The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police

In McCrary's 2007 AER paper, he argued that the series of court-ordered racial hiring quotas imposed on municipal police departments resulted in an increase of 14-percentage-points in the fraction of African American new hires. But, the evidence on police performance after the change in the composition of police, is mixed.

In this HW, we investigate the first research question, and analyze the relationship between the court-orders and the percentage of African American police officers. Because the data for this AER paper is confidential, this homework uses a simulated dataset prepared by an RA for this Signature Course (Xiaochen Sun). To simplify the settings from the original paper, we set 1970 as the year for the court-order, and provide 5 data points for each city (1960, 1970, 1980, 1990, 1999).

The names and descriptions of variables in the data file `HW3_Data` are

Name	Description
`city_no`	Index of city
`year`	Five data points are provided here: 1960, 1970, 1980, 1990, 1999
`order`	Variable indicating whether a city has the court-ordered hiring quotas (equal to 1) or not (equal to 0)
`police_size`	Number of sworn officers (in logrithm format)
`pop`	Population of the city (Unit is in 100K)
`prop_black`	Proportion of Africen American in population
`prop_police_blk`	Proportion of African American among police officers
`pcnt`	Number of police precincts
`south`	Variable indicating whether a city is in the south (equal to 1) or not (equal to 0)
`region`	Census region

##Functions you will need * table() * prop.table() * group_by() * summarize() * quantile() * mean()

