Instrumental Variable Estimation

Lingguo Cheng

Nanjing University

November 24, 2025

$Y_i = X_i + \varepsilon_i$
$\uparrow$
Z_i

① relevance. $Z_i \to X_i$

[② exclusion restriction $\mathrm{cov}(Z_i, \varepsilon_i) = 0$]

③ exogeneity Z_i 只通过 X_i 影响 Y_i.

Outline

A brief Introduction of Instrumental Variables (IV)

- In the 1920s, the father-and-son research team of Phillip and Sewall Wright: how to estimate the slope of supply and demand curves when observed data on prices and quantities are determined by the intersection of these two curves.
- Wright (1928) solves the statistical simultaneous equations problem by using variables that appear in one equation to shift this equation and trace out the other.
- The variables that do the shifting became known as instrumental variables (Reiersol, 1941).
- Simultaneous equations models (SEMs) have been enormously influential in the history of econometric thought. The technical language used to discuss IV methods still comes from this framework.
- However, the most important contemporary use of IV methods is to solve the problem of omitted variables bias (OVB).

Outline

A simple starting point: one endogenous variable, no covariates

- Suppose, as before, that potential outcomes can be written

$$Y_{si} \equiv f_i(s) = \alpha + \rho s + \eta_i$$

Then,

$$Y_i = \alpha + \rho s_i + \eta_i$$

- Imagine that there is a vector of control variables, A_i, called "ability," that gives a selection-on-observables story:

$$\eta_i = A_i'\gamma + v_i$$

Where γ is again a vector of population regression coefficients, so that v_i and A_i are uncorrelated by construction. For now, the variables A_i, are assumed to be the only reason why η_i and s_i are correlated, so that

$$E\left[s_i v_i\right] = 0$$

A simple starting point: one endogenous variable, no covariates

- In other words, if A_i were observed, we would be happy to include it in the regression of wages on schooling; thereby producing a long regression that can be written (Causal regression function or structural function)

$$Y_i = \alpha + \rho s_i + A_i'\gamma + v_i$$

- Then how to estimate the long regression coefficient, ρ, when A_i is unobserved?

ILS IV Estimator

- When the researcher has access to a variable (the instrument, which we'll call Z_i), that is *correlated with the endogenous variable of interest*, s_i, but *uncorrelated with any other determinants of the dependent variable*: $Cov(\eta_i, Z_i) = 0$, or, equivalently, Z_i is uncorrelated with both A_i and v_i.
- This statement is called an exclusion restriction, since Z_i can be said to be excluded from the causal model of interest.
- Given the exclusion restriction, it follows that

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(s_i, Z_i)} = \frac{Cov(Y_i, Z_i)/V(Z_i)}{Cov(s_i, Z_i)/V(Z_i)} = \frac{\text{reg } Y_i \text{ on } Z_i}{\text{reg } s_i \text{ on } Z_i}$$

ILS IV Estimator

- Proof:

$$\begin{aligned} Cov(Y_i, Z_i) &= Cov(\alpha + \rho s_i + A_i'\gamma + v_i, Z_i) \\ &= \rho Cov(s_i, Z_i) \end{aligned}$$

- The coefficient of interest, ρ, is the ratio of the population regression of Y_i on Z_i (called the *reduced form*) to the population regression of s_i on Z_i (called the *first stage*).
- An intuitive way to understand the IV estimator:

$$\begin{aligned} \rho = \frac{dY_i}{ds_i} &= \frac{dY_i}{dZ_i}\frac{dZ_i}{ds_i} = \frac{dY_i/dZ_i}{ds_i/dZ_i} = \frac{Cov(Y_i, Z_i)/Var(Z_i)}{Cov(s_i, Z_i)/Var(Z_i)} \\ &= \frac{Cov(Y_i, Z_i)}{Cov(s_i, Z_i)} \end{aligned}$$

Recap of IV Assumptions

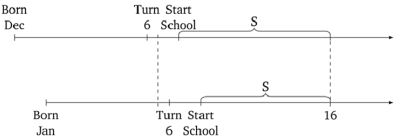
- It's worth recapping the assumptions needed for the ratio of covariances to equal the causal effect, ρ.
- The instrument must have a clear effect on s_i ***(First Stage)***
  - The first stage is not zero, but this is something you can check in the data
  - If the first stage is only marginally significantly different from zero, the resulting IV estimates are unlikely to be informative, a point we return to later
- The only reason for the relationship between Y and Z is the first stage ***(Exclusion Restriction)***
  - First, the instrument is as good as randomly assigned (i.e., Independent of potential outcomes, conditional on covariates, like the CIA)
  - Second, the instrument has no effect on outcomes other than through the first-stage channel

Where to find an instrumental variable?

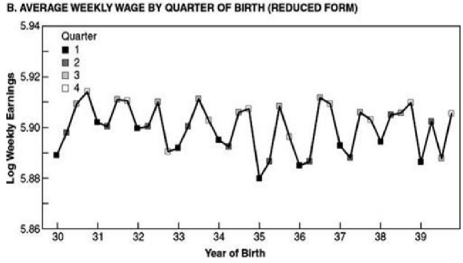
- Good instruments come from a combination of institutional knowledge and ideas about the processes determining the variable of interest.
  - For example, the economic model of education suggests that schooling decisions are based on the costs and benefits of alternative choices.
  - Thus, one possible source of instruments for schooling is differences in costs due to loan policies or other subsidies that vary independently of ability or earnings potential.
- A second source of variation in schooling is institutional constraints. A set of institutional constraints relevant to schooling is compulsory schooling laws.

Example: Angrist and Krueger (1991)

- The starting point for the Angrist and Krueger (1991) quarter-of-birth strategy is the observation that most states require students to enter school in the calendar year in which they turn 6. School start age is therefore a function of date of birth.
- Specifically, those born late in the year are young for their grade. In states with a December 31 birthday cutoff, children born in the fourth quarter enter school shortly before they turn 6, while those born in the first quarter enter school at around age.
- Furthermore, because compulsory schooling laws typically require students to remain in school only until their 16th birthday, these groups of students will be in different grades, or through a given grade to a different degree, when they reach the legal dropout age.
- The combination of school start-age policies and compulsory schooling laws creates a natural experiment in which children are compelled to attend school for different lengths of time, depending on their birthdays.



- If children were born on or before December 31, then they were assigned to the first grade. But if their birthday was on or after January 1, they were assigned to kindergarten.
- For most of the twentieth century, the US had compulsory schooling laws that forced a person to remain in high school until age 16. After age 16, one could legally stop going to school.



Outline

A little more generalization: one endogenous variable, with covariates

- Now, incorporate covariates X_i into the causal regression function, it becomes

$$Y_i = \alpha' X_i + \rho s_i + \eta_i$$

  Where

$$\eta_i = A_i' \gamma + v_i$$

- Then

$$s_i = X_i' \pi_{10} + \pi_{11} Z_i + \xi_{1i} \quad \text{(First-stage regression equation)}$$
$$Y_i = X_i' \pi_{20} + \pi_{21} Z_i + \xi_{2i} \quad \text{(Reduced-form regression equation)}$$

- In the language of the SEM the dependent variables in these two equations are said to be the **endogenous variables** (determined jointly within the system)
- Variables on the right-hand side are said to be the **exogenous variables** (determined outside the system)
- The instruments z_i are a subset of the exogenous variables.
- The exogenous variables that are not instruments are said to be **exogenous covariates**.

