Week 9: Difference-in-differences II

Marcelo Coca Perraillon

University of Colorado
Anschutz Medical Campus

Health Services Research Methods I
HSMP 7607
2020

Outline

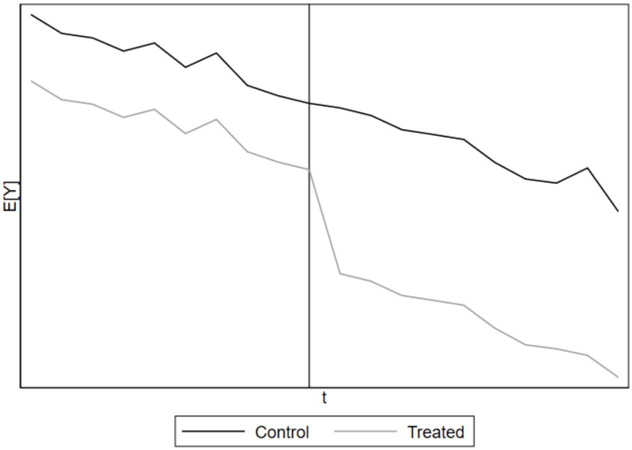
- Key elements
- Assumptions
- Two periods
- Estimating each separately
- Regression approach
- Adding “fixed effects”
- Testing parallel trends assumption

Key elements

- Remember the features of DiD:
  1. **Time**. There is always time in DiD – or said another way, events take place in time
  2. **Policy change or treatment** occurs at a point in time, which defines a before and after period. We always talk about "treatment" but in DiD a policy change is more common (epidemiologists talk about "exposure")
  3. **Comparison groups**. In DiD, one group receives the intervention or is subjected to the policy change *only in the post-period*. These groups do not need to be comparable
     (Aside: we won't cover triple DiD, but we could add another comparison group or another factor resulting three differences of differences)
  4. **Fixed factors**: We assume that important factors that area associated with with the outcome Y are fixed during the pre and post periods (more on this in a bit)
  5. **Time invariant factors**. If observed, we can control for those factors that could affect trends and vary over time
- Ideally, we observe outcomes for several periods because then we can verify a key assumption: parallel trends or constant bias

Example

- Simulated trends before and after a treatment or policy change



Review of derivation

- We assume that the outcome is determined by $Y_{it} = c_i + d_t + \delta D_{it} + \eta_{it}$, where i indexes the unit of observation and t indexes time
- c are d are variables, not coefficients. η_{it} is an unexplained, random error
- So the outcome depends on **constant** (**fixed**, **time-invariant**) factors at the unit of observation level (c_i) and factors that depend on time (d_t) but not on unit of observation i (plus randomness)
- Again. Factor(s) c do not change by time but do change by unit. Factor(s) d do not change by unit but can change by time
- We are implicitly assuming homogenous treatment effects – no covariates yet, but also homogeneous respect to time. We saw too that we could define treatment effect as depending on time as well (in the potential outcomes version we did so)

Differencing

- **(1)** Treated group after and before: $E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1] = c_i + d_1 + \delta_1 - (c_i + d_0 + \delta_0) = d_1 - d_0 + (\delta_1 - \delta_0) = d_1 - d_0 + \delta$
- That's a before and after comparison. The difference depends on factors that could change over time. If they don't change, then $d_1 = d_0$ and thus $E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1] = \delta$. *A before and after assumption must assume that nothing else changed*

  (2) Control group after and before: $E[Y_{i1}|D_i = 0] - E[Y_{i0}|D_i = 0] = c_i + d_1 - (c_i + d_0) = d_1 - d_0$
- The difference of the differences **(1)-(2)** is $= d_1 - d_0 + \delta - (d_1 - d_0) = \delta$
- The within group differencing got rid of factor c, the between group differencing got rid of d
- Note that we could have added more than one factor c or more factors d. The point is that they were fixed either within group or between group

What about other factors (covariates)?

- If we follow the logic of differencing, then we do not need to account for any other observed or unobserved constant (fixed, **time-invariant**) factors
- But we could take into account factors that vary at the unit of observation and by time (**time-varying covariates**)
- This means that we can extend our notation to condition for a vector of covariates $\mathbf{X}_{it}$, although we will imposed some assumptions when using regression analysis (exogeneity)
- So now we have: $Y_{it} = c_i + d_t + \delta D_{it} + \mathbf{X}'_{it}\beta + \eta_{it}$
- We could also "fixed effects" variables: for example, dummy indicators for state, hospital, county, person. It's way of using longitudinal data to get at causality. There is an obvious connection with DiD (more on this soon)

Estimation

- We often do a before and after comparison, even when we have more years (more on this in later classes)
- So we only need four means to estimate a DiD design
- A before and after comparison of outcome Y for the treated is: $E[Y_{tpost}] - E[Y_{tpre}]$. We want to compare that difference with the difference in the control: $E[Y_{cpost}] - E[Y_{cpre}]$
- The estimate of interest is:

  $$\Delta_{DiD} = E[Y_{tpost}] - E[Y_{tpre}] - \{E[Y_{cpost}] - E[Y_{cpre}]\}$$
- No regression model here yet, but we could estimate those four means parametrically or nonparametrically or semiparametrically
- The difference above is the same as:

  $$\Delta_{DiD} = E[Y_{tpost}] - E[Y_{cpost}] - \{E[Y_{tpre}] - E[Y_{cpre}]\}$$

Assumptions

- Following Lechner (2010)
  1. **SUTVA**. No interference (spillovers) and variation in treatment. Imagine a DiD using CO counties as treated and controls – clearly a problem
  2. **Exogeneity**: The covariates **X** are not influenced by the treatment. We saw similar assumptions in regression adjustment, although we don't make it explicit
  3. **Common trends** or constant bias. If the treated had not been treated, both treatment and control groups would have the same trends over time (possibly conditioning for other factors). Constant bias is the same assumption. Treated and control groups are not equivalent/comparable, but that difference remains constant over time
- The last assumption could be divided into an assumption about observed parallel trends before the intervention and the idea that "shocks" have a common effect in both groups

Example

- We are going to use the minimum wage example of Angrist and Pischke (from Card and Krueger's 1994 paper) since it's intuitive and current – plus we have the data. Florida just voted to increase the minimum wage to $15 per hour
- Here is the issue in a nutshell: in theory, a firm makes hiring decisions based on wages and the contribution of employees to revenue. In the most basic framework of a perfectly competitive market – the non-existent unicorn – a higher minimum wage implies that firms will demand fewer workers (or hours)
- Thus, a policy that helps those who can get jobs at the higher wage may harm some workers that won't find employment *because* of the higher minimum wage
- This has always been controversial. Crystal clear in theory, but real life may deviate from the model, so obtaining empirical would be important to revise the theory (in an ideal economics world)

