External Validity

Erick Gong (thanks to K. Ajayi)

UC Berkeley & CEGA

October 19

Agenda

- Internal Validity
- External Validity
- Case Studies

What is validity?

Two types

- Internal Validity
- External Validity

What is validity?

Two types
- Internal Validity
- External Validity

Internal Validity: Accurate estimate of the causal effect <u>within</u> the sample.

External Validity: Accurate estimate of the causal effect <u>outside</u> the sample.

Emphasis of course so far has been on internal validity

Why do we care?

External validity

Why do we care?

External validity

- Program Expansion or Scalability
- Replication
- Economic Theory
- Example: Progressa & CCT

Why do we care?

External validity

- Program Expansion or Scalability
- Replication
- Economic Theory
- Example: Progressa & CCT

Answer: We care because we want our research/findings to lead to more effective policies.

Goal: To convince policymakers, NGOs, governments to adopt effective policies.

Internal Validity

- Randomized Evaluations
  Method: Random assignment of treatment and control arm

Examples: Deworming medication, monetary incentives for HIV tests, scholarships for students

- What are we measuring?
  (ATE) Average Treatment Effects
- Randomized Evaluations have two types of effects

2 Effects:

ITT (Intent to Treat) = People made eligible for treatment / intervention

TOT (Treatment on the Treated) = People who actually took the treatment / intervention

Policy Implications?

Internal Validity

- Regression Discontinuity
  Compare two groups on each side of the threshold

Internal Validity

- Regression Discontinuity
  Compare two groups on each side of the threshold

Examples: Effect of winning scholarship, class size
Effect: LATE (Local Average Treatment Effect)
Policy Implication?

Internal Validity

Using methods we can measure

- Average Treatment Effect (ATE): Effect of treatment on randomly selected person
- Intent to Treat Effect (ITT): Effect of being made <u>eligible</u> for treatment
- Treatment on Treated Effect (TOT): Effect of <u>taking</u> treatment
- Local Average Treatment Effect (LATE): Effect of treatment for population near threshold

All our <u>causal</u> effects.

External Validity

- Can we now recommend intervention to other areas?
- Example: Scholarship Program in Kenya (Busia & Teso)
  Scholarships awarded to highest scoring 15% of girls
  Overall scholarship program has positive effect on test scores
- Policy implication: scholarships nationwide?

External Validity

- Can we now recommend intervention to other areas?
- Example: Scholarship Program in Kenya (Busia & Teso)
  Scholarships awarded to highest scoring 15% of girls
  Overall scholarship program has positive effect on test scores
- Policy implication: scholarships nationwide?

Wait! How would scholarship program affect these other groups:
1) Girls in urban schools
2) Girls in high income areas (Nairobi suburbs)
3) Girls who had dropped out of school
4) First graders in sample region (not just 6th graders)
5) Boys in 6th grade
6) Girls in Uganda

External Validity

- Can we now recommend intervention to other areas?
- Example: Scholarship Program in Kenya (Busia & Teso)
  Scholarships awarded to highest scoring 15% of girls
  Overall scholarship program has positive effect on test scores
- Policy implication: scholarships nationwide?

Wait! How would scholarship program affect these other groups:
1) Girls in urban schools
2) Girls in high income areas (Nairobi suburbs)
3) Girls who had dropped out of school
4) First graders in sample region (not just 6th graders)
5) Boys in 6th grade
6) Girls in Uganda
Heterogeneity in treatment effects.

External Validity

- Can we now recommend intervention to other areas?
- Example: Scholarship Program in Kenya (Busia & Teso)
  Scholarships awarded to highest scoring 15% of girls
  Overall scholarship program has positive effect on test scores
- Policy implication: scholarships nationwide?

Wait! How would scholarship program affect these other groups:
1) Girls in urban schools
2) Girls in high income areas (Nairobi suburbs)
3) Girls who had dropped out of school
4) First graders in sample region (not just 6th graders)
5) Boys in 6th grade
6) Girls in Uganda
Heterogeneity in treatment effects.
Goal: Want intervention to have an effect - OR - no reason to
motivate people who will already do something.

