Cluster Robust Inference with Heterogeneous Clusters

joint work with Chang Lee and Drew Carter

Douglas G. Steigerwald

UC Santa Barbara

July 2018

Empirical Framework

Kuhn et alia AER 2011

- measure consumption impact from a shock to neighbor's income
- 410 postal codes (g) : 4 to 105 households (i) : g grows with n

$$c_{gi} = \alpha_0 + \alpha_{fe} + \beta_1 \cdot win_g + \beta_2 \cdot income_{gi} + u_{gi}$$

- V covariance matrix of OLSE for coefficients
- $\widehat{V}$ cluster-robust variance estimator

- baseline beliefs for this empirical setting

1. $\widehat{V}$ is known to be consistent
2. $\widehat{V}$ removes downward bias in OLS estimator of V
3. degrees-of-freedom for hypothesis testing at least 410
    1. 410 for t test of $H_0 : \beta_1 = 0$
    2. n for t test of $H_0 : \beta_2 = 0$

Research Response

Our findings

$$c_{gi} = \alpha_0 + \alpha_{fe} + \beta_1 \cdot win_g + \beta_2 \cdot income_{gi} + u_{gi}$$

- V covariance matrix of OLSE for coefficients
- $\widehat{V}$ cluster-robust variance estimator

- our findings for this empirical setting

1. $\widehat{V}$ is known to be consistent - false
    1. previously established when group designs (cluster sizes) are equal
    2. we establish consistency when group designs (cluster sizes) vary
    3. inconsistent for α_{fe}
2. $\widehat{V}$ removes downward bias in OLS estimator of V - false
    1. $\widehat{V}$ may have downward bias
3. degrees-of-freedom for hypothesis testing at least 410 - false
    1. $\widehat{V}$ a function only of between cluster variation
    2. d-o-f at most 410 for either t test of $H_0 : \beta_1 = 0$ or $H_0 : \beta_2 = 0$
    3. variation in designs (cluster sizes) reduces d-o-f below 410

Road Map

- Data sets with growing number of clusters
- Interest focuses on cluster invariant regressor
  - no cluster fixed effects
- Consistency with cluster homogeneity White (1984)
- Finite sample behavior Cameron, Gelbach and Miller (2008)

1. Consistency with cluster heterogeneity
   1. allow cluster sizes to vary
   2. number of clusters tends to infinity
2. Guide to finite sample behavior - reflects cluster heterogeneity
   1. *Effective* number of clusters
   2. smaller than number of clusters
3. Guidelines for Empirical Research

Cluster Structure

data generating process

$$y_{gi} = \beta_0 + \beta_1 x_g + \beta_2 z_{gi} + u_{gi}$$

- y_{gi} observation i in cluster g
- n_g number of observations in cluster g $\qquad \sum_g n_g = n$
- G number of clusters
- Error covariance matrix

$$\Omega = \begin{bmatrix} \Omega_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Omega_G \end{bmatrix}$$

Ω_g unrestricted (positive definite)

Robust Test Statistic

Shah, Holt and Folsom 1977

a selection vector

$$H_0\colon\ a^{\mathrm{T}}\beta = 0$$

$\hat{\beta}$ OLSE for β with variance $\quad V = \sum_{g=1}^{G} Var\left[\left(X^{\mathrm{T}}X\right)^{-1} X_g^{\mathrm{T}} u_g\right]$

- test statistic

$$Z = \frac{a^{\mathrm{T}}\hat{\beta}}{\sqrt{a^{\mathrm{T}}\hat{V}a}}$$

- cluster robust variance estimator

$$\widehat{V} = \left(X^{\mathrm{T}}X\right)^{-1} \cdot \sum_{g=1}^{G} X_g^{\mathrm{T}}\hat{u}_g\hat{u}_g^{\mathrm{T}}X_g \cdot \left(X^{\mathrm{T}}X\right)^{-1}$$

robust to arbitrary structure of Ω_g

allows n_g to vary

Consistency

Theorem 1

Assumptions

- Ω_g not identical over g
- X_g not identical over g
- n_g not constant over g

If, as $n \longrightarrow \infty$: $G \longrightarrow \infty$

$$\frac{a^{\mathrm{T}}\widehat{V}a}{a^{\mathrm{T}}Va} \overset{MS}{\longrightarrow} 1$$

which leads directly to

$$Z \underset{H_0}{\rightsquigarrow} \mathcal{N}(0,1)$$

Remark 1

Convergence governed by G not n

$A_g = \left(X^{\mathrm{T}} X\right)^{-1} X_g^{\mathrm{T}} X_g$

$\hat{\beta}_g$ OLSE based only on X_g

$$\widehat{V} = \sum_g A_g \left(\hat{\beta}_g - \hat{\beta}\right) \left(\hat{\beta}_g - \hat{\beta}\right)^{\mathrm{T}} A_g^{\mathrm{T}}$$

