PROBLEM SET 1

Due on Monday, February 5, 2024

I - INSTRUCTIONS

To successfully complete this problem set, please follow these steps:

1. Download this Word document file into your computer.
2. Insert all your answers into this Word document. Guidance here on how to insert non-Word objects such as handwritten work or screenshot images in your answers.
3. **Once your document is complete, please save it as a PDF**. This is important to make sure all your work is preserved in the process of submission to Canvas.
4. Please submit an electronic copy of the PDF and your **replicable Stata or R script** to the Canvas assignment page.

II - IDENTIFICATION

(1) Your information

Your Last Name:	*Boochever*
Your First Name:	*Oscar*

(2) Group Members (please list below the classmates you worked with on this problem set):

n/a

(3) Compliance with Harvard Kennedy School Academic Code[1] (mark with an X below)

	Yes	**No**
I certify that my work in this problem set complies with the Harvard Kennedy School Academic Code	X	

[1] We abide by the Harvard Kennedy School Academic code for all aspects of the course. In terms of problem sets, unless explicitly written otherwise, the norms are the following: You are free (and encouraged) to discuss problem sets with your classmates. However, you must hand in your own unique written work and code in all cases. Any copy/paste of another's work is plagiarism. In other words, you can work with your classmate(s), sitting side-by-side and going through the problem set question-by-question, but you must each type your own answers and your own code. For more details, please see syllabus.

For this problem set, we will be examining the methods used in the following paper:

Fafchamps, Marcel, David McKenzie, Simon Quinn and Christopher Woodruff. 2014. "Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana." Journal of Development Economics, 106: 211-226.

Conceptual Questions (30 points + 8 extra points)

1. Read the paper. Clearly state the primary research question that the authors are trying to answer. What makes this an interesting question? (2 points)

   Does the method of capital investment into a business (cash versus in-kind grants) affect the profitability of the business?

   This is interesting because it has economic policy relevance for governments, as well as individual impact investors looking to have the biggest stimulus effect of their lending.

2. Explain the main finding of the paper (including what the 'flypaper effect' is) using non-technical jargon, as if you were writing a brief policy memo. Provide two versions:

   a. In 3-5 sentences (without copying the article abstract or the text): (2 points)

      In Ghana, investment provided via in-kind donations stick (flypaper effect) and generate significantly higher returns than unrestricted cash transfers. In general, the flypaper effect means that money allocated to a certain area or need tends to stay in that specific area. In this case, while we might expect different investment mechanisms of the same total value to have similar impact, this is not observed in the data, which can be explained by the flypaper effect.

   b. In a tweet, i.e., 280 characters or less: (1 point)

      Small-business owners: ever been told how to spend money from investors? You may be in luck! In Ghana, researchers find that in-kind donations increase profitability more than unrestricted cash transfers. See thread below.

3. Summarize the specific details of the treatment that participating businesses in this country underwent. (2 points)

Firms were randomly sorted into control group (n = 396) and two treatment groups (cash, n = 198; in-kind, n = 198), by stratifying into 16 strata based on gender, industry, capital stock, and "high capture" (constructed via survey responses), ranking based on profits, and then forming matching quadruples. Within these quadruplets, two were assigned control, and one of each treatment type was assigned. Participating businesses in the study received either unrestricted cash grants or in-kind donations, with the specific equipment or materials chosen by the business owner.

The researchers randomly selected when the businesses would receive their grants (by quadruplet), to incentivize continued participation in the study, and for logistic reasons. These grants were framed as prizes for participation.

4. The authors note that their experimental design is very similar to an earlier study in Sri Lanka. What reasons do they give to conduct a separate study, rather than expecting the findings of that research to apply in this context? (2 points)

1. The Ghanaian sample contains more than double the number of firms as were in the Sri Lankan study. This provides additional power to parse out the effects of providing capital via different methods.

2. De Mel et al. (2009a) found a lack of returns to capital in Sri Lankan female-owned microenterprises. In Ghana, and the African context, the situation may be different: there are more female-owned businesses above subsistence level. Thus, studying this in a new context investigates the effect of grants to women specifically in a country "known for its vibrant tradition of female entrepreneurship" (e.g. Hill, 1984). So, this adds to the potential external validity of grant form studies, and also allows the researchers to dive deeper into heterogenous treatment effects.

