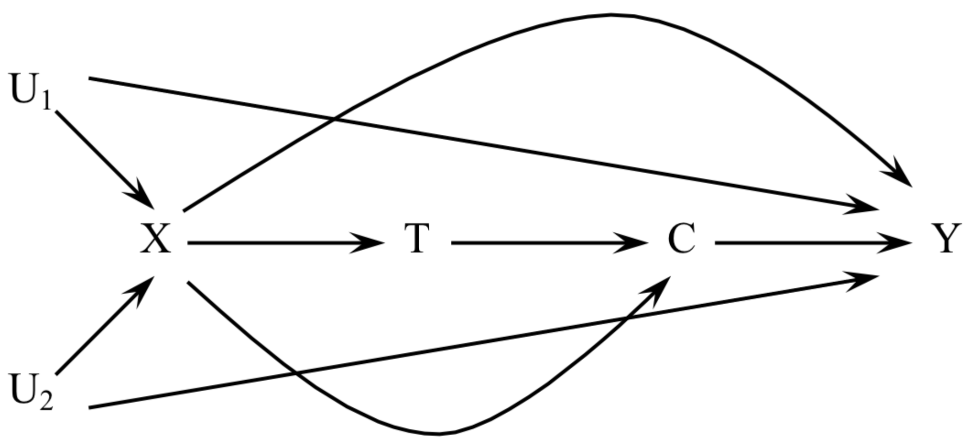
2. DAGs: Overview and Motivation

Felix Elwert, Ph.D.
elwert@wisc.edu
University of Wisconsin–Madison

Please Cite: Elwert, Felix. 2013. “Graphical Causal Models.” Pp. 245-273 in S.L. Morgan (ed.), *Handbook of Causal Analysis for Social Research*. New York: Sage Publications.

Directed Acyclic Graphs

DAGs graphically represent non-parametric structural equation models. They may look like the path models of yore, but they are far more general.



This is a DAG

Origins

DAGs have roots in

1. Structural equation models (1930s+)
2. Social science path models (1960s+)
3. Bayesian networks (1980s+).

Judea Pearl and colleagues synthesized and generalized these approaches to develop a powerful graphical syntax for causal inference.

We will follow Pearl (1995, 2009) and read DAGs as nonparametric structural equation models (NPSEM), which gives them a causal interpretation.

Compatible with the potential outcomes (Neyman-Rubin) framework.

Some Key Players

Computer Science:

Judea Pearl, Jin Tian, Thomas Verma

Philosophy:

Peter Spirtes, Clark Glymour, Richard Scheines

Biostatistics:

Jamie Robins, Sander Greenland, Tyler VanderWeele, Miguel Hernan

Why DAGs?

- Rigorous mathematical objects, support proofs
- Very general (nonparametric)
- For many purposes, DAGs are more accessible than potential outcomes notation
  - All pictures, no algebra
  - Focus attention on causal assumptions (language of applied scientists)
  - Great for deriving (nonparametric) identification results
  - Great for deriving the testable implications of a causal model
  - Intuition for understanding many problems in causal inference.
  - Particularly helpful for complex causal models
- Limitations
  - Don't display the parametric assumptions that are often necessary for estimation in practice.
  - Generality can obscure important distinctions between estimands.

Main Goals of This Course

- DAGs are useful for many topics
- This course aims to introduce you to DAGs' three main uses:
  1. Deriving testable implications of a causal model
  2. Understanding causal identification requirements
  3. Informing the use of some popular statistical techniques

Taking the Long View for Today

- We'll have to get through some terminological and conceptual setup before we can get to the meat of things.
- Therefore, I'd like to preview what you may find to be the most empowering lesson of today's sessions: using DAGs to identify causal effects

Identification

- When analyzing data, the analyst needs to ask: "Is the association I observe causal or spurious?"
- Identification: The possibility of separating causal from noncausal associations with ideal data == reducing the observed (conditional) association between treatment and outcome to its purely causal (i.e., non-spurious) component.

Today's Goal

At a basic level, identification usually means asking:

"What control variables, X, must be included —and which variables mustn't be included— in the analysis to achieve identification?"

Today's Goal

- We already know that the total causal effect of T on Y is identifiable from observable data {Y, T, X} if treatment is conditionally ignorable:

$$\{Y_1, Y_0\} \perp T \mid X$$

- But ignorability gives little guidance as to what variables should be in X beyond stating that X should make T and $\{Y_T\}$ conditionally independent.
    - Most people have poor intuition for unobservable counterfactuals, $\{Y_T\}$.
    - Indeed, the most popular substantive interpretation (that X is the assignment mechanism for T), though sufficient, is seriously incomplete.

- Instead of dealing in counterfactuals and conditional independences, it may be easier to talk about identification in terms of cause-effect statements–the common currency of scientific discourse.
- DAGs are powerful because they encode causal assumptions and permit the analyst to translate between (substantive) causal statements and (statistical) independence statements.
- <u>The promise</u>: once we lay out our causal assumptions, it's easy to derive all implied marginal and conditional independences in the system.
- This permits judgment about when ignorability is met, and what variables should be (or mustn't be) included in the analysis.

3. DAGs
Elements and Interpretation

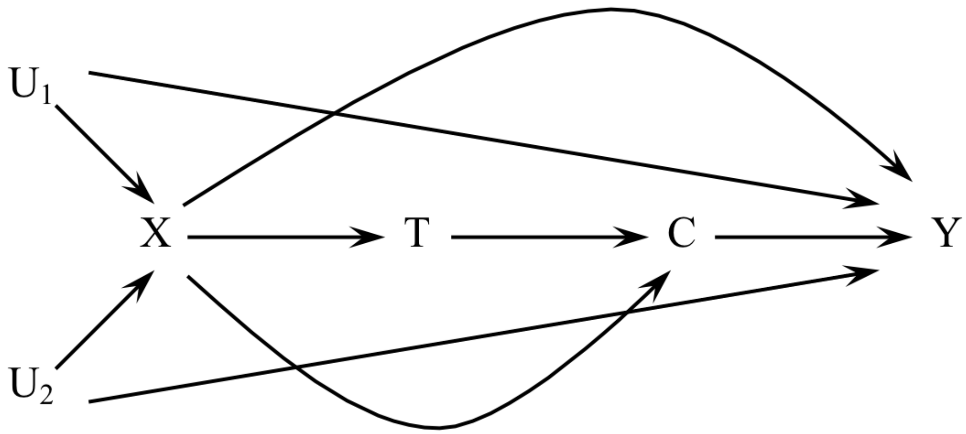
(Lots of terminology. It's worth the effort.)

Elwert@wisc.edu. Version 5/2013

Section Overview

- Basic elements
- Interpretation: encoding the qualitative causal assumptions of the data generating model
- DAGs as NPSEM

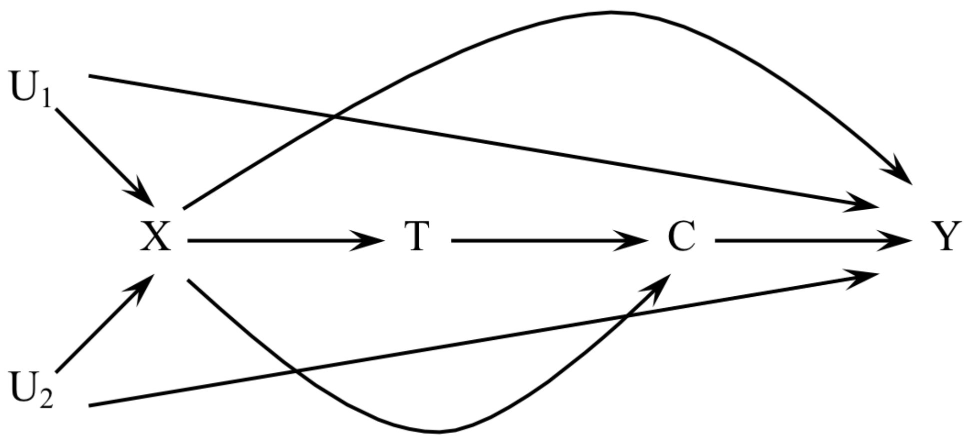
Building Blocks



DAGs consist of three elements:

1. Variables (nodes, vertices)
2. Arrows (directed edges, arcs): possible direct causal effects. The arrows order the variables in time.
3. Missing arrows; sharp assumptions about absent direct causal effects.

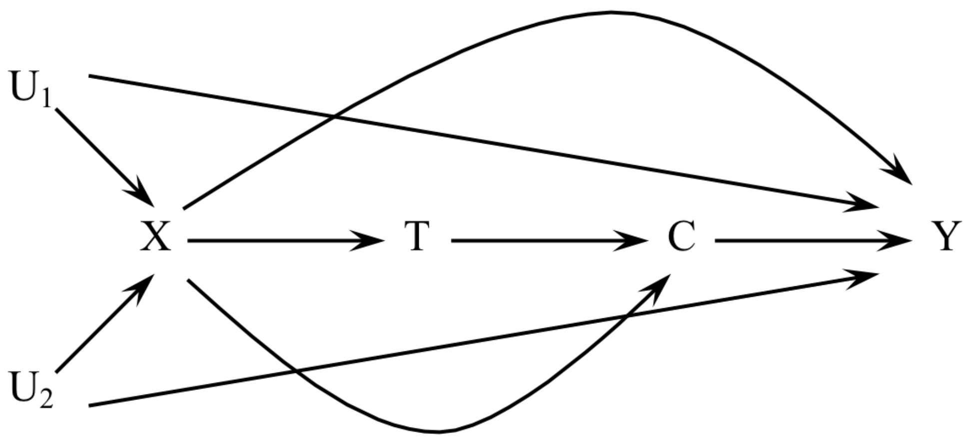
Nonparametric



DAGs are <u>non-parametric</u>, i.e. they make no assumption about

1. The distribution of the variables (nodes) in the DAG
2. The functional form of the direct causal effects (arcs)

Present and Missing Arcs



DAGs encode the analyst's qualitative causal knowledge/beliefs/ assumptions:

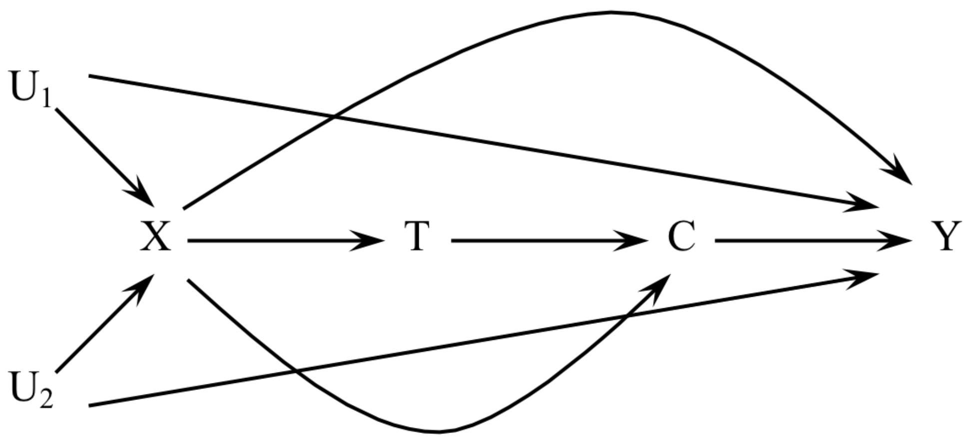
Directed arcs represent possible direct causal effects.

- E.g., C may or may not directly cause Y

Missing arcs represent sharp Nulls of no-effect.

- E.g., U_2 does not directly cause T; and T does not directly cause Y for anybody.

Missing Arcs Encode Assumptions

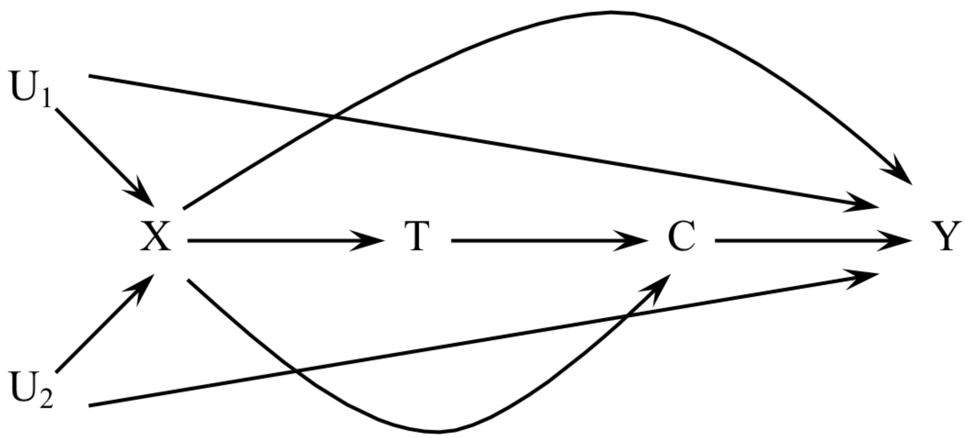


Thus, only <u>missing arcs encode causal assumptions</u>, whereas directed arcs represent ignorance!

- (As a sloppy shortcut, we often read arcs as existing causal effects)

In assessing whether a particular DAG is theoretically plausible/ valid, the debate should be about which arrows can be assumed to be <u>absent</u>.