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts -------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# ## Load the data with this chunk
data <- read.csv("data/HW3_Data.csv", header = TRUE)
```

Question 1 [10 pts]

Assume either a pre-post intervention only or post-only cross sectional study design for this question. What is the specific causal question of interest for this study? What are the potential outcomes for a single police department? For cities experiencing the intervention what is the average factual outcome? For cities experiencing the intervention what is the average missing counterfactual outcome? Is this study a randomized experiment or an observational study? Explain why in a sentence or two. How might you estimate the missing counterfactual for the cities experiencing the intervention? What assumption would need to hold for this to be an unbiased estimate?

Answer 1

There are multiple right answers to some parts of this question depending on which study design you had in mind. Full credit will be awarded for answers that are consistent with one of the three possible study designs.

Assume either a pre-post intervention only or post-only cross sectional study design for this question. What is the specific causal question of interest for this study?

CF: Assuming a pre-post intervention, What is the impact of the racial court order hiring quotas imposed on the regions hiring police officers?

What are the potential outcomes for a single police department?

The potential outcomes: With the court order the percentage of African American in the police department could increase/decrease or not having the court order will increase/decrease the percentage of African Americans in the police department.

For cities experiencing the intervention what is the average factual outcome?

The percentage of African American police officer after the court order was issued.

For cities experiencing the intervention what is the average missing counterfactual outcome?

What would happen to the number of hired police officers that are African American if the court order was not established.

Is this study a randomized experiment or an observational study? Explain why in a sentence or two.

Observational: This is an observational study because we are observing if the hiring of African Americans is increased or decreased if the court order is established.

How might you estimate the missing counterfactual for the cities experiencing the intervention?

We may estimate the missing counterfactual for the cities experiencing the intervention using the statistics and comparing the rates of African American in the regions before the court order and after the court order.

What assumption would need to hold for this to be an unbiased estimate?

The assumption we are making is that the current political affiliation or factors similar to this are not going to affect our study.

Question 2 [6 pts]

The court-ordered hiring quotas were imposed to the treated cities in 1970. Describe the average percentage of African American police officers in the post-period (1980) in cities with and without regulation? Give the average proportions and the corresponding standard deviations in tables. Then, explain what the tables show. Do cities without the regulation have higher, lower, or similar proportions of African American police officers, compared to the untreated cities? What about the standard deviation? Use `ggplot`to create boxplots to show the distributions of the percentage of African American police officers in the intervention and non-intervention cities (still just for 1980). What study design did you use here?

Answer 2

The court-ordered hiring quotas were imposed to the treated cities in 1970. Describe the average percentage of African American police officers in the post-period (1980) in cities with and without regulation? Give the average proportions and the corresponding standard deviations in tables. Then, explain what the tables show.

Without the order, The mean is 11.06% and the std Deviation is 11.49 % With the order, The mean is 13.00% and the std Deviation is 0.81 %

The tables show that the std deviation is higher for the case study with the order.

We can observe that the data has deviated from the data set more without the order than it has with the order. The data is more close together when the order is set.

Do cities without the regulation have higher, lower, or similar proportions of African American police officers, compared to the untreated cities? What about the standard deviation?

What study design did you use here?

The study design we used here is Post only, we are only looking at the cities when the order was set in place.

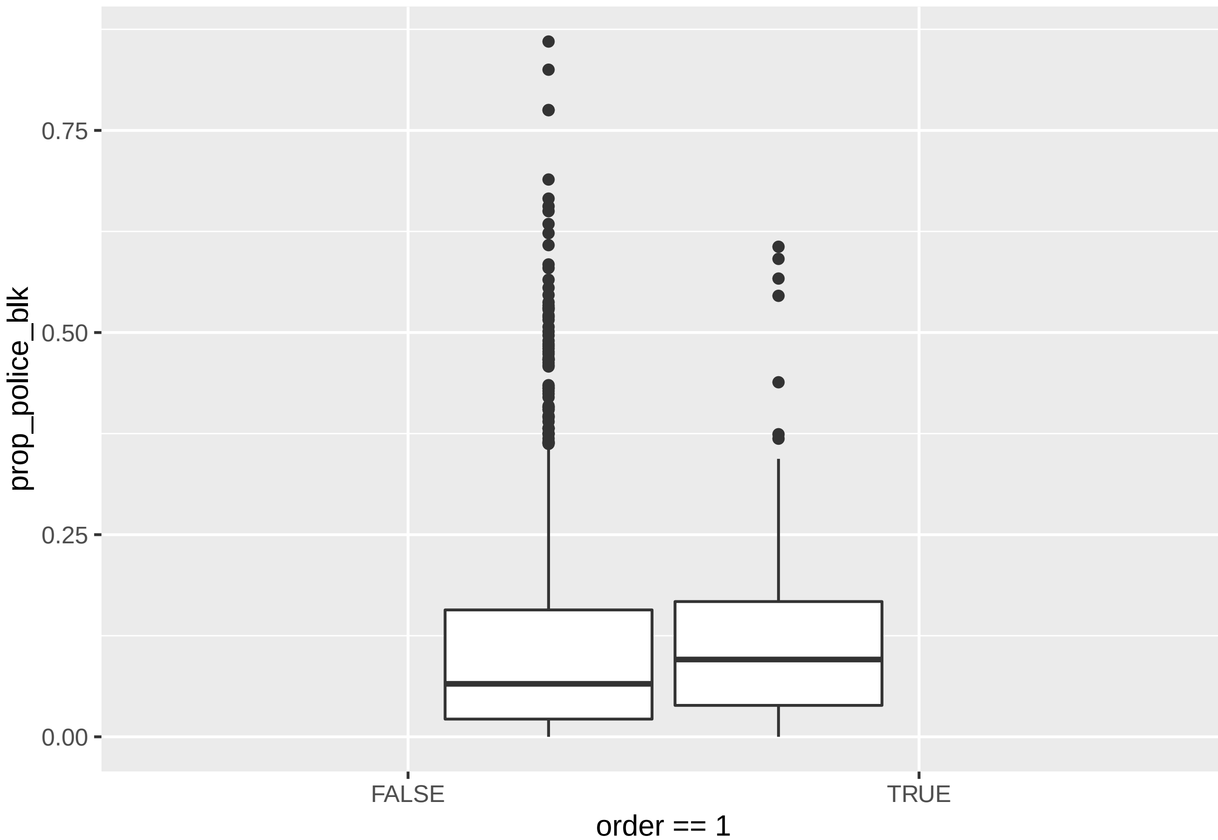
```
#Still 1980's, do we filter by this

data %>%
  filter(year == 1980) %>%
  group_by(order) %>%
  summarize(mean(prop_police_blk), sd(prop_police_blk))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 2 x 3
##   order `mean(prop_police_blk)` `sd(prop_police_blk)`
##   <int>                   <dbl>                 <dbl>
## 1     0                   0.111                0.115
## 2     1                   0.130                0.0810
```