5. The authors used a randomized control trial because they believed an observational analysis of similar policies would be insufficient. Imagine that another country implemented the investment program without randomizing treatment, and that you were trying to understand the effect of this program on food security in that country. What are **two possible confounders** (omitted variables) that would bias the results from your observational analysis? **Explain the mechanism** of the omitted variable and **use the omitted variable bias formula** to argue whether it would lead to an ***understatement* or *overstatement*** of the true effect. (3 points)

*In this example, I imagine the program as providing households capital in the form of cash or in-kind donations (which in this case would be chosen food items), with the outcome of interest being the **reported food security level of the business-owner's household** (as opposed to*

profits). Without randomization, groups in the control and treatment groups could vary on baseline characteristics, meaning omitting certain variables would bias the results.

*1. **Income** of households receiving the treatment. Households with higher income are typically more food secure to begin with.*

Food_security = a0 + a1Treatment_i + vi

Food_security = b0 + b1Treatment_i + b2Income_i + ei

Income_i = l + dTreatment_i = n_i

*Bias: a1 = b1 + b2*d*

*Treatment increases food security so b1: **positive***

*Income increases food security so b2: **positive***

Suppose* higher income households **more likely to get treatment → **d: positive***

*Since the treatment effect would be positive, and bias is positive, failing to randomize would lead to an **overstatement** of the true effect.*

*2. Household **size.** Bigger households have a greater food requirement, and thus equivalent capital injections would likely improve food security to a lesser (or negative) degree.*

Food_security = a0 + a1Treatment_i + vi

Food_security = b0 + b1Treatment_i + b2Size_i + ei

Size_i = l + dTreatment_i = n_i

*Bias: a1 = b1 + b2*d*

*Treatment increases food security so b1: **positive***

*Size decreases food security so b2: **negative***

Suppose* bigger households **more likely to get treatment → **d: positive***

*Since the treatment effect would be positive, and bias is negative, failing to randomize would lead to an **understatement** of the true effect.*

6. Let Y_{0i} be the business profits in the absence of the grants program and let $D_i = 1$ denote participation in the program. If there were no RCT and individuals were allowed to opt-in to the program, critics might point out that: (3 points)

$$E(Y_{0i}|D_i = 1) < E(Y_{0i}|\ D_i = 0)$$

Explain this equation in words, explain why it is a problem, and give a plausible scenario in which that may be the case.

This equation implies the expected profits of not receiving the treatment (grants) given you receive the grants (opt-in) is lower than the expected profits of not receiving the grants given you in fact do not receive the grants. This is a problem because it means the counterfactual for people who receive the treatment (the world in which they do not receive the treatment) is not equivalent to the people who actually do not receive the treatment. This creates selection bias, and thus any results would not be measuring the true ATE. This may be the case because people who need the grants (ie., lower current or projected profits) would opt-in to receiving the grants. These people are fundamentally different than the people who would not opt-in and might be in a better position financially. Any observed differences in profits between treatment and control may be confounded by pre-existing differences in profitability/financial status.

7. Defining treatment as being assigned to the grants program, what is the difference between the ITT and TOT in this context? Which do the authors report and why do they make this decision? Write an equation (using potential outcomes notation) that shows what the authors are trying to estimate. (3 points)

The ITT would be the effect of being assigned to a treatment group versus control group, whereas the TOT would be the effect of receiving the treatment (actually receiving the cash or in-kind grant).

The authors report the ITT because 9 (2%) of those assigned to treatment did not receive the treatment due to death, refusal, or attrition. Practically, this was very similar to the TOT given compliance was almost 100%.

ITT=E[Yi(Offered treatment) - Yi(Not offered)]

8. At what level do the authors cluster their standard errors of the main results of the paper (if at all)? Briefly note why the authors cluster the standard errors and why this is the appropriate level to cluster at. (2 extra points)

At the firm level: since data are collected from the same firms over multiple time periods, there are correlations in the outcomes within each firm across time. Clustering at the firm level helps to address this issue by adjusting standard errors to reflect the dependency structure within firms. This approach acknowledges that observations within the same firm may be correlated due to unobserved factors or persistent characteristics of the firm.

9. To assess whether treatment was actually randomly assigned, we can examine the results of a balance test, presented in Table 2. Do the results in this table make you more or less confident about the validity of the paper's results? Interpret one of the *p*-values from column (5) for the full sample. (3 points)

The results in the table make me confident in the validity of the paper's results, and it appears randomization was successful. Across all F-test of equality of means' p-values, none are statistically significant. This means the probability of observing the results in the data, if truly equal means, are well above 0.05, and the null hypothesis of equality cannot be rejected.