Node Terminology



Descendants of a node: all nodes directly or indirectly caused by the node

desc(T) = {C, Y}

Children of a node: all nodes directly caused by the node

child(T) = {C}

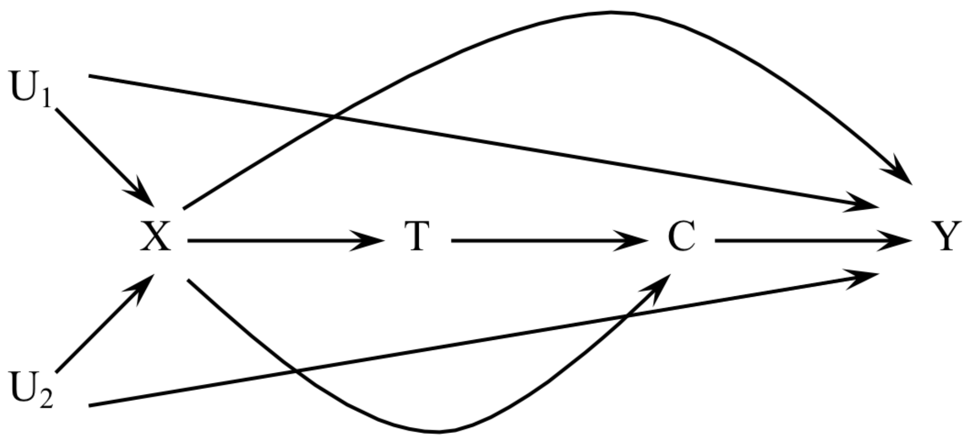
Ancestors of a node: all nodes directly or indirectly causing the node

an(T) = {X, U_1, U_2}

Parents of a node: all direct causes of the node

pa(T) = {X}

Paths



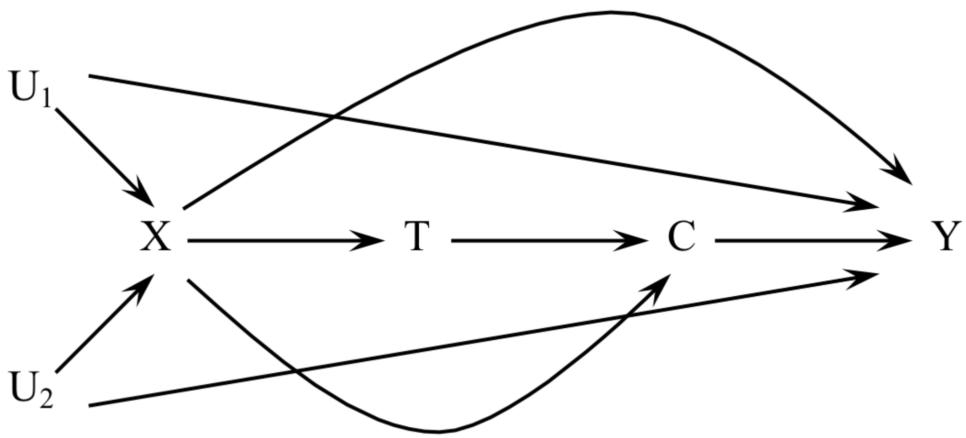
A path is a sequence of non-intersecting adjacent edges

E.g., X→T→C or U_2→Y←C←T

Note: (1) The direction of the arrows doesn't matter.

(2) Non-intersecting = a path cannot cross a node more than once

Colliders



Perhaps the most important concept!

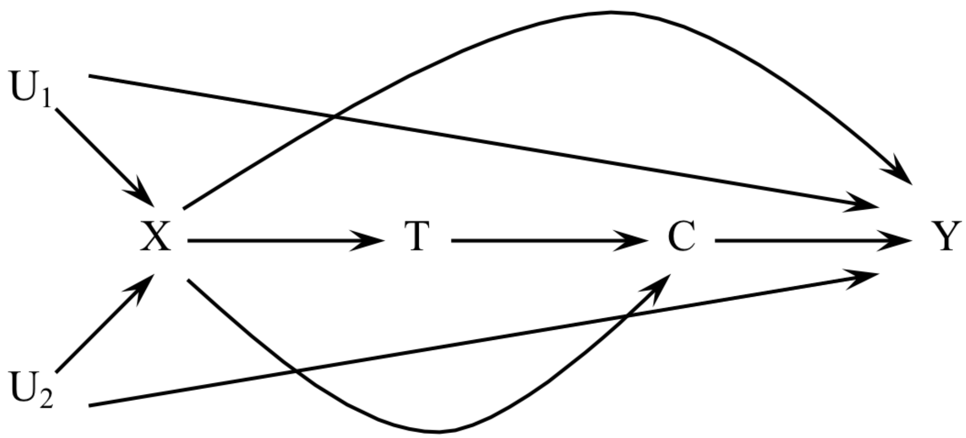
A collider variable is a variable along a path with two arrows pointing in

E.g.,X is a collider on the path $U_1 \rightarrow X \leftarrow U_2$.

But X is not a collider on the path $U_1 \rightarrow X \rightarrow T$.

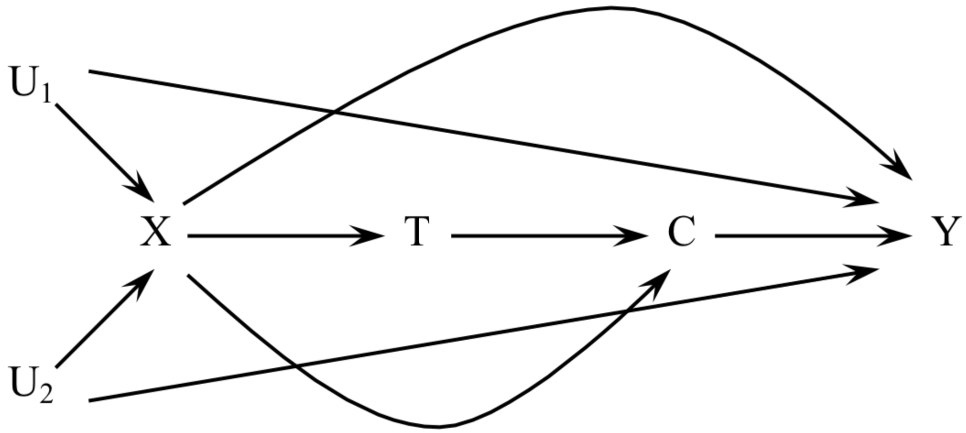
=>Colliders are specific to a path.

Directed Acyclic Graphs



DAGs are "directed" in that each arrow is single headed, expressing a single causal statement, e.g. T directly causes C. (We'll meet bi-headed arrows later.)

Directed Acyclic Graphs



DAGs are "acyclic" in that they contain no directed cycles: one cannot trace a sequence of arcs in the direction of the arrows and arrive whence one started.

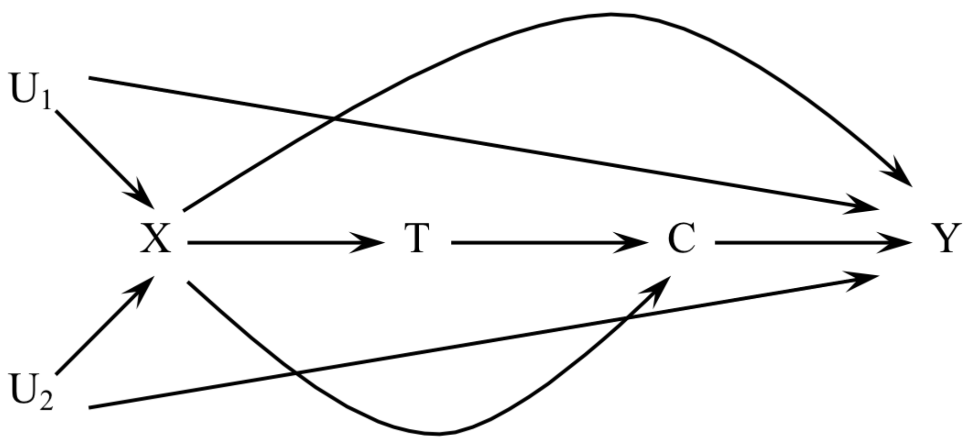
"The future cannot directly or indirectly cause the past."

Apparent counterexamples ('schooling and wages cause each other') are usually resolved by redrawing the DAG with a finer temporal articulation.

No "simultaneity."

(There's theory for cyclical graphs, too, see Pearl 2009.)

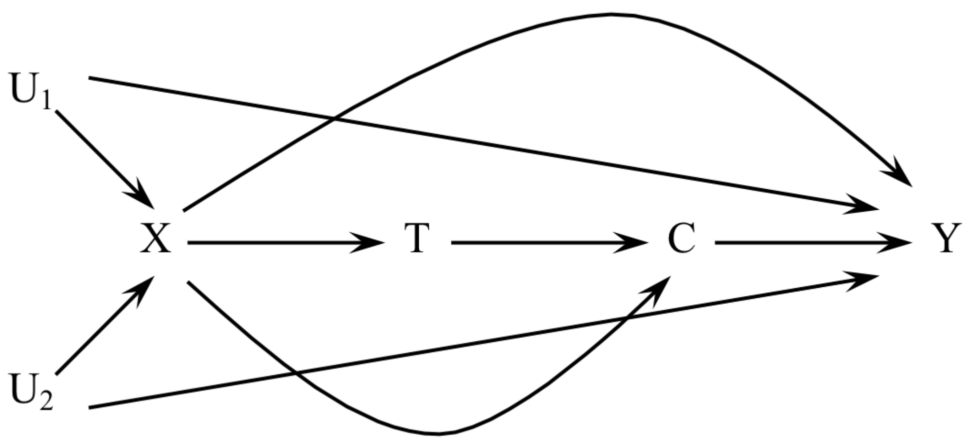
Causal DAGs



Definition: Causal DAGs include all common causes of any pair of variables already included in the DAG.

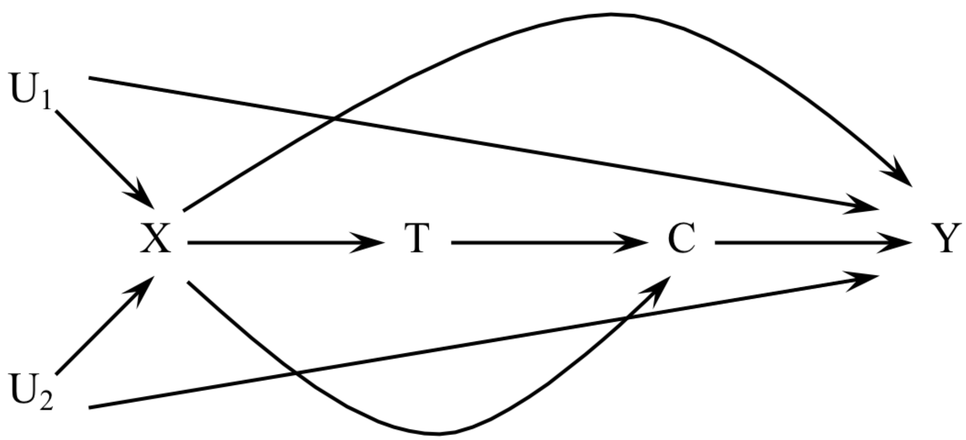
- E.g., there is no variable U_3 with direct effects into U_2 and T

Causal DAGs



Causal DAGs may include additional nodes that are not common causes (e.g. T, C), as long as all common causes involving these additional variables are also included.

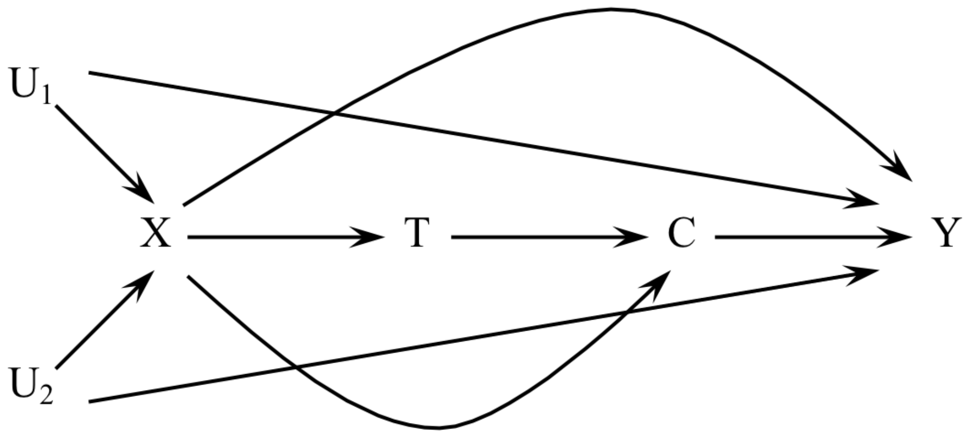
Causal DAGs



In everything that follows, we assume that the DAGs are <u>causal</u>.

- This is a very strong assumption, but it's often necessary for deriving identification results.
- Things get bad enough without doubting theory.

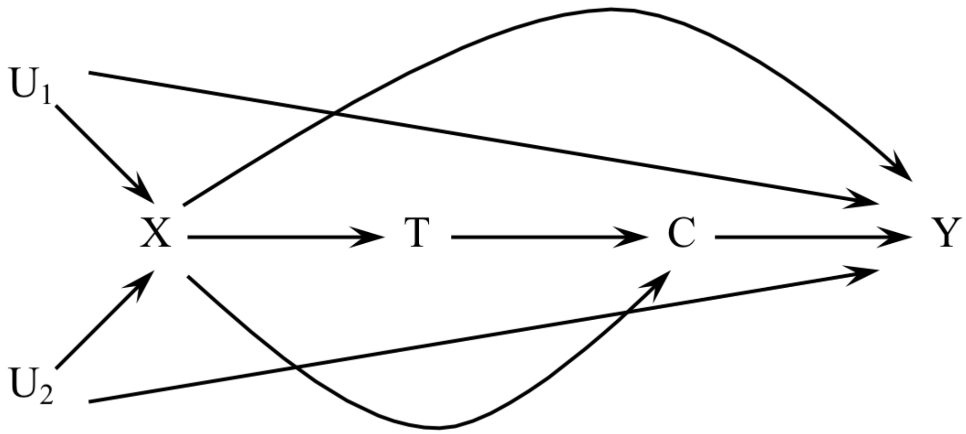
Causal DAGs



It can be shown that complicating a DAG by adding arrows to a given node set (i.e. relaxing assumptions) never helps non-parametric identification.

