Instrumental Variables: Extension and Application

Huan Deng

April 29, 2025

Outline

Characterizing Compliers

Overview

- LATE is specific to the instrument: a different IV might identify LATEs for for different subpopulations.
- We would therefore like to learn as much as we can about the compliers for different instruments.
- We can compute the proportion of compliers and characterize the distributions of compliers.

Counting

- We know that with monotonicity, the first stage identifies the size of the compliers:

$$\begin{aligned} P[D_i(1) > D_i(0)] &= E[D_i(1) - D_i(0)] \\ &= E[D_i(1)] - E[D_i(0)] \\ &= E[D_i|Z_i = 1] - E[D_i|Z_i = 0] \end{aligned} \tag{1}$$

- We can also count what proportion of the treated are compliers:

$$\begin{aligned} P[D_i(1) > D_i(0)|D_i = 1] &= \frac{P[D_i = 1|D_i(1) > D_i(0)]P[D_i(1) > D_i(0)]}{P[D_i = 1]} \\ &= \frac{P[Z_i = 1](E[D_i|Z_i = 1] - E[D_i|Z_i = 0])}{P[D_i = 1]} \end{aligned} \tag{2}$$

Characterizing

- We can't observe who are compliers because we never observe $D_i(1)$ and $D_i(0)$ at the same time.
- But luckily, we can still describe the distribution of compliers' characteristics.
- For example, whether the proportion of college graduates is higher among the compliers than in the population. This will help us assess to what extent our causal estimates from IV can be generalized to other settings (Internal Validity V.S. External Validity)
- Let X be a binary variable indicating college education, we have:

$$\begin{aligned}\frac{P[X_i = 1|D_i(1) > D_i(0)]}{P[X_i = 1]} &= \frac{P[D_i(1) > D_i(0)|X_i = 1]}{P[D_i(1) > D_i(0)]} \\ &= \frac{E[D_i|Z_i = 1, X_i = 1] - E[D_i|Z_i = 0, X_i = 1]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}\end{aligned} \tag{3}$$

Abadie's Kappa

- Abadie (2003) offers a general method for constructing any features of the joint distribution of (Y, D, X)

Theorem (Abadie's Kappa)

Suppose the assumptions of the LATE hold conditional on covariates, X_i. Let $g(\cdot)$ be any measurable real function of (Y, D, X) such that $\mathbb{E}[g(Y, D, X)] < \infty$. Define

$$\kappa_0 = (1 - D)\frac{(1 - Z) - P(Z = 0|X)}{P(Z = 0|X)P(Z = 1|X)},$$

$$\kappa_1 = D\frac{Z - P(Z = 1|X)}{P(Z = 0|X)P(Z = 1|X)},$$

$$\kappa = \kappa_0 P(Z = 0|X) + \kappa_1 P(Z = 1|X) = 1 - \frac{D(1 - Z)}{P(Z = 0|X)} - \frac{(1 - D)Z}{P(Z = 1|X)}.$$

Abadie's Kappa

(a) $\mathbb{E}[g(Y,D,X)|D(1)>D(0)] = \frac{1}{P(D(1)>D(0))}\mathbb{E}[\kappa g(Y,D,X)]$.

(b) $\mathbb{E}[g(Y(0),X)|D(1)>D(0)] = \frac{1}{P(D(1)>D(0))}\mathbb{E}[\kappa_0 g(Y,X)]$.

(c) $\mathbb{E}[g(Y(1),X)|D(1)>D(0)] = \frac{1}{P(D(1)>D(0))}\mathbb{E}[\kappa_1 g(Y,X)]$.

Moreover, (a–c) also hold conditional on X.

Proof as exercise.

Hint: any expectation is a weighted average of means for always-takers, never-takers, and compliers.

Applying Abadie's Kappa

- We can apply Abadie's Kappa formulas to construct a variety of interesting outcomes for *compliers*
  - Size of compliers: let $g(Y, D, X) = 1$, we know that $P(D(1) > D(0)) = E(\kappa)$
  - LATE: let $g(Y(0), X) = Y(0)$ and $g(Y(1), X) = Y(1)$, we get LATE by (b) and (C)
  - Means of coviariates: let $g(Y, D, X) = X$, we can compute the mean of X among compliers
  - CDF of potential outcomes: let $g(Y(0), X) = \mathbf{1}(Y(0) \leq y)$ and $g(Y(1), X) = \mathbf{1}(Y(1) \leq y)$

Research Designs That Exploit IVs

Research Designs That Exploit IVs

Fuzzy RD

Overview

- Regression Discontinuity (RD) exploits the situation where treatment is determined by a threshold crossing rule.
- Two types of RD: sharp RD and fuzzy RD
  - Sharp RD: treatment is a discontinuous *deterministic* rule of an observable.
  - Fuzzy RD: the *probability* of treatment is discontinuous in an observable.
- The fuzzy RD is by nature an IV.
- Here, we briefly discuss how RD design works.

