



Figure 6.4: Z is associated with both X and Y , yet the effect of X on Y is not confounded (when $r = -\alpha\gamma$).

where $\text{cov}(x, \epsilon) = 0$. Thus, whenever the equality $r = -\alpha\gamma$ holds, the regression coefficient of $r_{YX} = \beta + r + \alpha\gamma$ is an unbiased estimate of β , meaning that the effect of X on Y is unconfounded (no adjustment is necessary). Yet the associational conditions (U_1) and (U_2) are both violated by the variable Z ; Z is associated with X (if $\alpha \neq 0$) and conditionally associated with Y , given X (except for special values of γ for which $\rho_{yz \cdot x} = 0$).

This example demonstrates that the condition of unbiasedness (Definition 6.2.1) does not imply the modified criterion of Definition 6.3.2. The associational criterion might falsely classify some unconfounded situations as confounded and, worse yet, adjusting for the false confounder (Z in our example) will introduce bias into the effect estimate.¹⁵

6.4 Stable versus Incidental Unbiasedness

6.4.1 Motivation

The failure of the associational criterion in the previous example calls for a reexamination of the notion of confounding and unbiasedness as defined in (6.10). The reason that X and Y were classified as unconfounded in Example 6.3.3 was that, by setting $r = -\alpha\gamma$, we were able to make the spurious association represented by r *cancel* the one mediated by Z . In practice, such perfect cancelation would be an incidental event specific to a peculiar combination of study conditions, and it would not persist when the parameters of the problem (i.e., α , γ , and r) undergo slight changes—say, when the study is repeated in a different location or at a different time. In contrast, the condition of no-confounding found in Example 6.3.1 does not exhibit such volatility. In this example, the unbiasedness expressed in (6.10) would continue to hold regardless of the strength of connection between education and exposure and regardless on how education and age influence the type of car that a patient owns. We call this type of unbiasedness *stable*, since it is robust to change in parameters and remains intact as long as the configuration of causal connections in the model remains the same.

¹⁵Note that the Stone-Robins modifications of Definition 6.3.2 would also fail in this example, unless we can measure the factors responsible for the correlation between ϵ_1 and ϵ_2 .

In light of this distinction between stable and incidental unbiasedness, we need to reexamine whether we should regard a criterion as inadequate if it misclassifies (as confounded) cases that are rendered unconfounded by mere incidental cancelation and, more fundamentally, whether we should insist on including such peculiar cases in the definition of unbiasedness (given the precarious conditions under which (6.10) would be satisfied in these cases). Although answers to these questions are partly a matter of choice, there is ample evidence that our intuition regarding confounding is driven by considerations of stable unbiasedness, not merely incidental ones. How else can we explain why generations of epidemiologists and biostatisticians would advocate confounding criteria that fail in cases involving incidental cancelation? On the pragmatic side, failing to detect situations of incidental unbiasedness should not introduce appreciable error in observational studies because those situations are short-lived and are likely to be refuted by subsequent studies, under slightly different conditions.¹⁶

Assuming that we are prepared to classify as unbiased only cases in which unbiasedness remains robust to changes in parameters, two questions remain: (1) How can we give this new notion of “stable unbiasedness” a formal, nonparametric formulation? (2) Are practical statistical criteria available for testing stable unbiasedness? Both questions can be answered using structural models.

Chapter 3 describes a graphical criterion, called the “back-door criterion,” for identifying conditions of unbiasedness in a causal diagram.¹⁷ In the simple case of no adjustment (for measured covariates), the criterion states that X and Y are unconfounded if every path between X and Y that contains an arrow pointing into X must also contain a pair of arrows pointing head-to-head (as in Figure 6.3); this criterion is valid whenever the missing links in the diagram represent absence of causal connections among the corresponding variables. Because the causal assumptions embedded in the missing links are so explicit, the back-door criterion has two remarkable features. First, no statistical information is needed; the topology of the diagram suffices for reliably determining whether an effect is unconfounded (in the sense of Definition 6.2.1) and whether an adjustment for a set of variables is sufficient for removing confounding when one exists. Second, any model that meets the back-door criterion would in fact satisfy (6.10) for an infinite class of models (or situations), each generated by assigning different parameters to the causal connections in the diagram.

