Chapter 4: Directed Acyclic Graphs

Frontier Topics in Empirical Economics

Lecture notes based on slides by Zibin Huang
College of Business, Shanghai University of Finance and Economics

October 2025

Contents

1 Introduction

Causal inference is the central topic of applied economics. Throughout this book, we have relied almost exclusively on the *potential outcome* (PO) framework, proposed by Rubin (1974) and developed extensively by Imbens and Rubin (2015), and Angrist and Pischke (2009). This framework is sometimes called the "Rubin Causal Model," and it has become the dominant language of causal inference in economics.

But is this the only statistical framework for dealing with causal questions? Of course not. *Graphical models*, and in particular *Directed Acyclic Graphs* (DAGs), constitute another important and influential approach to causality. This framework is most closely associated with the work of Judea Pearl, an Israeli-American computer scientist and philosopher who was awarded the 2011 Turing Award "for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning." Pearl's approach has deep roots in computer science and artificial intelligence, and it provides a strikingly different, yet complementary, perspective on how to think about cause and effect.

In this chapter, we introduce the DAG framework and its key concepts: graphs, Bayesian networks, causal edges, d-separation, the *do*-operator, and the backdoor and frontdoor adjustment formulas. We then turn to a critical comparison of DAGs and potential outcomes, drawing heavily on Imbens (2020), who provides a thoughtful and balanced assessment of where each framework excels and where it falls short. The goal is not to declare a winner, but to understand what DAGs can offer to applied economists, and what limitations remain.

Before diving in, it is worth noting the key references for this chapter:

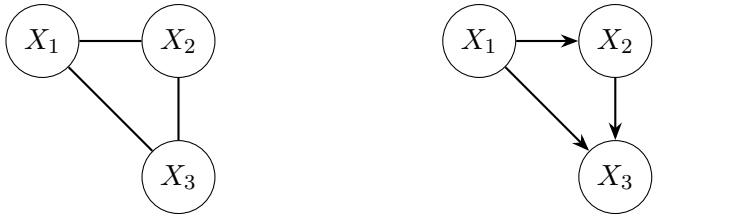
- Pearl (2009), *Causality*, Cambridge University Press—the definitive technical treatment.
- Pearl and Mackenzie (2018), *The Book of Why*, Allen Lane—a highly accessible introduction for a general audience.
- Neal (2020), *Introduction to Causal Inference*—an excellent online course and textbook (freely available at `https://www.bradyneal.com/causal-inference-course`).
- Imbens (2020)—a critical and constructive assessment from an economist's perspective.

2 The DAG Approach: Graphs

We begin with the mathematical object at the heart of this framework: the *graph*.

Definition 2.1 (Graph)**.** A **graph** is a collection of **nodes** (vertices) and **edges** that connect them. Two nodes are called **adjacent** if they are connected by an edge.

There are two basic types of graphs. In an *undirected graph*, edges have no direction. They simply indicate that two nodes are connected. In a *directed graph*, each edge has a direction: it goes out of one node (the **parent**) and into another (the **child**).



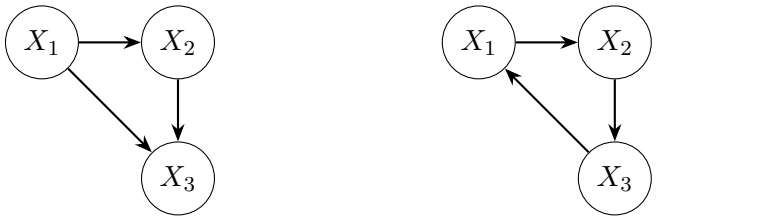
(a) Undirected Graph

(b) Directed Graph

Figure 1: Undirected and directed graphs.

A **path** is any sequence of adjacent nodes, regardless of the direction of the edges. A **directed path** is a path where all edges point in the same direction, from the first node toward the last. If there is a directed path from X to Y, we say X is an **ancestor** of Y, and Y is a **descendant** of X.

Definition 2.2 (Directed Acyclic Graph)**.** A **directed acyclic graph (DAG)** is a directed graph that contains no cycles. That is, there is no directed path that starts and ends at the same node.



(a) No Cycle (DAG)

(b) Cycle (Not a DAG)

Figure 2: A directed acyclic graph and a directed graph with a cycle.

The "acyclic" restriction is crucial. It rules out feedback loops: if X causes Y and Y causes X, there is a cycle, and the graph is not a DAG. As we will discuss later, this is one of the key limitations of the DAG framework for economics—many economic models involve equilibrium and simultaneity, which are inherently cyclical.

3 Bayesian Networks

How do we connect graphs to statistics? The first step is to connect graphs to probability distributions. This is the idea behind *Bayesian networks*.

3.1 From Joint Distributions to Factorization

Recall that any joint distribution can be decomposed using the chain rule of probability:

$$P(x_1, x_2, \ldots, x_n) = P(x_1) \prod_{i \neq 1} P(x_i \mid x_{i-1}, \ldots, x_1). \tag{1}$$

For example, with three variables:

$$P(x_1, x_2, x_3) = P(x_1) \cdot P(x_2 \mid x_1) \cdot P(x_3 \mid x_2, x_1).$$

This decomposition is always valid. It follows from the definition of conditional probability and involves no assumptions. But it is also unwieldy: every variable depends on

all preceding variables.

