ECON 251 Discussion

Instrumental Variables (IV) and Returns to Schooling

Elird Haxhiu

Fall 2022


khaled eltokhy
@eltokh7
Econ PhD in a nutshell
This terminology is not your fault
But being able to follow it is your problem
Sorry
9:43 AM · Oct 20, 2022 · Twitter for iPhone
2 Retweets
2 Quote Tweets
74 Likes

Outline

1. Review some key concepts
   - OLS assumptions and main results
   - Mincer (1974) regression framework (see Heckman et al., 2003)
   - Omitted variable bias (OVB) formula

2. Instrumental variables (IV) logic + intuition
   - Examples of instruments for schooling in the wild
   - Method of moments (MM) estimator consistency under IV assumptions
   - In practice: when doing IV is worse than just doing OLS

3. Sensitivity Analysis: building towards Cinelli and Hazlett (2020)

Usual Assumptions

- MLR1 (linear outcome model) $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i$
- MLR2 (random sampling) $\{Y_i, X_{i1}, \ldots, X_{ik}\}_{i=1}^{N}$ is random draw
- MLR3 (no collinearity) no X_{ij} linear function of any other X_{il}

- MLR4 (independence) $E[U_i | X_{i1}, \ldots, X_{ik}] = 0$

- MLR5 (homoskedasticity) $\mathrm{Var}(U_i | X_{i1}, \ldots, X_{ik}) = \sigma^2$
- MLR6 (normality) $U_i \sim N(0, \sigma^2)$

$$\Rightarrow Y_i \sim N(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}, \sigma^2)$$

Ordinary Least Squares (OLS) Estimator + Results

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i$$

$$\min_{\{\beta_0,\beta_1,\ldots,\beta_k\}} \frac{1}{N}\sum_{i=1}^{N}(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2 \qquad \Rightarrow \qquad \hat{\beta}_j^{OLS}$$

Ordinary Least Squares (OLS) Estimator + Results

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i$$

$$\min_{\{\beta_0,\beta_1,\ldots,\beta_k\}} \frac{1}{N}\sum_{i=1}^{N}(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2 \quad \Rightarrow \quad \hat{\beta}_j^{OLS}$$

- T1 (unbiased) MLR1+2+3+4 $\Rightarrow$ $E\left[\widehat{\beta}_j^{\,OLS}\right] = \beta_j \quad \forall j = \{0,1,\ldots,k\}$
- T2 (efficient) MLR1+2+3+4+5 $\Rightarrow$ $E\left[\widehat{\beta}_j^{\,OLS}\right] = \beta_j \quad \forall j = \{0,1,\ldots,k\}$

  (Gauss-Markov) $\mathrm{Var}\left[\widehat{\beta}_j^{\,OLS}\right] \leq \mathrm{Var}\left[\widehat{\beta}_j^{\,\text{other linear}}\right]$

Ordinary Least Squares (OLS) Estimator + Results

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + U_i$$

$$\min_{\{\beta_0,\beta_1,\ldots,\beta_k\}} \frac{1}{N}\sum_{i=1}^{N}(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2 \quad \Rightarrow \quad \hat{\beta}_j^{OLS}$$

- T3 (efficient) MLR1+2+3+4+5+6 $\Rightarrow$ $\widehat{\beta_j}^{OLS} \sim N\left(\beta_j, \mathrm{Var}\left[\beta_j\right]\right) \quad \forall j = \{0,1,\ldots,k\}$

  (Classical)

$$\frac{\widehat{\beta_j}^{OLS} - \beta_j}{\mathrm{sd}[\beta_j]} \sim N(0,1) \qquad \frac{\widehat{\beta_j}^{OLS} - \beta_j}{\mathrm{se}[\beta_j]} \sim t(N-k-1)$$

Mincer (1974) regression framework

- $\log Y_i \geq 0$ denotes log earnings (outcome)
- ~~$S_i \in \{0,1\}$ is whether i finished college~~ (treatment) in general, $S_i \in \{0,1,2,\ldots,S_{\max}\}$
- U_i is unobserved error term (ex: ability)
- simple linear population regression function (PRF) $\log Y_i = \alpha + \beta \cdot S_i + U_i$
- β = returns to extra year of schooling

