Difference in Differences

Christopher Taber

Department of Economics
University of Wisconsin-Madison

February 1, 2012

Difference Model

Lets think about a simple evaluation of a policy.

If we have data on a bunch of people right before the policy is enacted and on the same group of people after it is enacted we can try to identify the effect.

Suppose we have two years of data 0 and 1 and that the policy is enacted in between

We could try to identify the effect by simply looking at before and after the policy

That is we can identify the effect as

$$\bar{Y}_1 - \bar{Y}_0$$

We could formally justify this with a fixed effects model.

Let

$$Y_{it} = \beta_0 + \alpha T_{it} + \theta_i + u_{it}$$

We have in mind that

$$T_{it} = \begin{cases} 0 & t = 0 \\ 1 & t = 1 \end{cases}$$

We will also assume that u_{it} is orthogonal to the other stuff

We don't need to make any assumptions about θ_i

Background on Fixed effect.

Lets forget about the basic problem and review fixed effect more generally

Assume that we have T_i observations for each individual numbered $1, ..., T_i$

We write the model as

$$Y_{it} = X_{it}\beta + \theta_i + u_{it}$$

and assume u_{it} is uncorrelated with other stuff in the model.

For a generic variable Z_{it} define

$$\bar{Z}_i \equiv \frac{1}{T_i} \sum_{i=1}^{N} Z_{it}$$

then notice that

$$\bar{Y}_i = \bar{X}_i' \beta + \theta_i + \bar{u}_i$$

So

$$\left(Y_{it} - \bar{Y}_i\right) = \left(X_{it} - \bar{X}\right)' \beta + (u_{it} - \bar{u}_i)$$

We can get a consistent estimate of β by regressing $\left(Y_{it} - \bar{Y}_i\right)$ on $\left(X_{it} - \bar{X}\right)$.

The key thing is we didn't need to assume anything about the relationship between θ_i and X_i

This is numerically equivalent to putting a bunch of individual fixed effects into the model and then running the regressions

To see why let D_{it} be a $N \times 1$ vector of dummy variables so that for the j^{th}element:

$$D_{it}^{(j)} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

and write the regression model as

$$Y_{it} = X_{it}\widehat{\beta} + D_{it}'\widehat{\delta} + \widehat{u}_{it}$$

It will again be useful to think about this as a partitioned regression

For a generic variable Z_{it}, think about a regression of Z_{it} onto D_{it}

Abusing notation somewhat, the least squares estimator for this is

$$\widehat{\delta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T_i} D_{it} D_{it}' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T_i} D_{it} Z_{it}$$

- The matrix $\sum_{i=1}^{N} \sum_{t=1}^{T_i} D_{it} D_{it}'$ is an $N \times N$ diagonal matrix with each (i, i) diagonal element equal to T_i.
- The vector $\sum_{i=1}^{N} \sum_{t=1}^{T_i} D_{it} Z_{it}$ is an $N \times 1$ vector with j^{th} element $\sum_{t=1}^{T_i} Z_{it}$
- Thus $\widehat{\delta}$ is an $N \times 1$ vector with generic element $\bar{Z}_i$
- $D_{it}' \widehat{\delta} = \bar{Z}_i$
- Or using notation from the previous lecture notes we can write
$$\widetilde{Z}_{it} = M_D Z_{it} = Z_{it} - \bar{Z}_i$$

Thus we can see that $\widehat{\beta}$ just comes from regressing $\left(Y_{it} - \bar{Y}_i\right)$ on $\left(X_{it} - \bar{X}\right)$ which is exactly what fixed effects is

First Differencing

The other standard way of dealing with fixed effects is to "first difference" the data so we can write

$$Y_{it} - Y_{it-1} = (X_{it} - X_{it-1})' \beta + u_{it} - u_{it-1}$$

Note that with only 2 periods this is equivalent to the standard fixed effect because

$$\begin{aligned} Y_{i2} - \bar{Y}_i &= Y_{i2} - \frac{Y_{i1} + Y_{i2}}{2} \\ &= \frac{Y_{i2} - Y_{i1}}{2} \end{aligned}$$

This is not the same as the regular fixed effect estimator when you have more than two periods

To see that lets think about a simple “treatment effect” model with only the regressor T_{it}.

Assume that we have T periods for everyone, and that also for everyone

$$T_{it} = \begin{cases} 0 & t \leq \tau \\ 1 & t > \tau \end{cases}$$

Think of this as a new national program that begins at period $\tau + 1$

The standard fixed effect estimator is

$$\begin{aligned}\widehat{\alpha}_{FE} =& \frac{scov(\left(T_{it} - \bar{T}_i\right), \left(Y_{it} - \bar{Y}_i\right)}{\left(T_{it} - \bar{T}_i\right)} \\ =& \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(T_{it} - \bar{T}_i\right)\left(Y_{it} - \bar{Y}_i\right)}{\left(\sum_{i=1}^{N} \sum_{t=1}^{T} \left(T_{it} - \bar{T}_i\right)^2\right)}\end{aligned}$$

Let

$$\bar{Y}_A = \frac{1}{N(T-\tau)} \sum_{i=1}^{N} \sum_{t=\tau+1}^{T} Y_{it}$$

$$\bar{Y}_B = \frac{1}{N\tau} \sum_{i=1}^{N} \sum_{t=1}^{\tau} Y_{it}$$

The numerator is

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T}\left(T_{it}-\frac{T-\tau}{T}\right)\left(Y_{it}-\bar{Y}_i\right)\\
&=\sum_{i=1}^{N}\left[\sum_{t=1}^{\tau}\left(T_{it}-\frac{T-\tau}{T}\right)Y_{it}+\sum_{t=\tau+1}^{T}\left(T_{it}-\frac{T-\tau}{T}\right)Y_{it}\right]\\
&=-\tau\left(\frac{T-\tau}{T}\right)N\bar{Y}_B+(T-\tau)\frac{\tau}{T}N\bar{Y}_A\\
&=\tau\left(\frac{T-\tau}{T}\right)N\left[\bar{Y}_A-\bar{Y}_B\right]
\end{aligned}$$

The denominator is

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T}\left(T_{it}-\frac{T-\tau}{T}\right)^2 \\
&\quad=\sum_{i=1}^{N}\left[\sum_{t=1}^{\tau}\left(-\frac{T-\tau}{T}\right)^2+\sum_{t=\tau+1}^{T}\left(1-\frac{T-\tau}{T}\right)^2\right] \\
&\quad=N\left[\tau\frac{T-\tau}{T}\frac{T-\tau}{T}+(T-\tau)\frac{\tau}{T}\frac{\tau}{T}\right] \\
&\quad=N\left[\frac{\tau T^2-2\tau^2T+\tau^3}{T^2}+\frac{T\tau^2-\tau^3}{T^2}\right] \\
&\quad=N\left[\frac{\tau T^2-\tau^2T}{T^2}\right] \\
&\quad=N\tau\left[\frac{T-\tau}{T}\right]
\end{aligned}
$$

So the fixed effects estimator is just

$$\bar{Y}_A - \bar{Y}_B$$

Next consider the first differences estimator

$$\begin{aligned}
&\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(T_{it}-T_{it-1}\right)\left(Y_{it}-Y_{it-1}\right)}{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(T_{it}-T_{it-1}\right)^2} \\
=&\frac{\sum_{i=1}^{N}\left(Y_{i\tau}-Y_{i\tau-1}\right)}{N} \\
=&\bar{Y}_{\tau}-\bar{Y}_{\tau-1}
\end{aligned}$$

Notice that you throw out all the data except right before and after the policy change.

You can also see that these correspond in the two period case

Thus we have shown in the two period model-or multi-period model that the fixed effects estimator is just a difference in means, before and after the policy is implemented

This is sometimes called the “difference model”

The problem is that this attributes any changes in time to the policy

That is suppose something else happened at time τ other than just the program.

We will attribute whatever that is to the program.

