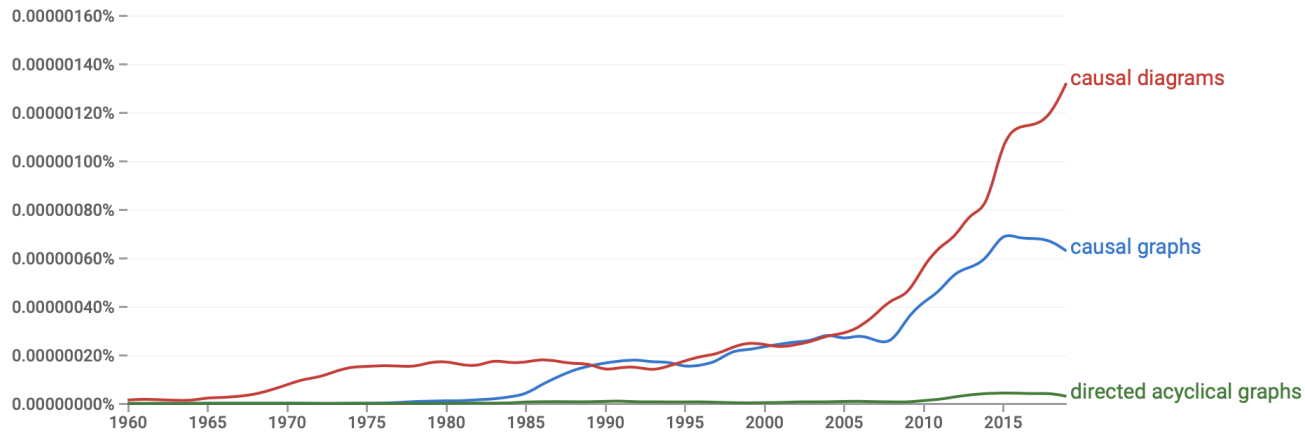
Warm-Up

`Coding Time: simpson_paradox`

Introduction

A graphical approach to causal modeling is not new. For example, in the early twentieth century Sewall Wright, one of the fathers of modern genetics, invented the path analysis. It was later adapted to economic modeling by Wright's *father* Philip Wright. The use of graphical modeling for causal inference has been largely ignored by economists, with a few exceptions. However, it was revitalized after a 2000 book titled *Causality: Models, Reasoning, and Inference* by computer scientist and Turing Award winner Judea Pearl, who adapted graphical modeling for studying artificial intelligence. Graphical models are immensely helpful for designing a credible identification strategy. We will not cover the entire book, of course, but only the most basic elements of Pearl's theory.



Causal diagrams grow in popularity

Pearl's work provides a language and a framework for thinking about causality that differs from the potential outcome framework. Pearl proves that the fundamental concepts underlying the potential outcome framework and his perspective are equivalent. In this lecture, we suppress potential outcome random variables and use only observed outcome variables. Pearl has shown that graphs nonetheless provide a direct and powerful way of thinking about causality and the identification strategies. We also implicitly focus on only the unconditional average treatment effect.

Pearl shows that there are three basic strategies for identifying a causal effect. In this lecture, we will briefly preview all three and talk more specifically about the first one. The other two will be covered in later lectures. The strategies are

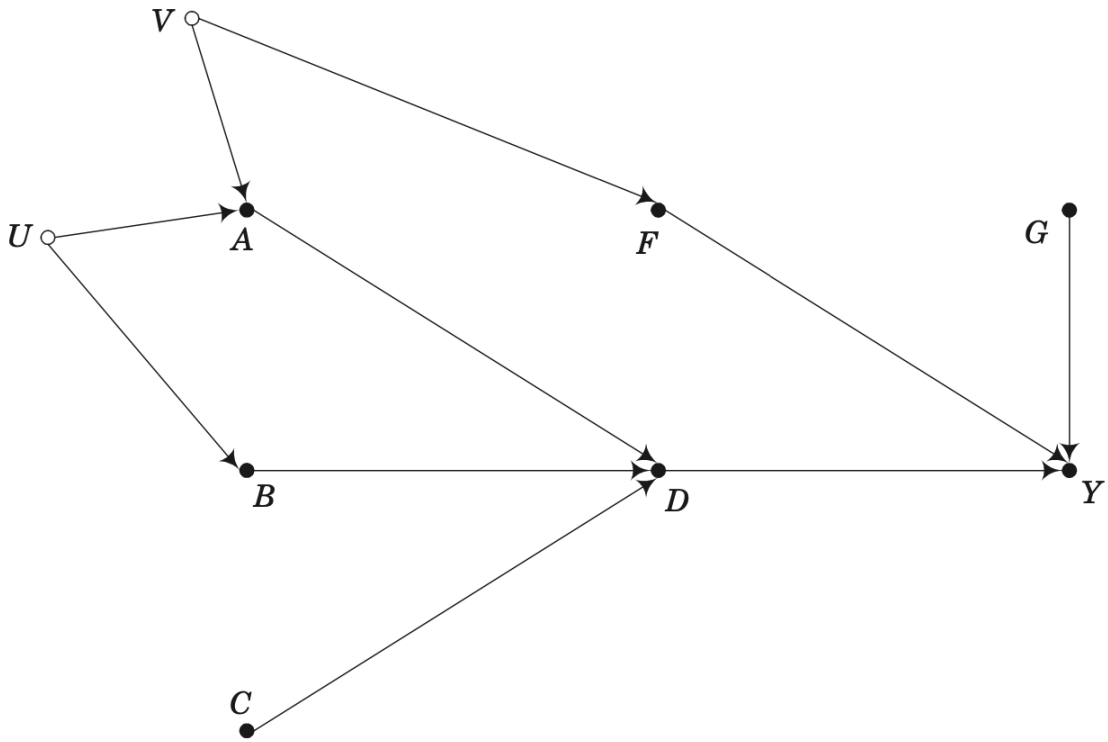
1. Conditioning on variables that block all back-door paths
2. Conditioning on variables that allow for estimation by a mechanism