To illustrate, consider the diagram depicted in Figure 6.3. The back-door criterion will identify the pair (X, Y) as unconfounded, because the only path ending with an arrow into X is the one traversing (X, E, Z, A, Y) , and this path contains two arrows pointing head-to-head at Z . Moreover, since the criterion is based only on graphical relationships, it is clear that (X, Y) will continue to be classified as unconfounded regardless of the strength or type of causal relationships that are represented by the arrows in the diagram. In contrast, consider Figure 6.4 in Example 6.3.3, where two paths end with arrows into X . Since none of these paths contains head-to-head arrows, the back-door criterion will fail to classify the effect of X on Y as unconfounded, acknowledging that an equality $r = -\alpha\gamma$ (if it prevails) would not represent a stable

¹⁶As we have seen in Example 6.3.3, any statistical test capable of recognizing such cases would require measurement of *all* variables in T .

¹⁷A gentle introduction to applications of the back-door criterion in epidemiology can be found in Greenland et al. (1999a).

case of unbiasedness.

The vulnerability of the back-door criterion to causal assumptions can be demonstrated in the context of Figure 6.3. Assume the investigator suspects that variable Z (car type) has some influence on the outcome variable Y . This would amount to adding an arrow from Z to Y in the diagram, classifying the situation as confounded, and suggesting an adjustment for E (or $\{A, Z\}$). Yet no adjustment is necessary if, owing to the specific experimental conditions in the study, Z has in fact no influence on Y . It is true that the adjustment suggested by the back-door criterion would introduce no bias, but such adjustment could be costly if it calls for superfluous measurements in a non-confounding situation.¹⁸ The added cost is justified in light of (i) the causal information at hand (i.e., that Z may potentially influence Y) and (ii) insistence on ensuring stable unbiasedness—that is, avoiding bias in all situations compatible with the information at hand.

6.4.2 Formal Definitions

To formally distinguish between *stable* and *incidental* unbiasedness, we use the following general definition.

Definition 6.4.1 (Stable Unbiasedness)

Let A be a set of assumptions (or restrictions) on the data-generating process, and let C_A be a class of causal models satisfying A . The effect estimate of X on Y is said to be stably unbiased given A if $P(y|do(x)) = P(y|x)$ holds in every model M in C_A . Correspondingly, we say that the pair (X, Y) is stably unconfounded given A .

The assumptions commonly used to specify causal models can be either parametric or topological. For example, the structural equation models used in the social sciences and economics are usually restricted by the assumptions of linearity and normality. In this case, C_A would consist of all models created by assigning different values to the unspecified parameters in the equations and in the covariance matrix of the error terms. Weaker, nonparametric assumptions emerge when we specify merely the topological structure of the causal diagram but let the error distributions and the functional form of the equations remain undetermined. We now explore the statistical ramifications of these nonparametric assumptions.

Definition 6.4.2 Structurally Stable No-Confounding

Let A_D be the set of assumptions embedded in a causal diagram D . We say that X and Y are stably unconfounded given A_D if $P(y|do(x)) = P(y|x)$ holds in every parameterization of D . By “parameterization” we mean an assignment of functions to the links of the diagram and prior probabilities to the background variables in the diagram.

¹⁸On the surface, it appears as though the Stone-Robins criterion would correctly recognize the absence of confounding in this situation, since it is based on associations that prevail in the probability distribution that actually generates the data (according to which $\{E, Z\}$ should be independent of Y given $\{A, X\}$). However, these associations are of no help in deciding whether certain measurements can be *avoided*; such decisions must be made prior to gathering the data and must rely therefore on subjective assumptions about the disappearance of conditional associations. Such assumptions are normally supported by causal, not associational, knowledge (see Section 1.3).

Explicit interpretation of the assumptions embedded in a causal diagram are given in Chapters 3 and 5. Put succinctly, if D is the diagram associated with the causal model, then:

1. every missing arrow (between, say X and Y) represents the assumption that X has no effect on Y once we intervene and hold the parents of Y fixed;
2. every missing bidirected link between X and Y represents the assumption that there are no common causes for X and Y , except those shown in D .

Whenever the diagram D is acyclic, the back-door criterion provides a necessary and sufficient test for stable no-confounding, given A_D . In the simple case of no adjustment for covariates, the criterion reduces to the nonexistence of a common ancestor, observed or latent, of X and Y .¹⁹ Thus, we have our next theorem.