Mincer (1974) regression framework

- $\log Y_i \geq 0$ denotes log earnings (outcome)
- ~~$S_i \in \{0,1\}$ is whether i finished college~~ (treatment) in general, $S_i \in \{0,1,2,\ldots,S_{\max}\}$
- U_i is unobserved error term (ex: ability)
- simple linear population regression function (PRF) $\log Y_i = \alpha + \beta \cdot S_i + U_i$
- β = returns to extra year of schooling

- Potential outcomes $\log Y_i(s)$ + treatment effects $\text{ATE} \coloneqq E[\log Y_i(s) - \log Y_i(s-1)]$

- Need independence to *identify* ATE $S_i \perp \log Y_i(s) \quad \Leftrightarrow \quad E[U_i | S_i] = 0$

  with simple comparisons in $\hat{\beta}^{OLS}$
- Recall how independence implies exogeneity $\Rightarrow \text{Cov}(S_i, U_i) = 0$

Omitted variable bias (OVB)

"True" model $\log Y_i = \alpha + \beta \cdot S_i + \delta^{X \to Y} \cdot X_i + U_i$ $\quad \mathrm{Cov}(S_i, U_i) = 0$

Our model $\log Y_i = a + b \cdot S_i + E_i$

Auxiliary model $X_i = c + \gamma^{S \to X} \cdot S_i + \eta_i$

Omitted variable bias (OVB)

“True” model $\log Y_i = \alpha + \beta \cdot S_i + \delta^{X \to Y} \cdot X_i + U_i$ $\mathrm{Cov}(S_i, U_i) = 0$

Our model $\log Y_i = a + b \cdot S_i + E_i$

Auxiliary model $X_i = c + \gamma^{S \to X} \cdot S_i + \eta_i$

In week 5 we proved that naively assuming $\mathrm{Cov}(S_i, E_i) = 0$ in our model implies

$$b = \frac{\mathrm{Cov}(S_i, \log Y_i)}{\mathrm{Var}(S_i)}$$

$$= \beta + \gamma^{S \to X} \cdot \delta^{X \to Y}$$

= causal effect + (var in S related to X) · (var in X related to Y)

What if $\text{Cov}(S_i, U_i) = 0$ is also suspect?

1. Despair ⇒ intellectual nihilism, true reality hidden to little humans in the world

2. One answer ⇒ sensitivity analysis + new tools in Cinelli and Hazlett (2020)

3. Traditional approach ⇒ find an instrumental variable Z_i which generates some exogenous variation in the treatment S_i but does not affect $\log Y_i$ directly

Outline

1. Review some key concepts
    - OLS assumptions and main results
    - Mincer (1974) regression framework (see Heckman et al., 2003)
    - Omitted variable bias (OVB) formula

2. Instrumental variables (IV) logic + intuition
    - Examples of instruments for schooling in the wild
    - Method of moments (MM) estimator consistency under IV assumptions
    - In practice: when doing IV is worse than just doing OLS

3. Sensitivity Analysis: building towards Cinelli and Hazlett (2020)

IV logic and intuition

- Variation in treatment S_i includes both endogenous and exogenous parts
- Goal: isolate exogenous variation and rely on it exclusively
- How can we decompose S_i into S_i^{eXog} and S_i^{eNdog} using an instrument Z_i?

IV logic and intuition

- Variation in treatment S_i includes both endogenous and exogenous parts
- Goal: isolate exogenous variation and rely on it exclusively
- How can we decompose S_i into S_i^{eXog} and S_i^{eNdog} using an instrument Z_i?