Adding nodes (variables) to a DAG, however, may help non-parametric identification.

Causal DAGs Encode the Data-Generating Model

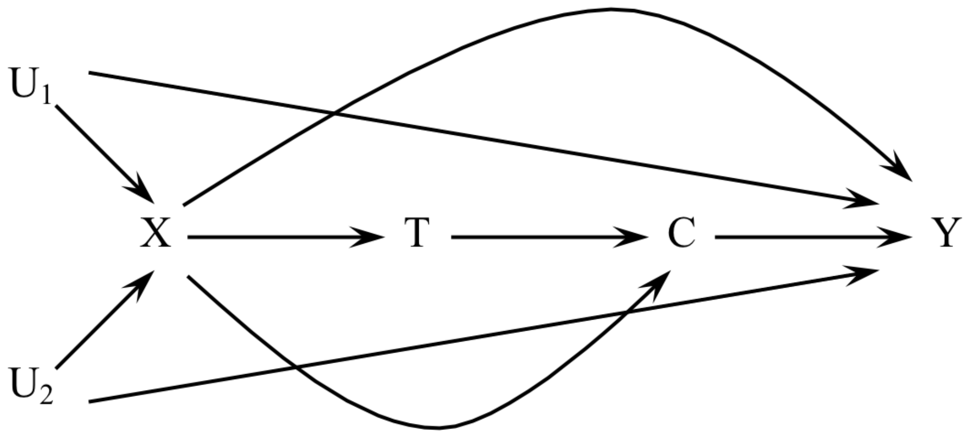


Causal DAGs encode the qualitative causal assumptions of the data-generating model ("model-of-how-the-world-works") against which all inferences must be judged.

Specifically, the DAG must capture the causal structure of

1. How the variables take their values in "nature"
2. What variables and values are collected

Consider Observed and Unobserved Variables

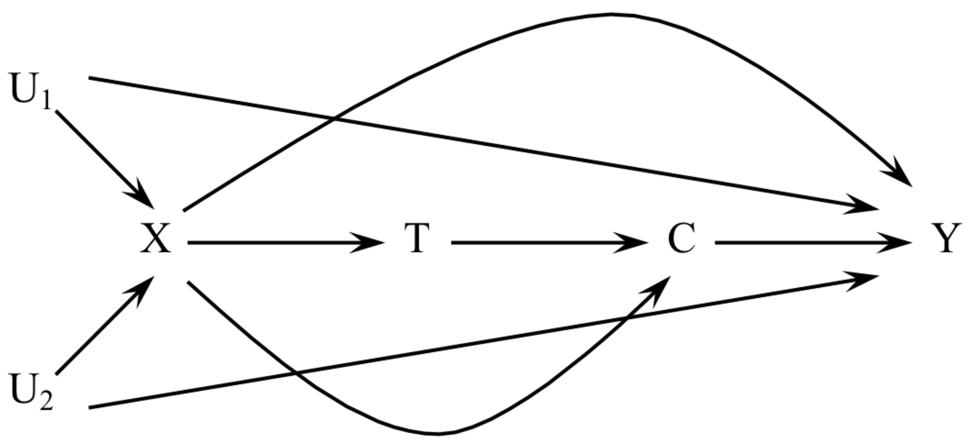


Note! When building a model (= drawing a DAG), you must consider <u>all</u> factors/variables that play a role in data generation, <u>regardless of whether they are observed or unobserved</u>.

Also include all variables on which we explicitly or implicitly condition in the course of analysis (including a variable for “data collection” itself if data collection is affected by any variable in the DAG). (Examples later.)

It is a common mistake to reason ‘backward” from the observed variables in a dataset rather than considering how the world really works.

Assumptions Are Unavoidable



If we are not willing to make assumptions, we cannot point-identify causal effects.

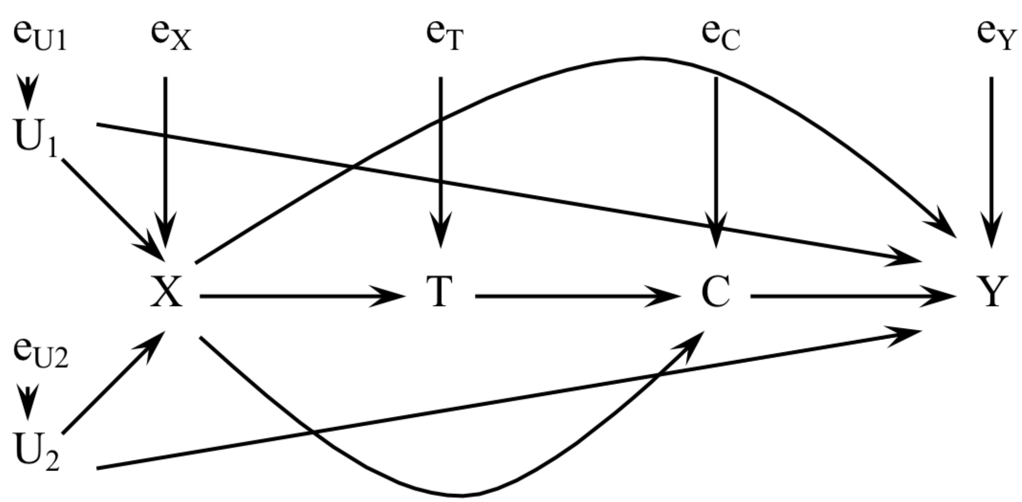
Assumption-free causal inference is impossible.

If our assumptions are wrong, our inference may be wrong.

We should make the least onerous assumptions possible.

"Garbage in, garbage out" (Nancy Cartwright).

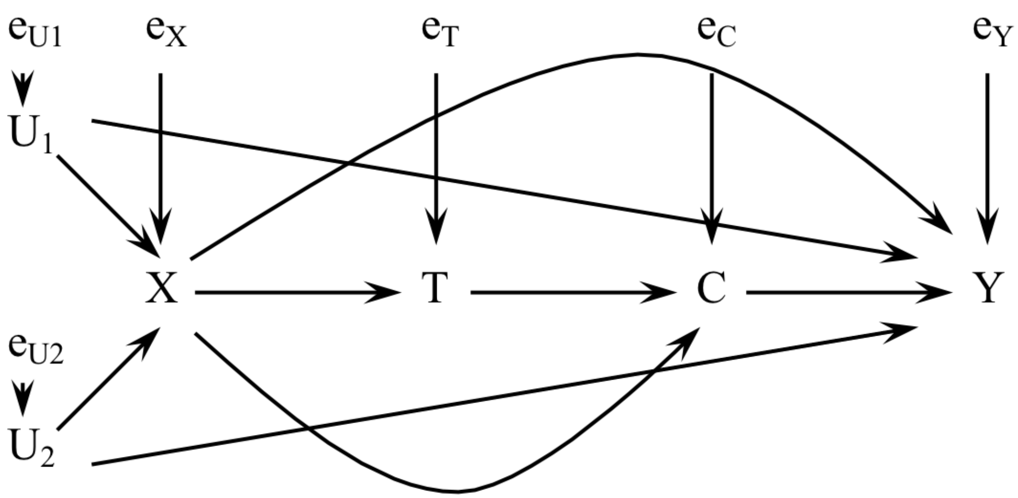
Suppressing Independent Errors



We usually omit the set of independent “error term” (idiosyncratic direct causes) on each variable from the causal DAG because they don’t help with non-parametric identification.

Sometimes, however, these independent error terms are useful for showing why a particular effect is *not* identified (examples later).

DAGs as NPSEM



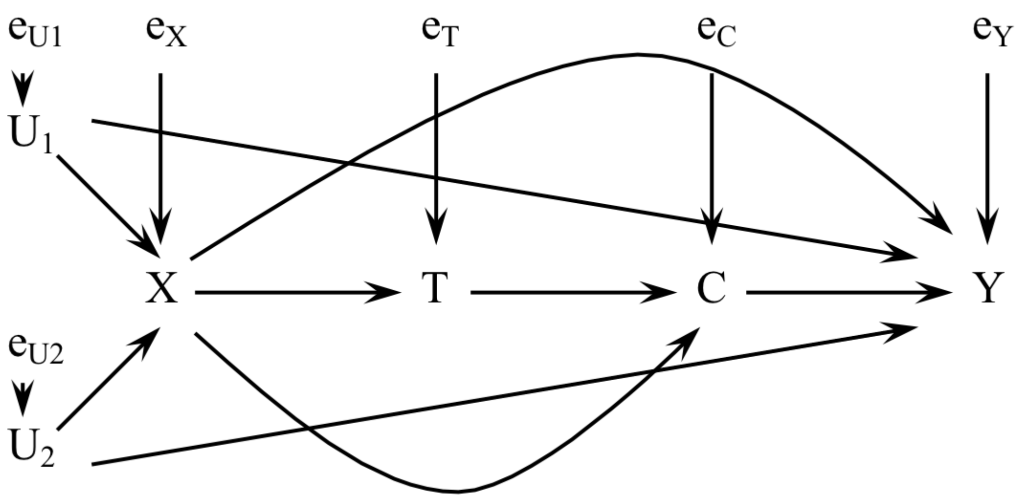
Technically, we interpret DAGs as visual representations of nonparametric structural equation models (NPSEM).

Each variable, V, is generated by a NPSE that relates the variable to its parents in the causal DAG, *pa(V)*, and to its idiosyncratic error, e_V, via some arbitrary deterministic function, f_V.

$$V = f_V(pa(V), e_V)$$

Since the error term is stochastic, the NPSEM is stochastic. Given the error terms, everything else is determined.

DAGs as NPSEM



The following equations give the NPSEM for the above DAG:

{e} are independently distributed

$U_1 = f_{U_1}(e_{U_1})$

$U_2 = f_{U_2}(e_{U_2})$

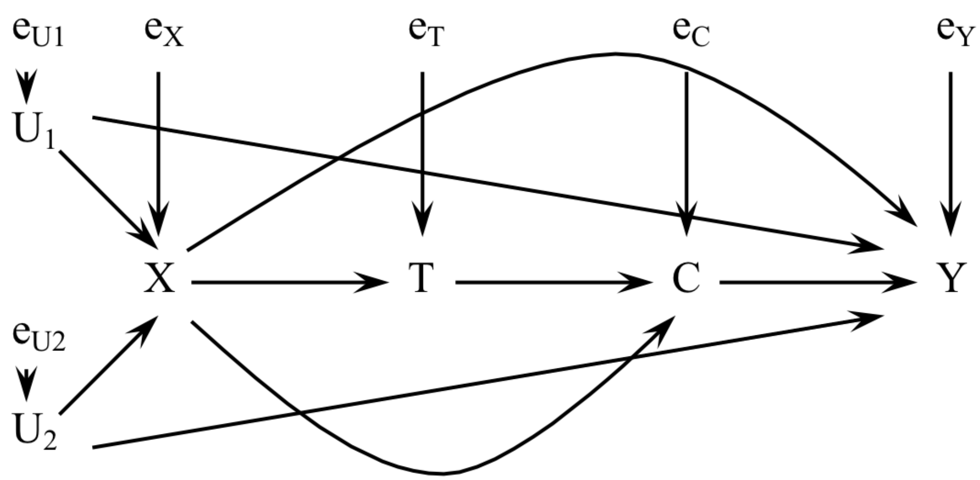
$X = f_X(U_1, U_2, e_X)$

$T = f_T(X, e_T)$

$C = f_C(T, X, e_C)$

$Y = f_Y(U_1, U_2, X, C, e_Y)$

A Qualitative DAG is Consistent With Multiple Quantitative Models



Since DAGs represent a NPSEM with arbitrary nonparametric functions *f*, a given DAG is compatible with many specific parametric models, e.g.,

$$C = f_C(T, X, e_C)$$

is compatible with

$$C = T + X + e_c$$

and with

$$C = \alpha * \frac{\beta_i \sqrt{T}}{X! * \sin e_C}$$

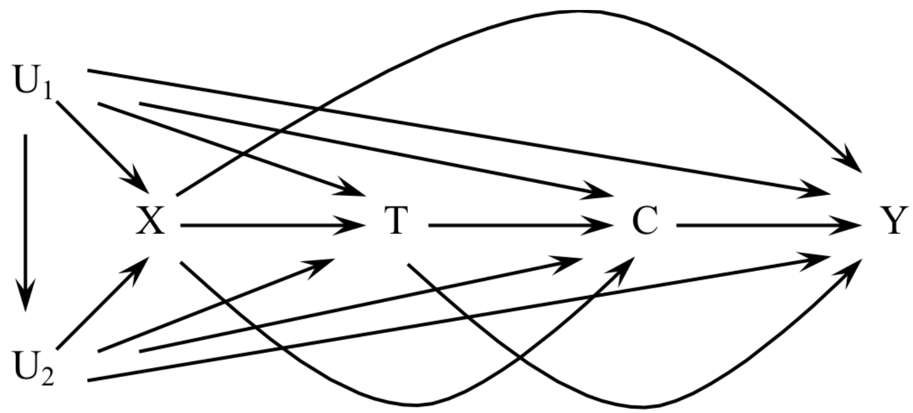
among others.

Elwert@wisc.edu. Version 5/2013

Therefore, every DAG is compatible with multiple joint probability densities.

Note: the parameters of *f* may vary across individuals, and *f* permits interactions.

Complete DAGs



A "complete DAG" (no missing arcs) is compatible with all possible probability density functions over the node set, $P(Y,C,T,X,U_1,U_2)$.

