ECON 121 FA23 Problem Set 4

Robert Tso

Question 1

Verbal: list group members.

Robert Tso - A13829791

Tomas Lopez - A16798775

Stephanie Nguyen - A16215540

Akash Juwadi - A16372772

ChatGPT

Question 2

Code: Load packages and dataset, summarize data.

Verbal: Interpret the summary statistics.

Notable mean differences between the headstart participating children are sibdiff of 0.350750326, showing that children who are in headstart have a third likelihood to have siblings also in headstart. There are also 25% more black children but 4.7% less hispanic children. The largest impact is comp_score_5to6 with a mean difference of -5.560310055 and median difference of-6.0, showing that headstart children have lower test scores in this bracket. Logged income or lninc_0to3 with mean of -0.421336775 and median of -.5775361, showing that headstart children have 0.5 less logged income.

```
# The PDF will show the code you write here but not the output.
# Load packages and dataset here.

#install.packages("plm")

rm(list=ls())
library(tidyverse)
library(fixest)
library(plm)

load(url("https://github.com/tvogl/econ121/raw/main/data/nlsy_kids.Rdata"))
```

```
# The PDF will show the code AND output here.
# Summarize the data here.



#looking at the amount of missing data to determine which ones are worth dropping.
colSums(is.na(nlsy_kids))/nrow(nlsy_kids)
```

```
##        head_start           sibdiff            mom_id          hispanic
##        0.00000000        0.00000000        0.00000000        0.00000000
##             black              male         firstborn        lninc_0to3
##        0.00000000        0.00000000        0.00000000        0.05111372
##             momed      dadhome_0to3           ppvt_3              lnbw
##        0.00140680        0.37584994        0.84196952        0.03399766
##   comp_score_5to6  comp_score_7to10 comp_score_11to14            repeat
##        0.43259086        0.23892145        0.32450176        0.24056272
##          learndis            hsgrad          somecoll              idle
##        0.02837046        0.25252052        0.25252052        0.25275498
##          fphealth
##        0.25252052
```

```
#ppvt_3 is missing 80%+, dropping would result in most of the data being missing

nlsy_kids<-nlsy_kids %>% drop_na(lninc_0to3, momed, lnbw,
                      comp_score_5to6 ,learndis)

view(nlsy_kids)

summary(nlsy_kids)
```

```
##    head_start        sibdiff           mom_id         hispanic
##  Min.   :0.0000   Min.   :0.0000   Min.   :    3   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 3197   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median : 6104   Median :0.0000
##  Mean   :0.2097   Mean   :0.2249   Mean   : 5996   Mean   :0.2032
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.: 8673   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :12667   Max.   :1.0000
##
##      black             male          firstborn        lninc_0to3
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   : 3.909
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 9.643
##  Median :0.0000   Median :1.0000   Median :0.0000   Median :10.189
##  Mean   :0.3254   Mean   :0.5117   Mean   :0.3198   Mean   :10.132
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:10.649
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :13.423
