Quant 2, Lab 3

DAGs, Sensitivity Analysis

Sylvan Zheng

2025-02-13

DAGs

- **DAG** stands for Directed Acyclic Graph

DAGs

- **DAG** stands for Directed Acyclic Graph
  - *Directed*: No reverse causality or simultaneity;

DAGs

- **DAG** stands for Directed Acyclic Graph
  - *Directed*: No reverse causality or simultaneity;
  - *Acyclic*: No cycles

DAGs

- **DAG** stands for Directed Acyclic Graph
  - *Directed*: No reverse causality or simultaneity;
  - *Acyclic*: No cycles
- Representation of the data generating process (DGP)

DAGs

- **DAG** stands for Directed Acyclic Graph
  - *Directed*: No reverse causality or simultaneity;
  - *Acyclic*: No cycles
- Representation of the data generating process (DGP)
  - Nodes (X,D,Y etc.) are random variables

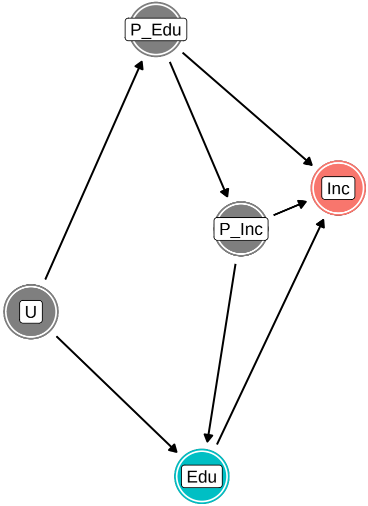
DAGs

- **DAG** stands for Directed Acyclic Graph
  - *Directed*: No reverse causality or simultaneity;
  - *Acyclic*: No cycles
- Representation of the data generating process (DGP)
  - Nodes (X,D,Y etc.) are random variables
  - Edges ($X \rightarrow Y$) denote a direct causal effect of X on Y

DAGs

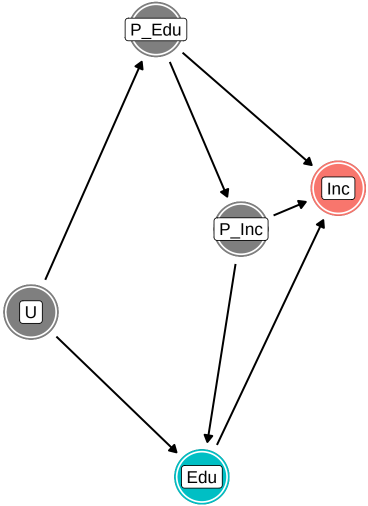
- **DAG** stands for Directed Acyclic Graph
  - *Directed*: No reverse causality or simultaneity;
  - *Acyclic*: No cycles
- Representation of the data generating process (DGP)
  - Nodes (X,D,Y etc.) are random variables
  - Edges ($X \rightarrow Y$) denote a direct causal effect of X on Y
- Tools to help understand whether a research design can identify a causal relationship

Example: Becker, 1994



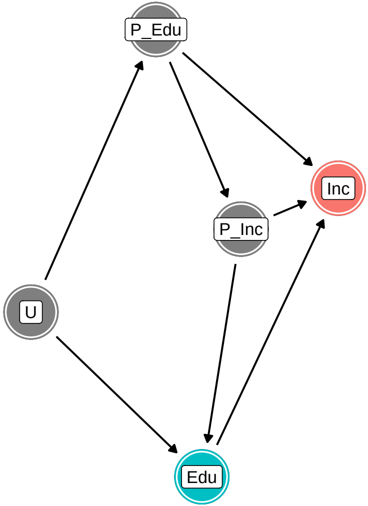
- Main relationship of interest: Education effect on Income

Example: Becker, 1994



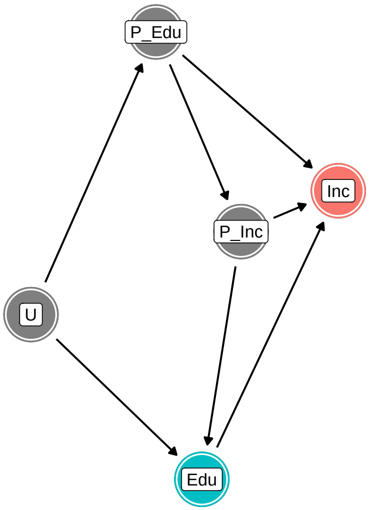
- Main relationship of interest: Education effect on Income
- Parental effects (income, education) affect both child income and education

Example: Becker, 1994



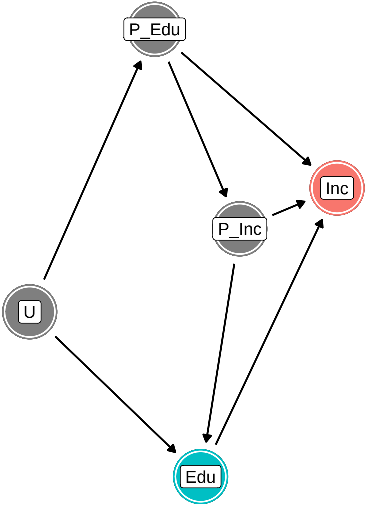
- Main relationship of interest: Education effect on Income
- Parental effects (income, education) affect both child income and education
- Unobserved **family specific** factors (ie, genetics) affect parent and child education

DAGs and Identification



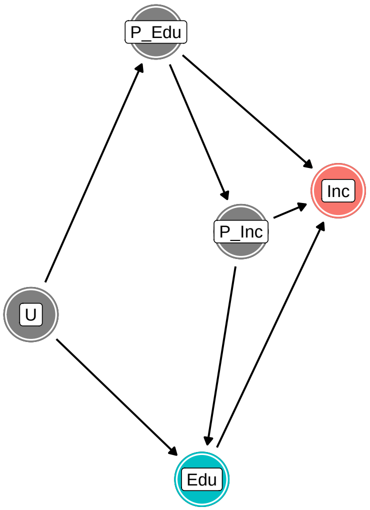
- To identify the effect of some D on Y

DAGs and Identification



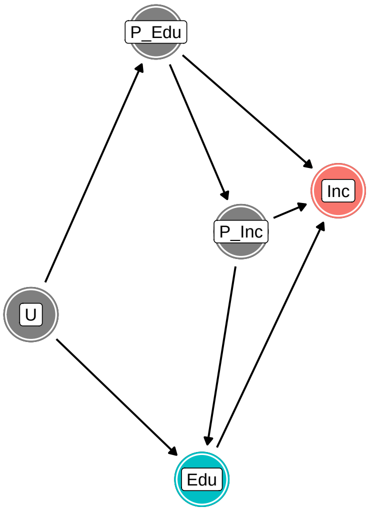
- To identify the effect of some D on Y
- DAG **must** satisfy the **backdoor criterion** (no backdoor paths)