The key insight of Bayesian networks is that we can *simplify* this factorization if we are willing to make assumptions about the dependence structure. For instance, if $X_1 \perp X_3 \mid X_2$, then $P(x_3 \mid x_2, x_1) = P(x_3 \mid x_2)$, and the factorization simplifies to:

$$P(x_1, x_2, x_3) = P(x_1) \cdot P(x_2 \mid x_1) \cdot P(x_3 \mid x_2).$$

By making more independence assumptions, we can simplify further. The idea is to use a graph to *represent* the assumed dependence structure, creating a one-to-one mapping between the graph G and the probabilistic relations in P.

3.2 The Minimality Assumption and Bayesian Network Factorization

Minimality Assumption

1. **Local Markov Assumption:** Given its parents in the DAG, a node X is independent of all its non-descendants.
2. **Minimal Independence:** Adjacent nodes in the DAG are dependent.

The Local Markov Assumption says that the dependence structure is "local": once you know a node's parents, you do not need to know anything else about its non-descendants.

Bayesian Network Factorization

Given a probability distribution P and a DAG G satisfying the Minimality Assumption, P factorizes according to G as:

$$P(x_1, x_2, \ldots, x_n) = P(x_1) \prod_i P(x_i \mid \text{pa}_i) \tag{2}$$

where pa_i denotes the set of parents of node i in the graph G.

The Bayesian Network Factorization tells us something powerful: if the probability distribution P has the statistical structure shown in graph G, then each variable x_i depends *only* on its parents pa_i in the graph. We say that "G represents P," or equivalently that "G and P are compatible," or that "P is Markov relative to G."

3.3 A Worked Example

Suppose we have four variables x_1, x_2, x_3, x_4. The full chain-rule decomposition is:

$$P(x_1, x_2, x_3, x_4) = P(x_1) \cdot P(x_2 \mid x_1) \cdot P(x_3 \mid x_2, x_1) \cdot P(x_4 \mid x_3, x_2, x_1).$$

Now suppose our DAG is:

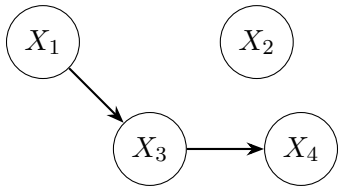


Figure 3: A DAG with four variables, where X_3 depends on X_1 and X_4 depends on X_3.

In this graph, X_3 has one parent (X_1), X_4 has one parent (X_3), and X_1, X_2 have no parents. By the Bayesian Network Factorization, we can simplify the full decomposition as:

$$P(x_1, x_2, x_3, x_4) = P(x_1) \cdot P(x_2) \cdot P(x_3 \mid x_1) \cdot P(x_4 \mid x_3). \tag{3}$$

The edges in the graph encode exactly which statistical dependencies exist.

4 Basics of Causal Graphs

4.1 Causal Graphs

Up until now, we have discussed only *statistical* dependencies. The arrows in the graphs above encode conditional independence relations, but they do not necessarily carry a causal interpretation. To move from statistical associations to causal claims, we need one more assumption.

Causal Edges Assumption

In a directed graph, every parent is a **direct cause** of all its children.

By adding this assumption, the arrows in the DAG acquire a causal meaning. Directed paths correspond to *causation*: if there is a directed path from T to Y, then T (possibly indirectly) causes Y. The DAG now represents not just statistical dependencies, but

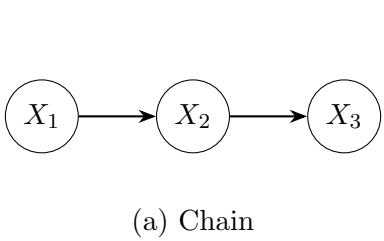
a full *causal model*.

A more mathematically rigorous formulation of this idea uses the framework of *structural equation models* (SEMs), where each variable is written as a function of its parents and an independent error term.

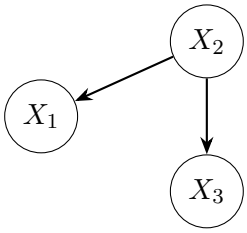
4.2 Graphical Building Blocks

There are three fundamental building blocks in any DAG. Understanding these three structures is the key to understanding how association and causation flow through a causal graph.

4.2.1 Chains, Forks, and Colliders



(a) Chain



(b) Fork

Figure 4: Chain and fork structures.

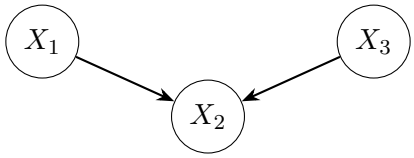


Figure 5: Collider (immorality) structure.

Two key principles govern the flow of information through these structures:

- **Association is symmetric:** X_1 and X_3 are associated in both chains and forks (but *not* in colliders).
- **Causation is asymmetric:** in the chain, X_1 causes X_3 (through X_2), but X_3 does not cause X_1.

4.2.2 Conditioning Blocks Chains and Forks

In chains and forks, conditioning on the middle variable X_2 *blocks* the flow of association between X_1 and X_3.

