# Impact Evaluation: Quasi-Experimental Methods

Focus on Instrumental Variables

Christopher Udry
Yale University

# What is Impact Evaluation?

- IE assesses how a program affects the well-being or welfare of individuals, households or communities (or businesses). In agriculture, we may care specifically about profits

- Well-being at the individual level can be captured by income, consumption, or broader welfare measures

- At the community level, poverty levels or growth rates may be appropriate, depending on the question

# Outline

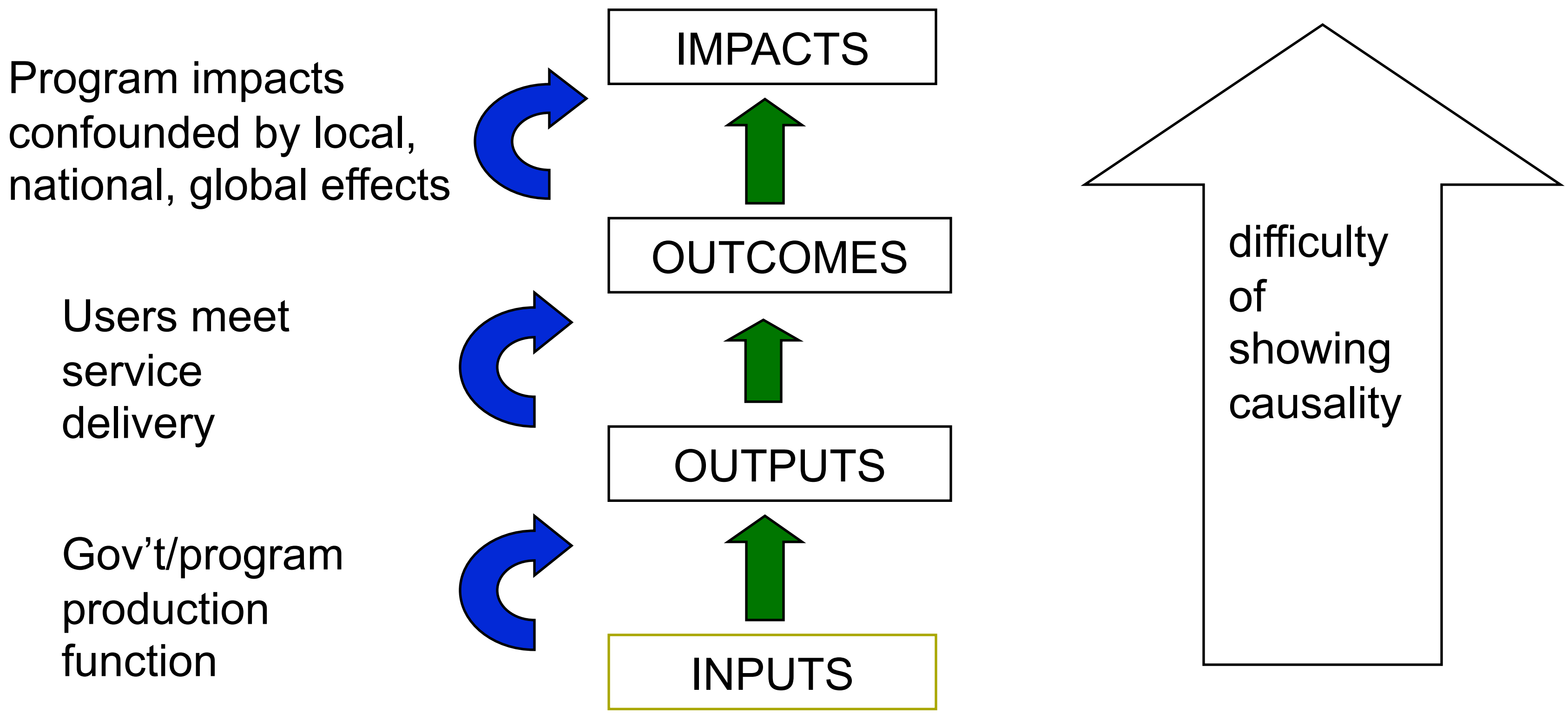
- Advantages of Impact Evaluation
- Challenges for IE: Need for a Counterfactual
- Methods for Constructing Comparison Groups

# IE Versus other M&E Tools

- The key distinction between impact evaluation and other M&E tools is the focus on discerning the impact of the program from all other confounding effects

- IE seeks to provide evidence of the causal link between an intervention and outcomes

# Monitoring and IE



# Logic Model: An Example

- Consider a program of providing advice on a new technology to farmers
- What are:
  - Inputs?
  - Outputs?
  - Outcomes?
  - Impacts?