DAGs and Identification



- To identify the effect of some D on Y
- DAG **must** satisfy the **backdoor criterion** (no backdoor paths)
  - A **backdoor path** is an alternate path between D and Y that does not go through a collider (more on these later)

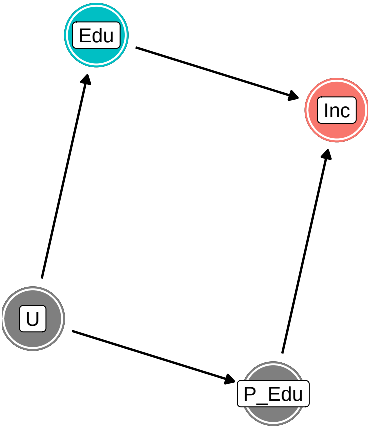
DAGs and Identification

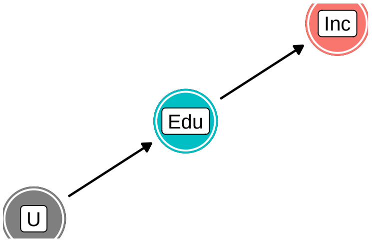


- To identify the effect of some D on Y
- DAG **must** satisfy the **backdoor criterion** (no backdoor paths)
  - A **backdoor path** is an alternate path between D and Y that does not go through a collider (more on these later)
- Eg, we cannot identify the effect of `Edu` on `Inc` because there is a **backdoor path**, eg through `P_Inc`

Controlling for a variable

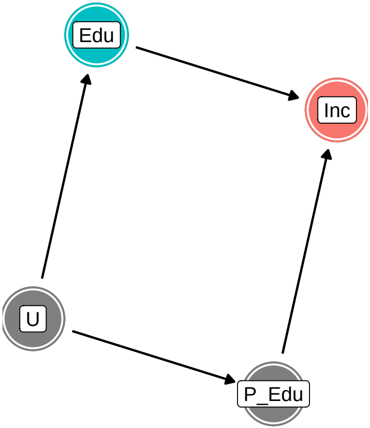
- If we *control for a variable* in a DAG, we **remove its node and corresponding edges**

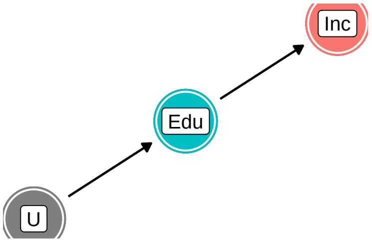




Controlling for a variable

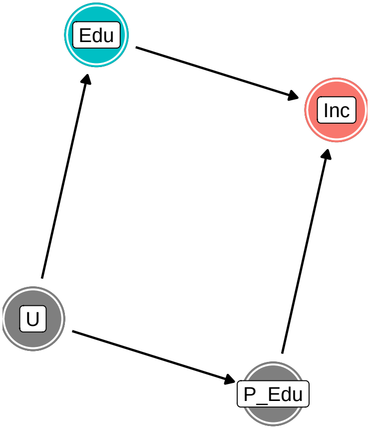
- If we *control for a variable* in a DAG, we **remove its node and corresponding edges**
  - Unless it's a collider

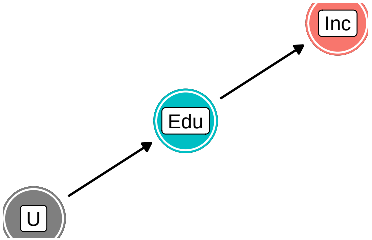




Controlling for a variable

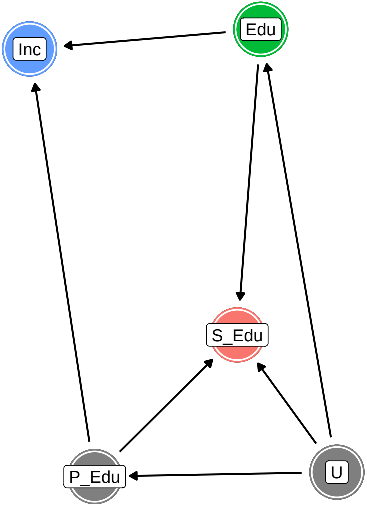
- If we *control for a variable* in a DAG, we **remove its node and corresponding edges**
  - Unless it's a collider
- Ex, if we control for `P_Inc` and `P_Edu`, we get the following DAGs:





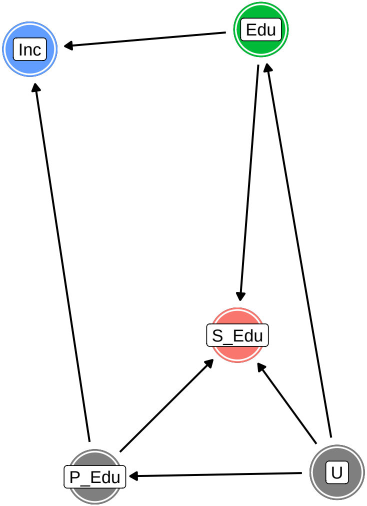
Colliders

- A **collider** is a node that has multiple arrows leading into it



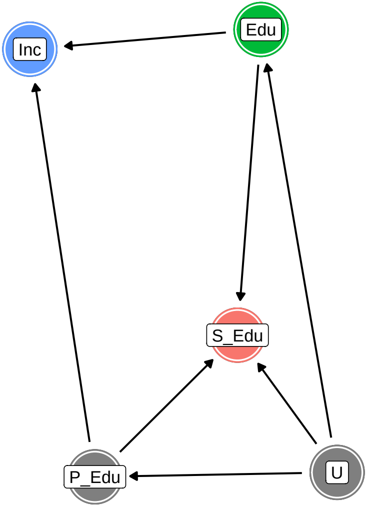
Colliders

- A **collider** is a node that has multiple arrows leading into it
- Consider the following DAG that includes a sibling's education `S_Edu`.



