Judea Pearl and DAGs

- Judea Pearl and colleagues in Artificial Intelligence at UCLA developed DAG modeling to create a formalized causal inference methodology
- They causality concepts extremely clear, they provide a map to the estimation strategy, and maybe best of all, they communicate to others what must be true about the data generating process to recover the causal effect

Judea Pearl, 2011 Turing Award winner, drinking his first IPA

Further reading

1. Pearl (2018) The Book of Why: The New Science of Cause and Effect, Basic Books (*popular*)
2. Morgan and Winship (2014) Counterfactuals and Causal Inference: Methods and Principles for Social Research, Cambridge University Press, 2nd edition (*excellent*)
3. Pearl, Glymour and Jewell (2016) Causal Inference In Statistics: A Primer, Wiley Books (*accessible*)
4. Pearl (2009) Causality: Models, Reasoning and Inference, Cambridge, 2nd edition (*difficult*)
5. Cunningham (2021) Causal Inference: The Mixtape, Yale, 1st edition (*best choice, no question*)

Causal model

- The causal model is sometimes called the structural model, but for us, I prefer the former as it's less alienating
- It's the system of equations describing the relevant aspects of the world
- It necessarily is filled with causal effects associated with some particular comparative statics
- To illustrate, I will assume a Beckerian human capital model

Human capital model: statements *and* graphs

Let's describe my simplified Beckerian human capital model.

- Individuals maximize utility by choosing consumption and schooling (D) subject to multi-period budget constraint
- Education has current costs but longterm returns
- But people choose different levels of schooling based on a number of things we will call "background" (B) which won't be in the dataset ("unobserved")
- And own-schooling will also be because of parental schooling (PE)
- Finally, wages (Y) are a function of parental schooling

Becker's human capital causal model

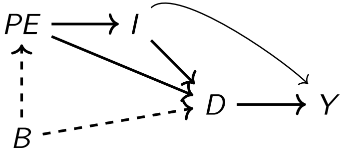
We can represent that causal model visually

$$
\begin{tikzcd}
PE \arrow[r] \arrow[drr] & I \arrow[dr] \arrow[drr, bend left] & & \\
 & & D \arrow[r] & Y \\
B \arrow[uu, dashed] \arrow[urr, dashed] & & &
\end{tikzcd}
$$

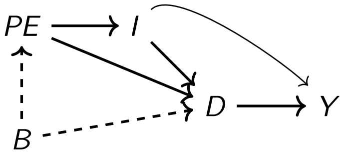
PE is parental education, *B* is "unobserved background factors (i.e., "ability")", *I* is family income, *D* is college education and *Y* is log wages. The DAG is an approximation of Becker's underlying (causal) human capital model.

Arrows, but also *missing arrows*

Before we dive into all this notation, couple of things

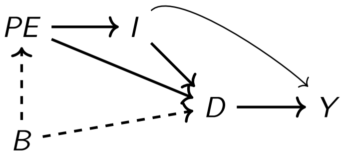


PE and *D* are caused by *B*. But why doesn't *B* cause *Y*?? Do you believe this? Why/why not? We can dispute this, but notice – we can see the assumption, which is transparent and communicates the author's beliefs, as well as the needed assumptions in their forthcoming *empirical* model. Every empirical strategy makes assumptions, but oftentimes they are not as transparent to us as this is.



- B is a **parent** of PE and D
- PE and D are **descendants** of B
- There is a **direct (causal) path** from D to Y
- There is a **mediated (causal) path** from B to Y through D
- There are four **paths** from PE to Y but none are direct, and one is unlike the others

Colliders



Notice anything different with this DAG? Look closely.

