Causal Inference

Lecture #5

Juraj Medzihorsky

SGIA & RMC
Durham University

5 March 2024

IWD

Susan Athey

FOLLOW

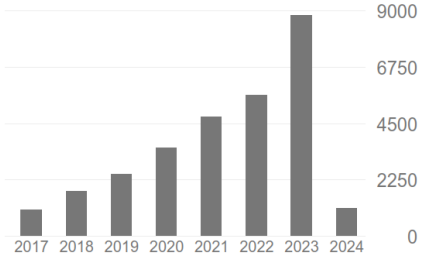
The Economics of Technology Professor, Stanford University Graduate School of Business
Verified email at stanford.edu - Homepage

Market design econometrics auctions news media online advertising

TITLE	CITED BY	YEAR
Estimation and inference of heterogeneous treatment effects using random forests S Wager, S Athey Journal of the American Statistical Association 113 (523), 1228-1242	2945	2018
When should you adjust standard errors for clustering? A Abadie, S Athey, GW Imbens, JM Wooldridge The Quarterly Journal of Economics 138 (1), 1-35	2676	2023
Recursive partitioning for heterogeneous causal effects S Athey, G Imbens Proceedings of the National Academy of Sciences 113 (27), 7353-7360	2053 *	2016
Generalized random forests S Athey, J Tibshirani, S Wager	1909	2019
The state of applied econometrics: Causality and policy evaluation S Athey, GW Imbens Journal of Economic perspectives 31 (2), 3-32	1607	2017

Cited by VIEW ALL

	All	Since 2019
Citations	37613	26851
h-index	75	65
i10-index	125	114



Public access VIEW ALL

1 article 33 articles

not available available

Based on funding mandates

IWD

Tamar Sofer

FOLLOW

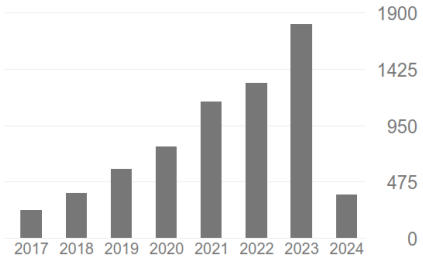
Associate Professor, Beth Israel Deaconess Medical Center, Harvard Medical School
Verified email at bidmc.harvard.edu - Homepage

Statistical Genetics Genetic Epidemiology Sleep Genetics + 'Omics Diverse populations High

TITLE	CITED BY	YEAR
Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models H Chen, C Wang, MP Conomos, AM Stilp, Z Li, T Sofer, AA Szpiro, ... The American Journal of Human Genetics 98 (4), 653-666	396	2016
The trans-ancestral genomic architecture of glycemic traits J Chen, CN Spracklen, G Marenne, A Varshney, LJ Corbin, J Luan, ... Nature genetics 53 (6), 840-860	337	2021
Genetic diversity and association studies in US Hispanic/Latino populations: applications in the Hispanic Community Health Study/Study of Latinos MP Conomos, CA Laurie, AM Stilp, SM Gogarten, CP McHugh, ... The American Journal of Human Genetics 98 (1), 165-184	301	2016
Genetic association testing using the GENESIS R/Bioconductor package SM Gogarten, T Sofer, H Chen, C Yu, JA Brody, TA Thornton, KM Rice, ... Bioinformatics 35 (24), 5346-5348	280	2019
Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for	238	2022

Cited by VIEW ALL

	All	Since 2019
Citations	6885	6040
h-index	45	41
i10-index	101	100



Public access VIEW ALL

3 articles 154 articles

not available available

Based on funding mandates

Esther Duflo

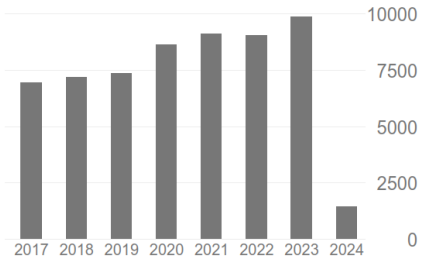
Economics, MIT
Verified email at mit.edu
Economics

FOLLOW

TITLE	CITED BY	YEAR
How much should we trust differences-in-differences estimates? M Bertrand, E Duflo, S Mullainathan The Quarterly journal of economics 119 (1), 249-275	13849	2004
Poor economics: A radical rethinking of the way to fight global poverty AV Banerjee, E Duflo Public Affairs	4912	2011
Women empowerment and economic development E Duflo Journal of Economic literature 50 (4), 1051-1079	3987	2012
The miracle of microfinance? Evidence from a randomized evaluation A Banerjee, E Duflo, R Glennerster, C Kinnan American economic journal: Applied economics 7 (1), 22-53	3167	2015
Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment E Duflo American economic review 91 (4), 795-813	3088	2001

