

Instrumental Variables and LATE

PP5001 · Week 3

Professor David A. Jaeger

Recall: Omitted Variables Bias

The true population relationship is

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 A_i + \nu_i,$$

where Y_i is earnings, D_i is **years of schooling**, and A_i is ability. β_1 is the causal effect of an additional year of schooling.

Recall the population regression of ability on schooling:

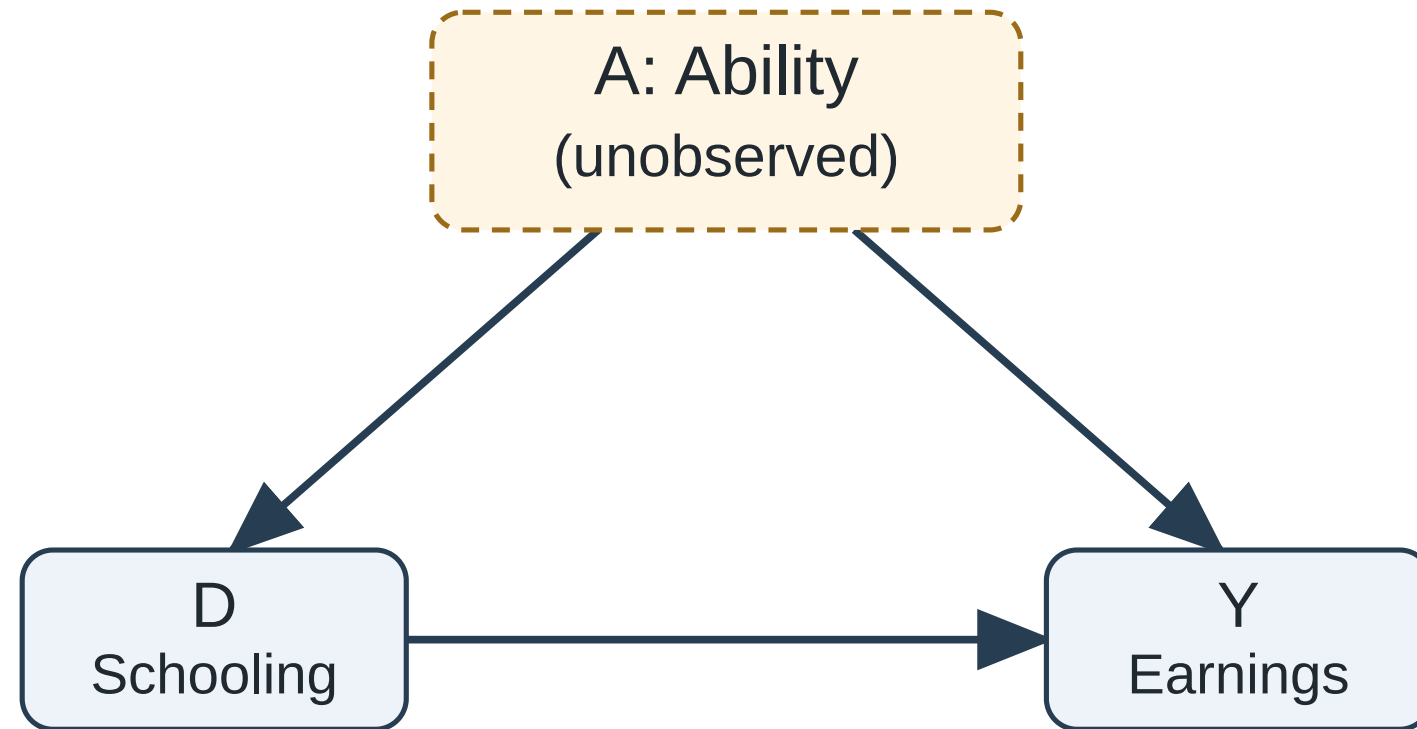
$$A_i = \gamma_0 + \gamma_1 D_i + v_i.$$

Omitting ability gives a short-regression slope of

$$\beta_1^S = \beta_1 + \beta_2 \gamma_1.$$

If ability raises earnings and is positively related to schooling, the short regression **overstates the causal effect of schooling**.

Omitted Variables Bias: The DAG



Ability affects both schooling and earnings, creating the backdoor path $D \leftarrow A \rightarrow Y$.

Our simple regression combines the causal effect along $D \rightarrow Y$ with the association arising through unobserved ability.

Bias and Unbiasedness of OLS

Across repeated samples, the estimated slope varies. Its **bias** is

$$\text{Bias} \left(\hat{\beta}_1^S \right) = E \left[\hat{\beta}_1^S \right] - \beta_1.$$

Imagine repeatedly observing outcomes for the **same schooling values**, $\mathbf{D} = (D_1, \dots, D_n)$. From $Y_i = \beta_0 + \beta_1 D_i + u_i$,

$$\hat{\beta}_1^S = \beta_1 + \frac{\sum_i (D_i - \bar{D}) u_i}{\sum_i (D_i - \bar{D})^2}.$$

Conditional on $\mathbf{D}$, the schooling values and denominator are fixed:

$$E \left[\hat{\beta}_1^S \mid \mathbf{D} \right] = \beta_1 + \frac{\sum_i (D_i - \bar{D}) E[u_i \mid \mathbf{D}]}{\sum_i (D_i - \bar{D})^2}.$$

If $E[u_i \mid \mathbf{D}] = 0$, OLS is **unbiased**. When omitted ability is related to schooling, this condition generally fails.

