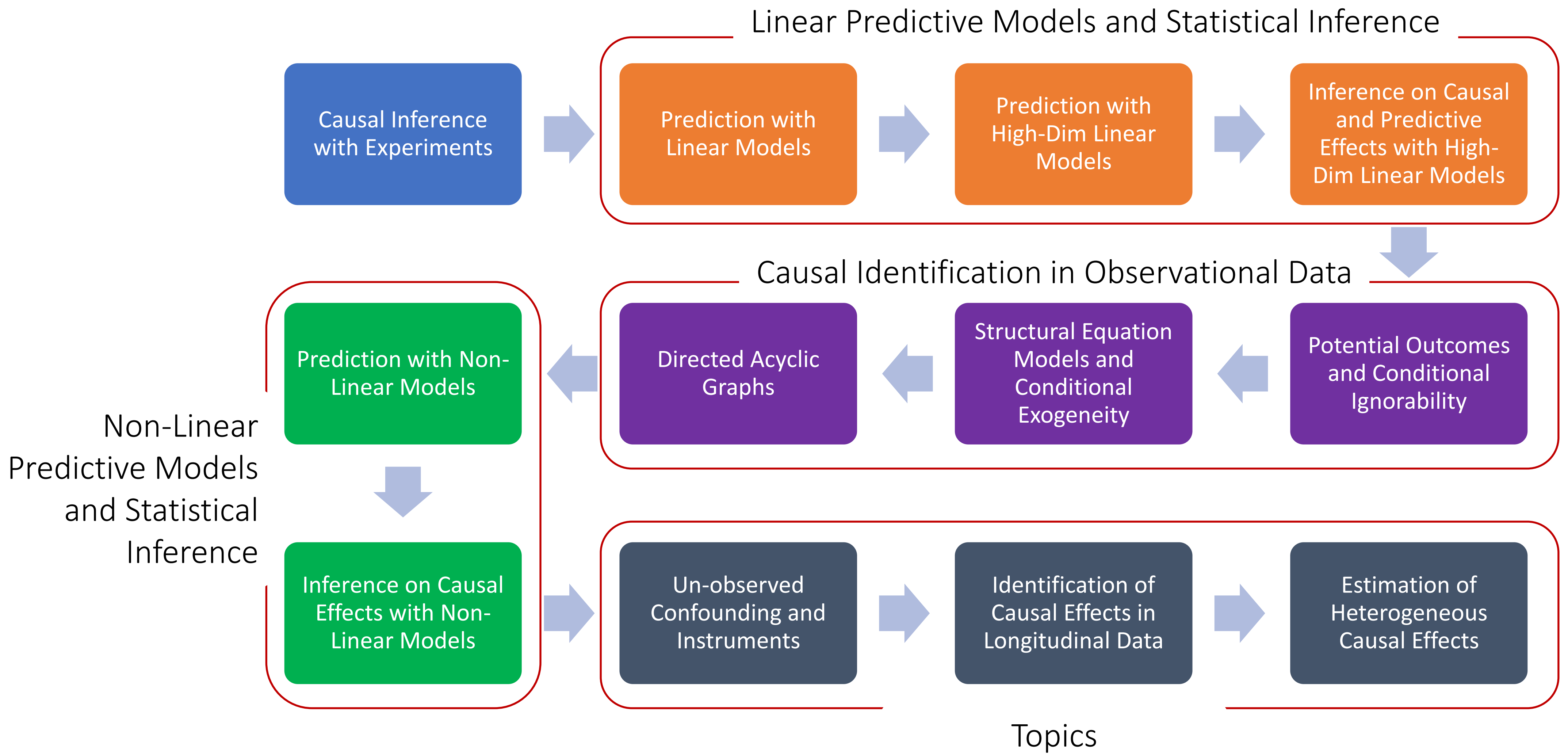
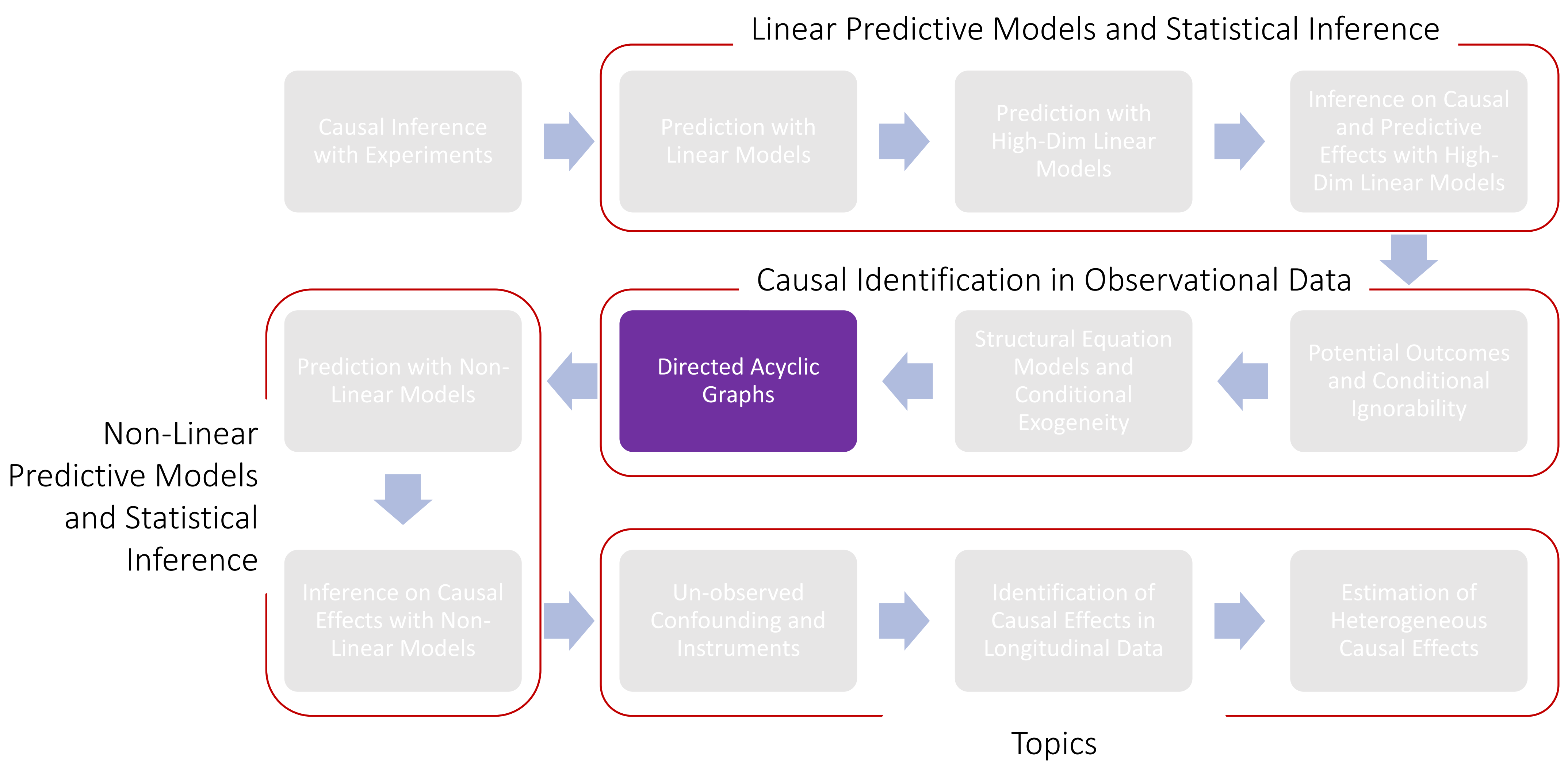
MS&E 228: Directed Acyclic Graphs and Non-Linear SEMs

