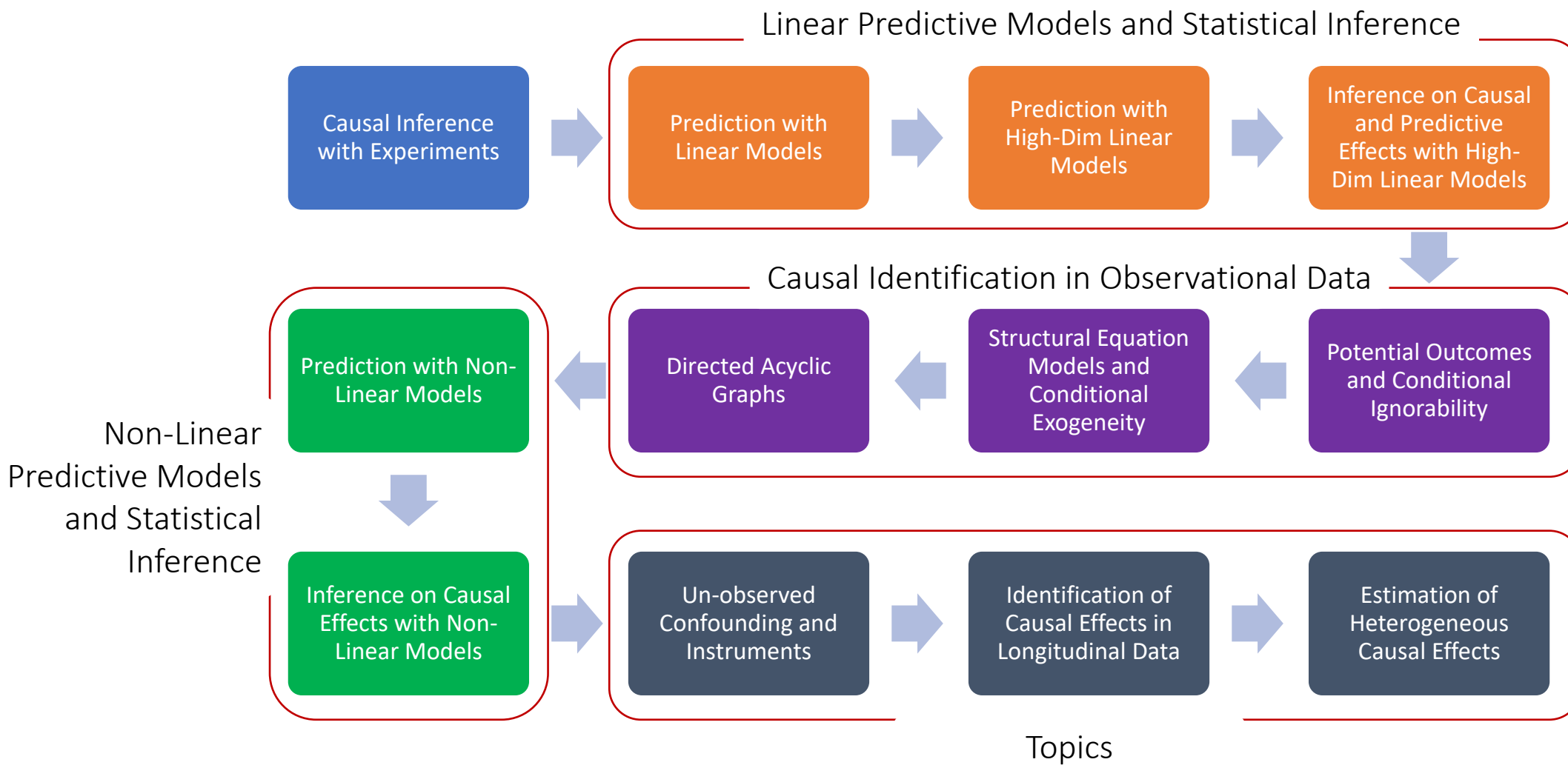
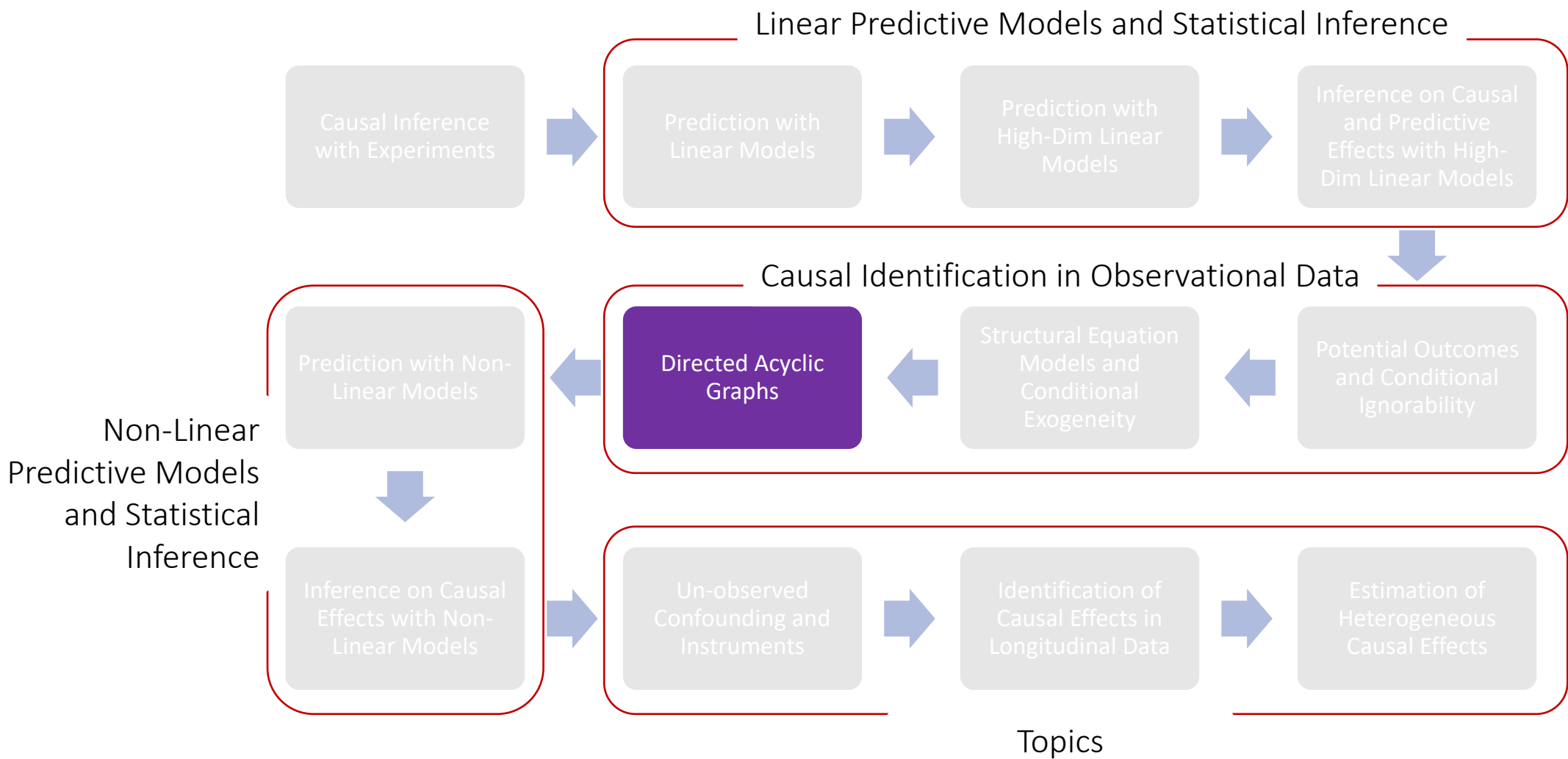


MS&E 228: Directed Acyclic Graphs and Non-Linear SEMs

Vasilis Syrgkanis

MS&E, Stanford





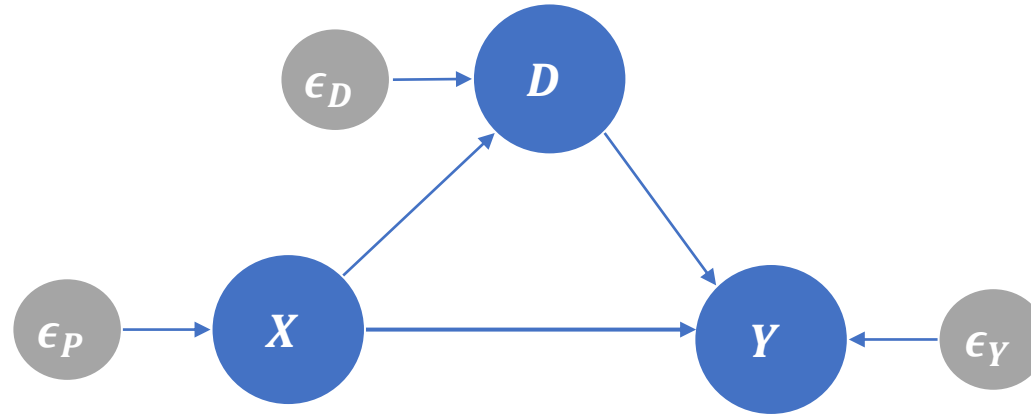
Goals for Today

- Graphical criteria for selection of adjustment set
- Crash course on good and bad “controls”

