From *Causality,* pages 154-163 (Sections 5.3.2 - 5.4.1)
Important Topic: On the meaning of structural equations

[Author's note: This topic, which I often call "The Confusion of the Century" has been lingering in the literatures of statistics, econometrics, social science and psychology since the early 1900's. With the exception of very few, researchers using structural equations models are still debating the meaning of the equations, what makes them different from regression equations, the assumptions embodied in each equation, and the utility of the parameters they labor to estimate.

A major contributor to this lingering confusion has been the lack of mathematical notation to define structural concepts; the answers to the questions above cannot be articulated in the language of textbook statistics. The sections posted below resolve, so I hope, one of the most embarrassing confusion in the history of data analysis.]

(Scroll down to start of Section 5.3.2)

5.3.2 Comparison to Nonparametric Identification

The identification results of the previous section are significantly more powerful than those obtained in Chapters 3 and 4 for nonparametric models. Nonparametric models should nevertheless be studied by parametric modelers for both practical and conceptual

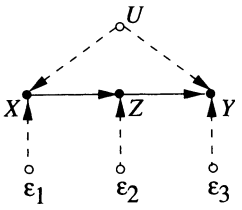

Figure 5.12 Path diagram corresponding to equations (5.4)–(5.6), where $\{X, Z, Y\}$ are observed and $\{U, \varepsilon_1, \varepsilon_2, \varepsilon_3\}$ are unobserved.

reasons. On the practical side, investigators often find it hard to defend the assumptions of linearity and normality (or other functional-distributional assumptions), especially when categorical variables are involved. Because nonparametric results are valid for nonlinear functions and for any distribution of errors, having such results allows us to gauge how sensitive standard techniques are to assumptions of linearity and normality. On the conceptual side, nonparametric models illuminate the distinctions between structural and algebraic equations. The search for nonparametric quantities analogous to path coefficients forces explication of what path coefficients really mean, why one should labor at their identification, and why structural models are not merely a convenient way of encoding covariance information.

In this section we cast the problem of nonparametric causal effect identification (Chapter 3) in the context of parameter identification in linear models.

Parametric versus Nonparametric Models: An Example

Consider the set of structural equations

$$x = f_1(u, \varepsilon_1), \tag{5.4}$$

$$z = f_2(x, \varepsilon_2), \tag{5.5}$$

$$y = f_3(z, u, \varepsilon_3), \tag{5.6}$$

where X, Z, Y are observed variables, f_1, f_2, f_3 are unknown arbitrary functions, and U, ε_1, ε_2, ε_3 are unobservables that we can regard either as latent variables or as disturbances. For the sake of this discussion, we will assume that U, ε_1, ε_2, ε_3 are mutually independent and arbitrarily distributed. Graphically, these influences can be represented by the path diagram of Figure 5.12.

The problem is as follows. We have drawn a long stream of independent samples of the process defined by (5.4) – (5.6) and have recorded the values of the observed variables X, Z, and Y; we now wish to estimate the unspecified quantities of the model to the greatest extent possible.

To clarify the scope of the problem, we consider its linear version, which is given by

$$x = u + \varepsilon_1, \tag{5.7}$$

$$z = \alpha x + \varepsilon_2, \tag{5.8}$$

$$y = \beta z + \gamma u + \varepsilon_3, \tag{5.9}$$

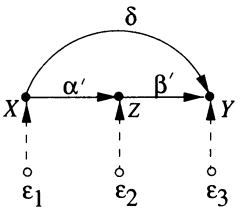

Figure 5.13 Diagram representing model M' of (5.12)–(5.14).

where U, ε_1, ε_2, ε_3 are uncorrelated, zero-mean disturbances.[14] It is not hard to show that parameters $\alpha, \beta,$ and γ can be determined uniquely from the correlations among the observed quantities X, Z, and Y. This identification was demonstrated already in the example of Figure 5.7, where the back-door criterion yielded

$$\beta = r_{YZ \cdot X}, \quad \alpha = r_{ZX}, \tag{5.10}$$

and hence

$$\gamma = r_{YX} - \alpha\beta. \tag{5.11}$$

Thus, returning to the nonparametric version of the model, it is tempting to generalize that, for the model to be identifiable, the functions $\{f_1, f_2, f_3\}$ must be determined uniquely from the data. However, the prospect of this happening is unlikely, because the mapping between functions and distributions is known to be many-to-one. In other words, given any nonparametric model M, if there exists one set of functions $\{f_1, f_2, f_3\}$ compatible with a given distribution $P(x, y, z)$, then there are infinitely many such functions (see Figure 1.6). Thus, it seems that nothing useful can be inferred from loosely specified models such as the one given by (5.4)–(5.6).

