A Note on Testing Exogeneity of Instrumental Variables

(DRAFT PAPER)

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

1 Introduction

It is common in the literature on instrumental variables to remark upon the difficulty of knowing or demonstrating that a potential instrument is exogenous, in the sense of being uncorrelated with the disturbances [Bartels, 1991, Johnston, 1972]. It is also widely recognized that exogeneity is an assumption embedded in the model specification [Engle, et al, 1984], hence, it rests on subjective judgment and, like other structural assumptions of causation and "zero-restrictions", it cannot be tested in purely observational studies. The purpose of this note is to show that despite its elusive nature, exogeneity can nevertheless be given some empirical test. The test is not guaranteed to detect all violations of exogeneity but it can, in certain circumstances, screen away real bad choices of would-be instruments.

2 An Instrumental Inequality

Definition 2.1 *(exogeneity) A variable z is said to be exogenous relative to an ordered pair of variables (x, y) if the relation between x, y and z is governed by the following two equations:*

$$\begin{aligned} x &= f_1(z, v) \\ y &= f_2(x, u) \end{aligned} \tag{1}$$

with the restriction that u and z are mutually independent. f_1, f_2 are arbitrary deterministic functions, and u and v represent unobserved, possibly correlated disturbances.

The definition above captures the two main features associated with exogeneity; externality and locality. The required independence between z and u rules out the possibility that z be influenced by some latent cause which also influences y. Thus, in this sense z acts as an instrument external to the system ("randomized", in statistical terminology). The same independence also rules out z having any direct effect on y, unmediated by x, thus capturing the notion of locality, whereby an external intervention is presumed to "affect x *only*". Note that no restrictions are posed on the domains of u and v; each may be finite or unbounded, discrete or continuous, ordered or unstructured.

Our problem is to determine, from the observed joint probability distribution $P(x, y, z)$ whether z can be exogenous relative to (x, y), that is, whether there exist two functions f_1 and f_2 and a probability distribution on u, v, and z (with z and u independent), such that the distribution generated by the two equations corresponds precisely to the observed distribution $P(x, y, z)$.

Theorem 2.2 *If the following inequality holds:*

$$\max_x \sum_y [\max_z P(x, y|z)] > 1 \tag{2}$$

then z is non-exogenous relative to (x, y). Otherwise, z may or may not be exogenous.

Proof: If the probability distribution $P(x, y, z)$ is generated by the process defined in Eq. (1), then it can be expressed in the form

$$P(x, y, z) = \sum_u \sum_v P(y|x, u)P(x|z, v)P(v|z, u)P(u)P(z)$$

This can be seen by decomposing $P(x, y, z, u, v)$ into product form along the order (y, x, v, u, z), and using the independence relations imposed by the model of Eq. (1). Therefore,

$$\begin{aligned} P(x, y|z) &= \textstyle\sum_v \sum_u P(y|x, u)P(x|z, v)P(u)P(v|z, u) \\ &= E_u[P(y|x, u)g(x, z, u) \end{aligned} \tag{3}$$

where

$$g(x, z, u) = \sum_v P(x|z, v)P(v|z, u) \tag{4}$$

We now choose an arbitrary function $F : dom(X) \times dom(Y) \rightarrow dom(Z)$, replace z by $F(x, y)$ and sum Eq. (3) over y.

$$\sum_y P(x, y|F(y, x)) = E_u \sum_y P(y|x, u)g(x, F(x, y), u) \tag{5}$$

For any fixed value of x, y, and u, Eq. (4) gives

$$0 \leq g(x, F(x, y), u) \leq 1 \tag{6}$$

The r.h.s. of Eq. (5) represents a convex combination of such g terms and should, likewise, satisfy

$$0 \leq E_u \sum_y P(y|x, u)g(x, F(x, y), u) \leq 1 \tag{7}$$

thus constraining the l.h.s. of Eq. (5) to

$$\sum_y P(x, y|F(x, y)) \leq 1 \tag{8}$$

Since this inequality holds for any choice of F, we might as well choose a z that maximizes the value of P in each term, which yields

$$\sum_y \max_z P(x, y|z) \leq 1 \tag{9}$$

Moreover, since this inequality must hold for every x, we can write

$$\max_x \sum_y [\max_z P(x, y|z)] \leq 1 \tag{10}$$

which proves the theorem. $\Box$

We call the inequality above an *Instrumental Inequality* because it constitutes a necessary condition for any instrumental variable z to qualify as exogenous relative to (x, y). In the same fashion, if x, y, and z are continuous variables characterized by a density function $f(x, y, z)$ we get

$$\max_i \sum_j \max_z P(i, j|z) \leq 1 \tag{11}$$

where

$$P(i, j|z) = \int_{y \in Y_j} \int_{x \in X_i} f(x, y|z) dx\ dy \tag{12}$$

and $\{Y_j\}, \{X_i\}$ are any partitions of the domains of y and x, respectively.

