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

```r
nsw <- read_csv("~/Desktop/nsw_mixtape.csv")
```

Part A

```r
# 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

```r
# 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
  )
```
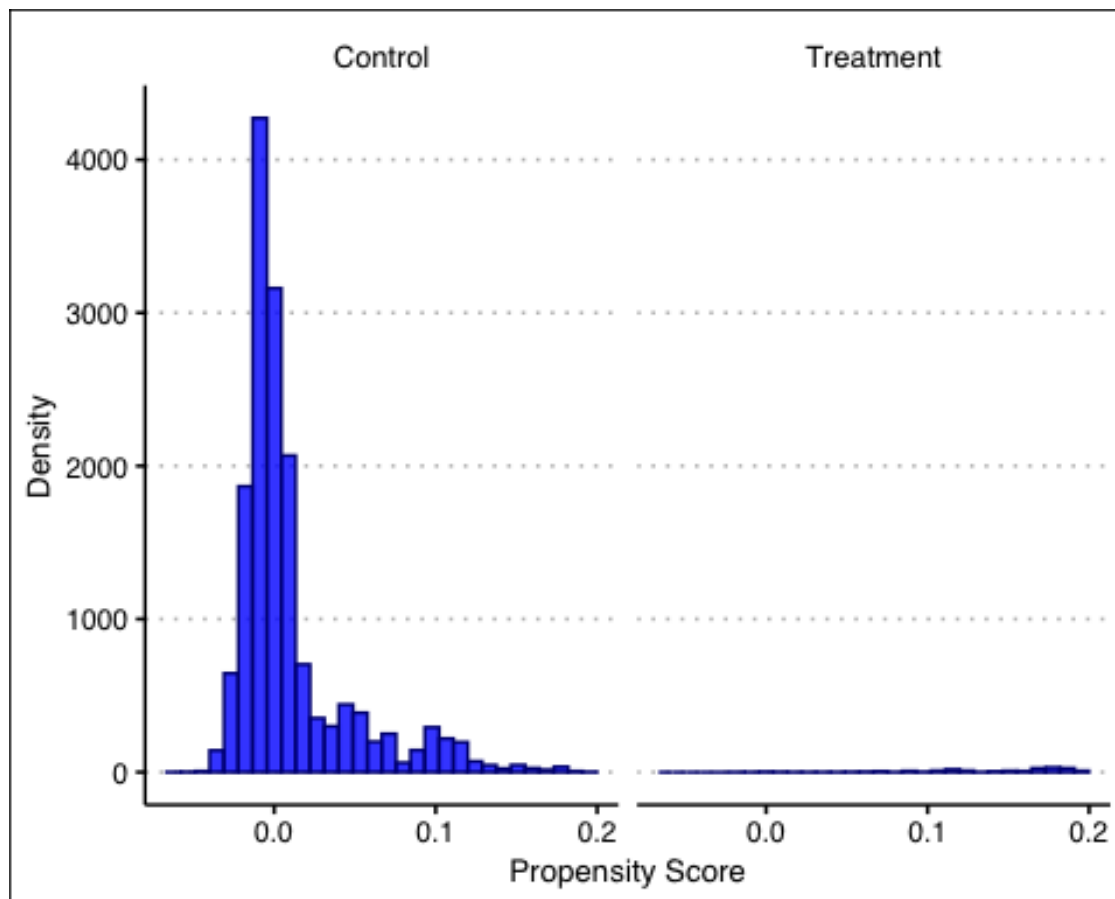
Part C

