Part 8: Policy Evaluation
Event Studies and Difference in Difference

Chris Conlon

April 5, 2020

Applied Econometrics

Motivation: Recap Matching

Matching estimators had some advantages:

- Limited assumptions on functional forms
- We could do nearest neighbor matching and use kernels to compute treatment effects

Matching estimators had some drawbacks:

- Treated patients were "matched" to control patients based only on observable characteristics
  - Ignored selection on unobservables.
- Relied on cross sectional variation to construct a control group.

Motivation

IV estimators resolve some of those issues but

- Good IV are in short supply!

Often (in this course at least) we have access to panel data.

- What if we could use panel data to control for unobserved heterogeneity within a treated individual/group?

Difference in Difference estimators are like the opposite of matching

- Strong assumptions on functional form
- but... allow for unobservable heterogeneity in outcomes.

A Famous Example: Card and Krueger (AER 1994)

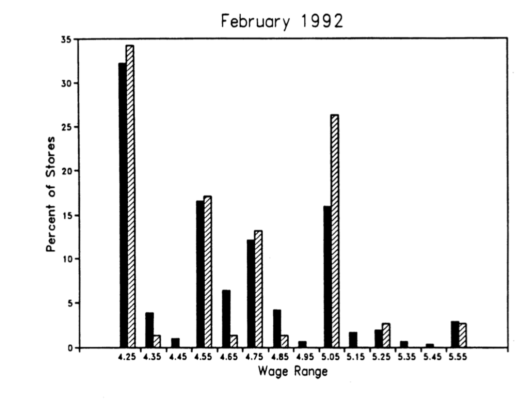
A Famous Example: Card and Krueger (AER 1994)

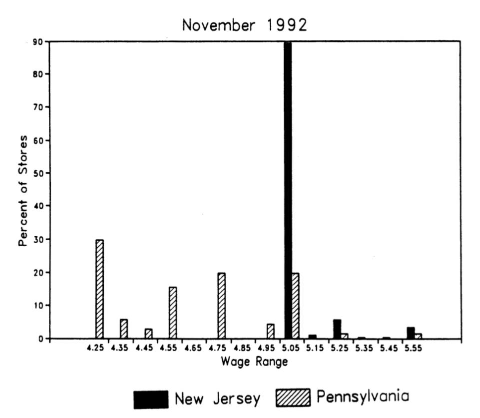
- On April 1, 1992 NJ raises its minimum wage from $\$4.25 \rightarrow \$5.05$ per hour.
- Question: Econ 101 predicts this will reduce demand for low wage workers
  - Focus on fast food restaurants (since they pay min wage)
  - Focus on starting wage (avoid tenure, high turnover)
- Survey 410 restaurants in NJ (treated group) and eastern PA (control group).
- Idea: Compare change in wages in NJ to PA: $\Delta_{DD} = \Delta_{NJ} - \Delta_{PA}$
  - Wave 1: February 15-March 4, 1992
  - Wave 2: November 5 - December 31, 1992

Balance Table: Covariates

Variable	Stores in: NJ	Stores in: PA	t[a]
1. *Distribution of Store Types (percentages):*			
a. Burger King	41.1	44.3	−0.5
b. KFC	20.5	15.2	1.2
c. Roy Rogers	24.8	21.5	0.6
d. Wendy's	13.6	19.0	−1.1
e. Company-owned	34.1	35.4	−0.2
2. *Means in Wave 1:*			
a. FTE employment	20.4 (0.51)	23.3 (1.35)	−2.0
b. Percentage full-time employees	32.8 (1.3)	35.0 (2.7)	−0.7
c. Starting wage	4.61 (0.02)	4.63 (0.04)	−0.4
d. Wage = \$4.25 (percentage)	30.5 (2.5)	32.9 (5.3)	−0.4
e. Price of full meal	3.35 (0.04)	3.04 (0.07)	4.0
f. Hours open (weekday)	14.4 (0.2)	14.5 (0.3)	−0.3
g. Recruiting bonus	23.6 (2.3)	29.1 (5.1)	−1.0
3. *Means in Wave 2:*			
a. FTE employment	21.0 (0.52)	21.2 (0.94)	−0.2
b. Percentage full-time employees	35.9 (1.4)	30.4 (2.8)	1.8
c. Starting wage	5.08 (0.01)	4.62 (0.04)	10.8
d. Wage = \$4.25 (percentage)	0.0	25.3 (4.9)	—
e. Wage = \$5.05 (percentage)	85.2 (2.0)	1.3 (1.3)	36.1
f. Price of full meal	3.41 (0.04)	3.03 (0.07)	5.0
g. Hours open (weekday)	14.4 (0.2)	14.7 (0.3)	−0.8
h. Recruiting bonus	20.3 (2.3)	23.4 (4.9)	−0.6

Distribution of Wages





Differences in Wages : 2 x 2 Table

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

Variable	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
	PA (i)	NJ (ii)	Difference, NJ − PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	−2.69 (1.37)	−2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	−2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores[c]	−2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	−2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	−2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	−2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

[a]Stores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ($N = 101$), is between \$4.26 and \$4.99 per hour ($N = 140$), or is \$5.00 per hour or higher ($N = 73$).