Example

- In 1992, New Jersey raised the state minimum wage by about 19%, a relatively large increase (from \$4.25 to \$5.05)
- (Aside: The current national minimum wage is \$7.25. A minimum wage of \$15 is \$31,200 per year)
- Fast food chains are large employers that usually pay minimum wages, so card and Krueger obtained data from February 1992 (before or pre period) and November 1992 (the after or post period)
- They used data from similar fast food restaurants in Pennsylvania, which did not change the minimum wage (\$4.25)
- So we have the key elements of a basic, two-period DiD: time, a policy change or treatment, a treated group, and a treatment that is applied to only one group in the post-period
- (Notation: in Chapter 5, Angrist and Pischke use γ instead of c and λ instead of d. They are variables, not parameters. They use three indexes: store i, state s, and time t. But there are only two states, one treated and one not treated, so they could have used only two indexes: i and t)

Data

- We will use the same dataset (CardKrueger1994.dta), although some numbers do not match. I removed stores with missing values (see do file). The outcome for stores that closed is 0

```
rename t post
describe
Contains data from H:\Teaching\Methods 2020\lectures\Week 9 difference-in-differences\DiD\code\CardK
> rueger1994.dta
  obs:          820                        Dataset from Card&Krueger (1994)
 vars:            8                        27 May 2011 20:36
----------------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
----------------------------------------------------------------------------------------------------
id              int     %8.0g                 Store ID
post            byte    %8.0g                 Feb. 1992 = 0; Nov. 1992 = 1
treated         long    %8.0g      treated    New Jersey = 1; Pennsylvania = 0
fte             float   %9.0g                 Output: Full Time Employment
bk              byte    %8.0g                 Burger King == 1
kfc             byte    %8.0g                 Kentuky Fried Chiken == 1
roys            byte    %8.0g                 Roy Rogers == 1
wendys          byte    %8.0g                 Wendy's == 1
----------------------------------------------------------------------------------------------------
Sorted by: id  post
```

Structure

- We have longitudinal (as opposed to cross-sectional data)
- For each store i, we have two observations (balanced data). So two pairs of observation are treated or controls for both periods

```
list id fte treated post  bk kfc roys wendys in 1/10, sep(2) nolabel

     +-----------------------------------------------------------+
     | id    fte   treated   post   bk   kfc   roys   wendys |
     |-----------------------------------------------------------|
  1. |  1     31         1      0    1     0      0        0 |
  2. |  1     40         1      1    1     0      0        0 |
     |-----------------------------------------------------------|
  3. |  2     13         1      0    1     0      0        0 |
  4. |  2   12.5         1      1    1     0      0        0 |
     |-----------------------------------------------------------|
  5. |  3   12.5         1      0    0     1      0        0 |
  6. |  3    7.5         1      1    0     1      0        0 |
     |-----------------------------------------------------------|
  7. |  4     16         1      0    0     0      1        0 |
  8. |  4     20         1      1    0     0      1        0 |
     |-----------------------------------------------------------|
  9. |  5     20         1      0    0     0      1        0 |
 10. |  5     25         1      1    0     0      1        0 |
     +-----------------------------------------------------------+
```

Baseline

- Characteristics before the policy change

```
sum treated bk kfc roys wendys if treated ==1 & post ==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     treated |        315           1           0          1          1
          bk |        315    .4095238    .4925283          0          1
         kfc |        315     .215873    .4120812          0          1
        roys |        315     .247619    .4323161          0          1
      wendys |        315    .1269841     .333485          0          1

sum treated bk kfc roys wendys if treated ==0 & post ==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     treated |         76           0           0          0          0
          bk |         76    .4473684     .500526          0          1
         kfc |         76    .1578947    .3670652          0          1
        roys |         76    .2236842    .4194817          0          1
      wendys |         76    .1710526     .379057          0          1
```

Estimator

- Estimate and compare means

```
* Treated
sum fte if treated ==1 & post==1
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         fte |        315    17.49921    8.809567          0       55.5
scalar y_tpost = r(mean)
sum fte if treated ==1 & post==0
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         fte |        315    17.04683    8.812696          3         80
scalar y_tpre = r(mean)

* Control
sum fte if treated ==0 & post==1
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         fte |         76    17.52303    7.960023          0      38.25
scalar y_cpost = r(mean)
sum fte if treated ==0 & post==0
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         fte |         76    20.01316     11.8232        4.5       67.5
scalar y_cpre = r(mean)

* DiD estimator
di y_tpost - y_tpre - (y_cpost - y_cpre)
2.9425125
```

Intuition

- Full-time equivalent in the treated state (NJ) actually *increased* (slightly) before and after (17.50 vs 17.04). But this could be due to other factors that changed that we are not controlling for
- That's why we have a control. Employment *decreased* in the control state (PA; from 20.01 to 17.52)
- The key assumption here are that at the state-level, we control for state-specific **fixed effects** c_i because of the first differencing
- NJ and PA are different states with different baseline levels of employment (17.04 vs 20.01), but we assume that whichever factors are different between the states reminded constant. **Not the factors themselves, but the difference** $(d_1 - d_0)$ (THIS IS IMPORTANT - sorry for yelling)
- That's the constant bias part or the parallel trend that we can't test in this example
- If we suspect that there are factors that changed within states or between states that would affect the constant bias, we should control for them adding them to $\mathbf{X}$ in a regression approach

Estimation

- Remember regression adjustment using the command -teffects ra- approach, semiparametrically. We could do the same here. However, the most common strategy is a parametric model:

  $FTE_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \times Post_t) + \epsilon_{it}$
- Treatment D doesn't depend on time (a fast food restaurant i is treated or control in both periods), while *Post* depends on time but it's the same by store i
- This is a saturated model; we will get four predicted means, which replicates what we did with the -summarize- command
- But now we can test if the DiD estimator is statistically significant: $H_0 : \beta_3 = 0$
- **Is this model correct**? Maybe. We haven't checked residuals. Don't forget this. What is the consequence? Maybe SEs are not correct

Regression approach

- DiD estimator is not statistically significant at 0.05

```
reg fte i.treated##i.post, robust

Linear regression                               Number of obs     =        782
                                                F(3, 778)         =       1.42
                                                Prob > F          =     0.2341
                                                R-squared         =     0.0084
                                                Root MSE          =     9.0693

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |
         NJ  |  -2.966332   1.439264    -2.06   0.040    -5.791633   -.1410318
      1.post |  -2.490132   1.628318    -1.53   0.127    -5.686548    .7062852
             |
treated#post |
       NJ#1  |   2.942513   1.773501     1.66   0.097    -.5389025    6.423928
             |
       _cons |   20.01316   1.350721    14.82   0.000     17.36167    22.66465
------------------------------------------------------------------------------
```