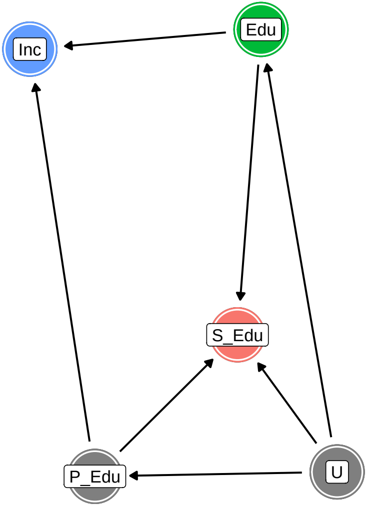
Colliders

- A **collider** is a node that has multiple arrows leading into it
- Consider the following DAG that includes a sibling's education `S_Edu`.
- Suppose we are interested in understanding the relationship between sibling education (`Edu -> S_Edu`)



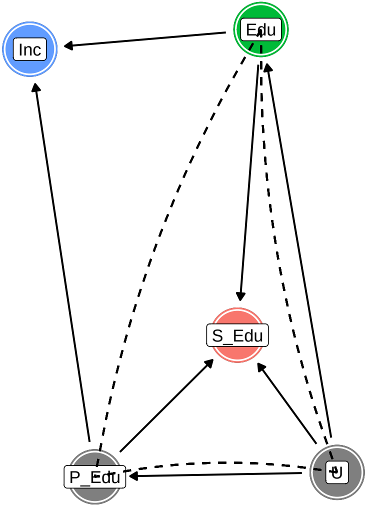
Colliders

- A **collider** is a node that has multiple arrows leading into it
- Consider the following DAG that includes a sibling's education `S_Edu`.
- Suppose we are interested in understanding the relationship between sibling education (`Edu -> S_Edu`)
- Should we control for `Inc`?



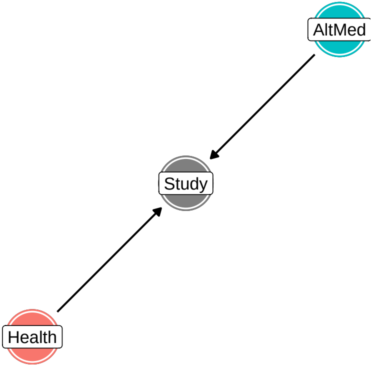
Colliders

- A **collider** is a node that has multiple arrows leading into it
- Consider the following DAG that includes a sibling's education `S_Edu`.
- Suppose we are interested in understanding the relationship between sibling education (`Edu -> S_Edu`)
- Should we control for `Inc`?
  - No. Because `Inc` is a collider, the backdoor path is closed.



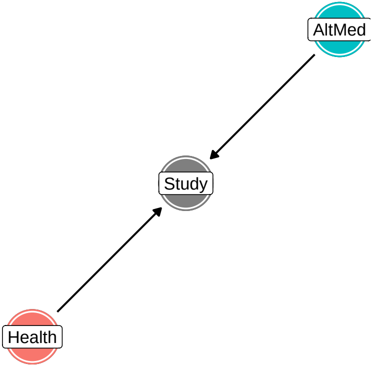
Colliders | Sample Selection

- Collider bias often discussed in the context of sample selection



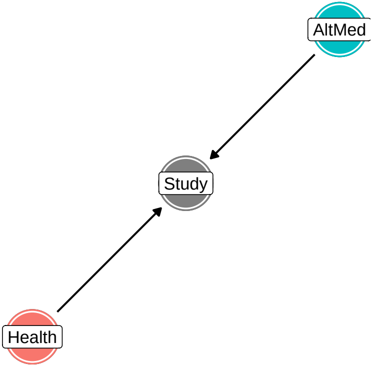
Colliders | Sample Selection

- Collider bias often discussed in the context of sample selection
- Super simple example



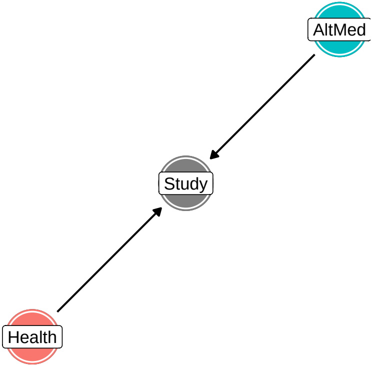
Colliders | Sample Selection

- Collider bias often discussed in the context of sample selection
- Super simple example
- Suppose alternative medicine has no true effect on health outcomes



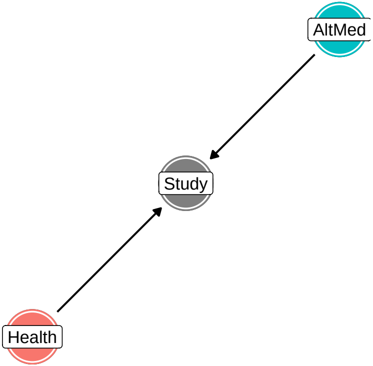
Colliders | Sample Selection

- Collider bias often discussed in the context of sample selection
- Super simple example
- Suppose alternative medicine has no true effect on health outcomes
- But, the study selects (conditions) on experience with alternative medicine



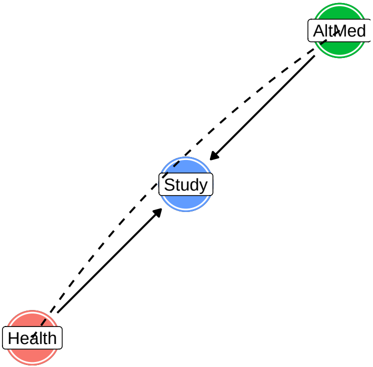
Colliders | Sample Selection

- Collider bias often discussed in the context of sample selection
- Super simple example
- Suppose alternative medicine has no true effect on health outcomes
- But, the study selects (conditions) on experience with alternative medicine
- If health has any effect on study participation, conditioning on the study induces a relationship



