Exogeneity Revisited

Judea Pearl

University of California, Los Angeles

Computer Science Department

Los Angeles, CA, 90095-1596, USA

judea@cs.ucla.edu

July 21, 2010

Abstract

1 Exogeneity and Causal Language

In communicating with colleagues in econometrics, I am often asked how concepts based on classical econometric models fit into modern vocabulary of causal reasoning. One of the issues that is brought up in such discussions is the notion of exogeneity, which seems to have played a major role in the history of econometric thought and which received a formal, albeit problematic treatment in the classical paper of Engel, Hendry and Richard (EHR) (Engle et al., 1983).

Since EHR's paper is quoted in almost every advanced textbook in econometrics, it is natural to ask whether the EHR concept of exogeneity, especially its causal version called "super-exogeneity," corresponds to an analogous concept in the modern language of causal inference, particularly the language of causal diagrams, *do*-calculus, structural models, potential outcomes, and counterfactual logic (Pearl, 2009). The answer, of course, is Yes, because causal models, to earn their title, must give precise definition and characterization to any concept involving cause-effect relationships, not to mention notions such as exogeneity, which have captured sustained interest among econometricians for several decades.

There are problems however in translating the classical works of EHR into modern vocabulary, stemming from profound conceptual differences as to the role of models in econometric research, the causal interpretation of such models, and the probabilistic rather than causal vocabulary that was used to encode that interpretation.

As one colleague wrote to me recently:

> "I was struck by your observation that knowing the joint probability distribution is not enough to dis-entangle the causal structure. I do not think this is known to econometricians. Hendry's methodology makes the DGP (Data Generating Process) central to inference and assumes that it contains all information possible – it is the holy grail, as some critics have said."

Indeed, the DGP used in the influential writings of Hendry and EHR is defined as a parameterized set the joint distribution functions which contains, presumably, all information needed for economic analysis, including prediction, explanation and policy evaluation. It is hard to reconcile this view with modern understanding of causation, according to which joint distribution functions do not, and cannot, carry causal information, since probabilistic relations are but descriptive epi-phenomena of the underlying econometric model which, at its core, conveys counterfactual relationships (Haavelmo, 1943, 1944; Balke and Pearl, 1994; Heckman, 2005; Pearl, 2000).

Not surprisingly, modern textbooks in econometrics tend to be cryptic when discussing the implications of the EHR formulation and, while stressing the importance of the topic, they systematically evade the two core questions that readers wish to ask:

1. What would we be able to do, infer, or conclude if told that variable X is super-exogenous that we would not be able to do, infer, or conclude if told that X is NOT super-exogenous?

2. Suppose someone gives us a completely specified econometric model with all parameters given to 10 digits accuracy and nothing left to imagination; can one decide, given such a model, whether X is super-exogenous or not? If so, how?

Before writing *Causality* (Pearl, 2000), I posed these two questions to many econometricians and the answers received fell into two distinct categories:

1. The EHR paper is ambiguous and super-exogeneity no longer plays any role in econometrics.

2. The EHR paper is one of the most important papers in econometrics, but to answer your questions would take volumes to detail; it depends on many conditions not all are encoded in the model.

In short, those who thought that super-exogeneity is an important concept could not bring themselves to tell me why, or how to discern it, even given a fully specified econometric model (namely, an oracle for all econometric questions, in which all conditions, including Lucas critics and rational expectations are already encoded).

Encouraged by Ed Leamer, I wrote a brief section on exogeneity in my book (Pearl, 2000, pp. 165–8, 246), and I argue that super-exogeneity is none other but the "back-door" criterion of non-confoundedness which generates the classical requirement that X be independent of all error terms affecting Y. (This independence is named "strong ignorability" in the language of potential outcomes (Rubin, 1974).) Therefore, super-exogeneity can be determined from the model by inspection, and be given a simple formula and a simple experimental test:

$$P(y|do(x)) = P(y|x) \tag{1}$$

In words, the probability of $Y = y$ under experimental conditions where X is held constant at $X = x$ is equal to the conditional probability $P(y|x)$ that one estimates in observational studies by regressing Y on X.