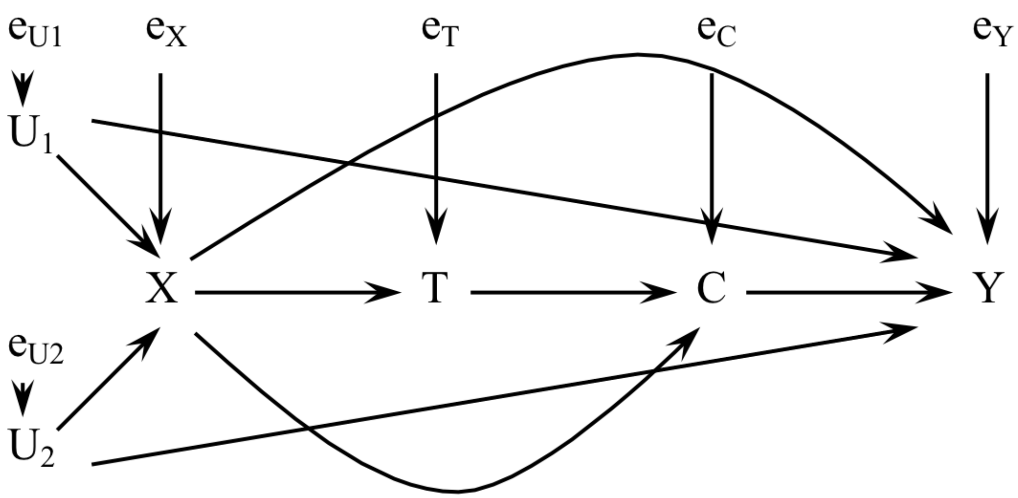
Incomplete DAGs are compatible with all possible pdfs over the node set subject to the constraints placed on the system by the missing arrows.

This makes DAGs a very general tool.

Drawbacks:

1. Interactions and effect heterogeneity are implied, not notated.
2. DAGs are primarily designed to assist with nonparametric identification (that's a good thing)

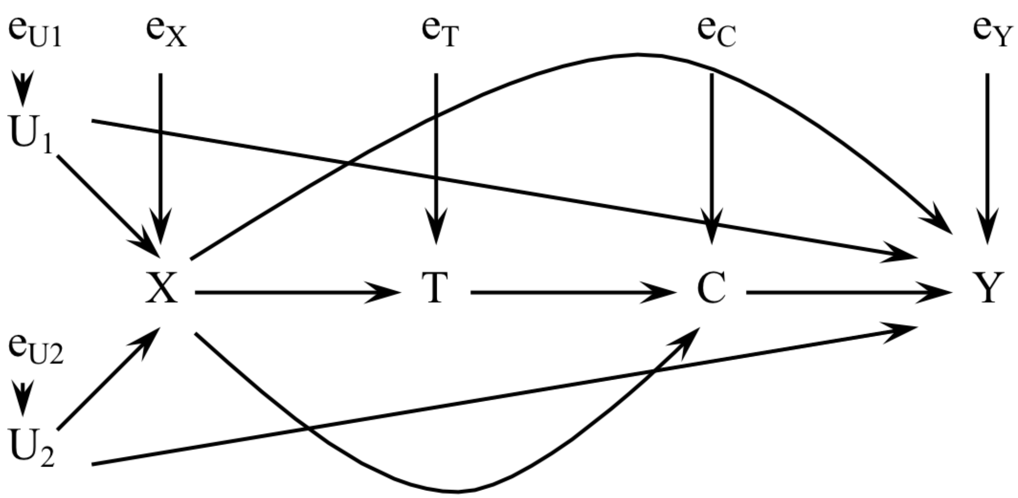
DAGs vs. Algebra



$$V = f_V(pa(V), e_V)$$

Note that it's easy to misread the equality signs in NPSEM. The r.h.s. is the causal inputs for the l.h.s.: Intervening in the world to change the r.h.s. will change the l.h.s. But intervening on the l.h.s. will not change the r.h.s—the future cannot change the past. (The symbols of conventional algebra weren't designed for causation—graphs are clearer.)

DAGs Represent the Population



Think of DAGs as population-level representations of the data (probability distributions) generated by a model

=> No sampling variability–identification is about bias, not efficiency

4. Deriving Testable Implications of a DAG

(d-Separation)

Moving from Causation to Association

Elwert@wisc.edu. Version 5/2013

Section Overview

- DAGs encode causal assumptions about the relationships between variables (i.e. a qualitative causal model of data generation).
- From these causal assumptions, one can deduce all associations (marginal and conditional dependences and independences) in the system.

$\Rightarrow$ In this section, we use DAGs to deduce the testable implications of causal models.

$\Rightarrow$ Later, we will use the DAG's causal assumptions and implied associations to determine which observable associations are causal, and which ones are not causal (spurious), i.e. which associations identify causal effects

Adages

1. "Association does not imply causation."
   - True: Just because two variables are associated does not mean that one causes the other.
2. "No association without causation"
   - Absent chance (sampling variation) all associations are assumed to have a causal origin.

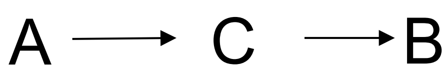
$\Rightarrow$#2 is arguably more helpful for our purposes

Reading Associations Off of DAGs

When are two variables in a DAG associated?

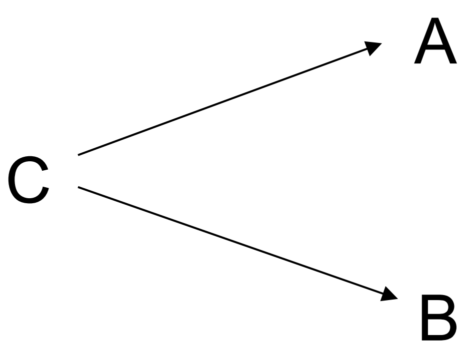
⇒We can determine all observable associations implied by the assumptions encoded in a DAG from just three elementary rules.

3 Sources of Association Between Two Variables A & B



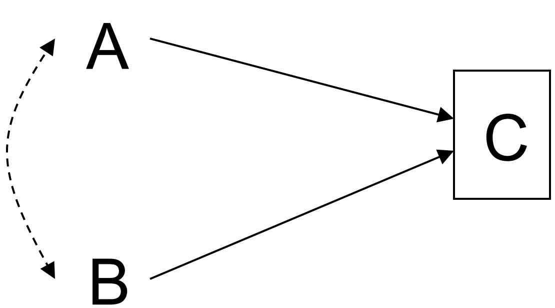
(1) Direct and indirect causation

$A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B|C$



(2) Common cause confounding

$A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B|C$



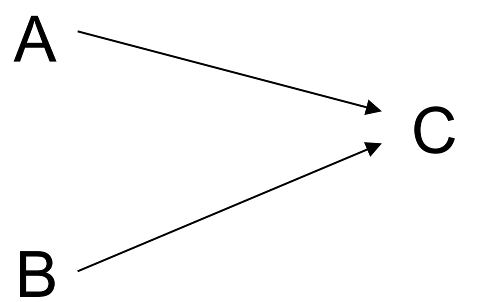
(3) Conditioning on a common effect (“collider”): Selection

$A \perp\!\!\!\perp B$ and $A \not\perp\!\!\!\perp B|C$

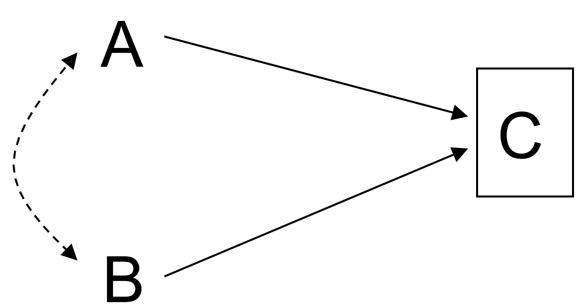
$\dashleftarrow\dashrightarrow$: non-causal (spurious) association. $\square$: conditioning.

Conditioning on a Collider

Notice: No causal effect of A on B



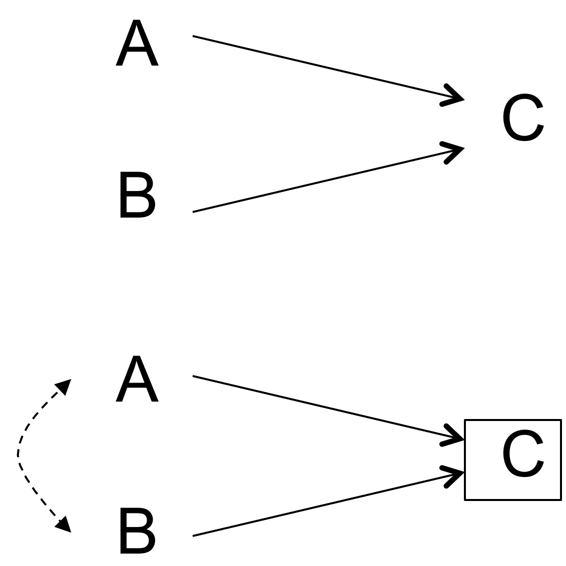
$A \perp\!\!\!\perp B$: marginally independent



$A \not\perp\!\!\!\perp B|C$: conditionally dependent
$\Rightarrow$ The association is <u>biased</u> for the true causal effect.

True under very mild conditions.

Conditioning on a Collider



Pearl's Sprinkler Example

A: It rains

B: The sprinkler is on

C: The lawn is wet

Hollywood Success

A: Good looks

B: Acting skills

C: Fame

Academic Tenure Example

A: Productivity

B: Originality

C: Tenure

In all three examples, conditioning on the collider C induces a spurious association between two variables, A and B, that don't cause each other and share no common cause, i.e. that are marginally independent in the population.

Conditioning on a Collider

Marginal Associations

All	B=0	B=1	
A=0	27	27	54
A=1	73	73	146
	100	100	OR=1.00

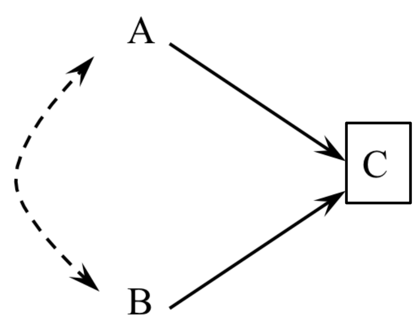
All	C=0	C=1	
A=0	35	19	54
A=1	68	78	146
	103	97	OR=2.11

All	C=0	C=1	
B=0	55	45	100
B=1	48	52	100
	103	97	OR=1.32

Conditional Associations

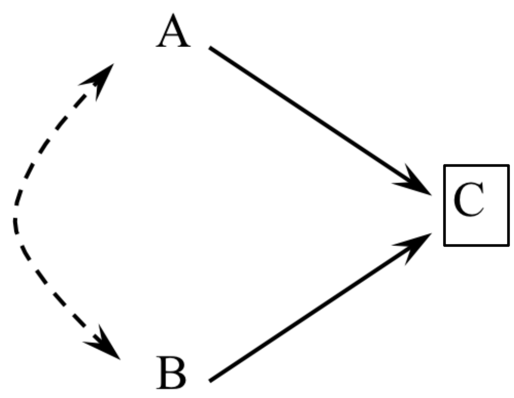
C=0	B=0	B=1	
A=0	20	15	35
A=1	35	33	68
	55	48	OR=1.26

C=1	B=0	B=1	
A=0	7	12	19
A=1	38	40	78
	45	52	OR=0.61



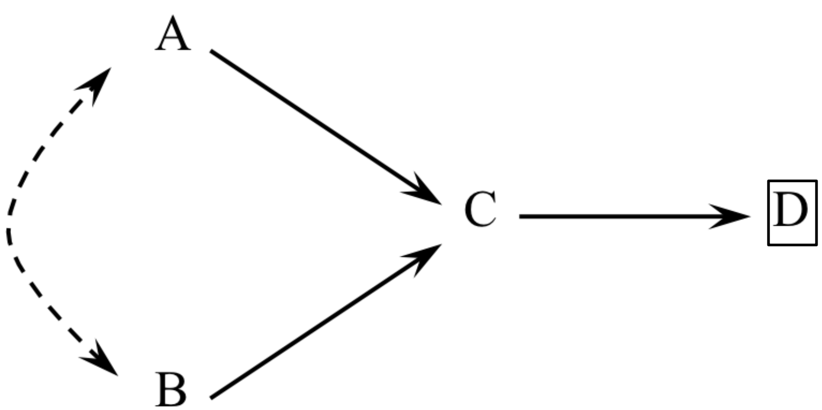
Notice: $A \perp\!\!\!\perp B$, but $A \not\!\perp\!\!\!\perp B$. The induced association between A and B flips signs across levels of C. (The two conditional associations do not cancel out in a logistic regression of Y on A and B.)

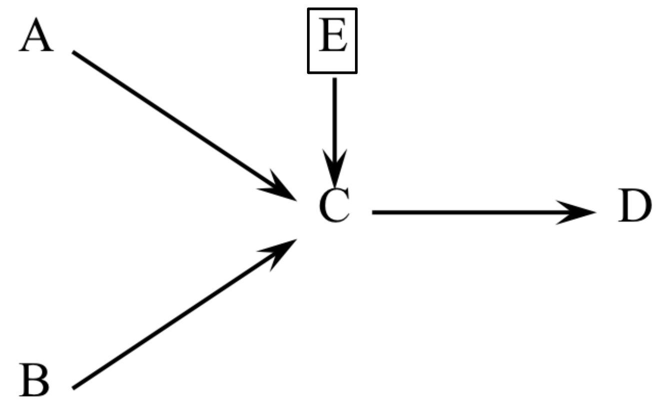
Conditioning on a Collider FAQ



1. True under very mild conditions (“faithfulness”: essentially, non-cancellation; exact cancellation has probability zero)
2. Adding arrows doesn’t help—conditioning on the collider would still distort the association between its ancestors.
3. Conditioning on a collider = Berkson’s Bias/Paradox (1946)

Conditioning on a Descendant of a Collider





Same problem as outright conditioning on the collider itself

Intuition: D carries information about C, so conditioning on D qualitatively amounts to conditioning on C itself.

