The effect of direct admissions programs on college enrollment

Ryan Blake

CLEAN

LOAD FILES AND PACKAGES Note: This analysis relies on ACS data provided by IPUMS. The ACS extract exceeds the 100 MB limit for GitHub and is not attached to the git repo. The file is stored locally and is available on request.

This chunk loads dependencies, the ACS microdata, and the ACS metadata. It then loads a sheet containing binary variables that outline the policy timeline for each state–i.e., 1 for states/years with direct admissions and 0 otherwise. The last line loads state-level NHGIS data from 2010; this data is for DiNardo Fortin Lemieux (DFL) reweights.

```r
#rm(list = ls(all.names = TRUE))
library(tidyverse)
library(ipumsr)
library(fixest)
library(miceadds)
library(tinytex)

setwd("/Users/ryan/DataspellProjects/direct_admissions/")
ddi <- read_ipums_ddi("usa_00019.xml", lower_vars = TRUE)
ipums <- read_ipums_micro(ddi)
```

```
## Use of data from IPUMS USA is subject to conditions including that users should
## cite the data appropriately. Use command 'ipums_conditions()' for more details.
```

```r
policysheet <- read.csv("/Users/ryan/DataspellProjects/direct_admissions/policysheet.csv")

nhgis <- read.csv("/Users/ryan/DataspellProjects/direct_admissions/nhgis.csv")
```

CALCULATE DFL REWEIGHTS This chunk creates the variable necessary for DFL reweighting. It first cleans the NHGIS data to create new variables that describe a states urban/rural mix, high school graduation rate, college attendance rate, youth unemployment rate, and portion of adults serving in the military. Note that these variables are taken from states at a baseline period prior to adoption (2010, in this case). This sheet is then merged with the policy implementation timeline.

Next, a probit regression is run to predict the likelihood that a state has adopted the policy at some point during the observation period. These fitted values are then assigned to states that have adopted the policy; a value of 0 is assigned to state that have not adopted the policy.

Finally, this chunk runs a series of checks to assess the validity of each chosen variable.

```r
nhgis <- nhgis %>%
  mutate(
    #percent living in rural area
    rural_pct = rural / total_region,
    #percent w/ high school degree
    hs_complete_pct =
      (malehs+malecollege1+malecollege2+maleassoc+malebach+malemaster+maleprof+maledoc+
        femalehs+femalecollege1+femalecollege2+femaleassoc+femalebach+femalemaster+femaleprof+femaledoc)
        total_attainment,
    #percent college go on
    college_pct =
      (malecollege1+malecollege2+maleassoc+malebach+malemaster+maleprof+maledoc+
        femalecollege1+femalecollege2+femaleassoc+femalebach+femalemaster+femaleprof+femaledoc)/
        (malehs +femalehs),
    #percent young adults <25 unemployed in labor force
    youth_unemployment =
      (male16_labor_civ_unemployed + male20_labor_civ_unemployed + male22_labor_civ_unemployed +
        female16_labor_civ_unemployed + female20_labor_civ_unemployed + female22_labor_civ_unemployed)/
        (male16_labor_civ + male20_labor_civ + male22_labor_civ +
          female16_labor_civ + female20_labor_civ + female22_labor_civ),
    #percent young adults <25 in labor force who serve in the military
    vet_pct =
      (male16_labor_vet + male20_labor_vet + male22_labor_vet +
        female16_labor_vet + female20_labor_vet + female22_labor_vet)/
        (male16_labor_total + male20_labor_total + male22_labor_total +
          female16_labor_total + female20_labor_total + female22_labor_total),
  )%>%
  select(state, statefip, rural_pct, hs_complete_pct, youth_unemployment, vet_pct, median_income)

#MERGE POLICY AND NHGIS DATA
states <- merge(policysheet, nhgis, by=c("state","statefip"))


#CALCULATE DFL
reg_nhgis <- glm(da2020 ~ rural_pct + hs_complete_pct + youth_unemployment + vet_pct +  median_income,
                 family=binomial(link='probit'), data = states)

#PRINT DFL REGRESSION RESULTS
summary(reg_nhgis)
```

```
##
## Call:
## glm(formula = da2020 ~ rural_pct + hs_complete_pct + youth_unemployment +
##     vet_pct + median_income, family = binomial(link = "probit"),
##     data = states)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.27626  -0.05796  -0.00066   0.00000   1.83852
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         1.211e+01  3.285e+01   0.369    0.712
```

```