Specifically, for owner's age, the p-value of 0.429 means that if there were truly no differences in the ages of owners across groups, there is a 42.9% chance of observing the observed differences in means in a sample of that size. Since this is well above 5%, it suggests the differences could have occurred due to chance alone, and there is insufficient evidence to reject the null hypothesis that the age of owners in each group are equal.

10. Attrition in experiments like this one is often a concern for internal validity. Does the particular sort of attrition mentioned in this article give you reason to be concerned about the validity in this study? Describe using particular aspects of the experiment or its implementation. (1 point)

I don't feel concerned about the validity of the study. For one, the authors eliminated firms that closed or refused to answer the round 2 survey before randomization. Through the rest of the waves, 6% of firms closed (equal between treatment and control), only 8% of sample is not present in wave 6, and 11% do not report profit data.

One possible issue to note is that the attrition rates are higher for the control group, but this is explainable by the encouragement (or lack thereof) that receiving grants would foster, and robustness checks reveal that results don't seem driven by attrition.

11. What other threats to internal validity may have affected this experiment? Choose one threat and explain how it might bias the coefficient of interest. (2 points)

I would worry about contamination or spillover effects of the treatment onto control groups. For example, even with the enumeration area design, perhaps regional context is such that business owners have informal communication through social networks, even across gender and sector. In this case, maybe those receiving grants are better able to vocalize what additional capital allows their business in terms of profit-maximizing. If communicating with control group individuals who did not receive the grants, perhaps this knowledge could influence business decisions that affects profits. If so, the coefficient of interest (on receiving the treatment) would be understated as the control group would benefit from the treatment indirectly, thus shrinking the difference in gained profits between treatment and control groups.

12. Describe four specific problems with generalizing the results of this study due to using an RCT. Hint: review Banerjee et al. (2017). (4 points)

1. Location context – the RCT was conducted in Ghana, and the results are specific to that context. There are a number of potential confounding variables (eg., sectors, business ownership by gender, regional variations, initial profits) that would prevent the generalizability of the study to another context – for example, the United States.

2. Randomization or site-selection bias, which means that the individuals who agree to participate in a study may be fundamentally different from the rest of the population. This means, any results from the RCT may not be applicable to populations who would not agree to this kind of intervention. For example, in a government funded program aimed to increase profits, stringently libertarian individuals would not accept government funding, let alone participate in a government experiment.

3. Political conditions – if a program becomes big enough, one possible side effect of political reactions in a different environment could be corrupt government officials seeking to skim some of the funding for business-owners. Alternatively, there is no guarantee for support for the policy in general. These factors inhibit generalizability.

4. Implementation challenges at scale: this RCT demonstrated results at the particular level conducted – there is no guarantee this results would hold at a much larger scale, particularly due to issues of implementation and logistical concerns.

13. List at least two strategies the authors use to address some of the concerns you described above. (2 points)

1. Diverse sample selection: researchers intentionally selected a diverse sample of microenterprises across enumeration areas and sectors. By including businesses with a range of experiences and qualities, they enhanced the external validity of this experiment to multiple types of contexts.

2. Multiple wave study design: instead of only focusing on short-term outcomes, the researchers collected data in multiple waves over time. This allowed them to track changes in profits over time, as well as account for attrition, and alleviates some concern about the randomization bias – with minimal non-compliance and attrition, it appears reasonable that these conditions would hold elsewhere.

14. Why do the authors include Table 4? (2 extra points)

The authors include table 4 because it explains the hetergenous treatment effects of the program by various strata. For example, industry makeup, "capture," capital, and profit level.

15. Do you think the main results would be the same if this experiment were expanded to slightly larger businesses? Describe plausible scenarios in which providing the same transfer to these different businesses could both lead to either (i) a larger increase in profits, or (ii) a smaller increase in profits. (2 extra points)

I think the main results would be less strong at larger businesses, but its plausible that the results could be larger or smaller.

Larger: bigger businesses have a higher level of staffing and experience, and could likely utilize grants more effectively to serve their needs.

