11.5.4 Where Is Economic Modeling Today? – Courting Causes with Heckman

Section 5.2 of this book decries the decline in the understanding of structural equation modeling in econometric in the past three decades (see also Hoover 2003, "Lost Causes") and attributes this decline to a careless choice of notation which blurred the essential distinction between algebraic and structural equations. In a series of articles (Heckman 2000, 2003, 2005; Heckman and Vytlacil 2007), James Heckman has set out to overturn this perception, reclaim causal modeling as the central focus of economic research, and reestablish economics as an active frontier in causal analysis. This is not an easy task by any measure. To adopt the conceptual and technical advances that have emerged in neighboring disciplines would amount to admitting decades of neglect in econometrics, while to dismiss those advances would necessitate finding them econometric surrogates. Heckman chose the latter route, even though most modern advances in causal modeling are rooted in the ideas of economists such as Haavelmo (1943), Marschak (1950), and Strotz and Wold (1960).

One step in Heckman's program was to reject the *do*-operator and the "surgery" semantics upon which it is based, thus depriving economists of the structural semantics of counterfactuals developed in this book (especially Chapter 7), which unifies traditional econometrics with the potential-outcome approach. Heckman's reasons for rejecting surgery are summarized thus:

> Controlled variation in external (forcing) variables is the key to defining causal effects in nonrecursive models … Pearl defines a causal effect by 'shutting one equation down' or performing 'surgery' in his colorful language. He implicitly assumes that 'surgery,' or shutting down an equation in a system of simultaneous equations, uniquely fixes one outcome or internal variable (the consumption of the other person in my example). In general, it does not. Putting a constraint on one equation places a restriction on the entire set of internal variables. In general, no single equation in a system of simultaneous equations uniquely determines any single outcome variable. Shutting down one equation might also affect the parameters of the other equations in the system and violate the requirements of parameter stability. (Heckman and Vytlacil 2007)

Clearly, Heckman's objections are the same as Cartwright's (Section 11.4.6):

1. Ideal surgery may be technically infeasible,
2. Economic systems are nonmodular.

We have repudiated these objections in four previous subsections (11.4.3–11.4.6) which readers can easily reapply to deconstruct Heckman's arguments. It is important to reemphasize, though, that, as in the case of Cartwright, these objections emanate from conflating the task of definition (of counterfactuals) with those of identification and practical estimation, a frequent confusion among researchers which Heckman (2005) sternly warns readers to avoid.

This conflation is particularly visible in Heckman's concern that "shutting down one equation might also affect the parameters of the other equations in the system." In the physical world, attempting to implement the conditions dictated by a "surgery" may sometimes affect parameters in other equations, and, as we shall see, the same applies to Heckman's proposal of "external variation." However, we are dealing here with symbolic, not physical, manipulations. Our task is to formulate a meaningful mathematical definition of "the causal effect of one variable on another" in a symbolic system called a "model." This permits us to manipulate symbols at will, while ignoring the technical feasibility of these manipulations. Implementational considerations need not enter the discussion of *definition*.

A New Definition of Causal Effects: "External Variation"

Absent surgery semantics, Heckman and Vytlacil (HV) set out to configure a new definition of causal effects, which, hopefully, would be free of the faults they discovered in the surgery procedure, by basing it on "external-variations," instead of shutting down equations. It is only unfortunate that their new definition, the cornerstone of their logic of counterfactuals, is not given an explicit formal exposition: it is relegated to a semiformal footnote (HV, p. 77) that even a curious and hard-working reader would find difficult to decipher. The following is my extrapolation of HV's definition as it applies to multi-equations and nonlinear systems.