## rural_pct          -6.011e+00  1.106e+01  -0.544    0.587
## hs_complete_pct     2.554e+01  3.309e+01   0.772    0.440
## youth_unemployment -6.318e+01  5.313e+01  -1.189    0.234
## vet_pct            -1.381e+02  1.077e+02  -1.282    0.200
## median_income      -5.083e-04  4.533e-04  -1.121    0.262
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 16.7944  on 49  degrees of freedom
## Residual deviance:  8.6432  on 44  degrees of freedom
## AIC: 20.643
##
## Number of Fisher Scoring iterations: 12
```

```
with(summary(reg_nhgis), 1 - deviance/null.deviance) #print r squared
```

```
## [1] 0.4853502
```

```
#FIT DFL REGRESSION RESULTS TO STATES
predict_da <- fitted(reg_nhgis)
states  <- states %>%
  mutate(pda = predict_da,
         dfl = ifelse(da2020 == 1, 1, (predict_da/(1-predict_da))))

#DFL VARIABLE CHECK
lm(rural_pct ~ da2020, data = states)
```

```
##
## Call:
## lm(formula = rural_pct ~ da2020, data = states)
##
## Coefficients:
## (Intercept)       da2020
##      0.2525       0.1113
```

```
lm(rural_pct ~ da2020, data = states, weights = dfl)
```

```
##
## Call:
## lm(formula = rural_pct ~ da2020, data = states, weights = dfl)
##
## Coefficients:
## (Intercept)       da2020
##     0.33432      0.02953
```

```
lm(hs_complete_pct ~ da2020, data = states)
```

```
##
## Call:
## lm(formula = hs_complete_pct ~ da2020, data = states)
##
```

```
## Coefficients:
## (Intercept)       da2020
##     0.86361      0.02358
```

```r
lm(hs_complete_pct ~ da2020, data = states, weights = dfl)
```

```
##
## Call:
## lm(formula = hs_complete_pct ~ da2020, data = states, weights = dfl)
##
## Coefficients:
## (Intercept)       da2020
##    0.891528    -0.004329
```

```r
lm(youth_unemployment ~ da2020, data = states)
```

```
##
## Call:
## lm(formula = youth_unemployment ~ da2020, data = states)
##
## Coefficients:
## (Intercept)       da2020
##     0.15928     -0.03719
```

```r
lm(youth_unemployment ~ da2020, data = states, weights = dfl)
```

```
##
## Call:
## lm(formula = youth_unemployment ~ da2020, data = states, weights = dfl)
##
## Coefficients:
## (Intercept)       da2020
##    0.121122     0.000962
```

```r
lm(vet_pct ~ da2020, data = states)
```

```
##
## Call:
## lm(formula = vet_pct ~ da2020, data = states)
##
## Coefficients:
## (Intercept)       da2020
##    0.021707    -0.008214
```

```r
lm(vet_pct ~ da2020, data = states, weights = dfl)
```

```
##
## Call:
## lm(formula = vet_pct ~ da2020, data = states, weights = dfl)
##
## Coefficients:
## (Intercept)       da2020
##    0.012068     0.001425
```

```r
lm(median_income ~ da2020, data = states)
```

```
##
## Call:
## lm(formula = median_income ~ da2020, data = states)
##
## Coefficients:
## (Intercept)       da2020
##       52061        -5665
```

```r
lm(median_income~ da2020, data = states, weights = dfl)
```

```
##
## Call:
## lm(formula = median_income ~ da2020, data = states, weights = dfl)
##
## Coefficients:
## (Intercept)       da2020
##     47008.4       -612.4
```

MERGE STATE DATA WITH IPUMS DATA This chunk converts the wide state-level data to long data, and then merges with the ACS microdata. The data is then filtered to include only 19-year-old high school graduates who have lived in the same state for more than 1 year (a distinction which is necessary for defining who received treatment).

The outcome variable (enroll) is then created to include those who are currently attending a post-secondary institution. A series of independent variables are also created to be used as controls.

```r
#WIDE DATA TO LONG DATA
states_long <- states %>%
  pivot_longer(
    starts_with("da"),
    names_to = c(".value", "year"),
    names_pattern = "^(.*?)(\\d+)$",
  ) %>%
  mutate(year = as.numeric(year))

#MERGE STATE LEVEL POLICY DATA ONE TO MANY WITH ACS DATA
df <- inner_join(states_long, ipums, by = c("statefip", "year"), multiple = "all")

#FILTER TO INCLUDE ONLY HIGH SCHOOL GRADUATES
#MUTATE TO CREATE BINARY OUTCOME VARIABLE 'ENROLL' & ADD CONTROLS
df <- df %>%
        filter(educd >= 63, age == 19, school == 1 | school == 2, migrate1 == 1 | migrate1 == 2) %>%
        mutate(enroll = ifelse(educd >= 65, 1, 0), #OUTCOME VARIABLE
               female = ifelse(sex == 2, 1,0),
               hispanic = ifelse(hispand > 0,1,0),
               black = ifelse(race == 2 & hispand == 0,1,0),
               othrace = ifelse(race > 2 & hispand == 0,1,0),
               citizen = ifelse(citizen == 3, 0, 1),
               english = ifelse(speakeng == 1, 0, 1),
               insurance = ifelse(hcovany == 1, 0, 1),
               dflrewt = perwt*dfl)
```

RESULTS

SUMMARY STATISTICS This chunk prints a series of summary statistics.

