

Chapter 16: Instrumental Variable Estimation

Contents

1	16.1 The Three Instrumental Conditions (pp. 227-231)	2
1.1	Condition 1: Relevance	2
1.2	Condition 2: Exchangeability	2
1.3	Condition 3: Exclusion Restriction	2
2	16.2 The Usual IV Estimand (pp. 231-234)	3
2.1	IV Estimand for Binary Z and A	3
2.2	Why This Works	3
2.3	Example: Randomized Encouragement	3
3	16.3 Instrumental Variable Estimation versus Randomized Experiments (pp. 234-237)	4
3.1	Perfect Compliance	4
3.2	Imperfect Compliance	4
3.3	Compliance Types	4
3.4	Local Average Treatment Effect (LATE)	4
4	16.4 Two-Stage Least Squares Estimation (pp. 237-240)	4
4.1	2SLS Algorithm	5
4.2	Why This Works	5
4.3	Including Covariates	5
5	16.5 Instrumental Variable Estimation with Measured Confounders (pp. 240-242)	5
5.1	Two Scenarios	5
5.2	Conditional IV Estimation	6
6	16.6 Instrumental Variable Estimation versus Regression (pp. 242-244)	6
6.1	Comparison	6
6.2	When Estimates Differ	6
7	16.7 The Survivor Average Causal Effect (pp. 244-246)	7
7.1	Challenges with Survival Outcomes	7
7.2	Survivor Average Causal Effect (SACE)	7
7.3	Identification	7
8	Summary	7

This chapter introduces **instrumental variable (IV) estimation**, a method for identifying causal effects when there is **unmeasured confounding**. Unlike the methods in previous chapters, IV estimation does not rely on conditional exchangeability. Instead, it uses a special variable (the instrument) that affects treatment but not the outcome directly.

This chapter is based on Hernán and Robins (2020, chap. 16, pp. 227-246).

Key innovation: IV methods allow causal inference even with unmeasured confounding, but require strong assumptions about the instrument. The causal effect identified is often for a specific subgroup (compliers), not the entire population.

1 16.1 The Three Instrumental Conditions (pp. 227-231)

An **instrumental variable** Z must satisfy three conditions to identify causal effects.

Definition 1.1 (Instrumental Variable). A variable Z is an **instrumental variable** for the effect of A on Y if:

1. **Relevance:** Z is associated with A
2. **Exchangeability:** Z is independent of unmeasured confounders U (i.e., $Y^{a,z} \perp\!\!\!\perp Z$)
3. **Exclusion restriction:** Z affects Y only through A (i.e., $Y^{a,z} = Y^a$ for all a, z)

Where $Y^{a,z}$ denotes the potential outcome under treatment $A = a$ and instrument $Z = z$.

1.1 Condition 1: Relevance

Statement: Z is associated with A

Meaning: The instrument must actually affect treatment assignment.

Example: Randomized encouragement

- $Z = 1$ if encouraged to take treatment, $Z = 0$ if not encouraged
- Relevance requires that encouragement increases the probability of treatment

Testing: Relevance can be tested empirically by checking $\Pr[A = 1 | Z = 1] \neq \Pr[A = 1 | Z = 0]$

Weak instruments: If the association between Z and A is very weak, the IV estimator will have:

- Large variance (imprecise estimates)
- Potential bias even in large samples

Common rule of thumb: F-statistic > 10 in first-stage regression (for continuous outcomes).

1.2 Condition 2: Exchangeability

Statement: $Y^{a,z} \perp\!\!\!\perp Z$ (or $Y^a \perp\!\!\!\perp Z$ under exclusion)

Meaning: The instrument is “as good as randomly assigned” with respect to potential outcomes.

Example: Randomized encouragement

- If encouragement is randomized, exchangeability holds by design
- $Z \perp\!\!\!\perp U$ where U are unmeasured confounders

Testing: Exchangeability generally **cannot** be tested (involves unmeasured confounders)

1.3 Condition 3: Exclusion Restriction

Statement: $Y^{a,z} = Y^a$ for all a, z

Meaning: The instrument affects the outcome **ONLY** through its effect on treatment.

Example: Randomized encouragement

- Encouragement affects outcome only by changing treatment received
- NOT through psychological effects, information effects, etc.

