alpha \beta_o.
\]

Nevertheless, \( \beta_o \), the effect of \( X \) on \( Y \), is identified from (3) as \( \beta_o = E(XZ')^{-1} E(ZY) \). To verify this from the expressions above, we write

\[
E(XZ')^{-1} E(ZY) = \{ [E(ZZ')]^{-1} E(XZ') \}^{-1} [E(ZZ')]^{-1} E(ZY)
\]

\[
= \{ \gamma + [E(ZZ')]^{-1} E(ZU_x') \alpha \beta_o \}^{-1} [\pi_o + [E(ZZ')]^{-1} E(ZU_x') \alpha \beta_o]
\]

\[
= \{ \gamma + [E(ZZ')]^{-1} E(ZU_x') \alpha \beta_o \}^{-1} \{ \gamma + [E(ZZ')]^{-1} E(ZU_x') \alpha \beta_o \} \beta_o
\]

\[
= \beta_o.
\]

The above expression also clearly shows that we may have \( \gamma \) equal to zero, so that in \( S_5 \), the P XI \( Z \) do not have to cause \( X \), whereas \( \gamma \) 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 statistical association required to identify the effect of interest. We provide an example below. When \( \gamma = 0 \), \( Z \) can be thought of as a “pure predictive proxy” for \( U_z \), the true causal instrumental variable that allows the recovery of the effect of \( X \) on \( Y \). (We further discuss predictive proxies below.) The P XI case thus makes use of a causally meaningful instrument that satisfies the relevancy and validity moment conditions but that does not satisfy the conditions of Angrist, Imbens and Rubin (1996.) In particular, P XI provides a causal account of the method of instrumental variables that relaxes two of A IM’s assumptions, namely, random assignment (or more weakly the ignorability) of \( Z \) (assumption (b)) and nonzero average causal effect of \( Z \) on \( X \) (assumption (d)).

We note that in the P XI case, a function of two inconsistent estimators, the OLS estimators of \( \gamma_x \) and \( \pi_o \) for structural equations (2) and (3'), is itself a consistent estimator for the effect of interest, \( \beta_o \). Thus, identification strategies that advocate the recovery of causal effects as functions only of identifiable effects, as in Pearl (2000, p.153-154), miss recovering certain identifiable causal effects.
A number of applied papers in economics that use the standard method of instrumental variables to estimate the effect of a potentially endogenous $X$ on $Y$ implicitly employ the PXI method to justify the validity and relevancy of their instruments. Consider, for example, measuring the effect of the number of years of education an individual receives on his/her 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 that individual’s ability, making the variable “years of education” potentially endogenous. To avoid this problem, BC employ the numbers 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 their future wages by means other than their educational attainment. In our framework, BC attempt to use a statistical association between the number of sisters and the level of education that the daughters attain without necessarily having that one causes the other. For instance, we can postulate that aspects of the parents’ socioeconomic background and capacity to help finance the daughters’ college education is what is generating the statistical association between the number of sisters and the education level (see $G_6$). If the data are indeed generated as in $G_6$, then the number of sisters is a proxy for the unobserved exogenous instrument(s) “parents’ socioeconomic background.”

![Graph 6 ($G_6$)](image)

The OXI and PXI methods allow the identification of the effect of an endogenous $X$ on the response of interest $Y$, as they employ instruments $Z$ that satisfy the standard validity and relevance conditions. We thus refer to such variables $Z$ as “proper standard instruments.” Of equal importance is that $Z$ provides a source of variation that precedes $X$ and that affects $Y$ only via $X$ if at all. We thus also refer to OXI or PXI $Z$ as pre-cause instrumental variables, as they causally precede the cause of interest $X$. We also refer to any $Z$ satisfying XI as an unconditional instrumental variable, to distinguish it from the conditional instrumental variables discussed below.

### 3.3 Failures of Identification
In this section, we examine how structural identification of $\beta_o$ via the XI method fails in the standard “irrelevant,” “invalid,” and “under-identified” cases. We thus demonstrate how, under A.1 and A.2, our causal framework accounts for not only the successes of the standard method of instrumental variables but also its failures.

### 3.3.1 Irrelevant Exogenous Instruments

Not only must proper instruments $Z$ be valid, they must also be relevant, i.e. correlated with $X$, in order to ensure the identification of the effect of $X$ on $Y$. Structural equation system $S_7$ and its associated causal graph $G_7$ depict the irrelevant XI case and demonstrate how an irrelevant XI satisfies neither CP:OXI nor CP:PXI.

Let $S_7$ be given by:

1. $Z = \alpha_z U_z$
2. $X = \alpha_x U_x$
3. $Y = X' \beta_o + U_y$

where $U_x \perp Y$, $U_x \perp Z$ and $U_y \perp Z$.

Even though $Z$ is a valid standard instrument satisfying XI, it cannot be used to identify the effect of $X$ on $Y$, because it is no longer true that the effect of $X$ on $Y$ can 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 of these effects are zero, and their ratio is hence undetermined. In $S_7$, neither CP:OXI(i) nor CP:PXI(i) hold, since neither $Z$ nor $U_z$ cause $X$. We thus call these irrelevant exogenous variables. Observe that when $\ell = k$ (as assumed here), the presence of irrelevant exogenous variables causes condition (ii) of Proposition 3.2.1 (stochastic identification) to fail. The causal effect $\beta_o$ fails to be identified in this case.

### 3.3.2 Endogenous Instruments

We next examine the failure of XI, condition (ii) of Proposition 3.2.1. In this case $Z \perp U_y$; in accord with standard terminology we call such $Z$ endogenous instruments. There are a number of ways that this can occur, which we now consider in some detail.

We first note that a potential instrument $Z$ does not need to be relevant in order to be endogenous. An example of an irrelevant and endogenous instrument $Z$ is one such that $Z$ doesn’t cause $X$ and $U_z \perp U_x$, but both $U_x$ and $U_z$ cause $U_y$. Turning now to relevant instruments, consider the system $S_8$:
(1) $Z^c = \alpha^c U^c_z$

(2) $X^c = \gamma^c_x Z^c + \alpha^c_U U^c_x$

(3) $Y^c = X^c' \beta^c_o + U^c_y$

where $U^c_x \perp/ U^c_y$, $U^c_x \perp/ U^c_z$, and $U^c_y \perp/ U^c_z$. Substituting (2) into (3) with $\pi_o \equiv \gamma^c_x' \beta^c_o$ gives

$$(3') Y^c = Z^c' \pi^c_o + U^c_x' \alpha^c_x' \beta^c_o + U^c_y.$$

We have endogenous instruments because $U^c_i \perp/ U^c_z$, implying that XI fails. This can occur in several ways. For example, correlation between $U^c_z$ and $U^c_y$ can arise because either $U^c_y$ causes $U^c_z$ (see $G_{8d}$, $G_{8b}$) or $U^c_x$ causes both $U^c_z$ and $U^c_y$ (see $G_{8c}$). In this case, CP:OXI(ii) does not hold, as $Z$ and $Y$ are confounded; and CP:PXI(ii) does not hold, as $U^c_z$ and $Y$ are confounded, implying that the full effect of $U^c_z$ on $Y$ would not be identified had $U^c_z$ been observed. In fact, CP:PXI(i) also fails here, as $U^c_z$ and $X$ are also confounded.

Alternatively, an endogenous instrument occurs when $U^c_z$ affects $U^c_y$ via a channel other than X. Since by assumption $Z$ can't cause $U^c_y$, we need consider only the case where $U^c_z$ causes $Y$ via an intermediate cause other than $X$ (see $G_{8d}$ and $G_{8e}$). In this case, CP:OXI(ii) does not hold as $Z$ and $Y$ are confounded; and CP:PXI(iii) does not hold as $U^c_z$ causes $Y$ via an intermediate cause other than $X$. The effect of $X$ on $Y$ can thus no longer be expressed as the ratio of the effect of $U^c_z$ on $Y$ and the effect of $U^c_z$ on $X$. 

Graph 8a ($G_{8a}$)  Endogenous Instruments

Graph 8b ($G_{8b}$)  Endogenous Instruments

Graph 8c ($G_{8c}$)  Endogenous Instruments

Graph 8d ($G_{8d}$)  Endogenous Instruments

Graph 8e ($G_{8e}$)  Endogenous Instruments
The conclusion of Proposition 3.2.1 fails in this case because structural identification fails. From structural equation (3) we have

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

but \( E(ZU_y) \) does not vanish, precluding structural identification of \( \beta_o \) as \( E(ZX')^{-1}E(ZY) \).

### 3.3.3 Under-Identified Exogenous Instruments

We now consider what happens when instruments \( Z \) are valid and relevant, but condition (i) of Proposition 3.2.1 fails. Specifically, consider the system \( S_0 \):

\[
\begin{align*}
(1) \ Z &= \alpha_z U_z \\
(2) \ X &= \alpha_x U_x \\
(3) \ Y &= X' \beta_o + Z' \gamma_o + U_y
\end{align*}
\]

where \( U_x \perp U_y, U_x \perp U_z, \) and \( U_y \perp U_z \).