Proposition 4.1. *In the chain* $X_1 \to X_2 \to X_3$*:*

$$P(x_1, x_3 \mid x_2) = P(x_1 \mid x_2) \cdot P(x_3 \mid x_2).$$

That is, $X_1 \perp X_3 \mid X_2$*.*

Proof. By Bayesian factorization and the Local Markov Assumption:

$$P(x_1, x_2, x_3) = P(x_1) \cdot P(x_2 \mid x_1) \cdot P(x_3 \mid x_2).$$

Therefore:

$$P(x_1, x_3 \mid x_2) = \frac{P(x_1, x_2, x_3)}{P(x_2)} = \frac{P(x_1) \cdot P(x_2 \mid x_1)}{P(x_2)} \cdot P(x_3 \mid x_2) = P(x_1 \mid x_2) \cdot P(x_3 \mid x_2).$$

□

The same result holds for forks: if X_2 is a common cause of X_1 and X_3, then conditioning on X_2 renders X_1 and X_3 independent. Intuitively, the "spurious" association between X_1 and X_3 is entirely due to their shared cause X_2; once we know X_2, there is no residual association.

4.2.3 Conditioning Opens Colliders

Colliders behave in exactly the opposite way. In the collider $X_1 \to X_2 \leftarrow X_3$, the variables X_1 and X_3 are *marginally independent*:

$$P(x_1, x_3) = \int_{x_2} P(x_1) \cdot P(x_3) \cdot P(x_2 \mid x_1, x_3) = P(x_1) \cdot P(x_3) \int_{x_2} P(x_2 \mid x_1, x_3) = P(x_1) \cdot P(x_3). \tag{4}$$

But *conditioning on* the collider X_2 *creates* a dependence between X_1 and X_3:

$$P(x_1, x_3 \mid x_2) \neq P(x_1 \mid x_2) \cdot P(x_3 \mid x_2) \quad \text{in general.}$$

Collider Bias: The Bad Control Problem

Consider a simple example. Let X_1 be "good-looking," X_3 be "kindness," and X_2 be "marriage availability." Suppose being good-looking and being kind both independently make someone more likely to be married ($X_1 \to X_2 \leftarrow X_3$).

Among unmarried people (conditioning on $X_2 = 0$), you observe a *negative* correlation between attractiveness and kindness: "Why are all the good-looking ones so unkind?" This is pure selection bias. It arises because girls who are both good-looking and kind already get married. You are conditioning on a collider, or a post-determined outcome (the marriage status). There is no causal relationship between attractiveness and kindness.

This is precisely the **bad control problem** in econometrics: controlling for a variable that is itself affected by the treatment (or by the treatment's causes) can introduce bias rather than remove it.

Remark 4.2. The homework at the end of this chapter asks you to prove formally that conditioning on the collider X_2 creates a dependence between X_1 and X_3.

4.3 Blocked Paths and d-Separation

The three building blocks (chains, forks, and colliders) give us the vocabulary to describe when and how association flows through a DAG. The general concept is *d-separation*.

Definition 4.3 (Blocked Path)**.** A path between X and Y is **blocked** by a conditioning set Z if **either** of the following is true:

1. Along the path, there is a chain ($\to W \to$) or a fork ($\leftarrow W \to$) where $W \in Z$.
2. There is a collider W on the path such that neither W nor any descendant of W is in Z.

The first condition says: conditioning on a non-collider blocks the path. The second condition says: *not* conditioning on a collider (or its descendants) also blocks the path. Association flows along unblocked paths and does *not* flow along blocked paths.

Definition 4.4 (d-Separation)**.** Two sets of nodes X and Y are **d-separated** by a set of nodes Z if *all* paths between any node in X and any node in Y are blocked by Z.

d-Separation is the graphical criterion for conditional independence. When X and Y

are d-separated by Z, all association flows between X and Y are blocked by Z.

d-Separation and Statistical Independence

Theorem (Pearl, 2009, Theorem 1.2.4–1.2.5); (Neal, 2020, Theorem 3.1).
If X and Y are d-separated in a DAG G conditional on Z, then X and Y are independent conditioned on Z in every distribution compatible with G:

$$X \perp_G Y \mid Z \;\Rightarrow\; X \perp_P Y \mid Z, \quad \forall\, P \text{ compatible with } G.$$

Conversely, if X and Y are independent conditional on Z in all P compatible with G, then X and Y are d-separated in G conditional on Z:

$$(\forall\, P \text{ compatible with } G,\; X \perp_P Y \mid Z) \;\Rightarrow\; X \perp_G Y \mid Z.$$

This theorem is a *bridge* between graphs and probability. It tells us exactly how to translate statistical independence statements into graphical statements, and vice versa.

The notion of d-separation leads directly to the question of causal identification. Recall:

- **Associations** flow along all unblocked paths.
- **Causation** flows along *directed* unblocked paths.

To isolate the causal effect of T on Y, we need to ensure that there is no *non-causal* association between T and Y. Graphically, this means: in the augmented graph where we remove all outgoing edges from T, the variables T and Y should be d-separated. In other words, all non-causal ("backdoor") paths must be blocked.

4.4 The *do*-Operator

The DAG framework formalizes causal interventions using the *do*-operator, introduced by Pearl (2009). We define $\text{do}(T = t)$ as an intervention that forces the entire population to receive treatment level t.

In terms of potential outcomes, the *do*-operator corresponds to:

$$P(y \mid \text{do}(t)) = P(Y = y \mid \text{do}(T = t)) = P(Y(t) = y), \tag{5}$$

where $Y(t)$ is the potential outcome under treatment t, exactly as defined in Chapter 1.