If we added time dummy variables into our model we could not separate the time effect from T_{it} (in the case above)

To solve this problem, suppose we have two groups:

- People who are affected by the policy changes ($\diamondsuit$)
- People who are not affected by the policy change ($\clubsuit$)

We can think of using the controls to pick up the time changes:

$$\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0}$$

Then we can estimate our policy effect as a difference in difference:

$$\widehat{\alpha} = \left(\bar{Y}_{\diamondsuit 1} - \bar{Y}_{\diamondsuit 0}\right) - \left(\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0}\right)$$

To put this in a regression model we can write it as

$$Y_{it} = \beta_0 + \alpha T_{it} + \delta t + \theta_i + \varepsilon_{it}$$

Now think about what happens if we run a fixed effect regression in this case

Let $S(i)$ indicate and individual's suit (either $\diamondsuit$ or $\clubsuit$)

Further we will assume that

$$T_{it} = \begin{cases} 0 & S(i) = \clubsuit \\ 0 & S(i) = \diamondsuit, t = 0 \\ 1 & S(i) = \diamondsuit, t = 1 \end{cases}$$

Identification

Lets first think about identification in this case notice that

$$
\begin{aligned}
&\left[E(Y_{i,1} \mid S(i) = \diamondsuit) - E(Y_{i,0} \mid S(i) = \diamondsuit)\right] \\
&- \left[E(Y_{i,1} \mid S(i) = \clubsuit) - E(Y_{i,0} \mid S(i) = \clubsuit)\right] \\
=& [(\beta_0 + \alpha + \delta + E(\theta_i \mid S(i) = \diamondsuit)) - (\beta_0 + E(\theta_i \mid S(i) = \diamondsuit))] \\
&- [(\beta_0 + \delta + E(\theta_1 \mid S(i) = \clubsuit)) - (\beta_0 + E(\theta_i \mid S(i) = \clubsuit))] \\
=&\alpha + \delta \\
&- \delta \\
=&\alpha
\end{aligned}
$$

Fixed Effects Estimation

Doing fixed effects is equivalent to first differencing, so we can write the model as

$$(Y_{i1} - Y_{i0}) = \delta + \alpha\,(T_{i1} - T_{i0}) + (\varepsilon_{i1} - \varepsilon_{i0})$$

Let $N_{\diamondsuit}$ and $N_{\clubsuit}$ denote the number of diamonds and clubs in the data

Note that for $\diamondsuit$s, $T_{i1} - T_{i0} = 1$, but for $\clubsuit$s, $T_{i1} - T_{i0} = 1$

This means that

$$\bar{T}_1 - \bar{T}_0 = \frac{N_{\diamondsuit}}{N_{\diamondsuit} + N_{\clubsuit}}$$

and of course

$$1 - (\bar{T}_1 - \bar{T}_0) = \frac{N_{\clubsuit}}{N_{\diamondsuit} + N_{\clubsuit}}$$

So if we run a regression

$$\widehat{\alpha} = \frac{\sum_{i=1}^{N} \left((T_{i1} - T_{i0}) - (\bar{T}_1 - \bar{T}_0) \right) (Y_{i1} - Y_{i0})}{\sum_{i=1}^{N} \left(T_{i1} - T_{i0} - \bar{T}_1 + \bar{T}_0 \right)^2}$$

$$= \frac{N_{\diamondsuit} \left(\frac{N_{\clubsuit}}{N_{\clubsuit}+N_{\diamondsuit}} \right) \left(\bar{Y}_{\diamondsuit 1} - \bar{Y}_{\diamondsuit 0} \right) - N_{\clubsuit} \frac{N_{\diamondsuit}}{N_{\clubsuit}+N_{\diamondsuit}} \left(\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0} \right)}{N_{\diamondsuit} \left(\frac{N_{\clubsuit}}{N_{\clubsuit}+N_{\diamondsuit}} \right)^2 + N_{\clubsuit} \left(\frac{N_{\diamondsuit}}{N_{\clubsuit}+N_{\diamondsuit}} \right)^2}$$

$$= \frac{\frac{N_{\diamondsuit} N_{\clubsuit}}{N_{\clubsuit}+N_{\diamondsuit}} \left(\bar{Y}_{\diamondsuit 1} - \bar{Y}_{\diamondsuit 0} \right) - \frac{N_{\clubsuit} N_{\diamondsuit}}{N_{\clubsuit}+N_{\diamondsuit}} \left(\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0} \right)}{\frac{N_{\diamondsuit} N_{\clubsuit} (N_{\clubsuit}+N_{\diamondsuit})}{(N_{\clubsuit}+N_{\diamondsuit})^2}}$$

$$= \left(\bar{Y}_{\diamondsuit 1} - \bar{Y}_{\diamondsuit 0} \right) - \left(\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0} \right)$$

Actually you don't need panel data, but could do just fine with repeated cross section data.

In this case we add a dummy variable for being a $\diamondsuit$, let this be $\diamondsuit_i$

Then we can write the regression as

$$Y_{it} = \beta_0 + \alpha T_{it} + \delta t + \gamma \diamondsuit_i + \varepsilon_{it}$$

To show this works, lets work with the GMM equations (or Normal equations)

$$
\begin{aligned}
0 &= \sum_{i=1}^{N} \sum_{t=0}^{1} \widehat{\varepsilon}_{it} \\
&= \sum_{\color{red}\diamondsuit} \widehat{\varepsilon}_{i0} + \sum_{\color{red}\diamondsuit} \widehat{\varepsilon}_{i1} + \sum_{\color{blue}\clubsuit} \widehat{\varepsilon}_{i0} + \sum_{\color{blue}\clubsuit}^{N} \widehat{\varepsilon}_{i1} \\
0 &= \sum_{i=1}^{N} \sum_{t=0}^{1} T_{it} \widehat{\varepsilon}_{it} \\
&= \sum_{\color{red}\diamondsuit} \widehat{\varepsilon}_{i1}
\end{aligned}
$$

$$0 = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=0}^{1} t\widehat{\varepsilon}_{it}$$

$$= \sum_{\diamondsuit} \widehat{\varepsilon}_{i1} + \sum_{\clubsuit}^{N} \widehat{\varepsilon}_{i1}$$

$$0 = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=0}^{1} \diamondsuit_i \widehat{\varepsilon}_{it}$$

$$= \sum_{\diamondsuit} \widehat{\varepsilon}_{i0} + \sum_{\diamondsuit} \widehat{\varepsilon}_{i1}$$

We can rewrite these equations as

$$0 = \sum_{\diamondsuit} \widehat{\varepsilon}_{i0}$$

$$0 = \sum_{\diamondsuit} \widehat{\varepsilon}_{i1}$$

$$0 = \sum_{\clubsuit} \widehat{\varepsilon}_{i0}$$

$$0 = \sum_{\clubsuit}^{N} \widehat{\varepsilon}_{i1}$$

Which we can write as

$$\bar{Y}_{\diamondsuit 0} = \widehat{\beta}_0 + \widehat{\gamma}$$
$$\bar{Y}_{\diamondsuit 1} = \widehat{\beta}_0 + \widehat{\alpha} + \widehat{\delta} + \widehat{\gamma}$$
$$\bar{Y}_{\clubsuit 0} = \widehat{\beta}_0$$
$$\bar{Y}_{\clubsuit 1} = \widehat{\beta}_0 + \widehat{\delta}$$

We can solve for the parameters as

$$
\begin{aligned}
\widehat{\beta}_0 &= \bar{Y}_{\clubsuit 0} \\
\widehat{\gamma} &= \bar{Y}_{\diamondsuit 0} - \bar{Y}_{\clubsuit 0} \\
\widehat{\delta} &= \bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0} \\
\widehat{\alpha} &= \bar{Y}_{\diamondsuit 1} - \bar{Y}_{\clubsuit 0} - \left(\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0}\right) - \left(\bar{Y}_{\diamondsuit 0} - \bar{Y}_{\clubsuit 0}\right) \\
&= \left(\bar{Y}_{\diamondsuit 1} - \bar{Y}_{\diamondsuit 0}\right) - \left(\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0}\right)
\end{aligned}
$$

Now more generally we can think of “difference in differences” as

$$Y_{it} = \beta_0 + \alpha T_{it} + \delta_t + \theta_{g(i)} + \varepsilon_{it}$$

where $g(i)$ is the individual’s group

(I like to separate the underlying econometric model from the way in which we estimate it)

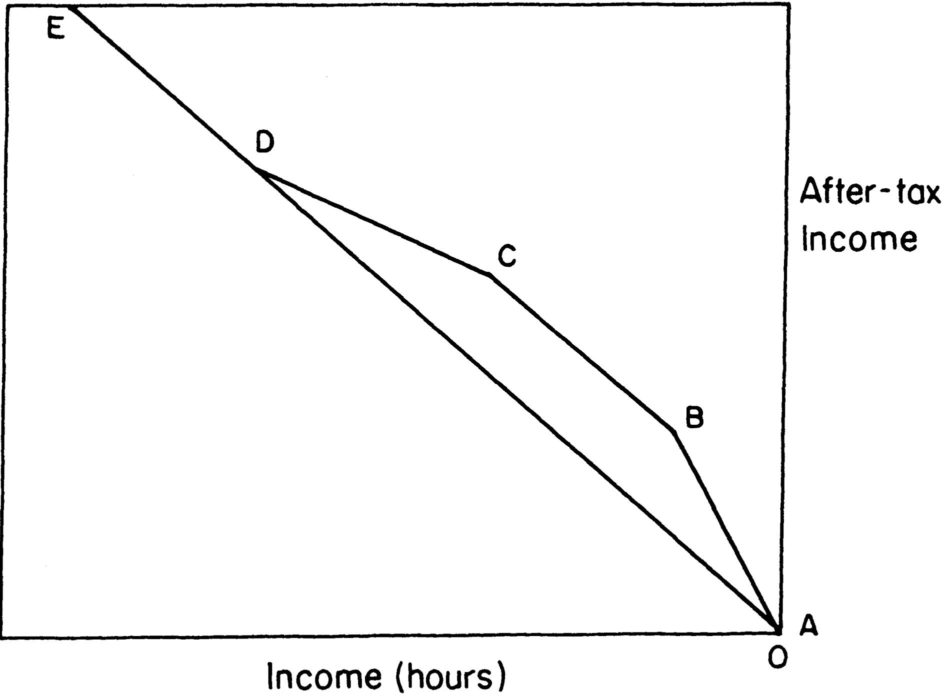
There are many papers that do this basic sort of thing

Eissa and Liebman "Labor Supply Response to the Earned Income Tax Credit" (QJE, 1996)

They want to estimate the effect of the earned income tax credit on labor supply of women

The EITC is a subsidy that goes mostly to low income women who have children

It looks something like this:


E
D
C
B
A
O
After-tax
Income
Income (hours)

Eissa and Liebman evaluate the effect of the effect on EITC from the Tax Reform Act of 1986.

