Methods for Measuring Firm Outcomes

David McKenzie
September 2019
Outline

• Why is measurement a problem?
• Why does this matter?
• What can we do about it?
Why is it a problem?

Measuring Firm Outcomes is Difficult

• Many Small Firms Don’t Keep Records
• Others Misreport in their records for tax purposes
• Higher growth/larger firms may consider financial outcomes commercially sensitive and not want to report at all
• Program participants may have incentives to lie about outcomes to show program has benefited them
• Attrition can be a problem
Proportion of firms keeping written records

- Bangladesh
- Chile
- Ghana
- Kenya
- Mexico
- Nigeria
Example 1: Nigeria Business Plan Competition

• Firms received grant payments in 4 tranches, linked to showing progress on employment and sales

• Program officials collected employment data from firms AFTER all payments made

• We then went in as an independent survey team about a month later, and collected employment data also
Employment Reported by Owner in Second Round Survey

Reporting to Survey vs Reporting to Program

75% report more employees to the program than they do to the survey,
11.5% report the same to both, only 13.5% report more in the survey.

Distribution of Difference in Reporting of Employees to Program vs Survey
Why does it matter?

Accurate measurement is critical to our ability to learn whether firm programs are working or not.

- True effect: 10% increase in profits
- Standard deviation of profits is 100, control mean is 300
- Statistical power = 56%
Why does it matter?

Accurate measurement is critical to our ability to learn whether firm programs are working or not.

Statistical power is 38% instead of 56%.

Need to boost sample size by 110 firms to achieve the same power as without the measurement error.

i.e. need to go from 200 firms to 310 firms!!!!
What can we do to improve measurement?

**Surveys:**
- Getting firms to provide data
  - How do we get control group to respond if they don’t get our program?
  - How can we stop “please the experimenter” effects/firms not telling the truth?
- How can we collect more accurate measures?
- What else should we ask about?

**Beyond pure survey measures:**
- Objective information
- Administrative data
- Action-oriented/Task-based approaches

**Agreeing Ex-ante on what measures are key:**
Surveys: Getting firms to provide data

• Common concern of policymakers “How can I get firms that don’t get my program to answer follow-up questions?”

“I did not take part in the face to face interview because I was sent a letter that I wasn’t qualified for grant”

“I did not take part in the face to face interview because for how many years I haven’t been given the grant”

“Respondent said he is tired of being asked questions without getting help”
How much of a problem is this?

• Using 57 rounds of surveys from 21 different impact evaluations in 19 different countries:
  • Average treatment-control response difference is only 3 percentage points
  • Matters more when what they have missed out on is really valuable - moving from a program worth $100 to one worth $10,000 is associated with a 3.4 percentage point increase in the treatment-control response gap
First way of dealing with this – exert more effort

<table>
<thead>
<tr>
<th>Interview Attempt Number</th>
<th>Treatment Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Firms</td>
<td>Number of Firms</td>
</tr>
<tr>
<td>1</td>
<td>176</td>
<td>233</td>
</tr>
<tr>
<td>2</td>
<td>102</td>
<td>185</td>
</tr>
<tr>
<td>3</td>
<td>94</td>
<td>112</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
<td>34</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>11+</td>
<td>19</td>
<td>34</td>
</tr>
<tr>
<td>Web+SMS+CATI</td>
<td>109</td>
<td>169</td>
</tr>
</tbody>
</table>
Solution 2: Separate Survey from Project as Much as Can

• Don’t go in as the government saying you want to see whether the project they took part in benefited them

• Two options:
  • Option 1: include firms as part of other official general purpose surveys taking part in country – e.g. add as booster sample to survey of manufacturing, etc.
    • Advantages can be high response rates, and collected by different agency
    • Firms may still be concerned that gets linked to tax obligations
  • Option 2: come in as independent survey “this survey is being conducted by TNSRMS. Its purpose is to help better understand the growth process of enterprises being run by Nigerian youth, as well as the development process of youths thinking about starting a business. The results of this survey will be kept confidential and will be used for research purposes only. No information will be presented from this research which could be used to identify any individual business.”
How can we collect more accurate measures?

