Research Replication: Migration

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

This project 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.

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

Set Up:

Code and Answer:

```
library(haven)
library(dplyr)
library(ggplot2)
library(lfe)
library(stargazer)

R2<- read_dta('C:\\Users\\roryq\\Downloads\\Round2.dta')
R1<- read_dta('C:\\Users\\roryq\\Downloads\\Round1_Controls_Table1.dta')
obs<-nrow(R1)
```

Each observation is a household. There are 1900 observations in the data set.

Code:

```
# I would like to include total food and non food to replicate column 6 in table 3 exactly

r1<- R1 %>% select(incentivized,q9pdcalq9,exp_total_pc_r1,hhmembers_r1,tsaving_hh_r1,village,hhid)

r2<- R2 %>% select(incentivized, average_exp2,upazila,village,migrant,total_fish,hhid,total_food,total_n

r<- merged_df <- merge(r1, r2, by = c("hhid", "village", "incentivized"))
```

Because the effects of the cash and credit 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. Cluster standard errors by village. This is equivalent to Table 1, the difference column.

Code:

```
reg_q9pdcalq9 <- felm(q9pdcalq9 ~ incentivized|0|0|village, data = r1)
reg_exp_total_pc_r1 <- felm(exp_total_pc_r1 ~ incentivized |0|0|village, data = r1)
reg_hhmembers_r1 <- felm(hhmembers_r1 ~ incentivized |0|0|village, data = r1)
reg_tsaving_hh_r1 <- felm(tsaving_hh_r1 ~ incentivized |0|0|village, data = r1)

stargazer(reg_q9pdcalq9, reg_exp_total_pc_r1, reg_hhmembers_r1, reg_tsaving_hh_r1,
  type = "latex"
  ,
  covariate.labels = c("Difference"
  ),
  dep.var.labels = c(
    "Cal/Person/Day",    # Dependent variable for reg_q9pdcalq9
    "Monthly HH Expenditures", # Dependent variable for reg_exp_total_pc_r1
    "HH Size",               # Dependent variable for reg_hhmembers_r1
    "HH Savings"             # Dependent variable for reg_tsaving_hh_r1
  ),digits = 2, omit.stat=c( "ser"))
)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Dec 06, 2024 - 7:00:46 PM

Table 3:

	Cal/Person/Day	Monthly HH Expenditures	HH Size	HH Savings
	Dependent variable:			
	(1)	(2)	(3)	(4)
Difference	20.25	28.06	−0.07	−160.56
	(36.99)	(38.29)	(0.10)	(140.09)
Constant	2,060.18***	1,036.43***	4.02***	1,515.25***
	(31.20)	(28.13)	(0.08)	(113.85)
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
Note:				*p<0.1; **p<0.05; ***p<0.01

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

Being incentivized (cash or credit) is associated with an additional 20.25 calories per day per person.

Being incentivized (cash or credit) is associated with an additional 28.06 increase in total monthly household expenditures per capita.

Being incentivized (cash or credit) is associated with a 160.55 decrease in total household savings.

Being incentivized is associated with a decrease in house size by.07, this particular interpretation doesnt make much senese, however it is not significantly different than zero, so it appears that household size does not change with incentivizatation.

In this table we should look at the sign, size and significance of these relationships. For all three relationships, none of them are significant at any conventional significance levels.

Using the round 2 data, regress migrant on the incentivized treatment indicator. The equivalent table in the paper is Table V row 2, column 2

Code:

```
model<- felm(migrant~incentivized, r2)
stargazer(model)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Dec 06, 2024 - 7:00:46 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

How should the coefficients in the table above be interpreted? Why is this table important? Is the constant meaningful?

Being incentivized is associated with a .192 increase in the likelihood of having a household member migrate (seasonally). This table is important because it shows that is strong relationship with being offered treatment and migration, so that later we can use incentivation as an IV (it is significant at the .01 confidence level).

the constant shows that without incentivization there is still a .583 likelihood that a household will have a seasonal migrant, it is meaningfull because it shows that there is already a presence of migration before the experiment is run.

What is the underlying migration rate in the non-incentivized group? How might this change our interpretation of the results for the information treatment arm? Whose decision is impacted by the intervention?

The underlying migration rate is reflected by the constant of the previous regression approximately 58% migration rate of those that werent incentivized. because the migration rate is so high initial we have to understand that there are a significant amount of always takers in our sample. In control we have to worry about effects because these people are choosing to migrate probably because they are better off when they do. Only those who are sometimes takers in the treatment group are affected by the intervention.

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. Drop a single outlier due to fish consumed, and cluster standard error by village.