This raises the question whether the EHR paper of 1983 (which is one of the most cited articles in econometrics, with 1,189 citations according to

Google Scholar) could not have been reduced to a couple of pages, given modern vocabulary and modern interpretation of structural equations. I further argued in that section that the definition given by EHR is flawed, in that it does not distinguish between structural and statistical parameters (Pearl, 2009, p. 167). Although no economist has thus far objected to my critics, it is hard to know whether this indicates agreement or confusion; the latter seems more likely, given the schism in which the field currently finds itself (e.g., (Pearl, 2009, pp. 171, 379–80; Pearl, 2010; Hoover, 2004).

In this paper I will first provide a brief account of the ambiguities in the definition and interpretation of EHR's conceptualization of exogeneity and then attempt to remove those ambiguities and position exogeneity in the framework of modern theories of causation.

2 Ambiguities in the understandings of super-exogeneity

Hendry and Santos (2006) summarize the essence of exogeneity thus: "super-exogeneity of the parameters of conditional models under changes in the distributions of conditioning variables is of paramount importance."

This summary casts exogeneity as a relationship between two mathematical objects, a "conditional model" and a "distribution of conditioning variables." Super-exogeneity is said to exist when the parameters of the former remain invariant to changes in the latter. This conception, variants of which appear in all discussions of exogeneity, highlights four basic ambiguities.

1. What is a "conditional model"? Is it the conditional distribution $P(y|x)$? Is it the post-intervention distribution $P(y|do(x))$ or the structural equation $y = f(x, \epsilon)$? As is well known to modern students of causality, the three entities carry different types of information, of increasing level of details.

2. What are the "parameters" of the a "conditional model"? For example, is $P(Y = 1|X = 0)$ a parameter of the conditional model? Would the conditional expectation $E(Y|x)$ or the conditional variance $E(Y^2|x)$ qualify as parameters of the conditional model?

3. What kind of "changes in the distributions of conditioning variables" should we consider when testing exogeneity? Should we consider "changes due to new observations"? or merely changes due to external intervention? or perhaps all kind of changes, including a complete model

restructuring?

4. Regardless of the answers to 1-3, how are we to test for invariance between the conditional and marginal model? Specifically, can we achieve invariance by a clever re-parameterization of the model? If so, which parametric representation is the correct one?

Many of these ambiguities could have been avoided had the DGP been defined in structural, rather than probabilistic vocabulary. Interestingly, four years after the publication of EHR's paper, Hendry (1987) made one overture in this direction and equated the DGP with the structural equation $y = z\beta + \epsilon$, which he describes as "a Monte Carlo experiment conducted for correct specifications." Unfortunately, he quickly abandoned this approach (ibid p. 32) in favor of a conditional density description, saying: "Denote the DGP of all relevant economics variables $\{x_i\}$ by $D(X_T|X_0;\theta_T)$."

The structural description has since disappeared from all subsequent writings on exogeneity. For example, in Hendry (2004) we find: "the joint density $D_w(w|F_t - 1, \lambda)$ is the data generation process (DGP)." And in Hendry and Santos (2006) we find again DGP defined as a sequential decomposition of joint density.

The inevitable confusion that probabilistic vocabulary induces has been particularly pronounced in econometric textbooks. For example, in the highly popular textbook by Greene (2003) (to which Google Scholar ascribes over 30,000 citations) we find the following quotes:

1. "exogeneity is an assumption in regression."
2. "what constitutes an 'exogenous' variable becomes ambiguous." p. 591.
3. "exogeneity is not an absolute concept at all; it is defined in the context of the model."
4. "we define a variable as exogenous in the context of our model if the joint density may be written $f(y, x) = f(y|\beta, x)f(\theta, x)$ where the parameters in the conditional distribution do not appear in and are functionally unrelated to those in the marginal distribution of x."

