Midterm Review. Power analysis

Pablo Balán

GOV 1263, March 8 2018
Housekeeping

- **Assignment #2**: Due seven minutes ago
- **Assignment #3**: Due in three weeks
The menu

- “Midterm review”
- Power analysis
WHAT HAS SECTION BEEN ABOUT?
What has section been about?

- Anonymous student: “There’s a huge disconnect between what we do in section and in class.” 😞
- Goal of section:
  - Not a watered-down summary of lecture!
  - Give you the tools to:
    - Fully understand the papers from lecture from the perspective of research design
    - Help you with your own research design
- We have only had 5 hours of section
What has section been about? Topic 1: Potential outcomes model

- Q: Why do we learn the potential outcomes model?
- We need a language to talk about counterfactuals
- Causality is about counterfactual states of the world
- We are interested in causal relations between phenomena (not correlations)
Potential outcomes model
Potential outcomes model
Q: Why do we learn about selection bias?

Comparisons between treated and untreated groups (without randomization) are most likely to be wrong.

Everything we observe is wrong, because people “select into” treatment and control.

Randomization is the solution to this problem.
Because observers tend to go to locations that are more accessible, are conveniently located to the last observed location, and are likely to have problems with voter registration, it is difficult to determine what portion of any observed difference in voter registration outcomes should be attributed to the presence of observers and what portion to differences in underlying characteristics, even in the absence of spillover effects. We substantially reduce concerns about confounding by adopting an experimental approach and randomizing which electoral areas should be observed.
Ichino and Schündeln

Because observers tend to go to locations that are more accessible, are conveniently located to the last observed location, and are likely to have problems with voter registration, it is difficult to determine what portion of any observed difference in voter registration outcomes should be attributed to the presence of observers and what portion to differences in underlying characteristics, even in the absence of spillover effects. We substantially reduce concerns about confounding by adopting an experimental approach and randomizing which electoral areas should be observed.
In order to examine spillover effects, we used a two-stage randomized design with blocking in the first stage in a design similar to Miguel and Kremer (2004). As noted earlier, these spillovers are forms of interference across units, a violation of the stable unit treatment value assumption (SUTVA) (Rubin 1978). A simple comparison of means of registration between treatment and control electoral areas will therefore be a biased estimate of the primary effect of a registration observer, even if the assignment of observers to registration centers is randomized. Consequently, we design our experiment to explicitly take into account the possible strategic response of political parties in a way that allows us to detect both localized and general spillover effects.
In order to examine spillover effects, we used a two-stage randomized design with blocking in the first stage in a design similar to Miguel and Kremer (2004). As noted earlier, these spillovers are forms of interference across units, a violation of the stable unit treatment value assumption (SUTVA) (Rubin 1978). A simple comparison of means of registration between treatment and control electoral areas will therefore be a biased estimate of the primary effect of a registration observer, even if the assignment of observers to registration centers is randomized. Consequently, we design our experiment to explicitly take into account the possible strategic response of political parties in a way that allows us to detect both localized and general spillover effects.
Q: Why do we learn about regression?
- Most used tool in statistical analysis
- Used to analyze both observational and experimental data
- Many times, it is similar to a correlation
- Used to predict the value of an outcome variable given values of several independent variables
- We are interested in when regression has a causal interpretation
### Dependent Variable: Percentage change in number of registered voters from 2004 to 2008

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>IV</td>
<td>IV</td>
<td>IV</td>
</tr>
<tr>
<td>Treatment constituency ($T^*$)</td>
<td>-0.006</td>
<td>-0.042*</td>
<td>-0.041*</td>
<td>-0.003</td>
<td>-0.036</td>
<td>-0.036</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.022)</td>
<td>(0.024)</td>
<td>(0.017)</td>
<td>(0.022)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Electoral area assigned registration observer ($T$)</td>
<td>-0.016</td>
<td>-0.030*</td>
<td>-0.035*</td>
<td></td>
<td></td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td>(0.015)</td>
</tr>
<tr>
<td>ELA visited by registration observer ($V$)</td>
<td></td>
<td></td>
<td></td>
<td>-0.022</td>
<td>-0.044*</td>
<td>-0.042*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.022)</td>
</tr>
<tr>
<td># Electoral areas in 5 km assigned registration observer</td>
<td>0.028*</td>
<td>0.027**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Electoral areas in 5–10 km assigned registration observer</td>
<td>0.010</td>
<td>0.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T^*$ # Electoral areas in 5 km assigned registration observer</td>
<td>-0.007</td>
<td>-0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T^*$ # Electoral areas in 5–10 km assigned registration observer</td>
<td>0.019***</td>
<td>0.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q: Why do we learn about omitted variable bias?
For the same reason of why difference in means is likely to be wrong, so is regression if we omit relevant variables.
For regression have a causal interpretation, we need no omitted variables (no unobserved confounders).
It is also a good way to think about the world.
### Table 3—Vote Changes during Aggregation by Candidate Type

