# ECON 1190 Problem Set 4: Randomized Control Trials

Claire Duquennois

**Name:**

## 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:**

```
library(haven)
```

```
## Warning: package 'haven' was built under R version 4.1.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2

##
## 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 (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(lfe)
```

```
## Warning: package 'lfe' was built under R version 4.1.2

## Loading required package: Matrix
```

```
datar1<-read_dta("Round1_Controls_Table1.dta")
datar1<-as.data.frame(datar1)
nrow(datar1)
```

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

```
datar2<-read_dta("Round2.dta")
datar2<-as.data.frame(datar2)
nrow(datar2)
```

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

The data for round one contains 1900 observations and the data for round two contains 1907 observations. In both datasets, the observations are at the household level. The second dataset contains more observations because of split-off households.

## 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:**

```
datar1<-datar1[, c("incentivized","exp_total_pc_r1","q9pdcalq9",
                   "hhmembers_r1","tsaving_hh_r1", "village")]

datar2<-datar2[, c("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 incentivised 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:**

```
balangereg1<-felm(exp_total_pc_r1~incentivized|0|0|village, datar1)
balangereg2<-felm(q9pdcalq9~incentivized|0|0|village, datar1)
balangereg3<-felm(hhmembers_r1~incentivized|0|0|village, datar1)
balangereg4<-felm(tsaving_hh_r1~incentivized|0|0|village, datar1)

stargazer(balangereg1,balangereg2,balangereg3,balangereg4, type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Mar 26, 2024 - 4:37:37 PM

Table 3:

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | exp_total_pc_r1 | q9pdcalq9 | hhmembers_r1 | tsaving_hh_r1 |
| | (1) | (2) | (3) | (4) |
| incentivized | 28.062 | 20.247 | −0.069 | −160.555 |
| | (38.287) | (36.990) | (0.096) | (140.093) |
| Constant | 1,036.432*** | 2,060.175*** | 4.021*** | 1,515.250*** |
| | (28.132) | (31.204) | (0.081) | (113.855) |
| Observations | 1,892 | 1,893 | 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 | 720.798 (df = 1890) | 545.854 (df = 1891) | 1.333 (df = 1890) | 2,007.345 (df = 995) |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.01 |

**Answer:** The results presented in this table are equivalent to those presented in table 1 of the paper.

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

**Answer:**

The coefficient on the constant gives the mean value of that variable for the omitted group, which in this case is the non-incentivized households. The other coefficient gives the mean difference between household in an incentivized treatment arm as compared to the non-incentivized households. Looking at total household consumption for instance, we see that the mean for the non-incentivized group is 1036.432 (Takkas per person per month, I believe) while the mean for the incentivized group is 1036.432+28.062=1064.494. This information is equivalent to the information presented in the fifth column of table 1, just presented in a different way. As with table 1, we are interested in seeing if the randomization of households between the incentivized and non-incentivized groups was done correctly. We want to make sure that the incentivized and non-incentivized group look similar in terms of their baseline characteristics. Since almost all the estimated coefficients are small and not statistically significant, the incentivized and non-incentivized group are balanced on observables.

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

**Code:**

```
regtakeup1<-felm(migrant~incentivized, datar2)


stargazer(regtakeup1,  type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Mar 26, 2024 - 4:37:37 PM

Table 4:

| | *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:** These results are equivalent to those reported in row 1 of table 2.

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

**Answer:** The intercept coefficient gives us the migration rate for households in the non-incentivized group, which is 36%. The other coefficient then tell us the difference between the non-incentivized group and the incentivized group. Thus we can see that those in the incentivized treatment arms (cash and credit) are much more likely to migrate, by 22 percentage points. Knowing these values is important because it allows us to think carefully about how much the treatment is actually changing behaviors: there are many households in the incentivized arms that do not take up the treatment (never takers) and there are also many households in the non-incentivized group who get treated (always takers), thus it will be important for us to distinguish between the average treatment effects (ATE) versus the treatment on the treated (TOT).

## 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 fact that 36% of the non-incentivized group migrates is significant. It suggest that this migration "technology" is not novel and is a well known economic opportunity. It is thus not surprising that the information treatment had no significant effect. This also means that the people who are moved to migrate by the incentive are marginal migrants: people for whom the benefits to migrating are probably particularly low since they would not have chosen to incur the migration cost on their own.

## 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:**

```
#finding the outlier observation:
summary(datar2$total_fish)
```

```
Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.00    86.91   173.81   256.47   325.89 16359.82
```

```
datar2noout<-datar2[datar2$total_fish<16350, ]

regconscol4<-felm(average_exp2~incentivized|upazila|0|village, datar2noout)

stargazer(regconscol4, type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Mar 26, 2024 - 4:37:37 PM

Table 5:

| | *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:**

Among households who were in an incentivized treatment arm, total consumption increased by about 98 Takka per household member compared to the non-incentivized group.

