Replication and Extension of "Trading frictions in Indian village economies"

Yuto Yamaoka

2021/12/8

The original study

Emerick, Kyle, 2018, "Trading frictions in Indian village economies", Journal of Development Economics 132 (2018) 32–56.
https://doi.org/10.1016/j.jdeveco.2017.12.010

Results I work on
Table 2 and Table 5

Data I use
I cite data from The Abdul Latif Jameel Poverty Action Lab Dataverse (Massachusetts Institute of Technology) on Harvard Dataverse. (*https://bit.ly/3pzVjyR*)

In this paper, I replicated and extended some results from Emerick's paper "Trading frictions in Indian village economies".
The original paper focuses on trading frictions in rural Indian villages where people from various backgrounds like castes and surnames reside. In the case of rural villages in developing countries, trades tend to occur informally within social networks, unlike standard markets. To test whether trading frictions exist, the author observed the diffusion of a new seed variety in 82 villages. After introducing the new variety to all 82 villages, he allowed the new variety to diffuse through farmer-to-farmer trading in a random half of villages while farmers in the other half of villages. He estimated and compared the demand for the new seed variety between the former(Network villages) and the latter(Door-to-door villages).

In this replication and extension paper,

(i) It is a severe issue that the diffusion of new technology in rural Indian villages can be encouraged or discouraged by a traditionally strong social relationship such as caste and surname.

(ii) To deal with this issue is crucial because this research can contribute to understanding how trade friction and eventually inequality are generated and alleviated within social networks in rural Indian villages.

(iii) I examine how much the intervention of door-to-door sales increases the adoption of a new seed variety in each caste.

(iv) I find that Other Backward Classes(OBC) record more adoption rate than not only Scheduled Castes(SC), but also General class.

```r
library(BalanceR)
library(sandwich)
library(lmtest)
library(broom)
library(car)
library(tidyverse)
library(stargazer)
library(ggplot2)
```

Table 2

```r
#load data and create data frames for Table2
df <- read.delim("tradefriction.tab", header = TRUE, sep = "\t", quote =
"")

df <- df%>%
  mutate(treat = case_when(exper_group == "\"treatment\"" ~ 1, TRUE ~ 0),
         price10 = case_when(priceofswarnasub1 == 10 ~ 1, TRUE ~ 0),
         price12 = case_when(priceofswarnasub1 == 12 ~ 1, TRUE ~ 0),
         price14 = case_when(priceofswarnasub1 == 14 ~ 1, TRUE ~ 0),
         landcult = (df$areal_01 + df$areal_02 + df$areal_03 + df$areal_0
4 + df$areal_05 +
                     df$areal_06 + df$areal_07 + df$areal_08 + df$areal
_09 + df$areal_10 +
                     df$areal_11 + df$areal_12 + df$areal_13 + df$areal
_14 + df$areal_15))

#transform NA to zero
df <- mutate_at(df, c("areal_01", "areal_02", "areal_03", "areal_04", "ar
eal_05",
                      "areal_06", "areal_07", "areal_08", "areal_09", "ar
eal_10",
                      "areal_11", "areal_12", "areal_13", "areal_14", "ar
eal_15"),
                ~replace(., is.na(.), 0))

#create a data frame for column 1 and column 2
df_analytic <- df%>%
  filter(!is.na(adopt))%>%
  select(village_id, adopt, treat, price10, price12, price14, strata)
```

```
#create a data frame for column 3
df_covariate <- df%>%
  mutate(sc = case_when(caste == 2 ~ 1, TRUE ~ 0),
         coop_member = case_when(partclub_3 == 1 ~ 1, TRUE ~ 0),
         landcult = (df$areal_01 + df$areal_02 + df$areal_03 + df$areal_0
4 + df$areal_05 +
                     df$areal_06 + df$areal_07 + df$areal_08 + df$areal_0
9 + df$areal_10 +
                     df$areal_11 + df$areal_12 + df$areal_13 + df$areal_1
4 + df$areal_15))%>%
    select(village_id, adopt, treat, price10, price12, price14, sc, bplca
rd, landcult, coop_member, swarnaadopt, strata)

df_covariate <- df_covariate%>%
  filter(!is.na(adopt) & !is.na(coop_member) &
           !is.na(landcult) & !is.na(bplcard))
```

