A econometric perspective of instrumental variable method

Xinzhou Guo

HKUST

(Credited to Zhichao Jiang)

April 14, 2024

Linear regression models

- Outcome Y; regressor D
- Linear model:
$$\underline{Y_i = D_i^\top \beta + \epsilon_i, \mathbb{E}(\epsilon_i \mid D_i) = 0}$$
$$\rightsquigarrow \underline{\mathbb{E}(D_i \epsilon_i) = 0}$$
- OLS estimation
  - $\mathbb{E}\left\{D_i\left(Y_i - D_i^\top \beta\right)\right\} = 0$
  - $\beta = \left\{\mathbb{E}\left(D_i D_i^\dagger\right)\right\}^{-1} \mathbb{E}(D_i Y_i)$
  - univariate linear regression:

$$\operatorname{cov}(Y_i, D_i) = \operatorname{cov}(\beta_0 + \beta_1 D_i + \epsilon_i, D_i) = \beta_1 \operatorname{cov}(D_i, D_i)$$

$Y \sim D \qquad \hat{\beta} \to \beta$

$$Y_i = D_i^T \beta + \varepsilon_i$$

$$E(\varepsilon_i | D_i) = 0$$

$$E\,Y(1) = E\,E(Y | U, D=1)$$

$$Y = D^T \beta + U^T r + \varepsilon \qquad E(\eta | D) \neq 0$$

$$Y \sim D \qquad \hat{\beta} \xrightarrow{?} \beta$$

$$\underline{Y = D^T \beta + \eta} \qquad \eta = U^T r + \varepsilon$$

Endogenous variable

$Y \sim D$ consistent

- Regressor D is exogenous: $\mathbb{E}(D_i \epsilon_i) = 0$

$\mathbb{E}(\epsilon \mid D) \neq 0$

- Regressor D is endogenous: $\mathbb{E}(D_i \epsilon_i) \neq 0$

$Y \sim D$ $\quad \hat{\beta} \not\to \beta$

- Suppose the true model is $Y_i = \beta_0 + \beta_1 D_i + \beta_2 U_i + \eta_i$,

$$\mathbb{E}(\eta_i \mid D_i, U_i) = 0$$

How is the model related to estimation of ACE on observational data?

- Linear model with observed data: $Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$, where $\epsilon_i = \beta_2 U_i + \eta_i$
  - $\mathbb{E}(\epsilon_i \mid D_i) = \mathbb{E}(\beta_2 U_i + \eta_i \mid D_i) = \beta_2 \mathbb{E}(U_i \mid D_i)$
  - $\beta_2 \neq 0$ and $\mathbb{E}(U_i \mid D_i) \neq 0 \rightsquigarrow D_i$ is an endogenous variable

Omitted variable bias

- OLS estimation of $Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$

$$\beta_1^{\text{ols}} = \frac{\text{cov}\,(Y_i, D_i)}{\text{cov}\,(D_i, D_i)} = \beta_1 + \frac{\text{cov}\,(\epsilon_i, D_i)}{\text{cov}\,(D_i, D_i)} = \beta_1 + \beta_2 \cdot \frac{\text{cov}\,(U_i, D_i)}{\text{cov}\,(D_i, D_i)}$$

$Y \sim D$

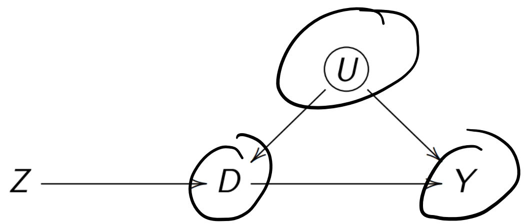
$\|$

ACE

- Bias $\neq 0$ if U_i associated with both D_i and Y_i – why?

Latent confounders

Instrumental variable in simple linear regression



- Target: estimate β_1 in $Y_i = \beta_0 + \beta_1 D_i + \epsilon_i, \epsilon_i = \beta_2 U_i + \eta_i, \mathbb{E}(\epsilon_i) = 0$
- Instrumental variable $Z_i : \mathbb{E}(Z_i \epsilon_i) = 0$ – what does it imply?
    - requires $\text{cov}(Z_i, D_i) \neq 0$ – can't be white noise

$$\text{cov}(Y_i, Z_i) = \text{cov}(\beta_0 + \beta_1 D_i + \epsilon_i, Z_i) = \beta_1 \text{cov}(D_i, Z_i)$$

$$\beta_1 = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)}$$

- Wald estimator: $\frac{\widehat{\text{cov}}(Y_i,Z_i)}{\widehat{\text{Cov}}(D_i,Z_i)}$ - ratio of two regression coefficients

D, Z are 1-dimension

$$\beta_1 = \frac{\operatorname{cov}(Y, Z)}{\operatorname{cov}(D, Z)}$$

$$\beta_1 = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)}$$

$Y \sim Z$

$$\hat{\beta}_{YZ} = \frac{\text{cov}(Y, Z)}{\text{cov}(Z, Z)}$$

$D \sim Z$

$$\hat{\beta}_{DZ} = \frac{\text{cov}(D, Z)}{\text{cov}(Z, Z)}$$

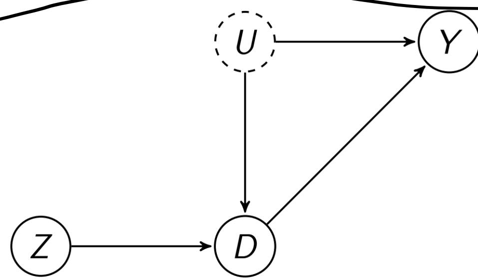
Least square estimator

$$\beta_1 = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)} = \frac{\text{cov}(Y_i, Z_i) / \text{cov}(Z_i, Z_i)}{\text{cov}(D_i, Z_i) / \text{cov}(Z_i, Z_i)}$$