The dangers of relying on syntactical criteria such as parameters "appearing" or "not appearing" in distributions are exemplified in Pearl (2000, p. 168). But the dangers of expecting causal notions such exogeneity to emerge from probabilistic definitions are more subtle.

3 new section 3

To witness, consider a joint distribution $P(x, y)$ defined for binary variables X and Y. The natural parameterization of $P(x, y)$ would be in terms of the four parameter vector

$$\overline{p} = (p_{00}, p_{10}, p_{01}, p_{11}) \tag{2}$$

where

$$p_{ij} = P(X = i, Y = j) \qquad i, j = 0, 1 \tag{3}$$

using the parameterization, we have the decomposition:

$$P(x, y) = P(x)P(y|x)$$

$$P(x) = \begin{cases} p_{00} + p_{01} & x = 0 \\ p_{10} + p_{11} & x = 1 \end{cases} \tag{4}$$

$$P(y|x) = \begin{cases} \dfrac{p_{00}}{p_{00} + p_{01}} & x = 0 \quad y = 0 \\ \dfrac{p_{01}}{p_{00} + p_{01}} & x = 0 \quad y = 1 \\ \dfrac{p_{10}}{p_{10} + p_{11}} & x = 1 \quad y = 0 \\ \dfrac{p_{11}}{p_{10} + p_{11}} & x = 1 \quad y = 0 \end{cases} \tag{5}$$

We see that *every* parameter that appears in $p(y|x)$ also appears in $P(x)$. Therefore, taking literally the criterion that "the parameters in the conditional distribution do not appear in and are functionally unrelated to those in the marginal distribution of x" would behoove us to conclude that no variable could possibly be deemed exogenous relative to another. Obviously, some restrictions be imposed on the choice of parameters as well as on what we mean by "parameters of a distribution appearing in another distribution," or "functionally related" to other parameters.

Indeed if we choose to describe $p(x, y)$ using the following parameterization:

$$\underline{q} = (q_1, q_{01}, q_{11}) \tag{6}$$

where

$$P(x;\underline{q}) = \begin{cases} 1-q_1 & x=0 \\ q_1 & x=1 \end{cases} \tag{7}$$

$$P(y|x;\underline{q}) = \begin{cases} 1-q_{01} & x=0, y=0 \\ q_{01} & x=0, y=1 \\ 1-q_{11} & x=1, y=0 \\ q_{11} & x=1, y=1 \end{cases} \tag{8}$$

We see that non of the parameters in the conditional distribution appears in the marginal distribution of X, and we might be tempted to conclude that *every* variable X is exogenous relative to any other.

Perhaps this is what Green meant by stating: "exogeneity is not a absolute concept at all, it is defined in the context of the model." On the other hand, it is inconceivable that exogeneity, a concept deemed important for estimation and decision, would be so utterly sensitive to notation. After all, whether we can conduct inference ignoring $P(x)$ without loss of information is an objective empirical question and cannot be dependent, as Green states, on "the context of the model."

To witness, there are features of $P(x, y)$ that are invariant to choice of parameters. For example, regardless of whether we choose to describe $P(x, y)$ using $\underline{p}$ (Eq. 5) or $\underline{q}$ (Eq. 6) the fact remains that knowledge of $P(x)$ is unneeded for estimating the ratio $\eta = \frac{p_{00}p_{11}}{p_{01}p_{10}}$, and is needed for estimating the ratio $\eta' = \frac{p_{00}p_{01}}{p_{11}p_{10}}$.[1]

The fact also remains that this basic distinction between η and η' cannot be revealed by simple syntactic means such as testing whether "the parameters in the conditional distribution do not appear in the marginal distribution." We will now define this notion of invariant formally.