Recap of Last Lecture



Non-Linear versions of structural equation models are equivalent to Directed Acyclic Graphs



Exogenously determined
“outside” of the model

Endogenously
determined by
the structural
model

For any DAG, we can write ASEM

$$X_j := f_j(\text{Parents}_j, \epsilon_j) = f_j(\text{Pa}_j, \epsilon_j)$$

Shocks ϵ_j are jointly independent and independent of $\{X_j\}$



Corresponding structural response functions

$$X_j(pa_j) := f_j(pa_j, \epsilon_j)$$

Shocks can be multi-dimensional
e.g. separate shock variable per
parental value

Potential/Counterfactual
Outcome Processes

Structural Response
Function

Potential values of
parents

DAGs encode conditions on factorization of probability law

$$p(\{x_\ell\}_{\ell \in V}) = \prod_{\ell \in V} p(x_\ell | pa_\ell)$$



DAGs encode conditional independencies: S d-separates X from Y in DAG G implies $X \perp\!\!\!\perp Y \mid S$

$$(X \perp\!\!\!\perp_d Y \mid S)_G \Rightarrow X \perp\!\!\!\perp Y \mid S$$



Implies testable restrictions we can use to refute DAG from data;
e.g. for linear ASEM, BLP of Y using X, S should have zero on X

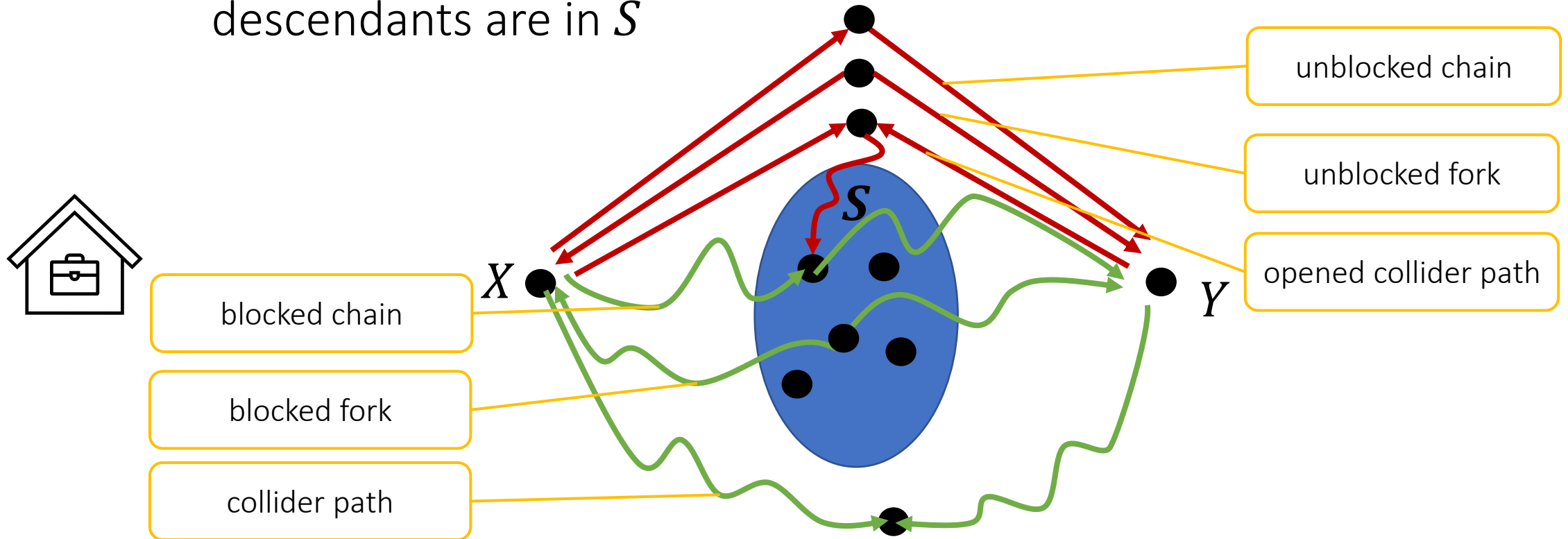
$$Y = \alpha X + \beta' S + \epsilon, \quad \epsilon \perp (X, S)$$

Test whether it
is non-zero!

X is d-separated from Y by S if **every path** from X to Y is **blocked**.

S **blocks a path** if one of the following holds:

- path contains chain $X \rightarrow M \rightarrow Y$ or fork $X \leftarrow M \rightarrow Y$ and $M \in S$
- path contains collider $X \rightarrow M \leftarrow Y$ and neither M nor its descendants are in S



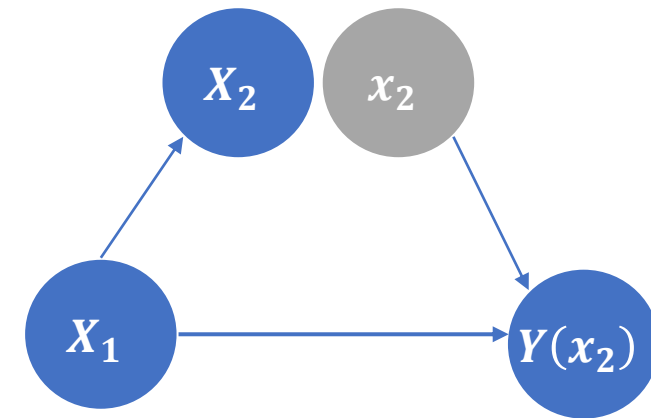
Fix Interventions $\text{fix}(X_j = x_j)$

Locally replace X_j in every RHS of a structural equation with x_j . Leave as-is structural response of X_j . Also measures potential outcome $Y(x_j)$

Fix intervention visually represented as **SWIG** $\tilde{G}(x_j)$. Depicts potential outcome $Y(x_j)$ and original variable X_j on the same graph

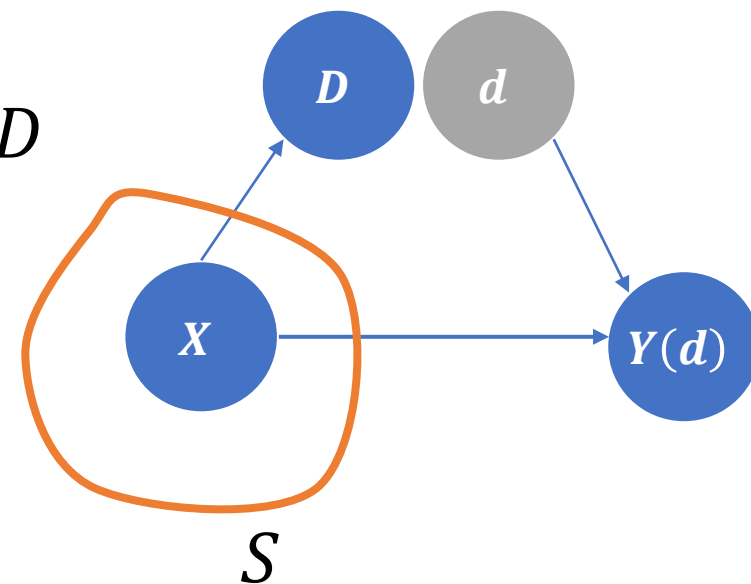


If we can check $Y(x_j) \perp X_j \mid S$ based on the SWIG, we can identify $E[Y(x_j)]$ via conditioning





Conditional ignorability between treatment D and outcome Y conditional on set S holds if $Y(d)$ is d-separated from D on SWIG $\tilde{G}(d)$ induced by $\text{fix}(D = d)$ by the set S



Useful Adjustment Strategies

Adjustment Strategies

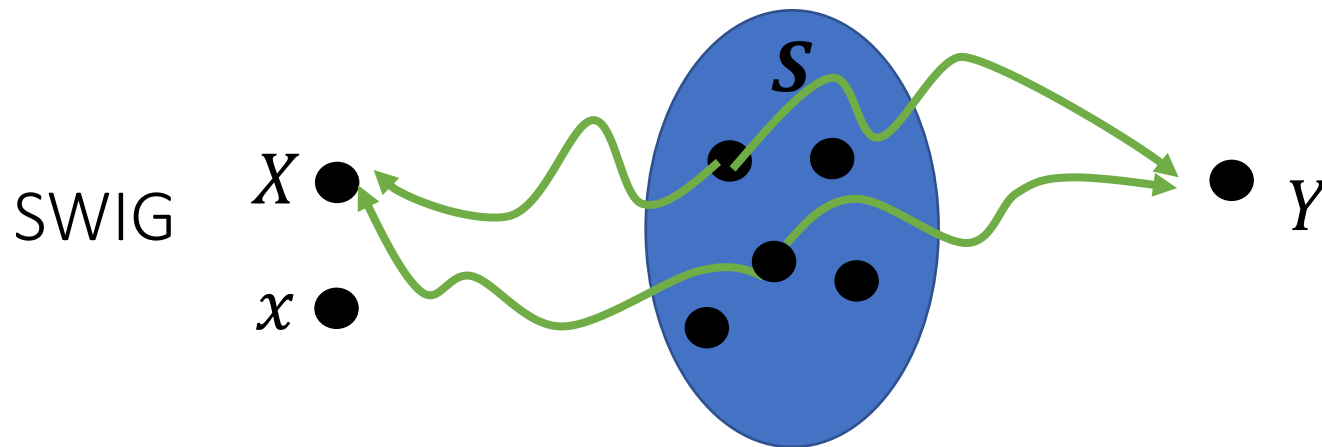
Assuming Y is a descendant of D (otherwise effect is zero)