In this case, the regressors \( X \) are endogenous, as \( U_x \perp U_y \), but we have that \( Z \) is relevant since \( X \perp Z \), and valid since \( Z \perp U_y \). It follows from (3), however, that

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

Once again, structural identification of \( \beta_o \) fails, this time due to the presence of the unknown (non-zero) \( \gamma_o \). In terms of CP:OXI and CP:PXI, the problem is that \( Z \) enters the structure that determines \( Y \) directly, and not solely via \( X \). This violates property (iii) of CP:OXI and property (iv) of CP:PXI, since \( Z \) affects \( Y \) directly instead of via \( X \).

Viewed in this way, the lack of structural identification appears as a kind of “omitted variables bias,” resulting from the failure to include \( Z \) in the instrumental variables regression. This difficulty cannot, however, be resolved by including \( Z \) in the IV regression as then one is attempting to identify both \( \beta_o \) and \( \gamma_o \), and there are not enough proper instruments to do this. The standard order condition for identification requires the availability of at least one valid instrument for each right-hand side variable of a structural equation. Attempting to include \( Z \) in the IV regression places us in the classical “under-identified” case in which there are more right-hand side variables than valid
instruments. In addition to condition (i) of Proposition 3.2.1 not holding, this causes condition (ii) to fail too for the IV regression that includes both $X$ and $Z$ as regressors and uses only $Z$ as instruments.

4 Extended Instruments

As the example in the introduction demonstrates, it is possible to identify causal effects of interest even in the absence of exogenous regressors or instruments. We now investigate situations in which vectors $Z$ or $W$ are not valid instruments in the standard sense, as they are correlated with the error term $U_y$, but are nevertheless instrumental in identifying the effect of $X$ on $Y$. We therefore call such variables extended instrumental variables (EIV.). In particular, we introduce the notions of conditional or conditioning EIV. In each case, we explicitly reference two key conditions that together ensure the full identification of the causal effect of interest: (i) a conditional independence relationship that parallels the validity condition for the standard instrumental variables method, ensuring what White and Chalak (2006a) call structural identification; and (ii) the analog of the standard relevance, order, and rank conditions for identification, ensuring what White and Chalak (2006a) call stochastic identification.

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 results of Section 3.2 concern identification of causal effects using a single vector of what we have called “unconditional” instruments $Z$. We now consider single EIV methods that employ a vector of conditioning instruments $W$ to proxy for the effects of unobserved confounding variables for $X$ and $Y$. We extend our notation by writing $W \equiv [W_1, \ldots, W_m]'$, $U_w \equiv [U_{w_1}, \ldots, U_{w_m}]'$, and letting $W$ denote an $n \times m$ matrix of identically distributed observations on $W$.

The treatment effect literature has introduced two central methods to treat the problem of confoundedness: randomization and matching (see for e.g. Rubin, 1974; Rosenbaum, 2002.) R.A. Fisher (1949, p. 12) argues that randomization is the “reasoned basis” for inference (see Rosenbaum, 2002, p. 21.) If randomization is feasible, randomly assigning agents to treatment and control groups ensures that there aren’t systematic confounding variables for the cause and effect of interest. Since the process is random, the distribution of the assignment mechanism is known: it gives equal probability to every possible treatment assignment. We saw above in cases $S_1$ and $S_3$ that randomization permits identification of causal effects. Randomization, however, is uncommon in observational studies, where the researcher usually lacks the ability to control the variables of interest.

In non-randomized studies, matching units that share common causes or attributes from the treatment and control groups provides a way forward. By conditioning on the information in the true confounding variables, it is possible to interpret the remaining
conditional association between the putative cause and effect 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), the “ignorability condition” and “propensity score” (Rubin, 1974; Rosenbaum and Rubin, 1983), the “back-door” method (Pearl, 1995), and “predictive proxies” (White, 2006; White and Chalak, 2006a.) In labor economics, matching methods are well established and have been discussed in the contexts of the distribution of earnings, policy evaluation, and the return to education and training programs, for example (see Roy, 1951; Heckman and Robb, 1985; Heckman, Ichimura, and Todd, 1998.)

We now investigate causal structures in which conditioning instruments \( W \) permit the identification of the effect of an endogenous \( X \) on \( Y \). Thus, consider structural equations system \( S_2 \), in which \( X \) is endogenous because \( U_x \perp U_y \). Suppose that this dependence arises from the presence of a common cause for both \( U_x \) and \( U_y \). It is instructive to start with the extreme case where we actually observe the true common causes or confounding variables \( W \) that jointly determine \( U_x \) and \( U_y \). Of course, this violates our assumption that observables do not cause unobservables (A.1(a)), so this is only a temporary expedient adopted to provide useful insight. We will remove this shortly. To proceed, consider structural equations system \( S_{10a} \) and its associated causal graph \( G_{10a} \):

\[
\begin{align*}
(1) \quad & W = \alpha_w U_w \\
(2) \quad & U_{x_1} = \gamma_{x_1} W \\
(3) \quad & U_{y_1} = \gamma_{y_1} W \\
(4) \quad & X = \alpha_{x_1} U_{x_1} + \alpha_{x_2} U_{x_2} \\
(5) \quad & Y = X' \beta_o + U_{y_1} + U_{y_2}
\end{align*}
\]

so that \( U_x \perp U_{y_1}, U_x \perp U_{y_2}, \) and \( U_y \perp W \), 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 \( W \).

\(^3\) In subsequent structural equations systems of Section 4, we sometimes 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 \( S_{10b} \) (and \( S_{12} \) below), necessary for stochastic identification.
In $S_{10a}$, 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 $S_{10a}$ that parallels $XR$ and $XI$ above is

\[(CXR|I) \quad \text{Conditionally Exogenous Regressors given Conditioning Instruments:} \quad X \perp U_y | W\]

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

The role of $S_{10a}$ is merely to motivate $CXR|I$. We emphasize that $S_{10a}$ is by no means a necessary structure for $CXR|I$ to hold. As we discuss below, $CXR|I$ can also hold for properly chosen $W$ even when the true confounding variables for $X$ and $Y$ cannot be observed.

Just as $XR$ and $XI$ can deliver structural identification of $\beta_o$, so can $CXR|I$. Specifically, the key moment condition resulting from $CXR|I$ in our linear separable framework is:

$$E(XU_y | W) = E(X | W) \times E(U_y | W). \quad (M3)$$

To see how this condition structurally identifies $\beta_o$, rewrite $(M3)$ as

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

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 that

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

so that $\beta_o$ is structurally identified as

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

Note that this derivation relies only on $Y = X' \beta_o + U_{y_1} + U_{y_2}$, the linear regression representation $E(XW')[E(WW')]^{-1}W$ for $E(X | W)$, and $CXR|I$. The specific structure of $S_{10a}$ is not required.

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

$$\beta_o = \{E(XX') - E(XW')[E(WW')]^{-1} E(WX')\}^{-1} \times \{E(XY) - E(XW')[E(WW')]^{-1}E(WY)\}.$$
We defer a formal statement of this result until we have further explored the causal structures relevant to this case. From a causal perspective, identification holds because after conditioning on $W$, the association remaining between $X$ and $Y$ can only be explained as a response in $Y$ due to variation in $X$.

Under mild conditions, a consistent and asymptotically normal plug-in estimator for $\beta_0$ is

$$\hat{\beta}_n^{CXRIL} = \{X'(I - W(W'W)^{-1}W')X\}^{-1}\{X'(I - W(W'W)^{-1}W')Y\}.$$ 

Even though $W$ plays an instrumental role in identifying $\beta_0$, there is no requirement that $W$ be exogenous. For example, in $S_{10a}$ we clearly have that $W$ is endogenous, as $W \perp U_y$. Conditioning instruments are thus not standard instruments, motivating their description as extended instrumental variables (EIV). We call $\hat{\beta}_n^{CXRIL}$ an EIV estimator.

Inspecting $\hat{\beta}_n^{CXRIL}$, we see that it has the form of a standard instrumental variables estimator using as derived standard instruments the residuals of the regression of $X$ on $W$, $X - E(XW')(E(WW'))^{-1}W$. Nevertheless, we do not put these derived instruments on an equal footing with $W$, as it is $W$ that provides the natural causal explanation enabling the recovery of the effect of $X$ on $Y$. Of even greater significance, however, is the fact that as White and Chalak (2006a) show, when A.2 is relaxed to permit non-separable structures, these derived instruments no longer play an essential role, whereas $W$ (the vector of “predictive proxies” in White and Chalak’s (2006a) terminology) continues to play the instrumental role in identifying the causal effects of interest.

The estimator $\hat{\beta}_n^{CXRIL}$ is easily recognized as the Frisch-Waugh (1933) partial regression estimator, obtained by regressing $Y$ on the residuals $(I - W(W'W)^{-1}W')X$ from a regression of $X$ on $W$. This can also be obtained as the coefficient estimator associated with $X$ from a simple linear regression of $Y$ on both $X$ and $W$.

