Difference-in-Differences (DID)

Outline: Difference-in-Differences

a. Motivating example
b. Pre/post analyses and comparisons across groups
c. Relationship to difference-in-differences
d. Regression representation
e. Assumptions
f. Standard errors
g. Extensions
h. Results from example

Motivating example: DID

- Suppose we want to know the causal effect of neighborhood school quality on housing prices.

$$\text{School Quality} \xrightarrow{?} \text{Housing Prices}$$

- Problems in estimation:

  Selection bias likely -- e.g. schools in affluent neighborhoods may appear to be of higher quality because the children attending them have greater resources

Motivation: example

- Now suppose we observe a "natural experiment" in which redistricting occurs in a particular town (Shaker Heights, Ohio)
- The redistricting was accompanied by busing and several schools were also closed
- Plan was announced January 1987, approved March 1987, first affected the 1987-88 school year
- Redistricting could affect school quality (or *perceived* school quality) via the following mechanisms (next slide).

Possible mechanisms for redistricting to affect school quality (or *perceived* school quality)

1. Neighborhood school effects
   - harder for parents to get involved, harder for children to participate in after-school programs.
2. Racial composition effects
   - one of the goals of redistricting in this example was to foster greater racial integration. This could affect *perceived* school quality or the "taste" for school for those with integrationist or segregationist preferences.
3. Transportation effect
   - busing could be seen as a positive for busy parents.

This example from Bogart & Cromwell (2000) "How much Is a Neighborhood School Worth" J. Urban Economics, v.47.

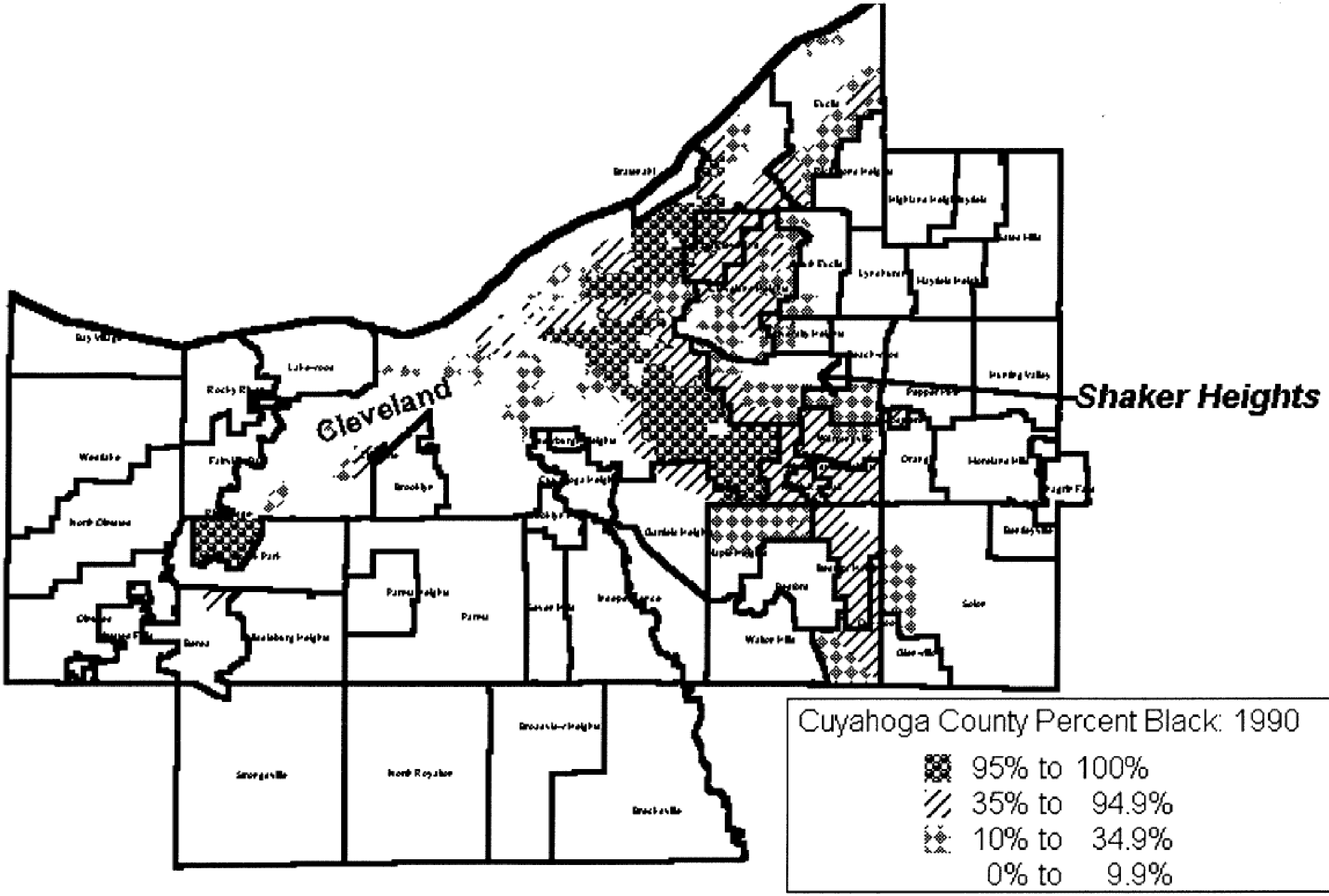
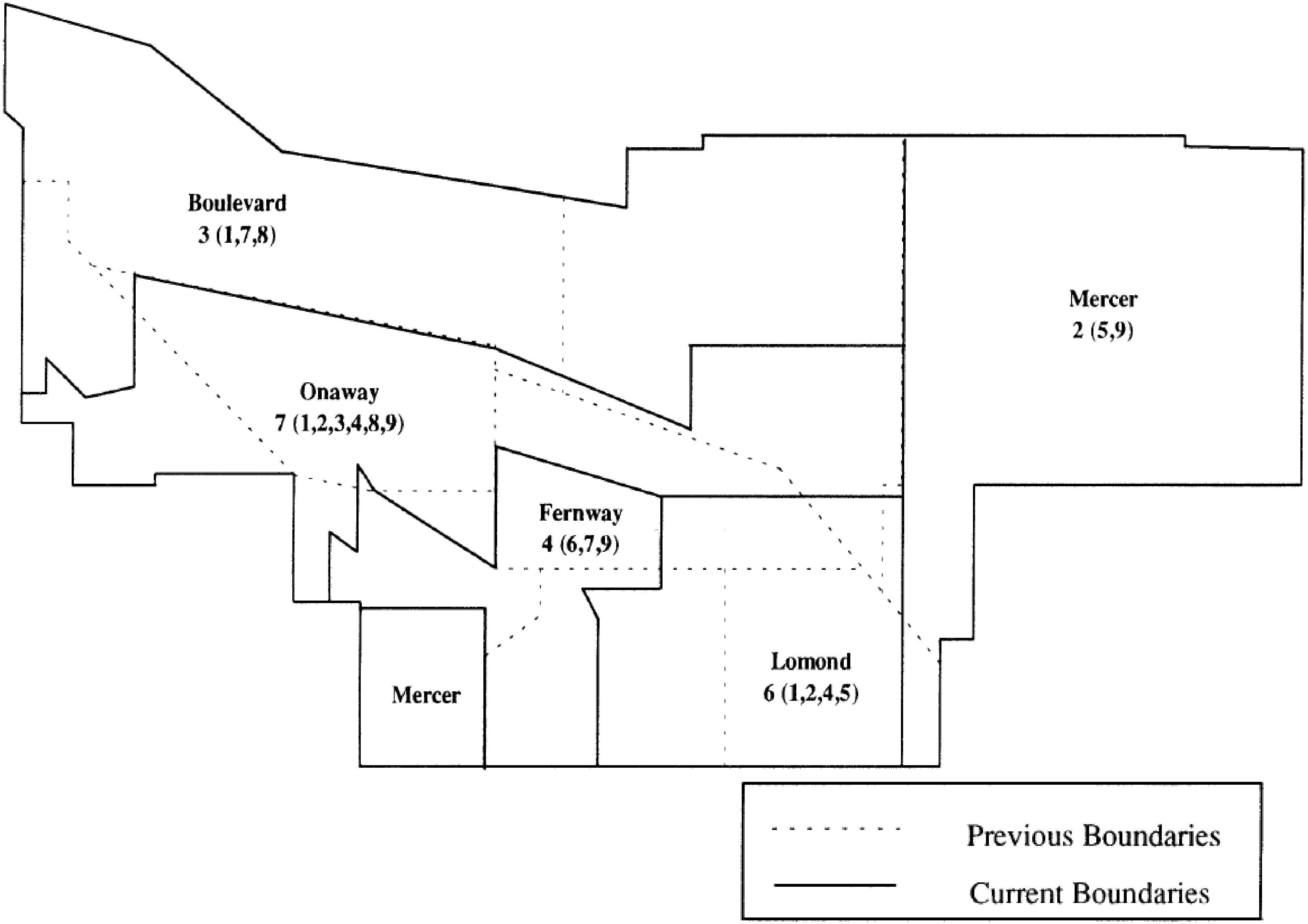