Given a system of equations:

$$Y_i = f_i(Y, X, U) \; i = 1, 2, \dots, n,$$

where X and U are sets of observed and unobserved external variables, respectively, the causal effect of Y_j on Y_k is computed in four steps:

1. Choose any member X_t of X that appears in f_j. If none exists, exit with failure.
2. If X_t appears in any other equation as well, consider excluding it from that equation (e.g., set its coefficient to zero if the equation is linear or replace X_t by a constant).[15]
3. Solve for the reduced form

   $$Y_i = g_i(X, U) \; i = 1, 2, \dots, n \tag{11.23}$$

   of the resulting system of equations.
4. The causal effect of Y_j on Y_k is given by the partial derivative:

   $$dY_k/dY_j = dg_k/dX_t : dg_j/dX_t. \tag{11.24}$$

Example 11.5.2 Consider a system of three equations:

$$\begin{aligned} Y_1 &= aY_2 + cY_3 + eX + U_1 \\ Y_2 &= bY_1 + X + U_2 \\ Y_3 &= dY_1 + U_3. \end{aligned}$$

[15] It is not clear what conditions (if any) would forbid one from setting $e = 0$, in example 11.5.2, or ignoring X altogether and adding a dummy variable X' to the second equation. HV give the impression that deciding on whether e can be set to 0 requires deep understanding of the problem at hand; if this is their intention, it need not be.

Needed: the causal effect of Y_2 *on* Y_1.

The system has one external variable, X, which appears in the first two equations. If we can set $e = 0$, x will appear in the equation of Y_2 only, and we can then proceed to Step 3 of the "external variation" procedure. The reduced form of the modified model yields:

$$dY_1/dX = a/(1 - ba - cd) \qquad dY_2/dX = (1 - cd)/(1 - ab - cd),$$

and the causal effect of Y_1 on Y_2 calculates to:

$$dY_1/dY_2 = a/(1 - cd).$$

In comparison, the surgery procedure constructs the following modified system of equations:

$$Y_1 = aY_2 + cY_3 + eX + U_1$$
$$Y_2 = y_2$$
$$Y_3 = dY_1 + U_3,$$

from which we obtain for the causal effect of Y_2 on Y_1;

$$dY_1/dy_2 = a/(1 - cd),$$

an expression identical to that obtained from the "external variation" procedure.

It is highly probable that the two procedures always yield identical results, which would bestow validity and conceptual clarity on the "external variation" definition.

11.5.5 External Variation versus Surgery

In comparing their definition to the one provided by the surgery procedure, HV write (p. 79): "Shutting down an equation or fiddling with the parameters … is not required to define causality in an interdependent, nonrecursive system or to identify causal parameters. The more basic idea is exclusion of different external variables from different equations which, when manipulated, allow the analyst to construct the desired causal quantities."

I differ with HV on this issue. I believe that "surgery" is the more basic idea, more solidly motivated, and more appropriate for policy evaluation tasks. I further note that basing a definition on exclusion and external variation suffers from the following flaws:

1. In general, "exclusion" involves the removal of a variable from an equation and amounts to "fiddling with the parameters." It is, therefore, a form of "surgery" – a modification of the original system of equations – and would be subject to the same criticism one may raise against "surgery." Although we have refuted such criticism in previous sections, we should nevertheless note that if it ever has a grain of validity, the criticism would apply equally to both methods.
2. The idea of relying exclusively on external variables to reveal internal cause–effect relationships has its roots in the literature on *identification* (e.g., as in the studies of "instrumental variables") when such variables act as "nature's experiments." This restriction, however, is unjustified in the context

of *defining* causal effect, since "causal effects" are meant to quantify effects produced by *new* external manipulations, not necessarily those shown explicitly in the model and not necessarily those operating in the data-gathering phase of the study. Moreover, every causal structural equation model, by its very nature, provides an implicit mechanism for emulating such external manipulations, via surgery.

Indeed, most policy evaluation tasks are concerned with *new* external manipulations which exercise direct control over endogenous variables. Take, for example, a manufacturer deciding whether to double the current price of a given product after years of letting the price track the cost, i.e., $price = f(cost)$. Such a decision amounts to removing the equation $price = f(cost)$ from the model at hand (i.e., the one responsible for the available data) and replacing it with a constant equal to the new price. This removal emulates faithfully the decision under evaluation, and attempts to circumvent it by appealing to "external variables" are artificial and hardly helpful.

As another example, consider the well-studied problem (Heckman 1992) of evaluating the impact of terminating an educational program for which students are admitted based on a set of qualifications. The equation $admission = f(qualifications)$ will no longer hold under program termination, and no external variable can simulate the new condition (i.e., $admission = 0$) save for one that actually neutralizes (or "ignores," or "shuts down") the equation $admission = f(qualifications)$.

