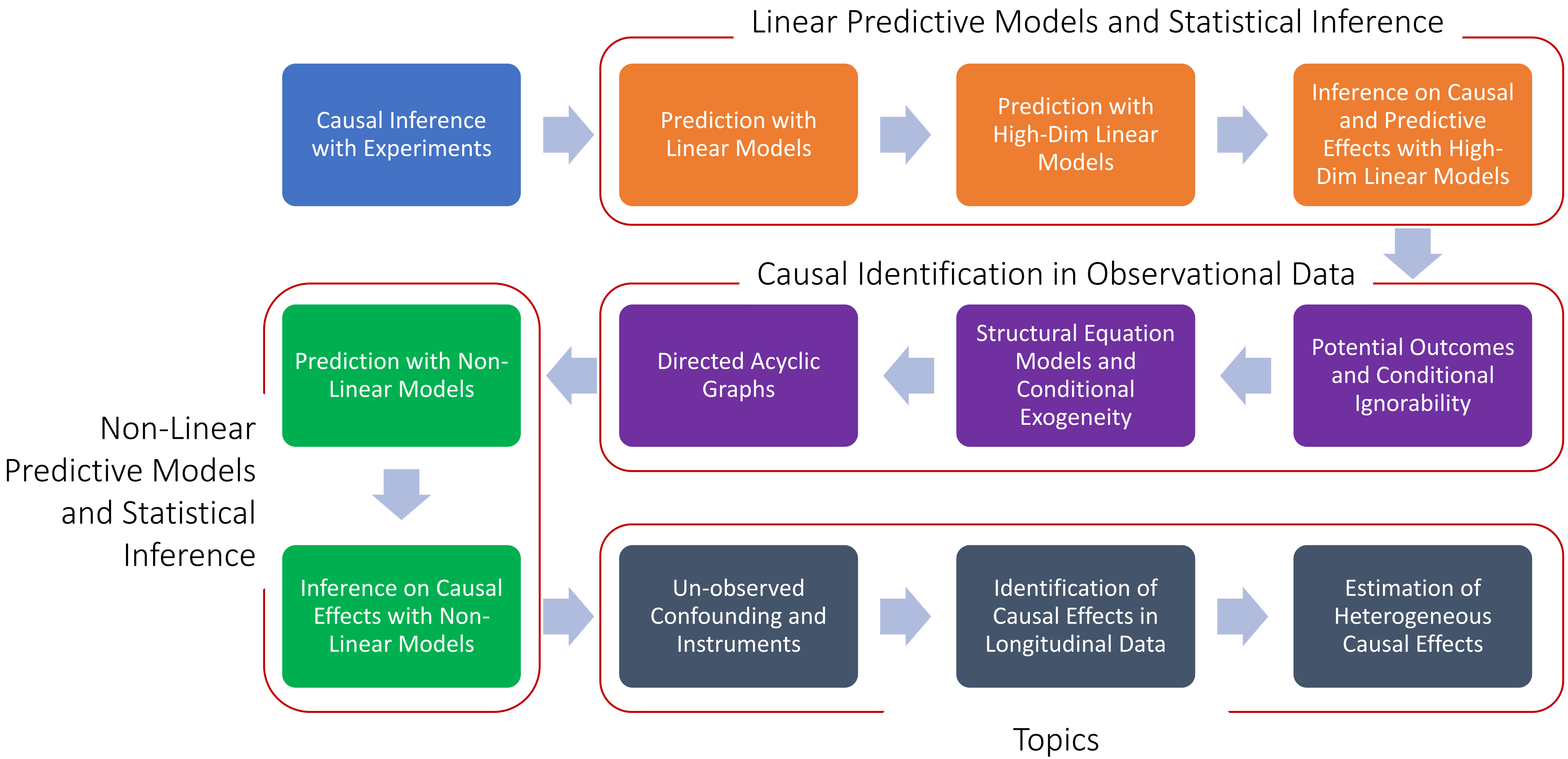
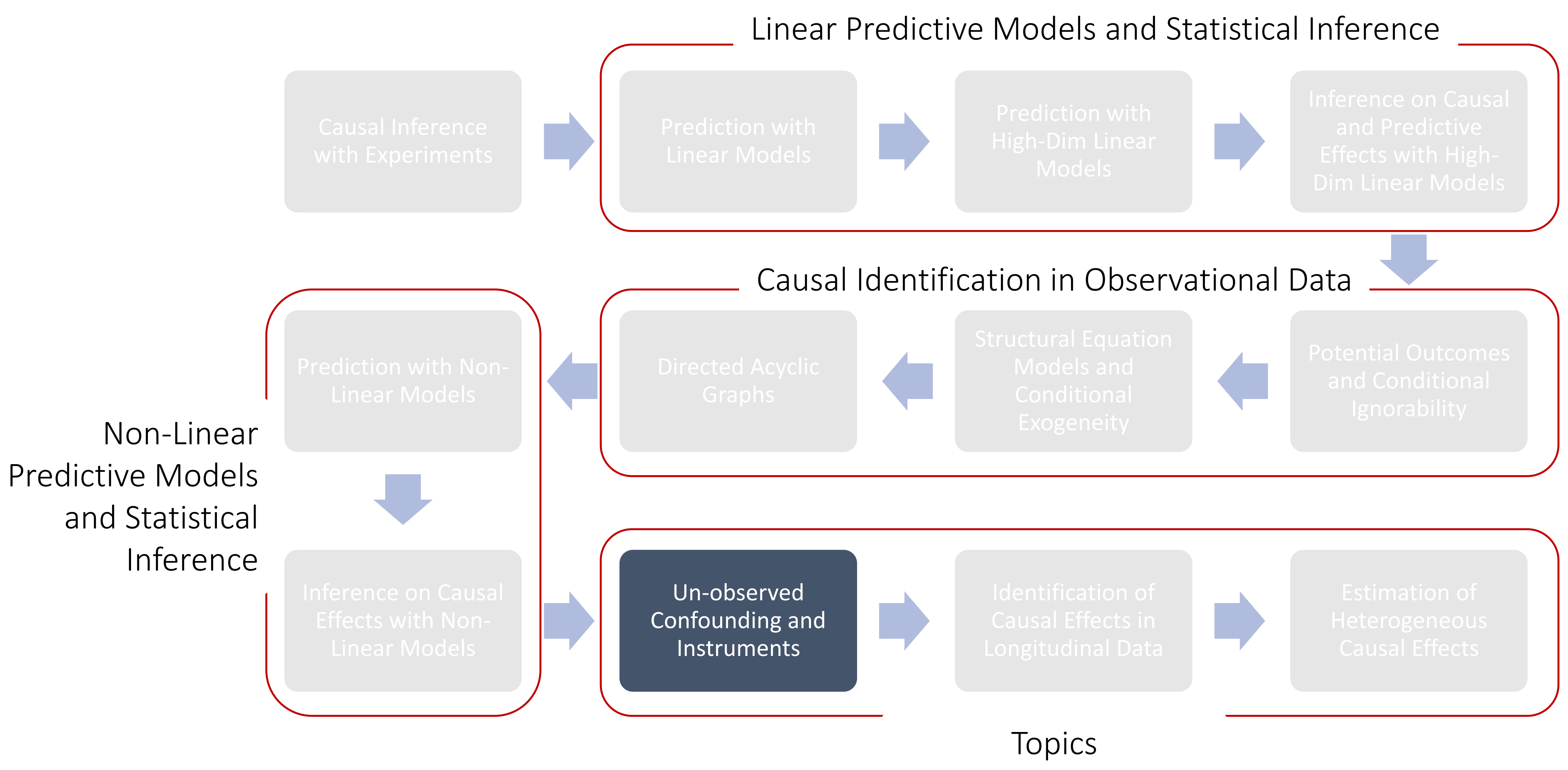
MS&E 228: Unobserved Confounding