Vasilis Syrgkanis

MS&E, Stanford



Linear Predictive Models and Statistical Inference
Causal Inference with Experiments
Prediction with Linear Models
Prediction with High-Dim Linear Models
Inference on Causal and Predictive Effects with High-Dim Linear Models
Causal Identification in Observational Data
Potential Outcomes and Conditional Ignorability
Structural Equation Models and Conditional Exogeneity
Directed Acyclic Graphs
Prediction with Non-Linear Models
Non-Linear Predictive Models and Statistical Inference
Inference on Causal Effects with Non-Linear Models
Un-observed Confounding and Instruments
Identification of Causal Effects in Longitudinal Data
Estimation of Heterogeneous Causal Effects
Topics

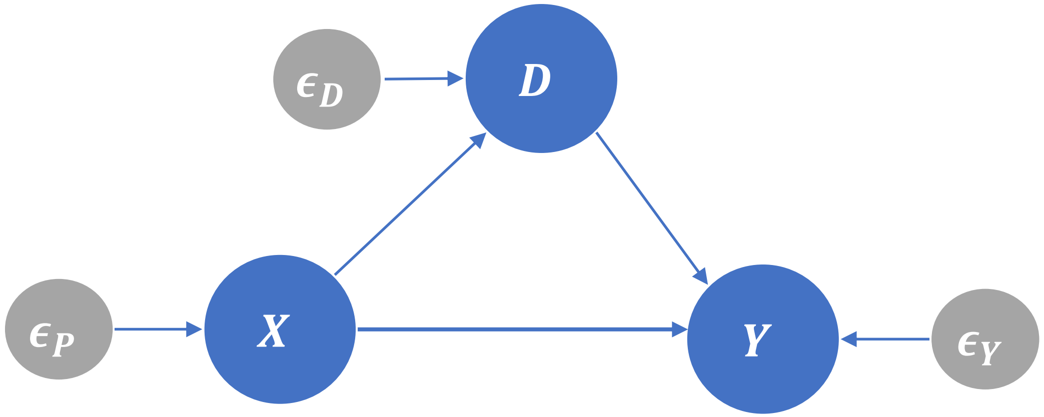

Linear Predictive Models and Statistical Inference
Causal Inference with Experiments
Prediction with Linear Models
Prediction with High-Dim Linear Models
Inference on Causal and Predictive Effects with High-Dim Linear Models
Causal Identification in Observational Data
Prediction with Non-Linear Models
Directed Acyclic Graphs
Structural Equation Models and Conditional Exogeneity
Potential Outcomes and Conditional Ignorability
Non-Linear Predictive Models and Statistical Inference
Inference on Causal Effects with Non-Linear Models
Un-observed Confounding and Instruments
Identification of Causal Effects in Longitudinal Data
Estimation of Heterogeneous Causal Effects
Topics