3. Using an instrumental variable that is an exogenous shock to the cause

Basics of DAGs

No, not these DAGs.

Consider the causal relationships depicted on the following figure. This is known as a *causal diagram* or a *directed acyclical graph*.



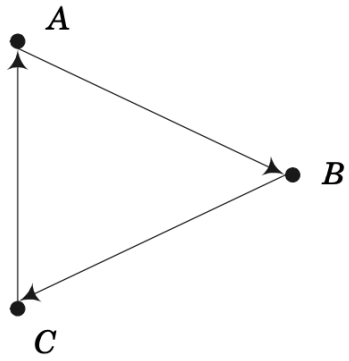
Each node represents a random variable and is labeled by a letter (or an explicit name of the variable). Nodes that are represented by a solid circle • are observed random variables. Nodes that are represented by a hollow circle ◦ are unobserved random variables. Causes are represented by directed edges → (single-headed arrows). An arrow from one node to another signifies that the variable at the origin causes the variable at the terminus.

A DAG does not permit a representation of simultaneous causation. Only directed edges are permissible, and direct causation can run in only one direction, as in $X \to Y$. The arrows do not tell anything about the size or the shape of the effect. $X \to Y$ simply means that X causes Y, without specifying whether the effect is, e.g., linear or quadratic.

A path is any any sequence of edges pointing in any direction that connects one variable to another. Notice that a path does not have to follow the direction of arrows. A *path*, in other words, means an *undirected* path.

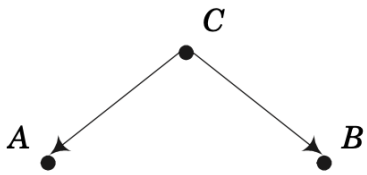
A causal diagram is an acyclic graph. No directed paths emanating from a causal variable also terminate at the same causal variable. You cannot, e.g., have graphs like

this one:

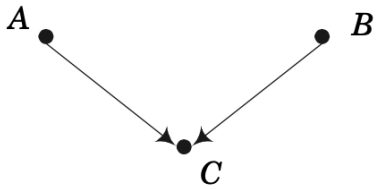


The figure below presents the three basic patterns of causal relationships among any three variables that are related to each other: a chain of mediation, a fork of mutual dependence, and an inverted fork of mutual causation.

(a) Mediation



(b) Mutual dependence



(c) Mutual causation

For the graph in panel (a), A affects B through A's causal effect on C and C's causal effect on B. This type of a causal chain renders the variables A and B unconditionally associated. The unconditional association between A and B means that knowing the value of A gives one some information on the likely value of B.

For the graph in panel (b), A and B are both caused by C. Here, A and B are also unconditionally associated, but now it is because they mutually depend on C. Here the

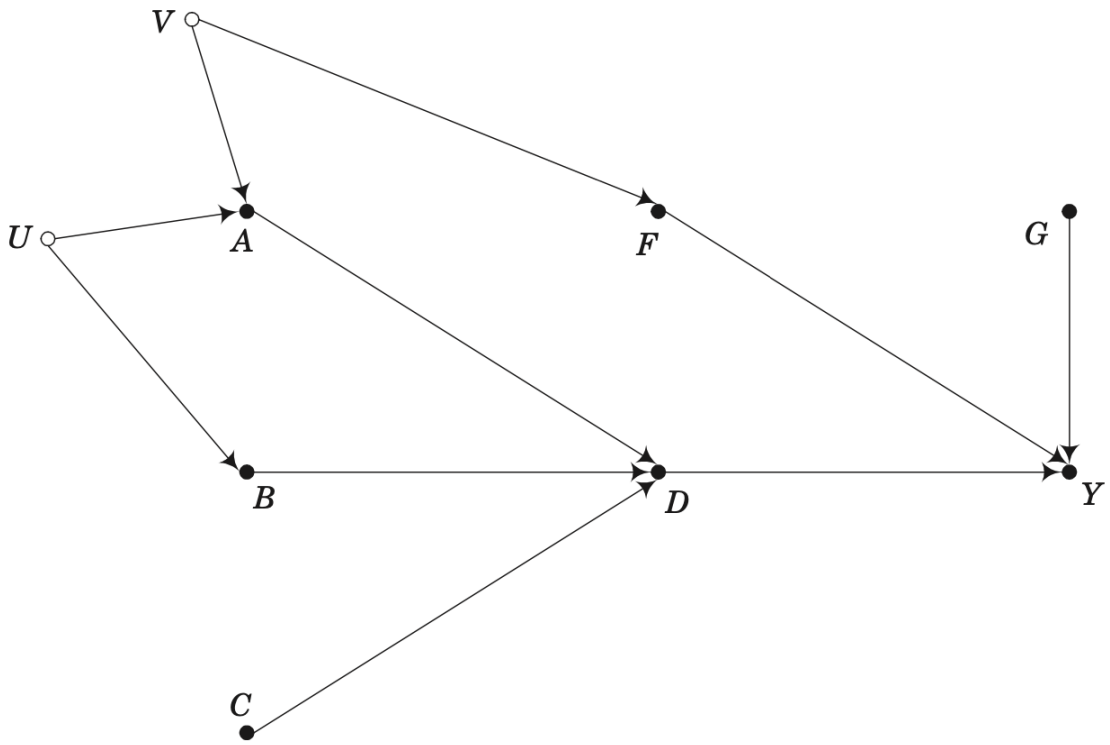
unconditional association between A and B is indirect, as neither A nor B has a direct causal effect on each other.