<table>
<thead>
<tr>
<th>Panel A. Votes</th>
<th>Post-aggregation votes — pre-aggregation votes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Investigated (= 1)</td>
<td>1.870*</td>
</tr>
<tr>
<td>(0.980)</td>
<td>(1.208)</td>
</tr>
<tr>
<td>Provincial aggregator connection (= 1)</td>
<td>0.942</td>
</tr>
<tr>
<td>(1.651)</td>
<td></td>
</tr>
<tr>
<td>Prov. + district aggregator connection (= 1)</td>
<td>2.207</td>
</tr>
<tr>
<td>(1.848)</td>
<td></td>
</tr>
<tr>
<td>Karzai connection (= 1)</td>
<td>-0.745</td>
</tr>
<tr>
<td>(2.216)</td>
<td></td>
</tr>
<tr>
<td>Government service (= 1)</td>
<td></td>
</tr>
<tr>
<td>(1.918)</td>
<td></td>
</tr>
<tr>
<td>Incumbent (= 1)</td>
<td>0.245</td>
</tr>
<tr>
<td>Constant</td>
<td>0.104***</td>
</tr>
<tr>
<td>R²</td>
<td>0.013</td>
</tr>
<tr>
<td>Number candidates</td>
<td>1.783</td>
</tr>
<tr>
<td>Number polling stations</td>
<td>149</td>
</tr>
<tr>
<td>Number candidate–polling station</td>
<td>48,008</td>
</tr>
<tr>
<td>observations</td>
<td>0.056</td>
</tr>
<tr>
<td>Connection(s) + investigated = 0 (p-value)</td>
<td>0.131</td>
</tr>
</tbody>
</table>
of connection is most important for rigging an election. Second, connections are likely correlated with other candidate attributes that facilitate fraud, and we observe only a very limited number of candidate characteristics making it difficult to rule out omitted variables. Last, we omit the five largest and smallest observations of the
Q: Why do we learn about bad control?
Because including some variables in a regression might make the results further away from the truth
All control variables should be unaffected by the treatment (pre-treatment)
What has section been about? Topic 6: R and painful coding

Q: Why do we learn R?

Pragmatically
- To do power analysis, which is a component of any research design, including yours
- All the experiments covered in lecture did power analysis when their author designed them
- To be able to replicate any results from any published papers we read, on your own
- This is important and it’s part of being a critical consumer of research

For life
- It’s an incredibly useful skill for life, regardless of what you end up doing
Second half of class

- Fewer new things
- More critical (technical) analysis of the experiments from lecture (but you couldn’t do that without the tools)
- More about the final research design
POWER ANALYSIS
Motivation

- Two meanings of “zero”: 1) no effect (true zero) or 2) I wasn’t able to find an effect.
- Supposing our treatment has an effect, how often will I be able to detect it?
- Implies making guesses about: sample size, effect magnitude, etc.
- Power usually set to 0.8
Type I and Type II Error

<table>
<thead>
<tr>
<th>STUDY FINDINGS</th>
<th>REALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NULL HYPOTHESIS</td>
</tr>
<tr>
<td></td>
<td>TRUE</td>
</tr>
<tr>
<td>TRUE</td>
<td></td>
</tr>
<tr>
<td>FALSE</td>
<td>Type I error ($\alpha$) ‘False positive’</td>
</tr>
</tbody>
</table>

- **Power**: the probability of correctly rejecting the null hypothesis $= \Pr(\text{Reject } H_0|H_0 \text{ is False})$
- **Power** $= 1 - \Pr(\text{Accept } H_0|H_0 \text{ is False}) = 1 - \Pr(\text{Type II error})$
What does power depend on?