Identification is not an end in itself, however, even in linear models. Rather, it serves to answer practical questions of prediction and control. At issue is not whether the data permit us to identify the form of the equations but, instead, whether the data permit us to provide unambiguous answers to questions of the kind traditionally answered by parametric models.

When the model given by (5.4)–(5.6) is used strictly for prediction (i.e., to determine the probabilities of some variables given a set of observations on other variables), the question of identification loses much (if not all) of its importance; all predictions can be estimated directly from either the covariance matrices or the sample estimates of those covariances. If dimensionality reduction is needed (e.g., to improve estimation accuracy) then the covariance matrix can be encoded in a variety of simultaneous equation models, all of the same dimensionality. For example, the correlations among X, Y, and Z in the linear model M of (5.7)–(5.9) might well be represented by the model M' (Figure 5.13):

$$x = \varepsilon_1, \tag{5.12}$$

$$z = \alpha' x + \varepsilon_2, \tag{5.13}$$

$$y = \beta' z + \delta x + \varepsilon_3. \tag{5.14}$$

[14] An equivalent version of this model is obtained by eliminating U from the equations and allowing ε_1 and ε_3 to be correlated, as in Figure 5.7.

This model is as compact as (5.7)–(5.9) and is covariance equivalent to M with respect to the observed variables X, Y, Z. Upon setting $\alpha' = \alpha, \beta' = \beta$, and $\delta = \gamma$, model M' will yield the same probabilistic predictions as those of the model of (5.7)–(5.9). Still, when viewed as data-generating mechanisms, the two models are not equivalent. Each tells a different story about the processes generating X, Y, and Z, so naturally their predictions differ concerning the changes that would result from subjecting these processes to external interventions.

5.3.3 Causal Effects: The Interventional Interpretation of Structural Equation Models

The differences between models M and M' illustrate precisely where the structural reading of simultaneous equation models comes into play, and why even causally shy researchers consider structural parameters more "meaningful" than covariances and other statistical parameters. Model M', defined by (5.12)–(5.14), regards X as a direct participant in the process that determines the value of Y, whereas model M, defined by (5.7)–(5.9), views X as an indirect factor whose effect on Y is mediated by Z. This difference is not manifested in the data itself but rather in the way the data would change in response to outside interventions. For example, suppose we wish to predict the expectation of Y after we intervene and fix the value of X to some constant x; this is denoted $E(Y \mid do(X = x))$. After $X = x$ is substituted into (5.13) and (5.14), model M' yields

$$E[Y \mid do(X = x)] = E[\beta'\alpha' x + \beta'\varepsilon_2 + \delta x + \varepsilon_3] \tag{5.15}$$

$$= (\beta'\alpha' + \delta)x; \tag{5.16}$$

model M yields

$$E[Y \mid do(X = x)] = E[\beta\alpha x + \beta\varepsilon_2 + \gamma u + \varepsilon_3] \tag{5.17}$$

$$= \beta\alpha x. \tag{5.18}$$

Upon setting $\alpha' = \alpha, \beta' = \beta$, and $\delta = \gamma$ (as required for covariance equivalence; see (5.10) and (5.11)), we see clearly that the two models assign different magnitudes to the (total) causal effect of X on Y: model M predicts that a unit change in x will change $E(Y)$ by the amount $\beta\alpha$, whereas model M' puts this amount at $\beta\alpha + \delta$.