Omitted Ability and the Bias of OLS

Substitute $u_i = \beta_2 A_i + \nu_i$ and assume $E[\nu_i | \mathbf{D}] = 0$:

$$E[\hat{\beta}_1^S | \mathbf{D}] = \beta_1 + \beta_2 \frac{\sum_i (D_i - \bar{D}) E[A_i | \mathbf{D}]}{\sum_i (D_i - \bar{D})^2}.$$

Now assume the conditional mean of ability is linear in schooling:

$$E[A_i | \mathbf{D}] = \gamma_0 + \gamma_1 D_i.$$

Since $\sum_i (D_i - \bar{D}) = 0$, the numerator becomes

$$\sum_i (D_i - \bar{D}) (\gamma_0 + \gamma_1 D_i) = \gamma_1 \sum_i (D_i - \bar{D})^2.$$

Hence, using the population slope of the ability-on-schooling regression,

$$E[\hat{\beta}_1^S | \mathbf{D}] = \beta_1 + \beta_2 \gamma_1 = \beta_1 + \beta_2 \frac{\text{Cov}(A_i, D_i)}{\text{Var}(D_i)}.$$

What can we do if ability is unobserved?

The Wald Estimator: Setup

Write $Y_i = \beta_0 + \beta_1 D_i + u_i$, where $u_i = \beta_2 A_i + \nu_i$.

Suppose we have a binary instrument, Z_i , that has **two properties at the same time**:

$$E[u_i \mid Z_i = 1] = E[u_i \mid Z_i = 0] = 0,$$

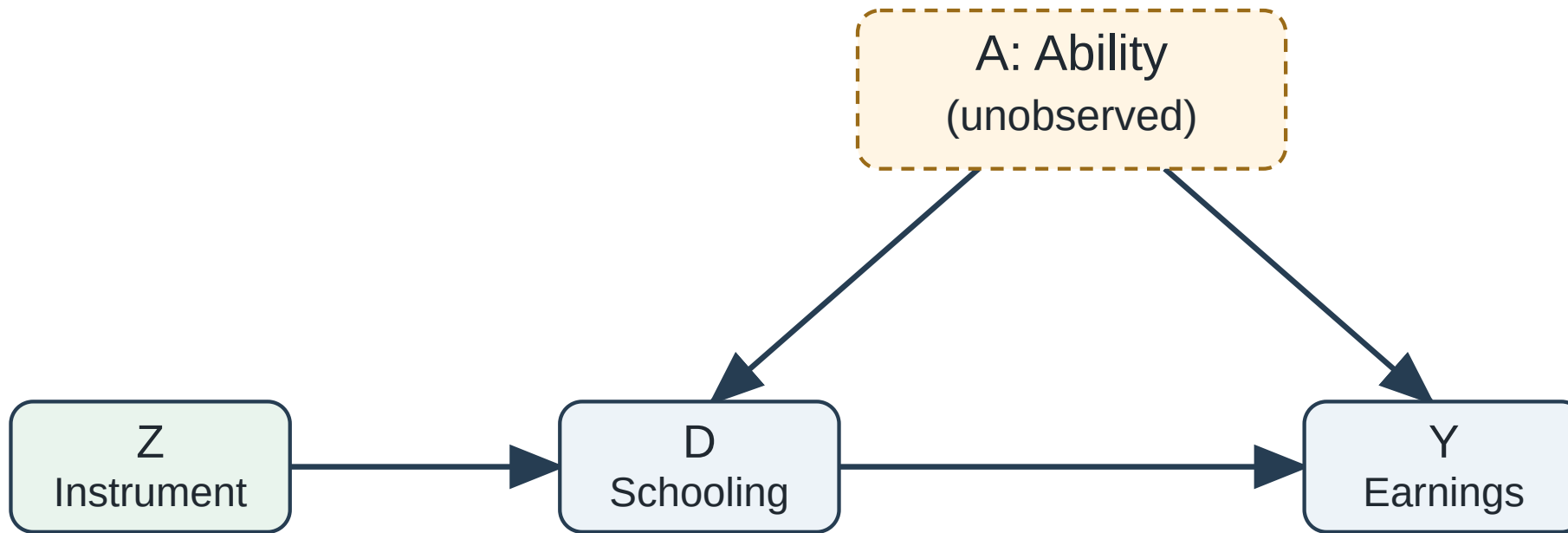
$$E[D_i \mid Z_i = 1] \neq E[D_i \mid Z_i = 0].$$

The instrument shifts schooling and affects earnings through schooling.

We call the first relationship the **orthogonality assumption** and the second relationship the **relevance assumption**.

Finding a variable with either property is easy. The challenge is finding one that satisfies **both at the same time**: it predicts years of schooling while remaining uncorrelated with the unobserved determinants of earnings.

The Instrument in the DAG



- **Relevance:** $Z \rightarrow D$ shifts schooling.
- **Exclusion:** the effect of Z on earnings operates through schooling.
- **Independence:** Z is independent of ability and the other unobserved determinants of earnings.

The last two conditions support the orthogonality assumption.