Cited by VIEW ALL

	All	Since 2019
Citations	99958	45772
h-index	105	82
i10-index	230	177



Public access VIEW ALL

0 articles — 36 articles

not available — available

Based on funding mandates

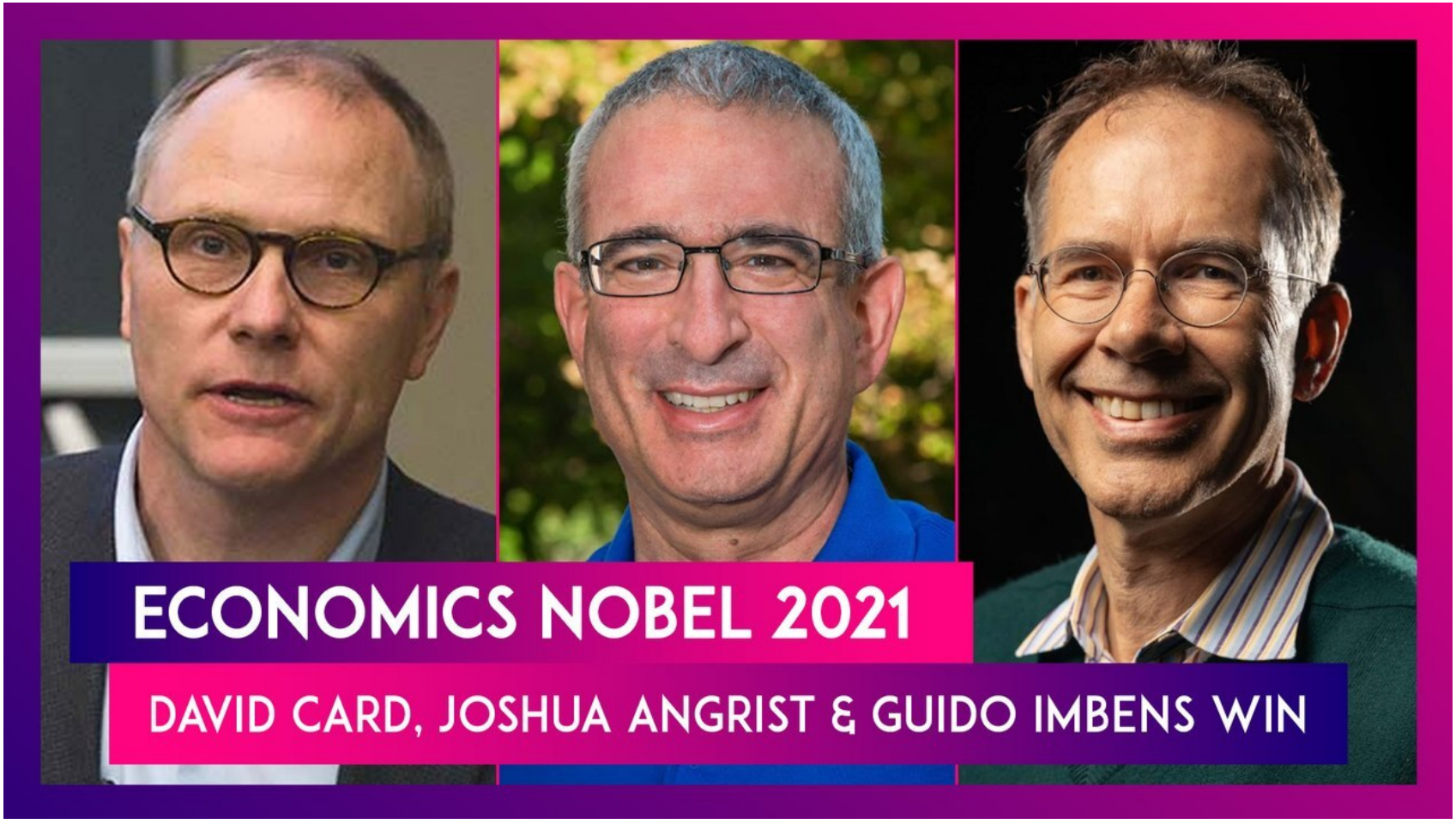
Today

1. Formative and summative?
2. Instrumental Variables
    - Side note 1: Fuzzy RDD
    - Side note 2: Heckman Selection Models
3. A glimpse into today's future
4. Partial Identification