Predictions with marginal effects

- Remember, marginal effects are predictions

```
* Predictive margins
margins treated, at(post=(0 1))
Adjusted predictions                            Number of obs     =        782
Model VCE    : Robust
Expression   : Linear prediction, predict()
1._at        : post            =           0
2._at        : post            =           1
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _at#treated |
       1#PA  |   20.01316   1.350721    14.82   0.000     17.36167    22.66465
       1#NJ  |   17.04683   .4970232    34.30   0.000     16.07116    18.02249
       2#PA  |   17.52303    .909379    19.27   0.000      15.7379    19.30815
       2#NJ  |   17.49921   .4968467    35.22   0.000     16.52389    18.47453
------------------------------------------------------------------------------
* Marginal effects
margins, dydx(treated) at(post=(0 1))
Conditional marginal effects                    Number of obs     =        782
Model VCE    : Robust
Expression   : Linear prediction, predict()
dy/dx w.r.t. : 1.treated
1._at        : post            =           0
2._at        : post            =           1
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0.treated    |  (base outcome)
-------------+----------------------------------------------------------------
1.treated    |
         _at |
          1  |  -2.966332   1.439264    -2.06   0.040    -5.791633   -.1410318
          2  |    -.02382   1.036256    -0.02   0.982    -2.058009    2.010369
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Model interpretation reminder

- In general, we can show that the model is a differences of differences (using a version without covariates; with covariates we need to hold them constant)

  $E[Y]$ Treated in post period: $\beta_0 + \beta_1 + \beta_2 + \beta_3$

  $E[Y]$ Treated in pre period: $\beta_0 + \beta_1$

  (1) Difference treated post - pre: $\beta_2 + \beta_3$

  $E[Y]$ Control in post period: $\beta_0 + \beta_2$

  $E[Y]$ Control in pre period: β_0

  (2) Difference control post - pre: β_2

  Difference of differences (1)-(2): $\Delta_{DiD} = \beta_3$
- **Caution**: Interacted models are not difference-in-differences **research designs**, but interactions with dummy variables are difference-in-differences

We can do the same with incremental effects

- The model is:
  $E[FTE_{it}|D_i, Post_t] = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3(D_i \times Post_t)$
- Incremental effect: $\frac{\Delta E[FTE_{it}|D_i, Post_t]}{\Delta Post_t} = \beta_2 + \beta_3 D_i$
- So β_2 is the effect or difference in average employment for the *control* (PA) before and after (17.52 - 20.01 = -2.96).
- $\beta_2 + \beta_3$ is the difference in average employment for the *treated* before and after: $17.49921 - 17.04683 = -2.490132 + 2.942513 = .452381$ (first is from sum command, second from coefficients)
- β_3 by itself is the difference of the difference
- Don't underestimate this. You could have done in the other way
  $\frac{\Delta E[FTE_{it}|D_i, Post_t]}{\Delta D_i} = \beta_1 + \beta_3 Post$

Adding covariates

- There is a difference between chain restaurants in PA and NJ, so we could control for this adding dummies for type of chain
- We will leave Burger King as reference since it's the largest (doesn't really matter since we care about the β_3). We can write the model as: $FTE_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \times Post_t) + \alpha_i + \epsilon_{it}$
- Note the α_i. **That's a shortcut for writing dummy variables**

```
reg fte i.treated##i.post kfc roys wendys, robust
Linear regression                               Number of obs     =        782
                                                F(6, 775)         =      56.25
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1898
                                                Root MSE          =     8.2137

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |
         NJ  |  -2.375187   1.281091    -1.85   0.064    -4.890007    .1396337
      1.post |  -2.490132    1.43552    -1.73   0.083    -5.308099    .3278363
             |
treated#post |
       NJ#1  |   2.942513   1.572405     1.87   0.062    -.1441661    6.029191
             |
         kfc |  -10.17891   .6112341   -16.65   0.000    -11.37878   -8.979042
        roys |  -1.902694    .821605    -2.32   0.021    -3.515529   -.2898592
      wendys |  -1.010944   .9793684    -1.03   0.302    -2.933473    .9115851
       _cons |   22.21888   1.327919    16.73   0.000     19.61214    24.82563
------------------------------------------------------------------------------
```

What just happened?

- That is what an economist would call adding "chain type fixed effects"
- Now, the treatment effect didn't change at all. But why would it change? We are comparing the same restaurants before and after, so chain type couldn't affect the difference-in-difference
- Chain type is not a confounder since chain type is not changing before and after

```
tab chaintype treated if post ==0
           |    New Jersey = 1;
           |  Pennsylvania = 0
 chaintype |        PA         NJ |     Total
-----------+----------------------+----------
         1 |        34        129 |       163
         2 |        12         68 |        80
         3 |        17         78 |        95
         4 |        13         40 |        53
-----------+----------------------+----------
     Total |        76        315 |       391
           |    New Jersey = 1;
           |  Pennsylvania = 0
 chaintype |        PA         NJ |     Total
-----------+----------------------+----------
         1 |        34        129 |       163
         2 |        12         68 |        80
         3 |        17         78 |        95
         4 |        13         40 |        53
-----------+----------------------+----------
     Total |        76        315 |       391
```

Variance explained

- But note that the precision of the estimate did change - p-value for β_3 is smaller now
- Note too that R^2 increased – we are explaining more of the variability in FTE because chain type does affect FTE
- Remember the traffic in Chicago example. The outcome by itself has a variance. We can explain some of this variance with a regression model
- The type of chain is correlated with FTE, so we will explain more of the variance by adding chain type fixed effects, which improves the model
- Yet, chain type is not going to change our estimate of treatment effects
- My former boss of three weeks used to say something like "Add XYZ to the model to soak up variance"
- Now, don't forget that we are talking about linear regression here. Careful with non-linear models. In logit/probit models, for example, the variance is fixed, so there is nothing to "soak up"

Making connections: variance explained

- Compare $\sqrt{var(fte)}$ observed (9.09) with model Root MSE $= 8.2137$. R^2 is the variance explained adjusted for degrees of freedom

