ECON 626: Applied Microeconomics

Lecture 8:

Permutations and Bootstraps

Professors: Pamela Jakiela and Owen Ozier

Part I: Randomization Inference

Randomization Inference vs Confidence Intervals

- See Imbens and Rubin, *Causal Inference*, first chapters.
- 100 years ago, Fisher was after a "sharp null," where Neyman and Gosset (Student) were concerned with average effects.

Randomization Inference

How can we do hypothesis testing without asymptotic approximations?

Randomization Inference

How can we do hypothesis testing without asymptotic approximations?
Begin with idea of a **sharp null**: $Y_{1i} = Y_{0i}\ \forall i$. (Gerber and Green, p.62)

Randomization Inference

How can we do hypothesis testing without asymptotic approximations?
Begin with idea of a **sharp null**: $Y_{1i} = Y_{0i}\ \forall i$. (Gerber and Green, p.62)

- If $Y_{1i} = Y_{0i}\ \forall i$, then if we observe either, we have seen both.

Randomization Inference

How can we do hypothesis testing without asymptotic approximations?
Begin with idea of a **sharp null**: $Y_{1i} = Y_{0i}\ \forall i$. (Gerber and Green, p.62)

- If $Y_{1i} = Y_{0i}\ \forall i$, then if we observe either, we have seen both.
- All possible treatment arrangements would yield the same Y values.

Randomization Inference

How can we do hypothesis testing without asymptotic approximations?
Begin with idea of a **sharp null**: $Y_{1i} = Y_{0i} \; \forall i$. (Gerber and Green, p.62)

- If $Y_{1i} = Y_{0i} \; \forall i$, then if we observe either, we have seen both.
- All possible treatment arrangements would yield the same Y values.
- We could then calculate all possible treatment effect estimates under the sharp null.

Randomization Inference

How can we do hypothesis testing without asymptotic approximations?
Begin with idea of a **sharp null**: $Y_{1i} = Y_{0i}\ \forall i$. (Gerber and Green, p.62)

- If $Y_{1i} = Y_{0i}\ \forall i$, then if we observe either, we have seen both.
- All possible treatment arrangements would yield the same Y values.
- We could then calculate all possible treatment effect estimates under the sharp null.
- The distribution of these possible treatment effects allows us to compute p-values:

Randomization Inference

How can we do hypothesis testing without asymptotic approximations?
Begin with idea of a **sharp null**: $Y_{1i} = Y_{0i}\ \forall i$. (Gerber and Green, p.62)

- If $Y_{1i} = Y_{0i}\ \forall i$, then if we observe either, we have seen both.
- All possible treatment arrangements would yield the same Y values.
- We could then calculate all possible treatment effect estimates under the sharp null.
- The distribution of these possible treatment effects allows us to compute p-values: The probability that, under the null, something this large or larger would occur at random. *(For the two sided test, "large" means in absolute value terms.)*

Randomization Inference

How can we do hypothesis testing without asymptotic approximations?
Begin with idea of a **sharp null**: $Y_{1i} = Y_{0i}\ \forall i$. (Gerber and Green, p.62)

- If $Y_{1i} = Y_{0i}\ \forall i$, then if we observe either, we have seen both.
- All possible treatment arrangements would yield the same Y values.
- We could then calculate all possible treatment effect estimates under the sharp null.
- The distribution of these possible treatment effects allows us to compute p-values: The probability that, under the null, something this large or larger would occur at random. *(For the two sided test, "large" means in absolute value terms.)*
- This extends naturally to the case where treatment assignments are restricted in some way. Recall, for example, the Bruhn and McKenzie (2009) discussion of the many different restrictions that can be used to yield balanced randomization.

Randomization Inference

It is often impractical to enumerate *all* possible treatment effects. Instead, we sample a large number of them:

Randomization Inference

It is often impractical to enumerate *all* possible treatment effects. Instead, we sample a large number of them:

- Regress Y on T. Note the absolute value of the coefficient on T.

Randomization Inference

It is often impractical to enumerate *all* possible treatment effects. Instead, we sample a large number of them:

- Regress Y on T. Note the absolute value of the coefficient on T.
- For a large number of iterations:
  - Devise an alternative random assignment of treatment. In the simplest, unrestricted case, this means scrambling the relationship between Y and T randomly, preserving the number of treatment and comparison units in T. Call this assignment *AlternativeTreatment*.