Vasilis Syrgkanis

MS&E, Stanford

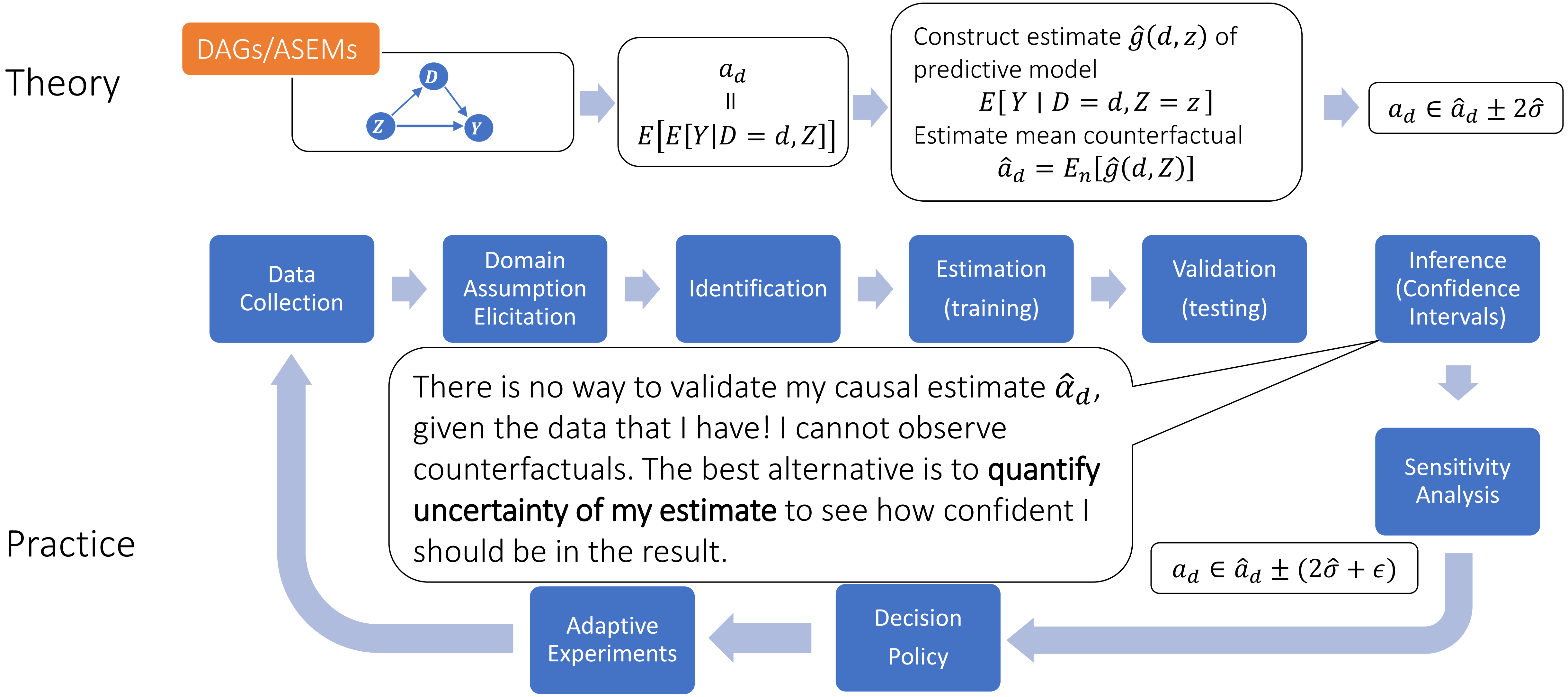

Linear Predictive Models and Statistical Inference
Causal Inference with Experiments
Prediction with Linear Models
Prediction with High-Dim Linear Models
Inference on Causal and Predictive Effects with High-Dim Linear Models
Causal Identification in Observational Data
Potential Outcomes and Conditional Ignorability
Structural Equation Models and Conditional Exogeneity
Directed Acyclic Graphs
Non-Linear Predictive Models and Statistical Inference
Prediction with Non-Linear Models
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
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

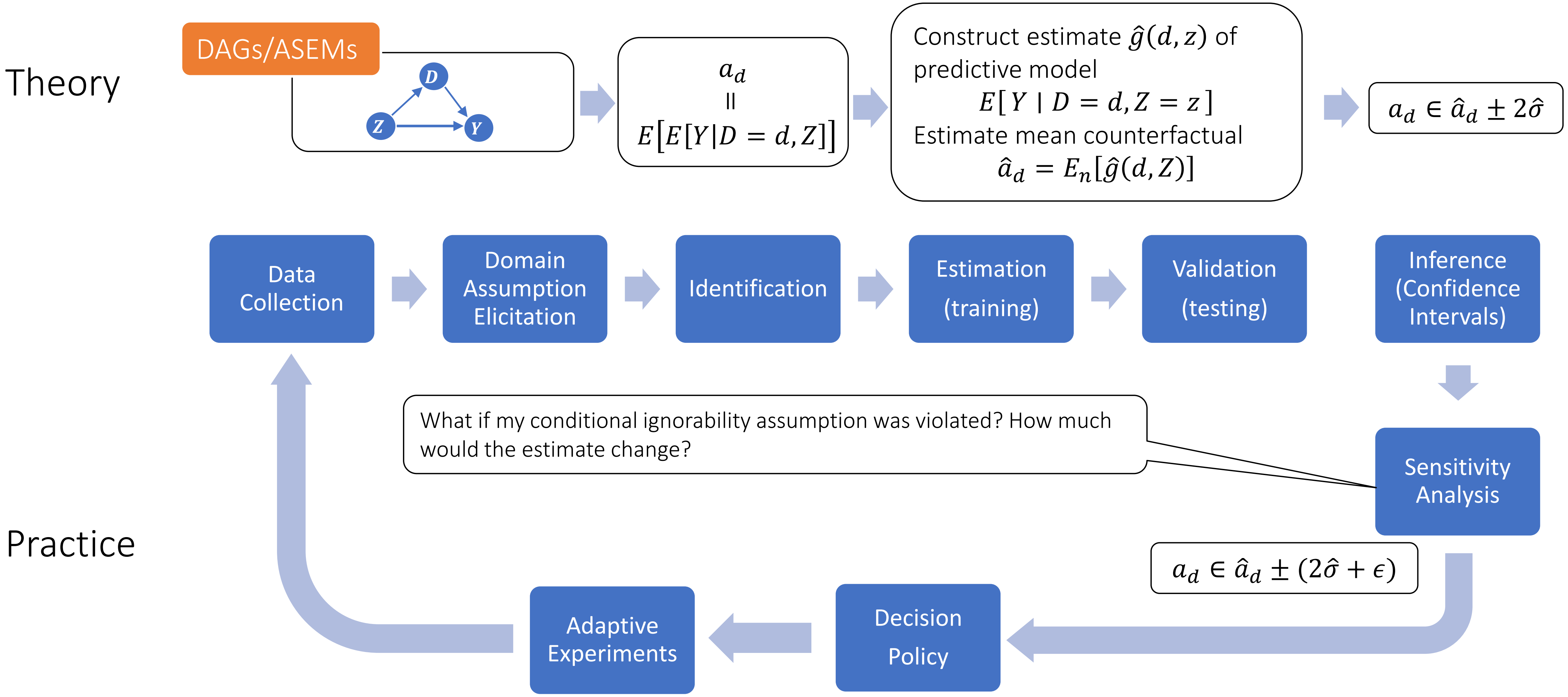
Goals for Today

- What can we do when we have un-observed confounding
- Omitted variable bias bounds
- Introduction to “Instruments”

