Bacon decomposition for understanding differences-in-differences with variation in treatment timing

July 11, 2019

Stata Conference

Andrew Goodman-Bacon (Vanderbilt University)
Austin Nichols (Abt Associates)
Thomas Goldring (University of Michigan)
Overview

• In canonical difference-in-differences (DD), the regression version = function of pre/post and treat/control means.

• When treatment turns on at different times, the regression DD coefficient is a weighted average of canonical “2x2” DDs (Goodman-Bacon 2018)
  – Shows where such DDs “come from”

• This command calculates the component DDs and their weights, plots them (ie. shows variation), compares specifications
  – Future: conducts balance tests, analyzes estimand,
What is Difference-in-Differences?
What is Difference-in-Differences?
What is Difference-in-Differences?

\[ \bar{y}_{POST} \]

\[ \bar{y}_{PRE} \]

\[ \bar{y}_{TREAT} \]

Units of \( y \)

0  10  20  30  40  

Time

\( t^*_k \)

\( y_{it} \)

\( y_{iu} \)

\( y^k \)

\( \Delta y_{PER} \)

\( \Delta y_{POST} \)
What is Difference-in-Differences?

$$(\bar{y}_{POST}^{TREAT} - \bar{y}_{PRE}^{TREAT})$$
What is Difference-in-Differences?

$\left( \bar{y}_{\text{TREAT}}^{\text{POST}} - \bar{y}_{\text{TREAT}}^{\text{PRE}} \right)$
What is Difference-in-Differences?

\[
(\bar{y}_{POST}^{TREAT} - \bar{y}_{PRE}^{TREAT})
\]
What is Difference-in-Differences?

\[(\bar{y}_{POST}^{TREAT} - \bar{y}_{PRE}^{TREAT}) \]

\[(\bar{y}_{POST}^{CONTROL} - \bar{y}_{PRE}^{CONTROL}) \]

Units of \(y\)

Time

\(t^*_k\)
What is Difference-in-Differences?

\[
\hat{\beta}_{DD} = (\bar{y}_{POST}^TREAT - \bar{y}_{PRE}^TREAT) - (\bar{y}_{POST}^{CONTROL} - \bar{y}_{PRE}^{CONTROL})
\]
What is Difference-in-Differences?

Wooldridge (2002):

\[
\hat{\delta}_1 = (\bar{y}_{B,2} - \bar{y}_{B,1}) - (\bar{y}_{A,2} - \bar{y}_{A,1})
\]

This estimator has been labeled the difference-in-differences (DID) estimator in the recent program evaluation literature, although it has a long history in analysis of variance.
What is Difference-in-Differences?

Cameron and Trivedi (2007):

Then the OLS estimator reduces to

$$\hat{\phi} = \Delta \bar{y}^{tr} - \Delta \bar{y}^{nt}. \quad (22.43)$$

This estimator is called the differences-in-differences (DID) estimator, since one estimates the time difference for the treated and untreated groups and then takes the difference in the time differences.
What is Difference-in-Differences?

Angrist and Pischke (2009):

The population difference-in-differences,

\[
\left\{ E[Y_{ist} | s = NJ, t = Nov] - E[Y_{ist} | s = NJ, t = Feb] \right\} \\
- \left( E[Y_{ist} | s = PA, t = Nov] - E[Y_{ist} | s = PA, t = Feb] \right) = \delta,
\]
What is Difference-in-Differences?

Imbens and Wooldridge (2007):

for those observations in the treatment group in the second period. The difference-in-differences estimate is

$$\hat{\delta}_1 = (\bar{y}_{B,2} - \bar{y}_{B,1}) - (\bar{y}_{A,2} - \bar{y}_{A,1}).$$  \hspace{1cm} (1.2)
What is Difference-in-Differences?

Angrist and Krueger (1999):

\[
\{E[Y_i | c = \text{Miami}, t = 1981] - E[Y_i | c = \text{Comparison}, t = 1981]\} 
- \{E[Y_i | c = \text{Miami}, t = 1979] - E[Y_i | c = \text{Comparison}, t = 1979]\} = \delta. \tag{21}
\]
What is Difference-in-Differences?

Heckman, LaLonde and Smith (1999):

then the difference-in-differences estimator given by

\[(\bar{Y}_{1t} - \bar{Y}_{0t'})_1 - (\bar{Y}_{0t} - \bar{Y}_{0t'})_0, \quad t > k > t'\]
What is Difference-in-Differences?