Randomization Inference

It is often impractical to enumerate *all* possible treatment effects. Instead, we sample a large number of them:

- Regress Y on T. Note the absolute value of the coefficient on T.
- For a large number of iterations:
  - Devise an alternative random assignment of treatment. In the simplest, unrestricted case, this means scrambling the relationship between Y and T randomly, preserving the number of treatment and comparison units in T. Call this assignment *AlternativeTreatment*.
  - Regress Y on *AlternativeTreatment*.

Randomization Inference

It is often impractical to enumerate *all* possible treatment effects. Instead, we sample a large number of them:

- Regress Y on T. Note the absolute value of the coefficient on T.
- For a large number of iterations:
  - Devise an alternative random assignment of treatment. In the simplest, unrestricted case, this means scrambling the relationship between Y and T randomly, preserving the number of treatment and comparison units in T. Call this assignment *AlternativeTreatment*.
  - Regress Y on *AlternativeTreatment*.
  - Note whether the absolute value of the coefficient on *AlternativeTreatment* equals or exceeds the absolute value of the original (true) coefficient on T. If so, increment a counter.

Randomization Inference

It is often impractical to enumerate *all* possible treatment effects. Instead, we sample a large number of them:

- Regress Y on T. Note the absolute value of the coefficient on T.
- For a large number of iterations:
  - Devise an alternative random assignment of treatment. In the simplest, unrestricted case, this means scrambling the relationship between Y and T randomly, preserving the number of treatment and comparison units in T. Call this assignment *AlternativeTreatment*.
  - Regress Y on *AlternativeTreatment*.
  - Note whether the absolute value of the coefficient on *AlternativeTreatment* equals or exceeds the absolute value of the original (true) coefficient on T. If so, increment a counter.
- Divide the counter by the number of iterations. You have a p-value!

Gerber and Green, p.63: "... the calculation of p-values based on an inventory of possible randomizations is called *randomization inference*."

Randomization Inference

Gerber and Green, p.64:

- "The sampling distribution of the test statistic under the null hypothesis is computed by simulating all possible random assignments. When the number of random assignments is too large to simulate, the sampling distribution may be approximated by a large random sample of possible assignments. p-values are calculated by comparing the observed test statistic to the distribution of test statistics under the null hypothesis."

Randomization Inference

Gerber and Green, p.64:

- "The sampling distribution of the test statistic under the null hypothesis is computed by simulating all possible random assignments. When the number of random assignments is too large to simulate, the sampling distribution may be approximated by a large random sample of possible assignments. p-values are calculated by comparing the observed test statistic to the distribution of test statistics under the null hypothesis."

NOTE: How large a random sample?
What is the standard deviation of a binary outcome with mean 0.05?

Randomization Inference

Gerber and Green, p.64:

- "The sampling distribution of the test statistic under the null hypothesis is computed by simulating all possible random assignments. When the number of random assignments is too large to simulate, the sampling distribution may be approximated by a large random sample of possible assignments. p-values are calculated by comparing the observed test statistic to the distribution of test statistics under the null hypothesis."

NOTE: How large a random sample?
What is the standard deviation of a binary outcome with mean 0.05?
About 0.22.

Randomization Inference

Gerber and Green, p.64:

- "The sampling distribution of the test statistic under the null hypothesis is computed by simulating all possible random assignments. When the number of random assignments is too large to simulate, the sampling distribution may be approximated by a large random sample of possible assignments. p-values are calculated by comparing the observed test statistic to the distribution of test statistics under the null hypothesis."

NOTE: How large a random sample?
What is the standard deviation of a binary outcome with mean 0.05?
About 0.22.
Standard error (of this estimated p-value) ...in a sample of size 100 alternative randomizations?

Randomization Inference

Gerber and Green, p.64:

- "The sampling distribution of the test statistic under the null hypothesis is computed by simulating all possible random assignments. When the number of random assignments is too large to simulate, the sampling distribution may be approximated by a large random sample of possible assignments. p-values are calculated by comparing the observed test statistic to the distribution of test statistics under the null hypothesis."

NOTE: How large a random sample?
What is the standard deviation of a binary outcome with mean 0.05? About 0.22.
Standard error (of this estimated p-value) ...in a sample of size 100 alternative randomizations? About 0.022.

