The Price of Prejudice

Replication by: Pawonee Khadka, University of Alabama

11/28/2022

The paper is based on a field experiment that goes on to investigate ethnic prejudice in the workplace. The authors want to see how potential discriminators respond to changes in the cost of discrimination.

```
library(dplyr)
library(tidyverse)
library(haven)
```

All analysis for the original paper was done using STATA and all data and necessary code instructions was made publicly available. With my replication work, I didn't have much to do with the raw data, but jumped into replication right away. Following is the chuck for replication of Table 2: Team Production Function.

```
Price <- read_dta("Price_of_Prejudice_Stata_data (1).dta") %>%
  mutate( productivity = case_when(ethnicity == 3 | couple == 4 ~ 0, TRUE ~ 1))%>%
  mutate (LnProd_1 = log(prod_1),
          LnProd_2 = log(own_prod_2),
          muslim_team = 0)
Price_1 <- Price %>%
  mutate(Muslim_sounding_team = case_when(ethnicity == 2 & couple == 1 ~ 1, TRUE ~ 0))

Price_2 <- Price_1 %>%
  mutate(Danish_sounding_team = case_when(ethnicity == 1 & couple == 1 ~ 1, TRUE ~ 0))

Price_2 <- Price_2 %>%
  mutate(Alone_2nd_round = case_when(couple == 3 ~ 1, TRUE ~ 0))

Price_2 <- Price_2 %>%
  mutate(team = 1)%>%
  mutate(team = case_when(Danish_sounding_team == 1 ~ 2, Muslim_sounding_team ==1 ~ 3, Alone_2nd_round =

Price_2 <- Price_2 %>%
  mutate(Team_type = case_when(team == 1 ~ "Heterogeneous", team ==2 ~ "Danish", team == 3 ~ "Muslim",
```

Dummies:

```
Price_2 <- Price_2 %>%
  mutate(danish = case_when(ethnicity == 1 ~ 1, TRUE ~ 0))%>%
  mutate(lnprodpartnertemp = case_when(Alone_2nd_round == 0 ~ log(prod_partner), TRUE ~ 0 ),
         decision_maker = case_when(type ==1 & couple != 3 ~ 1, TRUE ~0))
# Here, danish 0 "Muslim-sounding" 1 "Danish-sounding"
#temporary partner variable
```

Interaction -between being a decision-maker and being in a heterogeneous team

```
Price_2 <- Price_2 %>%
  mutate(decision_maker_mixed = case_when(muslim_team == 0 & Danish_sounding_team == 0 & Alone_2nd_roun
         lnprod1alone = LnProd_1 * Alone_2nd_round)
#name first one "D-M in heterogeneous team" and second lnprod1alone "ln(Prod_1) * alone"
```

Estimate production function

```
Price_3 <- Price_2 %>%
  filter(productivity == 1)

A <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male, data = Price_3)
# For vce(robust) :
library(sandwich)
library(lmtest)
AA <- vcovHC(A, type = "HC1")
robust_AA <- sqrt(diag(AA))

B <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male+ decision_maker, data = Price_3)
BB <- vcovHC(B, type = "HC1")
robust_BB <- sqrt(diag(BB))

C <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male+ Danish_sounding_team + Muslim_sounding
CC <- vcovHC(C, type = "HC1")
robust_CC <- sqrt(diag(CC))

D <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male+ Danish_sounding_team + Muslim_sounding
DD <- vcovHC(D, type = "HC1")
robust_se <- sqrt(diag(DD))

library (stargazer)
stargazer(A,B,C,D, type ="text",
          title = "Table 2:Team Production Function",
          dep.var.labels = c("Dependent Variable: ln(prod2)"),
          notes.label ="Significance Levels",
          covariate.labels = c("logprod1i","logprod1j","logprod1i * Alone","Male","Danish","Muslim","Al
          no.space = TRUE,
          se = list(robust_AA,robust_BB,robust_CC,robust_se),
          keep.stat = c("adj.rsq","n")

          )
```

```
##
## Table 2:Team Production Function
## ================================================================
##                                      Dependent variable:
##                                 ------------------------------
##                                   Dependent Variable: ln(prod2)
##                                  (1)      (2)      (3)      (4)
## ----------------------------------------------------------------
## logprod1i                      0.416*** 0.413*** 0.385*** 0.387***
```

```
##                                 (0.066)  (0.066)  (0.052)  (0.053)
## logprod1j                      0.416*** 0.421*** 0.385*** 0.387***
##                                 (0.066)  (0.067)  (0.052)  (0.053)
## logprod1i * Alone              0.398*** 0.401***  0.444    0.443
##                                 (0.064)  (0.065)  (0.285)  (0.287)
## Male                             0.027    0.026    0.020    0.020
##                                 (0.027)  (0.027)  (0.033)  (0.033)
## Danish                                   -0.008             -0.028
##                                          (0.024)            (0.039)
## Muslim                                                       0.038
##                                                             (0.047)
## Alone                                              0.047*  0.064**
##                                                   (0.025)  (0.032)
## Decision maker                                     -0.024   -0.025
##                                                   (0.035)  (0.035)
## Decision maker * Heterogeneous                     -0.354   -0.335
##                                                   (1.411)  (1.422)
## Constant                         0.849    0.848   1.135*** 1.117***
##                                 (0.517)  (0.519)  (0.369)  (0.373)
## ------------------------------------------------------------------
## Observations                      147      147      147      147
## Adjusted R2                      0.465    0.461    0.463    0.456
## ==================================================================
## Significance Levels                    *p<0.1; **p<0.05; ***p<0.01
```

In this table: *the Dependent variable is the log of the number of envelopes stuffed in round 2 by worker i* prod1i is the number of envelopes stuffed in round 1 by worker i *prod1j is the number of envelopes stuffed by i's coworker in round 2* Alone is a dummy set to 1 if worker i works alone in round 2 *Male is worker i's gender* Decision maker indicates if worker i makes a choice of coworker *The remaining dummies characterize team composition in round 2.

- Two main observations can be made here:
- The table replication is not exact. This is because of tiny differences between softwares. I have used the stargazer package which allows for professional tables, yet some differences remain as compared to original table.
- There is a key difference to mark here. The number of observations, in this replication is 147 unlike 140 in the paper. This has led to some minor chane in statistic. The code was explained fairly well, however, authors did not provide specifications on filtering the data. My assumption is that the error is stemming from filtering differences by "Or" vs "And"

Second table:

Estimating the cost of discrimination

```
Price_4 <- Price_2 %>%
  filter(type == 1)%>%
  filter(ethnicity != 3|type_day_1 != 9999|type_day_2 != 9999)%>%
  filter(main == 1 & info == 1)
```

Predict production with own type * Reset team composition dummies, interaction terms

```
Price_4 <- Price_4 %>%
  mutate(Alone_2nd_round == 0,
         lnprod1alone = 0,
         lnprodpartnertemp = log(prod_own))
```