- Condition on all parents of Y (that are not descendants of D)
- Condition on all parents of D
- Condition on the union of the above

- Condition using backdoor blocking criterion (minimal adjustment)
- Condition on all common causes of D and Y

Conditioning on Parents

- Empirically widely used strategy
- Requires only partial knowledge of the graph
- If we only know the parents of D or the parents of Y we are ok
- Adding any further set to the parent strategy maintains validity



Parents of X

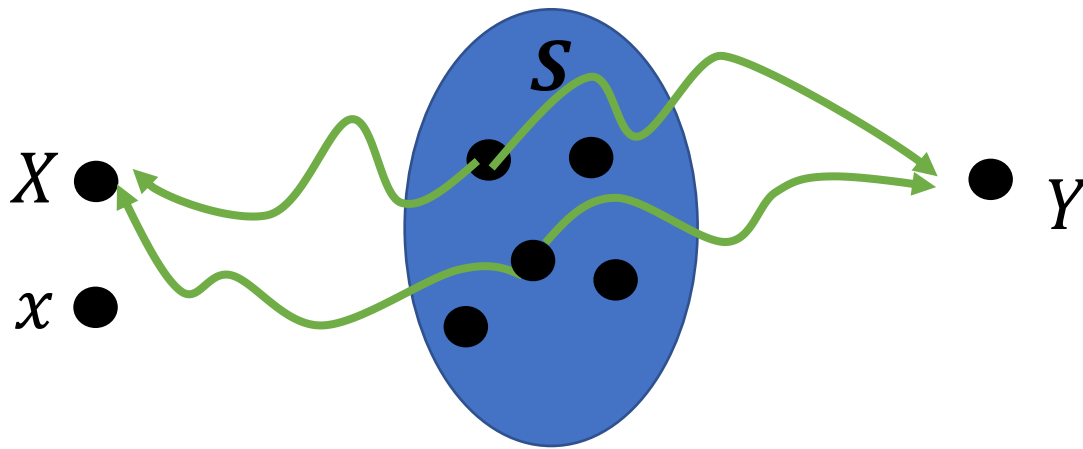
In SWIG, X has no descendants
All paths $X - Y$ must pass from
parent

Parent has to be a fork or chain

Conditioning on Parents

- Empirically widely used strategy
- Requires only partial knowledge of the graph
- If we only know the parents of D or the parents of Y we are ok
- Adding any further set to the parent strategy maintains validity

SWIG



Parents of Y

Parent of Y cannot be a collider

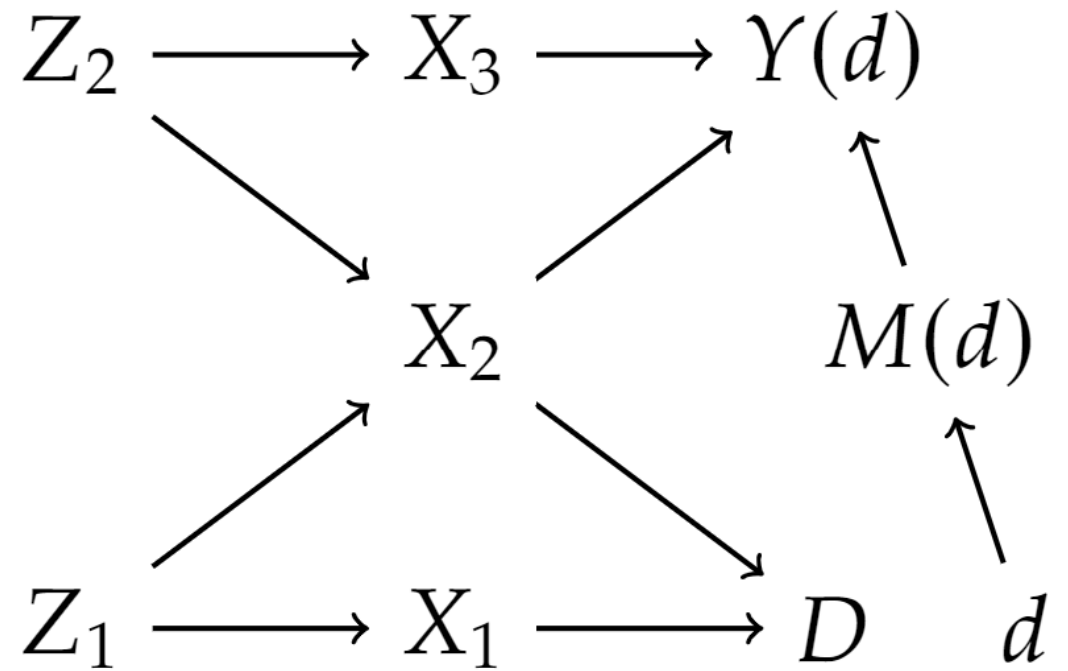
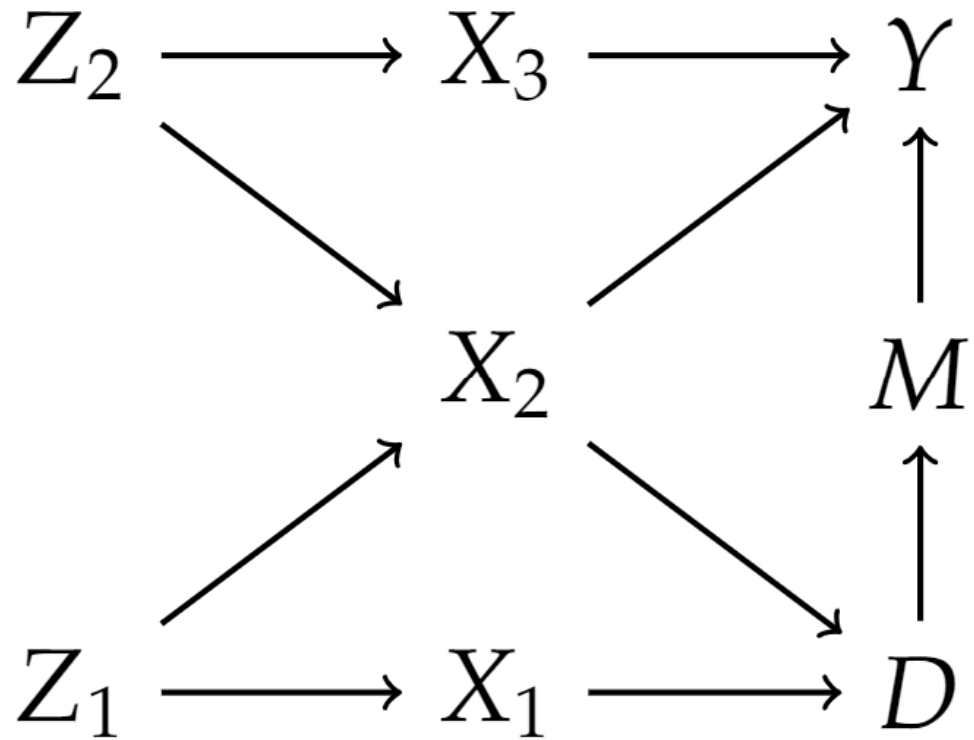
All paths $X - Y$ must pass from a parent of Y

If descendant of D , note that parent variable is not in the SWIG but its counterfactual value is

Backdoor Blocking

- Non-directed path: a path from Y to D that has some reverse edge \leftarrow
- Backdoor path: a non-directed path from Y to D that ends with a \rightarrow
- S is a valid adjustment set if it contains no descendant of D and all backdoor paths from Y to D are blocked
- Allows us to find minimal adjustment sets (small size)

Example



All Common Causes

- Consider all common ancestors of D and Y
- These are called the “common causes of D and Y ”
- Very common empirical practice
- More conservative version: union of ancestors of D and Y

- Any backdoor path from Y to D must either contain a common ancestor or contain a collider
- Conditioning on common ancestors blocks all paths that don't contain a collider

Good and Bad Controls

High Level Categorization of Variables

- **Pre-treatment variables:** variables whose value is determined before the assignment of the treatment
- **Post-treatment variables:** variables whose value is determined after the assignment of the treatment

Good and Bad Controls: High Level

- Pre-treatment variables that are ancestors of either D or Y are ok controls (worst-case can hurt precision/variance)
- There exist pre-treatment variables not of that sort, that can lead to wrong answer (M-bias)
- Most post-treatment variables are bad controls (or don't do much)
- Typically introduce either mediation bias or collider bias (Heckman selection)

“Pre-Treatment Variables”