Randomization Inference

Gerber and Green, p.64:

- "The sampling distribution of the test statistic under the null hypothesis is computed by simulating all possible random assignments. When the number of random assignments is too large to simulate, the sampling distribution may be approximated by a large random sample of possible assignments. p-values are calculated by comparing the observed test statistic to the distribution of test statistics under the null hypothesis."

NOTE: How large a random sample?
What is the standard deviation of a binary outcome with mean 0.05?
About 0.22.
Standard error (of this estimated p-value) ...in a sample of size 100 alternative randomizations? About 0.022.
...in a sample of size 10,000 alternative randomizations?

Randomization Inference

Gerber and Green, p.64:

- "The sampling distribution of the test statistic under the null hypothesis is computed by simulating all possible random assignments. When the number of random assignments is too large to simulate, the sampling distribution may be approximated by a large random sample of possible assignments. p-values are calculated by comparing the observed test statistic to the distribution of test statistics under the null hypothesis."

NOTE: How large a random sample?
What is the standard deviation of a binary outcome with mean 0.05? About 0.22.
Standard error (of this estimated p-value) ...in a sample of size 100 alternative randomizations? About 0.022.
...in a sample of size 10,000 alternative randomizations? About 0.0022.

Randomization Inference Confidence Intervals

Major drawback

- This doesn't give you a confidence interval automatically.
- Under assumptions, you can construct them (Gerber and Green, section 3.5):
  - "The most traightforward method for filling in missing potential outcomes is to assume that the treatment effect τ_i is the same for all subjects."

Randomization Inference Confidence Intervals

Gerber and Green, section 3.5:

Randomization Inference Confidence Intervals

Gerber and Green, section 3.5:

- "For subjects in the control condition, missing $Y_i(1)$ values are imputed by adding the estimated ATE to the observed values of $Y_i(0)$."

Randomization Inference Confidence Intervals

Gerber and Green, section 3.5:

- "For subjects in the control condition, missing $Y_i(1)$ values are imputed by adding the estimated ATE to the observed values of $Y_i(0)$."
- "Similarly, for subjects in the treatment condition, missing $Y_i(0)$ values are imputed by subtracting the estimated ATE from the observed values of $Y_i(1)$."

Randomization Inference Confidence Intervals

Gerber and Green, section 3.5:

- "For subjects in the control condition, missing $Y_i(1)$ values are imputed by adding the estimated ATE to the observed values of $Y_i(0)$."
- "Similarly, for subjects in the treatment condition, missing $Y_i(0)$ values are imputed by subtracting the estimated ATE from the observed values of $Y_i(1)$."
- "This approach yields a complete schedule of potential outcomes, which we may then use to simulate all possible random allocations."

Randomization Inference Confidence Intervals

Gerber and Green, section 3.5:

- "For subjects in the control condition, missing $Y_i(1)$ values are imputed by adding the estimated ATE to the observed values of $Y_i(0)$."
- "Similarly, for subjects in the treatment condition, missing $Y_i(0)$ values are imputed by subtracting the estimated ATE from the observed values of $Y_i(1)$."
- "This approach yields a complete schedule of potential outcomes, which we may then use to simulate all possible random allocations."
- "In order to form a 95% confidence interval, we list the estimated ATE from each random allocation in ascending order. The estimate at the 2.5th percentile marks the bottom of the interval, and the estimate at the 97.5th percentile marks the top of the interval."

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\begin{pmatrix} 4 \\ 3 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix} +$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\begin{pmatrix} 4 \\ 3 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} +$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\begin{pmatrix} 4 \\ 3 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 4 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} +$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\begin{pmatrix} 4 \\ 3 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 4 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \end{pmatrix}\begin{pmatrix} 4 \\ 4 \end{pmatrix}$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\binom{4}{3}\binom{4}{1} + \binom{4}{1}\binom{4}{3} + \binom{4}{4}\binom{4}{0} + \binom{4}{0}\binom{4}{4}$

  $= 4 \cdot 4 + 4 \cdot 4 + 1 \cdot 1 + 1 \cdot 1 =$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\binom{4}{3}\binom{4}{1} + \binom{4}{1}\binom{4}{3} + \binom{4}{4}\binom{4}{0} + \binom{4}{0}\binom{4}{4}$

  $= 4 \cdot 4 + 4 \cdot 4 + 1 \cdot 1 + 1 \cdot 1 = = 16 + 16 + 1 + 1 =$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\binom{4}{3}\binom{4}{1} + \binom{4}{1}\binom{4}{3} + \binom{4}{4}\binom{4}{0} + \binom{4}{0}\binom{4}{4}$

