Statistical Modeling and Causal Inference with R

Week 2: Potential Outcomes Framework

Manuel Bosancianu

Max Schaub

September 14, 2020

Hertie School of Governance

Outline

- ✓ Quick recap:
  - ✓ Fundamental terms: ITE, ATE, ATT, ATU, and NATE;
  - ✓ Biases in inference: selection bias.
- ✓ Example of Banerjee, Duflo, Glennerster, and Kinnan (2015)

Recap

Fundamental terms

In the Potential Outcomes Framework (POF),

$$ITE = \delta_i = y_{1i} - y_{0i} \tag{1}$$

We never see both states of the world at the same time ("fundamental problem" of causal inference).

$$y_i = d_i y_{1i} + (1 - d_i) y_{0i} \tag{2}$$

Fundamental terms

$$ATE = E[\delta_i] = E[y_{1i} - y_{0i}] = E[y_{1i}] - E[y_{0i}] \tag{3}$$

Impossible to compute for actual data.

$$\begin{aligned} ATT &= E[\delta_i | d_i = 1] \\ &= E[y_{1i} - y_{0i} | d_i = 1] \\ &= E[y_{1i} | d_i = 1] - E[y_{0i} | d_i = 1] \end{aligned} \tag{4}$$

Fundamental terms

$$\begin{aligned} ATU &= E[\delta_i | d_i = 0] \\ &= E[y_{1i} - y_{0i} | d_i = 0] \\ &= E[y_{1i} | d_i = 0] - E[y_{0i} | d_i = 0] \end{aligned} \tag{5}$$

One quantity we can compute is the "naive" estimate of the ATE.

$$NATE = E[y_{1i} | d_i = 1] - E[y_{0i} | d_i = 0] \tag{6}$$

Connections: NATE–ATE–ATT–ATU

$$ATE = p * ATT + (1 - p) * ATU, \ where \ p = prob(D = 1) \quad (7)$$

$$NATE = ATE + \underbrace{E[Y_0|D = 1] - E[Y_0|D = 0]}_{selection\ bias} + \underbrace{(1 - p)(ATT - ATU)}_{HTE\ bias} \quad (8)$$

Biases in causal inference

Selection bias: the difference in expected outcomes in the absence of treatment for the actual treatment and control group.

HTE bias: the difference in *returns to treatment* (the treatment effect) between the treatment and control group, multiplied by the share of the population in control.

NATE is a good substitute for *ATE* only in the absence of selection bias and HTE bias.

Effects of Microfinance

Why this article

- ✓ Question of significance . . .

Why this article

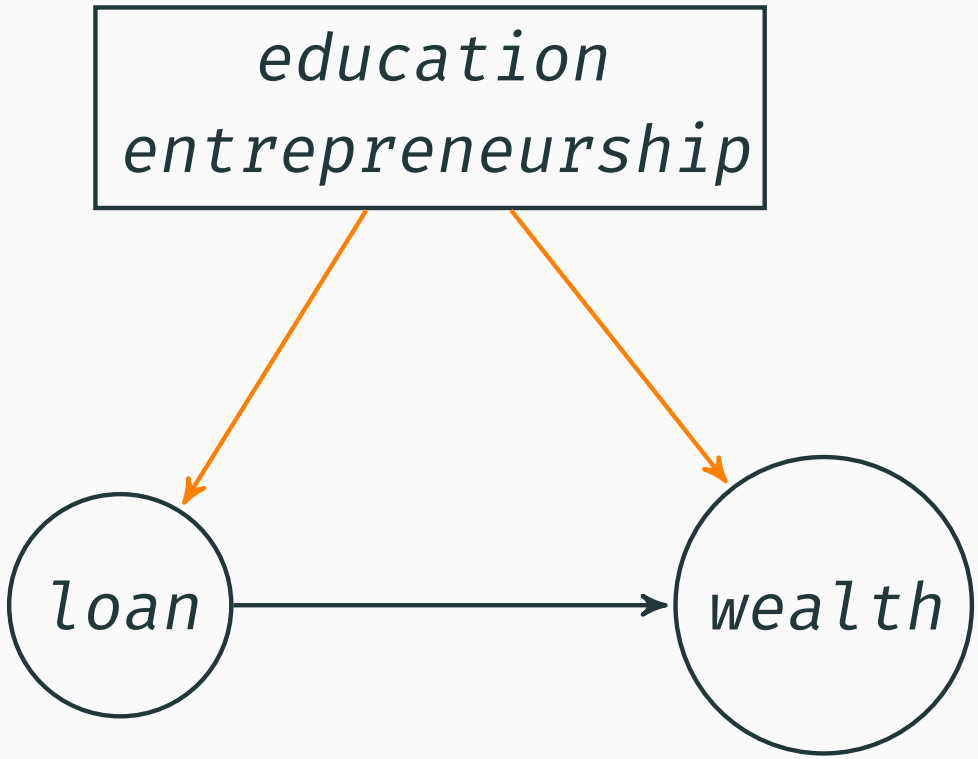
- ✓ Question of significance . . .
- ✓ . . . plagued by issues of *self-selection* . . .

