

# Cross-cutting issues in impact evaluations involving firms

David McKenzie  
September 2019

# Some key issues

- 1) Which firms should we target our programs to, and treatment heterogeneity?
- 2) How do we get firms to take-up our programs?
- 3) Choosing the right time frame to measure impacts
- 4) Dealing with Small Samples
- 5) *Bonus content* If what we are trying to do is so good, why aren't firms already doing it?
  - The role of information
  - The difficult of learning

Targeting

# Common approach version 1

- *Policy aim:* encourage entrepreneurship, innovation, exports, and job creation
- *Example instruments:* matching grants, business plan competitions, investment readiness programs, export facilitation
- *Common targeting approach:* try to pick the firms you think have the highest likelihood of success aka “the best” firms.

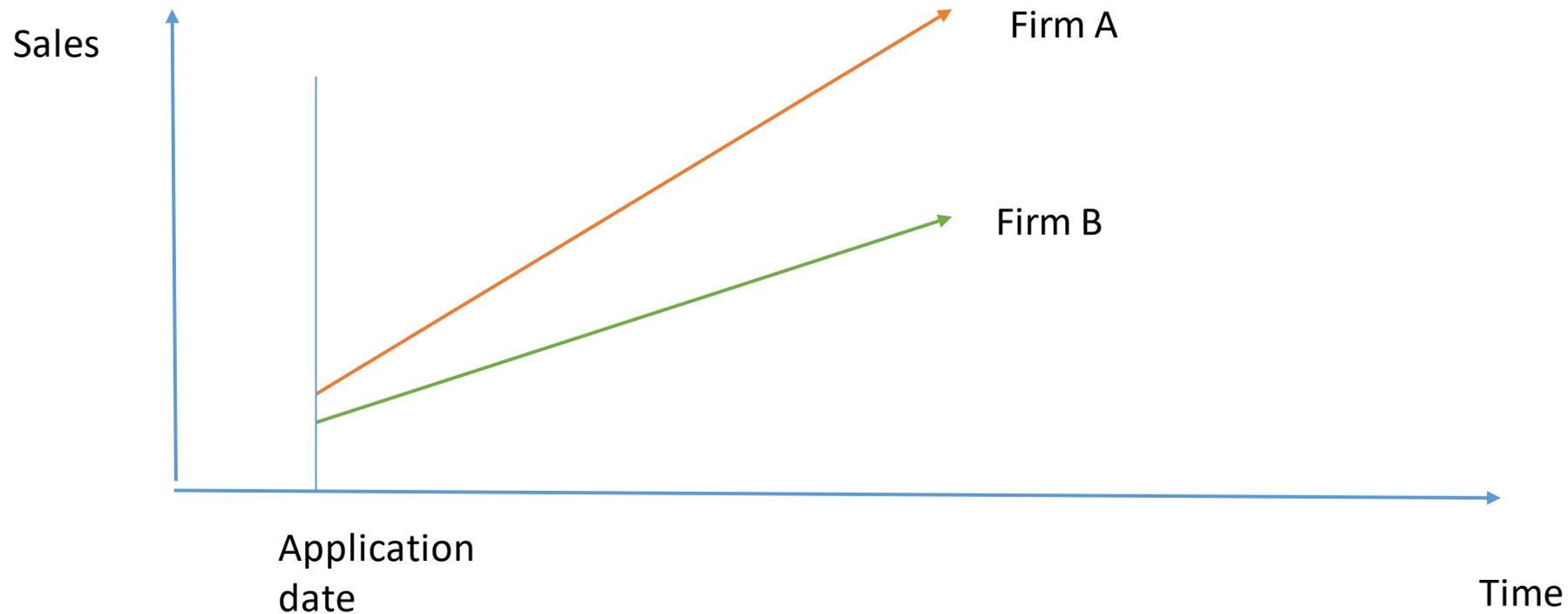
e.g. score all the applications for your grant program, and then award the program to the highest scoring

# Common approach version 2

- *Policy aim:* help microenterprises to grow and increase the incomes of the poor
- *Example instruments:* microfinance, business training programs, small grants
- *Common targeting approach:* target the most disadvantaged firms – e.g. female-owned firms in rural areas, threshold in terms of poverty or assets

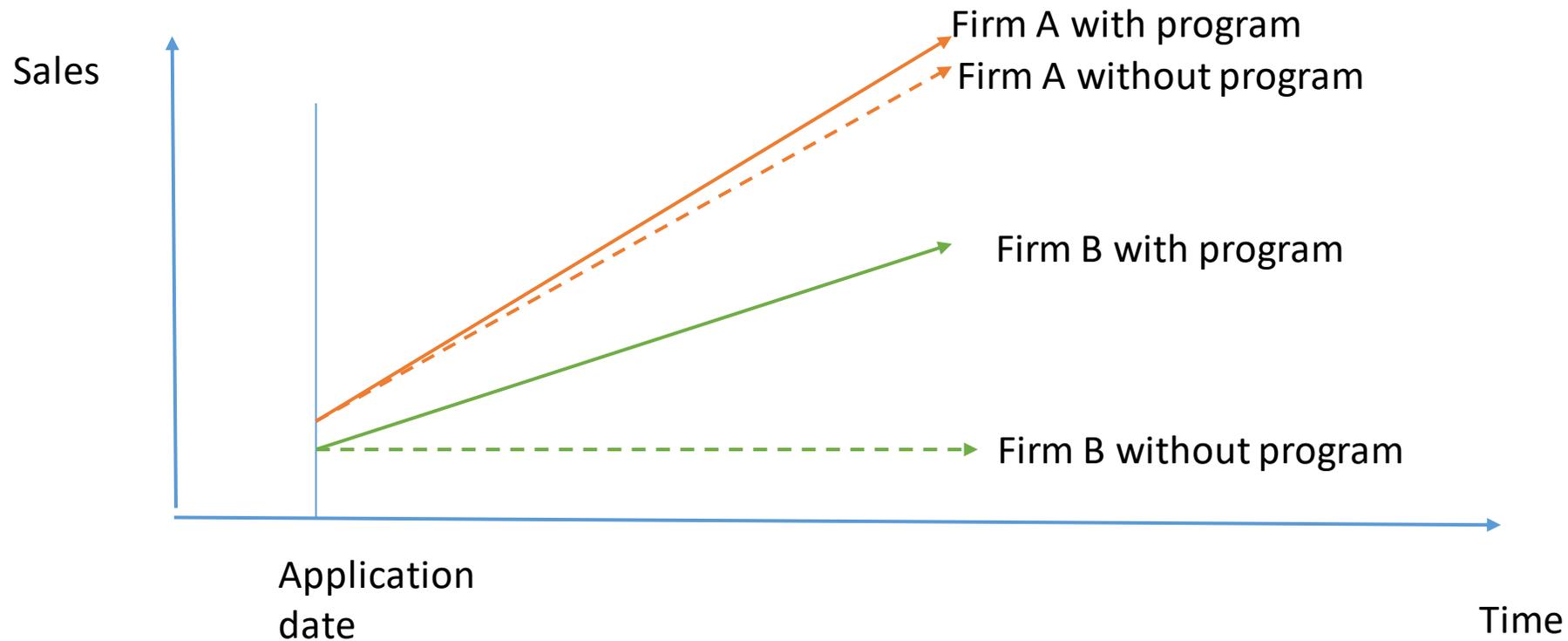
# What's wrong with these approaches?