Examples: Good Controls

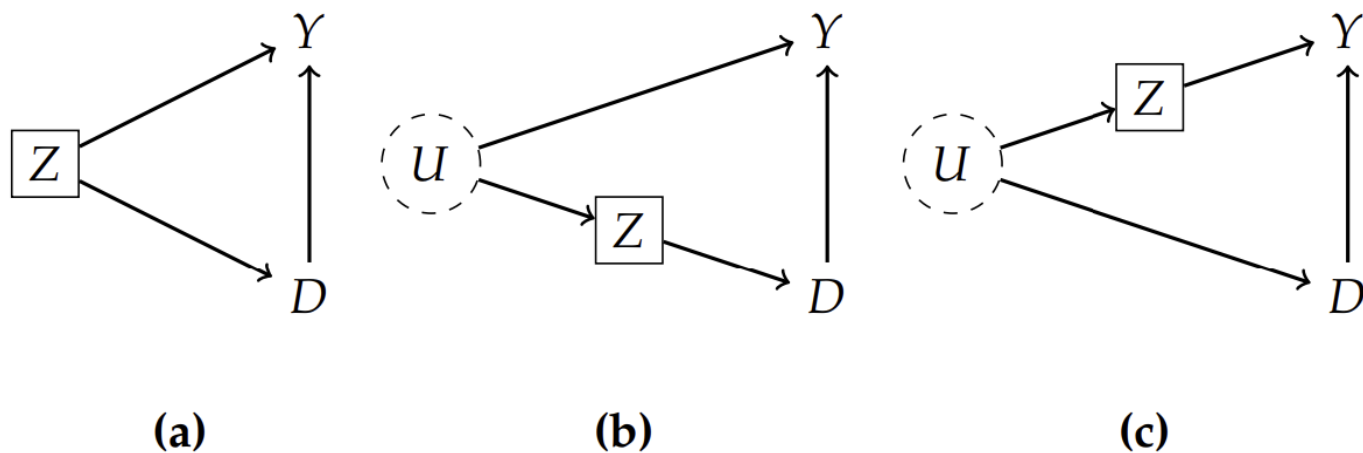
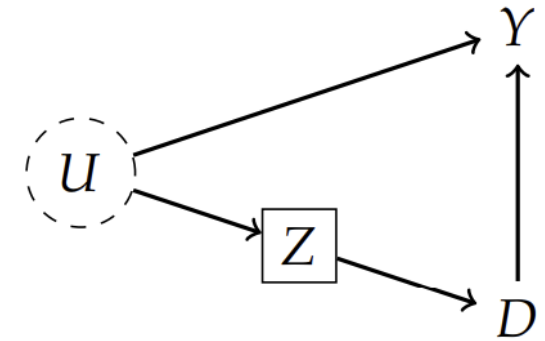


Figure 8.5: Good controls: (a) observed common cause, (b) complete treatment proxy control of unobserved common cause, (c) complete outcome proxy control of unobserved common cause.

Example of Proxy

- Effect of Prenatal Multivitamin consumption (D) on birth defects (Y)
 - Prior family history of birth defects (Z) can influence mother's decision for multivitamin consumption
 - Unmeasured genetic factors (U) cause family history (Z) of birth defects
 - Unmeasured genetic factors (U) have direct influence on birth defects (Y)
-
- Family history of birth defects is a good proxy control for unmeasured genetic factors



(b)

Examples: Good Controls

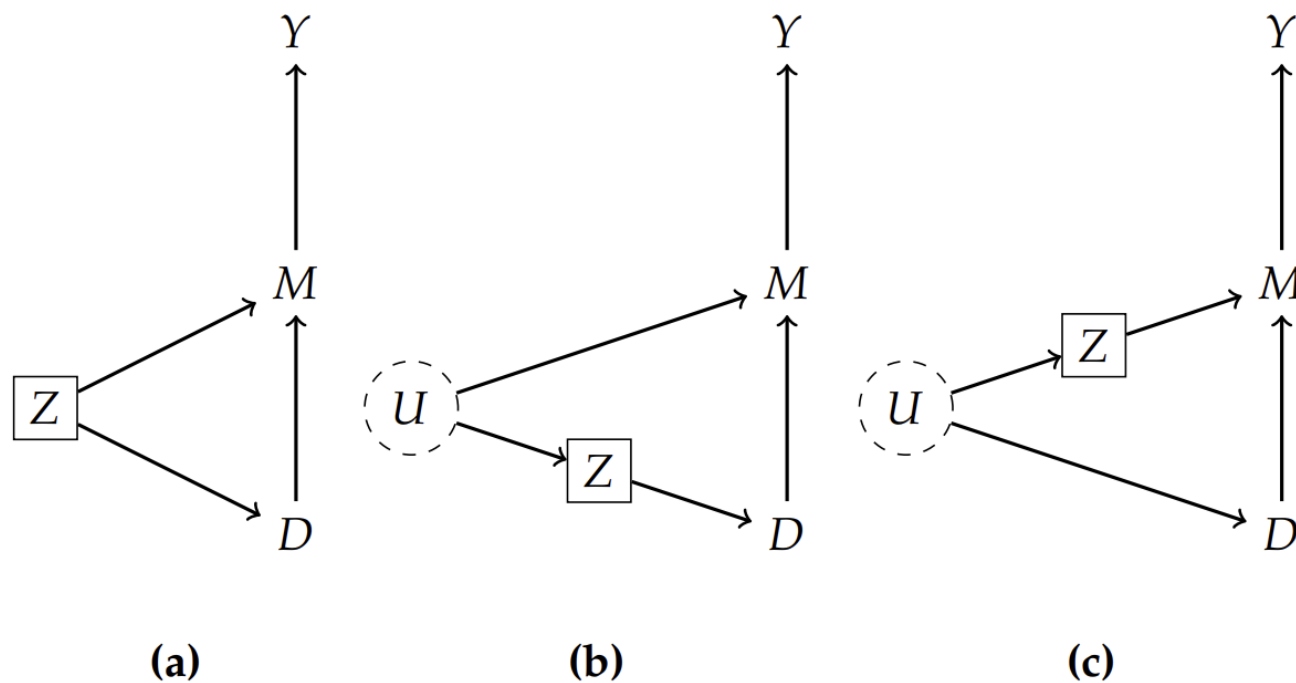


Figure 8.6: Good controls: (a) confounded mediator with observed common cause, (b) confounded mediator, with observed complete treatment proxy control of unobserved common cause, (c) confounded mediator with observed complete outcome proxy control of unobserved common cause.

Examples: Neutral Controls

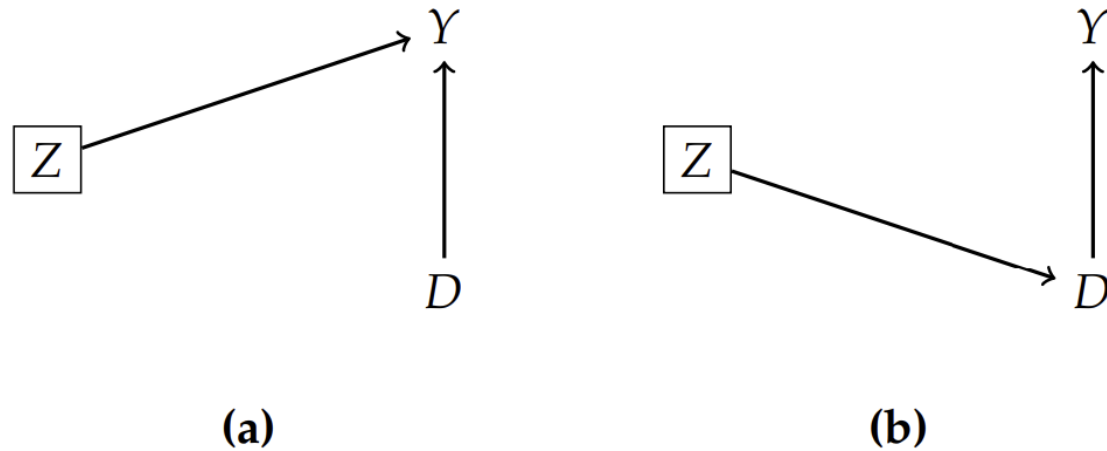


Figure 8.7: Neutral controls: (a) Outcome-only cause. Can improve precision; decrease variance. (b) Treatment-only cause. Can decrease precision; introduce variance.

Examples: Pre-Treatment Variable that is Bad Control

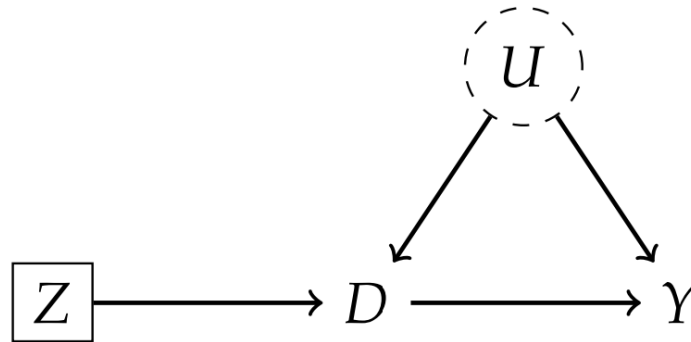


Figure 8.8: Bad control. Bias amplification by adjusting for an *instrument*. Treatment-only cause (*instrument*) that can amplify unobserved confounding bias.

