RDD

Areeya Aksornpan

4/15/2021

I had already reloaded the experimental group data, merged in the CPS controls in Stata, and exported it as nsw\_mixtape.csv. Please see attached my work in nsw\_pscore1.do file.

## Question 1

nsw <- read\_csv("~/Desktop/nsw\_mixtape.csv")

### Part A

# LPM Quadratic  
lpm <- lm(treat ~ age + agesq + educ + educsq + marr + nodegree +   
 black + hisp + re75 + u75 + re74 + u74, data = nsw)  
# Logit Cubic  
logit <- glm(treat ~ age + agesq + agecube + educ + educsq + marr + nodegree + black + hisp + re75 + u75 + re74 + u74, family = binomial(link = "logit"), data = nsw)

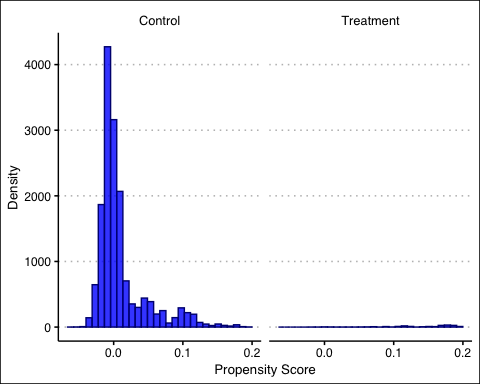
### Part B

# LPM Propensity Scores  
ps\_lpm <-   
 tibble(  
 p\_score = predict(lpm, type = "response"),  
 treat = lpm$model$treat  
 )  
# Logit Propensity Score  
ps\_logit <-   
 tibble(  
 p\_score = predict(logit, type = "response"),  
 treat = logit$model$treat  
 )

### Part C

ct\_tm <- c("Control", "Treatment")  
names(ct\_tm) <- c("0", "1")

# LPM Histogram  
ps\_lpm %>%  
 ggplot() +  
 geom\_histogram(aes(x = p\_score), fill = 'blue', color = 'navy', alpha = 0.8) +   
 labs(x = "Propensity Score", y = "Density") +  
 theme\_clean() +   
 facet\_grid(. ~ treat, labeller = labeller(treat = ct\_tm))

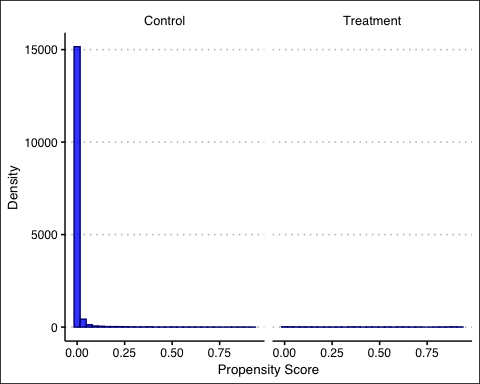


# LPM Max and Min  
ps\_lpm %>%  
 group\_by(treat) %>%  
 summarise(Max = round(max(p\_score), 4), Min = round(min(p\_score), 4)) %>%  
 mutate(treat = if\_else(treat == 0, "Control", "Treatment")) %>%  
 rename(Treat = treat) %>%  
 kbl("pipe")

|  |  |  |
| --- | --- | --- |
| Treat | Max | Min |
| Control | 0.1938 | -0.0637 |
| Treatment | 0.1939 | -0.0139 |

The max and min values of the propensity score for the treatment group are 0.1939 and -0.0139. For the control group, the max and min values are 0.1938 and -0.0637.

# Logit Histogram  
ps\_logit %>%  
 ggplot() +  
 geom\_histogram(aes(x = p\_score), fill = 'blue', color = 'navy', alpha = 0.8) +   
 labs(x = "Propensity Score", y = "Density") +  
 theme\_clean() +   
 facet\_grid(. ~ treat, labeller = labeller(treat = ct\_tm))