At that time only people with children were eligible

They use:

- For Treatments: Single women with kids
- For Controls: Single women without kids

They look before and after the EITC

Here is the simple model

TABLE II

LABOR FORCE PARTICIPATION RATES OF UNMARRIED WOMEN

	Pre-TRA86 (1)	Post-TRA86 (2)	Difference (3)	Difference-in-differences (4)
A. *Treatment group:*				
With children [20,810]	0.729 (0.004)	0.753 (0.004)	0.024 (0.006)	
Control group:				
Without children [46,287]	0.952 (0.001)	0.952 (0.001)	0.000 (0.002)	*0.024 (0.006)*
B. *Treatment group:*				
Less than high school, with children [5396]	0.479 (0.010)	0.497 (0.010)	0.018 (0.014)	
Control group 1:				
Less than high school, without children [3958]	0.784 (0.010)	0.761 (0.009)	−0.023 (0.013)	*0.041 (0.019)*
Control group 2:				
Beyond high school, with children [5712]	0.911 (0.005)	0.920 (0.005)	0.009 (0.007)	*0.009 (0.015)*
C. *Treatment group:*				
High school, with children [9702]	0.764 (0.006)	0.787 (0.006)	0.023 (0.008)	
Control group 1:				
High school, without children [16,527]	0.945 (0.002)	0.943 (0.003)	−0.002 (0.004)	*0.025 (0.009)*
Control group 2:				
Beyond high school, with children [5712]	0.911 (0.005)	0.920 (0.005)	0.009 (0.007)	*0.014 (0.011)*

Data are from the March CPS, 1985–1987 and 1989–1991. Pre-TRA86 years are 1984–1986. Post-TRA86 years are 1988–1990. Labor force participation equals one if annual hours are positive, zero otherwise. Standard errors are in parentheses. Sample sizes are in square brackets. Means are weighted with CPS March supplement weights.

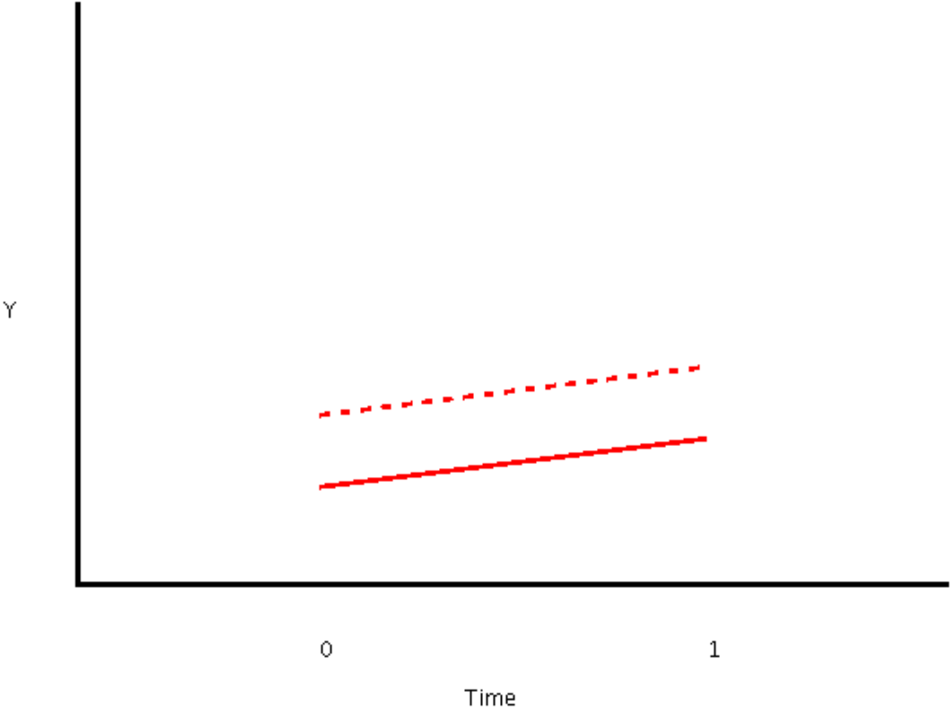
Note that this is nice and suggests it really is a true effect

As an alternative suppose the data showed

	Treatment	Control
Before	1.0	1.5
After	1.1	1.6

This would give a difference in difference estimate of 0.

However how do we know what the right metric is?


Y
0
1
Time

Take logs and you get

	Treatment	Control
Before	0.00	0.41
After	0.10	0.47

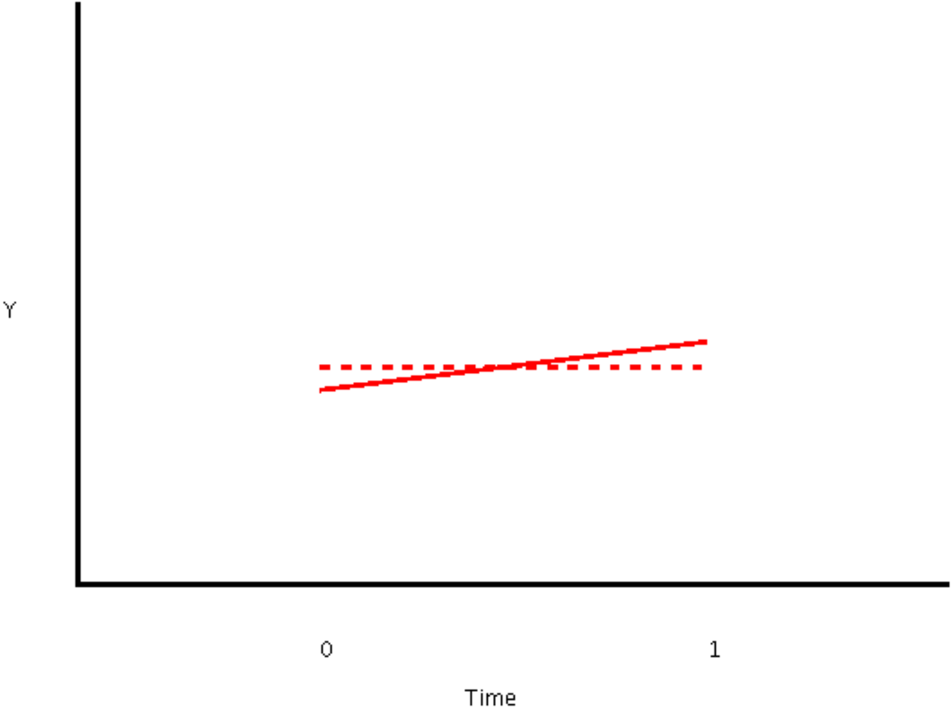
This gives diff-in-diff estimate of 0.04

But you could also take exponentials

	Treatment	Control
Before	2.71	4.48
After	3.00	4.95

This gives a diff-in-diff estimate of -0.18

However if the model looks like this, we have much stronger evidence of an effect


Y
0
1
Time

Eissa and Liebman estimate the model as a probit

$$Prob(Y_{it} = 1) = \Phi\left(\beta_0 + \alpha T_{it} + X_{it}'\beta + \delta_t + \theta_{g(i)}\right)$$

They also look at the effect of the EITC on hours of work

TABLE III

PROBIT RESULTS: CHILDREN VERSUS NO CHILDREN ALL UNMARRIED WOMEN

Variables	Sample: all unmarried women					
	Without covariates (1)	Demographic characteristics (2)	Unemployment and AFDC (3)	State dummies (4)	Second child dummy (5)	Separate year interactions (6)
Coefficient estimates						
Other income (1000s)	—	−0.035 (.001)	−0.034 (.001)	−0.034 (.001)	−0.034 (.001)	−0.039 (.001)
Number of preschool children	—	−0.395 (.016)	−0.279 (.018)	−0.281 (.018)	−0.278 (.018)	−0.279 (.018)
Nonwhite	—	−0.422 (.016)	−0.521 (.030)	−0.520 (.031)	−0.518 (.031)	−0.518 (.031)
Age	—	−0.237 (.059)	−0.209 (.060)	−0.195 (.060)	−0.194 (.060)	−0.193 (.060)
Age squared	—	0.007 (.002)	0.006 (.002)	0.006 (.002)	0.006 (.002)	0.006 (.002)
Education	—	−0.020 (.014)	−0.029 (.014)	−0.029 (.014)	−0.029 (.014)	−0.029 (.014)
Education squared	—	0.010 (.001)	0.010 (.001)	0.010 (.001)	0.010 (.001)	0.010 (.001)
Second child	—	—	—	—	−0.118 (.040)	−0.117 (.040)
State Unemployment rate	—	—	−0.096 (.007)	−0.063 (.012)	−0.064 (.012)	−0.064 (.012)
State Unemployment rate kids × kids	—	—	0.028 (.010)	0.029 (.010)	0.029 (.010)	0.030 (.010)
Maximum monthly AFDC benefit	—	—	−0.001 (.000)	−0.001 (.000)	−0.001 (.001)	−0.001 (.000)