FIG. 1. Cuyahoga County percent black: 1990 by Transportation Analysis Zone 199
Census data.


Boulevard
3 (1,7,8)
Mercer
2 (5,9)
Onaway
7 (1,2,3,4,8,9)
Fernway
4 (6,7,9)
Mercer
Lomond
6 (1,2,4,5)
Previous Boundaries
Current Boundaries

School Percent Nonwhite

Year → School district ↓	1983–1984	1984–1985	1985–1986	1986–1987	1987–1988	1988–1989
Boulevard	34.5	42.0	35.7	39.6	49.8	52.8
Fernway	23.4	21.6	25.2	22.6	48.2	45.8
Lomond	55.4	56.7	54.9	57.2	42.0	43.6
Mercer	26.8	27.9	28.3	27.9	45.0	45.0
Onaway	34.3	37.1	35.1	31.7	51.4	52.4
Ludlow	53.0	54.6	62.4	65.1		
Malvern	27.7	29.2	26.1	23.6		
Moreland	79.2	84.4	84.9	86.9		
Sussex	41.0	42.1	42.9	44.6		

Year → School district ↓	1989–1990	1990–1991	1991–1992	1992–1993	1993–1994	1994–1995
Boulevard	50.5	49.7	53.5	57.0	58.3	63.0
Fernway	51.6	54.8	57.2	59.9	63.3	59.0
Lomond	41.4	44.4	46.7	45.6	52.4	54.7
Mercer	46.9	50.0	48.3	51.8	50.9	52.7
Onaway	50.0	49.6	48.7	46.3	46.1	52.7

Source: Shaker Heights City School District.