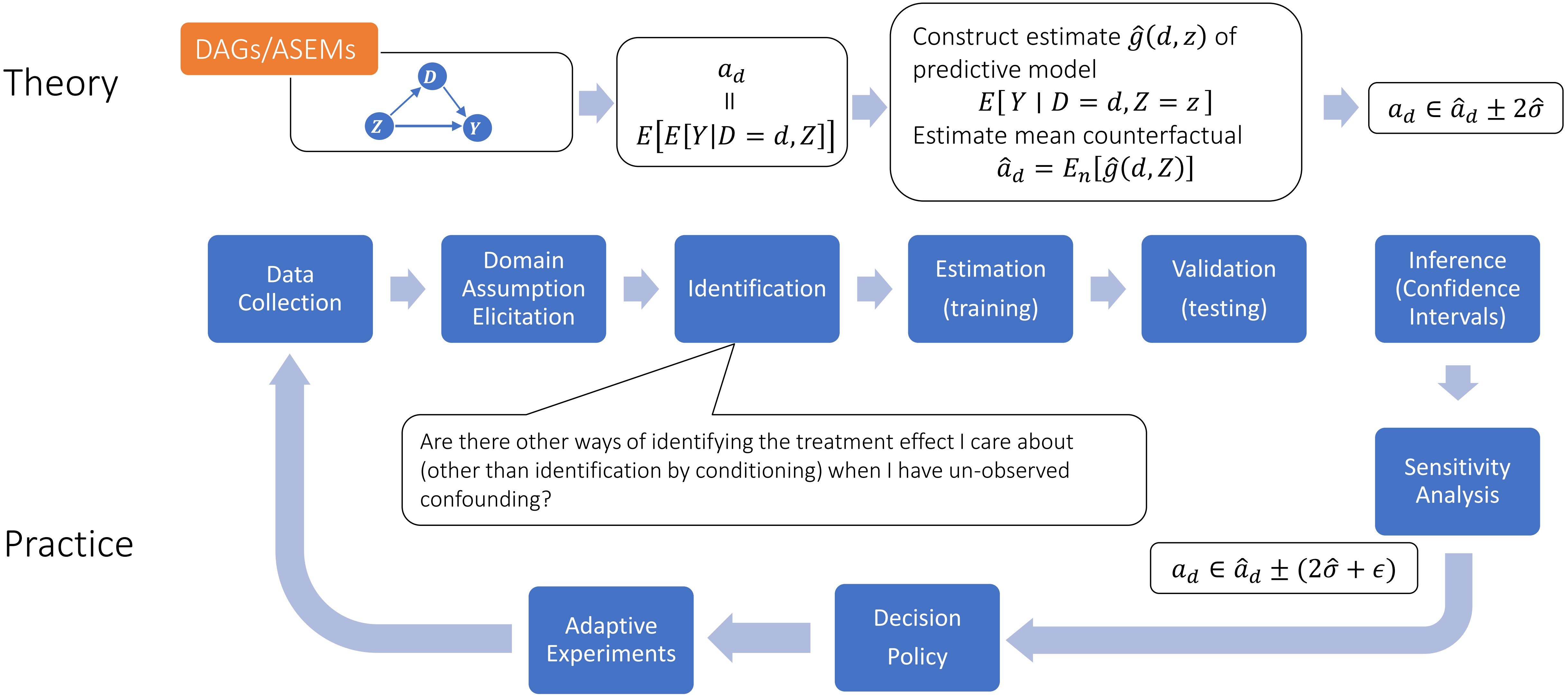
Causal Inference Pipeline



Causal Inference Pipeline

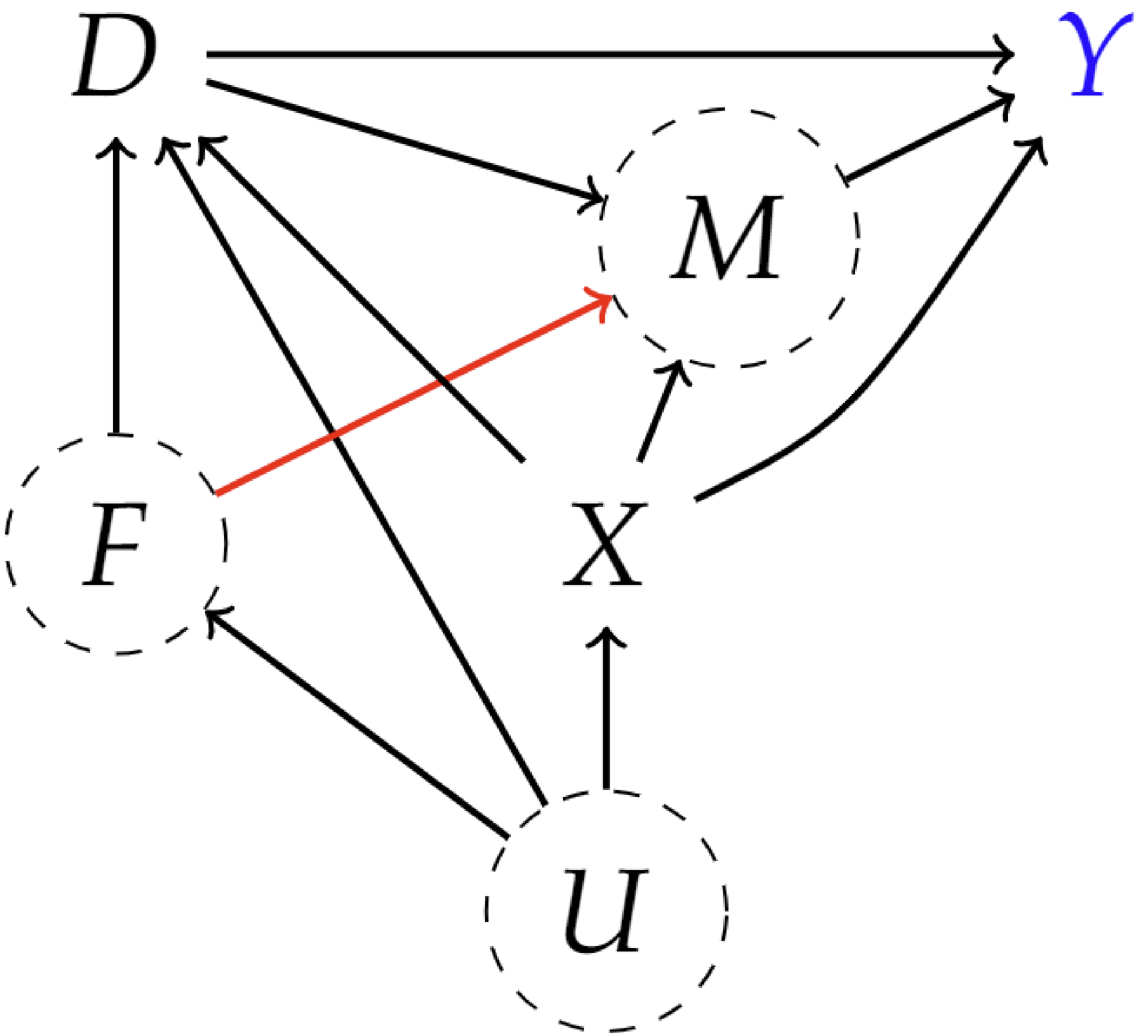


Causal Inference Pipeline



Possible Violations

- F: firm characteristics
- D: eligibility for 401k
- Y: net financial assets
- X: age, income, family size, years of education, a married indicator, a two-earner status indicator, a defined benefit pension status indicator, an IRA participation indicator, and a home ownership indicator
- M: match amount



Unobserved Confounding: Overview of Today’s Content

Unobserved Confounding

- If there are confounding factors that are un-observed, we cannot distinguish between an effect and a spurious correlation, without further structure

$$D \longrightarrow Y$$

Figure 4.1: D causes Y

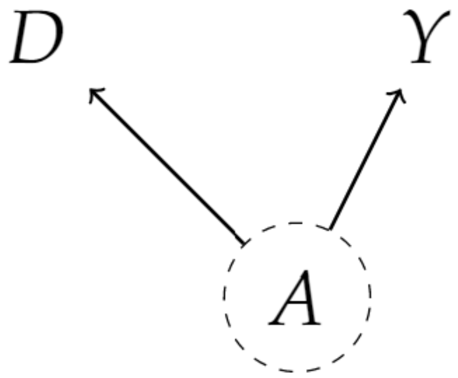


Figure 4.2: D and Y are caused by a latent factor A

Unobserved Confounding

- If there are confounding factors that are un-observed, we cannot distinguish between an effect and a spurious correlation, without further structure

$$D \longrightarrow Y$$

Figure 4.1: D causes Y

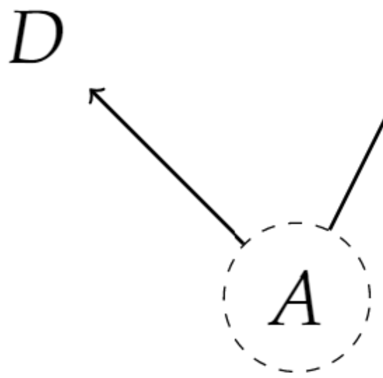


Figure 4.2: D and Y are caused by a latent factor A

One Type of Structure: Instruments

- We already saw one example of more structure in the assignments (which one?)

- The structure we will investigate today are instruments

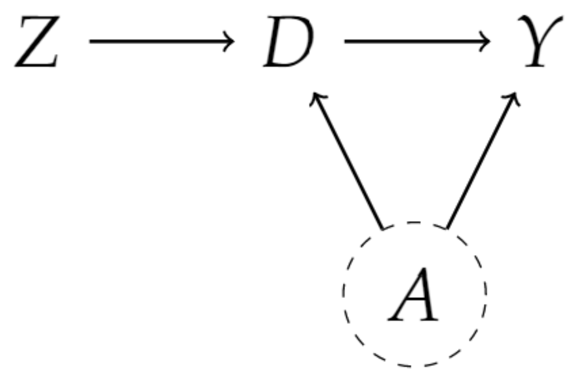


Figure 4.3: A DAG with Latent Confounder A and Instrument Z.