- They are attempting to target on *levels and predicted growth* rather than *policy impacts*



# What's wrong with these approaches?

- They are attempting to target on *levels and predicted growth* rather than *policy impacts*

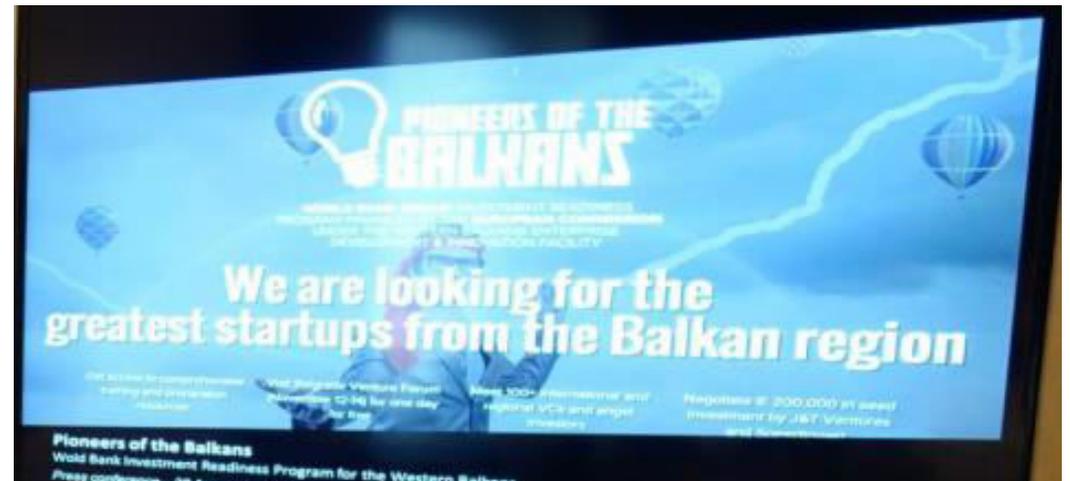


# The government is generally not a venture capitalist

- If you are a VC, want to identify the best firms, invest in them, and then you don't care whether they grow *because* of your money, or because you are just good at identifying who will grow fast anyway
- Policymaker should care more about who will benefit most from taking part in the program

# Example 1: Investment Readiness in the Western Balkans

- **Five country RCT** to test investment readiness program (IRP) in the Western Balkans:
  - Croatia, Serbia, Montenegro, Macedonia, Kosovo
- **Marketed as competitive program for innovative firms** seeking outside financing
- **346 start-ups/SMEs** randomized into:
  - **Treatment group** – high cost and intensive investment readiness program
  - **Control group** – gets inexpensive online only course



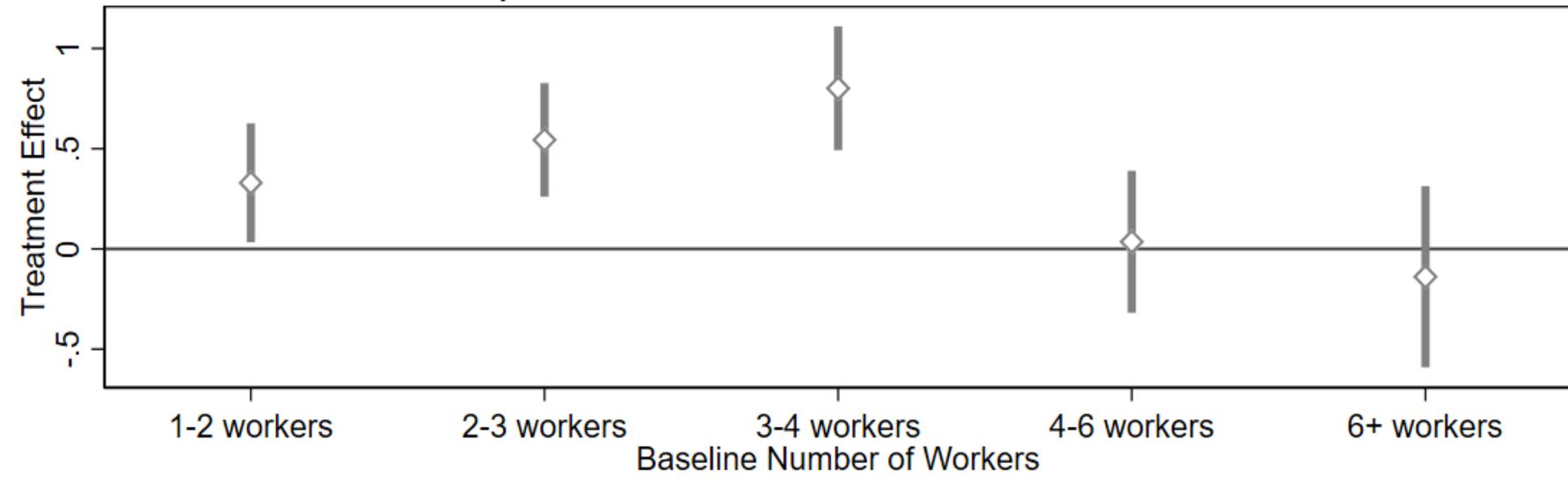
# What did we find?

- Firms pitched to judges, who scored them on investment readiness
  - We find treated firms **get significantly higher investment readiness scores**, by 0.3 s.d.
- Then track firms for 2 years, and measure whether they get external financing
  - Control group were more successful in making deals with outside investors than we had thought – 24% made a deal
  - Treatment effect was +5 percentage points, but not statistically significant (confidence interval is -4.7 p.p, + 14.7 p.p.)
- Reaction of country units/research community to imprecise impact...