The notation $P(y \mid \text{do}(t))$ is *not* the same as the conditional probability $P(y \mid t)$. The latter conditions on *observing* $T = t$, which may reflect both causation and selection. The former represents what would happen if we *intervened* to set $T = t$, breaking all arrows pointing into T.

Identification via the do-Operator

A causal quantity Q is **identifiable** if we can reduce an expression involving $\text{do}(\cdot)$ to one that involves only standard (do-free) probabilities of observed variables. This is precisely analogous to the PO framework, where identification means reducing expressions involving potential outcomes $Y(t)$ to expressions involving only observed data.

5 Causal Identification Method in Causal Graphs

5.1 Backdoor Adjustment

One of the most important identification results in the DAG framework is the *backdoor adjustment*. Non-directed unblocked paths from T to Y, paths that go "through the back" of T, are called **backdoor paths**. These paths create non-causal associations between treatment and outcome.

Definition 5.1 (Backdoor Criterion)**.** A set of variables W satisfies the **backdoor criterion** relative to (T, Y) if:

1. W blocks all backdoor paths from T to Y; and
2. W does not contain any descendant of T.

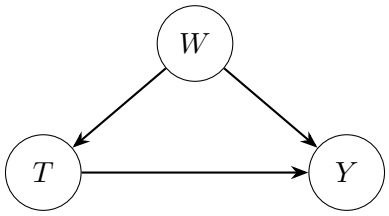


Figure 6: A backdoor path $T \leftarrow W \rightarrow Y$ through the confounder W.

Backdoor Adjustment Theorem

If W satisfies the backdoor criterion relative to (T, Y), then:

$$P(y \mid \text{do}(t)) = \int_w P(y \mid t, w)\, P(w). \tag{6}$$

This should look familiar. The variable set W is exactly what we usually call "control variables" in applied economics. The backdoor criterion is the graphical analogue of the "selection on observables" (or "unconfoundedness") assumption from the potential outcome framework (Chapter 1, Section 5). The backdoor adjustment formula is just the familiar formula for the average treatment effect under unconfoundedness:

$$\text{ATE} = E[Y(1)] - E[Y(0)] = \int \{E[Y \mid T = 1, W = w] - E[Y \mid T = 0, W = w]\}\, dF_W(w).$$

5.2 Frontdoor Adjustment

One of the most interesting identification strategies unique to the DAG framework is the *frontdoor adjustment*. This idea has no direct analogue in the standard PO toolkit, and it illustrates the creative power of graphical reasoning.

5.2.1 Setup

Consider the following DAG, where W is an *unobserved* confounder:

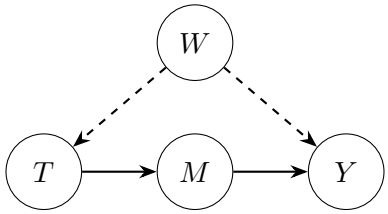


Figure 7: Frontdoor structure: T affects Y only through the mediator M, with W an unobserved confounder (dashed).

In this graph, T affects Y *only* through the mediator M. The confounder W creates a backdoor path $T \leftarrow W \rightarrow Y$, so we cannot simply regress Y on T. But we can identify the causal effect in three steps:

1. **Identify the effect of T on M:** There is no backdoor path from T to M (the

only path is the direct arrow $T \to M$), so $P(m \mid \text{do}(t)) = P(m \mid t)$.

2. **Identify the effect of M on Y:** There is a backdoor path from M to Y through W, but it can be blocked by conditioning on T. So $P(y \mid \text{do}(m)) = \sum_{t'} P(y \mid m, t')P(t')$.

3. **Combine:** Chain the two results together.

Definition 5.2 (Frontdoor Criterion)**.** A set of variables M satisfies the **frontdoor criterion** relative to (T, Y) if:

1. M completely mediates the causal effect of T on Y (i.e., all directed paths from T to Y go through M).

2. There is no unblocked backdoor path from T to M.

3. All backdoor paths from M to Y are blocked by T.

Frontdoor Adjustment Theorem

If T, M, Y satisfy the frontdoor criterion, then:

$$P(y \mid \text{do}(t)) = \sum_{m} P(m \mid t) \sum_{t'} P(y \mid m, t')\, P(t'). \tag{7}$$

This is a remarkable result: we can identify the causal effect of T on Y *even though there is an unobserved confounder*, as long as we observe a complete mediator M.

5.2.2 Limitations

The frontdoor adjustment is intellectually beautiful, but its practical applicability is limited. The requirement that M *completely* mediates the effect of T on Y is very strong. In most economic settings:

- What if T affects Y through other channels besides M?
- What if some unobserved U affects both M and Y?
- What if the confounder W also affects M directly?

Any of these violations breaks the frontdoor criterion. Finding a genuine complete mediator in real economic applications is extremely difficult. How to apply this method to economics remains an open question.

5.3 Non-parametric Identification: *do*-Calculus

The backdoor and frontdoor criteria are *sufficient* conditions for causal identification, but they are not *necessary*. That is, there may be causal effects that are identifiable even though neither criterion is satisfied. Can we find a *complete* set of rules that determines whether any causal effect is identifiable in any DAG?

The answer is yes: the *do*-calculus, developed by Pearl (2009).