Meyer (1995):

In this case, and unbiased estimate of $\beta$ can be obtained by difference in differences as

$$\hat{\beta}_{dd} = \Delta \bar{y}_0^1 - \Delta \bar{y}_0^0$$

$$= \bar{y}_1^1 - \bar{y}_0^1 - (\bar{y}_1^0 - \bar{y}_0^0),$$

(4)
What is Difference-in-Differences?

Abadie (2005):

$$D(i, 1) = 1,$$ and the individual-specific component, $$\eta(i)$$. This model is called “difference-in-differences” because under the identifying condition in equation (2) we have

$$\alpha = \{E[Y(i, 1) \mid D(i, 1) = 1] - E[Y(i, 1) \mid D(i, 1) = 0]\}
- \{E[Y(i, 0) \mid D(i, 1) = 1] - E[Y(i, 0) \mid D(i, 1) = 0]\}, \quad (5)$$
What is Difference-in-Differences?

Athey and Imbens (2006):

\[ \tau_{\text{DID}} = \left[ \mathbb{E}[Y_i | G_i = 1, T_i = 1] - \mathbb{E}[Y_i | G_i = 1, T_i = 0] \right] \\
- \left[ \mathbb{E}[Y_i | G_i = 0, T_i = 1] - \mathbb{E}[Y_i | G_i = 0, T_i = 0] \right]. \]

i.e., \( \varepsilon_i \perp (G_i, T_i) \), and is normalized to have mean zero. The standard DID estimand is
What is Difference-in-Differences?

DiNardo and Lee (2011):

First, let us simplify the problem by considering the situation where the program was made available at only one point in time $\tau$. This allows us to define $D = 1$ as those who were treated at time $\tau$, and $D = 0$ as those who did not take up the program at that time.
Keywords in NBER Papers Since 2012

Snow (1855) 430
Fisher (1926) 360
Thistlewaite and Campbell (1960) 277
Wright (1928), Wald (1940), Durbin (1954) 481

"difference-in-differences" | "diff-in-diff"
"randomization" & "randomize"
"regression discontinuity" | "regression kink"
"instrumental variables" & "instruments"
Keywords in NBER Papers Since 2012

The last arrow in the quasi-experimental quiver is differences-in-differences, probably the most widely applicable design-based estimator.

- Snow (1855)
- Fisher (1926)
- Thistlewaite And Campbell (1960)
- Wright (1928), Wald (1940), Durbin (1954)
Keywords in NBER Papers Since 2012

2014/2015 AER/QJE/JPE/ReStud/JHE/JDE published 93 DD papers:
49% had timing variation

Snow (1855)
Fisher (1926)
Thistlewaite and Campbell (1960)
Wright (1928), Wald (1940), Durbin (1954)
Variation in Timing

\[ y_u^k \]

\[ y_u^l \]

\[ y_u^U \]

\[ \text{PRE}(k) \]

\[ \text{MID}(k,l) \]

\[ \text{POST}(l) \]

Time

\[ t^*_k \]

\[ t^*_l \]
**Variation in Timing**

$D_{it}$ turns on at different times $t^*$

(Federalism, judicial enforcement, sub-federal funding process, natural disasters, mass layoffs…)

![Graph showing variation in timing with time periods PRE(k), MID(k,l), and POST(l)]
Two-Way Fixed Effects Estimator

\[ y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it} \]

- Unit fixed effects
- Time fixed effects
- Treatment dummy

What is $\hat{\beta}^{DD}$?
This has been unclear, but we do have good intuition for subsamples here where one group’s treatment changes and another’s does not.
A. Early Group vs. Untreated Group
\[ \hat{\beta}_{kU}^{2x2} = \left( \bar{y}_k^{\text{POST}(k)} - \bar{y}_k^{\text{PRE}(k)} \right) - \left( \bar{y}_U^{\text{POST}(k)} - \bar{y}_U^{\text{PRE}(k)} \right) \]
\[ \hat{\beta}_{\ell U}^{2 \times 2} \]

**B. Late Group vs. Untreated Group**

The figure depicts a comparison between the Late Group and the Untreated Group over time. The graph shows the units of \( y \) on the y-axis and time on the x-axis. The data points for each group are represented by different markers, with the Late Group shown by black circles and the Untreated Group by grey triangles. The graph includes a line indicating the trend for each group, with the Late Group showing an upward trend, while the Untreated Group remains relatively stable. The time intervals are marked as PRE(\( l \)) and POST(\( l \)).
$$\hat{\beta}^{2x2}_{\ell U} = \left( \bar{y}^{\text{POST}(\ell)}_{\ell} - \bar{y}^{\text{PRE}(\ell)}_{\ell} \right) - \left( \bar{y}^{\text{POST}(\ell)}_{U} - \bar{y}^{\text{PRE}(\ell)}_{U} \right)$$
\( \hat{\beta}_{2x2,k}^{2x2,k} \)

