17.802 – Recitation 5

Inference and Power Calculations

Yiqing Xu

MIT

March 7, 2014

Inference (mostly MHE Ch8)

- Inference in Asymptopia (and with Weak Null)
  - Robust standard error and biases
  - Clustering
  - Bootstrap

- Permutation inference (and with Sharp Null)

Robust Standard Errors

- We're interested in the large-sample (approximate) distributions of the estimators
- Robust standard errors are "robust" because they provide accurate hypothesis tests and confidence intervals in large samples

$$\sqrt{N}(\hat{\beta} - \beta) = \left[\frac{\sum X_i X_i'}{N}\right]^{-1} (\frac{1}{\sqrt{N}} \sum X_i e_i)$$

$$A.VarCov(\hat{\beta}) = E[X_i X_i']^{-1} E[X_i X_i' e_i^2]^{-1} E[X_i X_i']^{-1}$$

- Conventional: $E[X_i X_i' e_i^2] = E[X_i X_i']^{-1} \sigma^2$
- H. White: No! $E[X_i X_i' e_i^2]$ is pretty good, why bother?
- But they're biased (in finite samples)!

TABLE 8.1.1
Monte Carlo results for robust standard error estimates

Parameter Estimate	Mean (1)	Standard Deviation (2)	Empirical 5% Rejection Rates: Normal (3)	Empirical 5% Rejection Rates: t (4)
A. Lots of heteroskedasticity				
$\hat{\beta}_1$	−.001	.586		
Standard Errors				
Conventional	.331	.052	.278	.257
HC_0	.417	.203	.247	.231
HC_1	.447	.218	.223	.208
HC_2	.523	.260	.177	.164
HC_3	.636	.321	.130	.120
max(HC_0, *Conventional*)	.448	.172	.188	.171
max(HC_1, *Conventional*)	.473	.190	.173	.157
max(HC_2, *Conventional*)	.542	.238	.141	.128
max(HC_3, *Conventional*)	.649	.305	.107	.097
B. Little heteroskedasticity				
$\hat{\beta}_1$	.004	.600		
Standard Errors				
Conventional	.520	.070	.098	.084
HC_0	.441	.193	.217	.202
HC_1	.473	.207	.194	.179
HC_2	.546	.250	.156	.143
HC_3	.657	.312	.114	.104
max(HC_0, *Conventional*)	.562	.121	.083	.070
max(HC_1, *Conventional*)	.578	.138	.078	.067
max(HC_2, *Conventional*)	.627	.186	.067	.057
max(HC_3, *Conventional*)	.713	.259	.053	.045
C. No heteroskedasticity				
$\hat{\beta}_1$	−.003	.611		
Standard Errors				
Conventional	.604	.081	.061	.050
HC_0	.453	.190	.209	.193
HC_1	.486	.203	.185	.171
HC_2	.557	.247	.150	.136
HC_3	.667	.309	.110	.100
max(HC_0, *Conventional*)	.629	.109	.055	.045
max(HC_1, *Conventional*)	.640	.122	.053	.044
max(HC_2, *Conventional*)	.679	.166	.047	.039
max(HC_3, *Conventional*)	.754	.237	.039	.031

Notes: The table reports results from a sampling experiment with 25,000 replications. Columns 1 and 2 shows the mean and standard deviation of estimated *standard errors*, except for the first row in each panel which shows the mean and standard deviation of $\hat{\beta}_1$. The model is as described by (8.1.9), with $\beta_1 = 0$, $r = .1$, $N = 30$, and

Robust S.E.s are Biased!

- "Robust" S.E. has more (downward) bias in finte samples where heteroskedasticity is small than the conventional one
- Never to accept the result when robust S.E. is smaller than conventional
- Always be more conservative – take the largest S.E.

