Culture and Institutions: Data Task 2

Caroline Belka, Carmen Cano, Jonas Wallstein

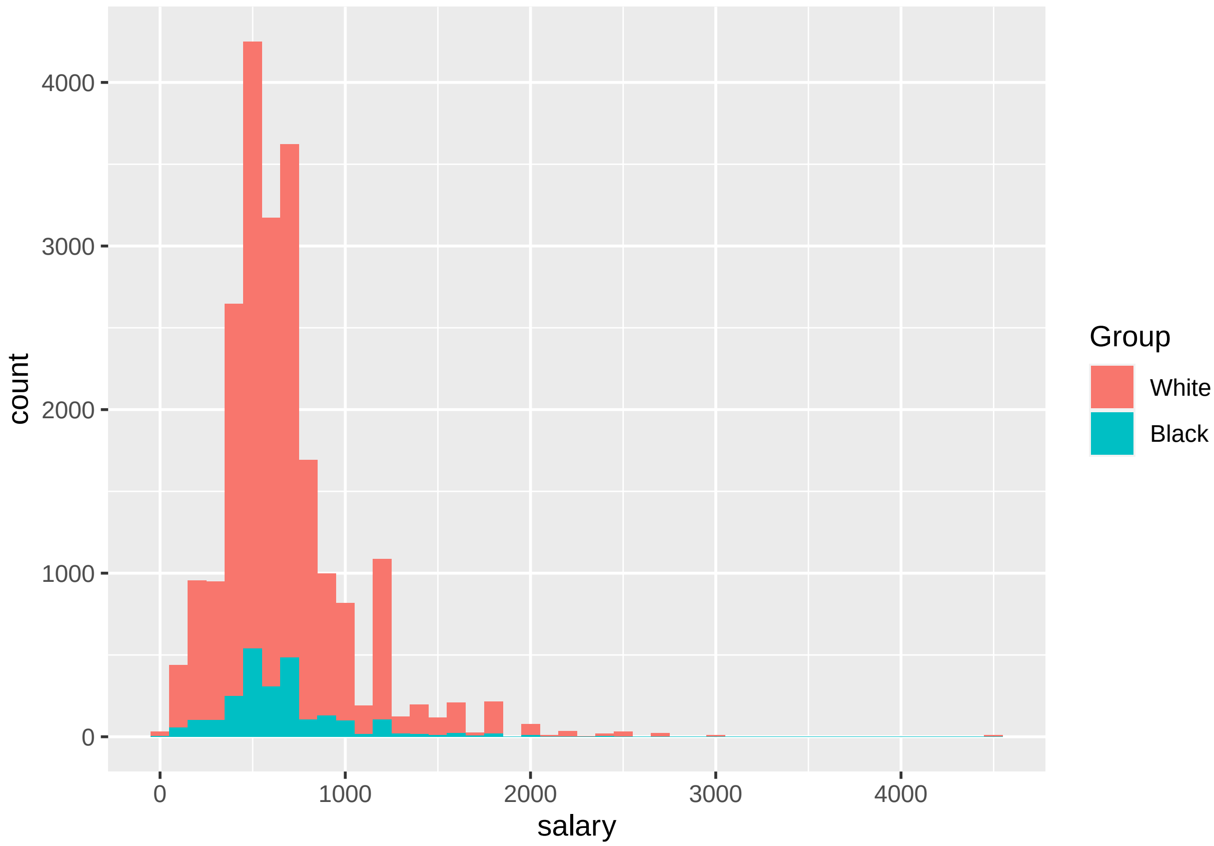
2023-06-04

```r
rm(list = ls())
#load packages
library(tidyverse)
library(ggplot2)
library(scales)
library(modelsummary)
library(fixest)
library(estimatr)


wage_panel <- read_delim("./data_task2.csv")
```