For the third graph in panel (c), A and B are connected by a pathway through C. But now A and B are both causes of C. C is called a "collider" variable. A variable is a collider along a particular path if it has two arrows running into it. Figuratively, the causal effects of A and B "collide" with each other at C.

A path that is connected by a collider variable does not generate an unconditional association between the variables that cause the collider variable. In panel (c), the pathway between A and B through C does not generate an unconditional association between A and B. If nothing is known about the value that C takes on, then knowing the value that A takes on yields no information about the value that B takes on. The path $A \rightarrow C \leftarrow B$ does not generate an association between A and B because the collider variable C “blocks” the possible causal effects of A and B on each other.

Even though collider variables do not generate unconditional associations between the variables that determine them, the incautious handling of colliders can create conditional dependence that can sabotage a causal analysis. The importance of considering collider variables is a key insight of Pearl’s framework.

Let's get back to our graph



Now, suppose that the causal variable of primary interest is D and that the causal effect that we wish to estimate is the effect of D on Y. Given the structure of causal

relationships represented in the graph, which variables must we observe and then use in our analysis to estimate the causal effect of D on Y?

> In Pearl's framework, the causal variable D has a probability distribution. The causal effects emanating from the variables A, B, and C are explicitly represented in the graph by directed edges, but the relative sizes of these effects are not represented in the graph. Other causes of D that are unrelated to A, B, and C are left implicit. The outcome variable, Y, is likewise caused by F , G, and D, but there are other implicit causes that are unrelated to F, G, and D that give Y its probability distribution.

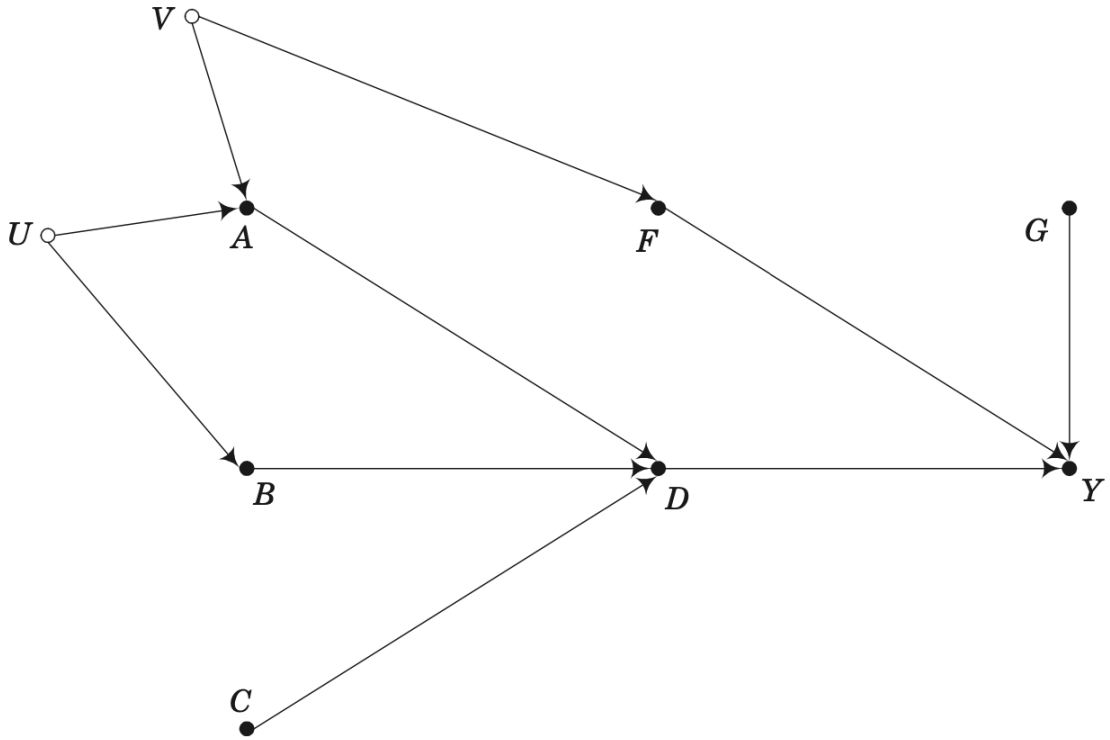
The Three Strategies

There are three basic strategies for estimating causal effects.