##
##      momed        dadhome_0to3        ppvt_3            lnbw
##  Min.   : 1.00   Min.   :0.0000   Min.   :  0.00   Min.   :1.792
##  1st Qu.:11.00   1st Qu.:0.0000   1st Qu.: 12.00   1st Qu.:4.644
##  Median :12.00   Median :1.0000   Median : 19.00   Median :4.754
##  Mean   :12.08   Mean   :0.6663   Mean   : 22.37   Mean   :4.728
##  3rd Qu.:13.00   3rd Qu.:1.0000   3rd Qu.: 31.00   3rd Qu.:4.852
##  Max.   :20.00   Max.   :1.0000   Max.   :101.00   Max.   :5.434
##                  NA's   :419      NA's   :1757
##  comp_score_5to6 comp_score_7to10 comp_score_11to14     repeat
##  Min.   : 0.00   Min.   : 0.00    Min.   : 0.6667   Min.   :0.0000
##  1st Qu.:29.67   1st Qu.:26.67    1st Qu.:25.3333   1st Qu.:0.0000
##  Median :44.67   Median :46.00    Median :44.5000   Median :0.0000
##  Mean   :45.75   Mean   :45.94    Mean   :45.3677   Mean   :0.2738
##  3rd Qu.:63.00   3rd Qu.:64.83    3rd Qu.:64.0000   3rd Qu.:1.0000
##  Max.   :98.50   Max.   :99.00    Max.   :99.0000   Max.   :1.0000
##                  NA's   :317      NA's   :589       NA's   :427
##     learndis           hsgrad          somecoll           idle
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :1.0000   Median :0.0000   Median :0.0000
##  Mean   :0.04506   Mean   :0.6902   Mean   :0.2995   Mean   :0.1557
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##                    NA's   :368      NA's   :368      NA's   :368
##     fphealth
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.0943
##  3rd Qu.:0.0000
##  Max.   :1.0000
##  NA's   :368
```

```r
nlsy_kids_HS <- subset(nlsy_kids, head_start == 1)
nlsy_kids_NHS <- subset(nlsy_kids, head_start == 0)


#comparing the difference of mean and median of the two subsets of head_start
```

```r
# Subset the dataset for head_start = 1
subset_head_start_1 <- subset(nlsy_kids, head_start == 1)

# Subset the dataset for head_start = 0
subset_head_start_0 <- subset(nlsy_kids, head_start == 0)

# Get the names of all variables in the dataset
all_variables <- names(nlsy_kids)

# Initialize vectors to store differences
mean_differences <- c()
median_differences <- c()

# Loop through variables
for (variable in all_variables) {
  # Skip non-numeric variables
  if (!is.numeric(nlsy_kids[[variable]])) {
    next
  }

  # Calculate mean and median for head_start = 1
  mean_head_start_1 <- mean(subset_head_start_1[[variable]])
  median_head_start_1 <- median(subset_head_start_1[[variable]])

  # Calculate mean and median for head_start = 0
  mean_head_start_0 <- mean(subset_head_start_0[[variable]])
  median_head_start_0 <- median(subset_head_start_0[[variable]])

  # Calculate differences
  mean_difference <- mean_head_start_1 - mean_head_start_0
  median_difference <- median_head_start_1 - median_head_start_0

  # Store differences in vectors
  mean_differences <- c(mean_differences, mean_difference)
  median_differences <- c(median_differences, median_difference)
}

# Create a data frame to display the results
results <- data.frame(
  Variable = all_variables,
  MeanDifference = mean_differences,
  MedianDifference = median_differences
)

# Print the results, head_start minus non_head_start
print(results)
```

```
##             Variable MeanDifference MedianDifference
## 1         head_start    1.000000000     1.000000e+00
## 2            sibdiff    0.350750326     1.000000e+00
## 3             mom_id  764.634986226     1.285000e+03
## 4           hispanic   -0.047978288     0.000000e+00
## 5              black    0.254929680     1.000000e+00
```

```
## 6              male    0.045327679     0.000000e+00
## 7         firstborn   -0.025522872     0.000000e+00
## 8        lninc_0to3   -0.421336775    -5.775361e-01
## 9             momed   -0.210132123     0.000000e+00
## 10     dadhome_0to3             NA               NA
## 11           ppvt_3             NA               NA
## 12            lnbw    0.005308237    -8.584023e-03
## 13  comp_score_5to6   -5.560310055    -6.000000e+00
## 14 comp_score_7to10             NA               NA
## 15 comp_score_11to14            NA               NA
## 16           repeat             NA               NA
## 17         learndis    0.003112766     0.000000e+00
## 18           hsgrad             NA               NA
## 19         somecoll             NA               NA
## 20             idle             NA               NA
## 21         fphealth             NA               NA
```