```
data %>%
  ggplot(aes(group=year==1980, y = prop_police_blk, x = order == 1 )) + #notice the + rather than %>%
  geom_boxplot()
```



prop_police_blk
0.75
0.50
0.25
0.00
FALSE
TRUE
order == 1

Question 3 [8 pts]

Consider just cities with court-ordered hiring quotas. Describe the distributions of the percentage of African American police officers before (1960) and after (1980) the imposition of the regulation. Conduct the comparison in terms of the mean, median, and quartiles. Create a table and a figure to show the difference (or lack of a difference) in those two distributions. Calculate the average treatment effect. Explain your results in words. Repeat using 1990 as the post year.

Given the results you estimated, what confounder(s) do you think are mot likely to bias this estimate of the causal effect?

Answer 3

Consider just cities with court-ordered hiring quotas. Describe the distributions of the percentage of African American police officers before (1960) and after (1980) the imposition of the regulation.

Before the 1960's our distributions was more close together and when we studied using the 1980's data we had more outliers compared the 1960's data set.

Conduct the comparison in terms of the mean, median, and quartiles. Create a table and a figure to show the difference (or lack of a difference) in those two distributions. Calculate the average treatment effect. Explain your results in words. Repeat using 1990 as the post year.

Min. 1st Qu. Median Mean 3rd Qu. Max. 0.0006154 0.0186315 0.0349142 0.0493877 0.0697692 0.1899384 Min. 1st Qu. Median Mean 3rd Qu. Max. 0.01014 0.06619 0.11356 0.13002 0.17107 0.36874 Min. 1st Qu. Median Mean 3rd Qu. Max. 0.01014 0.06619 0.11356 0.13002 0.17107 0.36874

The means for the 1980's and 1990's was higher than the results for 1960's. The results for post was showing a better turn out rate than the pre study.

Given the results you estimated, what confounder(s) do you think are most likely to bias this estimate of the causal effect?

The cofounders that could be most biased could be the size of the city or how much of the population of the city is African American or not.

```
#fair assessment
#a figure to show the difference (or lack of a difference) in those two distributions
#side by side box plot for both order




Sum1960 <- summary(data$prop_police_blk[data$year == 1960 & data$order == 1])
print(Sum1960)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0006154 0.0186315 0.0349142 0.0493877 0.0697692 0.1899384
```

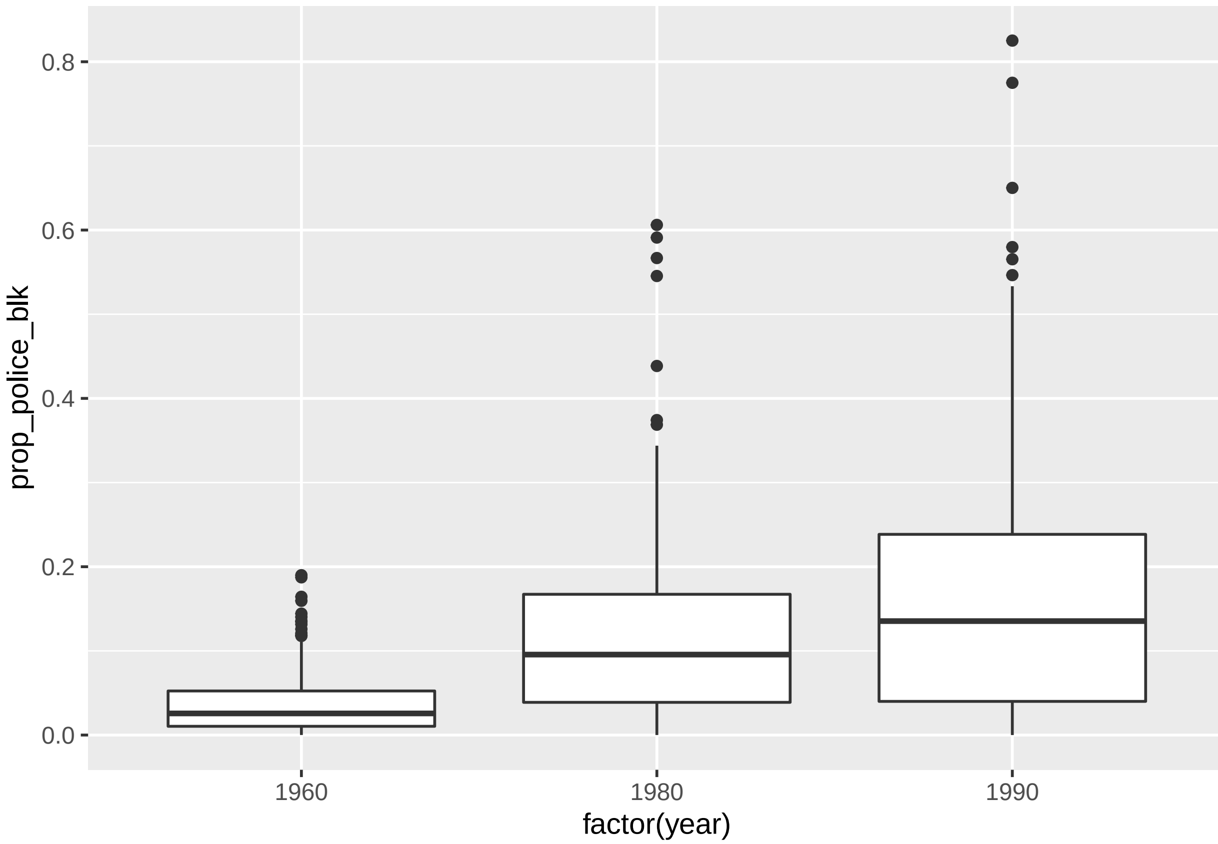
```
Sum1980 <- summary(data$prop_police_blk[data$year == 1980 & data$order == 1])
print(Sum1980)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01014 0.06619 0.11356 0.13002 0.17107 0.36874
```

