Inference in "Difference in Differences" with a Small Number of Policy Changes

Timothy Conley[1] Christopher Taber[2]

[1]Department of Economics
University of Western Ontario

[2]Department of Economics
and Institute for Policy Research
Northwestern University

February 13, 2012

Difference in Differences

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 dummes

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 $N_1 \rightarrow \infty$ but treat N_0 as fixed

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 $\left(\eta_{jt} - \overline{\eta}_j\right)$ can be approximated by using residuals from control groups

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: $\alpha = 0$

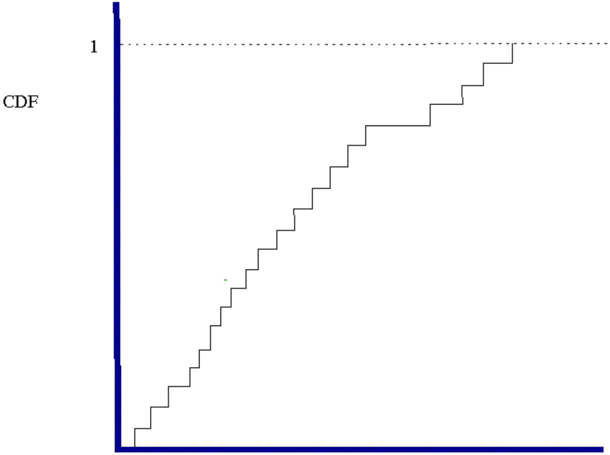
- First run DD regression of Y_{jt} on T_{jt}, X_{jt},time dummies and group dummies
- The estimated regression equation (abusing notation) can just be written as

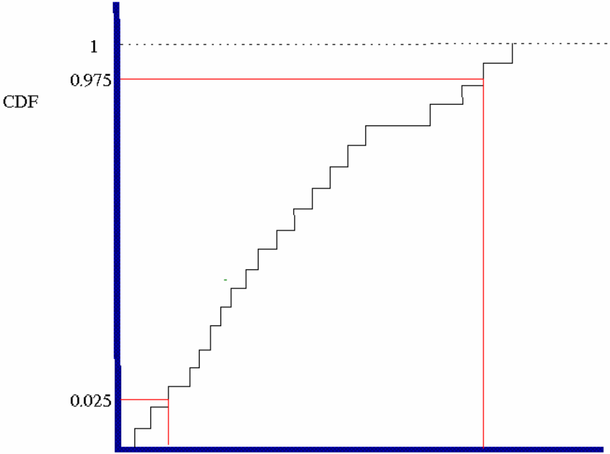
$$\Delta Y_j = \widehat{\gamma} + \widehat{\alpha}\Delta T_j + \Delta X_j'\widehat{\beta} + v_j$$

- 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):

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

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.

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

Application to Merit Aid programs

We start with Georgia only

Column (1)

As was discussed above:

- Run regression of Y_i on X_i and fully interacted state$\times$year dummies
- Then run regression of estimated state$\times$year dummies on d_{jt}, state dummies and time dummies
- Get estimate of $\hat{\alpha}$
- Using control states simulate distribution of $\hat{\alpha}$ under various null hypothesese
- Confidence intervals is the set of nulls that are not rejected

Estimates for

Effect of Georgia HOPE Program on College Attendance

	A	B	C
	Linear Probability	Logit	Population Weighted Linear Probability
Hope Scholarship	0.078	0.359	0.072
Male	-0.076	-0.323	-0.077
Black	-0.155	-0.673	-0.155
Asian	0.172	0.726	0.173
State Dummies	yes	yes	yes
Year Dummies	yes	yes	yes
95% Confidence intervals for Hope Effect			
Standard Cluster by State×Year	(0.025,0.130)	(0.119,0.600) [0.030,0.149]	(0.025, 0.119)
Standard Cluster by State	(0.058,0.097)	(0.274,0.444) [0.068,0.111]	(0.050,0.094)
Conley-Taber	(-0.010,0.207)	(-0.039,0.909) [-0.010,0.225]	(-0.015,0.212)
Sample Size			
Number States	42	42	42
Number of Individuals	34902	34902	34902

Estimates for

Merit Aid Programs on College Attendance

	A	B	C
	Linear Probability	Logit	Population Weighted Linear Probability
Merit Scholarship	0.051	0.229	0.034
Male	-0.078	-0.331	-0.079
Black	-0.150	-0.655	-0.150
Asian	0.168	0.707	0.169
State Dummies	yes	yes	yes
Year Dummies	yes	yes	yes
95% Confidence intervals for Merit Aid Program Effect			
Standard Cluster by State×Year	(0.024,0.078)	(0.111,0.346) [0.028,0.086]	(0.006,0.062)
Standard Cluster by State	(0.028,0.074)	(0.127,0.330) [0.032,0.082]	(0.008,0.059)
Conley-Taber	(0.012,0.093)	(0.056,0.407) [0.014,0.101]	(-0.003,0.093)
Sample Size			
Number States	51	51	51
Number of Individuals	42161	42161	42161

Estimates for

Effect of Georgia HOPE Program on College Attendance

	A	B	C
	Linear Probability	Logit	Population Weighted Linear Probability
Hope Scholarship	0.078	0.359	0.072
Male	-0.076	-0.323	-0.077
Black	-0.155	-0.673	-0.155
Asian	0.172	0.726	0.173
State Dummies	yes	yes	yes
Year Dummies	yes	yes	yes
95% Confidence intervals for Hope Effect			
Standard Cluster by State×Year	(0.025,0.130)	(0.119,0.600) [0.030,0.149]	(0.025, 0.119)
Standard Cluster by State	(0.058,0.097)	(0.274,0.444) [0.068,0.111]	(0.050,0.094)
Conley-Taber	(-0.010,0.207)	(-0.039,0.909) [-0.010,0.225]	(-0.015,0.212)
Sample Size			
Number States	42	42	42
Number of Individuals	34902	34902	34902

Estimates for

Merit Aid Programs on College Attendance

	A	B	C
	Linear Probability	Logit	Population Weighted Linear Probability
Merit Scholarship	0.051	0.229	0.034
Male	-0.078	-0.331	-0.079
Black	-0.150	-0.655	-0.150
Asian	0.168	0.707	0.169
State Dummies	yes	yes	yes
Year Dummies	yes	yes	yes
95% Confidence intervals for Merit Aid Program Effect			
Standard Cluster by State×Year	(0.024,0.078)	(0.111,0.346) [0.028,0.086]	(0.006,0.062)
Standard Cluster by State	(0.028,0.074)	(0.127,0.330) [0.032,0.082]	(0.008,0.059)
Conley-Taber	(0.012,0.093)	(0.056,0.407) [0.014,0.101]	(-0.003,0.093)
Sample Size			
Number States	51	51	51
Number of Individuals	42161	42161	42161

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