Difference in Differences

Christopher Taber

Department of Economics
University of Wisconsin-Madison

February 13, 2018

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 T_{it} after taking accounting for the fixed effect

We don't need to make any assumptions about θ_i

Background on Fixed effect.

Lets forget about the basic problem and review fixed effects 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 the vector of u_{it} is uncorrelated with the vector of X_{it} (though this is stronger than what we need)

Also one can think of θ_i as a random intercept, so there is no intercept included in X_{it}

For a generic variable Z_{it} define

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

then notice that

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

So

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

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

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

(From here you can see that what we need for consistency is that $E\left[(X_{it} - \bar{X})\left(u_{it} - \bar{u}_i\right)\right] = 0$)

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_i be a $N \times 1$ vector of dummy variables so that for the j^{th}element:

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

and write the regression model as

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

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_i

Abusing notation somewhat, the least squares estimator for this is

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

- The matrix $\sum_{i=1}^{N} \sum_{t=1}^{T_i} D_i D_i^{'}$ is an $N \times N$ diagonal matrix with each (i,i) diagonal element equal to T_i.
- The vector $\sum_{i=1}^{N} \sum_{t=1}^{T_i} D_i Z_{it}$ 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 $\bar{Z}_i$
- $D_i' \widehat{\delta} = \bar{Z}_i$

Or using notation from the previous lecture notes we can write

$$\widetilde{Z} = M_D Z$$

where a generic row of this matrix is

$$Z_{it} - D_i'\widehat{\delta} = Z_{it} - \bar{Z}_i$$

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

Model vs. Estimator

For me it is very important to distinguish the econometric model or data generating process from the method we use to estimate these models.

- The model is
$$Y_{it} = X_{it}\beta + \theta_i + u_{it}$$
- We can get consistent estimates of β by regressing Y_{it} on X_{it} and individual dummy variables

This is conceptually different than writing the model as

$$Y_{it} = X_{it}\beta + D_i'\theta + u_{it}$$

Technically they are the same thing but:

- The equation is strange because notationally the true data generating process for Y_{it} depends upon the sample
- More conceptually the model and the way we estimate them are separate issues-this mixes the two together

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

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

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

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

The denominator is

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

So the fixed effects estimator is just

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

Next consider the first differences estimator

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

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

and only two time periods before ($t = 0$) and after ($t = 1$)

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:

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

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

$$Y_{it} = \beta_0 + \alpha T_{s(i)t} + \delta t + \theta_i + \varepsilon_{it}$$

where $s(i)$ indicates persons suit

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

Let $s(i)$ indicate and individual's suit (either ♦ or ♣)

Further we will assume that

$$T_{st} = \begin{cases} 0 & s = \clubsuit \\ 0 & s = \diamondsuit, t = 0 \\ 1 & s = \diamondsuit, t = 1 \end{cases}$$

Identification

Lets first think about identification in this case notice that

$$
\begin{aligned}
&[E(Y_{i,1} \mid s(i) = \blacklozenge) - E(Y_{i,0} \mid s(i) = \blacklozenge)] \\
&- [E(Y_{i,1} \mid s(i) = \clubsuit) - E(Y_{i,0} \mid s(i) = \clubsuit)] \\
=&[(\beta_0 + \alpha + \delta + E(\theta_i \mid s(i) = \blacklozenge)) - (\beta_0 + E(\theta_i \mid s(i) = \blacklozenge))] \\
&- [(\beta_0 + \delta + E(\theta_i \mid s(i) = \clubsuit)) - (\beta_0 + E(\theta_i \mid s(i) = \clubsuit))] \\
=&\alpha + \delta \\
&- \delta \\
=&\alpha
\end{aligned}
$$

Fixed Effects Estimation

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

$$(Y_{i1} - Y_{i0}) = \delta + \alpha \left(T_{s(i)1} - T_{s(i)0}\right) + (\varepsilon_{i1} - \varepsilon_{i0})$$

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

Note that for $\diamondsuit$'s, $T_{s(i)1} - T_{s(i)0} = 1$, but for $\clubsuit$'s, $T_{s(i)1} - T_{s(i)0} = 0$

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

$$\widehat{\alpha} = \frac{\sum_{i=1}^{N}\left(\left(T_{s(i)1} - T_{s(i)0}\right) - \left(\bar{T}_1 - \bar{T}_0\right)\right)\left(Y_{i1} - Y_{i0}\right)}{\sum_{i=1}^{N}\left(T_{s(i)1} - T_{s(i)0} - \bar{T}_1 + \bar{T}_0\right)^2}$$

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

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

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

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 $\blacklozenge$, let this be $\blacklozenge_i$

Then we can write the regression as

$$Y_i = \widehat{\beta}_0 + \widehat{\alpha} T_{s(i)t(i)} + \widehat{\delta} t(i) + \widehat{\gamma} \blacklozenge_i + \widehat{\varepsilon}_i$$

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

$$
\begin{aligned}
0 &= \sum_{i=1}^{N} \widehat{\varepsilon}_i \\
&= \sum_{\blacklozenge,0} \widehat{\varepsilon}_i + \sum_{\blacklozenge,1} \widehat{\varepsilon}_i + \sum_{\clubsuit,0} \widehat{\varepsilon}_i + \sum_{\clubsuit,1} \widehat{\varepsilon}_i
\end{aligned}
$$

$$
\begin{aligned}
0 &= \sum_{i=1}^{N} T_{s(i)t(i)} \widehat{\varepsilon}_i \\
&= \sum_{\blacklozenge,1} \widehat{\varepsilon}_i
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N} t(i)\widehat{\varepsilon}_i \\
&= \sum_{\blacklozenge,1}\widehat{\varepsilon}_i + \sum_{\clubsuit,1}\widehat{\varepsilon}_i
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N} \blacklozenge_i\widehat{\varepsilon}_i \\
&= \sum_{\blacklozenge,0}\widehat{\varepsilon}_i + \sum_{\blacklozenge,1}\widehat{\varepsilon}_i
\end{aligned}
$$