Question 3

Code: Regression.

Verbal: Interpret.

Headstart has a -5.560 coefficient on test scores implying that by enrolling your child in a head start would result in a test score decrease by 5.6. With a t value of -4.888 the coefficient is statistically significant. It is reasonable to assume exogeneity, however it is likely there are omitted variables. We have to consider that the purpose of the headstart program is to help disadvantaged children with their education, so those enrolled in headstart may already have been on track to test lower scores down the line due to other environmental factors. This bias may be causing the coefficient to be much lower than it could be, since as it currently stands, the treatment seems worse than the symptom.

```
# All question 3 code here.

# Run OLS regression

ols_model <- feols(comp_score_5to6 ~ head_start, data = nlsy_kids)
summary(ols_model)
```

```
## OLS estimation, Dep. Var.: comp_score_5to6
## Observations: 2,308
## Standard-errors: IID
##             Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 46.91155   0.520966 90.04719 < 2.2e-16 ***
## head_start  -5.56031   1.137640 -4.88758 1.091e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 22.2   Adj. R2: 0.009824
```

```
#boxplot(comp_score_5to6 ~ head_start, data = nlsy_kids, main = "Comparison of Test Scores by Head Star
```

Question 4

Code: Regression.

Verbal: Interpret.

The coefficient if still in the same order of magnitude and direction, but it has become larger in effect, -7.2604 compared to -5.560, showing that when accounting for the same family under the same mother, the effect has become stronger.

```
# All question 4 code here


# Create a new data frame with family-level means
family_means <- nlsy_kids %>%
  group_by(mom_id) %>%
  summarise(mean_head_start = mean(head_start),
            mean_test_scores = mean(comp_score_5to6))

# Run OLS regression
ols_model_family <- feols(mean_test_scores ~ mean_head_start,
                          data = family_means)

# Summarize the model
summary(ols_model_family)
```

```
## OLS estimation, Dep. Var.: mean_test_scores
## Observations: 1,393
## Standard-errors: IID
##                 Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)     47.38902   0.617196 76.78117  < 2.2e-16 ***
## mean_head_start -7.26036   1.442797 -5.03214 5.4838e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 20.2   Adj. R2: 0.017173
```

Question 5

Code: Regression.

Verbal: Interpret.

By using OLS with fixed effects of mother ID, the coefficient of head start becomes positive with a magnitude similar to the one found in question 4. Under the assumption that the Head start program should improve test scores, we would take this coefficient to be one that most likely reflects the effect of the program, as in previous analyses it has been returning coefficients that seem counter-intuitive to the intentions of the treatment.

```
# All question 5 code here

ols_model_fe <- feols(comp_score_5to6 ~ head_start | mom_id,
                      data = nlsy_kids)
summary(ols_model_fe)
```

```
## OLS estimation, Dep. Var.: comp_score_5to6
## Observations: 2,308
## Fixed-effects: mom_id: 1,393
## Standard-errors: Clustered (mom_id)
##            Estimate Std. Error t value   Pr(>|t|)
## head_start  7.47425     2.0887 3.57841 0.00035755 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.4     Adj. R2: 0.449959
##              Within R2: 0.015901
```

Question 6

Code: Regression.

Verbal: Interpret.

The coefficient of head start becomes negative again, but with a much smaller magnitude of 1.3095 demonstrating that when accounting for the ppvt test taken at age 3, head start has a much weaker predictive power on the age 5-6 test scores, and again has a negative effect as shown in questions 3 and 4. As the coefficient has swung to such a wide degree and has a standard error of 15 the robustness of the original regression may be called into question. When accounting separately for learning disability, robustness remains similar.

```
# All question 6 code here

ols_model_cov_ppvt <- feols(comp_score_5to6 ~ head_start + ppvt_3 | mom_id,
                      data = nlsy_kids, vcov = "hetero")
```

```
## NOTE: 1,757 observations removed because of NA values (RHS: 1,757).
```

```
ols_model_cov_learndis <- feols(comp_score_5to6 ~ head_start + learndis | mom_id,
                      data = nlsy_kids, vcov = "hetero")

ols_model_cov_both <- feols(comp_score_5to6 ~ head_start + ppvt_3 + learndis | mom_id,
                      data = nlsy_kids, vcov = "hetero")
```

```
## NOTE: 1,757 observations removed because of NA values (RHS: 1,757).
```

```
summary(ols_model_cov_ppvt)
```

```
## OLS estimation, Dep. Var.: comp_score_5to6
## Observations: 551
## Fixed-effects: mom_id: 492
## Standard-errors: Heteroskedasticity-robust
##             Estimate Std. Error   t value Pr(>|t|)
## head_start -1.315816  15.296579 -0.086020  0.93175
## ppvt_3      0.212165   0.436988  0.485518  0.62917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 4.775     Adj. R2: 0.599702
##               Within R2: 0.021241
```

```
summary(ols_model_cov_learndis)
```

```
## OLS estimation, Dep. Var.: comp_score_5to6
## Observations: 2,308
## Fixed-effects: mom_id: 1,393
## Standard-errors: Heteroskedasticity-robust
##             Estimate Std. Error  t value   Pr(>|t|)
## head_start   7.45996    2.45221  3.04214 2.4162e-03 **
## learndis   -12.46645    3.13891 -3.97158 7.7014e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.3     Adj. R2: 0.461585
##              Within R2: 0.037756
```

```r
summary(ols_model_cov_both)
```

```
## OLS estimation, Dep. Var.: comp_score_5to6
## Observations: 551
## Fixed-effects: mom_id: 492
## Standard-errors: Heteroskedasticity-robust
##             Estimate Std. Error   t value Pr(>|t|)
## head_start -1.309506  15.435583 -0.084837  0.93269
## ppvt_3      0.213546   0.444067  0.480885  0.63247
## learndis    0.820930  24.169992  0.033965  0.97303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 4.77473     Adj. R2: 0.592601
##                 Within R2: 0.021353
```

Question 7

Code: Regressions.