Kids (γ_0)	−1.053 (.020)	−0.250 (.029)	−1.403 (.106)	−1.438 (.108)	−1.458 (.110)	−1.462 (.110)
Post86 (γ_1)	−0.001 (.028)	0.019 (.031)	−0.152 (.067)	−0.104 (.069)	−0.094 (.069)	
Kids × Post86 (γ_2)	*0.069 (.027)*	*0.074 (.030)*	*0.103 (.037)*	*0.113 (.037)*	*0.087 (.043)*	—
Kids × 1988						*0.033 (.057)*
Kids × 1989						*0.116 (.058)*
Kids × 1990						*0.112 (.057)*
Second child × post86					0.051 (.043)	—
Log likelihood	−20759	−17105	−16793	−16633	−16629	−16626
Predicted participation response for treatment group		*.019 (.008)*	*.026 (.010)*	*.028 (.009)*	*.022 (.009)*	*.008, .029, .028 (.014), (.015), (.015)*

Data are from survey years 1985–1987 and 1898–1991 of the March CPS. The dependent variable is labor force participation. It equals one if the woman worked at least one hour during the tax year. *Post86* equals one for tax years 1988, 1989, 1990. *Kids* equals one if the tax filing unit contained at least one child. In addition to the variables shown, all regressions include year dummies for 1984, 1985, 1989, and 1990. Columns (2) through (6) also include variables for the number of children in the tax filing unit age-cubed. Columns (3) through (6) also include interactions of *age* and *nonwhite* with *post86* and with *kids*. Columns (4) through (6) also include a full set of state dummies. Column (6) also includes interactions of *second child* with the year dummies for 1988, 1989, and 1990. The number of observations is 67, 097. Standard errors are in parentheses. Regressions are weighted with CPS March supplement weights.

TABLE V

HOURS AND WEEKS REGRESSIONS: CHILDREN VERSUS NO CHILDREN

Dependent variable:	Annual hours	Annual hours	Annual hours	Annual hours	Annual weeks	Annual weeks
Variables	All single women with hours > 0 (1)	Less than high school with hours > 0 (2)	All single women (3)	Less than high school (4)	All single women with hours > 0 (5)	All single women (6)
Coefficient estimates						
Other income (1000s)	−21.83 (.61)	−26.81 (2.93)	−29.92 (.62)	−56.65 (2.46)	−0.433 (.012)	−0.670 (.014)
Number of preschool children	−66.28 (10.42)	−72.21 (25.57)	−136.49 (9.18)	−107.94 (16.92)	−1.833 (.214)	−3.944 (.207)
Nonwhite	−140.94 (11.77)	−142.84 (41.29)	−209.80 (12.43)	−266.32 (36.14)	−2.680 (.241)	−4.788 (.281)
Age	786.82 (22.38)	475.01 (64.29)	576.16 (23.59)	211.04 (54.87)	13.743 (.459)	9.391 (.533)
Age squared	−21.45 (.75)	−12.62 (2.21)	−15.12 (.80)	−4.79 (1.89)	−0.385 (.015)	−0.252 (.018)
Education	56.69 (6.41)	14.22 (17.07)	114.90 (6.14)	−56.03 (15.03)	1.262 (.132)	3.086 (.139)
Education squared	−1.58 (.25)	−0.21 (1.22)	−2.22 (.24)	5.97 (1.05)	−0.041 (.005)	−0.068 (.006)
Unemployment rate	−9.98 (3.85)	−31.37 (14.58)	−15.94 (4.15)	−42.24 (13.00)	−0.130 (.079)	−0.304 (.094)
Unemployment rate × kids	5.27 (4.17)	33.60 (13.44)	1.33 (4.14)	34.40 (11.10)	0.054 (.086)	−.065 (.094)
Maximum monthly AFDC benefit	−0.22 (.06)	−0.10 (.18)	−0.54 (.06)	−0.14 (.14)	−0.005 (.001)	−.014 (.001)
Kids (γ_0)	−83.03 (47.82)	−249.44 (132.61)	−186.48 (46.65)	−327.07 (110.24)	−6.856 (.981)	−11.420 (1.054)
Post86 (γ_1)	−29.95 (23.61)	63.27 (78.03)	−45.33 (25.20)	−56.27 (69.26)	0.722 (.484)	0.222 (.569)
Kids × Post86 (γ_2)	*25.22 (15.18)*	*2.98 (46.04)*	*37.37 (15.31)*	*83.83 (39.42)*	*.126 (.311)*	*.560 (.346)*
Observations	59,474	5700	67,097	9354	59,474	67,097

Data are from survey years 1985–1987 and 1989–1991 of the March CPS. *Post86* equals one for tax years 1988, 1989, and 1990. *Kids* equals one if the tax filing unit contained at least one child. In addition to the variables shown, all regressions include year dummies for 1984, 1985, 1989, and 1990; variables for the number of children in the tax filing unit; age-cubed; interactions of *age* and *nonwhite* with *post86* and with *kids*; and a full set of state dummies. Standard errors are in parentheses. Regressions are weighted with CPS March supplement weights.

Donahue and Levitt “The Impact of Legalized Abortion on Crime” (QJE, 2001)

This was a paper that got a huge amount of attention in the press at the time

They show (or claim to show) that there was a large effect of abortion on crime rates

The story is that the children who were not born as a result of the legalization were more likely to become criminals

This could be either because of the types of families they were likely to be born to, or because there was differential timing of birth

Identification comes because 5 states legalized abortion prior to Roe v. Wade (around 1970): New York, Alaska, Hawaii, Washington, and California

In 1973 the supreme court legalized abortion with Roe v. Wade

What makes this complicated is that newborns very rarely commit crimes

They need to match the timing of abortion with the age that kids are likely to commence their criminal behavior

They use the concept of effective abortion which for state j at time t is

$$EffectiveAbortion_{jt} = \sum_{a} Abortionlegal_{it-1} \left(\frac{Arrests_{a}}{Arrests_{total}} \right)$$

The model is then estimated using difference in differences:

$$log(Crime_{jt}) = \beta_1 EffectiveAbortion_{jt} + X'_{jt}\Theta + \gamma_j + \lambda_t + \varepsilon_{jt}$$

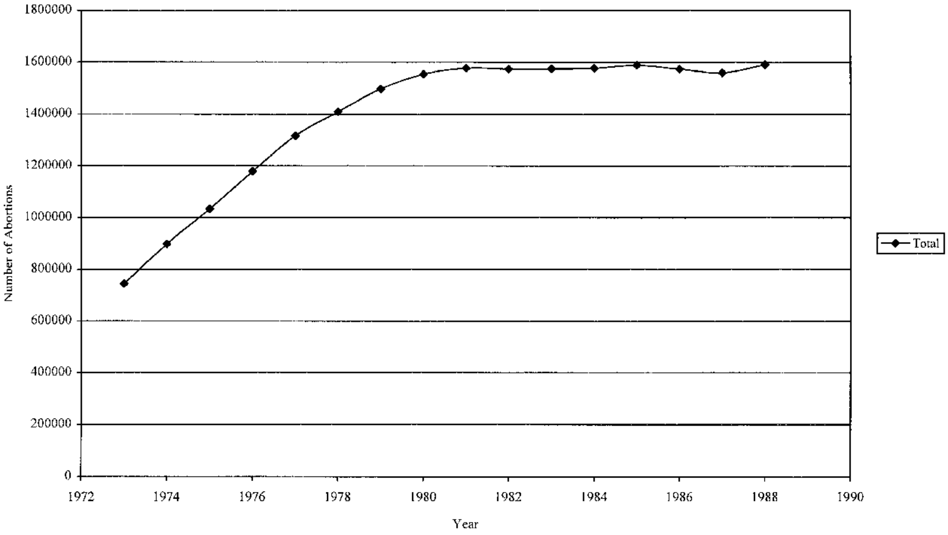


FIGURE I
Total Abortions by Year
Source: Alan Guttmacher Institute [1992].