Covariate-Adjusted IV Estimation

- The covariate-adjusted IV estimator is the sample analog of the ratio

$$\rho = \frac{\pi_{21}}{\pi_{11}} = \frac{\text{Cov}\left(Y_i, \tilde{Z}_i\right)}{\text{Cov}\left(s_i, \tilde{Z}_i\right)} \left(= \frac{\text{Cov}\left(Y_i, \tilde{Z}_i\right)/V\left(\tilde{Z}_i\right)}{\text{Cov}\left(s_i, \tilde{Z}_i\right)/V\left(\tilde{Z}_i\right)} \right)$$

- Where $\tilde{Z}_i$ is the residual from a regression of Z_i on the exogenous covariates, X_i.
- Econometricians call the sample analog of the equation above an indirect least squares (ILS) estimator of ρ in the causal model with covariates,

$$Y_i = \alpha' X_i + \rho s_i + \eta_i$$
$$\eta_i = A_i'\gamma + v_i$$

- **Proof:**
  You can easily show that $\text{Cov}\left(Y_i, \tilde{Z}_i\right) = \rho\,\text{Cov}\left(s_i, \tilde{Z}_i\right)$

Two-Stage Least Squares (2SLS)

- The reduced-form equation can be derived by substituting the first-stage equation into the causal relation of interest, which is also called a *"structural equation"* in simultaneous equations language.

$$\begin{aligned} Y_i &= \alpha' X_i + \rho s_i + \eta_i \\ &= \alpha' X_i + \rho \left(X_i' \pi_{10} + \pi_{11} Z_i + \xi_{1i} \right) + \eta_i \\ &= X_i' \left(\alpha + \rho \pi_{10} \right) + \rho \pi_{11} Z_i + \left(\rho \xi_{1i} + \eta_i \right) \\ &\equiv X_i' \pi_{20} + \pi_{21} Z_i + \xi_{2i} \end{aligned}$$

  Where $\pi_{20} \equiv \alpha + \rho \pi_{10}, \quad \pi_{21} \equiv \rho \pi_{11}, \quad \xi_{2i} \equiv \rho \xi_{1i} + \eta_i$

- It shows again,

$$\rho = \frac{\pi_{21}}{\pi_{11}}$$

Two-Stage Least Squares (2SLS)

- Note also that a slight rearrangement gives (This provides the basis for 2SLS)

$$Y_i = \alpha' X_i + \rho \left(X_i' \pi_{10} + \pi_{11} Z_i \right) + \left(\rho \xi_{1i} + \eta_i \right)$$

- Where $(X_i' \pi_{10} + \pi_{11} Z_i)$ is the population fitted value from the first-stage regression of s_i on X_i and Z_i.
- Because Z_i and X_i are uncorrelated with the reduced-form error, ξ_{2i}, the coefficient on $(X_i' \pi_{10} + \pi_{11} Z_i)$ in the population regression of Y_i on X_i and $(X_i' \pi_{10} + \pi_{11} Z_i)$ equals ρ.

Two-Stage Least Squares (2SLS) in Practice

- In practice, of course, we almost always work with data from samples. Given a random sample, the first-stage fitted values are consistently estimated by

$$\hat{s}_i = X_i'\hat{\pi}_{10} + \hat{\pi}_{11}Z_i$$

- The coefficient on $\hat{s}_i$ in the regression of Y_i on X_i and $\hat{s}_i$ is called the two-stage least squares (2SLS) estimator of ρ. In other words, 2SLS estimates can be constructed by OLS estimation of the "second-stage equation,"
- This is called 2SLS because it can be done in two steps, the first estimating $\hat{s}_i$ and the second estimating causal equation. The resulting estimator is consistent for ρ because the covariates and first-stage fitted values are uncorrelated with both η_i and $(s_i - \hat{s}_i)$.

$$Y_i = \alpha' X_i + \rho\hat{s}_i + [\eta_i + \rho\,(s_i - \hat{s}_i)]$$

Comments on 2SLS Estimation

- The 2SLS name notwithstanding, we don't usually construct 2SLS estimates in two steps.
  - For one thing, the resulting standard errors are wrong, as we discuss later.
  - Still, the fact that the 2SLS estimator can be computed by a sequence of OLS regressions is one way to remember why it works.
- Intuitively, *conditional on covariates*, 2SLS retains only the variation in s_i that is generated by quasi-experimental variation—that is, generated by the instrument Z_i.

2SLS Estimator Is Also an ILS Estimator

- 2SLS is a many-splendored thing. For one, it is an IV estimator: the 2SLS estimate of ρ in the causal equation is the sample analog of

$$\rho = \frac{\text{Cov}\left(Y_i, \hat{s}_i^*\right)}{\text{Cov}\left(s_i, \hat{s}_i^*\right)}$$

  where $\hat{s}_i^*$ is the residual from a regression of $\hat{s}_i$ on X_i.
- It is also easy to show that, in a model with a single endogenous variable and a single instrument, the 2SLS estimator is the same as the corresponding ILS estimator.
- **Proof:** (Show it)

2SLS Estimator Equivalence

- Remember that $\rho = \frac{\text{Cov}\left(Y_i, \tilde{Z}_i\right)}{\text{Cov}\left(s_i, \tilde{Z}_i\right)}$
- Now the second stage regression function

$$Y_i = \alpha' X_i + \rho \hat{s}_i + \text{error term}$$

- Then the 2SLS estimator can be written as

$$\hat{\rho} = \frac{\text{Cov}\left(Y_i, \hat{s}_i^*\right)}{\text{Var}\left(\hat{s}_i^*\right)} = \frac{\text{Cov}\left(Y_i, \hat{s}_i^*\right)}{\text{Cov}\left(s_i, \hat{s}_i^*\right)}$$

- Therefore, it is equivalent to using $\hat{s}_i$ as the instrument of s_i.
- Notice that

$$\begin{aligned}
s_i &= \hat{\pi}_{11} X_i + \hat{\pi}_{12} Z_i = \hat{s}_i + e_{1i}, \quad \text{Cov}\left(e_{1i}, \hat{s}_i\right) = 0 \\
\hat{s}_i &= \hat{\pi}_{21} X_i + \hat{s}_i^*, \quad \text{Cov}\left(\hat{s}_i^*, X_i\right) = 0 \\
s_i &= \hat{\pi}_{21} X_i + \hat{s}_i^* + e_{1i} \\
\text{Cov}\left(s_i, \hat{s}_i^*\right) &= \text{Cov}\left(\hat{\pi}_{21} X_i + \hat{s}_i^* + e_{1i}, \hat{s}_i^*\right) = \text{Var}\left(\hat{s}_i^*\right)
\end{aligned}$$

Multiple Instruments

- Assuming each instrument captures the same causal effect (a strong assumption that is relaxed below), we might want to combine these alternative IV estimates into a single more precise estimate. In models with multiple instruments, 2SLS accomplishes this by combining multiple instruments into a single instrument.
- Suppose, for example, we have three instrumental variables, Z_{1i}, Z_{2i}, and Z_{3i}. In the Angrist and Krueger (1991) application, these are dummies for first-, second-, and third-quarter births. The first-stage equation then becomes

$$s_i = X_i'\pi_{10} + \pi_{11}Z_{1i} + \pi_{21}Z_{2i} + \pi_{31}Z_{3i} + \xi_{1i}$$

  while the second stage remains the same.
- The IV interpretation of this 2SLS estimator is the same as before: the instrument is the residual from a regression of first-stage fitted values on exogenous covariates.