Deriving the Wald Estimator

Taking conditional expectations of the outcome equation:

$$\begin{aligned}E[Y_i \mid Z_i = 1] &= \beta_0 + \beta_1 E[D_i \mid Z_i = 1], \\E[Y_i \mid Z_i = 0] &= \beta_0 + \beta_1 E[D_i \mid Z_i = 0].\end{aligned}$$

Subtract the second equation from the first:

$$\underbrace{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]}_{\text{difference in outcomes}} = \beta_1 \underbrace{[E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]]}_{\text{difference in treatment}}.$$

Solve for β_1 and use the analogy principle to replace the population conditional means with the sample conditional means:

$$\hat{\beta}_{1,\text{Wald}} = \frac{\bar{Y}_{Z=1} - \bar{Y}_{Z=0}}{\bar{D}_{Z=1} - \bar{D}_{Z=0}}.$$

An Illustrative Schooling and Earnings Model

Generate 400 observations from these relationships:

$$D_i = 11 + 5Z_i + 1.8A_i + v_i, \quad Y_i = 100 + 5D_i + 20A_i + \nu_i.$$

Z_i is binary with probability $1/2$. Ability A_i and the disturbances v_i, ν_i are independent, mean-zero normal variables with standard deviations 1, 0.65, and 10; all are independent of Z_i .

The true schooling effect is **5**. Ability raises both schooling and earnings:

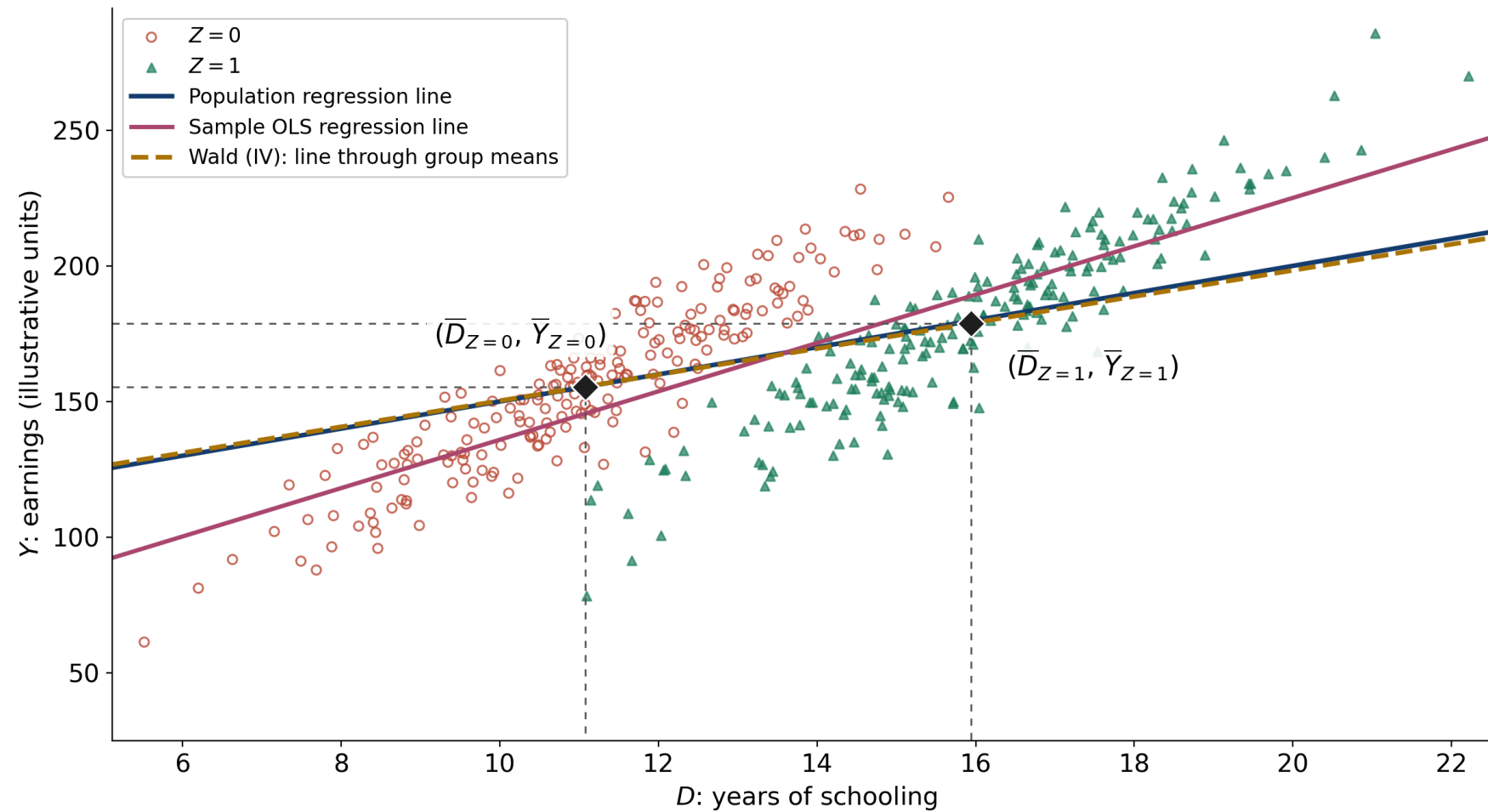
$$\text{Cov}(A_i, D_i) = 1.8, \quad \text{Var}(D_i) = 25(0.25) + 1.8^2 + 0.65^2 = 9.9125.$$

The population short-regression slope is therefore

$$\beta_1^s = 5 + 20 \frac{1.8}{9.9125} \approx 8.63.$$

Omitting ability gives an upward bias of approximately 3.63. The graph shows the sample OLS and Wald estimates alongside the true relationship.