Why this article

- ✓ Question of significance . . .
- ✓ . . . plagued by issues of *self-selection* . . .
- ✓ . . . highlighting challenges of doing good causal inference

In terms of design and analysis it is very sophisticated, but it allows us to discuss biases and assumptions.

Limits of observational data



Those taking loans and those with no loans are bound to be different.

The setup

104 disparate neighborhoods in Hyderabad (2005): 52 randomly allocated to opening a Spandana branch.

15–18 months after: household survey with 65 HHs in each neighborhood.

2 years after: follow-up survey with same HHs.

Balance

TABLE 1A—BASELINE SUMMARY STATISTICS

	Control group			Treatment − control	
	Obs. (1)	Mean (2)	SD (3)	Coeff. (4)	*p*-value (5)
Household composition					
Number members	1,220	5.038	(1.666)	0.095	0.303
Number adults (>=16 years old)	1,220	3.439	(1.466)	−0.011	0.873
Number children (<16 years old)	1,220	1.599	(1.228)	0.104	0.098
Male head	1,216	0.907	(0.290)	−0.012	0.381
Head's age	1,216	41.150	(10.839)	−0.243	0.676
Head with no education	1,216	0.370	(0.483)	−0.008	0.787
Access to credit					
Loan from Spandana	1,213	0.000	(0.000)	0.007	0.195
Loan from other MFI	1,213	0.011	(0.103)	0.007	0.453
Loan from a bank	1,213	0.036	(0.187)	0.001	0.859
Informal loan	1,213	0.632	(0.482)	0.002	0.958
Any type of loan	1,213	0.680	(0.467)	0.002	0.942
Amount borrowed from (in Rs)					
Spandana	1,213	0	(0.000)	69	0.192
Other MFI	1,213	201	(2,742)	170	0.568
Bank	1,213	7,438	(173,268)	−5,420	0.279
Informal loan	1,213	28,460	(65,312)	−570	0.856
Total	1,213	37,892	(191,292)	−5,879	0.343