Example: Pre-Treatment Variable that is Bad Control

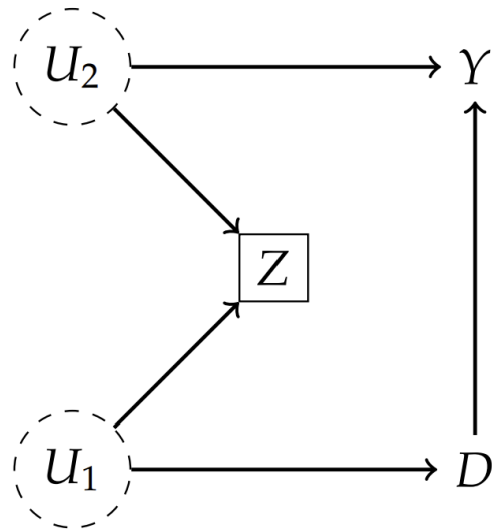
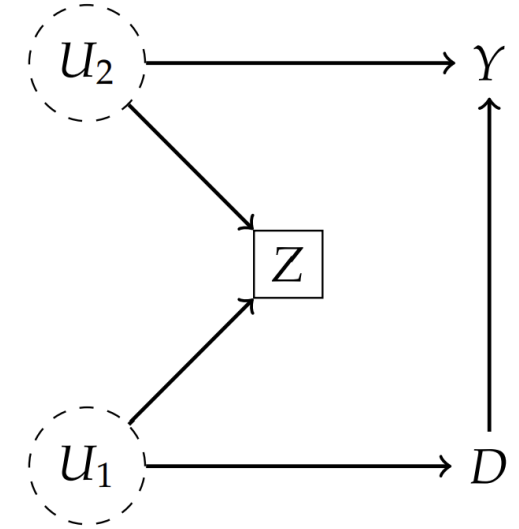


Figure 8.9: Bad control. M-Bias. Pre-treatment variable that introduces Heckman selection bias between two un-correlated unobserved causes.

Example: Homophily Bias in Peer Effects

- Peer effects on civic engagement level
- D is civic engagement of one person at time t
- Y is civic engagement of another person at time $t+1$
- Consider samples of pairs of friends
- Civic engagement can be driven by personal traits U_1 and U_2 that are independently drawn well before friendship and affect civic engagement (e.g. level of altruism)
- Friendship Z driven also by personal traits (homophily)
- By looking at pairs of friends we are implicitly conditioning on Z



Example: M-Bias not robust to perturbation

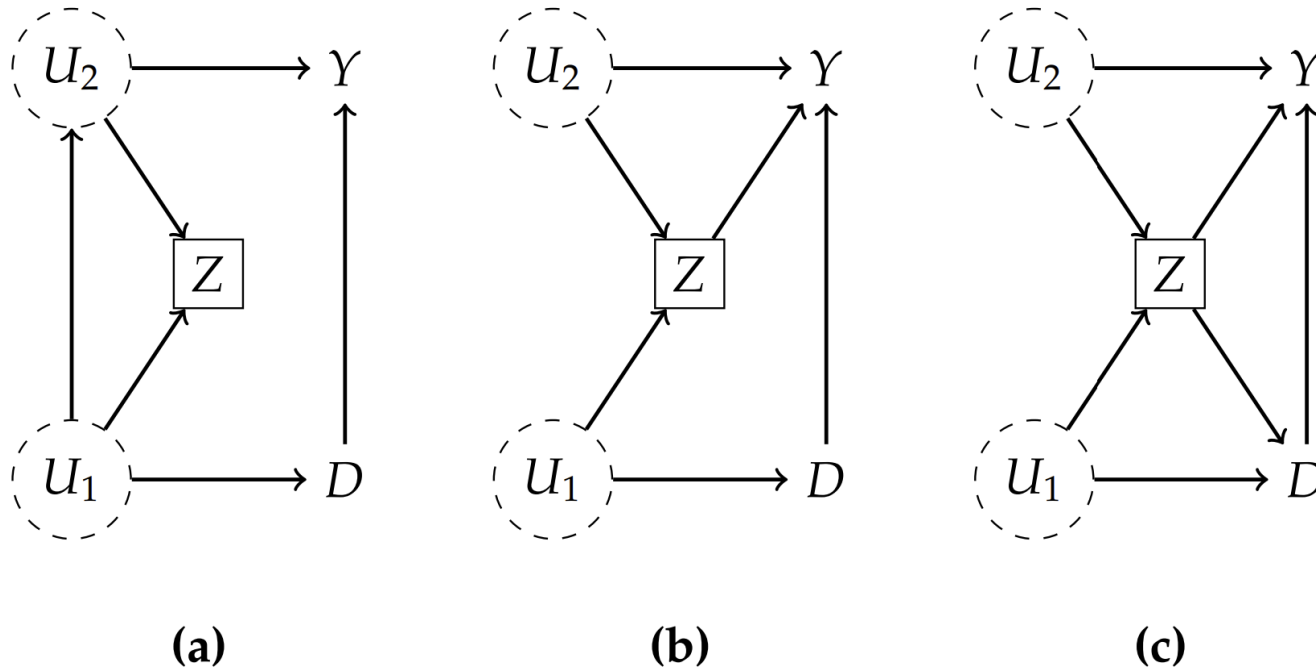


Figure 8.10: Unclear controls: **(a)** M-bias with correlated unobserved factors, **(b)** M-Bias with confounding. Pre-treatment variable that introduces Heckman selection between two un-correlated unobserved causes, but also is a confounder itself. No solution is perfect. **(c)** Butterfly Bias. M-bias with direct confounding.

“Post-Treatment Variables”

Example: Bad Controls, Mediation Bias

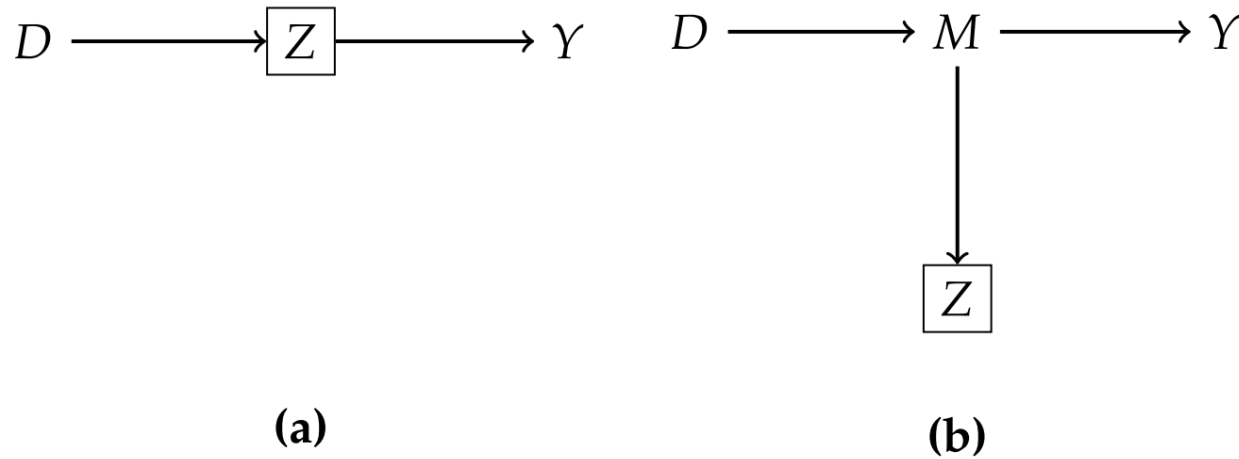


Figure 8.11: Bad controls: (a) over-control bias, by controlling on a mediator, (b) over-control bias, by controlling on some outcome caused by a mediator.

Example: Exception to Mediation Bias

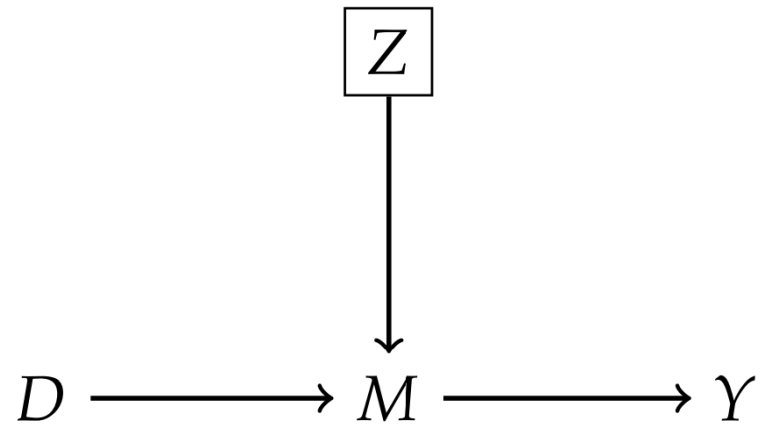


Figure 8.12: Neutral control. Cause of a mediator. Can potentially improve precision.