The Wald Estimator: A Graphical Illustration



Understanding the Wald Estimator

- **Population regression line:** the true causal slope is $\beta_1 = 5$.
- **Sample OLS regression line:** omitted ability affects both D_i and Y_i , making the estimated slope too steep in this example.
- **Instrument groups:** open circles have $Z_i = 0$; triangles have $Z_i = 1$.
- **Wald (IV) line:** passes through the two group means, $(\bar{D}_{Z=0}, \bar{Y}_{Z=0})$ and $(\bar{D}_{Z=1}, \bar{Y}_{Z=1})$.

$$\text{Slope of IV line} = \frac{\bar{Y}_{Z=1} - \bar{Y}_{Z=0}}{\bar{D}_{Z=1} - \bar{D}_{Z=0}} = \frac{\text{vertical difference between means}}{\text{horizontal difference between means}}.$$

The instrument shifts D_i independently of ability. Sampling variation means that the IV line need not coincide exactly with the population line.

First Stage and Reduced Form: Terminology

- **First stage:** regress the endogenous explanatory variable on **all the exogenous variables**.
- **Reduced form:** regress the outcome on **all the exogenous variables**.

Both include the excluded instrument(s) and any included exogenous covariates, together with an intercept.

The term **reduced form** comes from simultaneous-equations models: we solve the system to express each endogenous variable as a function of all the exogenous variables and a disturbance.

In IV applications, we call the equation for Y_i “the reduced form.” The equation for D_i , which we call the **first stage**, is also a reduced-form equation.

In our simplest example, both regressions use an intercept and Z_i .

First Stage and Reduced Form

$$D_i = \pi_0 + \pi_1 Z_i + v_i \quad \text{First stage}$$

$$Y_i = \delta_0 + \delta_1 Z_i + e_i \quad \text{Reduced form}$$

For a binary instrument:

$$\pi_1 = E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0],$$

$$\delta_1 = E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0].$$

With one endogenous explanatory variable and one excluded instrument, the model is **exactly identified**.

The IV estimate is the ratio of their reduced-form and first-stage coefficients:

$$\hat{\beta}_{1,\text{Wald}} = \frac{\hat{\delta}_1}{\hat{\pi}_1}.$$

From Wald to the IV Estimator

For binary Z_i , with $P(Z_i = 1) = p$,

$$\begin{aligned}\text{Cov}(Y_i, Z_i) &= [E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]] p(1 - p), \\ \text{Cov}(D_i, Z_i) &= [E[D_i | Z_i = 1] - E[D_i | Z_i = 0]] p(1 - p).\end{aligned}$$

The common factor cancels, giving

$$\beta_1 = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(D_i, Z_i)} = \frac{\text{Cov}(Y_i, Z_i) / \text{Var}(Z_i)}{\text{Cov}(D_i, Z_i) / \text{Var}(Z_i)}.$$

This is the reduced-form slope divided by the first-stage slope. The covariance expression also applies to a continuous instrument with $\text{Cov}(Z_i, u_i) = 0$ and a nonzero first stage.

IV as a Method-of-Moments Estimator

Orthogonality concerns the covariance between the instrument and the **outcome disturbance**:

$$\begin{aligned} 0 &= \text{Cov} (Z_i, u_i) \\ &= E [(Z_i - E [Z_i]) (u_i - E [u_i])] \\ &= E [(Z_i - E [Z_i]) u_i] . \end{aligned}$$

The “useful result” lets us drop the mean from the second term because $E [Z_i - E [Z_i]] = 0$.

Now substitute $u_i = Y_i - \beta_0 - \beta_1 D_i$:

$$E [(Z_i - E [Z_i]) (Y_i - \beta_0 - \beta_1 D_i)] = 0.$$

IV by Method of Moments: The Sample Estimator

Use the **analogy principle** to replace population moments with sample moments:

$$\frac{1}{n} \sum_i (Z_i - \bar{Z}) (Y_i - \hat{\beta}_0 - \hat{\beta}_1 D_i) = 0.$$

Since $\sum_i (Z_i - \bar{Z}) = 0$, the intercept drops out. Solving gives

$$\hat{\beta}_{1,IV} = \frac{\sum_i (Z_i - \bar{Z}) (Y_i - \bar{Y})}{\sum_i (Z_i - \bar{Z}) (D_i - \bar{D})} = \frac{\widehat{\text{Cov}}(Z_i, Y_i)}{\widehat{\text{Cov}}(Z_i, D_i)}.$$

This is the **method of moments**: choose the coefficients to satisfy the sample counterpart of the population condition. With a binary instrument, it gives the Wald estimator; the formula also applies to a continuous instrument.

Using the Instrument to Predict Treatment

Our goal is to isolate variation in D_i that is uncorrelated with u_i .

The population first-stage regression gives the instrument-predicted component:

$$D_i^* = \pi_0 + \pi_1 Z_i, \quad \pi_1 = \frac{\text{Cov}(D_i, Z_i)}{\text{Var}(Z_i)}.$$

Because D_i^* is a linear function of the instrument,

$$\text{Cov}(D_i^*, u_i) = \pi_1 \text{Cov}(Z_i, u_i) = 0.$$

We estimate this component using the fitted values

$$\widehat{D}_i = \widehat{\pi}_0 + \widehat{\pi}_1 Z_i.$$

Dividing Schooling into Two Components

The first-stage regression decomposes observed schooling exactly:

$$D_i = \underbrace{\widehat{D}_i}_{\text{exogenous part}} + \underbrace{\widehat{v}_i}_{\text{endogenous part}}, \quad \widehat{v}_i = D_i - \widehat{D}_i.$$

- $\widehat{D}_i = \widehat{\pi}_0 + \widehat{\pi}_1 Z_i$: schooling predicted by the instrument. A valid instrument isolates variation that is unrelated to the unobserved determinants of earnings.
- $\widehat{v}_i$: the first-stage residual. It contains the remaining variation in schooling, including variation associated with unobserved ability.