Verbal: Interpret.

The coefficients of head start on 7-10 is 0.166361 and 11-14 is 0.230717, both of which are lower than 5-6's 0.334273, showing that the effect seems to decrease with age, although not by much. the t-value for 7-10 is not high enough for statistical significance, so we can just focus on ages 11-14. The seemingly 33% drop in coefficient makes sense as the head start program was not a long lasting treatment and the students may forget what they have learned over time.

```
# All question 7 code here

# creating a function to standardize test scores across ages
standardize_scores <- function(scores) {
  return((scores - mean(scores, na.rm = TRUE)) / sd(scores, na.rm = TRUE))
}

# Standardize test scores for each age group
nlsy_kids$std_comp_score_5to6 <- standardize_scores(nlsy_kids$comp_score_5to6)
nlsy_kids$std_comp_score_7to10 <- standardize_scores(nlsy_kids$comp_score_7to10)
nlsy_kids$std_comp_score_11to14 <- standardize_scores(nlsy_kids$comp_score_11to14)

# Estimate fixed-effects models for each age group
fe_model_5to6 <- feols(std_comp_score_5to6 ~ head_start | mom_id, data = nlsy_kids, vcov = "iid")
fe_model_7to10 <- feols(std_comp_score_7to10 ~ head_start | mom_id, data = nlsy_kids, vcov = "iid")
```

```
## NOTE: 317 observations removed because of NA values (LHS: 317).
```

```
fe_model_11to14 <- feols(std_comp_score_11to14 ~ head_start | mom_id, data = nlsy_kids, vcov = "iid")
```

```
## NOTE: 589 observations removed because of NA values (LHS: 589).
```

```
# Print the summary of the models
summary(fe_model_5to6)
```

```
## OLS estimation, Dep. Var.: std_comp_score_5to6
## Observations: 2,308
## Fixed-effects: mom_id: 1,393
## Standard-errors: IID
##            Estimate Std. Error t value   Pr(>|t|)
## head_start 0.334273   0.086982 3.84301 0.00012994 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.466716     Adj. R2: 0.449959
##                  Within R2: 0.015901
```

```
summary(fe_model_7to10)
```

```
## OLS estimation, Dep. Var.: std_comp_score_7to10
## Observations: 1,991
```

```
## Fixed-effects: mom_id: 1,210
## Standard-errors: IID
##            Estimate Std. Error t value Pr(>|t|)
## head_start 0.166361   0.090219 1.84398 0.065565 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.443857     Adj. R2: 0.497121
##                  Within R2: 0.00434
```

```
summary(fe_model_11to14)
```

```
## OLS estimation, Dep. Var.: std_comp_score_11to14
## Observations: 1,719
## Fixed-effects: mom_id: 1,042
## Standard-errors: IID
##            Estimate Std. Error t value Pr(>|t|)
## head_start 0.230717   0.091539 2.52043  0.01195 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.425524     Adj. R2: 0.539556
##                  Within R2: 0.00931
```

Question 8

Code: Regressions.

Verbal: Interpret.

When regressing on self-rated health, high school graduate and some college, some college is statistically insignificant. In regards self-rated health, the head start program has a negative effect. In regards to education, the effect on the probability to graduate high school is double that of attending college,

```
# All question 8 code here

nlsy_kids <- nlsy_kids %>% drop_na(fphealth, hsgrad, somecoll)

# Estimate fixed-effects logit models for each outcome
fe_model_fphealth <- feols(fphealth ~ head_start | mom_id, data = nlsy_kids)
fe_model_hsgrad <- feols(hsgrad ~ head_start | mom_id, data = nlsy_kids)
fe_model_somecoll <- feols(somecoll ~ head_start | mom_id, data = nlsy_kids)

# Print the summary of the models
summary(fe_model_fphealth)
```

```
## OLS estimation, Dep. Var.: fphealth
## Observations: 1,940
## Fixed-effects: mom_id: 1,149
## Standard-errors: Clustered (mom_id)
##             Estimate Std. Error  t value Pr(>|t|)
## head_start -0.083621   0.040997 -2.03968 0.041611 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.181173     Adj. R2: 0.056989
##                  Within R2: 0.006847
```

```
summary(fe_model_hsgrad)
```

```
## OLS estimation, Dep. Var.: hsgrad
## Observations: 1,940
## Fixed-effects: mom_id: 1,149
## Standard-errors: Clustered (mom_id)
##            Estimate Std. Error t value Pr(>|t|)
## head_start 0.140696    0.05468 2.57309 0.010204 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.272736     Adj. R2: 0.146143
##                  Within R2: 0.008539
```

```
summary(fe_model_somecoll)
```

```
## OLS estimation, Dep. Var.: somecoll
## Observations: 1,940
## Fixed-effects: mom_id: 1,149
## Standard-errors: Clustered (mom_id)
##            Estimate Std. Error t value Pr(>|t|)
```

```
## head_start 0.076984   0.051825 1.48548  0.13769
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.264813     Adj. R2: 0.179577
##                  Within R2: 0.002728
```

Question 9

Code: Analysis of heterogeneity.