	OLS		2SLS					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years of education	.071 (.0004)	.067 (.0004)	.102 (.024)	.13 (.020)	.104 (.026)	.108 (.020)	.087 (.016)	.057 (.029)
Exogenous Covariates								
Age (in quarters)								✓
Age (in quarters) squared								✓
9 year-of-birth dummies		✓			✓	✓	✓	✓
50 state-of-birth dummies		✓			✓	✓	✓	✓
Instruments								
dummy for QOB = 1			✓	✓	✓	✓	✓	✓
dummy for QOB = 2				✓		✓	✓	✓
dummy for QOB = 3				✓		✓	✓	✓
QOB dummies interacted with year-of-birth dummies (30 instruments total)							✓	✓

Notes: The table reports OLS and 2SLS estimates of the returns to schooling using the Angrist and Krueger (1991) 1980 census sample. This sample includes native-born men, born 1930–39, with positive earnings and nonallocated values for key variables. The sample size is 329,509. Robust standard errors are reported in parentheses. QOB denotes quarter of birth.

Econometric Model Results Analysis

- Overall, the 2SLS estimates are mostly a bit larger than the corresponding OLS estimates — not driven by omitted variables such as ability and family background.
- The results from models including **year-of-birth** and **state-of-birth dummies** as **exogenous covariates** are similar, not surprisingly, since the quarter of birth is not closely related to either of these controls.
- Column 7 shows the results of adding interaction terms to the instrument list—Three quarter-of-birth dummies interacted with nine dummies for the year of birth (1930-39), **30** excluded instruments. The first-stage equation:

$$s_i = X_i'\pi_{10} + \pi_{11}Z_{1i} + \pi_{21}Z_{2i} + \pi_{31}Z_{3i} + \sum_j (B_{ij}Z_{1i})\kappa_{1i} + \sum_j (B_{ij}Z_{2i})\kappa_{2i} + \sum_j (B_{ij}Z_{3i})\kappa_{3i} + \xi_{1i}$$

where B_{ij} is a dummy equal to one if individual i was born in year j for j equal to 1931-39.

Methodological Notes on Econometric Models

- The last 2SLS model adds controls for linear and quadratic terms in age in quarters to the list of exogenous covariates. This finely coded age variable provides a partial control for the fact that small differences in age may be an omitted variable that confounds the quarter-of-birth identification strategy. As long as the effects of age are reasonably smooth, the quadratic age-in-quarters model will pick them up.
- Columns 7 and 8 illustrate the interplay between identification and estimation.

Recap of IV and 2SLS Lingo

- **Endogenous variables** The endogenous variables are the dependent variable and the independent variable(s) to be instrumented; in a simultaneous equations model, endogenous variables are determined by solving a system of stochastic linear equations.
- **Exogenous variables** The exogenous variables include the exogenous covariates that are not instrumented and the instruments themselves. In a simultaneous equations model, exogenous variables are determined outside the system.
- 2SLS aficionados live in a world of mutually exclusive labels:
  - In any empirical study involving IV, the random variables to be studied are either **dependent variables**, **independent endogenous variables**, **instrumental variables**, or **exogenous covariates**.
  - Sometimes we shorten this to dependent and endogenous variables, instruments, and covariates (fudging the fact that the dependent variable is also endogenous in a traditional SEM).

Recap of Assumptions on IV with Covariates

- The instrument must have a clear effect on s_i conditional on X_i ***(First Stage)***
  - The first stage is not zero, but this is something you can check in the data
  - If the first stage is only marginally significantly different from zero, the resulting IV estimates are unlikely to be informative, a point we return to later
- The only reason for the relationship between Y_i and Z_i is the first stage ***(Exclusion Restriction)***
  - First, the instrument is as good as randomly assigned conditional on X_i (i.e., Independent of potential outcomes, conditional on covariates, like the CIA)
  - Second, the instrument has no effect on outcomes other than through the first-stage channel

Outline

Wald Estimator

- Without covariates, the causal regression model is

$$Y_i = \alpha + \rho s_i + \eta_i$$

  where η_i and s_i may be correlated.
- As an IV estimator, we have known that

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(s_i, Z_i)} = \frac{Cov(Y_i, Z_i)/V(Z_i)}{Cov(s_i, Z_i)/V(Z_i)}$$

- Given that Z_i is a dummy variable that equals one with probability p, we can easily show that

$$\begin{aligned}
\operatorname{Cov}(Y_i, Z_i) &= [E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)]\, p \cdot (1 - p) \\
\operatorname{Cov}(s_i, Z_i) &= [E(s_i \mid Z_i = 1) - E(s_i \mid Z_i = 0)]\, p \cdot (1 - p)
\end{aligned}$$

- **Proof:**

$$\begin{aligned}
\operatorname{Cov}(Y_i, Z_i) &= E\left\{\left[Y_i - E(Y_i)\right]\left[Z_i - E(Z_i)\right]\right\} \\
&= E_{Z_i}\left\{E\left\{\left[Y_i - E(Y_i)\right]\left[Z_i - E(Z_i)\right] \mid Z_i\right\}\right\} \\
&= E\left\{\left[Y_i - E(Y_i)\right]\left[Z_i - E(Z_i)\right] \mid Z_i = 1\right\} \mathrm{P}(Z_i = 1) \\
&\quad + E\left\{\left[Y_i - E(Y_i)\right]\left[Z_i - E(Z_i)\right] \mid Z_i = 0\right\} \mathrm{P}(Z_i = 0) \\
&= E\left\{\left[Y_i - E(Y_i)\right] \mid Z_i = 1\right\} p \cdot (1-p) \\
&\quad - E\left\{\left[Y_i - E(Y_i)\right] \mid Z_i = 0\right\} p \cdot (1-p) \\
&= E\left[Y_i \mid Z_i = 1\right] p \cdot (1-p) - E\left[Y_i \mid Z_i = 0\right] p \cdot (1-p) \\
&= \left[E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)\right] p \cdot (1-p)
\end{aligned}$$

- Similarly,

$$Cov(s_i, Z_i) = \left[E(s_i \mid Z_i = 1) - E(s_i \mid Z_i = 0)\right] p \cdot (1-p)$$

- It therefore follows that (**Wald Formula**)

$$\rho = \frac{Cov\left(Y_i, Z_i\right)}{Cov\left(s_i, Z_i\right)} = \frac{\left[E\left(Y_i \mid Z_i = 1\right) - E\left(Y_i \mid Z_i = 0\right)\right]}{\left[E\left(s_i \mid Z_i = 1\right) - E\left(s_i \mid Z_i = 0\right)\right]}$$

- Another route to get this result

$$\begin{aligned} E\left(Y_i \mid Z_i\right) &= \alpha + \rho E\left(s_i \mid Z_i\right) + E\left(\eta_i \mid Z_i\right) \\ &= \alpha + \rho E\left(s_i \mid Z_i\right) \end{aligned}$$

- Proof:

$$\begin{aligned} E\left(Y_i \mid Z_i = 1\right) &= \alpha + \rho E\left(s_i \mid Z_i = 1\right) \\ E\left(Y_i \mid Z_i = 0\right) &= \alpha + \rho E\left(s_i \mid Z_i = 0\right) \\ E\left(Y_i \mid Z_i = 1\right) - E\left(Y_i \mid Z_i = 0\right) &= \rho\left[E\left(s_i \mid Z_i = 1\right) - E\left(s_i \mid Z_i = 0\right)\right] \end{aligned}$$

Comments:

- The principal claim that motivates IV estimation of causal effects is that the only reason for any relation between the dependent variable and the instrument is the effect of the instrument on the causal variable of interest.
- In the context of a dummy instrument, it therefore seems natural to divide—or rescale—the reduced-form difference in means by the corresponding first-stage difference in means.
- Straightforward and transparent

Example 1: Angrist and Krueger (1991),QJE

	(1) Born in 1st Quarter of Year	(2) Born in 4th Quarter of Year	(3) Difference (Std. Error) (1) – (2)
ln (weekly wage)	5.892	5.905	−.0135 (.0034)
Years of education	12.688	12.839	−.151 (.016)
Wald estimate of return to education			.089 (.021)
OLS estimate of return to education			.070 (.0005)

Notes: From Angrist and Imbens (1995). The sample includes native-born men with positive earnings from the 1930–39 birth cohorts in the 1980 census 5 percent file. The sample size is 162,515.

