Online Appendix

Appendix A.1: Deriving the Power Calculations Formula

This appendix relies on Wittes (2002) and McConnell and Vera-Hernández (2015). Researcher A is interested in estimating the effect of T (the treatment randomly assigned to a subset of the sample) and she has access to sample S that includes a set of m potential outcome variables $(y^k)_{k=1,...,m}$. The researcher decides to run a series of regressions:

$$y^k = a + b_k T + u \tag{A.1}$$

and carries out a series of tests: $H_0^k : b_k = 0$. The z-statistic associated with each test is given by:

$$Z^k = \frac{\bar{Y}_1^k - \bar{Y}_0^k}{\sigma_k \sqrt{1/n_0 + 1/n_1}} \tag{A.2}$$

Where $\bar{Y}_1^k$ ($\bar{Y}_0^k$) is the sample average of Y^k for observations with $T = 1$ ($T = 0$) and n_0 (n_1) is the number of observations with $T = 1$ ($T = 0$). Under H_0^k, $\bar{Y}_1^k = \bar{Y}_0^k$ and Z^k follows a normal distribution with mean zero and variance one.

The choice of α and β lead to the following set of equations:

$$Pr(|z| > Z_{1-\alpha/2}|H_0) < \alpha \tag{A.3}$$

$$Pr(|z| > Z_{1-\alpha/2}|H_A) > 1 - \beta \tag{A.4}$$

Assuming that, for non-true null hypotheses, the effect is δ_k and that $n_0 = n_1 = N/2$ leads to

$$Pr(\frac{\sqrt{N}|\bar{Y}_1^k - \bar{Y}_0^k|}{\sigma_k \sqrt{4}} > Z_{1-\alpha/2}|H_A) > 1 - \beta \tag{A.5}$$

Subtracting both sides by δ_k and dividing both sides by $\sigma_k \sqrt{4/N}$ leads to

$$Pr(\frac{\sqrt{N}(|\bar{Y}_1^k - \bar{Y}_0^k| - \delta_k)}{\sigma_k \sqrt{4}} > Z_{1-\alpha/2} - \frac{\sqrt{N}\delta_k}{\sigma_k \sqrt{4}}|H_A) > 1 - \beta \tag{A.6}$$

Given that under H_A, the expectation of $(\bar{Y}_1^k - \bar{Y}_0^k)$ is δ_k, $\frac{\sqrt{N}(|\bar{Y}_1^k - \bar{Y}_0^k| - \delta_k)}{\sigma_k \sqrt{4}}$ is normally distributed. It follows that

$$Z_{1-\alpha/2} - \frac{\sqrt{N}\delta_k}{\sigma_k \sqrt{4}} = Z_\beta = -Z_{1-\beta} \tag{A.7}$$

Rearranging the equation leads to:

$$Z_{1-\beta} = \delta_k \sqrt{\frac{N}{4\sigma_k^2}} - Z_{1-\alpha/2} \tag{A.8}$$

and so:

$$1 - \beta = \Phi(\delta_k \sqrt{\frac{N}{4\sigma_k^2}} - Z_{1-\alpha/2}) \tag{A.9}$$

If the researcher has access to m variables and plans to use Bonferroni corrections, power is:

$$1 - \beta^{Bonf} = \Phi(\delta_k \sqrt{\frac{N}{4\sigma_k^2}} - Z_{1-\alpha/(2m)}) \tag{A.10}$$