The estimating equation for column 1 is

$$adoption_{ij} = \beta_0 + \beta_1 Price10_j + \beta_2 Price12_j + \beta_3 Price14_j + \epsilon_{ij}$$

where $adoption_{ij}$ is an indicator for adoption by farmer i in $village_j$, and β_1, β_2, and β_3 measure the gaps in adoption at the three different price levels.

The estimating equations for column 2 and column 3 are

$$adoption_{ij} = \alpha_0 + \alpha_1 Door\ to\ door_j + \epsilon_{ij}$$

and

$$adoption_{ij} = \delta_0 + \delta_1 SC_i + \delta_2 BPL\ card_i + \delta_3 Cultivated_i + \delta_4 Member_i + \delta_5 Swarnauser_i + \epsilon_{ij}$$

```
reg_1 <- lm(adopt ~ price10 + price12 + price14 + factor(strata), data =
df_analytic)
vcov_1 <- vcovCL(reg_1, type = "HC1", cluster = ~ village_id)
robust_se_1 <- sqrt(diag(vcov_1))

reg_2 <- lm(adopt ~ treat + factor(strata), data = df_analytic)
vcov_2 <- vcovCL(reg_2, type = "HC1", cluster = ~village_id)
robust_se_2 <- sqrt(diag(vcov_2))

reg_3 <- lm(adopt ~ treat + sc + bplcard + landcult + coop_member + swarn
aadopt + factor(strata), data = df_covariate)
vcov_3 <- vcovCL(reg_3, type = "HC1", cluster = ~village_id)
robust_se_3 <- sqrt(diag(vcov_3))
```

```
stargazer(reg_1,reg_2,reg_3,
         se = list(robust_se_1, robust_se_2, robust_se_3),
         type = "text", omit.stat = c("f", "ser"),
         keep = c("price10", "price12", "price14", "treat", "sc", "bplca
rd", "landcult", "coop_member", "swarnaadopt"),
         dep.var.labels = "Adoption", dep.var.caption = "",
         covariate.labels = c("Door-to-door and Price = 10", "Door-to-do
or and Price = 12",
                              "Door-to-door and Price = 14", "Door-to-do
or treatment",
                              "Farmer is SC", "Farmer has BPL card", "La
nd cultivated in 2012",
                              "Ag. cooperative member", "Swarna user in
2012"),
         title  = "Estimated difference between adoption in networks and
 door-to-door sales.",
         font.size = "footnotesize",
         add.lines =list(c("Strata Fixed Effect", "Yes", "Yes", "Yes"),
                         c("Mean of Dep Variable: Networks", "0.07", "0.
07", "0.07")))
```

Estimated difference between adoption in networks and door-to-door sales.
===
                                            Adoption
                                   (1)   (2)   (3)

Door-to-door and Price = 10   0.380***
                                 (0.077)

Door-to-door and Price = 12   0.357***
                                 (0.066)

Door-to-door and Price = 14   0.275***
                                 (0.061)

Door-to-door treatment   0.336***   0.337***
                                           (0.043)   (0.043)

Farmer is SC   -0.060
                                                      (0.040)

Farmer has BPL card   -0.054*
                                                      (0.030)

```
Land cultivated in 2012                                   0.004
                                                         (0.007)

Ag. cooperative member                                   -0.019
                                                         (0.023)

Swarna user in 2012                                      0.090***
                                                         (0.033)


----------------------------------------------------------------
Strata Fixed Effect              Yes       Yes       Yes
Mean of Dep Variable: Networks   0.07      0.07      0.07
Observations                     1,150     1,150     1,134
R2                               0.190     0.185     0.203
Adjusted R2                      0.185     0.181     0.196
================================================================
Note:                            *p<0.1; **p<0.05; ***p<0.01
```

I replicate Table 2 perfectly. Although the author did not include adjusted R^2 , I add it in my replicated result.