Sharp Regression Discontinuity (SRD)

Treatment is a deterministic rule:

$$D_i = 1\{X_i \geq c\}$$

X_i is called the running variable
Conditional Response Surface:

$$m_1(x) = E[Y(1)|X = x]$$

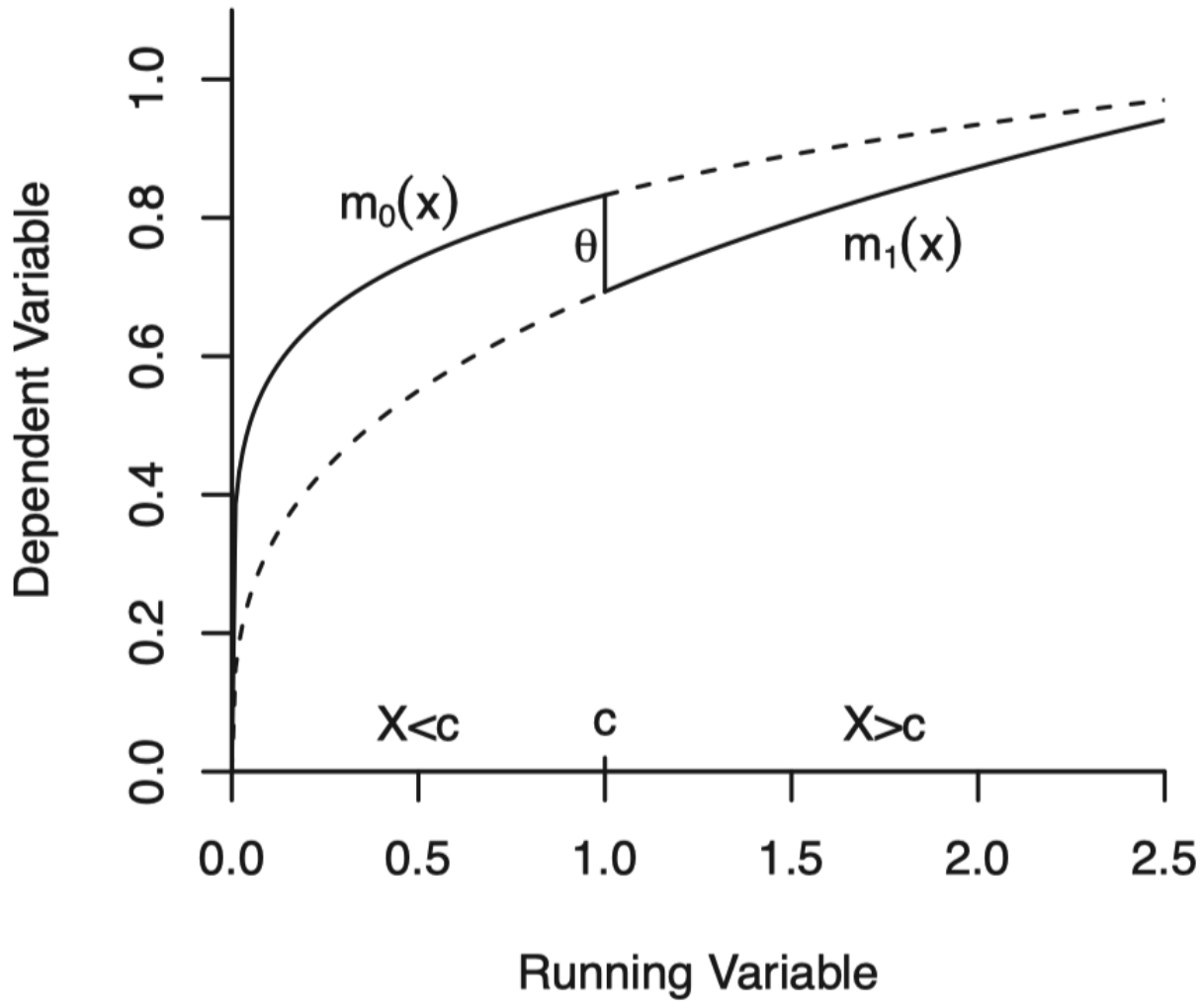
and

$$m_0(x) = E[Y(0)|X = x]$$

CATE at Cutoff:

$$\tau(c) = m_1(c) - m_0(c)$$

Graphical Illustration of Sharp RD (from Hansen's Textbook)