Note. Ludlow, Malvern, Moreland, and Sussex schools were closed in 1987.

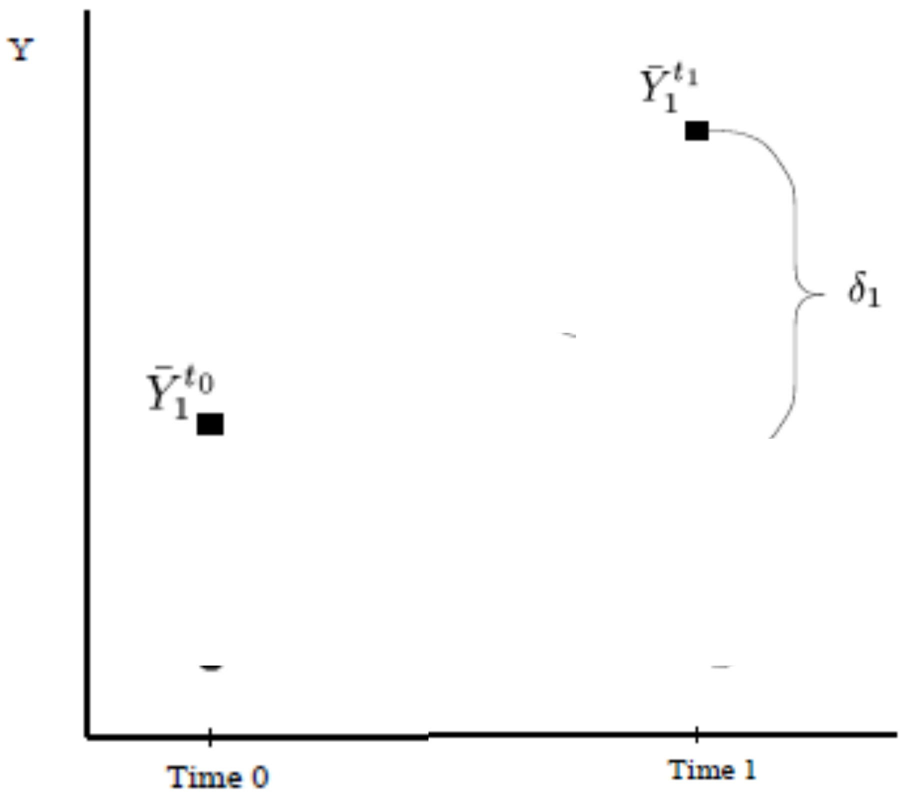
School Quality and Housing Prices

Now that we have a "natural experiment" caused by this shock to the system. How can we use this to get better causal estimates?

Strategy 1

- Strategy 1: pre-post design

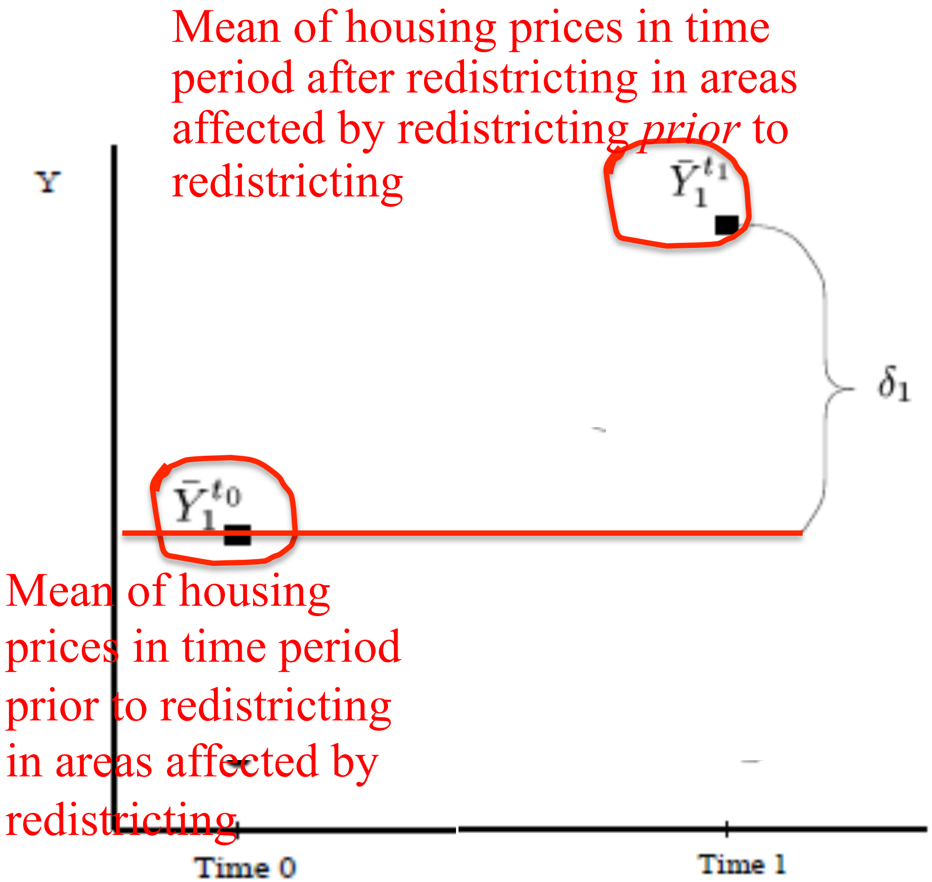
  compare housing prices for affected areas before and after redistricting occurred.