- D is a **collider** along the path $B \rightarrow D \leftarrow I$ (i.e., "colliding" at D)
- D is a **noncollider** along the path $B \rightarrow D \rightarrow Y$

Summarizing Value of DAGs imo

1. Facilitates the task of designing identification strategy for estimating average causal effects
2. Facilitates the task of testing compatibility of the model with your data
3. Visualizes the identifying assumptions which opens up the model to critical scrutiny

Creating DAGs

- The DAG is a *relevant* causal relationships describing the relationship between D and Y
- It will include:
  - All direct causal effects among the *relevant* variables in the graph
  - All common causes of any pair of *relevant* variables in the graph
- No need to model a dinosaur stepping on a bug causing in a million years some evolved created that impacted your decision to go to college
- We get ideas for DAGs from theory, models, observation, experience, prior studies, intuition
- Sometimes called the data generating process.

Confounding

- Omitted variable bias has a name in DAGs: "confounding"
- Confounding occurs when when the treatment and the outcomes have a common cause or parent which creates spurious correlation between D and Y

$$\begin{tikzcd} D \arrow[rr] & & Y \\ & X \arrow[ul] \arrow[ur] & \end{tikzcd}$$

The *correlation* between D and Y no longer reflects the causal effect of D on Y

Backdoor Paths

- Confounding creates **backdoor paths** between treatment and outcome ($D \leftarrow X \rightarrow Y$) – i.e., spurious correlations
- Not the same as mediation ($D \rightarrow X \rightarrow Y$)
- We can "block" backdoor paths by conditioning on the common cause X
- Once we condition on X, the correlation between D and Y estimates the causal effect of D on Y
- Conditioning means calculating $E[Y|D = 1, X] - E[Y|D = 0, X]$ for each value of X then combining (e.g., integrating)

$$\begin{tikzcd} D \arrow[rr] & & Y \\ & \boxed{X} \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Blocked backdoor paths

A backdoor path is blocked if and only if:

- It contains a noncollider that has been conditioned on
- Or it contains a collider that has not been conditioned on

Examples of blocked paths

Examples:

1. Conditioning on a noncollider blocks a path:

$$\begin{tikzcd} \boxed{X} \arrow[r] \arrow[rr, bend left] & Z & Y \end{tikzcd}$$

2. Conditioning on a collider opens a path (i.e., creates spurious correlations):

$$\begin{tikzcd} Z \arrow[r] & \boxed{X} & Y \arrow[l] \end{tikzcd}$$

3. *Not* conditioning on a collider blocks a path:

$$\begin{tikzcd} Z \arrow[r] & X & Y \arrow[l] \end{tikzcd}$$

Backdoor criterion

Backdoor criterion

Conditioning on X satisfies the backdoor criterion with respect to (D, Y) directed path if:

1. All backdoor paths are blocked by X
2. No element of X is a collider

In words: If X satisfies the backdoor criterion with respect to (D, Y), then controlling for or matching on X identifies the causal effect of D on Y

What control strategy meets the backdoor criterion?

- List all backdoor paths from D to Y. I'll wait.

$$
\begin{tikzcd}
X_1 \arrow[r] \arrow[rrr, bend left] & D \arrow[rr] & & Y \\
& & X_2 \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

- What are the necessary and sufficient set of controls which will satisfy the backdoor criterion?

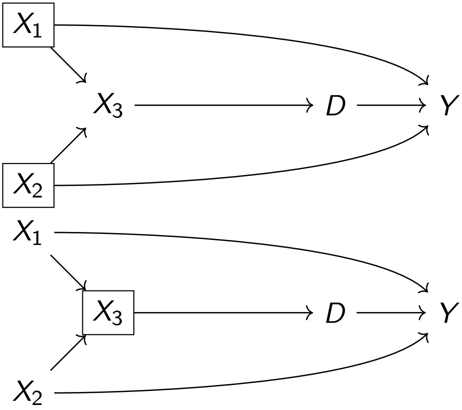
What if you have an unobservable?