1. Formative and summative?

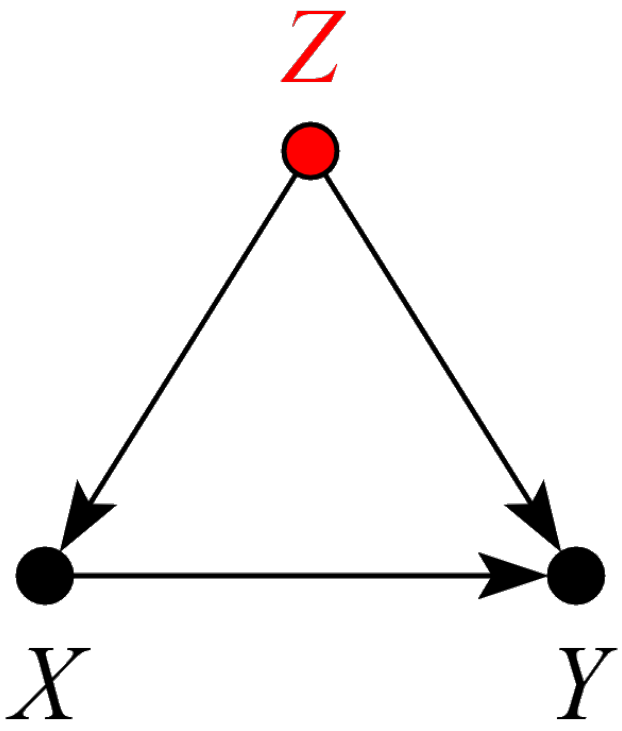
2. Instrumental Variables

Instrumental Variables



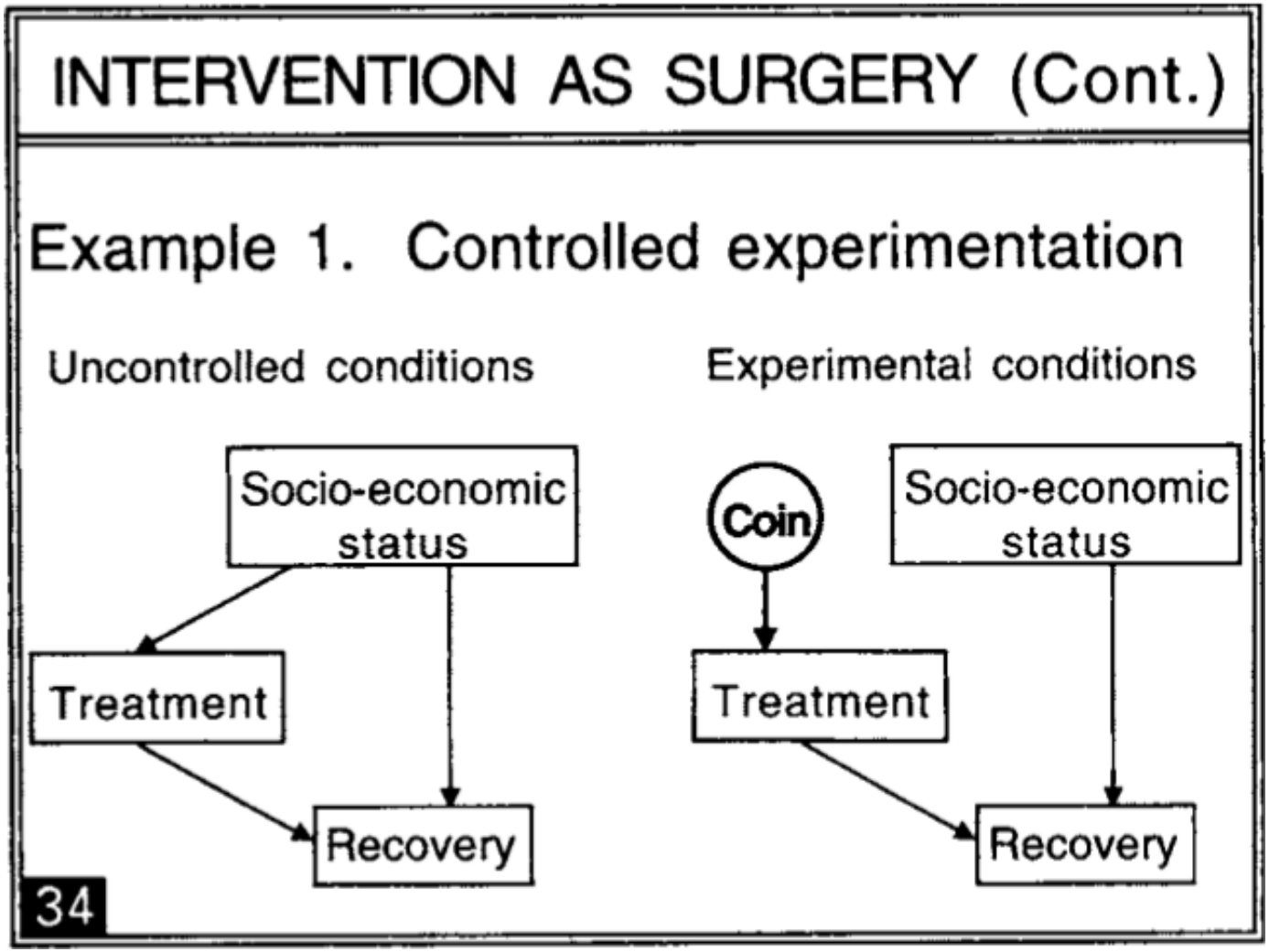
Source: *Dailymotion*

Recap again



Z is a *confounder* w.r.t. X's effect on Y

Recap again



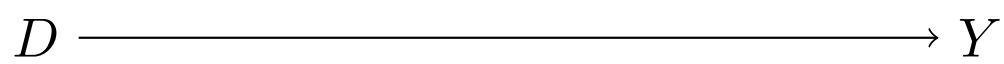
Source: Pearl, *Causality*, 2nd edition

Instrumental Variables: General Motivation

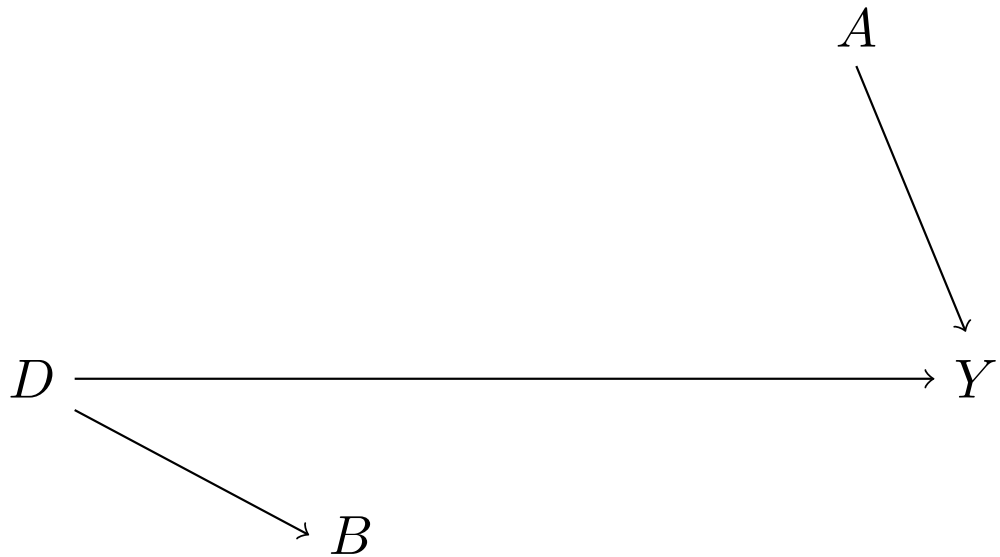
When and why would we use instrumental variables?

- We are interested in the causal of effect of D on Y: $D \longrightarrow Y$
- We have a classic omitted variables problem that we cannot solve via a selection-on-observables argument
  - We have a strong theory that some factor, U is correlated with both D and Y...
  - But U is unmeasured (and perhaps unmeasureable)
- We are unable to randomly assign D

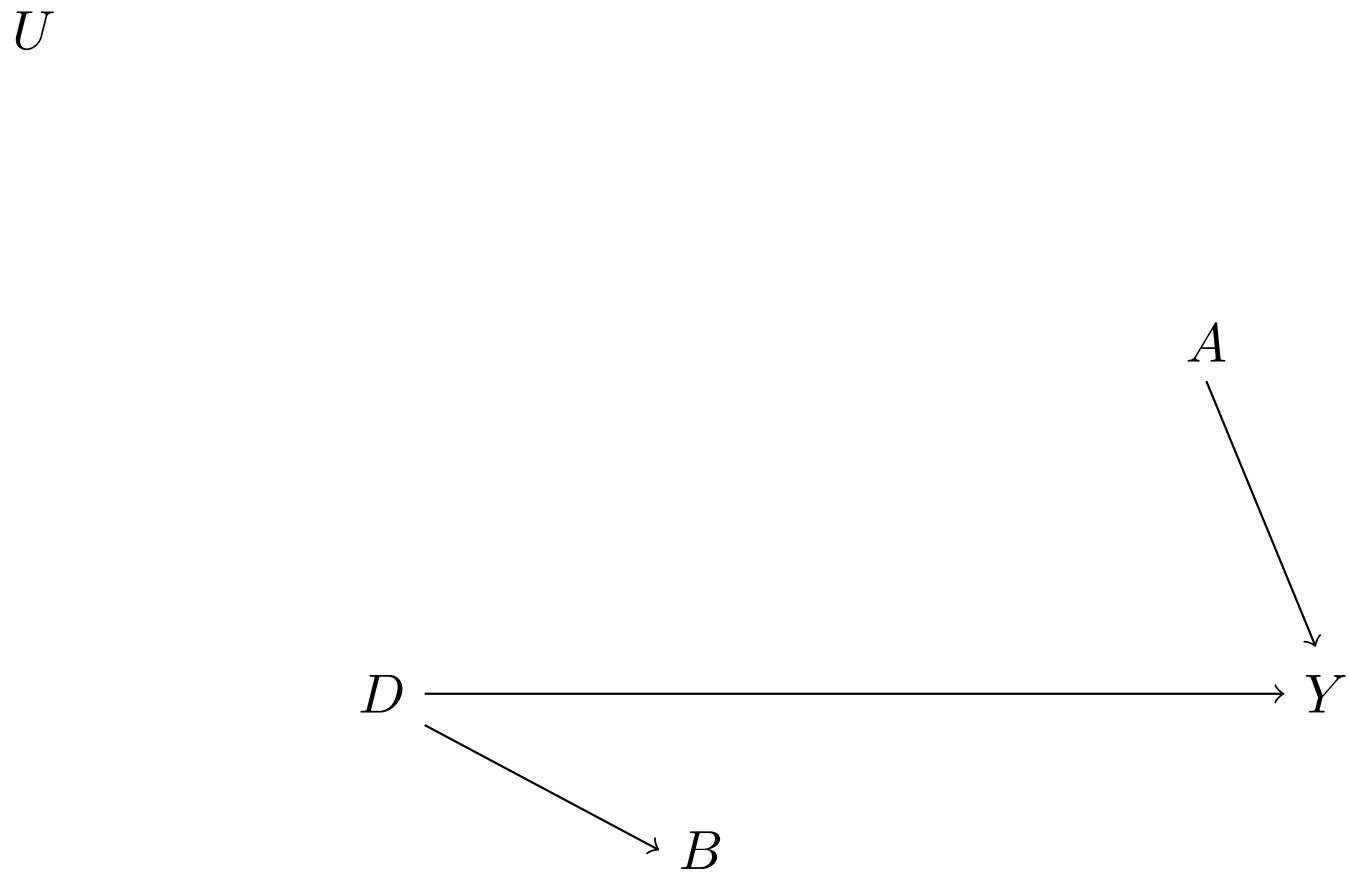
Instrumental Variables: General Motivation