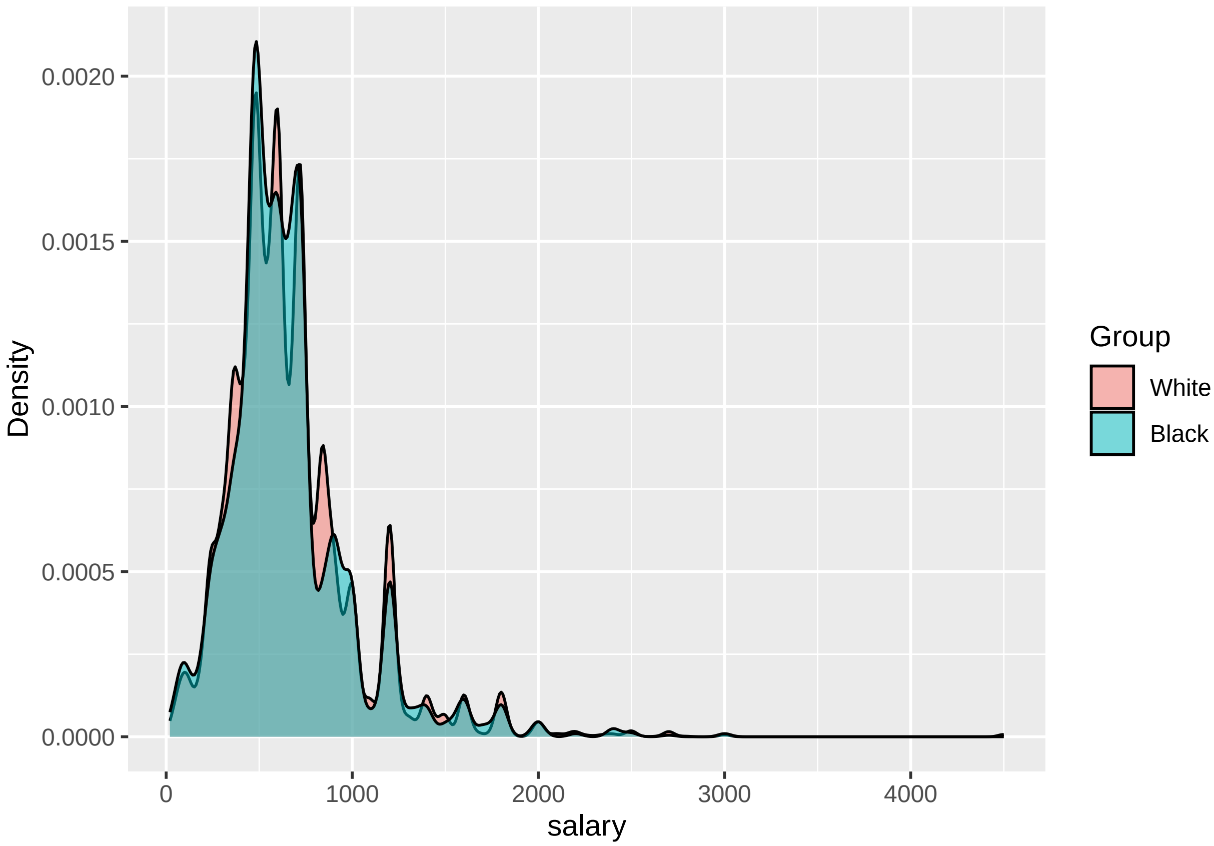
1

1a

```r
wage_panel_1911 <- filter(wage_panel, year == 1911 & salary < 10000)

# With absolute count as y axis
gg1 <-  ggplot(wage_panel_1911, aes(x=salary, fill = factor(black))) +
  geom_histogram(binwidth = 100) +
  scale_fill_discrete(name = "Group", labels = c("White", "Black"))
plot(gg1)
```




```
# Alternatively density relative to the own group
gg2 <- ggplot(wage_panel_1911, aes(x = salary, fill = factor(black))) +
  geom_density(alpha = 0.5) +
  scale_fill_discrete(name = "Group", labels = c("White", "Black")) +
  ylab("Density")
plot(gg2)
```



1b

```
wage_panel_1921 <- filter(wage_panel, year == 1921 & salary < 10000)

# With absolute count as y axis
gg3 <-  ggplot(wage_panel_1921, aes(x=salary, fill = factor(black))) +
  geom_histogram(binwidth = 100) +
  scale_fill_discrete(name = "Group", labels = c("White", "Black"))
plot(gg3)
```


```r
# Alternatively density relative to the own group
gg4 <- ggplot(wage_panel_1921, aes(x = salary, fill = factor(black))) +
  geom_density(alpha = 0.5) +
  scale_fill_discrete(name = "Group", labels = c("White", "Black")) +
  ylab("Density")
plot(gg4)
```

1c

In 1911, the distribution seems equal between black and white workers. In 1921, although all workers have a higher income on average, white workers reached a higher salary level relative to black workers.

2

```
sum1 <- datasummary_balance( ~ black, wage_panel_1911)
sum1
```

	0		1			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	Std. Error
year	1911.0	0.0	1911.0	0.0	0.0	0.0
salary	669.7	360.1	659.8	344.8	-9.9	7.5
ln_salary	6.4	0.5	6.4	0.6	0.0	0.0
age_in_1910	36.6	12.1	36.7	11.7	0.1	0.3
id	274732.1	156420.2	279373.6	157926.0	4641.5	3395.9
paid_per_annum	0.5	0.5	0.5	0.5	0.0	0.0
paid_per_month	0.1	0.3	0.1	0.3	0.0	0.0
paid_per_day	0.3	0.5	0.3	0.5	0.0	0.0
log_rank	7.7	0.9	7.6	1.0	0.0	0.0
female	0.2	0.4	0.2	0.4	0.0	0.0
works_in_DC	0.5	0.5	0.5	0.5	0.1	0.0

None of the variables seem to have significant differences.

