# The Wild Bootstrap with a Small Number of Large Clusters


Ivan A. Canay — Northwestern

Andres Santos — UCLA

Azeem M. Shaikh — U. Chicago


January 4, 2020

# The Question

### Wild Bootstrap

- Prevalent inference method in linear models with few clusters.
- Due to remarkable simulations by Cameron, Gelbach & Miller (2008).
- Simulations show size control with as few as five clusters.

### Examples

- Meng, Qian, and Yared (2015, REStud): 19 clusters.
- Acemoglu, Cantoni, Johnson, Robinson (2011, AER): 13 clusters.
- Giuliano and Spilimbergo (2014, REStud): 9 clusters.
- Kosfeld and Rustagi (2015, AER): 5 clusters.

### The Problem:

- Available theory requires $\#$ clusters $\rightarrow$ infinity.
- Asymptotic properties with few clusters remain unknown.

# The Question

## What We Know

- Simulations have shown wild bootstrap can fail to control size
  ... but not easy to find these designs.
- Justifications are asymptotic as number of clusters diverges
  ... but why does it work with as few as five clusters?
- Small changes to the procedure can affect simulation performance
  ... e.g. why do Rademacher weights do better than Mammen weights?

## This Paper

- Study the performance of the Wild bootstrap with few clusters.
- Study in asymptotic framework where number of clusters is fixed.
- Will Show Wild bootstrap can be valid with few clusters.
- Result requires clusters to be suitably "homogenous."

1. Setup and Notation

2. Main Result

3. Simulation Evidence

# The Model

$$Y_{i,j} = W_{i,j}'\gamma + Z_{i,j}'\beta + \epsilon_{i,j}$$

where $\gamma \in \mathbf{R}^{d_w}$, $\beta \in \mathbf{R}^{d_z}$ and $E[Z_{i,j}\epsilon_{i,j}] = 0$ and $E[W_{i,j}\epsilon_{i,j}] = 0$ ($\forall i, j$).

## Notation

- We index clusters by $j \in J$.
- We index number of clusters by $q = |J|$.
- We index units in the $j^{th}$ cluster by $i \in I_{n,j}$.
- We index number of units in cluster $j$ by $n_j = |I_{n,j}|$.

## Comment

- $\beta$ is main coefficient of interest (e.g. $Z_{i,j} \in \mathbf{R}$).
- $\gamma$ is a nuisance parameter (e.g. $W_{i,j}$ are fixed effects).

# The Test

For some $c \in \mathbf{R}^{d_z}$ and $\lambda \in \mathbf{R}$ we consider the hypothesis testing problem

$$H_0 : c'\beta = \lambda \qquad H_1 : c'\beta \neq \lambda$$

**Test Statistic**

$$T_n \equiv |\sqrt{n}(c'\hat{\beta}_n - \lambda)|$$

where $\hat{\beta}_n$ is the ordinary least squares estimator of $\beta$.

**Wild Bootstrap Test**

$$\phi_n = 1\{T_n > \hat{c}_n(1-\alpha)\}$$

where $\hat{c}_n(1-\alpha)$ is computed using a specific variant of the wild bootstrap.

**Note:** Will comment on properties of the Studentized test statistic later.

# Critical Values

Work with a very specific variant of the wild bootstrap.

**Step 1**

- Run a restricted regression of $Y_{i,j}$ on $(W_{i,j}, Z_{i,j})$ subject to $c'\beta = \lambda$.
- Let $\hat{\gamma}_n^{\mathsf{r}} \in \mathbf{R}^{d_w}$ and $\hat{\beta}_n^{\mathsf{r}} \in \mathbf{R}^{d_z}$ be restricted estimators.
- Let $\hat{\epsilon}_{i,j}^{\mathsf{r}}$ be the corresponding residuals from restricted regression.

