Sensitivity Analysis with the R^2-Calculus

Tobias Freidling   Qingyuan Zhao

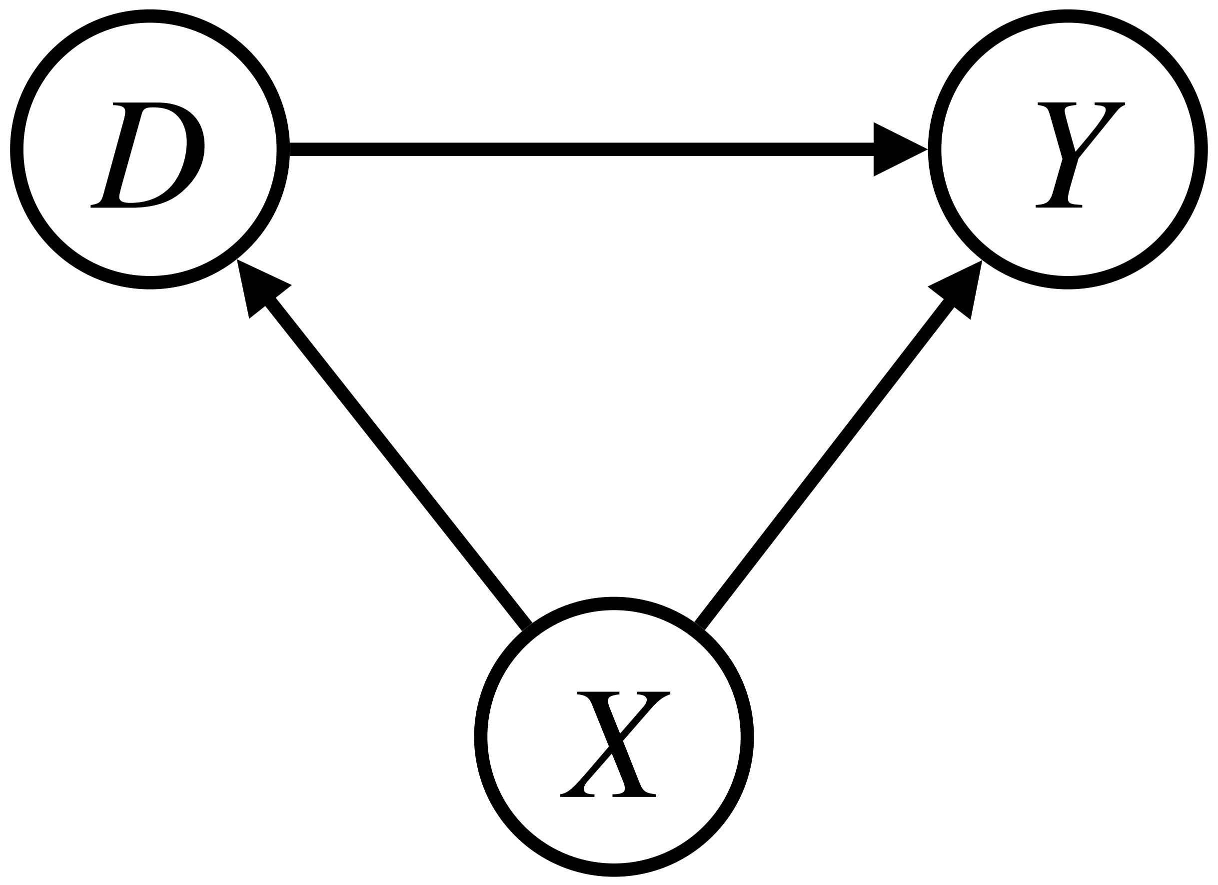
Statistical Laboratory, University of Cambridge