# What were we missing?

- We had thought impacts might differ according to initial level of investment readiness – found positive but not significant interaction with low readiness.
- But perhaps this is not the best metric of who will benefit most
- Referee suggested size

### Impact on Investment Readiness Score



### Impact on Making a Deal with an Investor



# Abadie et al. Endogenous stratification method

**Table 6: Heterogeneity in Treatment by Predicted Likelihood of Making a Deal with an Investor**

	Impact on Making a Deal within 2 years	
	Leave-one-out estimator	Repeated Split-Sample Estimator
Low Predicted Likelihood of Funding	0.143** (0.066)	0.124** (0.055)
High Predicted Likelihood of Funding	-0.081 (0.079)	-0.049 (0.074)

Notes: Bootstrap standard errors, based on 500 bootstrap replications, are reported in parentheses.

\*, \*\*, \*\*\* denotes significance at the 10, 5, and 1 percent levels respectively.

The repeated split-sample estimator uses 200 splits of the data.

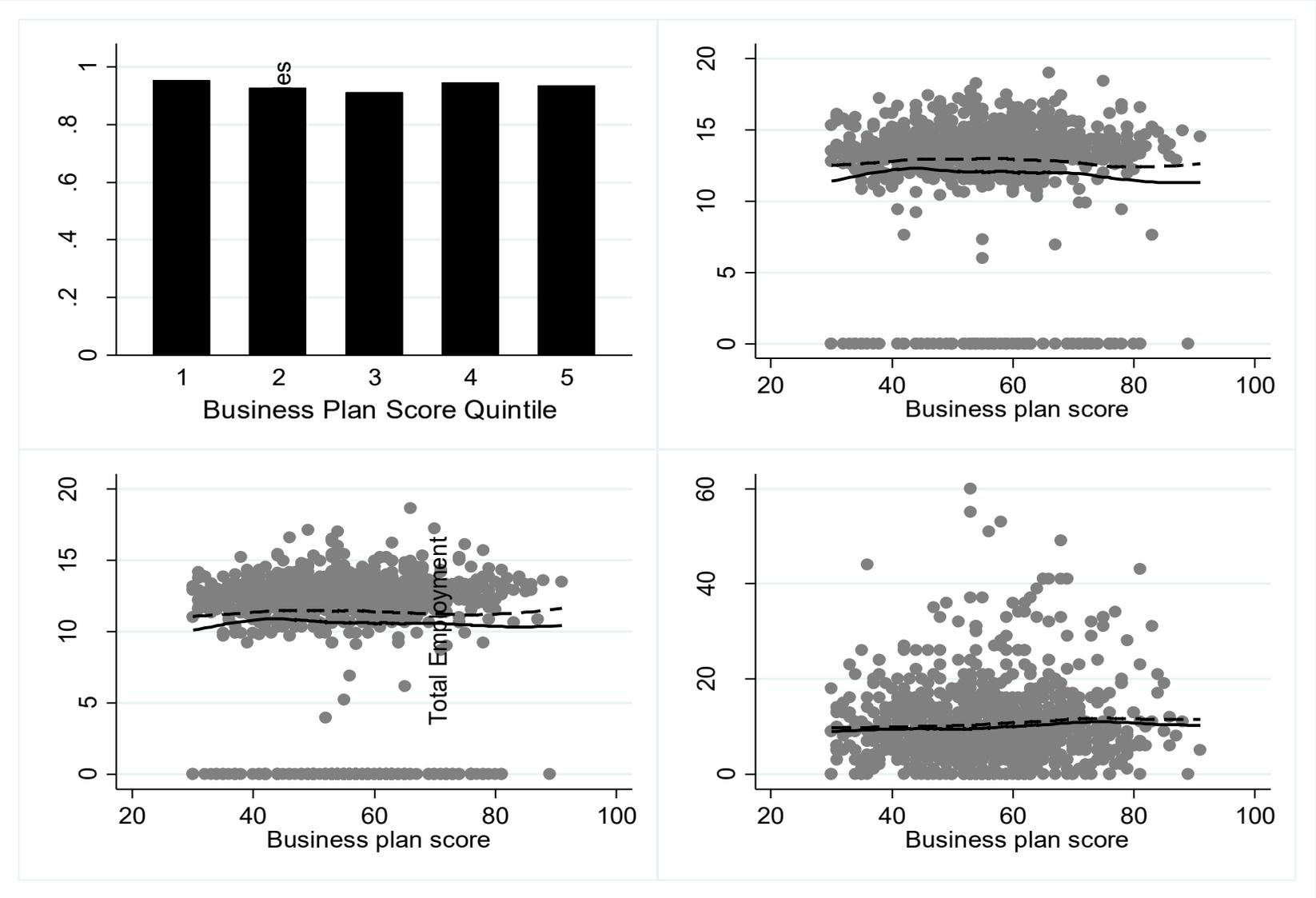
Predicted likelihood of funding based on the following baseline characteristics:

Employment above the median, initial investment readiness score, country, whether the firm has had a private investor, whether it classifies itself as early stage, sector, firm age, whether the main founder has post-graduate education, whether at least one founder is female, and whether the firm has previously received mentoring. Abadie et al. (2018) endogenous stratification approach used.

# Example 2: Business plan competition in Nigeria

- 24,000 firms apply to the program
- Top 2,400 selected as semi-finalists based on scores
- Then choose top 300 overall as winners with certainty, choose some regional winners based on top marks, then randomly choose 720 winners from among remaining 1900 semi-finalists
- Measure impacts on firm creation, employment, sales growth of winning the program

# Really hard to predict success



# Example 3: Small grants to firms in Sri Lanka

- Experiment: Gave grants of \$100 and \$200 to male and female-owned microenterprises in Sri Lanka – randomly choose which firms to give these to, and measure subsequent return on capital.
- What do we find:
  - Female-owned firms and firms whose owners have lower cognitive ability have lower profits to start with
  - Grants have much larger impacts on male-owned firms, and on firms whose owners have higher ability

# So what should you do?

- Often unclear who our program is going to work best for
  - Theory can offer some predictions, but often easy to think of reasons which could go either way
    - ⇒ Big role for impact evaluation to help in figuring out who to target, and to test targeting in pilot phase before scale-up.
- ⇒ For now:
- ⇒ Maybe easier to figure out who program will NOT work for than who it will work for
    - mounting evidence that firms that are small and old are not high-growth-potential firms
    - If firms can't get their act together to apply online for your program, and to exert enough effort to attend some introductory event/initial ordeal – then unlikely to make it all the way through a program.