- $\text{cov}(Y_i, Z_i) / \text{cov}(Z_i, Z_i)$: regression of Y_i on Z_i
- $\text{cov}(D_i, Z_i) / \text{cov}(Z_i, Z_i)$: regression of D_i on Z_i
- Wald estimator: $\frac{\text{lm}(Y \sim Z)\$\,\text{coef}[2]}{\text{lm}(D \sim Z)\$\,\text{coef}[2]}$
- With a binary IV Z and a binary treatment D, the IV estimator is equal to the CACE estimator – why?

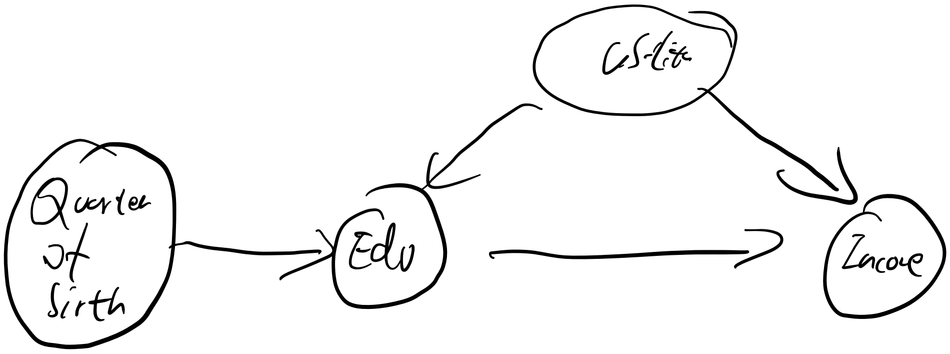
Assumptions for instrumental variable

- Three assumptions for a valid instrumental variable
  - exclusion: Z_i is excluded from the model of Y_i (outcome model)
  - exogeneity: Z_i is uncorrelated with unobserved predictors U_i (latent confounder)
  - rank condition: Z_i is correlated with D_i (treatment indicator)



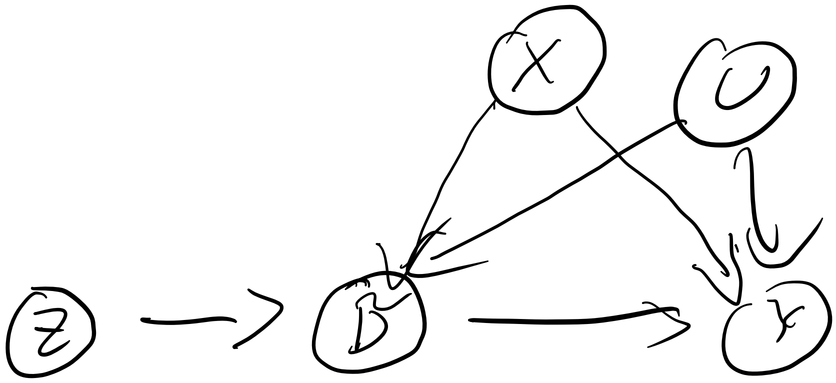
Examples of instrumental variable

- Can we find a valid instrumental variable in practice?
    - good instruments come from knowledge about the processes determining the variable of interest
- Examples
    - Encouragement design
    - Quarter of birth as instrument for educational attainment to address "ability bias" (Angrist and Krueger, 1991)
    - Gender difference as instrument for number of kids (Angrist and Evans, 1998)
    - Mendelian randomization: genotypes are assigned randomly when passed from parents to offspring during meiosis, if mate choice is not associated with genotype, then the population genotype distribution should be unrelated to the confounding factors Voight et al. (2012) studied the causal effect of plasma high-density lipoprotein (HDL) cholesterol on the risk of myocardial infarction



D and Z is a vector

$Y = D^T \beta + \varepsilon \quad , \quad E(\varepsilon | D) \neq 0$



X: observed confounders

$\hat{D}$: treatment indicator

$Y = \beta_0 + \beta_1 \hat{D} + \beta_2 X + \beta_3 U + \tilde{\varepsilon}$

$Y = \beta_0 + \beta_1 \hat{D} + \beta_2 X + \varepsilon$

$$D = (1, \tilde{D}, X)$$

$$Y = D^T \beta + \varepsilon$$

$$\beta = (\beta_0, \beta_1, \beta_2)$$

Just-identified case

What is the dimension of D and Z and what if we observe some confounders that D includes both treatment and some of the confounders?