Let $\underline{p} = (p_1, p_2, \ldots, p_n)$ be a vector of parameters that specifies the joint density $P(u)$ and let $\underline{q} = (q_1, q_2, \ldots, q_k)$ be the vector of parameter that specifies the conditional probability $P(y|x)$. Clearly $\underline{q} = \underline{f}(\underline{p})$. Assume we are interested in a parameter η of the $P(v)$. Define X to be exogenous relative to η, if $\exists$ a function g such that $y = g(\underline{q})$, or, in other words, for any $\underline{p}$ and $\underline{p'}$,

$$\underline{f}(\underline{p}) = \underline{f}(p') \Rightarrow \eta(p) = \eta(p')$$

[1]This can be verified by noting that $\eta = \frac{(1-q_{01})q_{11}}{(1-q_{11}q_{01})}$ while $\eta' = \frac{(1-q_{01})(q_{01})(1-p_1)}{(1-q_{11})q_{11}P_1^2}$; the latter depends on $p_1 = P(X = 1)$, the former does not.

Conversely, X is not exogenous for η if $\exists$ two vectors, p and p', such that

$$f(p) = f(p') \text{ yet } \eta(p) \neq \eta(p').$$

To illustrate this latter criterion, we show that, in the binary example of Eq. (2), X is not exogenous for $\eta' = \dfrac{p_{00}p_{01}}{p_{10}p_{11}}$

Consider two models

$$\underline{p} = (p_{00}, p_{01}, p_{10}, p_{11})$$

$$p' = (p'_{00}, p'_{01}, p'_{10}, p'_{11})$$

$$\underline{f}(\underline{p}) = (q_1(\underline{p}), q_{01}(\underline{p}), q_{11}(p))$$

$$= ((p_{10} + p_{11}), \frac{p_{01}}{p_{01} + p_{00}}, \frac{p_{11}}{p_{11} + p_{10}}$$

such that $q_1(\underline{p}) \neq q_1(\underline{p'}), q_{01}(\underline{p}) = q_{01}(\underline{p'}), q_{11}(\underline{p}) = q_{11}(\underline{p'})$

$$p_{10} + p_{11} \neq p'_{10} + p'_{11}, \frac{p_{01}}{p_{01} + p_{00}} = \frac{p'_{01}}{p'_{01} + p'_{00}} \frac{p_{11}}{p_{11} + p_{10}} = \frac{p'_{11}}{p'_{11}} p'_{10}$$

take $p_{11} + p_{10} = \frac{1}{2}(p'_{11} + p_{10}) \Rightarrow p_{01} + p_{00} = 1 - p_{11} - p_{10} = 1 - \frac{1}{2}(p'_{11} + p'_{10})$

$$= 1 - \frac{1}{2}(1 - p'_{00} - p'_{01}) = \frac{1}{2} + \frac{1}{2}(p'_{00} + p'_{01})$$

$$\text{but } \frac{p_{01}}{p_{01} + p_{00}} = \frac{p'_{01}}{p'_{01} + p'_{00}}$$

$$\text{and } \frac{p_{11}}{p_{11} + p_{10}} = \frac{p'_{11}}{p'_{11} + p'_{10}}$$

This will mark $\underline{q} = (q_{01}, q_{11}) = (q'_{01}, q'_{11})$ while $y(\underline{p}) \neq y(p')$

4 Disambiguating EHR exogeneity

EHR attempt to define exogeneity in term of invariance of the conditional probability $P(y|x)$ to changes in the process generating X is rest on healthy intuition, it is encapsulated indeed in Eq. (1), using the extreme case where the process generating X reduces to holding X constant at $X = x$. However, this definition invokes a specific intervention $do(X = x)$, which requires a causal model for its definition. EHR program prefers to define things in terms of marginal and conditional probabilities, seemingly with no underlying causal model, and thus requires a thorough disambiguation of terms such as "the parameters of the marginal and conditional distribution," "invariant to change in the marginal distribution."

The following constitutes an explication of these terms.