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

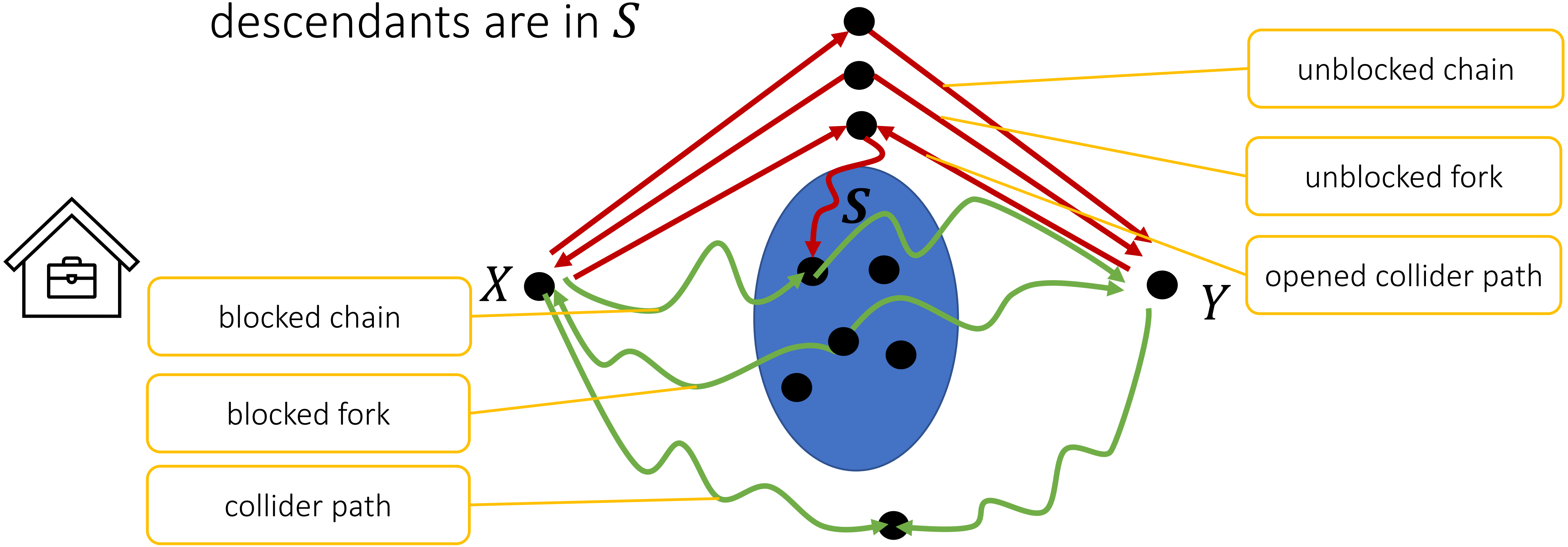
$$(X \perp\!\!\!\perp_d Y \mid S)_G \quad \Rightarrow \quad X \perp\!\!\!\perp Y \mid S$$

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

$$Y = \boxed{\alpha} X + \beta' S + \epsilon, \qquad \epsilon \perp (X, S)$$

Test whether it is non-zero!

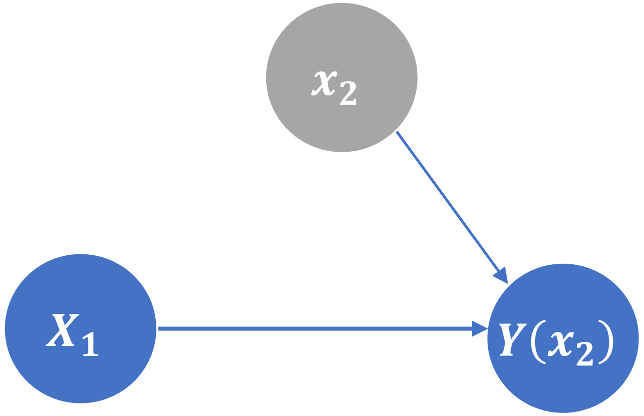
X is d-separated from Y by S if **every path** from X to Y is **blocked**.
S **blocks a path** if one of the following holds:
- path contains chain $X \to M \to Y$ or fork $X \leftarrow M \to Y$ and $M \in S$
- path contains collider $X \to M \leftarrow Y$ and neither M nor its descendants are in S



Do Interventions $\mathrm{do}(X_j = x_j)$

Replace structural response for X_j with $X_j := x_j$.

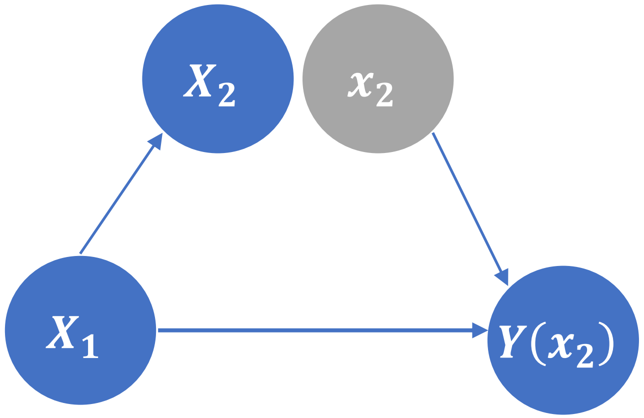
Measures potential outcome $Y(x_j)$.



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

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



Graphical Criteria for Valid Adjustment Sets

Conditional Ignorability

- Recall to do identification by conditioning we need for a set S

$$Y(d) \perp\!\!\!\perp D \mid S$$

- Then predictive response equals structural response

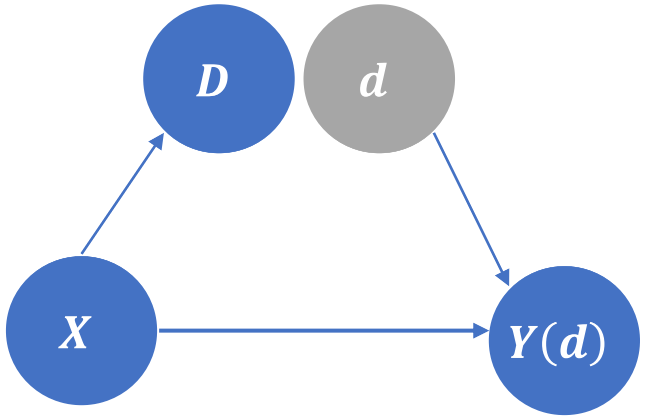
$$E[Y(d) \mid S] = E[\, Y \mid D = d, S\,]$$

