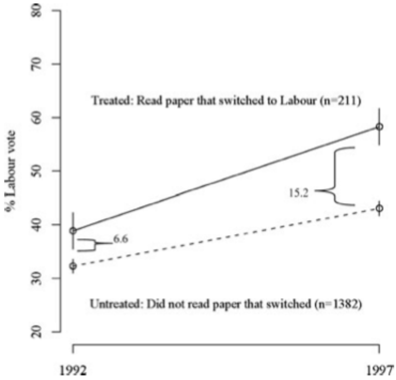
Panel Data

March 31, 2010

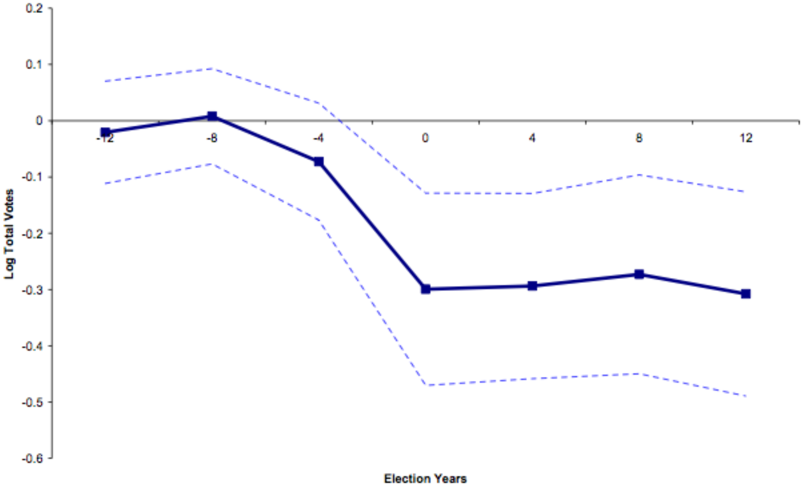
The Effect of the Media

FIGURE 1 Persuasive Effect of Endorsement Changes on Labour Vote Choice between 1992 and 1997



This figure shows that reading a paper that switched to Labour is associated with an (15.2 − 6.6 =) 8.6 percentage point shift to Labour between the 1992 and 1997 UK elections. Paper readership is measured in the 1996 wave, before the papers switched, or, if no 1996 interview was conducted, in an earlier wave. Confidence intervals show one standard error.

The Effect of Jim Crow



Notes: Points are cumulative (4-year) leads and lags (starting from 12 years before) in the dynamic version of the main specification together with 95% confidence intervals.

Time is On Our Side, Yes it Is...

- Notation With Time:
  - Y_{it}^0, Y_{it}^1 are the potential outcomes with a time subscript
  - D_{it} is a time varying dummy variable that indicates whether or not the individual i receives treatment at time t.
  - D_i^* is a time-constant dummy variable that indicates whether individual i ever receives the treatment at any point.
- To estimate causal effects we exploit the distinction between D_{it} (treatment exposure) vs D_i^* (treatment group membership).

	$D_i^* = 0$	$D_i^* = 1$
$t = 0$	Y_{i0}^0	Y_{i0}^0
$t = 1$	Y_{i1}^0	Y_{i1}^1

Table: Potential Outcomes

Counterfactual Possibilities

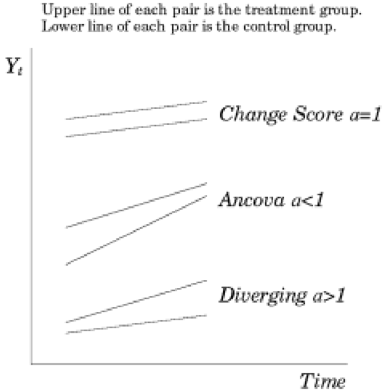


Figure 9.5: Three examples of possible trajectories for the treatment and control groups of the outcome in the absence of treatment.

How to Adjust?

1. *"Change Score"* or *"fixed effects"* :

$$\begin{aligned} E[\delta_{it+1}] &= (E[Y_{it+1}|D_i^* = 1] - E[Y_{it+1}|D_i^* = 0]) \\ &\quad -(E[Y_{it}|D_i^* = 1] - E[Y_{it}|D_i^* = 0]) \end{aligned}$$