- $\widehat{V}$ is a function only of between cluster variation consistency requires $G \to \infty$

$$y_{gi} = \beta_0 + \beta_1 x_g + \beta_2 z_{gi} + u_{gi}$$

even for test of β_2 behavior of Z is governed by G

- if there is no cluster correlation each observation is a cluster $G = n$

Remark 2

Inconsistent Testing

$\widehat{V}$ is a function only of between cluster variation

- consistency of $\widehat{V}$ depends on G growing
- inconsistent test for
  - coefficient estimator that depends on fixed subset of clusters
- leading examples
  - controls that correspond to a group of clusters
  - cluster specific controls (cluster fixed effects)

Cluster Heterogeneity and Asymptotic Approximation

What gives rise to cluster heterogeneity?
For example:

1. unequal cluster sizes
2. equal cluster sizes, but variation in Ω_g
3. equal cluster sizes and constant Ω_g, but variation in X_g

- the majority of empirical studies have cluster heterogeneity

convergence of Z requires $G \longrightarrow \infty$

Is G an accurate guide to performance under heterogeneity?

Cluster Heterogeneity Measure

analysis leads to a natural measure of heterogeneity for each cluster

$$\gamma_g = a^{\mathrm{T}} \left(X^{\mathrm{T}} X \right)^{-1} X_g^{\mathrm{T}} \Omega_g X_g \left(X^{\mathrm{T}} X \right)^{-1} a$$

- depends on which coefficients are under test through a
- measure of heterogeneity for entire sample

$$\Gamma = \frac{\frac{1}{G} \sum_{g=1}^{G} \left(\gamma_g - \bar{\gamma} \right)^2}{\bar{\gamma}^2}$$

  - (squared) coefficient of variation for γ_g

Finite Sample Behavior of Cluster Robust Estimator

- leading term in asymptotic behavior of Z is governed by
- G under homogeneity
  - number of clusters is a guide to inference
- $\frac{G}{1+\Gamma}$ under heterogeneity

- inference is guided by the effective number of clusters

$$ENC = \frac{G}{1+\Gamma}$$

Magnitude of Cluster Correction

example: if $\Gamma = 2$

$$ENC = \frac{G}{3}$$

different order of magnitude than standard bias correction

$$G - k$$

As $n \longrightarrow \infty$: ENC governs the mean-squared error of $\widehat{V}$

- cluster heterogeneity increases
  - variation in $\widehat{V}$
  - bias in $\widehat{V}$

Laboratory Performance

Framework

$$y_{gi} = \beta_0 + \beta_1 x_g + \beta_2 z_{gi} + u_{gi}$$

error components model

$$u_{gi} = \varepsilon_g + v_{gi}$$
$$\varepsilon_g \overset{iid}{\sim} \mathcal{N}(0,1) \text{ independently of } v_{gi} | X \sim \mathcal{N}\left(0, cz_{gi}^2\right)$$

correlation matrix for cluster g

$$\begin{bmatrix} 1 & & \rho_{ij} \\ & \ddots & \\ \rho_{ij} & & 1 \end{bmatrix} \text{ where } \rho_{ij} = \frac{1}{\sqrt{1+cz_{gi}^2} \cdot \sqrt{1+cz_{gj}^2}}$$

$c = 500$ nearly uncorrelated (heteroskedastic)
$c = 0$ perfectly correlated (homoskedastic)

Design Variation

- 2500 observations divided into 100 groups
- $x_g \overset{iid}{\sim} Bernoulli(.5) \quad z_{gi} \overset{iid}{\sim} U(0,1)$

1. Cluster Sizes
   1. design 1 : $n_1 = 25 \quad n_2 = \cdots = n_{100} = 25$
   2. design 2 : $n_1 = 124 \quad n_2 = \cdots = n_{100} = 24$
   3. $\vdots$
   4. design 10 : $n_1 = 916 \quad n_2 = \cdots = n_{100} = 16$
2. Error Cluster Correlation
   1. $c = 500$: correlation ≈ 0 heteroskedastic
   2. $\vdots$
   3. $c = 0$: correlation $=1$ homoskedastic

Impact of Design on Effective Number of Clusters

Effective Number of Clusters:

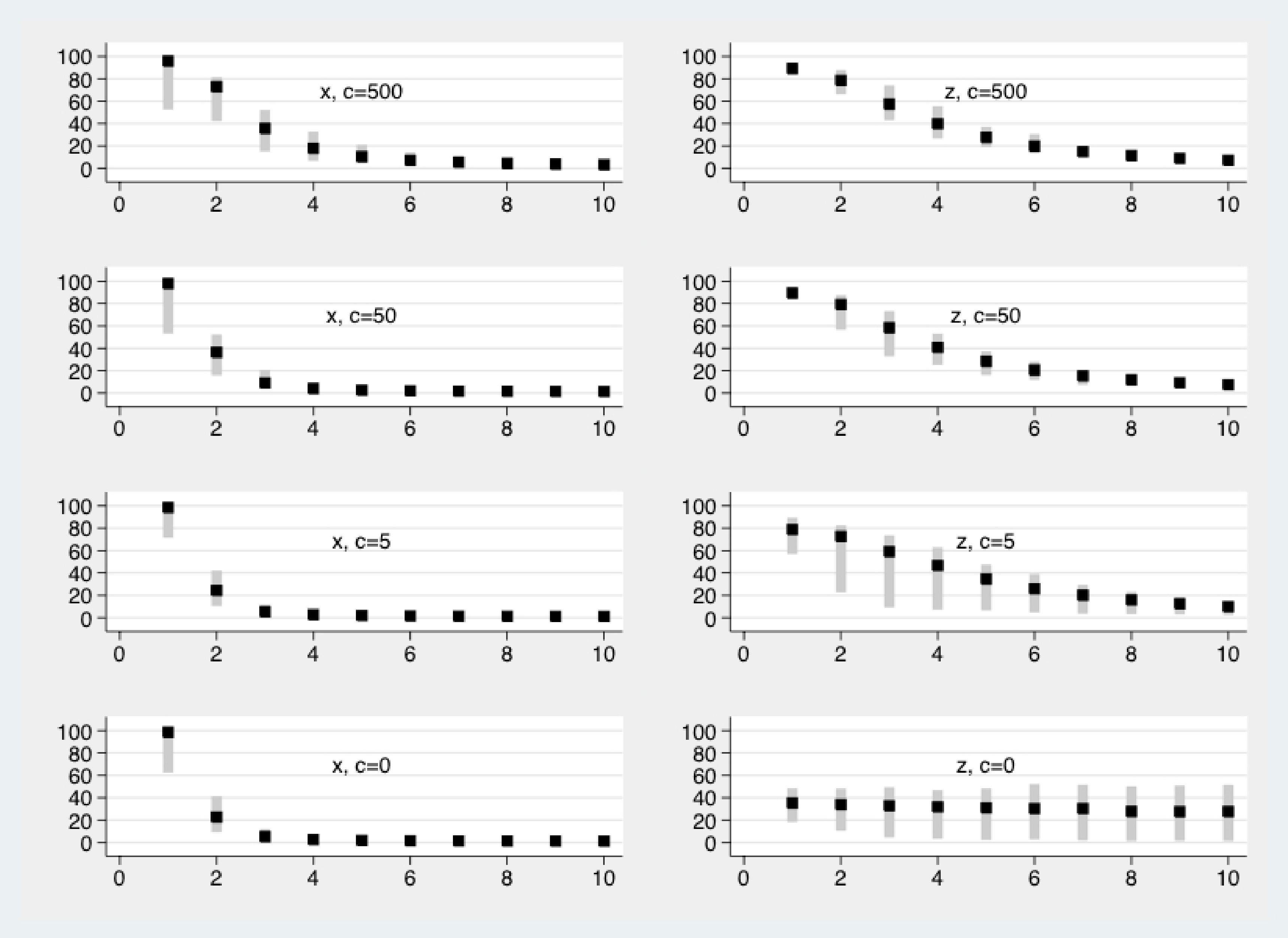
$$\frac{G}{1+\Gamma(\Omega, X)}$$

$\Gamma(\Omega, X)$: measure of cluster heterogeneity

1. cluster size variation
   1. Increasing cluster size variation reduces ENC
2. realized values for X
   1. Data sets with unequal values for x_g reduce ENC
3. cluster error correlation
   1. As the cluster error correlation increases, ENC is more sensitive to variation in x_g