(a) Sharp Regression Discontinuity

Identification in RD

Theorem: (Hahn, Todd, Van der Klaauw, 2001)

- Assume that $m_1(x)$ and $m_0(x)$ are continuous at the cutoff point.
- Then, $\tau(c)$ is identified as:

$$\tau(c) = \lim_{x \to c^+} E[Y|X = x] - \lim_{x \to c^-} E[Y|X = x]$$

- Proof:

$$\begin{aligned}
\tau(c) &= m_1(c) - m_0(c) \\
&= \lim_{x \to c^+} E[Y(1)|X = x] - \lim_{x \to c^-} E[Y(0)|X = x] \\
&= \lim_{x \to c^+} E[Y(1)|X = x, 1\{x \geq c\}] - \lim_{x \to c^-} E[Y(0)|X = x, 1\{x \geq c\}] \\
&= \lim_{x \to c^+} E[Y|X = x] - \lim_{x \to c^-} E[Y|X = x] \qquad (4)
\end{aligned}$$

Fuzzy Regression Discontinuity (FRD)

- Treatment probability jumps at cutoff, but not perfectly determined.
- If there $\exists\ \delta > 0$, such that $D \perp \{Y(0), Y(1)\} | X \in (c - \delta, c + \delta)$, then:

$$\tau(c) = \frac{\lim_{x \to c^+} E[Y|X = x] - \lim_{x \to c^-} E[Y|X = x]}{\lim_{x \to c^+} P[D|X = x] - \lim_{x \to c^-} P[D|X = x]}$$

- This is a Wald/IV estimator using $T = 1\{X \geq c\}$ as an instrument.

Fuzzy Regression Discontinuity (FRD)

- Consider $Y = Y(0) + (Y(1) - Y(0))D$, take conditional expectation at x which is within a small neighborhood of the cutoff point:

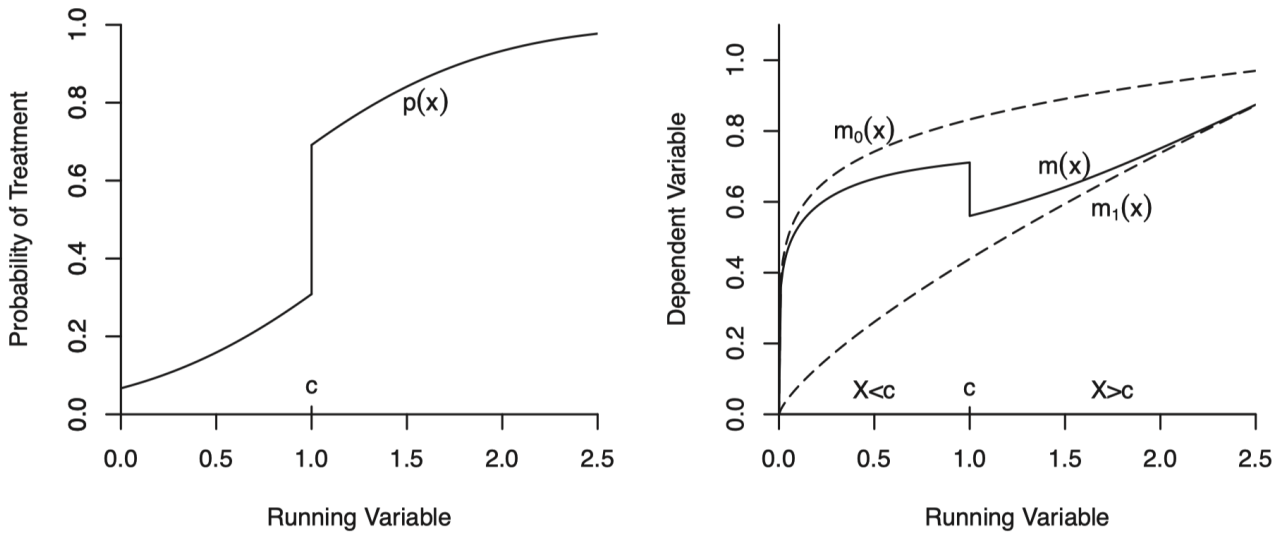
$$E[Y|X = x] = E[Y(0)|X = x] + E[(Y(1) - Y(0))D|X = x] \tag{5}$$