  $= 4 \cdot 4 + 4 \cdot 4 + 1 \cdot 1 + 1 \cdot 1 = = 16 + 16 + 1 + 1 = 34$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\begin{pmatrix} 4 \\ 3 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 4 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \end{pmatrix}\begin{pmatrix} 4 \\ 4 \end{pmatrix}$

  $= 4 \cdot 4 + 4 \cdot 4 + 1 \cdot 1 + 1 \cdot 1 = = 16 + 16 + 1 + 1 = 34$ out of...
  $\begin{pmatrix} 8 \\ 4 \end{pmatrix} =$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\binom{4}{3}\binom{4}{1} + \binom{4}{1}\binom{4}{3} + \binom{4}{4}\binom{4}{0} + \binom{4}{0}\binom{4}{4}$

  $= 4 \cdot 4 + 4 \cdot 4 + 1 \cdot 1 + 1 \cdot 1 == 16 + 16 + 1 + 1 = 34$ out of...
  $\binom{8}{4} = 70.$

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\begin{pmatrix} 4 \\ 3 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 4 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \end{pmatrix}\begin{pmatrix} 4 \\ 4 \end{pmatrix}$

  $= 4 \cdot 4 + 4 \cdot 4 + 1 \cdot 1 + 1 \cdot 1 = = 16 + 16 + 1 + 1 = 34$ out of...
  $\begin{pmatrix} 8 \\ 4 \end{pmatrix} = 70.$ p-value just under 50 percent.

Activity 1

Remember the Lady Tasting Tea from the first class?
Suppose she gets a certain number right. For example:

- Eight cups, four of which had milk added first.
- After tasting, suppose she correctly says there are four cups which had milk added first, but while she correctly identifies three cups, she gets one wrong. What is the probability of being that correlated with the truth, or better? (in absolute value terms?)
- $\binom{4}{3}\binom{4}{1} + \binom{4}{1}\binom{4}{3} + \binom{4}{4}\binom{4}{0} + \binom{4}{0}\binom{4}{4}$

  $= 4 \cdot 4 + 4 \cdot 4 + 1 \cdot 1 + 1 \cdot 1 = = 16 + 16 + 1 + 1 = 34$ out of...
  $\binom{8}{4} = 70$. p-value just under 50 percent.

Or what if it were ten cups, and she got 4 out of 5?
This becomes unwieldy to calculate exactly. Activity: randomly sample.