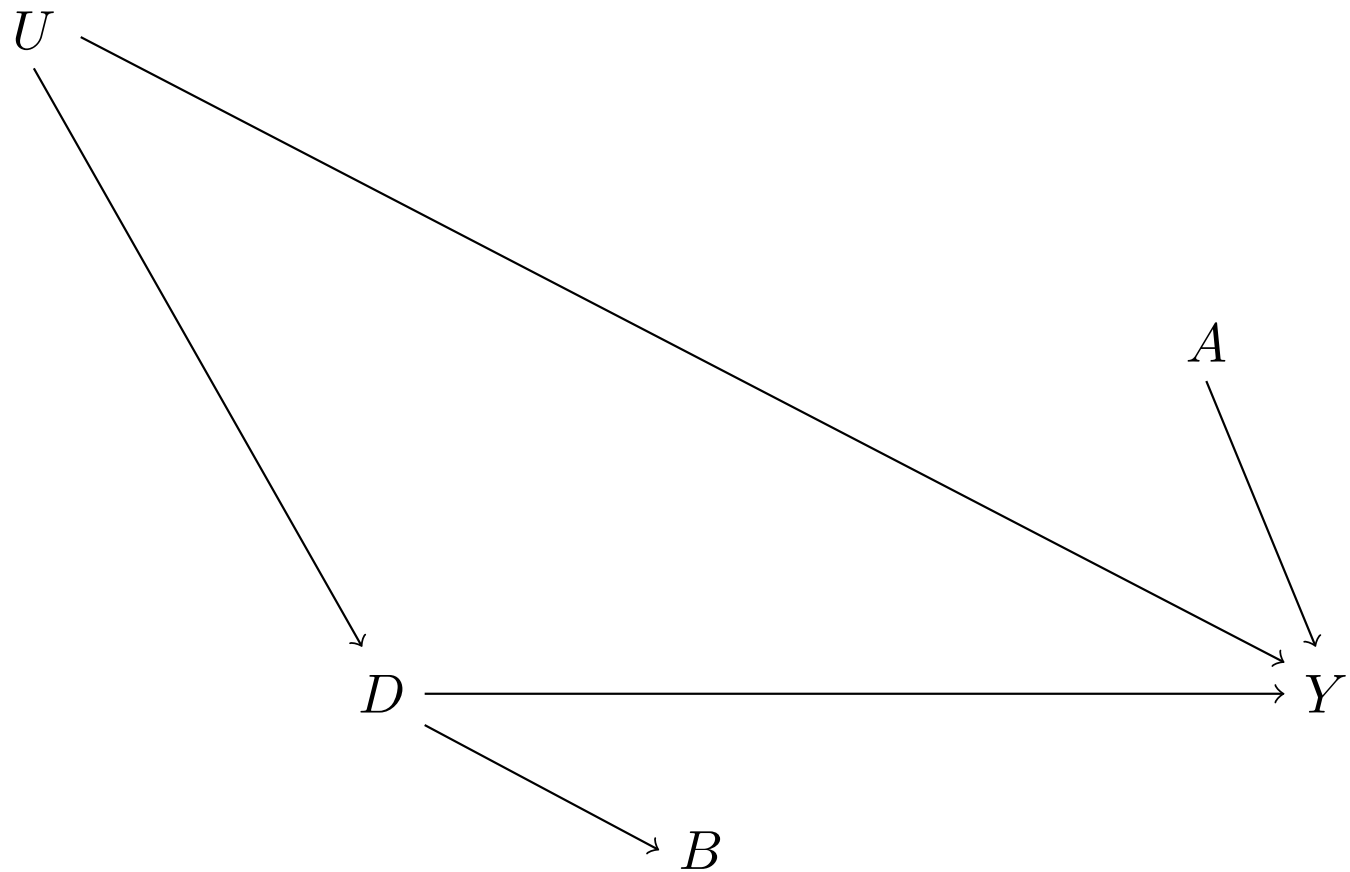
Instrumental Variables: General Motivation



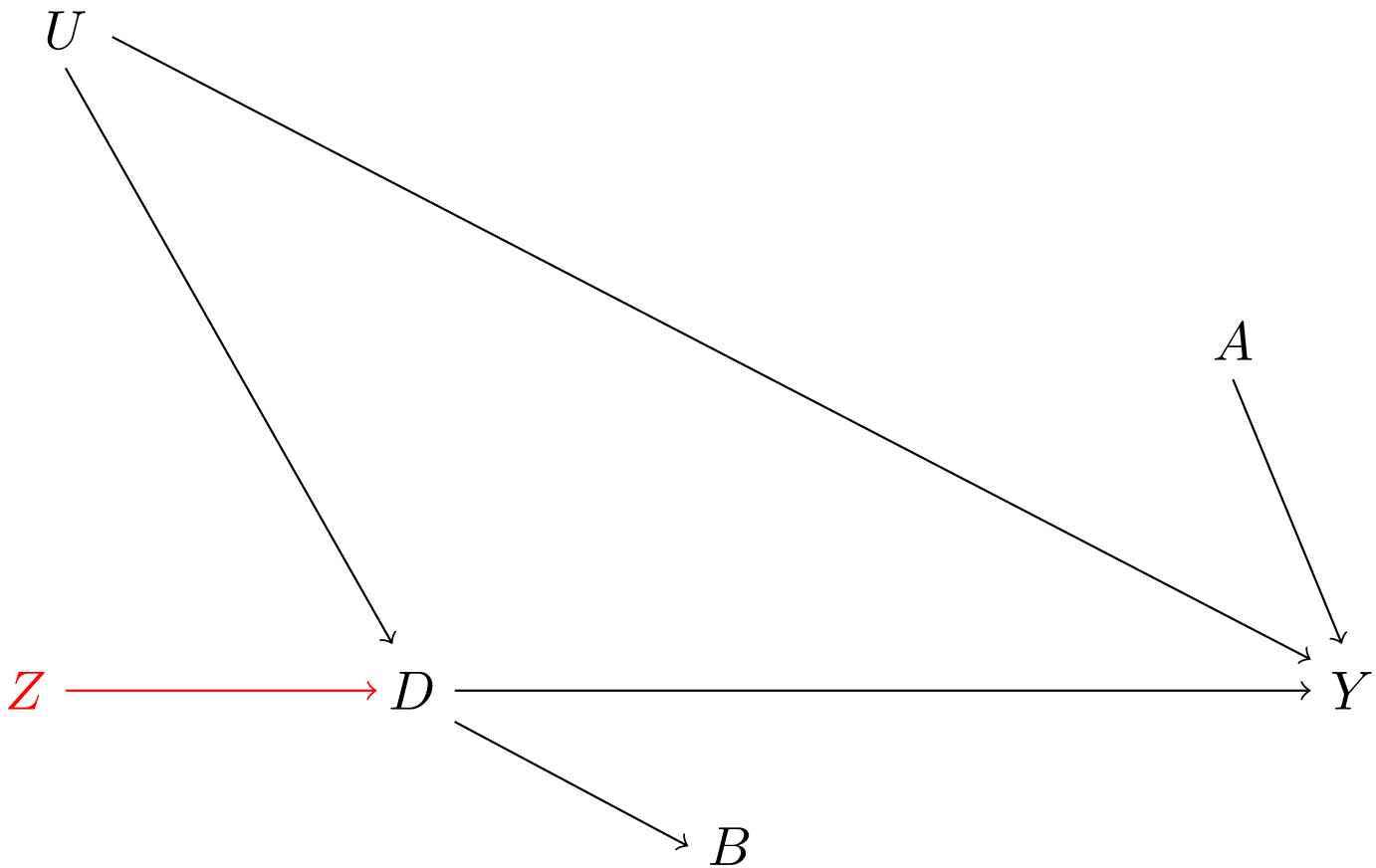
Instrumental Variables: General Motivation



Instrumental Variables: General Motivation



Instrumental Variables: General Motivation



Instrumental Variables: General Motivation

A valid instrument

- Z must be causally important to D

$$\mathbb{E}\big[D(Z=1) - D(Z=0)\big] \neq 0 \text{ for } \mathcal{Z} = \{0, 1\}$$

- Z is a "random shock" to the path between D and Y
  - Z must not be caused by D or Y
  - Z only affects Y through its effect on D

