Bootstrap-Based Improvements for Inference with Clustered Errors

Colin Cameron, Jonah Gelbach, Doug Miller
U.C. - Davis, U. Maryland, U.C. - Davis

May, 2008
1. Introduction

- OLS regression with **individual-level data** \((i)\)
  with **cluster or grouping** \((g)\)

\[ y_{ig} = x'_{ig} \beta + u_{ig}, \quad i = 1, \ldots, N_g, \quad g = 1, \ldots, G. \]

- E.g. \(y_{ig}\) is (log) wage of \(ith\) individual who lives in state \(g\)
  and \(x_{ig}\) includes regressor correlated within state \(g\)
  and \(u_{ig}\) is error correlated within state \(g\).
Default OLS standard errors based on \( s^2(X'X)^{-1} \) that ignore clustering can greatly understate true standard errors. See Moulton (1986, 1990).
Default OLS standard errors based on $s^2(\mathbf{X}'\mathbf{X})^{-1}$ that ignore clustering can greatly understate true standard errors. See Moulton (1986, 1990).

Standard solution is to get cluster-robust (CR) standard errors. This is the cluster option in Stata.
• Default OLS standard errors based on $s^2(X'X)^{-1}$ that ignore clustering can greatly understate true standard errors. See Moulton (1986, 1990).

• Standard solution is to get cluster-robust (CR) standard errors. This is the cluster option in Stata.

• If the number of clusters $G$ is small then:
  1. CR standard errors are downwards biased (too small)
  2. $t$ statistic is not standard normal distributed
  3. $t$-tests using standard normal critical values over-reject.
Default OLS standard errors based on $s^2(X'X)^{-1}$ that ignore clustering can greatly understate true standard errors. See Moulton (1986, 1990).

Standard solution is to get cluster-robust (CR) standard errors. This is the cluster option in Stata.

If the number of clusters $G$ is small then:
1. CR standard errors are downwards biased (too small)
2. t statistic is not standard normal distributed
3. t-tests using standard normal critical values over-reject.

This paper is concerned with **better inference when there are few clusters.**
Wrong standard errors and critical values: 5%
Outline of Talk

1. Introduction
2. Cluster-Robust Inference
3. Cluster Bootstrap (without and with refinement)
4. Monte Carlo Simulations
7. Conclusion
2.1 OLS with Cluster Errors

- Model for $G$ clusters with $N_g$ individuals per cluster:

$$y_{ig} = x'_{ig} \beta + u_{ig}, \quad i = 1, \ldots, N_g, \; g = 1, \ldots, G,$$
$$y_g = X_g \beta + u_g, \quad g = 1, \ldots, G,$$
$$y = X \beta + u,$$

- OLS estimator

$$\hat{\beta} = (\sum_{g=1}^{G} \sum_{i=1}^{N_g} x_{ig} x'_{ig})^{-1} (\sum_{g=1}^{G} \sum_{i=1}^{N_g} x_{ig} y_{ig})$$
$$= (\sum_{g=1}^{G} X'_g X_g)^{-1} (\sum_{g=1}^{G} X_g y_g)$$
$$= (X'X)^{-1} X' y.$$
OLS with Clustered Errors

As usual

\[
\hat{\beta} = \beta + (X'X)^{-1}X'u \\
= \beta + (X'X)^{-1} \left( \sum_{g=1}^{G} X_g u_g \right).
\]
As usual

\[ \hat{\beta} = \beta + (X'X)^{-1}X'u \]
\[ = \beta + (X'X)^{-1}(\sum_{g=1}^{G} X_g u_g). \]

Assume independence over \( g \) with

\[ u_g \sim [0, \Sigma_g = E[u_g u'_g]]. \]
As usual

\[ \hat{\beta} = \beta + (X'X)^{-1}X'u \]

\[ = \beta + (X'X)^{-1}(\sum_{g=1}^{G} X_g u_g) \].

Assume independence over \( g \) with

\[ u_g \sim [0, \Sigma_g = \text{E}[u_g u'_g]]. \]

Then \( \hat{\beta} \sim \mathcal{N}[\beta, \text{V}[\hat{\beta}]] \) with

\[ \text{V}[\hat{\beta}] = (X'X)^{-1}(\sum_{g=1}^{G} X_g \Sigma_g X'_g)(X'X)^{-1}. \]
If ignore clustering the default OLS variance estimate should be inflated by approximately

\[ \tau_j \simeq 1 + \rho_{xj} \rho_u (\bar{N}_g - 1), \]

where

- \( \rho_{xj} \) is the within cluster correlation of \( x_j \)
- \( \rho_u \) is the within cluster error correlation
- \( \bar{N}_g \) is the average cluster size.