Treatment Effect Heterogeneity

I attempt to extend the above result in this part where I examine how the treatment effect varies across different castes(Scheduled Caste, Other Backward Class, and General). This is because Caste System in India has a substantial influence on social relationships, especially in rural villages. In the case of the diffusion of a new seed variety, the adoption rate can be changed in each caste because it is possible that each class people belong to determines the attitude towards door-to-door sales and its price.

Firstly, I run a regression analysis based on a subsample.

```
df_caste <- df%>%
  filter(!is.na(caste) & !is.na(adopt))%>%
  select(village_id, adopt, treat, caste, strata)

reg_sc <- lm(adopt ~ treat + factor(strata), data = df_caste, subset = ca
ste == 2)
vcov_sc <- vcovCL(reg_sc, type ="HC1", cluster = ~village_id)
robust_sc <- sqrt(diag(vcov_sc))

reg_obc <- lm(adopt ~ treat + factor(strata), data = df_caste, subset = c
aste == 3)
vcov_obc <- vcovCL(reg_obc, type ="HC1", cluster = ~village_id)
robust_obc <- sqrt(diag(vcov_obc))
```

```
reg_gene <- lm(adopt ~ treat + factor(strata), data = df_caste, subset =
caste == 4)
vcov_gene <- vcovCL(reg_gene, type ="HC1", cluster = ~village_id)
robust_gene <- sqrt(diag(vcov_gene))


stargazer(reg_sc, reg_obc, reg_gene,
          se = list(robust_sc, robust_obc, robust_gene),
          keep = "treat",
          type = "text", omit.stat = c("ser","f"),
          dep.var.labels = "Adoption",
          column.labels = c("other backward class", "Genaral"))



====================================================
                    Dependent variable:
              --------------------------------------
                            Adoption
              other backward class Genaral
                      (1)             (2)       (3)
----------------------------------------------------
treat               0.207***                0.401*** 0.341***
                    (0.053)                 (0.062)   (0.057)


----------------------------------------------------
Observations          231             387       515
R2                   0.076           0.251     0.205
Adjusted R2          0.051           0.239     0.195
====================================================
Note:                       *p<0.1; **p<0.05; ***p<0.01
```

Then, I make an interaction model to test whether the treatment effects across castes on the above table are statistically significant. The equation I set is

$$adoption_ij = \beta_0 + \beta_1 OBC_i + \beta_2 General_i + \beta_3 door\ to\ door_j * SC_i + \beta_4 door\ to\ door_j * OBC_i + \beta_5 door\ to\ door_j * General_i + \epsilon_{ij}$$

where SC_i, OBC_i, and $General_i$ are dummy variables that equal 1 when farmer i belong to each caste.

```
#Interaction model
caste_interaction<- df_caste%>%
  mutate(sc = case_when(caste == 2 ~ 1, TRUE ~ 0),
         obc = case_when(caste == 3 ~ 1, TRUE ~ 0),
         general = case_when(caste == 4 ~ 1, TRUE ~ 0),
         treat_s = treat*sc,
         treat_o = treat*obc,
```

```
            treat_g = treat*general)%>%
  select(village_id, adopt, treat, sc, obc,
          general, treat_s, treat_o, treat_g, strata)

reg_int1 <- lm(adopt ~ obc + general + treat_s + treat_o +
                  treat_g + factor(strata) , data = caste_interaction)
vcov_int1 <- vcovCL(reg_int1, type ="HC1", cluster = ~village_id)
robust_int1 <- sqrt(diag(vcov_int1))


#Run F-test to see whether the coefficients on door to door*SC, door to d
oor*OBC, and door to door*General are zero.
linearHypothesis(reg_int1, c("treat_s=0", "treat_o=0", "treat_g=0"), whit
e.adjust = T)

Linear hypothesis test

Hypothesis:
treat_s = 0
treat_o = 0
treat_g = 0

Model 1: restricted model
Model 2: adopt ~ obc + general + treat_s + treat_o + treat_g + factor(str
ata)

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F     Pr(>F)
1   1125
2   1122  3 76.653 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now, it is possible to reject the hypothesis that at least one of the three coefficients on the interaction variables is not zero. Therefore, I can state that the treatment effects across castes on the above table are statistically significant.