Instrumental Variables: Practical Steps

- We want to estimate:

$$\mathbb{E}[Y(D=1) - Y(D=0)] = \mathbb{E}[Y(D=1)] - \mathbb{E}[Y(D=0)]$$

  but we have the problem of an unobserved confounder

- We first estimate treatment assignment with the instrument:

$$\hat{D} = \mathbb{E}[D|Z]$$

- We then estimate the effect of the fitted $\hat{D}$ instead of the true D

$$\mathbb{E}[Y(\hat{D}=1)] - \mathbb{E}[Y(\hat{D}=0)]$$

Behavioral Motivation: Non-Compliance

Problem

- *Often we cannot force subjects to take specific treatments*
- *Units choosing to take the treatment may differ in unobserved characteristics from units that refrain from doing so*

Example: Non-compliance in the Job Training Partnership Act (JTPA)

		Not Enrolled in Training	Enrolled in Training	Total
Experiment	Assigned to Control	3,663	54	3,717
	Assigned to Training	2,683	4,804	7,487
	Total	6,346	4,858	11,204

Potential Outcome Model for Instrumental Variables

Definition (Instrument)

Z_i: Binary instrument for *unit* i

$$Z_i = \begin{cases} 1 & \text{if unit } i \text{ “encouraged” to receive treatment} \\ 0 & \text{if unit } i \text{ “encouraged” to receive control} \end{cases}$$

Definition (Potential Treatments)

$D(z)$ indicates *potential* treatment status given $Z = z$

- $D(1) = 1$: unit encouraged to take treatment and takes treatment
- $D(1) = 0$: unit encouraged to take treatment and does not take treatment
- $D(0) = 1$: unit not encouraged to take treatment and takes treatment
- $D(0) = 0$: unit not encouraged to take treatment and does not take treatment

Assumption

Observed treatments are realized as

$$D = Z \cdot D(1) + (1 - Z) \cdot D(0) \text{ so } D_i = \begin{cases} D_{1i} & \text{if } Z_i = 1 \\ D_{0i} & \text{if } Z_i = 0 \end{cases}$$

Following the terminology of Angrist, Imbens, and Rubin (1996), using $D(Z = z)$ we can define:

Definition

- *Compliers*: $D(1) > D(0)$ ($D(0) = 0$ and $D(1) = 1$).
- *Always-takers*: $D(1) = D(0) = 1$.
- *Never-takers*: $D(1) = D(0) = 0$.
- *Defiers*: $D(1) < D(0)$ ($D(0) = 1$ and $D(1) = 0$).

Problem

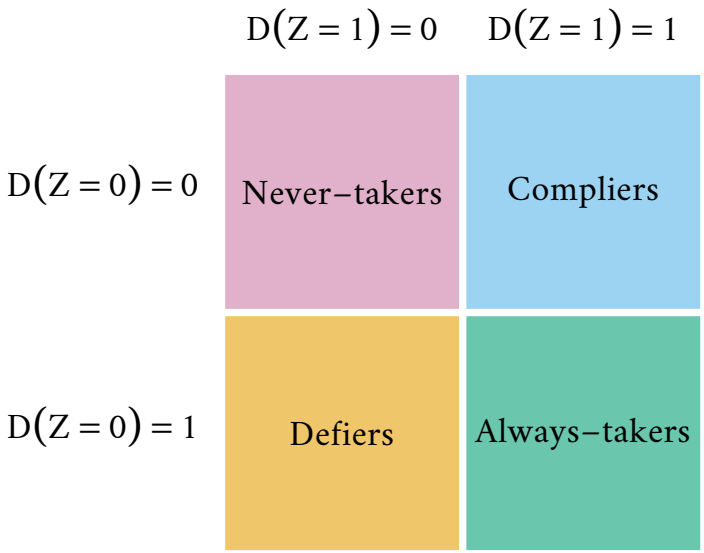
Only one of the potential treatment indicators $\{D(0), D(1)\}$ is observed, so we cannot identify which group any particular individual belongs to

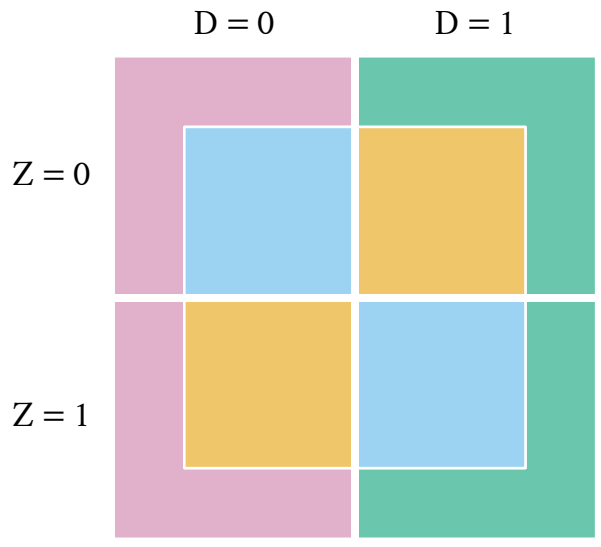
Principal Stratification: The IV Special Case

	Potential		**Observed**	
	$D(Z=0)$	$D(Z=1)$	Z	D
Compliers	0	1	0	0
			1	1
Defiers	1	0	0	1
			1	0
Always-takers	1	1	0	1
			1	0
Never-takers	0	0	0	0
			1	0

A special case of Principal Stratification (Frangakis and Rubin 2002), a much more general and highly useful framework.

Principal Stratification: The IV Special Case





Estimand (LATE)

$\alpha_{LATE} = \mathbb{E}[Y(1) - Y(0)|D(1) > D(0)]$ *is defined as the Local Average Treatment Effect for Compliers*

- *This estimand varies with the particular instrument* Z

Proposition (Special Cases)

- *When the treatment intake,* D*, is itself randomized, then* $Z = D$ *and every individual is a complier*
- *Given one-sided noncompliance,* $D(0) = 0$*:*

$$\begin{aligned}\mathbb{E}[Y(1)|D(1) > D(0)] &= \mathbb{E}[Y(1)|D(1) = 1] \\ &= \mathbb{E}[Y(1)|Z = 1, D(1) = 1] = \mathbb{E}[Y(1)|D = 1] \text{ and} \\ \mathbb{E}[Y(0)|D(1) > D(0)] &= \mathbb{E}[Y(0)|D = 1]\end{aligned}$$

so $\alpha_{LATE} = \mathbb{E}[Y(1) - Y(0)|D(1) > D(0)] = \mathbb{E}[Y(1) - Y(0)|D = 1] = \alpha_{ATT}$

Identification with Instrumental Variables

Identification Assumption

1. *Independence of the Instrument:*

$$\{Y(0), Y(1), D(0), D(1)\} \perp\!\!\!\perp Z$$

2. *Exclusion Restriction:*

$$\text{For } Y(D, Z),\ Y(D, 0) = Y(D, 1)$$

3. *First Stage:*

$$0 < \mathbb{P}[Z = 1] < 1 \text{ and } \mathbb{P}[D(1) = 1] \neq \mathbb{P}[D(0) = 1]$$

4. *Monotonicity (no defiers):*

$$D(1) \geq D(0)$$

Identification Result

$$\mathbb{E}[Y(1) - Y(0) | D(1) > D(0)] = \frac{\mathbb{E}[Y|Z = 1] - \mathbb{E}[Y|Z = 0]}{\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]} \left(= \frac{cov(Y, Z)}{cov(D, Z)} \right)$$

Identification Assumption

1. *Independence of the Instrument:*

$$\{Y(0), Y(1), D(0), D(1)\} \perp\!\!\!\perp Z$$

2. *Exclusion Restriction:*

$$\text{For } Y(D, Z),\ Y(D, 0) = Y(D, 1)$$

3. *First Stage:*

$$0 < \mathbb{P}[Z = 1] < 1 \text{ and } \mathbb{P}[D(1) = 1] \neq \mathbb{P}[D(0) = 1]$$