Moulton (1986, 1990) showed that could be large even if \( \rho_u \) small. e.g. \( N_G = 81, \rho_x = 1 \) and \( \rho_u = 0.1 \) then \( \rho = 9. \)

Should correct for clustering but \( \Sigma_g \) is unknown.
2.2 Moulton-type Standard Errors

Assume a random effects (RE) model

\[ u_{ig} = \alpha_g + \epsilon_{ig} \]

\[ \alpha_g \sim iid[0, \sigma_{\alpha}^2] \]

\[ \epsilon_{ig} \sim iid[0, \sigma_{\epsilon}^2]. \]
2.2 Moulton-type Standard Errors

- Assume a random effects (RE) model

\[ u_{ig} = \alpha_g + \varepsilon_{ig} \]
\[ \alpha_g \sim iid[0, \sigma^2_{\alpha}] \]
\[ \varepsilon_{ig} \sim iid[0, \sigma^2_{\varepsilon}] \]

- Then \( \Sigma_g = \sigma^2_u I_{N_g} + \sigma^2_{\alpha} e_{N_g} e'_{N_g} \) and we use

\[ \hat{V}_{RE}[\hat{\beta}] = (X'X)^{-1} \left( \sum_{g=1}^{G} X_g \hat{\Sigma}_g X'_g \right) (X'X)^{-1}, \]

where \( \hat{\Sigma}_g = \hat{\sigma}^2_u I_{N_g} + \hat{\sigma}^2_{\alpha} e_{N_g} e'_{N_g} \), and \( \hat{\sigma}^2_{\varepsilon} \) and \( \hat{\sigma}^2_{\alpha} \) are consistent.
Assume a **random effects (RE)** model

\[
\begin{align*}
    u_{ig} &= \alpha_g + \varepsilon_{ig} \\
    \alpha_g &\sim iid[0, \sigma_{\alpha}^2] \\
    \varepsilon_{ig} &\sim iid[0, \sigma_{\varepsilon}^2].
\end{align*}
\]

Then \( \Sigma_g = \sigma_u^2 I_{N_g} + \sigma_{\alpha}^2 e_{N_g} e'_{N_g} \) and we use

\[
\hat{V}_{RE}[\hat{\beta}] = (X'X)^{-1} \left( \sum_{g=1}^{G} X_g \hat{\Sigma}_g X'_g \right) (X'X)^{-1},
\]

where \( \hat{\Sigma}_g = \hat{\sigma}_u^2 I_{N_g} + \hat{\sigma}_{\alpha}^2 e_{N_g} e'_{N_g} \), and \( \hat{\sigma}_\varepsilon^2 \) and \( \hat{\sigma}_{\alpha}^2 \) are consistent.

**Weakness** is strong distributional assumptions.
The **cluster-robust variance estimate (CRVE)**

\[
\hat{V}_{CR}[\hat{\beta}] = (X'X)^{-1} \left( \sum_{g=1}^{G} X_g \tilde{u}_g \tilde{u}'_g X'_g \right) (X'X)^{-1},
\]

provides a consistent estimator of \( V_{CR}[\hat{\beta}] \) if

\[
\text{plim} \frac{1}{G} \sum_{g=1}^{G} X_g \tilde{u}_g \tilde{u}'_g X'_g = \text{plim} \frac{1}{G} \sum_{g=1}^{G} X_g \overset{°}{g} X'_g.
\]
2.3 Cluster-Robust Variance Estimates

- The **cluster-robust variance estimate (CRVE)**

$$
\hat{V}_{CR}[\hat{\beta}] = (X'X)^{-1} \left( \sum_{g=1}^{G} X_g \hat{u}_g \hat{u}_g' X_g' \right) (X'X)^{-1},
$$

provides a consistent estimator of $V_{CR}[\hat{\beta}]$ if

$$
\text{plim} \frac{1}{G} \sum_{g=1}^{G} X_g \hat{u}_g \hat{u}_g' X_g' = \text{plim} \frac{1}{G} \sum_{g=1}^{G} X_g \circ g X_g'.
$$