The population counterpart is $D_i = D_i^* + v_i$. Instrument validity implies

$$\text{Cov}(D_i^*, u_i) = 0, \quad \text{Cov}(D_i, u_i) = \text{Cov}(v_i, u_i).$$

IV uses the instrument-predicted component to estimate the causal effect of schooling.

Instrumental Variables as Two-Stage Least Squares

First stage: regress D_i on an intercept and Z_i ; obtain $\widehat{D}_i$.

Second stage: regress Y_i on an intercept and $\widehat{D}_i$.

$$Y_i = \hat{\beta}_{0,2SLS} + \hat{\beta}_{1,2SLS} \widehat{D}_i + \hat{r}_i.$$

Here $\hat{r}_i$ is the residual from this regression.

With a binary instrument, $\widehat{D}_i = \overline{D}_{Z=0}$ for the first group and $\widehat{D}_i = \overline{D}_{Z=1}$ for the second. The second-stage slope is therefore the slope through the two group means:

$$\hat{\beta}_{1,2SLS} = \frac{\overline{Y}_{Z=1} - \overline{Y}_{Z=0}}{\overline{D}_{Z=1} - \overline{D}_{Z=0}} = \hat{\beta}_{1,Wald}.$$

This is the graphical argument expressed as two regressions.

Including Covariates

$$Y_i = \beta_0 + \beta_1 D_i + \sum_{k=1}^K \beta_{k+1} X_{ik} + u_i,$$
$$D_i = \pi_0 + \pi_1 Z_i + \sum_{k=1}^K \pi_{k+1} X_{ik} + v_i.$$

Here X_{ik} is covariate k for person i , and K is the number of covariates. Include the same exogenous covariates in **both stages**.

With one excluded instrument, 2SLS equals the reduced-form coefficient divided by the first-stage coefficient when the regressions use the **same observations, weights, and covariates**.

In Oregon, indicators for the number of household members on the lottery list are part of the identifying comparison.

Estimating IV in Python

```
1 import linearmodels.iv as iv
2
3 model = iv.IV2SLS.from_formula(
4     "outcome ~ 1 + [insured ~ selected]", data=df
5 )
6 result = model.fit(cov_type="robust")
```

`[insured ~ selected]` instruments insurance with lottery selection; `1` includes an intercept.

Use the IV routine's standard errors. Running two OLS regressions manually reproduces the coefficient, but the second regression's ordinary standard errors are incorrect for IV.

With a binary outcome, this is a **linear probability model**. Interpret coefficients in probability units, and multiply by 100 for percentage points.

Finite Samples and Consistency

Unbiasedness concerns the mean of an estimator across repeated samples of a given size: $E \left[\hat{\beta}_1 \right] = \beta_1$.

Consistency concerns what happens as the sample size grows: the probability that the estimate is close to the true parameter approaches one.

For one instrument, the IV estimator is

$$\hat{\beta}_{1,IV} = \beta_1 + \frac{n^{-1} \sum_i (Z_i - \bar{Z}) u_i}{n^{-1} \sum_i (Z_i - \bar{Z}) D_i}.$$

With independent random sampling and finite second moments, the numerator converges to $\text{Cov}(Z_i, u_i) = 0$, and the denominator to $\text{Cov}(Z_i, D_i) \neq 0$. Hence

$$\hat{\beta}_{1,IV} \xrightarrow{p} \beta_1.$$

IV can be consistent without being unbiased in finite samples. We return to its finite-sample behaviour when studying weak instruments.

Heterogeneous Treatment Effects

So far, β_1 has been a constant effect. Now allow

$$Y_i^1 - Y_i^0$$

to differ across people.

Health insurance might have different effects for people with different health needs or access to care.

For whom does IV estimate an average treatment effect?

The answer depends on whose treatment status the instrument changes.

Potential Treatment Statuses

- D_i^1 : treatment status if the instrument equals one.
- D_i^0 : treatment status if the instrument equals zero.

$$D_i = Z_i D_i^1 + (1 - Z_i) D_i^0.$$

The superscript on Y_i^d denotes the treatment state.

The superscript on D_i^z denotes the instrument state.

For Oregon, imagine each person's Medicaid enrolment with and without the lottery offer.

Four Responses to the Instrument

Type	D_i^0	D_i^1	Response
Never-taker	0	0	No treatment under either assignment
Complier	0	1	Treatment when offered
Always-taker	1	1	Treatment under either assignment
Defier	1	0	Treatment only without the offer

These types are defined relative to a **particular instrument**.

We observe only one potential treatment status for each person. A treated lottery winner could be an always-taker or a complier.

Assumption 1: Instrument Independence

The instrument is independent of potential outcomes and treatment statuses:

$$Z_i \perp (Y_i^0, Y_i^1, D_i^0, D_i^1) .$$

Random assignment of an offer can support this assumption.

In observational applications, we need to explain why the instrument supplies an as-good-as-random comparison, possibly conditional on covariates.

The derivation first considers unconditional independence. In Oregon, independence is conditional on household-list size.

Assumption 2: Exclusion

Let $Y_i(d, z)$ denote an outcome under treatment d and instrument value z .

