RAND EDUCATION

Authentic Power Calculations for RD Studies

J.R. Lockwood

June 30, 2010

Power in a Randomized Experiment

$$\boxed{d\sqrt{n} \approx 4}$$

- d: **standardized effect size**
- n: **# units in each arm**
- **Can approximately handle most issues by fiddling with** n**; e.g.**
  - **Clustering: replace** n **with** $ESS = \frac{n}{DEFF}$
  - **Covariates: replace** n **with** $ESS = \frac{n}{1-R^2}$
  - **Imbalance: replace** n **with** $ESS = 4np(1-p)$

Why is Power in RD Worse?

- S = “forcing variable”
- T = treatment = $1\{S < 0\}$ (WLOG)
- Power degraded due to collinearity between S and T
  - e.g. if S is uniform and T is split at the midpoint, $R^2_{ST} = 0.75$
    - Variance inflation is $\frac{1}{1-R^2_{ST}} = 4$
    - $\rightarrow$ sample size required for power equivalent to randomized experiment is 4 times larger
    - Equivalently, minimum detectable effect for equivalent sample size is 2 times larger

Why is Assessing RD Power More Challenging?

Primarily because power is affected by:

- **Shape of distribution of S and where cutoff c determining T is in that distribution**
  - **Schochet (2008) provides clear description**
- **Estimators for $E(Y|c^-)$ and $E(Y|c^+)$ might be complex and may involve data-dependent data restrictions**
  - **E.g. Cross-validation choice of bandwidth (Ludwig and Miller, 2005; Imbens and Lemieux, 2008) or simultaneous choice of bandwidth and model complexity (Kirby, McCombs, and Mariano, 2009)**
- **Other complications like fuzziness and clustering exacerbate these issues**