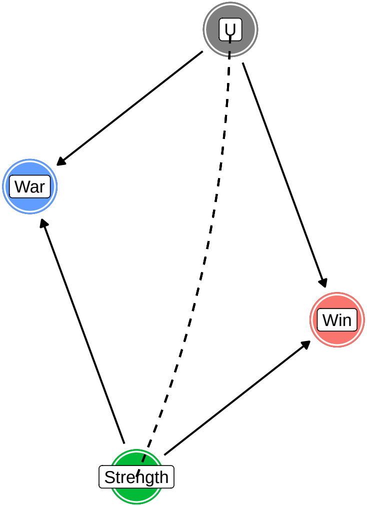
Colliders | Sample Selection

- Collider bias often discussed in the context of sample selection
- Super simple example
- Suppose alternative medicine has no true effect on health outcomes
- But, the study selects (conditions) on experience with alternative medicine
- If health has any effect on study participation, conditioning on the study induces a relationship



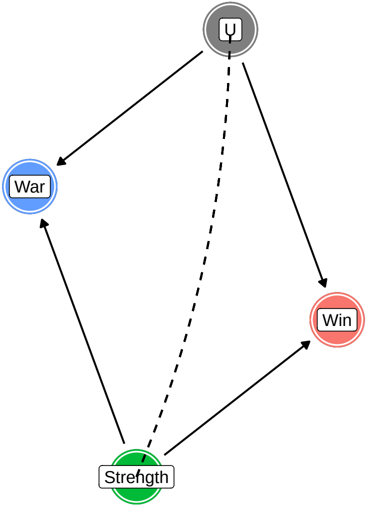
Colliders | Sample Selection

- Another example



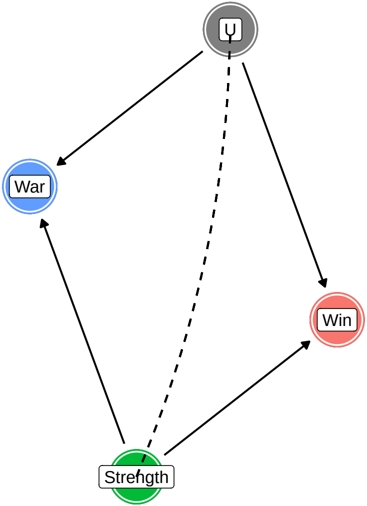
Colliders | Sample Selection

- Another example
- Country military strength appears to be uncorrelated with winning a war



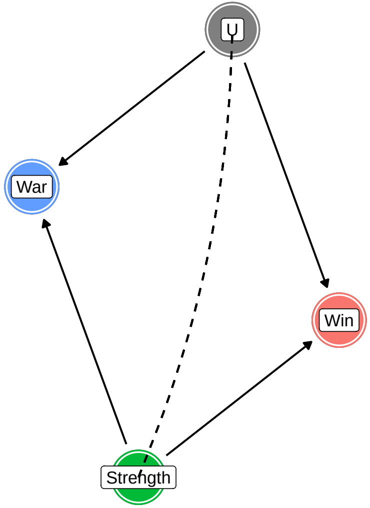
Colliders | Sample Selection

- Another example
- Country military strength appears to be uncorrelated with winning a war
  - But, unobserved factors `U` also affect whether countries get into wars in the first place and whether they win

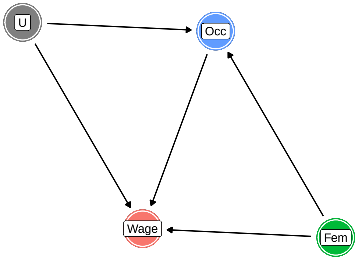


Colliders | Sample Selection

- Another example
- Country military strength appears to be uncorrelated with winning a war
  - But, unobserved factors `U` also affect whether countries get into wars in the first place and whether they win
  - Conditioning on `War` opens a backdoor path through `U`

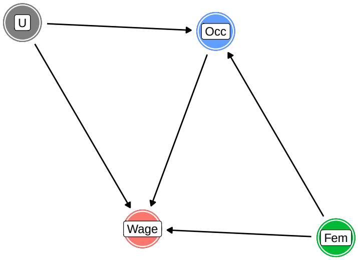


Colliders | Gender Wage Gap



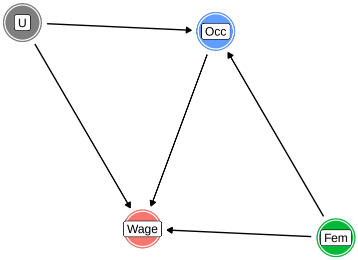
- Should we control for Occupation (`Occ`)?

Colliders | Gender Wage Gap



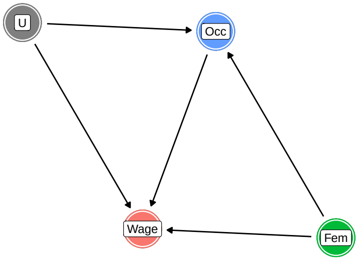
- Should we control for Occupation (`Occ`)?
- Let's use simulation to illustrate (`gender_wage_sim.R`)

Colliders | Gender Wage Gap



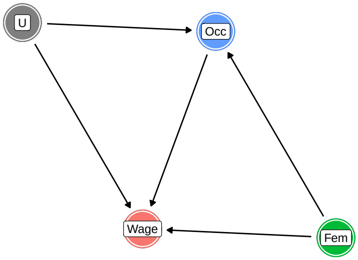
- Should we control for Occupation (`Occ`)?
- Let's use simulation to illustrate (`gender_wage_sim.R`)
- Suppose the following DGP

Colliders | Gender Wage Gap



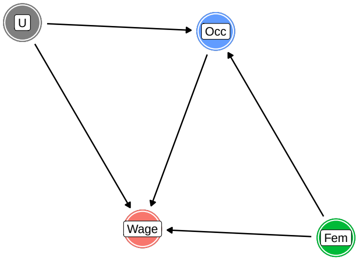
- Should we control for Occupation (`Occ`)?
- Let's use simulation to illustrate (`gender_wage_sim.R`)
- Suppose the following DGP
  - `Occ` = -0.1 * `Fem` + `u` + ϵ_1