Example: Controlled Direct Effect Gone Wrong

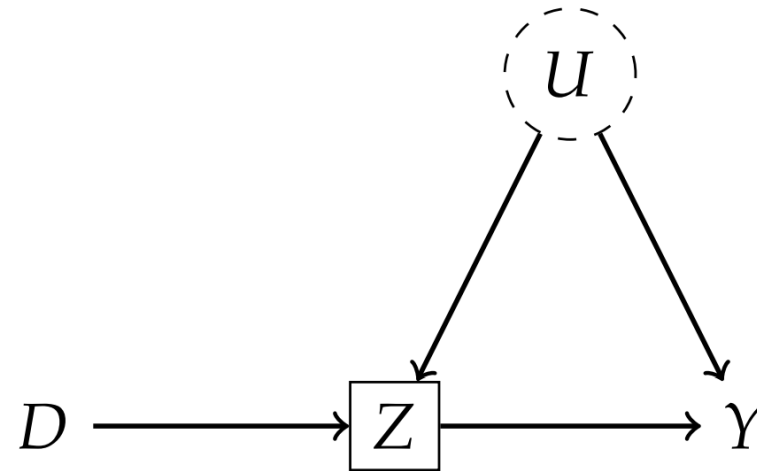


Figure 8.13: Bad control even for *controlled direct effect*. Confounded mediator bias.

Example: Bad Controls, Collider (Heckman Selection) bias

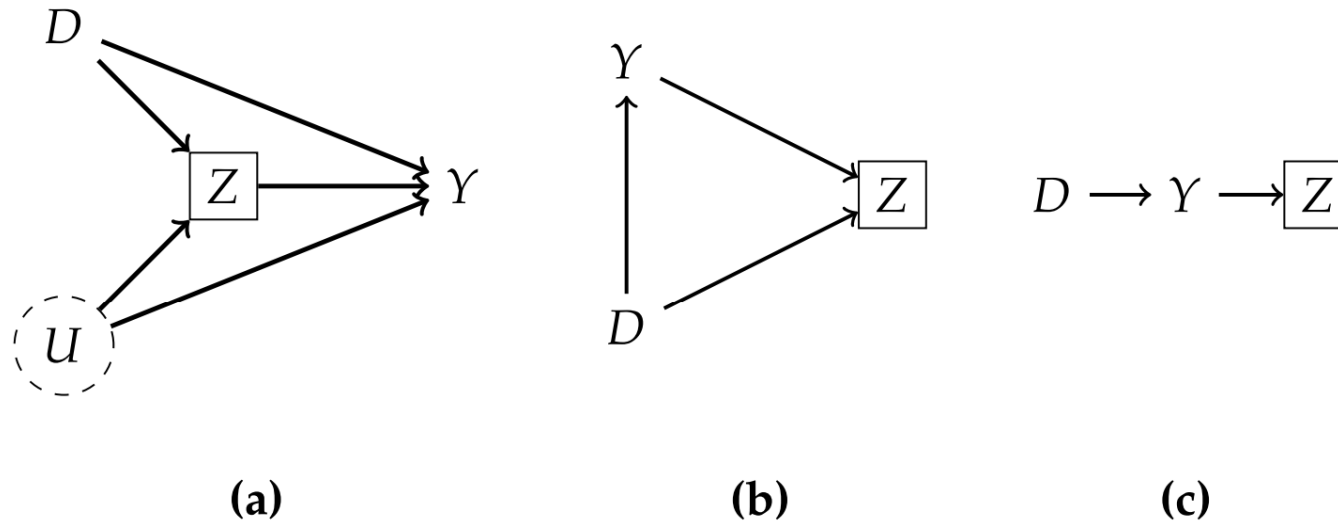


Figure 8.14: Bad controls: (a) collider stratification bias (e.g. low birth-weight “paradox” example), (b) collider stratification bias, (c) case control bias, by controlling on another outcome of the outcome of interest.

Direct and Indirect Effects for Analysis of Wage Discrimination

A Structural Equation Model of Wage Discrimination

$$\begin{aligned}Y &:= \kappa D_w + \theta H + \epsilon_Y \\D_w &:= G + \delta H + \epsilon_{D_w} \\H &:= \gamma G + \lambda D_h + \epsilon_H \\D_h &:= G + \epsilon_{D_h} \\G, &\end{aligned}$$

Existence of arrow $G \rightarrow H$ indicates there is potential heterogeneity in preferences for human capital acquisition (e.g. inherent preferences over different occupations)

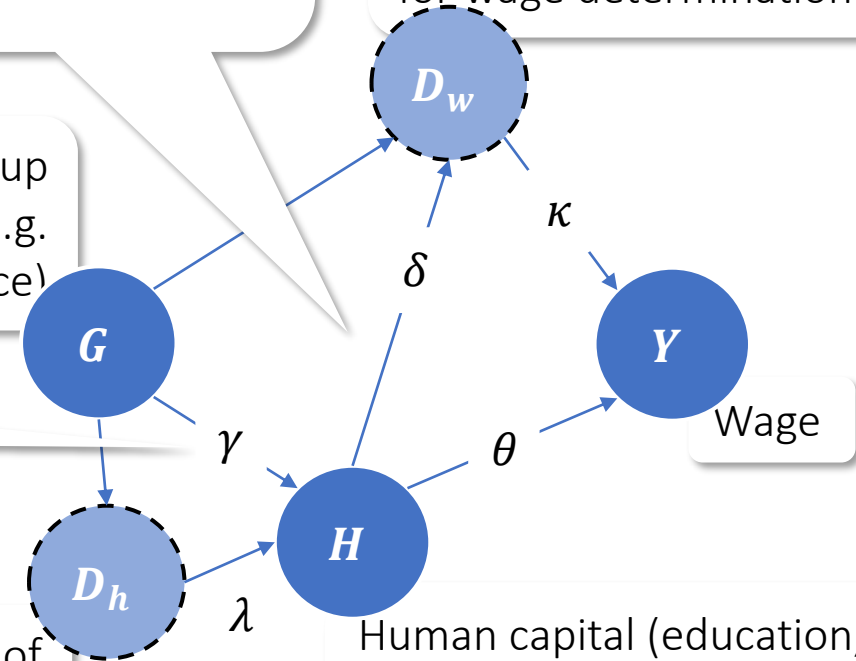
Absence of arrow $G \rightarrow Y$ indicates there is no inherent differentiation in productivity level driven by G ; only due to discrimination or differences in human capital

Sensitive Group Attribute (e.g. sex, race)

Abstraction Variable of “Discriminatory Behavior” for wage determination

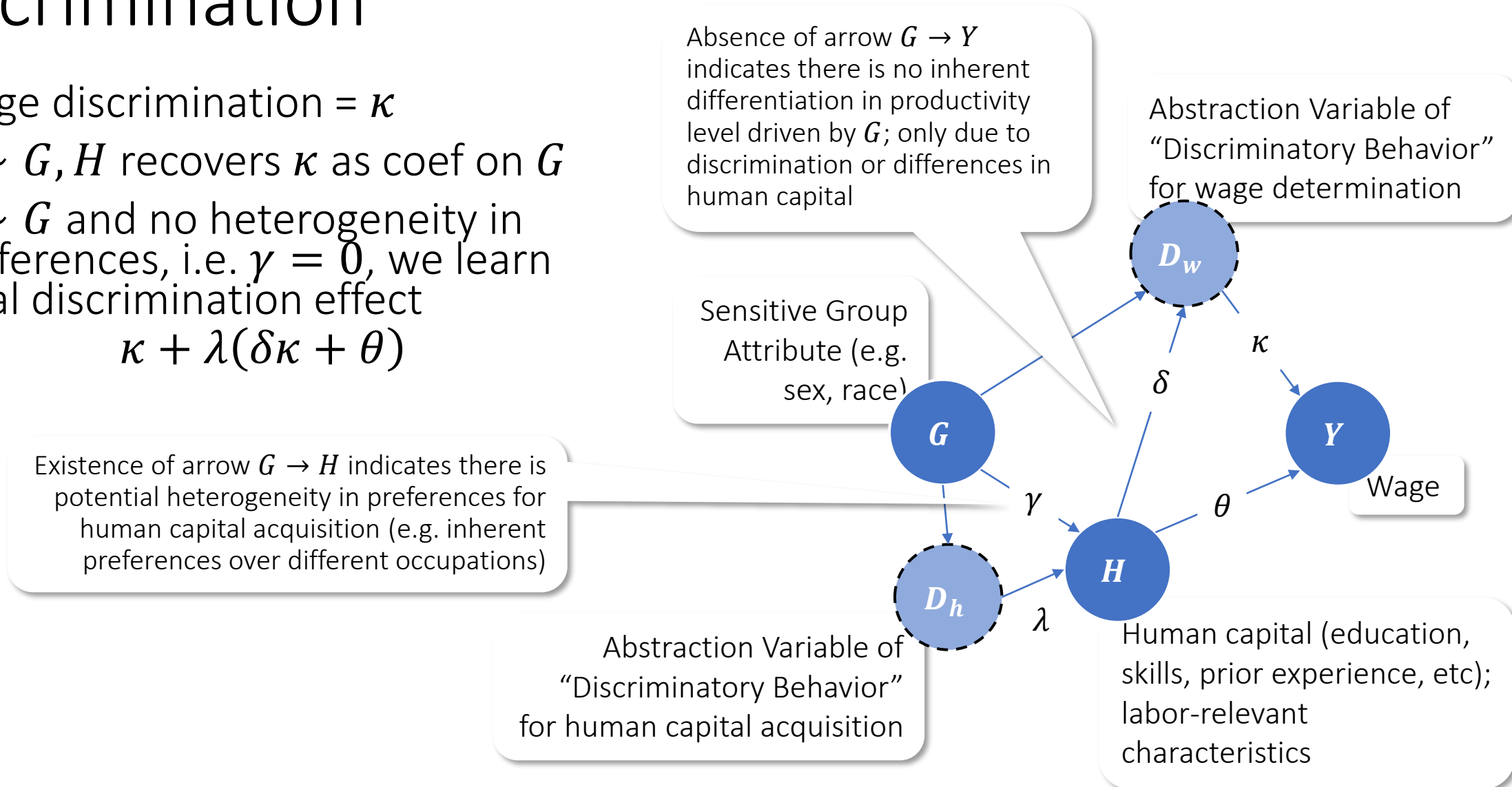
Abstraction Variable of “Discriminatory Behavior” for human capital acquisition