We can rewrite these equations as

$$0 = \sum_{\diamondsuit,0} \widehat{\varepsilon}_i$$

$$0 = \sum_{\diamondsuit,1} \widehat{\varepsilon}_i$$

$$0 = \sum_{\clubsuit,0} \widehat{\varepsilon}_i$$

$$0 = \sum_{\clubsuit,1} \widehat{\varepsilon}_i$$

Using

$$Y_i = \widehat{\beta}_0 + \widehat{\alpha} T_{s(i)t(i)} + \widehat{\delta} t(i) + \widehat{\gamma} \blacklozenge_i + \widehat{\varepsilon}_i$$

we can write as

$$\begin{aligned}
\bar{Y}_{\blacklozenge 0} &= \widehat{\beta}_0 + \widehat{\gamma} \\
\bar{Y}_{\blacklozenge 1} &= \widehat{\beta}_0 + \widehat{\alpha} + \widehat{\delta} + \widehat{\gamma} \\
\bar{Y}_{\clubsuit 0} &= \widehat{\beta}_0 \\
\bar{Y}_{\clubsuit 1} &= \widehat{\beta}_0 + \widehat{\delta}
\end{aligned}$$

We can solve for the parameters as

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

Now more generally we can think of "difference in differences" as

$$Y_i = \beta_0 + \alpha T_{g(i)t(i)} + \delta_{t(i)} + \theta_{g(i)} + \varepsilon_i$$

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

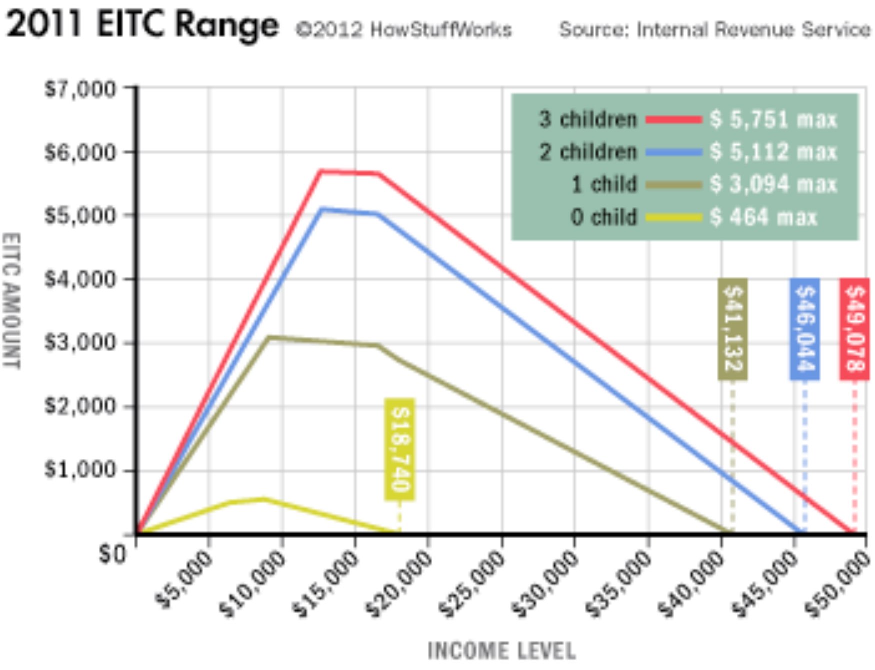
There are many papers that do this basic sort of thing

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:



2011 EITC Range
©2012 HowStuffWorks
Source: Internal Revenue Service
3 children $ 5,751 max
2 children $ 5,112 max
1 child $ 3,094 max
0 child $ 464 max
$18,740
$41,132
$46,044
$49,078
EITC AMOUNT
$7,000
$6,000
$5,000
$4,000
$3,000
$2,000
$1,000
$0
$5,000
$10,000
$15,000
$20,000
$25,000
$30,000
$35,000
$40,000
$45,000
$50,000
INCOME LEVEL

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	1.00	1.50
After	1.10	1.65

This would give a difference in difference estimate of -0.05.

However how do we know what the right metric is?

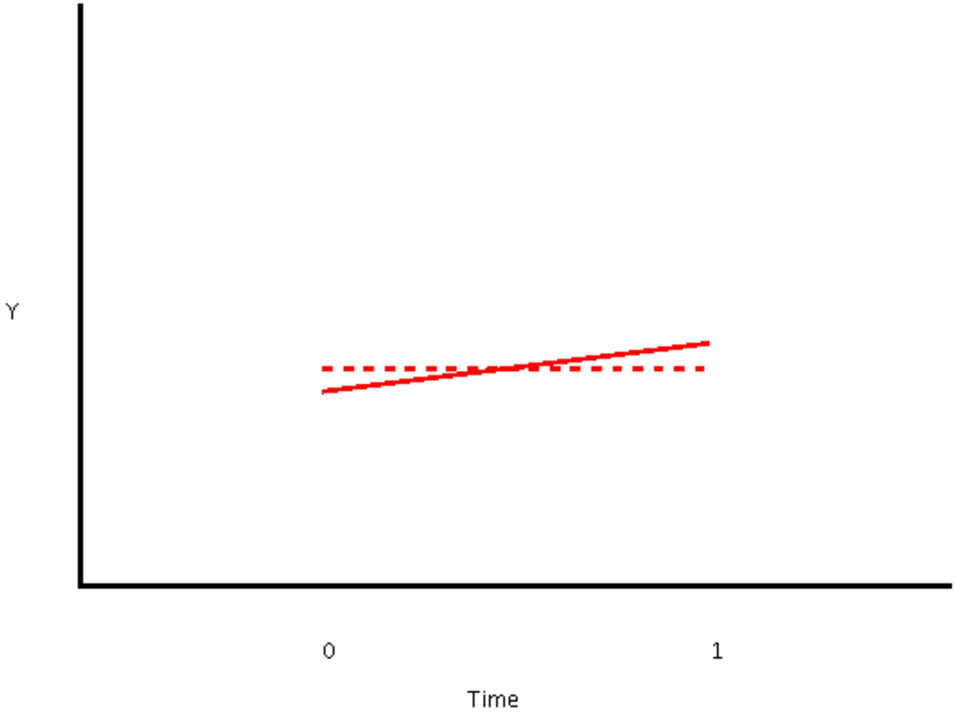
Take logs and you get

	Treatment	Control
Before	0.00	0.41
After	0.10	0.50

This gives diff-in-diff estimate of 0.01

So even the sign is not robust

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

$$Prob(Y_i = 1) = \Phi\left(\beta_0 + \alpha T_{g(i)t} + X_i'\beta + \delta_{t(i)} + \theta_{g(i)}\right)$$

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

TABLE III

PROBIT RESULTS: CHILDREN VERSUS NO CHILDREN ALL UNMARRIED WOMEN

	Sample: all unmarried women					
Variables	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.