Code:

```
max<- max(R2$total_fish)



df <- subset(R2, total_fish != max)

summary(df$total_fish)
```

Min. 1st Qu. Median Mean 3rd Qu. Max. 0.0 86.9 173.8 248.0 325.9 2194.3

```
model <- felm(average_exp2~incentivized | upazila | 0 | village, data = df)


# Generate the table with clustered standard errors
stargazer(model, se = list(model$rse), type = "latex")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Dec 06, 2024 - 7:00:46 PM

<div align="center">

Table 5:

	Dependent variable:
	average_exp2
incentivized	68.359***
	(23.991)
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

</div>

Answer:

Given that a person was incentivized, controlling for upazila, it is associated with a total consumption per person per month increase of 68.36.

This result is significant to the .01 confidence level and a moderately high increase compared to base line consumption.

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:

```
model1 <- felm(average_exp2~incentivized | 0 | 0 | village, data = df)

stargazer(model,model1)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Dec 06, 2024 - 7:00:46 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:

If you run the same regression without controlling for upazilas, then the consumption coefficient increases overall, but so does the standard error, and it becomes significant at the .05 level not the .01 level. This tells us that there are underlying consumption patterns in the upazilas, increaseing the standard errors, and some explanatory power of the upazilas is conflated with consumption in the second regression causing the coefficient to be higher.

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 is the header because each of those three columns is a treatment, cash, credit, or information. Therefore intention to treat (ITT) can be separated by each treatment. ITT shows us the effect getting the offer, where as the late shows the effect of actually taking the offer. When we interpret ITT we should remember that this shows the effect getting an offer has on the outcome variable. The ITT is the reduced form, so you can not draw casual inference from it.

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 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 will use the offer (incentives) as an instrumental variable to use in a 2sls to get at causality. Because who is given the offer is random, the exclusion restriction holds, even though not everybody who is incentivized takes the offer, from the below regression we can see that there is a relationship between this and migration, satisfying the first stage.

What is the first stage specification?

Answer:

$$Migrant_{ivj} = \lambda + \rho Z_v + \gamma X_{ivj} + \phi_j + \epsilon_{ivj}$$

Where Z is a set of instruments of incentiviazation, and X is a vector of household characteristics. In our regression we will only being using incentivized (binary variable referencing all treatment arms) regressed on migrant as the first stage for simplicity.

Estimate the first stage and check that we have a strong (not weak) instrument for migration.

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

Code:

```
model<- felm(migrant~incentivized|upazila|0|village, data=r2)

# Check if weak instrument with F stat

x<-summary(model)$fstat[1]

stargazer(model)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Dec 06, 2024 - 7:00:46 PM

Table 7:

	Dependent variable:
	migrant
incentivized	0.184***
	(0.030)
Observations	1,870
R^2	0.086
Adjusted R^2	0.079
Residual Std. Error	0.480 (df = 1855)

Note:	*p<0.1; **p<0.05; ***p<0.01

Answer:

The F statistic for this model is 12.4005567 which is larger than 10. The rule of thumb is that if your f statistics is larger than 10 then you do not have a weak instrument, therefore this is an acceptable instrument to use.

Use our 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, r2)
 stargazer(regiv1, type = "latex")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Fri, Dec 06, 2024 - 7:00:46 PM

<div align="center">

Table 8:

	Dependent variable:
	average_exp2
'migrant(fit)'	332.359*
	(171.813)
Observations	1,868
R^2	0.011
Adjusted R^2	0.004
Residual Std. Error	469.930 (df = 1853)
Note:	*p<0.1; **p<0.05; ***p<0.01

</div>

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.

Why is this result different from the result in columns 4?

Answer:

This results are different because column 4 is the ITT effect and column 6 is the LATE effect. Meaning that column 6 is our estimation of the effect of taking the offer/incentive (to migrate), where as column 4 was just the effect of being given the offer/incentive (to migrate).

Why is this value particularly relevant for policy decisions about the cost effectivness of the treatment in the context of this experiment.

Answer:

This value is relevant when evaluating policy decisions, because it gives a causal estimate on how taking the offer will affect participants if the policy were implemented. You could compare the amount of consumption increase with the cost of implementing the policy, to see if the programs causal effects, outweigh the cost of implementing it.

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:

if this were a national policy, and made migration mandatory, this could over saturate the job market in the urban industry, and could drive down wages during the migration period, or lead to unemployment, as there could be more migrants than jobs available.

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 dont think that there is much external validity in this paper, and that attempts to scale it will not be fruitful, or as fruitful as this study. For one since this study has been published there have been numerous studies that are similar (from the extra readings on canvas) and have scaled it, and the general concensus from these studies is that it doesnt have significant impacts on the participants.