# Research Questions of Interest

- Relative performance of different ways of predicting success
- What works to predict treatment impacts rather than grow in the absence of a program?
- Impact of different screening mechanisms on skewness of outcome:
  - Does your targeting mechanism target “go big or bust” firms, or firms with lower riskiness, but less upside?
- Role of price – does charging firms help target, or help effectiveness?
  - Jamaica business training evidence

Take-up

# Often even when our programs are free, many firms don't participate in them

- Examples

- Matching grant programs

- In review of 15 projects, only 70% of funding disbursed on average
    - Program in Puebla – only 80 out of 150 successful applicants actually took-up grants, even though all had applied and signed letters of intent

- Business training programs

- On average only 65% of firms invited attend, even when training offered for free

- Firm formalization programs

- Even after big reforms to make it easy to register for firms, majority of informal firms don't respond

# One possible reaction

- “Low take-up is not a problem – it helps sort firms into those who really value the program and those who don’t, and we don’t waste our time or resources on the latter”

# So why should you care about low take-up?

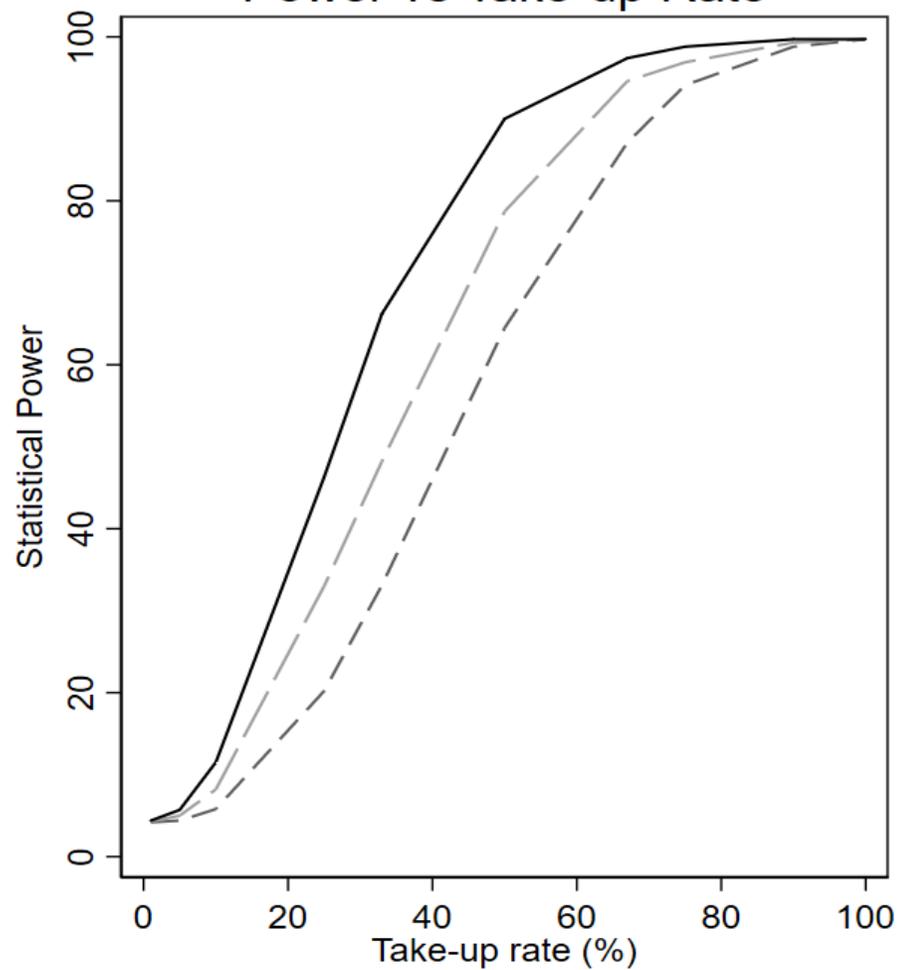
- 1) Project implementation reason: low take-up may reflect problems with the way you are implementing intervention
  - Not reaching the targeted groups (e.g. hold training in far off location and poor or women with kids can't travel there)
  - Bureaucratic red tape stops those who want program benefiting from it
  - May reflect political capture
- 2) Project effectiveness reason: sometimes the people least interested in program may be the ones who could benefit from it the most
  - E.g. business training/management improvement – badly managed firms often don't know they are badly managed
- 3) Project learning reason: really hard to know if your program is working if take-up is low for it

# Rule of the Inverse Square

- If the take-up rate is  $1/p$ , you need a sample of  $p^2$  times your original sample to have enough power to detect effect
- E.g. If only  $\frac{1}{2}$  the firms you invite take-up your program, you need 4 times the sample as if 100% took it up
- If only  $\frac{1}{4}$  the firms you invite take it up, you need 16 times the sample!

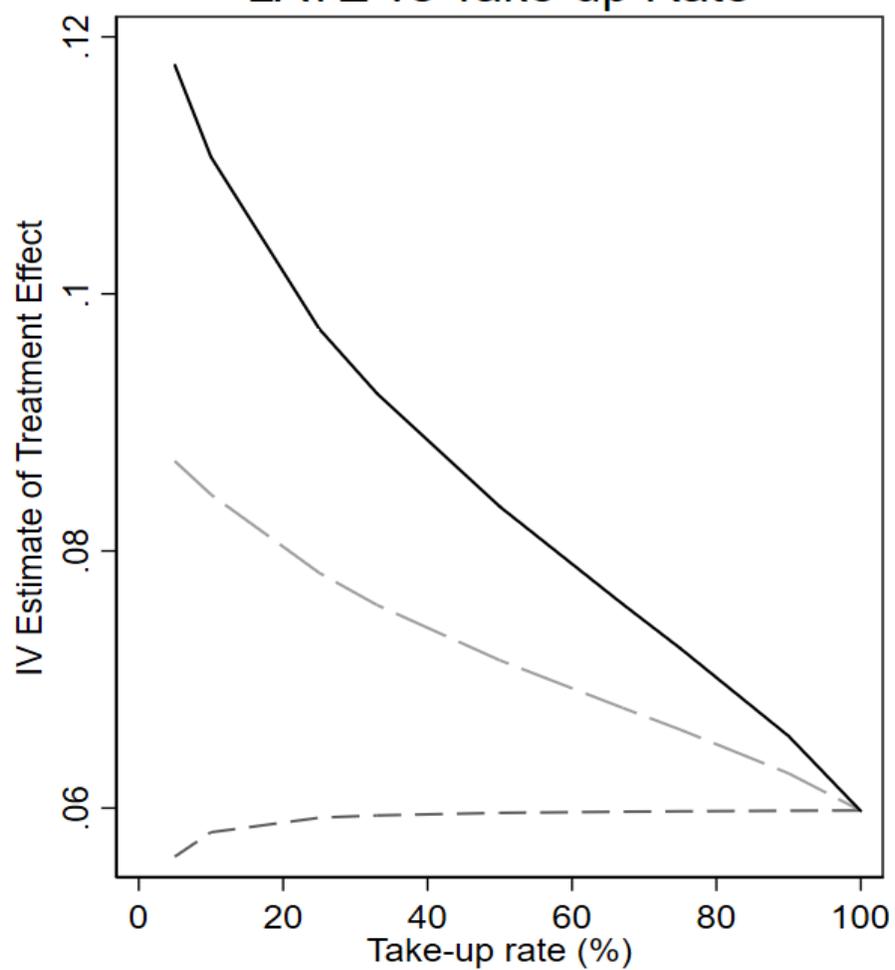
(Note: if there is treatment heterogeneity correlated with take-up, then rule not so severe).