- for each set of cluster sizes and value of c : generate 1000 values of X

Impact of Design on ENC



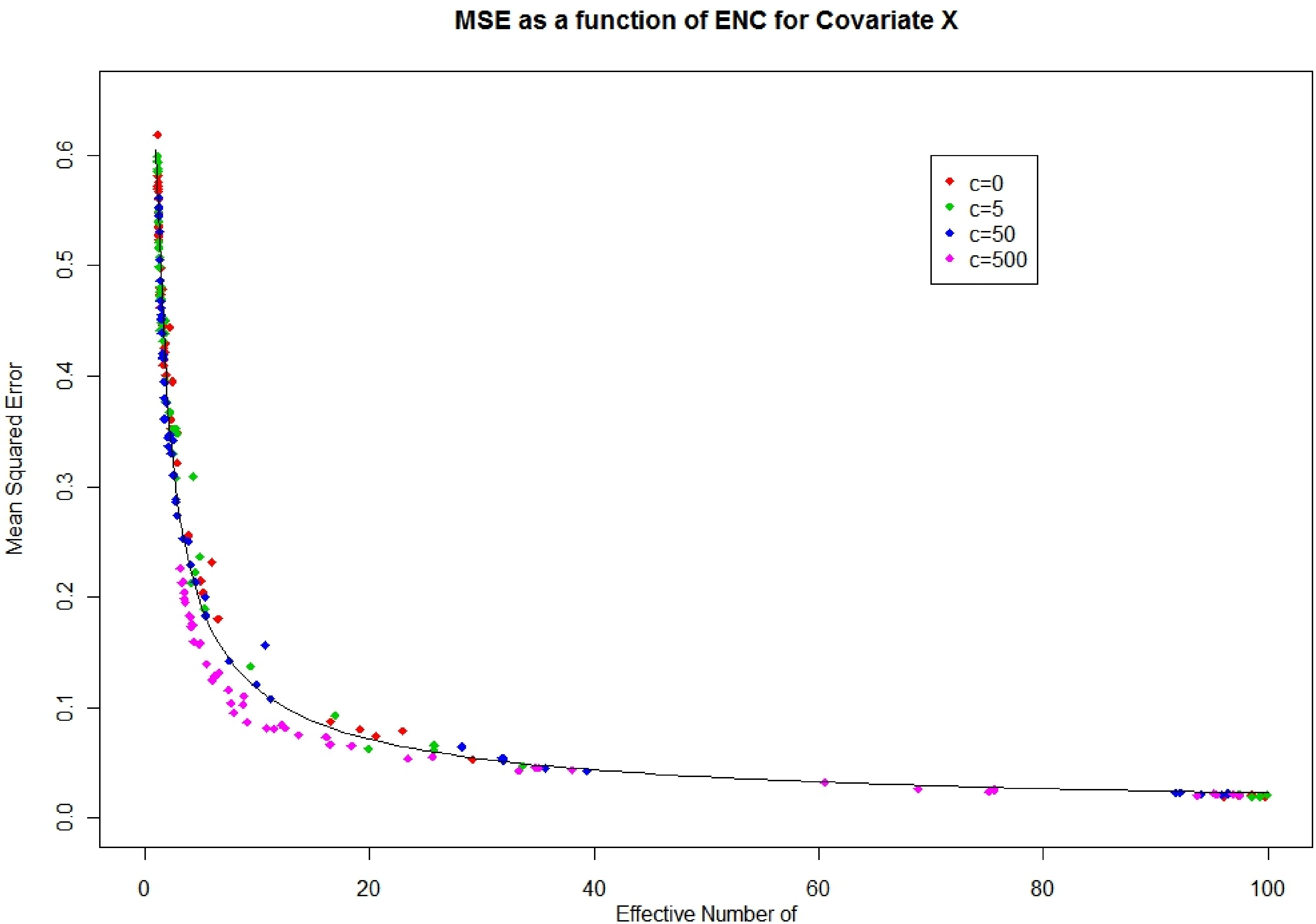
Impact of Effective Number of Clusters on MSE of Cluster-Robust Variance Estimator