- Average predictive response equals average structural response

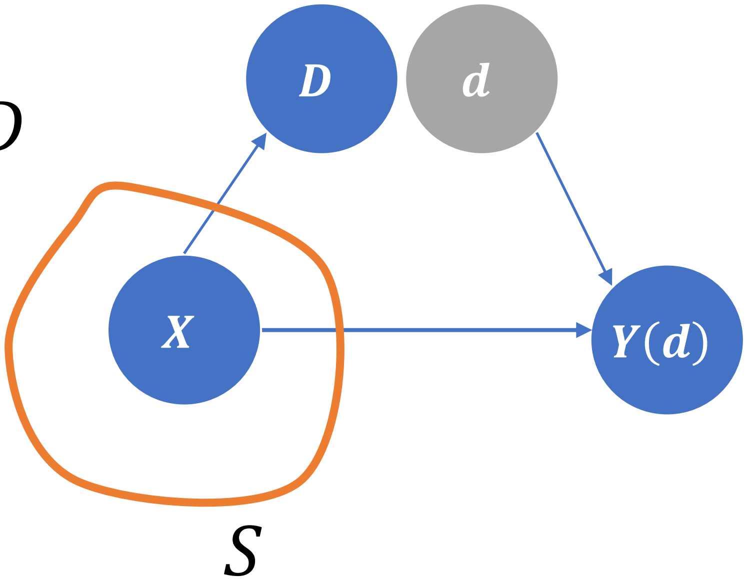
$$E[Y(d)] = E\big[E[Y|D = d, S]\big]$$

How can we check Conditional Ignorability

- Given a DAG, can we visually inspect if conditional ignorability holds
- Note that the SWIG graph contains both $Y(d)$ and D!
- We can simply check if $Y(d)$ is independent of D conditional on S on the SWIG graph!
- This is just a conditional independence statement on a DAG
- We can use d-separation!



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



Example

$$
\begin{tikzcd}
Z \arrow[r] & D & d \arrow[r] & Y(d) \\
 & & X \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

Example

$$
\begin{tikzcd}
Z_2 \arrow[r] \arrow[dr] & X_3 \arrow[r] & Y \\
 & X_2 \arrow[ur] \arrow[dr] & M \arrow[u] \\
Z_1 \arrow[r] \arrow[ur] & X_1 \arrow[r] & D \arrow[u]
\end{tikzcd}
$$

$$
\begin{tikzcd}
Z_2 \arrow[r] \arrow[dr] & X_3 \arrow[r] & Y(d) & \\
 & X_2 \arrow[ur] \arrow[dr] & & M(d) \arrow[ul] \\
Z_1 \arrow[r] \arrow[ur] & X_1 \arrow[r] & D & d \arrow[u]
\end{tikzcd}
$$

Side Note: Counterfactual Distributions

- We can also easily handle counterfactual distribution learning
$$\Pr(\, Y(d) \leq t \mid S \,)$$
- Redefine outcome to be $\widetilde{\mathrm{Y}}(d) \coloneqq 1(Y(d) \leq t)$

- Same statements hold for this binary outcome
$$\Pr(\, Y(d) \leq t \mid S \,) = E\left[\tilde{Y}(d)\middle|S\right] = E\left[\tilde{Y}\middle|D = d, S\right] = \Pr(Y \leq t|D = d, S)$$

$$\Pr(Y(d) \leq t) = E\left[E\left[\tilde{Y}\middle|D = d, S\right]\right] = E[\Pr(Y \leq t|D = d, S)]$$

Useful Adjustment Strategies

Adjustment Strategies

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

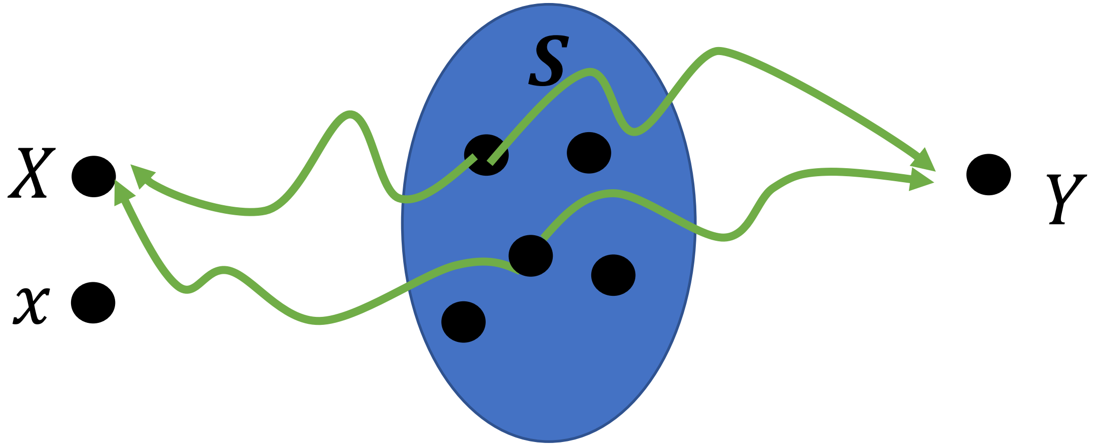
- Condition on all parents of Y (that are not descendants of D)
- Condition on all parents of D
- Condition on the union of the above