```
tabstat fte, by(chaintype) stats(N mean sd)
Summary for variables: fte
     by categories of: chaintype

chaintype |         N      mean        sd
----------+------------------------------
        1 |       326  20.25844  8.757839
        2 |       160  10.02656  4.822526
        3 |       190  18.32895  9.195105
        4 |       106  19.28066  8.754706
----------+------------------------------
    Total |       782  17.56362  9.090051
-----------------------------------------
* Reminder
qui reg fte i.treated##i.post kfc roys wendys
* Save root MSE
scalar rmse = e(rmse)
* Save outcome variance
qui sum fte
scalar yvar = r(sd)^2
* R^2
di 1-( (rmse^2)*(782-7)) / (yvar*(782-1))
.18978931
```

Checking assumptions

- As we have discussed before, there are some assumptions that are "exclusion restrictions" that you need to argue about
- Other assumptions can be tested with data, or at least you can see if the data seems consistent with an assumption, even though it may not guaranteed that the assumption is valid
- The parallel trends assumption is one of them
- Graphically showing that trends are parallel is a good first step, but we can test the assumption as well
- We will see adjusted and unadjusted plots – sometimes we may need to adjust for factors that affect the difference in trends

Data

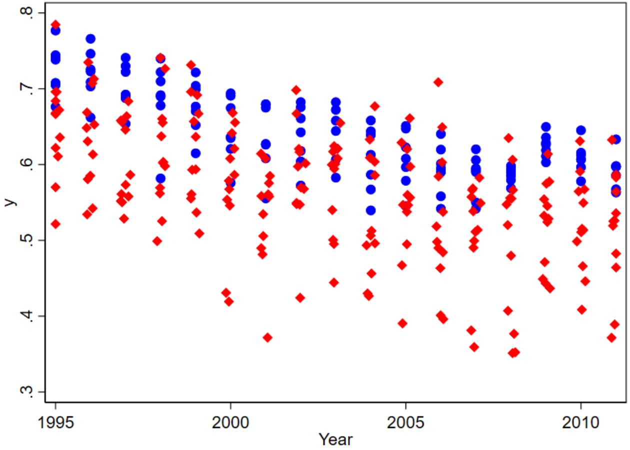
- We are going to use a sample dataset with economic data by country (the dataset itself is not that interesting; just an example)
- Each country has data from 1995 to 2011. I renamed some variables (and added some noise)

```
list year year2 country y treated in 1/20

     +---------------------------------------------+
     | year   year2   country          y   treated |
     |---------------------------------------------|
  1. | 1995       1       AUS   .7845408         0 |
  2. | 1996       2       AUS   .7145114         0 |
  3. | 1997       3       AUS   .6829197         0 |
  4. | 1998       4       AUS   .5660708         0 |
  5. | 1999       5       AUS   .5964009         0 |
     |---------------------------------------------|
  6. | 2000       6       AUS   .6703118         0 |
  7. | 2001       7       AUS    .602838         0 |
  8. | 2002       8       AUS   .6675223         0 |
  9. | 2003       9       AUS   .5423564         0 |
 10. | 2004      10       AUS   .5826657         0 |
     |---------------------------------------------|
 11. | 2005      11       AUS   .6592514         0 |
 12. | 2006      12       AUS   .6497246         0 |
 13. | 2007      13       AUS   .5503935         0 |
 14. | 2008      14       AUS   .6061184         0 |
 15. | 2009      15       AUS   .5784074         0 |
     |---------------------------------------------|
 16. | 2010      16       AUS   .6350716         0 |
 17. | 2011      17       AUS   .5618598         0 |
 18. | 1995       1       BRA   .7384017         1 |
 19. | 1996       2       BRA   .7255375         1 |
 20. | 1997       3       BRA   .7221702         1 |
     +---------------------------------------------+
```

Raw data

- Raw data by year for treated (blue) and controls (red)

```
scatter y year if treated ==1, mcolor(blue) legend(off) || ///
        scatter y year if treated ==0, mcolor(red) msize(small) jitter(2)
graph export raw.png, replace
```



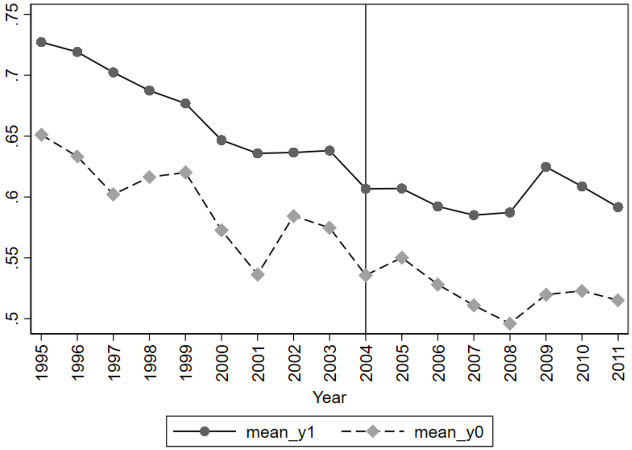
Plotting average trends

- We want to plot $E[Y]$ by year for each treated group as the first step
- Several ways to do this: first, the longish, "manual" way
- Pay attention to the prefix "by year:" (could be "bysort year")
- Check the label option "angle"

```
* By year, calculate means
sort year
by year: egen mean_y1 = mean(y) if treated==1
by year: egen mean_y0 = mean(y) if treated==0

* Plot
scatter mean_y1 year, connect(l) sort || scatter mean_y0 year, sort connect(l) ///
   xline(2004) xlabel(1995(1)2011, angle(vertical))
graph export trends1.png, replace
```

Plotting average trends

- Policy change takes place in 2004



Plotting average trends - using a regression model

- All we did was calculate the mean of Y by year and group
- We could do the same with a **saturated regression model**. The model would estimate many parameters: dummy variables for each year, one for treated, and their interactions
- Writing down the model can get messy:

  $Y_{it} = \alpha_0 + \sum_{j=1996}^{2011} \beta_j Year_j + \alpha D_i + \sum_{j=1996}^{2011} \gamma_j (Year_j \times D_i) + \epsilon_{it}$
- Not very elegant but as long as you can communicate what you did...
- Borrowing from ANOVA type notation, we could write: $Y_{it} = \alpha_i + \gamma_t + \delta_{it}$. Each Greek letter represent a set of dummies as before, δ_{it} is their interaction. More elegant

Regression model

```
reg y i.year##i.treated
      Source |       SS           df       MS      Number of obs   =       323
-------------+----------------------------------   F(33, 289)      =      8.71
       Model |  1.13068712        33  .034263246   Prob > F        =    0.0000
    Residual |  1.13743071       289  .003935746   R-squared       =    0.4985
-------------+----------------------------------   Adj R-squared   =    0.4413
       Total |  2.26811783       322  .007043844   Root MSE        =    .06274