```
df %>%
        summarize(
                n = n(),
                college_go_on = mean(enroll),
                college_go_on_sd = sd(enroll),
                female = mean(female),
                black = mean(black),
                hispanic = mean(hispanic),
                othrace = mean(othrace),
                citizen = mean(citizen),
                english = mean(english),
                insurance = mean(insurance))
```

```
## # A tibble: 1 x 10
##        n college_~1 colle~2 female black hispa~3 othrace citizen english insur~4
##    <int>      <dbl>   <dbl>  <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 316465      0.626   0.484  0.505 0.116   0.180   0.104   0.960   0.999   0.877
## # ... with abbreviated variable names 1: college_go_on, 2: college_go_on_sd,
## #   3: hispanic, 4: insurance
```

```
df %>%
        group_by(year) %>%
        summarize(
                n = n(),
                college_go_on = mean(enroll),
                college_go_on_sd = sd(enroll),
                female = mean(female),
                black = mean(black),
                hispanic = mean(hispanic),
                othrace = mean(othrace),
                citizen = mean(citizen),
                english = mean(english),
                insurance = mean(insurance))
```

```
## # A tibble: 9 x 11
##    year     n college_go_on colle~1 female black hispa~2 othrace citizen english
##   <dbl> <int>         <dbl>   <dbl>  <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1  2012 36016         0.635   0.481  0.506 0.122   0.166  0.0917   0.958   0.999
## 2  2013 35199         0.636   0.481  0.508 0.116   0.169  0.0978   0.959   0.999
## 3  2014 34878         0.627   0.484  0.501 0.120   0.173  0.0987   0.960   0.999
## 4  2015 35006         0.629   0.483  0.506 0.118   0.176  0.101    0.959   0.999
## 5  2016 34791         0.630   0.483  0.509 0.114   0.182  0.106    0.959   0.999
## 6  2017 35126         0.623   0.485  0.501 0.117   0.186  0.106    0.957   0.999
## 7  2018 35934         0.613   0.487  0.505 0.110   0.191  0.104    0.961   0.999
## 8  2019 35804         0.609   0.488  0.502 0.112   0.193  0.110    0.964   0.999
## 9  2020 33711         0.637   0.481  0.510 0.111   0.185  0.123    0.964   0.999
## # ... with 1 more variable: insurance <dbl>, and abbreviated variable names
## #   1: college_go_on_sd, 2: hispanic
```

```r
df %>%
        group_by(da) %>%
        summarize(
                n = n(),
                college_go_on = mean(enroll),
                college_go_on_sd = sd(enroll),
                female = mean(female),
                black = mean(black),
                hispanic = mean(hispanic),
                othrace = mean(othrace),
                citizen = mean(citizen),
                english = mean(english),
                insurance = mean(insurance))
```

```
## # A tibble: 2 x 11
##      da      n college_g~1 colle~2 female  black hispa~3 othrace citizen english
##   <int>  <int>       <dbl>   <dbl>  <dbl>  <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     0 315459       0.627   0.484  0.505 0.116    0.180  0.104    0.960   0.999
## 2     1   1006       0.589   0.492  0.479 0.0159   0.106  0.0845   0.978   0.999
## # ... with 1 more variable: insurance <dbl>, and abbreviated variable names
## #   1: college_go_on, 2: college_go_on_sd, 3: hispanic
```

OLS MODELS This chunk runs two separate OLS models and prints the results (including r^2). The first model uses person level weights supplied in the ACS data; the second model uses the DFL reweights calculated from NHGIS data. Both models control for gender, race, ethnicity, citizenship, health insurance, English proficiency, and state/year fixed effects. Standard errors are clustered at the state level.

```r
#OLS - PERSON LEVEL WEIGHTS
olsmod1 <- miceadds::glm.cluster(data=df,
                                 formula=enroll ~ da + female + hispanic + black + othrace + citizen + i
                                        factor(statefip) + factor(year),
                                 cluster="state",
                                 weights= df$perwt,
                                 family="gaussian")
summary(olsmod1$glm_res)
```

```
##
## Call:
## stats::glm(formula = formula, family = family, data = data, weights = wgt__)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -24.477   -4.445    2.235    3.510   23.931
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.210996   0.023520   8.971  < 2e-16 ***
## da                  0.019440   0.020985   0.926 0.354234
## female              0.094888   0.001718  55.228  < 2e-16 ***
## hispanic           -0.045656   0.002451 -18.626  < 2e-16 ***
## black              -0.042917   0.002666 -16.097  < 2e-16 ***
## othrace             0.025419   0.003079   8.257  < 2e-16 ***
```