Colliders | Gender Wage Gap



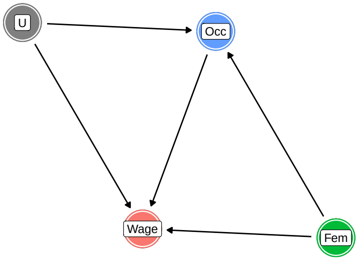
- Should we control for Occupation (`Occ`)?
- Let's use simulation to illustrate (`gender_wage_sim.R`)
- Suppose the following DGP
  - `Occ` = -0.1 * `Fem` + u + ϵ_1
  - `Wage` = -0.1 * `Fem` + `Occ` + 2 * u + ϵ_2

Colliders | Gender Wage Gap



- Should we control for Occupation (`Occ`)?
- Let's use simulation to illustrate (`gender_wage_sim.R`)
- Suppose the following DGP
  - `Occ` = -0.1 * `Fem` + u + ϵ_1
  - `Wage` = -0.1 * `Fem` + `Occ` + 2 * u + ϵ_2
  - `Fem` $\sim$ Bernoulli(0.5)

Colliders | Gender Wage Gap



- Should we control for Occupation (`Occ`)?
- Let's use simulation to illustrate (`gender_wage_sim.R`)
- Suppose the following DGP
    - `Occ` = -0.1 * `Fem` + u + ϵ_1
    - `Wage` = -0.1 * `Fem` + `Occ` + 2 * u + ϵ_2
    - `Fem` $\sim$ Bernoulli(0.5)
    - $\epsilon_1, \epsilon_2, u \sim N(0, 1)$

Colliders | Gender Wage Gap

Simulation Setup

```
N <- 10000
tb <- tibble(
    # Gender is exogenous
    female = sample(c(0, 1), N, replace = T),
    # U is exogenous
    u = rnorm(N),
    # Occupation choice a function of u and gender
    occupation = u - 0.1 * female + rnorm(N),
    # Wage is a function of u and occupation
    #  AND very slightly directly affected by gender
    wage = -0.1 * female + occupation + 2 * u + rnorm(N)
)
```

Colliders | Gender Wage Gap

Simulation Results

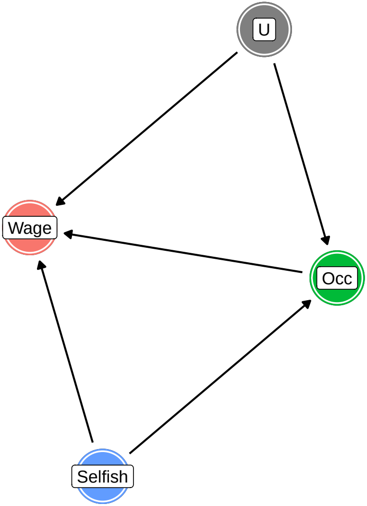
Dependent Variable:		wage	
Model:	(1)	(2)	(3)
Variables			
Constant	-0.0563	-0.0285	-0.0012
	(0.0461)	(0.0244)	(0.0142)
female	-0.1045	0.0197	-0.0986***
	(0.0653)	(0.0346)	(0.0202)
occupation		1.969***	0.9759***
		(0.0123)	(0.0101)
u			2.007***
			(0.0144)
Fit statistics			
Observations	10,000	10,000	10,000

IID standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

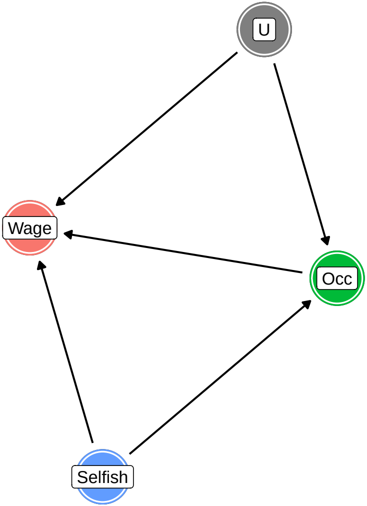
Bias Amplification

- Same DAG as before, but let's say we are now interested in `Occ->Wage`



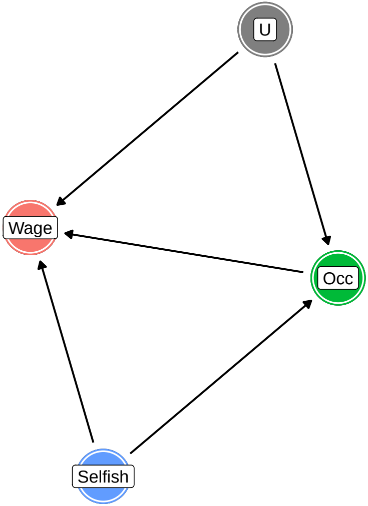
Bias Amplification

- Same DAG as before, but let's say we are now interested in `Occ->Wage`
- Suppose `Selfish` increases `Occ` but decreases `Wages`



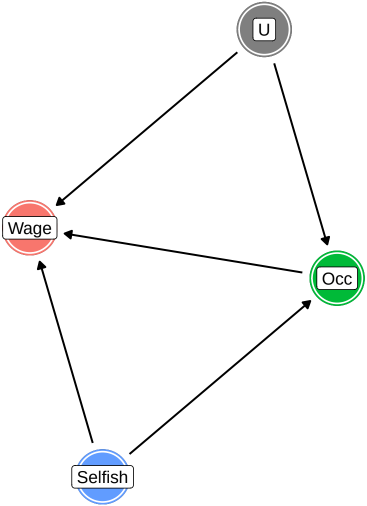
Bias Amplification

- Same DAG as before, but let's say we are now interested in `Occ->Wage`
- Suppose `Selfish` increases `Occ` but decreases `Wages`
- Suppose we observe `Selfish`



Bias Amplification

- Same DAG as before, but let's say we are now interested in `Occ->Wage`
- Suppose `Selfish` increases `Occ` but decreases `Wages`
- Suppose we observe `Selfish`
- Should we control for `Selfish`?