# Logic Model: An Example

- *Inputs*: visits by extension agents, perhaps physical inputs like seeds
- *Outputs*: knowledge of the new technology
- *Outcomes*: use of the new technology
- *Impact*: profit change, consumption change

# Advantages of IE

- In order to be able to determine which projects are successful, need a carefully designed impact evaluation strategy
- This is useful for:
  - Understanding if projects worked:
    - Justification for funding
    - Scaling up
    - Meta-analysis: Learning from Others
  - Cost-benefit tradeoffs across projects
  - Can test between different approaches of same program or different projects to meet national indicator

# Essential Methodology

- Difficulty is determining what would have happened to the individuals or communities of interest in absence of the project

- The key component to an impact evaluation is to construct a suitable comparison group to proxy for the "counterfactual"

- Problem: can only observe people in one state of the world at one time

# Before/After Comparisons

- Why not collect data on individuals before and after intervention (the Reflexive)? Difference in income, etc, would be due to project

- Problem: many things change over time, including the project
  - The country is growing and profits are rising.  Is this due to the program or would have occurred in absence of program?
  - Many factors affect yield in a given year

# Comparison Groups

- Instead of using before/after comparisons, we need to use comparison groups to proxy for the counterfactual
- Two Core Problems in Finding Suitable Groups:
  - Programs are targeted
    - Recipients receive intervention for particular reason
  - Participation is voluntary
    - Individuals who participate differ in observable and unobservable ways (selection bias)
- Hence, a comparison of participants and an arbitrary group of non-participants can lead to misleading or incorrect results

# Counterfactual: Methodology

- We need a comparison group that is as identical in observable and unobservable dimensions as possible, to those receiving the program, and a comparison group that will not receive spillover benefits.

- Number of techniques:
  - Randomization as gold standard
  - Various Techniques of Matching

# How to construct a comparison group – building the counterfactual

1. Randomization
2. Difference-in-Difference
3. Regression discontinuity
4. Matching
   - Pipeline comparisons
   - Propensity score
5. IV

# Instrumental Variables and IE

- Instrumental variables have many uses
- IV can be generated *ex ante*:
  - Randomized promotion (or encouragement design)
  - "Randomized offering" of a program
- IV can be used ex post to correct for non-compliance or conduct retrospective IE:
  - Correction for non-compliance to recover TOT from ITT
    - E.g. Randomized Assignment with non-compliers
    - E.g. Fuzzy Regression Discontinuity
  - Look for exogenous variation to evaluate the impact of a program in absence of a prospective design.
- Here:
  - General Principles behind IVs
  - Ex ante focus on randomized promotion
  - IV, non-compliance and randomized offering

# An example to start off with…

- Let's look at the productivity of a technology, say fertilizer or fallowing
  - Any farmer is eligible (Universal eligibility)
  - Some people choose to use (Participants)
  - Other people choose not to use (Non-participants)

- Some simple (but not-so-good) ways to evaluate the program:
  - Compare before and after situation in the participant group
  - Compare situation of participants and non-participants after the intervention
  - Compare situation of participants and non-participants before and after (DD).

# Fertilizer Use

Say we decide to compare outcomes for those who participate to the outcomes of those who do not participate:

- A simple model to do this:

$$y = \alpha + \beta_1 P + \beta_2 x + \varepsilon$$

$$P = \begin{cases} 1 & \text{If farmer uses fertilizer} \\ 0 & \text{If person does not} \end{cases}$$

$x$ = Control variables (exogenous & observed)

- Why is this not working? 2 problems:
  - Variables that we omit (for various reasons) but that are important
  - Decision to participate is endogenous.

# Problem #1: Omitted Variables

- Even if we try to control for "everything", we'll miss:

  (1) Characteristics that we didn't know they mattered, and
  (2) Characteristics that are too complicated to measure
  (not observables or not observed):
    - Talent, motivation, soil fertility
    - Level of information and access to services
    - Opportunity cost of participation

- Full model would be:

$$y = \gamma_0 + \gamma_1 x + \gamma_2 P + \gamma_3 M_1 + \eta$$

But we cannot observe $M_1$ , the "missing" and unobserved variables.

# Omitted variable bias

- True model is: $y = \gamma_0 + \gamma_1 x + \gamma_2 P + \gamma_3 M_1 + \eta$

- But we estimate: $y = \beta_0 + \beta_1 x + \beta_2 P + \varepsilon$

- If there is a correlation between $M_1$ and $P$, then the *OLS* estimator of $\beta_2$ will not be a consistent estimator of $\gamma_2$, the true impact of $P$.

- Why?