Human capital (education, skills, prior experience, etc); labor-relevant characteristics



A Structural Equation Model of Wage Discrimination

- Wage discrimination = κ
- $Y \sim G, H$ recovers κ as coef on G
- $Y \sim G$ and no heterogeneity in preferences, i.e. $\gamma = 0$, we learn total discrimination effect
 $\kappa + \lambda(\delta\kappa + \theta)$



Concluding Remarks

- Any study that claims it is estimating the effect of D on Y by conditioning on S must be based on a rigorous thought process
- The DAG/ASEM framework is a rigorous form of this process
- Enables explicit incorporation of domain knowledge in a domain expert friendly manner
- Automatic identification arguments and testable restrictions
- Effective in communicating assumptions of observational study

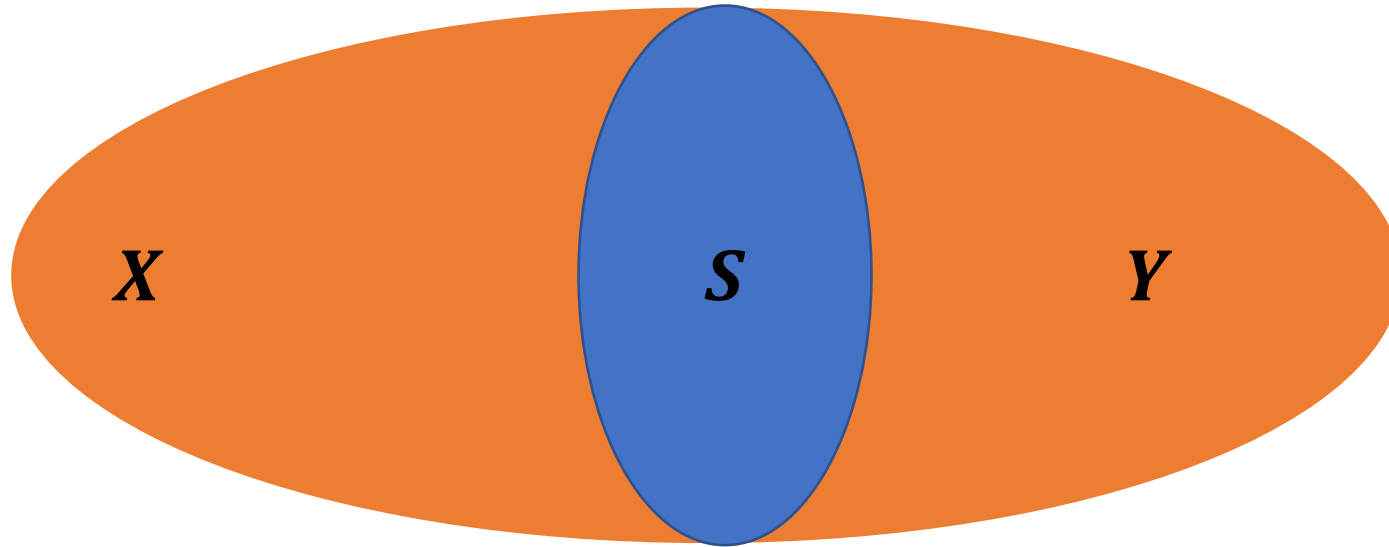
Proving the Main Theorem!

Proof Step 1

- A set of nodes \mathbf{X} is called ancestral if all ancestors of \mathbf{X} are in \mathbf{X}
- Removing all nodes outside of an ancestral set and looking at the resulting graph and ASEM, the probability law is the same as the probability law of \mathbf{X} in the original graph (exercise)

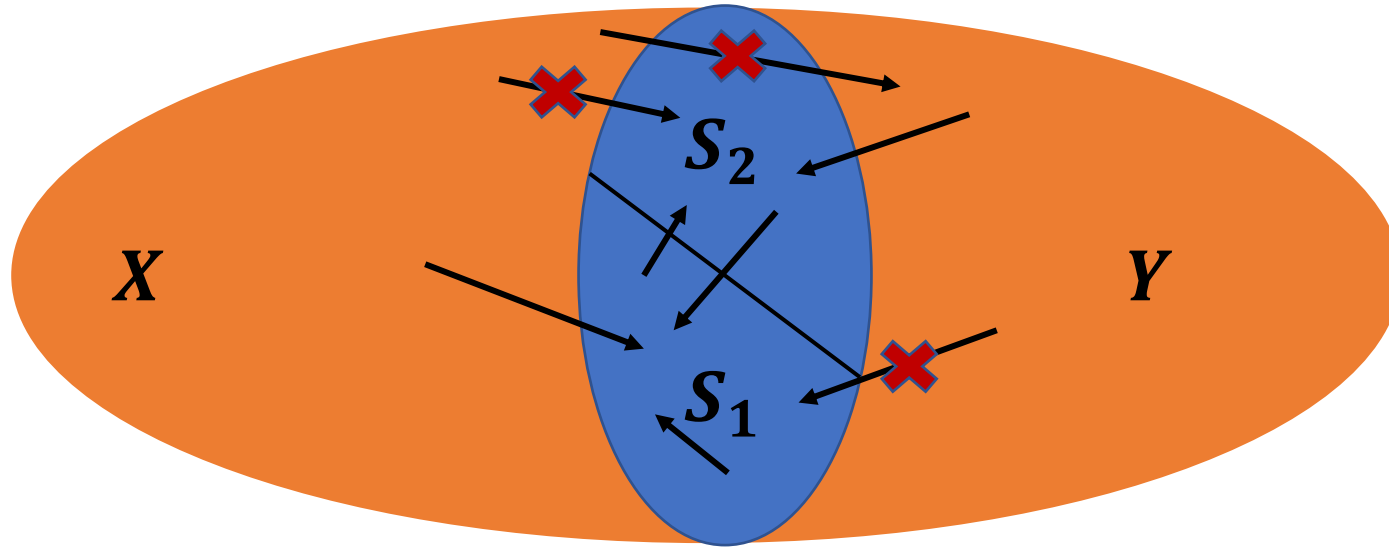
Proof Step 2

- Suppose that a set of nodes \mathbf{X} is D-separated from a set of nodes \mathbf{Y} by a set of nodes \mathbf{S}
- And that $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{S}$ is the set of all nodes



Proof Step 2

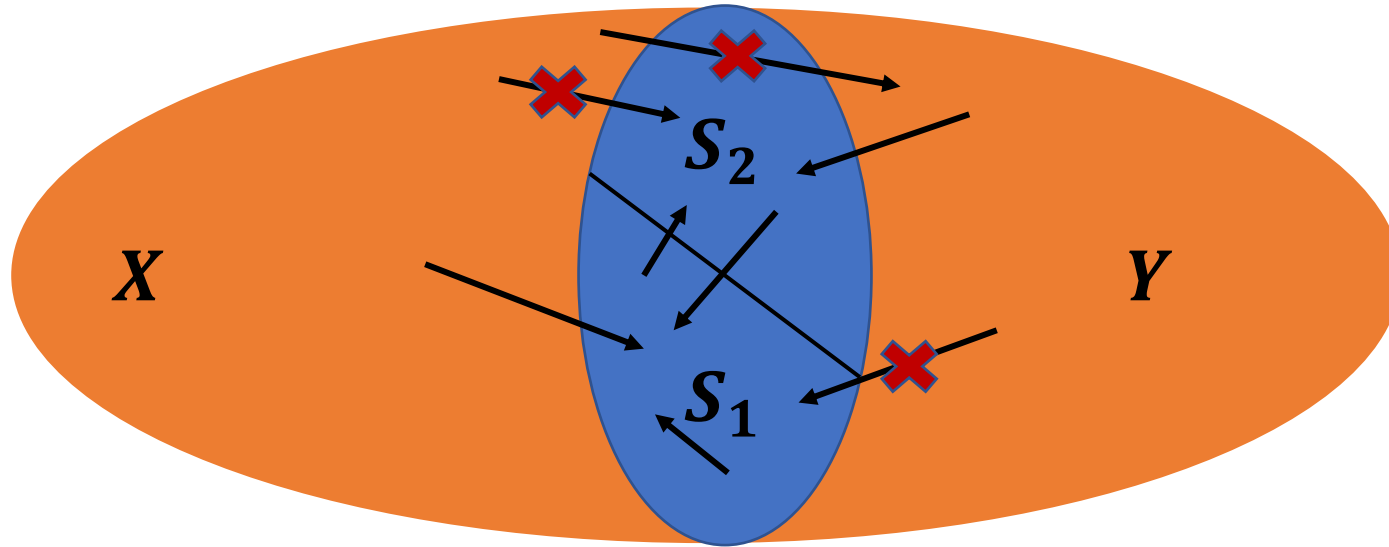
- Let S_1 the subset of S that has a parent in X . Let S_2 the remainder.
- It has to be that $Pa(X \cup S_1) \in X \cup S$
- It has to be that $Pa(Y \cup S_2) \in Y \cup S$