So we'd compare 2 means: one for houses sold in the affected area pre-redistricting, one for sold in the affected area post-redistricting

Strategy 1

- Strategy 1: pre-post design
  Compare housing prices for affected areas before and after redistricting occurred.

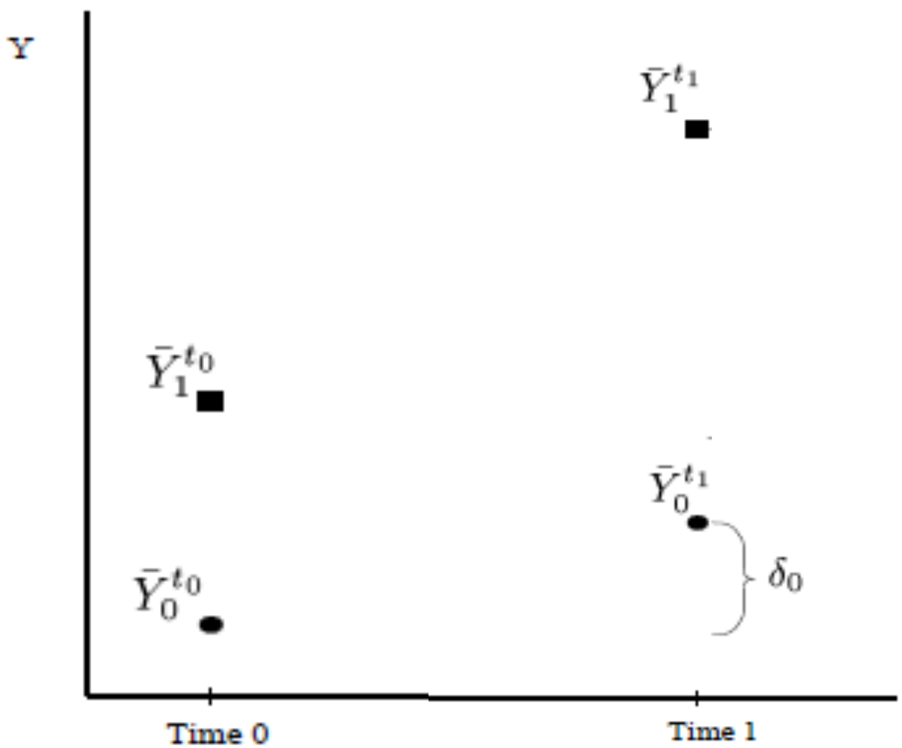


So we'd compare 2 means: one for houses sold in the affected area pre-redistricting, one for sold in the affected area post-redistricting

!!! Pop Quiz !!!

- What is the downside of this strategy?
- What does it assume?
- What is it ignoring?

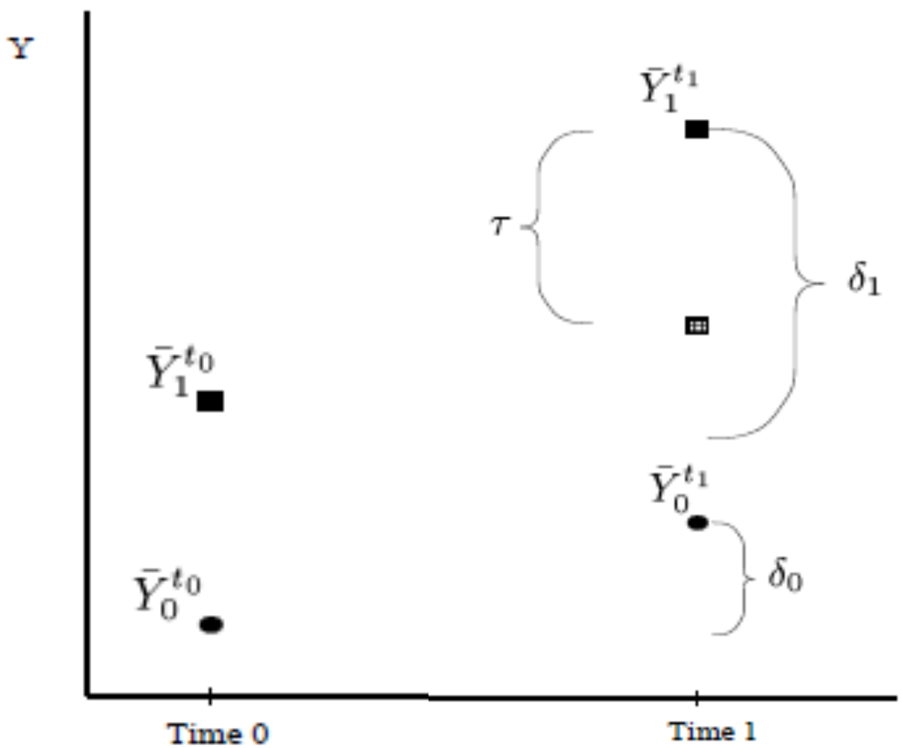
Strategy 1



Problem: that difference in housing prices over time could be a result of many other factors – We are not accounting for the change that might have occurred even in the absence of the program

Strategy 1 → Strategy 3

- Let's consider 4 means: those for each combination of time period (pre-redistricting, post-redistricting) and exposure group (living in affected area, not in affected area).