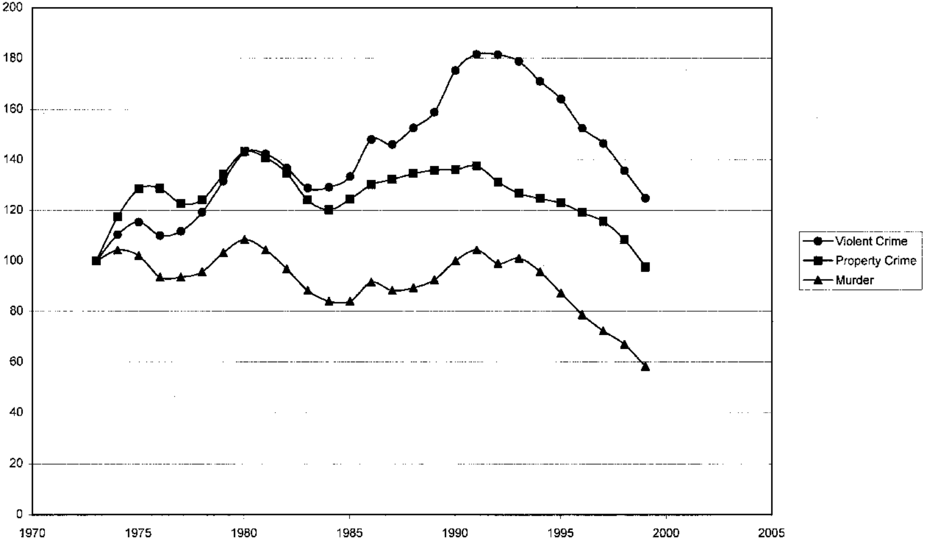


FIGURE II
Crime Rates from the Uniform Crime Reports, 1973–1999

Data are national aggregate per capita reported violent crime, property crime, and murder, indexed to equal 100 in the year 1973. All data are from the FBI's *Uniform Crime Reports,* published annually.

TABLE I

CRIME TRENDS FOR STATES LEGALIZING ABORTION EARLY VERSUS THE REST OF THE UNITED STATES

Crime category	Percent change in crime rate over the period				Cumulative,
	1976–1982	1982–1985	1988–1994	1994–1997	1982–1997
Violent crime					
Early legalizers	16.6	11.1	1.9	−25.8	−12.8
Rest of U. S.	20.9	13.2	15.4	−11.0	17.6
Difference	−4.3	−2.1	−13.4	−14.8	−30.4
	(5.5)	(5.4)	(4.4)	(3.3)	(8.1)
Property crime					
Early legalizers	1.7	−8.3	−14.3	−21.5	−44.1
Rest of U. S.	6.0	1.5	−5.9	−4.3	−8.8
Difference	−4.3	−9.8	−8.4	−17.2	−35.3
	(2.9)	(4.0)	(4.2)	(2.4)	(5.8)
Murder					
Early legalizers	6.3	0.5	2.7	−44.0	−40.8
Rest of U. S.	1.7	−8.8	5.2	−21.1	−24.6
Difference	4.6	9.3	−2.5	−22.9	−16.2
	(7.4)	(6.8)	(8.6)	(6.8)	(10.7)
Effective abortion rate at end of period					
Early legalizers	0.0	64.0	238.6	327.0	327.0
Rest of U. S.	0.0	10.4	87.7	141.0	141.0
Difference	0.0	53.6	150.9	186.0	186.0

a

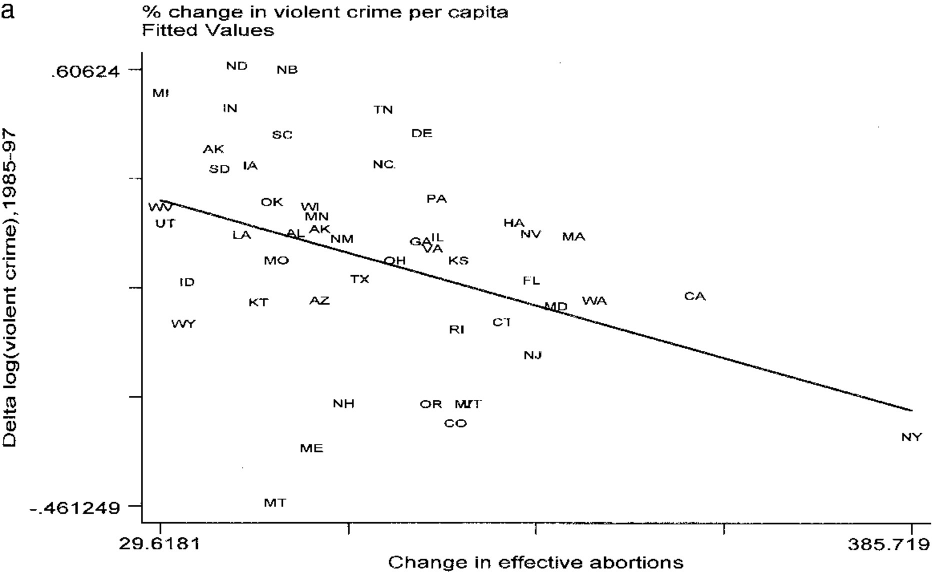

% change in violent crime per capita
Fitted Values
.60624
-.461249
Delta log(violent crime),1985-97
29.6181
385.719
Change in effective abortions
ND
NB
MI
IN
TN
SC
DE
AK
SD
IA
NC
PA
WV
OK
WI
MN
UT
AK
AL
LA
NM
HA
NV
MA
GA
IL
VA
MO
OH
KS
ID
TX
FL
KT
AZ
MD
WA
CA
WY
CT
RI
NJ
NH
OR
MT
CO
NY
ME
MT

TABLE IV

PANEL-DATA ESTIMATES OF THE RELATIONSHIP BETWEEN ABORTION RATES AND CRIME

	ln(Violent crime per capita)		ln(Property crime per capita)		ln(Murder per capita)	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
"Effective" abortion rate (× 100)	−.137 (.023)	−.129 (.024)	−.095 (.018)	−.091 (.018)	−.108 (.036)	−.121 (.047)
ln(prisoners per capita) ($t-1$)	—	−.027 (.044)	—	−.159 (.036)	—	−.231 (.080)
ln(police per capita) ($t-1$)	—	−.028 (.045)	—	−.049 (.045)	—	−.300 (.109)
State unemployment rate (percent unemployed)	—	.069 (.505)	—	1.310 (.389)	—	.968 (.794)
ln(state income per capita)	—	.049 (.213)	—	.084 (.162)	—	−.098 (.465)
Poverty rate (percent below poverty line)	—	−.000 (.002)	—	−.001 (.001)	—	−.005 (.004)
AFDC generosity ($t-15$) (× 1000)	—	.008 (.005)	—	.002 (.004)	—	−.000 (.000)
Shall-issue concealed weapons law	—	−.004 (.012)	—	.039 (.011)	—	−.015 (.032)
Beer consumption per capita (gallons)	—	.004 (.003)	—	.004 (.003)	—	.006 (.008)
R^2	.938	.942	.990	.992	.914	.918

Dynarski "The New Merit Aid", in *College Choices: The Economics of Where to Go, When to Go, and How to Pay for it*, 2002

(http://ideas.repec.org/p/ecl/harjfk/rwp04-009.html)

In relatively recent years many states have implemented Merit Aid programs

In general these award scholarships to people who go to school in state and maintain good grades in high school

Here is a summary

Table 2.1 Merit Aid Program Characteristics, 2003

State	Start	Eligibility	Award (in-state attendance only, exceptions noted)
Arkansas	1991	initial: 2.5 GPA in HS core and 19 ACT renew: 2.75 college GPA	public: $2,500 private: same
Florida	1997	initial: 3.0–3.5 HS GPA and 970–1270 SAT/20–28 ACT renew: 2.75–3.0 college GPA	public: 75–100% tuition/fees[a] private: 75–100% average public tuition/fees[a]
Georgia	1993	initial: 3.0 HS GPA renew: 3.0 college GPA	public: tuition/fees private: $3,000
Kentucky	1999	initial: 2.5 HS GPA renew: 2.5–3.0 college GPA	public: $500–3,000[a] private: same
Louisiana	1998	initial: 2.5–3.5 HS GPA and ACT > state mean renew: 2.3 college GPA	public: tuition/fees + $400–800[a] private: average public tuition/fees[a]
Maryland	2002	initial: 3.0 HS GPA in core renew: 3.0 college GPA	2-year school: $1,000 4-year school: $3,000
Michigan	2000	initial: level 2 of MEAP or 75th percentile of SAT/ACT renew: NA	in-state: $2,500 once out-of-state: $1,000 once
Mississippi	1996	initial: 2.5 GPA and 15 ACT renew: 2.5 college GPA	public freshman/sophomore: $500 public junior/senior: $1,000 private: same
Nevada	2000	initial: 3.0 GPA and pass Nevada HS exam renew: 2.0 college GPA	public 4-year: tuition/fees (max $2,500) public 2-year: tuition/fees (max $1,900) private: none
New Mexico	1997	initial: 2.5 GPA 1st semester of college renew: 2.5 college GPA	public: tuition/fees private: none
South Carolina	1998	initial: 3.0 GPA and 1100 SAT/24 ACT renew: 3.0 college GPA	2-year school: $1,000 4-year school: $2,000
Tennessee	2003	initial: 3.0–3.75 GPA and 890–1280 SAT/19–29 ACT renew: 3.0 college GPA	2-year school: tuition/fees ($1,500–2,500)[a] 4-year school: tuition/fees ($3,000–4,000)[a]
West Virginia	2002	initial: 3.0 HS GPA in core and 1000 SAT/21 ACT renew: 2.75–3.0 college GPA	public: tuition/fees private: average public tuition/fees

Note: HS = high school.