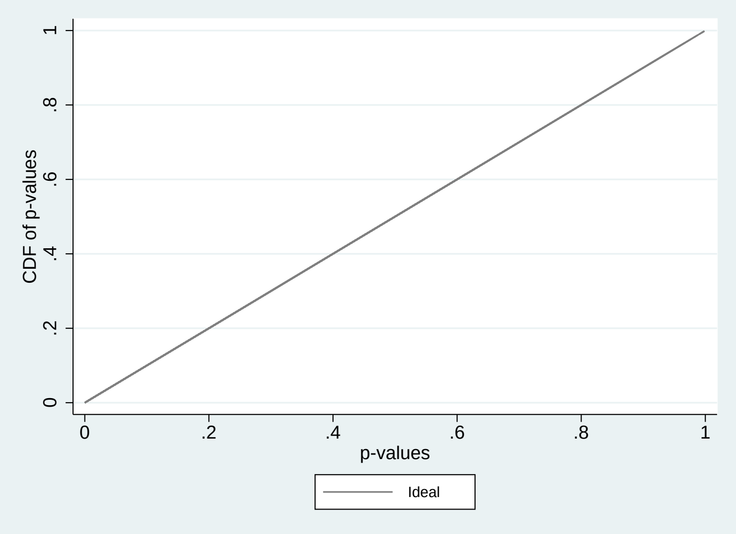
Part IIa: Bootstrap

Bootstrap basics

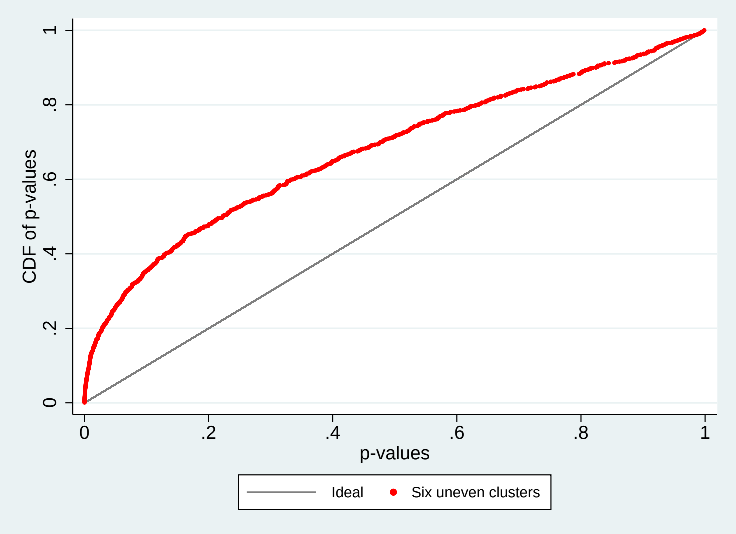
- See Angrist and Pischke, pp.300-301 (Bootstrap).
- Sampling $\{Y_i, X_i\}$ with replacement: "pairs bootstrap" or "nonparametric bootstrap."
- Keeping X_i fixed, sampling $\hat{e}_i$ with replacement, constructing new outcomes Y_i treating X_i as fixed using the original $\hat{\beta}$: one kind of "parametric bootstrap."
- Keeping X_i fixed, constructing new outcomes Y_i treating X_i as fixed using the original $\hat{\beta}$, but randomly flipping the sign of $\hat{e}_i$, preserving relationships between X_i and the variance of the residual: "wild bootstrap."

Part IIb:
Few Clusters;
Wild Cluster Bootstrap

What is the problem with having too few clusters?



What is the problem with having too few clusters?



How will "bootstrapping" solve the cluster problem?

Bootstrapping is drawing (often with replacement) from some aspect of the data to quantify variability of a statistic.

How will "bootstrapping" solve the cluster problem?

Bootstrapping is drawing (often with replacement) from some aspect of the data to quantify variability of a statistic.
We cluster standard errors because we are concerned that the error may be heteroskedastic and correlated within clusters. So we could not sensibly use a bootstrapping procedure that ignored covariates or the cluster structure.

How will "bootstrapping" solve the cluster problem?

Bootstrapping is drawing (often with replacement) from some aspect of the data to quantify variability of a statistic.
We cluster standard errors because we are concerned that the error may be heteroskedastic and correlated within clusters. So we could not sensibly use a bootstrapping procedure that ignored covariates or the cluster structure. Cameron, Gelbach, Miller (2008) and Cameron and Miller (2015) discuss a procedure that respects covariate structure ("wild")

How will "bootstrapping" solve the cluster problem?

Bootstrapping is drawing (often with replacement) from some aspect of the data to quantify variability of a statistic.
We cluster standard errors because we are concerned that the error may be heteroskedastic and correlated within clusters. So we could not sensibly use a bootstrapping procedure that ignored covariates or the cluster structure. Cameron, Gelbach, Miller (2008) and Cameron and Miller (2015) discuss a procedure that respects covariate structure ("wild") and cluster structure ("cluster")

How will "bootstrapping" solve the cluster problem?

Bootstrapping is drawing (often with replacement) from some aspect of the data to quantify variability of a statistic.
We cluster standard errors because we are concerned that the error may be heteroskedastic and correlated within clusters. So we could not sensibly use a bootstrapping procedure that ignored covariates or the cluster structure. Cameron, Gelbach, Miller (2008) and Cameron and Miller (2015) discuss a procedure that respects covariate structure ("wild") and cluster structure ("cluster") while drawing alternative residuals ("bootstrap").

The Wild Cluster Bootstrap

Cameron, Gelbach, and Miller procedure goes as follows (Cameron and Miller 2015, Section VI.C.2):

- "First, estimate the main model, imposing (forcing) the null hypothesis that you wish to test... For example, for test of statistical significance of a single variable regress y_{ig} on all components of x_{ig} except the variable that has regressor with coefficient zero under the null hypothesis."