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
       1996  |  -.0178561   .0256117    -0.70   0.486    -.0682651     .032553
       1997  |  -.0490813   .0256117    -1.92   0.056    -.0994904    .0013278
       1998  |  -.0347811   .0256117    -1.36   0.176    -.0851902     .015628
       1999  |  -.0310305   .0256117    -1.21   0.227    -.0814396    .0193786
       2000  |  -.0784285   .0256117    -3.06   0.002    -.1288376   -.0280195
       2001  |   -.114868   .0256117    -4.48   0.000    -.1652771    -.064459
       2002  |  -.0667116   .0256117    -2.60   0.010    -.1171206   -.0163025
       2003  |  -.0764574   .0256117    -2.99   0.003    -.1268665   -.0260484
       2004  |  -.1154181   .0256117    -4.51   0.000    -.1658272   -.0650091
       2005  |  -.1010258   .0256117    -3.94   0.000    -.1514349   -.0506167
       2006  |  -.1229533   .0256117    -4.80   0.000    -.1733624   -.0725443
       2007  |  -.1401936   .0256117    -5.47   0.000    -.1906026   -.0897845
       2008  |   -.15502    .0256117    -6.05   0.000     -.205429   -.1046109
       2009  |  -.1313859   .0256117    -5.13   0.000     -.181795   -.0809769
       2010  |  -.1281985   .0256117    -5.01   0.000    -.1786075   -.0777894
       2011  |  -.1359933   .0256117    -5.31   0.000    -.1864024   -.0855843
             |
   1.treated |   .0760852   .0298367     2.55   0.011     .0173605      .13481
             |
year#treated |
     1996 1  |   .0097807   .0421955     0.23   0.817    -.0732687      .09283
     1997 1  |   .0242146   .0421955     0.57   0.567    -.0588348    .1072639
     1998 1  |  -.0050124   .0421955    -0.12   0.906    -.0880617     .078037
     1999 1  |    -.01933   .0421955    -0.46   0.647    -.1023794    .0637193
     2000 1  |  -.0020724   .0421955    -0.05   0.961    -.0851218     .080977
     2001 1  |   .0234893   .0421955     0.56   0.578    -.0595601    .1065387
     2002 1  |  -.0239757   .0421955    -0.57   0.570    -.1070251    .0590737
...
```

Margins to the rescue –again

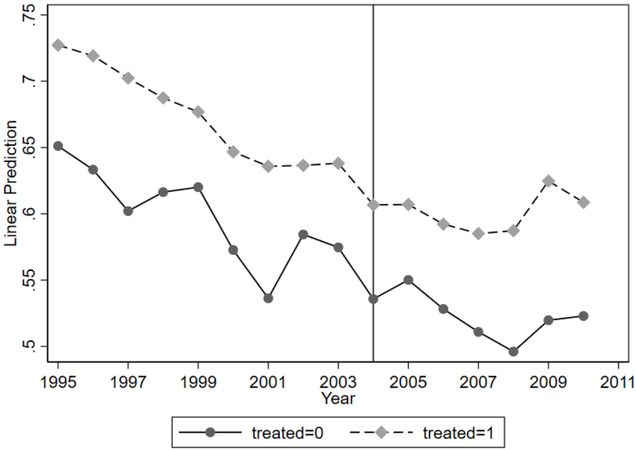
- Hope you start makings some connections and things start to click
- We could reproduce the mean of the outcome by treated group and year with the -margins- command. And then we can use -marginsplot- to make a graph. **This will became very helpful when we add covariates**

```
margins treated, at(year=(1995(1)2010)) vsquish
Adjusted predictions                              Number of obs     =        323
Model VCE    : OLS
Expression   : Linear prediction, predict()
1._at        : year            =        1995
2._at        : year            =        1996
...
...
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _at#treated |
        1 0  |   .6511415   .0181102    35.95   0.000     .6154969    .6867861
        1 1  |   .7272267   .0237118    30.67   0.000      .680557    .7738965
        2 0  |   .6332854   .0181102    34.97   0.000     .5976408      .66893
        2 1  |   .7191513   .0237118    30.33   0.000     .6724816     .765821
        3 0  |   .6020602   .0181102    33.24   0.000     .5664156    .6377048
        3 1  |     .70236   .0237118    29.62   0.000     .6556903    .7490297
        4 0  |   .6163604   .0181102    34.03   0.000     .5807158     .652005
        4 1  |   .6874333   .0237118    28.99   0.000     .6407635     .734103
        5 0  |    .620111   .0181102    34.24   0.000     .5844664    .6557556
        5 1  |   .6768662   .0237118    28.55   0.000     .6301965    .7235359
        6 0  |    .572713   .0181102    31.62   0.000     .5370684    .6083576
        6 1  |   .6467258   .0237118    27.27   0.000     .6000561    .6933956
        7 0  |   .5362735   .0181102    29.61   0.000     .5006289    .5719181
        7 1  |    .635848   .0237118    26.82   0.000     .5891783    .6825177
...
...
```

Marginsplot

- Note that -marginsplot- takes options to modify graphs

```
marginsplot, noci xlabel(1995(2)2011) title("") xline(2004)
graph export trends2.png, replace
```



Testing parallel trends - two pre-periods

- To test trends we need at least two pre-intervention observations
- Suppose that we only had two data points in the pre-period: 1995 and 2004 (we assume that the intervention took place at the end of 2004)
- A test of parallel trends would reduce to testing if the change in $E[Y]$ from 1995 to 2004 is the same in both groups. We don't know what happened in between those years
- We could estimate this model (with data for only 1995 and 2004):

  $Y_{it} = \beta_0 + \beta_1 Y2004_t + \beta_2 D_i + \beta_3 (Y2004_t \times D_i) + \epsilon_{it}$
- $Y2004$ is a dummy that equals 1 if 2004 and 0 if 1995
- If $\beta_3 = 0$, them the difference in the treated and control group outcome is the same in both years
- Confused? Write it down!: $\frac{\Delta E[Y]}{\Delta Y2004} = \beta_1 + \beta_3 D$
- (**Caution**: Same as DiD estimator, but it's not a DiD design. In the pre-period, there was no treatment)

Two periods

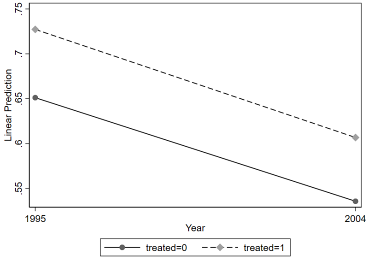
- We do not reject the null, change is the same (p-value: 0.908)

```
. reg y i.year##i.treated if inlist(year, 1995, 2004)

      Source |       SS           df       MS      Number of obs   =        38
-------------+----------------------------------   F(3, 34)        =     14.12
       Model |  .178571428         3  .059523809   Prob > F        =    0.0000
    Residual |  .143355272        34  .004216332   R-squared       =    0.5547
-------------+----------------------------------   Adj R-squared   =    0.5154
       Total |    .3219267        37  .008700722   Root MSE        =    .06493