At this point, it is tempting to ask whether we should substitute $x - \varepsilon_1$ for u in (5.9) prior to taking expectations in (5.17). If we permit the substitution of (5.8) into (5.9), as we did in deriving (5.17), why not permit the substitution of (5.7) into (5.9) as well? After all (the argument runs), there is no harm in upholding a mathematical equality, $u = x - \varepsilon_1$, that the modeler deems valid. This argument is fallacious, however.[15] Structural equations are not meant to be treated as immutable mathematical equalities. Rather, they are meant to define a state of equilibrium – one that is *violated* when the equilibrium is perturbed by outside interventions. In fact, the power of structural equation models is

[15] Such arguments have led to Newcomb's paradox in the so-called evidential decision theory (see Section 4.1.1).

that they encode not only the initial equilibrium state but also the information necessary for determining which equations must be violated in order to account for a new state of equilibrium. For example, if the intervention consists merely of holding X constant at x, then the equation $x = u + \varepsilon_1$, which represents the preintervention process determining X, should be overruled and replaced with the equation $X = x$. The solution to the new set of equations then represents the new equilibrium. Thus, the essential characteristic of structural equations that sets them apart from ordinary mathematical equations is that the former stand not for one but for many sets of equations, each corresponding to a subset of equations taken from the original model. Every such subset represents some hypothetical physical reality that would prevail under a given intervention.

If we take the stand that the value of structural equations lies not in summarizing distribution functions but in encoding causal information for predicting the effects of policies (Haavelmo 1943; Marschak 1950; Simon 1953), it is natural to view such predictions as the proper generalization of structural coefficients. For example, the proper generalization of the coefficient β in the linear model M would be the answer to the control query, "What would be the change in the expected value of Y if we were to intervene and change the value of Z from z to $z + 1$?", which is different, of course, from the observational query, "What would be the difference in the expected value of Y if we were to *find* Z at level $z + 1$ instead of level z?" Observational queries, as we discussed in Chapter 1, can be answered directly from the joint distribution $P(x, y, z)$, while control queries require causal information as well. Structural equations encode this causal information in their syntax by treating the variable on the left-hand side of the equality sign as the effect and treating those on the right as causes. In Chapter 3 we distinguished between the two types of queries through the symbol $do(\cdot)$. For example, we wrote

$$E(Y \mid do(x)) \triangleq E\,[Y \mid do(X = x)] \tag{5.19}$$

for the controlled expectation and

$$E(Y \mid x) \triangleq E(Y \mid X = x) \tag{5.20}$$

for the standard conditional or observational expectation. That $E(Y \mid do(x))$ does not equal $E(Y \mid x)$ can easily be seen in the model of (5.7)–(5.9), where $E(Y \mid do(x)) = \alpha\beta x$ but $E(Y \mid x) = r_{YX}x = (\alpha\beta + y)\,x$. Indeed, the passive observation $X = x$ should not violate any of the equations, and this is the justification for substituting both (5.7) and (5.8) into (5.9) before taking the expectation.

In linear models, the answers to questions of direct control are encoded in the path (or structural) coefficients, which can be used to derive the total effect of any variable on another. For example, the value of $E(Y \mid do(x))$ in the model defined by (5.7)–(5.9) is $\alpha\beta x$, that is, x times the product of the path coefficients along the path $X \rightarrow Z \rightarrow Y$. Computation of $E(Y \mid do(x))$ would be more complicated in the nonparametric case, even if we knew the functions f_1, f_2, and f_3. Nevertheless, this computation is well defined; it requires the solution (for the expectation of Y) of a modified set of equations in which f_1 is "wiped out" and X is replaced by the constant x:

$$z = f_2(x, \varepsilon_2), \tag{5.21}$$

$$y = f_3(z, u, \varepsilon_3). \tag{5.22}$$

Thus, computation of $E(Y \mid do(x))$ requires evaluation of

$$E(Y \mid do(x)) = E\{f_3\ [f_2(x, \varepsilon_2), u, \varepsilon_3]\},$$

where the expectation is taken over U, ε_2, and ε_3. Remarkably, graphical methods perform this computation without knowledge of f_2, f_3, and $P(\epsilon_2, \epsilon_3, u)$ (Section 3.3.2).

This is indeed the essence of identifiability in nonparametric models. The ability to answer interventional queries *uniquely*, from the data and the graph, is precisely how Definition 3.2.3 interprets the identification of the causal effect $P(y \mid do(x))$. As we have seen in Chapters 3 and 4, that ability can be discerned graphically, almost by inspection, from the diagrams that accompany the equations.