### Power vs Take-up Rate



--- Heterog = 0    -.- Heterog = 0.03  
— Heterog = 0.06

### LATE vs Take-up Rate

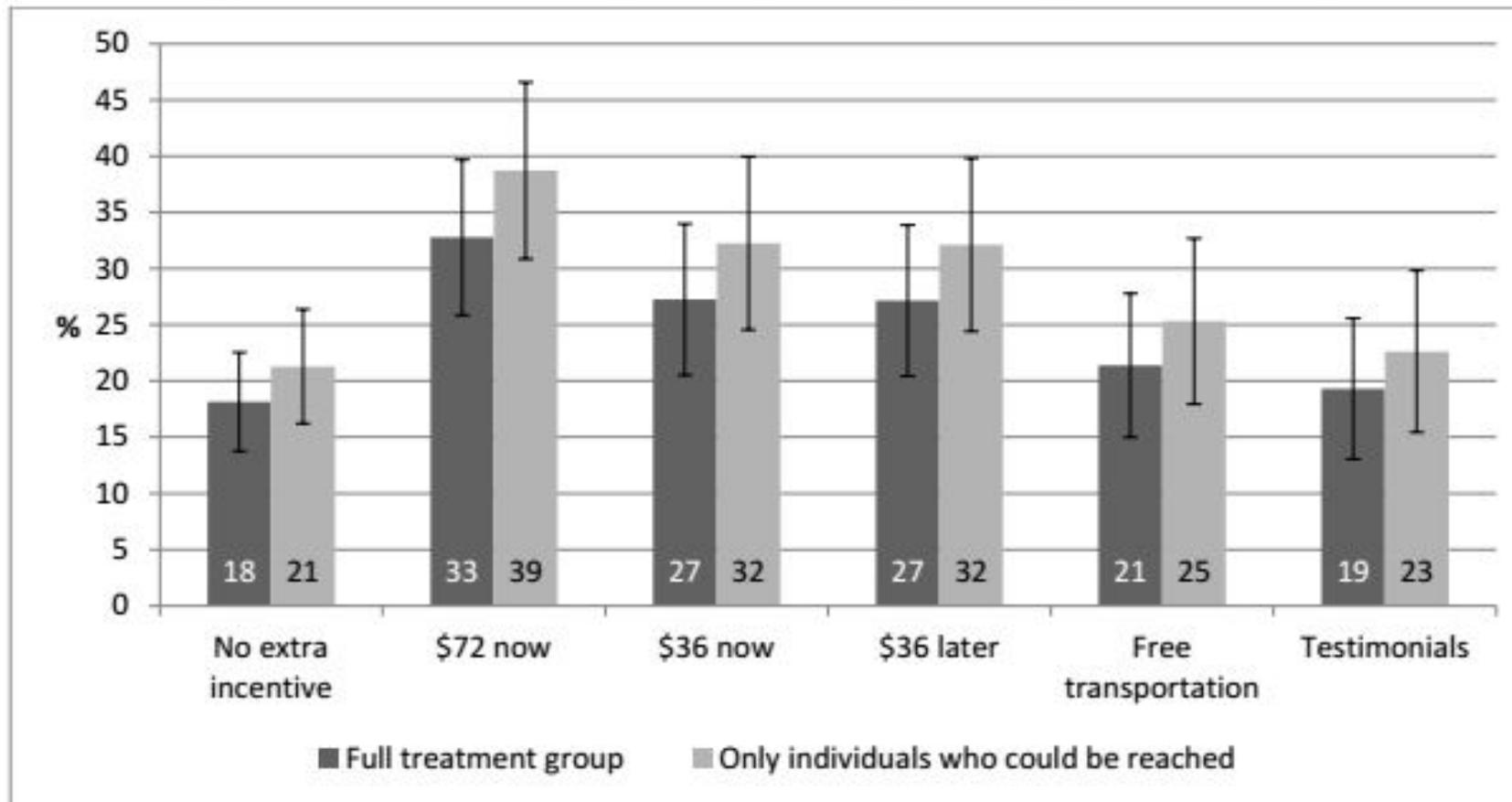


--- Heterog = 0    -.- Heterog = 0.03  
— Heterog = 0.06

# Implications for program design

1) Test different approaches to get people to take up program

**Figure 1: Financial Literacy Training Take-Up Rates by Incentive Group**



# Implications for program design

2) Low take-up can often reveal problems in the design or implementation of your program

3) Need to plan proactively for ways to ensure adequate take-up in order to have sufficient power to learn.