Appendix A.2: Additional Results

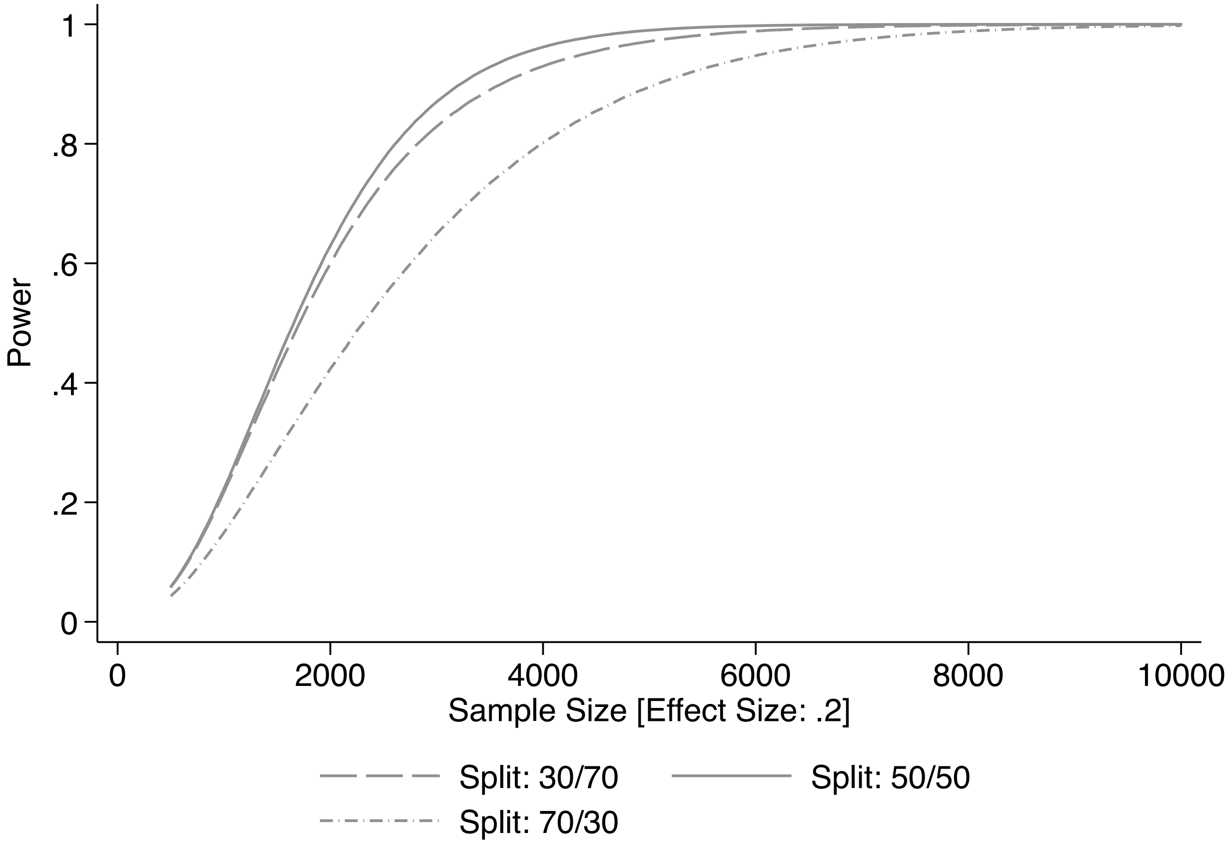

Figure A.1: Power Under the Sample Split Approach with Bonferroni Corrections: Share in the Training Sample