OIST, Machine Learning und Data Science Unit, July 2023

Causal Inference and DAGs

- Directed Acyclic Graph (DAG)
- Linear Structural Equation Model

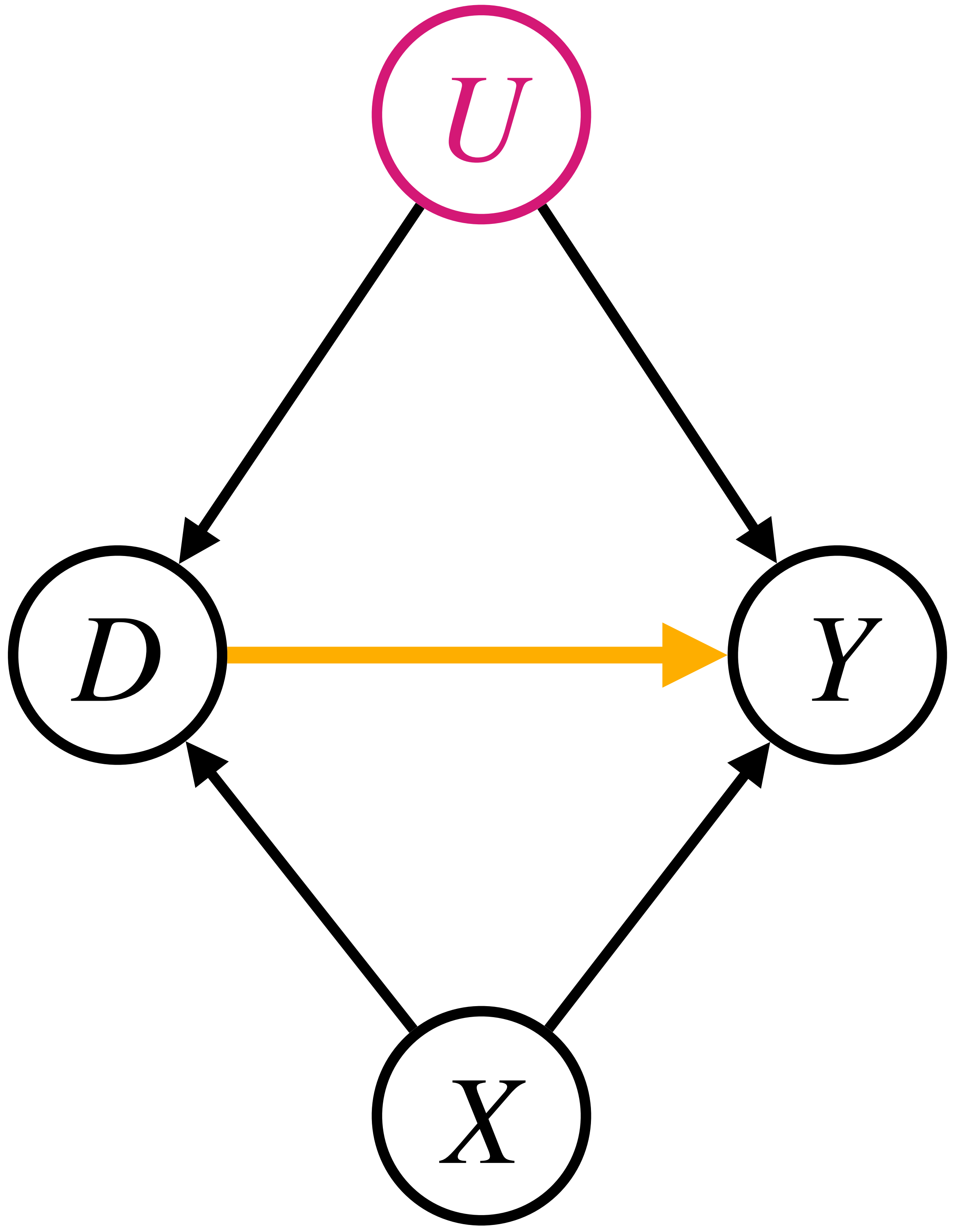
$$X \Leftarrow \varepsilon_X$$
$$D \Leftarrow \beta_{DX} X + \varepsilon_D$$
$$Y \Leftarrow \beta_{YD} D + \beta_{YX} X + \varepsilon_Y$$



- Goal: Estimate the causal effect of D on Y, i.e. estimate β_{YD} from data $(X_i, D_i, Y_i)_{i=1}^n$

Causal Inference and Sensitivity Analysis

- Goal: Estimate β_{YD}
- Assumptions:
  - Correction model specification, i.e. linearity
  - Correct DAG
  - No unmeasured confounders
- Sensitivity analysis explores how violations of assumptions affect estimation
- This work: focus on unmeasured confounders