This latter regression emerges naturally from $S_{10a}$, after performing the simple substitutions required to enforce our convention that observables do not cause unobservables. Substituting (2) into (4) and (3) into (5) in $S_{10a}$ gives the structure $S_{10b}$:

\begin{align*}
(1) \quad & W = \alpha_w U_w \\
(2) \quad & X = \gamma_x W + \alpha_x U_x \\
(3) \quad & Y = X'\beta_0 + W'\gamma_0 + U_y
\end{align*}

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

[Graph 10b ($G_{10b}$)]

Conditioning Instruments
In formulating $S_{10b}$, we have adjusted the notation in the natural way. With the given independence conditions, we are back to the case treated by Proposition 3.1.1 as $X$ and $W$ jointly satisfy XR. In $S_{10b}$, both $\beta_o$, the causal effect of $X$ on $Y$, and $\gamma_o$, the causal effect of $W$ on $Y$ are identified. Observe that the latter is only the direct causal effect of $W$ on $Y$. The full causal effect of $W$ on $Y$ is given by $\gamma_o + \gamma' \beta_o$, which is identified from a regression of $Y$ on $W$ only.

As noted above, $S_{10a}$ or $S_{10b}$ is sufficient but not necessary for CXRII. Structures satisfying Pearl’s (1995; 2000, pp. 79-81) “back-door” criterion, in which an observable (here $W$) mediates an indirect link between $X$ and $Y$ also ensure CXRII. In Pearl’s framework, $W$ is modeled either as the vector of common causes ($G_{10a}$, $G_{10b}$), or as an effect of the unobserved common cause and a cause of either $Y$ or $X$ ($G_{11a}$, $G_{11b}$). In $G_{11a}$ and $G_{11b}$ below, CXRII holds because either $U_x$ causes $Y$ via $W$ or $U_y$ causes $X$ via $W$. We do not observe the confounding variable, which is $U_x$ in $G_{11a}$ and $U_y$ in $G_{11b}$. Instead, $W$ acts as an observable proxy for the true confounding variables in each case.

Specifically, let $S_{11a}$ be given by

\begin{align*}
(1) \quad W^c &= \alpha_w U_w \\
(2) \quad X^c &= \alpha_x U_x \\
(3) \quad Y^c &= X' \beta_o + W' \gamma_o + U_y
\end{align*}

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

Similarly, let $S_{11b}$ be given by

\begin{align*}
(1) \quad W^c &= \alpha_w U_w \\
(2) \quad X^c &= \gamma_x W + \alpha_x U_x \\
(3) \quad Y^c &= X' \beta_o + U_y
\end{align*}

with $U_w \perp U_x$, $U_w \perp U_y$, and $U_x \perp U_y$. 
There are a number of noteworthy features about each of these structures. First consider $S_{11a}$. Note that the direction of causality between $U_x$ and $U_w$ is not specified in $G_{11a}$. $S_{11a}$ thus corresponds to three possible back door structures. For all of these structures, $X$ and $W$ jointly satisfy XR in (3), so Proposition 3.1.1 holds; the situation for $S_{11a}$ is completely parallel to that for $S_{10a}$. In $S_{11a}$, $W$ is a structurally relevant exogenous variable correlated with $X$, so omitting $W$ from the identifying regression results in the classical “omitted variable bias.” The inclusion of $W$ is thus crucial to identifying $\beta_0$.

Now consider $G_{11b}$. For concreteness, suppose $U_y$ causes $U_w$. Now $X$ is endogenous because $X \perp U_y$. Also, $W$ is endogenous because $W \perp U_y$. Nevertheless, given CXRII, $\beta_0$ is structurally identified. If stochastic identification also holds, $\beta_0$ is identified as

$$\beta_0 = \{E(XX') - E(XW')E(WW')\}^{-1}E(WY)$$

As we observed above, this is consistently estimated by the coefficients on $X$ from a regression of $Y$ on both $X$ and $W$. But this is truly a remarkable situation: here we have causally meaningful coefficients $\beta_0$ identified and consistently estimated using a regression that not only contains endogenous regressors $X$, but also structurally irrelevant and endogenous regressors $W$. (We call $W$ “structurally irrelevant” because $W$ does not appear in (3) of $S_{11b}$.) According to the textbooks, such a regression should yield nonsense. Nevertheless, the causal structure ensures structural identification of $\beta_0$.

What about the remaining coefficients in this regression, those associated with $W$? In the context of $S_{11b}$, these have no causal interpretation. Instead they have only a predictive interpretation, as discussed in detail by White (2006) and White and Chalak (2006a). We thus have an interesting situation in which some regression coefficients have a causal meaning (those associated with $X$), but others do not (those associated with $W$). That is to say, not all of the regression coefficients need to have signs and magnitudes that make causal sense. This constitutes an instance of what Heckman (2006) has termed “Marshak’s maxim,” which holds that we may identify certain economically meaningful components of a given structure (here $\beta_0$) without having to identify the entire structure.

Nor does Pearl’s back door method exhaust the possibilities for achieving CXRII. Another possibility, discussed in depth by White (2006) and White and Chalak (2006a) is the case of “predictive proxies.” In the present framework, predictive proxies arise from structures such as $S_{12}$, which violates Pearl’s back-door criterion:

$$\begin{align*}
(1) \quad W &= \alpha_{w_1} U_{w_1} + \alpha_{w_2} U_{w_2} \\
(2) \quad U_{x_i} &= \gamma_{x_i} U_{w_1} \\
(3) \quad X &= \alpha_x U_{x_1} + \alpha_x U_{x_2} \\
(4) \quad U_{y_i} &= \gamma_{y_i} U_{w_1} \\
(5) \quad Y &= X^\prime \beta_0 + U_{y_1} + U_{y_2}
\end{align*}$$

Graph 12 ($G_{12}$)
Conditioning Instruments
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 $U_w = (U_{w_1}', U_{w_2}')', U_x = (U_{x_1}', U_{x_2}')'$, $U_x = (U_{x_1}', U_{x_2}')'$, and $U_y = (U_{y_1}', U_{y_2}')'$, so that $W \perp U_{y},$ and $X \perp U_{y}$.

In $S_{12}$, we view $U_{w_1}$ as an unobserved common cause for $X$ and $Y$; the predictive proxy $W$ can be viewed as a measurement error-laden version of $U_{w_1}$. We see that both $X$ and $W$ are endogenous; however, Proposition 4.4 of White and Chalak (2006a) applies to $U_{w}$. The key to this is the ability of $W$ to predict $U_{w}$ (hence $X$) sufficiently well that $U_{y}$ contains no additional information useful in predicting $X$. Just as in $S_{11b}$, the causal effect of interest is identified from a regression containing endogenous $X$ and structurally irrelevant endogenous $W$. Our comments about $S_{11b}$ fully apply to $S_{12}$.

We may refer to unobserved confounding variables for $X$ and $Y$ as the *common causes* of $X$ and $Y$. Given its role as an observable proxy for the unobserved common causes, we may also refer to $W$ ensuring $CXR|I$ as a vector of *common cause instruments*.

The $CXR|I$ condition can play a central role in the treatment effects and matching literature. Using $Y_x$ to denote the value that $Y$ would take had $X$ been set to $x$ (the “potential outcome”), it can be shown that when $Y = X'\beta_0 + U_y$ and $CXR|I$ holds, then the key “ignorability” or “unconfoundedness” condition $Y_x \perp X | W$ of Rosenbaum and Rubin (1983) holds. (See White (2006, proposition 3.2.).)

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) $Y = X'\beta_0 + U_y$, 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) $CXR|I$: $X \perp U_y | W$ holds.

Then $\beta_0$, the average total causal effect of $X$ on $Y$, is fully identified as:

$$\beta_0 = \{E(XX') - E(XW')[E(WW')]^{-1}E(WX')\}^{-1}\{E(XY) - E(XW')[E(WW')]^{-1}E(WY)\}.$$  

In contrast to the XI method, we note that the $CXR|I$ method does not require $\ell = k$, either for structural or for stochastic identification.

White and Chalak (2006a) 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) there
is no need to provide a list of causal properties for CXRII parallel to CP:OXI or CP:PXI. Note, however, that because we are interested in the total effect of $X$ on $Y$, we do not permit $X$ to cause $W$.

### 4.1.2 Conditional Instruments

In Section 3.2, we discuss the use of standard exogenous instrumental variables $Z$ to identify the effect of the potentially endogenous $X$ on $Y$ as the ratio of the effect of $Z$ on $Y$ and that of $Z$ on $X$. In this section, we demonstrate how a single vector of extended instruments $Z$, that we refer to as conditional instruments, permits the identification of the causal effect of the endogenous $X$ on $Y$ as the product of the effects of $X$ on $Z$ and that of $Z$ on $Y$. We also refer to this class of extended instruments as intermediate cause instrumental variables as these variables mediate the effects of $X$ on $Y$.

To illustrate, consider structural equation system $S_{13}$ and its associated causal graph $G_{13}$:

\[
X^c = \alpha_x U_x \\
Z^c = \gamma X + \alpha_z U_z \\
Y^c = Z^c \delta + U_y
\]

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) we get:

\[
(3') Y^c = X' \beta_o + U_z \alpha_z \delta + U_y; \text{ with } \beta_o \equiv \gamma \delta.
\]

This is the structure described in the introduction. We immediately see that $X$ is endogenous, and it is also clear that $Z$ is endogenous, so neither standard regression nor standard instrumental variables methods can identify $\beta_o$, the effect of interest. Nor do we have CXRII, as $X \perp U_y \mid Z$, so there are no conditioning instruments available. Nevertheless, $\beta_o$ is structurally identified as a result of our next conditional independence relationship:

\[\text{(CXIIIR)} \quad \text{Conditionally Exogenous Instruments given Regressors: } Z \perp U_y \mid X\]

We refer to such $Z$ as conditional instruments and refer to methods that identify $\beta_o$ using these extended instrumental variables as CXIIIR methods.