3

```
wage_panel$post = ifelse(wage_panel$year >= 1913, 1, 0)

ln_salary = aggregate(x = wage_panel$ln_salary,
                      by = list(wage_panel$black, wage_panel$post), FUN = mean) %>%
  rename(black = Group.1, post = Group.2, ln_salary = x)
ln_salary
```

```
##   black post ln_salary
## 1     0    0  6.415327
## 2     1    0  6.389987
## 3     0    1  6.703792
## 4     1    1  6.602146
```

4

4a

```
first_diff = ln_salary[4,3] - ln_salary[2,3]
first_diff
```

```
## [1] 0.2121597
```

The salary increases for black workers after 1913. The simple difference estimator is not a suitable estimator because it does not isolate the policy effect but also considers other factors such as a general increase in wages for all groups. Consequently, it is not a good approach to estimate the causal effect of the policy intervention.

4b

```
did = (ln_salary[4,3] - ln_salary[2,3]) - (ln_salary[3,3] - ln_salary[1,3])
did
```

```
## [1] -0.07630512
```

For both groups the average salary increases, yet for black workers the wages increase was smaller than for white workers, due to the policy. As long as the trends are parallel before 1913 this seems to be a valid estimator.

5

5a

```
lm1 <- lm_robust(ln_salary ~ black*post, data = wage_panel)
summary(lm1)
```

```
##
## Call:
## lm_robust(formula = ln_salary ~ black * post, data = wage_panel)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error  t value  Pr(>|t|) CI Lower  CI Upper    DF
## (Intercept)  6.41533   0.002775 2312.033 0.000e+00  6.40989  6.420765 96095
## black       -0.02534   0.008118   -3.121 1.800e-03 -0.04125 -0.009429 96095
```

```
## post         0.28846   0.003798   75.946 0.000e+00  0.28102  0.295909 96095
## black:post  -0.07631   0.010740   -7.105 1.211e-12 -0.09735 -0.055256 96095
##
## Multiple R-squared:  0.06198 ,   Adjusted R-squared:  0.06195
## F-statistic:  2110 on 3 and 96095 DF,  p-value: < 2.2e-16
```

5b

For black workers the policy decreases the wage by 7.6% on average, relative to white workers. The effect is significant at the 1% level.

6

6a

Year fixed effects capture factors that affect both black and white workers in a given year. In the setting this could be the effect of world war I on the labor market or the change in industry practices that altered the labor market.

6b

```r
fem1 <- feols(ln_salary ~ black + black:post | year, wage_panel, vcov = "hetero")
etable(fem1, tex = T)
```

Dependent Variable:	ln_salary
Model:	(1)
Variables	
black	-0.0272***
	(0.0081)
black × post	-0.0796***
	(0.0105)
Fixed-effects	
year	Yes
Fit statistics	
Observations	96,099
R^2	0.12382
Within R^2	0.00241

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

7

7a

Individual fixed effects capture factors such as ability, motivation, health, and others that could explain why a workers salary deviates from the mean of all workers. In the regressions before where aggregated on the level of black and white workers such that we could only differentiate structural differences between the two groups. With individual fixed effect we cannot only capture structural but also individuals differences (yet we cannot separate the two)

7b

```
fem2 <- feols(ln_salary ~ black:post | year + id, wage_panel, vcov = "hetero")
etable(fem2, tex = T)
```

Dependent Variable:	ln_salary
Model:	(1)
Variables	
black × post	-0.0666***
	(0.0058)
Fixed-effects	
year	Yes
id	Yes
Fit statistics	
Observations	96,099
R^2	0.87323
Within R^2	0.00185