4. *Monotonicity (no defiers):*

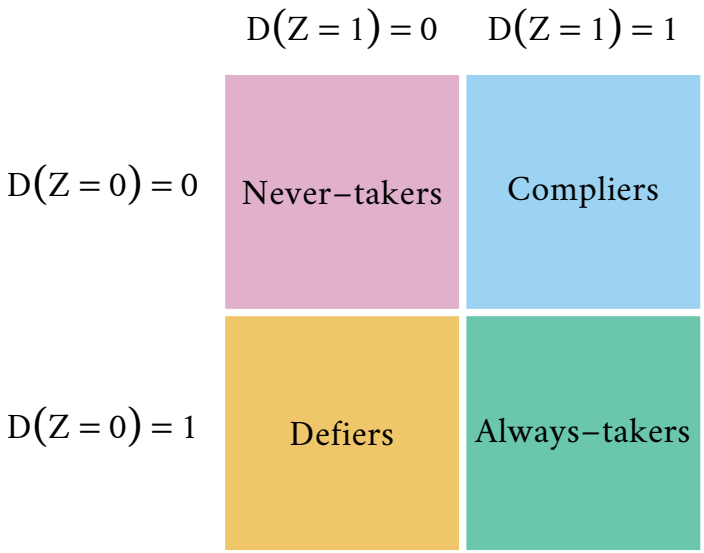
$$D(1) \geq D(0)$$

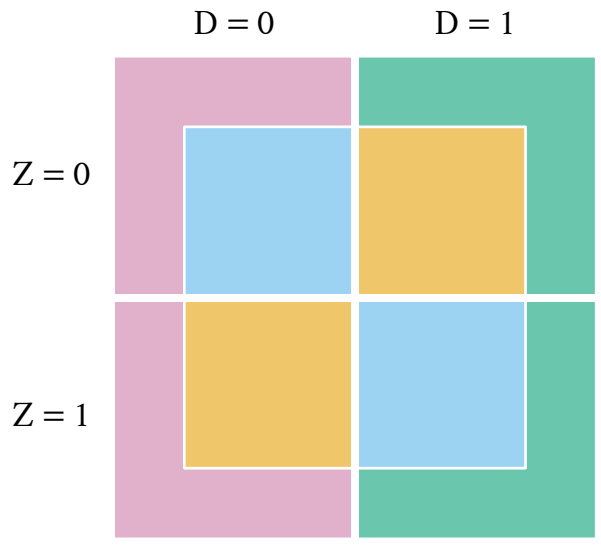
Proof.

$$\frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}$$

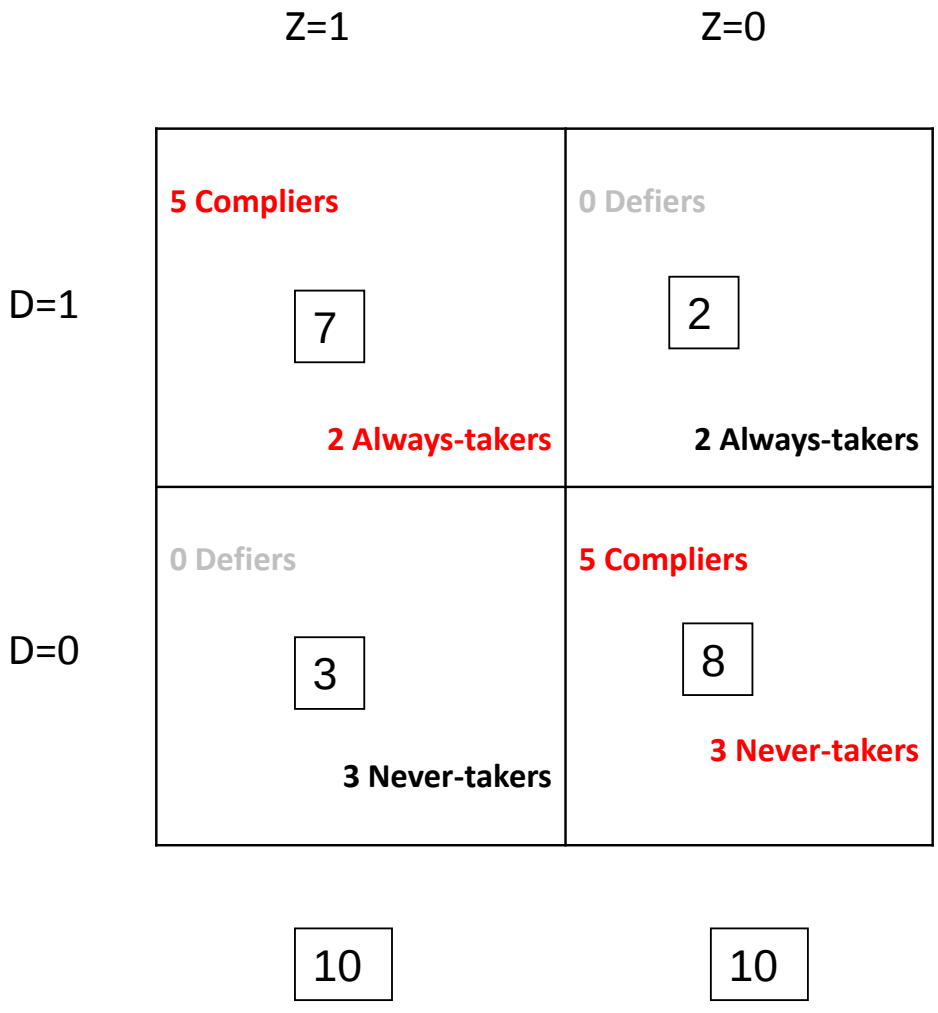
$$= \frac{\mathbb{E}[Y(1)D(1) + Y(0)(1 - D(1))|Z=1] - \mathbb{E}[Y(1)D(0) + Y(0)(1 - D(0))|Z=0]}{\mathbb{E}[D(1)|Z=1] - \mathbb{E}[D(0)|Z=0]}$$

Principal Stratification





Identification of the Proportion of Compliers



Pr(Complier) = E[D|Z=1] - E[D|Z=0] = 7/10 – 2/10 = .5

Identification Assumptions

1. Independence of the Instrument (a.k.a. Ignorability):

$$\{Y(0), Y(1), D(0), D(1)\} \perp\!\!\!\perp Z$$

- I.e., the instrument is independent of the vector of potential outcomes and potential treatment assignment
- Implies that the instrument Z is as good as randomly assigned
- Implies that the first stage captures the causal effect of Z on D:

$$\begin{aligned}\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0] &= \mathbb{E}[D(1)|Z=1] - \mathbb{E}[D(0)|Z=0] \\ &= \mathbb{E}[D(1) - D(0)]\end{aligned}$$

2. Exclusion Restriction:

$$\text{For } Y(D, Z),\ Y(D, 0) = Y(D, 1)$$

- Z can have no effect on Y except through its effect on D
- Allows to attribute correlation between Z and Y to the effect of D alone
- Random assignment of Z does not necessarily make your exclusion restriction claims more or less valid
- Assumption is not testable
- Usually, the believability of your entire argument in an IV framework will hinge on whether critics believe your exclusion restriction claims