For our linear separable system, CXIIIR implies the key moment condition
\[ E(ZU_y | X) = E(Z | X) \times E(U_y | X) \]  

\( (M4) \)

Parallel to our analysis of CXRII, it follows from this moment condition that

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

This and structural equation (3) imply

\[ E( [Z - E(ZX')] [E(XX')]^{-1} X ] [Y - Z' \delta_o ] ) = 0, \]

so that \( \delta_o \) is structurally identified as

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

That is, \( \delta_o \) is structurally identified by CXRII with regressors \( Z \) and conditioning instruments \( X \). Full identification of \( \delta_o \) holds given stochastic identification; here, this requires the non-singularity of \( \{ E(ZZ') - E(ZX')[E(XX')]^{-1} E(XZ') \} \).

If \( \gamma \) can be also be identified, then identification of \( \beta_o \) follows, as \( \beta_o \equiv \gamma' \delta_o \). In \( S_{13} \), we see that \( \gamma \) is structurally identified by XR, as \( X \perp U_z \). If \( \gamma \) is stochastically identified (i.e., \( E(XX') \) is non-singular), Proposition 3.1.1 gives \( \gamma' = [E(XX')]^{-1} E(ZX') \).

We have the following formal result.

**Proposition 4.1.2** Suppose A.1 and A.2 hold such that: (i) \( Z = \gamma X + \alpha_c U_z, \ Y = Z' \delta_o + U_y \), 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) CXIIR: \( Z \perp U_y \perp X \) hold.

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

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

In contrast to the XI method, we note that the CXIIR method does not require \( \ell = k \). This condition is required neither for structural nor for stochastic identification.

Section 7 gives straightforward conditions under which the plug-in estimator \( \hat{\beta}^{\text{CXIIR}}_n \) is a consistent and asymptotically normal estimator for \( \beta_o \) where:

\[ \hat{\beta}^{\text{CXIIR}}_n \equiv \gamma' \delta^{\text{CXIIR}}_n \]

\[ \equiv \dot{\gamma}^{\text{XR}}_n \delta^{\text{CXIIR}}_n . \]
Although this method uses a single vector of extended instrumental variables $Z$ to identify the causal effect of interest, both these instruments and the regressors $X$ play dual roles in the process. The extended instruments play the dual role of a response for $X$ and a cause for $Y$. The regressors serve as exogenous regressors with respect to $U_z$ in (2) and as conditioning instruments with respect to $U_y$ in (3), as is explicit in (iii)(a) and (iii)(b). We reflect these latter roles in our notation $\hat{f}_n^{XR}$ and $\hat{\delta}_n^{CXRI}$ above.

Analogous to the case of unconditional instruments, we can state a succinct set of causal properties required to ensure that the conditional instruments $Z$ identify the effect of interest when $X$ and $Y$ are confounded:


(i) The effect of $X$ on $Z$ is identified
(ii) The effect of $Z$ on $Y$ is identified
(iii) If $X$ causes $Y$, it does so only via $Z$

As is readily verified, $S_{13}$ satisfies CP:XI|R.

The CXI|R method corresponds to the “front-door” method introduced by Pearl (1995). Pearl (1995, 2000) discusses the front-door method primarily in relation to the back-door method discussed in Section 4.1.1. In particular, the treatment effect literature typically applies CXII (back door) to identify the effect of interest in the presence of confounding by conditioning on a covariate (the conditioning instrument $W$) that is *not* affected by the treatment. In the CXI|R (front-door) case, we condition on a variable that *is* affected by the treatment (indeed, that mediates the treatment) to allow the identification and consistent estimation of the causal effect of the treatment $X$ on $Y$.

The CXI|R method can play a particularly useful role in measuring the effects of certain policies as illustrated in $G_{14}$. In particular, we might be interested in 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.

---

Graph 14 ($G_{14}$)
Policy Evaluation by means of CXI|R
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 Vega, 2005) but suspect that the new education law is endogenous as it is correlated with unobserved causes of the students’ performance. For instance, one might argue that 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 might recover the effect of the legislation on student performance by employing intermediate causes that are affected by the new policy and that in turn affect student performance. In this case, these intermediate cause instruments should be implementation mechanisms that are responses only to the new policy and are not caused by the unobserved common confounding causes of the policy and the response of interest otherwise. In our 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

In Sections 3.1, 3.2, 4.1.1, and 4.1.2, we examine four cases where the identification of the effect of $X$ on $Y$ obtains:

$$
\begin{align*}
XR: & \quad X \perp U_y \\
XI: & \quad Z \perp U_y \\
CXRI: & \quad X \perp U_y \mid W \\
CXII: & \quad Z \perp U_y \mid X
\end{align*}
$$

In each case we have either the independence or conditional independence of a single vector of observed variables $X$ or $Z$ from $U_y$, the unobserved causes of $Y$. (In stating CXRI, we could have used the notation $Z$ instead of $W$, but we keep the notation distinct to make explicit the unique role played by conditioning instruments.) We therefore refer to XR, XI, CXRI, and CXII as *single* EIV methods.

The remaining possibilities of independence or conditional independence from $U_y$ that are so far unexplored when considering only a single vector of unconditional, conditional or conditioning EIV $Z$ or $W$, are those associated with $Y$. Clearly, $Y \perp U_y$, $Y \perp U_y \mid X$, and $Y \perp U_y \mid W$ since, by definition, $U_y$ is an immediate cause of $Y$. Similarly, $X \perp U_y \mid Y$ since conditioning on a common effect generally renders the possibly independent $X$ and $U_y$ necessarily dependent. The final possibility to consider is whether identification can be achieved in the case for which instruments $Z$ are conditionally independent of $U_y$ given $Y$.

A causal structure that generates this final conditional independence relationship is one where $Z$ is a *post-response* instrument, so that $X$ causes $Y$, which then causes $Z$. Such a structure is given by structural equations system $S_{15}$ and its associated graph $G_{15}$. $S_{15}$ perhaps looks promising, as one may consider the possibility of identifying the effect of the endogenous $X$ on $Y$ as the ratio of the effect of $X$ on $Z$ and that of $Y$ on $Z$, analogous to indirect least squares. Unfortunately, the identification of $\beta_0$ in this way is not possible, as we now demonstrate.
Let $S_{15}$ be given by:

1. $X^c = \alpha_x U_x$
2. $Y^c = X' \beta_o + U_y$
3. $Z^c = \gamma Y + \alpha_z U_z$

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 $\delta_o \equiv \beta_o \gamma$ we get:

$$\left(3'\right) Z^c = X' \delta_o + \gamma U_y + \alpha_z U_z.$$

Unfortunately, the conditional independence condition $Z \perp U_y \mid Y$ is not sufficient to identify $\beta_o$ in manner analogous to CXRII or CXIIIR. From this condition and algebra analogous to that above, we obtain the moment equations

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

The first term above is always zero, however, so even if $\{E(ZX') - E(ZY')[E(YY')]^{-1}E(YX')\}$ is non-singular, the solution for $\beta_o$ is also always zero.

Another way to see why the above conditional independence is not sufficient for the identification and consistent estimation of $\beta_o$ is that in $S_{15}$ the effect of $Y$ on $Z$, $\gamma$, is identified from Proposition 3.1.1 as $\gamma = E(ZY')[E(YY')]^{-1}$ since $Y \perp U_z$. However, the effect of $X$ on $Z$ is not identified as $X$ and $Z$ are confounded by either $U_x$ or $U_y$. Hence, $\beta_o$ is itself not identified as the ratio of these two effects.

As this exhausts the possibilities for identification of causal effects using a single vector of instruments, we now consider methods that make use of two vectors of extended instruments.

### 4.2 Double Extended Instrumental Variables Methods

Sections 3 and 4.1 discuss all the possible ways in which a single vector of unconditional, conditional or conditioning EIVs permits the identification and consistent estimation of the effect of $X$ on $Y$. We now turn our attention to EIV methods that make joint use of conditional instruments $Z$, and conditioning instruments $W$, for the identification of causal effects of interest.

For the remainder of this section, we let the random variable $Y$ be the response of interest, the elements of the $k_1 \times 1, \ldots , k_p \times 1$ random vectors $X_1, \ldots , X_p$ be the causes of interest, and the elements of the $\ell_1 \times 1, \ldots , \ell_q \times 1$ random vectors $Z_1, \ldots , Z_q$ and the $m_1 \times 1, \ldots , m_r \times 1$ random vectors $U_1, \ldots , U_r$. 