- Take limit from the right and left side of the cutoff and use the independence assumption:

$$\begin{aligned}
\lim_{x \to c^+} E[Y|X = x] &= \lim_{x \to c^+} E[Y(0)|X = x] + \lim_{x \to c^+} E[(Y(1) - Y(0))D|X = x] \\
&= m_0(c) + \lim_{x \to c^+} E[Y(1) - Y(0)|X = x] \lim_{x \to c^+} E[D|X = x] \\
&= m_0(c) + E[Y(1) - Y(0)|X = c] \lim_{x \to c^+} E[D|X = x] \qquad (6) \\
\lim_{x \to c^-} E[Y|X = x] &= \lim_{x \to c^-} E[Y(0)|X = x] + \lim_{x \to c^-} E[(Y(1) - Y(0))D|X = x] \\
&= m_0(c) + \lim_{x \to c^-} E[Y(1) - Y(0)|X = x] \lim_{x \to c^-} E[D|X = x] \\
&= m_0(c) + E[Y(1) - Y(0)|X = c] \lim_{x \to c^-} E[D|X = x] \qquad (7)
\end{aligned}$$

- Take the difference between the above two equations and rearrange terms, we can get the IV-like expression.

Graphical Illustration of Fuzzy RD (from Hansen's Textbook)



(a) Conditional Treatment Probability

(b) Fuzzy Regression Discontinuity

Figure 21.3: Fuzzy Regression Discontinuity Design

Research Designs That Exploit IVs

Simulated IV

overview

- Simulated IV dates back to Currie and Gruber (1996)
- This paper investigates the effects of eligibility of pregnant women for the Medicaid on infant mortality.
- Eligibility is a known but potentially complicated function of regional (for example, state-level) government policy and individual demographics such as income and family structure.
- The individual characteristics are potentially endogenous.
- Here, we discuss how Currie and Gruber (1996) deal with the endogeneity issue and also discuss the recent development of simulated IV by Borusyak and Hull (2023).

Setup

- We want to learn the effect of program participation on some outcome of interest:

$$Y_i = \beta X_i + \epsilon_i \tag{8}$$

  for a fixed population $i = 1, \ldots, N$.
- We have a candidate instrument in mind, which is the program eligibility:

$$Z_i = f(g, w_i) \tag{9}$$

  - g is a vector of policy shocks, which we assume to be exogenous.
  - w denote demographics.
  - $f(\cdot)$ are known mappings from policy and demographics to eligibility.
  - We can also let $Z_i = X_i$, which means the effect of eligibility is of direct interest.

Currie and Gruber (1996)

- Since w is endogenous, Z is not a valid IV because it is a function of w
- So a valid IV should take out w: we can either integrate it out or "subtract it away"
- Currie and Gruber (1996) propose the use of simulated instruments to integrate out w
- The CG paper simulates the generosity of each state's policy as the average eligibility of a simulated nationally representative sample of individuals, if they were to live in that state.
- $Z_i^{CG} = E_w[f(g_{s_i}, w)] = h(g_{s_i})$
- Z_i^{CG} is only a function of exogenous policy shocks, hence it can serve as an IV.

Borusyak and Hull (2023)

- Borusyak and Hull (2023) take a different approach and propose a recentered IV.
- The expected "instrument" $\mu_i = E[f(g, w_i) \mid w_i] \equiv \int f(g, w_i) dG(g \mid w_i)$ causes the endogeneity problem:

$$E\left[\frac{1}{N}\sum_i Z_i \varepsilon_i\right] = E\left[\frac{1}{N}\sum_i \mu_i \varepsilon_i\right] \neq 0 \tag{10}$$

- $\tilde{z}_i = Z_i - \mu_i$ satisfies:

$$E\left[\frac{1}{N}\sum_i \tilde{Z}_i \varepsilon_i\right] = 0. \tag{11}$$

- Thus, $\tilde{Z}_i$ is a valid instrument as well.