```
## citizen             0.013108   0.004250   3.084 0.002040 **
## insurance           0.157913   0.002558  61.733  < 2e-16 ***
## english             0.222745   0.022350   9.966  < 2e-16 ***
## factor(statefip)2  -0.145437   0.019580  -7.428 1.11e-13 ***
## factor(statefip)4  -0.020823   0.009275  -2.245 0.024775 *
## factor(statefip)5  -0.009224   0.011165  -0.826 0.408706
## factor(statefip)6   0.034729   0.007480   4.643 3.43e-06 ***
## factor(statefip)8  -0.008250   0.009799  -0.842 0.399793
## factor(statefip)9   0.024893   0.010376   2.399 0.016434 *
## factor(statefip)10 -0.031125   0.017905  -1.738 0.082146 .
## factor(statefip)11 -0.017459   0.020617  -0.847 0.397086
## factor(statefip)12  0.029122   0.007912   3.681 0.000233 ***
## factor(statefip)13 -0.004224   0.008382  -0.504 0.614274
## factor(statefip)15 -0.085925   0.015942  -5.390 7.05e-08 ***
## factor(statefip)16 -0.059817   0.019044  -3.141 0.001684 **
## factor(statefip)17  0.036886   0.008220   4.487 7.21e-06 ***
## factor(statefip)18 -0.021424   0.009136  -2.345 0.019036 *
## factor(statefip)20  0.026675   0.011003   2.424 0.015331 *
## factor(statefip)21 -0.045248   0.010069  -4.494 7.00e-06 ***
## factor(statefip)22 -0.019228   0.010076  -1.908 0.056364 .
## factor(statefip)23 -0.029157   0.015554  -1.875 0.060847 .
## factor(statefip)24  0.037114   0.009430   3.936 8.30e-05 ***
## factor(statefip)25  0.032042   0.008910   3.596 0.000323 ***
## factor(statefip)26  0.015170   0.008462   1.793 0.073018 .
## factor(statefip)27  0.013296   0.009842   1.351 0.176692
## factor(statefip)28  0.048235   0.010883   4.432 9.33e-06 ***
## factor(statefip)29  0.014232   0.009354   1.522 0.128123
## factor(statefip)30 -0.084254   0.017060  -4.939 7.87e-07 ***
## factor(statefip)31  0.036259   0.012850   2.822 0.004776 **
## factor(statefip)32 -0.084259   0.012238  -6.885 5.78e-12 ***
## factor(statefip)33 -0.001636   0.014700  -0.111 0.911387
## factor(statefip)34  0.031536   0.008779   3.592 0.000328 ***
## factor(statefip)35  0.030004   0.012749   2.354 0.018597 *
## factor(statefip)36  0.074200   0.007811   9.500  < 2e-16 ***
## factor(statefip)37  0.019991   0.008438   2.369 0.017829 *
## factor(statefip)38 -0.005520   0.020396  -0.271 0.786682
## factor(statefip)39 -0.020770   0.008299  -2.503 0.012327 *
## factor(statefip)40 -0.020806   0.010573  -1.968 0.049092 *
## factor(statefip)41 -0.024844   0.010755  -2.310 0.020891 *
## factor(statefip)42 -0.012142   0.008151  -1.490 0.136308
## factor(statefip)44  0.037561   0.014650   2.564 0.010352 *
## factor(statefip)45  0.023201   0.009774   2.374 0.017612 *
## factor(statefip)46  0.021022   0.019996   1.051 0.293119
## factor(statefip)47 -0.030155   0.009136  -3.301 0.000965 ***
## factor(statefip)48  0.014412   0.007595   1.897 0.057767 .
## factor(statefip)49 -0.059195   0.011193  -5.289 1.23e-07 ***
## factor(statefip)50  0.009613   0.018927   0.508 0.611507
## factor(statefip)51  0.012540   0.008661   1.448 0.147646
## factor(statefip)53 -0.015908   0.009323  -1.706 0.087971 .
## factor(statefip)54 -0.073570   0.013608  -5.406 6.44e-08 ***
## factor(statefip)55 -0.022024   0.009484  -2.322 0.020220 *
## factor(statefip)56  0.024595   0.022119   1.112 0.266148
## factor(year)2013   -0.010847   0.003652  -2.970 0.002978 **
## factor(year)2014   -0.022147   0.003651  -6.066 1.31e-09 ***
```

```
## factor(year)2015   -0.029053   0.003662  -7.935 2.12e-15 ***
## factor(year)2016   -0.028388   0.003660  -7.756 8.81e-15 ***
## factor(year)2017   -0.036506   0.003657  -9.982  < 2e-16 ***
## factor(year)2018   -0.048278   0.003641 -13.261  < 2e-16 ***
## factor(year)2019   -0.049351   0.003627 -13.606  < 2e-16 ***
## factor(year)2020   -0.057003   0.003642 -15.651  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 22.72608)
##
##     Null deviance: 7430205  on 316464  degrees of freedom
## Residual deviance: 7190509  on 316399  degrees of freedom
## AIC: 535080
##
## Number of Fisher Scoring iterations: 2
```

```
with(summary(olsmod1$glm_res), 1 - deviance/null.deviance) #print r squared
```

```
## [1] 0.03225976
```

```
# OLS - DFL REWEIGHTS
olsmod2 <- miceadds::glm.cluster(data=df,
                                 formula=enroll ~ da + female + hispanic + black + othrace + citizen + i
                                         factor(statefip) + factor(year),
                                 cluster="state",
                                 weights= df$dflrewt,
                                 family="gaussian")
summary(olsmod2$glm_res)
```

```
##
## Call:
## stats::glm(formula = formula, family = family, data = data, weights = wgt__)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -20.1307    0.0000    0.0000    0.0004   19.0464
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.793e-01  1.224e+00   0.228 0.819465