- $\mathbb{E}(Z_i\epsilon_i) = 0$

- Z_i and D_i have the same dimension and $\mathbb{E}(Z_i D_i^\top)$ has full rank

  $(\mathbb{E} Z_i D_i^\top)^{-1}$ exists

- $\mathbb{E}\{Z_i(Y_i - D_i^\top\beta)\} = 0 \rightsquigarrow \hat{\beta} = \{\mathbb{E}(Z_i D_i^\top)\}^{-1}\mathbb{E}(Z_i Y_i)$

- OLS is a special case when $\mathbb{E}(D_i\epsilon_i) = 0$: $Z_i = D_i$

  (no latent confounders)

$$\beta = \left[\Sigma (z_i D_i^T)\right]^{-1} \Sigma z_i \xi_i$$

$$= \left[\Sigma (z_i D_i^T)\right]^{-1} \Sigma z_i^T z_i \cdot \left(\Sigma z_i^T z_i\right)^{-1}$$

$$\Sigma z_i \xi_i$$

Over-identified case

- Under-identified case: $\mathbb{E}\left(Z_i D_i^\top\right)$ does not have full column rank
  - For example, Z_i has a lower dimension than D
  - $\mathbb{E}(ZY) = \mathbb{E}\left(ZD^\top\right)\beta$ has infinitely many solutions – not identifiable ACE
- Over-identified case: Z_i has a higher dimension than D_i and $\mathbb{E}\left(Z_i D_i^\top\right)$ has full column rank
  - the number of equations is larger than the number of parameters
  - many ways to estimate β
- A computational trick: two-stage least squares (TSLS)

Two-stage least squares (Theil, 1953; Basmann, 1957)

- Procedure
  - run OLS of D_i on Z_i, and obtain the fitted values $\widehat{D}_i$. If D_i is a vector, then we need to run component-wise OLS and put the fitted value together
  - run OLS of Y on $\widehat{D}_i$, and obtain the coefficient $\widehat{\beta}_{2sls}$

$(\hat{\beta}^\top, \hat{D})^\top$ exist

- We do not assume the linearity of D in Z
- The IVs "purge" endogeneity by projection:

$$\widehat{D}_i = \underbrace{Z_i^\top \mathbb{E}\left(Z_i Z_i^\top\right)^{-1} \mathbb{E}\left(Z_i D_i\right)}_{P_z D}$$

  - $Y_i = D_i^\top \beta + \epsilon_i$
  - $D_i = \widehat{D}_i + \widehat{\delta}_i$ with $\mathbb{E}\left(\widehat{D}_i \widehat{\delta}_l^\top\right) = 0$ and $\mathbb{E}\left(\widehat{D}_i \epsilon_i\right) = 0$ – why?
  - $Y_i = \widehat{D}_i^\top \beta + \widehat{\delta}_i^\top \beta + \epsilon_i$

$EZ\epsilon = 0$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$E \hat{D} \varepsilon$$

$$= E D^T P_Z \varepsilon$$

$$P_Z = Z (Z^T Z)^{-1} Z^T \varepsilon$$

Asymptotic inference

- Estimation error – why?

$$\widehat{\beta}_{2sls} - \beta = \left(\sum_{i=1}^{n} \widehat{D}_i \widehat{D}_i^\top\right)^{-1} \sum_{i=1}^{n} \widehat{D}_i \epsilon_i$$

- $\mathbb{E}\left(\widehat{D}_i \epsilon_i\right) = 0 \rightsquigarrow$ Consistency of $\widehat{\beta}_{2sls}$
- Variance estimator

$$\widehat{V}_{2sls} = \left(\sum_{i=1}^{n} \widehat{D}_i \widehat{D}_i^\top\right)^{-1} \left(\sum_{i=1}^{n} \widehat{\epsilon}_i^2 \widehat{D}_i \widehat{D}_i^\top\right) \left(\sum_{i=1}^{n} \widehat{D}_i \widehat{D}_i^\top\right)^{-1}$$

- $\widehat{\epsilon}_i = Y_i - D_i^\top \widehat{\beta}$, not the residual from the second stage least squares

Classical instrumental variables estimator

- Linear model: $Y_i = D_i^\top \beta + \epsilon$ where $\mathbb{E}(\epsilon_i) = 0$ and D_i is a vector with length K – how about D for Z?
- Endogeneity:

$$\mathbb{E}(\epsilon_i \mid D_i) \neq 0$$

- Instruments Z_i is a vector with length L :
  1. Exogeneity: $\mathbb{E}(Z_i \epsilon_i) = 0$
  2. Exclusion restriction: Z_i does not belong to the outcome model
  3. Rank condition: $\mathbb{E}\left(Z_i D_i^\top\right)$ have full column rank $\leftarrow$ $\mathrm{cov}(Z, D) \neq 0$
- Identification: $\mathbb{E}(Z_i Y) = \mathbb{E}\left(Z_i D_i^\top\right)\beta$
  - $K = L$: just-identified $\rightsquigarrow$ we can directly solve $\beta = \mathbb{E}\left(Z_i D_i^\top\right)^{-1}\mathbb{E}(Z_i Y_i)$
  - $K > L$: under-identified $\rightsquigarrow$ there are many β satisfy

$$\mathbb{E}(Z_i Y) = \mathbb{E}\left(Z_i D_i^\top\right)\beta$$

  - $K < L$: over-identified $\rightsquigarrow$ there are many ways to estimate β

$$Y = \beta_0 + \beta_1 \tilde{D} + \beta_2 X + \beta_3 U + \varepsilon$$

X: observed confounders

U: latent confounders

$$Y = \beta^T D + \varepsilon$$

$$D = (1, \hat{D}, X)$$

$$\varepsilon = \beta_3 U + \tilde{\varepsilon}$$

$$E(\varepsilon | X) = 0$$

$$E(\tilde{\varepsilon} | X) = 0$$

$$\beta_3 E(U | X) \overset{?}{=} 0$$

IV for a single endogenous treatment

$E(\varepsilon \mid D, X) \neq 0$

- Consider the linear models

① (linear Model) ⟶

$$Y_i = \beta_0 + \beta_1 D_i + X_i^\top \beta_X + \epsilon_i$$

$$D_i = \gamma_0 + \gamma_1 Z_i + X_i^\top \gamma_X + \eta_i$$

$E(\varepsilon \mid X) = 0$

②

- Endogenous variable D_i; IV Z_i; other exogenous regressors X_i – when can part of confounders be treated as X_i?
- Two-stage least squares estimator: $\widehat{\beta}_{1,2sls}$
- Reduced form