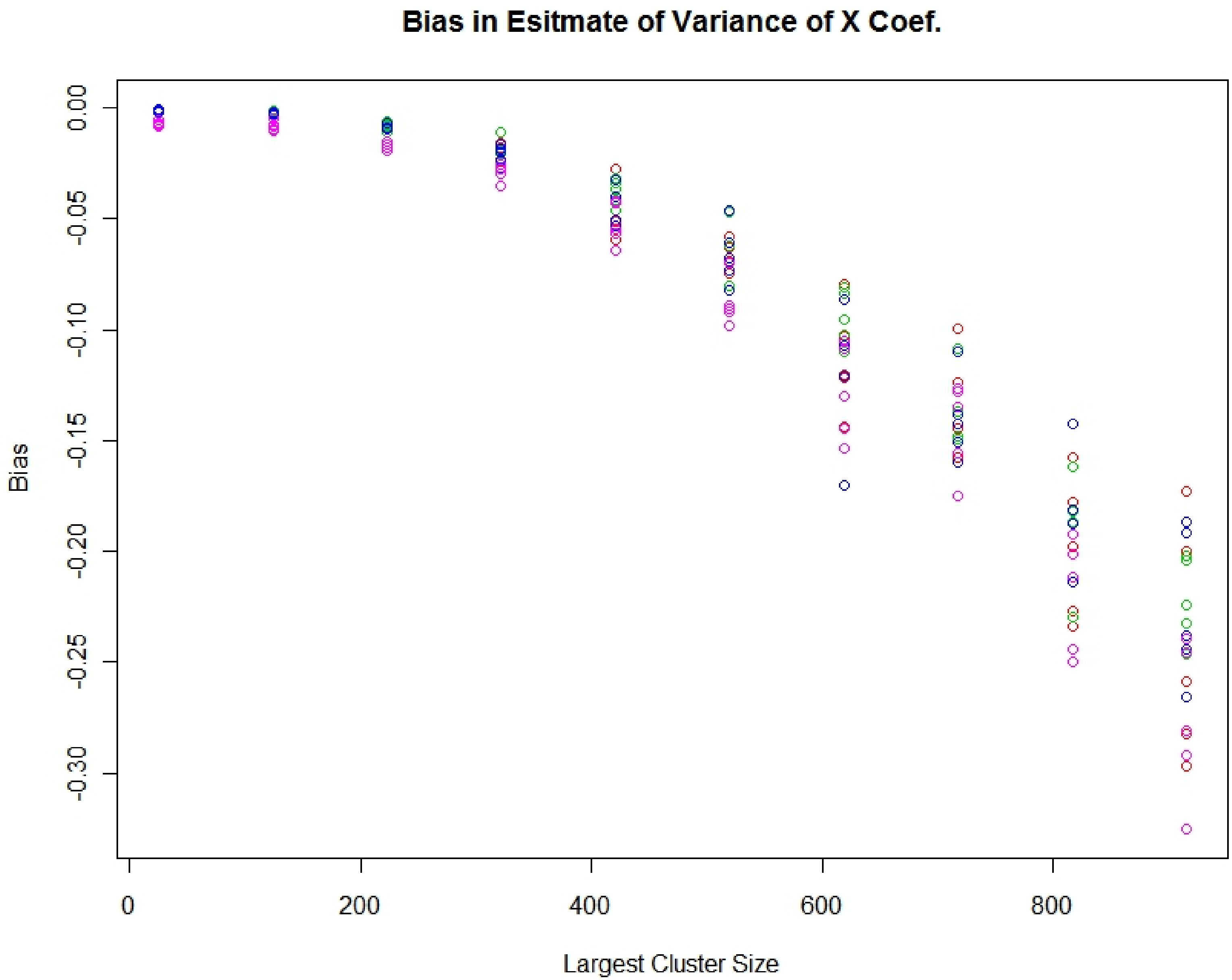
Mean-Squared Error:

$$MSE\left(\left.\frac{a^{\mathrm{T}}\widehat{V}a}{a^{\mathrm{T}}Va}\right| X\right) \approx \frac{2}{G}(1+\Gamma)$$

- Reducing the ENC increases the MSE for $\widehat{V}$

1. MSE is conditional on realization of X
    1. 5 values of X are generated for each set of cluster sizes and value of c
    2. for each value of X, 1000 values of u are generated


MSE as a function of ENC for Covariate X
c=0
c=5
c=50
c=500
Mean Squared Error
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100
Effective Number of


Bias in Esitmate of Variance of X Coef.
Bias
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
-0.30
0
200
400
600
800
Largest Cluster Size

MSE of Cluster-Robust Variance Estimator

Cluster-Invariant Regressor

$$y_{gi} = \beta_0 + \beta_1 x_g + \beta_2 z_{gi} + u_{gi}$$

- estimator of variance for $\hat{\beta}_1$
- MSE is impacted by bias
- bias is driven by variation in cluster size

With variation in cluster sizes, the cluster-robust standard error can be significantly downward biased for the cluster-invariant regressor.

MSE of Cluster-Robust Variance Estimator

Cluster-Varying Regressor

$$y_{gi} = \beta_0 + \beta_1 x_g + \beta_2 z_{gi} + u_{gi}$$

- estimator of variance for $\hat{\beta}_2$
- MSE impact depends on c
- if $c = 500$ (no error cluster correlation) : bias impacts
  - bias driven by variation in cluster size
- if $c < 500$ (error cluster correlation) : variation dominates

With error cluster correlation, the cluster-robust standard error can be highly variable for the cluster-varying regressor.