1. One can condition on variables (with procedures such as stratification, matching, weighting, or regression) that block all back-door paths from the causal variable to the outcome variable.
2. One can use exogenous variation in an appropriate instrumental variable to isolate covariation in the causal and outcome variables.
3. One can establish an isolated and exhaustive mechanism that relates the causal variable to the outcome variable and then calculate the causal effect as it propagates through the mechanism.

Conditioning

Consider again our graph and the opportunities it presents to estimate the causal effect of D on Y with the conditioning strategy.



First note that there are two back-door paths from D to Y in the graph. They generate a supplemental noncausal association between D and Y:

1. $D \leftarrow A \leftarrow V \rightarrow F \rightarrow Y$
2. $D \leftarrow B \leftarrow U \rightarrow A \leftarrow V \rightarrow F \rightarrow Y$

What is a *back-door path*? A back-door path is a path between any causally ordered sequence of two variables that includes a directed edge → that points to the first variable.

Both of these back-door paths can be blocked by conditioning on A and B *or* by conditioning on F. These two conditioning strategies will succeed in identifying the causal effect of D on Y under a variety a of conditioning techniques and in the presence of nonlinear effects. They are minimally sufficient in the sense that one can condition on any subset of the observed variables in $\{A, B, C, F, G\}$ as long as the subset includes either $\{A, B\}$ or $\{F\}$.

> One cannot identify the causal effect of D on Y by conditioning only on A. Conditioning only on A, which is a collider, creates dependence between B and F. As a result, conditioning only on A fails to block all back-door paths from D to Y.

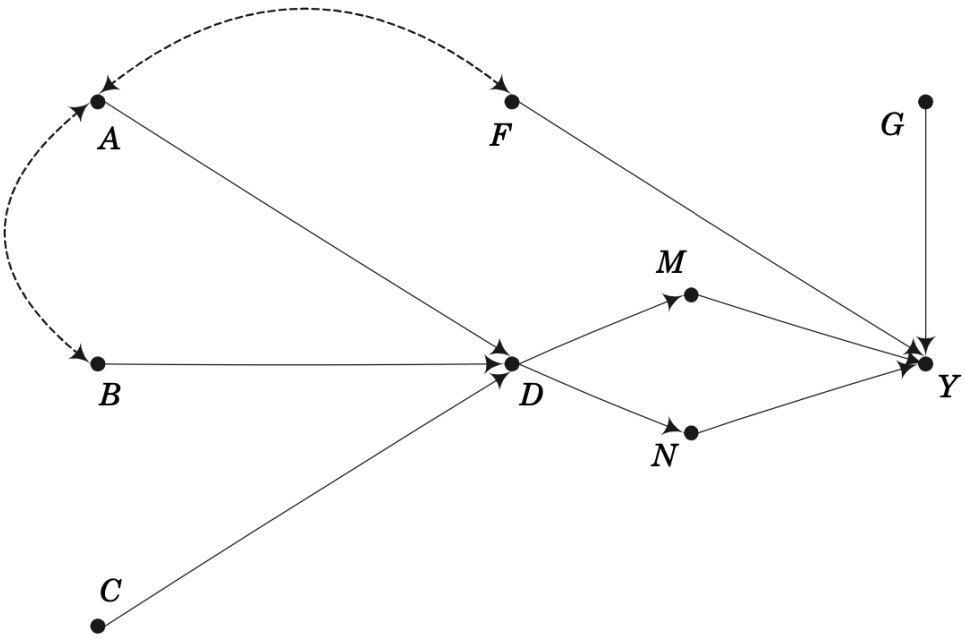
Instrumental Variables

The goal here is not to block back-door paths from the causal variable to the outcome variable but rather to use a localized exogenous shock to both the causal variable and the outcome variable in order to estimate indirectly the relationship between the two.

The variable C is a *valid* instrument for D because it causes D but does not have an effect on Y except though its effect on D. As a result, one can estimate the causal effect of D on Y by taking the ratio of the relationships between C and Y and between C and D. For this estimation strategy, A, B, F, G do not need to be observed.

Front Door Method

Consider this graph



The mediating variables M and N completely account for the causal effect of D on Y, and M and N are not determined by anything other than D. The causal effect of D on Y can also be calculated by estimation of the causal effect of D on M and N and then subsequently the causal effects of M and N on Y. The variables A, B, C, F, G can be ignored.

> There is nothing special about two mediating variables. It could be just one variable, or more than two.

Discussion

In an ideal scenario, all three strategies could be used to obtain the causal effects, and all three would generate equivalent estimates. If a causal effect estimate generated by conditioning on variables that block all back-door paths is similar to a causal effect estimate generated by a valid instrumental variable estimator, then each estimate is bolstered. Better yet, if a front door method then generates a third equivalent estimate, all three causal effect estimates would be even more convincing. And, in this case, an elaborated explanation of how the causal effect comes about is also available, as a

researcher could then describe how the causal effect is propagated through the intermediate variables M and N.

The presentation of causal effect estimation is, of course, a simplification. It is rare that one can specify causes as cleanly as in the causal diagrams in these figures. Estimating causal effects is typically more challenging.

