ECON 1190 Problem Set 4: Randomized Control Trials

Claire Duquennois

03/26/2024

Name: Rohan Krishnan

1 Empirical Analysis using Data from Bryan, G., Chowdury, S., Mobarak, A. M. (2014, Econometrica)

This exercise uses data from Bryan, Chowdhury, and Mobarak's paper, "Underinvestment in a Profitable Technology: the Case of Seasonal Migration in Bangladesh," published in *Econometrica* in 2014. This paper studies the effects of seasonal migration on household consumption during the lean season in rural Bangladesh by randomly subsidizing the cost of seasonal migration.

The data can be found by going to Mushfiq Mobarak's Yale faculty page, and then following the link to the data repository page on the Harvard dataverse.

2 Set Up:

2.1 Question: Loading the data - Load any packages you will need and the data contained in the following files `Round1_Controls_Table1.dta` and `Round2.dta`. How many observations are contained in each of these datasets. What is the level of an observation?

Code and Answer:

```
#Load necessary libraries
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(lfe)
```

```
## Loading required package: Matrix
```

```
library(ggplot2)

#Load aej_maindata.dta
r1 <- read_dta("~/Downloads/Round1_Controls_Table1.dta")
r2 <- read_dta("~/Downloads/Round2.dta")

#Check number of observations in each data set
nrow(r1)
```

```
## [1] 1900
```

```
nrow(r2)
```

```
## [1] 1907
```

```
#Check if data frames
is(r1)
```

```
## [1] "tbl_df"     "tbl"        "data.frame" "list"       "oldClass"
## [6] "vector"
```

```
r1 <- data.frame(r1)
is(r1)
```

```
## [1] "data.frame" "list"       "oldClass"   "vector"
```

```
is(r2)
```

```
## [1] "tbl_df"     "tbl"        "data.frame" "list"       "oldClass"
## [6] "vector"
```

```
r2 <- data.frame(r2)
is(r2)
```

```
## [1] "data.frame" "list"       "oldClass"   "vector"
```

There are 1900 observations in the first round data set and 1907 in the second round data set.

2.2 Question: Keep the variables listed below. A description of each variable should appear in the column headers of the loaded data.

For Round 1 data:

Name	Description
incentivized	Is 1 if the household is in either the cash or credit treatment groups, 0 if in the information or control group
q9pdcalq9	Total calories per person per day
exp_total_pc_r1	Total monthly household expenditures per capita
hhmembers_r1	Number of household members
tsaving_hh_r1	Total household savings

For Round 2 data:

Name	Description
incentivized	Is 1 if the household is in either the cash or credit treatment groups, 0 if in the information or control group
average_exp2	Total consumption per person per month in round 2
upazila	Sub-district name
village	Village name
migrant	Member of household migrates this season
total_fish	Total monthly household expenditures per capita on fish

Code:

```
#Select variables
r1.1 <- r1 %>%
  select(incentivized, q9pdcalq9, exp_total_pc_r1, hhmembers_r1, tsaving_hh_r1)

r2.1 <- r2 %>%
  select(incentivized, average_exp2, upazila, village, migrant, total_fish)
```

2.3 Question: Because the effects of the cash and treatment arms are similar and they find no effect of the information treatment, the authors choose to focus much of their analysis on the contrast between the incentivized group (cash and credit) and the not incentivized group (information and control). We will do the same. Regress all the baseline household characteristics still included in the round 1 data on the incentivized indicator. Present your results in a table. What is the equivalent table in the paper?

Code:

```
#Regress incentivized on all other variables in round 1
reg.1 <- felm(q9pdcalq9~ incentivized, data = r1.1)
reg.2 <- felm(exp_total_pc_r1~ incentivized, data = r1.1)
reg.3 <- felm(hhmembers_r1 ~ incentivized, data = r1.1)
reg.4 <- felm(tsaving_hh_r1 ~ incentivized, data = r1.1)

#Present in table
stargazer(reg.1, reg.2, reg.3, reg.4, se = list(reg.1$rse, reg.2$rse, reg.3$rse, reg.4$rse),
          header = FALSE, no.space = TRUE, type = "latex",
         title = "Incentivized group vs not incentivized group")
```