Testing: Exclusion generally **cannot** be tested (untestable assumption)

Common violations:

1. **Direct effects:** Instrument affects outcome through pathways other than treatment
2. **Defiers:** Some individuals do the opposite of what the instrument suggests (see Section 16.3)

Critical importance: Exclusion is the most controversial IV assumption. Subject-matter knowledge is essential for justifying it.

2 16.2 The Usual IV Estimand (pp. 231-234)

Under the three IV conditions, we can identify a causal effect.

2.1 IV Estimand for Binary Z and A

Setting: Binary instrument Z , binary treatment A , outcome Y

IV estimand:

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

This is the **Wald estimator** or **ratio estimator**.

Interpretation: The effect of Z on Y , divided by the effect of Z on A .

2.2 Why This Works

Numerator: $E[Y | Z = 1] - E[Y | Z = 0]$

- By exchangeability: equals $E[Y^{z=1}] - E[Y^{z=0}]$
- By exclusion: equals $E[Y^{A^{z=1}}] - E[Y^{A^{z=0}}]$

Denominator: $E[A | Z = 1] - E[A | Z = 0]$

- By exchangeability: equals $\Pr[A^{z=1} = 1] - \Pr[A^{z=0} = 1]$

Ratio: Under additional assumptions (see next section), this estimates the average causal effect in a specific subgroup.

Identification without conditional exchangeability:

The key insight: Even if A is confounded (unmeasured U affects both A and Y), we can use Z to identify causal effects because:

1. Z is randomized (or acts as if randomized)
2. Z affects Y only through A

This allows us to isolate the causal pathway $A \rightarrow Y$ without measuring all confounders.

2.3 Example: Randomized Encouragement

Design: Randomize individuals to receive encouragement to exercise (Z)

- Not everyone encouraged will exercise ($A = 1$)
- Some not encouraged will exercise anyway

IV estimate:

$$\frac{\text{Mean health in encouraged} - \text{Mean health in not encouraged}}{\Pr[\text{Exercise} | \text{Encouraged}] - \Pr[\text{Exercise} | \text{Not encouraged}]}$$

If 60% exercise when encouraged vs 30% when not, and mean health differs by 6 points:

$$\frac{6}{0.60 - 0.30} = \frac{6}{0.30} = 20$$

Effect of exercise on health (in a subgroup) is 20 points.

3 16.3 Instrumental Variable Estimation versus Randomized Experiments (pp. 234-237)

IV estimation is like an imperfect randomized experiment.

3.1 Perfect Compliance

If everyone complied with their assigned treatment ($A = Z$):

- IV estimand = average causal effect in the full population
- This is just a standard randomized experiment

3.2 Imperfect Compliance

When $A \neq Z$ for some individuals:

- IV estimand estimates effect in a **subgroup** (compliers)
- Not the average effect in the full population

3.3 Compliance Types

Definition 3.1 (Principal Strata). Individuals can be classified into **principal strata** based on potential treatments $A^{z=1}$ and $A^{z=0}$:

1. **Compliers:** $A^{z=1} = 1, A^{z=0} = 0$ (take treatment if and only if $Z = 1$)
2. **Always-takers:** $A^{z=1} = 1, A^{z=0} = 1$ (always take treatment)
3. **Never-takers:** $A^{z=1} = 0, A^{z=0} = 0$ (never take treatment)
4. **Defiers:** $A^{z=1} = 0, A^{z=0} = 1$ (do opposite of instrument)

Monotonicity assumption: No defiers exist.

Why monotonicity:

If defiers exist, the IV estimand can be severely biased. With monotonicity:

- Compliers: Instrument changes their treatment
- Always-takers: Get treatment regardless
- Never-takers: Never get treatment regardless

The IV estimand identifies the causal effect in **compliers only** (LATE = local average treatment effect).

3.4 Local Average Treatment Effect (LATE)

Under IV conditions plus monotonicity:

$$\text{IV estimand} = E[Y^{a=1} - Y^{a=0} \mid \text{Complier}]$$

This is the **average causal effect in compliers**, not in the full population.

Interpretation: IV tells us the effect of treatment for those who would comply with the instrument.

Limitation: We don't know who the compliers are (unobservable principal stratum).