The Wild Cluster Bootstrap

Cameron, Gelbach, and Miller procedure goes as follows (Cameron and Miller 2015, Section VI.C.2):

- "First, estimate the main model, imposing (forcing) the null hypothesis that you wish to test... For example, for test of statistical significance of a single variable regress y_{ig} on all components of x_{ig} except the variable that has regressor with coefficient zero under the null hypothesis."
- "Form the residual $\tilde{u}_{ig} = y_i g - x'_{ig}\tilde{\beta}_{H0}$"

The Wild Cluster Bootstrap

Cameron, Gelbach, and Miller procedure goes as follows (Cameron and Miller 2015, Section VI.C.2):

- In each resampling:

The Wild Cluster Bootstrap

Cameron, Gelbach, and Miller procedure goes as follows (Cameron and Miller 2015, Section VI.C.2):

- In each resampling:
  - "Randomly assign cluster g the weight $d_g = -1$ with probability 0.5 and the weight $d_g = 1$ with probability 0.5. All observations in cluster g get the same value of the weight." (Rademacher weights)

The Wild Cluster Bootstrap

Cameron, Gelbach, and Miller procedure goes as follows (Cameron and Miller 2015, Section VI.C.2):

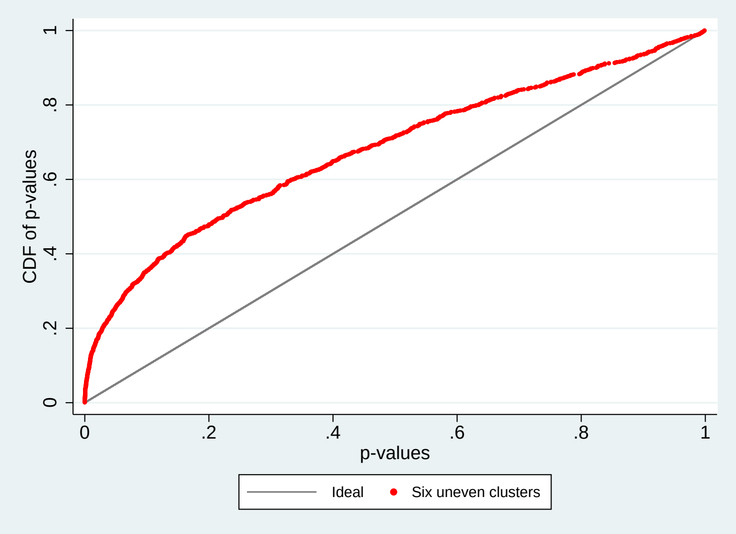
- In each resampling:
  - "Randomly assign cluster g the weight $d_g = -1$ with probability 0.5 and the weight $d_g = 1$ with probability 0.5. All observations in cluster g get the same value of the weight." (Rademacher weights)
  - "Generate new pseudo-residuals $u^*_{ig} = d_g \times \tilde{u}_{ig}$, and hence new outcome variables $y^*_{ig} = x'_{ig}\tilde{\beta}_{H0} + u^*_{ig}$. Then proceed with step 2 as before, regressing y^*_{ig} on x_{ig} [not imposing the null], and calculate w^* [the t-statistic from this regression, with clustered standard errors.]"

The Wild Cluster Bootstrap

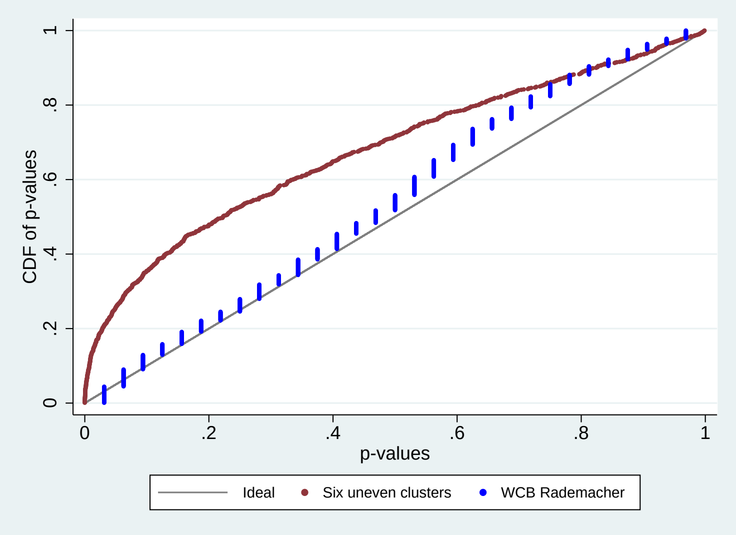
Cameron, Gelbach, and Miller procedure goes as follows (Cameron and Miller 2015, Section VI.C.2):

- In each resampling:
  - "Randomly assign cluster g the weight $d_g = -1$ with probability 0.5 and the weight $d_g = 1$ with probability 0.5. All observations in cluster g get the same value of the weight." (Rademacher weights)
  - "Generate new pseudo-residuals $u_{ig}^* = d_g \times \tilde{u}_{ig}$, and hence new outcome variables $y_{ig}^* = x_{ig}'\tilde{\beta}_{H0} + u_{ig}^*$. Then proceed with step 2 as before, regressing y_{ig}^* on x_{ig} [not imposing the null], and calculate w^* [the t-statistic from this regression, with clustered standard errors.]"
- The p-value for the the test based on the original sample statistic w equals the proportion of times that $|w| > |w_b^*|$.