Population Concepts

- What is the population of interest? or Who do we care about?
  Example: Cash rewards in this class
  Example: Scholarship Program in Kenya

Population Concepts

- What is the population of interest? or Who do we care about?
  Example: Cash rewards in this class
  Example: Scholarship Program in Kenya

1. All primary school students
2. Female/ Male Students
3. Urban / Rural Areas
4. Students from disadvantaged families or backgrounds

Population Concepts

- What is the population of interest? or Who do we care about?
  Example: Cash rewards in this class
  Example: Scholarship Program in Kenya

1. All primary school students
2. Female/ Male Students
3. Urban / Rural Areas
4. Students from disadvantaged families or backgrounds

- Helps sometimes to think about why we are conducting an intervention.
  Usually in development it is to help assist the poor out of poverty.
- Clarify who we are interested in and make sure it is an interesting population to study

Samples

- In an evaluation / study we usually analyze samples
- Define population of interest & then take sample

    Population > Sample > Treatment Group

- Treatment Group -> Sample
- Sample-> Population

Samples

- In an evaluation / study we usually analyze samples
- Define population of interest & then take sample

> Population > Sample > Treatment Group

- Treatment Group -> Sample
- Sample-> Population

Concern: How was sample selected?
Example: Scholarship program for young girls
Population of Interest: School girls from poor families
Selection of Sample: Flyers and radio ads asking students to sign up for program

Samples

- In an evaluation / study we usually analyze samples
- Define population of interest & then take sample

Population > Sample > Treatment Group

- Treatment Group -> Sample
- Sample-> Population

Concern: How was sample selected?
Example: Scholarship program for young girls
Population of Interest: School girls from poor families
Selection of Sample: Flyers and radio ads asking students to sign up for program
Problem: **Selection Problem** - when people in the sample are different from our population of interest in some unobserved way (i.e. motivation)

External Validity

Two Concerns about population / sample

- **Selection Problem**: Is the way the sample was chosen representative of the population of interest?
- **Heterogeneity**: Is the treatment effect going to be vary?

Example: CCT for safe sexual behavior to reduce HIV transmission

External Validity

Two Concerns about population / sample

- **Selection Problem**: Is the way the sample was chosen representative of the population of interest?

- **Heterogeneity**: Is the treatment effect going to be vary?

Example: CCT for safe sexual behavior to reduce HIV transmission
Population of Interest: Sexually Active Young People
Selection: Schools? Homes? Bars?
Heterogeneity in treatment: Female / Male

How does it work?

- In randomized evaluations, we have evidence IF intervention has an effect, but we don't know WHY
- Not just economics . . . medical evaluations (vaccinations)

How does it work?

Conditional Cash Transfers (CCT)

- Treatment: Conditioned cash payments on school enrollment
- How does it work? Is it because students adhering to the conditionality?
  OR - Is there an income effect? Would unconditional cash payments do the same?

How does it work?

Conditional Cash Transfers (CCT)

- Treatment: Conditioned cash payments on school enrollment
- How does it work? Is it because students adhering to the conditionality?
  OR - Is there an income effect? Would unconditional cash payments do the same?

Mechanism / Channel: How is the intervention actually working?

How does it work?

Conditional Cash Transfers (CCT)

- Treatment: Conditioned cash payments on school enrollment
- How does it work? Is it because students adhering to the conditionality?
  OR - Is there an income effect? Would unconditional cash payments do the same?

Mechanism / Channel: How is the intervention actually working?
Why is this important?
We can improve program if we know specifically how an intervention is working.
Example: CCTs can be very expensive to monitor.

General Equilibrium

- Most evaluations assume <u>partial equilibrium</u>
  Hypothetical Example: Vocational training program
  Research Design: Sample in rural Kenya in 10 villages
  Treatment: Training to become a skilled laborer (i.e. plumber, mechanic, electrician)
  Effect: People in program earn more money (+40% gain in wages)
- What if vocational training rolled out nationwide?

General Equilibrium

- Most evaluations assume <u>partial equilibrium</u>
  Hypothetical Example: Vocational training program
  Research Design: Sample in rural Kenya in 10 villages
  Treatment: Training to become a skilled laborer (i.e. plumber, mechanic, electrician)
  Effect: People in program earn more money (+40% gain in wages)
- What if vocational training rolled out nationwide?

As more people became skilled laborers, wages will go down, leading to decrease in wage gains

General Equilibrium

- Most evaluations assume <u>partial equilibrium</u>
  Hypothetical Example: Vocational training program
  Research Design: Sample in rural Kenya in 10 villages
  Treatment: Training to become a skilled laborer (i.e. plumber, mechanic, electrician)
  Effect: People in program earn more money (+40% gain in wages)
- What if vocational training rolled out nationwide?