- We first define an instrumental variable (IV) as any random variable satisfying

1. <u>Relevance</u> $\quad \mathrm{Cov}(Z_i, S_i) \neq 0$
2. <u>Exogeneity</u> $\quad \mathrm{Cov}(Z_i, U_i) = 0$
3. <u>Exclusion</u> $\quad$ no direct effect of Z_i on outcome $\log Y_i$

$\Leftrightarrow$ instrument Z_i does not appear in model of $\log Y_i$

IV intuition = decompose S_i into S_i^{X} and S_i^{N} using Z_i

- An instrumental variable (IV) satisfies
  1. Relevance — $\mathrm{Cov}(Z_i, S_i) \neq 0$
  2. Exogeneity — $\mathrm{Cov}(Z_i, U_i) = 0$
  3. Exclusion — no direct effect of Z_i on outcome Y_i

     $\Leftrightarrow$ instrument Z_i does not appear in model of Y_i

IV intuition = decompose S_i into S_i^{X} and S_i^{N} using Z_i

- An instrumental variable (IV) satisfies
  1. Relevance $\qquad \mathrm{Cov}(Z_i, S_i) \neq 0$
  2. Exogeneity $\qquad \mathrm{Cov}(Z_i, U_i) = 0$
  3. Exclusion $\qquad$ no direct effect of Z_i on outcome Y_i

     $\Leftrightarrow$ instrument Z_i does not appear in model of Y_i

- If we have such a Z_i for S_i then we can use a different linear model (first stage) to generate predicted values for the treatment $\hat{S}_i := \hat{\pi}_0 + \hat{\pi}_1 Z_i$
- Under IV exogeneity, this is equivalent to exogenous variation S_i^{X}
- We can then estimate returns β from model $\log Y_i = \alpha + \beta \cdot \hat{S}_i + U_i$

Examples of instruments Z_i for S_i in the wild

1. Distance to college when 16 years old
2. Month of birth interacted with compulsory school attendance laws
3. Natural disasters preventing some people from going to school
4. Number of siblings
5. Opportunities to emigrate (Haxhiu, 2022)

A subtle point

- The instrument you choose implicitly defines a “complier” group = the people moved to change the value of their treatment by the IV
- The estimator <u>relies only on these people</u> to construct an estimate of the β
- Different IVs often lead to different estimates of β if the sub-populations they induce into changing their value of S_i are somehow different

A subtle point

- The instrument you choose implicitly defines a "complier" group = the people moved to change the value of their treatment by the IV
- The estimator relies only on these people to construct an estimate of the β
- Different IVs often lead to different estimates of β if the sub-populations they induce into changing their value of S_i are somehow different

- Contrast with OLS, which relies on everyone to construct an estimate of β
- Therefore, we say OLS (= simple comparison) identifies the ATE
- The estimator under IV identifies a **local** average treatment effect (**L**ATE)

Derive MM estimator under IV assumptions

Start w/ IV exogeneity $\mathrm{Cov}(Z_i, U_i) = 0$ + substitute Mincer (1974) earnings model

$$\mathrm{Cov}(Z_i, \log Y_i - \alpha - \beta \cdot S_i) = 0$$

Derive MM estimator under IV assumptions

Start w/ IV exogeneity $\text{Cov}(Z_i, U_i) = 0$ + substitute Mincer (1974) earnings model

$$\text{Cov}(Z_i, \log Y_i - \alpha - \beta \cdot S_i) = 0$$
$$\text{Cov}(Z_i, \log Y_i) - \text{Cov}(Z_i, \alpha) - \beta \cdot \text{Cov}(Z_i, S_i) = 0$$
$$\text{Cov}(Z_i, \log Y_i) = \beta \cdot \text{Cov}(Z_i, S_i)$$
$$\beta = \frac{\text{Cov}(Z_i, \log Y_i)}{\text{Cov}(Z_i, S_i)}$$

Derive MM estimator under IV assumptions

Start w/ IV exogeneity $\mathrm{Cov}(Z_i, U_i) = 0$ + substitute Mincer (1974) earnings model

$$\begin{aligned}
\mathrm{Cov}(Z_i, \log Y_i - \alpha - \beta \cdot S_i) &= 0 \\
\mathrm{Cov}(Z_i, \log Y_i) - \mathrm{Cov}(Z_i, \alpha) - \beta \cdot \mathrm{Cov}(Z_i, S_i) &= 0 \\
\mathrm{Cov}(Z_i, \log Y_i) &= \beta \cdot \mathrm{Cov}(Z_i, S_i) \\
\beta &= \frac{\mathrm{Cov}(Z_i, \log Y_i)}{\mathrm{Cov}(Z_i, S_i)}
\end{aligned}$$

$$\Rightarrow \hat{\beta}^{MM} = \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)} := \frac{\sum_{i=1}^{N}(Z_i - \overline{Z})(\log Y_i - \overline{\log Y})}{\sum_{i=1}^{N}(Z_i - \overline{Z})(S_i - \overline{S})}$$

