Lavy: “Performance Pay and Teachers’ Productivity”

Policy question: *should* we use teacher bonuses to increase student performance?
- **pro:** creates greater incentives for teachers to work hard
  bonuses for individuals will reduce “free riding”
  (a problem with group bonuses)
- **con:** makes pay riskier, which employees don’t like, so pay **rate** will have to go up
  employees will emphasize only the behaviors that generate bonuses
  (e.g., “teaching to the test”)
  much work is complex, requires a lot of teamwork/cooperation,
  group can monitor behavior of individuals, (so use group bonuses?)

Use *tournaments* (based on rank order of finish, not absolute amount)?
- announce prizes in advance (easy to stay within budget)
- each employee has incentive to work incredibly hard
  (if employees are basically identical, huge incentive to exert greater effort)
  problem: sabotage?

Empirical question: *will* bonuses to teacher increase student performance?
Background:

high schools assigned to an individual-performance bonus experiment (bonus to a teacher based on ranking of his/her students’ test scores) assignment of school based on the school’s “matriculation rate” (“matriculation rate” = percent passing national exams) school assigned to the experiment if its matriculation rate was below 45%

BUT: “official” matriculation rate $S$ was based on incomplete data “true” matriculation rate $S^*$ only became known after schools were selected for the experiment $S = S^* + \varepsilon$, where $\varepsilon$ is measurement error

key insight: conditional on $S^*$, $\varepsilon$ is random! thus, conditional on $S^*$, selection of schools and students into the experiment is random thus, control for $S^*$ and pick schools with $S < 45$ not in the experiment these schools form a “control group” that can be compared with schools with the same $S^*$ that were in the experiment (“experimentals”)

effectively, we have random selection of experimental and controls (conditional on $S^*$, the “true” matriculation score)
basic model (data for two years: pre-experiment in 2000, experiment in 2001)

\[ Y_{ijt} = \alpha + X_{ijt} \beta + Z_{jt} \gamma + \delta T_{jt} + \phi_{ijt} + \eta D_t + \varepsilon_{ijt} \]

where \( Y_{ijt} \) = score of student \( i \) in school \( j \) at time \( t \)
\( X_{ijt} \) = characteristics of student \( i \) in school \( j \) at time \( t \)
\( Z_{jt} \) = characteristics of school \( j \) at time \( t \)
\( T_{jt} \) = interaction: dummy variable for “treated school”
\( \phi_{ijt} \) = set of dummy variables for school (\( j = 1, 2, \ldots \))
\( D_t \) = dummy variable for “year is 2001” (= year of experiment)

→ The estimated coefficient on \( T \), namely \( \delta \), is the “difference-in-difference” estimate of the effect of the experiment on student scores.
First consider the difference for the experimental schools between 2000 (no experiment) and 2001 (with experiment): improvement could have been due to unmeasured things that changed between 2000 and 2001, rather than due to the experiment itself.

(Beware the “post hoc” fallacy – “post hoc, ergo propter hoc.”)

Likewise, consider the difference for the control schools between 2000 and 2001: again, any change could have been due to unmeasured things that changed between 2000 and 2001, rather than due to the experiment itself.

But why is \( \delta \) an estimate of the effect of the experiment?

To simplify, assume that there is only one experimental school and one control school. (Argument doesn’t depend on this – allowing for multiple schools just makes the math more messy.) Likewise, ignore differences in X, Z, etc. (Again, argument doesn’t depend on this.)
So the equation is now $Y_{ijt} = \alpha + \delta T_{jt} + \phi_j + \eta D_t + \varepsilon_{ijt}$

where $T_{jt} = 1$ for the experimental school in 2001, and 0 otherwise;
$\phi_j = 1$ for the experimental school, 0 for the control
$D_t = 1$ for the year 2001, and 0 for 2000

For the experimental school, the *difference* between 2001 and 2000 is

$Y_{i11} = \alpha + \delta \times 1 + \phi_1 + \eta \times 1 + \varepsilon_{i11}$

$Y_{i10} = \alpha + \delta \times 0 + \phi_1 + \eta \times 0 + \varepsilon_{i10}$

$\therefore \quad Y_{i11} - Y_{i10} = \delta + \eta + \varepsilon_{i11} - \varepsilon_{i10}$

Likewise, for the control school, the 2001-2000 difference is

$Y_{i11} = \alpha + \delta \times 0 + \phi_0 + \eta \times 1 + \varepsilon_{i01}$

$Y_{i10} = \alpha + \delta \times 0 + \phi_0 + \eta \times 0 + \varepsilon_{i00}$

$\therefore \quad Y_{i01} - Y_{i00} = \eta + \varepsilon_{i01} - \varepsilon_{i00}$
So, finally, we have:

\[ Y_{i11} - Y_{i10} = \delta + \eta + \varepsilon_{i11} - \varepsilon_{i10} \quad \text{experimental difference} \]

\[ Y_{i01} - Y_{i00} = \eta + \varepsilon_{i01} - \varepsilon_{i00} \quad \text{control difference} \]

Provided both types of schools have the same trends over time, the \( \eta \)'s will cancel.

Provided the differences in the \( \varepsilon \)'s are the same, the terms involving the \( \varepsilon \)'s will also cancel.

In this case, the difference \textit{between} the two differences gives the effect of the experiment – and this is equal to \( \delta \)!

\[ [Y_{i11} - Y_{i10}] - [Y_{i01} - Y_{i00}] = \delta \quad ( = \text{experimental} - \text{control difference}) \]

Thus, the \textit{difference between} the two differences provides an estimate of the effect of the experiment. (Note that this estimate avoids the "post hoc" fallacy.)
Lavy's main results for the treatment effect, $\delta$ (Table 4, p. 1992)

<table>
<thead>
<tr>
<th>outcome</th>
<th>math (all quartiles)</th>
<th>English (all quartiles)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>limited control</td>
<td>full control</td>
</tr>
<tr>
<td>testing rate</td>
<td>0.046 (0.027)</td>
<td>0.041 (0.021)</td>
</tr>
<tr>
<td>control group mean</td>
<td>0.802</td>
<td></td>
</tr>
<tr>
<td>pass rate</td>
<td>0.110 (0.036)</td>
<td>0.087 (0.028)</td>
</tr>
<tr>
<td>control group mean</td>
<td>0.637</td>
<td></td>
</tr>
<tr>
<td>average score</td>
<td>5.469 (2.292)</td>
<td>5.307 (1.950)</td>
</tr>
<tr>
<td>control group mean</td>
<td>55.046</td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are in parentheses, underneath the coefficient estimates.
Table 5: Ignoring dummy variables for “school” produces estimates that are less statistically significant

Further looks at Table 4:
* The program helped students at/below median the most (if high-level student, might pass anyway?)
* Check to see how the program worked:
  improved scores of students who took the exam
  got more students to take the exam

Tables 6-7: regression-discontinuity estimates (probability of being treated drops sharply if matriculation rate goes above 45%) – results similar to the basic results

Table 8: effect of experiment on teaching methods, teaching effort effects evident mostly for English teaching, mostly via extra teaching time

Table 9: did the program cause teachers to inflate their grades? (part of the overall score was based on local test results) answer – apparently not (sanctions for discrepancies between local and national test scores!)