[a]Amount of award rises with GPA and/or test score.

Dynarski first looks at the Georgia Hope program (which is probably the most famous)

Her goal is to estimate the effect of this on college enrollment in Georgia

$$y_{iast} = \beta_0 + \beta_1 Hope_{st} + \delta_s + \delta_t + \delta_a + \varepsilon_{iast}$$

where i is an individual, a is age, s is state, and t is time

Table 2.2 **Estimated Effect of Georgia HOPE Scholarship on College Attendance of Eighteen-to-Nineteen-Year-Olds (Southern Census region)**

	(1)	(2)	(3)	(4)
HOPE Scholarship	.086	.085	.085	.069
	(.008)	(.013)	(.013)	(.019)
Merit program in border state			–.005	–.006
			(.013)	(.013)
State and year effects	Y	Y	Y	Y
Median family income		Y	Y	Y
Unemployment rate		Y	Y	Y
Interactions of year effects with black, metro, Hispanic		Y	Y	Y
Time trends				Y
R^2	.020	.059	.059	.056
No. of observations	8,999	8,999	8,999	8,999

Notes: Regressions are weighted by CPS sample weights. Standard errors (in parentheses) are adjusted for heteroskedasticity and correlation within state cells. Sample consists of eighteen-to-nineteen-year-olds in Southern Census region, excluding states (other than Georgia) that introduce merit programs by 2000. See table 2.1 for a list of these states.

She then looks at the broader set of Merit Programs

Table 2.5 **Effect of All Southern Merit Programs on College Attendance of Eighteen-to-Nineteen-Year-Olds**

	All Southern States (N = 13,965)			Southern Merit States Only (N = 5,640)		
	(1)	(2)	(3)	(4)	(5)	(6)
Merit program	.047			.052		
	(.011)			(.018)		
Merit program, Arkansas		.048			.016	
		(.015)			(.014)	
Merit program, Florida		.030			.063	
		(.014)			(.031)	
Merit program, Georgia		.074			.068	
		(.010)			(.014)	
Merit program, Kentucky		.073			.063	
		(.025)			(.047)	
Merit program, Louisiana		.060			.058	
		(.012)			(.022)	
Merit program, Mississippi		.049			.022	
		(.014)			(.018)	
Merit program, South Carolina		.044			.014	
		(.013)			(.023)	
Merit program, year 1			.024			.051
			(.019)			(.027)
Merit program, year 2			.010			.043
			(.032)			(.024)
Merit program, year 3 and after			.060			.098
			(.030)			(.039)
State time trends			Y			Y
R^2	.046	.046	.047	.035	.036	.036

Notes: Specification is that of column (3) in table 2.2, with the addition of state time trends where noted. Sample consists of eighteen-to-nineteen-year-olds in Southern Census region, with the last three columns excluding states that have not introduced a merit program by 2000. Standard errors in parentheses.

Table 2.6 Effect of All Southern Merit Programs on Schooling Decisions of Eighteen-to-Nineteen-Year-Olds (all Southern states; N = 13,965)

	College Attendance (1)	2-Year Public (2)	2-Year Private (3)	4-Year Public (4)	4-Year Private (5)
No time trends					
Merit program	.047	–.010	.004	.044	.005
	(.011)	(.008)	(.004)	(.014)	(.009)
R^2	.046	.030	.007	.030	.020
State time trends					
Merit program, year 1	.024	–.025	.009	.034	.010
	(.019)	(.012)	(.005)	(.012)	(.007)
Merit program, year 2	.010	–.015	.002	.028	–.001
	(.032)	(.018)	(.003)	(.035)	(.011)
Merit program, year 3	.060	–.037	.005	.065	.022
and after	(.030)	(.013)	(.003)	(.024)	(.010)
R^2	.047	.031	.009	.032	.022

Notes: Specification is that of column (3) in table 2.2, with the addition of state time trends where noted. Sample consists of eighteen-to-nineteen-year-olds in Southern Census region. Estimates are similar but less precise when sample is limited to Southern merit states. Standard errors in parentheses.

Event Studies

We have assumed that a treatment here is a static object

Suddenly you don't have a program, then you implement it, then you look at the effects

One might think that some programs take a while to get going so you might not see effects immediately

Others initial effects might be large and then go away

In general there are many other reasons as well why short run effects may differ from long run effects

The merit aid studies is a nice example they do two things:

- Provide a subsidy for people who have good grades to go to college
- Provide an incentive for students in high school to get good grades (and perhaps then go on to college)

The second will not operate in the short run as long as high school students didn't anticipate the program

Analyzing this is actually quite easy. It is just a matter of redefining the treatment.

In principal you could define the treatment as “being in the first year of a merit program" and throw out treatments beyond the second year

You could then define "being in the second year of a merit program" and throw out other treatments

etc.

It is better to combine them in one regression. You could just run the regression

$$Y_{it} = \beta_0 + \alpha_1 T_{it}^1 + \alpha_2 T_{it}^2 + \alpha_3 T_{it}^3 + \theta_i + \varepsilon_{it}$$

Dynarski does this as well

Table 2.5 Effect of All Southern Merit Programs on College Attendance of Eighteen-to-Nineteen-Year-Olds

	All Southern States (N = 13,965)			Southern Merit States Only (N = 5,640)		
	(1)	(2)	(3)	(4)	(5)	(6)
Merit program	.047			.052		
	(.011)			(.018)		
Merit program, Arkansas		.048			.016	
		(.015)			(.014)	
Merit program, Florida		.030			.063	
		(.014)			(.031)	
Merit program, Georgia		.074			.068	
		(.010)			(.014)	
Merit program, Kentucky		.073			.063	
		(.025)			(.047)	
Merit program, Louisiana		.060			.058	
		(.012)			(.022)	
Merit program, Mississippi		.049			.022	
		(.014)			(.018)	
Merit program, South Carolina		.044			.014	
		(.013)			(.023)	
Merit program, year 1			.024			.051
			(.019)			(.027)
Merit program, year 2			.010			.043
			(.032)			(.024)
Merit program, year 3 and after			.060			.098
			(.030)			(.039)
State time trends			Y			Y
R^2	.046	.046	.047	.035	.036	.036

Notes: Specification is that of column (3) in table 2.2, with the addition of state time trends where noted. Sample consists of eighteen-to-nineteen-year-olds in Southern Census region, with the last three columns excluding states that have not introduced a merit program by 2000. Standard errors in parentheses.

Table 2.6 Effect of All Southern Merit Programs on Schooling Decisions of Eighteen-to-Nineteen-Year-Olds (all Southern states; N = 13,965)

	College Attendance (1)	2-Year Public (2)	2-Year Private (3)	4-Year Public (4)	4-Year Private (5)
No time trends					
Merit program	.047	–.010	.004	.044	.005
	(.011)	(.008)	(.004)	(.014)	(.009)
R^2	.046	.030	.007	.030	.020
State time trends					
Merit program, year 1	.024	–.025	.009	.034	.010
	(.019)	(.012)	(.005)	(.012)	(.007)
Merit program, year 2	.010	–.015	.002	.028	–.001
	(.032)	(.018)	(.003)	(.035)	(.011)
Merit program, year 3	.060	–.037	.005	.065	.022
and after	(.030)	(.013)	(.003)	(.024)	(.010)
R^2	.047	.031	.009	.032	.022

Notes: Specification is that of column (3) in table 2.2, with the addition of state time trends where noted. Sample consists of eighteen-to-nineteen-year-olds in Southern Census region. Estimates are similar but less precise when sample is limited to Southern merit states. Standard errors in parentheses.