```
stargazer(reg_int1, se = list(robust_int1),
          title = "Estimated difference in adoptation between castes",
          type = "text", omit.stat = c("ser","f"),
          keep = c("obc", "general", "treat_s", "treat_o", "treat_g"),
          dep.var.labels = "Adoption",
          add.lines = c("F-test on the equality of coefficients", "2.2e-1
6***"))
```

```
Estimated difference in adoptation between castes
===================================================================
                                       Dependent variable:
                              ---------------------------
                                          Adoption
-------------------------------------------------------------------
obc                                        -0.002
                                           (0.047)


general                                    -0.017
                                           (0.049)


treat_s                                    0.185***
                                           (0.061)


treat_o                                    0.403***
                                           (0.064)


treat_g                                    0.345***
                                           (0.057)


-------------------------------------------------------------------
F-test on the equality of coefficients
2.2e-16***
Observations                                1,133
R2                                          0.200
Adjusted R2                                 0.193
===================================================================
Note:                              *p<0.1; **p<0.05; ***p<0.01
```

```r
#plot the estimates according to castes
robust_int_plot <- coeftest(reg_int1, vcov = vcovCL(reg_int1, type = "HC1
", cluster = ~ village_id))

df_plot <- tidy(robust_int_plot, conf.int = T)

filter(df_plot, term == "treat_s" | term == "treat_o" | term == "treat_g"
)%>%
  mutate(term = recode(term, treat_s = "Treatment * SC",
                      treat_o = "Treatment * OBC",
                      treat_g = "Treatment * General"))%>%
  ggplot(aes(x = term, y = estimate, ymin = conf.low, ymax = conf.high))+
  geom_point()+
  geom_pointrange()+
  labs(x = "Treatment Effects", y = "Estimates")+
  theme_bw()
```
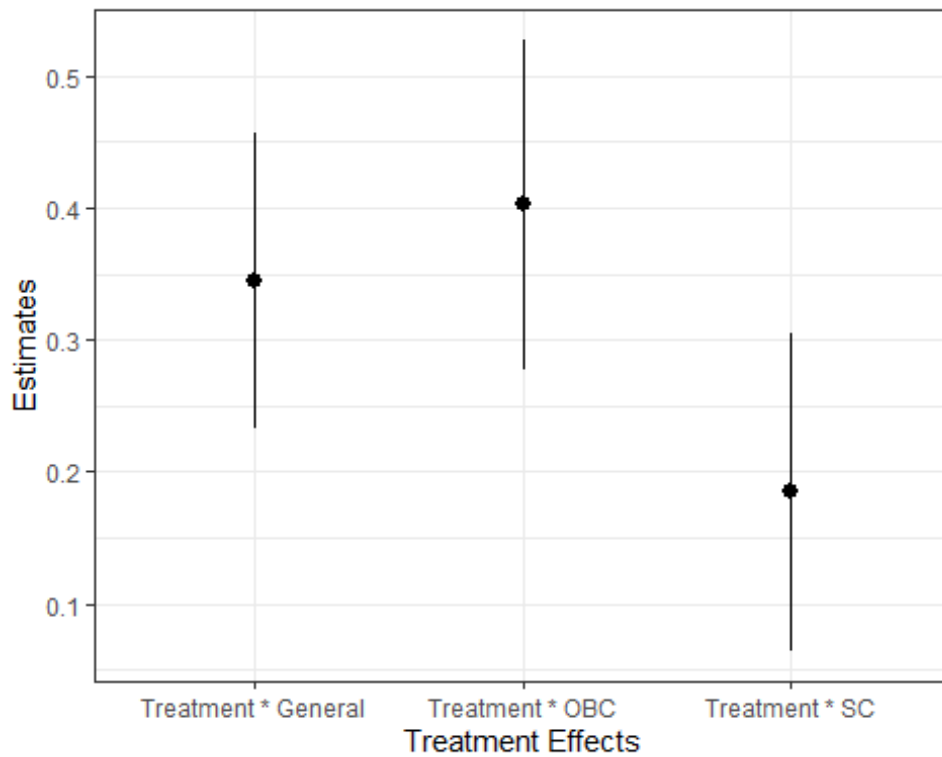
