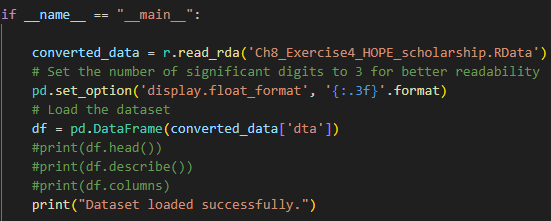
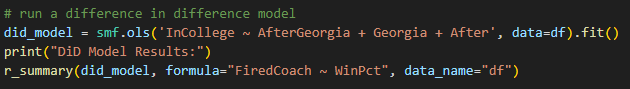
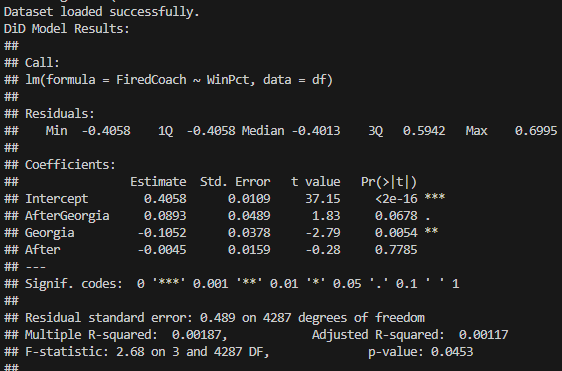
Lab 4 HOPE Scholarship in Python

## Preparation



#### (a) Run a basic difference-in-difference model. What is the effect of the program?

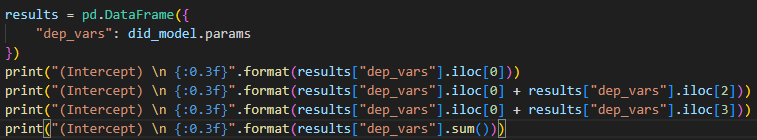




The “Georgia” variable indicates the difference in the college enrollment rate between Georgia and its neighboring states *before* the treatment (HOPE scholarship) went into effect. The coefficient implies that the college enrollment rate in Georgia was 10.5% lower than in surrounding states (states in the control group). The coefficient of the “After” variable which is 0.45% indicates that enrollment in the control group didn’t change much after HOPE into effect in Georgia, which makes logical sense. However, the “AfterGeorgia” variable coefficient, which is statistically significant at the 10% level, indicates that the HOPE scholarship program is associated with an 8.93% increase in enrollment.

#### (b) Calculate the percentage of people in the sample in college from the following four groups: (i) Before 1993/non-Georgia, (ii) Before 1993/Georgia, (iii) After 1992/non-Georgia, (iv) After 1992/Georgia. First, use the mean function (e.g., in Stata use mean Y if X1 == 0 & X2 == 0). Second, use the coefficients from the OLS output in part (a).

Using the OLS coefficients we can see that the



(i) Before 1993/non-Georgia fitted value was



1. Before 1993/Georgia fitted value was



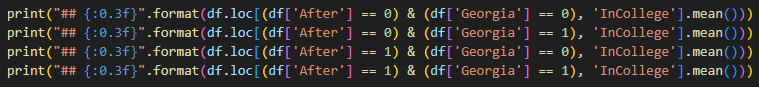
1. After 1992/non-Georgia fitted value was



1. After 1992/Georgia fitted value was



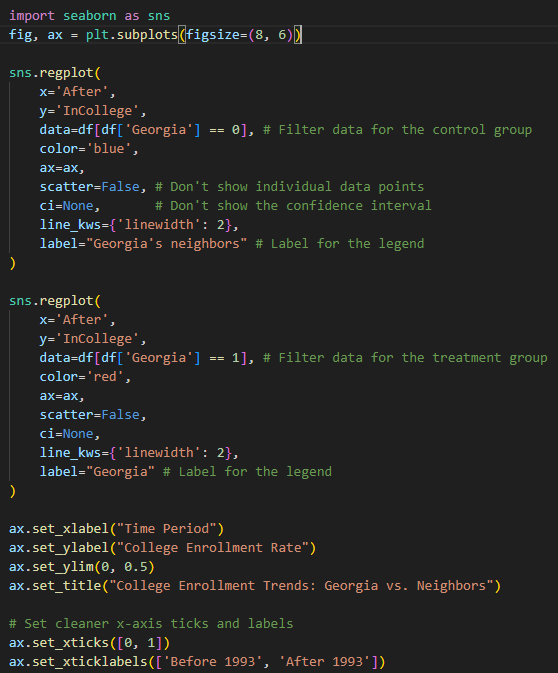
These are values are identical to what we get if we calculate them using the mean function. For example, the Before 1993/non-Georgia fitted value was