3. First Stage:

$$0 < \mathbb{P}[Z = 1] < 1 \text{ and } \mathbb{P}[D(1) = 1] \neq \mathbb{P}[D(0) = 1]$$

- Implies that the instrument Z induces variation in D
- Testable by regressing D on Z

4. Monotonicity:

$$D(1) \geq D(0)$$

- Rules out defiers
  - While the instrument is allowed to have no effect on some units, all those units that are affected are affected in the same way
  - Assignment to a value of $Z = 1$ could have no effect on a unit, or it could (hopefully) increase the probability that $D = 1$, but it should not increase the probability that $D = 0$
- Often easy to assess from institutional knowledge, but no direct tests
- Without monotonicity, IV estimators are not guaranteed to estimate a weighted average of the causal effect for compliers
  - If defiers were allowed to exist by assumption, their estimated effects may partially cancel out the estimated effects of compliers

Identification Assumptions

If all four of these identification assumptions are satisfied, an instrumental variables approach estimates the local average treatment effect (LATE)

$$\mathbb{E}\big[Y(1) - Y(0) | D(1) > D(0)\big]$$

LATE is the average effect of the treatment on the outcome for the subgroup of compliers.
I.e., those who are not defiers, *and* whose treatment status has been changed by the instrument

Instrumental Variable: Estimators

Estimand (LATE)

$$\mathbb{E}\big[Y(1) - Y(0) | D(1) > D(0)\big] = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} \left(= \frac{Cov(Y,Z)}{Cov(D,Z)} \right)$$

Instrumental Variable: Estimators

Estimand (LATE)

$$\mathbb{E}[Y(1) - Y(0)|D(1) > D(0)] = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} \left(= \frac{Cov(Y,Z)}{Cov(D,Z)} \right)$$

Estimator (Wald Estimator)

The sample analog estimator is:

$$\left(\frac{\sum_{i=1}^{N} Y_i Z_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} Y_i (1 - Z_i)}{\sum_{i=1}^{N} (1 - Z_i)} \right) \Bigg/ \left(\frac{\sum_{i=1}^{N} D_i Z_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} D_i (1 - Z_i)}{\sum_{i=1}^{N} (1 - Z_i)} \right)$$

Instrumental Variable: Estimators

Estimand (LATE)

$$\mathbb{E}[Y(1) - Y(0)|D(1) > D(0)] = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} \left(= \frac{Cov(Y,Z)}{Cov(D,Z)} \right)$$

Estimator (Wald Estimator as IV Regression)

Can also implement Wald Estimator using an IV regression:

$$Y = \mu + \alpha \widehat{D} + \varepsilon$$

where $\mathbb{E}[\varepsilon|Z] = 0$*, so* $\alpha = cov(Y,Z)/cov(D,Z)$*.*

- *In the special case of binary* Z *used to predict variation in* D*,*

$$\alpha = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]}$$

To estimate α *we run the simple IV regression of* Y *on a constant and* D *and instrument* D *with* Z*.*

Instrumental Variable: Estimators

Estimand (LATE)

$$\mathbb{E}[Y(1) - Y(0)|D(1) > D(0)] = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} \left(= \frac{Cov(Y,Z)}{Cov(D,Z)} \right)$$

Estimator (Two Stage Least Squares)

If identification assumptions only hold after conditioning on X, covariates are often introduced using 2SLS regression:

$$Y = \mu + \alpha \widehat{D} + X'\beta + \varepsilon,$$

where $\mathbb{E}[\varepsilon|X, Z] = 0$. Now α and β are computed regressing Y on D and X, and using Z and X as instruments.
In general, α estimated in this way does not have a clear causal interpretation (see Abadie (2003))

Identification with Traditional Instrumental Variables

Definition

Two equations:

- $Y = \gamma + \alpha \widehat{D} + \varepsilon$ (Second Stage)
- $D = \tau + \rho Z + \eta$ (First Stage)

Identification Assumption

1. *Exogeneity and Exclusion:* $Cov(Z, \eta) = 0$ *and* $Cov(Z, \varepsilon) = 0$
2. *First Stage:* $\rho \neq 0$
3. α *constant for all units* i

4.2 IV in Fuzzy RDD

Fuzzy RDD: Identification

Identification Assumption

- *Binary instrument* Z *with* $Z = \mathbb{I}\{X > c\}$
- *Restrict sample to observations close to discontinuity where* $\mathbb{E}[Y|X, D]$ *jumps so that* $X \approx c$ *and thus* $\mathbb{E}[X|Z = 1] - \mathbb{E}[X|Z = 0] \approx 0$
- *Usual instrumental variables assumptions (ignorability, first stage, monotonocity)*

Identification Result

$$
\begin{aligned}
\alpha_{FRDD} &= \mathbb{E}[Y(1) - Y(0)|X = c \text{ and } i \text{ is a complier}] \\
&= \frac{\lim_{X \downarrow c} \mathbb{E}[Y|X = c] - \lim_{X \uparrow c} \mathbb{E}[Y|X = c]}{\lim_{X \downarrow c} \mathbb{E}[D|X = c] - \lim_{X \uparrow c} \mathbb{E}[D|X = c]} \\
&= \frac{\textit{outcome discontinuity}}{\textit{treatment discontinuity}} \\
&\approx \frac{\mathbb{E}[Y|Z = 1] - \mathbb{E}[Y|Z = 0]}{\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]}
\end{aligned}
$$

4.2 Heckman Selection Models

Refresh: Different Kinds of Selection Bias

1. *Selection into treatment*

$$\{Y(1), Y(0)\} \not\perp\!\!\!\perp D$$

2. *Selection into the sample*

$$S \not\perp\!\!\!\perp \boldsymbol{X}$$

3. *Selection into reporting*

$$\mathbb{P}[V_i \text{ is observed}] \not\perp\!\!\!\perp V_i \quad \text{(but some } \boldsymbol{X}_i \text{ for the } i \text{ are observed)}$$

Refresh: Different Kinds of Selection Bias

1. *Selection into treatment*

$$\{Y(1), Y(0)\} \not\perp\!\!\!\perp D$$

2. *Selection into the sample*

$$S \not\perp\!\!\!\perp \boldsymbol{X}$$

3. *Selection into reporting*

$$\mathbb{P}[V_i \text{ is observed}] \not\perp\!\!\!\perp V_i \quad \text{(but some } \boldsymbol{X}_i \text{ for the } i \text{ are observed)}$$

Refresh: Different Kinds of Selection Bias

1. *Selection into treatment*

$$\{Y(1), Y(0)\} \not\perp\!\!\!\perp D$$

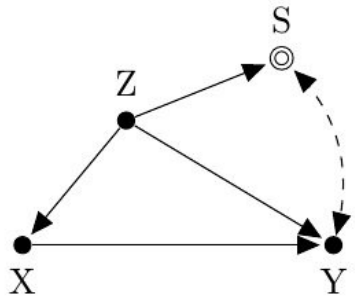
2. *Selection into the sample*

$$S \not\perp\!\!\!\perp \boldsymbol{X}$$

3. *Selection into reporting in Heckman Selection*

$$\mathbb{P}[Y_i \text{ is observed}] \not\perp\!\!\!\perp Y_i \quad \text{(but some } \boldsymbol{X}_i \text{ for the } i \text{ are observed)}$$

Heckman Selection Models a.k.a. Heckman Correction



(a)