Example 2: Angrist (1990), AER

- How does veteran status impact the labor market outcomes?
- In each year from 1970 to 1972, random sequence numbers (RSNs) were randomly assigned to each birth date in cohorts of 19-year-olds.
- Men with lottery numbers below a cutoff were eligible for the draft, while men with numbers above the cutoff could not be drafted. In practice, many draft-eligible men were still exempted from service for health or other reasons, while many men who were draft-exempt nevertheless volunteered for service.
- So veteran status was not completely determined by randomized draft eligibility, but draft eligibility provides a dummy instrument highly correlated with Vietnam-era veteran status.

Example 2: Angrist (1990), AER

- An important feature of the Wald/IV estimator is that the identifying assumptions are easy to assess and interpret.
- Let D_i denote Vietnam-era veteran status and Z_i indicate draft eligibility. The fundamental claim justifying our interpretation of the Wald estimator as capturing the causal effect of D_i is that the only reason why $E[Y_i|Z_i]$ changes as Z_i changes is the variation in $E[D_i|Z_i]$.
- Specification test of the exclusion condition:
  - A simple check on this is to look for an association between Z_i and personal characteristics that should not be affected by D_i, for example, race, sex, or any other characteristic that was determined before D_i was determined.
  - Another useful check is to look for an association between the instrument and outcomes in samples where there is no relationship between D_i and Z_i **(Placebo test)**. If the only reason for draft eligibility effects on earnings is veteran status, then draft eligibility effects on earnings should be zero in samples where draft eligibility status is unrelated to veteran status.

Wald estimates of the effects of military service on the earnings of white men born in 1950

	Earnings		Veteran Status		
Earnings Year	Mean (1)	Eligibility Effect (2)	Mean (3)	Eligibility Effect (4)	Wald Estimate of Veteran Effect (5)
1981	16,461	−435.8	.267	.159	−2,741
		(210.5)		(.040)	(1,324)
1971	3,338	−325.9			−2,050
		(46.6)			(293)
1969	2,299	−2.0			
		(34.5)			

Notes: Adapted from Angrist (1990), tables 2 and 3. Standard errors are shown in parentheses. Earnings data are from Social Security administrative records. Figures are in nominal dollars. Veteran status data are from the Survey of Income and Program Participation. There are about 13,500 individuals in the sample.

- It's comforting that the draft eligibility treatment effect on 1969 earnings is zero, since 1969 earnings predate the 1970 draft lottery.
- No significant relationship between earnings and draft eligibility status for men born in 1953, supporting that the only reason for draft eligibility effects is military service.

Example 3: Angrist and Evans (1998)

twins → family size → mother work
sex ratio →

- How does family size impact mothers' employment and work?
- **Twins instrument**, a dummy for a multiple second birth in a sample of mothers with at least two children.
- **Sibling sex composition**, an instrument motivated by the fact that American parents with two children are much more likely to have a third child if the first two are of the same sex than if the sex composition is mixed.
- The twins (sibling sex composition) instrument rests on the idea that the occurrence of a multiple birth (sibling sex composition) is essentially random, unrelated to potential outcomes or family background, and affects family labor supply solely by increasing fertility.

Table 4.1.4: Wald estimates of labor supply effects

			IV Estimates using:			
			Twins		Sex-composition	
Dependent variable	Mean	OLS	First stage	Wald estimates	First stage	Wald estimates
	(1)	(2)	(3)	(4)	(5)	(6)
Employment	0.528	-0.167	0.625	-0.083	0.067	-0.135
		(0.002)	(0.011)	(0.017)	(0.002)	(0.029)
Weeks worked	19.0	-8.05		-3.83		-6.23
		(0.09)		(0.758)		(1.29)
Hours/week	16.7	-6.02		-3.39		-5.54
		(0.08)		(0.637)		(1.08)

Note: The table reports OLS and Wald 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) extract including married women aged 21-35 with at least two children in the 1980 Census. OLS models include controls for mother's age, age at first birth, dummies for the sex of first and second births, and dummies for race.

- The twins first-stage is 0.625. What does it mean?
- This means that 37.5 percent of mothers with two or more children would have had a third birth anyway; a multiple third birth increases this proportion to 1.
- Why the two IV estimates differs so much? Which one should we trust?

Outline

IV with Heterogeneous Treatment Effect (LATE)

- The discussion of IV up to this point postulates a constant causal effect.
- In the case of a dummy variable such as veteran status, this means

$$Y_{1i} - Y_{0i} = \rho \quad \text{for all } i,$$

  while with a multivalued treatment such as schooling, this means

$$Y_{si} - Y_{s-1,i} = \rho \quad \text{for all } s \text{ and all } i.$$

- Why is treatment effect heterogeneity important? The answer lies in the distinction between the two types of validity that characterize a research design.

Internal vs External Validity

Internal Validity

is the question of whether a given design successfully uncovers causal effects for the population being studied. A randomized clinical trial, or for that matter a good IV study, has a strong claim to internal validity.

External Validity

is the predictive value of the study's findings in a different context. For example, if the study population in a randomized trial is especially likely to benefit from treatment, the resulting estimates may have little external validity.

Local Average Treatment Effects

- Let $Y_i(d, z)$ denote the potential outcome of individual i were this person to have treatment status $D_i = d$ and instrument value $Z_i = z$.
  - This tells us, for example, what the earnings of i would be given alternative combinations of veteran status and draft eligibility status.
- The causal effect of veteran status given i's realized draft eligibility status is

$$Y_i(1, Z_i) - Y_i(0, Z_i)$$

while the causal effect of draft eligibility status given i's veteran status is

$$Y_i(D_i, 1) - Y_i(D_i, 0)$$

- Instrumental variables initiate a causal chain where the instrument Z_i affects the variable of interest, D_i, which in turn affects outcomes, Y_i.
- Let D_{1i} be i's potential treatment status when $Z_i = 1$, while D_{0i} is i's potential treatment status when $Z_i = 0$. Observed treatment status is therefore

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i = \pi_0 + \pi_{1i}Z_i + \xi_i$$

  In random coefficients notation, $\pi_0 \equiv \mathbb{E}[D_{0i}]$ and $\pi_{1i} \equiv (D_{1i} - D_{0i})$, so π_{1i} is the heterogeneous causal effect of the instrument on D_i.
  - Only one of the potential treatment assignments, D_{1i} and D_{0i}, is ever observed for any one person. In the draft lottery example, D_{0i} tells us whether i would serve in the military if he drew a high (draft-ineligible) lottery number, while D_{1i} tells us whether i would serve if he drew a low (draft-eligible) lottery number.
  - We get to see one or the other of these potential assignments depending on Z_i. The average causal effect of Z_i on D_i is $\mathbb{E}[\pi_{1i}]$.
- Note: $D_{0i} = 1$ or 0, $D_{1i} = 1$ or 0; potential vs. realized; assigned vs. received.