C. Early Group vs. Late Group, before \( t^*_{k} \)

- Treatment
- Control

\( y^k_i \)

\( y^l_i \)

PRE(k) \( \rightarrow \) MID(k,l)
\[ \hat{\beta}_{2x2,k} = \left( \bar{y}_{k}^{\text{MID}(k,\ell)} - \bar{y}_{k}^{\text{PRE}(k)} \right) - \left( \bar{y}_{\ell}^{\text{MID}(k,\ell)} - \bar{y}_{\ell}^{\text{PRE}(k)} \right) \]
D. Late Group vs. Early Group, after $t^*_k$
\[ \hat{\beta}_{k\ell}^{2x2} = \left( \overline{y}_{\ell}^{POST} - \overline{y}_{\ell}^{MID(k,\ell)} \right) - \left( \overline{y}_{k}^{POST} - \overline{y}_{k}^{MID(k,\ell)} \right) \]
Difference-in-Differences Decomposition Theorem (3 Group Case)

\[ y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it} \]

For three groups:

\[ \hat{\beta}^{DD} = s_{kU} \hat{\beta}_{kU}^{2x^2} + s_{\ell U} \hat{\beta}_{\ell U}^{2x^2} + s_{k\ell}^{k} \hat{\beta}_{k\ell}^{2x^2,k} + s_{k\ell}^{k} \hat{\beta}_{k\ell}^{2x^2,\ell} \]
Difference-in-Differences Decomposition Theorem (3 Group Case)

\[ y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it} \]

For three groups:

\[ \hat{\beta}^{DD} = s_{kU} \hat{\beta}_{kU}^{2x2} + s_{\ell U} \hat{\beta}_{\ell U}^{2x2} + s_{k\ell} \hat{\beta}_{k\ell}^{2x2, k} + s_{k\ell} \hat{\beta}_{k\ell}^{2x2, \ell} \]
Difference-in-Differences Decomposition Theorem (3 Group Case)

\[ y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it} \]

For three groups:

\[ \hat{\beta}^{DD} = s_{kU} \hat{\beta}_{kU}^{2x2} + s_{\ell U} \hat{\beta}_{\ell U}^{2x2} + s_{k\ell} \hat{\beta}_{k\ell}^{2x2,k} + s_{k\ell} \hat{\beta}_{k\ell}^{2x2,\ell} \]

2x2 DDs: subsamples with two groups (treat/control) and two periods (pre/post)
What do we learn from the 2x2 DDs?

1. We didn’t know what comparisons were being made:
   “switchers vs untreated”?
   “early vs late”?
   “late vs early” (this is less obvious)?

   *It’s all of those.*

2. “What is the control group?”

   *Every group acts as a control (sometimes).*

3. Clarifies theory

   *We understand the estimand (ATET) and ID assumption (common trends) for each 2x2; late vs. early comparisons are biased if effects vary over time.*
Difference-in-Differences Decomposition Theorem (3 Group Case)

\[ y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it} \]

For three groups:

\[ \hat{\beta}^{DD} = s_{kU} \hat{\beta}_{kU}^{2x2} + s_{\ell U} \hat{\beta}_{\ell U}^{2x2} + s_{k\ell}^k \hat{\beta}_{k\ell}^{2x2,k} + s_{k\ell}^\ell \hat{\beta}_{k\ell}^{2x2,\ell} \]
Difference-in-Differences Decomposition Theorem (3 Group Case)

\[ y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it} \]

For three groups:

\[ \hat{\beta}^{DD} = \sum k U \hat{\beta}^{2x2}_{kU} + \sum \ell U \hat{\beta}^{2x2}_{\ell U} + \sum k \ell \hat{\beta}^{2x2,k\ell} + \sum k \ell \hat{\beta}^{2x2,\ell k} \]

A. Early Group vs. Untreated Group

Size: \((n_k + n_U)^2\) 

\[ n_{kU}(1 - n_{kU})D_k(1 - \bar{D}_k) \]

B. Late Group vs. Untreated Group

Size: \((n_\ell + n_U)^2\) 

\[ n_{\ell U}(1 - n_{\ell U})\bar{D}_\ell(1 - \bar{D}_\ell) \]