Nonetheless, beyond introducing the basic estimation techniques, these simple graphs convey two important sets of points. First, there is often more than one way to estimate a causal effect, and simple rules such as "control for all other causes of the outcome variable" can be poor guides for practice. For example, for our graph there are two completely different and plausible conditioning strategies: either condition on F or on A and B. The strategy to "control for all other causes of the outcome variable" is misleading because (1) it suggests that one should condition on G as well, which is unnecessary if all one wants to obtain is the causal effect of D on Y and (2) it does not suggest that one can estimate the causal effect of D on Y by conditioning on a subset of the variables that cause D. In this case, one can estimate the causal effect of D on Y without conditioning on any of the other causes of Y , but instead by conditioning on the variables that cause D. Even so, this last conditioning strategy should not be taken too far. One need not condition on C when also conditioning on both A and B. Not only is this unnecessary (just as for G with the other conditioning strategy), in doing so one fails to use C in its most useful way: as an instrumental variable.

Second, the strategies are not well suited to *discovering* the causes of outcomes and then comprehensively estimating the relative effects of all alternative causes. Consider again the question that we posed after introducing our graph. We asked a simpler version of the following question: Given the structure of causal relationships that relate A, B, C, D, F, G, and Y to each other, which variables must we observe and then use in our analysis to estimate the causal effect of D on Y? This sort of constrained question different from seeking to answer the more general question: What are the causes of Y? The methods that we will present are not irrelevant to this broader question, but they are designed to answer simpler subordinate questions.

If we had estimated the effect of D on Y by observing only A, B, D, Y and then conditioning on A and B, and if we then found that D had a small effect on Y, we would then want to observe both F and G and think further about whether what we considered to be common causes of both A and F might be known and observable after all. However, if we did not have a theory that suggested that F and G have causal effects on Y (i.e., and instead thought that D was the only systematic cause of Y), then determining that D has a small to nonexistent effect on Y would not help us to find any of the other causes of Y that may be important.

Where Does the Causal Diagram Come From

A causal diagram is a theoretical representation of the state-of-the-art knowledge about the phenomena you are studying. A causal diagram should describe all causal relationships relevant to the causal effect of interest. Drawing a diagram requires you to make the explicit commitment to a causal paths that might exist (what *is* on the diagram) and the complete commitment to the *lack of* causal paths (what *is not* on the diagram). Drawing a causal diagram requires you to make choices about which arrows to include and which arrows to exclude. The lack of an arrow necessarily means that you think there is meaningful causal effect in the data. Each choice, therefore, is an explicit assumption that you are making and which you should be able to defend.

To yield a credible identification strategy, a well-specified theory is needed to justify assumptions about underlying causal relationships. If theory is poorly specified, or divergent theories exist that support alternative assumptions about underlying causal relationships, then alternative causal effect estimates may be considered valid conditional on the validity of alternative maintained assumptions.

Conditioning on Observables

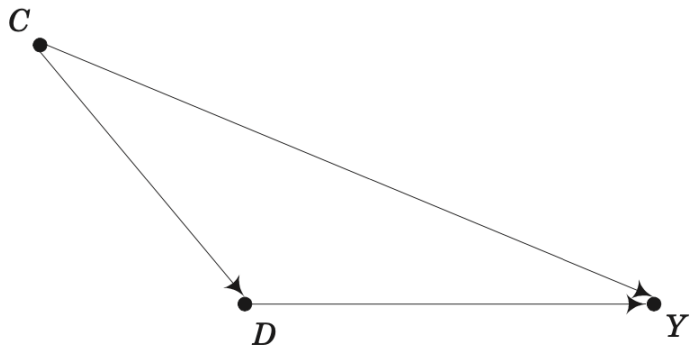
One of the most basic strategies to study causal questions in quantitative research is to analyze a relationship within groups defined by one or more variables. This approach is called subgroup analysis, subclassification, stratification, or tabular decomposition. The motivation is to analyze the data after conditioning on membership in groups identified by values of a variable that is thought to be related to both the causal variable and the outcome variable.

Suppose we are interested in the causal effect of A on B but are worried that C causes both. If analysis is carried out for a group in which all individuals have a particular value for the variable C, then the variable C is constant within the group and cannot therefore be associated with A or B. In practice, we will often use other conditioning methods, such as regression.

Conditioning is a powerful and general strategy. However, conditioning on a collider does not simplify the original graph but rather adds complications by creating new associations.
The reasoning here is not immediately intuitive. Consider a simple model in which $Y = A + B$. Thus A and B have a causal effect on Y, and for simplicity the relationship is deterministic. Suppose that A and B are independent in the population. Now if we fix Y at a level $\bar{Y}$, we get $\bar{Y} = A + B$. What does it do to A and B? It creates a deterministic and negative relationship between them! $A = \bar{Y} - B$. This example should remind you of level curves in math analysis or of iso-quants from your micro class.

The Back-Door Criterion

Perhaps the most general concern when studying causal effects is that the causal variable D and the outcome variable Y are mutually dependent on a common third variable C.



The total association between D and Y represents the genuine causal effect of D on Y and the common dependence of D and Y on C. In this case, it is often said that the causal effect of D on Y is confounded by C, or that C is a *confounder*. The causal effect of D on Y can be identified by conditioning on C.