Graph 15 ($G_{15}$)
$m \times 1$ random vectors $W_1, \ldots, W_s$ be extended instrumental variables, all with observed realizations as specified in A.1 and A.2. Their corresponding unobserved causes are given by the random variable $U_y$, and the elements of the random vectors $U_{x_1}, \ldots, U_{x_p}, U_{z_1}, \ldots, U_{z_q}, U_{w_1}, \ldots, U_{w_s}$. We put $X \equiv [X_1', \ldots, X_p']'$, $Z \equiv [Z_1', \ldots, Z_q']'$ and $W \equiv [W_1', \ldots, W_s']'$, where $X$ is of dimension $k \times 1$ with $k \equiv k_1 + \ldots + k_p$, $Z$ is of dimension $l \times 1$ with $l \equiv l_1 + \ldots + l_q$, and $W$ is of dimension $m \times 1$ with $m \equiv m_1 + \ldots + m_s$. Similarly, we put $U_x \equiv [U_{x_1}', \ldots, U_{x_p}]'$, $U_z \equiv [U_{z_1}', \ldots, U_{z_q}]'$, and $U_w \equiv [U_{w_1}', \ldots, U_{w_s}]'$. Boldface letters denote vectors and matrices of observations of $X, Y, Z$, and $W$, as above.

4.2.1 Conditional and Conditioning Instruments: CXII

Economic theory can suggest causal models that permit identification of causal effects using more than one type of instrumental variable. In this section, we discuss examples where both conditional and conditioning extended instrumental variables $Z$ and $W$ are needed to ensure structural identification.

4.2.1.a OCXII

Our first case is that of observed conditionally exogenous instruments given conditioning instruments (OCXII). To illustrate, consider structural equations system $S_{16a}$ with associated causal graph $G_{16a}$.

Let $S_{16a}$ be given by:

\begin{align*}
(1) & \quad W^c = \alpha_w U_w \\
(2) & \quad Z^c = \alpha_z U_z \\
(3) & \quad X^c = \chi Z + \alpha_x U_x \\
(4) & \quad Y^c = X' \beta_o + U_y,
\end{align*}

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 structural equation (4) and setting $\pi_o \equiv \chi' \beta_o$, we have:

\begin{align*}
(4') & \quad Y^c = Z' \pi_o + U_x' \alpha_x \beta_o + U_y
\end{align*}

The key conditional independence relationship that holds in $S_{16a}$ when $W$ is a sufficiently good predictor for $U_w$ (hence $Z$) is:
In contrast to CXRII, where the regressors are conditionally exogenous, here the conditional instruments \( Z \) are those for which the conditioning instruments, \( W \), ensure conditional exogeneity.

Given A.2, the key moment condition for structural identification resulting from CXIII is:

\[
E(ZU \mid W) = E(Z \mid W) \times E(U \mid W)
\]

(M5)

Algebra similar to that for CXRII delivers the structural identification of \( \beta_o \) under CXIII.

**Proposition 4.2.1** Suppose A.1 and A.2 hold such that: (i) \( Y = X' \beta_o + U \). Suppose further that (ii) there exist random vectors \( W \) and \( Z \) such that and that \( \ell = k \); \( E(ZY) \), \( E(ZW') \), \( E(WY) \), 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) CXIII: \( Z \perp U \mid W \) holds.

Then, \( \beta_o \), the average total causal effect of \( X \) on \( Y \), is fully identified as

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

Note the requirement that \( \ell = k \), analogous to the XI method. None of the methods discussed previously can identify \( \beta_o \) in this case, since none of the other admissible conditional independence relationships hold in \( S_{16a} \).

Section 7 provides straightforward conditions under which the plug-in CXIII estimator \( \hat{\beta}_n^{CXIII} \) is a consistent and asymptotically normal estimator for \( \beta_o \), where

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

This is a standard IV estimator with derived standard instruments \( Z - E(ZW')E(WW')^{-1} W \), the residuals from the regression of \( Z \) on \( W \). Nevertheless, we do not treat these residuals as the fundamental variables instrumental for the identification of \( \beta_o \) for reasons analogous to those discussed in connection with CXRII: it is \( Z \) and \( W \) that provide the natural causal explanation enabling the recovery of the effect of \( X \) on \( Y \); further, when A.2 is relaxed to permit non-separable structures, the derived instruments no longer play an essential role, whereas \( Z \) and \( W \) continue to play the instrumental role in identifying the causal effects of interest.

In \( S_{16a} \), \( Z \) satisfies the following causal properties that parallel CP:OXI and permit the identification of the effect of \( X \) on \( Y \) in a manner analogous to OXI.

(CP:OCXIII): 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 CXRII with conditioning instruments $W$

(ii) $Z$ indirectly causes $Y$, and the effect of $Z$ on $Y$ is identified via CXRII with conditioning instruments $W$

(iii) $Z$ causes $Y$ only via $X$

In contrast to CP:OXI, conditioning instruments $W$ are needed in CP:OCXII to ensure that the effect of $Z$ on $X$ and that of $Z$ on $Y$ are identified. Since $Z$ is observed, we refer to this case as observed conditionally exogenous instruments given conditioning instruments (OCXII). 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 PCXII

As in the case of XI and CXRII, we do not need to observe the true underlying cause; it suffices to observe a suitable proxy for it. In fact, this feature applies to all of the EIV methods that we discuss (see Theorem 5.1 and Corollary 5.2 below.) We illustrate this in $S_{16b}$ and associated causal graph $G_{16b}$.

Let $S_{16b}$ be given by:

1. $W^c = \alpha_w U_w$
2. $Z^c = \alpha_z U_z$
3. $X^c = \gamma Z + \alpha_x U_x$
4. $Y^c = X' \beta_o + U_y$

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 structural equation (4) and setting $\pi_o \equiv \gamma_x' \beta_o$, we have:

$Y^c = Z^c \pi_o + X^c \alpha_x' \beta_o + U_y$

Here again CXII holds and Proposition 4.2.1 applies to fully identify $\beta_o$ as for OCXII. However, in $S_{16b}$ the effect of $Z$ on $X$ and that of $Z$ on $Y$ are no longer identified and CP:OCXIII no longer holds. Instead, $U_z$ satisfies these properties and $Z$ plays the role of a proxy for the unobservables $U_z$. Parallel to PXI, we refer to $Z$ in this case as proxies for (unobserved) conditionally exogenous instruments given conditioning instruments (PCXIII). The causal properties that permit the identification of $\beta_o$ in this case are:

(CP:PCXIII) 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 CXRII 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 CXRII 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 its counterpart CP:PXI, but conditioning instruments $W$ are now needed for the effect of $U_z$ on $X$ and that of $U_z$ on $Y$ to be identified had $U_z$ been observed. Our comments about PXI fully apply here. In particular, the effect of $X$ on $Y$ can be represented as the “ratio” of the full effect of $U_z$ on $Y$ to the full effect of $U_z$ on $X$. But the effect of $Z$ on $Y$ and that of $Z$ on $X$ are not identified in $S_{16b}$ as they are in the OCXI|I case, even after employing conditioning instruments $W$; the resulting CXRII estimators are inconsistent. As before, these effects are confounded by the same variables, $U_z$, in just the right way to leave the ratio of the two CXRII estimators informative for the effect of interest, $\beta_o$. The PCXI|I case is thus another example in which a function of two inconsistent estimators, the CXRII estimators of $\chi'$ and $\pi_o$ from structural equations (3) and (4'), is itself a consistent estimator for the effect of interest, $\beta_o$. As in the PXI case, we may have $\chi$ equal to zero, so that in $S_{16b}$, $Z$ is not required to cause $X$. When $\chi = 0$, $Z$ acts as a “pure predictive proxy” for $U_z$.

4.2.2 Conditional and Conditioning Instruments: CXRII

It can happen that the conditioning instruments $W$ render only a subvector $X_2$ of $X \equiv [X_1', X_2']'$, conditionally exogenous. In this case, the methods discussed so far cannot structurally identify $\beta_o \equiv [\beta_1', \beta_2']'$. Nevertheless, $\beta_o$ can be structurally identified if there is another vector of conditional instruments $Z$ for $X_1$ that is conditionally exogenous given $W$. This situation is illustrated in $S_{17}$:

Let $S_{17}$ be given by:

1. $W = \alpha_w U_w$
2. $Z = \alpha_z U_z$
3. $X_1 = \gamma_{x_1} Z + \alpha_{x_1} U_{x_1}$
4. $X_2 = \alpha_{x_2} U_{x_2}$
5. $Y = X_1' \beta_1 + X_2' \beta_2 + U_y$

where $U_{x_1} \perp U_{x_2}$, $U_{x_1} \perp U_y$, $U_{x_2} \perp U_y$, $U_{x_1} \perp U_z$, $U_{x_2} \perp U_z$, $U_{x_1} \perp U_{w}$, $U_{x_2} \perp U_{w}$, $U_y \perp U_z$, $U_y \perp U_{w}$, and $U_z \perp U_{w}$.
In $S_{17}$, conditioning on $W$ alone does not permit the identification of $\beta_o$ as in Proposition 3.3.1 since $X_1$ is not conditionally exogenous after conditioning on $W$. Similarly, the use of $Z$ alone does not permit the identification of $\beta_o$ since $Z$ is endogenous. However, the joint use of $Z$ and $W$ permits the identification and consistent estimation of $\beta_o$ since conditioning on $W$ when it is a sufficiently good predictor for $U_w$ renders $X_2$ conditionally exogenous and $Z$ a vector of conditionally exogenous instruments for $X_1$. The key conditional independence relationship that holds in $S_{17}$ is:

(CXIRII) Conditionally Exogenous Instruments and Regressors given Conditioning Instruments: $(Z, X_2) \perp U_y \mid W$

In fact, CXIRII is the special case of CXIII in which $X_2$ plays the role of a conditionally exogenous instrument for itself. We refer to $\tilde{Z} = [Z', X_2']'$ as a vector of conditionally exogenous instruments and regressors given conditioning instruments and refer to the corresponding EIV method as the CXIRII method. The key moment condition in the linear separable case is given by:

$$E(\tilde{Z} U_y \mid W) = E(\tilde{Z} \mid W) \times E(U_y \mid W)$$  \hspace{1cm} (M6)

Proposition 4.2.2 establishes that $\beta_o$ is fully identified when such a $Z$ and $W$ are available and stochastic identification holds. The condition $\ell = k_1$ is necessary for stochastic identification.

