ECMA35550 HW2

Mia

2022-11-02

Part I. Randomization and Balance Check

1.1 Research Question Summary

This paper analyzed the impact of a prototypical adolescent empowerment program, as well as impact of financial incentive with cooking oil conditional on stay unmarried on reducing child marriage and increasing girls enrollment in school in rural Bangladesh, where large gains have been achieved in female education and employment while the rate of child marriage changed little. Specifically, the paper analyzed the effect based on a 10-year clustered randomization trail in 460 communities in rural Bangladesh, and tested the impact on untreated women living near treated villages as well. Further, the paper researched on whether changes in bride characteristics in experiment areas lead to different outcomes in marriage market in terms of price(dowry and denmeher) and match(husband quality)

1.2 Clustered Randomization

```
# Set seed
set.seed(35550)
# Read dataset
data <- read.csv("main.csv")
# Clean data
data <- data %>%
  filter(still_in_school != "") %>% # Remove 3 rows with empty school information
  filter(highest_class_passed != "Hafezi/Religious education" &
           highest_class_passed !=  "Other") %>%
  mutate(memberID = with_options( # Change the format of memberID
    c(scipen = 999),
    str_pad(memberID, 13, pad = 0))
    ) %>%
  mutate(highest_class_passed = ifelse(highest_class_passed == "HSC/ Equivalent",
      12,
      ifelse(highest_class_passed == "HSC/First year",
             11,
             ifelse(highest_class_passed == "SSC/Equivalent",
                    10,
                    ifelse(
                      highest_class_passed == "No Class Passed",
                      0,
                      str_sub(highest_class_passed, 7)
                    )
```

```
        )
      )
    )
  ) %>%
  mutate(highest_class_passed =
           as.numeric(words_to_numbers(highest_class_passed))
         ) %>%
  arrange(memberID)
count(data,marital_status)
```

```
##          marital_status     n
## 1     Currently married  3947
## 2              Divorced    15
## 3 Engaged to be married   172
## 4         Never married 19287
## 5             Separated    31
## 6               Widowed     7
```

```
data <- data %>%
  mutate(ever_married_dummy = ifelse(marital_status %in% c("Never married",
                                                           "Engaged to be married"),
                             0, #ever married=0 if they never married
                             1), #otherwise 1
         still_in_school_dummy = ifelse(still_in_school == "Yes",
                                        1,
                                        0)
         )
```

Note

- For the highest class passed column, since it's hard to define the exact education year they received with *Religious education*, and *Other*, we just remove those data to avoid ambiguity.
- Some data in this dataset has a memberID in 12 digits, while the expected digit should be 13. It's likely caused by that when loading numeric data with leading zero from csv file, R will automatically remove those leadign zeros. Therefore, we need to add those zeros back.
- Based on Bangladesh's education system, SSC is usually taken when studnets are on the 10th grade, and HSC is continued education following SSC, which takes 2 years in total. Therefore, we changed the highest class passed column with corresponding numeric data instead for Balance chaeck purpose.
- For the *Ever married* column, although there are some people indicated that they're *Divorced/Separated* or *Widowed*, based on the definition of UN, they're still considered as *ever married*; Similarly, *Engaged to be married* is still unmarried so we classify those data as *not married ever* as well. Another way is to remove those data to avoid unobserved differences in baseline, as people who engaged early probably have different belief in social norms and with different socioeconomic background.
  Reference

```
# Define cluster by village ID
cluster <- with(data, villageID)
Z_clust <- cluster_ra(clusters = cluster,
                      m_each = c(115, 115, 115, 115), # 1:1:1:1 ratio
                      conditions = c("Incentive", "Empowerment", # 4 arms
```

```
                                    "Incen*Empow", "Control")
                      )
# View Cluster Randomization Result
cluster_table <- as.data.frame(table(Z_clust, cluster))
cluster_table %>%
  group_by(cluster) %>%
  summarise(count = sum(Freq)) %>%
  arrange(desc(count))
```

```
## # A tibble: 460 x 2
##    cluster count
##    <fct>   <int>
##  1 9561559   154
##  2 9511206   149
##  3 1873409   143
##  4 1836011   138
##  5 3859560   137
##  6 9561104   137
##  7 1873554   134
##  8 1858242   132
##  9 3859038   129
## 10 6911156   129
## # ... with 450 more rows
```

Followed by the paper, we randomize by community at a ratio of 1:1:1:1. In this dataset, *villageID* represents the community they located in so we use this variable as clsuter.