Bias Bounds

Very Frequently Unobserved Confounders

- Average treatment effect not “identifiable”

$$\theta_0 \neq E\big[E[Y|T=1,X] - E[\,Y \mid T=0,X\,]\big]$$

- Realistic conditional exogeneity

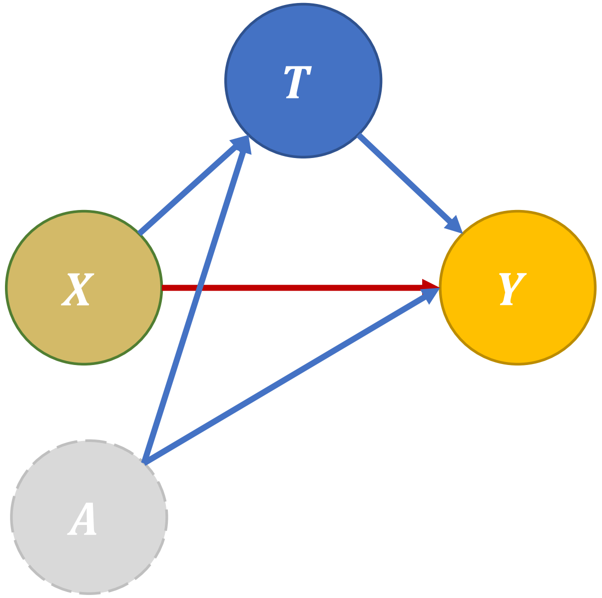
$$Y^{(d)} \perp T \mid X, A \qquad \text{(conditional exogeneity)}$$

- Ideal quantity: hypothetical g-formula

$$\theta_0 = E\big[E[Y|T=1,X,A] - E[\,Y \mid T=0,X,A\,]\big]$$

- Identifiable quantity:

$$\theta_s = E\big[E[Y|T=1,X] - E[\,Y \mid T=0,X\,]\big]$$



Omitted Variable Bias

- We want to estimate:

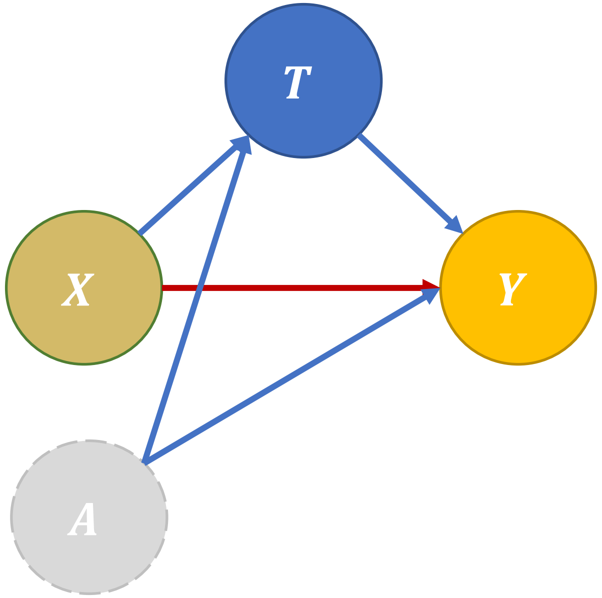
$$\theta_0 \coloneqq E[g(1,X,A) - g(0,X,A)]$$
$$g(T,X,A) \coloneqq E[\,Y \mid T,X,A\,]$$

- Which depends on an un-attainable “long regression”
- We can only estimate a “short regression”

$$g_s(T,X) \coloneqq E[Y|T,X] = E[g(T,X,A)|T,X]$$

- And compute “short estimate”

$$\theta_s = E[g_s(1,X) - g_s(0,X)]$$



Omitted Variable Bias Bounds

- Provide expression and construct bounds on Omitted Variable Bias (OMVB)
$$\theta_s - \theta_0$$
- Under interpretable assumptions that limit the strength of unobserved confounding
- Perform statistical inference on θ_0 allowing for ML regressions

- Sensitivity analysis has a long history:
  - Rosenbaum-Rubin'83: non-parametric bounds [non-sharp]
  - A lot of follow-up work making parametric assumptions

Paul R Rosenbaum and Donald B Rubin. 'Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome'. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1983), pp. 212–218 (cited on page 9).

James M Robins. 'Association, causation, and marginal structural models'. In: *Synthese* 121.1 (1999), pp. 151–179 (cited on page 9).

Kenneth A Frank. 'Impact of a confounding variable on a regression coefficient'. In: *Sociological Methods & Research* 29.2 (2000), pp. 147–194 (cited on page 9).

Paul R Rosenbaum. 'Observational studies'. In: *Observational studies*. Springer, 2002, pp. 1–17 (cited on page 9).

Guido W Imbens. 'Sensitivity to exogeneity assumptions in program evaluation'. In: *American Economic Review* 93.2 (2003), pp. 126–132 (cited on pages 9, 16).

Babette A Brumback et al. 'Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures'. In: *Statistics in medicine* 23.5 (2004), pp. 749–767 (cited on page 9).

Joseph G Altonji, Todd E Elder, and Christopher R Taber. 'Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools'. In: *Journal of political economy* 113.1 (2005), pp. 151–184 (cited on page 9).

Carrie A Hosman, Ben B Hansen, and Paul W Holland. 'The Sensitivity of Linear Regression Coefficients' Confidence Limits to the Omission of a Confounder'. In: *The Annals of Applied Statistics* (2010), pp. 849–870 (cited on page 9).

Kosuke Imai, Luke Keele, Teppei Yamamoto, et al. 'Identification, inference and sensitivity analysis for causal mediation effects'. In: *Statistical science* 25.1 (2010), pp. 51–71 (cited on page 9).

Tyler J. Vanderweele and Onyebuchi A. Arah. 'Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders.' In: *Epidemiology (Cambridge, Mass.)* 22.1 (Jan. 2011), pp. 42–52 (cited on page 9).

Matthew Blackwell. 'A selection bias approach to sensitivity analysis for causal effects'. In: *Political Analysis* 22.2 (2013), pp. 169–182 (cited on page 9).

Kenneth A Frank et al. ‘What would it take to change an inference? Using Rubin’s causal model to interpret the robustness of causal inferences’. In: *Educational Evaluation and Policy Analysis* 35.4 (2013), pp. 437–460 (cited on page 9).

Nicole Bohme Carnegie, Masataka Harada, and Jennifer L Hill. ‘Assessing sensitivity to unmeasured confounding using a simulated potential confounder’. In: *Journal of Research on Educational Effectiveness* 9.3 (2016), pp. 395–420 (cited on page 9).

Vincent Dorie et al. ‘A flexible, interpretable framework for assessing sensitivity to unmeasured confounding’. In: *Statistics in medicine* 35.20 (2016), pp. 3453–3470 (cited on page 9).

Joel A Middleton et al. ‘Bias amplification and bias unmasking’. In: *Political Analysis* 24.3 (2016), pp. 307–323 (cited on page 9).

Emily Oster. ‘Unobservable selection and coefficient stability: Theory and evidence’. In: *Journal of Business & Economic Statistics* (2017), pp. 1–18 (cited on page 9).

Tyler J VanderWeele and Peng Ding. 'Sensitivity analysis in observational research: introducing the E-value'. In: *Annals of Internal Medicine* 167.4 (2017), pp. 268–274 (cited on page 9).

Nathan Kallus and Angela Zhou. 'Confounding-robust policy improvement'. In: *arXiv preprint arXiv:1805.08593* (2018) (cited on page 9).

Nathan Kallus, Xiaojie Mao, and Angela Zhou. 'Interval estimation of individual-level causal effects under unobserved confounding'. In: *The 22nd international conference on artificial intelligence and statistics*. PMLR. 2019, pp. 2281–2290 (cited on page 9).

Carlos Cinelli et al. 'Sensitivity Analysis of Linear Structural Causal Models'. In: *International Conference on Machine Learning* (2019) (cited on page 9).

AlexanderM Franks, Alexander D'Amour, and Avi Feller. 'Flexible Sensitivity Analysis for Observational Studies Without Observable Implications'. In: *Journal of the American Statistical Association* 115.532 (2020), pp. 1730–1746 (cited on page 9).

Carlos Cinelli and Chad Hazlett. 'Making sense of sensitivity: Extending omitted variable bias'. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1 (2020), pp. 39–67 (cited on pages 9, 16).

Carlos Cinelli and Chad Hazlett. 'An omitted variable bias framework for sensitivity analysis of instrumental variables'. In: *Work. Pap* (2020) (cited on page 9).

Matteo Bonvini and Edward H Kennedy. 'Sensitivity analysis via the proportion of unmeasured confounding'. In: *Journal of the American Statistical Association* (2021), pp. 1–11 (cited on page 9).

Daniel O Scharfstein et al. 'Semiparametric Sensitivity Analysis: Unmeasured Confounding In Observational Studies'. In: *arXiv preprint arXiv:2104.08300* (2021) (cited on page 9).

Andrew Jesson et al. 'Quantifying Ignorance in Individual-Level Causal-Effect Estimates under Hidden Confounding'. In: *arXiv preprint arXiv:2103.04850* (2021) (cited on page 9).

Omitted Variable Bias: Partially Linear Models

- Let's consider a simpler structural equation model

$$
\begin{aligned}
Y &:= \alpha D + \delta A + f_Y(X) + \epsilon_Y \\
D &:= \gamma A + f_D(X) + \epsilon_D \\
A &:= f_A(X) + \epsilon_A \\
X &:= \epsilon_X
\end{aligned}
$$

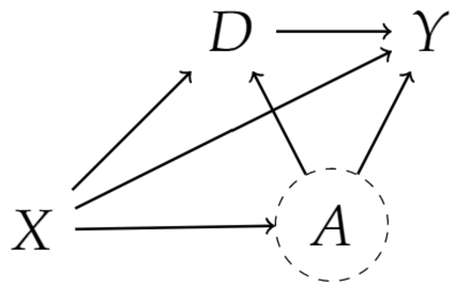


Figure 4.5: X are observed confounders, and A are unobserved confounders.