[b]Difference in employment between low-wage (\$4.25 per hour) and high-wage (≥ \$5.00 per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

[c]Subset of stores with available employment data in wave 1 and wave 2.

[d]In this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the subset of stores with available employment data in wave 1 and wave 2.

Outcome Equation

- Differences lack any covariates (different fast food chains).
- Also $\Delta_{PA} < 0$ and $\Delta_{NJ} > 0$ (!)
- Recall i denotes stores, $t \in 1, 2$. Run the following regression:

$$Y_{it} = \beta X_{it} + \alpha \cdot [i \in \text{NJ}] + \gamma \cdot \text{After}_t + \delta \cdot NJ_i \times After_t + u_i$$

$$Y_{it} = \beta X_{it} + \alpha \cdot [\text{wage gap}_i] + \gamma \cdot \text{After}_t + \delta \cdot \text{wage gap}_i \times After_t + u_i$$

- α is mean difference between NJ and PA
- γ is mean difference between period 1 and 2
- δ is the parameter of interest, the difference in difference
- wage gap$_i = [\text{min wage}_{i,2} - w_{i1}]_+ = \max\{0, \text{min wage}_{i,2} - w_{i1}\}$.
  (How much do you need to raise $t = 1$ wages to achieve minimum wage in $t = 2$?)

Differences in Wages

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model (i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap[a]	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership[b]	no	yes	no	yes	yes
4. Controls for region[c]	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls[d]	—	0.34	—	0.44	0.40

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are −0.237 and 8.825, respectively. All models include an unrestricted constant (not reported).

[a]Proportional increase in starting wage necessary to raise starting wage to new minimum rate. For stores in Pennsylvania the wage gap is 0.

[b]Three dummy variables for chain type and whether or not the store is company-owned are included.

[c]Dummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

[d]Probability value of joint F test for exclusion of all control variables.

Before and After

For now let's not bother with a control group.

- We look an outcome before or after an event $t \in 1, 2$
  - 1 before event happens
  - 2 after event happens.
  - A news event: the announcement of a merger or stock split.
  - A tax change, a new law, etc.
- Except under strong conditions $d_2 = d_1$ we shouldn't believe the results of the before and after estimator.
- Main Problem: we attribute changes to treatment that might have happened anyway trend.
- e.g: Cigarette consumption drops 4% after a tax hike. (But it dropped 3% the previous four years).
- Also worry about: anticipation, gradual rollout, etc.

A More General Method

General Case: Difference in Differences

- Sometimes we may feel we can impose more structure on the problem.
- Suppose in particular that we can write the outcome equation as

$$Y_{it} = \alpha_i + \gamma_t + \delta_i T_{it} + u_{it}$$

- Suppose that $T_{i1} = 0$ for all i and $T_{i2} = 1$ for a well defined group of individuals in our population.
- This framework allows us to identify the ATT effect under the assumption that the growth of the outcome in the non-treatment state is independent of treatment allocation:

$$E[Y_{i2}(0) - Y_{i1}(0) | T_{i1}, T_{i2}] = E[Y_{i2}(0) - Y_{i1}(0)]$$

- This is known as parallel trends.

Some Quick Algebra

Recall our regression equation:

$$Y_{it} = \alpha_i + \gamma_t + \delta_i T_{it} + u_{it}$$

This gives us:

$$\begin{aligned} E[Y_{i2} - Y_{i1} | T_{i2} = 1] &= E[Y_{i2}(1) - Y_{i1}(1) | T_{i2} = 1] \\ &= \underbrace{(\alpha_i - \alpha_i)}_{=0} + (\gamma_2 - \gamma_1) + \underbrace{E[\delta_i | T_{i2} = 1]}_{ATT} + E[u_{i2} - u_{i1} | T_{i2} = 1] \end{aligned}$$

Difference in Differences

Let's try and estimate $d_2 - d_1$ directly and then difference it out. Here we use parallel trends:

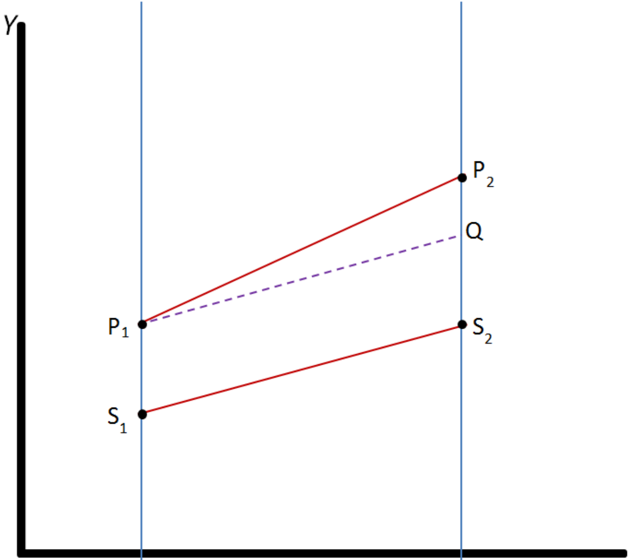
$$E[Y^0_{i2} - Y^0_{i1} | T_{i2} = 1] = E[Y^0_{i2} - Y^0_{i1} | T_{i2} = 0]$$
$$E[Y_{i2} - Y_{i1} | T_{i2} = 0] = d_2 - d_1$$

We now obtain an estimator for ATT:

$$E[\beta_i | T_{i2} = 1] = E[Y_{i2} - Y_{i1} | T_{i2} = 1] - E[Y_{i2} - Y_{i1} | T_{i2} = 0]$$

which can be estimated by the difference in the growth between the treatment and the control group.

Parallel Trends



Difference in Differences

Now consider the following problem:

- Suppose we wish to evaluate a training program for those with low earnings. Let the threshold for eligibility be B.
- We have a panel of individuals and those with low earnings qualify for training, forming the treatment group.
- Those with higher earnings form the control group.
- Now the low earning group is low for two reasons
  1. They have low permanent earnings (α_i is low) - this is accounted for by diff in diffs.
  2. They have a negative transitory shock (u_{i1} is low) - this is not accounted for by diff in diffs.

Difference in Differences

- #2 above violates the assumption $E[Y_{i2}^0 - Y_{i1}^0|T] = E[Y_{i2}^0 - Y_{i1}^0]$.
- To see why note that those participating into the program are such that $Y_{i0}^0 < B$. Assume for simplicity that the shocks u are iid. Hence $u_{i1} < B - \alpha_i - d_1$. This implies:

$$E[Y_{i2}^0 - Y_{i1}^0|T = 1] = d_2 = d_1 - E[u_{i1}|u_{i1} < B - \alpha_i - d_1]$$

For the control group:

$$E[Y_{i2}^0 - Y_{i1}^0|T = 1] = d_2 = d_1 - E[u_{i1}|u_{i1} > B - \alpha_i - d_1]$$

- Hence

$$E[Y_{i2}^0 - Y_{i1}^0|T = 1] - E[Y_{i2}^0 - Y_{i1}^0|T = 0] =$$
$$E[u_{i1}|u_{i1} > B - \alpha_i - d_1] - E[u_{i1}|u_{i1} < B - \alpha_i - d_1] > 0$$

- This is effectively regression to the mean: those unlucky enough to have a bad shock

Difference in Differences

Ashefelter (1978) was one of the first to consider difference in differences to evaluate training programs.

TABLE 1.—MEAN EARNINGS PRIOR, DURING, AND SUBSEQUENT TO TRAINING FOR 1964 MDTA CLASSROOM TRAINEES AND A COMPARISON GROUP

	White Males		Black Males		White Females		Black Females	
	Trainees	Comparison Group	Trainees	Comparison Group	Trainees	Comparison Group	Trainees	Comparison Group
1959	$1,443	$2,588	$ 904	$1,438	$ 635	$ 987	$ 384	$ 616
1960	1,533	2,699	976	1,521	687	1,076	440	693
1961	1,572	2,782	1,017	1,573	719	1,163	471	737
1962	1,843	2,963	1,211	1,742	813	1,308	566	843
1963	1,810	3,108	1,182	1,896	748	1,433	531	937
1964	1,551	3,275	1,273	2,121	838	1,580	688	1,060
1965	2,923	3,458	2,327	2,338	1,747	1,698	1,441	1,198
1966	3,750	4,351	2,983	2,919	2,024	1,990	1,794	1,461
1967	3,964	4,430	3,048	3,097	2,244	2,144	1,977	1,678
1968	4,401	4,955	3,409	3,487	2,398	2,339	2,160	1,920
1969	$4,717	$5,033	$3,714	$3,681	$2,646	$2,444	$2,457	$2,133
Number of Observations	7,326	40,921	2,133	6,472	2,730	28,142	1,356	5,192

Ashenfelter (1978) reports the following results.