(More precisely, D carries information about C and the information about A and B encoded in C. Information on C alone is not enough. E.g., conditioning on E would not induce an association between A and B because the information about C contained in E is independent of A and B.)

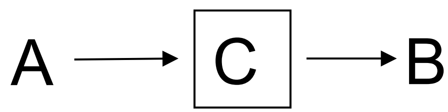
"Conditioning"

Conditioning may occur in the data analysis stage or in the data collection stage.

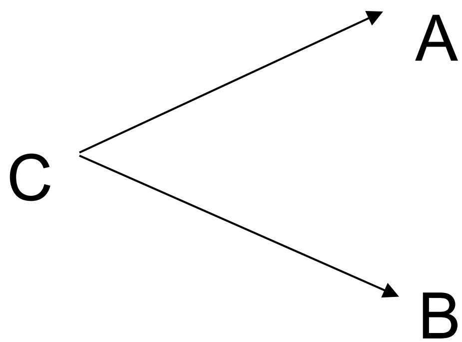
There are many ways to "condition" on a variable:

- Generally: introducing information about a variable into the analysis by some means.
  - Controlling (e.g. in regression)
  - Stratification (e.g. crosstabs, survival analysis, log-linear models)
  - Subgroup analysis (e.g. restrict analysis to employed women)
  - Sample selection (e.g. only collect data on poor whites)
  - Attrition, censoring, nonresponse (e.g., analyze only respondents or survivors)

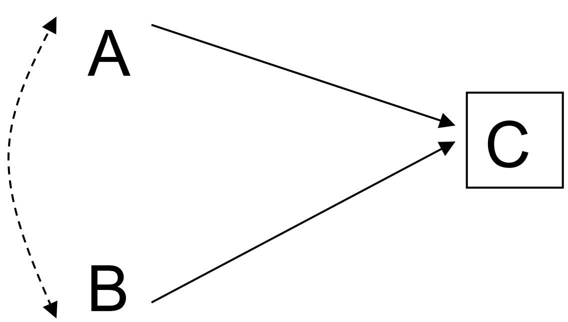
3 Sources of Bias in Estimating the Causal Effect of A on B



Overcontrol: intercepting the causal pathway



Confounding bias: failure to condition on a common cause



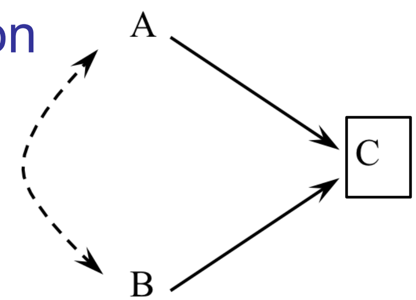
Endogenous selection bias: mistaken conditioning on a common effect.

Elwert@wisc.edu. Version 5/2013

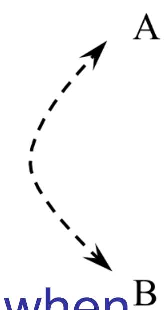
All three constitute analytic mistakes (Elwert 2013).

Bi-headed Arrows and Boxes

- Unless otherwise stated, we reserve bi-headed arrows for spurious association induced by conditioning on a collider—it acts like a real path (that exists only depending on what we condition on). (Note that this is just a handy shortcut for this workshop.)

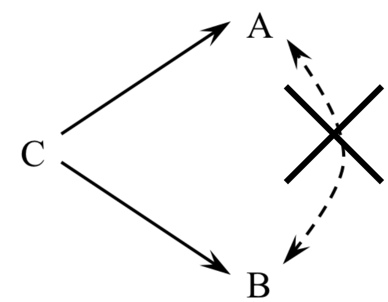


- In the literature, bi-headed arrows are also used to indicate confounding by an unobserved variable. In this workshop, I won't use that convention unless noted.



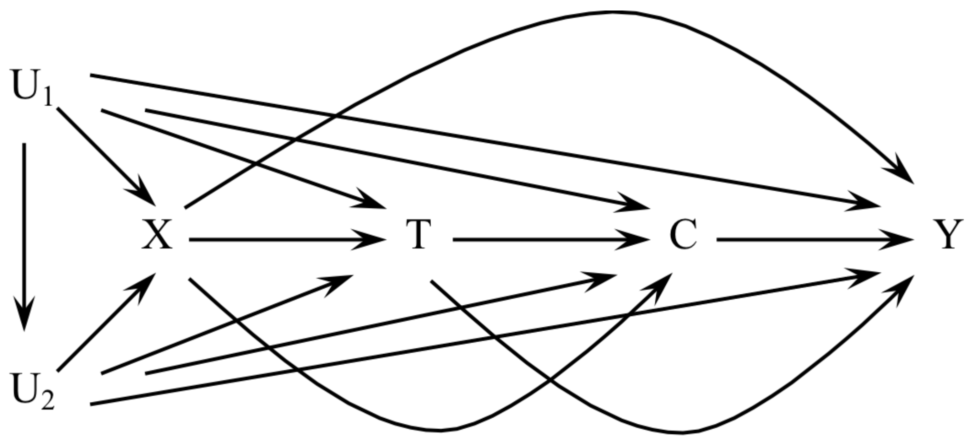
- We never use bi-headed arrows to indicate confounding when the confounding variable is explicitly drawn.

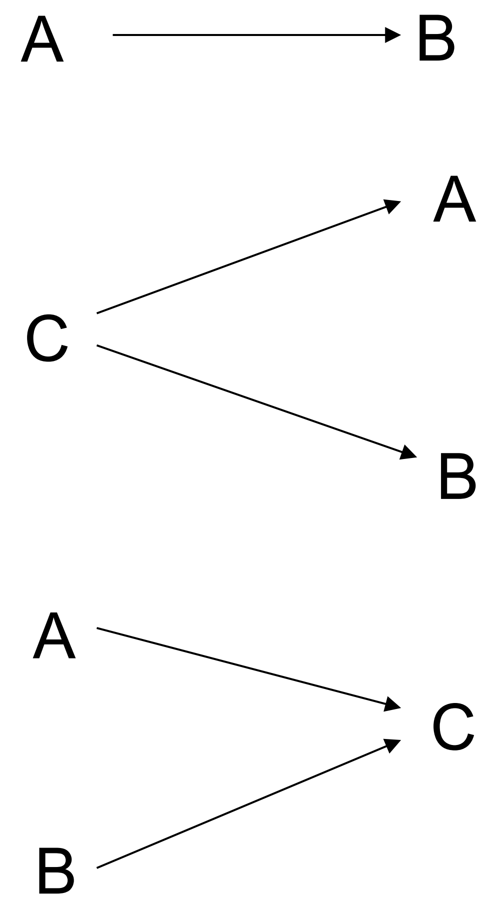
- Boxes around nodes indicate conditioning.



Putting It All Together: d-Separation

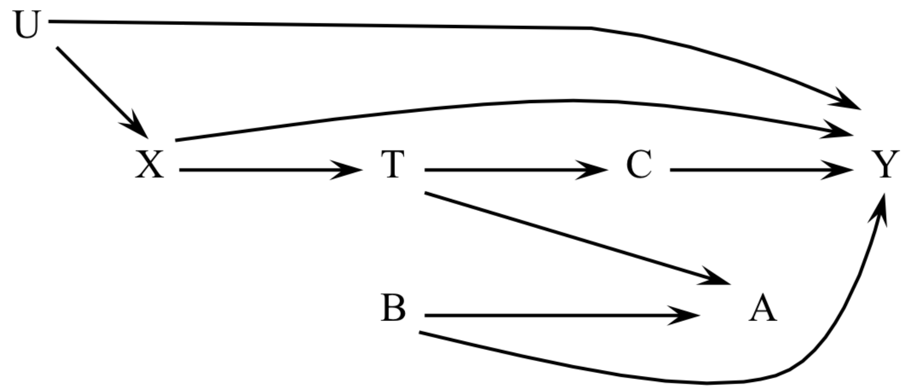
DAGs from the Elements





All DAGs can be constructed from just three elements—causal chains, forks, and inverted forks—the very elements that give rise to all associations via causation, confounding and endogenous selection.

Paths Transmit Association

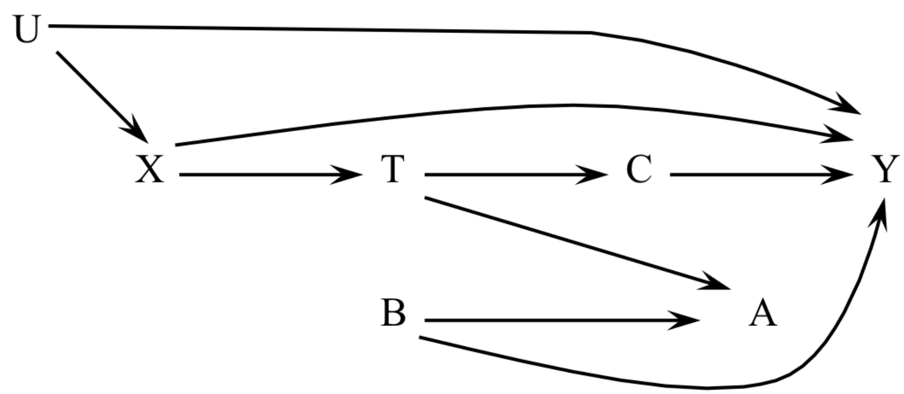


All associations are transmitted along paths.

But not all paths transmit association!

(Recall: A path is a non-intersecting sequence of adjacent arrows, regardless of the direction of the arrows; no path can pass through a given variable more than once.)

Causal and Non-Causal Paths



We distinguish two types of paths:

1. <u>Causal Path</u>: A path in which all arrows point away from the treatment, T, to the outcome, Y.

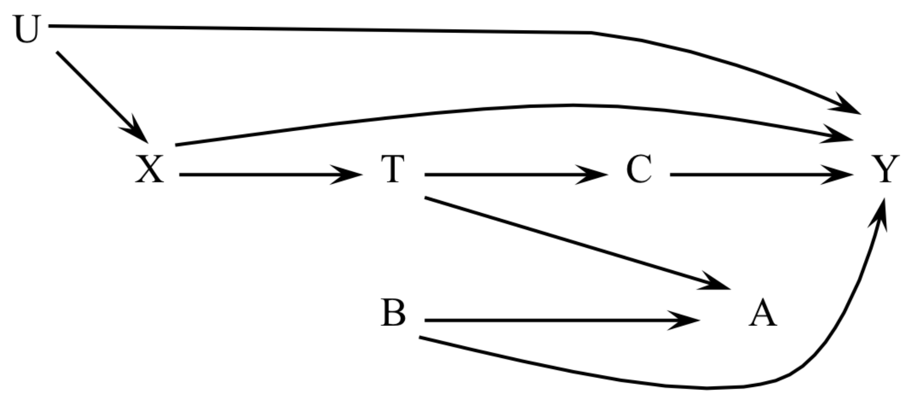
   Causal paths represent the causal effect of a treatment on the outcome. They give rise to the association of interest.

   The <u>total causal effect</u> of a treatment on an outcome consists of <u>all</u> causal paths connecting them.

2. <u>Non-causal path</u>: A path connecting T and Y in which at least one arrow points against the flow of time.

   Non-causal paths represent potential spurious sources of association between treatment and outcome.

Exercise: Paths



- List all causal paths from T to Y
- List all non-causal paths between T and Y

- List all causal paths from X to Y
- List all non-causal paths between X and Y

d-Separation

- The concept of d-separation ("directional separation;" Pearl 1988) subsumes the three structural sources of association and gives them a name.
- d-separation determines which paths transmit association, and which ones don't.
- We'll first state the definition formally and then explain it more accessibly.

d-Separation

- Definition: A path P is said to be "d-separated" (or "blocked") by a conditioning set of nodes {Z} iff
  1. P contains a chain I→M→J or a fork I←M→J such that the middle node M is in {Z}, or
  2. P contains a collider I→M←J such that neither the middle node M, nor any descendant of M, is in {Z}.
- Definition: A path P is said to be "d-connected" (or "unblocked" or "open") by a conditioning set of nodes {Z} iff it is not d-separated.
- Note: {Z} may be the empty set { }.

(See Pearl 2009)

Probabilistic Implications

Theorem: If two sets of variables X and Y are d-separated by Z along all paths in a DAG, then X is statistically independent of Y conditional on Z in every distribution compatible with the DAG.

Conversely, if X and Y are not d-separated by Z along all paths in the DAG, then X and Y are dependent conditional on Z in at least one distribution compatible with the DAG.

(Verma and Pearl 1988)

d-Separation:
Blocking and the "Flow" of Association

Here's the same phrased more accessibly:

- "Blocked" (d-separated) paths don't transmit association.
- "Unblocked" (d-connected) paths may transmit association.
- Three blocking criteria (key!!)
  1. Conditioning on a non-collider blocks a path
  2. Conditioning on a collider, or a descendent of a collider, unblocks a path
  3. Not conditioning on a collider leaves a path "naturally" blocked.

Probabilistic Implications

Two variables that are d-separated along all paths given {Z} are conditionally independent given {Z}.

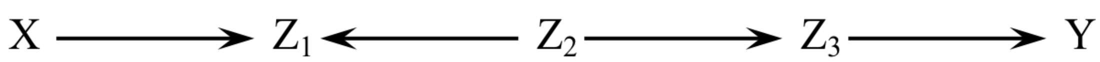
Conversely, two variables that are not d-separated along all paths given {Z} are potentially dependent given {Z}.