$$Y_i = \theta_0 + \theta_1 Z_i + X_i^\top \theta_X + \epsilon_i'$$

$$D_i = \gamma_0 + \gamma_1 Z_i + X_i^\top \gamma_X + \eta_i$$

$E(\varepsilon_i' \mid Z, X) = 0$

- $\theta_1 = \beta_1 \gamma_1$

$\varepsilon_i' = \beta_1 \eta_i + \varepsilon_i$

- Indirect least squares estimator: $\widehat{\beta}_{1,ils} = \widehat{\theta}_1 / \widehat{\gamma}_1$
- We can show that $\widehat{\beta}_{1,ils} = \widehat{\beta}_{1,2sls}$

$$Y_i = \beta_0 + \beta_1 (r_0 + r_1 z_i + X_i^T r_x + \eta_i) + X_i^T \beta_x + \varepsilon_i$$

$$= \overline{\beta_0 + \beta_1 r_0} + \overline{\beta_1 r_1} z_i + \overline{(\beta_1 r_x + \beta_x)}^T X_i + \overline{\beta_1 \eta_i + \varepsilon_i}$$

$$= \theta_0 + \theta_1 z_i + \theta_x^T X_i + \varepsilon_i'$$

$$\varepsilon_i' = \beta_1 \eta_i + \varepsilon_i$$

Anderson-Rubin confidence interval

$H_0 : \beta_1 = b \longrightarrow H_0 : \theta_1 - b\gamma_1 = 0$

- A more robust inference procedure

$$Y - bD_i = (\theta_0 - b\gamma_0) + (\theta_1 - b\gamma_1)\, Z_i + X_i^\top (\theta_X - b\gamma_X) + (\epsilon - b\eta_i)$$

  – why?

$\theta_1 - b\gamma_1$ $\qquad$ $\theta_1 = \beta\gamma_1$

- If $b = \beta_1$, then the coefficient of Z_i is zero

$Y - bD_i \sim Z_i + X_i$ $\quad \beta Z_i = 0$?

- Inverting tests for $H_0(b) : \beta_1 = b$ to obtain a confidence interval for β_1

$$\{b : |t_Z(b)| \leqslant z_\alpha\}$$

$t_Z(b)$ is the t-statistic for the coefficient of Z based on the OLS fit with the robust standard error

