The Wild Bootstrap with a Small Number of Large Clusters

Ivan A. Canay — Northwestern
Andres Santos — UCLA
Azeem M. Shaikh — U. Chicago

October 4, 2019

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 $\to$ 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.

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?

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?

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".

Related Literature

Wild Bootstrap

Liu (1988), Mammen (1993), Davidson & Mackinnon (1999), Cameron, Gelbach & Miller (2008), Davidson & Flachaire (2008), Kline & Santos (2012a, 2012b), Webb (2013), Mackinnon, Nielsen & Webb (2017).

$\Rightarrow$ These results do not explain performance with (as few as) five clusters.

Related Literature

Wild Bootstrap

Liu (1988), Mammen (1993), Davidson & Mackinnon (1999), Cameron, Gelbach & Miller (2008), Davidson & Flachaire (2008), Kline & Santos (2012a, 2012b), Webb (2013), Mackinnon, Nielsen & Webb (2017).

$\Rightarrow$ These results do not explain performance with (as few as) five clusters.

Fixed Number of Clusters

Ibragimov & Muller (2010, 2016), Bester, Conley & Hansen (2011), Canay, Romano & Shaikh (2017), Hagemann (2019).

$\Rightarrow$ Study alternative procedures with fixed number of clusters.

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 $E[W_{i,j}\epsilon_{i,j}] = 0$ $(\forall i, j)$.

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 $n_j = |I_{n,j}|$.

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

- β is main coefficient of interest (e.g. $Z_{i,j} \in \mathbf{R}$).
- γ 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

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

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 β.

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 β.

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: We will study properties of the Studentized test statistic later.

Critical Values

Work with a very specific variant of the wild bootstrap.

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.

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 ω 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.

Different Interpretation

Key: Under fixed number of clusters, distribution of $\{\omega_j\}_{j \in J}$ fixed with n.

Different Interpretation

Key: Under fixed number of clusters, distribution of $\{\omega_j\}_{j \in J}$ fixed with n.

Observations

- Let $\mathbf{G} \equiv \{-1, 1\}^q$, which corresponds to the support of $\omega = \{\omega_j\}_{j \in J}$.
- Every $(g_1, \ldots, g_q) = g \in \mathbf{G}$ is then a possible realization of $\omega = \{\omega_j\}_{j \in J}$.
- Note that $P(\omega = g) = 1/|\mathbf{G}|$ for every $g \in \mathbf{G}$ (all equally likely).

Different Interpretation

Key: Under fixed number of clusters, distribution of $\{\omega_j\}_{j \in J}$ fixed with n.

Observations

- Let $\mathbf{G} \equiv \{-1, 1\}^q$, which corresponds to the support of $\omega = \{\omega_j\}_{j \in J}$.
- Every $(g_1, \ldots, g_q) = g \in \mathbf{G}$ is then a possible realization of $\omega = \{\omega_j\}_{j \in J}$.
- Note that $P(\omega = g) = 1/|\mathbf{G}|$ for every $g \in \mathbf{G}$ (all equally likely).

Abuse Notation Write $\hat{\beta}_n^*(g)$ and $\hat{\gamma}_n^*(g)$ in place of $\hat{\beta}_n^*(\omega)$ and $\hat{\gamma}_n^*(\omega)$.

$$\begin{aligned} \hat{c}_n(1-\alpha) &\equiv \inf\{u \in \mathbf{R} : P(|\sqrt{n}(c'\hat{\beta}_n^*(\omega) - \lambda)| \le u | \text{Data}) \ge 1 - \alpha\} \\ &= \inf\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} 1\{|\sqrt{n}(c'\hat{\beta}_n^*(g) - \lambda)| \le u\} \ge 1 - \alpha\} \end{aligned}$$

Preliminary Notation

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

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

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.

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

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

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

Note: $\hat{\Pi}^{\mathsf{c}}_{n,j}$ 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

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

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

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

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)

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 $(W_{i,j})$

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.

