The Wild Bootstrap with a Small Number of Large Clusters

Ivan A. Canay                        Andres Santos                        Azeem M. Shaikh
Northwestern                        UCLA                                    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.

The Problem:

- Available theory requires \# clusters → 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 \( \gamma \in \mathbb{R}^{d_w}, \beta \in \mathbb{R}^{d_z} \) 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. \( Z_{i,j} \in \mathbb{R} \)).
- \( \gamma \) is a nuisance parameter (e.g. \( W_{i,j} \) are fixed effects).
The Test

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

$$H_0 : c' \beta = \lambda \quad \text{and} \quad H_1 : c' \beta \neq \lambda$$

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 \( \hat{\gamma}_r n \in \mathbb{R}^{d_w} \) and \( \hat{\beta}_r n \in \mathbb{R}^{d_z} \) be restricted estimators.
- Let \( \hat{\epsilon}_{i,j}^r \) 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
  \[
  Y_{i,j}^*(\omega) \equiv W_{i,j}' \hat{\gamma}_n^r + Z_{i,j}' \hat{\beta}_n^r + \omega_j \hat{\epsilon}_{i,j}^r
  \]
- 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

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

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

Comments

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

2 Main Result

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

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

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

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

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

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

\[
\frac{1}{\sqrt{n}} \sum_{j \in J} \sum_{i \in I_{n,j}} \left( \begin{array}{c} W_{i,j} e_{i,j} \\ Z_{i,j} e_{i,j} \end{array} \right)
\]

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

\[
\frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \left( \begin{array}{ccc} W_{i,j} W'_{i,j} & W_{i,j} Z'_{i,j} \\ Z_{i,j} W'_{i,j} & Z_{i,j} Z'_{i,j} \end{array} \right)
\]

**Comments**

- Requirements for showing \( \hat{\beta}_n \) and \( \hat{\beta}_n^r \) 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 \( \{ Z_j \}_{j \in J} \) with \( Z_j \sim N(0, \Sigma_j) \) and \( \Sigma_j > 0 \) we have

\[
\left\{ \frac{1}{\sqrt{n_j}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \, : \, j \in J \right\} \overset{d}{\rightarrow} \{ Z_j : j \in J \}
\]

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

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

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

Comments

• If $Z_{i,j} \in \mathbb{R}$, H(iii) means nonzero limit of $\sum_{i \in I_{n,j}} \tilde{Z}_{i,j}^2/n_j$.

• H(iv) requires convergence of full sample and cluster level projections.
For $\gamma \in \mathbb{R}$, $E[\epsilon_{i,j}] = 0$ 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

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

- 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).
For $\gamma \in \mathbb{R}$, $E[\epsilon_{i,j}] = 0$ 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 $(W_{i,j})$

$$\hat{\Pi}'_{n}W_{i,j} = \frac{1}{n_{j}} \sum_{i \in I_{n,j}} Z_{i,j}$$

$$(\hat{\Pi}^{c}_{n})'W_{i,j} = \frac{1}{n_{j}} \sum_{i \in I_{n,j}} Z_{i,j}$$

- 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

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

**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.
Main Conclusion

- Wild bootstrap provides size control with fixed \( \neq \) 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 \( \hat{\beta}_r_n \) 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

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

for known function \( F \) and unknown influence function \( \psi \).

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

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

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 \(\beta = (\beta_1, 2), \sigma(Z_{i,j}) = (Z^{(1)}_{i,j} + Z^{(2)}_{i,j})^2\) 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

<table>
<thead>
<tr>
<th>Test</th>
<th>Rade - with FEs</th>
<th>Rade - without FEs</th>
<th>Mammen - with FEs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$q=5$</td>
<td>$q=6$</td>
<td>$q=8$</td>
</tr>
<tr>
<td></td>
<td>Stud.</td>
<td>10.42</td>
<td>9.54</td>
</tr>
<tr>
<td>$n = 50$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>Non-Stud.</td>
<td>9.02</td>
<td>9.70</td>
</tr>
<tr>
<td></td>
<td>Stud</td>
<td>9.44</td>
<td>9.72</td>
</tr>
<tr>
<td>$n = 50$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stud</td>
<td>10.22</td>
<td>9.64</td>
</tr>
<tr>
<td>$n = 300$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stud</td>
<td>10.16</td>
<td>9.86</td>
</tr>
<tr>
<td>$n = 300$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Rejection prob. (in %) under $H_0$. 5,000 replications. $\alpha = 10\%$
### Size Without Homogeneity

<table>
<thead>
<tr>
<th>Test</th>
<th>Rade - with Fixed effects</th>
<th>Rade - without Fixed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$q$</td>
<td>4</td>
</tr>
<tr>
<td>$n = 50$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>Non-Stud</td>
<td>12.96</td>
</tr>
<tr>
<td></td>
<td>Stud</td>
<td>13.00</td>
</tr>
<tr>
<td>$n = 50$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>Non-Stud</td>
<td>12.26</td>
</tr>
<tr>
<td></td>
<td>Stud</td>
<td>12.32</td>
</tr>
<tr>
<td>$n = 300$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td>Non-Stud</td>
<td>13.54</td>
</tr>
<tr>
<td></td>
<td>Stud</td>
<td>13.40</td>
</tr>
</tbody>
</table>

**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 implementation with Rademacher weight and \( \hat{\beta}^r_n \).
- 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.