## 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:**

```
regconscol5<-felm(average_exp2~incentivized|0|0|village, datar2noout)

stargazer(regconscol4, regconscol5, type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Mar 26, 2024 - 4:37:37 PM

Table 6:

| | *Dependent variable:* | |
|---|---|---|
| | average_exp2 | |
| | (1) | (2) |
| incentivized | 68.359** | 73.620** |
| | (30.593) | (33.520) |
| Constant | | 950.805*** |
| | | (28.678) |
| Observations | 1,869 | 1,869 |
| $R^2$ | 0.044 | 0.006 |
| Adjusted $R^2$ | 0.036 | 0.005 |
| Residual Std. Error | 452.023 (df = 1854) | 459.341 (df = 1867) |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

**Answer:** The coefficient does change a bit but they are not statistically different from the estimates without the fixed effects . This is consistent with these results being from a randomized control trial. Because of the randomization, we do not need to be concerned with bias from omitted variables. Adding these controls should not change the coefficient.

### 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 stands for “Intent to Treat” since these are intent-to-treat estimates. These estimates tell us the average effect on all households that were randomized into that treatment arm, whether or not the household actually took up treatment and sent a migrant to the city. Based on the take up information we looked at earlier, this average reflect the outcome of several different types of households: a) never taker households (who’s outcome is unchanged by treatment), b) always taker households (who’s outcome is unchanged by treatment) and c) compliers (who are moved to migrate by treatment). Since, through randomization, our control group also includes these same types of households, any difference we observe is entirely driven by the difference in outcomes for the compliers, which are about 22% of the households. Thus these households must experience a very large treatment effect to generate such a large ITT treatment effect.

### 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 $\varphi_j$ are the subdistrict 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:** We exogenous variation generated by the randomization from the RCT as an instrument for Migration.

### 2.12 Question: What is the first stage specification?

**Answer:**

$$Migrant_{ivj} = \rho_0 + \rho_1 Incentivized_{ivj} + \varphi_j + \epsilon_{ivj}$$

where $Incentivized_{ivj}$ is an indicator for being in an incentivized treatment arms and the other variables are as defined above.

## 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:**

```
regfs1<-felm(migrant~incentivized|upazila|0|village, datar2noout)


stargazer(regfs1, type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Mar 26, 2024 - 4:37:37 PM

Table 7:

| | *Dependent variable:* |
|---|---|
| | migrant |
| incentivized | 0.183*** |
| | (0.030) |
| Observations | 1,869 |
| $R^2$ | 0.085 |
| Adjusted $R^2$ | 0.078 |
| Residual Std. Error | 0.480 (df = 1854) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

```
summary(regfs1)$fstat
```

[1] 12.36346

**Answer:** The f stat for both specifications is greater than 10 which passes the benchmark for a sufficiently strong instrument. Furthermore, because we know that this variation is explicitly random, violations of the exclusion restriction are less of a concern here, unless there is reason to believe that receiving a certain treatment may have affected consumption in a way other than through the migration mechanism.

### 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:**

```
regiv1<-felm(average_exp2~1|upazila|(migrant~incentivized)|village, datar2noout)



stargazer(regiv1, type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Mar 26, 2024 - 4:37:38 PM

Table 8:

| | *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:**

Total consumption in households that were induced to migrate by these treatments increased by about 374 takka. This is substantial since the mean of total consumption is about 1000 takka.

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

**Answer:** This value tell us the LATE, the estimated effect on the households who are actually affected by the treatment. The result in column 4 tell us the average effect of being in an incentivized village, but many households do not change their migration decision, so the effect is diluted by the non-compliance of never takers.

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

**Answer:** In this experiment, the LATE is a very relevant value if we think about costs and benefits. For this policy, the policymaker only needs to incur the cost of the policy for individuals who choose to migrate. This will be the always takers and the compliers. The benefits of the policy will be the increased consumption by the compliers. Thus the values we just calculated will be key to determining if the policy is cost effective.

## 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:** One critique of randomized control trials is that they usually tend to be fairly small scale experiments and thus make it difficult to estimate how the proposed policy might actually play out if it was implemented at scale. In this particular example there are many general equilibrium effects that would be of concern. If this policy was widely adopted, leading to a substantial increase in seasonal migration, we might expect to see important price adjustments that could substantially shift the returns experienced by migrants as well as the sending and receiving communities. Wages in the sending and receiving areas are of particular concern, and the prices of certain consumption goods could adjust as well. If migrant wages in the city, or the risk of unemployment in the city, changes substantially due to a large influx of additional migrants, the returns to migrating could easily be reduced to zero. In the sending villages, wages could increase due to the scarcity of labor. The authors explore there general equilibrium wage effects in a subsequent project.

### 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:**

No "correct" answer. We would like to see your opinion on this issue.

# 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.