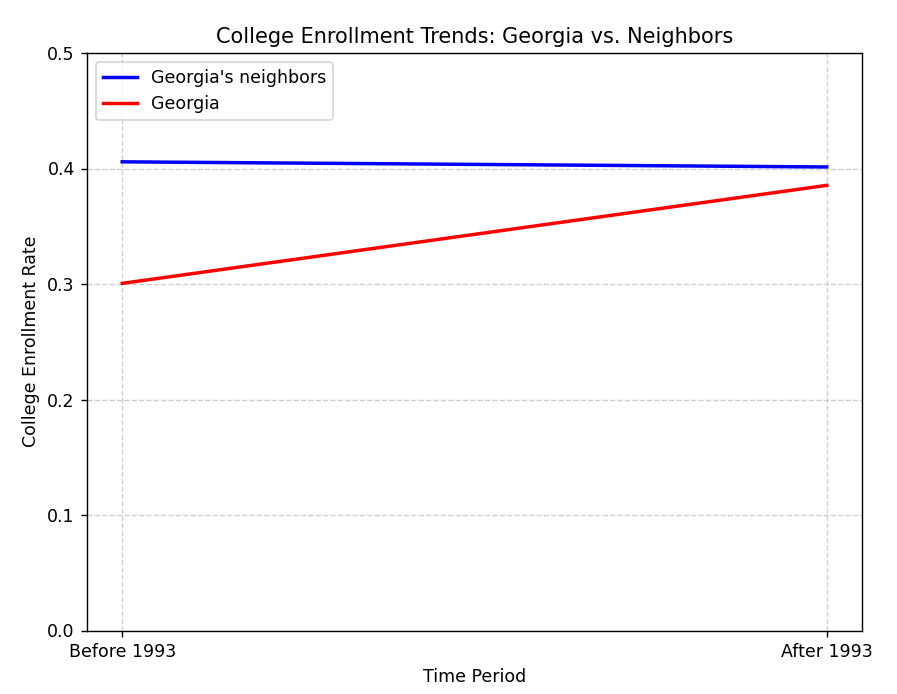




#### (c) Graph the fitted lines for the Georgia group and non-Georgia samples.

We can see that there was essentially no change in college attendance for young people in Georgia’s neighbors, but that attendance in Georgia rose considerably. This is a classic difference-in-difference figure.





#### (d) Use panel data formulation for a difference-in-difference model to control for all year and state effects.

When using the panel data formulation for a difference-in-difference model we have dummy variables for each state and for each year. This means that we no longer need to (or, can) control for the *Georgia* and *After* variables; one of the state dummies will be *Georgia* and the individual year dummies taken together account for the *After* dummy variable. Here we simply include time dummies as covariates in a one-way fixed effects model. The estimated effect remains similar as before: about 9.1 percentage points and statistically significant at =0.10.

result2 <-plm(InCollege ~ AfterGeorgia + factor(Year),  
 model = "within", index = c('StateCode'),  
 data = dta)  
summary(result2)

## Oneway (individual) effect Within Model  
##   
## Call:  
## plm(formula = InCollege ~ AfterGeorgia + factor(Year), data = dta,   
## model = "within", index = c("StateCode"))  
##   
## Unbalanced Panel: n = 6, T = 403-1280, N = 4291  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -0.493 -0.415 -0.334 0.569 0.736   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## AfterGeorgia 0.09142 0.04876 1.87 0.061 .  
## factor(Year)90 0.04661 0.02834 1.64 0.100   
## factor(Year)91 0.03228 0.02857 1.13 0.259   
## factor(Year)92 0.02354 0.02985 0.79 0.430   
## factor(Year)93 0.03016 0.03015 1.00 0.317   
## factor(Year)94 0.01451 0.03057 0.47 0.635   
## factor(Year)95 -0.00326 0.03170 -0.10 0.918   
## factor(Year)96 -0.02131 0.03226 -0.66 0.509   
## factor(Year)97 0.07534 0.03128 2.41 0.016 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 1020  
## Residual Sum of Squares: 1020  
## R-Squared: 0.00361  
## Adj. R-Squared: 0.000349  
## F-statistic: 1.72204 on 9 and 4276 DF, p-value: 0.0785

#### (e) Add covariates for 18-year-olds and African-Americans to the panel data formulation. What is the effect of the HOPE program?