**Step 2**

- Let $\{\omega_j\}_{j \in J}$ be i.i.d. with $P(\omega_j = 1) = P(\omega_j = -1) = 1/2$ for all $j \in J$.
- Define $\omega = \{\omega_j\}_{j \in J}$, and for each $\omega$ denote the new outcomes

$$Y_{ij}^*(\omega) \equiv W_{i,j}' \hat{\gamma}_n^{\mathsf{r}} + Z_{i,j}' \hat{\beta}_n^{\mathsf{r}} + \omega_j \hat{\epsilon}_{i,j}^{\mathsf{r}}$$

- Run an unrestricted regression of $Y_{i,j}^*(\omega)$ in $(W_{i,j}, Z_{i,j})$.
- Let $\hat{\gamma}_n^*(\omega)$ and $\hat{\beta}_n^*(\omega)$ be corresponding unrestricted coefficients.

# Critical Values

## Step 3

- Compute the $1-\alpha$ quantile of bootstrap statistic conditional on the data

$$\hat{c}_n(1-\alpha) \equiv \inf\{u \in \mathbf{R} : P(|\sqrt{n}(c'\hat{\beta}^*_n(\omega) - \lambda)| \leq u|\text{Data}) \geq 1-\alpha\}$$

- In practice $\hat{c}_n(1-\alpha)$ approximated via simulation of bootstrap samples.

## Comments

- Bootstrap uses $\hat{\beta}^{\mathsf{r}}_n$ satisfying $c'\hat{\beta}^{\mathsf{r}}_n = \lambda$ (impose the null).
- Use of Rademacher weights is essential for our results.
- Importance of Rademacher vs alternatives known from simulations.

# Preliminary Notation

- Let $\hat{\Pi}_n$ be the $d_w \times d_z$ matrix satisfying the orthogonality conditions

$$\sum_{j \in J} \sum_{i \in I_{n,j}} (Z_{i,j} - \hat{\Pi}_n' W_{i,j}) W_{i,j}' = 0$$

- $(Z_{i,j} - \hat{\Pi}_n' W_{i,j})$ is residual from regressing $Z_{i,j}$ on $W_{i,j}$ on whole sample.

$$\tilde{Z}_{i,j} \equiv (Z_{i,j} - \hat{\Pi}_n' W_{i,j})$$

- Let $\hat{\Pi}_{n,j}^{\mathsf{c}}$ be a $d_w \times d_z$ matrix satisfying the orthogonality conditions

$$\sum_{i \in I_{n,j}} (Z_{i,j} - (\hat{\Pi}_n^{\mathsf{c}})' W_{i,j}) W_{i,j}' = 0$$

**Note:** $\hat{\Pi}_{n,j}^{\mathsf{c}}$ may not be uniquely defined (e.g. include cluster fixed effects)

# Weak Assumption

## Assumption W

(i) The following statistic converges in distribution as $n$ diverges to infinity

$$\frac{1}{\sqrt{n}} \sum_{j \in J} \sum_{i \in I_{n,j}} \begin{pmatrix} W_{i,j}\epsilon_{i,j} \\ Z_{i,j}\epsilon_{i,j} \end{pmatrix}$$

(ii) The following statistic converges (in prob.) to a positive definite matrix

$$\frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \begin{pmatrix} W_{i,j}W'_{i,j} & W_{i,j}Z'_{i,j} \\ Z_{i,j}W'_{i,j} & Z_{i,j}Z'_{i,j} \end{pmatrix}$$

## Comments

- Requirements for showing $\hat{\beta}_n$ and $\hat{\beta}_n^{\mathsf{r}}$ converge in distribution.
- Implicit requirement dependence within cluster weak enough for CLT.
- Imply $\hat{\Pi}_n$ converges in probability to a well defined limit.