Sensitivity Analysis - History

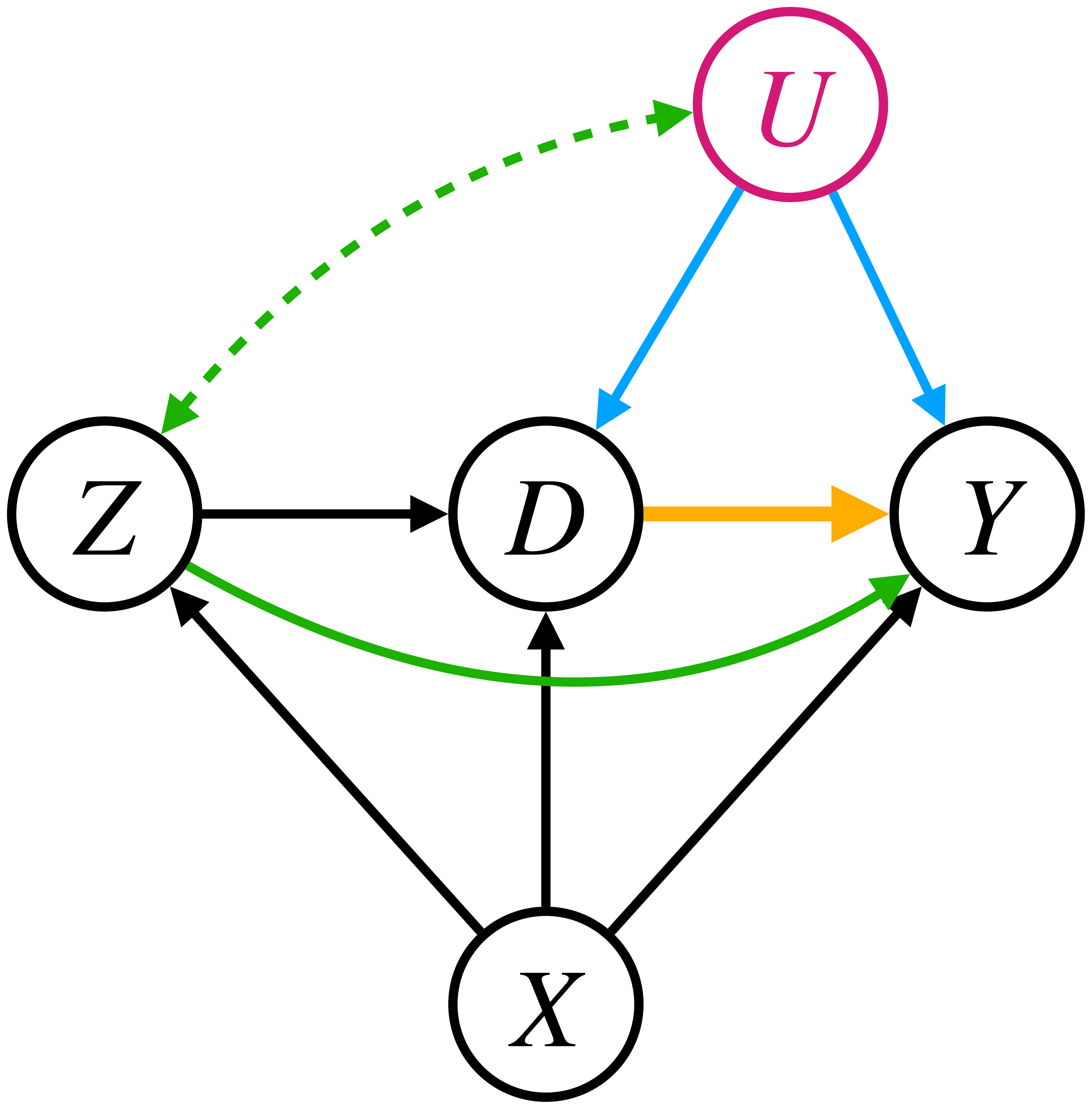
- Cornfield et al. (1959): Association between smoking and lung cancer
- Rosenbaum sensitivity model (Rosenbaum 1987), E-values (Ding and VanderWeele 2016), Marginal sensitivity model (Zhao et al. 2019, Dorn and Guo 2022), Instrumental variables (Altonji et al. 2005, Small 2007) and many more
- Lots of methods but **hardly used in practice**
- Reason: **simplistic** and **unintuitive**
- Proposal: **constrained optimization** $\rightarrow$ more flexibility and interpretability

Sensitivity Analysis - New Framework

- $(V_i, U_i)_{i=1}^n \sim \mathbb{P}_{V,U}$, but only $(V_i)_{i=1}^n$ observed
- Objective: $\beta(\theta, \psi)$
- θ are estimable parameters, i.e. only depend on $\mathbb{P}_V$
- ψ are sensitivity parameters, i.e. depend on $\mathbb{P}_{V,U}$
- Specify domain knowledge as constraints: $g(\theta, \psi) \leq 0$, $h(\theta, \psi) = 0$

$$\min_{\psi} / \max_{\psi} \beta(\hat{\theta}, \psi) \quad \text{subject to} \quad g(\hat{\theta}, \psi) \leq 0, \;\; h(\hat{\theta}, \psi) = 0.$$

Linear Model



Objective: $\beta = \beta_{Y \sim D|X,Z,U}$

Ordinary Least Squares (OLS)

$$\beta_{Y \sim D|X,Z} = \frac{\mathrm{cov}(Y^{\perp X,Z}, D^{\perp X,Z})}{\mathrm{var}(D^{\perp X,Z})}$$

Unbiased, if at least one of the arrows is absent.

Two-Stage Least Squares (TSLS)

$$\beta_{\mathrm{TSLS}} = \frac{\mathrm{cov}(Y^{\perp X}, Z^{\perp X})}{\mathrm{cov}(D^{\perp X}, Z^{\perp X})}$$

Unbiased, if both of the arrows are absent.

Parameters θ and ψ : R-values

- Let Y be a random variable; X, W, Z be random vectors.
- Residual of Y after regressing out X denoted by $Y^{\perp X}$

R^2-value : $$R^2_{Y\sim X} := 1 - \frac{\mathrm{var}(Y^{\perp X})}{\mathrm{var}(Y)}$$

Partial R^2-value : $$R^2_{Y\sim X|Z} := \frac{R^2_{Y\sim Z+X} - R^2_{Y\sim Z}}{1 - R^2_{Y\sim Z}}$$

Partial R-value : $$R_{Y\sim X|Z} := \mathrm{corr}(Y^{\perp Z}, X^{\perp Z})$$

Partial f-value : $$f_{Y\sim X|Z} := \frac{R_{Y\sim X|Z}}{\sqrt{1 - R^2_{Y\sim X|Z}}}$$

Cinelli and Hazlett (2020)

R^2-calculus

(i) *Independence*: If $Y \perp\!\!\!\perp X$, then $R^2_{Y\sim X} = 0$.

(ii) *Independent additivity*: If $X \perp\!\!\!\perp W$, then $R^2_{Y\sim X+W} = R^2_{Y\sim X} + R^2_{Y\sim W}$.