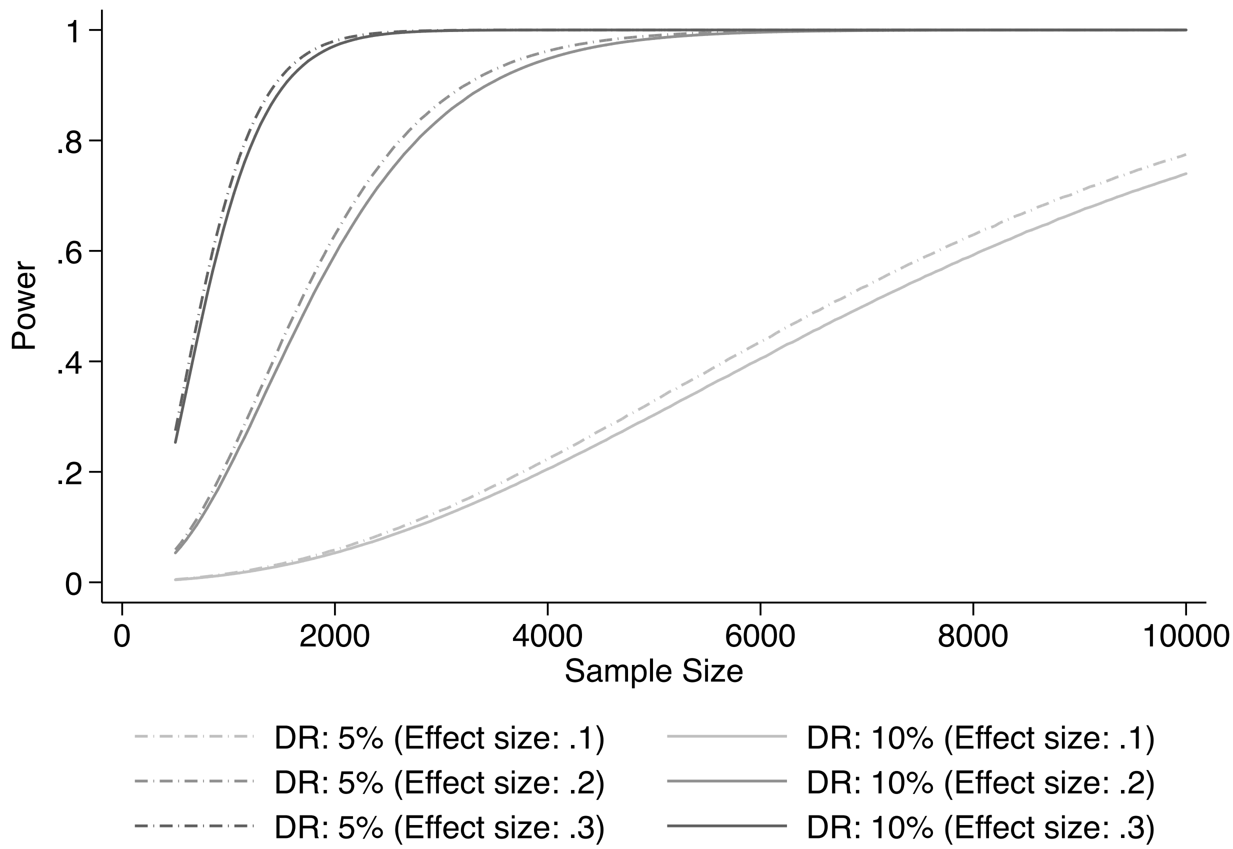

Figure A.2: Power Under the Sample Split Approach with Bonferroni Corrections: Decision Rule on Training Sample

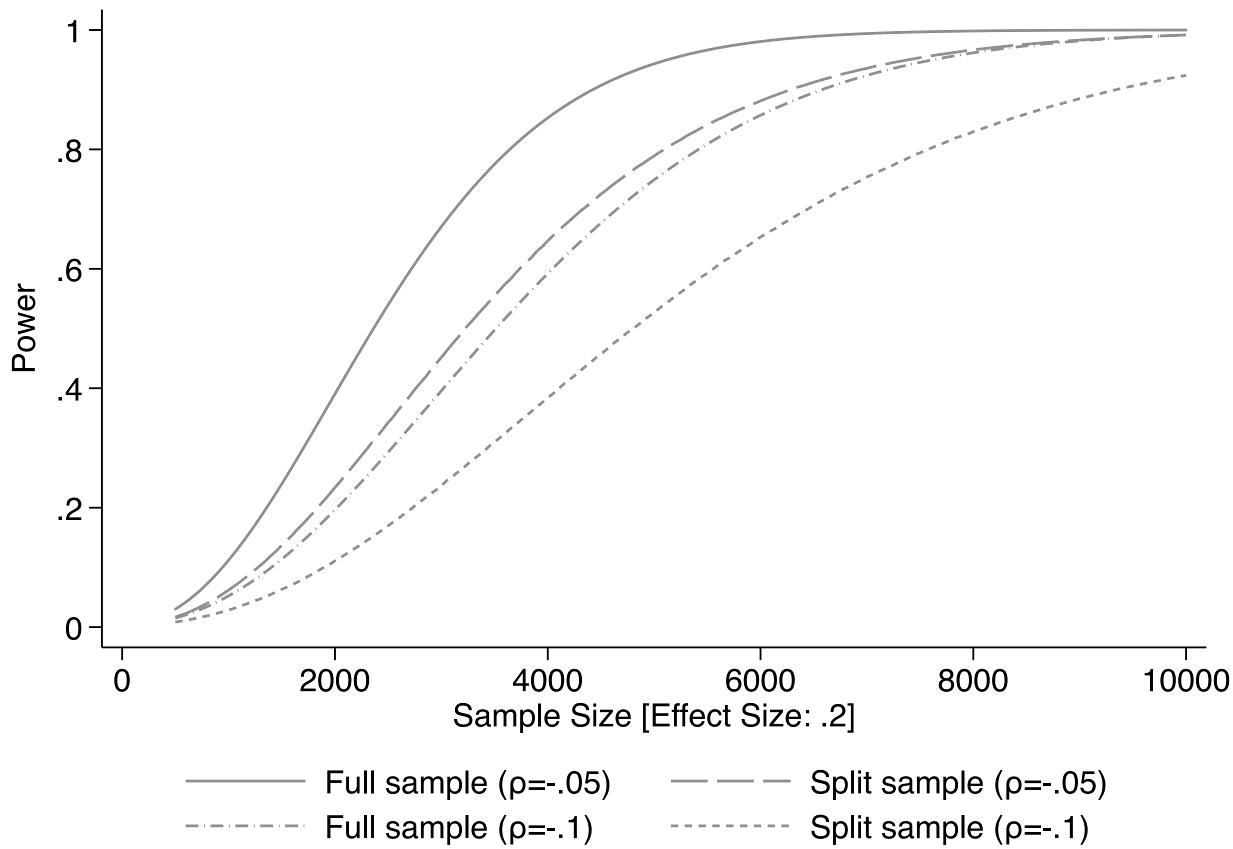

Figure A.3: Comparing Power with Clustered Samples & Bonferroni Corrections: Full Sample vs. Split Sample [Effect size = .2]