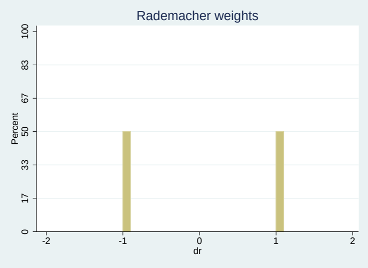
What happens with six clusters



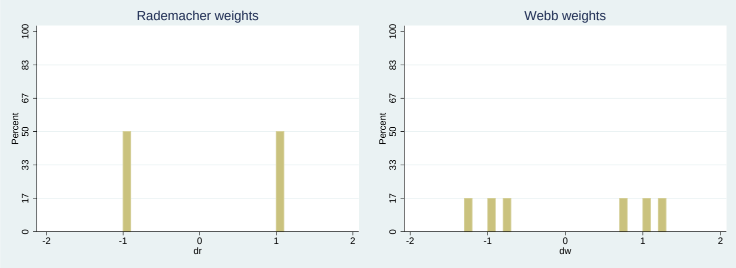
What happens with six clusters



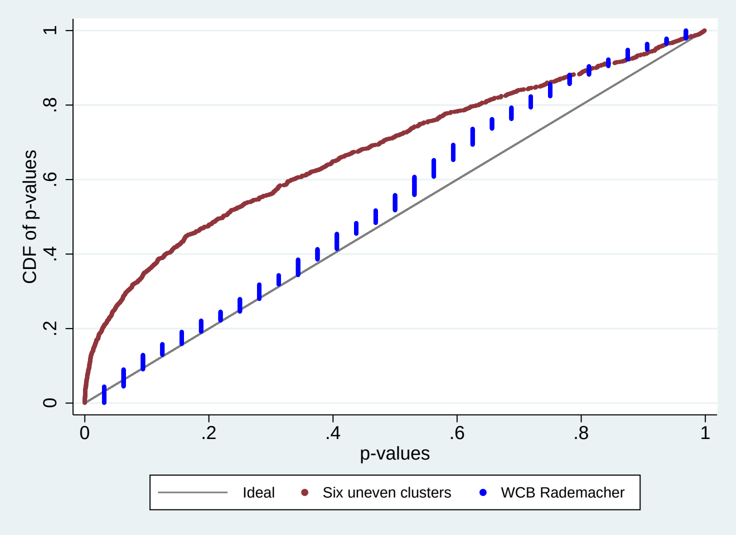
So-called Rademacher and Webb weights



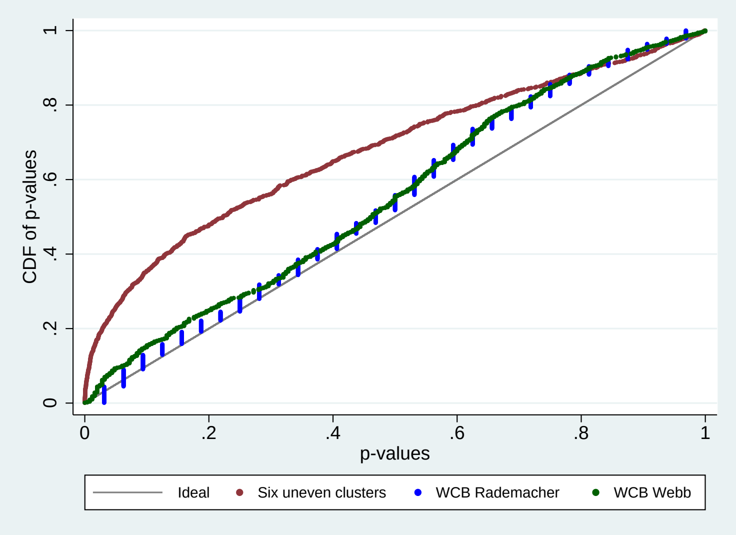
So-called Rademacher and Webb weights