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
       2004  |  -.1154181   .0265089    -4.35   0.000    -.1692907   -.0615455
             |
   1.treated |   .0760852   .0308819     2.46   0.019     .0133256    .1388449
             |
year#treated |
     2004 1  |  -.0050697   .0436737    -0.12   0.908    -.0938253    .0836858
             |
       _cons |   .6511415   .0187446    34.74   0.000     .6130478    .6892352
------------------------------------------------------------------------------
```

Two periods

- We can see this graphically with handy -maringsplot- as before

```
margins treated, at(year=(1995 2004)) vsquish
Adjusted predictions                              Number of obs     =         38
Model VCE    : OLS
Expression   : Linear prediction, predict()
1._at        : year            =        1995
2._at        : year            =        2004
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _at#treated |
        1 0  |   .6511415   .0187446    34.74   0.000     .6130478    .6892352
        1 1  |   .7272267   .0245425    29.63   0.000     .6773504     .777103
        2 0  |   .5357234   .0187446    28.58   0.000     .4976297     .573817
        2 1  |   .6067389   .0245425    24.72   0.000     .5568626    .6566152
------------------------------------------------------------------------------
marginsplot, noci
graph export twop.png, replace
```



Testing parallel trends - more periods

- Now, we could extend the same logic for all the periods, but there is one problem: that would be a lot to ask
- It's unrealistic to expect that the difference between the groups is going to be the same at every single point in time. We want it to be parallel, not the same
- Rather than such a stringent test, we could accept some year to year variability instead. We could "smooth" the trend by giving it some structure
- The most straightforward is a linear trend:

  $Y_{it} = \beta_0 + \beta_1 Year_t + \beta_2 D_i + \beta_3 (Year_t \times D_i) + \epsilon_{it}$
- It looks like the same model but it's not. Key difference is *Year* is not a dummy variable but the actual year (continuous: 1995, 1996, etc)
- Same result if year is 1, 2, 3...
- If $\beta_3 =$, then both groups have the **same slope**. That is, $E[Y]$ has been changing at the same rate

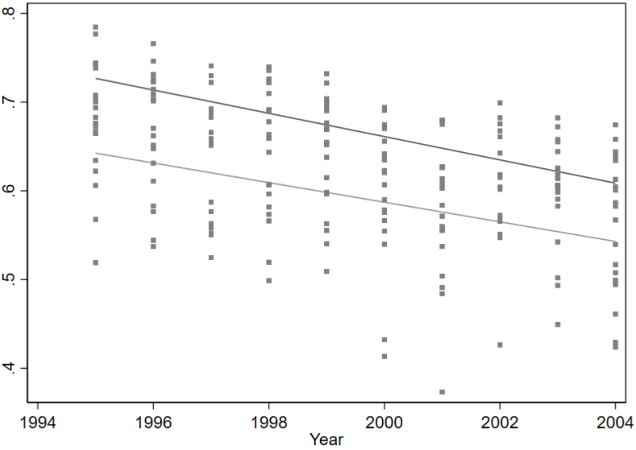
More periods

- We again don't reject the null (p-value: 0.54). Same slope
- Careful. We need to restrict estimation to the before period

```
reg y c.year##i.treated if year <= 2004

      Source |       SS           df       MS      Number of obs   =       190
-------------+----------------------------------   F(3, 186)       =     42.98
       Model |  .469401667         3  .156467222   Prob > F        =    0.0000
    Residual |   .67705273       186  .003640068   R-squared       =    0.4094
-------------+----------------------------------   Adj R-squared   =    0.3999
       Total |   1.1464544       189  .006065896   Root MSE        =    .06033

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |  -.0110593   .0019175    -5.77   0.000    -.0148421   -.0072764
   1.treated |   4.205181    6.31665     0.67   0.506    -8.256307    16.66667
             |
treated#c.year |
          1  |  -.0020656   .0031591    -0.65   0.514    -.0082979    .0041667
             |
       _cons |   22.70572   3.834061     5.92   0.000     15.14189    30.26956
------------------------------------------------------------------------------
```

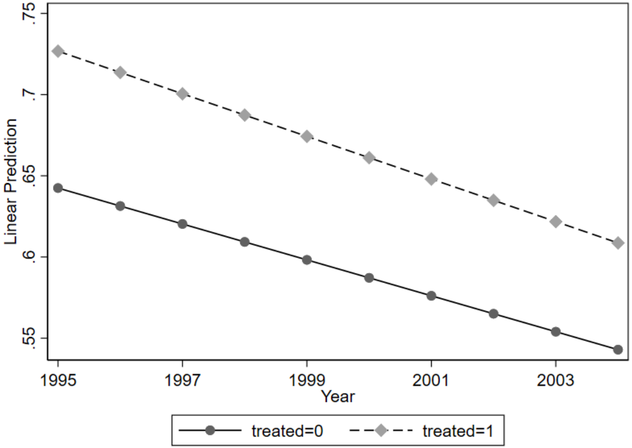
Graph model using predictions

```
line yhat year if year <= 2004 & treated ==1, sort ||  ///
     line yhat year if year <= 2004 & treated ==0, sort || ///
 scatter y year if year <=2004, msize(vsmall) legend(off)
graph export lin1.png, replace
```



Graph model using -marginsplot-

```
qui reg y c.year##i.treated if year <= 2004
qui margins treated, at(year=(1995(1)2004))
marginsplot, noci xlabel(1995(2)2004) title("")
graph export linmar.png, replace
```



Other specifications, covariates

- Trends could be non-linear. Maybe the best fitting model is a quadratic trends model or other functional form
- Remember that the difference between the groups may not be parallel in the raw, unadjusted data, but they could become parallel after "holding" other variables constant or after "taking into account" the effect of other variables
- Said another way, the trends could become parallel **conditional on other covariates**
- This is a common situation. The parallel trends test may fail with raw data (unadjusted) but it could pass when we control for covariates
- "Passing" here means that we **do not** reject the null
- We have all the tools to do this

Test parallel trends – adjusted

- We need to add covariates to our model:

  $Y_{it} = \beta_0 + \beta_1 Year_t + \beta_2 D_i + \beta_3(Year_t \times D_i) + \mathbf{X}'\beta + \epsilon_{it}$
- Those covariates could be time-varying or "fixed effects" so we could have a model like:

  $Y_{it} = \beta_0 + \beta_1 Year_t + \beta_2 D_i + \beta_3(Year_t \times D_i) + \mathbf{X}'_{it}\beta + \lambda_i + \epsilon_{it}$
- Remember, λ_i is a shortcut for a set of dummy/indicator variables. **It's not another intercept**
- That is, $\lambda_i = \sum_{j=2}^{k} \gamma_j Z_{ij} = \gamma_2 Z_{i2} + \cdots + \gamma_k Z_{ik}$, where k are the levels or categories of the variable Z
- In the minimum wage example, $k = 4$ because they were 4 chain brands. We need to leave one out as the reference category
- In some models we could add "time fixed effects" (that is, dummy for each year), so we would write λ_t instead of λ_i

Test parallel trends – adjusted

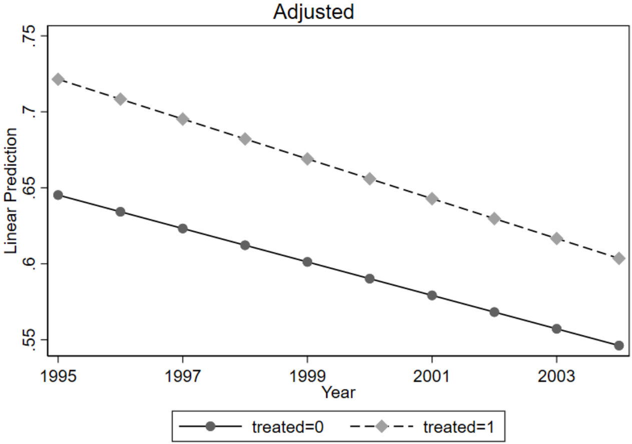
- The parallel trends tends “passes” again