Bias Amplification

Simulation Setup

```r
tb <- tibble(
    # U and Selfish exogenous
    u = rnorm(N),
    selfish = rnorm(N),
    # Selfish positively affects occupation
    occupation = u + selfish + rnorm(N),
    # Selfish negatively affects wages
    wage = occupation + 2 * u - 0.5 * selfish + rnorm(N)
)
```

Bias Amplification

Simulation Results

Dependent Variable:	wage	
Model:	(1)	(2)
Variables		
Constant	-0.0374*	-0.0313*
	(0.0213)	(0.0175)
occupation	1.495***	1.985***
	(0.0123)	(0.0124)
selfish		-1.485***
		(0.0215)
Fit statistics		
Observations	10,000	10,000

IID standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Post-treatment bias

- Homola, Pereira, and Tavits (2020, APSR) argue that living closer to a Nazi era concentration camp increases modern day far right support.

Post-treatment bias

- Homola, Pereira, and Tavits (2020, APSR) argue that living closer to a Nazi era concentration camp increases modern day far right support.
- Pepinsky, Goodman, Ziller (2023, APSR) argue that "state-level differences confound the relationship between distance to camps and out-group intolerance"

Post-treatment bias

- Homola, Pereira, and Tavits (2020, APSR) argue that living closer to a Nazi era concentration camp increases modern day far right support.
- Pepinsky, Goodman, Ziller (2023, APSR) argue that "state-level differences confound the relationship between distance to camps and out-group intolerance"
  - They add state level fixed effects and show that the original effect disappears.

Post-treatment bias

- Homola, Pereira, and Tavits (2020, APSR) argue that living closer to a Nazi era concentration camp increases modern day far right support.
- Pepinsky, Goodman, Ziller (2023, APSR) argue that "state-level differences confound the relationship between distance to camps and out-group intolerance"
  - They add state level fixed effects and show that the original effect disappears.
  - "Länder cannot be posttreatment variables unless we assume that the creation of Länder was caused by their distance from concentration camps."

Post-treatment bias

- Homola, Pereira, and Tavits (2020, APSR) argue that living closer to a Nazi era concentration camp increases modern day far right support.
- Pepinsky, Goodman, Ziller (2023, APSR) argue that "state-level differences confound the relationship between distance to camps and out-group intolerance"
  - They add state level fixed effects and show that the original effect disappears.
  - "Länder cannot be posttreatment variables unless we assume that the creation of Länder was caused by their distance from concentration camps."
- HPT (2024, APSR) rebuttal. "contemporary state fixed effects induce post-treatment bias if any factor (observable or not) that varies across German Länder is a direct or indirect descendant of proximity to concentration camps." "

Conclusions

- Should we give up on observational regression analysis?
  Maybe

Conclusions

- Should we give up on observational regression analysis? Maybe
  - Don't control enough $=>$ Omitted variable bias

Conclusions

- Should we give up on observational regression analysis? Maybe
  - Don't control enough => Omitted variable bias
  - Control on a collider => Collider bias

Conclusions

- Should we give up on observational regression analysis?
  Maybe
  - Don't control enough => Omitted variable bias
  - Control on a collider => Collider bias
  - Control on post treatment => Post treatment bias

Conclusions

- Should we give up on observational regression analysis? Maybe
  - Don't control enough => Omitted variable bias
  - Control on a collider => Collider bias
  - Control on post treatment => Post treatment bias
  - Control on something innocuous => Bias Amplification, sometimes

Conclusions

- Should we give up on observational regression analysis? Maybe
  - Don't control enough => Omitted variable bias
  - Control on a collider => Collider bias
  - Control on post treatment => Post treatment bias
  - Control on something innocuous => Bias Amplification, sometimes
- Always use a DAG and domain knowledge to justify your control strategy

Conclusions

- Should we give up on observational regression analysis? Maybe
  - Don't control enough => Omitted variable bias
  - Control on a collider => Collider bias
  - Control on post treatment => Post treatment bias
  - Control on something innocuous => Bias Amplification, sometimes
- Always use a DAG and domain knowledge to justify your control strategy
- Sensitivity analysis and Double ML can help, kind of

Conclusions

- Should we give up on observational regression analysis? Maybe
  - Don't control enough => Omitted variable bias
  - Control on a collider => Collider bias
  - Control on post treatment => Post treatment bias
  - Control on something innocuous => Bias Amplification, sometimes
- Always use a DAG and domain knowledge to justify your control strategy
- Sensitivity analysis and Double ML can help, kind of
  - (Double ML next week)

Sensitivity analysis

- General idea - quantify **how large** an omitted variable would have to be to mess up your results

Sensitivity analysis

- General idea - quantify **how large** an omitted variable would have to be to mess up your results
  - Roman and D'Urso show a correlation between anti LGBTQ attitudes and dislike for "Latinx" group label, controlling for several factors

Sensitivity analysis

- General idea - quantify **how large** an omitted variable would have to be to mess up your results
  - Roman and D'Urso show a correlation between anti LGBTQ attitudes and dislike for "Latinx" group label, controlling for several factors
  - Sensitivity analysis: Omitted variable would have to have as large an effect on "Latinx" favorability as partisanship

Sensitivity analysis: Attitudes in Darfur (Hazlett, 2019)

- 2003-2004 government violence against civilians

```r
library(sensemakr)
data("darfur")
darfur.model <- feols(
    peacefactor ~ directlyharmed + female +
        age + farmer_dar + herder_dar + pastvoted +
        hhsize_darfur | village,
    data = darfur
)
```

Sensitivity analysis: Attitudes in Darfur (Hazlett, 2019)

- 2003-2004 government violence against civilians
- Outcome (Y): attitudes toward peace

```r
library(sensemakr)
data("darfur")
darfur.model <- feols(
    peacefactor ~ directlyharmed + female +
        age + farmer_dar + herder_dar + pastvoted +
        hhsize_darfur | village,
    data = darfur
)
```

Sensitivity analysis: Attitudes in Darfur (Hazlett, 2019)

- 2003-2004 government violence against civilians
- Outcome (Y): attitudes toward peace
- Treatment (D): exposure to violence

```r
library(sensemakr)
data("darfur")
darfur.model <- feols(
    peacefactor ~ directlyharmed + female +
        age + farmer_dar + herder_dar + pastvoted +
        hhsize_darfur | village,
    data = darfur
)
```

Sensitivity analysis: Attitudes in Darfur (Hazlett, 2019)