- List all the backdoor paths from D to Y.

$$
\begin{tikzcd}
 & X_1 \arrow[dr] \arrow[drr] & & \\
U \arrow[r, dashed] \arrow[rrr, dashed, bend right] & X_2 \arrow[r] & D \arrow[r] & Y
\end{tikzcd}
$$

- What are the necessary and sufficient set of controls which will satisfy the backdoor criterion?
- What about the unobserved variable, U?

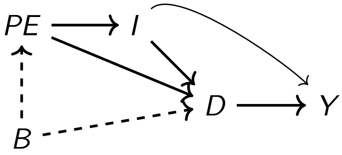
Multiple strategies

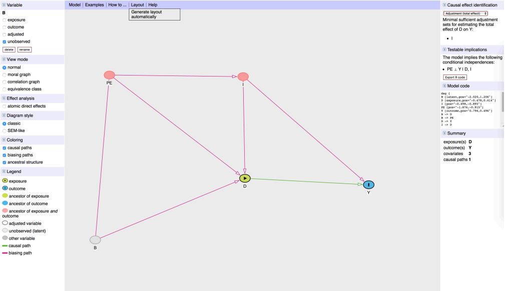


- Conditioning on the common causes, X_1 and X_2, is sufficient
- ... but so is conditioning on X_3

Testing the Validity of the DAG

- The DAG makes testable predictions
- Conditional on D and I, parental education (PE) should no longer be correlated with Y
- Can be hard to figure this out by hand, but software can help (e.g., Daggity.net is browser based)




Variable
B
exposure
outcome
adjusted
unobserved
delete
rename
View mode
normal
moral graph
correlation graph
equivalence class
Effect analysis
atomic direct effects
Diagram style
classic
SEM-like
Coloring
causal paths
biasing paths
ancestral structure
Legend
exposure
outcome
ancestor of exposure
ancestor of outcome
ancestor of exposure *and* outcome
adjusted variable
unobserved (latent)
other variable
causal path
biasing path
Model | Examples | How to ... | Layout | Help
Generate layout automatically
PE
I
D
Y
B
Causal effect identification
Adjustment (total effect)
Minimal sufficient adjustment sets for estimating the total effect of D on Y:
• I
Testable implications
The model implies the following conditional independences:
• PE ⊥ Y | D, I
Export R code
Model code
dag {
B [latent,pos="-2.020,1.206"]
D [exposure,pos="-0.478,0.414"]
I [pos="-0.498,-0.895"]
PE [pos="-1.874,-0.915"]
Y [outcome,pos="0.796,0.496"]
B -> D
B -> PE
D -> Y
I -> D
Summary
exposure(s) D
outcome(s) Y
covariates 3
causal paths 1

Collider bias

- **Conditioning on a collider introduces spurious correlations; can even mask causal directions**
  - There is only one backdoor path from D to Y

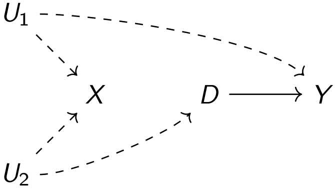
$$
\begin{tikzcd}
\boxed{X_1} \arrow[r] \arrow[rrr, bend left] & D \arrow[rr] \arrow[dr] & & Y \arrow[dl] \\
& & X_2 &
\end{tikzcd}
$$

  - Conditioning on X_1 blocks the backdoor path
  - But what if we also condition on X_2?

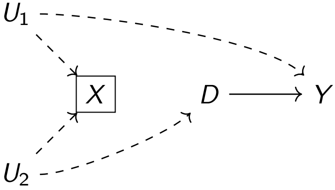
$$
\begin{tikzcd}
\boxed{X_1} \arrow[r] \arrow[rrr, bend left] & D \arrow[rr] \arrow[dr] & & Y \arrow[dl] \\
& & \boxed{X_2} &
\end{tikzcd}
$$

  - Conditioning on X_2 opens up a new path, creating new spurious correlations between D and Y

- **Even controlling for pretreatment covariates can create bias**
  - Name the backdoor paths. Is it open or closed?