- Strategy 1 looks at δ_1 and but ignores δ_0 which represents what we think would have been the change for the treatment group if they hadn't been exposed to the treatment.

$$\hat{\tau} = (\overline{Y}_1^{t_1} - \overline{Y}_1^{t_0}) - (\overline{Y}_0^{t_1} - \overline{Y}_0^{t_0})$$

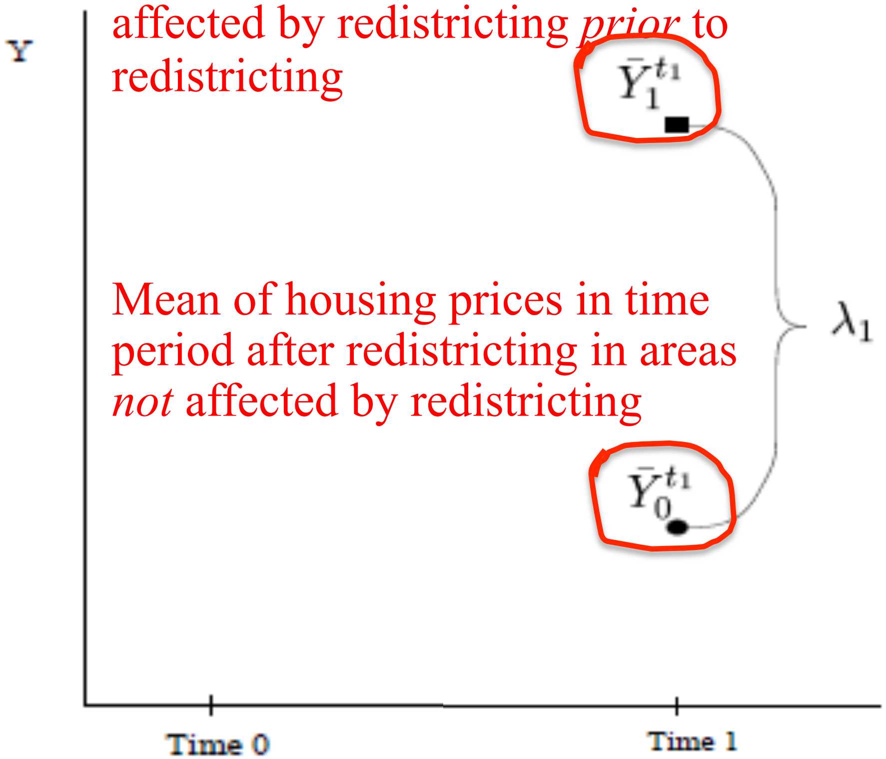
Impact = "treatment group change" – "control group change"

Strategy 2

- Strategy 2: compare housing prices for affected areas after the redistricting with housing prices for unaffected areas after redistricting occurred.

Mean of housing prices in time period after redistricting in areas affected by redistricting *prior* to redistricting

Mean of housing prices in time period after redistricting in areas *not* affected by redistricting

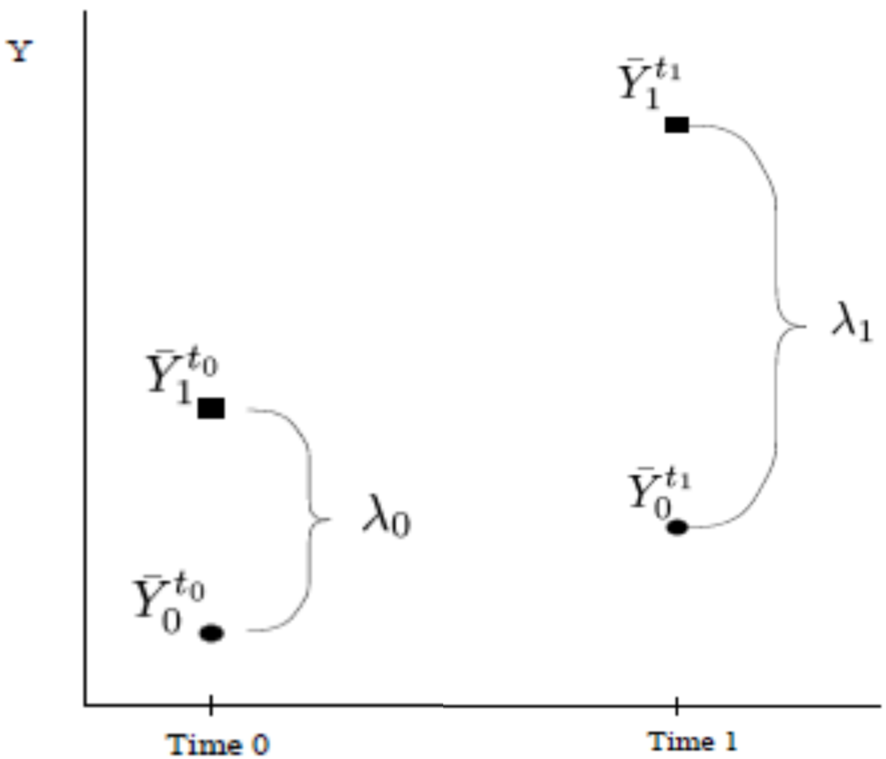


So we would compare 2 sample means: one for houses sold in the affected area post-redistricting, one for the houses sold in the unaffected area post-redistricting

!!! Pop Quiz !!!