Randomizing Individuals or Randomizing Policies?

- Borusyak and Hull (2023) argue that $\tilde{z}_i$ is more efficient than Z_i^{CG} because the former has a better first-stage prediction of X
- For those people whose eligibility don't vary across states (too rich or too poor), $\tilde{z}_i = 0$, so the first-stage of the recentering approach effectively removes these infra-marginal individuals.
- On the contrary, Z_i^{CG} actually can't affect those people's eligibility, poor first-stage.

Application

Overview

- IVs are widely applied in economic research.
- Here, we will discuss some of the applications.
- For those IVs that I have previously discussed, I won't repeat here.
- This is by no means exhaustive and only reflects my selective reading.

Application

Education

Class Size and Students Performance

- The effects of smaller class size have long been investigated by labor economists interested in the education production function.
- Krueger (1999) exploits the Tennessee STAR class size RCT and finds strong effects of class size reductions on learning.
- Angrist and Lavy (1999) exploit the Maimonides' Rule in Israeli elementary schools.
- "The number of pupils assigned to each teacher is twenty-five. If there are fifty, we appoint two teachers. If there are forty, we appoint an assistant, at the expense of the town"
- if more than 40 students are enrolled, the class is more likely to be split into two. So, the fuzzy RD applies to this situation.

Class Size as a Function of Enrollment (from Angrist and Lavy (1999))

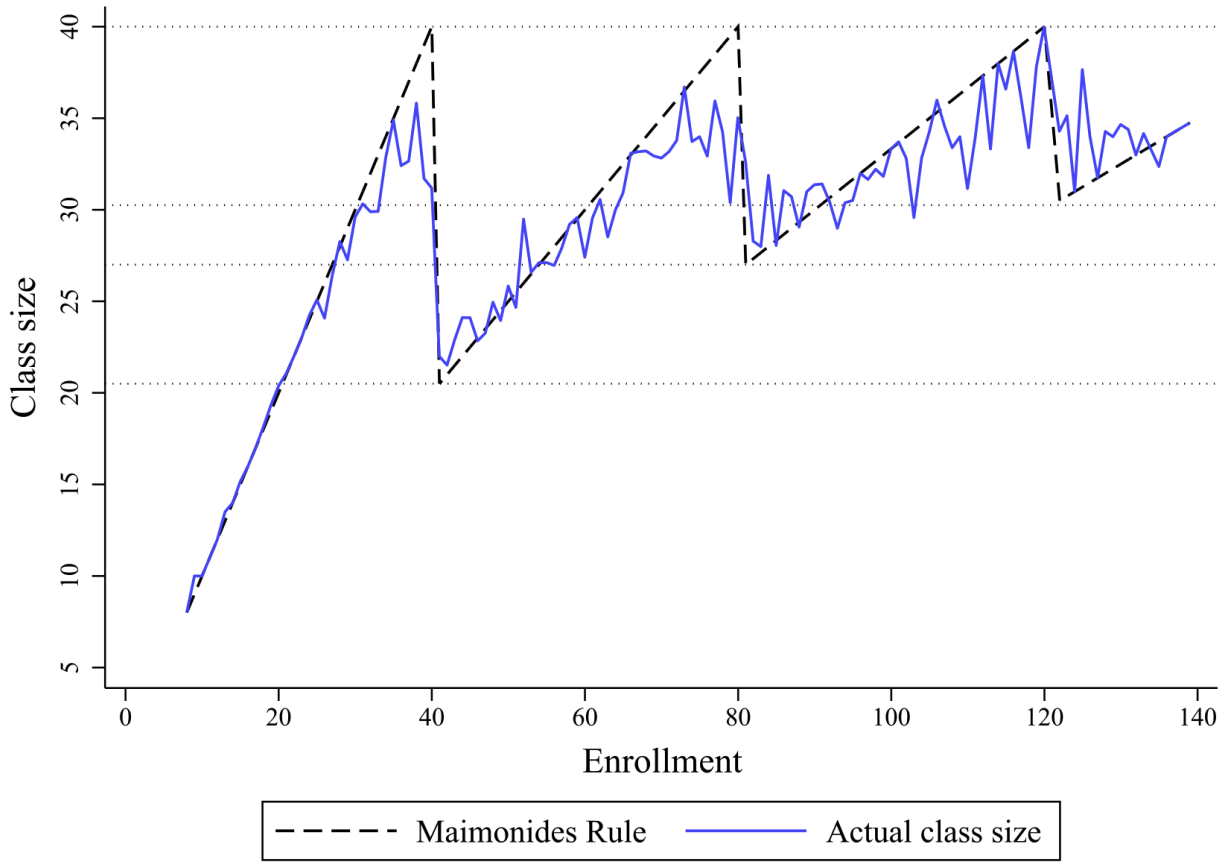


FIGURE 1.—Fourth grade class size by initial enrollment, actual average size and as predicted by Maimonides' Rule. *Notes*: Adapted from Angrist and Lavy (1999); data from Israel in 1991.