- $\hat{V}_{CR}[\hat{\beta}]$ yields **cluster-robust standard errors**.
The **cluster-robust variance estimate (CRVE)**

\[ \hat{V}_{CR}[\hat{\beta}] = (X'X)^{-1} \left( \sum_{g=1}^{G} X_g \tilde{u}_g \tilde{u}_g' X'_g \right) (X'X)^{-1}, \]

provides a consistent estimator of \( V_{CR}[\hat{\beta}] \) if

\[ \text{plim} \left\{ \frac{1}{G} \sum_{g=1}^{G} X_g \tilde{u}_g \tilde{u}_g' X'_g \right\} = \text{plim} \left\{ \frac{1}{G} \sum_{g=1}^{G} X_g \circ g X'_g \right\}. \]

\( \hat{V}_{CR}[\hat{\beta}] \) yields **cluster-robust standard errors**.

If OLS residuals are used then \( \tilde{u}_g = \hat{u}_g = y_g - X_g \hat{\beta} \).
Then \( \hat{V}_{CR}[\hat{\beta}] \) is downwards-biased for \( V[\hat{\beta}] \).
2.4 Alternative Cluster-Robust Variance Estimates

- Stata cluster options uses

\[ \tilde{u}_g = \sqrt{c} \hat{u}_g \text{ where } c = \frac{G}{G - 1} \times \frac{N - 1}{N - k} \sim \frac{G}{G - 1}. \]

- Bell and McCaffrey (2002) instead use

\[ \tilde{u}_g = \sqrt{\frac{G - 1}{G}} \left[ I_{N_g} - H_{gg} \right]^{-1} \hat{u}_g. \]

- Then \( \hat{V}_{CR}[\hat{\beta}] \) is the jackknife estimate of the variance of the OLS estimator.
  This leads to downwards-biased CRVE if in fact \( \Sigma_g = \sigma^2 I \).
  Angrist and Lavy (2002) use this.

- We call this CR3 as generalizes HC3 of MacKinnon and White (1985).
Two-sided Wald tests of $H_0: \beta_1 = \beta_1^0$ against $H_a: \beta_1 \neq \beta_1^0$
where $\beta_1$ is a scalar component of $\beta$.

Use the **Wald test statistic**

$$w = \frac{\hat{\beta}_1 - \beta_1^0}{s_{\hat{\beta}_1}},$$

where $s_{\hat{\beta}_1}$ is standard error from $\hat{V}_{CR}[\hat{\beta}]$.

Asymptotically $\mathcal{N}[0, 1]$ and reject at $\alpha = 0.05$ if $|w| > 1.960$.

In finite samples $w$ has much fatter tails than standard normal. Stata uses $T(G - 1)$ for critical values. And even this may over-reject.

3. Cluster Bootstraps

- **Bootstrap** due to Efron (1979)
  - Alternative asymptotic approximation for distribution of a statistic.
  - View the sample as the population and obtaining $B$ resamples leading to $B$ realizations of the statistic.
  - There are many, many ways to bootstrap.

- A bootstrap without asymptotic refinement
  - e.g. **bootstrap estimate of standard error**.
  - No better than regular asymptotic theory.
  - Popular as it may be simpler to implement.

- A bootstrap with asymptotic refinement
  - e.g. **bootstrap-t method**.
  - Asymptotically better than regular asymptotic theory.
  - Hopefully better in finite samples (use Monte Carlos to confirm).

- **Applied microeconometricians** rarely use asymptotic refinement.
3.1 Pairs Cluster Bootstrap-T Procedure

1. Do $B$ iterations of this step. On the $b^{th}$ iteration:
   1. Form a sample of $G$ clusters $\{(y_1^*, X_1^*), \ldots, (y_G^*, X_G^*)\}$ by resampling with replacement $G$ times from the original sample.
   2. Calculate the Wald test statistic $w_b^* = \frac{\hat{\beta}_{1,b}^* - \hat{\beta}_1}{s_{\hat{\beta}_{1,b}^*}}$, where
      - $\hat{\beta}_{1,b}^*$ is OLS estimate using the $b^{th}$ pseudo-sample
      - $s_{\hat{\beta}_{1,b}^*}$ is its standard error estimated using CR method
      - $\hat{\beta}_1$ is the original sample OLS estimate.

2. The empirical distribution of $w_1^*, \ldots, w_B^*$ estimates distribution of $w$. Reject $H_0$ at level $\alpha$ if and only if

   $$w < w_{[\alpha/2]}^* \text{ or } w > w_{[1-\alpha/2]}^*,$$

   where $w_{[q]}^*$ denotes the $q^{th}$ quantile of $w_1^*, \ldots, w_B^*$.