Let M_λ be a model indexed by a vector λ of parameters $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)$. We leave the precise representation of M_λ open at this point and require only that M_λ be sufficiently powerful to specify, for each λ, a complete probability function $P_{M_\lambda}(v)$ on the variables of interest. Now consider any property η of $P_{M_\lambda}(v)$ (say, the mean $E_{M_\lambda}(X)$, variance $E_{M_\lambda}(X - E(X))^2$, or the conditional probability $P_{M_\lambda}(y|x)$) which we write as $\eta[P_{M_\lambda}(v)]$. To capture the notion of "parameters of the marginal distribution $P(x)$, we define the *indexical projection* of η, $J(\eta)$, to be the (unique) subset of indices

$$J(\eta) \in \{1, 2, 3, \ldots, n\}$$

such that, for any two vectors λ' and λ'' we have

$$\eta[P_{M_{\lambda'}}(v)] = \eta[P_{M_{\lambda''}}(v)]$$

whenever $\lambda_i = \lambda_i'' \;\; \forall \;\; i \in J(\eta)$. In words, we say that a subset $\{\lambda_i | i \in J(\eta)\}$ of parameters "determines" η if any two models that agree on this set also generate the same η. The rational is clear; the other parameters are redundant, they are not needed for the determination of η.

We are now ready to define the notion of "invariance" between two quantities, η' and η'', which we will later apply to explicate the invariance of the conditional density $P(y|x)$ to changes in the marginal $P(x)$.

To this end, we need to define the preliminary notion of "image."

Definition 1 *(Image)*

We say that model $M_{\lambda'}$ is an image of model $M_{\lambda''}$ relative to η, just in case

$$\lambda_i' = \lambda_i'' \;\; \forall \;\; i \in J(\eta).$$

In words, $M_{\lambda'}$ and $M_{\lambda''}$ may differ only in parameters that determine η. Intuitively, if some of the parameters that determine η change, any image of $M_{\lambda'}$ can be considered "closest" model to $M_{\lambda'}$ in the sense or retaining the original values of all parameters that do not determine η.

We now define invariance

Definition 2 *(Invariance)*

We say that $\eta'(P)$ is invariant to changes in $\eta''(P)$ iff

$$\eta'[P_{M_{\lambda'}}] = \eta'[P_{M_{\lambda''}}]$$

whenever $M_{\lambda'}$ is an image of $M_{\lambda''}$

Note: Invariance is not a symmetric relation, i,e,. the marginal $P(x)$ is usually invariant to the conditional $P(y|x)$, but not vice versa.

Example 1

Acknowledgments

This research was partially supported by grants from NIH #1R01 LM009961-01, NSF #IIS-0914211, and ONR #N000-14-09-1-0665.

References

Balke, A. and Pearl, J. (1994). Counterfactual probabilities: Computational methods, bounds, and applications. In *Uncertainty in Artificial Intelligence 10* (R. L. de Mantaras and D. Poole, eds.). Morgan Kaufmann, San Mateo, CA, 46–54.

Engle, R., Hendry, D. and Richard, J. (1983). Exogeneity. *Econometrica* **51** 277–304.

Greene, W. (2003). *Econometric Analysis.* 5th ed. Prentice Hall, New York.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11** 1–12. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, 477–490, 1995.

Haavelmo, T. (1944). The probability approach in econometrics (1944)*. Supplement to *Econometrica* **12** 12–17, 26–31, 33–39. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, New York, 440–453, 1995.

Heckman, J. (2005). The scientific model of causality. *Sociological Methodology* **35** 1–97.

Hendry, D. (1987). Econometric methodology: A personal perspective (chapter 10). In *Advances in Econometrics: Fifth World Congress* (T. F. Bewley, ed.). Cambridge University Press, New York, 29–40.

Hendry, D. (2004). Causality and exogeneity in non-stationary economic time series. Tech. rep., Economics Department, University of Oxford, United Kingdom.

Hendry, D. and Santos, C. (2006). Automatic tests for super exogeneity. Tech. Rep. 11/2007, Portuguese Catholic University, Porto, Portugal.

Hoover, K. (2004). Lost causes. *Journal of the History of Economic Thought* **26** 149–164.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. 2nd edition, 2009.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

Pearl, J. (2010). Review of *hunting causes and using them*, N. Cartwright (2008). *Economics and Philosophy* **26** 69–94, <http://ftp.cs.ucla.edu/pub/stat_ser/r342.pdf>.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688–701.