```
reg y c.year##i.treated comlang log_distw if year <= 2004

      Source |       SS           df       MS      Number of obs   =       190
-------------+----------------------------------   F(5, 184)       =     26.89
       Model |   .48402909         5  .096805818   Prob > F        =    0.0000
    Residual |  .662425307       184  .003600138   R-squared       =    0.4222
-------------+----------------------------------   Adj R-squared   =    0.4065
       Total |   1.1464544       189  .006065896   Root MSE        =       .06

--------------------------------------------------------------------------------
             y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
          year |  -.0110056   .0019072    -5.77   0.000    -.0147683   -.0072429
     1.treated |   4.259519   6.282601     0.68   0.499    -8.135678    16.65472
               |
treated#c.year |
            1  |  -.0020969   .0031421    -0.67   0.505    -.0082961    .0041023
               |
       comlang |   .0330957   .0570804     0.58   0.563    -.0795205    .1457119
     log_distw |   .0109464   .0078496     1.39   0.165    -.0045403    .0264331
         _cons |   22.50697   3.814596     5.90   0.000       14.981    30.03294
--------------------------------------------------------------------------------
```

Adjusted plot

- Leaving covariates as observed

```
reg y c.year##i.treated comlang log_distw if year <= 2004
margins treated, at(year=(1995(1)2004))
marginsplot, noci xlabel(1995(2)2004) title("Adjusted")
graph export paradj.png, replace
```



Adjusted plot

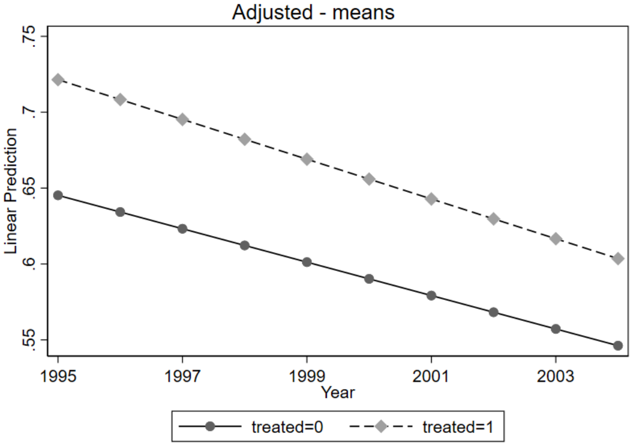
- Holding them at means (careful with syntax, you don't want the atmeans option; don't want to hold year at the mean)

```
margins treated, at(year=(1995(1)2004) (mean) comlang log_distw)
Adjusted predictions                                 Number of obs     =        190
Model VCE    : OLS
Expression   : Linear prediction, predict()

1._at        : year            =        1995
               comlang         =    .0804251 (mean)
               log_distw       =    8.392976 (mean)
2._at        : year            =        1996
               comlang         =    .0804251 (mean)
               log_distw       =    8.392976 (mean)
...
...
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _at#treated |
       1 0   |   .6452383   .0103865    62.12   0.000     .6247463    .6657303
       1 1   |   .7214602   .0138074    52.25   0.000      .694219    .7487014
       2 0   |   .6342326   .0088813    71.41   0.000     .6167104    .6517549
..
..
```

At means graph

- For graph, same syntax as before

```
marginsplot, noci xlabel(1995(2)2004) title("Adjusted - means")
graph export paradj_mean.png, replace
```



More on parametrization

- So far we have seen the standard parametric DiD regression model:
  $Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \times Post_t) + \epsilon_{it}$
- We can extend the model adding fixed effects (favorite not so magic powder):
  $$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \times Post_t) + \alpha_i + \epsilon_{it}$$
- We can also control for variables that we think could affect the evolution of the trends between treatment and control groups (remember, variables that affect the difference between trends):
  $$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \times Post_t) + \alpha_i + \mathbf{X}_{it}\beta + \epsilon_{it}$$
- If some variables in $\mathbf{X}$ do not change over time, they won't affect the DiD estimator (like chain type)

More on parametrization

- But we could define variables in a different way, with an **implicit** interaction term – this will be more useful when we have multiple periods
- So far we defined the treatment variable (or policy change) D_i as 1 if the unit (chain, country, state, person, etc) was treated and 0 otherwise
- D_i is the **same in both periods**. If the unit is treated, then its treatment indicator D_i is 1 in the pre and the post, even though we know that in the pre period no unit is treated
- We could instead define treatment in a different way. Let's define $D1_i$ as 1 for the treated group only in the post period, and 0 otherwise. The DiD parametric model becomes:

  $$Y_{it} = \gamma_0 + \gamma_1 D_i + \gamma_2 Post_t + \beta_3 D1_{it} + \epsilon_{it}$$
- In some papers, authors could write $D1_i$ instead of $D1_{it}$, which would make it clearer that $D \neq D1$
- There is no explicit interaction, but it's the same DiD estimator, we just defined the interaction as $D1$ since here $D1 \equiv D \times Post$. **Kind of trivial**

The model as before

- Model we estimated before

```
reg fte i.treated##i.post, robust

Linear regression                               Number of obs     =        782
                                                F(3, 778)         =       1.42
                                                Prob > F          =     0.2341
                                                R-squared         =     0.0084
                                                Root MSE          =     9.0693