It is also interesting to note that the method used in Haavelmo (1943) to define causal effects is mathematically equivalent to surgery, not to external variation. Instead of replacing the equation $Y_j = f_j(Y, X, U)$ with $Y_j = y_j$, as would be required by surgery, Haavelmo writes $Y_j = f_j(Y, X, U) + x_j$, where X_j is chosen so as to make Y_j constant, $Y_j = y_j$. Thus, since X_j liberates Y_j from any residual influence of $f_j(Y, X, U)$, Haavelmo's method is equivalent to that of surgery. Heckman's method of external variation leaves Y_j under the influence f_j.

3. Definitions based on external variation have the obvious flaw that the target equation may not contain any observable external variable. In fact, in many cases the set of observed external variables in the system is empty (e.g., Figure 3.5). Additionally, a definition based on a ratio of two partial derivatives does not generalize easily to nonlinear systems with discrete variables. Thus, those who seriously accept Heckman's definition would be deprived of the many identification techniques now available for instrumentless models (see Chapters 3 and 4) and, more seriously yet, would be unable to even ask whether causal effects are identified in any such model – identification questions are meaningless for undefined quantities.

   Fortunately, liberated by the understanding that definitions can be based on purely symbolic manipulations, we can modify Heckman's proposal and *add* fictitious external variables to any equation we desire. The added variables can then serve to define causal effects in a manner similar to the steps in equations (11.23) and (11.24) (assuming continuous variables). This brings us closer to surgery, with the one basic difference of leaving Y_j under the influence of $f_j(Y, X, U)$.

Having argued that definitions based on "external variation" are conceptually ill-motivated, we now explore whether they can handle noncausal systems of equations.

Equation Ambiguity in Noncausal Systems

Several economists (Leroy 2002; Neuberg 2003; Heckman and Vytlacil 2007) have criticized the *do*-operator for its reliance on *causal*, or directional, structural equations, where we have a one-to-one correspondence between variables and equations. HV voice this criticism thus: "In general, no single equation in a system of simultaneous equations uniquely determines any single outcome variable" (Heckman and Vytlacil 2007, p. 79).

One may guess that Heckman and Vytlacil refer here to systems containing nondirectional equations, namely, equations in which the equality sign does not stand for the non-symmetrical relation "is determined by" or "is caused by" but for symmetrical algebraic equality. In econometrics, such noncausal equations usually convey equilibrium or resource constraints; they impose equality between the two sides of the equation but do not endow the variable on the left-hand side with the special status of an "outcome" variable.

The presence of nondirectional equations creates ambiguity in the surgical definition of the counterfactual Y_x, which calls for replacing the equation *determining* X with the constant equation $X = x$. If X appears in several equations, and if the position of X in the equation is arbitrary, then each one of those equations would be equally qualified for replacement by $X = x$, and the value of Y_x (i.e., the solution for Y after replacement) would be ambiguous.

Note that symmetrical equalities differ structurally from reciprocal causation in directional nonrecursive systems (i.e., systems with feedback, as in Figure 7.4), since, in the latter, each variable is an "outcome" of precisely one equation. Symmetrical constraints can nevertheless be modeled as the solution of a dynamic feedback system in which equilibrium is reached almost instantaneously (Lauritzen and Richardson 2002; Pearl 2003a).

Heckman and Vytlacil create the impression that equation ambiguity is a flaw of the surgery definition and does not plague the exclusion-based definition. This is not the case. In a system of nondirectional equations, we have no way of knowing which external variable to exclude from which equation to get the right causal effect.

For example: Consider a nonrecursive system of two equations that is discussed in HV, p. 75:

$$Y_1 = a_1 + c_{12}Y_2 + b_{11}X_1 + b_{12}X_2 + U_1 \tag{11.25}$$

$$Y_2 = a_2 + c_{21}Y_1 + b_{21}X_1 + b_{22}X_2 + U_2. \tag{11.26}$$

Suppose we move Y_1 to the l.h.s. of (11.26) and get:

$$Y_1 = [a_2 - Y_2 + b_{21}X_1 + b_{22}X_2 + U_2]/c_{21}. \tag{11.27}$$

To define the causal effect of Y_1 on Y_2, we now have a choice of excluding X_2 from (11.25) or from (11.27). The former yields c_{12}, while the latter yields $1/c_{21}$. We see that the ambiguity we have in choosing an equation for surgery translates into ambiguity in choosing an equation and an external variable for exclusion.