Exclusion requires

$$Y_i(d, 1) = Y_i(d, 0) = Y_i^d.$$

The instrument affects the outcome **through treatment**. We can then write

$$Y_i = Y_i^0 + (Y_i^1 - Y_i^0) D_i.$$

For Oregon: could winning affect health care use or well-being through a route other than obtaining Medicaid?

Assumption 3: Monotonicity

Orient the instrument so it encourages treatment. Assume

$$D_i^1 \geq D_i^0 \quad \text{for every } i.$$

The instrument increases treatment for compliers and leaves treatment unchanged for always-takers and never-takers. **There are no defiers.**

For Oregon: could winning cause someone to lose Medicaid coverage they would otherwise have obtained?

The assumption concerns each person's response to the instrument, beyond its average first-stage effect.

Assumption 4: Relevance

The instrument must change the probability of treatment:

$$E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0] > 0.$$

Under independence and monotonicity, this difference is the proportion of **compliers**.

If the instrument changes nobody's treatment, it supplies no treatment comparison.

As in Week 1, we assume that treatment is well defined and that one person's outcome does not depend on other people's treatment assignments (SUTVA).

The LATE Theorem

For binary treatment and a binary instrument, under **independence**, **exclusion**, **monotonicity** ($D_i^1 \geq D_i^0$), and **relevance**,

$$\frac{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]} = E[Y_i^1 - Y_i^0 | D_i^1 > D_i^0] = \tau_{\text{LATE}}.$$

The Wald ratio identifies the **average treatment effect for compliers**: people whose treatment status is changed by the instrument.

Imbens and Angrist (1994). We now prove the result.

Proof of LATE (1)

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

The first equality uses **exclusion**: Z_i affects Y_i through treatment. The second uses $D_i = D_i^1$ when $Z_i = 1$ and **independence** to remove the conditioning on Z_i .

Similarly,

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

Therefore the numerator of the Wald ratio is

$$E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0] = E[(Y_i^1 - Y_i^0) (D_i^1 - D_i^0)].$$

Proof of LATE (2)

Monotonicity implies that $D_i^1 - D_i^0$ can equal **zero or one**. It excludes -1 as a possible value.

- For compliers, $D_i^1 - D_i^0 = 1$.
- For always-takers and never-takers, $D_i^1 - D_i^0 = 0$.

The product is zero for everyone except compliers. Its population mean is therefore the mean effect for compliers multiplied by their population share:

$$\begin{aligned} E \left[(Y_i^1 - Y_i^0) (D_i^1 - D_i^0) \right] \\ = E \left[Y_i^1 - Y_i^0 \mid D_i^1 > D_i^0 \right] P(D_i^1 > D_i^0) . \end{aligned}$$

The outcome difference between the instrument groups equals the **average effect for compliers multiplied by their share of the population**.

Proof of LATE (3)

Now consider the denominator of the Wald ratio:

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

The first equality follows from **independence**.

The second follows from **monotonicity**: the change in treatment equals one for compliers and zero for everyone else.

The first stage is therefore the **proportion of compliers**. **Relevance** ensures this proportion is positive, so we can divide by it.

Proof of LATE (4)

Putting the numerator and denominator together,

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

The complier share cancels. The Wald ratio identifies the **local average treatment effect**.

The LATE Theorem: Proved

We have established the following result.

For binary treatment and a binary instrument, under **independence**, **exclusion**, **monotonicity** ($D_i^1 \geq D_i^0$), and **relevance**,

$$\frac{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]} = E[Y_i^1 - Y_i^0 | D_i^1 > D_i^0] = \tau_{\text{LATE}}.$$

The Wald ratio identifies the **average treatment effect for compliers**: people whose treatment status is changed by the instrument.

“Local” refers to this group of compliers.

With conditional random assignment, the argument applies within covariate groups; 2SLS with group indicators combines those comparisons. The notes explain the weighting.

LATE, ATT, and External Validity

The **ATT** averages effects among those observed receiving treatment. Under monotonicity, recipients can include always-takers and compliers assigned the offer.

The **LATE** averages effects among compliers.

When there are no always-takers, random assignment and the LATE assumptions also allow IV to identify the ATT.

Different valid instruments can change treatment for different people. To generalise an IV result, consider who responds to the instrument and whether their effects are informative for the policy population.

The Oregon Health Insurance Experiment

For discussion

- Read Sections II and III.
- Explain the policy problem, the waiting list, the lottery, and what winning allowed a household to do.
- Distinguish lottery assignment from obtaining Medicaid coverage.

The Lottery and Medicaid Coverage

In 2008, Oregon used a lottery to offer people on a waiting list the opportunity to apply for **OHP Standard**.

Selected people still had to apply and meet eligibility requirements. Some enrolled; others did not. Some people in the control group obtained Medicaid through other routes.

The analysis sample contains **74,922 individuals**, including **29,834 selected by the lottery**.

The offer was assigned through the lottery. Medicaid enrolment reflects both the offer and people's circumstances and choices.

Data

For discussion

- Read Section III.
- Identify the administrative and survey data and the outcomes each measures.
- Why do the analysis samples differ?
- Explain survey nonresponse and why it matters for the comparison.

Data Sources and Analysis Samples

- **Lottery list:** assignment, household-list size, and baseline characteristics.
- **Administrative records:** Medicaid enrolment, hospital use, credit reports, and mortality.
- **Mail survey:** health care use, financial strain, and self-reported health.