- **Profits and Sales** – *ultimately what the firms themselves care most about*
  - Optimal approach, and recall period depends on size of firm, and whether keep records
  - For smallish firms, triangulation method of Anderson, Lazicky and Zia (2019)
    - **Monthly recall**: recall their total sales or all the money collected into the business during the last month
    - **Estimate based on weekly recall**: the respondent is asked for the best week and worst week of sales in the past month – the software then averages and multiplies by 4.25.
    - **Estimate based on daily recall**: the respondent is asked for last day’s sales, best day’s sales, and worst day’s sales in the last month – the three are then displayed together as anchors, with the respondent then asked for sales on a typical day. This is then multiplied by number of days per week the business transacts with customers, and then by 4.25 weeks per month.
Question 11.

Let me confirm what you have told me about your sales last month:

- Estimate #1: Best Guess = 600,000.00 / month
- Estimate #2: Average of Best and Worst weeks = 1,275,000.00 / month
- Estimate #3: Sum of Typical Daily Sales = 1,041,250.00 / month

Using these 3 numbers to guide you, what is your FINAL BEST GUESS of how much money you got selling your products/services to customers last month? Show your screen to the respondent and ensure that the respondent understands the 3 Monthly Sales estimates based on the information they gave previously. Check that the respondent understands the question is focused on total sales made in THE MOST RECENT MONTH.

1,275,000.00
Watch out for too many zeros in the currency

• Often lots of variation among firms, so one might report 100,000 in sales, and another 1,000,000. But very easy for enumerator to miss a digit.
  • Use natural bigger units where applicable (e.g. lakh or crore in India/Sri Lanka; luca/palo in Chile; thousands/millions perhaps) – danger is that this may be bad at capturing low amounts
  • For key variables:

![Table image]

TOTAL MONTHLY SALES LAST MONTH

- Less than N250,000
- N250,000 - N500,000
- N500,000 - N750,000
- N751,000 - N1,000,000
- N1,000,001 - N1,250,000
- N1,250,001 - N1,500,000
- N1,500,001 - N2,000,000
- N2,000,001 - N3,000,000
- N3,000,001 - N4,000,000
- N4,000,001 - N5,000,000
- N5,000,001 - N7,500,000
- N7,500,001 - N10,000,000
- Above N10,000,000

CHECK: ALSO WRITE THE AMOUNT IN WORDS (E.G. 5.5 MILLION NAIRA)
Measuring Employment

• Social benefit of creating employment depends not only on *how many* jobs created, but *types of jobs* and *who is hired*

• So want to ask about:
  • Characteristics of workers hired – are they youth, women, coming from poor backgrounds etc?
  • What were these workers doing before they were hired – are firms just hiring away workers from competitors vs training up the unemployed?
  • Wages paid, and whether firm is providing training
  • Whether jobs come with benefits, whether full-time/permanent vs temporary
What else should we ask about?

- Intermediate outcomes along the causal chain

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

<table>
<thead>
<tr>
<th>Sample of Firms</th>
<th>Interested in Investment</th>
<th>Try to pitch to investors</th>
<th>Actually pitch</th>
<th>Get Investment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control 1000</td>
<td>250</td>
<td>150</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>Treat 1000</td>
<td>500</td>
<td>300</td>
<td>100</td>
<td>20</td>
</tr>
</tbody>
</table>

Power 1.000 1.000 0.986 0.380
Min N/2 85 174 621 3300
Key intermediate outcomes

• E.g. Tech adoption:
  • Do they know about the technology and how it works?
    • E.g. knowledge questions about how Facebook for businesses works
  • Extensive margin: have they bought/adopted the technology?
    • E.g. have they set up a Facebook webpage for their business
  • Intensive margin: how much are they using the technology?
    • E.g. how frequently are they refreshing content? Are they doing digital advertising?
  • Immediate outcomes of the technology – even if it doesn’t show up yet in total sales, is the business seeing some value from it?
    • E.g. have they seen new customers who have come in because of a promotion on their Facebook?
What else should we ask about?