Key Assumptions for LATE Theorem

Assumption 1: Independence Assumption

The instrument is as good as randomly assigned: it is independent of the vector of potential outcomes and potential treatment assignments. Formally, this can be written as

$$[\{Y_i(d,z); \forall d, z\}, D_{1i}, D_{0i}] \perp Z_i$$

Comments:

This independence assumption is sufficient for a causal interpretation of the reduced form, that is, the regression of Y_i on Z_i. Specifically,

$$\begin{aligned} E\left(Y_i \mid Z_i = 1\right) - E\left(Y_i \mid Z_i = 0\right) &= E\left(Y_i\left(D_{1i}, 1\right) \mid Z_i = 1\right) - E\left(Y_i\left(D_{0i}, 0\right) \mid Z_i = 0\right) \\ &= E\left[Y_i\left(D_{1i}, 1\right) - Y_i\left(D_{0i}, 0\right)\right] \end{aligned}$$

represents the causal effect of the instrument on Y_i.

Note: Independence, Zero Conditional Mean, and Uncorrelated

- In a regression function, for the error term u_i, satisfying $E(u_i) = 0$ and a random variable Z_i, we have:
  - If $u_i \perp Z_i$ (u_i is independent of Z_i), then $E[u_i \mid Z_i] = 0$.
  - If $E[u_i \mid Z_i] = 0$, then $\text{Cov}(Z_i, u_i) = 0$
- **Proof:**
  (Part 1) u_i and Z_i are statistically independent means knowing the value of Z_i provides no information about the value of u_i. Mathematically, this means the conditional distribution of u_i given Z_i is identical to its marginal (unconditional) distribution. That is, $E[u_i \mid Z_i] = E[u_i] = 0$
  (Part 2)

$$\text{Cov}(Z_i, u_i) = E[Z_i u_i] - E[Z_i]\,E[u_i] = E[Z_i u_i]$$

Using the law of iterated expectations:

$$E[Z_i u_i] = E[E[Z_i u_i \mid Z_i]] = E[Z_i E[u_i \mid Z_i]] = E[Z_i \times 0] = 0$$

- Independence also means that

$$\begin{aligned} & E\left(D_i \mid Z_i = 1\right) - E\left(D_i \mid Z_i = 0\right) \\ & = E\left(D_{1i} \mid Z_i = 1\right) - E\left(D_{0i} \mid Z_i = 0\right) \\ & = E\left[D_{1i} - D_{0i}\right] \end{aligned}$$

- In other words, the first stage from our earlier discussion of 2SLS captures the causal effect of Z_i on D_i.

Assumption 2: Exclusion Restriction

- $Y_i(d, z)$ is a function only of d
- In general, the claim that an instrument operates through a single known causal channel is called an **exclusion restriction**. To be specific, while draft eligibility clearly affects veteran status, an individual's potential earnings as a veteran or nonveteran are assumed to be unchanged by draft eligibility status.
- Formally, the exclusion restriction says that

$$Y_i(d, 0) = Y_i(d, 1) \equiv Y_{di} \text{ for } d = 0, 1.$$

- In a linear model with constant effects, the exclusion restriction is expressed by the omission of instruments from the causal equation of interest and by saying that $E[z_i \eta_i] = 0$
- We need z_i and η_i to be uncorrelated, but the reasoning that lies behind this assumption is unclear until we consider the independence and exclusion restrictions as distinct propositions.

Independence vs. Exclusion Restriction

- The exclusion restriction is distinct from the claim that the instrument is (as good as) randomly assigned. Rather, it is a claim about a unique channel for causal effects of the instrument.
- The fact that the lottery number is randomly assigned (and therefore satisfies the independence assumption) does not make this possibility less likely.
  - The exclusion restriction fails for draft lottery instruments if men with low draft lottery numbers were affected in some way other than through an increased likelihood of military service.
  - For example, Angrist and Krueger (1992) looked for an association between draft lottery numbers and schooling. Their idea was that educational draft deferments could have led men with low lottery numbers to stay in college longer than they would have otherwise desired.

More Implications of Exclusion Restriction

- Using the exclusion restriction, we can define potential outcomes indexed solely against treatment status using the single index (Y_{1i}, Y_{0i}) notation:

$$Y_{1i} \equiv Y_i(1,1) = Y_i(1,0)$$
$$Y_{0i} \equiv Y_i(0,1) = Y_i(0,0)$$

- The observed outcome, Y_i, can therefore be written as:

$$\begin{aligned} Y_i &= Y_i(0, Z_i) + [Y_i(1, Z_i) - Y_i(0, Z_i)]D_i \\ &= Y_{0i} + (Y_{1i} - Y_{0i})\, D_i \end{aligned}$$

- Random coefficients notation for this is

$$Y_i = \alpha_0 + \rho_i D_i + \eta_i$$

Where $\alpha_0 \equiv E[Y_{0i}]$ and $\rho_i \equiv Y_{1i} - Y_{0i}$

More Key Assumptions

- **Assumption 3: the existence of the First Stage**

$$E\left(D_{1i} - D_{0i}\right) \neq 0$$

- **Assumption 4: Monotonicity Assumption (Imbens and Angrist, 1994)**

$$D_{1i} \geq D_{0i} \text{ or } D_{1i} \leq D_{0i} \text{ for all } i$$

  - Means that while the instrument may have no effect on some people, all those who are affected are affected in the same way.
  - In what follows, we assume monotonicity holds with $D_{1i} \geq D_{0i}$. In the draft lottery example, this means that although draft eligibility may have had no effect on the probability of military service for some men, there is no one who was actually kept out of the military by being draft-eligible.

The LATE Theorem

- **Suppose**
  - (A1, Independence): $[\{Y_i(d,z); \forall d, z\}, D_{1i}, D_{0i}] \perp Z_i$
  - (A2, Exclusion): $Y_i(d,0) = Y_i(d,1) \equiv Y_{di}$ for $d = 0, 1$;
  - (A3, First stage): $\mathbb{E}(D_{1i} - D_{0i}) \neq 0$
  - (A4, Monotonicity): $D_{1i} - D_{0i} \geq 0$ for all i, or vice versa;
- **Then**

$$\frac{[\mathbb{E}(Y_i \mid Z_i = 1) - \mathbb{E}(Y_i \mid Z_i = 0)]}{[\mathbb{E}(D_i \mid Z_i = 1) - \mathbb{E}(D_i \mid Z_i = 0)]} = \mathbb{E}[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}]$$
$$= \mathbb{E}[\rho_i \mid \pi_{1i} > 0]$$

Proof of LATE Theorem

- Use the exclusion restriction to write:

$$\begin{aligned} E(Y_i \mid Z_i = 1) &= E(Y_{0i} + (Y_{1i} - Y_{0i})D_i \mid Z_i = 1) \\ &= E[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}] \\ E(Y_i \mid Z_i = 0) &= E(Y_{0i} + (Y_{1i} - Y_{0i})D_i \mid Z_i = 0) \\ &= E[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}] \end{aligned}$$

- So the numerator of the Wald estimator is:

$$E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0) = E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})]$$

- Monotonicity means $D_{1i} - D_{0i}$ equals one or zero, so:

$$E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})] = E[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}]P(D_{1i} > D_{0i})$$

Proof of LATE Theorem

- A similar argument shows:

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

- **Verified.**

Comments

- This theorem says that an instrument that is *as good as randomly assigned*, *affects the outcome through a single known channel*, *has a first stage*, and *affects the causal channel of interest only in one direction* can be used to estimate **the average causal effect on the affected group**.
- IV estimates of effects of military service using the draft lottery capture the effect of military service on men who served because they were draft eligible but who would not otherwise have served.
- This excludes:
  - **Volunteers** (always takers)
  - **Men exempted from military service for medical reasons** (never takers)
- It includes men for whom the draft policy was binding.