The effective weighted survey response rate was approximately **50%**. The survey used intensive follow-up of a subsample of nonrespondents.

Different sources measure different outcomes and yield different analysis samples. Read the sample definitions and table notes.

Balance: Table II

For discussion

TABLE II
TREATMENT: CONTROL BALANCE

	Control mean (std. dev.) for full sample	Difference between treatment and control		
		Full sample (2)	Credit report subsample (3)	Survey respondents subsample (4)
	(1)			
Panel A: Match/response rates				
Matched in September 2009 credit data	0.663 (0.473)		−0.0043 (0.0037) [0.247]	
Responded to survey	0.506 (0.500)			−0.016 (0.0066) [0.014]
Response time (in days)	53.0 (57.8)			1.638 (1.088) [0.132]
Panel B: Prerandomization characteristics				
Lottery list variables				
<i>F</i> -statistic		1.286	0.553	0.574
[<i>p</i> -value]		[0.239]	[0.836]	[0.820]
Pre-randomization outcomes				
<i>F</i> -statistic		0.543	0.921	1.266
[<i>p</i> -value]		[0.844]	[0.518]	[0.281]
Both of the above				
<i>F</i> -statistic		0.915	0.793	0.782
[<i>p</i> -value]		[0.56]	[0.726]	[0.680]
<i>N</i>		74,922	49,980	23,741

First Stage: Table III

For discussion

TABLE III
FIRST-STAGE ESTIMATES

	Full sample		Credit report subsample		Survey respondents	
	Control mean (1)	Estimated FS (2)	Control mean (3)	Estimated FS (4)	Control mean (5)	Estimated FS (6)
(1) Ever on Medicaid	0.141	0.256 (0.0035)	0.135	0.255 (0.0042)	0.135	0.290 (0.0066)
(2) Ever on OHP Standard	0.027	0.264 (0.0029)	0.028	0.264 (0.0036)	0.026	0.302 (0.0055)
(3) # of months on Medicaid	1.408	3.355 (0.045)	1.352	3.366 (0.055)	1.509	3.943 (0.090)
(4) On Medicaid, end of study period	0.106	0.148 (0.0031)	0.101	0.151 (0.0038)	0.105	0.189 (0.0061)
(5) Currently have any insurance (self-report)					0.325	0.179 (0.0077)
(6) Currenty have private insurance (self-report)					0.128	−0.0076 (0.0053)
(7) Currently on Medicaid (self-report)					0.117	0.197 (0.0063)
(8) Currently on Medicaid					0.105	0.191 (0.0060)
(9) Ever on TANF	0.031	0.0011 (0.0013)	0.028	0.0021 (0.0016)	0.023	0.0019 (0.0025)
(10) TANF benefits (\$)	124	−1.659 (5.813)	111	1.543 (6.571)	100	−4.991 (10.884)
(11) Ever on food stamps	0.606	0.017 (0.0029)	0.594	0.018 (0.0035)	0.622	0.023 (0.0054)
(12) Food stamp benefits (\$)	1776	61.3 (15.0)	1787	60.0 (18.8)	2202	122.4 (33.4)
<i>N</i>		74,922		49,980		23,741

Estimation in the Study

Lottery selection instruments **Medicaid coverage**.

- Include indicators for the number of household members on the lottery list: a household could apply if any listed member won.
- For survey outcomes, include survey-wave indicators and their interactions with household-list size, and use the survey weights.
- **Cluster standard errors by household:** allow heteroskedasticity and correlation among household members.

Use the same sample, controls, and weights for the reduced form and first stage when calculating their ratio.

Hospital Use: Table IV, Panel A

For discussion

TABLE IV
HOSPITAL UTILIZATION

	Control mean (1)	ITT (2)	LATE (3)	<i>p</i> -values (4)
Panel A: Extensive margin				
All hospital admissions	0.067 (0.250)	0.0054 (0.0019)	0.021 (0.0074)	[0.004]
Admissions through ER	0.048 (0.214)	0.0018 (0.0016)	0.0070 (0.0062)	[0.265]
Admissions not through ER	0.029 (0.167)	0.0041 (0.0013)	0.016 (0.0051)	[0.002]

Health Care Use: Table V

For discussion

TABLE V
HEALTH CARE UTILIZATION (SURVEY DATA)

	Extensive margin (any)				Total utilization (number)			
	Control mean (1)	ITT (2)	LATE (3)	<i>p</i> -values (4)	Control mean (5)	ITT (6)	LATE (7)	<i>p</i> -values (8)
Prescription drugs currently	0.637 (0.481)	0.025 (0.0083)	0.088 (0.029)	[0.002] {0.005}	2.318 (2.878)	0.100 (0.051)	0.347 (0.176)	[0.049] {0.137}
Outpatient visits last six months	0.574 (0.494)	0.062 (0.0074)	0.212 (0.025)	[<0.0001] {<0.0001}	1.914 (3.087)	0.314 (0.054)	1.083 (0.182)	[<0.0001] {<0.0001}
ER visits last six months	0.261 (0.439)	0.0065 (0.0067)	0.022 (0.023)	[0.335] {0.547}	0.47 (1.037)	0.0074 (0.016)	0.026 (0.056)	[0.645] {0.643}
Inpatient hospital admissions last six months	0.072 (0.259)	0.0022 (0.0040)	0.0077 (0.014)	[0.572] {0.570}	0.097 (0.4)	0.0062 (0.0062)	0.021 (0.021)	[0.311] {0.510}
Standardized treatment effect		0.050 (0.011)	0.173 (0.036)	[<0.0001]		0.040 (0.011)	0.137 (0.038)	[0.0003]
Annual spending ^a					3,156	226 (108)	778 (371)	[0.037]