# Homogeneity Assumption

## Assumption H

(i) For independent $\{\mathcal{Z}_j\}_{j \in J}$ with $\mathcal{Z}_j \sim N(0, \Sigma_j)$ and $\Sigma_j > 0$ we have

$$\{\frac{1}{\sqrt{n_j}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \epsilon_{i,j} : j \in J\} \xrightarrow{d} \{\mathcal{Z}_j : j \in J\}$$

(ii) For each $j \in J$, $n_j/n \to \xi_j > 0$.

## Comments

- Requirement (i) requires convergence of cluster level "score".
- Requirement (ii) requires clusters not be "too" imbalanced.

# Homogeneity Assumption

### Assumption H

(iii) There are $a_j > 0$ and $\Omega_{\tilde{Z}}$ positive definite such that for each $j \in J$

$$\frac{1}{n_j} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \xrightarrow{p} a_j \Omega_{\tilde{Z}}$$

(iv) For each $j \in J$ it follows that

$$\frac{1}{n_j} \sum_{i \in I_{n,j}} \|W'_{i,j}(\hat{\Pi}_n - \hat{\Pi}^{\mathsf{c}}_{n,j})\|^2 \xrightarrow{p} 0$$

### Comments

- If $Z_{i,j} \in \mathbf{R}$, H(iii) means nonzero limit of $\sum_{i \in I_{n,j}} \tilde{Z}^2_{i,j} / n_j$.
- H(iv) requires convergence of full sample and cluster level projections.

# Some Discussion

For $\gamma \in \mathbf{R}$, $E[\epsilon_{i,j}] = 0$ and $E[Z_{i,j}\epsilon_{i,j}] = 0$ for all $i \in I_{n,j}$ and $j \in J$ suppose

$$Y_{i,j} = \gamma + Z'_{i,j}\beta + \epsilon_{i,j}$$

**Note:** Since here $W_{i,j} = 1$ for all $i \in I_{n,j}$ and $j \in J$ we therefore we have

$$\hat{\Pi}'_n W_{i,j} = \frac{1}{n}\sum_{j \in J}\sum_{i \in I_{n,j}} Z_{i,j} \qquad (\hat{\Pi}^{\mathsf{c}}_n)' W_{i,j} = \frac{1}{n_j}\sum_{i \in I_{n,j}} Z_{i,j}$$

- Hence, Assumption H(iv) (asymptotic equivalence of projections) needs

  Cluster level means are the same (asymptotically)

- While, Assumption H(iii) needs same covariance matrices (up to scaling).

# Some Discussion

For $\gamma \in \mathbf{R}$, $E[\epsilon_{i,j}] = 0$ and $E[Z_{i,j}\epsilon_{i,j}] = 0$ for all $i \in I_{n,j}$ and $j \in J$ suppose

$$Y_{i,j} = \gamma + Z'_{i,j}\beta + \epsilon_{i,j}$$

**Note:** Same model, but estimate with cluster level fixed effects $(W_{i,j})$

$$\hat{\Pi}'_n W_{i,j} = \frac{1}{n_j} \sum_{i \in I_{n,j}} Z_{i,j} \qquad (\hat{\Pi}^{\mathsf{c}}_n)' W_{i,j} = \frac{1}{n_j} \sum_{i \in I_{n,j}} Z_{i,j}$$

- Hence, Assumption H(iv) (equivalence of projections) is automatic.
- While, Assumption H(iii) needs same covariance matrices (up to scaling).

# Main Result

**Theorem** If Assumptions $W$ and $H$ hold and $c'\beta = \lambda$, then it follows that

$$\begin{aligned} \alpha - \frac{1}{2^{q-1}} &\leq \liminf_{n\to\infty} P(T_n > \hat{c}_n(1-\alpha)) \\ &\leq \limsup_{n\to\infty} P(T_n > \hat{c}_n(1-\alpha)) \\ &\leq \alpha \end{aligned}$$

**Comments**

- Wild bootstrap controls size for any number of clusters.
- Conservative, but difference decreases exponentially with $\#$ of clusters.
- Because $q$ fixed, $\hat{c}_n(1-\alpha)$ is not consistent.
- Theorem valid for IV under similar assumptions.