**Proposition 4.2.2** Suppose A.1 and A.2 hold such that: (i) $Y = X_1' \beta_1 + X_2' \beta_2 + U_y$, 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 and that $\ell = k_1$; 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) CXIRII: $(Z, X_2) \perp U_y \mid W$ holds.

Then $\beta_o$, the average total causal effect of $X$ on $Y$, is fully identified as

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

Under mild conditions provided in Section 7, the CXIRII plug-in estimator

$$\hat{\beta}_n^{\text{CXIRI}} = [\tilde{Z}' (I - W(W'W)^{-1}W')X]^{-1} [\tilde{Z}' (I - W(W'W)^{-1}W')Y]$$

is a consistent and asymptotically normal estimator for $\beta_o$.

**4.2.3 Conditional and Conditioning Instruments: CXIRI**

A generalization of the CXIR method occurs when CXIR fails but conditioning instruments $W$, together with regressors $X$, render the extended instruments $Z$ conditionally exogenous, as in $S_{17}$. 

41
Let $S_{18}$ be given by:

1. $W = \alpha_w U_w$
2. $X = \alpha_x U_x$
3. $Z = \gamma_z X + \alpha_z U_z$
4. $Y = Z' \delta_o + U_y$

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 structural equation (3) into structural equation (4) with $\beta_o \equiv \gamma_z' \delta_o$, we have:

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

The key conditional independence relationship that holds in $S_{18}$ is:

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

CXII|RI is weaker than CXII|R, its counterpart from Section 4.1. We refer to $Z$ as a vector of *conditionally exogenous instruments given regressors and conditioning instruments* and refer to this EIV method as the CXII|RI method. Now $X$ and $W$ jointly play the role of conditioning instruments for $Z$. This of course requires that $W$ is a sufficiently good predictor for $U_w$.

Given linearity and separability, the key moment condition that results from CXII|RI is:

$$E(ZU_j \mid \tilde{W}) = E(Z \mid \tilde{W}) \times E(U_j \mid \tilde{W}), \quad (M7)$$

where $\tilde{W} = [X', W']'$. Similar to the CXII|R method, the CXII|RI method identifies $\beta_o$ as the product of the effect of $X$ on $Z$ and that of $Z$ on $Y$.

**Proposition 4.2.3** Suppose A.1 and A.2 hold such that: (i) $Z = \gamma X + \alpha U_z$, $Y = Z' \delta_o + U_y$, 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 $\tilde{W} = [X', W']'$, $E(ZW')$, $E(\tilde{W} \mid \tilde{W})$, and $E(\tilde{W} \mid \tilde{W})$ exist and are finite; $E(\tilde{W} \mid \tilde{W})$ is non-singular and $E(Z \mid \tilde{W}) = E(Z \tilde{W}')(E(\tilde{W} \mid \tilde{W})^{-1} \tilde{W}$; (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: $X \perp U_z$ holds and (b) CXII|R: $Z \perp U_y \mid \tilde{W}$ holds.

Then $\beta_o \equiv \gamma' \delta_o$, the average total causal effect of $X$ on $Y$, is fully identified as:
\[
\beta_0 = [E(XX')]^{-1}E(XZ') \times \\
\{E(ZZ') - E(Z\tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} Z')\}^{-1}\{E(ZY) - E(Z\tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} Y)\}
\]

Here again, none of the previous EIV methods identify \(\beta_0\), since their required conditional independence relationships do not hold.

A key feature of \(S_{18}\) is that \(X \perp U_z\). It should now be clear that this independence relationship could in turn be relaxed to a conditional independence relationship, such as

\[
\text{CXR|I: } X \perp U_z | W_1,
\]

with \(W_1\) a suitable vector of conditioning instruments. This follows because just as for \(\text{CXI|R}\), the effect of \(X\) on \(Y\) is identified as the product of the effects of \(X\) on \(Z\) and of \(Z\) on \(Y\), given the conditional independence relationships and causal structures ensuring that these two effects are identified, and provided that the effect of \(X\) on \(Y\) is fully mediated by \(Z\). Pearl (1995, 2000) provides graphical criteria for structural identification of effects of interest to obtain in such a manner via his “front door” method.

Section 7 gives straightforward conditions under which the plug-in estimator \(\hat{\beta}_{\text{CXIIRI}}\) is a consistent and asymptotically normal estimator for \(\beta_0\), where:

\[
\hat{\beta}^\text{CXIIRI}_n \equiv [(X'X)^{-1}(X'Z) \times [Z'(I - \tilde{W}'(\tilde{W}'\tilde{W})^{-1}\tilde{W}')]^{-1}[Z'(I - \tilde{W}'(\tilde{W}'\tilde{W})^{-1}\tilde{W}')Y]
\]

\[= \hat{\beta}_{\text{XR}}^n \cdot \hat{\delta}^\text{CXIIRI}_n.
\]

In writing this last expression, we engage in a slight abuse of notation, as we do not make explicit the use of \(\tilde{W}\) in the \(\text{CXRII}\) estimator \(\hat{\delta}^\text{CXIIRI}_n\).

### 4.2.4 Further Comments on Double Extended Instrumental Variables

Like the single EIV case, conditions of the form \(Z \perp U_y \mid (Y, W)\) do not permit structural identification of \(\beta_0\). To see this, let \(\tilde{W} \equiv [W', Y']'\), and proceed as in Section 4.1.3. This yields the expression

\[
\{E(ZY) - E(Z\tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} Y)\} - \{E(ZX') - E(Z\tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} X')\} \beta_0
\]

\[= 0.
\]

The first term on the right above is identically zero, however, as it represents the covariance between \(Y\) and the residuals of the regression of \(Z\) on \(Y\) and \(W\). Just as in Section 4.1.3, this provides no useful information for structurally identifying \(\beta_0\).

The double EIV methods \(\text{CXIII}, \text{CXIIIR}, \text{and CXRII}\), 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 \(\text{XR, XI, CXRII, CXIII, and CXRIII}\) constitute an exhaustive set of
“primitive” methods, since other EIV methods, such as CXIIIR and CXIIIIR, identify causal effects as functions of effects identified by use of one or more of these primitives.

5. A Master Theorem for EIV Identification

The results of Sections 3 and 4 provide a variety of means for identifying the effects of potentially endogenous causes on the response of interest via the use of standard or extended instrumental variables. In this section we summarize these results by stating a “master theorem” that either contains our previous identification results as special cases, as for XI or CXRII, or delivers them as immediate corollaries, as for CXIIIR or CXIIIRI. Our master theorem 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) $Y = X'\beta_o + U$, where $X$ is $k \times 1$, $k > 0$, and $\beta_o$ 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 $[\tilde{Z}', \tilde{W}'] = A [X', Z', W']'$, for a given $(k + \tilde{m}) \times (k + \ell + m)$ matrix $A$, and that $E(\tilde{Z} X')$, $E(\tilde{Z} \tilde{W}')$, $E(\tilde{W} X')$, $E(\tilde{Z} Y)$, $E(\tilde{W} Y)$ exist and are finite; if $\tilde{m} > 0$, suppose $E(\tilde{W} \tilde{W}')$ is non-singular and that $E(\tilde{Z} | \tilde{W}) = E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1} \tilde{W}$. If $\tilde{m} = 0$, put $[E(\tilde{W} \tilde{W}')]^{-1} = 0$. 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 $\beta^*$ such that

$$\{E(\tilde{Z} X') \quad E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1} E(\tilde{W} X')\} \beta^* = \{E(\tilde{Z} Y) \quad E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1} E(\tilde{W} Y)\} = E\{[\tilde{Z} - E(\tilde{Z} | \tilde{W})]U_y\}$$

if and only if $\{E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1} E(\tilde{W} X')\}$ is non-singular.