- What is the downside of this strategy?
- What does it assume?
- What is it ignoring?

Strategy 2



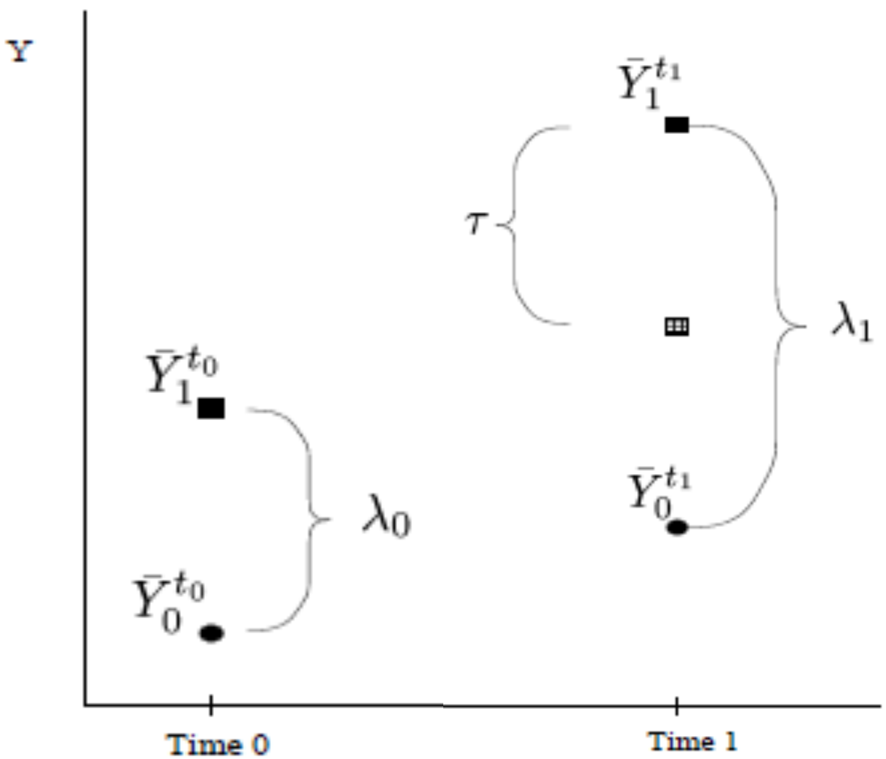
There could be differences in the types of houses affected and unaffected by redistricting – especially since we know the redistricting was motivated for a desire for greater racial integration – these differences in characteristics would be confounded with the effect of redistricting.

!!! Pop Quiz !!!

- What is the downside of this strategy?
- What does it assume?
- What is it ignoring?

Strategy 2 → Strategy 3

- Let's consider 4 means: those for each combination of time period (pre-redistricting, post-districting) and exposure group (living in affected area, not in affected area).



- Strategy 2 looks at λ_1 but ignores λ_0. Instead it could account for the pre-treatment differences between groups by calculating

$$\hat{\tau} = (\overline{Y}_1^{t_1} - \overline{Y}_0^{t_1}) - (\overline{Y}_1^{t_0} - \overline{Y}_0^{t_0})$$

- Impact = "Post-treatment comparison" – "pre-treatment comparison"
- Note that this equation is equivalent to the one in the previous slide. Both represent the difference-in-differences (DID) estimate

Strategy 3

- Strategy 3: compare *changes* in average housing prices across time periods (i.e. pre- vs. post-treatment) in affected groups with changes observed across the same time periods for unaffected groups (so combine strategies 1 and 2 in essence).
- Alternatively this could be described as comparing difference in housing prices across exposure groups at a pre-exposure time point to these same differences at a post-exposure time point