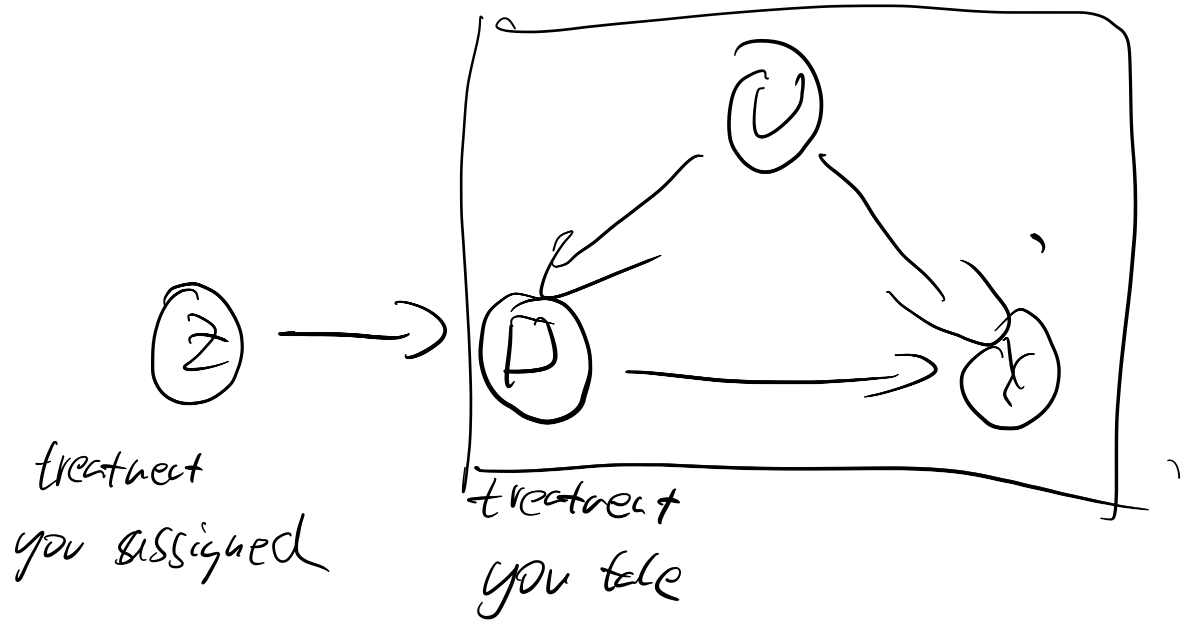
① $Y = \beta^T D + \varepsilon$

Assumption: $E Z \varepsilon = 0$ and $E Z D^T$ full column rank

② $Y = \beta_0 + \beta_1 D + \beta_X^T X + \varepsilon$

$D = \gamma_0 + \gamma_1 Z + \gamma_X^T X + \eta$

Assumption: $E X \varepsilon = 0$



$$E(Y(1) - Y(0))$$

$Y(1)$, take treatment 1

$Y(0)$, take treatment 0

$E Z \varepsilon = 0$ ✓

$cov(Z, D) \neq 0$

$$E(Y'(1) - Y'(0))$$

$Y'(1)$: assigned treatment 1

$Y'(0)$: ... 0

Connection between design and econometric perspectives

- Two-stage least squares estimates the CACE in non-compliance setting
- Implications of $Y_i(z, d) = \beta_0 + \beta_1 d + \beta_2 U_i + \eta_i$
  - $Y_i(z, d) = Y_i\left(z', d\right)$
  - $E\left(Y_i\left(D_i = 1\right) - Y_i\left(D_i = 0\right)\right)$ is the same for different people
  - $Z_i \perp Y_i(z, d) \Longleftrightarrow Z_i \perp U_i$

  Here, Z is the assigned treatment which can be often used as IV and D is the treatment received.
- Comparison
  - without modeling assumption, we can only estimate the treatment effect for compliers
  - generalization to other populations comes from the model

Quarter of Birth (Angrist and Krueger. 1991.)

- Instrument for educational attainment to address "ability bias"
  - Outcome: men's log weekly earnings in 1980
  - Compulsory education law in US: students must attend school until they reach age 16
  - Those born in the third or fourth quarter typically finish tenth grade before reaching age 16
  - Instrument at most decreases years of education by one year
- Instrument: first quarter vs. 2nd to 4th quarter $D \sim Z$
  - 1920s cohorts: est. $= -0.126$, s.e. (HC) $= 0.016$, corr $= -0.016$
  - 1930s cohorts: est. $= -0.109$, s.e. (HC) $= 0.013$, corr $= -0.014$
- IV estimates: two-stage estimate
  - 1920 cohorts: est. $= 0.072$, s.e. (HC) $= 0.022$
  - 1930 cohorts: est. $= 0.102$, s.e. $(HC) = 0.024$
- OLS estimates: $Y \sim D$
  - 1920 cohorts: est. $= 0.080$, s.e. (HC) $= 0.0004$
  - 1930 cohorts: est. $= 0.071$, s.e. (HC) $= 0.0004$

Discussion

- Weekly earnings Y_j, years of education D_i
  - causal quantity: $\mathbb{E}\{\log Y_i(d)\} - \mathbb{E}\left\{\log Y_i\left(d'\right)\right\}$
  - difference from $\log \mathbb{E}\{Y_i(d)\} - \log \mathbb{E}\left\{Y_i\left(d'\right)\right\}$

- $\log Y_i = D_i\beta + \mathbf{X}_i\beta_X + \epsilon_i$
  - equivalent to $Y_i = \exp\left(D_i\beta + \mathbf{X}_i\beta_X + \epsilon_i\right) = \exp\left(D_i\beta + \mathbf{X}_i\beta_X\right)\exp\left(\epsilon_i\right)$
    $\rightsquigarrow \log Y_i(d) - \log Y_i\left(d'\right) = \beta\left(d - d'\right)$

  - different from $Y_i = D_i\beta + \mathbf{X}_i\beta_X + \epsilon_i \rightsquigarrow Y_i(d) - Y_i\left(d'\right) = \beta\left(d - d'\right)$

  - different from $Y_i = \exp\left(D_i\beta + \mathbf{X}_i\beta_X\right) + \epsilon_i$
    $\rightsquigarrow Y_i(d) - Y_i\left(d'\right) = \exp\left(d\beta + \mathbf{X}_i\beta_X\right) - \exp\left(d'\beta + \mathbf{X}_i\beta_X\right)$

Fuzzy RD design

- Sharp regression discontinuity design: $Z_i = \mathbf{1}\{X_i \geqslant c\}$
- What happens if we have noncompliance?
- Forcing variable as an instrument: $Z_i = \mathbf{1}\{X_i \geqslant c\}$
- Potential outcomes: $D_i(z)$ and $Y_i(z)$
- Assumptions:
  - Monotonicity: $D_i(1) \geqslant D_i(0)$
  - Exclusion restriction: $Y_i(0) = Y_i(1)$ for units with $D_i(1) = D_i(0)$
  - $\mathbb{E}\{D_i(z) \mid X_i = x\}$ and $\mathbb{E}\{Y_i(z) \mid X_i = x\}$ are continuous in x
- No randomization is needed

$D_i(z)$ $D_i(1)$

$Y_i(z)$ $Y_i(1)$

compliance

$(D_i(1), D_i(0)) = (1, 0)$

$D_i(1) \geq D_i(0)$ excludes

$(D_i(1), D_i(0)) = (0, 1)$

$D_i(1) = D_i(0) \implies Y_i(1) = Y_i(0)$

includes $(D_i(1), D_i(0)) = (1, 1)$

$(D_i(1), D_i(0)) = (0, 0)$

$\implies Y_i(1) = Y_i(0)$

Examples

- Prime Minister's Village Road Program in India
  - program prioritized larger villages
  - X_i : population size; D_i : village i received a road
- Effect of grant on dropout rate
  - students were eligible for a university grant if their standardized family income was below 15,000 euros
  - X_i : family income; D_i : student i received a grant

Identification

- CACE – why and how is it related to estimator under IV?:

$$\mathbb{E}\left\{Y_i(1)-Y_i(0) \mid \text{ Complier }, X_i=c\right\}=\frac{\mathbb{E}\left\{Y_i(1)-Y_i(0) \mid X_i=c\right\}}{\mathbb{E}\left\{D_i(1)-D_i(0) \mid X_i=c\right\}}$$

why conditioning on X?