  - But what if we condition on X?



Living in reality - he doesn't love you

- **Fact #1**: We can't know if we have a collider bias (confounder) problem without making assumptions about the causal model (i.e. not in the codebook)
- **Fact # 2**: You can't just haphazardly throw in a bunch of controls on the RHS (i.e., "the kitchen sink") bc you may inadvertently be conditioning on a collider which can lead to massive biases
- **Fact # 3**: You have no choice but to leverage economic theory, intuition, intimate familiarity with institutional details and background knowledge for research designs.
- **Fact #4**: You can only estimate causal effects with **data** and **assumptions**.

Examples of collider bias

Bad controls

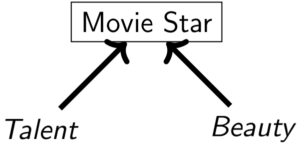
- Angrist and Pischke in MHE talk about a specific type of danger associated with controlling for an outcome – "bad controls"
- The problem is not controlling for an outcome;
- The problem is controlling for a collider and don't correct for *that*
- This has implications for when you work with non-random administrative data, too

Sample selection example of collider bias

Important: Since unconditioned colliders block back-door paths, what exactly does conditioning on a collider do? Let's illustrate with a fun example and some made-up data

- CNN.com headline: Megan Fox voted worst – but sexiest – actress of 2009 (link)
- Are these two things actually negatively correlated in the world?
- Assume talent and beauty are independent, but each causes someone to become a movie star. What's the correlation between talent and beauty for a sample of movie stars compared to the population as a whole (stars and non-stars)?

- What if the sample consists *only* of movie stars?



Stata code

```
clear all
set seed 3444

* 2500 independent draws from standard normal distribution
set obs 2500
generate beauty=rnormal()
generate talent=rnormal()

* Creating the collider variable (star)
gen score=(beauty+talent)
egen c85=pctile(score), p(85)
gen star=(score>=c85)
label variable star "Movie star"

* Conditioning on the top 15%
twoway (scatter beauty talent, mcolor(black) msize(small) msymbol(smx)),
ytitle(Beauty) xtitle(Talent) subtitle(Aspiring actors and actresses)
by(star, total)
```

Figure: Top left figure: Non-star sample scatter plot of beauty (vertical axis) and talent (horizontal axis). Top right right figure: Star sample scatter plot of beauty and talent. Bottom left figure: Entire (stars and non-stars combined) sample scatter plot of beauty and talent.

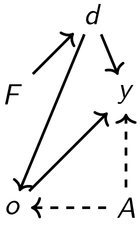
Stata

- Run Stata file star.do

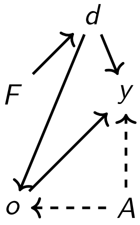
Occupational sorting and discrimination example of collider bias

- Let's look at another example: very common for think tanks and journalists to say that the gender gap in earnings disappears once you control for occupation.
- But what if occupation is a collider, which it could be in a model with occupational sorting
- Then controlling for occupation in a wage regression searching for discrimination can lead to all kinds of crazy results *even in a simulation where we explicitly design there to be discrimination*

DAG



F is female, d is discrimination, o is occupation, y is earnings and A is ability. Dashed lines mean the variable cannot be observed. Note, by design, being a female has no effect on earnings or occupation, and has no relationship with ability. So earnings is coming through discrimination, occupation, and ability.



