Problem Set 11

Daniel Shapiro

11/30/2022

Question 1 Background:

In this problem you will be replicating Jens Hainmueller, Jonathan Mummolo, and Yiqing Xu (2019), "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice," Political Analysis (HMX). I will ask you to recreate the plots and analyses using your own original code.

Your results will not be exactly the same, due to choices of bins, your random seed, and other modeling choices. You do not have to include controls, for example. However, you should show the same broad results, and explain the significance of those results.

1a) Recreate the simulated datasets introduced by HMX on page 168. Run a regression of your Y variables on X, D, and an interaction of X and D. Report the results in a nicely formatted table, and briefly describe what we can learn from the tables.

HMX produce three datasets on page 168. The first two follow this formula:

$$Y_i = 5 - 4X_i - 9D_i + 3D_iX_i + \epsilon_i$$

The last one follows this formula:

$$Y_i = 2.5 - X_i^2 - 5D_i + 2D_iX_i^2 + \zeta_i$$

Now, let's define the parameters for the first two equations, as reported by the article.

```
# There are 200 rows.

X1 <- rnorm(200, mean = 3, sd = 1)
Error1 <- rnorm(200, mean = 0, sd = 4)
D1 <- rbinom(200, 1, 0.5)

first <- as.data.frame(cbind(X1, D1, Error1)) %>%

# Add an extra empty vector.

  mutate(Y1 = 1)

# Run a for loop.

for(i in 1:200){
  first$Y1[i] <- 5 - 4*first$X1[i] - 9*first$D1[i] + 3*first$D1[i]*first$X1[i] + first$Error1[i]
```

```
}

# Setting up next equation

X2 <- rnorm(200, mean = 3, sd = 1)
Error2 <- rnorm(200, mean = 0, sd = 4)
D2 <- rnorm(200, mean = 3, sd = 1)

second <- as.data.frame(cbind(X2, D2, Error2)) %>%
  mutate(Y2 = 1)

for(i in 1:200){
  second$Y2[i] <- 5 - 4*second$X2[i] - 9*second$D2[i] + 3*second$D2[i]*second$X2[i] + second$Error2[i]
}
```

Now, we'll set up the next formula.

```
X3 <- runif(200, min = -3, max = 3)
Error3 <- rnorm(200, mean = 0, sd = 4)
D3 <- rbinom(200, 1, 0.5)

third <- as.data.frame(cbind(X3, D3, Error3)) %>%
  mutate(Y3 = 1) %>%
  mutate(xsquared = X3^2)

for(i in 1:200){
  third$Y3[i] <- 2.5 - third$xsquared[i] - 5*third$D3[i] + 2*third$D3[i]*third$xsquared[i] + third$Erro
}
```

Now we will run regressions.

```
# Ran robust because it's good practice

reg1 <- lm(Y1 ~ X1 + D1 + X1*D1, data = first)
reg2 <- lm(Y2 ~ X2 + D2 + X2*D2, data = second)
reg3 <- lm(Y3 ~ X3 + D3 + X3*D3, data = third)

sum1 <- summary(reg1)
sum2 <- summary(reg2)
sum3 <- summary(reg3)

stargazer(reg1, reg2, reg3, type = "text")
```

```
##
## ===============================================================
##                                     Dependent variable:
##                                 ----------------------------
##                                     Y1         Y2       Y3
##                                    (1)        (2)       (3)
## ---------------------------------------------------------------
## X1                                -3.881***
##                                    (0.439)
```

```
##
## D1                              -11.738***
##                                  (2.068)
##
## X1:D1                            3.761***
##                                  (0.651)
##
## X2                                          -4.524***
##                                              (0.852)
##
## D2                                          -9.338***
##                                              (0.919)
##
## X2:D2                                        3.218***
##                                              (0.273)
##
## X3                                                      0.087
##                                                        (0.282)
##
## D3                                                      0.397
##                                                        (0.654)
##
## X3:D3                                                  -0.007
##                                                        (0.404)
##
## Constant                         4.910***    5.643*   -0.119
##                                  (1.369)    (2.916)   (0.445)
##
## ------------------------------------------------------------
## Observations                       200        200       200
## R2                                0.287      0.704     0.003
## Adjusted R2                       0.276      0.699    -0.012
## Residual Std. Error (df = 196)    4.237      4.154     4.591
## F Statistic (df = 3; 196)       26.301***  155.217***  0.198
## ============================================================
## Note:                            *p<0.1; **p<0.05; ***p<0.01
```