The estimated effect remains similar as before: about 8.5 percentage points and statistically significant at =0.10.

result3 <-plm(InCollege ~ AfterGeorgia + Age18 + Black + factor(Year),  
 model = "within", index = c('StateCode'),  
 data = dta)  
summary(result3)

## Oneway (individual) effect Within Model  
##   
## Call:  
## plm(formula = InCollege ~ AfterGeorgia + Age18 + Black + factor(Year),   
## data = dta, model = "within", index = c("StateCode"))  
##   
## Unbalanced Panel: n = 6, T = 403-1280, N = 4291  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -0.553 -0.412 -0.298 0.545 0.872   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## AfterGeorgia 0.08531 0.04835 1.76 0.07770 .   
## Age18 -0.05006 0.01477 -3.39 0.00071 \*\*\*  
## Black -0.13908 0.01704 -8.16 4.3e-16 \*\*\*  
## factor(Year)90 0.04864 0.02809 1.73 0.08343 .   
## factor(Year)91 0.03707 0.02832 1.31 0.19064   
## factor(Year)92 0.02992 0.02959 1.01 0.31207   
## factor(Year)93 0.03584 0.02990 1.20 0.23066   
## factor(Year)94 0.01350 0.03030 0.45 0.65603   
## factor(Year)95 0.00365 0.03143 0.12 0.90762   
## factor(Year)96 -0.01747 0.03198 -0.55 0.58488   
## factor(Year)97 0.07079 0.03101 2.28 0.02248 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 1020  
## Residual Sum of Squares: 998  
## R-Squared: 0.0217  
## Adj. R-Squared: 0.018  
## F-statistic: 8.61242 on 11 and 4274 DF, p-value: 2.99e-15

#### (f) The way the program was designed, Georgia high school graduates with a B or higher average and annual family income over $50,000 could qualify for HOPE by filling out a simple one-page form. Those with lower income were required to apply for federal aid with a complex four-page form and had any federal aid deducted from their HOPE scholarship. Run separate basic difference-in-difference models for these two groups and comment on the substantive implication of the results.

The effect is higher than we’ve yet seen and statistically significant for Whites. The effect is negative and statistically insignificant for African-Americans. There are many possible explanations, but one is that the program was more helpful to higher income people than lower income people in getting them into college.

result4 <- lm(InCollege ~ AfterGeorgia + Georgia, data = dta[dta$Black==0,])  
summary(result4)

##   
## Call:  
## lm(formula = InCollege ~ AfterGeorgia + Georgia, data = dta[dta$Black ==   
## 0, ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.458 -0.435 -0.435 0.565 0.670   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.43497 0.00926 46.97 <2e-16 \*\*\*  
## AfterGeorgia 0.12728 0.05791 2.20 0.028 \*   
## Georgia -0.10453 0.04710 -2.22 0.027 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.495 on 3173 degrees of freedom  
## Multiple R-squared: 0.00172, Adjusted R-squared: 0.00109   
## F-statistic: 2.74 on 2 and 3173 DF, p-value: 0.0649

result5 <- lm(InCollege ~ AfterGeorgia + Georgia, data = dta[dta$Black==1,])  
summary(result5)

##   
## Call:  
## lm(formula = InCollege ~ AfterGeorgia + Georgia, data = dta[dta$Black ==   
## 1, ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.310 -0.310 -0.310 0.690 0.782   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.3104 0.0148 21.01 <2e-16 \*\*\*  
## AfterGeorgia -0.0316 0.0741 -0.43 0.67   
## Georgia -0.0604 0.0575 -1.05 0.29   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.458 on 1112 degrees of freedom  
## Multiple R-squared: 0.00365, Adjusted R-squared: 0.00186   
## F-statistic: 2.04 on 2 and 1112 DF, p-value: 0.131