Charter School

- Charter schools operate independently from public school districts but are publicly funded.
- KIPP (Knowledge Is Power Program) is a network of public charter schools.
- Core principles:
  - Extended school days and academic years.
  - Selective teacher hiring and data-driven instruction.
  - Strict student behavior norms.
- KIPP primarily serves low-income, minority students.

Randomized Admission via Lottery

- Massachusetts law mandates lottery-based admissions for oversubscribed charter schools.
- Lottery assignment creates random variation in KIPP attendance.
- This allows Angrist et. al. (2010) and Angrist et. al. (2012) to estimate the causal impact of attending KIPP using Instrumental Variables (IV).

Estimating the Effect of KIPP Attendance

$$\text{Effect of KIPP Attendance} = \frac{\text{Effect of Lottery Offer on Scores}}{\text{Effect of Lottery Offer on Attendance}} \tag{12}$$

- Reduced-form estimate: 0.355 standard deviations increase in math scores.
- First-stage effect: 0.74 increase in attendance.
- IV estimate: 0.48 standard deviations improvement. A very large effect! Few education-related interventions have such large effects.

LATE and Complier Interpretation

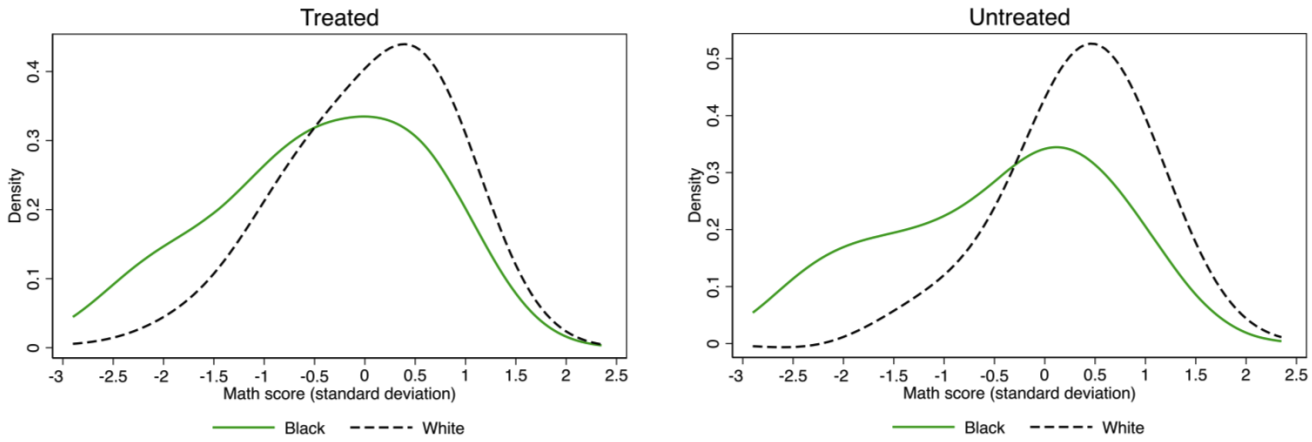
- LATE applies to "compliers"—students whose attendance is influenced by the lottery.
- Always-takers (students who attend KIPP regardless) and never-takers (students who don't attend even if they win) do not contribute to the IV estimate.
- Compliers are children likely to be seated at KIPP were the school to expand and offer additional seats in a lottery.
- The is exactly the subpopulation that relevant policies want to target.