Hastings, “Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California” (*AER*, 2004)

IO uses these methods as well.

Lets look at an important example (not something I know well).

There is huge variation in the price of gas across different geographic areas

One explanation is that increases in gas prices are due to vertical integration-retail stations are often owned by refiners.

Types of ownership of stations

- Independent
  - use unbranded gas
  - can shop among refiners to find best deal
- Branded
  - must used branded gas and have their signage
  - Three types
    - company operated station
    - lessee dealer (company owned but leased)
    - dealer owned

In 1997 ARCO (branded) took over most of the Thriftys (independent) in southern california.

Hastings does a difference in differences design to see if having an indpendent or company operated station near by.

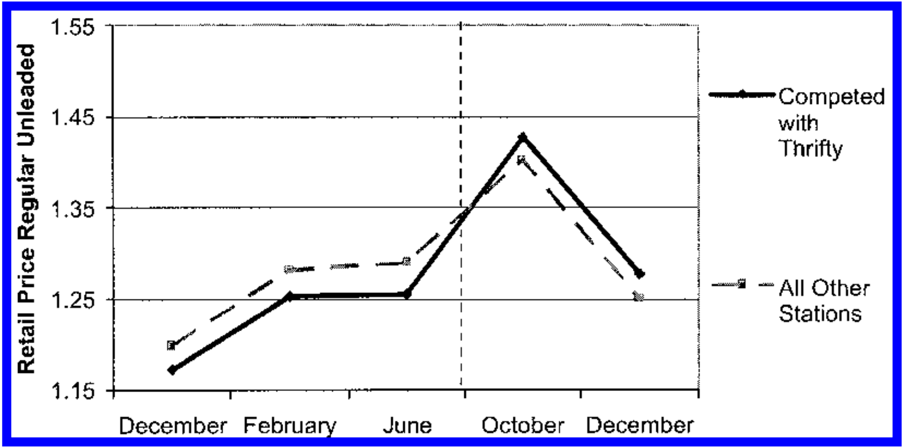
Nearby is described as within one mile

TABLE 1—SUMMARY STATISTICS OF RETAIL PRICE SAMPLE

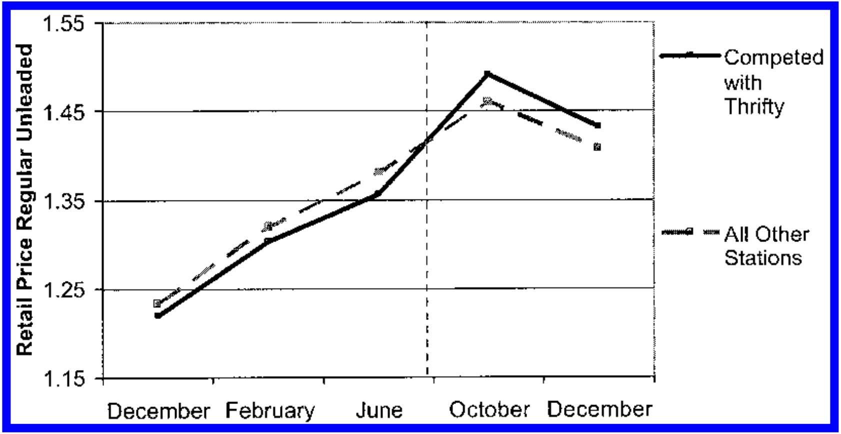
Panel A		
Percent of stations in sample	Los Angeles	San Diego
ARCO	19.41	13.21
Chevron	17.84	17.61
Mobil	15.88	13.21
Shell	14.12	17.61
Texaco	8.43	12.58
Unocal	12.55	11.95
Minor brands	5.25	8.18
Independents	6.52	5.66
Number of observations	$N = 510$	$N = 159$

Panel B

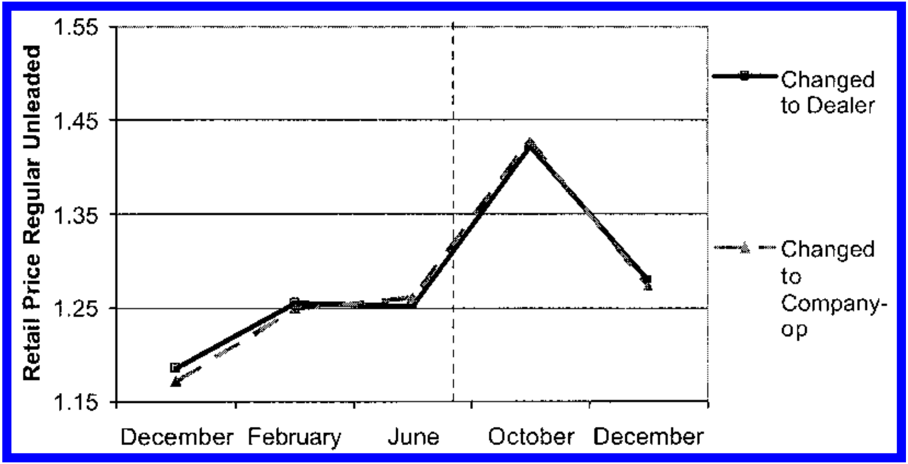
Average price (Standard deviation)	Los Angeles	San Diego
February, 1997	1.273	1.320
	(0.060)	(0.035)
June, 1997	1.285	1.375
	(0.068)	(0.049)
October, 1997	1.405	1.468
	(0.070)	(0.056)
December, 1997	1.266	1.414
	(0.073)	(0.0610)