Clustering – Moulton factor

$$y_{ig} = \beta_0 + \beta_1 x_g + e_{ig}; \quad E[e_{ig} e_{jg}] = \rho \sigma_e^2$$

- ρ is the intra-class correlation coefficient; if $e_{ig} = v_g + \eta_{ig}$,

$$\rho = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}$$

- The Moulton factor (Moulton 1986) is the square root of

$$\frac{V(\hat{\beta}_1)}{V_{OLS}(\hat{\beta}_1)} = 1 + (n-1)\rho$$

  n is the size of each group

- $n = 100, \rho = 0.2$, the Moulton factor is 4.56!

Clustering – Fix

- Liang and Zeger (1986)

$$\begin{aligned}\hat{\Omega}_{cl} &= (X'X)^{-1}\left(\sum_g X_g \hat{\Psi}_g X_g\right)(X'X)^{-1} \\ \Psi_g &= \alpha \hat{e}_g \hat{e}_g'\end{aligned}$$

  in which α adjusts degrees of freedom
- Same old, same old: clustered standard errors is downward biased with few clusters
- Try regression of group averages to test robustness

$$\bar{Y}_g = \beta_0 + \beta_1 x_g + \bar{e}_g$$

  (1) actual normality; (2) covariate-adjustment
- Block bootstrap

Bootstrap

- "Theory"
  - Suppose $\mathcal{W} = (W_1, \cdots, W_n)$ is an i.i.d. random sample of the distribution F
  - We are interested in $\theta(F)$
  - We consider an estimator, $\hat{\theta} = \hat{\theta}(\hat{F}_n)$
  - $\hat{F}_n$ is the empirical distribution, $\hat{F}(w) = \frac{1}{n}\sum_{j=1}^{n} I(W_j \leq w)$
  - $n \to \infty$, $\hat{F}_n \to F$ and $\hat{\theta} \xrightarrow{d} \theta(F)$ (in Asymptopia!)

- How does it work?
  - Generate bootstrap samples $\mathcal{W}^*$ (non-parametric; parametric)

    e.g. $y_i = x_i\beta + \epsilon_i$; sampling$\{x_i, y_i\}$ or $\{\hat{\epsilon}_i\}$

  - Calculate the bootstrap replications $\hat{\theta}(\mathcal{W}^*)$
  - Summarize the bootstrap replications
  - Often doing so by blocks

Inference

- Inference in Asymptopia (and with Weak Null)
  - Robust standard error and biases
  - Clustering
  - Bootstrap

- Permutation inference (and with Sharp Null)

Permutation Inference - Motivations

- Benefits:
  - No longer in Aymptopia
  - No needs for limiting theories
- Costs:
  - Research money on buying more powerful computers
  - Sharp Null
- Sharp Null
  - p-values in the absence of S.E.
  - Minimal assumptions (not even SUTVA, why?)
  - Applicable when things getting complicated

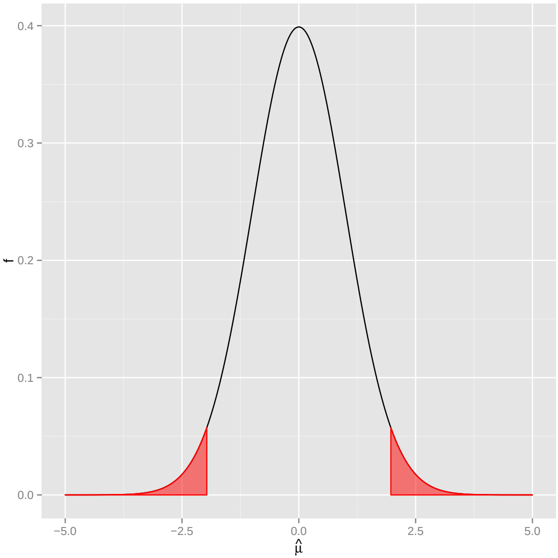
Summary

- Take-away
  - Robust S.E. has finite sample bias; can be even worse then the conventional
  - HC_1-HC_3 helps but no panacea
  - Clustering is huge, but clustered S.E. also is downward-biased
  - Block bootstrap improves things but we're not sure
- If you're tired of the world of Aymptopia
  - come back to real world and try permutation inference, if
    (1) you have powerful computers
    (2) you enjoy explaining stuff to people