(iii) *Decomposition of unexplained variance*: $1 - R^2_{Y\sim X+W} = (1 - R^2_{Y\sim X})(1 - R^2_{Y\sim W|X})$

(iv) *Recursion of partial correlation*:

$$R_{Y\sim X|W} = \frac{R_{Y\sim X} - R_{Y\sim W}R_{X\sim W}}{\sqrt{1 - R^2_{Y\sim W}}\sqrt{1 - R^2_{X\sim W}}}$$

(v) *Reduction of partial correlation*: If X and W is one-dimensional and $Y \perp\!\!\!\perp W$, then

$$R_{Y\sim X|W} = \frac{R_{Y\sim X}}{\sqrt{1 - R^2_{X\sim W}}}$$

(vi) *Three-variable restriction*: If X and W is one-dimensional, then

$$f_{Y\sim X|W}\sqrt{1 - R^2_{Y\sim W|X}} = f_{Y\sim X}\sqrt{1 - R^2_{X\sim W}} - R_{Y\sim W|X}R_{X\sim W}.$$

All statements are also true when Z is partialed out, i.e. add "$|\,Z$".

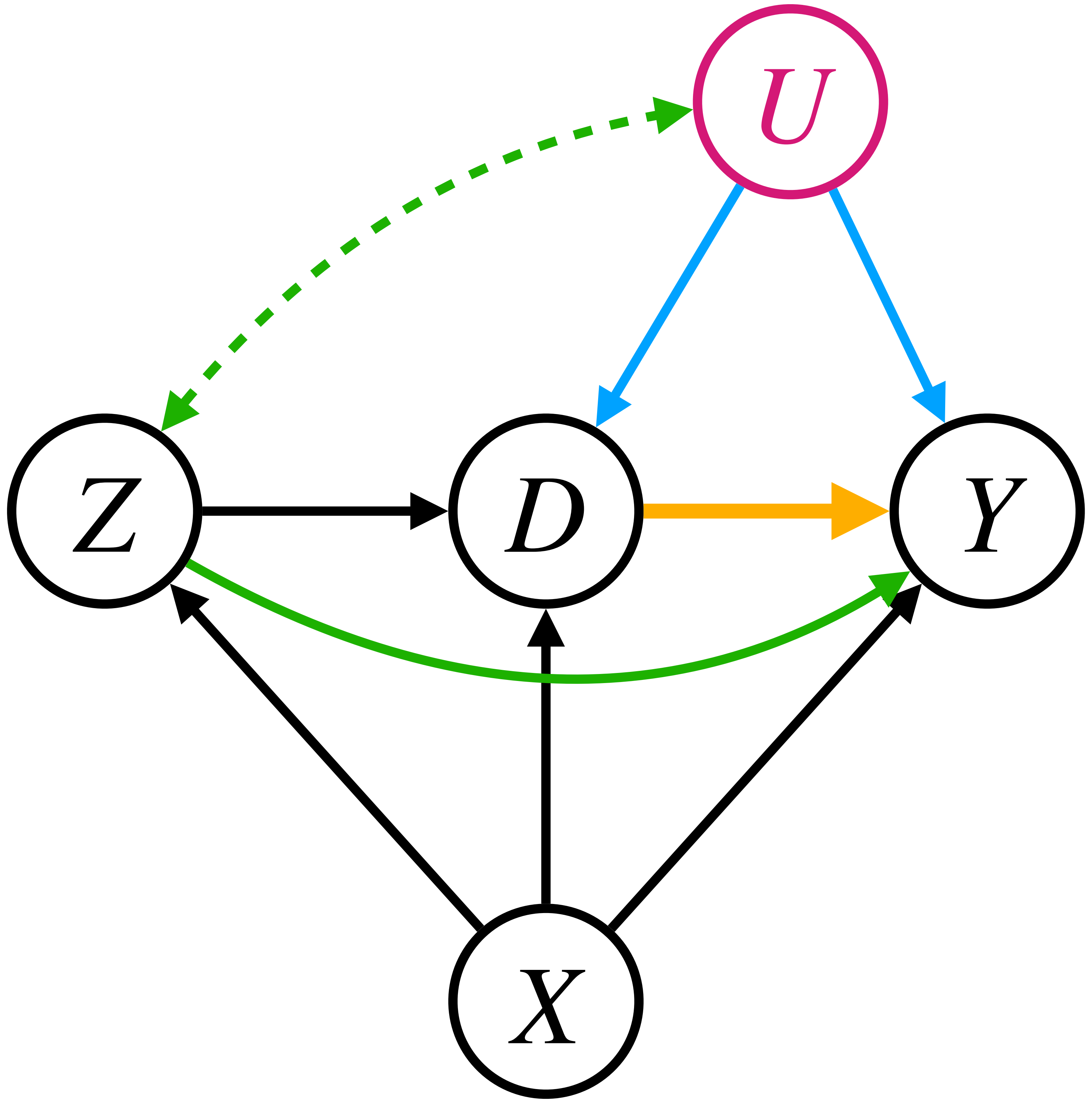
Anderson (1958)

Identifying objective β in terms of θ and ψ

Causal effect: $\beta = \beta_{Y\sim D|X,Z,U}$

$$\beta = \beta_{Y\sim D|X,Z} - R_{Y\sim U|X,Z,D}\, f_{D\sim U|X,Z} \frac{\mathrm{sd}(Y^{\perp X,Z,D})}{\mathrm{sd}(D^{\perp X,Z})}$$