Timeline

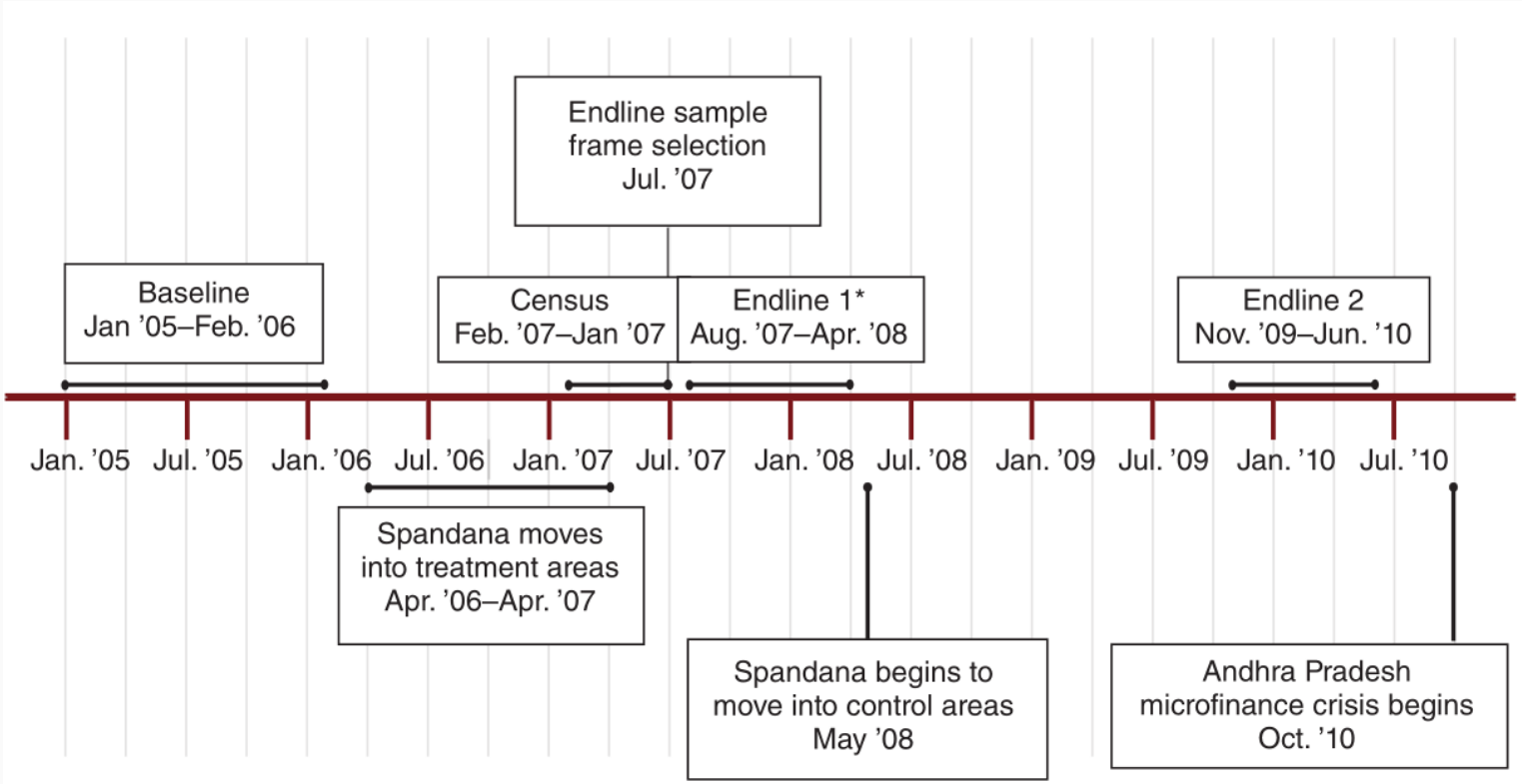


FIGURE 1. TIMELINE OF INTERVENTION AND DATA COLLECTION

Specification

$$Y_{ia} = \alpha + \beta \times Treat_{ia} + \boldsymbol{X}_a\gamma + \epsilon_{ia} \tag{9}$$

$\boldsymbol{X}_a$ is a vector of controls:

- ✓ area population
- ✓ total businesses
- ✓ average expenditure p.c.
- ✓ household literacy

Difference-in-means, with complications (controls, clustered SEs, weights).

Assumptions

SUTVA (non-interference)

”Treatment value” is the same across all population units.

- ✓ treatment is of uniform intensity
- ✓ no **externalities**

SUTVA (non-interference)

”Treatment value” is the same across all population units.

- ✓ treatment is of uniform intensity
- ✓ no **externalities**

People move in/out of areas because of accessibility of lending.

SUTVA (non-interference)

”Treatment value” is the same across all population units.

- ✓ treatment is of uniform intensity
- ✓ no **externalities**

People move in/out of areas because of accessibility of lending.

People loan to neighbors.

SUTVA (non-interference)

”Treatment value” is the same across all population units.

- ✓ treatment is of uniform intensity
- ✓ no **externalities**

People move in/out of areas because of accessibility of lending.

People loan to neighbors.

Increased economic activity benefits non-borrowers too.

Excludability

The only reason for the change in potential outcomes is the treatment.

Excludability

The only reason for the change in potential outcomes is the treatment.

Other organizations move into the control areas and start lending.

The treatment

Difficulties with these result in re-defining treatment: exposure to Spandana borrowing.

Analysis carried out at individual level, but key estimate is at area level, averaging over borrowers and non-borrowers.

