Replication Study

ASDS Stats Spring 2022

Kathryn Glen | 17331061

11 April 2022

Chosen study for replication:

ENOS, R., KAUFMAN, A., & SANDS, M. (2019). Can Violent Protest Change Local Policy Support? Evidence from the Aftermath of the 1992 Los Angeles Riot. *American Political Science Review,* *113*(4), 1012-1028. doi:10.1017/S0003055419000340

Studying the effect of violent protests on political behaviour. Found the 1992 Los Angeles riot led to a liberal policy support shift.

Rodney King as beaten in March 1991 by four white Los Angeles police officers who were videotaped. An al-white jury acquitted them on all charges on April 29th 1992. This resulted in violent riots began at intersection of Florence Avenue and Normandie Avenue in south central Los Angeles, a predominantly African American neighbourhood. Police withdrew completely like during the Watts Riots thirty years previously until May 3rd 1992 when federal troops arrived. Drew widespread political discussion that living conditions for African Americans in these areas of LA needed improvement.

Load packages:

if(!require(devtools)) install.packages("devtools") # devtools to download packages from github

library(devtools)  
# Load packages   
# (note: you may need to install certain packages first using install.packages(""))  
packages <- c("apsrtable",  
 "car",  
 "dplyr",  
 "ggplot2",  
 "ebal",  
 "ei",  
 "gmm", # I added  
 "eiPack", # I added   
 "foreign",  
 "gdata",  
 "geosphere",  
 "Hmisc",  
 "Matching",  
 "mosaic",  
 "questionr",   
 "optmatch",  
 "reshape2",  
 "RItools",  
 "stargazer",  
 "tidyverse",  
 "weights",  
 "xtable",  
 "gtsummary",  
 "imager")  
lapply(packages, require, character.only = TRUE)

I replicated their tests on the effect of the LA riots on change in public policy support based on race and distance from riots. The section marked mobilization data in the paper

They also tested attitude change, and long-term political patterns which I did not think necessary to address the main subject of their research question and by focusing on mobilization I could keep the presentation straightforward.

I realised during replicating this study:

This study was conducted in 2017, though only published in 2019. As such, a lot of their code I couldn’t run on the packages they used, so needed to adjust and find my own packages since their ones were no longer compatible with this version of R. Should do replications swiftly after publication to prevent this issue arising.

Here is the table of the votes they analyse:

Table

Description automatically generated

They include higher ed as a way of controlling for an increase in general support of education, so EdDiff is change in support for public schools. Public schools were chosen for being more closely linked to African American interests than higher education, which would be “ a very indirect way to address the policy concerns of the rioters.”

Their formula for calculating:

EdDiffi = (PubSchooli1992 - PubSchooli1990) - (HigherEdi1992 - HigherEdi1990)

“PubSchooli1992 indicates precinct i’s support for the public school initiative in 1992, measured as the votes cast in support of that ballot initiative divided by the total ballots cast. The same convention holds for PubSchooli1990, HigherEdi1992, and HigherEdi1990.”

“EdDiffi is the change in support for public schools in precinct i between 1990 and 1992, net the change in support for universities.”

EdDiff is important – the population-weighted mean of EdDiffi for all voters, and for white and African American voters separately.

Most of the work in this went into the preprocessing the data before I could replicate their results.

They used five different datasets - three of which are used in the section of the paper I replicated

1. loading voting data and then processing it by: Calculating the difference in difference

* merging election data into single df
* calculating percentage that support public school (based on Yes votes for progressive policy)
* calculating percentage that support highered
* calculating their difference between 1990 and 1992
* calculating education support difference by taking public school support away from highered support (diff in diff)

# Loading my data from saved CSVs, code can be found in R file.   
  
merged\_final <- read.csv("/Users/Kate/Desktop/Hacker/Stats - HT/Replication/Data/merged\_final.csv")  
out <- read.csv("/Users/Kate/Desktop/Hacker/Stats - HT/Replication/Data/finalout.csv")  
dat <- read.csv("/Users/Kate/Desktop/Hacker/Stats - HT/Replication/Data/dat.csv")

merged\_final$pct\_pubschool\_90 = merged\_final$P123Y/merged\_final$ballots\_cast\_90  
merged\_final$pct\_pubschool\_92 = merged\_final$P152Y/merged\_final$ballots\_cast\_92  
merged\_final$pct\_highered\_90 = merged\_final$P121Y/merged\_final$ballots\_cast\_90  
merged\_final$pct\_highered\_92 = merged\_final$P153Y/merged\_final$ballots\_cast\_92  
  
merged\_final$pubschool\_dif = merged\_final$pct\_pubschool\_92 - merged\_final$pct\_pubschool\_90  
merged\_final$highered\_dif = merged\_final$pct\_highered\_92 - merged\_final$pct\_highered\_90  
merged\_final$edu\_dif = merged\_final$pubschool\_dif - merged\_final$highered\_dif

ECOLOGICAL INFERENCE

Know the total vote outcomes and the racial demographics for each precinct, to understand the distinct voting patterns of different racial groups requires individual-level vote data This presents an “Ecological Inference Problem” (Robinson 1950). We use the Ecological Inference methods developed by King (1997) to isolate behavior by racial group.