   [e.g. $w = -2.10$, $w_{[0.025]}^* = -2.32$, $w_{[0.975]}^* = 1.87$ do not reject.]
3.2 Asymptotic Refinement

- Consider two-sided nonsymmetric test of nominal size 0.5.
- Actual size = \(0.05 + O(G^{-0.5})\) by conventional asymptotics.
- Actual size = \(0.05 + O(G^{-1})\) using preceding bootstrap -t.
- So bootstrap-t offers an asymptotic refinement.
- Key is to bootstrap an asymptotically pivotal statistic, one whose distribution does not depend on unknown parameters.
- \(w\) is asymptotically pivotal but \(\hat{\beta}\) is not.
3.3 Wild Cluster Bootstrap-T Procedure

1. Obtain the OLS estimator $\hat{\beta}$ and OLS residuals $\hat{u}_g$, $g = 1, \ldots, G$. [Best to use residuals that impose $H_0$].

2. Do $B$ iterations of this step. On the $b^{th}$ iteration:

   1. For each cluster $g = 1, \ldots, G$, form $\hat{u}^*_g = \hat{u}_g$ or $\hat{u}^*_g = -\hat{u}_g$ each with probability 0.5 and hence form $\hat{y}^*_g = X'_g \hat{\beta} + \hat{u}^*_g$.

   2. This yields wild cluster bootstrap resample $\{(\hat{y}^*_1, X_1), \ldots, (\hat{y}^*_G, X^*_G)\}$.

3. Calculate the OLS estimate $\hat{\beta}_{1,b}$ and its standard error $s_{\hat{\beta}_{1,b}}$ and given these form the Wald test statistic $w_b^* = (\hat{\beta}_{1,b} - \hat{\beta}_1) / s_{\hat{\beta}_{1,b}}$.

4. Reject $H_0$ at level $\alpha$ if and only if

   $w < w_{[\alpha/2]}$ or $w > w_{[1-\alpha/2]}$,

   where $w^*_q$ denotes the $q^{th}$ quantile of $w^*_1, \ldots, w^*_B$. 
Wild bootstrap proposed by Wu (1986) for regression in the nonclustered case (heteroskedastic errors).


Horowitz (1997, 2001) provides a Monte Carlo demonstrating good size properties.

We use Rademacher weights that provide refinement if $\hat{\beta}$ is symmetrically distributed.

Mammen weights can be used if $\hat{\beta}$ is asymmetrically distributed.

Davidson and Flachaire (2001) provide theory and simulation to support using Rademacher weights even in the asymmetric case.

Here we have extended the wild bootstrap to a clustered setting.

A brief application by Brownstone and Valletta (2001) also does this.

Davidson and MacKinnon (1999) advocate imposing the null hypothesis.
3.4 Bootstraps with Few Clusters

- For cluster pairs bootstrap-$t$
  - $\binom{2G-1}{G-1}$ possible unordered recombinations of the data
  - 126 combinations for $G = 5$ and 92,378 combinations for $G = 10$
  - implementation problems can still arise e.g. with binary regressor.

- For wild cluster bootstrap-$t$
  - $2^G$ possible recombinations
  - 32 combinations for $G = 5$ and 1,024 combinations for $G = 10$
  - but implementation problems less likely to arise e.g. with binary regressor.
### 3.6 Test Methods Used in This Paper

<table>
<thead>
<tr>
<th>Method</th>
<th>Bootstrap?</th>
<th>Refinement?</th>
<th>$H_0$ imposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Assume iid (default se’s)</td>
<td>No</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>2. Moulton-type correction se</td>
<td>No</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>3. Cluster-robust se</td>
<td>No</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>4. CR3 cluster-robust se</td>
<td>No</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>5. Pairs Cluster Bootstrap-se</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>6. Residual Cluster Bootstrap se</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>7. Wild Cluster Bootstrap-se</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>8. Pairs Cluster BCA</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>9. BDM Bootstrap-t</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>10. Pairs Cluster Bootstrap-t</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>11. Pairs Cluster CR3 Bootstrap-t</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>12. Residual Cluster Bootstrap-t</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>13. Wild Cluster Bootstrap-t</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
4. Monte Carlo Simulations

- Dgp has $\beta_0 = 1$ and $\beta_1 = 1$ in

$$y_{ig} = \beta_0 + \beta_1 x_{ig} + u_{ig}, \ i = 1, \ldots, 30, \ g = 1, \ldots, G.$$ 

We consider $G = 5, 10, 15, 20, 25$ and $30$. 