- 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

SWIG



Parents of $\boldsymbol{X}$

In SWIG, X has no descendants

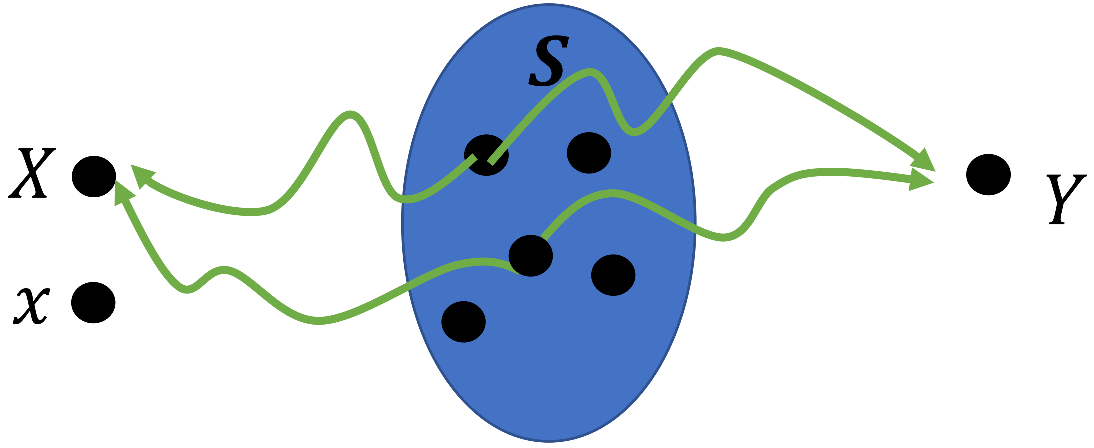
All paths $X - Y$ must pass from parent

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

$$
\begin{tikzcd}
Z_2 \arrow[r] \arrow[dr] & X_3 \arrow[r] & Y \\
 & X_2 \arrow[ur] \arrow[dr] & M \arrow[u] \\
Z_1 \arrow[r] \arrow[ur] & X_1 \arrow[r] & D \arrow[u]
\end{tikzcd}
$$

$$
\begin{tikzcd}
Z_2 \arrow[r] \arrow[dr] & X_3 \arrow[r] & Y(d) & \\
 & X_2 \arrow[ur] \arrow[dr] & & M(d) \arrow[ul] \\
Z_1 \arrow[r] \arrow[ur] & X_1 \arrow[r] & D & d \arrow[u]
\end{tikzcd}
$$

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

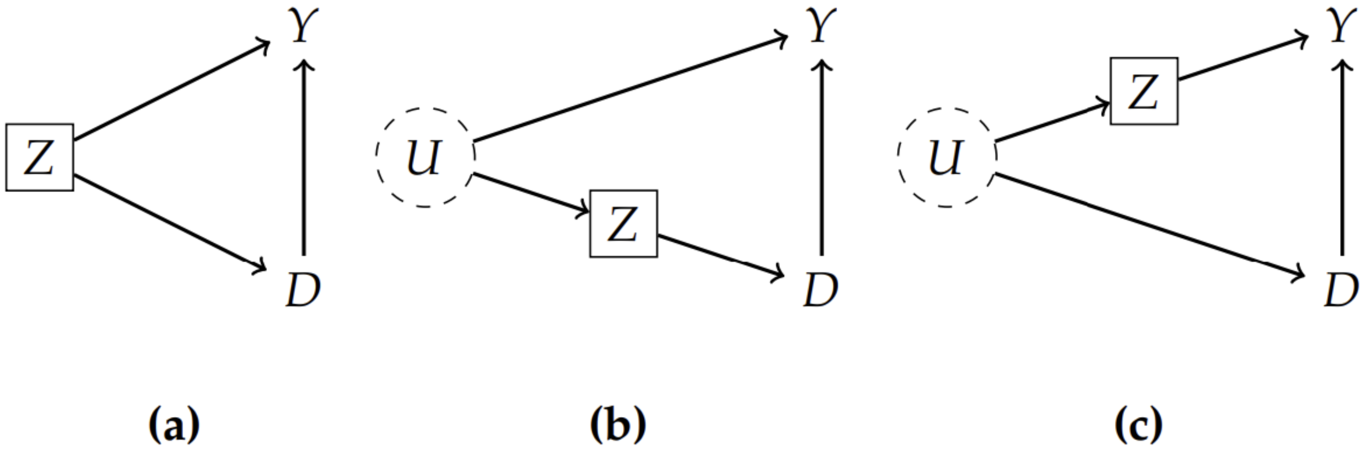
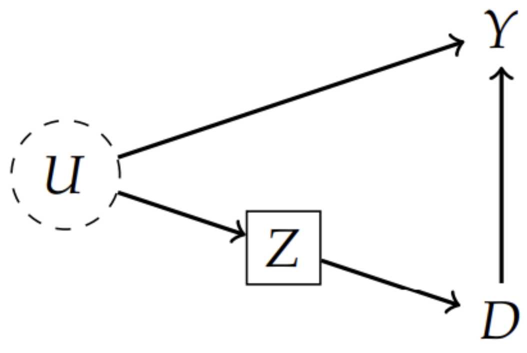


Figure 8.5: Good controls: **(a)** observed common cause, **(b)** complete treatment proxy control of unobserved common cause, **(c)** complete outcome proxy control of unobserved common cause.