Key Assumption

Lets think about the unbiasedness of DD

Going to the original model above we had

$$Y_{it} = \beta_0 + \alpha T_{it} + \delta t + \gamma \diamondsuit_i + \varepsilon_{it}$$

so

$$\begin{aligned}
\widehat{\alpha} &= \left(\bar{Y}_{\diamondsuit 1} - \bar{Y}_{\diamondsuit 0}\right) - \left(\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0}\right) \\
&= \left(\beta_0 + \alpha + \delta + \gamma + \bar{\varepsilon}_{\diamondsuit 1} - \beta_0 - \gamma - \bar{\varepsilon}_{\diamondsuit 0}\right) \\
&\quad - \left(\beta_0 + \delta + \bar{\varepsilon}_{\clubsuit 1} - \beta_0 - \bar{\varepsilon}_{\clubsuit 0}\right) \\
&= \alpha + \left(\bar{\varepsilon}_{\diamondsuit 1} - \bar{\varepsilon}_{\diamondsuit 0}\right) - \left(\bar{\varepsilon}_{\clubsuit 1} - \bar{\varepsilon}_{\clubsuit 0}\right)
\end{aligned}$$

So what you need is

$$E\left[(\bar{\varepsilon}_{\diamondsuit 1} - \bar{\varepsilon}_{\diamondsuit 0}) - (\bar{\varepsilon}_{\clubsuit 1} - \bar{\varepsilon}_{\clubsuit 0})\right] = 0$$

States that change their policy can have different *levels* of the error term

ButMullainathan it must be random in terms of the *change* in the error term

This can be a problem (Ashenfelter's dip is clear example), but generally is not that big a deal as states tend to not operate that quickly

However you might be a bit worried that those states are special

People do two things to adjust for this

Placebo Policies

If a policy was enacted in say 1990 you could pretend it was enacted in 1985 in the same place and then only use data through 1989

This is done occasionally

The easiest (and most common) is in the Event framework: include leads as well as lags in the model

Sort of the basis of Bertrand, Duflo, Mullainathan that I will talk about

Time Trends

This is really common

One might be worried that states that are trending up or trending down are more likely to change policy

One can include group$\times$time dummy variables in the model to fix this problem

Lets go back to the base example but now assume we have three years of data and that the policy is enacted between periods 1 and 2

Our model is now:

$$Y_{it} = \beta_0 + \alpha T_{it} + \delta_{\diamondsuit} t \diamondsuit_i + \delta_{\clubsuit} t(1 - \diamondsuit_i) + \delta_2 1(t = 2) + \gamma \diamondsuit_i + \varepsilon_{it}$$

Notice that this is 6 parameters in 6 unknowns

We can write it as a Difference in difference in difference:

$$
\begin{aligned}
\widehat{\alpha} =& \left(\bar{Y}_{\diamondsuit 2} - \bar{Y}_{\diamondsuit 1}\right) - \left(\bar{Y}_{\clubsuit 2} - \bar{Y}_{\clubsuit 1}\right) \\
& - \left(\bar{Y}_{\diamondsuit 1} - \bar{Y}_{\diamondsuit 0}\right) + \left(\bar{Y}_{\clubsuit 1} - \bar{Y}_{\clubsuit 0}\right) \\
\approx & (\alpha + \delta_{\diamondsuit} + \delta_2) - (\delta_{\clubsuit} + \delta_2) \\
& - (\delta_{\diamondsuit}) + (\delta_{\clubsuit}) \\
=& \alpha
\end{aligned}
$$

So that works

You can also just do fixed effects with time effects

Again it is useful to think about this in terms of a two staged regression

For regular fixed effects you just take the sample mean out of X,T, and Y

For fixed effects with a group trend, for each group you regress X,T, and Y on a time trend with an intercept and take the residuals

This has become a pretty standard thing to do and both Donohue and Levitt and also Dynarski did it

TABLE V

SENSITIVITY OF ABORTION COEFFICIENTS TO ALTERNATIVE SPECIFICATIONS

Specification	Coefficient on the "effective" abortion rate variable when the dependent variable is		
	ln (Violent crime per capita)	ln (Property crime per capita)	ln (Murder per capita)
Baseline	−.129 (.024)	−.091 (.018)	−.121 (.047)
Exclude New York	−.097 (.030)	−.097 (.021)	−.063 (.045)
Exclude California	−.145 (.025)	−.080 (.018)	−.151 (.054)
Exclude District of Columbia	−.149 (.025)	−.112 (.019)	−.159 (.053)
Exclude New York, California, and District of Columbia	−.175 (.035)	−.125 (.017)	−.273 (.052)
Adjust "effective" abortion rate for cross-state mobility	−.148 (.027)	−.099 (.020)	−.140 (.055)
Include control for flow of immigrants	−.115 (.024)	−.063 (.018)	−.103 (.047)
Include state-specific trends	−.078 (.080)	.143 (.033)	−.379 (.105)
Include region-year interactions	−.142 (.033)	−.084 (.023)	−.123 (.053)
Unweighted	−.046 (.029)	−.022 (.023)	.040 (.054)
Unweighted, exclude District of Columbia	−.149 (.029)	−.107 (.015)	−.140 (.055)
Unweighted, exclude District of Columbia, California, and New York	−.157 (.037)	−.110 (.017)	−.166 (.075)
Include control for overall fertility rate ($t - 20$)	−.127 (.025)	−.093 (.019)	−.123 (.047)

Table 2.3 Effect of Georgia HOPE Scholarship on Schooling Decisions (October CPS, 1988–2000; Southern Census region)

	College Attendance (1)	2-Year Public (2)	2-Year Private (3)	4-Year Public (4)	4-Year Private (5)
No time trends					
Hope Scholarship	.085	–.018	.015	.045	.022
	(.013)	(.010)	(.002)	(.015)	(.007)
R^2	.059	.026	.010	.039	.026
Add time trends					
Hope Scholarship	.069	–.055	.014	.084	.028
	(.019)	(.013)	(.004)	(.023)	(.016)
R^2	.056	.026	.010	.029	.026
Mean of dependent variable	.407	.122	.008	.212	.061

Notes: Specification in "No time trends" is that of column (3) in table 2.2. Specification in "Add time trends" adds trends estimated on pretreatment data. In each column, two separate trends are included, one for Georgia and one for the rest of the states. Sample consists of eighteen-to-nineteen-year-olds in Southern Census region, excluding states (other than Georgia) that introduce a merit program by 2000. No. of observations = 8,999. Standard errors in parentheses.

Inference

In most of the cases we were thinking before we had individual data and state variation

Lets think about this in terms of “repeated cross sectional” data so that

$$Y_i = \alpha T_{j(i)t(i)} + Z_i'\delta + X_{j(i)t(i)} + \theta_{j(i)} + \gamma_{t(i)} + u_i$$

Note that one way one could estimate this model would be in two stages:

- Take sample means of everything in the model by j and t
- Using obvious notation one can now write the regression as:

$$\overline{Y}_{jt} = \alpha T_{jt} + \overline{Z}_{jt}'\delta + X_{jt} + \theta_j + \gamma_t + \overline{u}_{jt}$$

- You can run this second regression and get consistent estimates

This is a pretty simple thing to do, but notice it might give very different standard errors

We were acting as if we had a lot more observations than we actually might

Formally the problem is if

$$u_i = \eta_{j(i)t(i)} + \varepsilon_i$$

If we estimate the big model via OLS, we are assuming that u_i is i.i.d.

However, if there is an η_{jt} this is violated

Since it happens at the same level as the variation in T_{jt} it is very important to account for it (Moulton, 1990)

The standard thing is to "cluster" by state$\times$year

Clustering

To review clustering lets avoid all this fixed effect notation and just think that we have G groups and N_j persons in each group.

$$Y_{gi} = X'_{gi}\beta + u_{gi}.$$

Let

$$N^T = \sum_{g=1}^{G} N_g$$

the total number of observations

We get asymptotics from the expression

$$\sqrt{N^T}\left(\widehat{\beta} - \beta\right) \approx \left(\frac{1}{N^T}\sum_{g=1}^{G}\sum_{i=1}^{N_g} X_{gi}X'_{gi}\right)^{-1} \frac{1}{\sqrt{N^t}}\sum_{g=1}^{G}\sum_{i=1}^{N_g} X_{gi}u_{gi}$$

The standard OLS estimate (ignoring degree of freedom corrections) would use:

$$\frac{1}{\sqrt{N^T}}\sum_{g=1}^{G}\sum_{i=1}^{N_g} X_{gi}u_{gi} \approx N(0, E(X_{gi}X'_{gi}u^2_{gi}))$$
$$= N(0, E(X_{gi}X'_{gi})\sigma^2_u)$$

The White heteroskedastic standard errors just use

$$\frac{1}{\sqrt{N^T}}\sum_{g=1}^{G}\sum_{i=1}^{N_g} X_{gi}u_{gi} \approx N(0, E(X_{gi}X'_{gi}u^2_{gi}))$$

And approximate

$$E(X_{gi}X'_{gi}u_{gi}^2) \approx \frac{1}{\sqrt{N^T}} \sum_{g=1}^{G} \sum_{i=1}^{N_g} X_{gi}X'_{gi}\widehat{u}_{gi}^2$$

Clustering uses the approximation:

$$\frac{1}{\sqrt{G}} \sum_{g=1}^{G} \left(\sum_{i=1}^{N_g} X_{gi}u_{gi} \right) \approx N\left(0, E\left[\left(\sum_{i=1}^{N_g} X_{gi}u_{gi} \right) \left(\sum_{i=1}^{N_g} X'_{gi}u_{gi} \right) \right] \right)$$

And we approximate the variance as

$$E\left[\left(\sum_{i=1}^{N_g} X_{gi}u_{gi} \right) \left(\sum_{i=1}^{N_g} X'_{gi}u_{gi} \right) \right] \approx \frac{1}{G} \sum_{g=1}^{G} \left(\sum_{i=1}^{N_g} X_{gi}\widehat{u}_{gi} \right) \left(\sum_{i=1}^{N_g} X'_{gi}\widehat{u}_{gi} \right)$$