Methods of breaking this ambiguity were proposed by Simon (1953) and are discussed on pages 226–8.

Summary – Economic Modeling Reinvigorated

The idea of constructing causal quantities by exclusion and manipulation of external variables, while soundly motivated in the context of identification problems, has no logical basis when it comes to model-based definitions. Definitions based on surgery, on the other hand, enjoy generality, semantic clarity, and computational simplicity.

So, where does this leave econometric modeling? Is the failure of the "external variable" approach central or tangential to economic analysis and policy evaluation?

In almost every one of his recent articles James Heckman stresses the importance of counterfactuals as a necessary component of economic analysis and the hallmark of econometric achievement in the past century. For example, the first paragraph of the HV article reads: "they [policy comparisons] require that the economist construct counterfactuals. Counterfactuals are required to forecast the effects of policies that have been tried in one environment but are proposed to be applied in new environments and to forecast the effects of new policies." Likewise, in his *Sociological Methodology* article (2005), Heckman states: "Economists since the time of Haavelmo (1943, 1944) have recognized the need for precise models to construct counterfactuals… The econometric framework is explicit about how counterfactuals are generated and how interventions are assigned…"

And yet, despite the proclaimed centrality of counterfactuals in econometric analysis, a curious reader will be hard pressed to identify even one econometric article or textbook in the past 40 years in which counterfactuals or causal effects are formally defined. Needed is a procedure for computing the counterfactual $Y(x, u)$ in a well-posed, fully specified economic model, with X and Y two arbitrary variables in the model. By rejecting Haavelmo's definition of $Y(x, u)$, based on surgery, Heckman commits econometrics to another decade of division and ambiguity, with two antagonistic camps working in almost total isolation.

Economists working within the potential-outcome framework of the Neyman-Rubin model take counterfactuals as primitive, unobservable variables, totally detached from the knowledge encoded in structural equation models (e.g., Angrist 2004; Imbens 2004). Even those applying propensity score techniques, whose validity rests entirely on the causal assumption of "ignorability," or unconfoundedness, rarely know how to confirm or invalidate that assumption using structural knowledge (see Section 11.3.5). Economists working within the structural equation framework (e.g., Kennedy 2003; Mittelhammer et al. 2000; Intriligator et al. 1996) are busy estimating parameters while treating counterfactuals as metaphysical ghosts that should not concern ordinary mortals. They trust leaders such as Heckman to define precisely what the policy implications are of the structural parameters they labor to estimate, and to relate them to what their colleagues in the potential-outcome camp are doing.[16]

The surgery semantics (pp. 98–102) and the causal theory entailed by it (Chapters 7–10) offer a simple and precise unification of these two antagonistic and narrowly focused schools of econometric research – a theorem in one approach entails a theorem in the other, and vice versa. Economists will do well resurrecting the basic

---

[16] Notably, the bibliographical list in the comprehensive review article by economist Hoover (2008) is almost disjoint from those of economists Angrist (2004) and Imbens (2004) – the cleavage is culturally deep.

ideas of Haavelmo (1943), Marschak (1950), and Strotz and Wold (1960) and re-invigorating them with the logic of graphs and counterfactuals presented in this book.

For completeness, I reiterate here explicitly (using parenthetical notation) the two fundamental connections between counterfactuals and structural equations.

1. The structural definition of counterfactuals is:

   $$Y_M(x, u) = Y_{M_x}(u).$$

   Read: For any model M and background information u, the counterfactual conditional "Y if X had been x" is given by the solution for Y in submodel M_x (i.e., the mutilated version of M with the equation determining X replaced by $X = x$).

2. The empirical claim of the structural equation $y = f(x, e(u))$ is:

   $$Y(x, z, u) = f(x, e(u)),$$

   for any set Z not intersecting X or Y.

   Read: Had X and Z been x and z, respectively, Y would be $f(x, e(u))$, independently of z, and independently of other equations in the model.