The result from my extension indicates that people belonging to Other Backward Classes, classically located hierarchically between Scheduled Castes and General classes, are more willing to adopt the new seed variety than the other two castes. In my view, the possible possession of close social connections with the SC and General can be a convincing explanation of the high adaptation rate of OBC. However, I have to review previous research and conduct more detailed analysis to prove my expectation.

Table 5

I also replicate Table 5 that represents estimated peer effects among social relationships in both door-to-door villages and network villages. The author focused specifically on two measures of social links: sub-castes and surnames.

The regression specification the author wrote is

$$adoption_{ij} = \beta_0 + \beta_1 door\ to\ door_j + \beta2 linksOR_{ij} + \beta_3 links_{ij} + \beta_4 linksOR_{ij} * door\ to\ door_j + \beta5 links_{ij} * door\ to\ door_j + \epsilon_{ij}$$

where where $linksOR_{ij}$ is the number of peers of farmer i that were selected as original recipients and $links_{ij}$ is the total number of links of farmer i.

In Table 5, the author considered the equation in terms of surname and sub-caste. $linksOR_{ij}$ and $links_{ij}$ are respectively denoted as "Original recipients w/same surname" and "Total number w/same surname" in column 1 and column 2 and represented as "Original recipients same sub-caste" and "Total number same sub-caste" in column 3 and column 4.

```
df_tf <- df%>%
  mutate(sc = case_when(caste == 2 ~ 1, TRUE ~ 0),
         coop_member = case_when(partclub_3 == 1 ~ 1, TRUE ~ 0),
         pepsurname = orwithsurname + nrwithsurname,
         pepsamesub = samesubOR + samesubNR)%>%
  select(village_id, adopt, caste, treat,orwithsurname, pepsurname, sames
ubOR, pepsamesub, strata)

reg_tf_1 <-lm(adopt ~ treat + treat*orwithsurname + treat*pepsurname +orw
ithsurname +
              pepsurname + factor(strata), data = df_tf)
vcovtf_1 <-vcovCL(reg_tf_1, type = "HC1", cluster = ~village_id)
robust_tf_1 <- sqrt(diag(vcovtf_1))

reg_tf_2 <-lm(adopt ~ treat*orwithsurname + treat*pepsurname +
              orwithsurname + pepsurname + factor(village_id) - treat,
data = df_tf)
vcovtf_2 <-vcovCL(reg_tf_2, type = "HC1", cluster = ~village_id)
robust_tf_2 <- sqrt(diag(vcovtf_2))

reg_tf_3 <-lm(adopt ~ treat + treat*samesubOR + treat*pepsamesub +samesub
OR +
              pepsamesub + factor(strata), data = df_tf)
vcovtf_3 <-vcovCL(reg_tf_3, type = "HC1", cluster = ~village_id)
robust_tf_3 <- sqrt(diag(vcovtf_3))