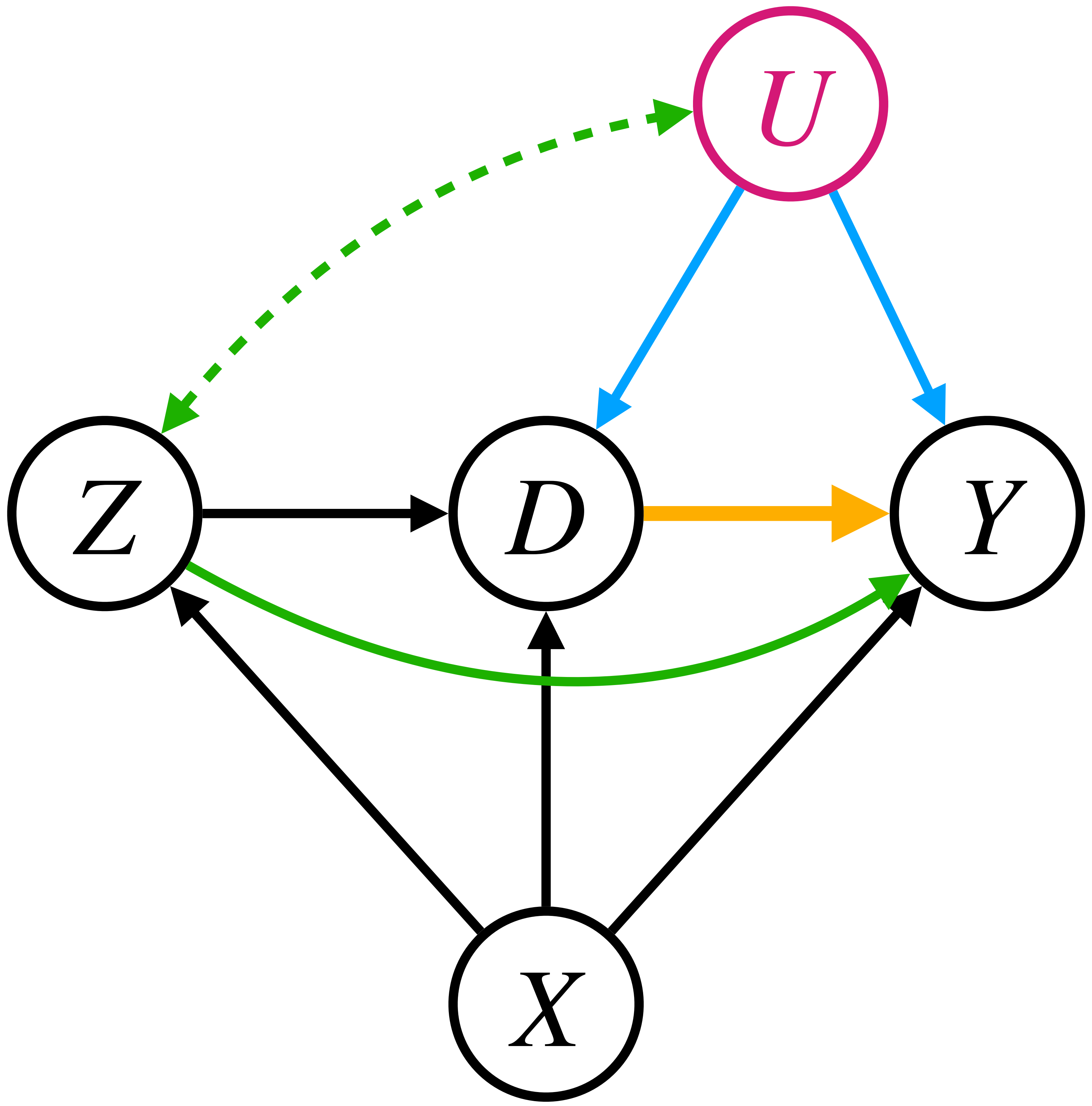
Sensitivity parameters ψ : $R_{Y\sim U|X,Z,D}$ and $R_{D\sim U|X,Z}$



Nagar (1959), Theil (1961)

Table of Bounds

$U \to D$	$R_{D\sim U\|X,Z} \in [B^l_{UD}, B^u_{UD}]$
	$R^2_{D\sim U\|\tilde{X},\dot{X}_I,Z} \leq b_{UD}\, R^2_{D\sim \dot{X}_J\|\tilde{X},\dot{X}_I,Z}$
$U \to Y$	$R_{Y\sim U\|X,D,Z} \in [B^l_{UY}, B^u_{UY}]$
	$R^2_{Y\sim U\|\tilde{X},\dot{X}_I,Z} \leq b_{UY}\, R^2_{Y\sim \dot{X}_J\|\tilde{X},\dot{X}_I,Z}$
	$R^2_{Y\sim U\|\tilde{X},\dot{X}_I,Z,D} \leq b_{UY}\, R^2_{Y\sim \dot{X}_J\|\tilde{X},\dot{X}_I,Z,D}$
$U \leftrightarrow Z$	$R_{Y\sim Z\|X,U,D} \in [B^l_{ZY}, B^u_{ZY}]$
	$R^2_{Z\sim U\|\tilde{X},\dot{X}_{-j}} \leq b_{UZ}\, R^2_{Z\sim \dot{X}_j\|\tilde{X},\dot{X}_{-j}}$
$Z \to Y$	$R_{Y\sim Z\|X,U,D} \in [B^l_{ZY}, B^u_{ZY}]$
	$R^2_{Y\sim Z\|X,U,D} \leq b_{ZY}\, R^2_{Y\sim \dot{X}_j\|\tilde{X},\dot{X}_{-j},Z,U,D}$



The bounds can be combined in any way.