"Intent-to-treat" effect on likely borrowers:

- ✓ lived in area > 3 years
- ✓ HH includes woman aged 18–55

Results

Results

TABLE 3B—SELF-EMPLOYMENT ACTIVITIES: REVENUES, ASSETS AND PROFITS (*Households with old businesses*)

	Assets (stock) (1)	Investment in last 12 months (2)	Revenue (3)	Expenses (4)	Profit (5)	Employees (6)	Index of dependent variables (7)
Panel A. Endline 1							
Treated area	898	1,119	5,266	1,620	2,105*	−0.05	0.09
	(1,063)	(698)	(3,720)	(3,257)	(1,100)	(0.0824)	(0.0406)
Observations	2,083	2,083	1,955	2,020	1,624	2,088	2,088
Control mean	6,757	678	14,505	12,325	2,038	0.41	0.00
Hochberg-corrected *p*-value							0.057
Panel B. Endline 2							
Treated area	1,682	−948	343	−2,644*	839	−0.12	−0.007
	(1,412)	(588)	(1,263)	(1,491)	(945)	(0.099)	−0.0263
Observations	1,878	1,878	1,859	1,862	1,844	1,878	1,878
Control mean	10,301	2,292	12,564	12,418	1,948	0.46	0.00
Hochberg-corrected *p*-value							>0.999

Results

TABLE 4—INCOME

	Self employment (profit) (1)	Daily labor/salaried (2)	Index of dependent variables (3)
Panel A. Endline 1			
Treated area	354	−526	−0.0501
	(314)	(358)	(0.0459)
Observations	6,239	6,827	6,832
Control mean	745	2,988	0.000
Hochberg-corrected *p*-value			>0.999
Panel B. Endline 2			
Treated area	542	−141	0.0114
	(372)	(212)	(0.0261)
Observations	6,090	6,142	6,142
Control mean	953	5,514	0.000
Hochberg-corrected *p*-value			>0.999

Results

TABLE 6—CONSUMPTION (*Per capita, per month*)

	Total (1)	Durables (2)	Nondurable (3)	Food (4)	Health (5)	Education (6)	Temptation goods (7)	Festivals and celebrations (8)	Home durable good index (9)
Panel A. Endline 1									
Treated area	10.24	19.73*	−6.50	−12.11	−3.7	−2.061	−8.785*	−14.16*	−0.051
	(37.22)	(11.35)	(31.81)	(12.06)	(11.51)	(9.865)	(4.92)	(8.09)	(0.057)
Observations	6,827	6,781	6,781	6,827	6,827	5,415	6,827	6,827	6,841
Control mean	1,419	116	1,305	525	140	168	84	69	2.37
Hochberg-corrected *p*-value	>0.999								
Panel B. Endline 2									
Treated area	−48.83	0.42	−45.45	−11.20	−22.54	12.16	−10.07	6.17	−0.0127
	(51.53)	(9.88)	(46.92)	(17.88)	(17.50)	(15.19)	(6.61)	(4.12)	(0.0426)
Observations	6,142	6,140	6,142	6,142	6,141	4,910	6,142	6,103	6,142
Control mean	1,914	131	1,755	687	187	206	118	90	2.66
Hochberg-corrected *p*-value	0.691								

Results

TABLE 7—SOCIAL EFFECTS

	Share of children aged 5–15 in school		Hours worked per child aged 5–15 over the past 7 days:		Share of teenagers (aged 16–20) in school		Index of women's independence/ empowerment	Number new self-employ. activities managed by women (all HHs)	Index of dependent variables
	Girls (1)	Boys (2)	Girls (3)	Boys (4)	Girls (5)	Boys (6)	(7)	(8)	(9)
Panel A. Endline 1									
Treated area	−0.016	−0.012	−0.028	0.613	−0.037	−0.007	0.007	0.0143***	−0.008
	(0.013)	(0.011)	(0.202)	(0.743)	(0.024)	(0.028)	(0.023)	(0.005)	(0.0097)
Observations	3,035	3,073	3,035	3,073	2,174	1,866	6,862	6,762	6,862
Control mean	0.919	0.918	0.594	0.577	0.338	0.429	−0.001	0.026	0.000
Hochberg-corrected *p*-value									>0.999
Panel B. Endline 2									
Treated area	0.015	0.007	0.092	−0.531*	0.021	−0.021	−0.011	−0.005	0.005
	(0.011)	(0.011)	(0.133)	(0.269)	(0.024)	(0.027)	(0.021)	(0.006)	(0.009)
Observations	2,755	2,746	2,755	2,746	1,789	1,665	6,142	6,142	6,142
Control mean	0.923	0.928	0.286	1.379	0.329	0.474	−0.003	0.047	0.000
Hochberg-corrected *p*-value									>0.999

Thank you for the kind attention!

References

Banerjee, A., Duflo, E., Glennerster, R., & Kinnan, C. (2015). The Miracle of Microfinance? Evidence from a Randomized Evaluation. *American Economic Journal: Applied Economics*, *7*(1), 22–53.