The Logic of Power Calculations

$$Y_i = \gamma + \delta D_i + \epsilon_i, \quad i = 1, \cdots, N, \ \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

- From data structure to Minimum Detectable Effect (MDE)
  - You know σ_ϵ (from previous research, or your gut)
  - You know N
  - You know the Null hypothesis, e.g., $\delta = 0$
  - You calculate $SE(\hat{\delta})$, $CI(\hat{\delta})$
  - You know the significance level you want, e.g., $\alpha = 0.05$ (Type I error)
  - $\Rightarrow MDE(\delta)$
- From a presumed effect to sample size, and ultimately, money

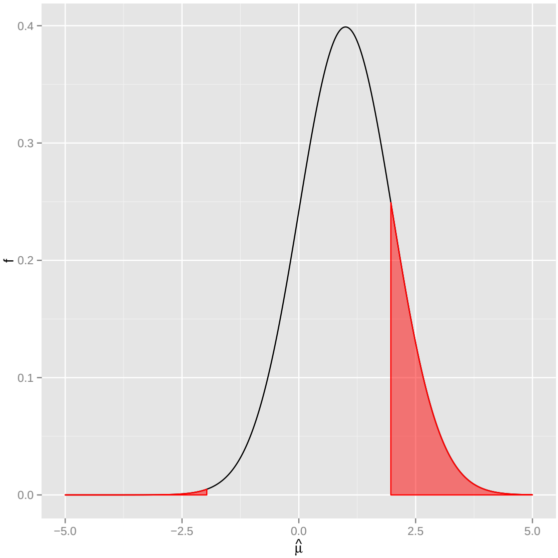


0.4
0.3
0.2
0.1
0.0
f
−5.0
−2.5
0.0
2.5
5.0
μ̂

The Logic of Power Calculations

$$Y_i = \gamma + \delta D_i + \epsilon_i, \quad i = 1, \cdots, N, \ \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

- From data structure to Minimum Detectable Effect (MDE)
- From a presumed effect to sample size, and ultimately, money
  - You know σ_ϵ
  - You know the size of the effect δ
  - You know the sampling distribution of $\hat{\delta}$, $F(\hat{\delta})$
  - (1) You know N, you calculate power
    (2) You know how much power you want, e.g., $\beta = 0.9$ (Type II error), You calculate N, which leads to $


0.4
0.3
0.2
0.1
0.0
f
−5.0
−2.5
0.0
2.5
5.0
$\hat{\mu}$

The Logic of Power Calculations

$$Y_i = \gamma + \delta D_i + \epsilon_i, \quad i = 1, \cdots, N, \ \epsilon_i \sim N(0, \sigma_\epsilon^2)$$

- From data structure to Minimum Detectable Effect (MDE)
  - You know σ_ϵ (from previous research, or your gut)
  - You know N
  - You know the Null hypothesis, e.g., $\delta = 0$
  - You calculate $SE(\hat{\delta})$, $CI(\hat{\delta})$
  - You know the significance level you want, e.g., $\alpha = 0.05$ (Type I error)
  - $\Rightarrow MDE(\delta)$
- From a presumed effect to sample size, and ultimately, money
  - You know σ_ϵ
  - You know the size of the effect δ
  - You know the sampling distribution of $\hat{\delta}$, $F(\hat{\delta})$
  - (1) You know N, you calculate power
    (2) You know how much power you want, e.g., $\beta = 0.9$ (Type II error), You calculate N, which leads to $

In Summary

- Inference of frequentists
  - Inference in Aymptopia ("robust," clustering, bootstrap)
  - Permutation inference
- Power calculations
  - From data structure to MDE
  - From power to sample size and money