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

The Test Statistic

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

The Test Statistic

- Suppose the null hypothesis is true so that $c'\beta = \lambda$. Then it follows that

$$T_n = \sqrt{n}|c'\hat{\beta}_n - \lambda| = \sqrt{n}|c'(\hat{\beta}_n - \beta)| = |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{1}{\sqrt{n}} \sum_{i \in I_j} Z_{i,j}\epsilon_{i,j}|$$

where $\hat{\Omega}_n \equiv \sum_{j \in J} \sum_{i \in I_{n,j}} Z_{i,j} Z'_{i,j}/n$ is usual $d_z \times d_z$ matrix.

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

The Test Statistic

- Suppose the null hypothesis is true so that $c'\beta = \lambda$. Then it follows that

$$T_n = \sqrt{n}|c'\hat{\beta}_n - \lambda| = \sqrt{n}|c'(\hat{\beta}_n - \beta)| = |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{1}{\sqrt{n}} \sum_{i \in I_j} Z_{i,j}\epsilon_{i,j}|$$

where $\hat{\Omega}_n \equiv \sum_{j \in J} \sum_{i \in I_{n,j}} Z_{i,j}Z'_{i,j}/n$ is usual $d_z \times d_z$ matrix.

- Therefore, for an appropriate function T we can write T_n as

$$T_n = T(S_n) \qquad S_n = (\hat{\Omega}_n, \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} Z_{i,j}\epsilon_{i,j}\}_{j \in J})$$

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

The Bootstrap Statistic

- Since $\hat{\beta}_n^{\mathsf{r}}$ satisfies $c'\hat{\beta}_n^{\mathsf{r}} = \lambda$ by construction, it then follows that

$$\sqrt{n}|c'\hat{\beta}_n^*(g) - \lambda| = \sqrt{n}|c'(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathsf{r}})| = |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{1}{\sqrt{n}} \sum_{i \in I_j} g_j Z_{i,j} \hat{\epsilon}_{i,j}^{\mathsf{r}}|$$

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

The Bootstrap Statistic

- Since $\hat{\beta}_n^{\mathsf{r}}$ satisfies $c'\hat{\beta}_n^{\mathsf{r}} = \lambda$ by construction, it then follows that

$$\sqrt{n}|c'\hat{\beta}_n^*(g) - \lambda| = \sqrt{n}|c'(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathsf{r}})| = |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{1}{\sqrt{n}} \sum_{i \in I_j} g_j Z_{i,j} \hat{\epsilon}_{i,j}^{\mathsf{r}}|$$

- Therefore, for the same function T characterizing T_n it follows that

$$\sqrt{n}|c'\hat{\beta}_n^*(g) - \lambda| = T(gS_n^*) \qquad gS_n^* = (\hat{\Omega}_n, \{\frac{g_j}{\sqrt{n}} \sum_{i \in I_{n,j}} Z_{i,j} \hat{\epsilon}_{i,j}^{\mathsf{r}}\}_{j \in J})$$

for any $(g_1, \ldots, g_q) = g \in \mathbf{G}$, where recall $\mathbf{G} = \{-1, 1\}^q$.

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

The Critical Value

- Since $\hat{\beta}_n^{\mathsf{r}}$ satisfies $c'\hat{\beta}_n^{\mathsf{r}} = \lambda$ by construction, it then follows that

$$\begin{aligned} \hat{c}_n(1-\alpha) &= 1-\alpha \text{ quantile of } |\sqrt{n}c'(\hat{\beta}_n^*(g)-\lambda)| \text{ over } g \in \mathbf{G} \\ &= 1-\alpha \text{ quantile of } T(gS_n^*) \text{ over } g \in \mathbf{G} \end{aligned}$$

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

The Critical Value

- Since $\hat{\beta}_n^{\mathsf{r}}$ satisfies $c'\hat{\beta}_n^{\mathsf{r}} = \lambda$ by construction, it then follows that

$$\begin{aligned}\hat{c}_n(1-\alpha) &= 1-\alpha \text{ quantile of } |\sqrt{n}c'(\hat{\beta}_n^*(g)-\lambda)| \text{ over } g \in \mathbf{G} \\ &= 1-\alpha \text{ quantile of } T(gS_n^*) \text{ over } g \in \mathbf{G}\end{aligned}$$