Omitted Variable Bias: Partially Linear Models

- Suppose we run the residual-on-residual process and first partial out X
- For convenience $E\left[\epsilon_A^2\right] = 1$

$$\tilde{Y} := a\tilde{D} + \delta\tilde{A} + \epsilon_Y$$
$$\tilde{D} := \gamma\tilde{A} + \epsilon_D$$
$$\tilde{A} := \epsilon_A$$

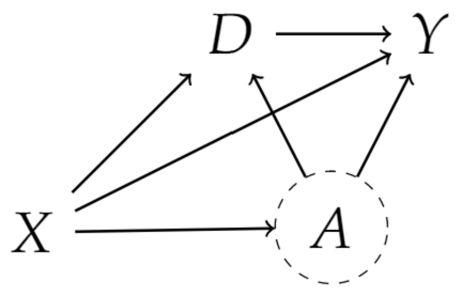


Figure 4.5: X are observed confounders, and A are unobserved confounders.

- Then the double ML method would run OLS of $\tilde{Y}$ on $\tilde{D}$

$$\theta_s = \frac{E\left[\tilde{Y}\tilde{D}\right]}{E\left[\tilde{D}^2\right]} = \frac{E\left[a\,\tilde{D}^2 + \delta\tilde{A}\tilde{D}\right]}{E\left[\tilde{D}^2\right]} = a + \delta\frac{E\left[\gamma\tilde{A}^2\right]}{E\left[\tilde{D}^2\right]} = a + \frac{\delta\gamma}{\gamma^2 + E\left[\epsilon_D^2\right]}$$

Bias Bounds

- If the analyst can provide bounds on the strength of each relationship (e.g. $\delta, \gamma, E\left[\epsilon_D^2\right]$) then we can provide bounds on the true effect

$$\theta_0 = \theta_s \pm \left|\frac{\delta\gamma}{\gamma^2 + E\left[\epsilon_D^2\right]}\right|$$

- More interpretable bounds:

$$\left(\frac{\delta\gamma}{\gamma^2 + E\left[\epsilon_D^2\right]}\right)^2 = \boxed{R^2_{\tilde{Y}\sim\tilde{A}|\tilde{D}}}\frac{R^2_{\tilde{D}\sim\tilde{A}}}{1-\boxed{R^2_{\tilde{D}\sim\tilde{A}}}}\boxed{\frac{\mathrm{E}\left[\left(\tilde{Y}-\theta_s\tilde{D}\right)^2\right]}{E\left[\tilde{D}^2\right]}}$$

Measurable from the data

R^2 in linear regression of $\tilde{Y}$ on $\tilde{A}$ after linearly partialling out $\tilde{D}$

R^2 in linear regression of $\tilde{D}$ on $\tilde{A}$

Bias Bounds

- If the analyst can provide bounds on the strength of each relationship (e.g. $\delta, \gamma, E\left[\epsilon_D^2\right]$) then we can provide bounds on the true effect

$$\theta_0 = \theta_s \pm \left|\frac{\delta\gamma}{\gamma^2 + E\left[\epsilon_D^2\right]}\right|$$

- More interpretable bounds:

$$\left(\frac{\delta\gamma}{\gamma^2 + E\left[\epsilon_D^2\right]}\right)^2 = \boxed{R^2_{Y\sim A|D,X}} \frac{R^2_{D\sim A|X}}{1 - \boxed{R^2_{D\sim A|X}}} \boxed{\frac{\mathrm{E}\left[\left(\tilde{\mathrm{Y}} - \theta_s\tilde{D}\right)^2\right]}{E\left[\tilde{D}^2\right]}}$$

Measurable from the data

Reduction in unexplained variance of Y when adding A in the model that predicts Y from treatment and controls

Reduction in unexplained variance of D when adding A in the model that predicts D from controls

For more details:
Making Sense of Sensitivity: Extending Omitted Variable Bias

For more general analysis see:
Long Story Short: Omitted Variable Bias in Causal Machine Learning

Bias Bounds

- The analyst provides bounds on the partial R^2

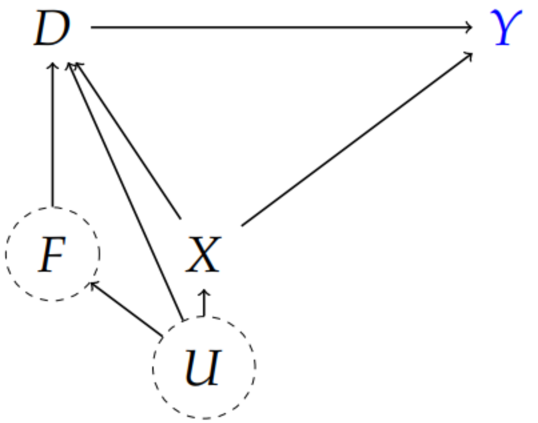
$$R^2_{Y\sim A|D,X} \leq C_Y^2, \qquad R^2_{D\sim A|X} \leq C_D^2$$

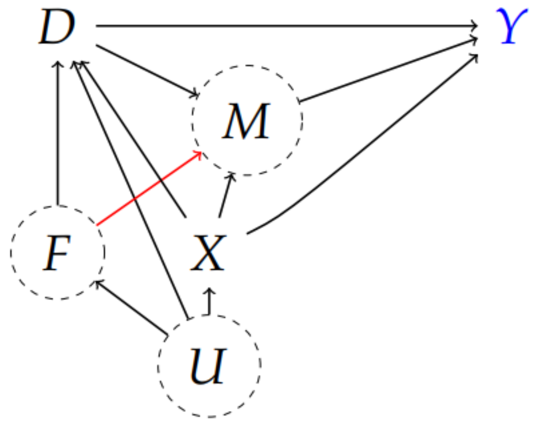
- Based on these bounds we can conclude that

$$\theta_0 \in \theta_s \pm \sqrt{C_Y^2 \frac{C_D^2}{1-C_D} \frac{E\left[\left(\tilde{Y} - \theta_s \widetilde{D}\right)^2\right]}{E\left[\widetilde{D}^2\right]}}$$

Application: 401k eligibility

- Y=net financial assets
- D=eligibility to enroll in 401(k) program
- X=pre-treatment worker-level covariates (observed)
- F=pre-treatment firm-level covariates (unobserved)
- M=amount of contribution matched by employer
- U=general latent factors