TABLE 2.—CRUDE ESTIMATES (AND ESTIMATED STANDARD ERRORS), ASSUMING $B=0$ AND $\beta_j'=0$ FOR $j>1$, OF THE EFFECT OF TRAINING ON EARNINGS DURING AND AFTER TRAINING, WHITE MALE MDTA 1964 CLASSROOM TRAINEES

Effect in (value of t)	Value of Effects for $t-s=1963$	$t-s=1962$	$t-s=1961$
1962	—	—	91 (13)
1963	—	−179 (14)	−88 (17)
1964	−426 (16)	−605 (18)	−514 (20)
1965	763 (20)	584 (22)	675 (23)
1966	697 (25)	518 (27)	609 (28)
1967	833 (28)	655 (30)	746 (31)
1968	745 (34)	566 (35)	657 (36)

Difference in Differences

- The assumption on growth of the non-treatment outcome being independent of assignment to treatment may be violated, but it may still be true conditional on X.
- Consider the assumption

$$E[Y_{i2}^0 - Y_{i1}^0 | X, T] = E[Y_{i2}^0 - Y_{i1}^0 | X]$$

- This is just matching assumption on a redefined variable, namely the growth in the outcomes. In its simplest form the approach is implemented by running the regression

$$Y_{it} = \alpha_i + d_t + \beta_i T_{it} + \gamma_t' X_i + u_{it}$$

which allows for differential trends in the non-treatment growth depending on X_i. More generally one can implement propensity score matching on the growth of outcome variable when panel data is available.

Difference in Differences with Repeated Cross Sections

- Suppose we do not have available panel data but just a random sample from the relevant population in a pre-treatment and a post-treatment period. We can still use difference in differences.
- First consider a simple case where $E[Y_{i2}^0 - Y_{i1}^0|T] = E[Y_{i2}^0 - Y_{i1}^0]$.
- We need to modify slightly the assumption to

$$
\begin{aligned}
E[Y_{i2}^0|_{\text{Group receiving training}}] - E[Y_{i1}^0|_{\text{Group receiving training in the next period}}] \\
= E[Y_{i2}^0 - Y_{i1}^0]
\end{aligned}
$$

which requires, in addition to the original independence assumption that conditioned on particular individuals that population we will be sampling from does not change composition.

- We can then obtain immediately an estimator for ATT as

$$E[\beta_i|T_{i2} = 1]$$

Difference in Differences with Repeated Cross Sections

- More generally we need an assumption of conditional independence of the form

$$E[Y^0_{i2}|X, \text{Group receiving training}] - E[Y^0_{i1}|X, \text{Group receiving training next period}]$$
$$= E[Y^0_{i2}|X] - E[Y^0_{i1}|X]$$

- Under this assumption (and some auxiliary parametric assumptions) we can obtain an estimate of the effect of treatment on the treated by the regression

$$Y_{it} = \alpha_g + d_t + \beta T_{it} + \gamma' X_{it} + u_{it}$$

Difference in Differences with Repeated Cross Sections

- More generally we can first run the regression

$$Y_{it} = \alpha_g + d_t + \beta(X_{it})T_{it} + \gamma' X_{it} + u_{it}$$

  where α_g is a dummy for the treatment of comparison group, and $\beta(X_{it})$ can be parameterized as $\beta(X_{it}) = \beta' X_{it}$. The ATT can then be estimated as the average of $\beta' X_{it}$ over the (empirical) distribution of X.
- A non parametric alternative is offered by Blundell, Dias, Meghir and van Reenen (2004).

Difference in Differences and Selection on Unobservables

- Suppose we relax the assumption of *no selection* on unobservables.
- Instead we can start by assuming that

$$E[Y^0_{i2}|X, Z] - E[Y^0_{i1}|X, Z] = E[Y^0_{i2}|X] - E[Y^0_{i1}|X]$$

  where Z is an instrument which determines training eligibility say but does not determine outcomes in the non-training state. Take Z as binary (1,0).
- Non-Compliance: not all members of the eligible group ($Z = 1$) will take up training and some of those ineligible ($Z = 0$) may obtain training by other means.
- A difference in differences approach based on grouping by Z will estimate the impact of being allocated to the eligible group, but not the impact of training itself.

Difference in Differences and Selection on Unobservables

- Now suppose we still wish to estimate the impact of training on those being trained (rather than just the effect of being eligible)
- This becomes an IV problem and following up from the discussion of LATE we need stronger assumptions
  - Independence: for $Z = a, \{Y_{i2}^0 - Y_{i1}^0, Y_{i2}^1 - Y_{i1}^1, T(Z = a)\}$ is independent of Z.
  - Monotonicity $T_i(1) \geq T_i(0) \, \forall \, i$
- In this case LATE is defined by

$$[E(\Delta Y|Z = 1) - E(\Delta Y|Z = 0)]/[Pr(T(1) = 1) - Pr(T(0) = 1)]$$

assuming that the probability of training in the first period is zero.