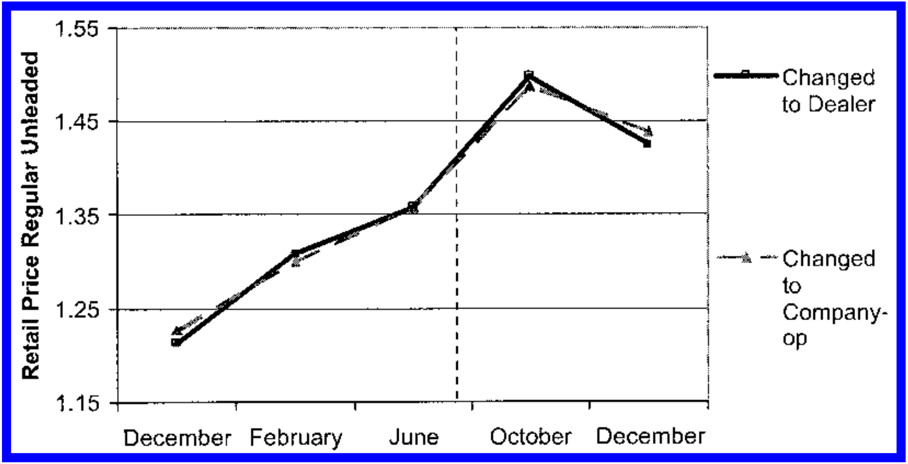
(a) Los Angeles


1.55
1.45
1.35
1.25
1.15
Retail Price Regular Unleaded
December
February
June
October
December
Competed with Thrifty
All Other Stations

(b) SAN DIEGO



(a) Los Angeles


1.55
1.45
1.35
1.25
1.15
Retail Price Regular Unleaded
December
February
June
October
December
Changed to Dealer
Changed to Company-op

(b) SAN DIEGO

The basic specification (in my notation)

$$p_{it} = \alpha_i + \sigma_{\ell(i)t} + \phi c_{it} + \theta Z_{it} + \varepsilon_{it}$$

Where

- i is station
- $\ell(i)$ location of station
- t quarter
- c_{it} company operated competitor
- Z_{it} Independent competitor

TABLE 2—FIXED-EFFECTS ESTIMATION

	Dependent variable: Retail price for regular unleaded		
Variable	(1)	(2)	(3)
Intercept	1.3465	1.3465	1.3617
	(0.0421)	(0.0415)	(0.0287)
Company operated	0.1080	−0.0033	−0.0033
	(0.0107)	(0.0178)	(0.0122)
Independent	—	−0.1013	−0.0500
		(0.0143)	(0.0101)
LA*February	—	—	0.0180
			(0.0065)
LA*June	—	—	0.0243
			(0.0065)
LA*October	—	—	0.1390
			(0.0064)
SD*February	—	—	−0.0851
			(0.0036)
SD*June	—	—	−0.0304
			(0.0036)
SD*October	—	—	0.0545
			(0.0036)

TABLE 3—FIXED-EFFECTS ESTIMATION, INDEPENDENT COEFFICIENT BY BRAND GROUP

Dependent variable: Retail price for regular unleaded (Standard errors are in parentheses)		
Variable	(1) Parameter estimate	(2) Parameter estimate
Intercept	1.3622	1.3620
	(0.0287)	(0.0287)
Company operated	−0.0018	−0.0008
	(0.0124)	(0.0124)
Independent · High-share brands	−0.0273	−0.0362
	(0.0125)	(0.0156)
Independent · Middle-share brands	−0.0530	−0.0617
	(0.0154)	(0.0179)
Independent · Low-share brands	−0.0700	−0.0741
	(0.0185)	(0.0190)
Independent · ARCO	−0.0731	−0.0741
	(0.0149)	(0.0149)
Independent · *N*-decreased	—	0.0130
		(0.0136)
City-time effects	Yes	Yes
Adjusted R^2	0.7183	0.7187

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

$$EffectiveAbortion_{jt} = \sum_a Abortionlegal_{jt-a} \left(\frac{Arrests_a}{Arrests_{total}} \right)$$

The model is then estimated using difference in differences:

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

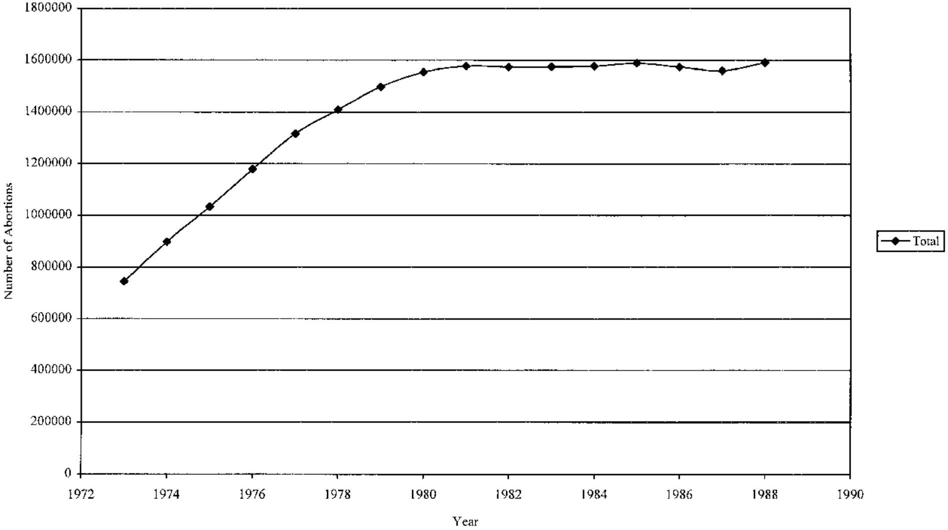


FIGURE I
Total Abortions by Year
Source: Alan Guttmacher Institute [1992].