Note: “Conditioning” on Z here refers to perfect stratification on Z. Conditioning on some parsimonious function of a non-collider Z (e.g. in regression) may not fully block a path and let some residual association pass.

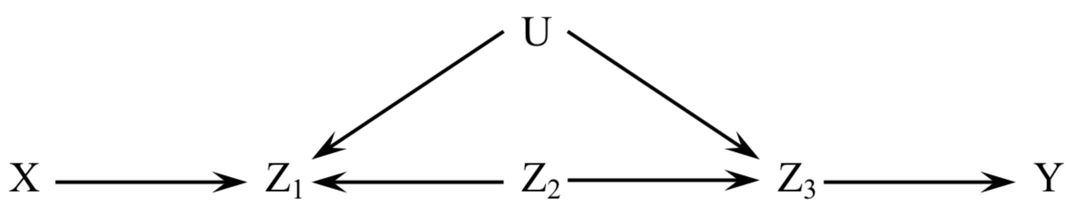
Examples 1 (d-separation)

- Which conditioning sets {Z} d-separate X and Y? List all permissible conditioning sets (including, if appropriate the empty set, { }). All U are unobserved.

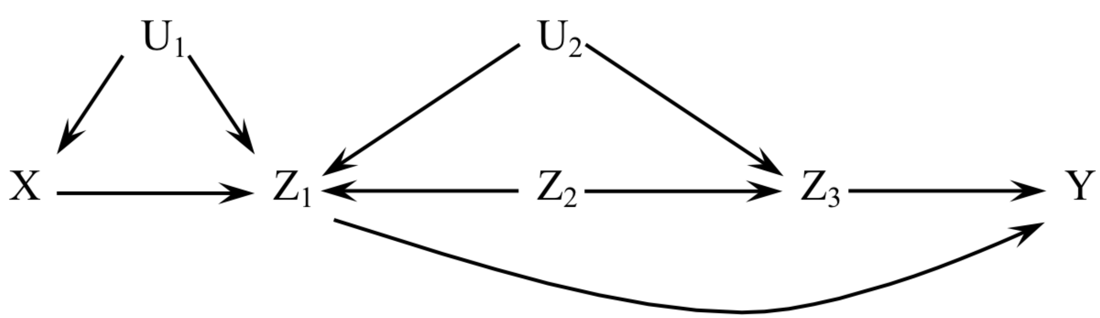
1.



2.



3.



Testable Implications

- One important use of DAGs is that they support the derivation of all testable (structural) implications of a model.
  - Since we know that causal inference requires causal assumptions, it's important to test one's assumptions empirically as far as possible.
- Using the d-separation/blocking criterion, we can read all implied marginal and conditional dependences and independences off the DAG.
- The implications involving only observed variables are testable!
  - Note: <u>independences</u> are considered a stronger test of a model because independences are implied under weaker conditions.

Testable Implications

These are the associational implications of the causal assumptions embedded in the DAG (Elwert 2013) (assuming "strong faithfulness"). The relations involving only observed variables are the <u>testable implications</u> of the model.

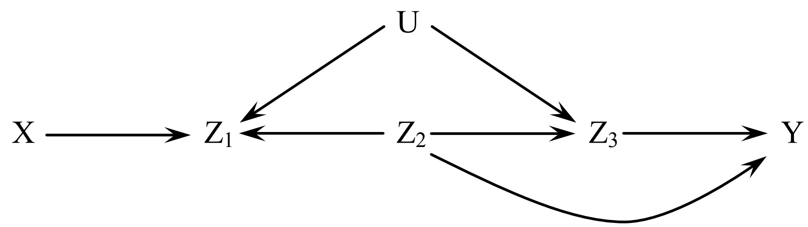


Table 13.1 All pairwise marginal and conditional independences and dependences implied by the causal DAG in Fig. 13.6

Independences		Dependences	
Marginal	Conditional	Marginal	Conditional
X and Z_2	X and Z_2 given (Z_3 or Y or U)	X and Z_1	X and Z_1 given (any other)
X and Z_3	X and Z_3 given (Z_2 or Y or U)	Z_1 and U	X and Z_2 given (Z_1 and (any other))
			X and U given (Z_1 and (any other))
X and Y	X and Z_3 given (Z_1 and Z_2 and U and (Y or ()))	Z_1 and Z_2	X and Z_3 given (Z_1 and (() or (Z_2 eo U) or Y)
	X and U given (Z_2 or Z_3 or Y)		
X and U	X and Y given (U or Z_2 or Z_3)	Z_1 and Z_3	X and Y given (Z_1 and (() or (Z_2 eo (U or Z_3))))
Z_2 and U	X and Y given (Z_1 and Z_2 and (U or Z_3))	Z_1 and Y	Z_1 and U given (any others)
	Z_1 and Z_3 given (U and Z_2 and (() or X or Y))	U and Z_3	Z_1 and Z_2 given (any others)
	Z_1 and Y given (Z_2 and (U or Z_3) and (() or X))	U and Y	Z_1 and Z_3 given (X or Y or (Z_2 eo U))
	Z_2 and U given X	Z_2 and Z_3	Z_1 and U given (X or Z_3 or (U eo Z_2))
			Z_2 and U given ((Z_1 or Z_3) and (any other))
	U and Y given ((Z_2 and Z_3) and (any other))	Z_2 and Y	Z_2 and Z_3 given (any others)
			Z_2 and Y given (any others)
			U and Z_3 given (any others)
			U and Y given (X or Z_1 or (Z_2 eo Z_3))
			Z_3 and Y given (any others)

Notes: "Any other" is any combination of other variables not already named, including the empty set; "or" is the inclusive "either one or both"; "eo" is the exclusive "either one but not both"; () is the empty set

Limitations of Model Testing

- Even under ideal conditions (infinite sample size, no measurement error), one can never test all assumptions in a model.

1. Recall that we must assume that the DAG is “causal,” i.e. that it includes all common causes. One can never test out the absence of <u>all</u> possible common causes, known and unknown, observed and unobserved.
2. Some models that are causally distinct (different DAGs) are “observationally equivalent,” i.e. they have identical observational implications, e.g. (if U is unobserved),

(Observational equivalence is a big topic in (so far largely theoretical) work dealing with the recovery of latent causal structures from observational data.)

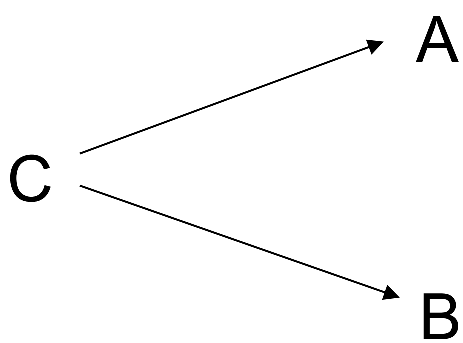
Elwert@wisc.edu. Version 5/2013

Examples 2: Testable Implications

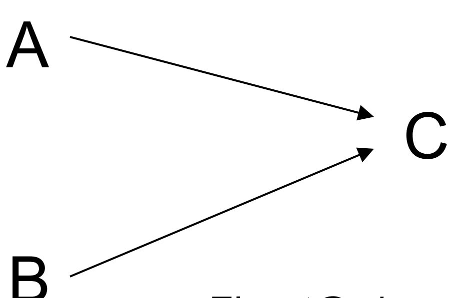
- List the testable independences implied by these DAGs.
- Can we observationally distinguish between DAGs 1, 2, and 3; or between DAGs 4 and 5?

1.

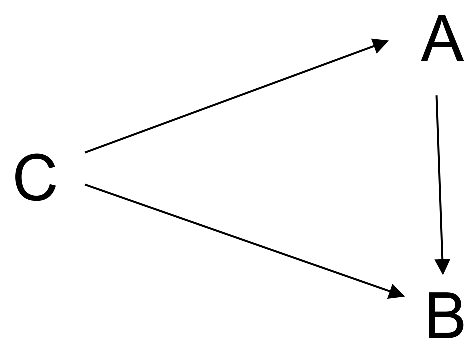
2.



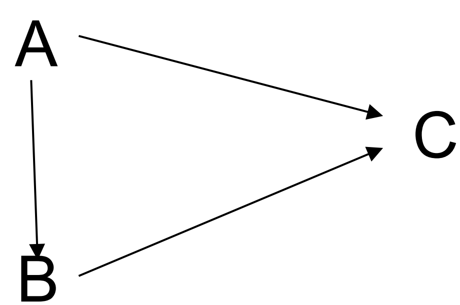
3.



4.



5.



5. Graphical Identification Criteria

Section Overview

- Graphical identification criteria
- Adjustment criterion
- Backdoor criterion
- Some sufficient shortcuts

Identification

- The causal effect of T on Y is said to be "identified" if it is possible, with ideal data (infinite sample size, perfect measurement), to purge all non-causal association from the observed association between T and Y such that only the causal association remains.

- Question: How can one tell with DAGs whether a total causal effect is nonparametrically identified? (There are many answers to this question—stay tuned!)

Graphical Identification Criteria

- Pearl and others have discovered a great number of graphical identification criteria.

- A graphical identification criterion is a set of rules that use DAGs (“graphs”) to detect identifiability in a class of models represented by a DAG.

Graphical Identification Criteria

- Most graphical identification criteria give sufficient conditions (i.e., “if this criterion is met, then the model is nonparametrically identifiable”). Note that a model may fail a given sufficient criterion and still be identified by some other criterion.
- Only Pearl’s (1995) “do-calculus” is a complete identification criterion (i.e., can tell for all models whether or not they are nonparametrically identified).
- Here, we focus on arguably the most important of sufficient criterion: the adjustment criterion (which generalizes Pearl’s famous backdoor criterion)

Adjustment Criterion

Adjustment Criterion (Shpitser et al. 2010)

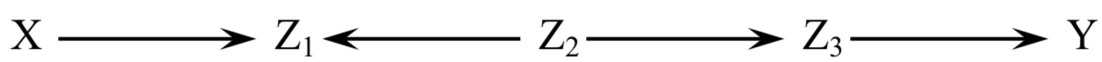
- Recall that the ACE of T on Y is said to be “identified” if it is possible, with ideal data to purge all non-causal association from the observed association between T and Y such that only the causal association remains.
- One way to interpret this with DAGs, is to note that the total causal effect of T on Y is identifiable if one can condition on (“adjust for”) a set of variables {Z} that
  1. blocks all non-causal paths between T and Y,
  2. without blocking any causal paths between T and Y.
- (Equivalently: d-separate T and Y along all noncausal paths while leaving all causal paths d-connected.)

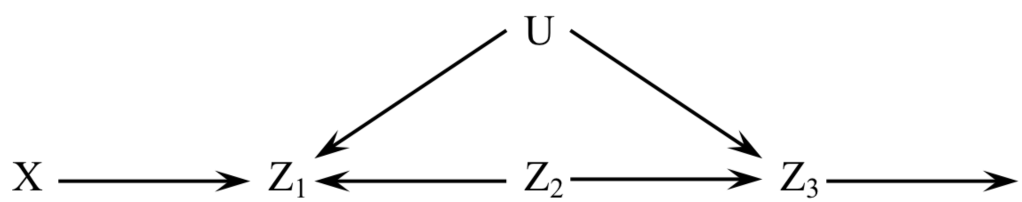
Adjustment Criterion (Shpitser et al. 2010)

- Formally:
- A set of variables {Z} (which may be empty) fulfills the adjustment criterion relative to the total causal effect of T on Y iff
  1. {Z} blocks all noncausal paths from T to Y, and
  2. No element of {Z} is on a causal path from T to Y or descends from a variable on a causal path from T to Y.

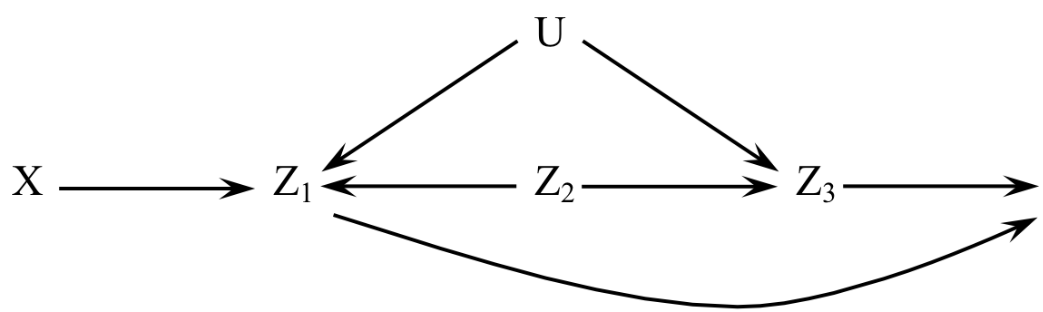
Examples 3 (identification)

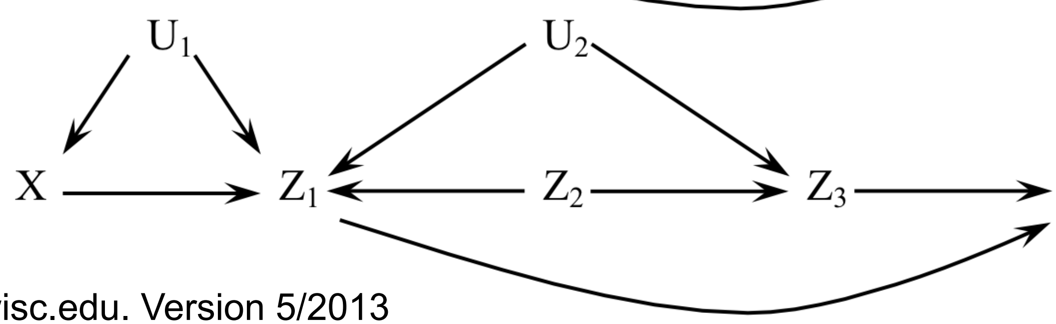
- Can we identify the ACE of X on Y by conditioning on some set {Z} (which may be empty)?