  When $M_1$ is missing from the regression, the coefficient of $P$ will "pick up" some of the effect of $M_1$

  Best solution: Measure M!

# This is Exactly the Problem in Production Func. Estimates

$$y = \alpha K^{\beta} L^{\gamma} \varepsilon$$

$$MP_L = \gamma \frac{y}{L} = w$$

$L^*$ depends on $\varepsilon$

$$\ln(y) = \ln(\alpha) + \beta \ln(K) + \gamma \ln(L) + \ln(\varepsilon)$$

# Problem #2: Endogenous Decision to Participate

- True model is: $$y = \gamma_0 + \gamma_1 x + \gamma_2 P + \eta$$

  with $$P = \pi_0 + \pi_1 x + \pi_2 M_2 + \xi$$

  $M_2$ = Vector of unobserved / missing characteristics

  (i.e. we don't fully know why people decide to participate)

- Since we don't observe $M_2$ , we can only estimate a simplified model:

$$y = \beta_0 + \beta_1 x + \beta_2 P + \varepsilon$$

- Is $\beta_{2,\, OLS}$ an unbiased estimator of $\gamma_2$?

# Problem #2: Endogenous Decision to Participate

- We estimate: $$y = \beta_0 + \beta_1 x + \beta_2 P + \varepsilon$$
- But true model is: $$y = \gamma_0 + \gamma_1 x + \gamma_2 P + \eta$$
  with $$P = \pi_0 + \pi_1 x + \pi_2 M_2 + \xi$$
- Is $\beta_{2,\,OLS}$ an unbiased estimator of $\gamma_2$?

$$\begin{aligned} Corr(\varepsilon, P) &= corr(\varepsilon, \pi_0 + \pi_1 x + \pi_2 M_2 + \xi) \\ &= \pi_1 \, corr(\varepsilon, x) + \pi_2 \, corr(\varepsilon, M_2) \\ &= \pi_2 \, corr(\varepsilon, M_2) \end{aligned}$$

- If there is a correlation between the missing variables that determine participation (e.g. soil fertility) and outcomes not explained by observed characteristics, then the OLS estimator will be biased.

# What can we do to solve this problem?

- We estimate: $y = \beta_0 + \beta_1 x + \beta_2 \boldsymbol{P} + \boldsymbol{\varepsilon}$

- So the problem is the correlation between ***P*** and $\boldsymbol{\varepsilon}$

- How about we replace ***P*** with “something else”, call it ***Z***:
    - ***Z*** needs to be similar to ***P***
    - But is not correlated with $\boldsymbol{\varepsilon}$

# Back to Fertilizer

- ***P*** = participation
- $\boldsymbol{\varepsilon}$ = that part of outcomes that is not explained by program participation or by observed characteristics
- I'm looking for a variable ***Z*** that is:
  (1) Closely related to participation *P*
  (2) but doesn't directly affect people's outcomes *Y, other than through its effect on participation.*
- So this variable must be coming from outside.

# Generating an outside variable for fertilizer use

- Say that an extension officer visits farmers to encourage them to participate.
  - She only visits 50% of persons on her roster, and
  - She randomly chooses whom she will visit
- If she is effective, some she visits will use fert.

  There will be a correlation between receiving a visit and fertilizer use
- But visit does not have direct effect on outcomes (e.g. income) apart from its effect through inducing fertilizer use.
- Randomized "encouragement" or "promotion" visits are an Instrumental Variable.

# Characteristics of an instrumental variable

- Define a new variable $Z$

$$Z = \begin{cases} 1 & \text{If person was randomly chosen to receive the encouragement visit from extension worker} \\ 0 & \text{If person was randomly chosen not to receive the encouragement visit from the social worker} \end{cases}$$

- $Corr\,(Z, P) > 0$

  People who receive the encouragement visit are more likely to participate than those who don't

- $Corr\,(Z, \varepsilon) = 0$

  No correlation between receiving a visit and benefit to the program apart from the effect of the visit on participation.

- $Z$ is called an instrumental variable

# Two-stage least squares *(2SLS)*

Remember the original model with endogenous *P*:

$$y = \beta_0 + \beta_1 x + \beta_2 P + \varepsilon$$

## Step 1

Regress the endogenous variable *P* on the instrumental variable(s) *Z* and other exogenous variables

$$P = \delta_0 + \delta_1 x + \delta_2 Z + \tau$$

- Calculate the predicted value of *P* for each observation: *P*
- Since *Z* and *x* are not correlated with ε, neither will be *P*.
- You will need one instrumental variable for each potentially endogenous regressor.