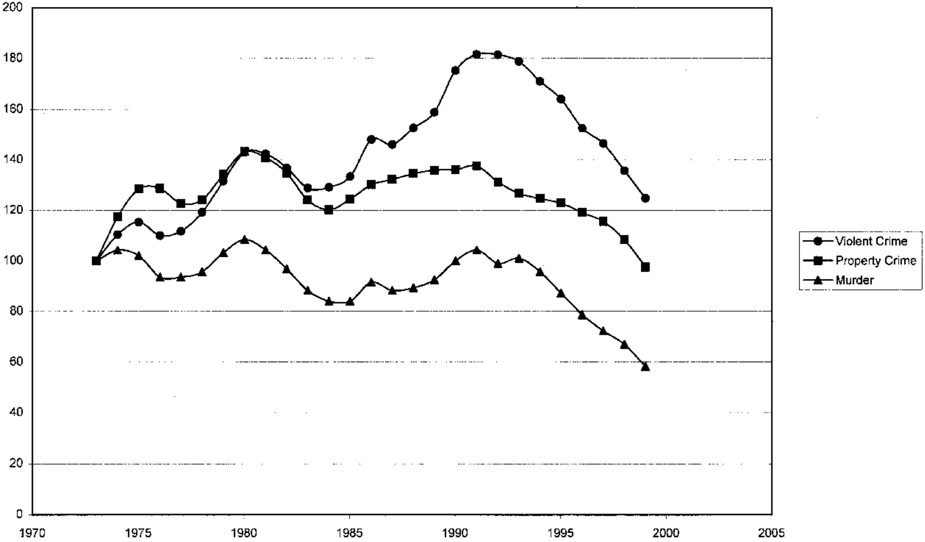


FIGURE II

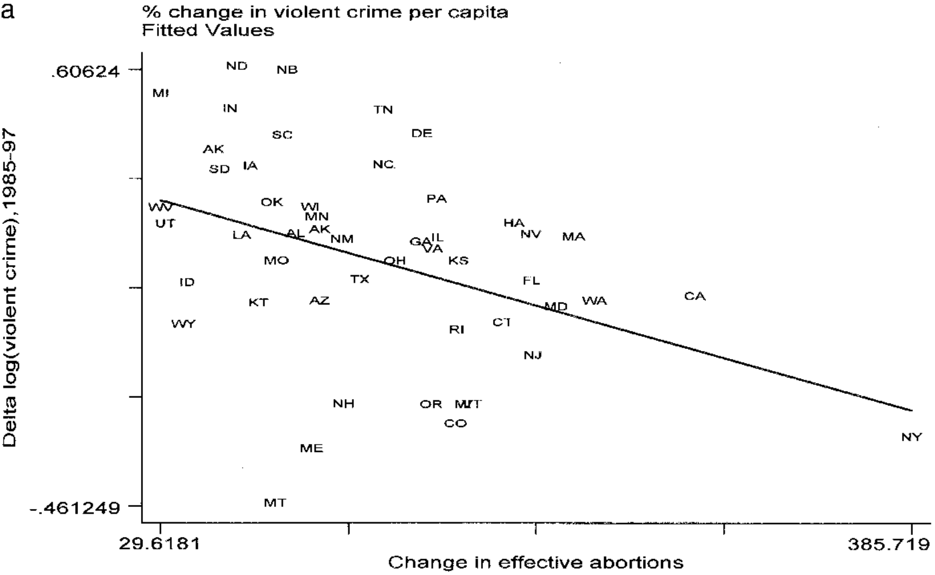
Crime Rates from the Uniform Crime Reports, 1973–1999

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: 1976–1982	1982–1985	1988–1994	1994–1997	Cumulative, 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
WV
OK
WI
PA
MN
UT
HA
LA
AL
AK
NM
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

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

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

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

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

etc.

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

$$Y_i = \beta_0 + \alpha_1 T_{1g(i)t(i)} + \alpha_2 T_{2g(i)t(i)} + \alpha_3 T_{3g(i)t(i)} + \delta_{g(i)} + \rho_{t(i)} + \varepsilon_i$$

Key Assumption

Lets think about the unbiasedness of DD

Going to the original model above we had

$$Y_i = \beta_0 + \alpha T_{s(i)t(i)} + \delta t(i) + \gamma \blacklozenge_i + \varepsilon_i$$