- Identification formula:

$$\frac{\lim_{x \downarrow c} \mathbb{E}\left(Y_i \mid X_i=x\right)-\lim_{x \uparrow c} \mathbb{E}\left(Y_i \mid X_i=x\right)}{\lim_{x \downarrow c} \mathbb{E}\left(D_i \mid X_i=x\right)-\lim_{x \uparrow c} \mathbb{E}\left(D_i \mid X_i=x\right)}$$

if $D_i(1) - D_i(0) = 1$

$$E[Y_i(1) - Y_i(0) \mid X = c]$$

$$= E[Y_i(1) - Y_i(0) \mid X = c]$$

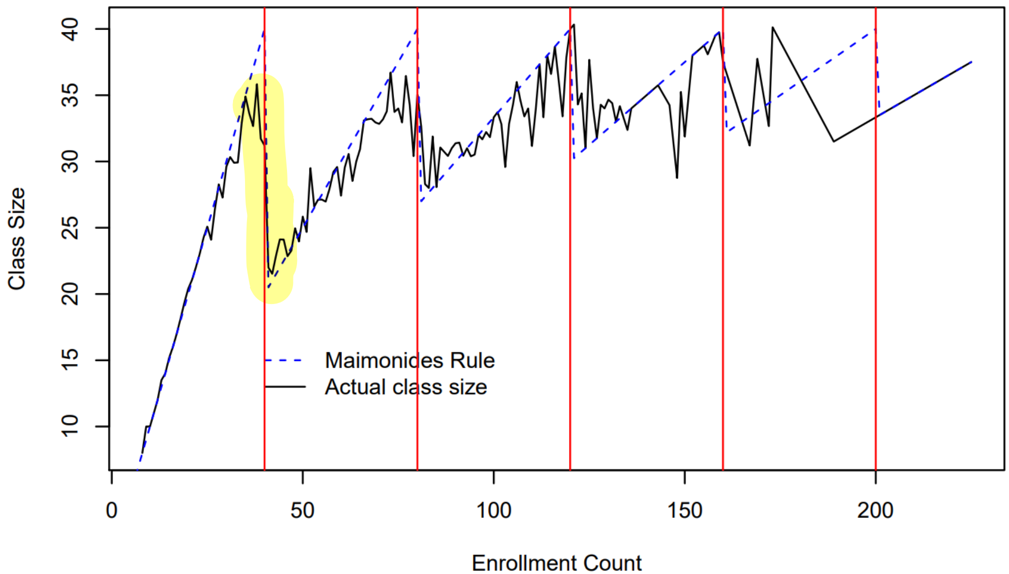
Inference

- Determine the neighborhood of c. Focus on the data with $X_i \in [c-h, c+h]$
- Estimate $\mathbb{E}\{D_i(1) - D_i(0) \mid X_i = c\}$ by the regression coefficient of D_i on $\{1, Z_i, R_i, L_i\}$, where $L_i = \min(X_i - c, 0)$ and $R_i = \max(X_i - c, 0)$
- Estimate $\mathbb{E}\{Y_i(1) - Y_i(0) \mid X_i = c\}$ by the regression coefficient of Y_i on $\{1, Z_i, R_i, L_i\}$
- The estimator is the ratio of the two coefficients
- It is equal to the TSLS fit of the regression of Y_i on $\{1, D_i, R_i, L_i\}$ with D_i instrumented by Z_i
- rdrobust package selects the bandwidth automatically

Class size effect (Angrist and Lavy. 1999.)

- Effect of class-size on student test scores
- Maimonides' Rule: Maximum class size = 40

$$f(\text{ count }) = \frac{\text{count}}{\left\lfloor \frac{\text{count-1}}{40} \right\rfloor + 1}$$



Empirical analysis

- Y_i : class average verbal test score
- D_i : class size; X_i : enrollment at the beginning of the year
- Window size: h
- Construction of forcing variable:

$$X_i = \begin{cases} 40 - \text{ count} & \text{if } 40 - h/2 \leqslant \text{ count } \leqslant 40 + h/2 \\ 80 - \text{ count} & \text{if } 80 - h/2 \leqslant \text{ count } \leqslant 80 + h/2 \\ \vdots & \vdots \end{cases}$$

- Linear models (cluster standard errors by schools):

$$D_i = \delta_1 + \alpha_1 \times I\left(X_i \geqslant 0\right) + \beta_1 X_i + \gamma_1 X_i \times I\left(X_i \geqslant 0\right) + \epsilon_{1i}$$
$$Y_i = \delta_2 + \alpha_2 \times I\left(X_i \geqslant 0\right) + \beta_2 X_i + \gamma_2 X_i \times I\left(X_i \geqslant 0\right) + \epsilon_{2i}$$

where $\hat{\alpha}_1 = -7.90($ s.e. $= 1.90)$ and $\hat{\alpha}_2 = -0.056$ (s.e. $= 2.08$)