Causal diagrams characterize the confounding using the language of *back-door paths*. In our graph, there are two paths that connect D and Y: $D \rightarrow Y$ and $D \leftarrow C \rightarrow Y$. The path $D \leftarrow C \rightarrow Y$ is a back-door path because it includes a directed edge pointing to D. The path $D \rightarrow Y$ is not a back-door path because it does not include a directed edge pointing to D.

> Note that sometimes you might have a direct causal effect of D on Y, $D \rightarrow Y$ and an indirect effect, mediated by another variable, say M: $D \rightarrow M \rightarrow Y$. This path is not considered a back-door path. But whether you want to block it or not depends on your research question.

The problem with back-door paths is that they may contribute to the association between D and Y. As a result, the observed association between D and Y may not consistently estimate the causal effect of D on Y. The observed association between D and Y does not identify the causal effect because the total association is a composite of the true causal effect $D \rightarrow Y$ and the back-door path $D \leftarrow C \rightarrow Y$.

We can use the so-called "*back-door criterion*" for determining whether or not conditioning on a given set of variables identifies the causal effect of interest. The causal effect is identified by conditioning on a set of variables Z if and only if all back-door paths between the causal variable and the outcome variable are blocked after conditioning on Z. All back-door paths are blocked by Z if and only if each back-door path (back-door criterion)

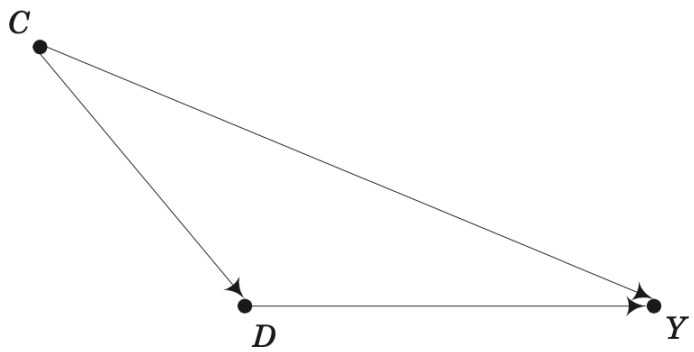
1. contains a chain of mediation $A \rightarrow C \rightarrow B$, where the middle variable C is in Z, or

2. contains a fork of mutual dependence $A \leftarrow C \rightarrow B$, where the middle variable C is in Z, or
3. contains an inverted fork of mutual causation $A \rightarrow C \leftarrow B$, where the middle variable C and all of C's descendants are not in Z.

Conditions 1 and 2 of the back-door criterion imply that back-door associations between the causal variable and the outcome variable can be eliminated by conditioning on observed variables that block each back-door path. Condition 3, however, is quite different and is not intuitive. It states that the set of conditioning variables Z cannot include collider variables that lie along back-door paths.

> Because the "or" in the back-door criterion is inclusive, one can condition on colliders and still satisfy the back-door criterion if the back-door paths along which the colliders lie are otherwise blocked because Z satisfies condition 1 or condition 2 with respect to another variable on the same back-door path.

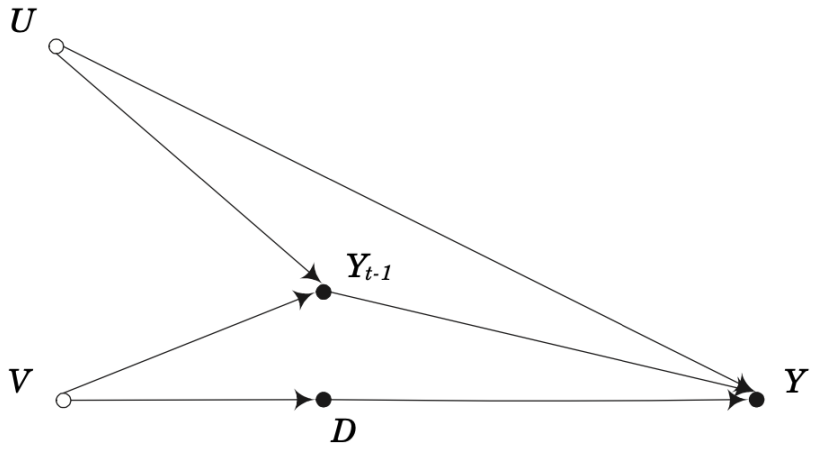
Consider our simple example again.



There is a single back-door path, which is a fork of mutual dependence where C causes both D and Y. Conditioning on C blocks $D \leftarrow C \rightarrow Y$ because C is the middle variable in a fork of mutual dependence. As a result, conditioning on C satisfies the back-door criterion, and identifies the causal effect of D on Y.