As more people became skilled laborers, wages will go down, leading to decrease in wage gains

- General Equilibrium Effects

Summary

- Using impact evaluation method for internal validity
  => find a causal effect (average treatment, intent to treat)
- External Validity
  1) What is population of interest?
  2) How are we selecting sample?
  3) Heterogeneity in treatment effects?
  4) How is the treatment actually working (the mechanism)?
  5) Are there general equilibrium effects?
- Replications where careful sampling of population of interest using theory.
- Example: Cash Transfers (Income vs. Conditionality Effect)

Intermission

Stretch, take a break, think about external validity

Corruption

- How can we reduce corruption?
- Community Based Monitoring
  Idea: People using the services are best ones to monitor
- Three studies that look at this

Uganda Healthcare

- Bjorkman & Svensson (2007)
  Does community monitoring increase the quality and quantity of health service provision?
- Randomized Evaluation:
  Treatment: Report cards on health clinics (service delivery), encouraged community to develop own monitoring methods.
  Outcomes: Health outcomes for children, utilization of health clinics, service quality: waiting time, attendance of health care workers
- What effect is it?
  What is population of interest?
  Heterogeneous Treatment Effects?
  How is the treatment actually working?

Uganda Healthcare

- Bjorkman & Svensson (2007)
  Does community monitoring increase the quality and quantity
  of health service provision?
- Randomized Evaluation:
  Treatment: Report cards on health clinics (service delivery),
  encouraged community to develop own monitoring methods.
  Outcomes: Health outcomes for children, utilization of health
  clinics, service quality: waiting time, attendance of health care
  workers
- What effect is it?
  What is population of interest?
  Heterogeneous Treatment Effects?
  How is the treatment actually working?

Number of different ways: 1) Monitoring and enforcing higher
quality, 2) Awareness of health issues, 3) others?
Unclear on the mechanism that this is working through

Uganda Healthcare

- Bjorkman & Svensson (2007)
  Does community monitoring increase the quality and quantity of health service provision?
- Randomized Evaluation:
  Treatment: Report cards on health clinics (service delivery), encouraged community to develop own monitoring methods.
  Outcomes: Health outcomes for children, utilization of health clinics, service quality: waiting time, attendance of health care workers
- What effect is it?
  What is population of interest?
  Heterogeneous Treatment Effects?
  How is the treatment actually working?

Number of different ways: 1) Monitoring and enforcing higher quality, 2) Awareness of health issues, 3) others?
Unclear on the mechanism that this is working through
Should we recommend this program to other areas?

Brazil & Government Audits

- Ferraz & Finan (2008)
  Question: Do voters punish/reward politicians for corrupt practices?
- Randomization
  Treatment: Municipalities randomly audited
  Outcomes: Electoral performance of incumbent mayors
- How is treatment working?

Brazil & Government Audits

- Ferraz & Finan (2008)
  Question: Do voters punish/reward politicians for corrupt practices?
- Randomization
  Treatment: Municipalities randomly audited
  Outcomes: Electoral performance of incumbent mayors
- How is treatment working?

The channel might be informing the electorate through media (local radio)

Brazil & Government Audits

- Ferraz & Finan (2008)
  Question: Do voters punish/reward politicians for corrupt practices?
- Randomization
  Treatment: Municipalities randomly audited
  Outcomes: Electoral performance of incumbent mayors
- How is treatment working?

The channel might be informing the electorate through media (local radio)
Should we recommend this to other areas?

Brazil & Government Audits

- Ferraz & Finan (2008)
  Question: Do voters punish/reward politicians for corrupt practices?
- Randomization
  Treatment: Municipalities randomly audited
  Outcomes: Electoral performance of incumbent mayors
- How is treatment working?

The channel might be informing the electorate through media (local radio)
Should we recommend this to other areas?

- similar levels of medical coverage
- similar levels of civic engagement (voting mandatory in Brazil)
- similar levels of corruption

Indonesia Road Construction

- Olken (2007)
  Question: Effects of audits and monitoring on corruption in
  road construction?
- Randomization
  Treatment: Audit of construction project (top down)
  Invitations to increase participation at meetings, comment box
  (bottom up)
  Outcome: Core samples of road to measure corruption
- Channel: Top down monitoring more effective then community
  based

Summary

- Community Based Monitoring May Work in Different contexts
- Evidence in support (Uganda Health Clinics, Electoral Outcomes in Brazil)
- Evidence not in favor (Indonesia Road Construction)
- Understand differences, do additional evaluations
- What policies are you comfortable with now to reduce corruption?

Research Proposals

- Worth spending 20 to 30 minutes over next week thinking about it
- Group Formation
- Questions?