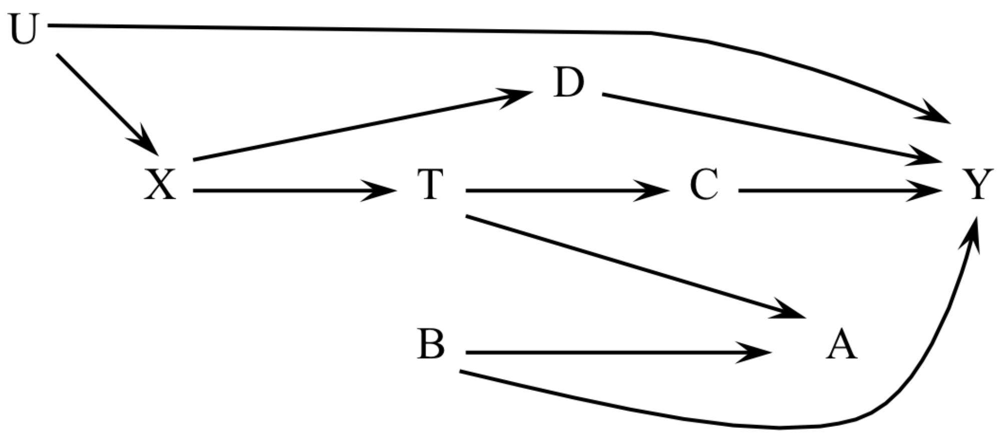
U is unobserved





Examples 4

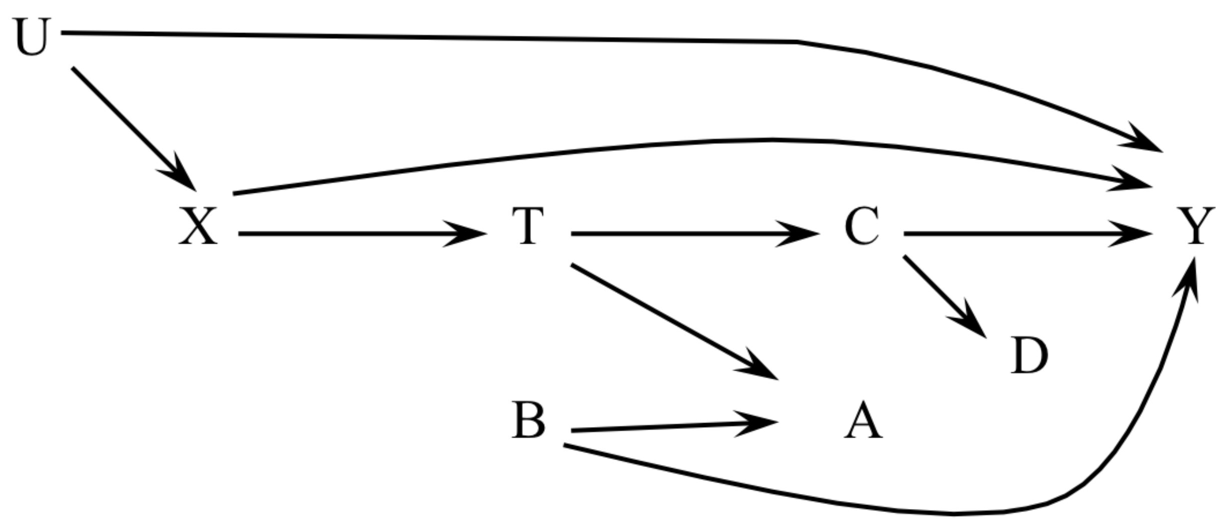
- Can we identify the ACE of T on Y by conditioning?
- Can we identify the ACE of X on Y by conditioning?



U is unobserved

Examples 5

- Why does conditioning on {X, D} violate the adjustment criterion for the causal effect of T on Y?



U is unobserved

The Adjustment Criterion is Complete

- The adjustment criterion is “complete,” meaning that it detects all, and only those, sets of variables {Z} that identify the effect of T on Y by simply conditioning on {Z}. (Shpitser et al. 2010)

DAGs, Ignorability, Counterfactuals

Shpitser, VanderWeele, and Robins (2010) prove:

Iff {Z} satisfies the adjustment criterion for the effect of T on Y, then T is ignorable given {Z},

$$\{Y_T\} \perp T \mid Z.$$

Question: Which variables should or can one control for to get ignorability?

Answer: The variables that satisfy the adjustment criterion relative to the DAG that generated the data!

DAGs, Ignorability, Counterfactuals

"Reducing ignorability conditions to [a graphical criterion] replaces judgments about counterfactual dependencies with ordinary judgments about cause-effect relationships" (Pearl 2009:80).

Excursus: Nonparametric estimation

If the adjustment criterion is met for the total causal effect of T on Y given Z then it implies a non-parametric estimator:

The distribution of the counterfactuals Y_T, $P(Y_T)$, is given by

$$P(Y_T) = \sum_Z P(Y \mid T, Z)\, P(Z).$$

The total causal effect of T on C for a binary Y is non-parametrically estimated (with discrete variables) as:

$$\Pr(Y_1=1) - \Pr(Y_0=1) = \sum_z \Pr(Y=1 \mid T=1, Z=z)\, P(Z=z) - \sum_z \Pr(Y=1 \mid T=0, Z=z)\, P(Z=z).$$

Unfortunately, we rarely have datasets large enough to implement this estimator (because Z tends to be high dimensional). (For continuous variables, replace sums with integrals.)

The mechanics are explained in Pearl (2009) Ch1 & Ch3.

Backdoor Criterion

Elwert@wisc.edu. Version 5/2013

Adjustment Criterion vs. Backdoor Criterion

- The adjustment criterion gives all identifying conditioning sets. But it sometimes includes variables in the conditioning set that, though valid, are not necessary.
  - E.g., in this DAG, conditioning on Z is permissible by the adjustment criterion for identifying the effect of T on Y, but it is not necessary:

$$Z \leftarrow T \rightarrow Y$$

  - In fact, it is never necessary to adjust for a descendant of treatment, and often harmful.
- The backdoor criterion is a narrower version of the adjustment criterion that omits some unnecessary conditioning sets.
- The backdoor criterion is easily the most famous graphical identification criterion, and it underlies all regression-based identification for observational data.

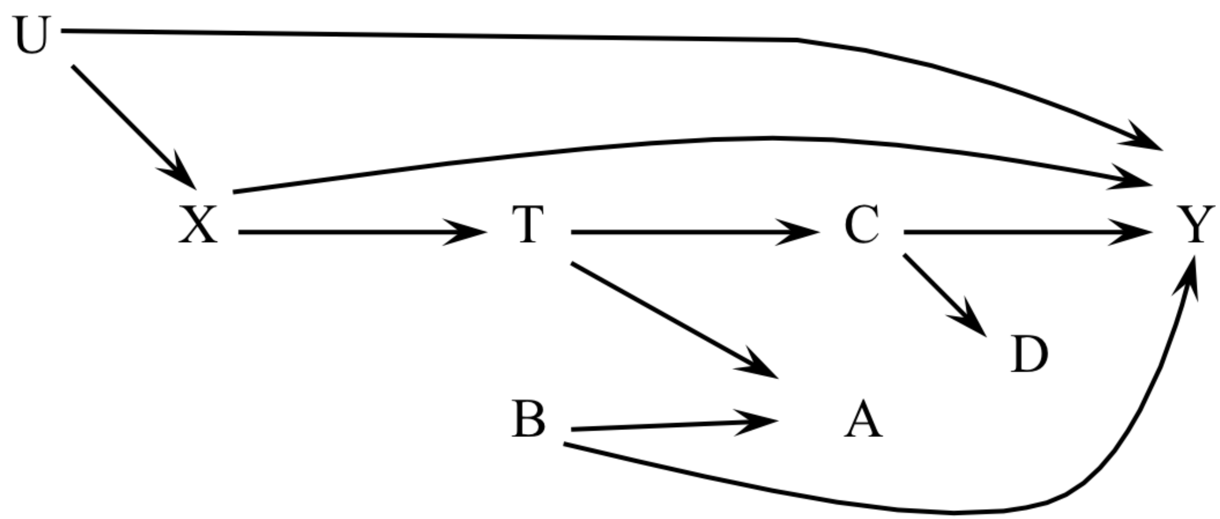
Backdoor Criterion (Pearl 1995)

- Definition: A set of variables {Z} satisfies the backdoor criterion relative to an ordered pair of variables (T,Y) in a DAG if:
  1. no node in {Z} is a descendant of T, and
  2. {Z} blocks (d-separates) every path between T and Y that contain an arrow into T (so-called "backdoor paths").
- Theorem: The total causal effect of T on Y is non-parametrically identifiable given {Z} if {Z} meets the backdoor criterion.

Backdoor Criterion (With Comments)

Claim: The total causal effect of T on Y is non-parametrically identifiable given {Z} if {Z} meets the backdoor criterion..

- We need to block all non-causal paths between T and Y.
  - Paths starting "T→" are either causal paths of interest or naturally blocked non-causal paths. Therefore, we never need to condition on a descendant of T.
  - Backdoor Paths starting "→T", however, are always noncausal, and they may be open. Thus, we may need to block them.
- What should be in the conditioning set {Z}?
  - A (possibly empty) set of variables that blocks all backdoor paths.
- What should not be in the conditioning set {Z}?
  - Any descendant of T because conditioning on any variable downstream from T is either unnecessary or harmful (by blocking a causal path or unblocking a non-causal path). [Exercise: verify this against our past examples]


U
X
T
C
Y
D
B
A

Backdoor Identification In Practice

Here's an algorithm for the backdoor criterion (less cumbersome algorithms exist, but this one is good for practice).

To check identification:

1. List all backdoor paths connecting T and Y.
2. Check whether all backdoor paths are naturally (unconditionally) blocked. [Yes: identified. No: move on.]
3. Check whether the unblocked paths can be blocked by conditioning on non-descendants of T. [Yes: move on. No: not identified.]
4. Check whether Step 3 unblocked any non-causal paths and then check if those can be blocked. [Yes: move on. No: not identified.]
5. Check whether any of the variables that must be conditioning on to block backdoor paths are on the causal pathway from T to Y or are descendants of a variable on the causal pathway. [Yes: Not identified. No: identified.]

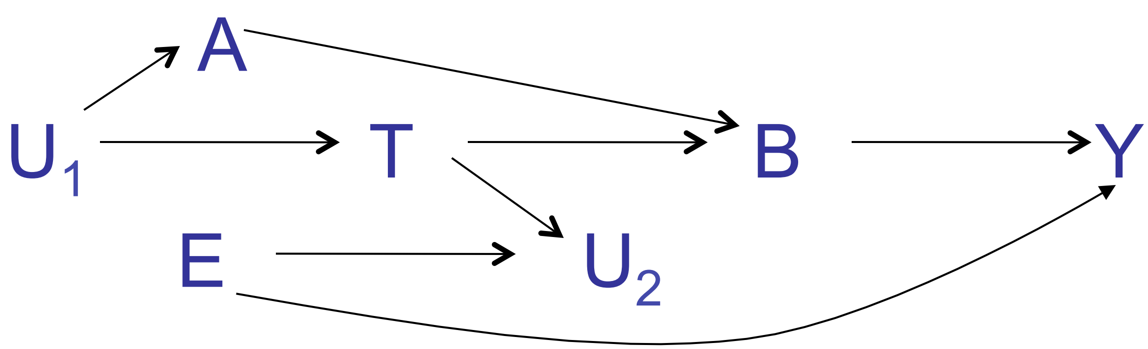
Instructions

Instructions for all examples below:

1. Assume that the DAG is causal
2. Check identification via the adjustment criterion (or, equivalently, the backdoor criterion)
3. All variables U are unobserved. All other variables are observed.
4. If an effect is identifiable, write down a permissible conditioning set.
5. If an effect is not identifiable, list the non-causal path(s) that cannot be blocked.

Example 6

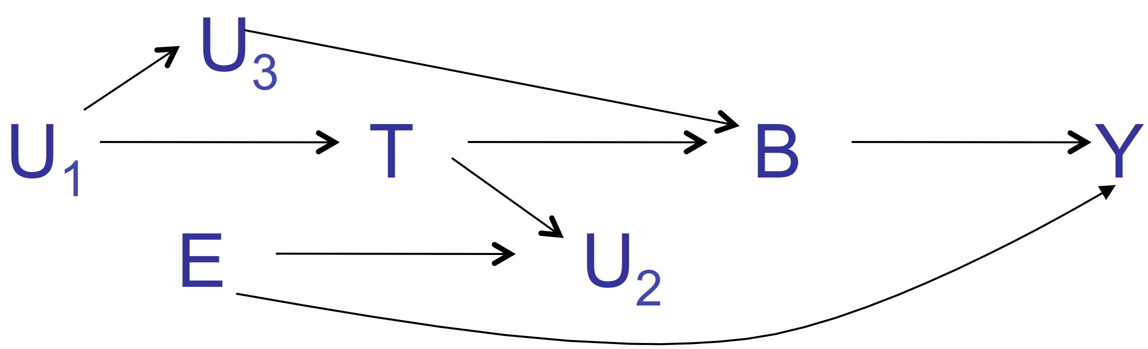
Is the causal effect of T on Y identifiable by adjustment?



Note: {U} are unobserved, all other observed.