Verbal: Interpret.

When accounting for race/ethnicity, the interaction coefficients for Hispanic is 0.031051, while for black is 0.077025 and male is 0.016949, none of which are statistically significant. Male alone has a negative coefficient of -0.089386 which is statistically significant. This shows that being male reaps the lowest benefit of the head start program compared to being black or hispanic.

```
# All question 9 code here


#using fixed effects OLS
fe_model_hsgrad <- feols(hsgrad ~ head_start + black + hispanic + male +
                                  head_start*(black + hispanic + male) |
                                  mom_id, data = nlsy_kids)
```

```
## The variables 'black' and 'hispanic' have been removed because of collinearity (see $collin.var).
```

```
# Summarize the model
summary(fe_model_hsgrad)
```

```
## OLS estimation, Dep. Var.: hsgrad
## Observations: 1,940
## Fixed-effects: mom_id: 1,149
## Standard-errors: Clustered (mom_id)
##                      Estimate Std. Error   t value Pr(>|t|)
## head_start           0.085950   0.115965  0.741171 0.458741
## male                -0.089386   0.033402 -2.676075 0.007555 **
## head_start:black     0.077025   0.134832  0.571264 0.567932
## head_start:hispanic  0.031051   0.162160  0.191485 0.848180
## head_start:male      0.016949   0.077804  0.217847 0.827587
## ... 2 variables were removed because of collinearity (black and hispanic)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.271223     Adj. R2: 0.151294
##                  Within R2: 0.01951
```

```
#using the Panel Linear Model(PLM) method

# Estimate the model with interaction terms
interaction_model <- plm(hsgrad ~ head_start + black + hispanic + male +
                                  head_start*(black + hispanic + male),
                         data = nlsy_kids,
                         index = c("mom_id"),
                         model = "within",
                         effect = "individual")

# Summarize the model
summary(interaction_model)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = hsgrad ~ head_start + black + hispanic + male +
##     head_start * (black + hispanic + male), data = nlsy_kids,
##     effect = "individual", model = "within", index = c("mom_id"))
##
## Unbalanced Panel: n = 1149, T = 1-5, N = 1940
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -0.876180 -0.044693  0.000000  0.044693  0.750000
##
## Coefficients:
##                      Estimate Std. Error t-value Pr(>|t|)
## head_start           0.085950   0.116132  0.7401 0.459456
## male                -0.089386   0.033406 -2.6758 0.007611 **
## head_start:black     0.077025   0.130667  0.5895 0.555712
## head_start:hispanic  0.031051   0.154006  0.2016 0.840264
## head_start:male      0.016949   0.074721  0.2268 0.820610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    145.55
## Residual Sum of Squares: 142.71
## R-Squared:      0.01951
## Adj. R-Squared: -1.4188
## F-statistic: 3.12795 on 5 and 786 DF, p-value: 0.0083719
```

Question 10

Verbal: Policy implications.

In regards to the effect of high school graduation, when controlling for gender, head start does not seem to correct the difference between male and female graduation rates. In the dataset provided, since black and hispanic share collinearity with mother fixed effect in the head start program, this data set alone is unable to determine if the headstart program is an effective treatment in the context of high school graduation. When accounting for costs, there may be more effective programs for later childhood stages, as we notice that the headstart treatment has a drop-off on test scores 4-6 years down the line, it might be better to allocate resources from head start to elementary or middle school instead to account for this "treatment decay". Overall, we would have to compare treatment effects between black and hispanic graduation rates without head start while keeping mother fixed effects. My concern is by going forward with the expansion of this program without further information, this policy could result in larger education gaps in minority communities, or worse, a stagnation of education that gets completely ignored because the treatment effects aren't bad enough to warrant investigation. I would also like to see an up-to-date dataset that include other minority groups such as Asians and Native Americans, as this program has an unknown effect on minority groups smaller than 5-7% of the population.