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

8

8a

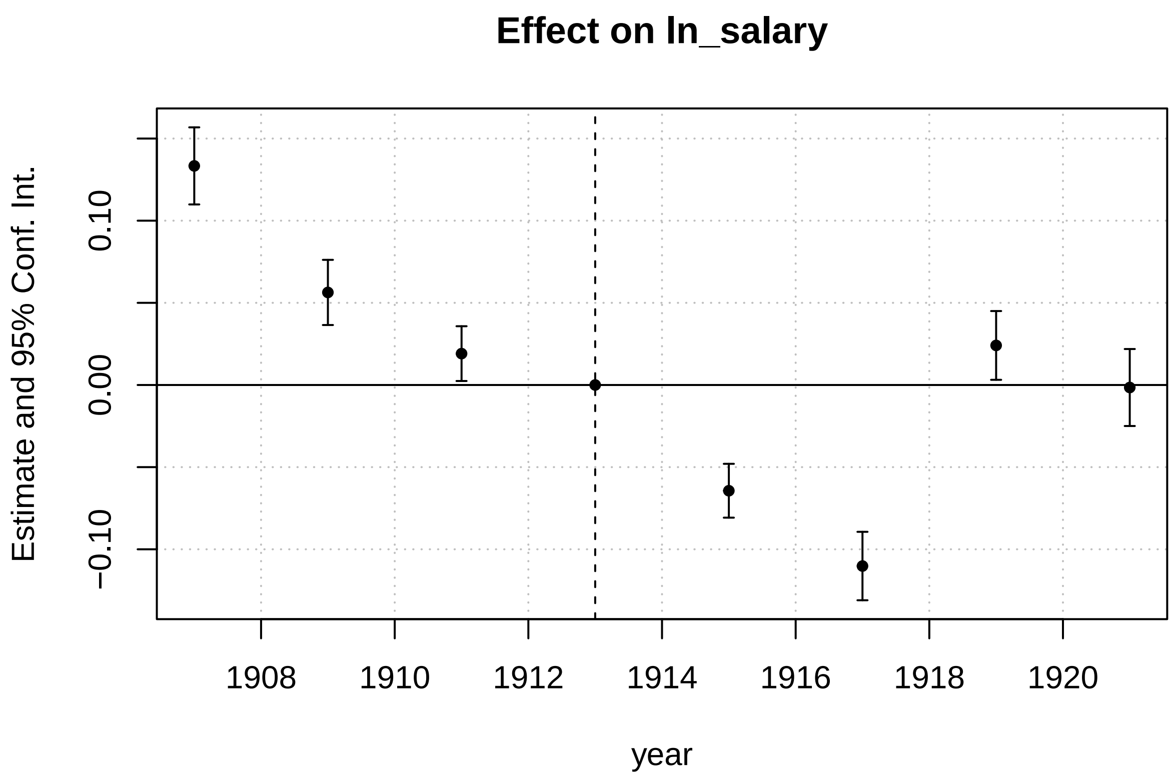
```
library(estimatr)
fem3 <- feols(ln_salary ~ i(year, black, ref = 1913) + year | id, wage_panel, vcov = "hetero")
etable(fem3, tex = TRUE)
```

Dependent Variable:	ln_salary
Model:	(1)
Variables	
black × year = 1907	0.1333***
	(0.0120)
black × year = 1909	0.0563***
	(0.0101)
black × year = 1911	0.0191**
	(0.0085)
black × year = 1915	-0.0644***
	(0.0084)
black × year = 1917	-0.1102***
	(0.0106)
black × year = 1919	0.0241**
	(0.0107)
black × year = 1921	-0.0015
	(0.0120)
year	0.0381***
	(0.0003)
Fixed-effects	
id	Yes
Fit statistics	
Observations	96,099
R^2	0.86713
Within R^2	0.28471

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

8b

```
iplot(fem3)
```



8c

The year specific wage effects for black individuals seem to decrease over the years. This decrease seems to take place even before 1913. This could be explained by a general dismay in reaction to more black workers in factories that eventually lead to the discriminatory policy. In the four years after the policy this negative effect seems to be particularly harsh, relative to 1913. Starting in 1917, the "Harlem Renaissance" could have lead to more resistance against the discriminatory policy and the wages could have increased again compared to 1913.

9

```
etable(fem1, fem2, fem3, tex = TRUE)
```

Dependent Variable:	ln_salary		
Model:	(1)	(2)	(3)
Variables			
black	-0.0272***		
	(0.0081)		
black × post	-0.0796***	-0.0666***	
	(0.0105)	(0.0058)	
black × year = 1907			0.1333***
			(0.0120)
black × year = 1909			0.0563***
			(0.0101)
black × year = 1911			0.0191**
			(0.0085)
black × year = 1915			-0.0644***
			(0.0084)
black × year = 1917			-0.1102***
			(0.0106)
black × year = 1919			0.0241**
			(0.0107)
black × year = 1921			-0.0015
			(0.0120)
year			0.0381***
			(0.0003)
Fixed-effects			
year	Yes	Yes	
id		Yes	Yes
Fit statistics			
Observations	96,099	96,099	96,099
R^2	0.12382	0.87323	0.86713
Within R^2	0.00241	0.00185	0.28471

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*