It is interesting to note that any tri-variate normal distribution satisfies the Instrumental Inequality, regardless of whether it was actually generated by a process defined in Eq. (1). This can be seen from the fact that for any correlation parameters R_{xy} R_{yz} and $R_{zx} > 0$, we can always find a unique solution to the structural coefficients a and b in the equations

$$\begin{aligned} x &= az + cu \\ y &= bx + u \end{aligned} \tag{13}$$

(the linear version of Eq. (1)) with z uncorrelated with u. In particular, this solution yields the celebrated "instrumental-variable" estimator

$$\hat{b} = R_{zy}/R_{zx} \tag{14}$$

for the parameter b in Eq. (13), which is known to be consistent only when z is truly uncorrelated with u [Bartels, 1991]. Thus, the Instrumental Inequality cannot weed out bad instruments z if all measured variables are normally distributed.

The effectiveness of the Instrumental Inequality is realized in non-normal distributions, especially over discrete variables. For example, if all observed variables are binary we obtain the inequalities

$$\begin{aligned} P(y = 0, x = 0|z = 0) + P(y = 1, x = 0|z = 1) \leq 1 \\ P(y = 0, x = 1|z = 0) + P(y = 1, x = 1|z = 1) \leq 1 \end{aligned} \tag{15}$$

which were derived in the analysis of non-compliance in experimental studies[Pearl, 1993].

3 Remarks

We see that the Instrumental Inequality is violated when the controlling instrument z manages to produce significant changes in the response variable y while the regressor x remains constant. Although such changes could in principle be explained by strong correlation between u and v (since x does not screen off z), the Instrumental Inequality sets a limit on the magnitude of the changes. The similarity to Bell's Inequality in quantum physics [Cushing & McMullin, 1989, Suppes, 1988] is not accidental; both inequalities delineate a class of observed correlations that cannot be explained by hypothesizing latent common causes (in our formulation, common causes are modeled by setting $u = v$). The Instrumental Inequality can, in fact, be viewed as a generalization of Bell's Inequality for cases where direct causal connection is permitted to operate between the correlated observables x and y.

Of special interest to experimenters is the prospect of applying the Instrumental Inequality for the detection of undesirable side-effects in clinical trials. In such trials dependencies between the treatment assignment (z) and factors (u) affecting the response process (y) can be attributed to one of two possibilities; either there is a direct causal effect of the assignment (z) on the response (y), unmediated by the treatment (x), or there is a common causal factor correlating the two. If the assignment is carefully randomized, then the latter possibility is ruled out and any violation of the Instrumental Inequality (even under conditions of imperfect compliance) can safely be attributed to some direct influence the assignment process has on subjects' response.

Alternatively, if one can eliminate direct effects of z on y, say through an effective use of placebo, then any observed violation of the Instrumental Inequality can safely be attributed to spurious correlation between z and u, namely, to selection bias.

References

[Bartels, 1991] Bartels, L.M. (1991) "Instrumental and 'Quasi-Instrumental' Variables." *American Journal of Political Science*, **35**, 777-800.

[Cushing & McMullin, 1989] Cushing, J.T. and McMullin, E. (Eds.). (1989) *Philosophical Consequences of Quantum Theory*. University of Notre Damm Press, Notre Damm, IA.

[Engle, et al, 1984] Engle, R.F., Hendry, D.F., and Richard, J.F. (1984) "Exogeneity." *Econometrica*, **51** (2), 277-304.

[Johnston, 1972] Johnston, J. (1972) *Econometric Methods*, (2nd ed.) Cambridge: MIT Press.

[Pearl, 1993] Pearl, J. (1993) "Aspects of Graphical Models Connected with Causality." In *Proceedings of the 49th Session of the International Statistical Institute*, Tome LV, Book1, Florence, Italy, 391-401.

[Suppes, 1988] Suppes, P. (1988) "Probabilistic Causality in Space and Time." In Skyrms, B. and Harper, W.L. (Eds.), *Causation, Chance, and Credence*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 135-151.