- **Opportunity cost/next best alternative?**
  - What would the firms be doing if not for our intervention?
  - E.g. Access to SME finance – want to ask specific questions about our new loan product, but then questions about other types of finance – see whether we are substituting/crowding out these.

- **Nature of competition and spillovers**
  - Who are main competitors, and can we measure impacts on them?
  - Who are main customers, and are we benefiting them?

- **Innovation/Externalities**
  - Justification for supporting private firms larger if firms are introducing new products/innovating; and/or if they are creating positive externalities

- **Productivity** – yes, but hard
What else should we ask about?

- How do firms learn whether trying some intervention is worthwhile or not?
  - Measure subjective expectations before intervention
  - Then measure counterfactual beliefs at different points throughout intervention and see how beliefs get updated.
  - Important for getting at question of “if this is so good, why weren’t firms already doing it?”
  - See McKenzie (2016) and Lewis and Rao (2015)
Beyond pure survey measures:

• Objective information
• Administrative data
• Action-oriented/Task-based approaches
Get Physically Verifiable Measures

1) Physical observation by enumerator of whether a business is in existence or not

2) Physical count of the number of workers observed on the business premises at the time of the survey
   - Obvious issues: misses workers at other locations, or other schedules, or those who are sick etc. so undercount – but still a useful check

3) Collect website addresses of firms, and then verify whether they still have a functioning URL
   - Advantage is you can track this over time
   - Need to take care in how this is recorded
Table 8: Impacts on Non-Self-Reported Outcomes

<table>
<thead>
<tr>
<th></th>
<th>New Firms</th>
<th></th>
<th>Existing Firms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Employees</td>
<td>Functioning</td>
<td>Employees</td>
<td>Functioning</td>
</tr>
<tr>
<td></td>
<td>Counted</td>
<td>Website</td>
<td>Counted</td>
<td>Website</td>
</tr>
<tr>
<td>Impact at Third-Follow-up</td>
<td>2.044***</td>
<td>0.040***</td>
<td>1.746***</td>
<td>0.081***</td>
</tr>
<tr>
<td></td>
<td>(0.263)</td>
<td>(0.014)</td>
<td>(0.439)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Long-Term Impact</td>
<td>1.171***</td>
<td>0.023*</td>
<td>1.059***</td>
<td>0.097***</td>
</tr>
<tr>
<td></td>
<td>(0.253)</td>
<td>(0.013)</td>
<td>(0.369)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Sample Size: Third Follow-up</td>
<td>860</td>
<td>1300</td>
<td>398</td>
<td>541</td>
</tr>
<tr>
<td>Sample Size: Long Term Follow-up</td>
<td>908</td>
<td>1300</td>
<td>386</td>
<td>541</td>
</tr>
<tr>
<td>Control Mean: Third Follow-up</td>
<td>1.077</td>
<td>0.047</td>
<td>2.175</td>
<td>0.099</td>
</tr>
<tr>
<td>Control Mean: Long Term Follow-up</td>
<td>1.220</td>
<td>0.045</td>
<td>2.209</td>
<td>0.072</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses, *, **, *** indicate significance at the 10, 5, and 1 percent levels respectively.

Long-Term Follow-up is Five Years after applying, and more than 3 years after all funding has been received.
Example 2: Objective Measures in High-Growth Firms

- Western Balkans Investment Readiness Project
- Firms are Innovative Start-ups
- Concern is attrition – especially that control group may not want to keep giving us information
Solution: Media Buzz

• Number of Media Mentions:

Contracted media intelligence specialist firm Meltwater to collect online media mentions of the firms in our sample over the six month period March 1 to August 31, 2015. Meltwater tracks more than 250,000 global news sources in 190 countries in 25 languages

- pre-program: 13% mentioned in media, can classify as positive, neutral, or negative mention.

- We find firms with higher investment readiness scores from judges more likely to get subsequent media attention

• Measures of social media activity/buzz:

  • Two-thirds had social media presence: Facebook most common – track Facebook likes, twitter followers, etc.

• Inclusion in AngelList – popular website for companies to raise funding

• Caution: majority of firms may not have any media or online presence in some countries.
Example 3: Getting Objective Data from Smaller Firms

• This is really tough – again you can verify whether the business is operating or not.