4 16.4 Two-Stage Least Squares Estimation (pp. 237-240)

Two-stage least squares (2SLS) is the most common IV method for continuous outcomes.

4.1 2SLS Algorithm

Stage 1: Regress treatment on instrument (and covariates if present)

$$A_i = \alpha_0 + \alpha_1 Z_i + \epsilon_i$$

Obtain predicted treatment: $\hat{A}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i$

Stage 2: Regress outcome on predicted treatment

$$Y_i = \beta_0 + \beta_1 \hat{A}_i + \eta_i$$

The coefficient $\hat{\beta}_1$ is the 2SLS estimate of the causal effect.

4.2 Why This Works

Intuition:

- Stage 1 extracts the variation in A that is “caused by” Z
- \hat{A} is the part of A that is free of confounding (because Z is randomized)
- Stage 2 estimates the effect of this “clean” variation on Y

Mathematical equivalence: For binary Z and A , 2SLS equals the Wald estimator.

Important:

- Standard errors from Stage 2 are WRONG (they ignore Stage 1 uncertainty)
- Use specialized IV software or bootstrap for correct standard errors
- R packages: `ivreg`, `AER::ivreg`
- Stata command: `ivregress 2sls`

4.3 Including Covariates

With measured confounders L that confound $A \rightarrow Y$ but not $Z \rightarrow A$:

Stage 1:

$$A_i = \alpha_0 + \alpha_1 Z_i + \alpha_2^\top L_i + \epsilon_i$$

Stage 2:

$$Y_i = \beta_0 + \beta_1 \hat{A}_i + \beta_2^\top L_i + \eta_i$$

Including L can improve efficiency even if not necessary for identification.

5 16.5 Instrumental Variable Estimation with Measured Confounders (pp. 240-242)

IV estimation can be combined with adjustment for measured confounders.

5.1 Two Scenarios

Scenario 1: Confounders of $A \rightarrow Y$ that don't affect Z

- Adjust by including L in both stages of 2SLS
- Improves efficiency but not necessary for identification

Scenario 2: Confounders of $Z \rightarrow Y$

- More problematic - threatens IV exchangeability assumption
- Need $Y^a \perp\!\!\!\perp Z \mid L$ (conditional exchangeability of instrument)
- Use conditional IV methods

5.2 Conditional IV Estimation

Modified IV conditions:

1. **Conditional relevance:** $Z \not\perp\!\!\!\perp A \mid L$
2. **Conditional exchangeability:** $Y^a \perp\!\!\!\perp Z \mid L$
3. **Exclusion restriction:** $Y^{a,z} = Y^a$ (still unconditional)

Estimation: Use 2SLS with L as covariates, then standardize over L .

When to adjust for L :

- **Include** if L confounds $A \rightarrow Y$ (for efficiency)
- **Include** if L confounds $Z \rightarrow Y$ (for identification)
- **Don't include** if L is affected by Z (would bias estimates)

Subject-matter knowledge is crucial for deciding which covariates to include.

6 16.6 Instrumental Variable Estimation versus Regression (pp. 242-244)

How do IV estimates compare to regression-based estimates?

6.1 Comparison

Regression (e.g., outcome regression or IP weighting):

- Assumes no unmeasured confounding: $Y^a \perp\!\!\!\perp A \mid L$
- Identifies average causal effect: $E[Y^{a=1}] - E[Y^{a=0}]$
- Efficient when assumptions hold

IV estimation:

- Allows unmeasured confounding of $A \rightarrow Y$
- Assumes valid instrument with IV conditions
- Identifies LATE: $E[Y^{a=1} - Y^{a=0} \mid \text{Complier}]$
- Less efficient (larger standard errors)

6.2 When Estimates Differ

If IV and regression give different estimates:

1. **Unmeasured confounding:** Regression is biased, IV may be valid
2. **Effect heterogeneity:** IV estimates LATE, regression estimates ATE
3. **IV violations:** IV assumptions may be violated
4. **Both wrong:** Both methods could have issues

Interpretation: Differences suggest either unmeasured confounding or effect heterogeneity (or both).