Table 3: Incentivized group vs not incentivized group

	Dependent variable:			
	q9pdcalq9	exp_total_pc_r1	hhmembers_r1	tsaving_hh_r1
	(1)	(2)	(3)	(4)
incentivized	20.247	28.062	−0.069	−160.555
	(26.988)	(30.445)	(0.067)	(137.514)
Constant	2,060.175***	1,036.432***	4.021***	1,515.250***
	(22.323)	(20.675)	(0.057)	(114.401)
Observations	1,893	1,892	1,892	997
R^2	0.0003	0.0003	0.001	0.001
Adjusted R^2	−0.0002	−0.0002	0.0001	0.0004
Residual Std. Error	545.854 (df = 1891)	720.798 (df = 1890)	1.333 (df = 1890)	2,007.345 (df = 995)
Note:				*p<0.1; **p<0.05; ***p<0.01

Answer:

The corresponding table is table 1.

2.4 Question: How should the coefficients in the table above be interpreted? What should we look for in this table?

Answer:

This table is conducting a balance test to see if there is a significant difference between treatment and control. Since none of the variables are statistically significant, we can say that there is no significant difference between treatment and control.

2.5 Question: Using the round 2 data, regress migrant on the treatment indicator. What is the equivalent table in the paper?

Code:

```
#Regress migrant on treatment indicator
reg.5 <- felm(migrant ~ incentivized, data = r2.1)
#Present in table
stargazer(reg.5, se = list(reg.5$rse),
          header = FALSE, no.space = TRUE, type = "latex",
          title = "Migrant regressed on treatment indicator")
```

Table 4: Migrant regressed on treatment indicator

	Dependent variable:
	migrant
incentivized	0.220***
	(0.024)
Constant	0.360***
	(0.020)
Observations	1,871
R^2	0.042
Adjusted R^2	0.042
Residual Std. Error	0.490 (df = 1869)
Note:	*p<0.1; **p<0.05; ***p<0.01

Answer: Table 2 is the equivalent table.

2.6 Question: How should the coefficients in the table above be interpreted? Why is this table important?

Answer: The coefficients should be interpreted as being incentivized predicting a statistically significant 22% increase in the probability of an individual migrating.

2.7 Question: What is the underlying migration rate in the non-incentivized group and how might this change our interpretation of the results?

Answer: The underlying migration rate in the non-incentivized group is 36%. This indicates that there is already some level of migration, indicating that we should more closely examine what types of households are migrating.

2.8 Question: Replicate the (exact) results presented in the third row of the fourth column of table 3. Present your result in a table and interpret this result.

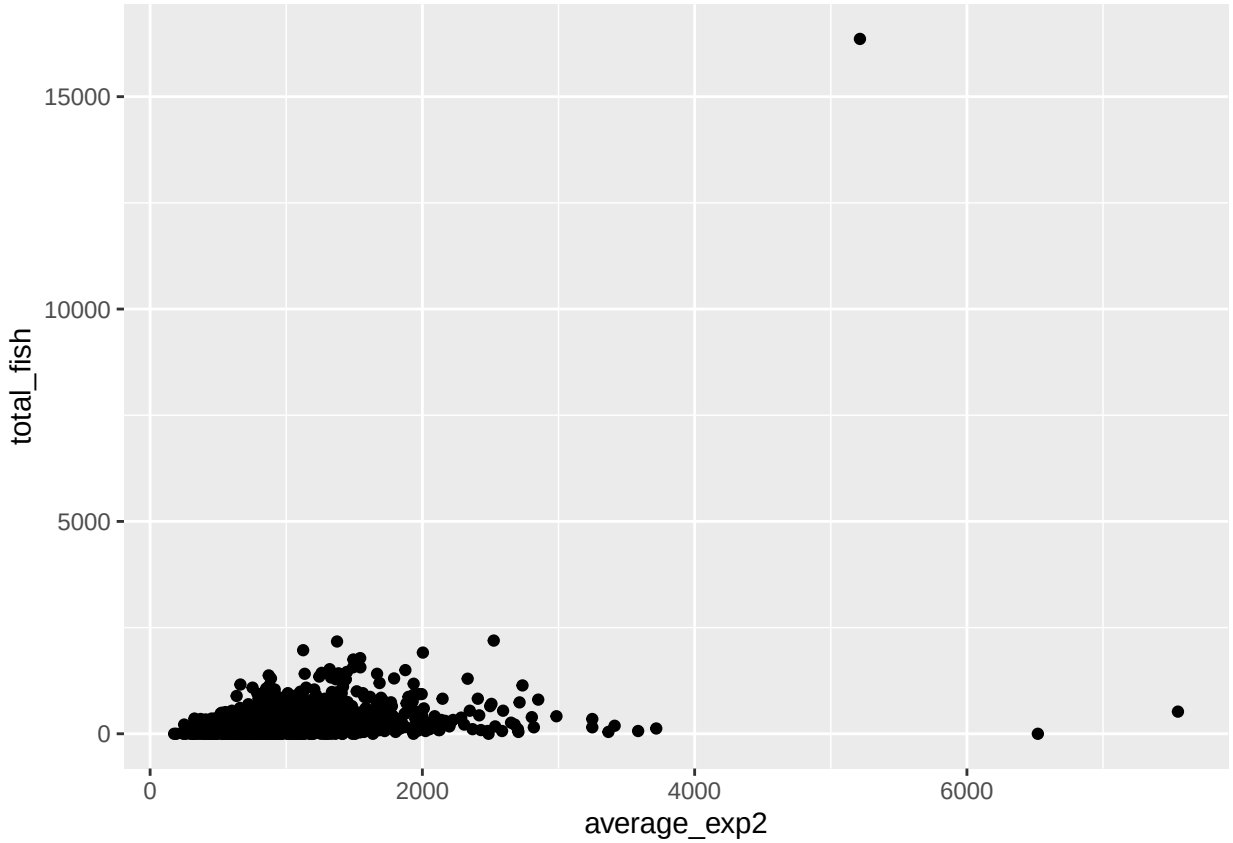
Hint 1: The authors elect to drop one household observation because the reported value of total fish consumed in the household is very high.