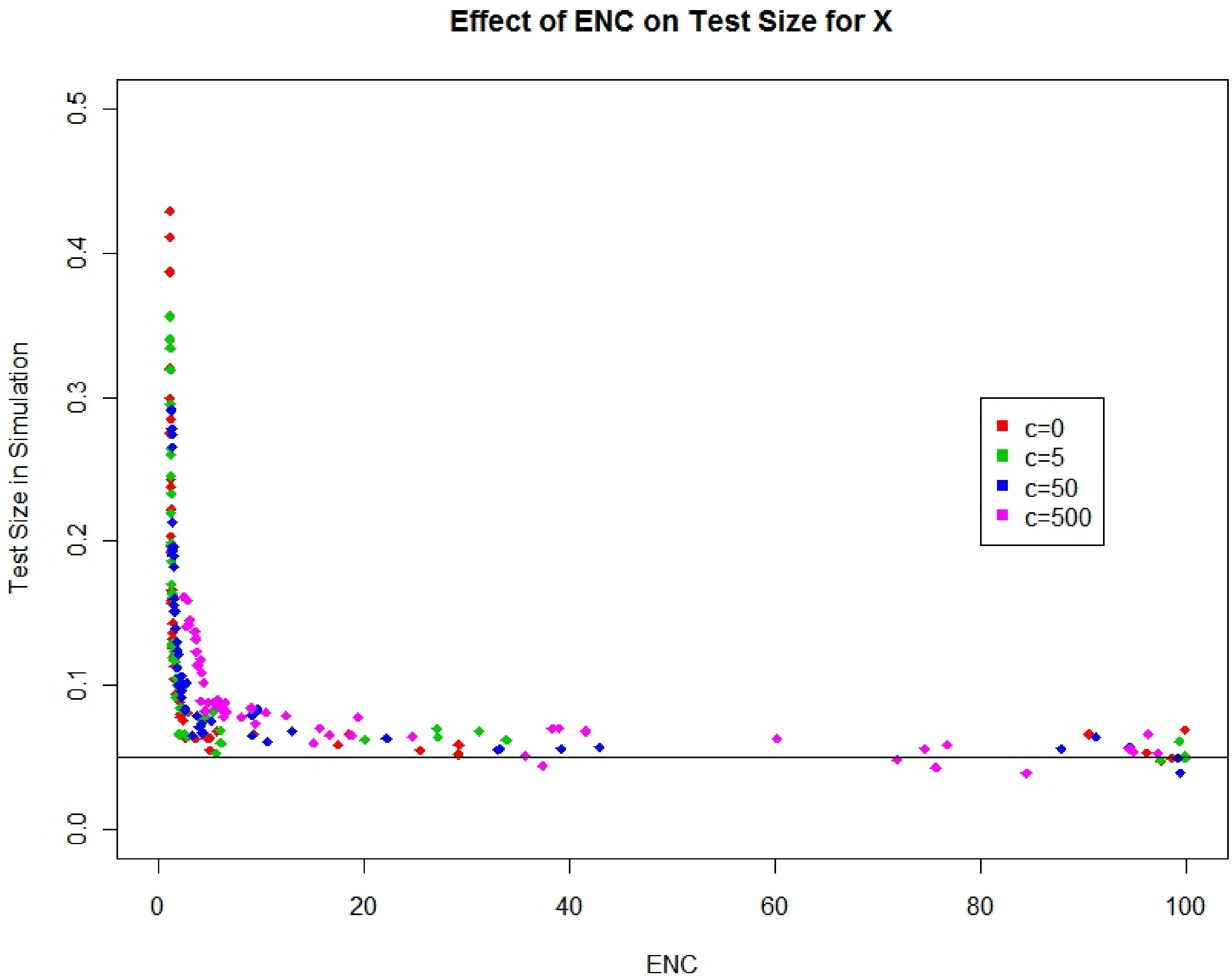
Empirical Test Size for Cluster-Robust t Test