Figure 5: *(a) A model of female labor supply (Heckman, 1976, 1979). Variables: hours worked (X), earnings (Y), socio-economic factors (Z), sampling mechanism (S).*

in 1967. The challenge to valid inference in this setting arises due to the fact that market wages are only observable for women who actually choose to work. His model is described by the following two equations

$$s_i = \quad \mathbb{1}[Z_i'\delta - \eta_i > 0] \tag{4.1}$$

$$y_i = \begin{cases} x_i\beta + Z_i'\gamma + \varepsilon_i & \text{if } s_i = 1, \\ \text{unobserved} & \text{if } s_i = 0. \end{cases} \tag{4.2}$$

Source: P. Hünermund and E. Barenboim

Heckman Selection Models

Context:

- Which Y_is are observed may depend on some unobseved confounder
- And we also observe at least one thing that affects which Y_is are observed but not Y itself
- Generalized as Generalized Joint Response Model (GJRM)

The basic idea: We model two processes as regression:

- *Outcome model:*

$$Y_i = f(D_i, \boldsymbol{X}_i, \ldots, \epsilon_i)$$

- *Selection model:*

$$\mathbb{I}[Y_i \text{ is observed}] = g(D_i, \boldsymbol{X}_i, Z_i, \ldots, \eta_i)$$

- connected via the joint distribution of residuals

$$\{\epsilon, \eta\} \sim c(\ldots)$$

such as

$$\{\epsilon, \eta\} \sim MultivariateNormal\big(\{0, 0\}, \Sigma\big)$$

3. A glimpse into today's future

A glimpse into today's future

- Matrix Completion
- Adaptive experiments
- Causal inference under interference
- Algorithmizing causal inference

Matrix Completion

- For panel data, where a set of units is observed at multiple points in time, but receive (or take) the treatment only on some of them.
- For example:

i	$Y(0)$ observed $t=1$	$t=2$	$t=3$	$Y(1)$ observed $t=1$	$t=2$	$t=3$
1	?	✓	?	✓	?	✓
2	?	✓	✓	✓	?	?

- Based on ML, technically quite complex.
- The basics:
- Predict (truly predict, not just fit) $Y(1)$ and $Y(0)$ using a set of predictors $\boldsymbol{X}$ conditional on which treatment assignment is independent of the potential outcomes

$$\{Y(1), Y(0)\} \perp\!\!\!\perp D|\boldsymbol{X}$$

using a regularized prediction model/algorithm.

Adaptive Experiments

- Suppose we want to simultaneously estimate which $\mathbb{E}[Y(D = d)]$ is the best of all $d \in \mathcal{D}$

  and some function $u(Y)$ that rewards us based on the generated Y,

  and we can run randomized experiments in which we allocate the incoming units into $d \in \mathcal{D}$.

  Purely random allocation may be good for the *estimation* part, but sub-optimal for the *reward* part.
- Very common in business settings. But also health, public services etc.
- Not a new idea, but took off massively recently.
- So-called 'multi-armed bandits'.

Conventionally, we always assume *no interference*, SUTVA: Stable Unit Treatment Value Assumption

$\forall i \in \{1, \ldots, N\}$, i's $Y_i(D)$ only depends on its D_i.

This is obviously false in many settings where units interact, such as markets.

Simple example, two units $i \in \{1, 2\}$, binary D:

	$Y(D_1, D_2)$			
	$Y(0,0)$	$Y(1,0)$	$Y(0,1)$	$Y(1,1)$
$i = 1$				
$i = 2$				

In short, for N units and $|\mathcal{D}|$ treatment options, there may be up to $|\mathcal{D}|^N$ potential outcomes *for each unit*.

A lot of progress has been made in how to make this manageable.
Experimental designs for settings such as marketplaces are one such area.

Algorithmizing Causal Inference

An old goal: to build algorithms that give causal answers from inputing some data and assumptions about the data.

Practically usable *causal discovery* remains out of reach for now.

But there are advances or attempts such as

- `https://doi.org/10.1080/01621459.2023.2216909`
- `https://github.com/microsoft/causica`

4. Partial Identification

Partial Identification

Source: https://www.charlesmanski.com/p/about-charles-manski.html

Partial Identification

- A different way of looking at estimation problems.
- Founded in the 1980s by Charles F. Manski,

  based on some older and simple, but powerful ideas.
- Can be summed-up as:

  If we're trying to estimate some quantity, let's figure out the smallest and the largest possible value it can take based only on assumptions that we know to be true.

  And we can check how the bounds change if we start adding assumptions that may not be true.
- Such bounds are known as *Manski bounds*.
- Applications in descriptive as well as causal inference.

Conventional Approaches vs. Partial Identification in Causal Inference

Shared:

- We are interested in some potential outcomes $Y(D)$ and comparisons between their values for different values that D can take.

Different:

- Conventional approaches estimate $Y(D = d)$ by re-weighting $Y|D = d$
- Partial identification instead takes $Y|D = d$ as is, and looks at the possible values of $Y(D = d)$ for those, among whom $D \neq d$.

Partial Identification: Example

Combined	E	$\neg E$		Recovery Rate
drug (C)	20	20	40	50%
no-drug ($\neg C$)	16	24	40	40%
	36	44	80	

Males	E	$\neg E$		Recovery Rate
drug (C)	18	12	30	60%
no-drug ($\neg C$)	7	3	10	70%
	25	15	40	

Females	E	$\neg E$		Recovery Rate
drug (C)	2	8	10	20%
no-drug ($\neg C$)	9	21	30	30%
	11	29	40	

Partial Identification: Example

	Data		$Y(d)=1$ cases		
Combined	$Y=0$	$Y=1$	*Worst*	*Best*	$\mathbb{B}\big[\mathbb{E}[Y(1)-Y(0)]\big]$
$D=1$	20	20	$[20,$	$20+40=60]$	$\big[-{}^{36}/_{80},\quad {}^{44}/_{80}\big]$
$D=0$	24	16	$[16,$	$16+40=56]$	
Males					
$D=1$	12	18	$[18,$	$18+10=28]$	$\big[-{}^{19}/_{40},\quad {}^{21}/_{40}\big]$
$D=0$	3	7	$[7,$	$7+30=37]$	
Females					
$D=1$	8	2	$[2,$	$2+30=32]$	$\big[-{}^{17}/_{40},\quad {}^{23}/_{40}\big]$
$D=0$	21	9	$[9,$	$9+10=19]$	

Partial Identification: Example

	Data		$Y(d)=1$ cases		
Combined	$Y=0$	$Y=1$	*Worst*	*Best*	$\mathbb{B}\big[\mathbb{E}[Y(1)-Y(0)]\big]$
$D=1$	20	20	$[20,$	$20+40=60]$	$\big[-{}^{36}/_{80},\quad {}^{44}/_{80}\big]$
$D=0$	24	16	$[16,$	$16+40=56]$	
Males					
$D=1$	12	18	$[18,$	$18+10=28]$	$\big[-{}^{19}/_{40},\quad {}^{21}/_{40}\big]$
$D=0$	3	7	$[7,$	$7+30=37]$	
Females					
$D=1$	8	2	$[2,$	$2+30=32]$	$\big[-{}^{17}/_{40},\quad {}^{23}/_{40}\big]$
$D=0$	21	9	$[9,$	$9+10=19]$	

But how to choose now?

- That depends. Are you avoiding worst-case scenarios or aiming for best-case scenarios? Or say minimizing maximum regret: the largest possible difference between what we've chosen and what was really the best.
- We can also start adding assumptions that may not always hold and see how that changes the bounds.

?

Thank you!