Inference for RCTs
How to get your p-values right

Simon Heß

November 15, 2017

Intro

- What do we need standard error (SE) estimates for?
- What econometric choices influence SE estimates?
- When to do what?

Table of contents

Covariate balancing

Why?

- reduce imbalances
- subgroup analysis

How?

- blocked, paired (lagged DV, Mahalanobis), re-randomize

Consequences?

- increased precision of $\hat{\beta}_{\mathsf{OLS}}$ and treatment correlated across observations and with covariates.

$\Rightarrow$ usual $\hat{V}[\hat{\beta}_{\mathsf{OLS}}]$ doesn't account for this: too large

$\Rightarrow$ CI too large, t-stat too small, p-values too large (*on average*)

Clustered assignment to treatment

Why?

- organizational, costs
- group TE study objective.

How?

- natural clusters (village, school class, state)
- artificial clusters (e.g. lab sessions)

Consequences?

- reduced precision of $\hat{\beta}_{\mathsf{OLS}}$

$\Rightarrow$ usual $\hat{V}[\hat{\beta}_{\mathsf{OLS}}]$ doesn't 'know' this: too small

$\Rightarrow$ CI too small, t-stat too large, p-values too small (*on average*)

Other designs aspects to consider

- sample size, sizes and n^o of clusters, sizes and n^o of strata
- multi-stage treatment assignments
- survey sampling
- attrition
- ...

Estimating TEs

$$y_i = \alpha + \tau D_i + x_i\beta + \varepsilon_i$$

```
regress y treatment x1 x2 x3
```

Stata time (1)

[introductory simulation]

Data analysis

Standard solutions

Stratified assignment ⇒ allow for strata FX[1]

```
regress y treatment x1 x2 x3 i.strata_id
```

Clustered assignment ⇒ adjust inference using CRSE

```
regress y treatment x1 x2 x3, cluster(cluster_id)
```

Issues:

- few clusters
- small strata (few clusters per strata)
- clusters of varying sizes
- ...

[1]conventional wisdom, *but* Bugni et al. (2017) show only valid if $\pi = 1/2$. SE adjustments made available but not used.

Data analysis

Resampling methods

When standard methods fail:

- bootstrap, and
- randomization inference

can quantify estimator uncertainty. Issues:

- computationally intensive
- no standard [software]

Useful methods (1) – Randomization inference (RI)

RI obtains the distribution of any test statistic under the null of "No TE" by resampling the treatment assignment.

1. define suitable test statistic (e.g. $\hat{\tau}_{\mathsf{OLS}}$)
2. R times: draw a treatment vector D^r and compute $\hat{\tau}^r_{\mathsf{OLS}}$
   $\Rightarrow$ distribution of the test statistic under the H_o "No TE"

Testing step: compare original $\hat{\tau}_{\mathsf{OLS}}$ to this distribution.

Useful methods (2) – [Wild-]Bootstrap

Bootstrap obtains the distribution of a test-statistic under the null of "No TE" by resampling the data.

1. estimate $$y_{ic} = \alpha + \tau D_c + x_{ic}\beta + \varepsilon_{ic}$$
   with imposed H_o to obtain $\hat{\alpha}^o, \hat{\tau}^o, \hat{\beta}^o, \hat{\varepsilon}^o_{ic}$.
2. B times: generate bootstrap data set (y^b_{ic}, D_c, x_{ic}):
   $$y^b_{ic} := \hat{\alpha}^o + \hat{\tau}^o D_c + x_{ic}\hat{\beta}^o + w_c \cdot \hat{\varepsilon}^o_{ic}$$
   $$w_c = \begin{cases} -1 & \text{with prob} = 1/2 \\ +1 & \text{with prob} = 1/2 \end{cases}$$
3. on each data set, estimate $\hat{t}^b_{\text{wald}}$
   $\Rightarrow$ distribution of the test statistic under the H_o "No [A]TE"

Testing step: compare original $\hat{t}_{\text{wald}}$ to the obtained distribution

Issues (1)

Few treatment clusters

Problem:

- Estimation of CRSEs relies on N_{clusters} going to infinity, not sample size N. (Cameron and Miller, 2015)
  - $\Rightarrow$ Over-rejection, with CRSE

Possible solutions:

- get more data
- bootstrap
- randomization inference

Stata time (2)

[cluster simulation]

Issues (2)

Many small strata + clustered treatment

Problem:

- CRSEs over-reject, even with many clusters (unclear why)

Possible solutions:

- increase clusters/strata
- bootstrap (?)
- randomization inference

Issues (3)
"Wildly varying cluster-sizes"

Problem:

- Over-rejection, even with 'many' clusters.
  (MacKinnon and Webb, 2017)

Possible solutions:

- weight or drop observations
- cluster SEs at stratum-level
- bootstrap
- randomization inference

Issues (4)

Baseline balance achieved through re-randomization

Problem:

- Under-rejection, but no way to control for strata-fixed effects

Possible solutions:

- control for testing variables
- randomization inference

Takeaways

1. plan your inference as early as possible
   - simulations can help foresee and understand issues
2. experimental design affects what is 'correct' (Abadie et al., 2017; Bruhn and McKenzie, 2009)
   - control variables can matter (strata fixed effects)
   - clusters matter
3. resampling methods can help when the rest fails
   - bootstrap
   - randomization inference

References I

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge (2017): "When Should You Adjust Standard Errors for Clustering?" *arXiv preprint arXiv:1710.02926*.

Bruhn, M. and D. McKenzie (2009): "In Pursuit of Balance: Randomization in Practice in Development Field Experiments," *American Economic Journal: Applied Economics*, 1, 200–232.

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2017): "Inference under covariate-adaptive randomization," *Journal of the American Statistical Association*.

Cameron, A. C. and D. L. Miller (2015): "A practitioner's guide to cluster-robust inference," *Journal of Human Resources*, 50, 317–372.

MacKinnon, J. G. and M. D. Webb (2017): "Wild bootstrap inference for wildly different cluster sizes," *Journal of Applied Econometrics*, 32, 233–254.