5.4 SOME CONCEPTUAL UNDERPINNINGS

5.4.1 What Do Structural Parameters Really Mean?

Every student of SEM has stumbled on the following paradox at some point in his or her career. If we interpret the coefficient β in the equation

$$y = \beta x + \varepsilon$$

as the change in $E(Y)$ per unit change of X, then, after rewriting the equation as

$$x = (y - \varepsilon)/\beta,$$

we ought to interpret $1/\beta$ as the change in $E(X)$ per unit change of Y. But this conflicts both with intuition and with the prediction of the model: the change in $E(X)$ per unit change of Y ought to be *zero* if Y does not appear as an independent variable in the original, structural equation for X.

Teachers of SEM generally evade this dilemma via one of two escape routes. One route involves denying that β has any causal reading and settling for a purely statistical interpretation, in which β measures the reduction in the variance of Y explained by X (see, e.g., Muthen 1987). The other route permits causal reading of only those coefficients that meet the "isolation" restriction (Bollen 1989; James et al. 1982): the explanatory variable must be uncorrelated with the error in the equation. Because ε cannot be uncorrelated with both X and Y (or so the argument goes), β and $1/\beta$ cannot both have causal meaning, and the paradox dissolves.

The first route is self-consistent, but it compromises the founders' intent that SEM function as an aid to policy making and clashes with the intuition of most SEM users. The second is vulnerable to attack logically. It is well known that every pair of bivariate normal variables, X and Y, can be expressed in two equivalent ways,

$$y = \beta x + \varepsilon_1 \quad \text{and} \quad x = \alpha y + \varepsilon_2,$$

where $\text{cov}(X, \varepsilon_1) = \text{cov}(Y, \varepsilon_2) = 0$ and $\alpha = r_{XY} = \beta\sigma_X^2/\sigma_Y^2$. Thus, if the condition $\text{cov}(X, \varepsilon_1) = 0$ endows β with causal meaning, then $\text{cov}(Y, \varepsilon_2) = 0$ ought to endow α with causal meaning as well. But this too conflicts with both intuition and the intentions

behind SEM; the change in $E(X)$ per unit change of Y ought to be zero, not r_{XY}, if there is no causal path from Y to X.

What then *is* the meaning of a structural coefficient? Or a structural equation? Or an error term? The interventional interpretation of causal effects, when coupled with the $do(x)$ notation, provides simple answers to these questions. The answers explicate the operational meaning of structural equations and thus should end, I hope, an era of controversy and confusion regarding these entities.

Structural Equations: Operational Definition

Definition 5.4.1 (Structural Equations)

An equation $y = \beta x + \varepsilon$ *is said to be* structural *if it is to be interpreted as follows: In an ideal experiment where we control X to x and any other set Z of variables (not containing X or Y) to z, the value y of Y is given by* $\beta x + \varepsilon$, *where* ε *is not a function of the settings x and z.*

This definition is operational because all quantities are observable, albeit under conditions of controlled manipulation. That manipulations cannot be performed in most observational studies does not negate the operationality of the definition, much as our inability to observe bacteria with the naked eye does not negate their observability under a microscope. The challenge of SEM is to extract the maximum information concerning what we wish to observe from the little we actually can observe.

Note that the operational reading just given makes no claim about how X (or any other variable) will behave when we control Y. This asymmetry makes the equality signs in structural equations different from algebraic equality signs; the former act symmetrically in relating observations on X and Y (e.g., observing $Y = 0$ implies $\beta x = -\varepsilon$), but they act asymmetrically when it comes to interventions (e.g., setting Y to zero tells us nothing about the relation between x and ε). The arrows in path diagrams make this dual role explicit, and this may account for the insight and inferential power gained through the use of diagrams.

The strongest empirical claim of the equation $y = \beta x + \varepsilon$ is made by excluding other variables from the r.h.s. of the equation, thus proclaiming X the *only* immediate cause of Y. This translates into a testable claim of *invariance*: the statistics of Y under condition $do(x)$ should remain invariant to the manipulation of any other variable in the model (see Section 1.3.2).[16] This claim can be written symbolically as

$$P(y \mid do(x), do(z)) = P(y \mid do(x)) \tag{5.23}$$

for all Z disjoint of $\{X \cup Y\}$.[17] In contrast, regression equations make no empirical claims whatsoever.

---

[16] The basic notion that structural equations remain invariant to certain changes in the system goes back to Marschak (1950) and Simon (1953), and it has received mathematical formulation at various levels of abstraction in Hurwicz (1962), Mesarovic (1969), Sims (1977), Cartwright (1989), Hoover (1990), and Woodward (1995). The simplicity, precision, and clarity of (5.23) is unsurpassed, however.