Smaller: bigger businesses already have a higher level of profit, so additional capital injection produces decreasing marginal returns

16. If you were a researcher at the World Bank interested in scaling up one or several of these treatments, what follow-up study would you propose to expand on these findings? Explain in 4-5 sentences as if you were trying to convince a policymaker of the need for additional research. (2 extra points)

I would suggest performing this type of study in an additional context – for example, in a Scandinavian country. I think the main question here is the external validity of the Ghana study. The Sri Lanka study, plus the Ghana study, and one final study in a different political and cultural setting, would leave us convinced of the efficacy of this program around the world. Additionally, I'd want to focus the attention on slightly larger businesses and bigger capital injections, to investigate the effectiveness of the program at a larger scale.

Data Analysis Questions (22 points + 2 extra points)

In this part, you will replicate the central results from the paper. You will be asked to analyze the data, present results, and submit a replicable Stata or R script separately.

In the problem set link, we have provided a lightly cleaned version of their main analysis files: `ReplicationDataGhanaJDE.dta`. Keep the file in a subfolder called `data` of your problem set Stata directory or Rstudio Project; this will facilitate our submission verification described at the end of this part. The data we are using comes from the public study page in the World Bank microdata library.

There are several new programming topics that this part will require:

i. Defining your own function
ii. Fixed effects and clustered standard errors
iii. String concatenation and extraction

i. To get started on writing your own function, please go through:

- RStudio Primer for writing functions, and
- An additional screencast on functions tailored to this problem

ii. And to familiarize with fixed effects and clustered standard errors, please make use of our screencasts walking through `lfe::felm()`, which is the function we recommend in the HKS cheat-sheet.

- Fixed effects with `felm`
- Clustered standard errors with `felm`

Please also make use of the Appendix of this problem set for more details about the data.

17. Produce a well-organized descriptive statistics table that includes (i) the **number of households**, (ii) the **number of units of randomization**, alongside (iii) the **sample mean and standard error** of the *real final profit* in the **control group**, and (iv) the same sample mean and standard error of the **same variable but in the treatment group**. In other words, the table should have one row (countries), and five columns (including country). To replicate the paper, remember to filter for Wave 2.

    a. Print your table below. (6 points)

Country	Number of Households	Number of Units of Randomization	Real Final Profit Sample Mean (Treatment)	Real Final Profit Sample SD (Treatment)	Real Final Profit Sample Mean (Control)	Real Final Profit Sample SD (Control)
Ghana	793	195	131.54	312.22	128.12	245.45

    b. Are the differences in baseline profits between the control and treatment groups significant at the 0.05 level? (3 points)

No: by testing the null hypothesis that the means are equivalent, we get a p-value of 0.86, so we cannot reject the null hypothesis.

```
        Welch Two Sample t-test

data:  treatment_data and control_data
t = 0.16994, df = 735.38, p-value = 0.8651
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -36.05666  42.89064
sample estimates:
mean of x mean of y
 131.5369  128.1199
```

18. Reproduce the coefficient estimate and standard error estimates in columns (1-2) of Table 3: these correspond to the main pooled OLS specifications (Equation 5 in the paper) without and with sample trimming respectively. (3 points)

Without trimming:

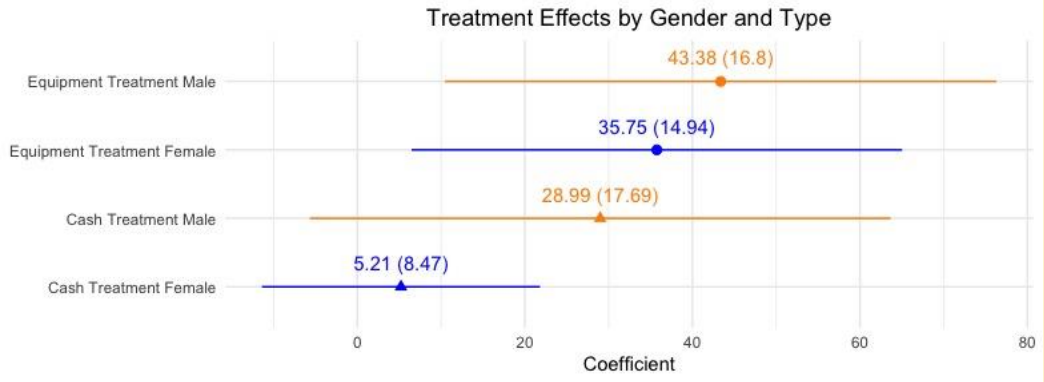
	term	estimate	std.error	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	atreatcash	14.5	8.68	1.67	0.0948
2	atreatequip	38.6	11.2	3.44	0.000578