Controlling for X is sufficient in top figure but not bottor

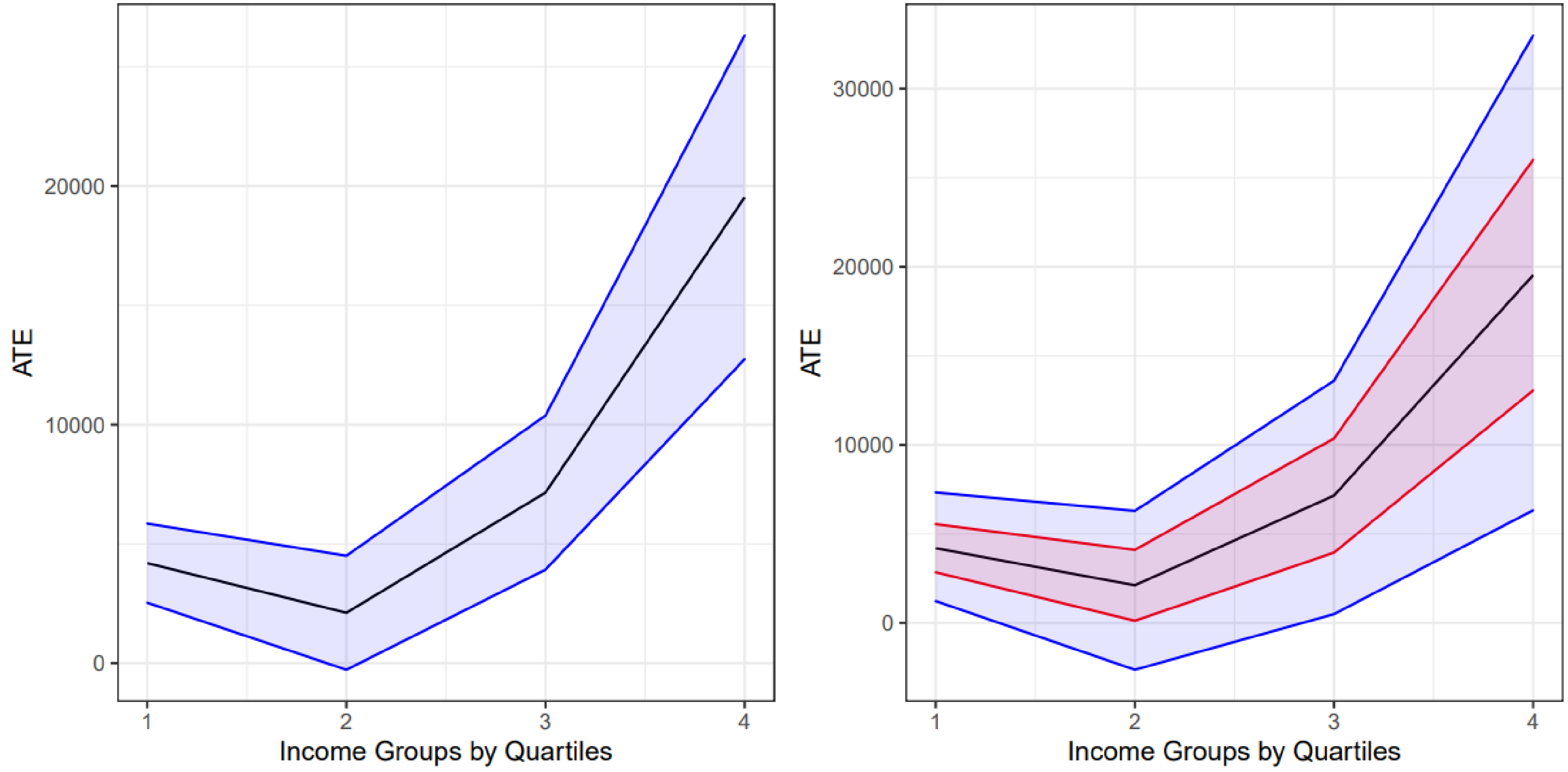
Confounding Scenario

- Posit that F explains as much variation in net financial assets as the total variation of maximal matched percentage (5%) of income over period of three years
- Posit that F explains an additional 2.5% of the variation in 401k eligibility, a 20% relative increase in the baseline R^2 of the treatment of 13%

- In PLR: translates to $\mathrm{C}_{\mathrm{Y}}^2 \coloneqq R_{Y\sim A|D,X}^2 \approx 4\%, C_D^2 \coloneqq R_{D\sim A|X}^2 \approx 3\%$

Robustness value (RV) = minimal equal strength of C_Y^2 and C_D^2 s.t. bound includes zero

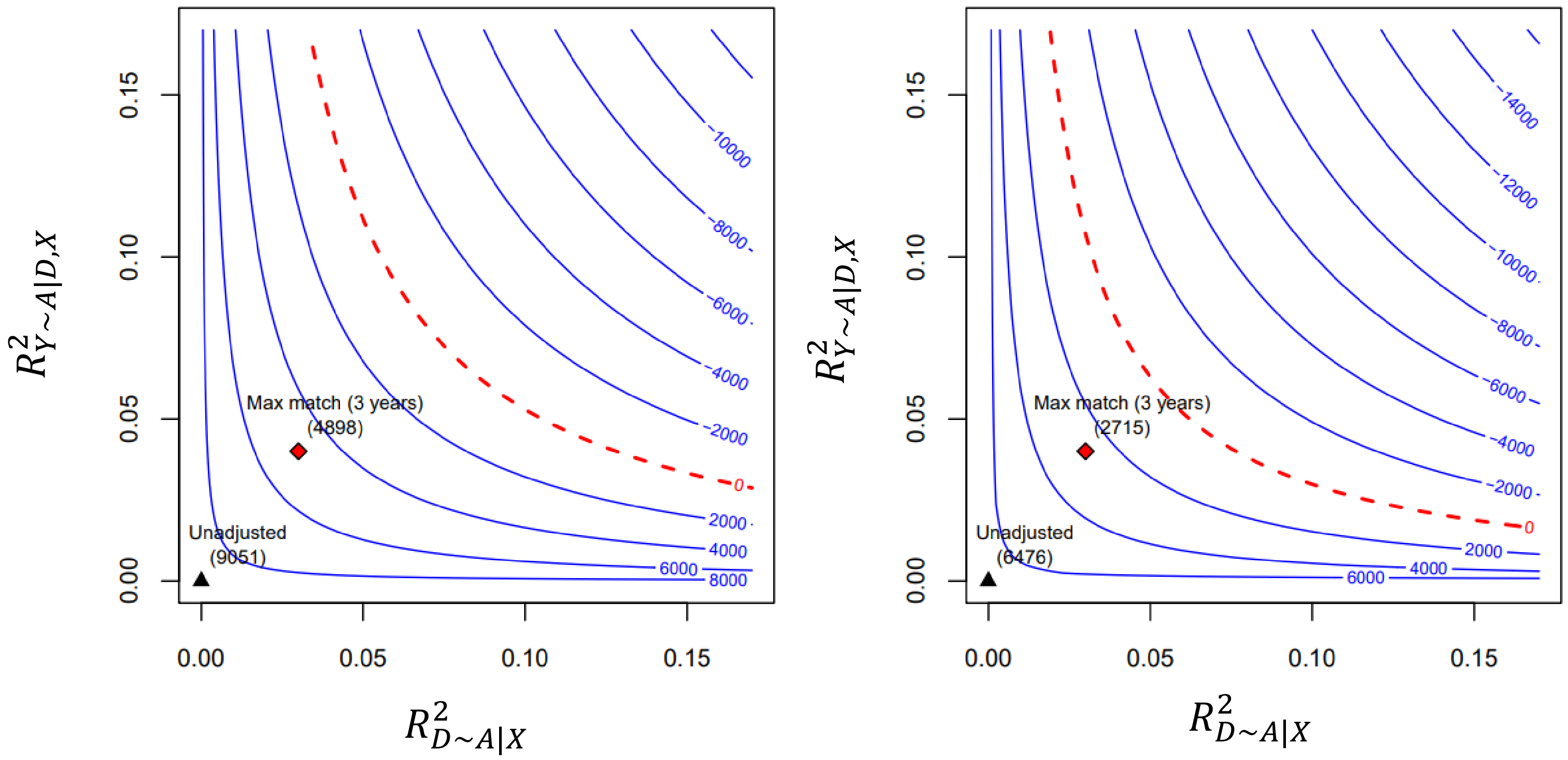
- RV=5.5% (at 95% significance level) > 4%,3%
- Finding that 401k eligibility has positive effect is robust to this confounding scenario



(a) Estimates under no confounding.

(b) Bounds under posited confounding.

Note: Estimate (black), bounds (red), and confidence bounds (blue) for the ATE. Confounding scenario: $\rho^2 = 1$; $C_Y^2 \approx 0.04$; $C_D^2 \approx 0.031$. Significance level of 5%.



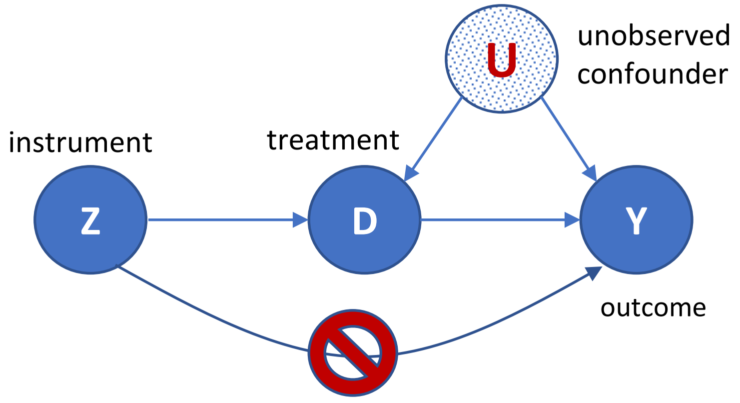
(a) Contours for $\theta_{-} = \theta_s - |B|$.