# Two-stage least squares *(2SLS)*

## Step 2

Regress *y* on the predicted variable *P* and the other exogenous variables

$$y = \beta_0 + \beta_1 x + \hat{\beta}_2 P + \varepsilon$$

- Note: The standard errors of the second stage *OLS* need to be corrected because *P* is not a fixed regressor.
- In Practice: Use *STATA* ivreg command, which does the two steps at once and reports correct standard errors.
- Intuition: By using *Z* for *P*, we cleaned *P* of its correlation with $\eta$
- It can be shown that (under certain conditions) $\beta_{2,IV}$ yields a consistent estimator of $\gamma_2$ (large sample theory)

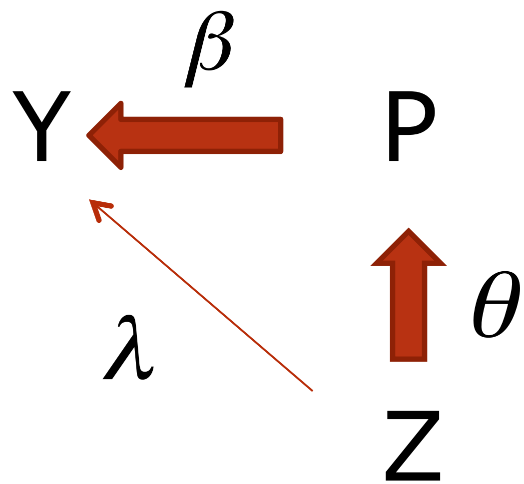
# Why does IV work?

$$\text{plim } \hat{\beta}_{IV} = \frac{\text{cov}(Y,Z)}{\text{cov}(P,Z)} = \frac{\beta\theta}{\theta} = \beta$$

$$\begin{tikzcd} Y & P \arrow[l, "\beta"'] \\ & Z \arrow[u, "\theta"'] \end{tikzcd}$$

# When is IV biased?

$$\text{plim } \hat{\beta}_{IV} = \frac{\text{cov}(Y,Z)}{\text{cov}(P,Z)} = \frac{\beta\theta + \lambda}{\theta} = \beta + \frac{\lambda}{\theta}$$



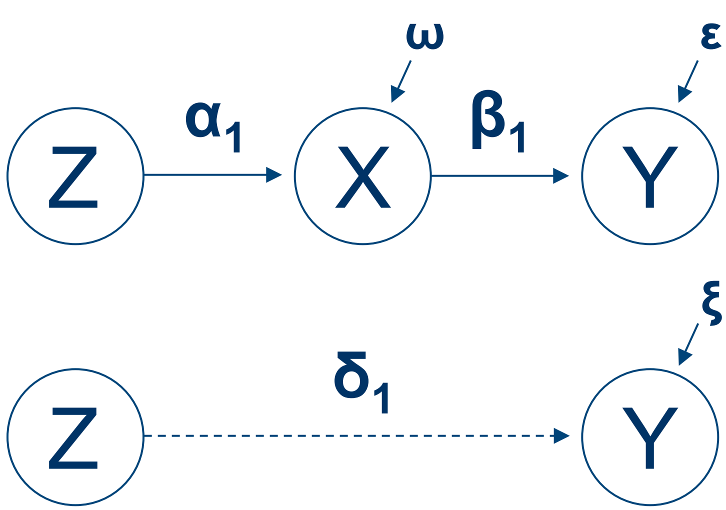
# Instrumental Variables Terminology

- Three different models to be familiar with
  - First stage: $\mathbf{X = \alpha_0 + \alpha_1 Z + \omega}$
  - Structural model: $\mathbf{Y = \beta_0 + \beta_1 X + \varepsilon}$
  - Reduced form: $\mathbf{Y = \delta_0 + \delta_1 Z + \xi}$
- An interesting equality:

  $$\delta_1 = \alpha_1 \times \beta_1$$

  so…

  $$\beta_1 = \delta_1 / \alpha_1$$



# Where do we find instrumental variables?

- Searching for an IV *ex post* …
- Generating an IV with information campaign designed *ex ante*
  - If everyone is eligible to participate in treatment
  - But some have more information than others
    (Who has more information will be more likely to participate)
  - Provision of “additional information” on a random basis

# Example: the productivity of fallowing

$$Y_{iht} = X_{iht}\alpha + F_{iht}\beta + \lambda_{ht} + \varepsilon_{iht}$$

| | 1 | |
|---|---|---|
| | OLS | |
| dependent variable | profit x1000 | |
| | estimate | std error |
| fallow duration (years)* | 163 | 48 |
| gender: 1=woman | -356 | 397 |