so

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

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

But 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

Figure 3: Effect of Switch to FDLP on Federal Borrowing Rate

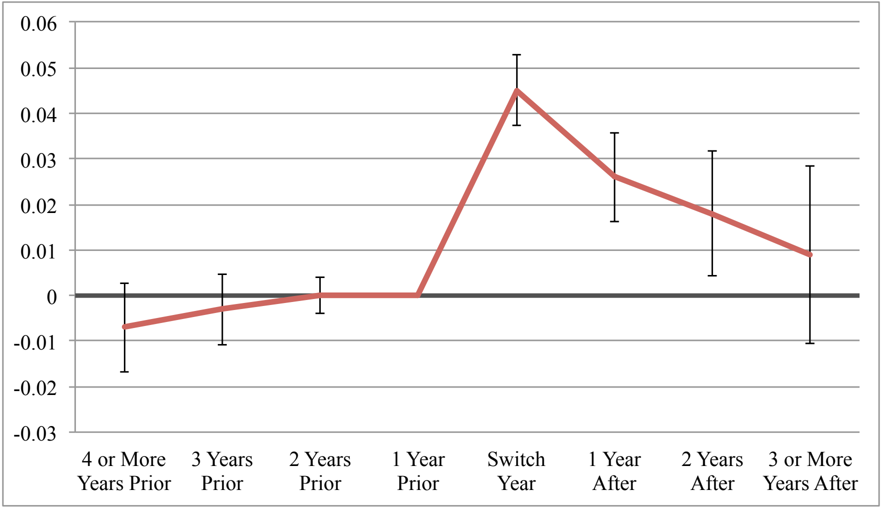
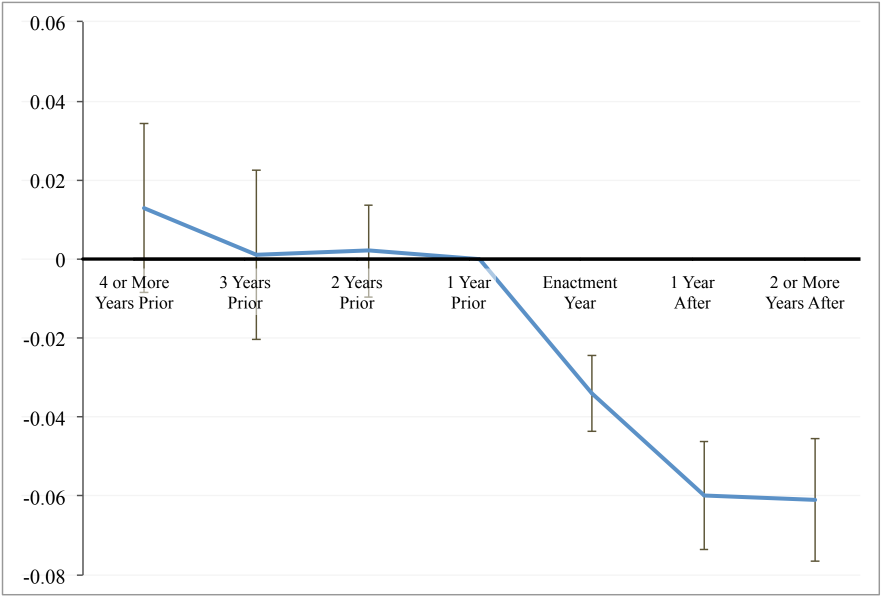


Figure 5: Effect of Lost Eligibility on Ln(Sticker Price)



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:

$$Y_i = \beta_0 + \alpha T_{s(i)t(i)} + \delta_{\diamondsuit} t(i) \diamondsuit_i + \delta_{\clubsuit} t(i)[1 - \diamondsuit_i] + \delta_2 1(t(i) = 2) + \gamma \diamondsuit_i + \varepsilon_{it}$$

Notice that this is 6 parameters in 6 unknowns

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

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

So that works

You can also just do this with state specific time trends

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 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)

Inference

In most of the cases discussed above, the authors had individual data and state variation

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

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

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:
$$\overline{Y}_{jt} = \alpha T_{jt} + \overline{Z}_{jt}'\delta + X_{jt}'\beta + \theta_j + \gamma_t + \overline{u}_{jt}$$
- 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) because

$$\overline{u}_{jt} = \eta_{j(i)t(i)} + \overline{\varepsilon}_{jt}$$

The variance of η_{jt} might be small relative to the variance of ε_i, but might be large relative to the variance of $\overline{\varepsilon}_{jt}$

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:

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

The White heteroskedastic standard errors just use

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

And approximate

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

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	Rejection rate: 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	Rejection rate: 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 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)

Conley and Taber

"Inference with Difference in Differences with a Small Number of Policy Changes," with T. Conley, (RESTAT, Feb., 2011)

We want to address one particular problem with many implementations of Difference in Differences

Often one wants to evaluate the effect of **a single state** or **a few states** changing/introducing a policy

A nice example is the Georgia HOPE Scholarship Program-a single state operated as the treatment

Simple Case

Assuming simple case (one observation per state$\times$year no regressors):

$$Y_{jt} = \alpha T_{jt} + \theta_j + \gamma_t + \eta_{jt}$$

Run regression of Y_{jt} on presence of program (T_{jt}), state dummies and time dummies

Simple Example

Suppose there is only one state that introduces the program at time t^*

Denote that state as $j = 1$

It is easy to show that (with balanced panels)

$$\widehat{\alpha}_{FE} = \alpha + \left(\frac{1}{T - t^*} \sum_{t=t^*+1}^{T} \eta_{1t} - \frac{1}{t^*} \sum_{t=1}^{t^*} \eta_{1t} \right)$$
$$- \left(\frac{1}{(N-1)} \sum_{j=2}^{N} \frac{1}{(T - t^*)} \sum_{t=t^*+1}^{T} \eta_{jt} - \frac{1}{(N-1)} \sum_{j=2}^{N} \frac{1}{t^*} \sum_{t=1}^{t^*} \eta_{jt} \right).$$

If

$$E\left(\eta_{jt} \mid d_{jt}, \theta_j, \gamma_t, X_{jt}\right) = 0.$$

it is unbiased.

However, this model is not consistent as $N \to \infty$ because the first term never goes away.

On the other hand, as $N \to \infty$ we can obtain a consistent estimate of the distribution of $\left(\frac{1}{T-t^*}\sum_{t=t^*+1}^{T}\eta_{1t} - \frac{1}{t^*}\sum_{t=1}^{t^*}\eta_{1t}\right)$ so we can still do inference (i.e. hypothesis testing and confidence interval construction) on α.

This places this work somewhere between small sample inference and Large Sample asymptotics

Base Model

Most straightforward case is when we have 1 observation per group$\times$year as before with

$$Y_{jt} = \alpha T_{jt} + X_{jt}'\beta + \theta_j + \gamma_t + \eta_{jt}$$

Generically define $\widetilde{Z}_{jt}$ as residual after regressing S_{jt} on group and time dummies

Then

$$\widetilde{Y}_{jt} = \alpha \widetilde{T}_{jt} + \widetilde{X}'_{jt}\beta + \widetilde{\eta}_{jt}.$$

"Difference in Differences" is just OLS on this regression equation

We let N_0 denote the number of “treatment” groups that change the policy $($i.e. d_{jt} changes during the panel)

We let N_1 denote the number of “control” groups that do not change the policy (i.e. T_{jt} constant)