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |
         NJ  |  -2.966332   1.439264    -2.06   0.040    -5.791633   -.1410318
      1.post |  -2.490132   1.628318    -1.53   0.127    -5.686548    .7062852
             |
treated#post |
       NJ#1  |   2.942513   1.773501     1.66   0.097    -.5389025    6.423928
             |
       _cons |   20.01316   1.350721    14.82   0.000     17.36167    22.66465
------------------------------------------------------------------------------
```

Coding the interaction

- New model with new coded variable that is trivially the same as the interaction

```
gen     treatp = treated
replace treatp = 0 if treated ==1 & post ==0
reg fte post treated treatp, robust
Linear regression                               Number of obs     =        782
                                                F(3, 778)         =       1.42
                                                Prob > F          =     0.2341
                                                R-squared         =     0.0084
                                                Root MSE          =     9.0693
------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        post |  -2.490132   1.628318    -1.53   0.127    -5.686548    .7062852
     treated |  -2.966332   1.439264    -2.06   0.040    -5.791633   -.1410318
      treatp |   2.942513   1.773501     1.66   0.097    -.5389025    6.423928
       _cons |   20.01316   1.350721    14.82   0.000     17.36167    22.66465
------------------------------------------------------------------------------
```

Coding the interaction

- Better when you want to use margins

```
reg fte i.post i.treated i.post#i.treated, robust nofvlabel

Linear regression                               Number of obs     =        782
                                                F(3, 778)         =       1.42
                                                Prob > F          =     0.2341
                                                R-squared         =     0.0084
                                                Root MSE          =     9.0693

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      1.post |  -2.490132   1.628318    -1.53   0.127    -5.686548    .7062852
   1.treated |  -2.966332   1.439264    -2.06   0.040    -5.791633   -.1410318
             |
post#treated |
        1 1  |   2.942513   1.773501     1.66   0.097    -.5389025    6.423928
             |
       _cons |   20.01316   1.350721    14.82   0.000     17.36167    22.66465
------------------------------------------------------------------------------
```

More on parametrization

- This is all somewhat trivial, but it can be confusing
- We could also say that we are adding "time fixed effects" in the model, which is of course the same as the *post* variable, with time coded with different values. So the model is:

  $Y_{it} = \gamma_0 + \gamma_1 D_i + \delta_t + \beta_3 D1_i + \epsilon_{it}$
- Furthermore, in this example, we could add "state fixed effects" as well. That's (trivially, again) just D_i. So:

  $Y_{it} = \gamma_0 + \alpha_i + \delta_t + \beta_3 D1_{it} + \epsilon_{it}$
- **NOT TRIVIAL**(!): The part that is not trivial is that when we have **multiple periods** not just two, δ_t would control for "time trends" and $D1$ is more general because it could accommodate **different timing of treatment** for some units
- Furthermore, with more periods, we could interact α_s and δ_t to have "state-specific time trends," which is often described as a robustness check when there are multiple periods: the DiD estimator shouldn't change (see Angrist and Pischke page 238)

Cat hair, bear claws, and a pinch of fixed effects

- First, last, and only cartoon to appear in my class notes

Time and state fixed effects

- Note below that I arbitrarily gave states an id of 4 and 8. Stata of course will code dummies so we go back to the same model
- Again, trivial in the sense that we are essentially recoding variables that are equivalent, but they are different ways of understanding DiD. It does make a difference with more time periods or states. Model below has time and state fixed effects: $Y_{it} = \gamma_0 + \alpha_i + \lambda_t + \delta D_{it} + \eta_{it}$

```
gen     time = 1 if post ==0
replace time = 2 if post ==1

gen     stateid = 4 if treated == 1
replace stateid = 8 if treated == 0

* Time and state fixed effects
reg fte i.time i.stateid treatp, robust

Linear regression                               Number of obs     =        782
                                                F(3, 778)         =       1.42
                                                Prob > F          =     0.2341
                                                R-squared         =     0.0084
                                                Root MSE          =     9.0693

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      2.time |  -2.490132   1.628318    -1.53   0.127    -5.686548    .7062852
   2.stateid |   2.966332   1.439264     2.06   0.040     .1410318    5.791633
      treatp |   2.942513   1.773501     1.66   0.097    -.5389025    6.423928
       _cons |   17.04683   .4970232    34.30   0.000     16.07116    18.02249
------------------------------------------------------------------------------
```

Big picture

- Careful with words and notation. Make sure you understand what fixed effects mean and what they are doing (or not) in the model
- They are not magic powders and the notation can be confusing. Sometimes papers are not clear and you need to infer what they did
- Same model could appear completely different. For example, $Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 Post_t + \beta_3 (D_i \times Post_t) + \epsilon_{it}$ versus $Y_{it} = \gamma_0 + \alpha_i + \delta_t + \beta_3 D_{it} + \eta_{it}$

  $Y_{ist} = \alpha_s + \delta_t + \beta_3 D_{ist} + \eta_{ist}$
- Hopefully at some point in a paper the authors explain that D is defined as $D_i \times Post_t$, but writing D_{it} rather than D_i helps
- The second model could be written in a different way too. Something like:

  $Y_{it} = \gamma_0 + \sum_s State_i + \sum_t Year_t + \gamma D_{it} + \eta_{it}$ or:

  $Y_{it} = \sum_s State_i + \sum_t Year_t + \gamma D_{it} + \eta_{it}$

Going back to basics

- You can put all dummies in a model as long as you don't estimate a constant

```
* Generate dummies by state
qui tab stateid, gen(st)

reg fte st1 st2 i.time treatp, robust noconstant

Linear regression                               Number of obs     =        801
                                                F(4, 797)         =     783.71
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7937
                                                Root MSE          =      9.003

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         st1 |   17.06518   .4838665    35.27   0.000     16.11538    18.01499
         st2 |   19.94872   1.317281    15.14   0.000     17.36297    22.53447
      2.time |   -2.40651   1.594091    -1.51   0.132    -5.535623    .7226031
      treatp |   2.913982   1.736818     1.68   0.094    -.4952963    6.323261
------------------------------------------------------------------------------
```

Extensions

- **Estimating models with many fixed effects** (too many dummies): You'll see next semester that fixed-effects models in longitudinal data estimate "within" type estimators without having to estimate many dummy coefficients
- In this class, we just use the -reg- command. But a better alternative with many dummies is the -xtreg- command that estimates random- and fixed-effects models
- **Nonlinear models**: Logit/Probit, Poisson. Some tricky problems estimating fixed effects, and some vexing identification issues
- **Synthetic controls**: Essentially, chose controls so pre-trends are similar using a weighted set of controls
- **DiD with inverse propensity score weights:** Use IPW to make groups more equivalent – same issues about overlap, doubly robust
- **Combining DiD and RDD**: more comparable groups
- **Nonparametric estimation**: Rather than making parametric assumptions and assume homogeneity, we cold use more flexible models