5.3.1 Notation

We need some notation for manipulated graphs:

- $G_{\overline{X}}$: take graph G and *remove all incoming edges* to X.
- $G_{\underline{X}}$: take graph G and *remove all outgoing edges* from X.
- $Z(W)$: the set of nodes of Z that are not ancestors of any node of W in $G_{\overline{T}}$.

5.3.2 The Three Rules

Rules of do-Calculus

1. **Rule 1 (Deletion of variable):**

$$P(y \mid \text{do}(t), z, w) = P(y \mid \text{do}(t), w) \quad \text{if } Y \perp_{G_{\overline{T}}} Z \mid T, W.$$

2. **Rule 2 (*do*-variable exchange):**

$$P(y \mid \text{do}(t), \text{do}(z), w) = P(y \mid \text{do}(t), z, w) \quad \text{if } Y \perp_{G_{\overline{T}\underline{Z}}} Z \mid T, W.$$

3. **Rule 3 (Deletion of action):**

$$P(y \mid \text{do}(t), \text{do}(z), w) = P(y \mid \text{do}(t), w) \quad \text{if } Y \perp_{G_{\overline{T}\overline{Z(W)}}} Z \mid T, W.$$

5.3.3 Intuition Behind the Rules

What do these three rules mean intuitively?

Rule 1 is an extension of d-separation. If we erase the do-operator from $\text{do}(t)$, the condition becomes $Y \perp_G Z \mid W$—just standard conditional independence under the

Markov assumption. In words: if Z provides no additional information about Y once we know T and W (in the manipulated graph), then we can drop Z from the expression.

Rule 2 is an extension of the backdoor adjustment. If we erase $\text{do}(t)$, the condition becomes: W blocks all non-causal paths between Z and Y in the graph where outgoing edges from Z are removed. In other words, we can replace $\text{do}(z)$ with simply observing z, because W already controls for all confounding between Z and Y.

Rule 3 says that if, in the appropriate manipulated graph, Z and Y are d-separated, then the intervention $\text{do}(z)$ has no effect on Y and can be dropped entirely.

Completeness of do-Calculus

Theorem (Pearl, 2009). A causal effect Q is identifiable in a model characterized by a graph G if and only if there exists a finite sequence of transformations, each conforming to one of Rules 1, 2, or 3, that reduces Q into a standard (do-free) probability expression involving observed quantities.

The do-calculus is *complete*: these three rules are sufficient to identify *every* identifiable causal effect. If a causal effect cannot be identified using these rules, then it is genuinely not identifiable from observational data given the assumed graph.

Non-parametric Only

An important caveat: the do-calculus concerns *non-parametric* identification only. It asks whether the causal effect can be expressed in terms of the observed joint distribution without any functional form assumptions. In many economic applications, parametric or semi-parametric assumptions (e.g., linearity, monotonicity) provide additional identifying power that the do-calculus does not capture.

5.4 An Example: College Returns on Wages

Let us consider a concrete example from the slides. Suppose we are interested in the causal effect of college education (D) on wages (Y). A plausible DAG might include:

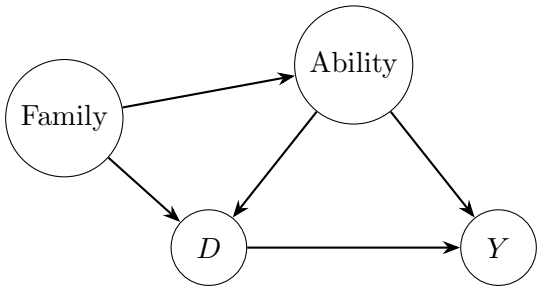


Figure 8: A causal DAG for college education and wages, with Ability and Family as background factors.

The question is: which variables do we need to control for? The DAG framework gives us a systematic answer through the backdoor criterion. The backdoor paths from D to Y are:

- $D \leftarrow \text{Ability} \rightarrow Y$
- $D \leftarrow \text{Family} \rightarrow \text{Ability} \rightarrow Y$

Controlling for Ability blocks both paths. Controlling for Family alone blocks only the second path but not the first (since Ability remains an unblocked non-collider). The DAG tells us precisely what we need to control for, and what we should *not* control for.

6 DAGs in Economics: Pros and Cons

Having introduced the DAG framework, we now turn to the central question of this chapter: what can DAGs offer to applied economists, and where do they fall short? We follow the structure of Imbens (2020), who provides a balanced and thoughtful assessment.

6.1 Advantages of DAGs

Pro 1: Clarity of Illustration. DAGs make causal assumptions *explicit and visual*. Consider two of the most common identification strategies in economics:

Unconfoundedness. The assumption that treatment is independent of potential outcomes conditional on observables $(Y(0), Y(1) \perp T \mid W)$ can be illustrated by a simple DAG: $W \rightarrow T$, $W \rightarrow Y$, $T \rightarrow Y$. The backdoor path $T \leftarrow W \rightarrow Y$ is blocked

by conditioning on W. This is immediately transparent in the graph.

Instrumental variables. The IV assumptions—relevance (Z affects T), exclusion (Z does not directly affect Y), and independence (Z is as good as randomly assigned)—correspond to a DAG: $Z \to T \to Y$, with no direct edge $Z \to Y$ and no common cause of Z and Y. The exclusion restriction, which can be hard to state precisely in words, becomes visually obvious: there is simply no arrow from Z to Y.