- Equivalently, let $T^{(k)}(S_n^*|\mathbf{G})$ be the k^{th} smallest value of $\{T(gS_n^*)\}_{g\in\mathbf{G}}$

$$T^{(1)}(gS_n^*|\mathbf{G}) \le \cdots \le \underbrace{T^{(\lceil|\mathbf{G}|(1-\alpha)\rceil)}(gS_n^*|\mathbf{G})}_{\hat{c}_n(1-\alpha)} \le \cdots \le T^{(|\mathbf{G}|)}(gS_n^*|\mathbf{G})$$

Sketch of Proof

Step 1 Rewrite the test to show (asymptotic) connection to randomized test.

$$T_n > \hat{c}_n(1-\alpha) \text{ or equivalently } T(S_n) > T^{(|\mathbf{G}|(1-\alpha))}(gS_n^*|\mathbf{G})$$

Comments

- If S_n equaled S_n^*, it would resemble a randomization test.
- Since the number of clusters is fixed, $\mathbf{G}$ is not changing.
- Showing bootstrap validity needs "non-standard" arguments.

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

$$|T(gS_n) - T(gS_n^*)| \leq |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{n_j}{n} \frac{1}{n_j} \sum_{i \in I_{n,j}} g_j Z_{i,j} Z_{i,j}' \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}|$$

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

$$|T(gS_n) - T(gS_n^*)| \leq |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{n_j}{n} \frac{1}{n_j} \sum_{i \in I_{n,j}} g_j Z_{i,j} Z_{i,j}' \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}|$$

$$(\Omega_j = E[Z_{i,j} Z_{i,j}'] + \text{LLN}) = |c'(\sum_{j \in J} \frac{n_j}{n} \Omega_j)^{-1} \sum_{j \in J} \frac{n_j}{n} \Omega_j g_j \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

$$|T(gS_n) - T(gS_n^*)| \leq |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{n_j}{n} \frac{1}{n_j} \sum_{i \in I_{n,j}} g_j Z_{i,j} Z_{i,j}' \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}|$$

$$(\Omega_j = E[Z_{i,j}Z_{i,j}'] + \text{LLN}) = |c'(\sum_{j \in J} \frac{n_j}{n} \Omega_j)^{-1} \sum_{j \in J} \frac{n_j}{n} \Omega_j g_j \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

$$(\text{Homogeneity}) = |c' \underbrace{(\sum_{j \in J} \frac{n_j}{n} \Omega_{\tilde{Z}})^{-1} \Omega_{\tilde{Z}}}_{\approx \text{ Identity}} \sum_{j \in J} \frac{n_j}{n} g_j \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

$$|T(gS_n) - T(gS_n^*)| \leq |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{n_j}{n} \frac{1}{n_j} \sum_{i \in I_{n,j}} g_j Z_{i,j} Z_{i,j}' \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}|$$

$$(\Omega_j = E[Z_{i,j}Z_{i,j}'] + \text{LLN}) = |c'(\sum_{j \in J} \frac{n_j}{n}\Omega_j)^{-1} \sum_{j \in J} \frac{n_j}{n}\Omega_j g_j \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

$$(\text{Homogeneity}) = |c' \underbrace{(\sum_{j \in J} \frac{n_j}{n}\Omega_{\tilde{Z}})^{-1}\Omega_{\tilde{Z}}}_{\approx \text{ Identity}} \sum_{j \in J} \frac{n_j}{n} g_j \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

$$(\text{Push } c' \text{ through}) \approx |\sum_{j \in J} \frac{n_j}{n} g_j c' \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

$$|T(gS_n) - T(gS_n^*)| \leq |c'\hat{\Omega}_n^{-1} \sum_{j \in J} \frac{n_j}{n} \frac{1}{n_j} \sum_{i \in I_{n,j}} g_j Z_{i,j} Z'_{i,j} \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}|$$

$$(\Omega_j = E[Z_{i,j}Z'_{i,j}] + \text{LLN}) = |c'(\sum_{j \in J} \frac{n_j}{n} \Omega_j)^{-1} \sum_{j \in J} \frac{n_j}{n} \Omega_j g_j \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