Consistency of MM estimator under IV assumptions

- Start from definition of estimator, and compute the probability limit

$$\hat{\beta}^{MM} = \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)}$$

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{MM} = \operatorname*{plim}_{N\to\infty} \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)}$$

Consistency of MM estimator under IV assumptions

- Start from definition of estimator, and compute the probability limit

$$\hat{\beta}^{MM} = \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)}$$

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{MM} = \operatorname*{plim}_{N\to\infty} \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)} = \frac{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Cov}}(Z_i, S_i)}$$

$$= \frac{\mathrm{Cov}(Z_i, \log Y_i)}{\mathrm{Cov}(Z_i, S_i)} = \frac{\mathrm{Cov}(Z_i, \alpha + \beta \cdot S_i + U_i)}{\mathrm{Cov}(Z_i, S_i)}$$

Consistency of MM estimator under IV assumptions

- Start from definition of estimator, and compute the probability limit

$$\hat{\beta}^{MM} = \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)}$$

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{MM} = \operatorname*{plim}_{N\to\infty} \frac{\widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\widehat{\mathrm{Cov}}(Z_i, S_i)} = \frac{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Cov}}(Z_i, \log Y_i)}{\operatorname*{plim}_{N\to\infty} \widehat{\mathrm{Cov}}(Z_i, S_i)}$$

$$= \frac{\mathrm{Cov}(Z_i, \log Y_i)}{\mathrm{Cov}(Z_i, S_i)} = \frac{\mathrm{Cov}(Z_i, \alpha + \beta \cdot S_i + U_i)}{\mathrm{Cov}(Z_i, S_i)}$$

$$= \frac{\mathrm{Cov}(Z_i, \alpha) + \beta \cdot \mathrm{Cov}(Z_i, S_i) + \mathrm{Cov}(Z_i, U_i)}{\mathrm{Cov}(Z_i, S_i)} = \beta + \frac{\mathrm{Cov}(Z_i, U_i)}{\mathrm{Cov}(Z_i, S_i)}$$

In practice: when doing IV worse than OLS

- The OVB formula for OLS implies that it converges to

$$\underset{N\to\infty}{\text{plim}}\,\hat{\beta}^{OLS} = \beta + \underbrace{\frac{\text{Cov}(S_i, X_i)}{\text{Var}(S_i)} \cdot \delta^{X\to Y}}_{:=\frac{\text{Cov}(S_i, E_i)}{\text{Var}(S_i)}}$$

- We have just shown that the MM estimator converges to

$$\underset{N\to\infty}{\text{plim}}\,\hat{\beta}^{MM} = \beta + \frac{\text{Cov}(Z_i, U_i)}{\text{Cov}(Z_i, S_i)}$$

In practice: when doing IV worse than OLS

- The OVB formula for OLS implies that it converges to

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{OLS} = \beta + \underbrace{\frac{\text{Cov}(S_i, X_i)}{\text{Var}(S_i)} \cdot \delta^{X\to Y}}_{:=\frac{\text{Cov}(S_i, E_i)}{\text{Var}(S_i)}}$$

- We have just shown that the MM estimator converges to

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{MM} = \beta + \frac{\text{Cov}(Z_i, U_i)}{\text{Cov}(Z_i, S_i)}$$

- Sensitivity analysis asks what happens if the relevant exogeneity conditions $\text{Cov}(S_i, U_i) = 0$ or $\text{Cov}(Z_i, U_i) = 0$ to prove consistency are not exactly $= 0$
- Not clear if $\text{Cov}(S_i, U_i) = 0$ or $\text{Cov}(Z_i, U_i) = 0$ more likely to hold…
- Note: IV always less precise than OLS, but gap shrinks with relevance