Hint 2: To replicate the standard errors in the paper make sure to cluster your standard errors as the authors do.

Code:

```
#Graph to see outlier
r2.1 %>%
  ggplot(aes(x = average_exp2, y = total_fish)) +
  geom_point()
```

```
## Warning: Removed 37 rows containing missing values (`geom_point()`).
```




```
#Find max total_fish value and filter it out of data
max(r2.1$total_fish)
```

[1] 16359.82

```
r2.2 <- r2.1 %>%
  filter(total_fish < 16000)

#Generate regression
reg.6 <- felm(average_exp2 ~ incentivized |upazila|0|village, data = r2.2)
```

```r
#Present in table
stargazer(reg.6, se = list(reg.6$cse),
          header = FALSE, no.space = TRUE, type = "latex",
          title = "Table 3 Row 3 Column 4 Replication")
```

Table 5: Table 3 Row 3 Column 4 Replication

	Dependent variable:
	average_exp2
incentivized	68.359**
	(30.593)
Observations	1,869
R^2	0.044
Adjusted R^2	0.036
Residual Std. Error	452.023 (df = 1854)
Note:	*p<0.1; **p<0.05; ***p<0.01

Answer: Incentivization predicts a statistically significant increase of 68.593 Takas in total consumption per month in round 2 controlling for differences across sub-districts.

2.9 Question: Run the same estimate without fixed effects and present your results in a table. What happens to the coefficient and standard errors? Is this surprising? What does this tell us?

Code:

```
#Generate regression
reg.7 <- felm(average_exp2 ~ incentivized |0|0|village, data = r2.2)

#Present in a table
stargazer(reg.7, se = list(reg.7$rse),
          type = "latex", header = FALSE, no.space = TRUE,
          title = "No Fixed Effect Regression")
```

Table 6: No Fixed Effect Regression

	Dependent variable:
	average_exp2
incentivized	73.620***
	(22.605)
Constant	950.805***
	(18.514)
Observations	1,869
R^2	0.006
Adjusted R^2	0.005
Residual Std. Error	459.341 (df = 1867)
Note:	*p<0.1; **p<0.05; ***p<0.01

Answer: Both the coefficient and standard error increases. This indicates that some villages may be more inclined to migrate than others (perhaps depending on their relative wealth). This is not surprising as the authors mention this as an issue.

2.10 Question: Why is the header of the first five columns of table 3 "ITT". What is meant by this and what does this tell us about how we should interpret these results?

Answer: ITT refers to "intent to treat". The columns examine the relation between being assigned treatment or not (regardless of participation) and each dependent variable. This allows us to compare the average difference between control and treated.