- Specification with lots of controls shows a positive relationship

Dependent Variable:	peacefactor
Model:	(1)
Variables	
directlyharmed	0.0973***
	(0.0238)
female	-0.2321***
	(0.0244)
age	-0.0021***
	(0.0007)
farmer_dar	-0.0404
	(0.0296)
herder_dar	0.0143
	(0.0365)
pastvoted	-0.0480*
	(0.0269)
hhsize_darfur	0.0012
	(0.0022)
Fixed-effects	
village	Yes

Sensitivity analysis: Attitudes in Darfur (Hazlett, 2019)

```
darfur.sensitivity <- sensemakr(
    model = darfur.model,
    treatment = "directlyharmed",
    benchmark_covariates = "female",
    kd = 1:3,
    ky = 1:3,
)
```

- `sensemakr` package lets us conduct sensitivity analysis **relative to a covariate of choice**

Sensitivity analysis: Attitudes in Darfur (Hazlett, 2019)

```r
darfur.sensitivity <- sensemakr(
    model = darfur.model,
    treatment = "directlyharmed",
    benchmark_covariates = "female",
    kd = 1:3,
    ky = 1:3,
)
```

- `sensemakr` package lets us conduct sensitivity analysis **relative to a covariate of choice**
- Ex, gender played an important role in exposure to violence: women were specifically targeted

Sensitivity analysis: Attitudes in Darfur (Hazlett, 2019)

```r
darfur.sensitivity <- sensemakr(
    model = darfur.model,
    treatment = "directlyharmed",
    benchmark_covariates = "female",
    kd = 1:3,
    ky = 1:3,
)
```

- `sensemakr` package lets us conduct sensitivity analysis **relative to a covariate of choice**
- Ex, gender played an important role in exposure to violence: women were specifically targeted
- `kd` and `ky` arguments: we investigate a hypothetical confounder **1-3x** as strong as `female`

Sensitivity Analysis

```
ovb_minimal_reporting(darfur.sensitivity)
```

Outcome: *peacefactor*

Treatment:	$R^2_{Y\sim D\vert\mathbf{X}}$	$RV_{q=1}$	$RV_{q=1,\alpha=0.05}$
directlyharmed	2.2%	13.9%	7.6%
Bound (1x female): $R^2_{Y\sim Z\vert\mathbf{X},D} = 12.5\%$, $R^2_{D\sim Z\vert\mathbf{X}} = 0.9\%$			

Sensitivity Analysis

```
summary(darfur.sensitivity)
```

- Partial R2 of the treatment with the outcome: an extreme confounder (orthogonal to the covariates) that explains 100% of the residual variance of the outcome, would need to explain at least 2.19% of the residual variance of the treatment to fully account for the observed estimated effect.

Sensitivity Analysis

```
summary(darfur.sensitivity)
```

- Partial R2 of the treatment with the outcome: an extreme confounder (orthogonal to the covariates) that explains 100% of the residual variance of the outcome, would need to explain at least 2.19% of the residual variance of the treatment to fully account for the observed estimated effect.
- Robustness Value, $q = 1$: unobserved confounders (orthogonal to the covariates) that explain more than 13.88% of the residual variance of both the treatment and the outcome are strong enough to bring the point estimate to 0.

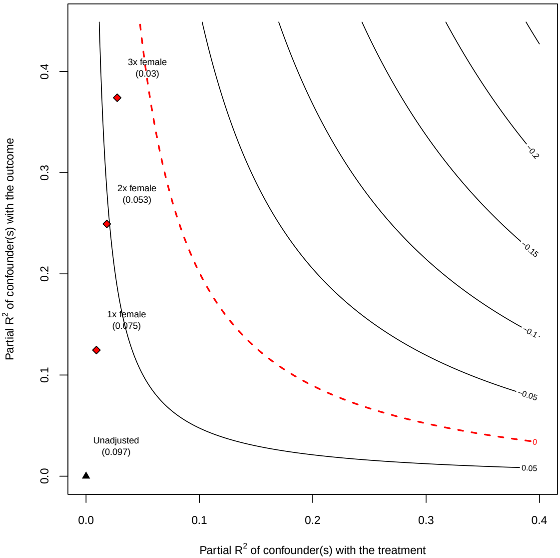
Sensitivity Analysis

```
summary(darfur.sensitivity)
```

- Partial R2 of the treatment with the outcome: an extreme confounder (orthogonal to the covariates) that explains 100% of the residual variance of the outcome, would need to explain at least 2.19% of the residual variance of the treatment to fully account for the observed estimated effect.
- Robustness Value, q = 1: unobserved confounders (orthogonal to the covariates) that explain more than 13.88% of the residual variance of both the treatment and the outcome are strong enough to bring the point estimate to 0.
- Robustness Value, q = 1, alpha = 0.05: unobserved confounders (orthogonal to the covariates) that explain more than 7.63% of the residual variance of both the treatment and the outcome are strong enough to bring the estimate to a range where it is no longer 'statistically different' from 0 at the significance level of alpha = 0.05.

Sensitivity: Plots

```
plot(darfur.sensitivity)
```



HW1

- Grades coming soon

HW1

- Grades coming soon
- Q2.3: If $\bar{Y_N} \to \mu$ and $\bar{S_N} \to 1$ then by Slutsky $\hat{\mu} \to \frac{\mu}{1} = \mu$

HW1

- Grades coming soon
- Q2.3: If $\bar{Y_N} \to \mu$ and $\bar{S_N} \to 1$ then by Slutsky $\hat{\mu} \to \frac{\mu}{1} = \mu$
  - But how do we know that $\bar{Y_N} \to \mu$?

HW1

- Grades coming soon
- Q2.3: If $\bar{Y_N} \to \mu$ and $\bar{S_N} \to 1$ then by Slutsky $\hat{\mu} \to \frac{\mu}{1} = \mu$
  - But how do we know that $\bar{Y_N} \to \mu$?
  - Technically, not sufficient to say that $E(\bar{Y_N}) = \mu$. Unbiased $! =$ consistency!