Time-frame for measuring  
impacts

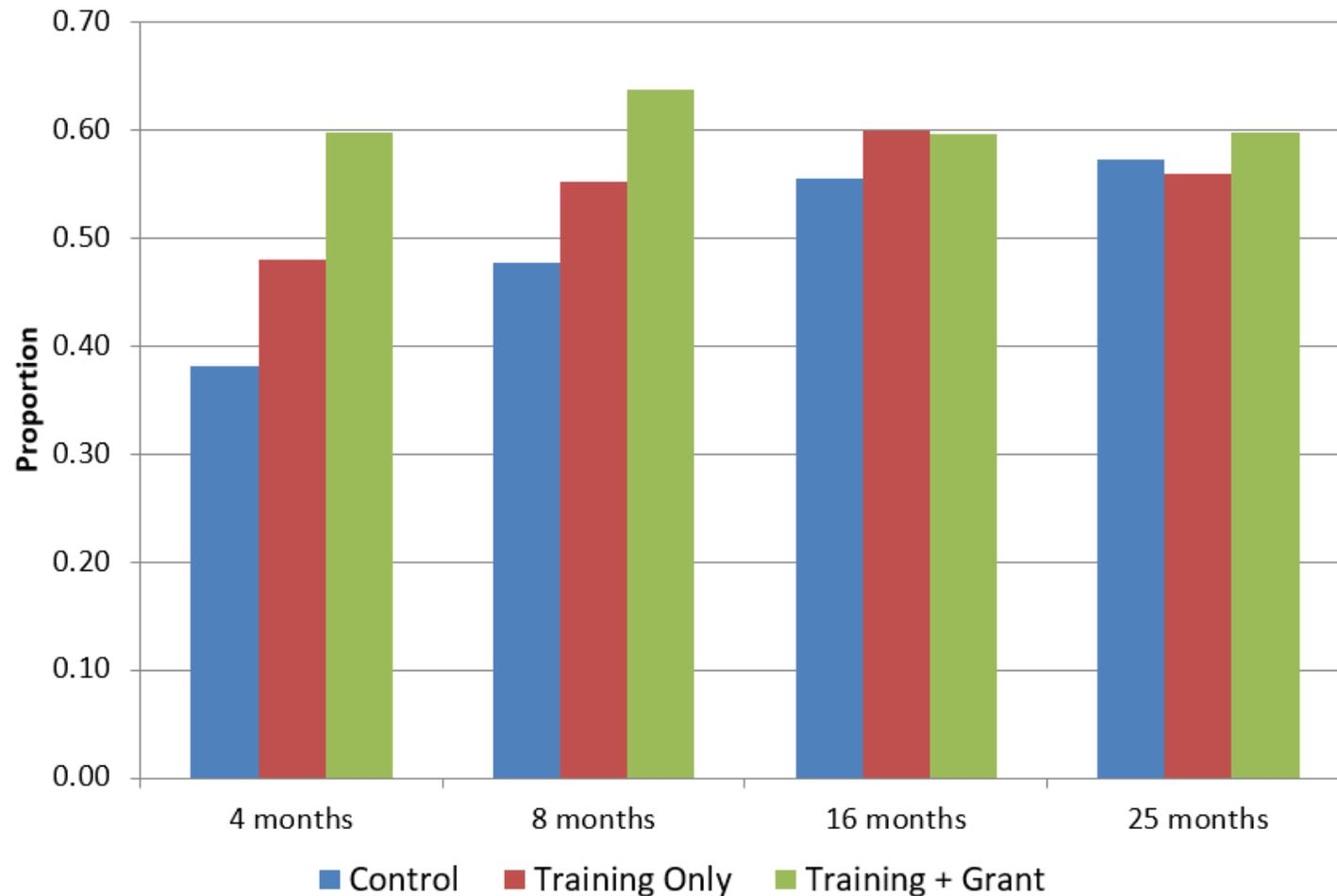
# Choosing right time frame to measure impacts

- Right timing depends on the intervention, but really want to trace out trajectory of impacts
- E.g. business training

# Story 1: Catch-up and failure

- De Mel et al. (2012) in Sri Lanka

Proportion owning a business by survey round



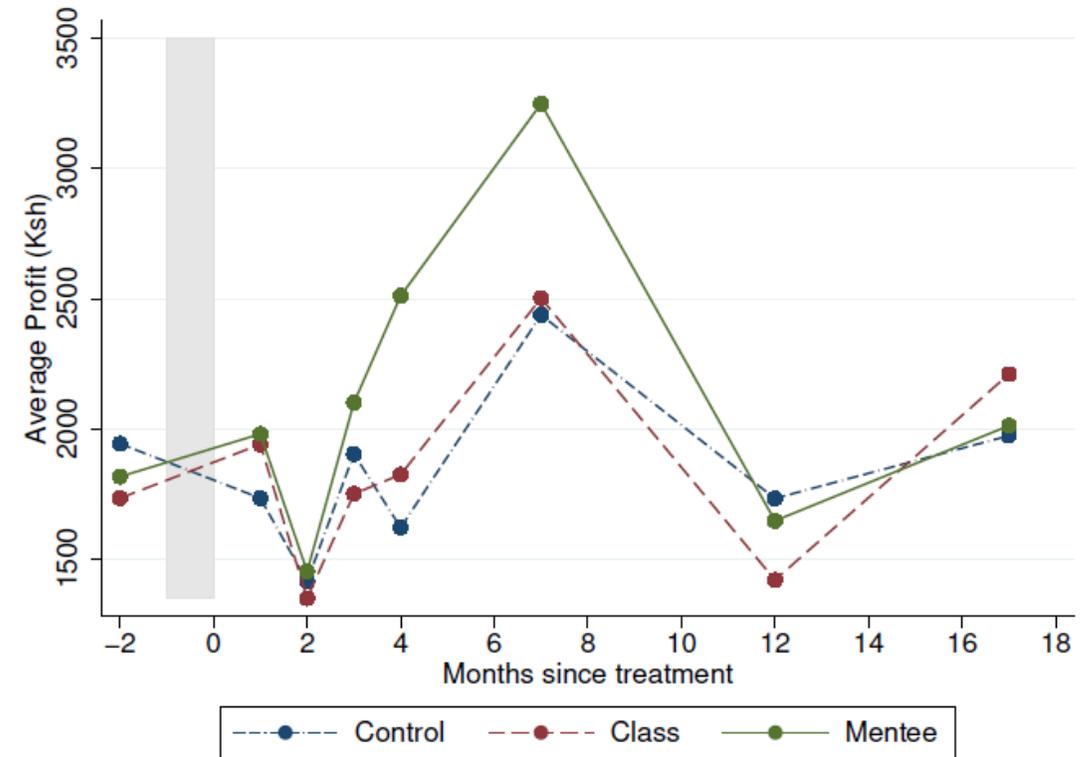
# Story 1: Catch-up and failure

## Mentors or Teachers? Microenterprise Training in Kenya

By WYATT BROOKS, KEVIN DONOVAN, AND TERENCE R. JOHNSON\*

*We use a randomized controlled trial to demonstrate that inexperienced female microenterprise owners in a Kenyan slum benefit from mentorship by an experienced entrepreneur in the same community. Mentorship increases profits by 20 percent on average with initially large effects that fade as matches dissolve. We conduct a formal business education intervention, which has no effect on profits despite changes in business practice. Our results demonstrate that missing information is a salient barrier to profitability, but the type of information matters: access to the localized, specific knowledge of mentors increases profit while abstract, general information from the class does not.*