With trimming:

	term	estimate	std.error	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	atreatcash	9.59	7.32	1.31	0.190
2	atreatequip	36.8	10.7	3.44	0.000578

19. Represent the values of column (5), rows 3 onward in Table 3 as a well-labeled coefficient plot. This corresponds to the main (untrimmed) OLS analysis finding treatment effects by gender and treatment type (Equation 6 in the paper). The figure must: (6 points)

    - Print the rounded coefficients and standard errors next to each point.
    - Be estimated from the regression specification described in the paper. Therefore, all the numbers should match exactly as reported in Table 3.[2]
    - Be clean and well-labeled (i.e., have clear axis labels, no chart junk, understandable to a reader who has not read the paper).



Try implementing the fixed effects specification either for question 18 or 19 above, and report your coefficients as a table or coefficient plot, respectively. (2 extra points)

Table 3 Column 3 with fixed effects, without trimming

term	estimate	std.error	statistic	p.value
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 atreatcash	3.96	13.9	0.285	0.775
2 atreatequip	43.2	12.3	3.51	0.000469

Table 3 Column 4 with fixed effects, with trimming

term	estimate	std.error	statistic	p.value
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 atreatcash	0.482	8.22	0.0587	0.953
2 atreatequip	30.9	10.7	2.88	0.00411

20. Submit your Stata or R script to the Canvas assignment as a separate .do or .R file (or .Rmd file, if you used RMarkdown). (4 points)

[2] Rounded to the third decimal place.

RCTs in Your Own Work (8 points)

21. Propose a specific policy question that could best be answered using an RCT. Explain the question in non-technical terms in no more than 3-5 sentences. Write out the empirical specification you would use. (4 points)

Question: *Can a mentorship program reduce recidivism rates among individuals upon release from prison compared to standard reentry support programs? For a specific prison, or set of prisons, all scheduled releases would be randomly assigned to either the mentorship program plus baseline supports, or just the baseline supports. Basically, the jurisdiction wants to know if a mentorship program is best to minimize recidivism.*

Outcome Variable: *Recidivism rate (measured as the proportion of individuals who reoffend within a specified time frame).*

Empirical Model: $Recidivism\_rate = B\_0 + B\_1 * Treatment\_group + B\_x * control\_variables + \varepsilon$

22. Describe your treatment group. Propose a comparison group and explain why you chose that group. (2 points)

Treatment Group: *The treatment group comprises individuals who participate in a mentorship program designed to support their reintegration into the community upon release from prison. These individuals are paired with mentors who provide guidance, emotional support, and practical assistance in navigating challenges post-release. They also receive the standards supports.*

Comparison Group*: The comparison group consists of individuals who receive the standard reentry support services, but they do not receive personalized mentorship as part of their reentry program.*

Rationale: *By comparing recidivism rates between individuals enrolled in the mentorship program and those receiving standard reentry support services, we can assess the impact of mentorship on reducing the likelihood of reoffending post-release. Critical here is the success of randomization – that individuals receiving the mentorship program are on average similar at baseline to individuals who did not receive mentorship. Confounders to note are sentence type, income, family/relationship status, and education.*

23. Is it possible that the control group could be "contaminated" as a result of interacting with the treatment group? Explain which measures you would take to limit this contamination. (2 points)

Yes – individuals receiving the mentorship could communicate with others not receiving mentorship. This risk seems very plausible given the social network and community built within one particular facility. To mitigate these risks I would provide strict communication guidelines, and consider having participants sign confidentiality agreements. I would also have the mentors play a supervisory role, ensuring physical separation between control and treatment groups.

Appendix for Data Analysis

Packages to install

Code for the data analysis portion will necessarily vary by student, and there are multiple packages available which could be used to construct solutions. However, the following packages have been tested by the teaching staff and are used in the sample solutions:

```
library(haven)
library(broom)
library(lfe)
library(tidyverse)
library(forcats)
library(glue)
library(ggplot2)
```

Documentation embedded in Stata .dta files

The dataset is a Stata .dta file. Stata .dta files often encode the description of each variable in a separate attribute, which are often very useful to understand what each variable represents. When one reads in a .dta file via `haven::read_dta()`, these variable attributes are stored as separate attributes associated with each variable. These can be checked by, for example, applying `str()` to each vector of a variable, but other packages provide convenience functions to more easily tabulate this data. We recommend the `labelled::lookfor()` function.

Variables in Regression

Equation (5) in the paper defines the main regression to be run.