[17] This claim is, in fact, only part of the message conveyed by the equation; the other part consists of a dynamic or counterfactual claim: If we were to control X to x' instead of x, then Y would attain

Note that this invariance holds relative to manipulations, not observations, of Z. The statistics of Y under condition $do(x)$ given the measurement $Z = z$, written $P(y \mid do(x), z)$, would certainly depend on z if the measurement were taken on a consequence (i.e., descendant) of Y. Note also that the ordinary conditional probability $P(y \mid x)$ does not enjoy such a strong property of invariance, since $P(y \mid x)$ is generally sensitive to manipulations of variables other than X in the model (unless X and ε are independent). Equation (5.23), in contrast, remains valid regardless of the statistical relationship between ε and X.

Generalized to a set of several structural equations, (5.23) explicates the assumptions underlying a given causal diagram. If G is the graph associated with a set of structural equations, then the assumptions are embodied in G as follows: (1) every missing arrow – say, between X and Y – represents the assumption that X has no causal effect on Y once we intervene and hold the parents of Y fixed; and (2) every missing bidirected link between X and Y represents the assumption that the omitted factors that (directly) influence X are uncorrected with those that (directly) influence Y. We shall define the operational meaning of the latter assumption in (5.25)–(5.27).

The Structural Parameters: Operational Definition

The interpretation of a structural equation as a statement about the behavior of Y under a hypothetical intervention yields a simple definition for the structural parameters. The meaning of β in the equation $y = \beta x + \varepsilon$ is simply

$$\beta = \frac{\partial}{\partial x} E[Y \mid do(x)], \tag{5.24}$$

that is, the rate of change (relative to x) of the expectation of Y in an experiment where X is held at x by external control. This interpretation holds regardless of whether ε and X are correlated in nonexperimental studies (e.g., via another equation $x = \alpha y + \delta$).

We hardly need to add at this point that β has nothing to do with the regression coefficient r_{YX} or, equivalently, with the conditional expectation $E(Y \mid x)$, as suggested in many textbooks. The conditions under which β coincides with the regression coefficient are spelled out in Theorem 5.3.1.

It is important nevertheless to compare the definition of (5.24) with theories that acknowledge the invariant character of β but have difficulties explicating which changes β is invariant to. Cartwright (1989, p. 194), for example, characterizes β as an invariant of nature that she calls "capacity." She states correctly that β remains constant under change but explains that, as the statistics of X changes, "it is the ratio $[\beta = E(YX)/E(X^2)]$ which remains fixed no matter how the variances shift." This characterization is imprecise on two accounts. First, β may in general not be equal to the stated ratio nor to any other combination of statistical parameters. Second – and this is the main point of Definition 5.4.1 – structural parameters are invariant to local interventions (i.e., changes in

---

the value $\beta x' + \varepsilon$. In other words, plotting the value of Y under various hypothetical controls of X, and under the same external conditions (ε), should result in a straight line with slope β. Such deterministic dynamic claims concerning system behavior under successive control conditions can be tested only under the assumption that ε, representing external conditions or properties of experimental units, remains unaltered as we switch from x to x'. Such counterfactual claims constitute the empirical content of every scientific law (see Section 7.2.2).

specific equations in the system) and not to general changes in the statistics of the variables. If we start with $\text{cov}(X, \varepsilon) = 0$ and the variance of X changes because we (or Nature) locally modify the *process* that generates X, then Cartwright is correct; the ratio $\beta = E(YX)/E(X^2)$ will remain constant. However, if the variance of X changes for any other reason – say, because we observed some evidence $Z = z$ that depends on both X and Y or because the process generating X becomes dependent on a wider set of variables – then that ratio will not remain constant.

The Mystical Error Term: Operational Definition

The interpretations given in Definition 5.4.1 and (5.24) provide an operational definition for that mystical error term

$$\varepsilon = y - E[Y \mid do(x)], \tag{5.25}$$

which, despite being unobserved in nonmanipulative studies, is far from being metaphysical or definitional as suggested by some researchers (e.g. Richard 1980; Holland 1988, p. 460; Hendry 1995, p. 62). Unlike errors in regression equations, ε measures the deviation of Y from its controlled expectation $E[Y \mid do(x)]$ and not from its conditional expectation $E[Y \mid x]$. The statistics of ε can therefore be measured from observations on Y once X is controlled. Alternatively, because β remains the same regardless of whether X is manipulated or observed, the statistics of $\varepsilon = y - \beta x$ can be measured in observational studies if we know β.