What is the use of the LATE?

- No theorem answers this question, but it is always worth discussing.
- On the other hand, while draft lottery instruments produce internally valid estimates of the causal effect of Vietnam-era conscription, the external validity – that is, the predictive value of these estimates for military service in other times and places – is not directly addressed by the IV framework.
- There is nothing in IV formulas to explain why Vietnam-era service affects earnings; for that, you need a theory (lost labor market experience).

Monotonicity, Defiers, and the LATE Theorem

- A failure of monotonicity means the instrument pushes some people into treatment while pushing others out ("defiers").
- Defiers complicate the link between LATE and the reduced form:

$$E\left(Y_i \mid Z_i = 1\right) - E\left(Y_i \mid Z_i = 0\right) = E\left[\left(Y_{1i} - Y_{0i}\right)\left(D_{1i} - D_{0i}\right)\right]$$

- Without monotonicity, this is equal to

$$\begin{aligned} E\left[\left(Y_{1i} - Y_{0i}\right)\left(D_{1i} - D_{0i}\right)\right] = {} & E\left[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}\right] P\left(D_{1i} > D_{0i}\right) \\ & - E\left[Y_{1i} - Y_{0i} \mid D_{1i} < D_{0i}\right] P\left(D_{1i} < D_{0i}\right) \end{aligned}$$

- We might therefore have a scenario where treatment effects are positive for everyone yet the reduced form is zero because effects on compliers are canceled out by effects on defiers.
- This doesn't come up in a constant effects model because the reduced form is always the constant effect times the first stage, regardless of whether the first stage includes defiant behavior.

The Compliant Subpopulation

- **Compliers**
  The LATE framework partitions any population with an instrument into a set of four instrument-dependent subgroups, defined by the manner in which members of the population react to the instrument.
- **Definition:**
  - Compliers. The subpopulation with $D_{1i} = 1$ and $D_{0i} = 0$.
  - Always-takers. The subpopulation with $D_{1i} = D_{0i} = 1$.
  - Never-takers. The subpopulation with $D_{1i} = D_{0i} = 0$.
  - Defiers (assumed nonexistent). The subpopulation with $D_{1i} = 0$ and $D_{0i} = 1$.

- LATE is the effect of treatment on the population of compliers.
- The term "compliers" comes from an analogy with randomized trials where some experimental subjects comply with the randomly assigned treatment protocol (e.g., take their medicine) but some do not, while some control subjects obtain access to the experimental treatment even though they are not supposed to.
- Without adding further assumptions (e.g., constant causal effects), LATE is not informative about effects on never-takers and always-takers because, by definition, treatment status for these two groups is unchanged by the instrument.
- IV solves the problem of causal inference in a randomized trial with partial compliance.

LATE vs. ATT

- The average causal effect on compliers (LATE) is not usually the same as the average treatment effect on the treated (ATT).
- From the simple fact that $D_i = D_{0i} + (D_{1i} - D_{0i})Z_i$, we learn that the treated population consists of two non-overlapping groups.
  - By monotonicity, we cannot have both $D_{0i} = 1$ and $D_{1i} - D_{0i} = 1$, since $D_{0i} = 1$ implies $D_{1i} = 1$.
  - The treated therefore have either $D_{0i} = 1$ or $D_{1i} - D_{0i} = 1$ and $Z_i = 1$, and hence D_i can be written as the sum of two mutually exclusive dummies, D_{i0} and $(D_{1i} - D_{0i})Z_i$. In other words, the treated consist of either always-takers or of compliers with the instrument switched on.
- Since the instrument is as good as randomly assigned, compliers with the instrument switched on are representative of all compliers.

ATT Decomposition

- From definition of ATT, we get

$$\underbrace{E\left[Y_{1i} - Y_{0i} \mid D_i = 1\right]}_{\text{Treatment Effect on the treated}}$$
$$= \underbrace{E\left[Y_{1i} - Y_{0i} \mid D_{0i} = 1\right] P\left[D_{0i} = 1 \mid D_i = 1\right]}_{\text{Treatment Effect on always-takers}}$$
$$+ \underbrace{E\left[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}, Z_i = 1\right] P\left[D_{1i} > D_{0i}, Z_i = 1 \mid D_i = 1\right]}_{\text{Treatment Effect on compliers}}$$

Since $P\left[D_{0i} = 1 \mid D_i = 1\right]$ and $P\left[D_{1i} > D_{0i}, Z_i = 1 \mid D_i = 1\right]$ add up to one, this means that the effect of treatment on the treated is a weighted average of effects on always-takers and compliers.

LATE vs. ATNT

- Likewise, LATE is not the average causal effect of treatment on the nontreated (ATNT), $E[Y_{1i} - Y_{0i}|D_i = 0]$.
- In the draft lottery example, the average effect on the nontreated is the average causal effect of military service on the population of nonveterans from Vietnam-era cohorts. The average effect of treatment on the nontreated is a weighted average of effects on never-takers and compliers. In particular,

$$\underbrace{E\left[Y_{1i} - Y_{0i} \mid D_i = 0\right]}_{\text{Treatment Effect on the nontreated}}$$
$$= \underbrace{E\left[Y_{1i} - Y_{0i} \mid D_{1i} = 0\right] P\left[D_{1i} = 0 \mid D_i = 0\right]}_{\text{Treatment Effect on never-takers}}$$
$$+ \underbrace{E\left[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}, Z_i = 0\right] P\left[D_{1i} > D_{0i}, Z_i = 0 \mid D_i = 0\right]}_{\text{Treatment Effect on compliers}}$$

ATE vs. LATE

- Finally

$$\begin{aligned} E[Y_{1i} - Y_{0i}] =& E[Y_{1i} - Y_{0i} \mid D_i = 1]P[D_i = 1] \\ &+ E[Y_{1i} - Y_{0i} \mid D_i = 0]P[D_i = 0] \end{aligned}$$

  shows the unconditional average treatment effect to be a weighted average of effects on compliers, always-takers, and never-takers.
- Because an IV is not directly informative about effects on always-takers and never-takers, instruments do not usually capture the average causal effect on all of the treated or on all of the nontreated.

Special Cases When LATE Equals ATT or ATNT

- Angrist and Evans (1998) estimate the causal effect of having three children on earnings in the population of women with at least two children.
- Let T_i be a dummy variable indicating multiple second births, IV
- Let Y_{0i} denote potential earnings if a woman has only two children while Y_{1i} denotes her potential earnings if she has three, an event indicated by D_i.
- Assuming that T_i is as good as randomly assigned, that fertility increases by at most one child in response to a multiple birth, and that multiple births affect outcomes only by increasing fertility, LATE using the twins instrument T_i is also $E[Y_{1i} - Y_{0i}|D_i = 0]$, *the average causal effect on women who are not treated* (i.e., have two children only).
- This is because all women who have a multiple second birth end up with three children, that is, there are *no never-takers* in response to the twins instrument.