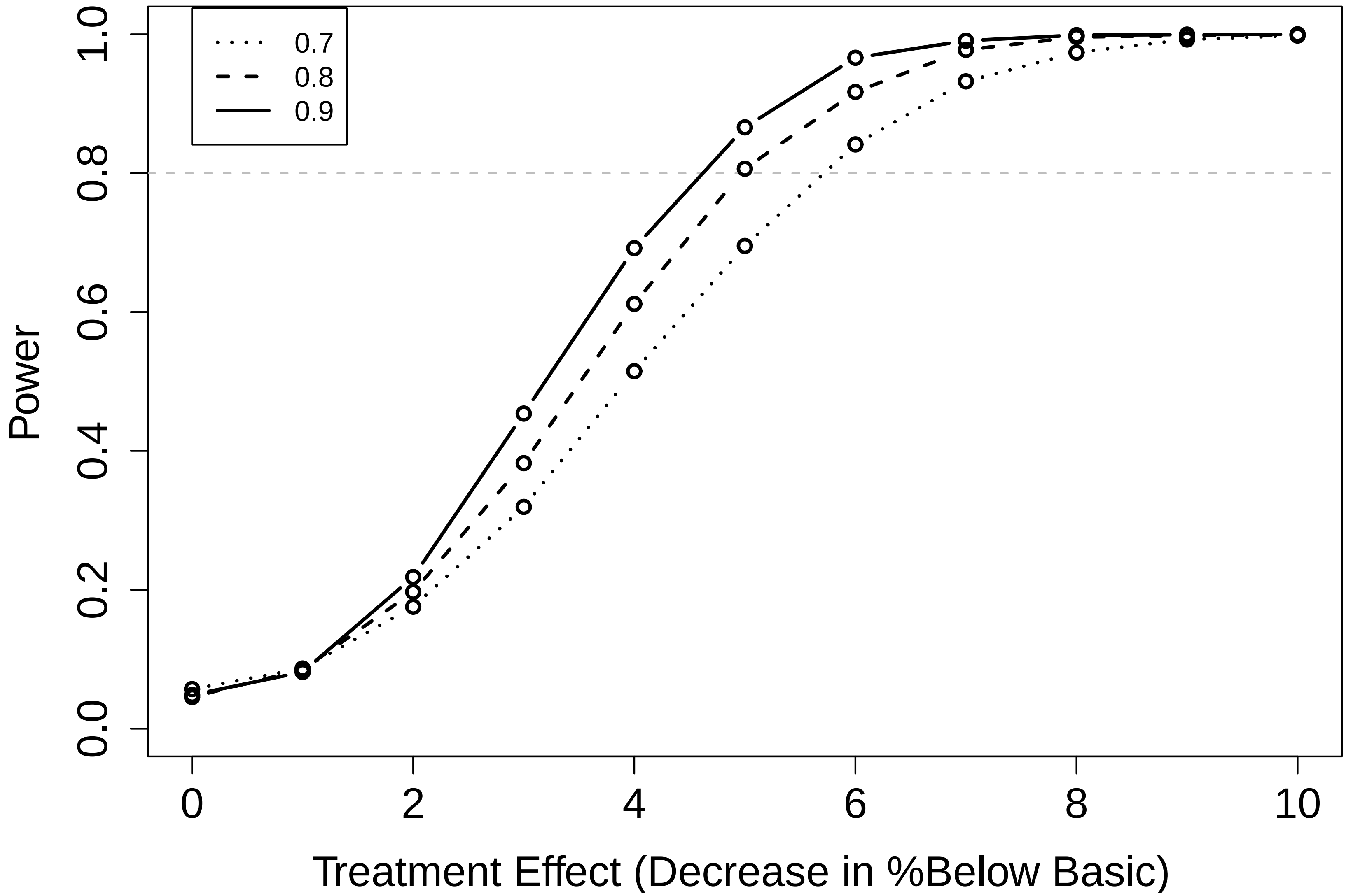
Simulation as an Alternative Approach

- Often know a lot about data during design of RD studies
  - "Happenstance RD": May have actual values of S and T and past values of Y (e.g. NCLB, RTTT)
  - "Designed RD": Will know how you intend to construct S and T and again probably have good proxies for Y
- Rather than trying to map knowledge about the data into power formulas, use knowledge about the data to simulate outcomes and analysis procedure

Sketch of Approach

- β: **True treatment effect**
- $D(\beta)$: **Simulated data, depends on** β
- $\hat{\beta}(D)$: **Estimated treatment effect, depends on** D
  - **"Black Box" - make it as complicated as analysis will be**
- **Step 1: Estimate distribution of** $\hat{\beta}(D)$ **given** $\beta = 0$
  - **Use this to determine rejection region** R
- **Step 2: Estimate** $Pr\{\hat{\beta}(D) \in R\}$ **for selected sequence of alternatives** β
  - **"Outer" loop: sequence of** β
  - **"Inner" loop:** M **Monte Carlo iterations and count how often estimated effect is in rejection region**

Example Output



Advantages of Simulation Approach

- Anything can be inserted in the analysis no matter how hard it would be to examine analytically; e.g.
  - Cluster corrections with imbalanced samples, including the use of random effects models to aid efficiency
  - Complex model selection criteria, such as bandwidth and functional form choice via cross-validation
- No need to agonize over what is meant by an “effect size” in RD - outcomes of simulation study get reported on the natural scale of the outcome measure
- Simulation approach naturally provides power curves rather than MDE at a single value of power (e.g. 0.80) which is more informative

Conclusions

- RD is unlike a randomized experiment because careful statistical model selection and specification is inherent to obtaining valid impact estimates
  - i.e. in RD there is generally not a simple, analytically tractable procedure that will provide a compelling estimate.
- As standard practice for RD becomes more sophisticated (e.g. by WWC standards setting a high bar), simple formulas are less likely to provide authentic assessments of power
- Simulation is a defensible and relatively easy alternative
  - And can benefit from the fact that very specific data is often available during the design phase