DID Estimation

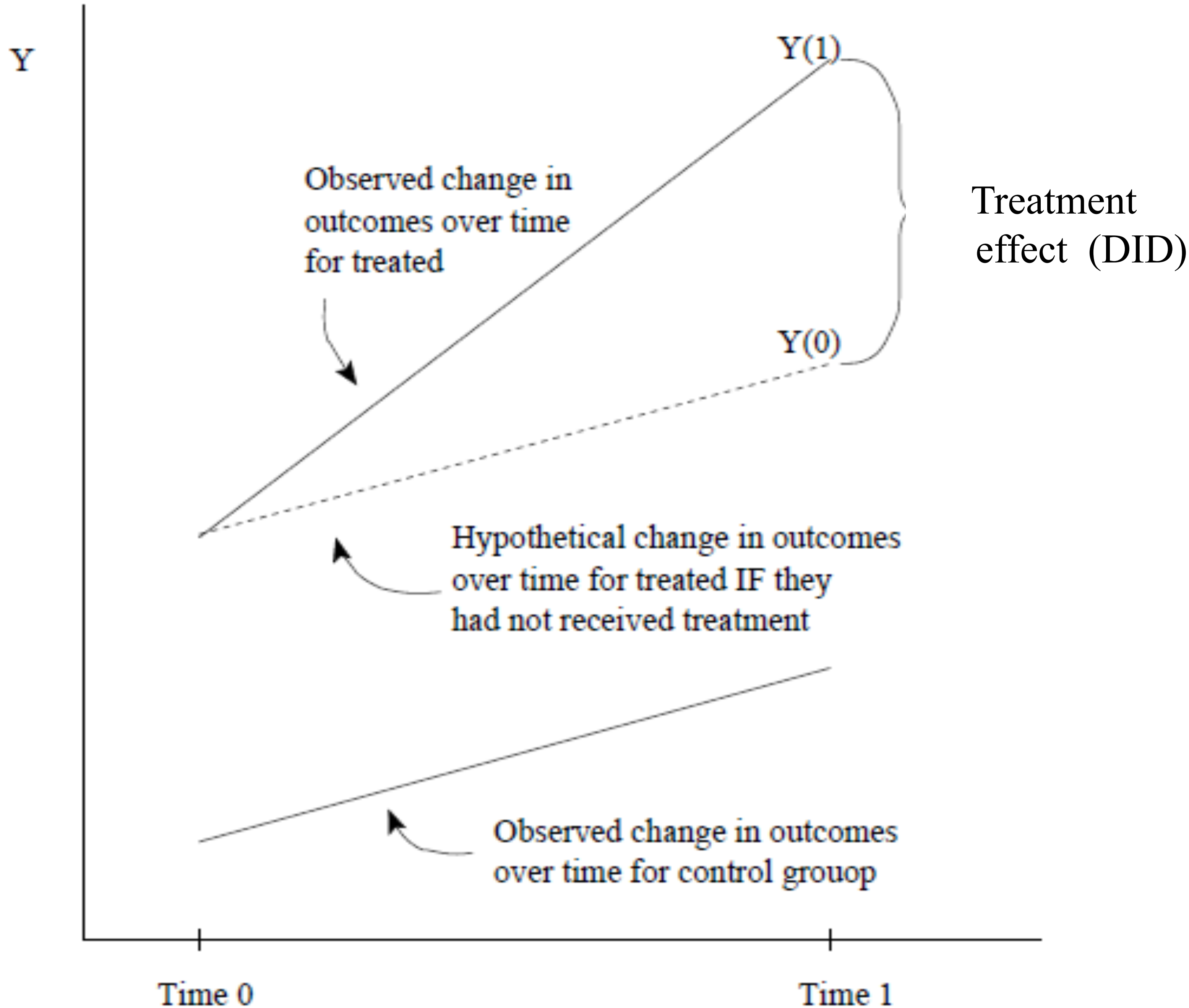
- In a regression framework, we could rewrite as

$$Y_i = \alpha_0 + \lambda_0 Z_i + \delta_0 T_i + \tau \; Z_i T_i + \varepsilon_i$$

Where Z_i denotes exposure group (house i is located in a neighborhood affected by redistricting or not) and T_i denotes time period (pre- and post-redistricting).

Here $\tau = \tau_{DID}$ Why? Here's some intuition:

1. $Z_i T_i$ is an interaction. Interactions measure precisely what we're looking at: the differences in "slopes" (here, difference in means) across two levels of another variable.
2. $Z_i T_i$ is a dummy variable for the only observations (house, people, etc.) that actually get treated.


Y
Y(1)
Observed change in outcomes over time for treated
Treatment effect (DID)
Y(0)
Hypothetical change in outcomes over time for treated IF they had not received treatment
Observed change in outcomes over time for control grouop
Time 0
Time 1

Why does the regression estimate work? Algebra.

- Algebraically we can show that τ is equal to the difference in mean differences equation we wrote above. Recall we want

$$\hat{\tau} = (\overline{Y}_1^{t_1} - \overline{Y}_1^{t_0}) - (\overline{Y}_0^{t_1} - \overline{Y}_0^{t_0})$$

- We can write out the expected value for each of the four populations, using our regression equation:

$$E[Y_1^{t_1}] = \alpha_0 + \lambda_0 + \delta_0 + \tau$$
$$E[Y_1^{t_0}] = \alpha_0 + \lambda_0$$
$$\Rightarrow \quad E[Y_1^{t_1} - Y_1^{t_0}] = \delta_0 + \tau$$

$$E[Y_0^{t_1}] = \alpha_0 + \delta_0$$
$$E[Y_0^{t_0}] = \alpha_0$$
$$\Rightarrow \quad E[Y_0^{t_1} - Y_0^{t_0}] = \delta_0$$

- Thus,

$$\tau_{DID} = E[Y_1^{t_1} - Y_1^{t_0}] - E[Y_0^{t_1} - Y_0^{t_0}] = \tau$$

What assumptions are necessary to believe that this estimate is a causal effect?

The biggest assumption is that the change in mean test scores that the control group experiences over time, reflects the same change that the treatment group would have experienced had they not been exposed to the treatment

→ We can formalize this assumption by defining a potential change for units if had they never been exposed to treatment as

$$D(0) = Y(0) - Y^{t_0}$$

Then perhaps we can assume that $D(0) \perp Z$

Would this be sufficient to identify ATE?

Which causal estimand would it be sufficient to identify?

assumptions (continued)

→ A stronger assumption would be to further assume that

$$D(0), D(1) \perp Z$$

this would identify what causal estimand?

→ The economists formalize the assumptions as

$$\varepsilon_i \perp Z_i, T_i$$

which says that unobserved characteristics have the same distribution across time points and across treatment groups.