Theorem 6.4.3 (Common-Cause Principle)

Let A_D be the set of assumptions embedded in an acyclic causal diagram D . Variables X and Y are stably unconfounded given A_D if and only if X and Y have no common ancestor in D .

Proof

The “if” part follows from the validity of the back-door criterion (Theorem 3.3.2). The “only if” part requires the construction of a specific model in which (6.10) is violated whenever X and Y have a common ancestor in D . This is easily done using linear models and Wright’s rules for path coefficients. \square

Theorem 6.4.3 provides a necessary and sufficient condition for stable no-confounding without invoking statistical data, since it relies entirely on the information embedded in the diagram. Of course, the diagram itself has statistical implications that can be tested (Sections 1.2.3 and 5.2.1), but those tests do not specify the diagram uniquely (see Chapter 2 and Section 5.2.3).

Suppose, however, that we do not possess all the information required for constructing a causal diagram and instead know merely for each variable Z whether it is safe to assume that Z has no effect on Y and whether X has no effect on Z . The question is now whether this more modest information, together with statistical data, is sufficient to qualify or disqualify a pair (X, Y) as stably unconfounded. The answer is positive.

6.4.3 Operational Test for Stable No-Confounding

Theorem 6.4.4 (Criterion for Stable No-Confounding)

Let A_Z denote the assumptions that (i) the data are generated by some (unspecified) acyclic model M and (ii) Z is a variable in M that is unaffected by X but may possibly

¹⁹The colloquial term “common ancestors” should exclude nodes that have no other connection to Y except through X (e.g., node E in Figure 6.3) and include latent nodes for correlated errors. In the diagram of Figure 6.4, for example, X and Y are understood to have two common ancestors; the first is Z and the second is the (implicit) latent variable responsible for the double-arrowed arc between X and Y (i.e., the correlation between ϵ_1 and ϵ_2).

affect Y .²⁰ If both of the associational criteria (U_1) and (U_2) of Definition 6.2.2 are violated, then (X, Y) are not stably unconfounded given A_Z .

Proof

Whenever X and Y are stably unconfounded, Theorem 6.4.3 rules out the existence of a common ancestor of X and Y in the diagram associated with the underlying model. The absence of a common ancestor, in turn, implies the satisfaction of either (U_1) or (U_2) whenever Z satisfies A_Z . This is a consequence of the d -separation rule (Section 1.2.3) for reading the conditional independence relationships entailed by a diagram.²¹ \square

Theorem 6.4.4 implies that the traditional associational criteria (U_1) and (U_2) could be used in a simple operational test for stable no-confounding, a test that does not require us to know the causal structure of the variables in the domain or even to enumerate the set of relevant variables. Finding just *any* variable Z that satisfies A_Z and violates (U_1) and (U_2) permits us to disqualify (X, Y) as stably unconfounded (though (X, Y) may be incidentally unconfounded in the particular experimental conditions prevailing in the study).

Theorem 6.4.4 communicates a formal connection between statistical associations and confounding that is not based on the closed-world assumption.²² It is remarkable that the connection can be formed under such weak set of added assumptions: the qualitative assumption that a variable may have influence on Y and is not affected by X suffices to produce a necessary statistical test for stable no-confounding.

6.5 Confounding, Collapsibility, and Exchangeability

6.5.1 Confounding and Collapsibility

Theorem 6.4.4 also establishes a formal connection between confounding and “collapsibility”—a criterion under which a measure of association remains invariant to the omission of certain variables.

Definition 6.5.1 (Collapsibility)

Let $g[P(x, y)]$ be any functional²³ that measures the association between Y and X in the joint distribution $P(x, y)$. We say that g is collapsible on a variable Z if

$$E_z g[P(x, y|z)] = g[P(x, y)].$$

²⁰By “possibly affecting Y ” we mean: A_Z does not contain the assumption that Z does not affect Y . In other words, the diagram associated with M must contain a directed path from Z to Y .

²¹It also follows from Theorem 7(a) in Robins (1997).

²²I am not aware of another such connection in the literature.

²³A *functional* is an assignment of a real number to any function from a given set of functions. For example, the mean $E(x) = \sum_x xP(x)$ is a functional, since it assigns a real number $E(X)$ to each probability function $P(x)$.