# Additional Comments

## Main Conclusion

- Wild bootstrap provides size control with fixed $\#$ clusters.
- Procedure also works if $q \uparrow \infty$, so Wild bootstrap is "robust" to $q$.

## Proof Comments

- The wild bootstrap is not consistent (i.e. $\hat{c}_n(1-\alpha)$ does not converge).
  ... instead show asymptotic equivalence to randomization test.
- Fundamental to use restricted estimator $\hat{\beta}_n^{\text{r}}$ and Rademacher weights
  ... both these observations are folklore from simulations.
- Similar arguments under studentization, but "ties" are not controlled
  ... instead can show size distortion bounded by $2^{1-q}$.

# Extension: Score Bootstrap

For non-linear models, score bootstrap applies to test statistics satisfying

$$T_n = F(\frac{1}{\sqrt{n}} \sum_{j \in J} \sum_{i \in I_{n,j}} \psi(X_{i,j})) + o_P(1)$$

for known function $F$ and unknown influence function $\psi$.

Using estimator $\hat{\psi}_n$ for $\psi$ obtain critical value from conditional quantile of

$$F(\frac{1}{\sqrt{n}} \sum_{j \in J} \omega_j \sum_{i \in I_{n,j}} \hat{\psi}_n(X_{i,j}))$$

**Comments**

- Asymptotically valid as $q \uparrow \infty$ without "homogeneity" assumptions.
- With "homogeneity" and $q$ fixed, size distortion bounded by $2^{1-q}$.
- Must use "restricted" estimator $\hat{\psi}_n$ and Rademacher weights.

1. Setup and Notation

2. Main Result

3. Simulation Evidence

# Simulation Design

$$Y_{i,j} = \gamma + Z'_{i,j}\beta + \sigma(Z_{i,j})(\eta_j + \epsilon_{i,j})$$

for $1 \leq i \leq n$ and $1 \leq j \leq q$ where we explore four parameter specifications.

**The Good Specifications**

- Model 1: $Z_{i,j} = A_j + \zeta_{i,j}$, $\sigma(Z_{i,j}) = Z^2_{i,j}$, $\gamma = 1$. All variables $N(0,1)$.
- Model 2: As in M.1, but $Z_{i,j} = \sqrt{j}(A_j + \zeta_{i,j})$.

**Note:** Models 1 and 2 need fixed effects to satisfy our assumptions.

# Simulation Design

$$Y_{i,j} = \gamma + Z'_{i,j}\beta + \sigma(Z_{i,j})(\eta_j + \epsilon_{i,j})$$

for $1 \leq i \leq n$ and $1 \leq j \leq q$ where we explore four parameter specifications.

**The Bad Specifications**

- Model 3: As in M.1, but $A_j \sim N(0, I_3)$, $\zeta_{i,j} \sim N(0, \Sigma_j)$, $\beta = (\beta_1, 1, 1)$.
- Model 4: As in M.1, but $\beta = (\beta_1, 2)$, $\sigma(Z_{i,j}) = (Z_{i,j}^{(1)} + Z_{i,j}^{(2)})^2$ with

$$Z_{i,j} \sim N(\mu_1, \Sigma_1) \text{ for } j > q/2$$
$$Z_{i,j} \sim N(\mu_2, \Sigma_2) \text{ for } j \leq q/2$$

where $\mu_1 = (-4, -2)$, $\mu_2 = (2, 4)$, $\Sigma_1 = I_2$ and $\Sigma_2 = \begin{pmatrix} 10 & 0.8 \\ 0.8 & 1 \end{pmatrix}$.