2. *"Lagged Dependent Variable"*:

$$\begin{aligned} E[\delta_{it+1}] &= (E[Y_{it+1}|D_i^* = 1] - E[Y_{it+1}|D_i^* = 0]) \\ &\quad -\alpha(E[Y_{it}|D_i^* = 1] - E[Y_{it}|D_i^* = 0]) \end{aligned}$$

where α where α is the correlation between Y_{it+1} and Y_{it+0},

In a Modeling Framework

- *"Change Score"* or *"fixed effects"* causal model:

$$
\begin{aligned}
Y_{it} &= \alpha + \gamma D_i^* + \epsilon_{it} \\
Y_{it+1} &= \alpha + \tau + \gamma D_i^* + \delta D_i + \epsilon_{it+1}
\end{aligned}
$$

- τ is change over time that applies to both treatment and control groups. The coefficient γ is the group difference that is constant over time.
- ϵ_{it} can be decomposed into 2 components a follows:

$$
\epsilon_{it} = U_i + V_{it}
$$

  where U_i can be thought of including all explanatory variables that are stable across time and *that have identical effects at both times*. V_{it} captures variation in explanatory variables that changes over time.

Identification in the Modeling Framework

- We can estimate the model for time t, but not for time $t+1$ because D_{it}^* and D_{it} are perfectly collinear.
- If we subtract the equation for Y_{it+1} from the equation Y_{it} we get:

$$Y_{it+1} - Y_{it} = \tau + \delta D_i + \epsilon_i^*$$

  where $\epsilon_i^* = \epsilon_{it+1} - \epsilon_{it}$
- The differencing takes care of the invariant U_i component of ϵ_i, but not the time varying V_{it}.
- To identify δ, we need to assume that $E(V_{it}|D_i) = 0$ for all i and t.

Assumptions?

1. *"Change Score"* or *"fixed effects"* : In the absence of treatment, any difference in the expectation of Y for those in treatment and control remains constant over time. In our notation:

$$E[Y^0_{it+1} - Y^0_{it} | D^* = 1] = E[Y^0_{it+1} - Y^0_{it} | D^* = 0]$$

2. *"Lagged Dependent Variable"*: In the absence of treatment, any difference between the expectations of Y for those in the treatment and control groups shrink by a multiplicative factor α between each time period.

Semi-parametric Diff-in-Diff

- In the typical parametric setup, we can't include time-invariant covariates in a fixed effect model, as they are "differenced-out". So does that mean that traits like race, gender, and geography are irrelevant?
- Not if we think that differences in invariant omitted variables is an indication that the assumption of no differential time trends across treatment and control is likely to hold.
- How to deal with this? We might be willing to make the weaker assumption that:

$$E[Y^0_{it+1} - Y^0_{it} | D^* = 1, X_i] = E[Y^0_{it+1} - Y^0_{it} | D^* = 0, X_i]$$

- In words, conditional on time invariant covariates, we assume that the difference in the expectation of Y for those in treatment and control remains constant over time.
- There are a variety of ways to implement this, including matching before estimation (using Genmatch or some other method) or propensity score weighting.

Testing Assumptions

- Given that the key assumption for difference-in-differences models is no differential time trends, the best validity check is to use pre-treatment outcome data to test whether this is true before treatment is applied.
- Of course, this requires more than one period of pre-treatment outcome data. One simply estimates the following:

$$\begin{aligned} E[\delta_{it}] &= (E[Y_{it}|D_i^* = 1] - E[Y_{it}|D_i^* = 0]) \\ &\quad -(E[Y_{it-1}|D_i^* = 1] - E[Y_{it-1}|D_i^* = 0]) \end{aligned}$$

- If that quantity does not equal 0, then you likely have a problem. Of course, even if your design passes this placebo test, it's possible that differential trends not due to treatment begin after the treatment is applied.

Green and Middleton (2008)

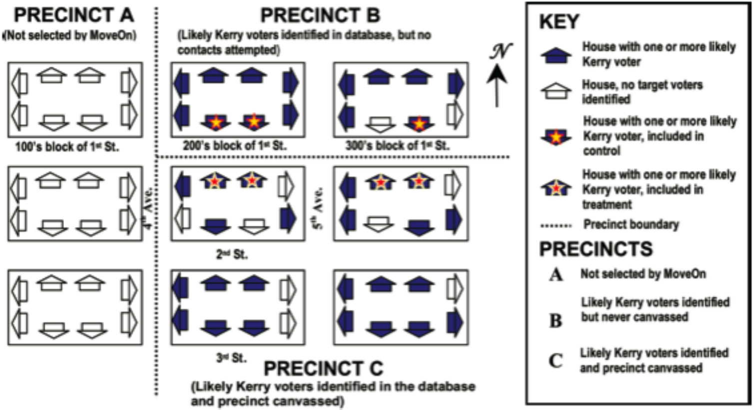


Figure: Green and Middleton (2009)