2.11 Question: We are interested in estimating how migration affects total expenditures for the households that were induced to migrate by the cash and credit treatments as follows,

$$TotExp_{ivj} = \alpha + \beta_1 Migrate_{ivj} + \varphi_j + \nu_{ivj}$$

where $Migrate_{ivj}$ is dummy indicator for if a member of household i in village v in subdistrict j migrated, and φ_j are the sub district fixed effects. However it is not possible to identify in the data which households were induced by the treatment vs those who would have migrated either way. Furthermore, there is likely substantial selection between the households that select into migration versus those that do not. Propose a source of exogenous variation that can be used as an instrument to isolate "good" exogenous variation in migration.

Answer: A good instrument could be the random assignment of being placed into the incentivized group. As this is a random allocation, there is no selection.

2.12 Question: What is the first stage specification?

Answer: $Migration = \beta_0 + \beta_1 Incentivized + \epsilon$

2.13 Question: Estimate the first stage and check that you have a strong instrument for migration.

Note: The first stage results reported in the paper appendix may differ slightly as explained in the table footnote.

Code:

```
#Generate first stage regression
reg.fs <- felm(migrant ~ incentivized, data = r2.2)

#Present in table
stargazer(reg.fs, se = list(reg.fs$rse),
          type = "latex", header = FALSE, no.space = TRUE,
          title = "First Stage Regression")
```

Table 7: First Stage Regression

	Dependent variable:
	migrant
incentivized	0.220***
	(0.024)
Constant	0.360***
	(0.020)
Observations	1,870
R^2	0.042
Adjusted R^2	0.041
Residual Std. Error	0.490 (df = 1868)
Note:	*p<0.1; **p<0.05; ***p<0.01

Answer: The p-value is highly significant at the 99%, meaning we can reject the null that there is no relationship.

2.14 Question: Use your instrument to estimate the LATE (Local Average Treatment Effect), the impact of migration on total consumption for those induced to migrate by the treatment, as in columns 6 of table 3 in the paper. Interpret your results.

Note: If you just use Incentivized as your instrument, your estimates will not be exactly the same. If you wish to replicate the paper's coefficients exactly, you will need to use multiple instruments, one for each treatment arm.

Code:

```
#Generate regression
reg.iv <- felm(average_exp2 ~ 1|upazila|(migrant ~ incentivized)|village, data = r2.2)

#Present in table
stargazer(reg.iv, se = list(reg.iv$cse),
          type = "latex", header = FALSE, no.space = FALSE,
          title = "LATE Using Instrumental Variable")
```

Table 8: LATE Using Instrumental Variable

	Dependent variable:
	average_exp2
'migrant(fit)'	374.857**
	(163.808)
Observations	1,867
R^2	−0.009
Adjusted R^2	−0.017
Residual Std. Error	464.584 (df = 1852)
Note:	*p<0.1; **p<0.05; ***p<0.01

Answer: Using incentivized as an instrument for migration and including fixed effects for sub-district and standard error clustering by village, the coefficient of migration is 374.9. This indicates that a household migrant statistically significantly predicts a 374.9 increase in total consumption.

2.15 Question: Why is this result different from the result in columns 4?

Answer: Our results looks at those who actually migrated (meaning we scale our IV to adjust for non-compliance) whereas column 4 only looks at those who were incentivized.

2.16 Question: Why is this value particularly relevant for policy decisions in the context of this experiment.

Answer: It appears that migration has a significant predictive increase in a households consumption, which opens avenues for policy decisions encouraging migration to boost local economies.

2.17 Question: Suppose a policy maker found these results so compelling that they decided to make this a national policy. How would general equilibrium effects potentially change the impacts of this policy if it was implemented in a very large scale way?

Answer: This experiment only has 1900 observations, of which only a fraction migrated. If there was large scale migration, there would likely be larger equilibrium shifts in the local economy including lower wages because of increased competition.

2.18 Question: One major concern that is often brought up in discussions about RCT's is the problem of external validity. It is not always clear how informative the findings from a small scale research project in one context are for policy makers working on a different scale and in different contexts. What are your thoughts on the external validity of this particular project and RCT's in general?

Answer: I think RCTs in isolation provide little benefit for policymakers. However, as a large amount of studies are done on a topic, their collective results could be useful for policy decisions.

3 Submission instructions:

1) Knit your assignment in PDF (It should be 20 pages long).
2) Make sure you have ONE question and answer per page (this allows gradescope to easily find your answers).
3) Upload your assignment PDF to gradescope.