Proof Step 2

- We can factorize:

$$p(x, y, s) = \prod_{W \in X \cup S_1} p(w|pa_w) \prod_{W \in Y \cup S_2} p(w|pa_w) = f(x, s_1)g(y, s_2)$$



Proof Step 2

- We can factorize:

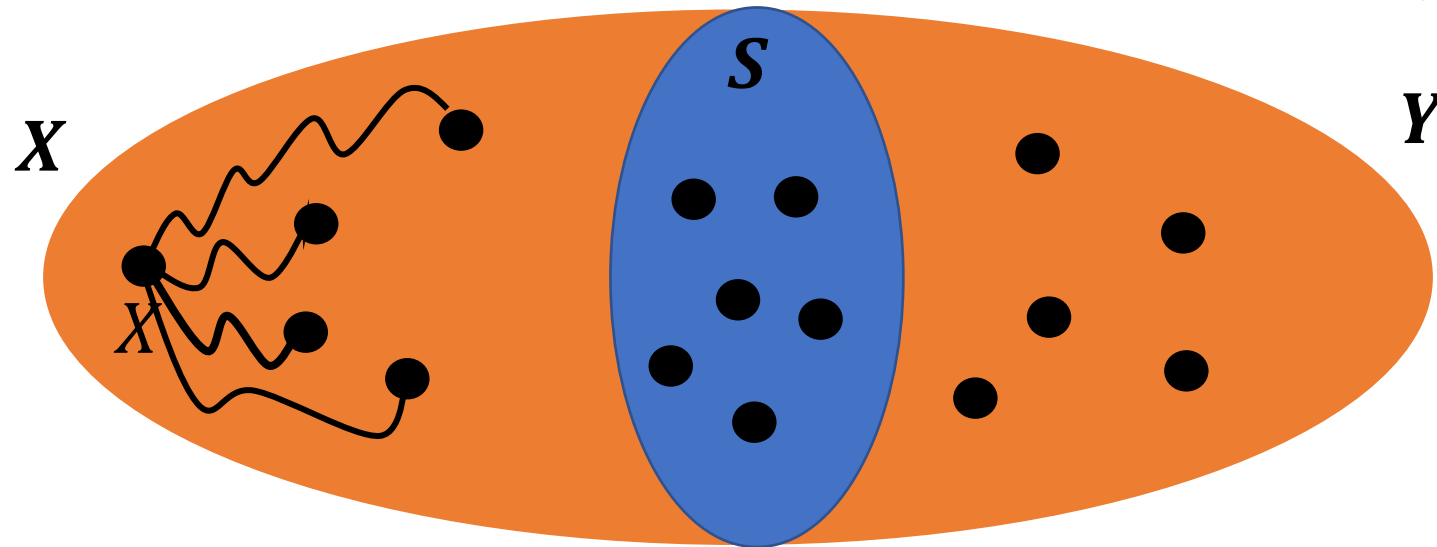
$$p(x, y, s) = \prod_{W \in X \cup S_1} p(w|pa_w) \prod_{W \in Y \cup S_2} p(w|pa_w) = f(x, s_1)g(y, s_2)$$

- Implies that:

$$X \perp\!\!\!\perp Y \mid S$$

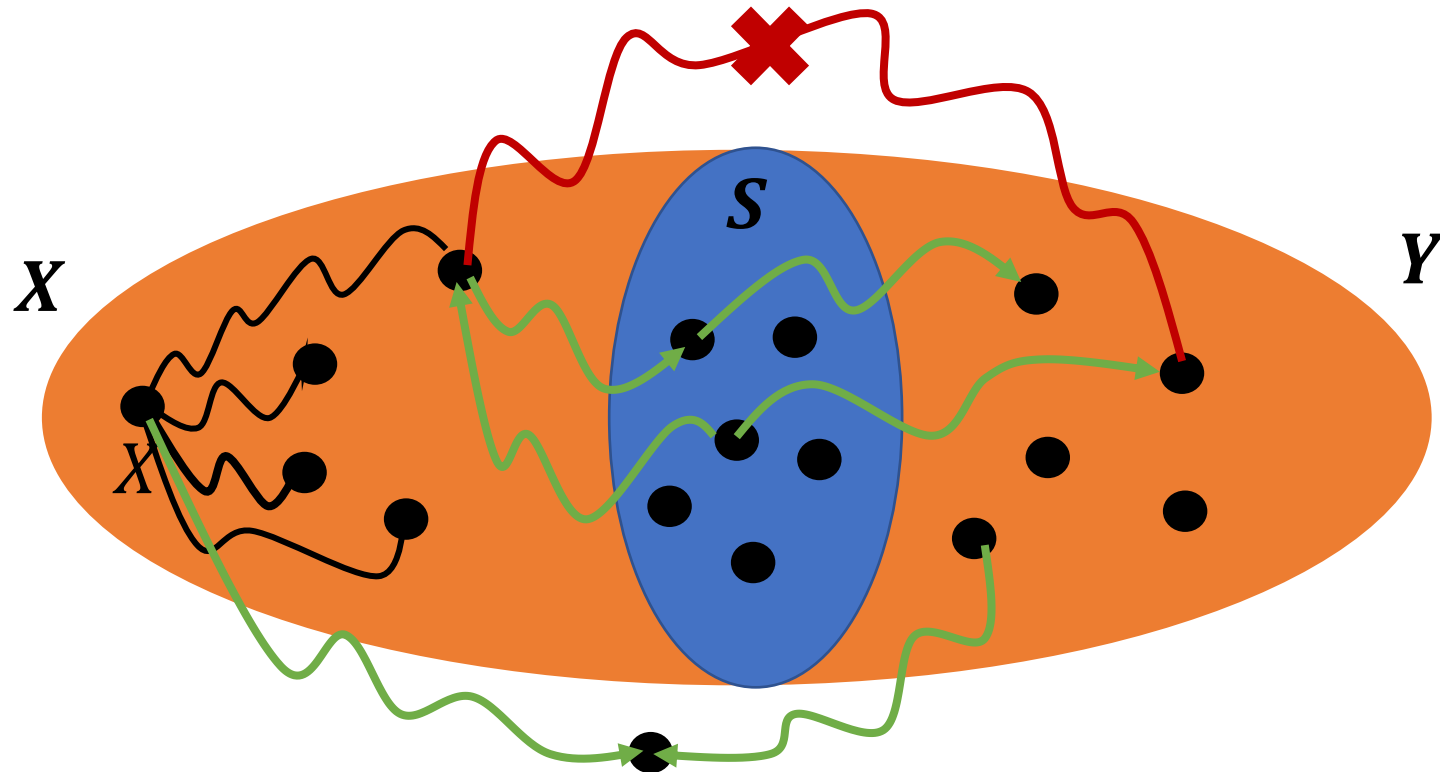
Final Step

- By first step, we can restrict to ancestral set of $X \cup Y \cup S$
- Does not change conditional independence relations (exercise)
- Does not change d-separation relations (exercise)
- Define **X** nodes in ancestral set of $X \cup Y \cup S$ not d-separated from X
- Define **Y** the remainder of nodes in ancestral set not in **X, S** .



Final Step

- By definition of d-separation, S must d-separate X from Y (exercise)
- We can invoke previous critical lemma



Final Step

- By marginalization

$$p(x, y, \mathbf{s}) = \int \int p(x, \mathbf{x}', y, \mathbf{y}', \mathbf{s}) d\mathbf{x}' d\mathbf{y}'$$

- By step 2

$$p(x, y, \mathbf{s}) = \int \int f(x, \mathbf{x}', \mathbf{s}) g(y, \mathbf{y}', \mathbf{s}) d\mathbf{x}' d\mathbf{y}'$$

- We can split integrals

$$p(x, y, \mathbf{s}) = \int f(x, \mathbf{x}', \mathbf{s}) d\mathbf{x}' \int g(y, \mathbf{y}', \mathbf{s}) d\mathbf{y}'$$

- Thus

$$p(x, y, \mathbf{s}) = \bar{f}(x, \mathbf{s}) \bar{g}(y, \mathbf{s}) \Rightarrow X \perp\!\!\!\perp Y \mid S$$