We allow $N_1 \to \infty$ but treat N_0 as fixed

Assumption

$((X_{j1}, \eta_{j1}), ..., (X_{jT}, \eta_{jT}))$ *is IID across groups;* $(\eta_{j1}, ..., \eta_{jT})$ *is expectation zero conditional on* $(d_{j1}, ..., d_{jT})$ *and* $(X_{j1}, ..., X_{jT})$; *and all random variables have finite second moments.*

Assumption

$$\frac{1}{N_1 + N_0} \sum_{j=1}^{N_1+N_0} \sum_{t=1}^{T} \widetilde{X}_{jt} \widetilde{X}_{jt}' \xrightarrow{p} \Sigma_x$$

where Σ_x *is finite and of full rank.*

Proposition

Under Assumptions 1.1-1.2, As $N_1 \to \infty : \widehat{\beta} \xrightarrow{p} \beta$ *and* $\widehat{\alpha}$ *is unbiased and converges in probability to* $\alpha + W$*, with:*

$$W = \frac{\sum_{j=1}^{N_0} \sum_{t=1}^{T} \left(T_{jt} - \overline{T}_j\right)\left(\eta_{jt} - \overline{\eta}_j\right)}{\sum_{j=1}^{N_0} \sum_{t=1}^{T} \left(T_{jt} - \overline{T}_j\right)^2}.$$

Bad thing about this: **Estimator of α is not consistent**

Good thing about this: **We can identify the distribution of $\widehat{\alpha} - \alpha$.**

As a result we can get consistent estimates of the distribution of $\widehat{\alpha}$ up to α.

To see how the distribution of $\left(\eta_{jt} - \overline{\eta}_j\right)$can be estimated, notice that for the controls

$$\begin{aligned} \widetilde{Y}_{jt} - \widetilde{X}'_{jt}\hat{\beta} &= \widetilde{X}'_{jt}(\hat{\beta} - \beta) + \left(\eta_{jt} - \overline{\eta}_j - \overline{\eta}_t + \overline{\eta}\right) \\ &\xrightarrow{p} \left(\eta_{jt} - \overline{\eta}_j\right) \end{aligned}$$

So the distribution of $\left(\eta_{jt} - \overline{\eta}_j\right)$ is identified using residuals from control groups with the following additional assumption

Assumption

$(\eta_{j1}, ..., \eta_{jT})$ *is independent of* $(d_{j1}, ..., d_{jT})$ *and* $(X_{j1}, ..., X_{jT})$, *with a bounded density.*

Let

$$\Gamma(a) \equiv \text{plim}\, \Pr((\widehat{\alpha} - \alpha) < a \mid \{T_{jt}, j = 1, .., N_0, t = 1, ..., T\}).$$

For the N_0=1 case we can estimate $\Gamma(a)$ using

$$\widehat{\Gamma}(a) \equiv \frac{1}{N_1} \sum_{\ell=N_0+1}^{N_0+N_1} 1 \left(\frac{\sum_{t=1}^{T} \left(T_{1t} - \overline{T}_1\right) \left(\widetilde{Y}_{\ell t} - \widetilde{X}'_{\ell t}\hat{\beta}\right)}{\sum_{t=1}^{T} \left(T_{1t} - \overline{T}_1\right)^2} < a \right).$$

More generally

$$\widehat{\Gamma}(a) \equiv$$

$$\left(\frac{1}{N_1}\right)^{N_0} \sum_{\ell_1=N_0+1}^{N_0+N_1} \cdots \sum_{\ell_{N_0}=N_0+1}^{N_0+N_1} 1 \left(\frac{\sum_{j=1}^{N_0} \sum_{t=1}^{T} \left(T_{jt} - \overline{T}_j\right) \left(\widetilde{Y}_{\ell_j t} - \widetilde{X}'_{\ell_j t}\hat{\beta}\right)}{\sum_{j=1}^{N_0} \sum_{t=1}^{T} \left(T_{jt} - \overline{T}_j\right)^2} < a \right)$$

Proposition

Under Assumptions 1.1 and 1.2, $\widehat{\Gamma}(a)$ *converges uniformly to* $\Gamma(a)$.

To see why this is useful, first consider testing

$$H_0 : \alpha = \alpha_0$$

If $\widehat{\Gamma}$ were continuous we would 95% acceptance region by $[A_{\text{lower}}, A_{\text{upper}}]$ such that

$$\begin{aligned}\widehat{\Gamma}\left(A_{\text{upper}}-\alpha_0\right) &= 0.975\\ \widehat{\Gamma}\left(A_{\text{lower}}-\alpha_0\right) &= 0.025.\end{aligned}$$

Reject if $\widehat{\alpha}$ is outside $\left[A_{\text{lower}}, A_{\text{upper}}\right].$

(In practice since $\widehat{\Gamma}$ is not continuous, we need to approximate this)

As $N_1 \to \infty$,the coverage probability of this interval will converge to 95%.

Practical Example

To keep things simple suppose that:

- There are two periods $(T = 2)$
- There is only one “treatment state”
- Binary treatment $(T_{11} = 0, T_{12} = 1)$

Now consider testing the null: $\alpha = 0$

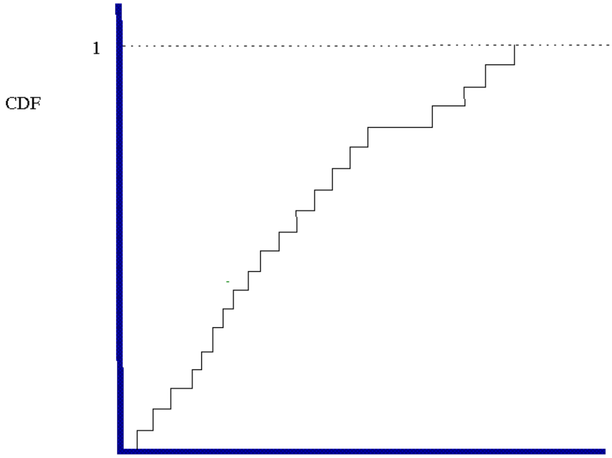
- First run DD regression of Y_{jt} on T_{jt}, X_{jt},time dummies and group dummies
- The estimated regression equation (abusing notation) can just be written as