$$(\text{Homogeneity}) = |c' \underbrace{(\sum_{j \in J} \frac{n_j}{n} \Omega_{\tilde{Z}})^{-1} \Omega_{\tilde{Z}}}_{\approx \text{ Identity}} \sum_{j \in J} \frac{n_j}{n} g_j \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

$$(\text{Push } c' \text{ through}) \approx |\sum_{j \in J} \frac{n_j}{n} g_j c' \sqrt{n}\{\beta - \hat{\beta}_n^{\mathsf{r}}\}| + o_p(1)$$

$$(\text{Use } c'\hat{\beta}_n^{\mathsf{r}} = c'\beta) = o_p(1)$$

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

So Far We have shown $T(gS_n) = T(gS_n^*) + o_p(1)$ for any $g \in \mathbf{G}$.

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

So Far We have shown $T(gS_n) = T(gS_n^*) + o_p(1)$ for any $g \in \mathbf{G}$.

In addition If $g = \pm(1, \dots, 1)$, then $T(gS_n) = T(gS_n^*)$ (same arguments)

Sketch of Proof

Step 2 Employ the homogeneity assumptions to relate $T(gS_n^*)$ to $T(gS_n)$

So Far We have shown $T(gS_n) = T(gS_n^*) + o_p(1)$ for any $g \in \mathbf{G}$.

In addition If $g = \pm(1, \ldots, 1)$, then $T(gS_n) = T(gS_n^*)$ (same arguments)

Therefore

$T_n > \hat{c}_n(1-\alpha)$ or equivalently $T(S_n) > T^{(|\mathbf{G}|(1-\alpha))}(gS_n^*|\mathbf{G})$

or w.p.a. one $T(S_n) > T^{(|\mathbf{G}|(1-\alpha))}(gS_n|\mathbf{G})$

Comments

- Using restricted estimator $\hat{\beta}_n^{\mathrm{r}}$ plays fundamental role.
- Ensuring $T(gS_n) = T(gS_n^*)$ for $g = \pm(1, \ldots, 1)$ fundamental for ties.

Sketch of Proof

Step 3 Establish asymptotic connection to randomization test to conclude.

Sketch of Proof

Step 3 Establish asymptotic connection to randomization test to conclude.

$$S_n \equiv (\hat{\Omega}_n, \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} Z_{i,j} \epsilon_{i,j}\}_{j \in J}) \stackrel{d}{\to} (\Omega_{\tilde{Z}}, \{\mathcal{Z}_j\}_{j \in J}) \equiv S$$

Sketch of Proof

Step 3 Establish asymptotic connection to randomization test to conclude.

$$S_n \equiv (\hat{\Omega}_n, \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} Z_{i,j}\epsilon_{i,j}\}_{j \in J}) \xrightarrow{d} (\Omega_{\tilde{Z}}, \{\mathcal{Z}_j\}_{j \in J}) \equiv S$$

Therefore

$$\begin{aligned} P(T_n > \hat{c}_n(1-\alpha)) &= P(T(S_n) > T^{(|\mathbf{G}|(1-\alpha)|)}(gS_n|\mathbf{G})) + o(1) \\ &\to P(T(S) > T^{(|\mathbf{G}|(1-\alpha))}(gS|\mathbf{G})) \end{aligned}$$

Sketch of Proof

Step 3 Establish asymptotic connection to randomization test to conclude.

$$S_n \equiv (\hat{\Omega}_n, \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} Z_{i,j}\epsilon_{i,j}\}_{j\in J}) \xrightarrow{d} (\Omega_{\tilde{Z}}, \{\mathcal{Z}_j\}_{j\in J}) \equiv S$$

Therefore

$$\begin{aligned} P(T_n > \hat{c}_n(1-\alpha)) &= P(T(S_n) > T^{(\lceil|\mathbf{G}|(1-\alpha)\rceil)}(gS_n|\mathbf{G})) + o(1) \\ &\to P(T(S) > T^{(\lceil|\mathbf{G}|(1-\alpha)\rceil)}(gS|\mathbf{G})) \end{aligned}$$

Finally since $gS \stackrel{d}{=} S$ for all $g \in \mathbf{G}$, properties of randomization tests imply

$$P(T(S) > T^{(\lceil|\mathbf{G}|(1-\alpha)\rceil)}(gS|\mathbf{G})) \leq \alpha$$

Additional Comments

Main Conclusion

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

Procedure Comments

- Fundamental to use restricted estimator $\hat{\beta}_n^{\mathrm{r}}$.
- Fundamental to use Rademacher weights.
- Both these observations are folklore from simulations.