## da                    2.311e-02  3.539e-03   6.531 6.56e-11 ***
## female                1.186e-01  1.705e-03  69.573  < 2e-16 ***
## hispanic             -6.370e-02  2.807e-03 -22.695  < 2e-16 ***
## black                -6.281e-02  4.492e-03 -13.983  < 2e-16 ***
## othrace              -8.534e-02  3.233e-03 -26.399  < 2e-16 ***
## citizen              -5.890e-02  5.671e-03 -10.387  < 2e-16 ***
## insurance             2.121e-01  2.422e-03  87.583  < 2e-16 ***
## english               1.582e-01  4.390e-02   3.604 0.000313 ***
## factor(statefip)2    -1.112e-01  1.745e+05   0.000 0.999999
## factor(statefip)4    -1.559e-02  1.239e+00  -0.013 0.989961
## factor(statefip)5    -1.092e-02  1.223e+00  -0.009 0.992877
## factor(statefip)6     4.408e-02  2.291e+04   0.000 0.999998
```

```
## factor(statefip)8  -1.270e-02  6.529e+04   0.000 1.000000
## factor(statefip)9   1.758e-02  7.321e+04   0.000 1.000000
## factor(statefip)10 -3.468e-02  1.576e+05   0.000 1.000000
## factor(statefip)11 -2.054e-02  1.854e+05   0.000 1.000000
## factor(statefip)12  3.055e-02  1.234e+00   0.025 0.980244
## factor(statefip)13  5.157e-04  4.445e+04   0.000 1.000000
## factor(statefip)15 -2.813e-02  1.357e+05   0.000 1.000000
## factor(statefip)16 -6.297e-02  1.223e+00  -0.051 0.958941
## factor(statefip)17  3.483e-02  8.371e+03   0.000 0.999997
## factor(statefip)18 -2.624e-02  1.223e+00  -0.021 0.982885
## factor(statefip)20  2.510e-02  1.223e+00   0.021 0.983630
## factor(statefip)21 -5.153e-02  1.523e+00  -0.034 0.973008
## factor(statefip)22 -1.875e-02  1.231e+00  -0.015 0.987852
## factor(statefip)23 -3.554e-02  1.224e+00  -0.029 0.976832
## factor(statefip)24  3.629e-02  6.059e+04   0.000 1.000000
## factor(statefip)25  2.497e-02  5.255e+04   0.000 1.000000
## factor(statefip)26  8.958e-03  5.733e+00   0.002 0.998753
## factor(statefip)27  8.034e-03  1.232e+00   0.007 0.994799
## factor(statefip)28  5.315e-02  4.670e+04   0.000 0.999999
## factor(statefip)29  1.172e-02  1.225e+00   0.010 0.992366
## factor(statefip)30 -8.024e-02  1.223e+00  -0.066 0.947694
## factor(statefip)31  3.228e-02  1.223e+00   0.026 0.978942
## factor(statefip)32 -7.538e-02  7.270e+03   0.000 0.999992
## factor(statefip)33 -9.543e-03  6.467e+04   0.000 1.000000
## factor(statefip)34  3.295e-02  5.062e+04   0.000 0.999999
## factor(statefip)35  3.966e-02  1.223e+00   0.032 0.974136
## factor(statefip)36  7.046e-02  1.369e+03   0.000 0.999959
## factor(statefip)37  1.848e-02  4.539e+04   0.000 1.000000
## factor(statefip)38 -1.073e-02  1.223e+00  -0.009 0.993001
## factor(statefip)39 -2.781e-02  1.223e+00  -0.023 0.981860
## factor(statefip)40 -4.913e-04  1.223e+00   0.000 0.999680
## factor(statefip)41 -2.416e-02  1.223e+00  -0.020 0.984244
## factor(statefip)42 -1.986e-02  1.223e+00  -0.016 0.987047
## factor(statefip)44  2.927e-02  1.947e+02   0.000 0.999880
## factor(statefip)45  2.110e-02  6.550e+04   0.000 1.000000
## factor(statefip)46  1.847e-02  1.223e+00   0.015 0.987954
## factor(statefip)47 -3.541e-02  1.224e+00  -0.029 0.976924
## factor(statefip)48  2.259e-02  2.067e+01   0.001 0.999128
## factor(statefip)49 -6.198e-02  1.223e+00  -0.051 0.959587
## factor(statefip)50 -7.255e-03  1.230e+00  -0.006 0.995294
## factor(statefip)51  1.346e-02  4.900e+04   0.000 1.000000
## factor(statefip)53 -1.033e-02  5.868e+04   0.000 1.000000
## factor(statefip)54 -8.207e-02  1.223e+00  -0.067 0.946498
## factor(statefip)55 -3.003e-02  1.223e+00  -0.025 0.980411
## factor(statefip)56  1.413e-02  1.857e+00   0.008 0.993929
## factor(year)2013   -8.888e-03  3.619e-03  -2.456 0.014048 *
## factor(year)2014    4.142e-02  3.728e-03  11.110  < 2e-16 ***
## factor(year)2015    2.466e-02  3.737e-03   6.599 4.16e-11 ***
## factor(year)2016   -1.600e-02  3.708e-03  -4.316 1.59e-05 ***
## factor(year)2017   -2.067e-02  3.811e-03  -5.424 5.84e-08 ***
## factor(year)2018    1.385e-02  3.868e-03   3.581 0.000343 ***
## factor(year)2019   -5.757e-02  3.849e-03 -14.958  < 2e-16 ***
## factor(year)2020   -4.612e-02  3.856e-03 -11.960  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.4611807)
##
##     Null deviance: 154963  on 316464  degrees of freedom
## Residual deviance: 145917  on 316399  degrees of freedom
## AIC: 6429020
##
## Number of Fisher Scoring iterations: 2
```

```r
with(summary(olsmod2$glm_res), 1 - deviance/null.deviance) #print r squared
```

```
## [1] 0.05837302
```

LOGIT MODELS This chunk runs two separate logit models and prints the results (including r^2). The first model uses person level weights supplied in the ACS data; the second model uses the DFL reweights calculated from NHGIS data. Both models control for gender, race, ethnicity, citizenship, health insurance, English proficiency, and state/year fixed effects. Standard errors are clustered at the state level.

