**Replication and Extension of “Trading frictions in Indian village economies”**

10A19169 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.

**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*](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 estimate 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,

1. 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.
2. 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.
3. I examine how much the intervention of door-to-door sales increases the adoption of a new seed variety in each caste.
4. I find that Other Backward Classes(OBC) record more adoption rate than Scheduled Castes(SC) and even General class.

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

**Table 2**

*#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\_04 + 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$areal\_15))  *#transform NA to zero*df <- mutate\_at(df, c("areal\_01", "areal\_02", "areal\_03", "areal\_04", "areal\_05",  "areal\_06", "areal\_07", "areal\_08", "areal\_09", "areal\_10",  "areal\_11", "areal\_12", "areal\_13", "areal\_14", "areal\_15"),  ~replace(., is.na(.), 0))  *#create a data frame for column 1 and column 2*df\_analytic <- df%>%  filter(!is.na(adopt))%>% select(village\_id, adopt, treat, price10, price12, price14, strata) *#create a data frame for column 3*df\_covariate <- df%>% mutate(sc = case\_when(caste == 2 ~ 1, TRUE ~ 0), coop\_member = case\_when(partclub\_3 == 1 ~ 1, TRUE ~ 0), landcult = (df$areal\_01 + df$areal\_02 + df$areal\_03 + df$areal\_04 + df$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))%>% select(village\_id, adopt, treat, price10, price12, price14, sc, bplcard, 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

where is an indicator for adoption by farmer in , and , , and measure the gaps in adoption at the three different price levels.

The estimating equations for column 2 and column 3 are

and

Since the author used vce(cl) in STATA for robust standard error, I use vcovCL( ) instead.

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 + swarnaadopt + 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", "bplcard", "landcult", "coop\_member", "swarnaadopt"), dep.var.labels = "Adoption", dep.var.caption = "", covariate.labels = c("Door-to-door and Price = 10", "Door-to-door and Price = 12", "Door-to-door and Price = 14", "Door-to-door treatment", "Farmer is SC", "Farmer has BPL card", "Land 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 , 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 which caste 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 = caste == 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 = caste == 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 estimating equation I set is

where , , and are dummy variables that equal 1 when farmer 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 door\*OBC, and door to door\*General are zero.*linearHypothesis(reg\_int1, c("treat\_s=0", "treat\_o=0", "treat\_g=0"), white.adjust = T)

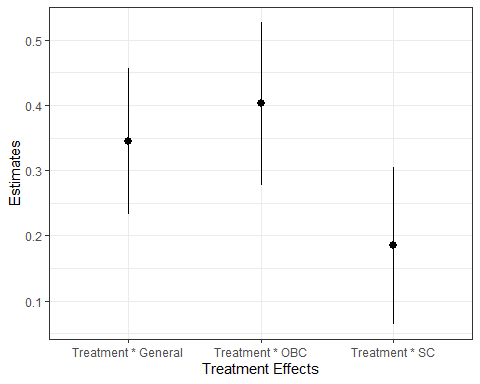
Linear hypothesis testHypothesis:treat\_s = 0treat\_o = 0treat\_g = 0Model 1: restricted modelModel 2: adopt ~ obc + general + treat\_s + treat\_o + treat\_g + factor(strata)Note: Coefficient covariance matrix supplied. Res.Df Df F Pr(>F) 1 1125 2 1122 3 76.653 < 2.2e-16 \*\*\*---Signif. codes: 0 '\*\*\*' 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-16\*\*\*"))

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

*#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 treatment effect heterogeneity indicates that people belonging to Other Backward Classes, classically located 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 willingness of OBC. However, I have to review previous research and conduct more detailed research 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

where where is the number of peers of farmer that were selected as original recipients and is the total number of links of farmer .

In Table 5, the author considered the equation in terms of surname and sub-caste. and 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, samesubOR, pepsamesub, strata)reg\_tf\_1 <-lm(adopt ~ treat + treat\*orwithsurname + treat\*pepsurname +orwithsurname + 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 +samesubOR +  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","pepsurname",  "treat:orwithsurname", "treat:pepsurname", "treat:samesubOR", "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 surname", "d-t-d \* links w/same surname", "d-t-d \* linksOR same sub-caste", "linksOR same sub-caste","links same sub-caste", "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 villages.", add.lines =list(c("Strata Fixed Effect", "Yes", "No", "Yes", "No"), c("Village Fixed Effect", "No", "Yes", "No", "Yes")))

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 problems 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.