HW1

- Grades coming soon
- Q2.3: If $\bar{Y_N} \to \mu$ and $\bar{S_N} \to 1$ then by Slutsky $\hat{\mu} \to \frac{\mu}{1} = \mu$
  - But how do we know that $\bar{Y_N} \to \mu$?
  - Technically, not sufficient to say that $E(\bar{Y_N}) = \mu$. Unbiased !$=$ consistency!
  - In Q2.5 we derive $V(\bar{Y_N}) = (\frac{1}{N} - \frac{1}{n})\frac{1}{n}\sum y_i^2$

HW1

- Grades coming soon
- Q2.3: If $\bar{Y_N} \to \mu$ and $\bar{S_N} \to 1$ then by Slutsky $\hat{\mu} \to \frac{\mu}{1} = \mu$
  - But how do we know that $\bar{Y_N} \to \mu$?
  - Technically, not sufficient to say that $E(\bar{Y_N}) = \mu$. Unbiased $!=$ consistency!
  - In Q2.5 we derive $V(\bar{Y_N}) = (\frac{1}{N} - \frac{1}{n})\frac{1}{n}\sum y_i^2$
  - As N,n go to infinity $(\frac{1}{N} - \frac{1}{n})$ approaches 0

HW1

- Grades coming soon
- Q2.3: If $\bar{Y_N} \to \mu$ and $\bar{S_N} \to 1$ then by Slutsky $\hat{\mu} \to \frac{\mu}{1} = \mu$
  - But how do we know that $\bar{Y_N} \to \mu$?
  - Technically, not sufficient to say that $E(\bar{Y_N}) = \mu$. Unbiased != consistency!
  - In Q2.5 we derive $V(\bar{Y_N}) = (\frac{1}{N} - \frac{1}{n})\frac{1}{n}\sum y_i^2$
  - As N,n go to infinity $(\frac{1}{N} - \frac{1}{n})$ approaches 0
  - $\frac{1}{n}\sum y_i^2$ approaches a finite value

HW1

- Grades coming soon
- Q2.3: If $\bar{Y_N} \to \mu$ and $\bar{S_N} \to 1$ then by Slutsky $\hat{\mu} \to \frac{\mu}{1} = \mu$
  - But how do we know that $\bar{Y_N} \to \mu$?
  - Technically, not sufficient to say that $E(\bar{Y_N}) = \mu$. Unbiased != consistency!
  - In Q2.5 we derive $V(\bar{Y_N}) = (\frac{1}{N} - \frac{1}{n})\frac{1}{n}\sum y_i^2$
  - As N,n go to infinity $(\frac{1}{N} - \frac{1}{n})$ approaches 0
  - $\frac{1}{n}\sum y_i^2$ approaches a finite value
  - So $V(\bar{Y_N}) \to 0$ and now we can say unbiased = consistency for $\bar{Y_n}$

HW1 | Q8 (Simulation question)

N	$\text{Var}\left[\overline{Y}_N\right]$	$\text{Var}[\hat{\mu}]$	$\text{Var}\left[\overline{Y}_N\right] - \text{Var}[\hat{\mu}]$	$(\frac{1}{N} - \frac{1}{n})\mu^2$	μ
20	0.2097426	0.07680564	0.132937	0.132937	1.647119
50	0.08132877	0.02978178	0.05154699	0.05154699	1.647119
100	0.03852416	0.01410716	0.024417	0.024417	1.647119

- Table printed with `knitr::kable`. Good! But...

HW1 | Q8 (Simulation question)

N	$\text{Var}\left[\overline{Y}_N\right]$	$\text{Var}[\hat{\mu}]$	$\text{Var}\left[\overline{Y}_N\right] - \text{Var}[\hat{\mu}]$	$(\frac{1}{N} - \frac{1}{n})\mu^2$	μ
20	0.2097426	0.07680564	0.132937	0.132937	1.647119
50	0.08132877	0.02978178	0.05154699	0.05154699	1.647119
100	0.03852416	0.01410716	0.024417	0.024417	1.647119

- Table printed with `knitr::kable`. Good! But...
- The theoretical difference in variances should NOT be identical to the empirical difference in variances

HW1 | Q8 (Simulation question)

N	$\text{Var}\left[\overline{Y}_N\right]$	$\text{Var}[\hat{\mu}]$	$\text{Var}\left[\overline{Y}_N\right] - \text{Var}[\hat{\mu}]$	$(\frac{1}{N} - \frac{1}{n})\mu^2$	μ
20	0.2097426	0.07680564	0.132937	0.132937	1.647119
50	0.08132877	0.02978178	0.05154699	0.05154699	1.647119
100	0.03852416	0.01410716	0.024417	0.024417	1.647119

- Table printed with `knitr::kable`. Good! But...
- The theoretical difference in variances should NOT be identical to the empirical difference in variances
- Please indent your code (can do this automatically with Ctrl/Cmd+I in Rstudio)

HW1 | Q8 (Simulation question)

N	Var $[\overline{Y}_N]$	Var$[\hat{\mu}]$	Var $[\overline{Y}_N]$ − Var$[\hat{\mu}]$	$(\frac{1}{N} - \frac{1}{n})\mu^2$	μ
20	0.2097426	0.07680564	0.132937	0.132937	1.647119
50	0.08132877	0.02978178	0.05154699	0.05154699	1.647119
100	0.03852416	0.01410716	0.024417	0.024417	1.647119

- Table printed with `knitr::kable`. Good! But...
- The theoretical difference in variances should NOT be identical to the empirical difference in variances
- Please indent your code (can do this automatically with Ctrl/Cmd+I in Rstudio)

HW1 | Q8 (Simulation question)

N	$\text{Var}[\overline{Y}_N]$	$\text{Var}[\hat{\mu}]$	$\text{Var}[\overline{Y}_N] - \text{Var}[\hat{\mu}]$	$(\frac{1}{N} - \frac{1}{n})\mu^2$	μ
20	0.2097426	0.07680564	0.132937	0.132937	1.647119
50	0.08132877	0.02978178	0.05154699	0.05154699	1.647119
100	0.03852416	0.01410716	0.024417	0.024417	1.647119

- Table printed with `knitr::kable`. Good! But...
- The theoretical difference in variances should NOT be identical to the empirical difference in variances
- Please indent your code (can do this automatically with Ctrl/Cmd+I in Rstudio)

```
for (N in sample_sizes) {
    do stuff
    for (j in 1:nsims) {
        do more stuff
    }
}
```

```
for (N in sample_sizes) {
do stuff
for (j in 1:nsims) {
do more stuff
}
}
```