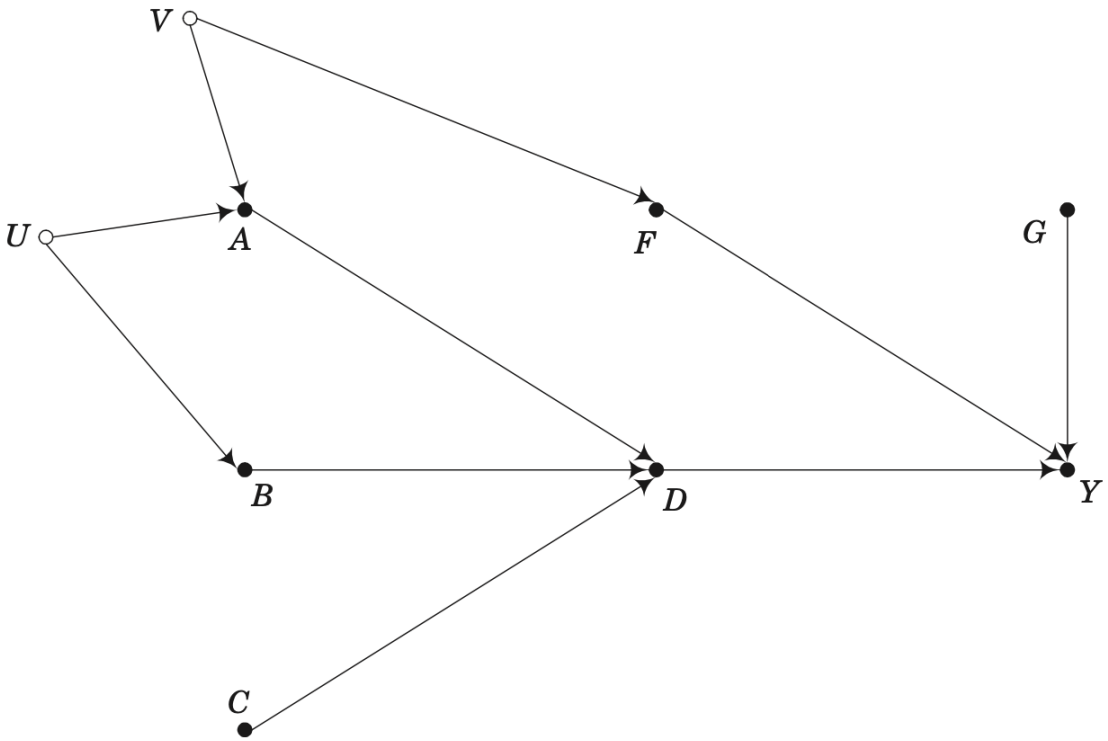
Consider another example. Suppose now that the confounder variable is unobserved, and thus cannot be conditioned on. Some researchers would argue that the effects of the unobserved confounder can be decomposed into a lagged process, using a prior

variable for the outcome, Y_{t-1}, and two separate unobserved variables, U and V



There are two back-door paths: $D \leftarrow V \rightarrow Y_{t-1} \rightarrow Y$ and $D \leftarrow V \rightarrow Y_{t-1} \leftarrow U \rightarrow Y$. The lagged outcome variable Y_{t-1} lies along both of these back-door paths, but Y_{t-1} does not satisfy the back-door criterion. It blocks the first back-door path $D \leftarrow V \rightarrow Y_{t-1} \rightarrow Y$ because, for this path, Y_{t-1} is the middle variable of a chain of mediation $V \rightarrow Y_{t-1} \rightarrow Y$. For the second path $D \leftarrow V \rightarrow Y_{t-1} \leftarrow U \rightarrow Y$, Y_{t-1} is a collider because it is the middle variable in an inverted fork of mutual causation $V \rightarrow Y_{t-1} \leftarrow U$. After conditioning on Y_{t-1}, at least one back-door path will remain unblocked. In other words, conditioning on Y_{t-1} blocks one back-door path but opens another one.

Consider our old example again.



Suppose, again, that we wish to estimate the causal effect of D on Y. For this DAG, there are two back-door paths between D and Y:

1. $D \leftarrow A \leftarrow V \rightarrow F \rightarrow Y$

2. $D \leftarrow B \leftarrow U \rightarrow A \leftarrow V \rightarrow F \rightarrow Y$

Notice that A is a collider variable in the second back-door path but not in the first back-door path. As a result, the first back-door path contributes to the association between D and Y, but the second back-door path does not. Whatever conditioning we enact to eliminate the confounding effect of the first back-door path does not unblock the second back-door path.

There are two different conditioning strategies available that will identify the causal effect (numbers 1 and 3) and a third one that may appear to work but that will fail (number 2):

1. F is the middle variable of a chain of mediation for both back-door paths $V \rightarrow F \rightarrow Y$. As a result, F satisfies the back-door criterion and conditioning on it identifies the causal effect.
2. A is a middle variable of a chain of mediation for the first back-door path $D \leftarrow A \leftarrow V$. But A is a collider variable for the second back-door path, as in $U \rightarrow A \leftarrow V$. As a result, A alone does not satisfy the back-door criterion. Conditioning on A would unblock the second back-door path.
3. A is a middle variable of a chain of mediation for the first back-door path, as in $D \leftarrow A \leftarrow V$. Likewise, B is a middle variable of a chain of mediation for the second back-door path, as in $D \leftarrow B \leftarrow U$. Thus, even though A blocks only the first back-door path and unblocks the second back-door path, conditioning on B blocks the second back-door path. As a result, A and B together satisfy the back-door criterion, and conditioning on them together identifies the causal effect of D on Y.

In sum, for this example the causal effect can be identified by conditioning in one of two minimally sufficient ways: either condition on F or condition on both A and B. One can of course condition in three additional ways that also satisfy the back-door criterion: F and A, F and B, and F, A, and B. Although these conditioning sets include unnecessary and redundant conditioning.