(c) Structural identification holds, that is $\beta_o$ satisfies

$$\{E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1} E(\tilde{W} X')\} \beta_o - \{E(\tilde{Z} Y) - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1} E(\tilde{W} Y)\} = 0$$

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

(d) The average causal effect $\beta_o$ is fully identified as

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

if and only if stochastic and structural identification jointly hold. ■
If stochastic identification holds but not structural identification, then we have

$$\beta^* = \beta_o + \{E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} X')\}^{-1} \{E\{\tilde{Z} - E(\tilde{Z} | \tilde{W})\}U_y\}. \tag{3}$$

This expresses the probability limit \( \beta^* \) of the plug-in EIV estimator

$$\hat{\beta}_{EIV}^n \equiv [\tilde{Z}' (I - \tilde{W} (\tilde{W}' \tilde{W}^{-1})\tilde{W}')X]^{1/2} [\tilde{Z}' (I - \tilde{W} (\tilde{W}' \tilde{W}^{-1})\tilde{W}')Y]$$

as the true average causal effect, \( \beta_o \), plus a “causal discrepancy,”

$$\delta^* = \{E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} X')\}^{-1} \{E\{\tilde{Z} - E(\tilde{Z} | \tilde{W})\}U_y\}. \tag{4}$$

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

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

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

Theorem 5.1 contains as special cases the XR, CXRI, XI, CXIII, and CXRIII methods. For these cases, an exclusion restriction acts to ensure that when \( Z \) is present, \( Z \) causes \( Y \) only via \( X \). It is straightforward to verify that conditional independence (\( \tilde{Z} \perp U_y | \tilde{W} \)), conditional mean independence (\( E(U_y | \tilde{Z}, \tilde{W}) = E(U_y | \tilde{W}) \)), and conditional non-correlation (\( E(\tilde{Z} U_y | \tilde{W}) = E(\tilde{Z} | \tilde{W}) E(U_y | \tilde{W}) \)) each imply the necessary structural identification condition \( E\{\tilde{Z} - E(\tilde{Z} | \tilde{W})\}U_y\} = 0 \). We refer to \( \tilde{Z} = E(\tilde{Z} | \tilde{W}) \) as derived standard instruments since they satisfy this moment condition. Thus, the causal structure determines the availability of extended instrumental variables. These, in turn dictate the form of derived standard instruments that satisfy moment conditions supporting estimation of identified causal effects.

The next Corollary provides an extension of Theorem 5.1 to cover cases such as CXIIIR and CIXIIRI, where direct identification of causal effects of interest is not possible but obtains instead as a function of identifiable effects as in Theorem 5.1.

**Corollary 5.2** Suppose A.1 and A.2 hold for a structural system \( S \) such that \( Y = X' \beta_o + U_y \), where \( X \) is \( k \times 1 \), \( k > 0 \), and \( \beta_o \) 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, \( \beta_o \) is fully identified as \( b(\theta_1, \ldots, \theta_H) \). \( \blacksquare \)

This covers the intermediate cause instrument case, in which an exclusion restriction acts to ensure that \( X \) causes \( Y \) only via \( Z \). 

---

The next Corollary provides an extension of Theorem 5.1 to cover cases such as CXIIIR and CIXIIRI, where direct identification of causal effects of interest is not possible but obtains instead as a function of identifiable effects as in Theorem 5.1.
6. Characterization of Structural Identification via Causal Matrices: An Illustration with Single EIV

Causal matrices are a powerful way to characterize the causal structures in which the identification of given causal effects of interest obtains. In particular, in other work (in progress), we provide a procedure to generate conditional independence matrices from causal matrices. These matrices characterize the conditional independence relationships that hold among the variables of given system $S$, conditioning on a given subset of variables in system $S$. (For the empty set, this gives the independence matrix.) Thus, given a causal matrix $C_S$, by inspecting the associated conditional independence matrices it is straightforward to determine whether the necessary exogeneity or conditional exogeneity relationships hold for the identification of given causal effects of interest.

Every causal matrix $C_S$ also has an associated path matrix $P_S$. The $k^{th}$ row and $l^{th}$ column entry of $P_S$, $p_{kl}$, takes the value 1 if there is a $(V_k, V_l)$-path in $G_S$ and equals 0 otherwise. Hence $P_S$ summarizes all direct and indirect causal relationships between the variables of $S$. The matrices $C_S$ and $P_S$ are related by the operation $P_S = p(C_S)$ where :

$$p_{kl} = 1 \quad \text{if there exists } h > 0 \text{ and a set } \{g_1, \ldots, g_h\} \text{ with elements in } \{1, \ldots, G\} \text{ such that } c_{k_{g_1}} \times \ldots \times c_{g_h} = 1;$$

$$p_{kl} = c_{kl} \quad \text{otherwise.}$$

Thus, $p$ maps an entry $c_{kl} = 1$ in $C_S$ to an entry $p_{kl} = 1$ in $P_S$, and changes an entry $c_{kl} = 0$ in $C_S$ to an entry $p_{kl} = 1$ in $P_S$ if and only if $V_k$ does not directly cause $V_l$ but there exists a sequence of intermediate variables that mediate an effect of $V_k$ on $V_l$. These path matrices, in conjunction with their corresponding causal matrices, express concisely the exclusion restrictions necessary for the identification of causal effects.

By examining the conditional independence and path matrices, one can determine whether structural identification of given causal effects of interest obtains. We describe this for the case of the single extended instrumental variable $Z$ and the single cause and response variables $X$ and $Y$. In work in progress, we discuss the identification of causal effects via causal matrices more generally for single and double EIV methods.

Under A.1 and A.2, the causal matrix for the single EIV case has the form

$$C_S = \begin{bmatrix}
C_{S_1} & C_{S_2} \\
C_{S_1} & C_{S_2}
\end{bmatrix} = \begin{bmatrix}
X & Y & Z & U_x & U_y & U_z \\
X & 0 & 0 & 0 & 0 & 0 \\
Y & 0 & 0 & 0 & 0 & 0 \\
Z & 0 & 0 & 0 & 0 & 0 \\
U_x & 1 & 0 & 0 & 0 & 0 \\
U_y & 0 & 1 & 0 & 0 & 0 \\
U_z & 0 & 0 & 1 & 0 & 0
\end{bmatrix}$$
The specified entries in the off-diagonal blocks follow by our conventions, as do the diagonal elements. We also have \( c_{21} = 0 \) by the acyclicity assumption and the fact the effect of interest is that of \( X \) on \( Y \). Further, the assumed acyclicity of \( S \) imposes on \( C_{S_i} \) 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 \( j, k, l = 1, 2, 3 \) as well as three constraints of the form \( c_{jk} \times c_{kj} = 0 \) and two constraints of the form \( c_{jk} \times c_{kl} \times c_{lj} = 0 \) on \( C_{S_4} \) for \( j, k, l = 4, 5, 6 \).

Under A1, \( C_{S_i} \) admits 9 possible values that we label in relation to \( Z \) as illustrated by the graphs in Table I.

<table>
<thead>
<tr>
<th>Non Causal, Joint Cause, and Joint Response</th>
<th>Pre-Cause</th>
<th>Intermediate Cause</th>
<th>Post-Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

Table I displays all possible acyclic causal structures that can relate \( X, Y \), and a single extended instrument \( Z \). These include the pre-cause, intermediate cause, and post-response instrument cases previously discussed. The (1, 1) entry of Table I, the “non-causal” case depicts the potentially valid common cause instrument case, provided that appropriate causal relationships hold among the unobserved variables. Other structures not obeying the exclusion restrictions for identification appear in the second column of the second, third and fourth rows of Table I.

Table I also displays the joint cause case where both \( X \) and \( Z \) cause \( Y \) as shown in the (1, 2) entry of Table I and the joint response case where \( X \) causes both \( Y \) and \( Z \) as shown in the (1, 3) entry of Table I.

Inspection of \( C_{S_i} \) reveals that for every entry of Table I, there are 25 possible acyclic causal structures that can relate the corresponding unobserved variables. Thus in total, \( C_S \)
can represent 225 (25×9) potential acyclic causal structures in the single EIV case. The analysis can be simplified by restricting attention to the presence or absence of statistical independence among the unobserved terms, as is standard practice in the structural equations literature. The 25 possible acyclic causal structures among the unobserved variables simplify to 8 possible sets of independence/dependence relationships among $U_x$, $U_y$ and $U_z$. Together with the 9 possible entries of Table I, these 8 sets generate 72 possible structural equations systems.

As can be verified, 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 $Y$ obtains in the single EIV case. Specifically, under our assumptions, 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 appropriate exclusion restrictions, exhaustively characterize all the acyclic causal structures in which the structural identification of the effect of $X$ on $Y$ is possible in the single EIV case. These are precisely the cases presented in Sections 3.1, 3.2, and 4.1.

For example, observe that 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, the identification of the effect of $X$ on $Y$ is not possible in these cases, even when the appropriate exogeneity or conditional exogeneity conditions hold.

7. Asymptotic Properties of EIV Estimators

Plug-in EIV estimators for causal coefficients identified by Theorem 5.1 have the form

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

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

**Theorem 7.1** Suppose the conditions of Theorem 5.1 hold and that

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

(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 $n^{1/2} (\hat{\beta}_n^{EIV} - \beta_0) \xrightarrow{d} N(0, Q^{-1} V Q^{-1})$. ■

Plug-in EIV estimators for average causal effects identified by Corollary 5.2 have the form

$$\hat{\beta}_n^{EIV} = b(\hat{\beta}_n^{EIV}),$$
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
\[
\hat{\theta}_{n}^{EIV} = \left[ \tilde{Z}_{h}(I - \tilde{W}_{h}(\tilde{W}_{h}'\tilde{W}_{h})^{-1}\tilde{W}_{h}')X_{h} \right]'[\tilde{Z}_{h}'(I - \tilde{W}_{h}(\tilde{W}_{h}'\tilde{W}_{h})^{-1}\tilde{W}_{h}')Y_{h}] \quad h = 1, \ldots, H,
\]
\[
\zeta_{h,i} \equiv [\tilde{Z}_{h,i} - E(\tilde{Z}_{h,i}|\tilde{W}_{h,j})]U_{y_{h,i}} \quad i = 1, \ldots, n; \; h = 1, \ldots, H,
\]
and put \( \zeta_{i} \equiv (\zeta_{1,i}', \ldots, \zeta_{H,i}')' \).