Likewise, correlations among errors can be estimated empirically. For any two nonadjacent variables X and Y, (5.25) yields

$$E[\varepsilon_Y \varepsilon_X] = E[YX \mid do(pa_Y, pa_X)] - E[Y \mid do(pa_Y)]E[X \mid do(pa_X)]. \tag{5.26}$$

Once we have determined the structural coefficients, the controlled expectations $E[Y \mid do(pa_Y)]$, $E[X \mid do(pa_X)]$, and $E[YX \mid do(pa_Y, pa_X)]$ become known linear functions of the observed variables pa_Y and pa_X; hence, the expectations on the r.h.s. of (5.26) can be estimated in observational studies. Alternatively, if the coefficients are not determined, then the expression can be assessed directly in interventional studies by holding pa_X and pa_Y fixed (assuming X and Y are not in parent–child relationship) and estimating the covariance of X and Y from data obtained under such conditions.

Finally, we are often interested not in assessing the numerical value of $E[\varepsilon_Y \varepsilon_X]$ but rather in determining whether ε_Y and ε_X can be assumed to be uncorrected. For this determination, it suffices to test whether the equality

$$E[Y \mid x, do(s_{XY})] = E[Y \mid do(x), do(s_{XY})] \tag{5.27}$$

holds true, where s_{XY} stands for (any setting of) all variables in the model excluding X and Y. This test can be applied to any two variables in the model *except* when Y is a parent of X, in which case the symmetrical equation (with X and Y interchanged) is applicable.

The Mystical Error Term: Conceptual Interpretation

The authors of SEM textbooks usually interpret error terms as representing the influence of omitted factors. Many SEM researchers are reluctant to accept this interpretation,

however, partly because unspecified omitted factors open the door to metaphysical speculations and partly because arguments based on such factors were improperly used as a generic, substance-free license to omit bidirected arcs from path diagrams (McDonald 1997). Such concerns are answered by the operational interpretation of error terms, (5.25), since it prescribes how errors are measured, not how they originate.

It is important to note, though, that this operational definition is no substitute for the omitted-factors conception when it comes to deciding whether pairs of error terms can be assumed to be uncorrected. Because such decisions are needed at a stage when the model's parameters are still "free," they cannot be made on the basis of numerical assessments of correlations but must rest instead on qualitative structural knowledge about how mechanisms are tied together and how variables affect each other. Such judgmental decisions are hardly aided by the operational criterion of (5.26), which instructs the investigator to assess whether two deviations – taken on two different variables under complex experimental conditions – would be correlated or uncorrected. Such assessments are cognitively unfeasible.

In contrast, the omitted-factors conception instructs the investigator to judge whether there could be factors that simultaneously influence several observed variables. Such judgments are cognitively manageable because they are qualitative and rest on purely structural knowledge – the only knowledge available during this phase of modeling.

Another source of error correlation that should be considered by investigators is *selection bias*. If two uncorrected unobserved factors have a common effect that is omitted from the analysis but influences the selection of samples for the study, then the corresponding error terms will be correlated in the sampled population; hence, the expectation in (5.26) will not vanish when taken over the sampled population (see discussion of Berkson's paradox in Section 1.2.3).

We should emphasize, however, that the arcs *missing* from the diagram, not those *in* the diagram, demand the most attention and careful substantive justification. Adding an extra bidirected arc can at worst compromise the identifiability of parameters, but deleting an existing bidirected arc may produce erroneous conclusions as well as a false sense of model testability. Thus, bidirected arcs should be assumed to exist, by default, between any two nodes in the diagram. They should be deleted only by well-motivated justifications, such as the unlikely existence of a common cause for the two variables and the unlikely existence of selection bias. Although we can never be cognizant of all the factors that may affect our variables, substantive knowledge sometimes permits us to state that the influence of a possible common factor is not likely to be significant.

Thus, as often happens in the sciences, the way we measure physical entities does not offer the best way of thinking about them. The omitted-factor conception of errors, because it rests on structural knowledge, is a more useful guide than the operational definition when building, evaluating, and thinking about causal models.