```r
tx_ctrl_labs <- c("Control", "Treatment")
names(tx_ctrl_labs) <- 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 = tx_ctrl_labs))
```
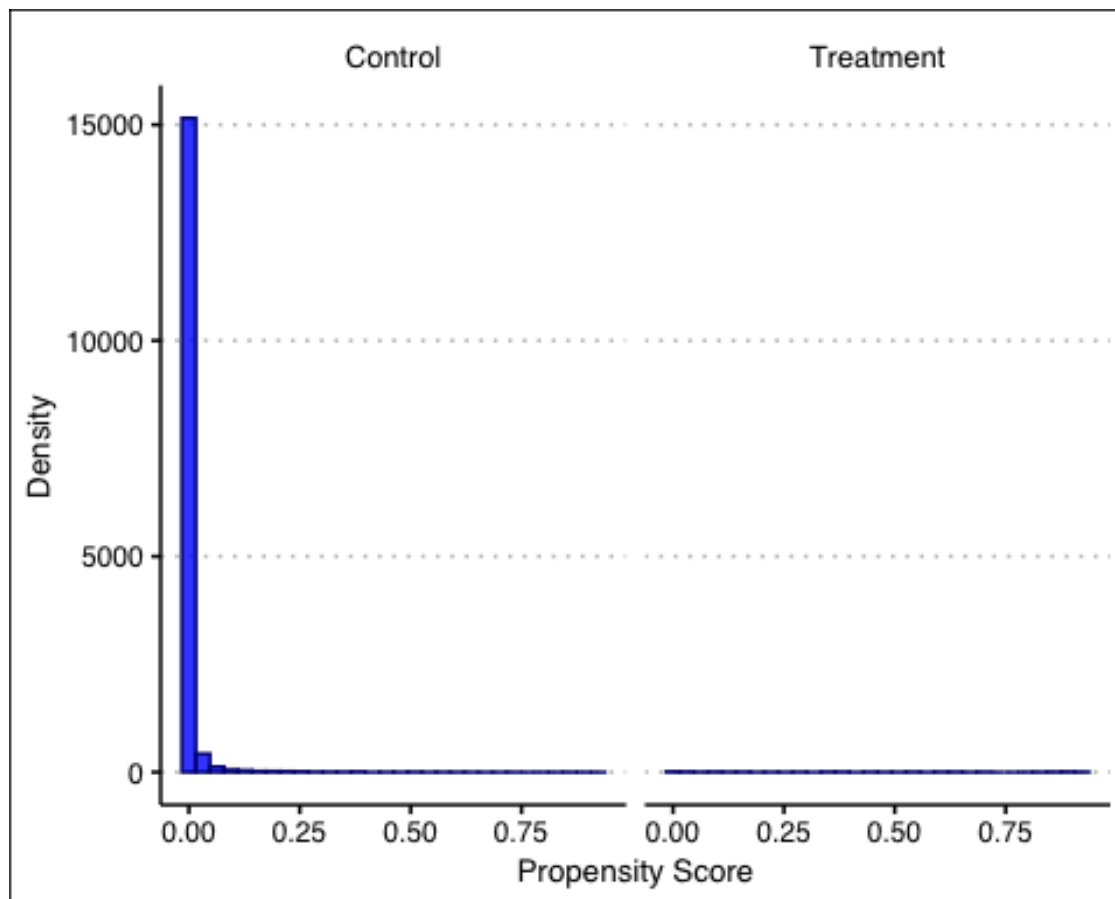
```
# LPM Max and Min
ps_lpm %>%
  group_by(treat) %>%
  summarise(Max = round(max(p_score), 4), Min = round(min(p_score), 4)) %>%
  mutate(treat = if_else(treat == 0, "Control", "Treatment")) %>%
  rename(Treat = treat) %>%
  kbl("pipe")
```

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 = tx_ctrl_labs))
```

```r
# 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")
```
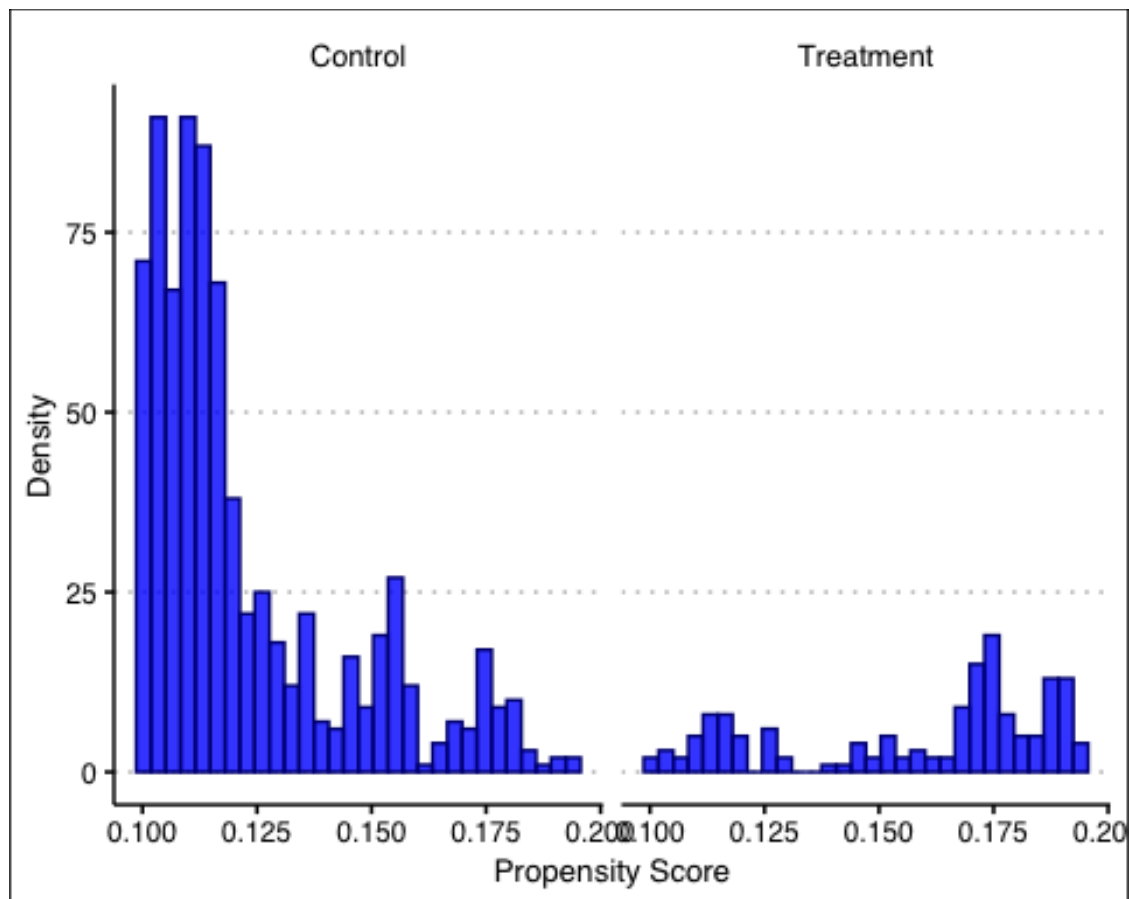
Treat	Max	Min
Control	0.9142	0.000
Treatment	0.9226	0.001