# Example: the productivity of fallowing

$$Y_{iht} = X_{iht}\alpha + F_{iht}\beta + \lambda_{ht} + \varepsilon_{iht}$$

$$F_{iht} = X_{iht}\gamma + Z_{iht}\delta + \upsilon_{ht} + \zeta_{iht}$$

| | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|
| | OLS | | OLS | | IV | |
| dependent variable | profit x1000 | | fallow duration | | profit x1000 cedis/hect | |
| | estimate | std error | estimate | std error | estimate | std error |
| fallow duration (years)* | 163 | 48 | | | 421 | 182 |
| gender: 1=woman | -356 | 397 | -0.58 | 0.32 | 19 | 433 |
| age | | | | | | |
| > 6 years of school | | | | | | |
| 1 if first of family in town | | | -0.44 | 0.47 | | |
| years family/resp lived in village | | | -0.01 | 0.01 | | |
| 1 if resp holds trad. office | | | 3.91 | 0.70 | | |
| number of wives of father | | | 0.39 | 0.18 | | |
| number of father's children | | | -0.08 | 0.03 | | |
| parity of mom in father's wives | | | -0.44 | 0.28 | | |
| 1 if fostered as child | | | 0.86 | 0.25 | | |
| size of inherited land | | | -0.29 | 0.27 | | |
| 1 if mother had any education | | | -0.87 | 0.46 | | |
| 1 if father had any education | | | -0.13 | 0.43 | | |

# Link back to the estimation formula

## Stage 1

- Regress fallow duration on instrumental variables
- Compute predicted value of fallow duration

## Stage 2

Regress profit on the predicted value of fallowing

# Reminder and a word of caution…

- *corr* $(Z,\varepsilon) = 0$
  - If *corr* $(Z, \varepsilon) \neq 0$, "Bad instrument"
  - "Finding" a good instrument is hard!
  - But you can build one yourself with a randomized encouragement design

- *corr* $(Z,P) \neq 0$
  - "Weak instruments": the correlation between $Z$ and $P$ needs to be sufficiently strong.
  - If not, the bias stays large even for large sample sizes.

# Recovering TOT from ATE in case of non-compliance

- Sometimes eligible units are selected randomly into the treatment group, are offered treatment, but not all of them accept it.
- Computing the Average Treatment Effect (ATE)
  Straight difference in average outcomes between the group to whom you offered treatment, and the group to whom you did not offer treatment
- Computing the Effect of Treatment on the Treated (TOT)
  Use the randomized offering as an instrumental variable *(Z)* for whether people accepted the treatment *(P)*

# Note: IV is a ‘local’ effect

- IV methods identify the average gains to persons induced to change their choice by a change of the instrument (referred to as compliers)
  - … however we cannot identify who these people are (“local average treatment effect” or LATE)
  - … different instruments will identify different parameters and answer different questions
- Caution in extrapolating to the whole population

# Instrumental Variables and Randomized Experiments

- Imperfect compliance in randomized trials
  - Some individuals assigned to treatment group will not receive $T_x$, and some assigned to control group will receive $T_x$
    - Assignment error; subject refusal; investigator discretion
  - Some individuals who receive $T_x$ will not change their behavior, and some who do not receive $T_x$ will change their behavior
    - A problem in randomized job training studies and other social experiments (e.g., housing vouchers)

# Instrumental Variables and Randomized Experiments

- Two different measures of treatment (X)
  - Treatment assigned = Exogenous
    - Intention-to-treat (ITT) analysis
      - Reduced-form model: $\mathbf{Y = \delta_0 + \delta_1 Z + \xi}$
    - Often leads to underestimation of treatment effect
  - Treatment delivered = Endogenous
    - Individuals who do not comply probably differ in ways that can undermine the study
    - Self-selection $\therefore$ bias and inconsistency

# References


- Angrist, J. D. and A. Krueger (2001). "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", *Journal of Economic Perspectives, 15(4)*.
- Angrist, J. D., G. W. Imbens and D. B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association,* Vol. 91, 434.
- Angrist, J., Bettinger, E., Bloom, E., King, E. and M. Kremer (2002). "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment", *American Economic Review, 92, 5.*
- Bradlow, E., (1998). "Encouragement Designs: An Approach to Self-Selected Samples in an Experimental Design", *Marketing Letters,* 9(4)
- Imbens, G. W. and J. D. Angrist, (1994). "Identification and Estimation of Local Average Treatment Effects." *Econometrica,* 62(2).
- Newman, J., M. Pradhan, L. B. Rawlings, G. Ridder, R. Coa, J. L. Evia, (2002). "An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund.", *World Bank Economic Review*, vol. 16(2).