- Two-stage least squares estimate: est. $= 0.007$ (s.e. $= 0.261$)

$$E\left(Y(X_i \geq 0) - Y(X_i < 0) \mid \text{complier}, X_i = 0 \right)$$

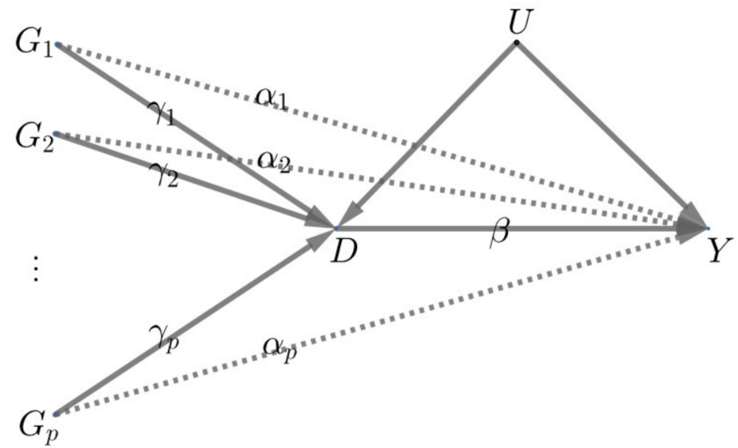
$$\begin{tikzcd} & U \arrow[dl] \arrow[dr] & \\ Z \arrow[r] & D \arrow[r] & Y \end{tikzcd}$$

Encouragement Trial

$D(Z)$, $Y(Z)$: potential Outcome

$$E\left(Y(1) - Y(0) \mid \text{complier} \right)$$

Mendelian randomization (MR)



- Use genetic information as IVs to estimate causal effects of exposures on outcomes
  - motivated by Mendel's second law, the law of random assortment: inheritance of one trait is independent of the inheritance of other traits

Mendelian randomization

- $G_{i1}, \ldots, G_{ip}$: single nucleotide polymorphisms (SNPs)
- Standard linear IV model with exclusion restriction

$$Y_i = \beta_0 + \beta D_i + \beta_u U_i + \epsilon_{i1}$$
$$D_i = \gamma_0 + \gamma_1 G_{i1} + \cdots + \gamma_p G_{ip} + \gamma_u U_i + \epsilon_{i2}$$

- Pleiotropy: one gene may influence more than one phenotypic trait
- Linear IV model without exclusion restriction

$$Y_i = \beta_0 + \beta D_i + \alpha_1 G_{i1} + \cdots + \alpha_p G_{ip} + \beta_u U_i + \epsilon_{i1}$$
$$D_i = \gamma_0 + \gamma_1 G_{i1} + \cdots + \gamma_p G_{ip} + \gamma_u U_i + \epsilon_{i2}$$

Mendelian randomization

- Reduced form with exclusion restriction

$$Y_i = \beta_0 + \beta\gamma_0 + \beta\gamma_1 G_{i1} + \cdots + \beta\gamma_p G_{ip} + \beta_u U_i + \epsilon_{i1}$$
$$D_i = \gamma_0 + \gamma_1 G_{i1} + \cdots + \gamma_p G_{ip} + \gamma_u U_i + \epsilon_{i2}$$

- Reduced form without exclusion restriction

$$Y_i = \beta_0 + \beta\gamma_0 + (\alpha_1 + \beta\gamma_1)\, G_{i1} + \cdots + (\alpha_p + \beta\gamma_p)\, G_{ip} + (\beta_u + \beta\gamma_u)\, U_i + \epsilon_{i1}$$
$$D_i = \gamma_0 + \gamma_1 G_{i1} + \cdots + \gamma_p G_{ip} + \gamma_u U_i + \epsilon_{i2}$$

- Denote the regression coefficient of Y_i on G_{ij} as Γ_j for $j = 1, \cdots, p$
  - $\Gamma_j = \beta\gamma_j$ with exclusion restriction
  - $\Gamma_j = \alpha_j + \beta\gamma_j$ without exclusion restriction

b on G

Quantity of Interest : β

Independence of G_j

$\Rightarrow$ $\hat{\tau}_1, \dots, \hat{\tau}_p$ uncorrelated

$\hat{\gamma}_1, \dots, \hat{\gamma}_p$

$\Rightarrow$ $\hat{\beta}_1, \dots, \hat{\beta}_p$ independent

$Y \sim G_1 + \dots + G_p$

$$\mathrm{cov}\begin{pmatrix}\hat{\tau}_1\\ \vdots \\ \hat{\tau}_p\end{pmatrix} = \begin{pmatrix} \ddots & 0 \\ 0 & \ddots \end{pmatrix}$$

$$\sigma^2 \Sigma^{-1}_{(G_1, \dots, G_p)}$$

$\hat{\beta}_1 = \frac{\hat{\tau}_1}{\hat{\gamma}_1}, \dots, \hat{\beta}_p = \frac{\hat{\tau}_p}{\hat{\gamma}_p}$ independent

$\hat{\beta} = \sum_{i=1}^{p} w_i \hat{\beta}_i$ $\quad \sum w_i = 1$ $\quad \widehat{se}_i^2$

$\min \mathrm{Var}(\hat{\beta}) = \min \sum w_i^2 \mathrm{Var}(\hat{\beta}_i^2)$

$$w_i = \frac{\frac{1}{se_i^2}}{\sum \frac{1}{se_i^2}}$$

Mendelian randomization based on summary statistics

- Most studies do not have individual data but rather summary statistics from genome-wide associate studies

- Data structure and assumptions
  - we have regression coefficients of D_i on G_{ij} 's: $\hat{\gamma}_1, \ldots, \hat{\gamma}_p$ and their standard errors $\text{se}_{D1}, \ldots, \text{se}_{Dp}$
  - we have regression coefficients of Y_i on G_{ij} 's: $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_p$ and their standard errors $\text{se}_{Y1}, \ldots, \text{se}_{Y_p}$
  - assume the coefficients $\hat{\gamma}_1, \ldots, \hat{\gamma}_p, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_p$ are consistent, jointly Normal, and independent, and ignore the uncertainty in the standard errors