Cinelli and Hazlett (2020)

Constraints $g(\theta, \psi) \leq 0$

- Direct bounds, e.g. $R_{D\sim U|X,Z} \in [-0.2, 0.4]$, $R^2_{Y\sim U|X,Z,D} \leq 0.5$
- Assumption: $X = (\dot{X}, \tilde{X})$ such that $U \perp\!\!\!\perp \dot{X} \,|\, \tilde{X}, Z$
- Comparative bound, e.g. $R^2_{Y\sim U|\tilde{X},\dot{X}_{-j},Z} \leq 2\, R^2_{Y\sim \dot{X}_j|\tilde{X},\dot{X}_{-j},Z}$

The unmeasured confounder U can explain at most twice as much variance in Y than $\dot{X}_j$ does - conditional on $(\tilde{X}, \dot{X}_{-j}, Z)$.

Constraints $g(\theta, \psi) \leq 0$ and $h(\theta, \psi) = 0$

- $\min / \max \; \beta_{Y\sim D|X,Z} - R_{Y\sim U|X,Z,D}\, f_{D\sim U|X,Z} \dfrac{\mathrm{sd}(Y^{\perp X,Z,D})}{\mathrm{sd}(D^{\perp X,Z})}$ subject to (1), (2), (3)

- Bounds: $R^2_{Y\sim U|\tilde{X},\dot{X}_{-j},Z} \leq 2\, R^2_{Y\sim \dot{X}_j|\tilde{X},\dot{X}_{-j},Z}\,,\; R_{D\sim U|X,Z} \in [-0.2, 0.4]$ (1)

- $R^2_{Y\sim U|X,Z} \overset{\text{(v)}}{=} \dfrac{R^2_{Y\sim U|\tilde{X},\dot{X}_{-j},Z}}{1 - R^2_{Y\sim \dot{X}_j|\tilde{X},\dot{X}_{-j},Z}}$ (2)

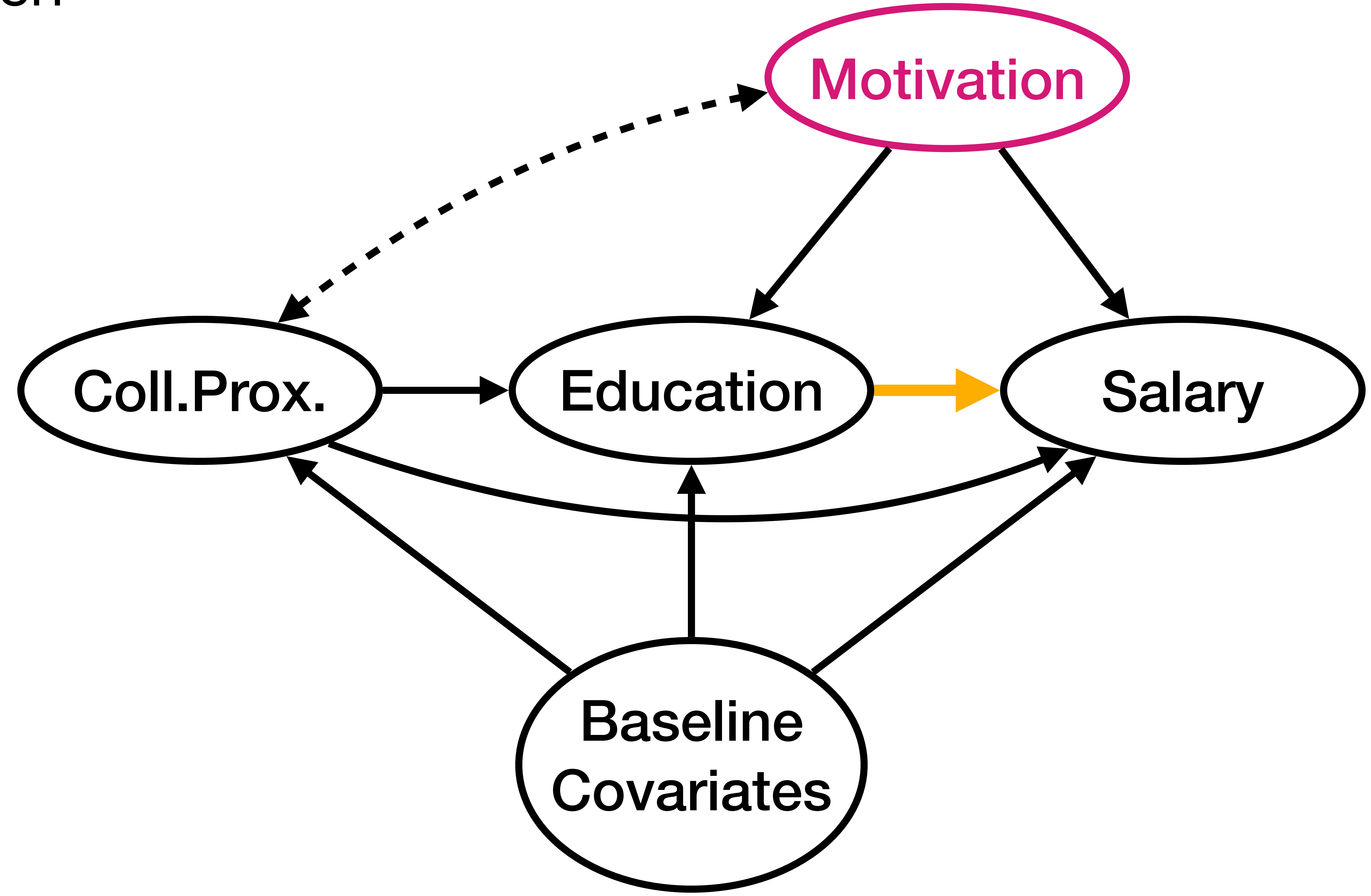
- $R_{Y\sim U|X,Z,D} \overset{\text{(iv)}}{=} \dfrac{R_{Y\sim U|X,Z} - R_{Y\sim D|X,Z}\, R_{D\sim U|X,Z}}{\sqrt{1 - R^2_{Y\sim D|X,Z}}\sqrt{1 - R^2_{D\sim U|X,Z}}}$ (3)