1.3 Clustered Randomization: Balance Check

```
# Add randomization result to original data
cluster_table <- cluster_table %>%
  filter(Freq != 0) %>% # Keep data with treatment assignment status only
  select(-Freq)
data$villageID_fct <- as.factor(data$villageID) # For merging purpose
# Merge
merged_cluster <- left_join(data, cluster_table,
                            by = c("villageID_fct" = "cluster")) %>%
  rename(treatment = Z_clust)
# Balance Check
balance_cluster <- merged_cluster %>%
  group_by(treatment) %>%
  summarise(ever_married_mean = (sum(ever_married_dummy) / n()) * 100,
            # Calculate pct of women ever married in each group
            ever_married_S.D. = (sd(ever_married_dummy)) * 100,
            still_in_school_mean = (sum(still_in_school_dummy) / n()) * 100,
            # Still in school percent by group
            still_in_school_S.D. = (sd(still_in_school_dummy)) * 100,
            highest_class_passed_mean = sum(highest_class_passed) / n(),
            # Highest class apssed by group
            highest_class_passed_S.D. = sd(highest_class_passed)
            ) %>%
```

Table 1: Balance Table

	Incentive		Empowerment		Incen*Empow		Control
	Incentive	Incentive_Diff	Empow	Empowt_Diff	Incen*Empow	Incen*Empow_Diff	Control
Ever married (%)							
ever_married_mean	16.53	-0.30	17.67	0.84	17.21	0.37	16.83
ever_married_S.D.	37.15		38.15		37.75		37.42
Still in school (%)							
still_in_school_mean	50.67	-0.65	49.81	-1.51	49.81	-1.51	51.32
still_in_school_S.D.	50.00		50.00		50.00		49.99
Highes education passed							
highest_class_passed_mean	6.34	-0.17	6.37	-0.13	6.41	-0.10	6.51
highest_class_passed_S.D.	2.97		2.86		2.88		2.81

```
  transpose_df()
# Calculate the difference in mena between each groupp and the Control group
diff <- balance_cluster %>%
  mutate_at(vars(-matches("name")), list(Diff = ~ . - Control)) %>%
  select(name, ends_with("Diff"), -Control_Diff) %>%
  filter(grepl("mean", name))
# Merge dataset
balance_cluster <- left_join(balance_cluster, diff, by = "name")
```

```
# Table
balance_cluster %>%
  select(
    name, Incentive, Incentive_Diff, Empowerment,
    Empowerment_Diff, `Incen*Empow`, `Incen*Empow_Diff`, Control
  ) %>%
  kable(
    col.names = c(
      "", "Incentive", "Incentive_Diff", "Empow",
      "Empowt_Diff", "Incen*Empow", "Incen*Empow_Diff", "Control"
    ),
    caption = "Balance Table",
    digits = 2, format.args = list(scientific = FALSE)
  ) %>%
  kable_classic() %>%
  kable_styling(
    position = "center", font_size = 8,
    full_width = T, html_font = "Cambria",
    c("stripend", "bordered")
  ) %>%
  add_header_above(c(" ", "Incentive" = 2, "Empowerment" = 2, "Incen*Empow" = 2,
                     "Control" = 1)) %>%
  group_rows(index = c(
    "Ever married (%)" = 2, "Still in school (%)" = 2,
    "Highes education passed" = 2
  ))
```

- Based on the Balance Table above, No significant difference is observed at the baseline.

1.4 Stratified Randomization

```
# Stratify
blocks <- with(data, unionID)
# Define stratification method
Z_stra <- block_and_cluster_ra(blocks = blocks, clusters = cluster,
                               prob_each = c(1/6, 1/3, 1/6,1/3),
                               conditions = c("Incentive", "Empowerment",
                                     "Incen*Empow", "Control")
                       )
# View Cluster Randomization Result
stra_cluster_table  <- as.data.frame(table(Z_stra, cluster)) %>%
  filter(Freq != 0) %>%
  select(-Freq)
head(stra_cluster_table, 20)
```

```
##         Z_stra cluster
## 1  Empowerment  313026
## 2      Control  313068
## 3  Empowerment  313078
## 4  Incen*Empow  313212
## 5  Empowerment  313223
## 6      Control  313243
## 7    Incentive  313253
## 8    Incentive  313364
## 9  Empowerment  313425
## 10     Control  313460
## 11 Empowerment  313470
## 12     Control  313495
## 13     Control  313527
## 14 Incen*Empow  313545
## 15 Empowerment  327020
## 16 Empowerment  327041
## 17     Control  327063
## 18 Incen*Empow  327069
## 19 Empowerment  327085
## 20     Control  327160
```

We stratify by union at a ratio of 1:2:1:2 and then still cluster randomize by community based on the paper, as all the randomization are conducted at a village level. Therefore, we use *block_cluster_ra* function here.