The inputs to the EI model are the proportion of whites, African Americans, Hispanics, Asians, and others in each precinct, and the proportion of Yes votes out of total ballots cast for each initiative. The outputs are estimates of what proportion of each group voted Yes on a given ballot initiative, for each precinct in our data.

1. Calculate the EIs

* this took a lot of processing power, so I saved it as a csv as I went as it kept using up vector memory before finishing running
* The outputs are estimates of what proportion of each group voted Yes on a given ballot initiative, for each precinct in our data
* once done bulk out the dataset with newly calculated proportions by setting the data to be proportionate, running EI on it,

1. Post EIs -recalculate -Calculate public difference, higher difference, take higher from pub to calculate education difference

* do this for each race group (white, black, asian, latino, other)

1. Used their code to generate the variance and find the standard errors

* so I would get the exact same dataset as the one they were using to run my own regressions on

Finally - able to calculate the weighted means

Riot as a positive shock for public school funding -> weighted mean positive Riot as negative shock for public school funding -> weighted mean negative

They find an EdDiffi weighted mean of 0.049 (95% confidence interval: [0.037, 0.061]) Positive shock

I find weighted mean 0.049 [0.037, 0.061]. Exactly the same.

]

all\_wtd.mean <- wtd.mean(out$edu\_dif, out$pop18) # pop18 is standard errors  
print(all\_wtd.mean)

## [1] 0.04905395

print(wtd.mean(out$edu\_dif, out$pop18) + 1.97\*sqrt(Hmisc::wtd.var(out$edu\_dif, out$pop18)/length(out))) #upper bound 0.06112077

## [1] 0.06104112

print(wtd.mean(out$edu\_dif, out$pop18) - 1.97\*sqrt(Hmisc::wtd.var(out$edu\_dif, out$pop18)/length(out))) #lower bound 0. 03706679

## [1] 0.03706679

Their weighted means: whites (0.028, CI: [0.018, 0.039]) African Americans (0.073, CI:[0.066, 0.081])

I find: whites (0.029, CI: [0.02, 0.039]) African Americans (0.072, CI:[0.04, 0.023])

These are slightly different results, but not extremely dissimilar. Most can be contributed to rounding, but the oddest one is the upper bounds of the African American Confidence Interval is 0.023 for me, while they found it to be 0.081. Possibly due to using different packages.

white\_wtd.mean <- (wtd.mean(out$white\_edudiff, out$white\_edudiff\_se)) # 0.02972858  
print(white\_wtd.mean)

## [1] 0.02972858

print(wtd.mean(out$white\_edudiff, out$white\_edudiff\_se) + 1.97\*sqrt(Hmisc::wtd.var(out$white\_edudiff, out$white\_edudiff\_se)/length(out))) # 0.03966345

## [1] 0.03959787

print(wtd.mean(out$white\_edudiff, out$white\_edudiff\_se) - 1.97\*sqrt(Hmisc::wtd.var(out$white\_edudiff, out$white\_edudiff\_se)/length(out))) # 0.01979371

## [1] 0.01985928

print(wtd.mean(out$black\_edudiff, out$black\_edudiff\_se))

## [1] 0.07162608

# 0.07162608  
  
black\_wtd.mean <- wtd.mean(out$black\_edudiff, out$black\_edudiff\_se)  
  
print(wtd.mean(out$black\_edudiff, out$black\_edudiff\_se) + 1.97\*sqrt(Hmisc::wtd.var(out$black\_edudiff, out$black\_edudiff\_se)/length(out))) # 0.07902265

## [1] 0.07897383

print(wtd.mean(out$black\_edudiff, out$black\_edudiff\_se) - 1.97\*sqrt(Hmisc::wtd.var(out$black\_edudiff, out$black\_edudiff\_se)/length(out))) # 0.06422951

## [1] 0.06427834

T-TEST

Weighted t test to see whether there is a statistically significant difference between white support change and black support change

print(weights::wtd.t.test(x=out$white\_edudiff, y = out$black\_edudiff, weight = out$white\_edudiff\_se, weighty = out$black\_edudiff\_se))

## $test  
## [1] "Two Sample Weighted T-Test (Welch)"  
##   
## $coefficients  
## t.value df p.value   
## -31.49325 3095.44823 0.00000   
##   
## $additional  
## Difference Mean.x Mean.y Std. Err   
## -0.041897503 0.029728578 0.071626080 0.001330364

# weighted t\_test: pvalue = 0.00000  
# Very statistically significant   
# Can reject the null hypothesis that there is no difference between   
# Sufficient evidence to think the mean levels in support for education is different between white and black populations in Los Angeles

Regressions

# lm(y~x)  
# x = causal (independent), y = effected (dependent)  
  
whiteed <- lm(out$edu\_dif~out$whitepct) # (Intercept) out$whitepct 0.0633 -0.0434   
blacked <- lm(out$edu\_dif~out$blackpct) # Intercept) out$blackpct 0.03398 0.05044

library(gtsummary)