没有 never-takers: LATE = ATNT

Oreopoulos (2006)

- His study estimates the economic returns to schooling using an increase in the British compulsory attendance age from 14 to 15.
- Compliance with the Britain's new compulsory attendance law was near perfect, though many teens would previously have dropped out of school at age 14. 没有 never-takers
- Since everybody in Oreopoulos's British sample finished an additional year when compulsory schooling laws were made stricter, there are no never-takers. Oreopoulos's IV strategy therefore captures the average causal effect of obtaining one more year of high school on all those who leave school at 14 (ATNT).
- Oreopoulos's IV strategy wouldn't estimate the effect of treatment on the nontreated in, say, Israel, where teenagers get more leeway when it comes to compulsory school attendance.

IV in Randomized Trials

- Sometimes, IV is generated by a randomized trial with *one-sided noncompliance*: Participation is voluntary among those randomly assigned to receive treatment; no one in the control group has access to the experimental intervention.
- Since the group that receives (i.e., complies with) the assigned treatment is a self-selected subset of those offered treatment, *a comparison between those actually treated and the control group is misleading*.
- IV using randomly assigned treatment intended as an instrumental variable for treatment received solves this sort of compliance problem.
- *LATE is the effect of treatment on the treated* in this case. Suppose the instrument Z_i is a dummy variable indicating random assignment to a treatment group, while D_i is a dummy indicating whether treatment was actually received. In practice, because of noncompliance, D_i is not equal to Z_i.

Bloom et al.(1997)

- Job Training Partnership Act (JTPA) training program: 1982, USA, to establish federal assistance programs to prepare youth and unskilled adults for entry into the labor force and to provide job training to economically disadvantaged and other individuals facing serious barriers to employment.
- Only 60 percent of those assigned to be trained received training, while roughly 2 percent of those assigned to the control group received training anyway (Bloom et al., 1997).
- Since the compliance problem in this case was largely confined to the treatment group, LATE using random assignment, Z_i, as an instrument for treatment received, D_i, is the effect of treatment on the treated.

Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status (OLS)		Comparisons by Assignment Status (ITT)		Instrumental Variable Estimates (IV)	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

Notes: Authors' tabulation of JTPA study data. The table reports OLS, ITT, and IV estimates of the effect of subsidized training on earnings in the JTPA experiment. Columns 1 and 2 show differences in earnings by training status; columns 3 and 4 show differences by random-assignment status. Columns 5 and 6 report the result of using random-assignment status as an instrument for training. The covariates used for columns 2, 4, and 6 are high school or GED, black, Hispanic, married, worked less than 13 weeks in past year, AFDC (for women), plus indicators for the JTPA service strategy recommended, age group, and second follow-up survey. Robust standard errors are shown in parentheses. There are 5,102 men and 6,102 women in the sample.

Intention-to-Treat (ITT) Analysis

- Since individuals assigned to the treatment group were free to decline (and 40 percent did so), this comparison throws away the random assignment unless the decision to comply was itself independent of potential outcomes.
- The contrast in columns 3 and 4 is known as an intention-to-treat (ITT) effect. Since Z_i was randomly assigned, ITT effects have a causal interpretation:
  - They tell us the causal effect of the offer of treatment, building in the fact that many of those offered have declined to participate. For this reason, the ITT effect is too small relative to the average causal effect on those who were in fact treated.
- Columns 5 and 6 put the pieces together and give us the most interesting effect: ITT divided by the difference in compliance rates between treatment and control groups as originally assigned (about 0.6).

The Bloom Result

- In this case, one relationship holds: ITT divided by compliance is the effect of treatment on the treated (ATT)
- We can recognize ITT as the reduced-form effect of the randomly assigned offer of treatment, our instrument in this case.
- The compliance rate is the first stage associated with this instrument, and the Wald estimand, as always, is the reduced form divided by the first stage.
- In general, this equals LATE, but because we have (almost) no always-takers, the treated population consists (almost) entirely of compliers. The IV estimates in columns 5 and 6 of the Table above are therefore consistent estimates of the effect of treatment on the treated.

The Bloom Theorem and Proof

Theorem (The Bloom Result)

Suppose the assumptions of the LATE theorem hold, and

$$E[D_i|Z_i = 0] = P[D_i = 1|Z_i = 0] = 0$$

Then

$$\frac{E\left(Y_i \mid Z_i = 1\right) - E\left(Y_i \mid Z_i = 0\right)}{P\left(D_i = 1 \mid Z_i = 1\right)} = E\left[Y_{1i} - Y_{0i} \mid D_i = 1\right]$$

The Bloom Result

- ***Proof:*** Since

$$E\left(Y_i \mid Z_i = 1\right) = E\left(Y_{0i} \mid Z_i = 1\right) + E\left[\left(Y_{1i} - Y_{0i}\right) D_i \mid Z_i = 1\right]$$

- While $E\left(Y_i \mid Z_i = 0\right) = E\left(Y_{0i} \mid Z_i = 0\right)$. Because $Z_i = 0$ means $D_i = 0$ when always-takers do not exist.

$$E\left(Y_i \mid Z_i = 1\right) - E\left(Y_i \mid Z_i = 0\right) = E\left[\left(Y_{1i} - Y_{0i}\right) D_i \mid Z_i = 1\right]$$

  Since $E[Y_{0i} \mid Z_i = 0] = E[Y_{0i} \mid Z_i = 1]$ by independence.

- But

$$E\left[\left(Y_{1i} - Y_{0i}\right) D_i \mid Z_i = 1\right] = E\left[\left(Y_{1i} - Y_{0i}\right) \mid D_i = 1, Z_i = 1\right] P\left(D_i = 1 \mid Z_i = 1\right)$$

  While $D_i = 1$ implies $Z_i = 1$, since no one with $Z_i = 0$ is treated. Hence

$$E\left[\left(Y_{1i} - Y_{0i}\right) \mid D_i = 1, Z_i = 1\right] = E\left[\left(Y_{1i} - Y_{0i}\right) \mid D_i = 1\right]$$

An Example: The Minneapolis Domestic Violence Experiment (MDVE)

- The MDVE was a pioneering effort to determine the best police response to domestic violence (Sherman and Berk, 1984).
- Whether a hardline response—arrest and at least temporary incarceration—is productive, especially in view of the fact that domestic assault charges are frequently dropped.
- The research design used randomly shuffled color-coded report forms telling the responding officers to arrest some perpetrators while referring others to counseling or merely separating the parties.
- In practice, however, the police were free to overrule the random assignment. For example, an especially dangerous or drunk offender was arrested no matter what.

MDVE Example

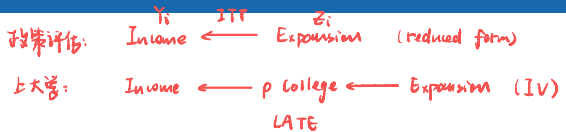
- As a result, the actual response often deviated from the randomly assigned response, though the two are highly correlated.
- Most published analyses of the MDVE data recognize this compliance problem and focus on ITT effects, that is, they use the original random assignment and not the treatment actually delivered.
- But the MDVE data can also be used to get the average causal effect on compliers, in this case those who were not arrested because they were randomly assigned to be treated differently but would have been arrested otherwise.