Mediation and Backdoor paths

1. $d \rightarrow o \rightarrow y$
2. $d \rightarrow o \leftarrow A \rightarrow y$

Stata model (Erin Hengel)

- Erin Hengel (`www.erinhengel.com`) and I worked out this code and she gave me permission to put in my Mixtape
- Let's look at collider_discrimination.do or collider_discrimination.R together

Table: Regressions illustrating collider bias with simulated gender disparity

Covariates:	**Unbiased combined effect**	**Biased**	**Unbiased wage effect only**
Female	-3.074***	0.601***	-0.994***
	(0.000)	(0.000)	(0.000)
Occupation		1.793***	0.991***
		(0.000)	(0.000)
Ability			2.017***
			(0.000)
N	10,000	10,000	10,000
Mean of dependent variable	0.45	0.45	0.45

- Recall we designed there to be a discrimination coefficient of -1
- If we do not control for occupation, then we get the combined effect of $d \rightarrow o \rightarrow y$ and $d \rightarrow y$
- Because it seems intuitive to control for occupation, notice column 2 - the sign flips!
- We are only able to isolate the direct causal effect by conditioning on ability and occupation, but ability is unobserved

Administrative data

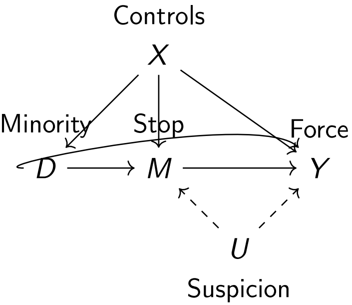
- Admin data has become extremely common, if not absolutely necessary
- But naive use of admin data can be dangerous if the drawing of the sample is itself a collider problem (Heckman 1979; Elwert and Winship 2014)
- Let's look at a new paper by Fryer (2019) and a critique by Knox, et al. (2019)

Collider bias and police use of force

- Claims of excessive and discriminator use of police force against minorities (e.g., Black Lives Matter, Trayvon Martin, Michael Brown, Eric Garner)
- Challenging to identify
  - Police-citizen interactions are conditional on interactions having already been triggered
  - That initial interaction is unobserved
- Fryer (2019) is a monumental study for its data collection and analysis: Stop and Frisk, Police-Public Contact Survey, and admin data from two jurisdictions
- Codes up almost 300 variables from arrest narratives which range from 2-100 pages in length – shoeleather!

Initial interaction

- Fryer finds that blacks and Hispanics were more than 50% more likely to have an interaction with the policy in NYC Stop and Frisk as well as Police-Public Contact survey
- It survives extensive controls – magnitudes fall, but still very large (21%)
- Moves to admin data
- Conditional on police interaction, *no* racial differences in officer-related shootings
- Fryer calls it one of the most surprising findings in his career
- Lots of eyes on this study as a result of the counter intuitive results; published in JPE
- Knox, et al (202) claim his data is itself a collider. What?



Fryer told us $D \rightarrow M$ exists from both Stop and Frisk and Police-Public. But note: admin data is instances of M stops, which is itself a collider. If this DAG is true, then spurious correlations enter between M and Y which may dilute our ability to estimate causal effects.

Knox, et al (2020)

- Move from DAG to more contemporary potential outcomes notation to design relevant parameters
- Use potential outcomes and bounds
- Even with lower bound estimates of the incidence of police violence against civilians is more than 5x higher than what Fryer (2019) finds
- Heckman (1979) – we *cannot* afford to ignore sample selection

Summarizing all of this

- Your dataset will not come with a codebook flagging some variables as "confounders" and other variables as "colliders" because those terms are always context specific
- Except for some unique situations that aren't generally applicable, you also don't always know statistically you have an omitted variable bias problem; but both of these are fatal for any application
- You only know to do what you're doing based on *knowledge about data generating process*.
- All identification must be guided by theory, experience, observation, common sense and knowledge of institutions
- DAGs absorb that information and can be then used to write out the explicit identifying model

DAGs are not panacea

- DAGs cannot handle, though, reverse causality or simultaneity
- So there are limitations. "All models are wrong but some are useful"
- They are also not popular (see Twitter ongoing debates which have descended into light hearted jokes as well as aggressive debates)
- But I think they are helpful and while not *necessary*, showcase what is necessary – assumptions
- Heckman (1979) can maybe provide some justification at times