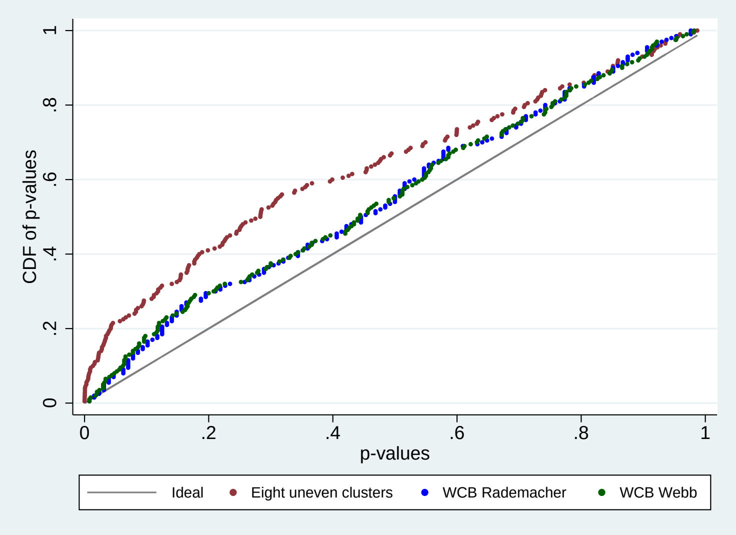
What happens with six clusters



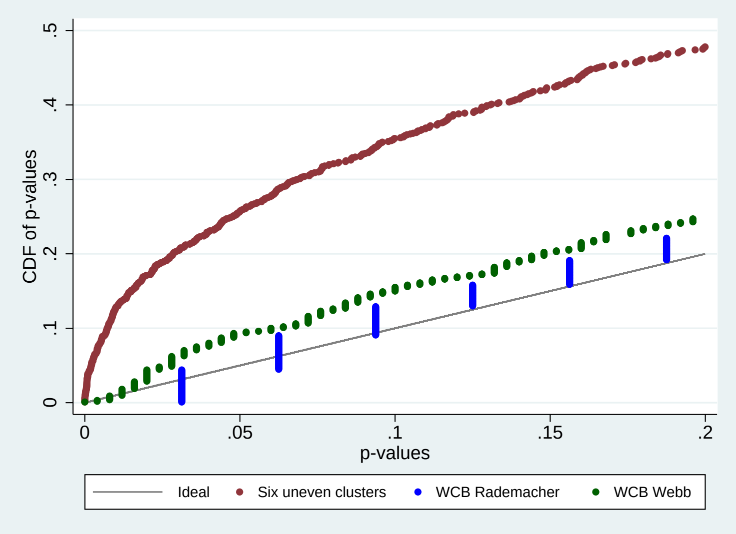
What happens with six clusters



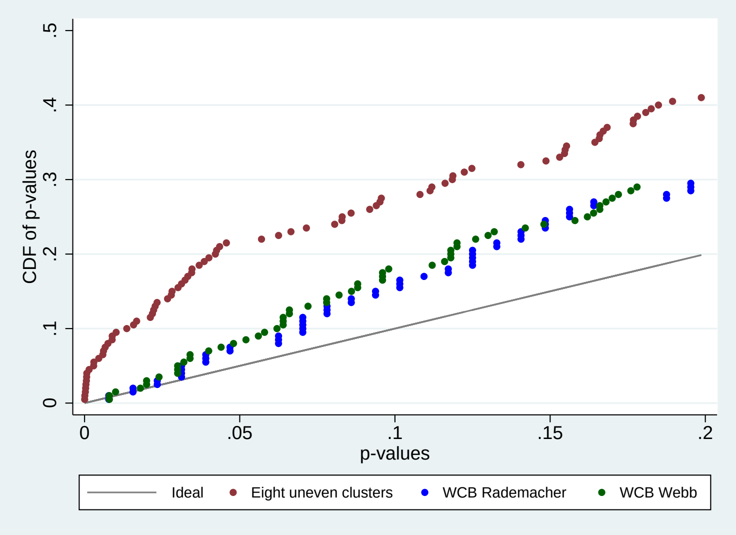
What happens with eight clusters



What happens with six clusters (zoom)



What happens with eight clusters (zoom)



Activity 2