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

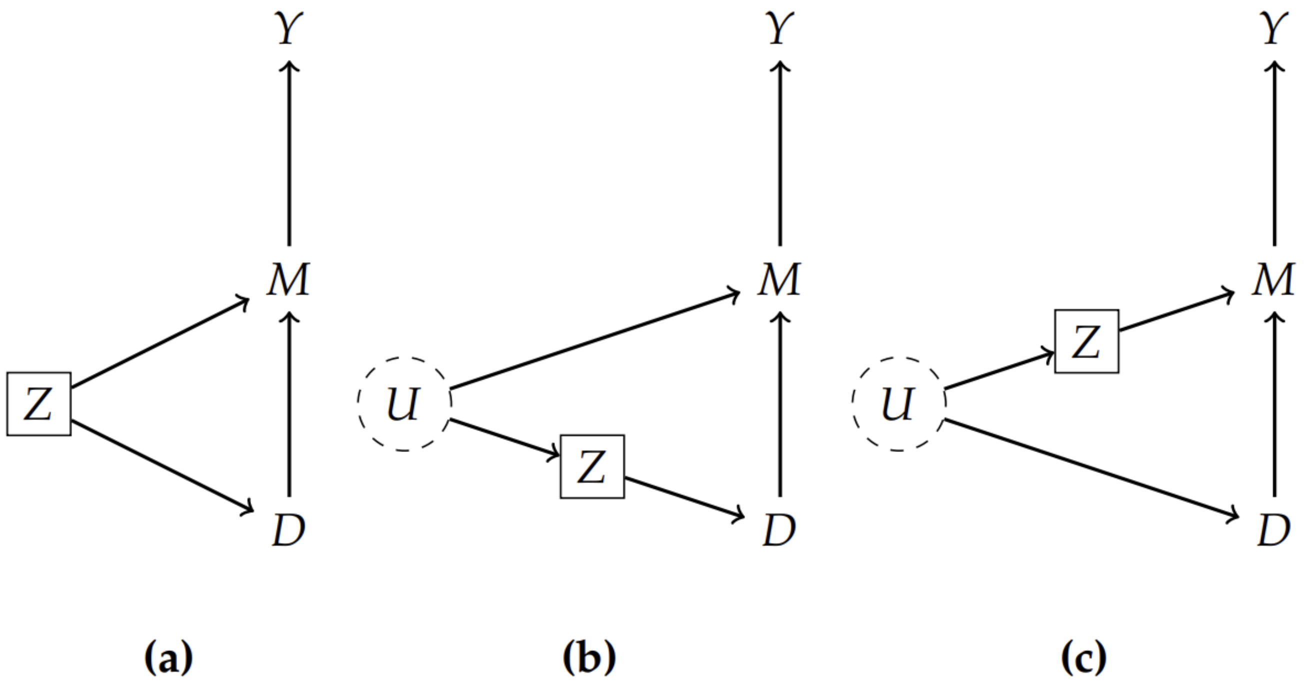


Figure 8.6: Good controls: **(a)** confounded mediator with observed common cause, **(b)** confounded mediator, with observed complete treatment proxy control of unobserved common cause, **(c)** confounded mediator with observed complete outcome proxy control of unobserved common cause.

Examples: Neutral Controls

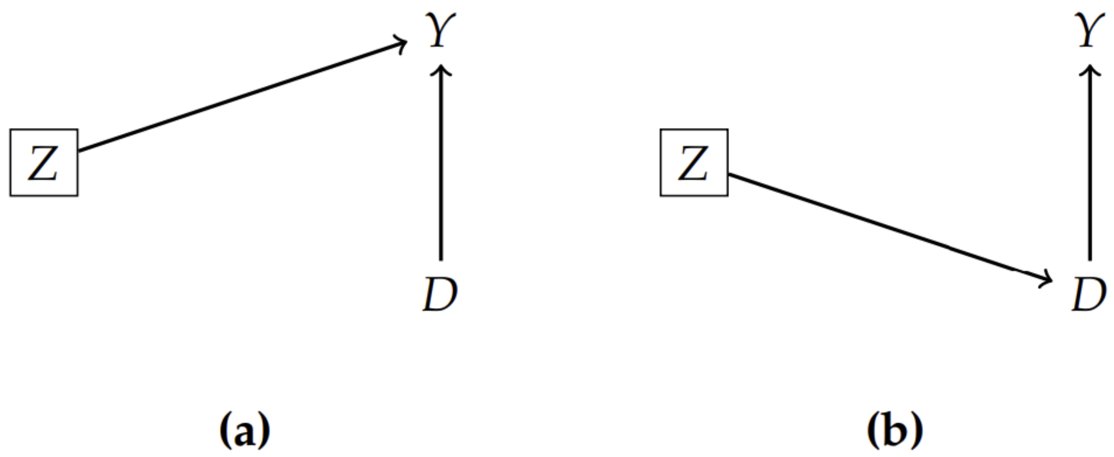


Figure 8.7: Neutral controls: **(a)** Outcome-only cause. Can improve precision; decrease variance. **(b)** Treatment-only cause. Can decrease precision; introduce variance.

Examples: Pre-Treatment Variable that is Bad Control

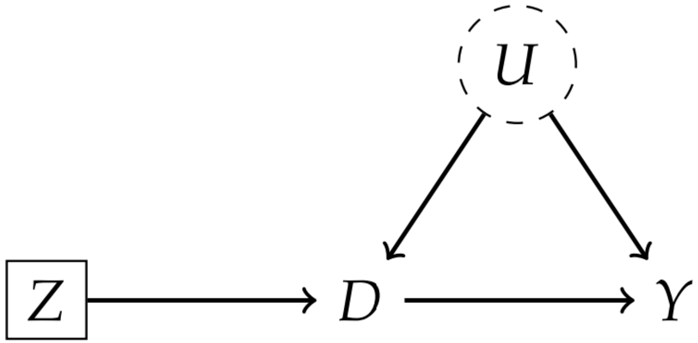


Figure 8.8: Bad control. Bias amplification by adjusting for an *instrument*. Treatment-only cause (*instrument*) that can amplify unobserved confounding bias.

Example: Pre-Treatment Variable that is Bad Control

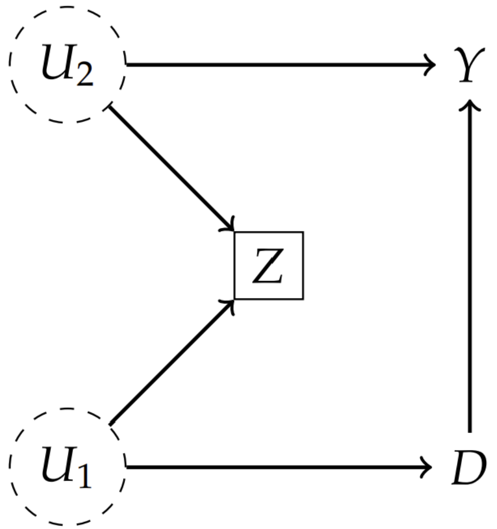
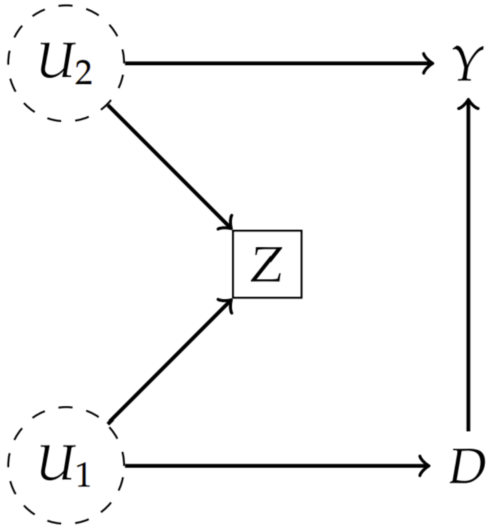


Figure 8.9: Bad control. M-Bias. Pre-treatment variable that introduces Heckman selection bias between two un-correlated unobserved causes.