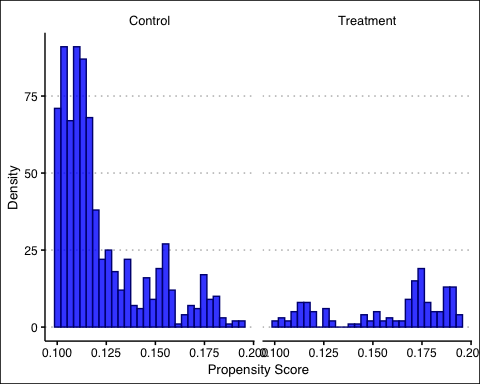
# Logit Max and Min  
ps\_logit %>%  
 group\_by(treat) %>%  
 summarise(Max = round(max(p\_score), 4), Min = round(min(p\_score), 4)) %>%  
 mutate(treat = if\_else(treat == 0, "Control", "Treatment")) %>%  
 rename(Treat = treat) %>%  
 kbl("pipe")

|  |  |  |
| --- | --- | --- |
| Treat | Max | Min |
| Control | 0.9142 | 0.000 |
| Treatment | 0.9226 | 0.001 |

The max and min values of the propensity score for the treatment group are 0.9226 and 0.001. For the control group, the max and min values are 0.9142 and 0.000.

### Part D

# LPM Histogram (between 0.1 and 0.9)  
ps\_lpm %>%  
 filter(between(p\_score, 0.1, 0.9)) %>%  
 ggplot() +  
 geom\_histogram(aes(x = p\_score), fill = 'blue', color = 'navy', alpha = 0.8) +   
 labs(x = "Propensity Score", y = "Density") +  
 theme\_clean() +   
 facet\_grid(. ~ treat, labeller = labeller(treat = ct\_tm))

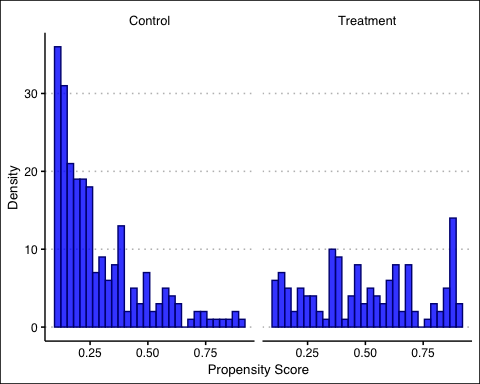


# LPM Max and Min (between 0.1 and 0.9)  
ps\_lpm %>%  
 filter(between(p\_score, 0.1, 0.9)) %>%  
 group\_by(treat) %>%  
 summarise(Max = round(max(p\_score), 4), Min = round(min(p\_score), 4)) %>%  
 mutate(treat = if\_else(treat == 0, "Control", "Treatment")) %>%  
 rename(Treat = treat) %>%  
 kbl("pipe")

|  |  |  |
| --- | --- | --- |
| Treat | Max | Min |
| Control | 0.1938 | 0.1001 |
| Treatment | 0.1939 | 0.1012 |

After dropping all units whose propensity scores are less than 0.1 and more than 0.9, the max and min values of the propensity score for the treatment group are 0.1939 and 0.1012. For the control group, the max and min values are 0.1938 and 0.1001.

# Logit - Histogram // filtered  
ps\_logit %>%  
 filter(between(p\_score, 0.1, 0.9)) %>%  
 ggplot() +  
 geom\_histogram(aes(x = p\_score), fill = 'blue', color = 'navy', alpha = 0.8) +   
 labs(x = "Propensity Score", y = "Density") +  
 theme\_clean() +   
 facet\_grid(. ~ treat, labeller = labeller(treat = ct\_tm))



# Logit -- max/min // filtered  
ps\_logit %>%  
 filter(between(p\_score, 0.1, 0.9)) %>%  
 group\_by(treat) %>%  
 summarise(Max = round(max(p\_score), 4), Min = round(min(p\_score), 4)) %>%  
 mutate(treat = if\_else(treat == 0, "Control", "Treatment")) %>%  
 rename(Treat = treat) %>%  
 kbl("pipe")

|  |  |  |
| --- | --- | --- |
| Treat | Max | Min |
| Control | 0.8948 | 0.1001 |
| Treatment | 0.8966 | 0.1016 |

After dropping all units whose propensity scores are less than 0.1 and more than 0.9, the max and min values of the propensity score for the treatment group are 0.8966 and 0.1016. For the control group, the max and min values are 0.8948 and 0.1001.

## Question 2

# LPM First Difference  
lpm1 <- ps\_lpm %>%   
 filter(treat == 1) %>%  
 pull(p\_score) %>%   
 mean()  
lpm0 <- ps\_lpm %>%   
 filter(treat == 0) %>%  
 pull(p\_score) %>%   
 mean()  
lpm\_sd <- round(lpm1 - lpm0, 4)  
glue('The before and after first difference for the LPM model is {lpm\_sd}.')