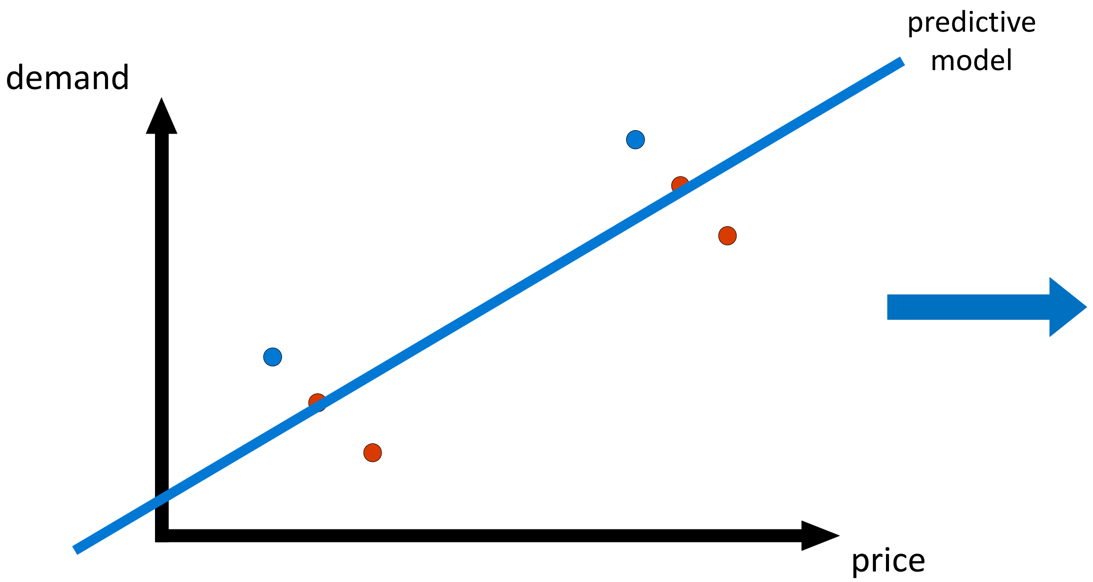
(b) Contours lower limit confidence bound.

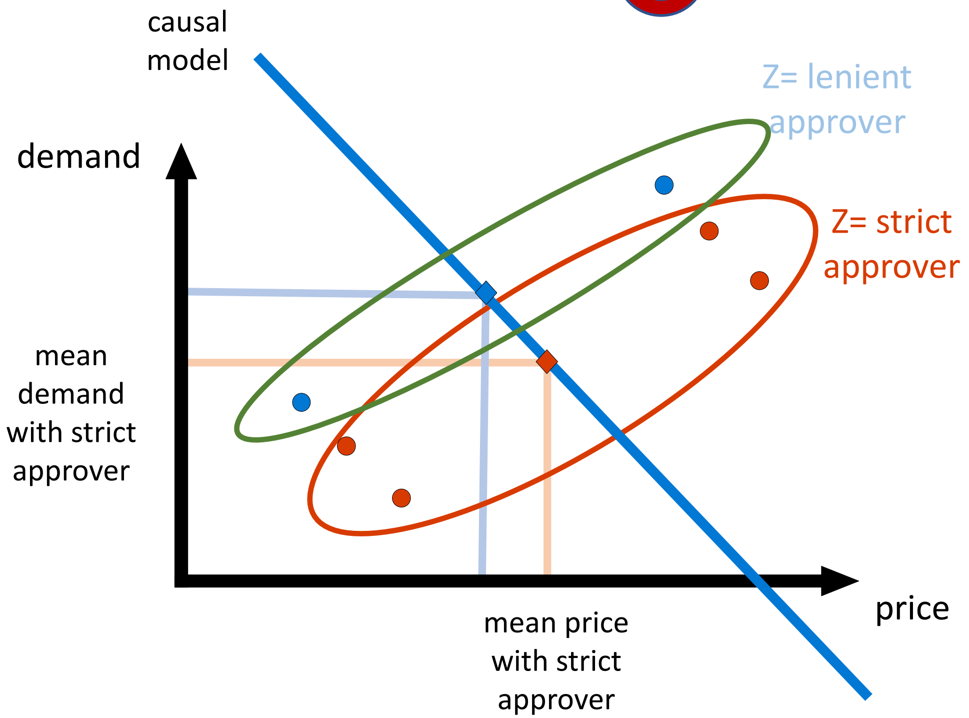
Can we recover the true effect?
Instrumental Variables

Instrumental Variables and 2SLS

Instrumental Variable: any random variable Z that affects the treatment (log-price) D but does not affect the outcome (log-demand) Y other than through the treatment [Wright'28, Bowden-Turkington'90, Angrist-Krueger'91, Imbens-Angrist'94]



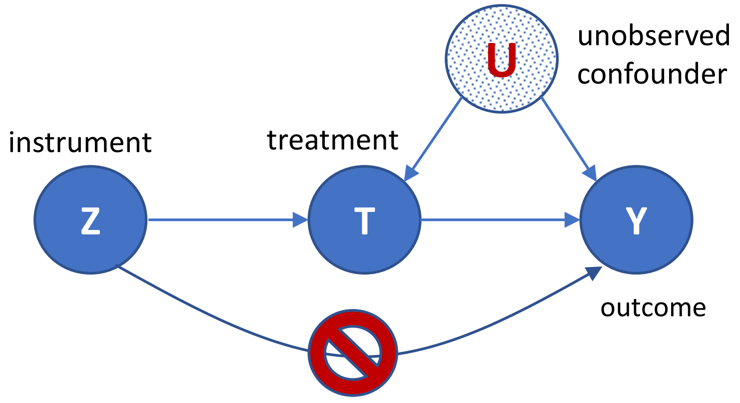




Instrumental Variables and 2SLS

Instruments are widely used

- In the discount example (see also [Kling AER06] for effects of incarceration)
  - Discounts are sent to an approver desk
  - Approver assignment is random and different approvers are more or less "lenient"
  - Approver leniency is an instrument

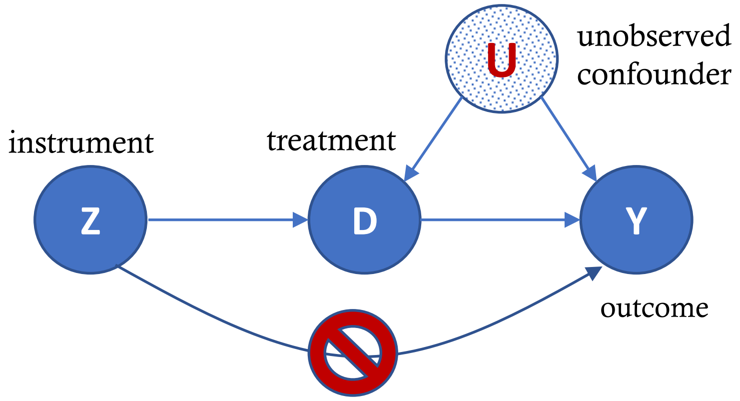


- In healthcare [Doyle et al., JPE15]
  - Random assignment to ambulance companies of nearby patients is an instrument for measuring hospital quality

- In Tech [S., NeurIPS19]
  - Recommendation A/B tests as instruments for the effects of downstream actions

Identification of Causal Effects via Instruments

Phillip Wright's idea (1928): the first causal path diagram analysis



- We can estimate effect of Z on y via a regression

$$\gamma = \frac{\mathbb{E}[\tilde{Z}\tilde{y}]}{\mathbb{E}[\tilde{Z}^2]}$$

- We can estimate the effect of Z on D via a regression

$$\delta = \frac{\mathbb{E}[\tilde{\mathrm{Z}}\tilde{\mathrm{D}}]}{\mathbb{E}[\tilde{\mathrm{Z}}^2]}$$

- The effect of Z on Y (γ) is the product of the effect of Z on T (δ) multiplied by the effect of T on y (θ)

$$\theta = \frac{\gamma}{\delta} = \frac{\mathbb{E}[\tilde{Z}\tilde{y}]}{\mathbb{E}[\tilde{Z}\tilde{D}]}$$

Partially Linear Instrumental Variable Model

- Typically for continuous treatment/instrument a partially linear structural equation assumed

$$
\begin{gathered}
Y \coloneqq \theta_0 D + f_Y(X) + \delta A + \epsilon_Y \\
D \coloneqq \beta Z + f_D(X) + \gamma A + \epsilon_D \\
Z \coloneqq f_Z(X) + \epsilon_Z \\
A \coloneqq f_A(X) + \epsilon_A
\end{gathered}
$$

All errors are exogenous and un-correlated

Partially Linear Instrumental Variable Model

- After partialling out the observed controls X

$$
\begin{aligned}
\tilde{Y} &:= \theta_0 \widetilde{D} + \delta \tilde{A} + \epsilon_Y \\
\widetilde{D} &:= \beta \tilde{Z} + \gamma \tilde{A} + \epsilon_D \\
\tilde{Z} &:= \epsilon_Z \\
\tilde{A} &:= \epsilon_A
\end{aligned}
$$

- We see immediately that:

$$\tilde{Y} := \theta_0 \widetilde{D} + U, \qquad U := \delta \tilde{A} + \epsilon_Y \perp \tilde{Z}$$

- Since $\epsilon_A, \epsilon_Y, \epsilon_Z$ are un-correlated: $E\left[\left(\delta \tilde{A} + \epsilon_Y\right)\tilde{Z}\right] = 0$
- Thus we have the moment restriction: $E\left[\left(\tilde{Y} - \theta_0 \widetilde{D}\right)\tilde{Z}\right] = 0$

Partially Linear Instrumental Variable Model

- After partialling out the observed controls X

$$\tilde{Y} \coloneqq \theta_0 \widetilde{D} + \delta \tilde{A} + \epsilon_Y$$
$$\widetilde{D} \coloneqq \beta \tilde{Z} + \gamma \tilde{A} + \epsilon_D$$
$$\tilde{Z} \coloneqq \epsilon_Z$$
$$\tilde{A} \coloneqq \epsilon_A$$