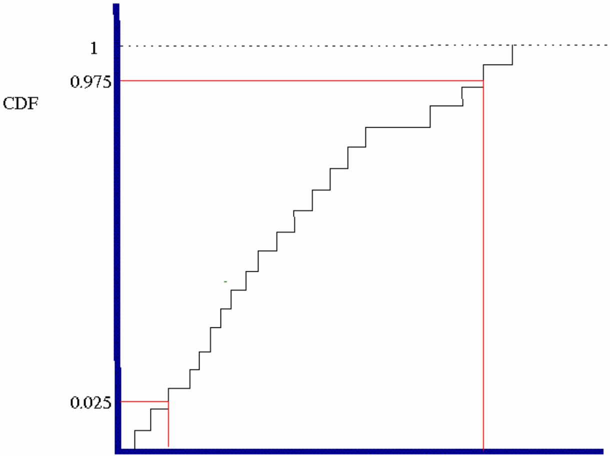
$$\Delta Y_j = \widehat{\gamma} + \widehat{\alpha}\Delta T_j + \Delta X_j'\widehat{\beta} + v_j$$

- Construct the empirical distribution of v_j using control states only
- now since the null is $\alpha = 0$ construct

$$v_1(0) = \Delta Y_1 - \widehat{\gamma} - \Delta X_1'\widehat{\beta}$$

- If this lies outside the 0.025 and 0.975 quantiles of the empirical distribution you reject the null


1
CDF


1
0.975
CDF
0.025

With two control states you would just get

$$v_1(\alpha^*) + v_2(\alpha^*)$$

and simulate the distribution of the sum of two objects

With $T > 2$ and different groups that change at different points in time, expression gets messier, but concept is the same

Model 2

More that 1 observation per state$\times$year

Repeated Cross Section Data (such as CPS):

$$Y_i = \alpha T_{j(i)t(i)} + X_i' \beta + \theta_{j(i)} + \gamma_{t(i)} + \eta_{j(i)t(i)} + \varepsilon_i.$$

Let $M(j,t)$ be the set of i in state j at time t

$|M(j(i),t)|$ be the size of that set

We can rewrite this model as

$$
\begin{aligned}
Y_i &= \lambda_{j(i)t(i)} + Z_i'\delta + \varepsilon_i \\
\lambda_{jt} &= \alpha T_{jt} + X_{jt}'\beta + \theta_j + \gamma_t + \eta_{jt}
\end{aligned}
$$

Suppose first that the number if individuals in a (j,t) cell is growing large with the sample size (i.e. $|M(j(i),t)| \to \infty$).

In that case one can estimate the model in two steps:

- First regress Y_i on Z_i and (j,t) dummies-this gives us a consistent estimate of λ_{jt}
- Now the second stage is just like our previous model

We show that one can ignore the first stage and do inference as in the previous section

This is just one example-we do a bunch more different cases in the paper

Monte Carlo Analysis

We also do a Monte Carlo Analysis to compare alternative approaches

The model we deal with is

$$
\begin{aligned}
Y_{jt} =& \alpha T_{jt} + \beta X_{jt} + \theta_j + \gamma_t + \eta_{jt} \\
\eta_{jt} =& \rho \eta_{jt-1} + u_{jt} \\
u_{jt} \sim& N(0,1) \\
X_{jt} =& a_x d_{jt} + \nu_{jt} \\
\nu_{jt} \sim& N(0,1)
\end{aligned}
$$

In base case

- $\alpha = 1$
- 5 Treatment groups
- $T = 10$
- T_{jt} binary
- turns on at 2,4,6,8,10
- $\rho = 0.5$
- $a_x = 0.5$
- $\beta = 1$

Monte Carlo Results

Size and Power of Test of at Most 5% Level[a]

Basic Model:

$$Y_{jt} = \alpha d_{jt} + \beta X_{jt} + \theta_j + \gamma_t + \eta_{jt}$$

$$\eta_{jt} = \rho\eta_{jt-1} + \varepsilon_{jt}, \alpha = 1, X_{jt} = a_x d_{jt} + \nu_{jt}$$

Percentage of Times Hypothesis is Rejected out of 10,000 Simulations

	Size of Test ($H_0 : \alpha = 1$)				Power of Test ($H_0 : \alpha = 0$)			
	Classic Model	Cluster	Conley Taber ($\widehat{\Gamma^*}$)	Conley Taber ($\widehat{\Gamma}$)	Classic Model	Cluster	Conley Taber ($\widehat{\Gamma^*}$)	Conley Taber ($\widehat{\Gamma}$)
Base Model[b]	14.23	16.27	4.88	5.52	73.23	66.10	54.08	55.90
Total Groups=1000	14.89	17.79	4.80	4.95	73.97	67.19	55.29	55.38
Total Groups=50	14.41	15.55	5.28	6.65	71.99	64.48	52.21	56.00
Time Periods=2	5.32	14.12	5.37	6.46	49.17	58.54	49.13	52.37
Number Treatments=1[c]	18.79	84.28	4.13	5.17	40.86	91.15	13.91	15.68
Number Treatments=2[c]	16.74	35.74	4.99	5.57	52.67	62.15	29.98	31.64
Number Treatments=10[c]	14.12	9.52	4.88	5.90	93.00	84.60	82.99	84.21
Uniform Error[d]	14.91	17.14	5.30	5.86	73.22	65.87	53.99	55.32
Mixture Error[e]	14.20	15.99	4.50	5.25	55.72	51.88	36.01	37.49
$\rho = 0$	4.86	15.30	5.03	5.57	82.50	86.42	82.45	83.79
$\rho = 1$	30.18	16.94	4.80	5.87	54.72	34.89	19.36	20.71
$a_x = 0$	14.30	16.26	4.88	5.55	73.38	66.37	54.08	55.93
$a_x = 2$	1418	16.11	4.82	5.49	73.00	65.91	54.33	55.76
$a_x = 10$	1036	9.86	11.00	11.90	51.37	47.78	53.29	54.59