• But if you want to move beyond self-reporting, need to think creatively.
  • Failed attempt 1: RFID tags
  • Attempt 2: Sri Lanka – random spot visits 15-16 times during a month at different times of day for 35 minutes per time, record sales that took place during visit, then aggregate up to get estimate of monthly revenue (expensive and intrusive – but suggested some underreporting)
Taking Photos to Value Inventories
$R^2 = 0.43$

$T$-value = 43.5
Verified business practices
Linking to Administrative Data

• Want to **plan for this proactively:**
  • Need to get firm tax IDs/other identifiers needed to be able to link to administrative datasets
  • May need to consider telling firms that a condition of applying for the program is that they agree for their outcomes to be linked/monitored for the next 5 years, and use this opt-in to get permission to link

• Example: Colombian Management Experiment
  • Link to Social Security Database to get Data on *Formal Employment*
    • Have to consider how important informal employment is in your context
  • Link to Export Database to get Data on Overseas Sales
  • Link to other surveys taken by governments

• Notes: admin data often messier than you think – and need to be cautious if firms underreporting for tax purposes
Action-oriented/Task-based approaches

• Concern is that survey answers may be just cheap talk – may want to have firms carry out tasks, or do supplementary experiments as way of better understanding mechanisms

• Great example: Cai and Szeidl (QJE, 2017) – business networking in China
  • Want to understand extent to which groups share information.
  • Provided randomly chosen managers with two bits of information: (i) information about government grant; and (ii) information on private saving opportunity
    • Can see extent to which provide rival information (about grant) varies with extent to which your group contains direct competitors – shows barrier to technology diffusion is product market competition
Action-oriented/Task-based approaches

• Can use lab-experiments to explore mechanisms. E.g. Berge et al. on business training in Tanzania
• Interested in whether training changed competitive mindset, and how this varied by gender
• Asked participants general knowledge questions, and paid for correct answers. Asked whether they thought they were better, equal, or worse than typical microfinance client, and then had choice:
  • Compete – and if do better than average, paid 750 TZ per correct answer, but 0 if perform worse than average
  • Don’t compete – fixed rate of 250 TZ per correct answer
  • Find training increased confidence of female entrepreneurs, but still much less willing to compete than male entrepreneurs
Don’t forget to measure costs

• Try to get this broken down into:
  • Line item costs with units and rates, so can think about costs of implementing elsewhere (e.g. with different venue costs, different manpower rates)
  • Fixed vs variable costs – to think of costs as scale-up

• Think also about opportunity costs to firms of participating:
  • Do they close their firms to attend training program?
  • Costs of transport etc. to participate
  • Time costs of dealing with application process, reporting, etc.
Some other things to consider measuring

• **Value Chains and Linkages**
  • Have focused so far on measuring individual firm outcomes, but for many projects here, also want to measure *linkages*
    • Try to measure how rents are distributed throughout value chain, and relative bargaining power of different parties
    • Measure information linkages across firms and new formation of linkages
    • I don’t know much on this – just flagging it for group working on these projects to brainstorm more over.
Agreeing Ex-ante on what measures are key

• Pre-analysis plan/open discussion with key policymakers – specify ex ante what are the key outcomes that the success of program will be judged on

• Otherwise easy ex post for supporters to look and find some result that sounds promising, or detractors to focus on another

• And ask about what
Example: Colombia export program

E.g.

• **Impact on Whether Firms Export in the Year After the Intervention Begins**

• The 190 hours of technical assistance will begin in the second half of 2018. 49 percent of the firms in the Benefits 2 (full intervention group) exported in 2017.

*What would be the minimum effect size you would want to see in order to consider scaling up this program to more firms?*
Conclusions

• Surveys will remain an important tool, but when you can, useful to try and supplement with other sources of information

• Objective/verifiable measures that you can get without the firm having to cooperate are great if they work

• But also need lots of experimentation and innovation here

• Bonus: good opportunity to get additional research paper out of experiments – especially if coordinate across multiple evaluations to draw general lessons