- Thus we have the moment restriction: $E\left[\left(\tilde{Y} - \theta_0 \widetilde{D}\right)\tilde{Z}\right] = 0$
- We re-derive a generalization of Wright's formula

$$\theta_0 = \frac{E\left[\tilde{Y}\tilde{Z}\right]}{E\left[\widetilde{D}\tilde{Z}\right]}$$

Partially Linear Instrumental Variable Model

- After partialling out the observed controls X

$$\tilde{Y} := \theta_0 \widetilde{D} + \tilde{A} + \epsilon_Y$$
$$\widetilde{D} := \beta \tilde{Z} + \beta \tilde{A} + \epsilon_D$$
$$\tilde{Z} := \epsilon_Z$$
$$\tilde{A} := \epsilon_A$$

- Setting falls into the general moment estimation framework

$$M(\theta, h, p, m) = E\left[\left(Y - h(X) - \theta\left(D - p(X)\right)\right)\ \left(Z - m(X)\right)\right] = 0$$

- Where $h(X) = E[Y|X]$, $p(X) = E[D|X]$, $m(Z) = E[Z|X]$

Orthogonal Method: Double ML for IV

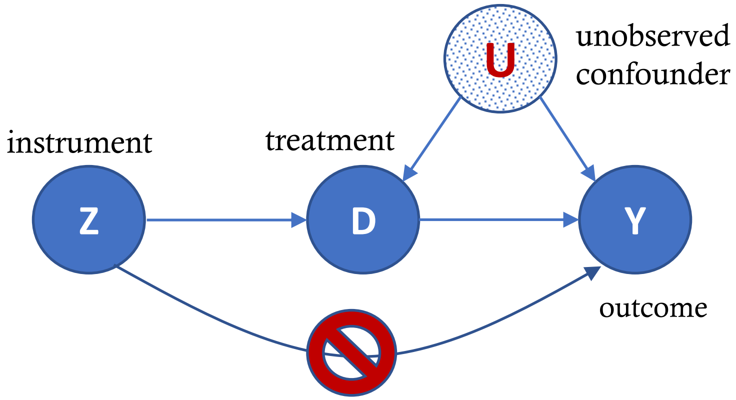
Double ML. Split samples in half

- Regress $Y \sim X$ with ML on first half, to get estimate $\hat{h}(S)$ of $E[Y|X]$
- Regress $D \sim X$ with ML on first half, to get estimate $\hat{p}(S)$ of $E[D|X]$
- Regress Z $\sim X$ with ML on first half, to get estimate $\widehat{m}(S)$ of $E[Z|X]$
- Construct residuals on other half, $\hat{Z} = Z - \widehat{m}(X), \widehat{D} \coloneqq D - \hat{p}(X)$ and $\hat{Y} \coloneqq Y - \hat{h}(X)$
- Solve moment condition:

$$E_n\left[\left(\hat{Y} - \theta\widehat{D}\right)\widehat{D}\right] = 0$$

```
from econml.iv.dml import OrthoIV
orthoiv = OrthoIV()
orthoiv.fit(y, D, Z, W=X).effect_inference()
```

The Binary Case



Imbens-Angrist (1994): core contribution of Nobel 2022 award

- Instrument/Treatment are binary (instrument=recommended treatment)
- Assume monotonicity: $D^{(1)} \geq D^{(0)}$
- Recommended treatment cannot reverse taken treatment
- Object of interest: Local Average Treatment Effect (ATE among compliers)

$$\theta_0 = E\left[\mathrm{Y}^{(1)} - \mathrm{Y}^{(0)} \middle| \mathrm{D}^{(1)} > \mathrm{D}^{(0)}\right]$$

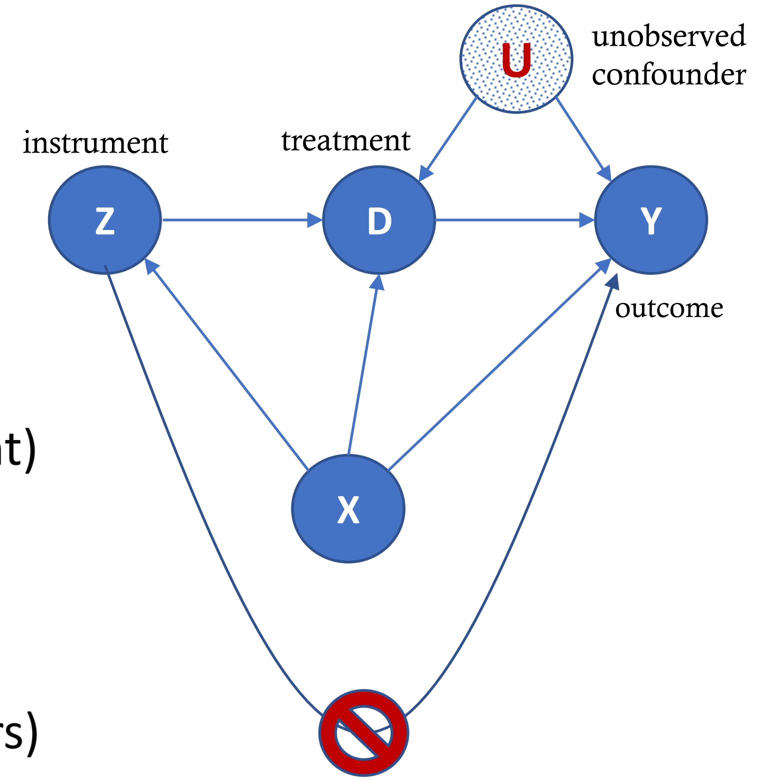
- Proof [Angrist-Imbens'94]:

$$\theta_0 = \frac{E\left[\left(Y^{(1)} - Y^{(0)}\right)1\left\{D^{(1)} > D^{(0)}\right\}\right]}{E\left[1\left\{D^{(1)} > D^{(0)}\right\}\right]} = \frac{E\left[Y^{(D(1))} - Y^{(D(0))}\right]}{E\left[D^{(1)} - D^{(0)}\right]} = \frac{\boxed{ATE(Z \to Y)}}{\boxed{ATE(Z \to D)}}$$

$$\boxed{\gamma = \frac{\mathbb{E}\left[\tilde{Z}\tilde{y}\right]}{\mathbb{E}\left[\tilde{Z}^2\right]}}$$

$$\boxed{\delta = \frac{\mathbb{E}\left[\tilde{\mathrm{Z}}\tilde{\mathrm{D}}\right]}{\mathbb{E}\left[\tilde{\mathrm{Z}}^2\right]}}$$

The Binary Case



Imbens-Angrist (1994): core contribution of Nobel 2022 award

- Instrument/Treatment are binary (instrument=recommended treatment)
- Assume monotonicity: $D^{(1)} \geq D^{(0)}$
- Recommended treatment cannot reverse taken treatment
- Object of interest: Local Average Treatment Effect (ATE among compliers)

$$\theta_0 = E\left[\mathrm{Y}^{(1)} - \mathrm{Y}^{(0)} \middle| \mathrm{D}^{(1)} > \mathrm{D}^{(0)}\right]$$

- Proof [Angrist-Imbens'94]:

$$E\left[E[Y|Z=1,X] - E[Y|Z=0,X]\right]$$

$$\theta_0 = \frac{E\left[\left(Y^{(1)} - Y^{(0)}\right)1\left\{D^{(1)} > D^{(0)}\right\}\right]}{E\left[1\left\{D^{(1)} > D^{(0)}\right\}\right]} = \frac{E\left[Y^{(D(1))} - Y^{(D(0))}\right]}{E\left[D^{(1)} - D^{(0)}\right]} = \frac{\boxed{ATE(Z \to Y)}}{\boxed{ATE(Z \to D)}}$$

$$E\left[E[D|Z=1,X] - E[D|Z=0,X]\right]$$

LATE in the Binary Case

- Under monotonicity

$$\theta_0 = \frac{E\big[E[\,Y \mid Z = 1, X\,] - E[Y|Z = 0, X]\big]}{E\big[E[D|Z = 1, X] - E[D|Z = 0, X]\big]}$$

- Moment formulation

$$E[E[Y|Z = 1, X] - E[Y|Z = 0, X] - \theta_0(E[D|Z = 1, X] - E[D|Z = 0, X])] = 0$$

$$+ \qquad\qquad\qquad\qquad +$$

$$\boldsymbol{a_{Z\to Y}(Z, X)(Y - E[Y|Z, X])} \qquad\qquad \boldsymbol{a_{Z\to D}(Z, X)(D - E[D|Z, X])}$$

- Orthogonal moment formulation: apply ATE debiasing twice