Distributions of Student Achievement

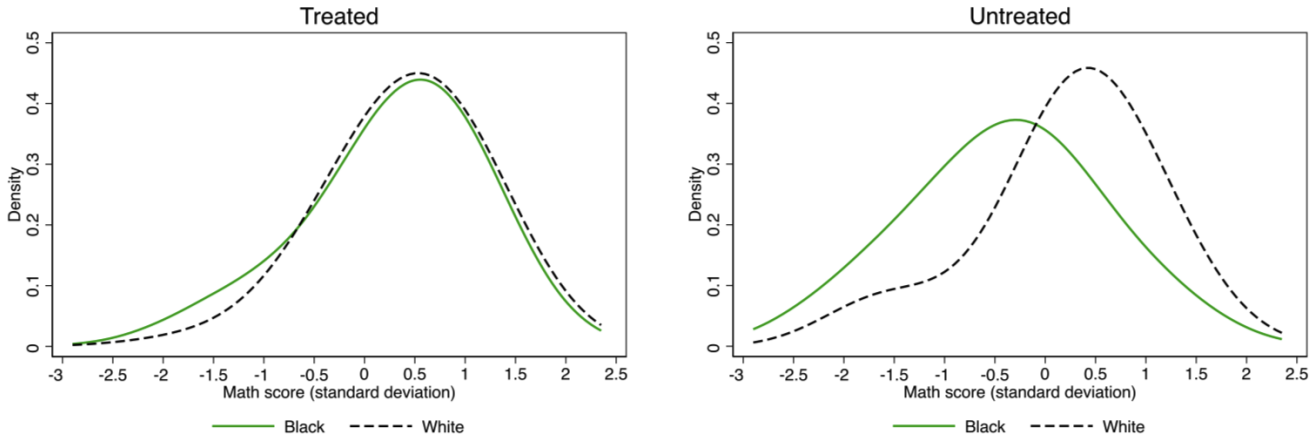
- Comparison of 4th and 8th grade test scores.
- Initial racial achievement gaps in 4th grade.
- By 8th grade, charter school students show substantial improvements.
- Black-White test score gap is significantly reduced among treated students.

Boston Charter Schools Close the Achievement Gap (from Angrist (2022))

A. Before Application (4th Grade Scores)



B. After Application (8th Grade Scores)



Application

Fertility and Female Labor Supply

Female Labor Supply

- Can declining fertility explains increasing female labor supply?
- Reverse causality: Career-oriented women may delay or limit childbirth.
- Angrist and Evans (1998) propose two IVs
- The two IVs identify LATEs for different compliers.
- Angrist and Fernández-Val (2013) offer an explanation for the different LATEs.

Two IVs for Fertility

- Twins IV: the occurrence of twins at second birth in samples of mothers with at least two children.
- Same-Sex IV: parents with at least two children of the same gender
- IV estimates constructed using the same-sex instrument are substantially more negative than the corresponding twins-IV estimates.

IV Estimates of Family Size on Female Labor Supply (from Angrist and Evans (1998))

TABLE II

IV ESTIMATES OF THE EFFECTS OF FAMILY SIZE ON LABOR SUPPLY.

			Twins Instrument		Same-Sex Instrument		Both
Dependent Variable	Mean	OLS (1)	First Stage (2)	IV Estimates (3)	First Stage (4)	IV Estimates (5)	2SLS Estimates (6)
Weeks worked	20.83	−8.98	0.603	−3.28	0.060	−6.36	−3.97
		(0.072)	(0.008)	(0.634)	(0.002)	(1.18)	(0.558)
	Overid: $\chi^2(1)$ (p-value)	—	—	—	—	—	5.3 (0.02)
Employment	0.565	−0.176		−0.076		−0.132	−0.088
		(0.002)		(0.014)		(0.026)	(0.012)
	Overid: $\chi^2(1)$ (p-value)	—	—	—	—	—	3.5 (0.06)

Note: This table reports OLS, IV, and 2SLS estimates of the effects of a third birth on labor supply using twins and sex composition instruments. Data are from the Angrist and Evans (1998) 1980 Census extract containing women aged 21–35 with at least two children. OLS models include controls for mother's age, age at first birth, ages of the first two children, and dummies for race. The sample size is 394,840.

Why Two LATEs Differ?

- Twins compliers are more likely to be college-educated. Highly-educated people have a higher wage and can afford out-of-home child care.
- Same-sex compliers are less likely to have a college degree, and are therefore more likely than twins compliers to use home child care.
- The Twins compliers have larger scope to deal with the third birth, so the negative impacts of a third birth on their labor supply are smaller.