Power depends on:

- Sample size
- True effect size
- Proportion of treated units
- $\alpha$: Prob(Type I error): TRADE-OFF
- Key design decisions: blocking, including covariates, etc.
Key trade-off

- Type I error and Type II errors are inversely related
- The more demanding in terms of minimizing Type I error (false positives), the less likely to find any results (but might end up with false negatives (higher Type II error)

<table>
<thead>
<tr>
<th>Lower Type I Error</th>
<th>Higher Type II Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Alpha</td>
<td>Lower Power</td>
</tr>
<tr>
<td>Lower False Positive Rate</td>
<td>Higher False Negative Rate</td>
</tr>
</tbody>
</table>
Key trade-off

Normal distribution under H0 and H1

- Distribution under H0
- Distribution under H1
- Critical value
- Power
- $\beta$
- $\alpha$

$x$ axis:
- 60
- 80
- 100
- 120
- 140
- 160
- 180

Density:
- 0.000
- 0.005
- 0.010
- 0.015
- 0.020
- 0.025
- 0.030
The optimal proportion of treated is a function of the variance of potential outcomes

- The variance of the Average Treatment Effect is given by:

\[
V(\hat{\tau}) = \frac{\sigma_1^2}{pN} + \frac{\sigma_0^2}{(1 - p)N}
\]

- \(\sigma_1^2\) and \(\sigma_0^2\) are the variances of potential outcomes under T and C
- To find the optimal \(p\), we take the derivative and set it to zero,

\[
\frac{\partial V(\hat{\tau})}{\partial p} = 0,
\]

which yields:

\[
p = \frac{\sigma_1}{\sigma_1 + \sigma_0} = \frac{1}{1 + \frac{\sigma_0}{\sigma_1}}
\]

- If \(\sigma_1 = \sigma_0\), the optimal proportion is \(p = 0.5\)
The proportion of treated that minimizes ATE variance is 0.5

- The variance of the treatment effect is minimized when the proportion of treated is 0.5
The formula for power is messy

\[ \text{POWER} = \Phi \left( \frac{|\mu_T - \mu_C| \sqrt{N}}{2\sigma} - \Phi^{-1}(1 - \frac{\alpha}{2}) \right) \]

- Power is a probability: will be between 0 and 1
- \( \Phi \) is the CDF of the normal distribution
- We have to plug in \( \mu_T \) and \( \mu_C \) (avg outcome for the treated/control)
- \( \sigma \) is the standard deviation of outcomes (assuming \( \sigma_T = \sigma_C \))
- \( \alpha \) is the significance level, \( P(\text{Type I error}) \).
- Go to the EGAP Power Calculator
The Minimum Detectable Effect (MDE)

- Given my choice of power (usually 0.8) and my sample size (N), what’s the smallest effect I can detect?

\[
MDE = M_{n-2} \sqrt{\frac{\sigma^2}{Np(1-p)}}
\]

- \(M_{n-2} = t_{1-\frac{\alpha}{2}} + t_{1-\psi}\)
- \(1 - \psi = 1 - P(\text{Type II Error})\)
- With \(\alpha = 0.05\) and \(\psi = 0.8\): \(M_{n-2} = t_{0.975} + t_{0.8} = 2 + 0.8 = 2.8\)
- MDE using Olken data
Minimum Detectable Effect as a function of N

- Minimum Detectable Effect (MDE) decreases as the sample size (N) increases.
- The graph shows a downward trend, indicating a lower MDE for larger values of N.
- The red dashed line represents a certain threshold for MDE, below which it is considered undetectable.
Take-away

YOU UNDERESTIMATE THE POWER

OF MY POWER ANALYSIS!
TYPES OF RANDOMIZATION
Types of randomization

- Complete randomization
- Block randomization
- Cluster randomization
- Factorial designs

Key question: How each of these design decisions affects power?
What can possibly go wrong?
Bernoulli trials

- Example: \( n = 4 \) with two males and two women
- Complete randomization will place two women in the same treatment group \( \frac{1}{3} \) of the time
- If that happens, how can we tell the treatment effect from gender differences?
- Downside: could end up with all treated or control
Complete randomization

Suppose we have 6 units and we want 3 treated and 3 control
We assign units 2, 4, 5 to treatment
Now unit 3 has a lower probability of being treated
Probability of treatment is no longer independent
Blocking to the rescue

What is blocking?
- Divide units into blocks defined by a particular variable (Example: gender)
- A block of men and a block of women
- These blocks will fully balanced on gender, not just on average
- Complete randomization within blocks

How does blocking help?
- Ensures balance between groups in terms of key factors and increases efficiency if the blocking variables predict the outcomes (i.e. they “remove” the effect of nuisance factors).
- Incorporates planned testing of different effects for specific subgroups

What to block on?
- The baseline of the outcome variable and the other main predictors
- Variables desired for subgroup analysis
- Stratification (e.g., by gender, age group, or region)
- Pair-Matching (e.g., use matching program to find most similar “twins”)

GOV 1263, March 8 2018
“BLOCK WHAT YOU CAN, RANDOMIZE WHAT YOU CANNOT”
(George Box)