Pro 2: Tool for Analyzing Complicated Causal Models. When the causal model involves many variables with complex interdependencies, DAGs provide systematic machinery (d-separation, do-calculus) for determining which causal effects are identifiable and which control sets are valid. This is the setting where DAGs truly shine: with a large number of variables, it becomes very difficult to track all the possible confounding paths by hand, and the graphical approach provides an algorithmic solution.

Pro 3: Frontdoor Criterion. As discussed earlier, the frontdoor adjustment provides an identification strategy that has no direct analogue in the PO framework. While its practical applicability remains limited (complete mediators are rare in economics), it demonstrates the creative power of graphical reasoning and remains an active area of research.

Pro 4: Systematic Analysis of Mediation Effects. DAGs can shed light on the identification of mediation (or mechanism) effects in a rigorous way. The standard "mediation effect test" (the Baron-Kenny approach) implicitly assumes a very simple causal structure, one that rules out any unobserved confounders between the treatment and the mediator, between the mediator and the outcome, or between the treatment and the outcome.

The Mediation Test Is Often a Regression Monkey Exercise

The standard mediation test requires you to accept a very simple causal structure just to implement an off-the-shelf test. But think about what this structure implies:

- What if there is an unobserved U affecting both X and Y?
- What if there is an unobserved U affecting both M and X?
- What if there is an unobserved U affecting both M and Y?

If you truly believe this simple structure (no confounders anywhere, full mediator), then you can directly identify the effect of X on Y using OLS or frontdoor method!

It means that when you use research design (DID, IV etc.) to rule out endogeneity issue but in the same time want to implement mediation test, you are doing research in an **inconsistent** way. If you think there are confounders, then you cannot run simple mediation test.
For most of the economists, the standard mediation test is, "a typical behavior of regression monkey." It is a super-simplified reduced DAG with very strong assumptions. Meanwhile, it leads to inconsistency when the authors use research designs to rule out confounders. We should think about mediation in a more general DAG framework.
Standard mediation test originates from psychology, where RCT could provide a clean environment without confounder. However, this is not the case in economics where we always struggle with observational data in real life. Therefore, in economics, standard mediation test can only give us some suggestive evidence at best. DO NOT put too much importance on the results of these tests.
Generally speaking, do not use off-the-shelf mediation test without thinking by yourself!

Then how *should* we investigate mechanisms in empirical research? I suggest several approaches:

1. Use **heterogeneity analysis** to provide indirect evidence of mechanisms.
2. Think from a **general DAG framework**: construct a causal graph based on your knowledge of the economics and test the implied relationships.
3. Construct a **theoretical model**, derive comparative statics, transform them into testable regressions.
4. Construct a **structural model**, illustrate the mechanism through decomposition and counterfactual analysis.

6.2 Disadvantages of DAGs

Con 1: DAGs Require Ex Ante Causal Structure. The DAG approach develops powerful machinery for identification, but it requires the whole structure of the causal model. Little is said about *why* we should believe a particular structure (the "before model" question) or about estimation and inference (the "after model" question). As

Imbens (2020) notes, the DAG literature sometimes appears to offer "a set of solutions in search of problems."

For economists, both the *before* and *after* are crucial. We want to know where the model comes from (economic theory? institutional knowledge?) and how to estimate the causal effect with valid inference in finite samples.

Con 2: DAGs Do Not Fit IV Very Well. The instrumental variables setting reveals several limitations of the DAG approach:

- **Shape restrictions:** The monotonicity assumption (central to the LATE theorem of Imbens and Angrist, 1994) cannot be naturally expressed in a DAG. DAGs encode conditional independence relations, not shape restrictions on functional forms.
- **Heterogeneity:** The LATE theorem is not easily derived in a DAG framework. The PO framework, with its explicit notation for heterogeneous potential outcomes $Y_i(0), Y_i(1)$, handles treatment effect heterogeneity much more naturally.
- **RDD:** Similarly, the regression discontinuity design does not gain much from the DAG approach.

More generally, the inability to fit IV well reveals two deeper weaknesses: DAGs are not convenient for expressing economic-related structural assumptions, and the PO framework can deal with heterogeneity issues better.

Con 3: DAGs Cannot Capture Simultaneity and Equilibrium. By definition, a DAG is *acyclic*, so it cannot contain feedback loops. But many economic models are built on equilibrium concepts that are inherently cyclical: supply determines price, and price determines supply. DAGs cannot naturally express simultaneity.

Imbens (2020) attempts to represent a supply-and-demand equilibrium in a DAG, but the result is "not so successful"—the time-indexed expansion required to avoid cycles is awkward and loses the economic content of the simultaneity.

6.3 Why Is PO Still More Popular in Economics?

Summarizing Imbens (2020), there are five features of the PO framework that explain its dominance in applied economics:

1. Some common economic assumptions (monotonicity, convexity) are easily captured in PO but not in DAGs.
2. Potential outcomes connect naturally to traditional economic models (supply and demand, utility maximization).
3. Applied economics typically involves models with relatively few variables, where the systematic machinery of d-separation adds less value.
4. PO handles treatment effect heterogeneity (ATE, ATT, LATE) more naturally.
5. PO connects closely to the implementation of the method and its inference—estimation, standard errors, confidence intervals are integral parts of the framework.