Financial Strain: Table VIII

For discussion

TABLE VIII
FINANCIAL STRAIN (SURVEY DATA)

	Control mean (1)	ITT (2)	LATE (3)	<i>p</i> -values (4)
Any out of pocket medical expenses, last six months	0.555 (0.497)	−0.058 (0.0077)	−0.200 (0.026)	[<0.0001] {<0.0001}
Owe money for medical expenses currently	0.597 (0.491)	−0.052 (0.0076)	−0.180 (0.026)	[<0.0001] {<0.0001}
Borrowed money or skipped other bills to pay medical bills, last six months	0.364 (0.481)	−0.045 (0.0073)	−0.154 (0.025)	[<0.0001] {<0.0001}
Refused treatment because of med- ical debt, last six months	0.081 (0.273)	−0.011 (0.0041)	−0.036 (0.014)	[0.01] {0.01}
Standardized treatment effect		−0.089 (0.010)	−0.305 (0.035)	[<0.0001]

Health: Table IX

For discussion

TABLE IX
HEALTH

	Control mean (1)	ITT (2)	LATE (3)	<i>p</i> -values (4)
Panel A: Administrative data				
Alive	0.992 (0.092)	0.00032 (0.00068)	0.0013 (0.0027)	[0.638]
Panel B: Survey data				
Self-reported health good/very good/excellent (not fair or poor)	0.548 (0.498)	0.039 (0.0076)	0.133 (0.026)	[<0.0001] {<0.0001}
Self-reported health not poor (fair, good, very good, or excellent)	0.86 (0.347)	0.029 (0.0051)	0.099 (0.018)	[<0.0001] {<0.0001}
Health about the same or gotten better over last six months	0.714 (0.452)	0.033 (0.0067)	0.113 (0.023)	[<0.0001] {<0.0001}
# of days physical health good, past 30 days*	21.862 (10.384)	0.381 (0.162)	1.317 (0.563)	[0.019] {0.018}
# days poor physical or mental health did not impair usual activity, past 30 days*	20.329 (10.939)	0.459 (0.175)	1.585 (0.606)	[0.009] {0.015}
# of days mental health good, past 30 days*	18.738 (11.445)	0.603 (0.184)	2.082 (0.64)	[0.001] {0.003}
Did not screen positive for depression, last two weeks	0.671 (0.470)	0.023 (0.0071)	0.078 (0.025)	[0.001] {0.003}
Standardized treatment effect		0.059 (0.011)	0.203 (0.039)	[<0.0001]

Interpreting the Results

For discussion

- Bring together Tables IV, V, VIII, and IX.
- Whose effect does IV identify?
- What conclusions about the policy are supported?

External Validity

For discussion

- Explain what you would need to know before applying the Oregon findings to a different insurance expansion.
- Consider the population, take-up, follow-up period and scale of the expansion.

Assessing the IV Assumptions

Independence: lottery assignment, conditional on household-list size.

Relevance: lottery selection substantially increases Medicaid enrolment.

Exclusion: consider other routes from lottery selection to outcomes, including changes in other programme participation and other household members' coverage.

Monotonicity: assess whether selection could reduce an individual's Medicaid coverage.

Internal validity concerns the credibility of this causal comparison. External validity concerns its applicability to other populations and policies.

Recap: The IV Assumptions

In the linear model, the instrument must satisfy **relevance** and **orthogonality**: it predicts treatment and is uncorrelated with the outcome disturbance.

For a **LATE interpretation**, we require:

- **Relevance**: the instrument changes the probability of treatment.
- **Independence**: assignment of the instrument is independent of potential outcomes and potential treatment choices, conditional on the appropriate controls.
- **Exclusion**: the instrument affects the outcome only through treatment.
- **Monotonicity**: changing the instrument moves treatment in the same direction for everyone it affects; there are no defiers.

We also maintain **SUTVA**: well-defined treatment and no interference between individuals.

Next week: what features of judge and examiner assignment make each assumption credible?

After-Class Exercise

Use the Oregon public-use data to study **outpatient visits**.

1. Estimate the first stage and the ITT with the supplied design controls and weights.
2. Calculate their ratio and verify it with 2SLS.
3. Compare the estimate with Table V; interpret the effect, uncertainty, and complier population.

The exercise supplies data-import and estimation guidance. Work in Quarto. This practice is **ungraded**.

[Exercise and Quarto setup file](#)

References

- Finkelstein, A., et al. (2012). “The Oregon Health Insurance Experiment: Evidence from the First Year.” *QJE* 127(3): 1057–1106. [Paper](#).
- Imbens, G. W., and J. D. Angrist (1994). “Identification and Estimation of Local Average Treatment Effects.” *Econometrica* 62(2): 467–475. [Paper](#).
- Angrist, J. D., and A. B. Krueger (2001). “Instrumental Variables and the Search for Identification.” *JEP* 15(4): 69–85. [Paper](#).
- Angrist and Pischke, *Mastering 'Metrics*, Chapter 3; *Mostly Harmless Econometrics*, Chapter 4.
- Huntington-Klein, *The Effect*, Chapter 19; Cunningham, *Causal Inference: The Remix*, Chapter 7.

[Reading links and full references in the notes.](#)