Example: Homophily Bias in Peer Effects



- Peer effects on civic engagement level
- Consider samples of pairs of friends
- D is civic engagement of one friend at time t
- Y is civic engagement of other friend at time t+1
- 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

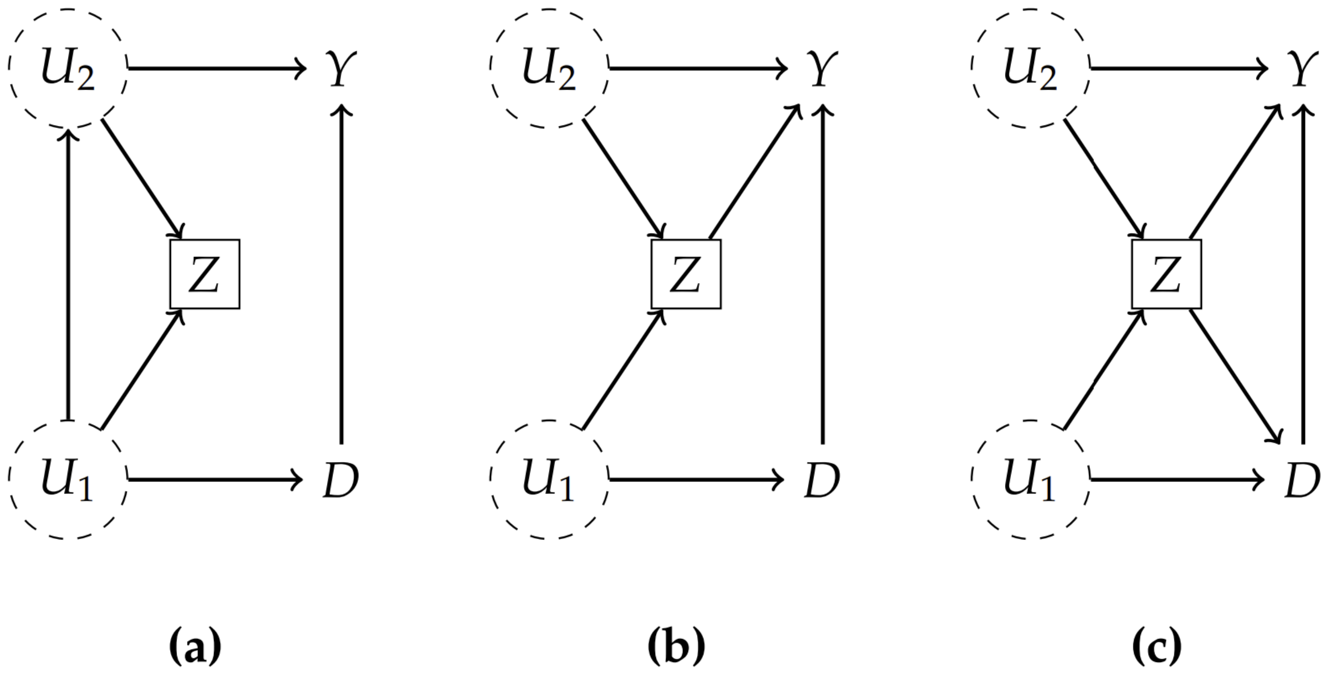


Figure 8.10: Unclear controls: **(a)** M-bias with correlated unobserved factors, **(b)** M-Bias with confounding. Pre-treatment variable that introduces Heckman selection between two un-correlated unobserved causes, but also is a confounder itself. No solution is perfect. **(c)** Butterfly Bias. M-bias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

“Post-Treatment Variables”

Example: Bad Controls, Mediation Bias

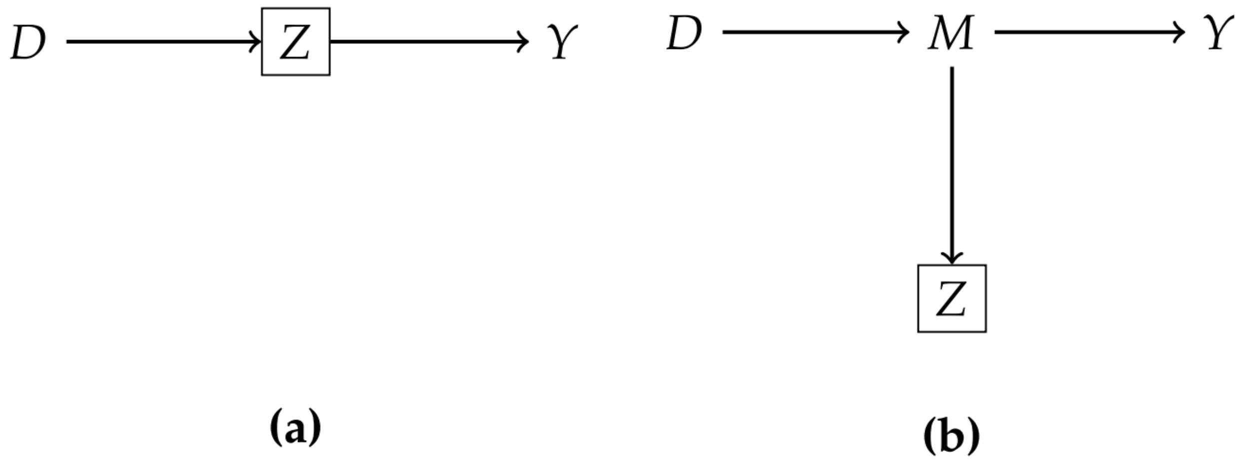


Figure 8.11: Bad controls: **(a)** over-control bias, by controlling on a mediator, **(b)** over-control bias, by controlling on some outcome caused by a mediator.