Application

Development

Dams

- Duflo and Pande (2007) study the productivity and distributional effects of large irrigation dams in India.
- Dam placement is not random and may be correlated with economic conditions.
  - Agricultural productivity influences dam placement decisions.
  - Wealthier states may build more dams, biasing OLS estimates.
- Instrumental Variables (IV) provide a solution by exploiting exogenous variation in dam placement.

Instrumental Variable: River Gradient

- Engineering literature shows that river gradient affects dam feasibility.
- Low (but nonzero) river gradient areas are most suitable for irrigation dams.
- Very steep gradients are suitable for hydroelectric dams.
- Regions where the river gradient is either flat or somewhat steep are the least likely to receive dams.

Constructing the IV: River Gradient and Dam Placement

- Use GIS data to measure river gradients within districts.
- Identify districts with gradients that favor dam construction.
- Predict likelihood of dam construction based on river gradient characteristics.

Empirical Findings

- Downstream districts experience increased agricultural production and irrigation coverage.
- Upstream districts see increased production volatility and displacement effects.
- Poverty decreases in downstream districts but rises in dam-hosting regions.
- Neither markets nor state institutions have alleviated the adverse distributional impacts of dam construction.

Slave Trade and Mistrust in Africa

- Slave trade, which occurred over a period of more than 400 years, had a significant negative effect on long-term economic development.
- Nunn and Wntchekon (2011) one of the channels through which the slave trade may affect economic development today: a culture of mistrust within Africa.
- This paper aims to answer a question: Whether ethnic groups that were heavily targeted by the slave trade in the past are less trusting of others today.

Endogeneity in the Relationship Between Slave Trade and Trust

- Trust levels today may be influenced by unobserved historical factors.
- Reverse causality: Societies with lower trust in the past might have been more vulnerable to the slave trade and these lower levels of trust continue to persist today.
- Need for an exogenous instrument to identify a causal effect.

Instrumental Variable: Historical Distance from the Coast

- Distance from the coast during the slave trade is used as an IV for slave trade exposure.
- Coastal regions were more exposed to slave trade due to proximity to European traders.
- Groups farther from the coast experienced lower intensity of slave raids.

Validity of the Instrument

- Relevance: Distance from the coast strongly predicts the number of slaves taken from an ethnic group.
- Exclusion restriction: Distance from the coast affects trust today only through its impact on the slave trade.
- Controls added for colonial influence, trade routes, and geographic factors.

Robustness Checks and Alternative Explanations

- Falsification test: No relationship between distance from the coast and trust outside Africa.
- Results robust to controlling for colonial history, trade routes, and economic conditions.
- Alternative explanations, such as historical political structures, do not fully explain the findings.

Application

Political Economy

AJR (2001)

- Maybe the most well-known paper that uses an IV design is Acemoglu, Johnson and Robinson (2001).
- AJR (2001) investigate the impact of colonial institutions on long-term economic development.
- Endogeneity issue: Richer economies may develop better institutions, leading to reverse causality.
- Historical factors influencing institutional development provide a potential instrument.

Instrumental Variable: European Settler Mortality

- European mortality rates in colonies are used as an instrument for institutions.
- High settler mortality deterred European settlement, leading to extractive institutions.
- Low settler mortality encouraged European settlement and establishment of inclusive institutions.

Validity of the Instrument

- Relevance: Settler mortality strongly predicts the type of institutions established.
- Exclusion restriction: Settler mortality affects current economic outcomes only through institutions.
- Robustness checks include controlling for geographic and climatic factors.
- Results show strong positive effects of inclusive institutions on economic development.

Channel Lineup and Fox News Viewership

- Fox News Channel (FNC) is the first news network to rank number one in total day and prime time ratings for 22 consecutive years
- FNC is also a frequent target of allegations of media bias, and has been researched extensively in the literature.
  - Voting behaviors: DellaVigna and Kaplan (2007), Martin and Yurukoglu (2017)
  - Judicial decisions: Ash and Poyker (2023)
  - Adherence to safety measures during the COVID-19 Pandemic: Bursztyn et al. (2020), Ananyev, Poyker, and Tian (2021), Simonov et al. (2022).

Channel Lineup and Fox News Viewership

- It is not surprising that Fox News viewership can be endogenous
- FNC channel lineup is basically the must-have IV for the strand of literature.
- Variation in channel positions causes some viewers to watch more or less of these channels.
- The exogeneity condition requires that channel positions for Fox News and its competitors are exogenous, and not chosen to accord with local political tastes.