Test $H_0: \beta_1 = 1$ vs $H_a: \beta_1 \neq 1$ at 5%:

Wald test statistic:

$$t = \frac{b_1}{s_b_1}.$$ 

Actual rejection rate computed using $N = 10,000$ Monte Carlo simulations.

Monte Carlo se $= \frac{1}{s_b_1}$ if $b_1 = 0.10.$
4. Monte Carlo Simulations

- Dgp has $\beta_0 = 1$ and $\beta_1 = 1$ in

$$y_{ig} = \beta_0 + \beta_1 x_{ig} + u_{ig}, \quad i = 1, \ldots, 30, \quad g = 1, \ldots, G.$$  

We consider $G = 5, 10, 15, 20, 25$ and $30$.

- Test $H_0 : \beta_1 = 1$ vs $H_a : \beta_1 \neq 1$ at 5%:

Wald test statistic: $t = \frac{\hat{\beta} - 1}{\hat{s_\beta}}$. 

Monte Carlo simulations and Cluster Bootstrap

[Colin Cameron, Jonah Gelbach, Doug Miller]
4. Monte Carlo Simulations

- Dgp has $\beta_0 = 1$ and $\beta_1 = 1$ in

  $$y_{ig} = \beta_0 + \beta_1 x_{ig} + u_{ig}, \ i = 1, \ldots, 30, \ g = 1, \ldots, G.$$  

  We consider $G = 5, 10, 15, 20, 25$ and 30.

- Test $H_0 : \beta_1 = 1$ vs $H_a : \beta_1 \neq 1$ at 5%:

  $$\text{Wald test statistic: } t = \frac{\hat{\beta} - 1}{\hat{s}_\beta}.$$  

- Actual rejection rate computed using $S = 1,000$ Monte Carlo simulations.

  Monte Carlo se = $\sqrt{\hat{a}(1 - \hat{a})/S} = 0.009$ if $\hat{a} = 0.10$. 
Table 1: Errors correlated within group but homoskedastic.

Dgp is

\[ y_{ig} = \alpha + \beta x_{ig} + u_{ig} = \alpha + \beta (z_g + z_{ig}) + (\varepsilon_g + \varepsilon_{ig}) \]
\[ z_g \sim \mathcal{N}[0, 1], \quad z_{ig} \sim \mathcal{N}[0, 1] \implies \rho_x = 0.5 \]
\[ \varepsilon_g \sim \mathcal{N}[0, 1], \quad \varepsilon_{ig} \sim \mathcal{N}[0, 1] \implies \rho_u = 0.5. \]

This is classic random effects model for both the error and the regressor.

Expect Moulton correction and residual bootstrap to do well.
Monte Carlo Simulations

1. iid errors: \( \sqrt{1 + (N_G - 1)\rho_x\rho_u} = \sqrt{8.25} = 2.87 \) and \( \Pr[|z| > 1.96/2.87] = 0.50 \) as \( G \to \infty \).

2. Moulton: rejection rate that if \( t \sim T(G - 2) \).

3. CR: Improvement, but not as good as Moulton.


5. Pairs cluster bootstrap-se: similar to CR.

6. Residual cluster bootstrap-se: close to 0.5 (why?)

7. Wild cluster bootstrap-se: close to 0.5 (why?)

9. BDM bootstrap-t uses iid $s_{\hat{\beta}}$ not cluster-robust $s_{\hat{\beta}}$. So no refinement. But still big improvement on 1.


12. Residual cluster bootstrap-t. Works well.

Table 3: Errors correlated within group and heteroskedastic.

Dgp is

\[ \begin{align*}
    y_{ig} &= \alpha + \beta x_{ig} + u_{ig} = \alpha + \beta (z_g + z_{ig}) + (\varepsilon_g + \varepsilon_{ig}) \\
    z_g &\sim \mathcal{N}[0, 1], \quad z_{ig} \sim \mathcal{N}[0, 1] \implies \rho_x = 0.5 \\
    \varepsilon_g &\sim \mathcal{N}[0, 1], \quad \varepsilon_{ig} \sim \mathcal{N}[0, 9 \times (z_g + z_{ig})^2] \implies \rho_u < 0.5.
\end{align*} \]

No longer classic random effects model for both the error and the regressor.

Find that Moulton-type correction breaks down.