The key point of this section is that conditioning on variables that lie along back-door paths can be an effective strategy to identify a causal effect. If all back-door paths between the causal variable and the outcome variable are blocked after the conditioning is enacted, then back-door paths do not contribute to the association between the causal variable and the outcome variable. And, as a result, the remaining association between the causal variable and outcome variable identifies the causal effect. However, conditioning on a collider variable has the opposite effect. It unblocks an already blocked back-door path.

Pearl's back-door criterion for evaluating conditioning strategies is a generalization of various previous solutions for how to solve the omitted-variable bias problem. The back-door criterion shows that researchers do not need to condition on all omitted

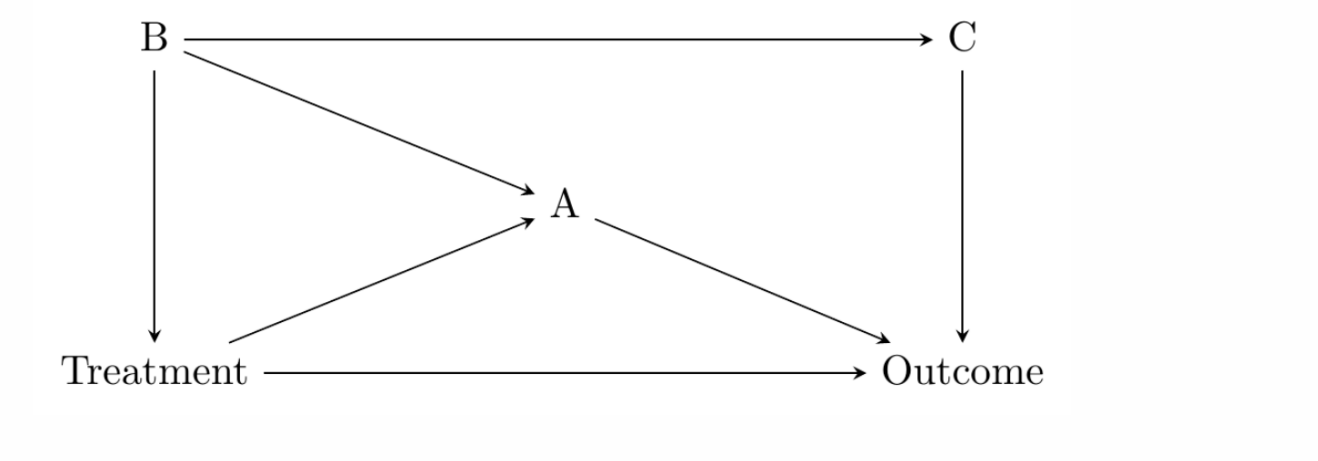
direct causes of an outcome variable. The back-door criterion shows clearly why researchers need to condition on only a minimally sufficient set of variables that renders all back-door paths blocked. Pearl's framework shows how to think clearly about the appropriateness of conditioning on endogenous variables. Just write down each back-door path and then determine whether or not each endogenous variable is a collider along any of these back-door paths.

How To Find Back-Door Paths

Sometimes finding all the back-door paths is easy. Sometimes it is not. How can we find every path from treatment to outcome? We can follow the following algorithm.

1. Start at the treatment variable
2. Follow one of the arrows coming in or out of the treatment variable to find another variable
3. Then, follow one of the arrows coming in or out of *that* variable
4. Keep repeating step 3 until you either come to a variable you've already visited (a loop) or find the outcome variable (a path, write it down)
5. Every time you either find a path or a loop, back up one and try a different arrow in/out until you have tried them all. Then, back up again and try all *those* arrows
6. Once you've tried all the ways out of the treatment variable and all the eventual paths, you've got all the paths!

Let's do a quick example with a simple diagram below



- Start at $Treatment$.
- Let's follow an arrow. Let's go straight to $Outcome$. $Treatment \rightarrow Outcome$ is a **path**.
- Back up to Treatment. Follow another arrow. This time to A.
- Now follow an arrow out of A. Let's go to $Outcome$. Done! $Treatment \rightarrow A \rightarrow Outcome$ is a **path**.
- Back up to A. Take the other arrow out to B.
- Where can we go from here? Only to C without repeating a variable.

- And from C we can only go to $Outcome$. $Treatment \rightarrow A \leftarrow B \rightarrow C \rightarrow Outcome$ is a **path**.
- Back up to B, nowhere else to go.
- Back up to A, and nowhere else to go.
- Back up to $Treatment$. The only arrow left is B.
- From B we can go to A, and then on to $Outcome$. $Treatment \leftarrow B \rightarrow A \rightarrow Outcome$ is a **path**.
- Back up to A, then back up to B. Only path remaining is C, then $Outcome$. $Treatment \leftarrow B \rightarrow C \rightarrow Outcome$ is a **path**.
- Back up to C, nowhere to go, back up to B, nowhere else to go, back up to $Treatment$, nowhere else to go. We've exhausted all the possibilities.

The full list of paths is:

- $Treatment \rightarrow Outcome$
- $Treatment \rightarrow A \rightarrow Outcome$
- $Treatment \rightarrow A \leftarrow B \rightarrow C \rightarrow Outcome$
- $Treatment \leftarrow B \rightarrow A \rightarrow Outcome$
- $Treatment \leftarrow B \rightarrow C \rightarrow Outcome$

> Homework suggestion: develop all possible conditioning strategies to identify the effect of Treatment on Outcome.