## The before and after first difference for the LPM model is 0.1309.

# Logit - First Difference  
logit1 <- ps\_logit %>%   
 filter(treat == 1) %>%  
 pull(p\_score) %>%   
 mean()   
logit0 <- ps\_logit %>%   
 filter(treat == 0) %>%  
 pull(p\_score) %>%   
 mean()   
logit\_sd <- round(logit1 - logit0, 4)  
glue('The before and after first difference for the logit model is {logit\_sd}.')

## The before and after first difference for the logit model is 0.414.

## Question 3

#Dehejia and Wahba (2002) referred from ipw.do  
nsw\_dw\_cps <- nsw %>% cbind(pscore = ps\_logit$p\_score)  
N <- nrow(nsw\_dw\_cps)  
#Manual with non-normalized weights using all the data  
nsw\_dw\_cps <- nsw\_dw\_cps %>%   
 mutate(d1 = treat/pscore,  
 d0 = (1 - treat)/(1 - pscore))  
s1 <- sum(nsw\_dw\_cps$d1)  
s0 <- sum(nsw\_dw\_cps$d0)  
nsw\_dw\_cps <- nsw\_dw\_cps %>%   
 mutate(y1 = treat \* re78/pscore,  
 y0 = (1 - treat) \* re78/(1 - pscore),  
 ht = y1 - y0)  
#Manual with normalized weights  
nsw\_dw\_cps <- nsw\_dw\_cps %>%   
 mutate(y1 = (treat\*re78/pscore)/(s1/N),  
 y0 = ((1 - treat)\*re78/(1 - pscore))/(s0/N),  
 norm = y1 - y0)  
nsw\_dw\_cps %>%   
 pull(ht) %>%   
 mean()

## [1] -11682.31

nsw\_dw\_cps %>%   
 pull(norm) %>%   
 mean()

## [1] -6784.387

The ATT under non-normalized weights is -$11,682 and the ATT under normalized weights is -$6,784.

#Trimming the propensity score  
nsw\_dw\_cps <- nsw\_dw\_cps %>%   
 select(-d1, -d0, -y1, -y0, -ht, -norm) %>%   
 filter(!(pscore >= 0.9)) %>%   
 filter(!(pscore <= 0.1))  
N <- nrow(nsw\_dw\_cps)  
#Manual with non-normalized weights using trimmed data  
nsw\_dw\_cps <- nsw\_dw\_cps %>%   
 mutate(d1 = treat/pscore,  
 d0 = (1 - treat)/(1 - pscore))  
s1 <- sum(nsw\_dw\_cps$d1)  
s0 <- sum(nsw\_dw\_cps$d0)  
nsw\_dw\_cps <- nsw\_dw\_cps %>%   
 mutate(y1 = treat \* re78/pscore,  
 y0 = (1 - treat) \* re78/(1 - pscore),  
 ht = y1 - y0)  
#Manual with normalized weights with trimmed data  
nsw\_dw\_cps <- nsw\_dw\_cps %>%   
 mutate(y1 = (treat\*re78/pscore)/(s1/N),  
 y0 = ((1 - treat)\*re78/(1 - pscore))/(s0/N),  
 norm = y1 - y0)  
nsw\_dw\_cps %>%   
 pull(ht) %>%   
 mean()

## [1] 1551.058

nsw\_dw\_cps %>%   
 pull(norm) %>%   
 mean()

## [1] 1350.894

After trimming the propensity scores which are less than 0.1 and more than 0.9, I found that the ATT under non-normalized weights is $1,551 and the ATT under normalized weights is $1,351.

The normalized weights after trimming are positive.

#Callway and Sant'Anna  
pdt <- mean(nsw\_dw\_cps$pscore)  
nsw\_dw\_cps %>%  
 mutate(estimator = (re78 - re75)/pdt \* (treat - pscore) / (1 - pscore)) %>%  
 summarise(mean(estimator),  
 sd(estimator)/sqrt(n()))

## mean(estimator) sd(estimator)/sqrt(n())  
## 1 2406.835 1038.758

Using Callaway & Sant’Anna, I find an ATT of $2,407.