The bootstrap-t methods work well (surprising for residual bootstrap).
If reject when $|t| > t_{8;0.025} = 2.306$ rather than 1.96 then methods 2-7 do better.

As increase cluster size then as expected default OLS does much worse while methods 2-13 change little.

If have four regressors rather than one (each $0.5 \times (z_g + z_{ig})$) then expect similar results though more noise.
Monte Carlo: Variations for $G=10$
Monte Carlo: Variations for G=10
Bootstrap for BDM Study

- Use BDM data on 21 years and up to $G = 50$ states.

$$y_{ig} = \alpha_g + \gamma_i + \beta_1 l_{ig} + u_{ig} : \text{i is year; g is state.}$$

$y_{ig}$ is a year-state measure of excess earnings (after control for age and education).
Use BDM data on 21 years and up to $G = 50$ states.

$$y_{ig} = \alpha_g + \gamma_i + \beta_1 I_{ig} + u_{ig}: \ i \text{ is year; } \ g \text{ is state.}$$

$y_{ig}$ is a year-state measure of excess earnings (after control for age and education).

$I_{ig}$ is a binary policy change indicator.
Randomly occurs in half the states.
When occurs between sixth and fifteenth year.
Once only change from 0 to 1.
Use BDM data on 21 years and up to \( G = 50 \) states.

\[
y_{ig} = \alpha_g + \gamma_i + \beta_1 I_{ig} + u_{ig} : \ i \text{ is year; } g \text{ is state.}
\]

\( y_{ig} \) is a year-state measure of excess earnings (after control for age and education).

\( I_{ig} \) is a binary policy change indicator.
Randomly occurs in half the states.
When occurs between sixth and fifteenth year.
Once only change from 0 to 1.

Actual rejection rate: fraction of times \( |w| = |(\hat{\beta}_1 - 0) / s_{\hat{\beta}_1}| > 1.96. \)
Bootstrap for BDM Study

- Use BDM data on 21 years and up to \( G = 50 \) states.

\[
y_{ig} = \alpha_g + \gamma_i + \beta_1 I_{ig} + u_{ig} : \ i \text{ is year}; \ g \text{ is state.}
\]

\( y_{ig} \) is a year-state measure of excess earnings (after control for age and education).

- \( I_{ig} \) is a binary policy change indicator.
  Randomly occurs in half the states.
  When occurs between sixth and fifteenth year.
  Once only change from 0 to 1.

- Actual rejection rate: fraction of times \( |w| = |(\hat{\beta}_1 - 0)/s_{\hat{\beta}_1}| > 1.96. \)

- Power against \( \beta_1 = 0.02 \): fraction of 400 times \( |w| > 1.96. \)
  [Do OLS of \( y_{ig} + (0.02 I_{ig}) \) on \( I_{tg} \).]
Get BDM’s result that assuming iid errors leads to massive over-rejection.

Get BDM’s result that cluster robust with $G = 6$ has over-rejection.

Moulton-type correction does surprisingly poorly.

Problems with pairs cluster bootstrap-t.
For 3 out of 6 states there was no policy change and $I_g = 0$:
$Pr[\text{all } 6 I_g^* = 0] = (1/2)^6 = 1/64.$

Find that wild cluster bootstrap-t does well.
Bootstrap for BDM Study: With individual data

- Similar to with aggregate data.
Model for whether have private insurance is

\[ y_{ijt} = \alpha_1 + \alpha_2 \text{SELF}_{ijt} + \alpha_3 \text{POST}_{ijt} + \beta_1 \text{SELF}_{ijt} \times \text{POST}_{ijt} + u_{jt}, \]

where \( i \) denotes individual, \( j \) denotes employer type, \( t \) denotes year
\( \text{SELF}_{ijt} = 1 \) if individual \( i \) is self-employed at time \( t \)
\( \text{POST}_{ijt} = 1 \) if the year is 1987, 1988 or 1999.

Difference-in-difference analysis controlling for clustered errors.
Treat years as clusters \( G = 8 \) (following Donald and Lang (2004)).

Aggregated employment-year data.
Then CR se is 0.0074 compared to 0.0044 without clustering.
Individual-level data results similar to aggregate.
7. Conclusion

- At the least should have small-sample correction of standard errors and use a $T$ distribution with $G$ or fewer degrees of freedom.
- Cluster pairs bootstrap-t is better than Cluster pairs BCA.
- But can run into problems implementing if binary cluster-invariant regressor and few clusters.
- Cluster Wild bootstrap-t works well, though is designed just for OLS.