MDVE Analysis in Angrist (2006)

- The MDVE is analyzed in this spirit in Angrist (2006). Because everyone in the MDVE who was assigned to be arrested was in fact arrested, there are no never-takers. This is an interesting twist and the flip side of the Bloom scenario: here we have $D_{1i} = 1$ for everybody.
- Consequently, LATE is the effect of treatment on the nontreated, that is,

$$E[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}] = E\left[(Y_{1i} - Y_{0i}) \mid D_i = 0\right]$$

How to Choose: ITT or LATE?



- Which question are you more interested in ?
  - Duflo (2001), AER: Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment
  - Angrist & Kruger (1991), QJE: Does Compulsory School Attendance Affect Schooling and Earnings?
- Statistically, it depends on whether exclusion restriction assumption is satisfied.

Counting and characterizing compliers

- We've seen that, except in special cases, each instrumental variable identifies a unique causal parameter, one specific to the subpopulation of compliers for that instrument.
- Different valid instruments for the same causal relation, therefore, estimate different things, at least in principle (an important exception being instruments that allow for perfect compliance on one side or the other).
- Differences in compliant subpopulations might explain variability in treatment effects from one instrument to another. We would therefore like to learn as much as we can about the compliers for different instruments.

- On the other hand, if the compliant subpopulations associated with two or more instruments are very different, yet the IV estimates they generate are similar, we might be prepared to adopt homogeneous effects as a working hypothesis.
- This revives the overidentification idea but puts it at the service of external validity. In fact, maintaining the hypothesis that all instruments in an overidentified model are valid, the traditional overidentification test statistic becomes a formal test for treatment effect homogeneity.
- Moreover, if the compliant subpopulation is similar to other populations of interest, the case for extrapolating estimated causal effects to these other populations is stronger.

The Size of a Complier Group

- The size of a complier group is easy to measure. This is just the Wald first stage, since, given monotonicity, we have

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

- We can also tell what proportion of the treated are compliers since, for compliers, treatment status is completely determined by Z_i. Start with the definition of conditional probability:

$$\begin{aligned} P[D_{1i} > D_{0i}|D_i = 1] &= \frac{P[D_i = 1|D_{1i} > D_{0i}]P[D_{1i} > D_{0i}]}{P[D_i = 1]} \\ &= \frac{P[Z_i = 1]\{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]\}}{P[D_i = 1]} \end{aligned}$$

The Size of a Complier Group

- Proof:

$$P[D_i = 1 | D_{1i} > D_{0i}] = P[Z_i = 1 | D_{1i} > D_{0i}]$$

and

$$P[Z_i = 1 | D_{1i} > D_{0i}] = P[Z_i = 1]$$

- In other words, the proportion of the treated who are compliers is given by the first stage, times the probability the instrument is switched on, divided by the proportion treated.

Table 4.4.2: Probabilities of compliance in instrumental variables studies

Source	Endogenous Variable (D)	Instrument (Z)	Sample	$P[D=1]$	1st Stage, $P[D_1 > D_0]$	$P[Z=1]$	$P[D_1 > D_0 \mid D=1]$	$P[D_1 > D_0 \mid D=0]$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Angrist (1990)	Veteran Status	Draft eligibility	White men born in 1950	0.267	0.159	0.534	0.318	0.101
			Non-white men born in 1950	0.163	0.060	0.534	0.197	0.033
Angrist and Evans (1998)	More than 2 children	Twins at second birth	Married women aged 21-35 with two or more children in 1980	0.381	0.603	0.008	0.013	0.966
		First two children are of the same sex	Married women aged 21-35 with two or more children in 1980	0.381	0.060	0.506	0.080	0.048
Angrist and Krueger (1991)	High school graduate	Third or fourth quarter birth	Men born between 1930 and 1939	0.770	0.016	0.509	0.011	0.034
Acemoglu and Angrist (2000)	High school graduate	State requires 11 or more years of school attendance	White men aged 40-49	0.617	0.037	0.300	0.018	0.068

Notes: The table shows an analysis of the absolute and relative size of the complier population for a number of instrumental variables. The first-stage, reported in column 6, gives the absolute size of the complier group. Columns 8 and 9 show the size of the complier population relative to the treated and untreated populations.

The characteristics of the compliers

- Unlike the size of the complier group, information on the characteristics of compliers seems like a tall order because the compliers cannot be individually identified. Because we can't see both D_{1i} and D_{0i} for each individual, we can't just list those with $D_{1i} > D_{0i}$ and then calculate the distribution of characteristics for this group.
- Nevertheless, in spite of the fact that compliers cannot be listed or named, it's easy to describe the distribution of complier characteristics.

Complier Characteristics Analysis

- Let x_{1i} be a Bernoulli-distributed characteristic, say a dummy indicating college graduates. Are sex composition compliers more or less likely to be college graduates than other women with two children?

$$\begin{aligned}\frac{P\left[x_{1i}=1 \mid D_{1i}>D_{0i}\right]}{P\left[x_{1i}=1\right]} &= \frac{P\left[D_{1i}>D_{0i} \mid x_{1i}=1\right]}{P\left[D_{1i}>D_{0i}\right]} \\ &= \frac{E\left[D_i \mid Z_i=1, x_{1i}=1\right]-E\left[D_i \mid Z_i=0, x_{1i}=1\right]}{E\left[D_i \mid Z_i=1\right]-E\left[D_i \mid Z_i=0\right]}\end{aligned}$$

- In other words, the relative likelihood a complier is a college graduate is given by the ratio of the first stage for college graduates to the overall first stage.

Proof:

$$P\left[x_{1i}=1 \mid D_{1i}>D_{0i}\right]=\frac{P\left[D_{1i}>D_{0i} \mid x_{1i}=1\right] P\left[x_{1i}=1\right]}{P\left[D_{1i}>D_{0i}\right]}$$

$$P\left[D_{1i}>D_{0i}\right]=E\left[D_i \mid Z_i=1\right]-E\left[D_i \mid Z_i=0\right]$$

$$P\left[D_{1i}>D_{0i} \mid x_{1i}=1\right]=E\left[D_i \mid Z_i=1, x_{1i}=1\right]-E\left[D_i \mid Z_i=0, x_{1i}=1\right]$$

Table 4.4.3: Complier-characteristics ratios for twins and sex-composition instruments

Variable	$E[x]$ (1)	Twins at second birth $E[x\|D_1 > D_0]$ (2)	Twins at second birth $P[x\|D_1 > D_0]/P[X]$ (3)	First two children are same sex $E[x\|D_1 > D_0]$ (6)	First two children are same sex $P[x\|D_1 > D_0]/P[X]$ (5)
Age 30 or older at first birth	0.00291	0.00404	1.39 (0.0201)	0.00233	0.995 (0.374)
Black or hispanic	0.125	0.103	0.822 (0.00421)	0.102	0.814 (0.0775)
High school graduate	0.822	0.861	1.048 (0.000772)	0.815	0.998 (0.0140)
College graduate	0.132	0.151	1.14 (0.00376)	0.0904	0.704 (0.0692)

Notes: The table reports an analysis of complier characteristics for twins and sex-composition instruments. The ratios in columns 3 and 5 give the relative likelihood compliers have the characteristic indicated in each row. Data are from the 1980 Census 5% sample, including married mothers age 21-35 with at least two children, as in Angrist and Evans (1998). The sample size is 254,654 for all columns.

Thanks for your Attention!

Any questions or comments please write to:

chenglingguo@nju.edu.cn