reg_tf_4 <-lm(adopt ~ treat*samesubOR + treat*pepsamesub +
              samesubOR + pepsamesub + factor(village_id) - treat, data
 = df_tf)
vcovtf_4 <-vcovCL(reg_tf_4, type = "HC1", cluster = ~village_id)
robust_tf_4 <- sqrt(diag(vcovtf_4))


stargazer(reg_tf_1, reg_tf_2, reg_tf_3, reg_tf_4,
          se = list(robust_tf_1, robust_tf_2, robust_tf_3, robust_tf_4),
          digits = 0.5, digits.extra = 0, align = TRUE,
          type = "text", omit.stat = c("f","ser"),
          keep = c("treat", "orwithsurname", "samesubOR","pepsamesub","pe
psurname",
                   "treat:orwithsurname", "treat:pepsurname", "treat:same
```

```
subOR", "treat:pepsamesub"),
         order = c("^treat$", "^treat:orwithsurname$", "^orwithsurname$"
, "^pepsurname$", "^treat:pepsurname$",
                   "^treat:samesubOR$", "^samesubOR$", "^pepsamesub$", "
^treat:pepsamesub$"),
         covariate.labels = c("d-t-d ","d-t-d * linksOR w/same surname",
                              "linksOR w/same surname","links w/same sur
name",
                              "d-t-d  * links w/same surname",
                              "d-t-d  * linksOR same sub-caste",
                              "linksOR same sub-caste","links same sub-c
aste",
                              "d-t-d  * links same sub-caste"),
         dep.var.caption = "Adoption", font.size = "tiny",
         column.labels = c("dtd surname", "nw surname", "dtd caste", "nw
 caste"),
         title = "Estimated peer effects in network and door-to-door vil
lages.",
         add.lines =list(c("Strata Fixed Effect", "Yes", "No", "Yes", "N
o"),
                         c("Village Fixed Effect", "No", "Yes", "No", "Y
es")))
```

```
Estimated peer effects in network and door-to-door villages.
===============================================================================
                                              Adoption
                              -------------------------------------------
                                               adopt
                             dtd surname nw surname dtd caste nw caste
                                 (1)        (2)         (3)       (4)
-------------------------------------------------------------------------------
d-t-d                           0.3***                 0.4***
                                (0.1)                  (0.1)

d-t-d * linksOR w/same surname  -0.1*      -0.1**
                                (0.0)      (0.0)

linksOR w/same surname           0.0       0.1***
                                (0.0)      (0.0)

links w/same surname            -0.0      -0.0***
                                (0.0)      (0.0)

d-t-d * links w/same surname     0.0       0.0**
                                (0.0)      (0.0)
```

```
d-t-d * linksOR same sub-caste                                    -0.1*      -0.1
                                                                  (0.0)     (0.0)

linksOR same sub-caste                                             0.0*      0.0**
                                                                  (0.0)     (0.0)

links same sub-caste                                              -0.0      -0.0*
                                                                  (0.0)     (0.0)

d-t-d * links same sub-caste                                       0.0       0.0
                                                                  (0.0)     (0.0)

---------------------------------------------------------------------
Strata Fixed Effect              Yes        No        Yes        No
Village Fixed Effect             No         Yes       No         Yes
Observations                     1,135      1,135     1,135      1,135
R2                               0.2        0.4       0.2        0.4
Adjusted R2                      0.2        0.4       0.2        0.4
=====================================================================
Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

I replicate Table 5 almost perfectly. However, I obtain slightly different statistical significance in the third row of column 2 and the eighth row in column 4. Although I cannot find why this difference happens, I may have an issue with the setting of standard robust SEs in lm() functions.

Overall, Table 5 indicates that sharing the same sub-castes and surnames with original recipients increases the adoption of the new seed variety in network villages where door-to-door sales are not implemented.