Bertrand, Duflo, and Mullainathan “How Much Should we Trust Difference in Differences” (QJE, 2004)

They notice that most (good) studies cluster by state$\times$year

However, this assumes that η_{jt} is iid, but if there is serial correlation in η_{jt} this could be a major problem

TABLE I

SURVEY OF DD PAPERS[A]

Number of DD papers	92	
Number with more than 2 periods of data	69	
Number which collapse data into before-after	4	
Number with potential serial correlation problem	65	
Number with some serial correlation correction	5	
GLS	4	
Arbitrary variance-covariance matrix	1	
Distribution of time span for papers with more than 2 periods	Average	16.5
	Percentile	Value
	1%	3
	5%	3
	10%	4
	25%	5.75
	50%	11
	75%	21.5
	90%	36
	95%	51
	99%	83
Most commonly used dependent variables	Number	
Employment	18	
Wages	13	
Health/medical expenditure	8	
Unemployment	6	
Fertility/teen motherhood	4	
Insurance	4	
Poverty	3	
Consumption/savings	3	
Informal techniques used to assess endogeneity	Number	
Graph dynamics of effect	15	
See if effect is persistent	2	
DDD	11	
Include time trend specific to treated states	7	
Look for effect prior to intervention	3	
Include lagged dependent variable	3	
Number with potential clustering problem	80	
Number which deal with it	36	

TABLE II
DD REJECTION RATES FOR PLACEBO LAWS

A. CPS DATA

Data	$\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$	Modifications	Rejection rate	
			No effect	2% effect
1) CPS micro, log wage			.675 (.027)	.855 (.020)
2) CPS micro, log wage		Cluster at state-year level	.44 (.029)	.74 (.025)
3) CPS agg, log wage	.509, .440, .332		.435 (.029)	.72 (.026)
4) CPS agg, log wage	.509, .440, .332	Sampling w/replacement	.49 (.025)	.663 (.024)
5) CPS agg, log wage	.509, .440, .332	Serially uncorrelated laws	.05 (.011)	.988 (.006)
6) CPS agg, employment	.470, .418, .367		.46 (.025)	.88 (.016)
7) CPS agg, hours worked	.151, .114, .063		.265 (.022)	.280 (.022)
8) CPS agg, changes in log wage	−.046, .032, .002		0	.978 (.007)

B. MONTE CARLO SIMULATIONS WITH SAMPLING FROM AR(1) DISTRIBUTION

Data	ρ	Modifications	Rejection rate	
			No effect	2% effect
9) AR(1)	.8		.373 (.028)	.725 (.026)
10) AR(1)	0		.053 (.013)	.783 (.024)
11) AR(1)	.2		.123 (.019)	.738 (.025)
12) AR(1)	.4		.19 (.023)	.713 (.026)
13) AR(1)	.6		.333 (.027)	.700 (.026)
14) AR(1)	−.4		.008 (.005)	.7 (.026)

They look at a bunch of different ways to deal with problem

TABLE IV
PARAMETRIC SOLUTIONS

Data	Technique	Estimated $\hat{\rho}_1$	Rejection rate: No effect	Rejection rate: 2% Effect
A. CPS DATA				
1) CPS aggregate	OLS		.49 (.025)	.663 (.024)
2) CPS aggregate	Standard AR(1) correction	.381	.24 (.021)	.66 (.024)
3) CPS aggregate	AR(1) correction imposing $\rho = .8$		.18 (.019)	.363 (.024)
B. OTHER DATA GENERATING PROCESSES				
4) AR(1), $\rho = .8$	OLS		.373 (.028)	.765 (.024)
5) AR(1), $\rho = .8$	Standard AR(1) correction	.622	.205 (.023)	.715 (.026)
6) AR(1), $\rho = .8$	AR(1) correction imposing $\rho = .8$		.06 (.023)	.323 (.027)
7) AR(2), $\rho_1 = .55$ $\rho_2 = .35$	Standard AR(1) correction	.444	.305 (.027)	.625 (.028)
8) AR(1) + white noise, $\rho = .95$, noise/signal = .13	Standard AR(1) correction	.301	.385 (.028)	.4 (.028)

TABLE V
BLOCK BOOTSTRAP

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
A. CPS DATA				
1) CPS aggregate	OLS	50	.43 (.025)	.735 (.022)
2) CPS aggregate	Block bootstrap	50	.065 (.013)	.26 (.022)
3) CPS aggregate	OLS	20	.385 (.022)	.595 (.025)
4) CPS aggregate	Block bootstrap	20	.13 (.017)	.19 (.020)
5) CPS aggregate	OLS	10	.385 (.024)	.48 (.024)
6) CPS aggregate	Block bootstrap	10	.225 (.021)	.25 (.022)
7) CPS aggregate	OLS	6	.48 (.025)	.435 (.025)
8) CPS aggregate	Block bootstrap	6	.435 (.022)	.375 (.025)
B. AR(1) DISTRIBUTION				
9) AR(1), $\rho = .8$	OLS	50	.44 (.035)	.70 (.032)
10) AR(1), $\rho = .8$	Block bootstrap	50	.05 (.015)	.25 (.031)

TABLE VI
IGNORING TIME SERIES DATA

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
A. CPS DATA				
1) CPS agg	OLS	50	.49 (.025)	.663 (.024)
2) CPS agg	Simple aggregation	50	.053 (.011)	.163 (.018)
3) CPS agg	Residual aggregation	50	.058 (.011)	.173 (.019)
4) CPS agg, staggered laws	Residual aggregation	50	.048 (.011)	.363 (.024)
5) CPS agg	OLS	20	.39 (.025)	.54 (.025)
6) CPS agg	Simple aggregation	20	.050 (.011)	.088 (.014)
7) CPS agg	Residual aggregation	20	.06 (.011)	.183 (.019)
8) CPS agg, staggered laws	Residual aggregation	20	.048 (.011)	.130 (.017)
9) CPS agg	OLS	10	.443 (.025)	.51 (.025)
10) CPS agg	Simple aggregation	10	.053 (.011)	.065 (.012)
11) CPS agg	Residual aggregation	10	.093 (.014)	.178 (.019)
12) CPS agg, staggered laws	Residual aggregation	10	.088 (.014)	.128 (.017)
13) CPS agg	OLS	6	.383 (.024)	.433 (.024)
14) CPS agg	Simple aggregation	6	.068 (.013)	.07 (.013)
15) CPS agg	Residual aggregation	6	.11 (.016)	.123 (.016)
16) CPS agg, staggered laws	Residual aggregation	6	.09 (.014)	.138 (.017)
B. AR(1) DISTRIBUTION				
17) AR(1), $\rho = .8$	Simple aggregation	50	.050 (.013)	.243 (.025)
18) AR(1), $\rho = .8$	Residual aggregation	50	.045 (.012)	.235 (.024)
19) AR(1), $\rho = .8$, staggered laws	Residual aggregation	50	.075 (.015)	.355 (.028)

TABLE VII
EMPIRICAL VARIANCE-COVARIANCE MATRIX

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
A. CPS DATA				
1) CPS aggregate	OLS	50	.49 (.025)	.663 (.024)
2) CPS aggregate	Empirical variance	50	.055 (.011)	.243 (.021)
3) CPS aggregate	OLS	20	.39 (.024)	.54 (.025)
4) CPS aggregate	Empirical variance	20	.08 (.013)	.138 (.017)
5) CPS aggregate	OLS	10	.443 (.025)	.510 (.025)
6) CPS aggregate	Empirical variance	10	.105 (.015)	.145 (.018)
7) CPS aggregate	OLS	6	.383 (.025)	.433 (.025)
8) CPS aggregate	Empirical variance	6	.153 (.018)	.185 (.019)
B. AR(1) DISTRIBUTION				
9) AR(1), $\rho = .8$	Empirical variance	50	.07 (.017)	.25 (.030)

TABLE VIII
ARBITRARY VARIANCE-COVARIANCE MATRIX

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
A. CPS DATA				
1) CPS aggregate	OLS	50	.49	.663
			(.025)	(.024)
2) CPS aggregate	Cluster	50	.063	.268
			(.012)	(.022)
3) CPS aggregate	OLS	20	.385	.535
			(.024)	(.025)
4) CPS aggregate	Cluster	20	.058	.13
			(.011)	(.017)
5) CPS aggregate	OLS	10	.443	.51
			(.025)	(.025)
6) CPS aggregate	Cluster	10	.08	.12
			(.014)	(.016)
7) CPS aggregate	OLS	6	.383	.433
			(.024)	(.025)
8) CPS aggregate	Cluster	6	.115	.118
			(.016)	(.016)
B. AR(1) DISTRIBUTION				
9) AR(1), $\rho = .8$	Cluster	50	.045	.275
			(.012)	(.026)
10) AR(1), $\rho = 0$	Cluster	50	.035	.74
			(.011)	(.025)