```r
#LOGIT - DFL REWEIGHT
logmod1 <- miceadds::glm.cluster(data=df,
                                 formula= enroll ~ da + female + hispanic + black + othrace + citizen +
                                         factor(statefip) + factor(year),
                                 cluster="state",
                                 weights= df$perwt,
                                 family=binomial(link= "logit"))
summary(logmod1$glm_res)
```

```
##
## Call:
## stats::glm(formula = formula, family = family, data = data, weights = wgt__)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -54.888  -10.212    6.004    8.981   53.889
##
## Coefficients:
##                      Estimate Std. Error  z value Pr(>|z|)
## (Intercept)        -1.2745301  0.0111254 -114.561  < 2e-16 ***
## da                  0.0846329  0.0090408    9.361  < 2e-16 ***
## female              0.4044736  0.0007469  541.563  < 2e-16 ***
## hispanic           -0.1943597  0.0010593 -183.487  < 2e-16 ***
## black              -0.1825911  0.0011495 -158.842  < 2e-16 ***
## othrace             0.1114097  0.0013582   82.025  < 2e-16 ***
## citizen             0.0556284  0.0018379   30.268  < 2e-16 ***
## insurance           0.6510482  0.0010920  596.199  < 2e-16 ***
## english             1.0010705  0.0106518   93.981  < 2e-16 ***
## factor(statefip)2  -0.6064782  0.0084105  -72.110  < 2e-16 ***
## factor(statefip)4  -0.0871571  0.0039908  -21.839  < 2e-16 ***
## factor(statefip)5  -0.0387600  0.0048066   -8.064 7.39e-16 ***
## factor(statefip)6   0.1488062  0.0032306   46.061  < 2e-16 ***
## factor(statefip)8  -0.0346721  0.0042222   -8.212  < 2e-16 ***
```

```
## factor(statefip)9   0.1066712  0.0045139   23.631  < 2e-16 ***
## factor(statefip)10 -0.1308292  0.0076744  -17.047  < 2e-16 ***
## factor(statefip)11 -0.0742393  0.0088669   -8.373  < 2e-16 ***
## factor(statefip)12  0.1244385  0.0034181   36.406  < 2e-16 ***
## factor(statefip)13 -0.0173595  0.0036118   -4.806 1.54e-06 ***
## factor(statefip)15 -0.3603539  0.0067950  -53.032  < 2e-16 ***
## factor(statefip)16 -0.2525435  0.0081782  -30.880  < 2e-16 ***
## factor(statefip)17  0.1586675  0.0035619   44.546  < 2e-16 ***
## factor(statefip)18 -0.0909095  0.0039331  -23.114  < 2e-16 ***
## factor(statefip)20  0.1140966  0.0047804   23.868  < 2e-16 ***
## factor(statefip)21 -0.1898514  0.0043182  -43.965  < 2e-16 ***
## factor(statefip)22 -0.0803894  0.0043315  -18.559  < 2e-16 ***
## factor(statefip)23 -0.1244874  0.0066780  -18.641  < 2e-16 ***
## factor(statefip)24  0.1596481  0.0041009   38.930  < 2e-16 ***
## factor(statefip)25  0.1385636  0.0038763   35.746  < 2e-16 ***
## factor(statefip)26  0.0642423  0.0036612   17.547  < 2e-16 ***
## factor(statefip)27  0.0564017  0.0042678   13.216  < 2e-16 ***
## factor(statefip)28  0.2064630  0.0047327   43.624  < 2e-16 ***
## factor(statefip)29  0.0603496  0.0040493   14.904  < 2e-16 ***
## factor(statefip)30 -0.3506602  0.0072617  -48.289  < 2e-16 ***
## factor(statefip)31  0.1562548  0.0056157   27.825  < 2e-16 ***
## factor(statefip)32 -0.3495726  0.0052450  -66.649  < 2e-16 ***
## factor(statefip)33 -0.0076196  0.0063661   -1.197  0.23134
## factor(statefip)34  0.1352641  0.0038062   35.538  < 2e-16 ***
## factor(statefip)35  0.1276112  0.0055118   23.152  < 2e-16 ***
## factor(statefip)36  0.3273757  0.0033994   96.303  < 2e-16 ***
## factor(statefip)37  0.0851775  0.0036495   23.340  < 2e-16 ***
## factor(statefip)38 -0.0248313  0.0088216   -2.815  0.00488 **
## factor(statefip)39 -0.0884359  0.0035751  -24.737  < 2e-16 ***
## factor(statefip)40 -0.0879025  0.0045500  -19.319  < 2e-16 ***
## factor(statefip)41 -0.1056231  0.0046240  -22.842  < 2e-16 ***
## factor(statefip)42 -0.0520946  0.0035143  -14.824  < 2e-16 ***
## factor(statefip)44  0.1625615  0.0064355   25.260  < 2e-16 ***
## factor(statefip)45  0.0991751  0.0042337   23.425  < 2e-16 ***
## factor(statefip)46  0.0884070  0.0087374   10.118  < 2e-16 ***
## factor(statefip)47 -0.1269012  0.0039282  -32.306  < 2e-16 ***
## factor(statefip)48  0.0621990  0.0032757   18.988  < 2e-16 ***
## factor(statefip)49 -0.2480599  0.0047955  -51.728  < 2e-16 ***
## factor(statefip)50  0.0405137  0.0082352    4.920 8.67e-07 ***
## factor(statefip)51  0.0533376  0.0037437   14.247  < 2e-16 ***
## factor(statefip)53 -0.0680781  0.0040170  -16.948  < 2e-16 ***