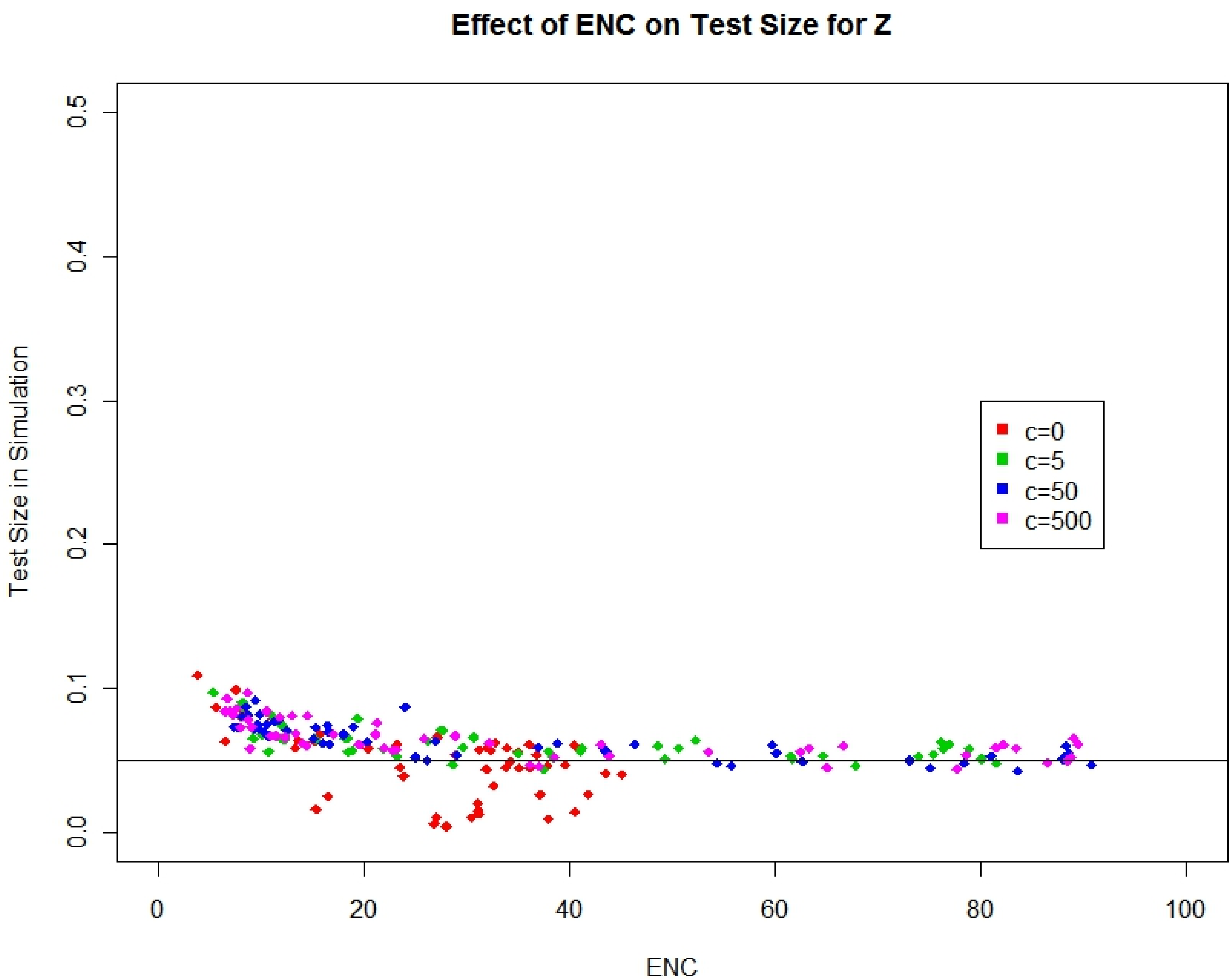
Cluster-Invariant Regressor

$$y_{gi} = \beta_0 + \beta_1 x_g + \beta_2 z_{gi} + u_{gi}$$

- test of $H_0 : \beta_1 = 0$
  - small ENC $\longrightarrow$ downward bias in cluster-robust s.e. $\longrightarrow$ large empirical test size
- test of $H_0 : \beta_2 = 0$
  - small ENC $\longrightarrow$ greater variation in cluster-robust s.e. $\longrightarrow$ variation in empirical test size

Most pronounced impact for hypothesis test of β_1


Effect of ENC on Test Size for X
Test Size in Simulation
0.0
0.1
0.2
0.3
0.4
0.5
ENC
0
20
40
60
80
100
c=0
c=5
c=50
c=500



Effect of ENC on Test Size for Z
Test Size in Simulation
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
100
ENC
c=0
c=5
c=50
c=500

Stata Program clustereff

Test $H_0 : \beta_1 = 0$

$$y_{gi} = \beta_0 + \beta_1 x_g + \beta_2 z_{gi} + u_{gi},$$

g indexes the variable *postcode*
stata command (uses the default value of $\rho = 1$)

clusteff x z, cluster(postcode) test(z)

- if the data is likely to have low within cluster correlation, select your own value

clusteff x z, cluster(postcode) test(z) cov(.2)

Field Performance : Effect of Unilateral Divorce on Married Women's Work

Voena (2015)

Are married women more likely to work once unilateral divorce is introduced in their state?