Some other practical matters

- We can always write estimator $\hat{\beta}^{MM}$ under IV assumptions as the
    1. ratio of two OLS estimators (reduced form $\div$ first stage)
    2. OLS coefficient in regression of outcome on “predicted” treatment

Some other practical matters

- We can always write estimator $\hat{\beta}^{MM}$ under IV assumptions as the
  1. ratio of two OLS estimators (reduced form $\div$ first stage)
  2. OLS coefficient in regression of outcome on “predicted” treatment

- We can include more than one instrument in the first stage predicting the endogenous variable, and then use any of the estimators above
- Generically called Two-Stage Least Squares (TSLS) estimator
- Potential costs, in addition to benefits, of having more IVs

Outline

1. Review some key concepts
    - OLS assumptions and main results
    - Mincer (1974) regression framework (see Heckman et al., 2003)
    - Omitted variable bias (OVB) formula

2. Instrumental variables (IV) logic + intuition
    - Examples of instruments for schooling in the wild
    - Method of moments (MM) estimator consistency under IV assumptions
    - In practice: when doing IV is worse than just doing OLS

3. Sensitivity Analysis: building towards Cinelli and Hazlett (2020)

Defining Sensitivity Analysis

- Estimands = real causal effects we wish to estimate, like the ATE or what we have referred to as the "true" slope in regression

- Estimators =

- Assumptions =

Defining Sensitivity Analysis

- Estimands = real causal effects we wish to estimate, like the ATE or what we have referred to as the "true" slope in regression
- Estimators = formulas/instructions for how to mix up the data to produce an estimate of something we hope captures the estimand, like $\hat{\beta}^{OLS}$ or $\hat{\beta}^{MM}$ under IV assumptions
- Assumptions =

Defining Sensitivity Analysis

- Estimands = real causal effects we wish to estimate, like the ATE or what we have referred to as the "true" slope in regression

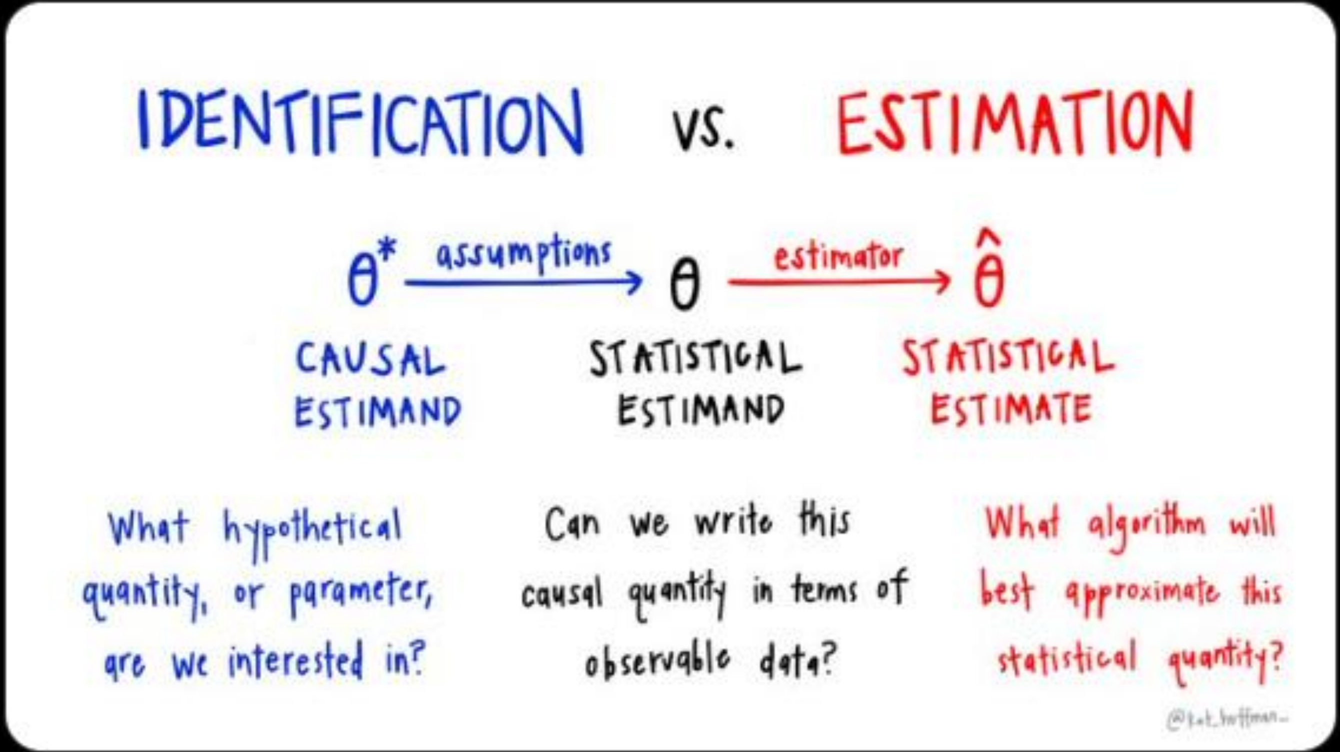
- Estimators = formulas/instructions for how to mix up the data to produce an estimate of something we hope captures the estimand, like $\hat{\beta}^{OLS}$ or $\hat{\beta}^{MM}$ under IV assumptions