Practical dilemma:

- If estimates agree: Reassuring (though both could be wrong)
- If estimates disagree: Difficult to know which to trust

Recommendation:

1. Carefully assess IV assumptions (especially exclusion)
2. Consider who the compliers are and whether LATE is the estimand of interest
3. Conduct sensitivity analyses
4. Report both estimates with clear statements about assumptions

7 16.7 The Survivor Average Causal Effect (pp. 244-246)

IV methods can be extended to handle **survival outcomes** and **time-to-event data**.

7.1 Challenges with Survival Outcomes

Issue: With time-to-event outcomes, some individuals are censored before the event.

Question: How do we interpret IV estimates when the outcome is survival time?

7.2 Survivor Average Causal Effect (SACE)

Definition 7.1 (Survivor Average Causal Effect). The **survivor average causal effect (SACE)** is:

$$E[Y^{a=1} - Y^{a=0} \mid S^{a=1} = 1, S^{a=0} = 1]$$

where S^a is an indicator for surviving (or remaining uncensored) under treatment a .

This is the effect in **always-survivors** - those who would survive under both treatment and control.

7.3 Identification

Setting: Survival S is affected by treatment A , and outcome Y is only observed if $S = 1$.

IV approach: Under IV conditions plus additional assumptions (monotonicity for survival), IV can identify SACE.

Interpretation: Effect of treatment on the outcome for those who would survive regardless of treatment.

SACE vs LATE:

- **LATE:** Effect in compliers (those whose treatment is affected by instrument)
- **SACE:** Effect in always-survivors (those who survive under both treatments)

Both are **principal stratification** approaches - effects in specific latent subgroups.

Challenge: SACE requires strong assumptions:

1. Monotonicity for survival (treatment doesn't hurt anyone's survival)
2. Exclusion restriction for both survival and outcome

These are often hard to justify.

8 Summary

Key concepts:

1. **Instrumental variable:** A variable Z that affects treatment but not the outcome directly
2. **IV conditions:** Relevance, exchangeability, exclusion restriction
3. **Wald estimator:** $\frac{E[Y|Z=1] - E[Y|Z=0]}{E[A|Z=1] - E[A|Z=0]}$ for binary Z, A
4. **Principal strata:** Compliers, always-takers, never-takers, defiers
5. **LATE:** Local average treatment effect in compliers
6. **2SLS:** Two-stage least squares for continuous outcomes
7. **SACE:** Survivor average causal effect for survival outcomes

When to use IV methods:

- **Unmeasured confounding** is a concern

- **Valid instrument** is available (strong assumptions)
- **LATE** is a meaningful estimand (compliers are of interest)
- **Efficiency loss** is acceptable (IV estimates have larger SEs)

Common instruments:

Setting	Instrument	Treatment	Outcome
Randomized encouragement	Encouragement	Behavior change	Health
Geographic variation	Distance to facility	Healthcare use	Health
Mendelian randomization	Genetic variant	Biomarker	Disease
Draft lottery	Lottery number	Military service	Earnings
Physician preference	Physician tendency	Treatment choice	Outcome

Assumptions to check:

1. **Relevance:** Test empirically (Z associated with A)
2. **Exchangeability:** Justify by design or argue plausibility
3. **Exclusion:** Requires subject-matter knowledge (cannot be tested)
4. **Monotonicity:** No defiers (often plausible, sometimes testable)

Advantages:

- Allows causal inference with unmeasured confounding
- Uses only treatment variation “caused by” instrument
- Provides a different estimand than regression methods

Limitations:

- Strong, untestable assumptions (especially exclusion)
- Estimates LATE, not ATE (interpretation challenge)
- Less efficient than regression (when regression assumptions hold)
- Weak instruments lead to bias and large variance

Practical advice:

1. **Think hard about exclusion:** Is it plausible the instrument only affects outcome through treatment?
2. **Check instrument strength:** Weak instruments are worse than no instrument
3. **Understand LATE:** Who are the compliers? Is their effect of interest?
4. **Sensitivity analyses:** Try different instruments if available
5. **Compare with regression:** Large differences suggest unmeasured confounding or heterogeneity

Looking ahead: Part III extends these methods to time-varying treatments, where IV ideas play a role in dealing with time-varying confounding and selection bias.

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. <https://miguelhernan.org/whatifbook>.