A final, pragmatic point: the DAG literature has not provided substantive empirical examples demonstrating clear advantages over the PO approach. Most examples from Pearl (2009) are "toy models." Until concrete applied examples emerge, adoption will remain slow.

7 Consider Bad Control Issue Using DAG

Despite these limitations, there are areas where DAGs offer genuine insight that the PO framework does not provide as easily.

7.1 Clarity in the Bad Control Problem

The *bad control problem* is illustrated with remarkable clarity in the DAG framework. Simpson's Paradox, which we discussed in Chapter 1, is immediately transparent when drawn as a graph: the "paradox" disappears once you see that conditioning on a collider opens a spurious path.

In the PO framework, the standard advice for selecting control variables is Angrist's rule of thumb: "only control for variables determined *before* the treatment." This is useful, but it is actually *wrong* in general—as the concept of M-bias shows.

7.2 M-bias

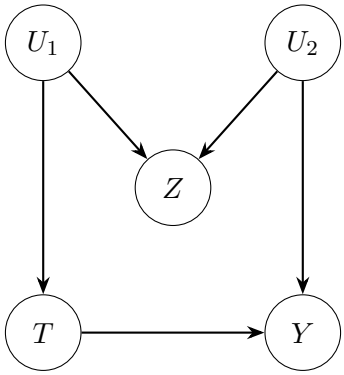


Figure 9: M-bias: Z is a pre-treatment variable, but it is a collider on the path $T \leftarrow U_1 \rightarrow Z \leftarrow U_2 \rightarrow Y$.

In this graph, Z is a *pre-treatment* variable. Angrist's rule of thumb would say: control for Z, since it is determined before T. But Z is a *collider* on the path $T \leftarrow U_1 \rightarrow Z \leftarrow U_2 \rightarrow Y$. Without conditioning on Z, this path is blocked (because Z is a collider). Conditioning on Z *opens* this path, creating a spurious association between T and Y.

The Lesson of M-bias

We should not control for colliders, and colliders do not necessarily occur *after* treatment. M-bias is the case where the collider is a pre-determined variable. The DAG framework provides a powerful and general tool for selecting control variables, given our assumptions about the causal structure. It forces us to state our causal assumptions explicitly and transparently.

8 Control Variable Issues: A Final Conclusion

Let us conclude this chapter by reviewing and discussing the issue of control variable selection. In earlier periods of empirical economics, economists often paid limited attention to the choice of control variables. Researchers frequently selected controls pretty arbitrarily, leaving substantial room for cherry-picking. As applied econometric methods have advanced, and as economists have increasingly incorporated insights from fields such as computer science and statistics, the discipline has become much more rigorous in its approach to control variable selection. Broadly speaking, there are two perspectives on how to select control variables. These perspectives reflect two different

scientific philosophies and two broad approaches to empirical economic research: theory or economic knowledge driven approaches versus data driven approaches.

First, control variables can be selected based on the underlying causal structure. The key idea is to propose a causal graph that represents the data generating process we believe to be true, and then determine which variables are confounders and which are colliders. To identify the treatment effect of interest, we should control for confounders while avoiding conditioning on colliders. The procedure is inherently more subjective and relies heavily on economic theory and institutional knowledge. We include certain control variables because we believe they play an important role within the causal framework we have specified.

Second, control variables can be selected based on model fit. This relates closely to the model selection problem discussed in Chapter 3. Once the basic identification assumptions implied by the causal structure are satisfied, researchers may seek to improve the goodness of fit of the model, especially its out of sample predictive performance. This is where machine learning methods become particularly useful. Such methods provide automated procedures for selecting control variables that optimize predictive accuracy, thereby reducing researcher discretion and limiting the scope for cherry-picking. This approach is more objective and data driven, relying less on prior human judgment and more on patterns revealed by the data.

One final point is particularly important. As economists, we have strong economic theory foundations and our primary concern is causal identification. Therefore, control variables should first be selected based on the causal structure we propose, and only afterward should considerations of model fit be incorporated. For example, one should always control for education in a wage equation, even if a LASSO procedure suggests otherwise.

9 Conclusion

The DAG approach to causality fully deserves the attention of all economists. It offers several genuine advantages:

- **Clarity:** DAGs force researchers to state their causal assumptions explicitly, making them transparent and debatable.
- **Control selection:** The backdoor criterion provides a principled, systematic

method for choosing control variables—one that avoids pitfalls like M-bias that simpler rules of thumb miss.

- **Complex models:** For models with many variables, the do-calculus provides algorithmic tools for determining identifiability.
- **Novel strategies:** The frontdoor criterion demonstrates that identification strategies exist beyond what the PO framework naturally suggests.

However, the DAG approach also has significant limitations for economics:

- It requires an ex ante causal structure, with little guidance on where this structure comes from.
- It cannot easily express shape restrictions (monotonicity, convexity) or handle treatment effect heterogeneity.
- It cannot represent simultaneity and equilibrium, which are central to many economic models.
- It lacks concrete empirical examples demonstrating its value over existing methods.

As Imbens (2020) concludes, the PO and DAG frameworks are *complementary*, with different strengths for different questions. The honest assessment is that, for the kinds of questions applied economists typically ask—with relatively few variables, strong institutional knowledge, and a focus on heterogeneity and inference—the PO framework remains the more natural choice. But understanding DAGs makes you a better applied researcher: it sharpens your thinking about what you are controlling for and why.