- Assumptions = how the variation we study (all S_i in OLS vs only S_i^{X} in IV) relates to potential confounding variables driving the observed relationship between treatment and outcome

Kat Hoffman
@kat_hoffman_

after describing all components of the target trial, we detailed its emulation using observational data

we consider measurement of data, confounding, loss-to-follow-up interventions, estimand identification (via the longitudinal g-formula) assumptions, etc.



8:48 AM · Oct 4, 2022 · Typefully

Defining Sensitivity Analysis

- <u>Estimands</u> = real causal effects we wish to estimate
- <u>Estimators</u> = formula for how to mix up data to produce an estimate hopefully capturing the estimand
- <u>Assumptions</u> = how the variation we study relates to potential confounding variables driving relationship

Defining Sensitivity Analysis

- <u>Estimands</u> = real causal effects we wish to estimate
- <u>Estimators</u> = formula for how to mix up data to produce an estimate hopefully capturing the estimand
- <u>Assumptions</u> = how the variation we study relates to potential confounding variables driving relationship

- Idea: how sensitive are estimated treatment effects (TE) to potentially minor violations of relevant exogeneity condition $\mathrm{Cov}(S_i, U_i) = 0$ or $\mathrm{Cov}(Z_i, U_i) = 0$?

Defining Sensitivity Analysis

- <u>Estimands</u> = real causal effects we wish to estimate
- <u>Estimators</u> = formula for how to mix up data to produce an estimate hopefully capturing the estimand
- <u>Assumptions</u> = how the variation we study relates to potential confounding variables driving relationship

- Idea: how sensitive are estimated treatment effects (TE) to potentially minor violations of relevant exogeneity condition $\mathrm{Cov}(S_i, U_i) = 0$ or $\mathrm{Cov}(Z_i, U_i) = 0$?

- Or: how much unobserved confounding would overturn TE estimates?
- The <u>less unobserved confounding</u> is needed to overturn a given TE estimate, the <u>less trustworthy</u> is the research design overall!

Smoking and lung cancer

- How we know smoking tobacco causes some increase in the probability of getting lung cancer?
- Correlation = 9 times more cancer among smokers…

Smoking and lung cancer

- How we know smoking tobacco causes some increase in the probability of getting lung cancer?
- Correlation = 9 times more cancer among smokers…

- Main idea: genetic arguments rule out selection bias driving all the observed relationship, since it is unlikely that one genetic mutation leads to such a huge difference in such a complex behavior in humans
- Why can’t we generalize this approach too much?

Cinelli and Hazlett (2020) general method

- Idea: use the estimated effect of a control variable to benchmark how strong of an effect some unobserved confounder would need to have to overturn an estimated treatment effect
- In the coming weeks, we are going to make this precise…

- For now, see Cinelli 2020 presentation of this paper to get ready!