C. Early Group vs. Late Group, before \(t^*\),

Size: \(((n_k + n_\ell)(1 - \bar{D}_\ell))^2\) 

\[ n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell 1 - \bar{D}_k}{1 - \bar{D}_\ell 1 - \bar{D}_\ell} \]

D. Late Group vs. Early Group, after \(t^*\),

Size: \(((n_k + n_\ell)\bar{D}_k)^2\) 

\[ n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_k 1 - \bar{D}_k}{\bar{D}_k - \bar{D}_k} \]

\[ \text{Variance:} \]

\[ n_{kU}(1 - n_{kU})D_k(1 - \bar{D}_k) \]

\[ n_{\ell U}(1 - n_{\ell U})\bar{D}_\ell(1 - \bar{D}_\ell) \]

\[ n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell 1 - \bar{D}_k}{1 - \bar{D}_\ell 1 - \bar{D}_\ell} \]

\[ n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_k 1 - \bar{D}_k}{\bar{D}_k - \bar{D}_k} \]
\[ y_{it} = \alpha_i + \alpha_t + \beta_{DD} D_{it} + u_{it} \]

For three groups:

\[ \hat{\beta}_{DD} = S_{kU} \hat{\beta}_{kU}^{2x2} + S_{\ell U} \hat{\beta}_{\ell U}^{2x2} + S_{k\ell} \hat{\beta}_{k\ell}^{2x2,k} + S_{\ell k} \hat{\beta}_{\ell k}^{2x2,\ell} \]

**A. Early Group vs. Untreated Group**
- Size: \((n_k + n_U)^2 \times \)
- Variance: \(n_{kU}(1 - n_{kU})\bar{D}_k(1 - \bar{D}_k)\)

**B. Late Group vs. Untreated Group**
- Size: \((n_\ell + n_U)^2 \times \)
- Variance: \(n_{\ell U}(1 - n_{\ell U})\bar{D}_\ell(1 - \bar{D}_\ell)\)

**C. Early Group vs. Late Group, before \(t^*\)**
- Size: \(((n_k + n_\ell)(1 - \bar{D}_\ell))^2 \times \)
- Variance: \(n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k - \bar{D}_\ell} \frac{1 - \bar{D}_k}{1 - \bar{D}_\ell} \frac{1 - \bar{D}_\ell}{1 - \bar{D}_\ell} \)

**D. Late Group vs. Early Group, after \(t^*\)**
- Size: \(((n_k + n_\ell)\bar{D}_k)^2 \times \)
- Variance: \(n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k}{\bar{D}_k} \frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k} \frac{1 - \bar{D}_k}{1 - \bar{D}_\ell} \frac{1 - \bar{D}_\ell}{1 - \bar{D}_\ell} \)

**Weights:** \(\frac{(subsample\ share)^2}{(subsample\ variance\ of\ FE-adj\ D)}\)

\(total\ variance\ of\ FE-adj\ D\)
What do we learn from the weights?

1. Relative importance of each kind of comparison.
   “switchers vs untreated”?
   “early vs late”?
   “late vs early” (this is less obvious)?
   More important if big group (bigger sample size) or treated closer to middle of the panel (bigger variance).
   “How much” comes from timing vs comparisons to untreated.

2. Importance of specific 2x2 DDs.
   Sometimes a few terms dominate.

3. Clarifies theory
   The estimand and ID assumption are “variance weighted”; can compare estimand to “parameters of interest” and conduct a proper balance test.
What will the command do?

- Describe where the DD “comes from”
  - Which 2x2s matter most? (sources of variation)
  - How different are the 2x2 DDs? (heterogeneity)
Replication: The Effect of Unilateral Divorce on Suicide (Stevenson and Wolfers 2006)