Meta-analysis

- With exclusion restriction, $\beta = \Gamma_j/\gamma_j$ for all j
- Fixed-effect meta-analysis: combine multiple estimates $\hat{\beta}_j = \hat{\Gamma}_j/\hat{\gamma}_j$ for the common β
- $\hat{\beta}_j$ has approximate variance

$$\mathrm{se}_j^2 = \left(\mathrm{se}_{Y_j}^2 + \hat{\beta}_j^2 \mathrm{se}_{Dj}^2\right)/\hat{\gamma}_j^2$$

- Best linear combination – why?

$$\hat{\beta}_{\text{fisher0}} = \frac{\sum_{j=1}^p \hat{\beta}_j/\mathrm{se}_j^2}{\sum_{j=1}^p 1/\mathrm{se}_j^2}$$

- Best linear combination ignoring the uncertainty in $\hat{\gamma}_j$

without exclusion

$$\sum_{j=1}^{p} \alpha_j \gamma_j = 0$$

$$\hat{\beta}_{\text{fisher1}} = \frac{\sum_{j=1}^p \hat{\beta}_j \hat{\gamma}_j^2/\mathrm{se}_{Y_j}^2}{\sum_{j=1}^p \hat{\gamma}_j^2/\mathrm{se}_{Y_j}^2}$$

$G_1, \ldots, G_p$ independent not for validity

① derive best linear combination

② based on summary statistics

③ marginal regression/correlation works

$$\hat{\beta}_j \hat{\gamma}_j^2 = \frac{\hat{\tau}_j}{\hat{\gamma}_j} \hat{\gamma}_j^2 = \hat{\tau}_j \hat{\gamma}_j$$
$$= \alpha_j \hat{\gamma}_j + \beta \hat{\gamma}_j^2$$

$\tau_j = \beta \gamma_j$ with exclusion

$\tau_j = \alpha_j + \beta \gamma_j$ without exclusion

$$Y = \begin{pmatrix} \hat{\tau}_1 \\ \vdots \\ \hat{\tau}_p \end{pmatrix} \qquad X = \begin{pmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_p \end{pmatrix}$$

$Y \sim X - 1$

Egger regression

- Use least squares to estimate β
- With exclusion restriction, run WLS of $\hat{\Gamma}_j$ on $\hat{\gamma}_j$ without an intercept

$$\hat{\beta}_{\text{egger1}} = \frac{\sum_{j=1}^{p} \hat{\gamma}_j \hat{\Gamma}_j w_j \cancel{/\mathrm{se}_j^2}}{\sum_{j=1}^{p} \hat{\gamma}_j^2 w_j}$$

which equals $\hat{\beta}_{\text{fisher 1}}$ if $w_j = 1/\mathrm{se}_{Y_j}^2$

- With exclusion restriction, run WLS of $\hat{\Gamma}_j$ on $\hat{\gamma}_j$ with an intercept

$$\hat{\beta}_{\text{egger0}} = \frac{\sum_{j=1}^{p} \left(\hat{\gamma}_j - \hat{\gamma}_w\right)\left(\hat{\Gamma}_j - \hat{\Gamma}_w\right) w_j/\mathrm{se}_j^2}{\sum_{j=1}^{p} \left(\hat{\gamma}_j - \hat{\gamma}_w\right)^2 w_j}$$

with $\hat{\gamma}_w = \sum_{j=1}^{p} w_j \hat{\gamma}_j / \sum_{j=1}^{p} w_j$ and $\hat{\Gamma}_w = \sum_{j=1}^{p} w_j \hat{\Gamma}_j / \sum_{j=1}^{p} w_j$

- Estimated intercept characterizes the violation of exclusion restriction to some extent

Discussion

- Exposures in MR studies are often not well defined
- IV assumptions may not hold
- Relies on strong modeling assumptions

Summary

- Instrumental variables as a general strategy for coping with selection bias (latent confounding)
  - exclusion restriction
  - randomization of instruments
  - rank condition
- Weak instruments $\rightsquigarrow$ transformed outcome $Y_i - bD_i$
- Fuzzy RD designs: instrumental variable is determined by the forcing variable
- Mendelian randomization: genetic information as IVs

Suggested readings

- Instrumental variable methods
  - Angrist and Pischke. Chapter 4
  - Wooldrige. Econometric Analysis of Cross Section and Panel Data. Chapter 5

- Fuzzy RD design
  - Angrist and Pischke. Chapter 6.2

- Mendelian randomization
  - Ding's book Chapter 25
  - VanderWeele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M., and Kraft, P. (2014). "Methodological challenges in Mendelian randomization"

IV for nonlinear models

- $Y_i = f(D_i, \mathbf{X}_i; \beta) + \epsilon_i$ with $\operatorname{cov}(Z_i, \epsilon) = 0 \rightsquigarrow$ moment estimation

- $Y_i = f(D_i, \mathbf{X}_i, \epsilon_i; \beta)$ with $\operatorname{cov}(Z_i, \epsilon) = 0 \rightsquigarrow$ needs distributional assumption

- Control function method for non-separable models: $\epsilon \perp D_i \mid C_i$, where $C_i = F_{D|Z}(D_i, Z_i)$