In addition, it is often assumed that the treatment effect is constant or additive (constant with noise) across individuals

Summary of Assumptions by estimand

- To identify the average treatment (ATE) we need

  $D(0), D(1) \perp Z$
- To identify the effect of the treatment on the treated (ATT or TOT) we need

  $D(0) \perp Z$

  But we also need to make our sample "look like" the treated (e.g. through matching or weighting)
- To identify the effect of the treatment on the controls (ATC or TOC) we need

  $D(1) \perp Z$

  But we also need to make our sample "look like" the controls (e.g. through matching or weighting)

Back to our example

Regression Results—Difference-in-Difference Estimator

Variable	Coefficient (standard error)
School district change	0.050 (0.016)
Sale in 1987 or later	0.062 (0.013)
School district change *and* sale in 1987 or later	– 0.104 (0.019)
% Nonwhite in school	– 0.051 (0.044)
ln (lot size)	0.223 (0.018)
ln (living area)	0.317 (0.026)
Construction grade AA or A +	0.193 (0.016)
Construction grade A	0.097 (0.019)
Construction grade B or C or D	0.026 (0.010)
ln (age of house)	– 0.079 (0.017)
Bad or fair condition	– 0.083 (0.012)
Excellent condition	0.089 (0.018)
Average room size	0.080 (0.009)
Plumbing fixtures	0.017 (0.002)
Heavy traffic	– 0.220 (0.020)
% Nonwhite in tract 1980	1.329 (0.271)
% Nonwhite in tract 1990	– 1.758 (0.257)
Intercept	6.632(0.228)
Adjusted R^2	0.65
Observations	4463
Dependent variable mean	10.97

Assumptions

- How do you feel about the plausibility of the assumptions in this example?
- Do you have any concerns?

When would we use? What would the data look like?

- This method requires some kind of natural experiment in the form of an "exogenous" shock that acts like a treatment.
- We need data from both exposed (treatment) and unexposed (control) groups in both the time period before the shock and the time period after it.
- An advantage of this method over the standard "change score" format is *we don't have to observe the same observational units at each time point*.
- So we can have standard *panel data* or more general longitudinal data in the form of different cross-sections of data of time.

Repeated cross-section

- Suppose we have data that is a repeated cross-section (i.e. the same people observed in two time periods)
- Then what if we fit the model as shown before (with separate observations for each person year)?

$$Y_{it} = \alpha_0 + \lambda_0 Z_{it} + \delta_0 T_{it} + \tau Z_{it} T_{it} + \varepsilon_{it}$$

- Is there any problem with this statistically?

Repeated cross-section

- Suppose we have data that is a repeated cross-section (i.e. the same people observed in two time periods)
- Then what if we fit the model as shown before (with separate observations for each person year)?

$$Y_{it} = \alpha_0 + \lambda_0 Z_{it} + \delta_0 T_{it} + \tau Z_{it} T_{it} + \varepsilon_{it}$$

- Is there any problem with this statistically?
- What if we instead fit the following model?

$$D_i = Y_{i2} - Y_{i1} = \delta_0 + \tau Z_{i2} + \varepsilon_i$$

Extensions: covariates

- Theoretically, we can include any covariates to make our assumptions seem more plausible. (and also to reduce our standard errors.)
- However it is important, as always, to avoid including covariate values that were measured post-treatment.
  - So, the extension of the change score model presented in the previous slide includes baseline levels of covariates for each unit.

$$D_i = \delta_0 + \tau Z_i + \sum_j \beta_j X_i + \varepsilon_i$$

  - Clearly it will not be possible to do this when our data comprise two distinct cross-sections.

Controlling for post-treatment vars

- Many social scientists who use these sorts of techniques ignore this distinction between of pre- and post-treatment covariates. This is only justifiable when controlling for covariates that we don’t expect to be influenced by the treatment (e.g. large-scale economic indicators when the treatment is a local intervention)
- But often not justifiable.

Regression Results—Difference-in-Difference Estimator

Variable	Coefficient (standard error)
School district change	0.050 (0.016)
Sale in 1987 or later	0.062 (0.013)
School district change *and* sale in 1987 or later	– 0.104 (0.019)
% Nonwhite in school	– 0.051 (0.044)
ln (lot size)	0.223 (0.018)
ln (living area)	0.317 (0.026)
Construction grade AA or A +	0.193 (0.016)
Construction grade A	0.097 (0.019)
Construction grade B or C or D	0.026 (0.010)
ln (age of house)	– 0.079 (0.017)
Bad or fair condition	– 0.083 (0.012)
Excellent condition	0.089 (0.018)
Average room size	0.080 (0.009)
Plumbing fixtures	0.017 (0.002)
Heavy traffic	– 0.220 (0.020)
% Nonwhite in tract 1980	1.329 (0.271)
% Nonwhite in tract 1990	– 1.758 (0.257)
Intercept	6.632(0.228)
Adjusted R^2	0.65
Observations	4463
Dependent variable mean	10.97

DID: more thoughts

Comparison of assumptions

- It seems that we have made substantial gains with little effort -- is this justified.
- Recall that to estimate TOT we need only assume that $D(0) \perp Z$

  (where $D(0) = Y(0) - Y^{t0}$)
- Is this a stronger or weaker assumption than the traditional ignorability assumption that

  $Y(0) \perp Z \mid Y^{t0}$?

Causes for concern?

- When is it a bad idea to assume $D(0) \perp Z$?
- Consider a traditional regression discontinuity set-up where Z is assigned based on X such that

  $Z=0$ for $X<a$

  $Z=1$ for $X>a$

- Would we expect this scenario to lead to bias if we implement DID? In what direction?

  (see Chay and Urquiola for a nice example of this)