Black pct in district and education support

plot(out$blackpct, out$edu\_dif)

Chart, scatter chart

Description automatically generated

plot(out$whitepct, out$edu\_dif)

Chart, scatter chart

Description automatically generated Plot Comparing Black and White Education Support Difference

plot(out$whitepct, out$edu\_dif, col="red", main = "Comparing Black and White Education Support Difference", xlab = "Race %", ylab = "Education Difference")  
points(out$blackpct, out$edu\_dif,col="blue")  
legend("bottomright", c("White %","Black %"),cex=.8,col=c("red","blue"),pch=c(1,2))

Chart, scatter chart

Description automatically generated

Clear division - black more likely to experience a positive change supporting education, while white less likely. Bunching on the left just tells me that most people don’t live in areas with a high percentage of black voters - this makes sense especially in the 1990s with urbanisation vs suburbanisation and black people tending to live in concentrated densely packed areas in the inner cities which were highly segregated by race.

Check this with coefficients

tbl\_regression(blacked)

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Beta** | **95% CI** | **p-value** |
| out$blackpct | 0.05 | 0.04, 0.06 | <0.001 |

for every pct increase in black population in a precinct, support for education increased by 0.05

tbl\_regression(whiteed)

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Beta** | **95% CI** | **p-value** |
| out$whitepct | -0.04 | -0.05, -0.04 | <0.001 |

# for every pct increase in white voters in a precinct, support for education decreased by -0.04

Figure 2 from the paper:

FIGURE 2. A Loess Line and Scatterplot Displaying the Relationship Between Each Precinct’s Distance From Florence and Normandie and that Precinct’s EdDiff Values for all Voters in 1,676 Precincts in the LA Basin. Points are Sized, and the Loess Line Weighted, by the Voting Age Population in Each Precinct

Chart, scatter chart

Description automatically generated

Figure C4 from their Appendix showing distance based on race.

Chart, scatter chart

Description automatically generated

Effect of Distance from Riot on support for public policy change

tbl\_regression(lm(edu\_dif ~ dist, data=dat, weight=dat$pop18))

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Beta** | **95% CI** | **p-value** |
| dist | -0.04 | -0.04, -0.03 | <0.001 |
|  |  |  |  |

# for every 10km away from riots support for public policy change decreases by 0.04

Measure in 10s of kilometers from riots

par(mfrow=c(3,1))  
plot(edu\_dif ~ dist, data=dat, weight=dat$pop18) # in area

plot(edu\_dif ~ dist + I(dist^2), data=dat, weight=dat$pop18) # further from area

plot(edu\_dif ~ dist + I(dist^2) + I(dist^3), data=dat, weight=dat$pop18) # further from area

Diagram

Description automatically generated

This also shows that most people live close to the riots. As shown in this Google Maps image LA is in a valley and bordered by the sea, most people live within 100km:

Map

Description automatically generated

# regression of effect of distance and % black pop in precinct on edu\_dif  
  
summary(lm(dat$edu\_dif ~ dat$dist + dat$blackpct))

## Call:  
## lm(formula = dat$edu\_dif ~ dat$dist + dat$blackpct)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.242832 -0.032545 0.001077 0.031269 0.177000   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.060855 0.005143 11.832 < 2e-16 \*\*\*  
## dat$dist -0.020913 0.003817 -5.478 4.95e-08 \*\*\*  
## dat$blackpct 0.029303 0.005461 5.366 9.19e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04863 on 1673 degrees of freedom  
## Multiple R-squared: 0.1069, Adjusted R-squared: 0.1059   
## F-statistic: 100.1 on 2 and 1673 DF, p-value: < 2.2e-16

# dat$dist -0.020913 pvalue 4.95e-08 \*\*\*  
# dat$blackpct 0.029303 pvaleu 9.19e-08 \*\*\*

# interaction of effect of distance and % black pop in precinct on edu\_dif  
# interaction - p value = 0.5977  
summary(lm(dat$edu\_dif ~ dat$dist \* dat$blackpct))

##   
## Call:  
## lm(formula = dat$edu\_dif ~ dat$dist \* dat$blackpct)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.244222 -0.032306 0.001228 0.031415 0.176916   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.062016 0.005595 11.084 < 2e-16 \*\*\*  
## dat$dist -0.021965 0.004307 -5.100 3.79e-07 \*\*\*  
## dat$blackpct 0.024654 0.010367 2.378 0.0175 \*   
## dat$dist:dat$blackpct 0.006638 0.012578 0.528 0.5977   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04864 on 1672 degrees of freedom  
## Multiple R-squared: 0.1071, Adjusted R-squared: 0.1055   
## F-statistic: 66.83 on 3 and 1672 DF, p-value: < 2.2e-16

While being in a precinct with a high black % and being close to the riots does have a positive effect on education support, they do not have an effect in combination. I suppose this might be because most precincts with high % of black people were close to the riots.

plot(dat$dist, dat$blackpct, main = "Black % and Distance from Riots")

Chart, scatter chart

Description automatically generated

There is a negative trend of black population percentage and distance from riots. Slightly obscured by again fact that most people don’t live in area with a high percentage of black people.