Example 7

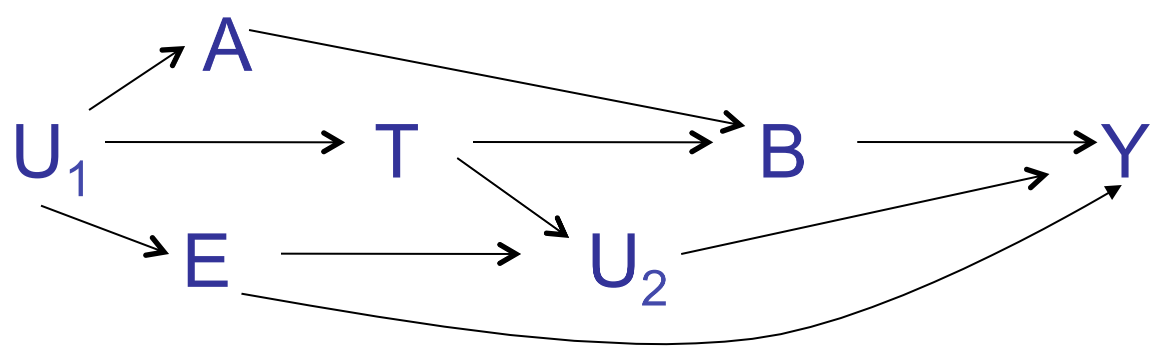
Is the causal effect of T on Y identifiable by adjustment?



Note: {U} are unobserved, all other observed.

Example 8

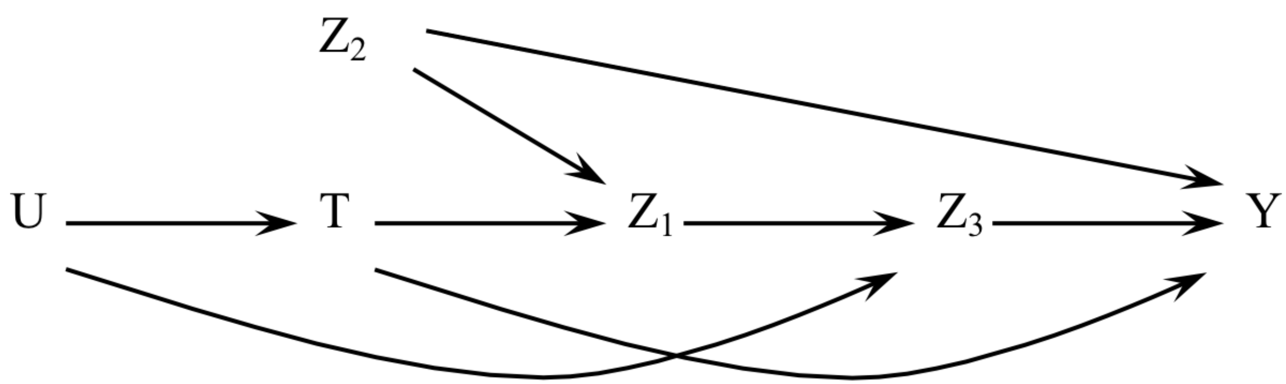
Is the causal effect of T on Y identifiable by adjustment?



Note: {U} are unobserved, all other observed.

Example 9

Is the causal effect of T on Y identifiable by adjustment?



Note: All but U are unobserved.

Some observations

- Note that more than one set {Z} may satisfy the adjustment (or backdoor) criterion.
- We need not necessarily control for all direct causes of T, nor for all direct causes of Y in order to identify the ACE of T on Y. (This is a fairly deep insight—think about it.)
- Please note that I've assumed a single treatment and a single outcome variable. The adjustment and backdoor criteria generalize to multiple treatment and outcome variables (see Pearl 2009 for glorious detail, see Elwert 2013 for an introduction).

If the DAG is Not Fully Known

- Applying the adjustment criterion is difficult if the DAG isn't fully known.
- Conventional practice suggests controlling for all pre-treatment variables ("kitchen sink" approach).
- As we will see in the next unit (on endogenous selection), controlling for the pre-treatment kitchen sink is dangerous.
- Sufficient and safe recommendation: If there is a set of observed pre-treatment variables that meet the adjustment criterion (i.e., if we can assume conditional ignorability), then controlling only for those observed variables that either cause treatment or outcome (directly or indirectly) is sufficient (VanderWeele and Shpitser 2011).

Helpful Weaker Criteria

- The backdoor and adjustment criteria can be somewhat unwieldy in practice. Weaker sufficient criteria (that apply to smaller classes of models) exist. Here are some:
- <u>“Parent criterion”</u>: If all parents of T are observed (no arrows U→T), then the effect of T on Y is identifiable by conditioning on its parents.
- <u>“Strong bow-pattern criterion”</u>: If some unobserved variable has arrows into T and into a child of T on a causal pathway to Y then the effect of T on Y is not nonparametrically identifiable by any criterion.
- <u>“Weak bow-pattern criterion”</u>: If some unobserved variable has arrows into T and any descendant of T on the causal pathway to Y then the effect of T on Y is not nonparametrically identifiable by the adjustment criterion.

Summary

- DAGs graphically encode data-generating non-parametric structural equation models.

- DAGs encode the analyst's qualitative causal assumptions.

- Three simple rules (causation, confounding, endogenous selection) translate between causation and association (d-separation).

- The d-separation criterion detects the testable implications in the model.

- The adjustment criterion (and the backdoor criterion) give the conditioning set(s) of variables that permit the non-parametric identification of a total causal effect by simple conditioning.

- This is the strategy implicitly invoked when we give a causal interpretation to matching or regression models.

- DAGs are compatible with the potential outcomes (counterfactual) framework of causality.

- The adjustment criterion reveals which variables give (conditional) ignorability.

- DAGs offer a rigorous yet intuitive and entirely algebra-free approach to really complicated causal inference challenges

- Consult Pearl's book *Causality* (2009) for more results, details, references, and proofs.

Cautionary Remark

- In observational social and medical science, we often lack theory to credibly delete arrows from a DAG (Greenland 2010). This is a real problem since it's the missing arrows that enable identification.
- In this unit, we have reviewed the basic tools required to derive <u>positive</u> identification results conditional on the assumption that the DAG is true. You may question whether the exercises (resting on sly exclusion restrictions) are particularly realistic. Fortunately, we can also use DAGs for another purpose: deriving <u>negative</u> identification results.
- I find that the practical value of DAGs often lies in revealing which identification strategies do <u>not</u> work. To show that something does not work given a theory (DAG), you needn't worry so much about strong assumptions—if it doesn't work in a sparse DAG, it won't work in a DAG with more arrows (though it may work with additional nodes). We'll see some scary examples in the next unit.

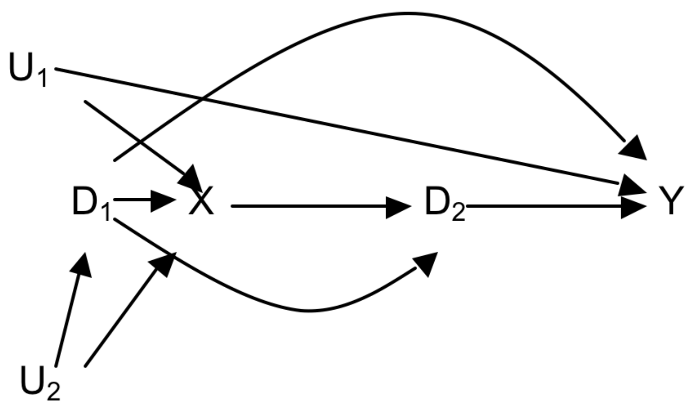
The End

Elwert@wisc.edu. Version 5/2013

More Exercises

Example 10

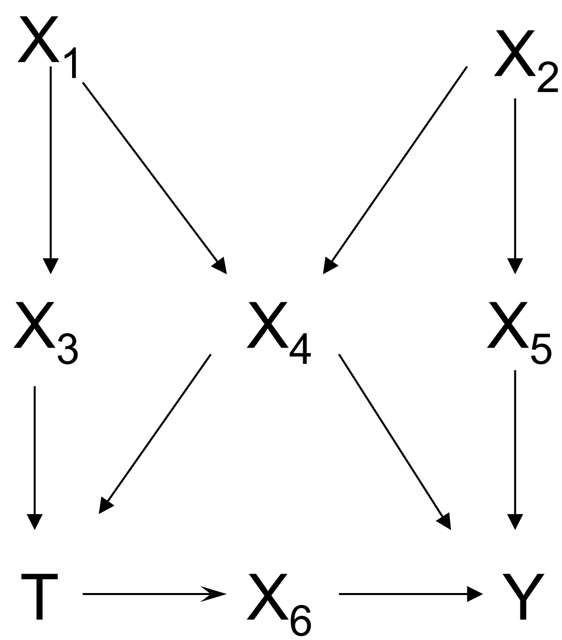
Can we identify the causal effect of D_1 on Y by adjustment? What about D_2 on Y?



Note: {U} are unobserved, all other observed.

Example 11

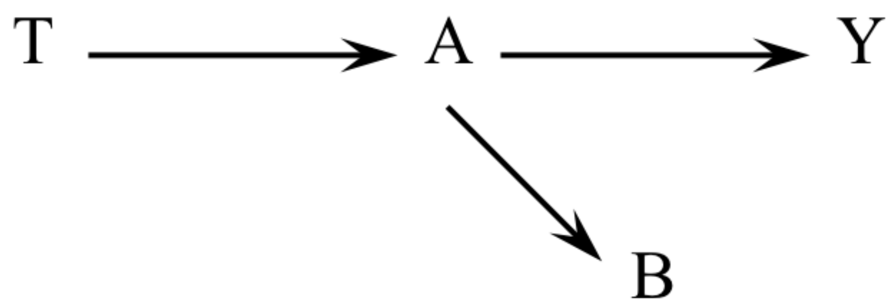
Which variables X_j need to be in {Z} to satisfy the backdoor Criterion for the causal effect of T on Y?



A canonic example from Pearl (1995)

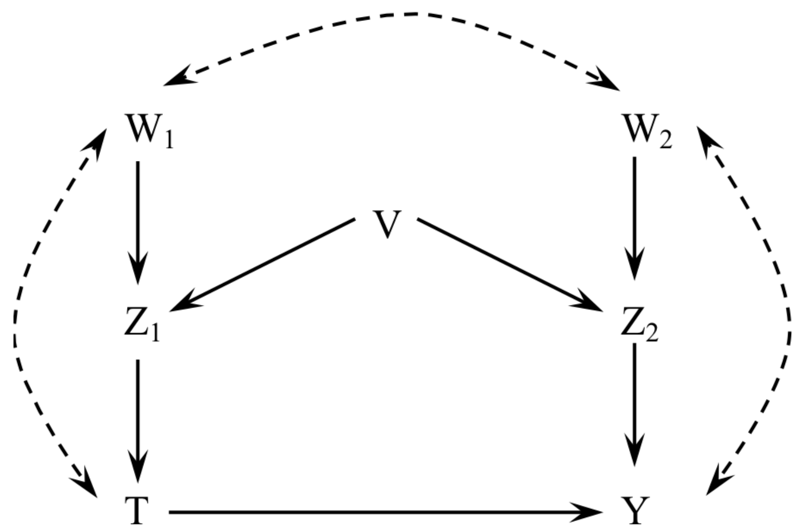
Example 12

Why would conditioning on B ruin identification of the causal effect of T on Y?



Example 13

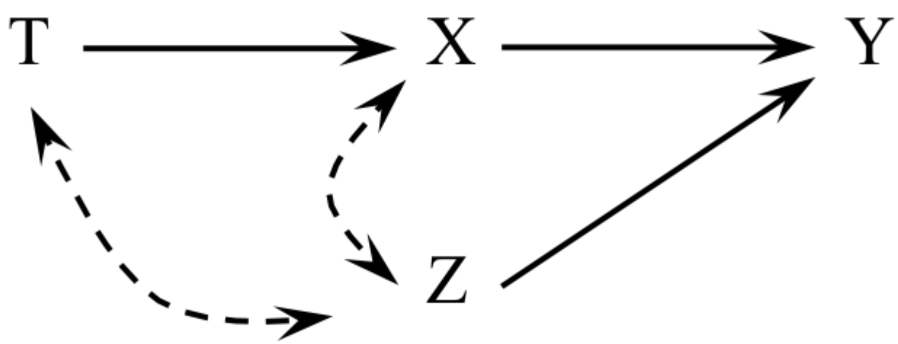
Can we identify the effect of T on Y by adjustment?



Bi-headed arrows here represent confounding by unobserved variables. (Pearl, 2009, p.345).

Example 14

Can we identify the effect of T on Y by adjustment?

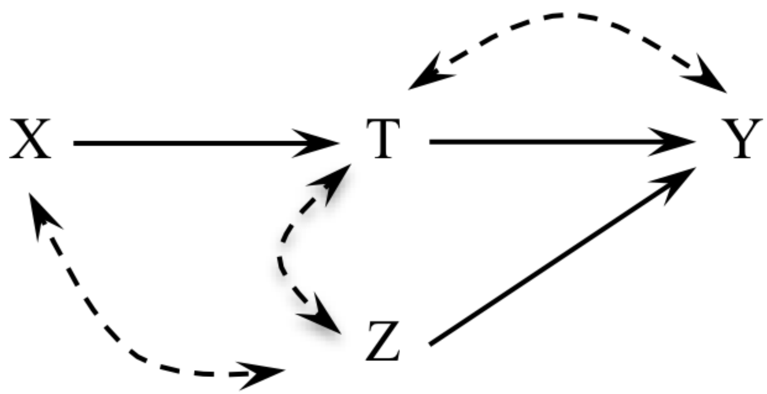


Note: Bi-headed arrows here represent confounding by unobserved variables.

(Pearl, 2009, p.90).

Example 15

Can you test the Null of no effect of T on Y?



Bi-headed arrows represent confounding by unobserved variables.

Tip: First draw the DAG representing the Null. Then check if it implies an independence that would vanish if the Null is false.

(This tricky-yet-beautiful example is due to Brito [2010].)