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

```r
# 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 = tx_ctrl_labs))
```



```
# 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")
```

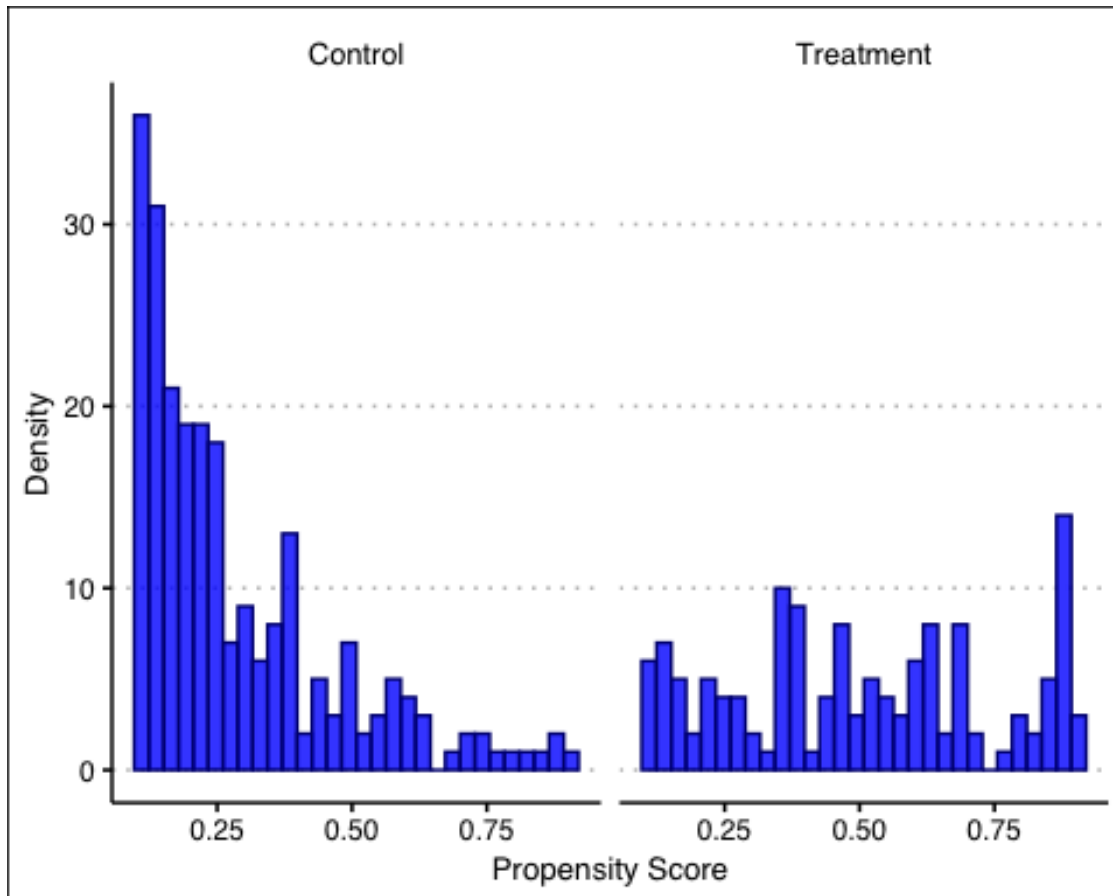| Treat	Max	Min
Control	0.1938	0.1001
Treatment	0.1939	0.1012

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 = tx_ctrl_labs))
```



```
# 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

```r
# 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

```r
#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.