Proof Comments

- The wild bootstrap is not consistent (i.e. $\hat{c}_n(1-\alpha)$ does not converge).
- Instead wild bootstrap behaves like randomization test.

Studentized Test

The Test Statistic

$$T_n^{\mathsf{s}} \equiv \frac{|\sqrt{n}(c'\hat{\beta}_n - \lambda)|}{\hat{\sigma}_n}$$

where $\hat{\sigma}_n$ are cluster robust s.e.; i.e. $\hat{\epsilon}_{i,j} \equiv (Y_{i,j} - W'_{i,j}\hat{\gamma}_n - Z'_{i,j}\hat{\beta}_n)$ and

$$\hat{\sigma}_n^2 = c'\hat{\Omega}_n^{-1}\hat{V}_n\hat{\Omega}_n^{-1}c \qquad \hat{V}_n = \frac{1}{n}\sum_{j\in J}\sum_{i\in I_j}\sum_{s\in I_j}\tilde{Z}_{i,j}\tilde{Z}'_{s,j}\hat{\epsilon}_{i,j}\hat{\epsilon}_{s,j}$$

The Bootstrap

- Wild bootstrap critical values adjusted accordingly.
- Wild bootstrap s.e. use wild bootstrap residuals from $\hat{\beta}_n^*(g)$.
- Write the resulting wild bootstrap critical value as $\hat{c}_n^{\mathsf{s}}(1-\alpha)$.

Studentized Test

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

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

Comments

- Problem: Unlike unstudentized version "ties" matter ($T_n^{\mathsf{s}} = \hat{c}_n^{\mathsf{s}}(1-\alpha)$).
- But: Probability of tie asymp. only $1/2^{q-1} \Rightarrow$ Small distortion.
- Similar intuition extends to nonlinear estimators and hypotheses.

Score Bootstrap (Sketch)

Test Statistic

$$T^{\mathrm{F}}(S_n) = F(S_n) + o_p(1)$$

where F is a known function, and S_n is the cluster level scores given by

$$S_n \equiv \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \psi(X_{i,j}) : j \in J\}$$

Critical Value

$$\hat{c}_n^{\mathrm{F}}(1-\alpha) \equiv \inf\{u : \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} 1\{F(g\hat{S}_n) \leq u\} \geq 1-\alpha\}$$

where $g\hat{S}_n$ are "perturbed" estimates for the cluster level scores given by

$$g\hat{S}_n \equiv \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} g_j \hat{\psi}_n(X_{i,j}) : j \in J\}$$

Score Bootstrap (Sketch)

Main Assumption (M)

$$\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \hat{\psi}_n(X_{i,j}) = \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \psi(X_{i,j}) + o_p(1)$$

To verify: use constrained estimator and "homogeneity" condition.

Example (GMM)

- For some $m(X_{i,j}, \cdot) : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_m}$ parameter $\beta \in \mathbf{R}^{d_\beta}$ satisfies

$$E[m(X_{i,j}, \beta)] = 0$$

- If T_n is Wald test-statistic based on GMM estimator, key condition is

$$\frac{1}{n} \sum_{i \in I_{n,j}} \nabla m(X_{i,j}, \hat{\beta}_n) \xrightarrow{p} a_j D(\beta)$$

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 $\Sigma_2 = \begin{pmatrix} 10 & 0.8 \\ 0.8 & 1 \end{pmatrix}$.

The Tests

The Tests We Consider

un-Stud: un-studentized test.

Stud: Studentized test.