1.5 Stratified Randomization: Balance Check

```
# Add randomization result to original data
merged_stra <- left_join(data, stra_cluster_table,
                         by = c("villageID_fct" = "cluster")) %>%
  rename(treatment = Z_stra)
```

Table 2: Balance Table

	Incentive		Empowerment		Incen*Empow		Control
	Incentive	Incentive_Diff	Empow	Empow_Diff	Incen*Empow	Incen*Empow_Diff	Control
Ever married (%)							
ever_married_mean	16.13	-0.30	17.08	0.84	14.81	0.37	18.65
ever_married_S.D.	36.79		37.64		35.52		38.96
Still in school (%)							
still_in_school_mean	50.66	-0.65	51.30	-1.51	52.64	-1.51	48.08
still_in_school_S.D.	50.00		49.99		49.94		49.97
Highes education passed							
highest_class_passed_mean	6.46	-0.17	6.45	-0.13	6.57	-0.10	6.23
highest_class_passed_S.D.	2.84		2.86		2.81		2.97

```r
# Balance Check
balance_stra <- merged_stra %>%
  group_by(treatment) %>%
  summarise(ever_married_mean = (sum(ever_married_dummy) / n()) * 100,
            ever_married_S.D. = (sd(ever_married_dummy)) * 100,
            still_in_school_mean = (sum(still_in_school_dummy) / n()) * 100,
            still_in_school_S.D. = (sd(still_in_school_dummy)) * 100,
            highest_class_passed_mean = sum(highest_class_passed) / n(),
            highest_class_passed_S.D. = sd(highest_class_passed)
            ) %>%
  transpose_df()

diff_stra <- balance_stra %>%
  mutate_at(vars(-matches("name")), list(Diff = ~ . - Control )) %>%
  select(name, ends_with("Diff"), -Control_Diff) %>%
  filter(grepl("mean", name))

balance_stra <- left_join(balance_stra, diff, by = "name")
```

```r
# Table
balance_stra %>%
  select(name, Incentive, Incentive_Diff, Empowerment,
         Empowerment_Diff, `Incen*Empow`, `Incen*Empow_Diff`, Control) %>%
  kable(col.names = c("", "Incentive", "Incentive_Diff", "Empow",
                      "Empow_Diff", "Incen*Empow", "Incen*Empow_Diff", "Control"),
        caption = "Balance Table",
        digits = 2, format.args = list(scientific = FALSE)) %>%
  kable_classic() %>%
  kable_styling(position = "center", font_size = 8,
                full_width = T, html_font = "Cambria",
                c("stripend", "bordered")) %>%
  add_header_above(c(" ", "Incentive" = 2, "Empowerment" = 2, "Incen*Empow" = 2,
                     "Control" = 1)) %>%
  group_rows(index = c("Ever married (%)" = 2, "Still in school (%)" = 2,
                       "Highes education passed" = 2))
```

Part II. Survey Design with SurveyCTO

2.1

Survey design Test survey

Calculation used for loading preload treatment data: *pulldata('preload', 'treatment', 'memberID', ${memberID})*

2.2

```
# Filter dataset
data_unmarried <- merged_stra %>%
  filter(ever_married_dummy == 0)
# Generate randomly selected dataset used for preload
set.seed(35550)
preload_df <- data_unmarried %>%
  group_by(treatment) %>%
  sample_n(5) %>%
  select(treatment, memberID, villageID) %>%
  mutate(insample = 1)
# Save to csv
write.csv(preload_df, "preload.csv")
```

```
# Load testing survey data
test_survey <- read_excel("survey_test_result.xlsx")
# Analysis
test_survey$still_in_school <- replace_na(test_survey$still_in_school,0)
test_survey %>%
  group_by(treatment_group) %>%
  summarise(pct_married = sum(marital_status) / n(),
            pct_still_in_school = sum(still_in_school) / n())
```

```
## # A tibble: 4 x 3
##   treatment_group pct_married pct_still_in_school
##   <chr>                 <dbl>               <dbl>
## 1 Control                 0.6                 0.2
## 2 Empowerment             0.6                 0.4
## 3 Incen*Empow             0.2                 0.6
## 4 Incentive               0.4                 0.6
```

By testing the follow-up survey, we observe:

- Control group: 60% of women is married, and 20% of women is still in school by the time of the follow-up survey;
- Empowerment Program group: 60% married, and 40% still in school;
- Incentive group: 40% married, and 60% of still in school;
- Incentive & Empowerment group: 20% married, and 60% still in school.