Optimization - Leveraging Monotonicity

- $a = R_{D\sim U|X,Z}$, $b = R_{Y\sim U|X,Z,D}$, $d = R_{Y\sim U|X,Z}$
- $\min / \max \ c_1 - b \dfrac{a}{\sqrt{1-a^2}} c_2, \quad$ s.t. $\quad a \in [-0.2, 0.4]$, $d^2 \leq c_3$, $b = \dfrac{d - c_4 a}{\sqrt{1 - c_4^2}\sqrt{1-a^2}}$
- Brute force grid search: create three-dimensional grid of (a, b, d)-values
- Adapted grid search:
  - For fixed a, b is monotone in d and the objective is monotone in b
  - One-dimensional grid of a-values
- $\mathcal{O}(m^3)$, but often $\mathcal{O}(m)$, where m is the number of points per grid dimension

Data Example

- National Longitudinal Survey of Young Men
- From 1966 until 1981
- 3010 individuals
- Y: log-earnings
- D: years of schooling
- X: years of labour force experience; indicators for living in the south, being black and living in a metropolitan area
- Z: growing up close to 4-year college



Card (1993)

Sensitivity Analysis

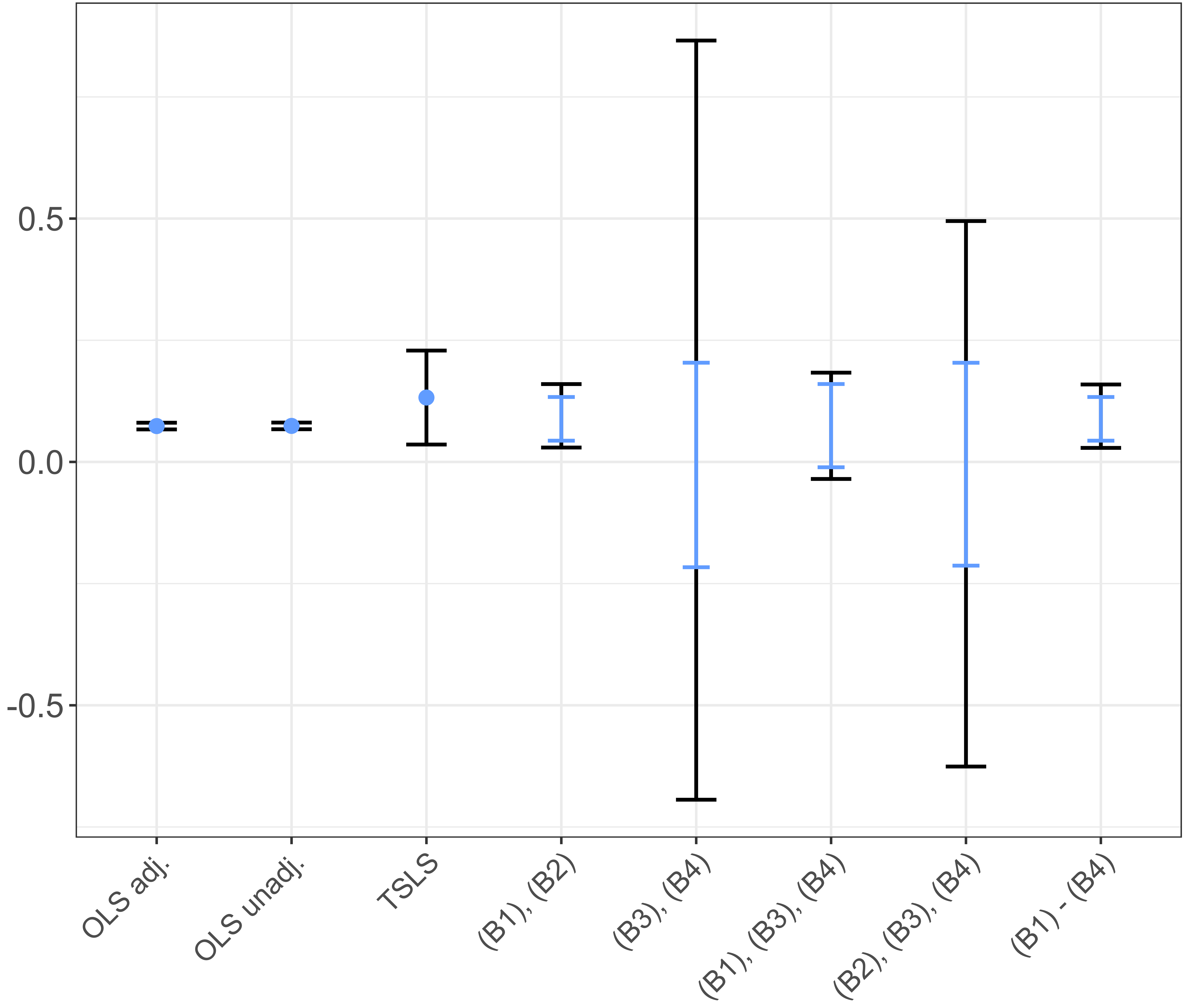
- Partition covariates X: $\dot{X}$ indicator for being black; $\tilde{X}$ remaining covariates
- Assumption: $U \perp\!\!\!\perp \dot{X} \mid \tilde{X}, Z$
- User specified bounds:

(B1) $R^2_{D \sim U | \tilde{X}, Z} \leq 4\, R^2_{D \sim \dot{X} | \tilde{X}, Z}$

(B2) $R^2_{Y \sim U | \tilde{X}, Z, D} \leq 5\, R^2_{Y \sim \dot{X} | \tilde{X}, Z, D}$

(B3) $R^2_{Z \sim U | \tilde{X}} \leq 0.5\, R^2_{Z \sim \dot{X} | \tilde{X}}$

(B4) $R^2_{Y \sim Z | X, U, D} \leq 0.1\, R^2_{Y \sim \dot{X} | \tilde{X}, Z, U, D}$



(B1): $U \to D$

(B2): $U \to Y$

(B3): $U \leftrightarrow Z$

(B4): $Z \to Y$

Things I did not talk about

- Extension of the R^2-calculus to Hilbert spaces
- Multiple unmeasured confounders
- Confidence statements: Sensitivity intervals via bootstrap
- Additional visualization tools

Conclusion

- New Framework: Sensitivity analysis as optimization problem
- Introduction of the R^2-calculus
- Sensitivity analysis in linear models with R-values (OLS and TSLS)
- Flexible and interpretable bounds
- Application on a dataset

Thanks for your attention!

arXiv: https://arxiv.org/abs/2301.00040

Website: https://tobias-freidling.onrender.com

References

Altonji, J.G., Elder, T.E. and Taber, C.R. (2005) 'An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling', Journal of Human Resources, XL(4), pp. 791–821.

Anderson, T.W. (1958) 'An introduction to multivariate statistical analysis'. Wiley Publications in Mathematical Statistics.

Cornfield, J. et al. (1959) 'Smoking and lung cancer: recent evidence and a discussion of some questions', Journal of the National Cancer institute, 22(1), pp. 173–203.

Ding, P. and VanderWeele, T.J. (2016) 'Sensitivity Analysis Without Assumptions', Epidemiology (Cambridge, Mass.), 27(3), pp. 368–377.

Dorn, J. and Guo, K. (2022) 'Sharp Sensitivity Analysis for Inverse Propensity Weighting via Quantile Balancing', Journal of the American Statistical Association, pp. 1–13.

Card, D. (1993) *Using Geographic Variation in College Proximity to Estimate the Return to Schooling*. w4483. Cambridge, MA: National Bureau of Economic Research, p. w4483.

Cinelli, C. and Hazlett, C. (2020) 'Making sense of sensitivity: extending omitted variable bias', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1), pp. 39–67.

Cochran, W.G. (1938) 'The Omission or Addition of an Independent Variate in Multiple Linear Regression', *Supplement to the Journal of the Royal Statistical Society*, 5(2), p. 171.

References

Frank, K.A. (2000) ‘Impact of a Confounding Variable on a Regression Coefficient’, *Sociological Methods & Research*, 29(2), pp. 147–194.

Hosman, C.A., Hansen, B.B. and Holland, P.W. (2010) ‘The sensitivity of linear regression coefficients’ confidence limits to the omission of a confounder’, *The Annals of Applied Statistics*, 4(2).

Nagar, A.L. (1959) ‘The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations’, *Econometrica*, 27(4), pp. 575–595.

Rosenbaum, P.R. (1987) ‘Sensitivity analysis for certain permutation inferences in matched observational studies’, Biometrika, 74(1), pp. 13–26.

Small, D.S. (2007) ‘Sensitivity Analysis for Instrumental Variables Regression With Overidentifying Restrictions’, Journal of the American Statistical Association, 102(479), pp. 1049–1058.

Theil, H. (1961) ‘Economic forecasts and policy’, North-Holland Publishing Company.

Zhao, Q., Small, D.S. and Bhattacharya, B.B. (2019) ‘Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(4), pp. 735–761.