**Theorem 7.2** Suppose the conditions of Corollary 5.2 hold with \( \theta_{0} \equiv (\theta_{1}', \ldots, \theta_{H}')' \), and suppose further that
\[
(i) \quad \tilde{Z}_{h}'(I - \tilde{W}_{h}(\tilde{W}_{h}'\tilde{W}_{h})^{-1}\tilde{W}_{h}')X_{h} / n \xrightarrow{p} Q_{h} \equiv E(\tilde{Z}_{h}X_{h}') - E(\tilde{Z}_{h}\tilde{W}_{h}')[E(\tilde{W}_{h}\tilde{W}_{h}')]^{-1}E(\tilde{W}_{h}X_{h}'), \; h = 1, \ldots, H;
\]
\[
(ii) \quad n^{-1/2} \sum_{i=1}^{n} \zeta_{i} \xrightarrow{d} N(0, V), \text{ where } V \text{ is finite and positive definite.}
\]
Then \( n^{1/2} (\hat{\theta}_{n}^{EIV} - \theta_{0}) \xrightarrow{d} N(0, Q^{-1}VQ^{-1}) \), where \( Q = \text{diag}(Q_{1}, \ldots, Q_{H}) \).

Suppose further that \( b \) is continuously differentiable at \( \theta_{0} \) such that \( \nabla b(\theta_{0}) \) (the gradient of \( b \) at \( \theta_{0} \)) has full column rank. Then with \( \hat{\beta}_{n}^{EIV} \equiv b(\hat{\theta}_{n}^{EIV}) \) and \( \beta_{0} \equiv b(\theta_{0}) \),
\[
\begin{align*}
    n^{1/2} (\hat{\beta}_{n}^{EIV} - \beta_{0}) & \xrightarrow{d} N(0, \nabla b(\theta_{0})'Q^{-1}VQ^{-1}\nabla b(\theta_{0})).
\end{align*}
\]

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 somewhat similar to the way in which optimal instruments are chosen in the standard IV framework. Just as GLS-like corrections for conditional heteroskedasticity may be involved in obtaining the optimal instruments in the standard case, such corrections will also play a role in the EIV case. We leave the analysis of the choice of optimal EIV to subsequent research.

**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 in observational studies with the aid of extended instrumental
variables (EIV). EIV methods make use of variables that are not necessarily “valid” instrumental variables in the traditional sense, but that emerge from a given causal structure to enable the recovery of causal effects of interest. In particular, 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 EIV permits the identification of causal effects of potentially endogenous causes on the response of interest. In particular, we analyze the exogenous regressors (XR), exogenous instruments (XI), conditionally exogenous regressors given conditioning instruments (CXRI), and conditionally exogenous instruments given regressors (CXRIR) methods. In the XI method, we discuss and provide a causal explanation for two subcategories: the observed exogenous instruments (OXI) and the proxies for unobserved exogenous instruments (PXI), thereby extending previous causal interpretations of IV methods, such as that of Angrist, Imbens and Rubin (1996.) Our framework also explains the failure of the XI method in the standard irrelevant, invalid, and under-identified cases. In the double EIV case, we demonstrate how the joint use of conditional and conditioning EIV permits the identification of causal effects of interest. In particular, we analyze the conditionally exogenous instruments given conditioning instruments (CXIRI), the conditionally exogenous instruments and regressors given conditioning instruments (CXIRII), and the conditionally exogenous instruments given regressors and conditioning instruments (CXRIRI) methods. We state a master theorem giving necessary and sufficient conditions for the identification of causal effects by means of extended instrumental variables methods and provide straightforward high-level conditions ensuring consistency and asymptotic normality for EIV plug-in estimators of the effects of interest.

By making use of causal matrices, path matrices, and conditional independence matrices it is possible to characterize the cases where the structural identification of causal effects of interest obtains. We illustrate such procedures in the single EIV case, demonstrating that the XR, XI, CXRI, and CXIRI methods exhaust the single EIV methods capable of structurally identifying causal effects. Work in progress analyzes a procedure for generating conditional independence matrices from causal matrices and establishing identification results for EIV methods more generally.

Here, we consider the identification of causal effects given causal structures specified a priori. In that same work in progress, we provide procedures for generating the class of causal matrices that are in agreement with a collection of given (observed) conditional independence matrices. This yields methods for suggesting or ruling out potential causal structures. We propose methods for causal inference based on those results and the identification results provided here.

Future work will analyze the asymptotically efficient choice of 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) provide several tests for conditional exogeneity. Other work in progress extends these tests and proposes 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.

Mathematical Appendix

**Proof of Proposition 3.1.1** From (iii), \( E(U_y) = 0 \). From (i), \( U_y = Y - X' \beta_o \). Substituting this into \( E(U_y) = 0 \) gives \( E(XY) - E(XX') \beta_o = 0 \). From (ii), \( E(XX') \) is non-singular. Thus \( \beta_o \) 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), \( E(U_y | W) = E(X | W) E(U_y | W) \). Equivalently,
\[
E([X - E(X | W)] U_y | W) = 0.
\]

From (ii), \( E(WW') \) is non-singular and \( E(X | W) = E(XW') [E(WW')]^{-1} W \), so
\[
E( [X - E(XW')[E(WW')]^{-1} W] U_y | W) = 0,
\]

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 \( \beta_o \) is fully identified as
\[
\beta_o = {E(XX') - E(XW')[E(WW')]^{-1} E(WX')}^{-1} {E(XY) - E(XW')[E(WW')]^{-1} E(WY)}
\]

**Proof of Proposition 4.1.2** From (iii)(a), \( E(U_z) = 0 \). From (i), \( \alpha_z U_z = Z - \gamma X \), and from (ii)(a) \( E(XX') \) is non-singular. Proposition 3.1.1 thus ensures that \( \gamma \) is fully identified as \( \gamma = [E(XX')]^{-1} E(XZ') \). Similarly, from (iii)(b), \( E(ZU_y | X) = E(Z | X) \times E(U_y | X) \); 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 \( E(Z | X) = E(ZX')[E(XX')]^{-1} X \), \( \delta_o \) 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 $\beta_o \equiv \gamma \delta_o$, $\beta_o$ is thus fully identified as:

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

**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 $Z'$. ■

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

**Proof of Theorem 5.1:**

(a) From (i) and (ii),

$$E[\{\tilde{Z} - E(\tilde{Z} | \tilde{W})\}U_Y]$$

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

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

$$- E(\tilde{Z} X') \beta_o + E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} X') \beta_o$$

Since $\beta_o$ is finite and $E(\tilde{Z} Y), E(\tilde{Z} \tilde{W}'), [E(\tilde{W} \tilde{W}')]^{-1}, E(\tilde{W} Y), E(\tilde{Z} X'),$ and $E(\tilde{W} X')$ exist and are finite, it follows that $E[\{\tilde{Z} - E(\tilde{Z} | \tilde{W})\}U_Y]$ exists and is finite.

(b) Consider the system of equations

$$\{E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} X')\} \beta$$

$$- \{E(\tilde{Z} Y) - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} Y)\} = E[\{\tilde{Z} - E(\tilde{Z} | \tilde{W})\}U_Y].$$

It is well known that this system admits a unique solution $\beta^*$ if and only if $\{E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} X')\}$ is non-singular.

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

(d) If stochastic and structural identification hold, we have that $\{E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} X')\}$ is non-singular and

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

$$- \{E(\tilde{Z} Y) - E(\tilde{Z} \tilde{W}')[E(\tilde{W} \tilde{W}')]^{-1}E(\tilde{W} Y)\} = 0.$$
It follows that $\beta_o$ is then fully identified as

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

To establish the converse, suppose that either stochastic or structural identification fails. If stochastic identification fails, then the inverse of $E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}') [E(\tilde{W} \tilde{W}')]^{-1} E(\tilde{W} X')$ does not exist, so $\beta_o$ cannot have the form given above. If structural identification fails, then $E[[\tilde{Z} - E(\tilde{Z} | \tilde{W})] U_y]$ is not zero. By (a), $\beta_o$ satisfies

$$
E[[\tilde{Z} - E(\tilde{Z} | \tilde{W})] U_y] \\
= E(\tilde{Z} Y) - E(\tilde{Z} \tilde{W}') [E(\tilde{W} \tilde{W}')]^{-1} E(\tilde{W} Y) \\
\quad - \left\{ E(\tilde{Z} X') - E(\tilde{Z} \tilde{W}') [E(\tilde{W} \tilde{W}')]^{-1} E(\tilde{W} X') \right\} \beta_o.
$$

But this is incompatible with the form given 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). ■

**References**