Example: Exception to Mediation Bias

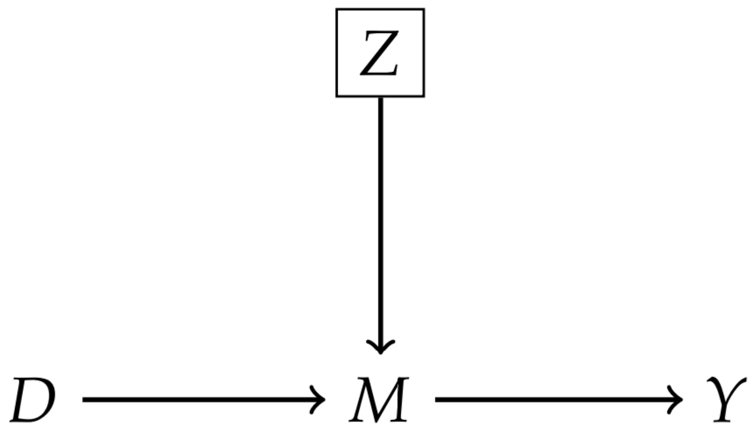


Figure 8.12: Neutral control. Cause of a mediator. Can potentially improve precision.

Example: Controlled Direct Effect Gone Wrong

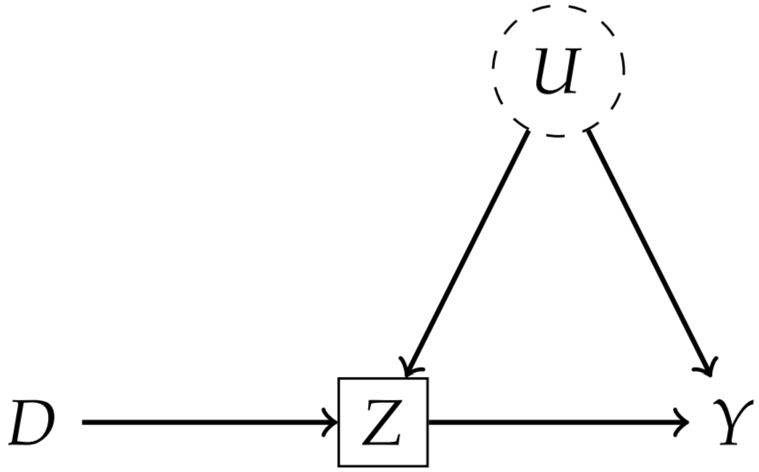


Figure 8.13: Bad control even for *controlled direct effect*. Confounded mediator bias.

Example: Bad Controls, Collider (Heckman Selection) bias

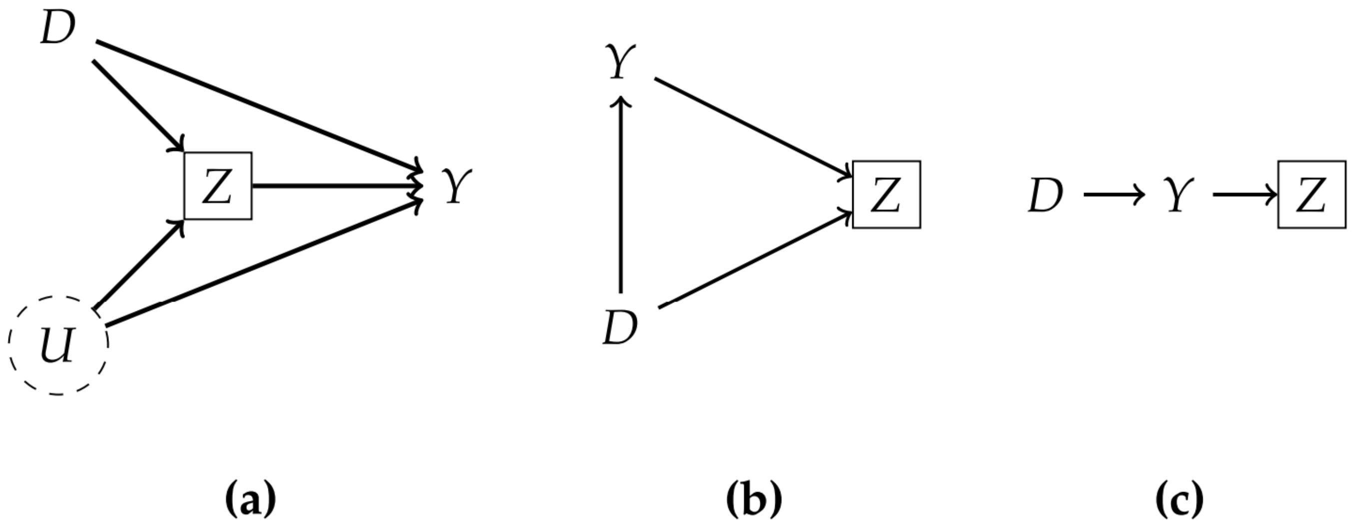


Figure 8.14: Bad controls: **(a)** collider stratification bias (e.g. low birth-weight "paradox" example), **(b)** collider stratification bias, **(c)** case control bias, by controlling on another outcome of the outcome of interest.

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