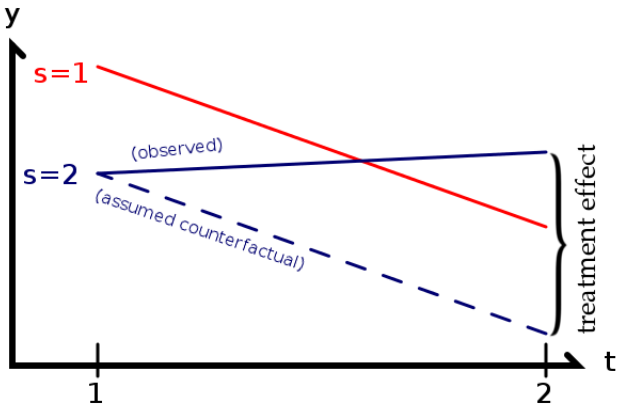
Causality 2: Difference-in-Difference (DID) Estimation

In Essence

- DID estimators are widely used in policy evaluations. This is a method of the natural experiments.
- Assume we have the following 4 observations:

	Treatment Group	Control Group
Before (Time 0)	$\bar{Y}_b^t$	$\bar{Y}_b^c$
After (Time 1)	$\bar{Y}_a^t$	$\bar{Y}_a^c$

DID $= (\bar{Y}_1^t - \bar{Y}_0^t) - (\bar{Y}_1^c - \bar{Y}_0^c)$



- DID estimation is basically a simple interaction effect. Mathematically and statistically, this is a very simple method.

$$y = \alpha + \beta T + \gamma G + \delta(T \times G) + e$$

where T is a time variable and G is the dummy of the treated group. DID estimator is δ.

- Assumption of DID
  - In the absence of the treatment, individual i's outcome at time t is given by:

$$y = \alpha + \beta T + \gamma G + e$$

  - There are two implicit identifying assumptions here:
    1. Time trend (T) is the same for treatment and control groups. This is the common trend assumption.
    2. Selection bias is related to fixed characteristics of the group (G). The magnitude of the selection bias term is not changing over time.
- What is not-so-easy is to acquire a right data with a good treatment variable.

Required Data

- Two+ units and 2+ observations over time.
- Policy change between two periods.

Card and Krueger 1994 Paper

- DID estimate is simply an interaction effect between `treated` and `t`.
- In the following estimate: `fte` refers to full-time equivalent employment; `treated` indicate the location (NJ increased the minimum wage); `t` is a time variable which controls for the general time trend.

```
. reg fte i.treated##i.t bk kfc roys

      Source |       SS           df       MS      Number of obs   =       780
-------------+----------------------------------   F(6, 773)       =     30.38
       Model |  12294.9359         6  2049.15599   Prob > F        =    0.0000
    Residual |  52144.1226       773   67.456821   R-squared       =    0.1908
-------------+----------------------------------   Adj R-squared   =    0.1845
       Total |  64439.0586       779  82.7202292   Root MSE        =    8.2132

------------------------------------------------------------------------------
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |
          NJ |  -2.339362   1.051575    -2.22   0.026    -4.403643    -.275082
         1.t |  -2.490132   1.332359    -1.87   0.062    -5.105602    .1253385
             |
   treated#t |
       NJ#1  |   2.939176   1.484871     1.98   0.048     .0243192    5.854033
             |
          bk |   .8497423   .9254713     0.92   0.359    -.9669927    2.666477
         kfc |  -9.331172   1.037195    -9.00   0.000    -11.36723   -7.295118
        roys |  -1.053964   1.003224    -1.05   0.294    -3.023332    .9154027
       _cons |   21.34211   1.185065    18.01   0.000     19.01578    23.66844
------------------------------------------------------------------------------
```

- You can use an user-written program, "`diff`".
- Install "`diff`" in your machine: `ssc install diff`

```
. diff fte, t(treated) p(t) cov(bk kfc roys) report

DIFFERENCE-IN-DIFFERENCES WITH COVARIATES

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS
Number of observations in the DIFF-IN-DIFF: 780
```

```
            Before          After
   Control: 76              76          152
   Treated: 314             314         628
            390             390
Report - Covariates and coefficients:
------------------------------------------------------------------------
 Variable(s)          |   Coeff.    | Std. Err. |    t     |  P>|t|
----------------------+-------------+-----------+----------+----------
bk                    | 0.850       | 0.925     | 0.918    | 0.359
kfc                   | -9.331      | 1.037     | -8.997   | 0.000
roys                  | -1.054      | 1.003     | -1.051   | 0.294
------------------------------------------------------------------------
------------------------------------------------------
 Outcome var.    | fte      | S. Err. |   |t|    |  P>|t|
-----------------+----------+---------+----------+---------
Before           |          |         |          |
   Control       | 21.342   |         |          |
   Treated       | 19.003   |         |          |
   Diff (T-C)    | -2.339   | 1.052   | -2.22    | 0.026**
After            |          |         |          |
   Control       | 18.852   |         |          |
   Treated       | 19.452   |         |          |
   Diff (T-C)    | 0.600    | 1.052   | 0.57     | 0.569
                 |          |         |          |
Diff-in-Diff     | 2.939    | 1.485   | 1.98     | 0.048**
------------------------------------------------------
R-square:    0.19
* Means and Standard Errors are estimated by linear regression
**Inference: *** p<0.01; ** p<0.05; * p<0.1
```