Figure 3. : Profit time series



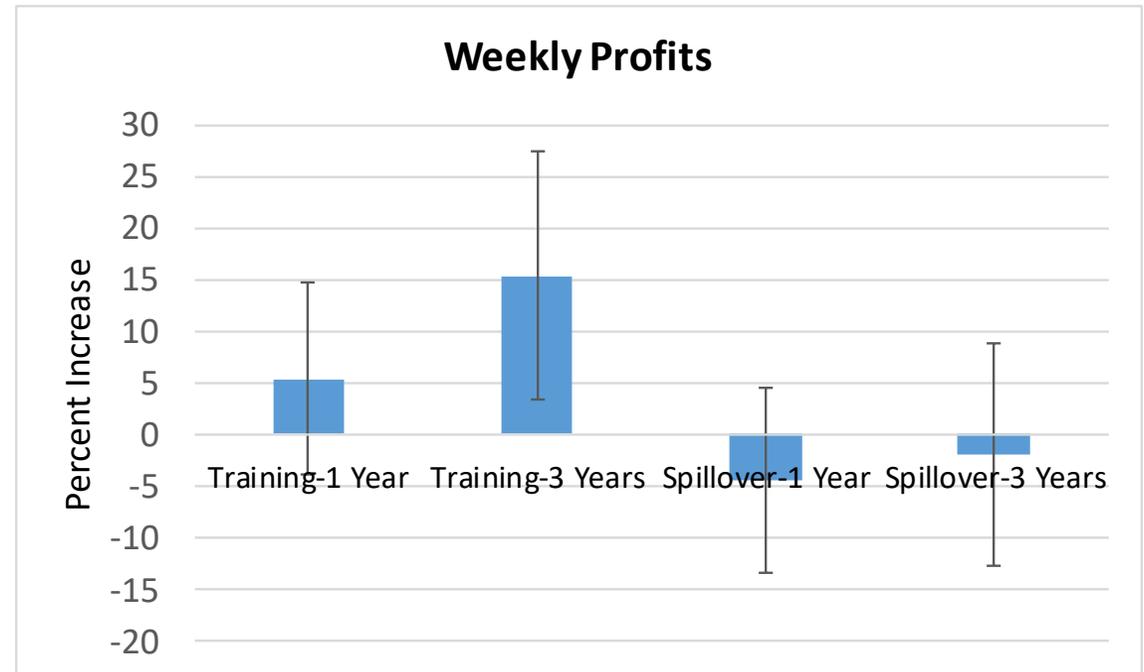
# Story 1 twist: failure and recovery

- Karlan et al. (JEBO) – consulting advice to firms worsened their outcomes in short-run – then they stop practices and recover

# Story 2: It takes time to see effects

- McKenzie and Puerto (2017) in Kenya

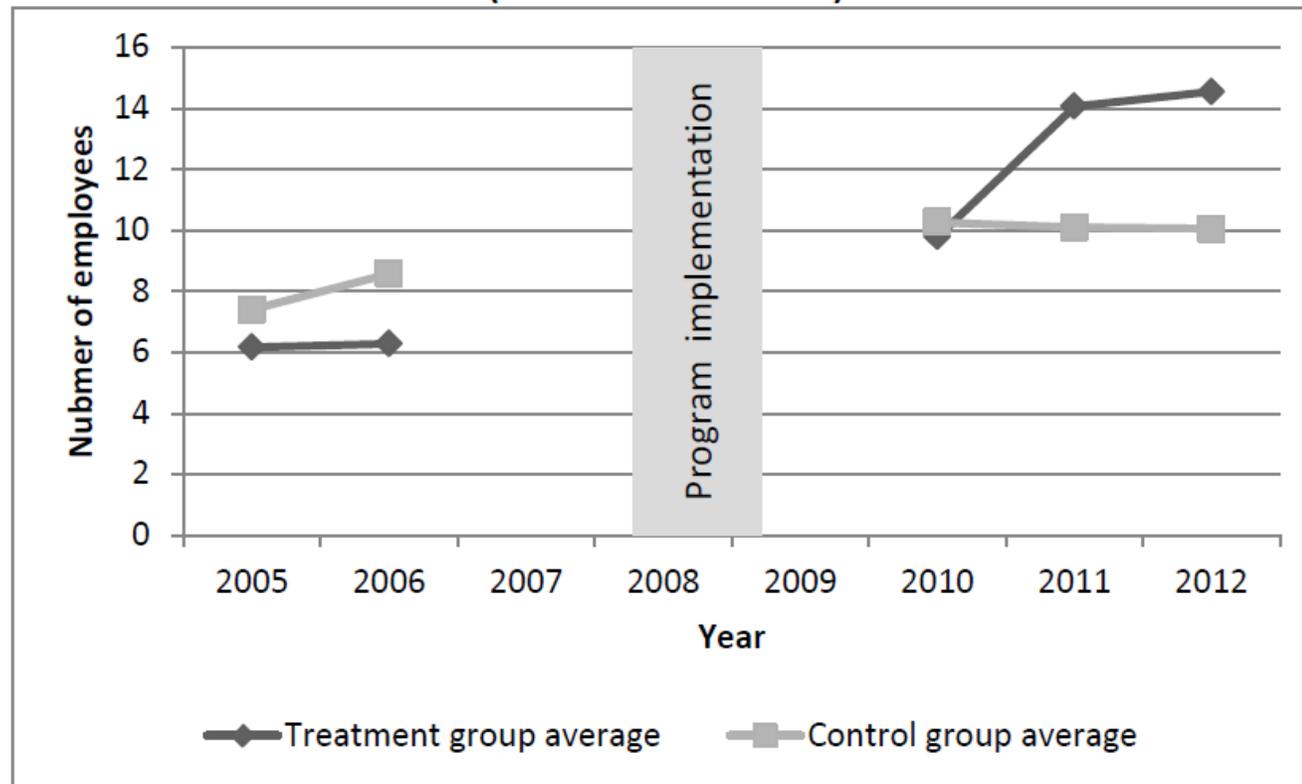
With no access to finance,  
only way to grow is slowly  
re-invest and bootstrap way  
up



# Story 2: It takes time to see effects

- Bruhn et al. (2018, JPE)

Figure 2: Average Number of Employees in the Treatment and Control Group over Time  
(Administrative Data)



## Story 2: It takes time to see effects

- Higuchi et al. (2017) – 312 small manufacturers in Vietnam, impact measured for 5 years
  - 5 year survival rate 52% in control
  - Treatment increases survival by 17 percentage points
- Higuchi and Sonobe (2016, JEBO) – cluster in Tanzania
  - Impacts stronger after 3 years than 1 year
  - Took time to make investments, and to learn through trial and error which practices were useful.