# Size Under Homogeneity

| | Test | Rade - with FEs | | | Rade - without FEs | | | Mammen - with FEs | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $q$ | | | $q$ | | | $q$ | |
| | | 5 | 6 | 8 | 5 | 6 | 8 | 5 | 6 | 8 |
| Model 1 | Non-Stud. | 9.90 | 9.34 | 9.42 | 14.48 | 13.80 | 12.48 | 14.42 | 13.06 | 12.16 |
| $n = 50$ | Stud. | 10.42 | 9.54 | 9.76 | 10.80 | 10.04 | 9.86 | 6.26 | 5.16 | 4.58 |
| Model 2 | Non-Stud. | 9.02 | 9.70 | 9.98 | 15.84 | 15.60 | 15.42 | 13.62 | 13.78 | 13.72 |
| $n = 50$ | Stud | 9.44 | 9.72 | 10.08 | 10.38 | 10.06 | 11.04 | 5.92 | 4.60 | 4.10 |
| Model 1 | Non-Stud. | 9.72 | 9.46 | 10.16 | 15.48 | 14.32 | 14.24 | 14.78 | 13.48 | 12.88 |
| $n = 300$ | Stud | 10.22 | 9.64 | 10.16 | 11.24 | 10.42 | 10.86 | 6.88 | 5.30 | 4.58 |
| Model 2 | Non-Stud. | 9.68 | 9.74 | 10.12 | 17.74 | 16.20 | 15.26 | 14.86 | 14.08 | 13.34 |
| $n = 300$ | Stud | 10.16 | 9.86 | 10.16 | 10.96 | 10.28 | 10.66 | 6.18 | 4.80 | 4.34 |

Table: Rejection prob. (in %) under $H_0$. 5,000 replications. $\alpha = 10\%$

# Size Without Homogeneity

| | Test | Rade - with Fixed effects | | | | Rade - without Fixed effects | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $q$ | | | | $q$ | | | |
| | | 4 | 5 | 6 | 8 | 4 | 5 | 6 | 8 |
| Model 3 | Non-Stud | 11.58 | 13.90 | 13.32 | 13.24 | 26.68 | 37.16 | 32.38 | 26.12 |
| $n = 50$ | Stud | 11.14 | 12.74 | 11.94 | 11.44 | 19.98 | 18.62 | 14.54 | 12.66 |
| Model 4 | Non-Stud | 12.96 | 17.70 | 16.30 | 12.96 | 12.44 | 22.64 | 18.00 | 14.22 |
| $n = 50$ | Stud | 13.00 | 16.34 | 14.62 | 10.88 | 15.24 | 22.68 | 17.22 | 12.84 |
| Model 3 | Non-Stud | 12.26 | 15.10 | 13.52 | 12.66 | 30.10 | 39.08 | 33.26 | 26.06 |
| $n = 300$ | Stud | 12.32 | 13.52 | 11.40 | 10.96 | 22.00 | 19.38 | 15.44 | 12.96 |
| Model 4 | Non-Stud | 13.54 | 17.18 | 15.94 | 12.84 | 14.72 | 24.38 | 17.56 | 13.78 |
| $n = 300$ | Stud | 13.40 | 15.78 | 14.94 | 11.72 | 17.12 | 25.10 | 17.66 | 12.58 |

Table: Rejection prob. (in %) under $H_0$. 5,000 replications. $\alpha = 10\%$

# Conclusion

## The Wild Bootstrap

- Valid under a fixed number of clusters (and still if $q \uparrow \infty$)
- Specific to implementatin with Rademacher weight and "$\hat{\beta}_n^{\mathrm{r}}$".
- Including cluster level fixed effects eases conditions.

## Related to Folklore

- Rademacher weights outperform Mammen despite large $q$ theory.
- "Imposing the null" has dramatic effects in simulations.
- Certain "heterogeneous" designs negatively affect wild bootstrap.

## Extensions

- Results apply to nonlinear models through the score bootstrap.
- Can be shown to over-reject by at most $2^{1-q}$.
- "Homogeneity" assumptions can be stringent due to nonlinearity.