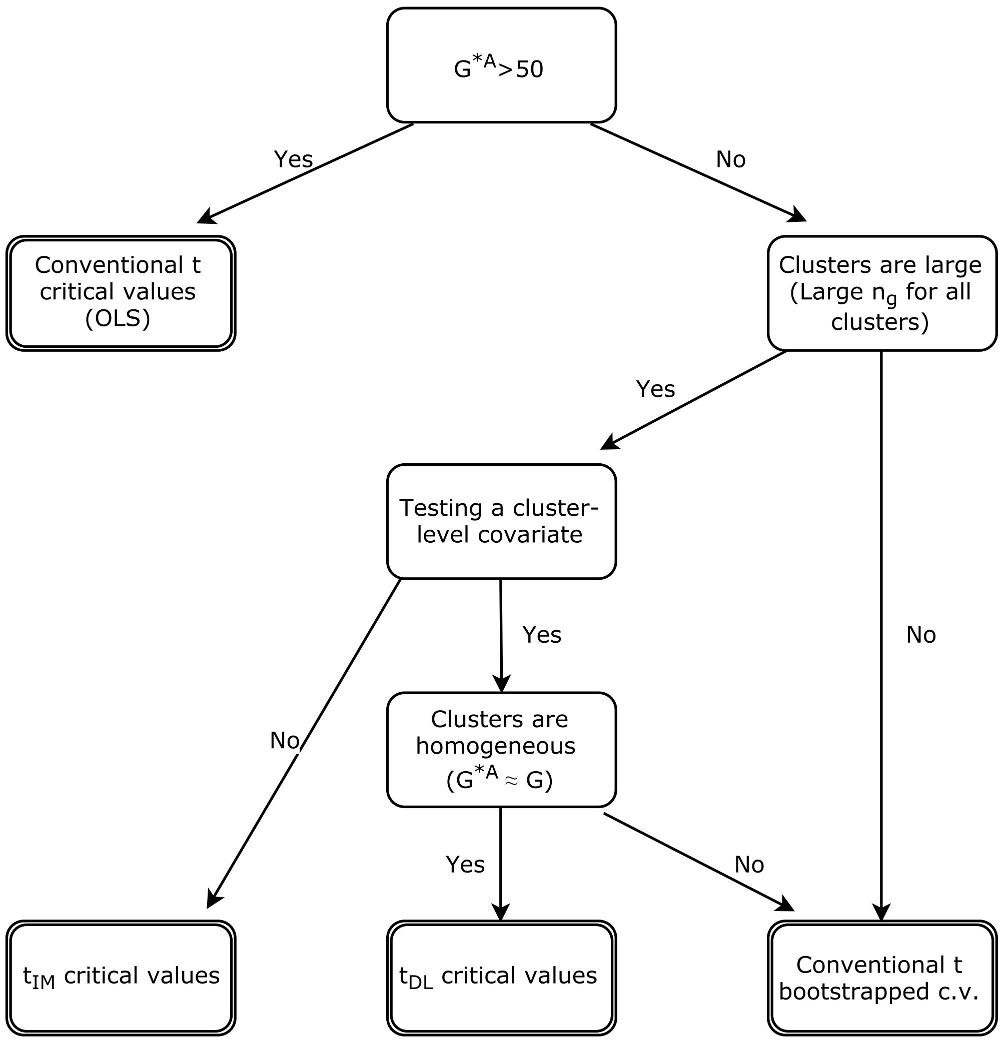
data on individuals indexed by state and time

- 51 clusters
- number of observations per cluster ranges from 3 to 3,552
- effective number of clusters

$$\begin{aligned} 13 \quad (\rho &= 1) \\ 20 \quad (\rho &= 0) \end{aligned}$$

- standard critical values ± 1.96 are not correct

Decision Tree



Bootstrap (Wild) Critical Values

$\widehat{u}_g$ the residuals for cluster g
w is a bivariate random variable that equally likely takes -1 and $+1$

- construct a bootstrap sample
  - draw (G times) with replacement from $\{\widehat{u}_1, \ldots, \widehat{u}_g\}$ and multiply by simultaneously drawn w
  - yields new bootstrap sample $\{\widehat{u}_1^*, \ldots, \widehat{u}_g^*\}$

construct $\left\{ y_g^* = x_g^{\mathrm{T}}\widehat{\beta} + \widehat{u}_g^* \right\}_{g=1}^{G}$
obtain OLSE $\widehat{\beta}^{*1}$ (repeat 999 times)
critical values from quantiles of $\widehat{\beta}^*$

Remarks

- Under Cluster Heterogeneity
  - cluster-robust variance estimator is consistent
  - cluster-robust $Z \rightsquigarrow \mathcal{N}\left(0,1\right)$
- Impact of heterogeneity
  - mean-squared error of $\widehat{V}$
- Effective Number of Clusters
  - single measure to capture cluster heterogeneity
  - can be well approximated
  - should be constructed for each sample
- Low Effective Number of Clusters
  - indicates substantial mean-squared error in $\widehat{V}$
  - indicates inflated empirical test size for cluster invariant regressor
  - use conservative critical values (bootstrap, student t)

Variability of Cluster-Robust Variance Estimator

Increase in Group Sizes

- Hold Design Constant (G, Γ)
  - increase group sizes to 220
- Allow $\widehat{V}$ to vary (variation in u across samples)

	ENC x	RV x	*ENC z*	RV z
Moderate Variation	88	.09	68	.11
Large Variation	17	.48	59	.14

- magnitude of relative variance is not approximately 2/ENC
- our asymptotic theory needs refinement

Student t Approximation

Satterthwaite 1946

Consider $\widetilde{V}$

$$Var\left(\frac{a^{\mathrm{T}}\widetilde{V}a}{a^{\mathrm{T}}Va}\right) = \frac{2}{G}\left(1+\Gamma\right)$$

If $\Gamma = 0$ (homogeneity)

$$G \cdot \frac{a^{\mathrm{T}}\widetilde{V}a}{a^{\mathrm{T}}Va} \sim \chi^2_G$$

- yet

$$\frac{a^{\mathrm{T}}\left(\hat{\beta}-\beta_0\right)}{\sqrt{a^{\mathrm{T}}\widetilde{V}a}} \text{ is not distributed as Student } t$$

- reason: numerator and denominator are correlated
- difficult to bound approximation error induced by correlation