How to bring the insights of DAGs more fully into applied economics remains a very open question. There are chances here for the next generation of researchers.

A Homework Problems

1. **Collider Bias Proof.**

   Consider the collider structure $X_1 \to X_2 \leftarrow X_3$.

   (a) Show that X_1 and X_3 are marginally independent: $P(x_1, x_3) = P(x_1) \cdot P(x_3)$.

   (b) Prove that conditioning on X_2 generally creates a dependence between X_1

and X_3. That is, show that $P(x_1, x_3 \mid x_2) \neq P(x_1 \mid x_2) \cdot P(x_3 \mid x_2)$ in general.

Hint: Replicate the proof strategy used for chains and forks in Section 5.2.

2. **d-Separation Practice.**

Consider the following DAG:

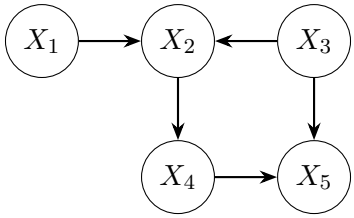


Figure 10: DAG for the d-separation practice problem.

For each of the following, determine whether the two variables are d-separated by the given conditioning set, and explain your reasoning:

(a) X_1 and X_3, conditioning on $\emptyset$.

(b) X_1 and X_3, conditioning on $\{X_2\}$.

(c) X_1 and X_5, conditioning on $\{X_3\}$.

(d) X_1 and X_5, conditioning on $\{X_2, X_3\}$.

3. **Backdoor Criterion Application.**

Consider the following DAG representing the effect of a job training program (T) on wages (Y):

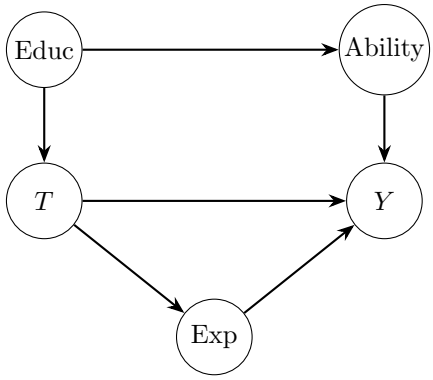


Figure 11: DAG for the backdoor criterion application problem.

Here Educ = education level, Ability = innate ability, Exp = work experience (post-training), T = training program, Y = wages.

(a) List all backdoor paths from T to Y.

(b) Does {Educ} satisfy the backdoor criterion? Why or why not?

(c) Does {Exp} satisfy the backdoor criterion? Why or why not?

(d) Find a minimal set that satisfies the backdoor criterion.

Practice Example: Backdoor Adjustment in a Simulated DAG

In this practice example, we simulate data from a known DAG and verify that the backdoor adjustment formula recovers the true causal effect, while a naive regression does not.

Data Generating Process. Consider the following DAG:

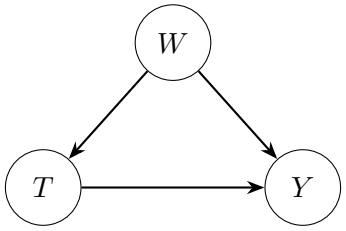


Figure 12: DAG for the practice example: W is the confounder of T and Y.

The structural equations are:

$$\begin{aligned} W_i &\sim \mathcal{N}(0,1), \\ T_i &= 0.8W_i + \varepsilon_{T,i}, \quad \varepsilon_{T,i} \sim \mathcal{N}(0,1), \\ Y_i &= \underbrace{2}_{\text{true causal effect}} \cdot T_i + 1.5W_i + \varepsilon_{Y,i}, \quad \varepsilon_{Y,i} \sim \mathcal{N}(0,1). \end{aligned}$$

The true causal effect of T on Y is $\beta = 2$. We generate $N = 2000$ observations.

Analysis. The Stata do-file (`src/chapter 4/ch4_dag.do`) performs the following steps:

1. **Simulate data** from the structural equations above.
2. **Naive regression:** Regress Y on T only (ignoring W). This produces a biased estimate because the backdoor path $T \leftarrow W \rightarrow Y$ is open.
3. **Backdoor adjustment:** Regress Y on T and W, controlling for the confounder. This blocks the backdoor path and recovers the true causal effect.
4. **Compare** the two estimates to the true value $\beta = 2$.

Expected results. The naive regression should overestimate the causal effect (because

W positively affects both T and Y, creating upward omitted variable bias). The controlled regression should produce an estimate close to 2.

Key lessons:

- The naive regression conflates the causal effect with the confounding association through W.
- The backdoor adjustment (controlling for W) isolates the causal channel $T \to Y$.
- This example illustrates the core logic of the backdoor criterion: identify all backdoor paths, find a valid conditioning set, and control for it.

The do-file is located at `src/chapter 4/ch4_dag.do` and the simulated dataset at `data/chapter 4/ch4_dag.dta`.

References

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Imbens, Guido W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, 58(4):1129–1179.

Imbens, Guido W. and Joshua D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Neal, Brady (2020). Introduction to Causal Inference. Course Lecture Notes (draft). Available at https://www.bradyneal.com/causal-inference-course.

Pearl, Judea (2009). *Causality: Models, Reasoning, and Inference*. 2nd edition. Cambridge University Press.

Pearl, Judea and Dana Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect*. Allen Lane.

Rubin, Donald B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688–701.