```
Sum1990 <- summary(data$prop_police_blk[data$year == 1990 & data$order == 1])
print(Sum1980)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01014 0.06619 0.11356 0.13002 0.17107 0.36874
```

```r
data %>%
  filter( year == "1960"| year == "1980" | year == "1990") %>%
  ggplot(aes(x = factor(year), y = prop_police_blk)) +
  geom_boxplot()
```



Question 4 [5 pts]

Now calculate the difference-in-differences estimate. Using a difference-in-differences study design, what is the the specific causal question? How is the average MCF estimated in a DiD design? What is the DiD treatment estimate? How does this estimate compare to your previous results?

Answer 4

Now calculate the difference-in-differences estimate. Using a difference-in-differences study design, what is the the specific causal question?

Using the DiD study design, What the change in African American new hiring rates will be in the police department pre-to-post when the court order isnt set before that time period (1960) and what the change in African American new hires is in the police department pre-to-post when the court order is set after that time period(1980) ?

How is the average MCF estimated in a DiD design?

Using the DiD study design, What is the average change in African American new hiring rates will be in the police department pre-to-post when the court order isnt set before that time period (1960) and what the average change in African American new hires is in the police department pre-to-post when the court order is set after that time period(1980) ?

What is the DiD treatment estimate?

The DID estimate says that more African Americans hired caused a 0.04 increase in the control group vs the treatment group.

How does this estimate compare to your previous results?

The estimate is about the same as to my previous result, th mean is around .04 and DiD values are also around .04.

```
data %>%
  group_by(order) %>%
  summarize(
    co1960 = mean(prop_police_blk[year == 1960]),
    co1980 = mean(prop_police_blk[year >= 1980])) %>%
  mutate(difference = co1980 - co1960) %>%
  mutate(did = difference[order ==1] - difference[order == 0])
```

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 2 x 5
##   order co1960 co1980 difference    did
##   <int>  <dbl>  <dbl>      <dbl>  <dbl>
## 1     0 0.0311  0.134      0.102 0.0428
## 2     1 0.0494  0.195      0.145 0.0428
```

Question 5 [5 pts]

Compare the distribution of region for the intervention group and the non-intervention group. Discuss the internal validity of your estimates of the impact of treatment from Question 2 and Question 4. Discuss the external validity of these estimates.

Answer 5

Compare the distribution of region for the intervention group and the non-intervention group. Discuss the internal validity of your estimates of the impact of treatment from Question 2 and Question 4. Discuss the external validity of these estimates.

The distribution of the groups for the intervention group and the non intervention groups are very close together, since this is not a random experiment and this is an observed experiment the internal validity could be weak and due to this fact, the external validity could also be weak. Having more cofounding variables also makes the internal validity weak.

```
# stem plot
# cat var -- region .. table prop table

# region should show upo
# as factor variable
# as.factor()

prop.table(table(data$region, data$order), margin=2)
```

```
##
##         0     1
##   1 0.200 0.190
##   2 0.215 0.180
##   3 0.175 0.130
##   4 0.205 0.320
##   5 0.205 0.180
```