State-year panel of female suicide rates 1964-1996

<table>
<thead>
<tr>
<th>Year ($t_k^*$)</th>
<th>Number of States</th>
<th>Share of States ($n_k$)</th>
<th>Treatment Share ($\bar{D}_k$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Reform States</td>
<td>5</td>
<td>0.10</td>
<td>.</td>
</tr>
<tr>
<td>Pre-64 Reform States</td>
<td>8</td>
<td>0.16</td>
<td>.</td>
</tr>
<tr>
<td>1969</td>
<td>2</td>
<td>0.04</td>
<td>0.85</td>
</tr>
<tr>
<td>1970</td>
<td>2</td>
<td>0.04</td>
<td>0.82</td>
</tr>
<tr>
<td>1971</td>
<td>7</td>
<td>0.14</td>
<td>0.79</td>
</tr>
<tr>
<td>1972</td>
<td>3</td>
<td>0.06</td>
<td>0.76</td>
</tr>
<tr>
<td>1973</td>
<td>10</td>
<td>0.20</td>
<td>0.73</td>
</tr>
<tr>
<td>1974</td>
<td>3</td>
<td>0.06</td>
<td>0.70</td>
</tr>
<tr>
<td>1975</td>
<td>2</td>
<td>0.04</td>
<td>0.67</td>
</tr>
<tr>
<td>1976</td>
<td>1</td>
<td>0.02</td>
<td>0.64</td>
</tr>
<tr>
<td>1977</td>
<td>3</td>
<td>0.06</td>
<td>0.61</td>
</tr>
<tr>
<td>1980</td>
<td>1</td>
<td>0.02</td>
<td>0.52</td>
</tr>
<tr>
<td>1984</td>
<td>1</td>
<td>0.02</td>
<td>0.39</td>
</tr>
<tr>
<td>1985</td>
<td>1</td>
<td>0.02</td>
<td>0.36</td>
</tr>
</tbody>
</table>

12 timing groups vs Non-reform: 12 2x2 DDs
12 timing groups vs Pre-64 reform: 12 2x2 DDs
12 timing groups vs 12 timing groups: 12×11 = 132 2x2 DDs
Graphing the Decomposition: Divorce Example

- Later Group Treatment vs. Earlier Group Control
  Weight = 0.26; DD = 3.51

- Treatment vs. Non-Reform States
  Weight = 0.24; DD = -5.33

- Treatment vs. Pre-1964 Reform States
  Weight = 0.38; DD = -7.04

- Earlier Group Treatment vs. Later Group Control
  Weight = 0.11; DD = -0.19

\[ DD \text{ Estimate} = -3.08 \]
What will the command do?

• Describe where the DD “comes from”
  – Which 2x2s matter most? (sources of variation)
  – How different are the 2x2 DDs? (heterogeneity)

• Calculate why estimates differ across specifications:
  – Is it the weights, the 2x2 DDs, or both?
Comparing two weighted averages

\[ \hat{\beta}^{DD} = s' \hat{\beta}^{2x2} \]

Now imagine an alternative specification that also has this form:

\[ \hat{\beta}_{alt}^{DD} = s_{alt}' \hat{\beta}_{alt}^{2x2} \]

If \( \hat{\beta}_{alt}^{DD} \neq \hat{\beta}^{DD} \), why? (Oaxaca/Blinder/Kitagawa decomposition)

\[
\frac{s'(\hat{\beta}_{alt}^{2x2} - \hat{\beta}^{2x2})}{s_{alt}} + \frac{(s'_{alt} - s')\hat{\beta}^{2x2}}{s_{alt}} + \frac{(s'_{alt} - s')(\hat{\beta}_{alt}^{2x2} - \hat{\beta}^{2x2})}{s_{alt}}
\]

Note (not for today): Goodman-Bacon (2018) now analyzes models with (any) controls, with an additional important nuance.
Plotting components: WLS vs. OLS

WLS = -0.35

OLS = -3.08

1970 States are the Control

1970 States are the Treatment
Plotting components: WLS vs. OLS

1970 States are the Control
1970 States are the Treatment

WLS = -0.35
OLS = -3.08

53% from 2x2 DDs
38% from weights
9% from interaction

Weighted 2x2 DDs
Unweighted 2x2 DDs
Why does WLS affect the 1970 states so much?

Only two states did no-fault divorce in 1970: Iowa and California.

CA has a huge downward trend and it matters a lot more in WLS than OLS.
Conclusion

• When treatment timing varies, the (two-way fixed effects) regression DD coefficient is a weighted average of simple 2x2 DDs (Goodman-Bacon 2018)

• This command will plot the 2x2 DDs against their weight to highlight where identification comes from and how heterogeneous are the 2x2 DDs.
  – “How much” variation comes from timing?
  – What is “the” control group?
  – Weights do NOT rely on outcome data (can apply to it to samples you don’t yet have)

• This command allows users to analyze why estimates change under different specifications (e.g. weights, controls, triple-diff)

• Future: test covariate balance (accounting for timing), compare estimand to other parameters of interest, adjust for bias from time-varying effects.
References


Textbooks and Survey Articles that describe 2x2 DD:
Recent Research on DD with Timing