ET-US: Equi-tail analog of the un-Stud test above.
Reject if $T_n < \hat{c}_n(\alpha/2)$ or $T_n > \hat{c}_n(1-\alpha/2)$.

ET-S: Same as ET-US but with studentized test statistic.

Variants of These Tests

- Implemented with or without cluster-lvl fixed effects
- Implemented with Rademacher or Mammen weights.

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
	Non-Stud.	9.90	9.34	9.42	14.48	13.80	12.48	14.42	13.06	12.16
Model 1	Stud.	10.42	9.54	9.76	10.80	10.04	9.86	6.26	5.16	4.58
$n = 50$	ET-NS	7.40	9.64	9.26	11.42	14.00	12.16	3.14	3.30	4.74
	ET-S	8.64	9.90	9.52	8.34	10.32	9.46	25.72	24.32	22.04
	Non-Stud.	9.02	9.70	9.98	15.84	15.60	15.42	13.62	13.78	13.72
Model 2	Stud	9.44	9.72	10.08	10.38	10.06	11.04	5.92	4.60	4.10
$n = 50$	ET-NS	6.68	9.88	9.72	12.44	15.68	15.00	1.54	2.22	3.58
	ET-S	7.60	10.34	9.88	8.30	10.24	10.80	25.42	25.26	25.40
	Non-Stud.	9.72	9.46	10.16	15.48	14.32	14.24	14.78	13.48	12.88
Model 1	Stud	10.22	9.64	10.16	11.24	10.42	10.86	6.88	5.30	4.58
$n = 300$	ET-NS	7.14	9.66	9.84	12.00	14.42	13.82	2.66	3.62	4.70
	ET-S	8.12	10.12	9.92	8.78	10.74	10.56	25.08	24.38	24.14
	Non-Stud.	9.68	9.74	10.12	17.74	16.20	15.26	14.86	14.08	13.34
Model 2	Stud	10.16	9.86	10.16	10.96	10.28	10.66	6.18	4.80	4.34
$n = 300$	ET-NS	7.26	10.00	9.96	13.60	16.24	14.74	1.80	2.36	3.40
	ET-S	8.16	10.42	9.88	8.00	10.44	10.40	26.80	26.66	25.42

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

Size Without Homogeneity

	Test	Rade - with Fixed effects q 4	5	6	8	Rade - without Fixed effects q 4	5	6	8
	Non-Stud	11.58	13.90	13.32	13.24	26.68	37.16	32.38	26.12
Model 3	Stud	11.14	12.74	11.94	11.44	19.98	18.62	14.54	12.66
$n = 50$	ET-NS	5.62	10.82	12.78	12.92	8.66	31.40	33.18	25.62
	ET-S	7.06	10.24	11.34	11.38	13.52	16.08	15.10	12.46
	Non-Stud	12.96	17.70	16.30	12.96	12.44	22.64	18.00	14.22
Model 4	Stud	13.00	16.34	14.62	10.88	15.24	22.68	17.22	12.84
$n = 50$	ET-NS	5.52	14.68	16.56	12.72	3.60	19.08	18.20	14.02
	ET-S	7.62	14.30	15.10	10.76	9.60	20.70	17.66	12.74
	Non-Stud	12.26	15.10	13.52	12.66	30.10	39.08	33.26	26.06
Model 3	Stud	12.32	13.52	11.40	10.96	22.00	19.38	15.44	12.96
$n = 300$	ET-NS	5.88	12.20	14.14	12.38	14.20	32.34	16.14	12.74
	ET-S	8.20	11.86	11.94	10.74	17.80	16.70	13.00	11.98
	Non-Stud	13.54	17.18	15.94	12.84	14.72	24.38	17.56	13.78
Model 4	Stud	13.40	15.78	14.94	11.72	17.12	25.10	17.66	12.58
$n = 300$	ET-NS	5.60	13.98	16.36	12.68	4.32	19.66	17.80	13.60
	ET-S	7.88	13.38	15.46	11.56	10.42	22.16	18.14	12.36

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 "$\hat{\beta}_n^{\mathsf{r}}$".
- Including cluster level fixed effects eases conditions.
- Studentized may over-reject (but negligible)

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.