# Dealing with Small Samples

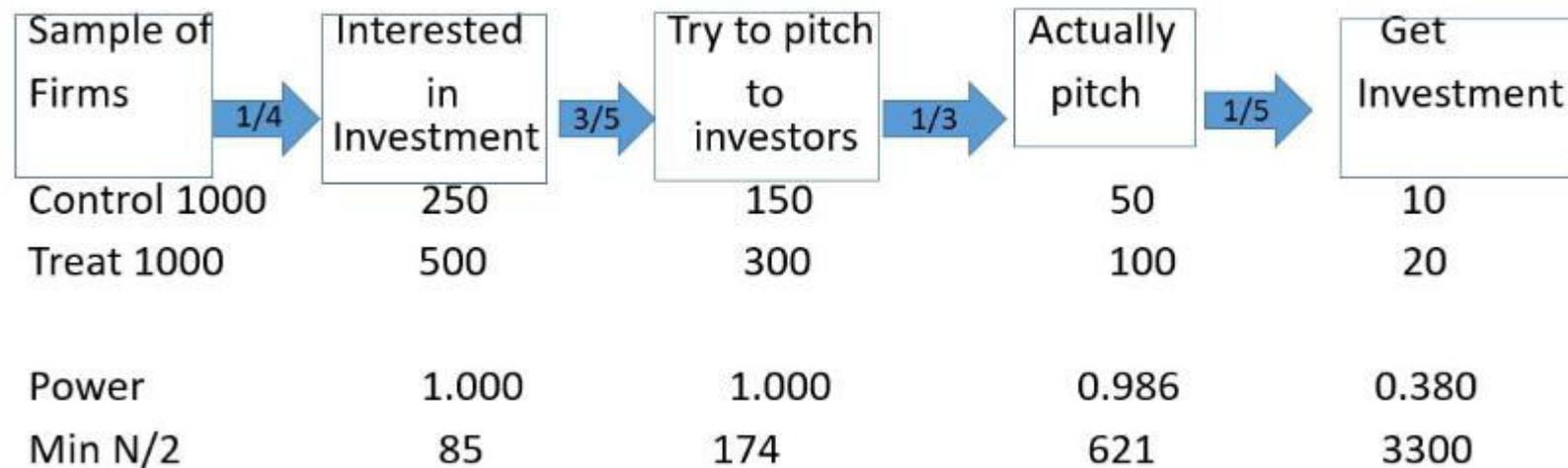
# Tips for Getting the Most out of Small Samples

- 1) Focus on measuring outcomes as accurately as possible (see Monday's talk)
- 2) More T – collect more time observations – can average out noise and boost power
  - Note: only works if your outcome is e.g. monthly profit, and you collect three months instead of one. No gain in splitting monthly into weekly or quarterly into monthly unless improves measurement.
- 3) Outlier reduction – not everyone in your program has to be in your impact evaluation
  - If have a few firms that are very different from the rest, set them aside before randomizing treatment assignment.

# Tips for getting the most out of small samples

4) Remember funnel of attribution – may not be able to accurately measure impact on end outcome (especially if they are rare), but may be able to detect effects on pathways towards it

**Figure 1: The Funnel of Attribution and Challenge of Measuring Impact**



If what we are doing is so good,  
why don't firms do it on their  
own?

# Is there really a market failure?

- E.g. Business Training
  - If it was really great, why don't firms pay for it themselves instead of us having to give it away?
    - Perhaps they are liquidity constrained
      - But then shouldn't the solution be to offer them loans- but then we don't see many firms borrowing in order to finance business training

# The failure of information

- Often firms overestimate the complexity of regulations, underestimate the returns to some types of activities
  - Experiments can test whether simply providing the correct information has any effect
    - Formalization experiments- has not
    - South Africa labor regulations – large effect
- But what stops firms from learning on their own?

# Some things I don't know are reasonably easily to learn if I want to

- Example: Registering a firm
- If you ask me today, I couldn't tell you how much it would cost to register a business in Greece, or what the steps I would have to go through are
- But if I really wanted to register a business, it would be pretty easy for me to find out
  - Look on government websites
  - Show up at a small business help center
  - Hire an accountant
  - Etc.

# Other things are really difficult for firms to learn on their own

Campaign Level Summa		
Cost	Assignment	
	Test	Control
\$128,750	1,257,756	300,000
\$40,234	1,257,756	300,000
\$68,398	1,257,756	300,000
\$260,000	957,706	300,000
\$81,433	2,535,491	300,000
\$150,000	2,175,855	1,087,924
\$191,750	3,145,790	3,146,420
\$191,750	3,146,347	3,146,420
\$191,750	3,145,996	3,146,420
\$9,964	281,802	161,163
\$16,549	483,015	277,751
\$25,571	292,459	169,024
\$18,234	311,566	179,709
\$18,042	259,903	452,983
\$27,342	355,474	204,034
\$33,840	314,318	182,223
\$90,000	1,075,828	1,075,827
\$180,000	2,321,606	244,432
\$180,000	600,058	3,555,971

E.g. Advertising experiments done by US retailers online

Only 3 out of the 25 experiments have sufficient power to be able to distinguish between a 0% return on investment, and a wildly profitable campaign which has a 50% return on investment over 2 weeks!

Problem is that individual-level sales are incredibly volatile and heterogeneous: the standard deviation of sales is typically ten times the mean over the advertising period. As a result, there is so much noise that it is very difficult for retailers to detect a treatment effect.

# Implication: Important additional knowledge benefit from our impact evaluations

- Only by comparing the experiences of many firms undertaking some new intervention may we be able to learn whether it works or not
  - It might be impossible for any single firm to test and learn from their own experiences whether there is any effect or not

# Of course there are other reasons why firms don't do good things by themselves

- Missing insurance markets
  - Intervention may have high expected return, but lots of risk
- Behavioral reasons
  - Self-control issues or other reasons stop firms from taking privately optimal decisions
- Coordination issues
  - For intervention to succeed, we need lots of parties to all act e.g. setting up new markets, linking together value chains, etc.

Etc.

# Implications for design of evaluations

- If it works
  - Want to try and understand why firms weren't doing it already
  - Why market wasn't already providing it
- If it doesn't work
  - Want to try and understand whether
    - If it could have worked, if firms could have easily done this on their own (i.e. whether there was a good reason the market wasn't already providing this service)
    - Some other market failure or constraint prevents firms from taking advantage of the program