## factor(statefip)54 -0.3071424  0.0058052  -52.908  < 2e-16 ***
## factor(statefip)55 -0.0939674  0.0040840  -23.009  < 2e-16 ***
## factor(statefip)56  0.1054025  0.0096690   10.901  < 2e-16 ***
## factor(year)2013   -0.0474506  0.0016016  -29.626  < 2e-16 ***
## factor(year)2014   -0.0964394  0.0015983  -60.337  < 2e-16 ***
## factor(year)2015   -0.1259655  0.0016020  -78.629  < 2e-16 ***
## factor(year)2016   -0.1231266  0.0016020  -76.856  < 2e-16 ***
## factor(year)2017   -0.1580440  0.0015971  -98.955  < 2e-16 ***
## factor(year)2018   -0.2079911  0.0015871 -131.050  < 2e-16 ***
## factor(year)2019   -0.2125043  0.0015807 -134.433  < 2e-16 ***
## factor(year)2020   -0.2448632  0.0015859 -154.400  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 41626448  on 316464  degrees of freedom
## Residual deviance: 40628731  on 316399  degrees of freedom
## AIC: 40628863
##
## Number of Fisher Scoring iterations: 6
```

```
with(summary(logmod1$glm_res), 1 - deviance/null.deviance) #print r squared
```

```
## [1] 0.02396834
```

```
#LOGIT - DFL REWEIGHT
logmod2 <- miceadds::glm.cluster(data=df,
                                 formula= enroll ~ da + female + hispanic + black + othrace + citizen +
                                         factor(statefip) + factor(year),
                                 cluster="state",
                                 weights= df$dflrewt,
                                 family=binomial(link= "logit"))
summary(logmod2$glm_res)
```

```
##
## Call:
## stats::glm(formula = formula, family = family, data = data, weights = wgt__)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -45.319    0.000    0.000    0.001   42.200
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -1.023e+00  3.751e+00  -0.273 0.785084
## da                   1.023e-01  1.087e-02   9.417  < 2e-16 ***
## female               5.134e-01  5.261e-03  97.589  < 2e-16 ***
## hispanic            -2.742e-01  8.543e-03 -32.091  < 2e-16 ***
## black               -2.718e-01  1.361e-02 -19.968  < 2e-16 ***
## othrace             -3.662e-01  9.836e-03 -37.235  < 2e-16 ***
## citizen             -2.494e-01  1.744e-02 -14.304  < 2e-16 ***
## insurance            8.887e-01  7.414e-03 119.863  < 2e-16 ***
## english              7.590e-01  1.460e-01   5.201 1.99e-07 ***
## factor(statefip)2   -4.752e-01  5.336e+05   0.000 0.999999
## factor(statefip)4   -6.667e-02  3.796e+00  -0.018 0.985988
## factor(statefip)5   -4.646e-02  3.748e+00  -0.012 0.990109
## factor(statefip)6    1.917e-01  7.086e+04   0.000 0.999998
## factor(statefip)8   -5.437e-02  1.990e+05   0.000 1.000000
## factor(statefip)9    7.680e-02  2.271e+05   0.000 1.000000
## factor(statefip)10  -1.487e-01  4.770e+05   0.000 1.000000
## factor(statefip)11  -8.976e-02  5.640e+05   0.000 1.000000
## factor(statefip)12   1.337e-01  3.781e+00   0.035 0.971792
## factor(statefip)13   2.871e-03  1.360e+05   0.000 1.000000
## factor(statefip)15  -1.163e-01  4.077e+05   0.000 1.000000
## factor(statefip)16  -2.719e-01  3.748e+00  -0.073 0.942160
```

```
## factor(statefip)17  1.532e-01  2.606e+04   0.000 0.999995
## factor(statefip)18 -1.137e-01  3.748e+00  -0.030 0.975806
## factor(statefip)20  1.097e-01  3.748e+00   0.029 0.976641
## factor(statefip)21 -2.204e-01  4.646e+00  -0.047 0.962166
## factor(statefip)22 -8.039e-02  3.773e+00  -0.021 0.982998
## factor(statefip)23 -1.542e-01  3.751e+00  -0.041 0.967212
## factor(statefip)24  1.589e-01  1.884e+05   0.000 0.999999
## factor(statefip)25  1.100e-01  1.639e+05   0.000 0.999999
## factor(statefip)26  3.818e-02  1.769e+01   0.002 0.998278
## factor(statefip)27  3.466e-02  3.777e+00   0.009 0.992678
## factor(statefip)28  2.342e-01  1.455e+05   0.000 0.999999
## factor(statefip)29  5.146e-02  3.755e+00   0.014 0.989066
## factor(statefip)30 -3.418e-01  3.748e+00  -0.091 0.927331
## factor(statefip)31  1.423e-01  3.748e+00   0.038 0.969706
## factor(statefip)32 -3.186e-01  2.203e+04   0.000 0.999988
## factor(statefip)33 -4.190e-02  1.986e+05   0.000 1.000000
## factor(statefip)34  1.444e-01  1.570e+05   0.000 0.999999
## factor(statefip)35  1.719e-01  3.749e+00   0.046 0.963427
## factor(statefip)36  3.170e-01  4.374e+03   0.000 0.999942
## factor(statefip)37  8.072e-02  1.402e+05   0.000 1.000000
## factor(statefip)38 -4.696e-02  3.748e+00  -0.013 0.990002
## factor(statefip)39 -1.205e-01  3.748e+00  -0.032 0.974349
## factor(statefip)40 -1.502e-03  3.748e+00   0.000 0.999680
## factor(statefip)41 -1.041e-01  3.748e+00  -0.028 0.977837
## factor(statefip)42 -8.646e-02  3.748e+00  -0.023 0.981598
## factor(statefip)44  1.298e-01  6.088e+02   0.000 0.999830
## factor(statefip)45  9.209e-02  2.023e+05   0.000 1.000000
## factor(statefip)46  8.103e-02  3.748e+00   0.022 0.982750
## factor(statefip)47 -1.520e-01  3.751e+00  -0.041 0.967671
## factor(statefip)48  9.914e-02  6.343e+01   0.002 0.998753
## factor(statefip)49 -2.651e-01  3.748e+00  -0.071 0.943616
## factor(statefip)50 -3.282e-02  3.770e+00  -0.009 0.993054
## factor(statefip)51  5.910e-02  1.509e+05   0.000 1.000000
## factor(statefip)53 -4.444e-02  1.791e+05   0.000 1.000000
## factor(statefip)54 -3.498e-01  3.748e+00  -0.093 0.925638
## factor(statefip)55 -1.304e-01  3.748e+00  -0.035 0.972247
## factor(statefip)56  6.296e-02  5.738e+00   0.011 0.991246
## factor(year)2013   -3.897e-02  1.113e-02  -3.501 0.000464 ***
## factor(year)2014    1.840e-01  1.160e-02  15.861  < 2e-16 ***
## factor(year)2015    1.093e-01  1.157e-02   9.446  < 2e-16 ***
## factor(year)2016   -7.148e-02  1.141e-02  -6.265 3.73e-10 ***
## factor(year)2017   -9.035e-02  1.171e-02  -7.717 1.19e-14 ***
## factor(year)2018    5.897e-02  1.193e-02   4.942 7.74e-07 ***
## factor(year)2019   -2.492e-01  1.178e-02 -21.161  < 2e-16 ***
## factor(year)2020   -2.005e-01  1.181e-02 -16.972  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 866037  on 316464  degrees of freedom
## Residual deviance: 828334  on 316399  degrees of freedom
## AIC: 822147
##
```

```
## Number of Fisher Scoring iterations: 6
```

```r
with(summary(logmod2$glm_res), 1 - deviance/null.deviance) #print r squared
```

```
## [1] 0.04353564
```

EVENT STUDY This chunk performs an event study using the feols function from the fixest package. The first call writes a new variable to comply with the requirements of the package. The second call runs an OLS regression (note: the package does not currently allow for binomial regression, which is more appropriate for this study). The last two calls print a summary of the regression and plot the results.

```r
#### EVENT STUDY
##https://lost-stats.github.io/Model_Estimation/Research_Design/event_study.html

event_study <- df %>%
        mutate(year_treated = case_when(state == "Idaho" ~ 2016,
                                        state == "South Dakota" ~ 2018,
                                        state != "Idaho" | state != "South Dakota" ~ 0))

es_dflrewt = feols(enroll ~ sunab(year_treated, year, ref.p = -1) +
        female + hispanic + black + othrace + citizen + insurance + english |
        state + year,
                   cluster = ~statefip,
                   weights = event_study$dflrewt,
                   data = event_study)

summary(es_dflrewt)
```

```
## OLS estimation, Dep. Var.: enroll
## Observations: 316,465
## Weights: event_study$dflrewt
## Fixed-effects: state: 50,  year: 9
## Standard-errors: Clustered (statefip)
##             Estimate Std. Error   t value   Pr(>|t|)
## year::-6    0.058046   0.019992  2.903469 5.5197e-03 **
## year::-5   -0.102623   0.034408 -2.982560 4.4450e-03 **
## year::-4   -0.032725   0.011129 -2.940407 4.9907e-03 **
## year::-3    0.013318   0.022239  0.598861 5.5202e-01
## year::-2   -0.069611   0.010956 -6.353773 6.6518e-08 ***
## year::0     0.025592   0.015875  1.612020 1.1338e-01
## year::1    -0.025596   0.006456 -3.964917 2.3863e-04 ***
## year::2    -0.049823   0.013552 -3.676528 5.8646e-04 ***
## year::3     0.035072   0.011748  2.985371 4.4106e-03 **
## year::4     0.037094   0.020975  1.768437 8.3213e-02 .
## female      0.118704   0.014514  8.178352 1.0210e-10 ***
## hispanic   -0.062736   0.044024 -1.425042 1.6048e-01
## black      -0.060991   0.027245 -2.238646 2.9754e-02 *
## othrace    -0.086316   0.024048 -3.589402 7.6486e-04 ***
## citizen    -0.052668   0.061846 -0.851607 3.9858e-01
## insurance   0.214399   0.014664 14.620594  < 2.2e-16 ***
## english     0.144878   0.096740  1.497598 1.4065e-01
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.677173     Adj. R2: 0.063284
##                  Within R2: 0.050614
```

```r
iplot(es_dflrewt,
      xlab = 'Time to treatment',
      main = 'Event study: Staggered treatment (TWFE)')
```