A Practitioner's Guide to Cluster-Robust Inference

A. C. Cameron and D. L. Miller

presented by
Federico Curci

March 4, 2015

In the previous episode

Variance inflation

$$\tau_k = 1 + \rho_{x_k}\rho_u \left(\bar{N}_g - 1\right) \tag{1}$$

Cluster-Robust estimate of Variance Matrix (CRVE)

$$V\left[\hat{\beta}\right] = \left(X'X\right)^{-1} B_{clu} \left(X'X\right)^{-1}$$
$$B_{clu} = \sum_{g=1}^{G} X_g' E\left[u_g u_g' | X_g\right] X_g$$

- Assumption $E\left[u_{ig}u_{jg'}|x_{ig},x_{jg'}\right] = 0$ unless $g = g'$
- Consistent as $G \to \infty$

Bias-Variance Trade-off

- Larger Clusters have less bias but higher variability

In the previous episode

Variance inflation

$$\tau_k = 1 + \rho_{x_k} \rho_u \left(\bar{N}_g - 1 \right) \tag{1}$$

Cluster-Robust estimate of Variance Matrix (CRVE)

$$V\left[\hat{\beta}\right] = \left(X'X\right)^{-1} B_{clu} \left(X'X\right)^{-1}$$
$$B_{clu} = \sum_{g=1}^{G} X_g' E\left[u_g u_g' | X_g\right] X_g$$

- Assumption $E\left[u_{ig} u_{jg'} | x_{ig}, x_{jg'}\right] = 0$ unless $g = g'$
- Consistent as $G \to \infty$

Bias-Variance Trade-off

- Larger Clusters have less bias but higher variability

Problems:

- Clusters not nested
- Few clusters

Multi-way clustering

- Example: state-year panel clustering both within years and states
- Not possible to cluster at intersection of two groupings
- Solutions:
    - Include sufficient regressors to eliminate error correlation in all but one dimension, and then do CRVE for remaining dimension
    - Multi-way CRVE

Multi-way CRVE

Model

- Model: $y_i = x_i'\beta + u_i$
- Assumption: $E\left[u_i u_j | x_i, x_j\right] = 0$ unless observations i and j share any cluster dimension
- Multi-way CRVE $V\left[\hat{\beta}\right] = \left(X'X\right)^{-1} \hat{B} \left(X'X\right)^{-1}$:
  $\hat{B} = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j' \hat{u}_i \hat{u}_j 1\left[\text{i,j share any cluster}\right]$
- Asymptotics relies on number of clusters of the dimension with the fewest number of clusters

Multi-way CRVE, cont'd

Implementation

- Identify two ways of clustering and unique "group 1 by group 2 combinations"
- Estimate the model clustering on "group 1" → $\hat{V}_1$
- Estimate the model clustering on "group 2" → $\hat{V}_2$
- Estimate the model clustering on "group 1 by group 2" → $\hat{V}_{1\cap 2}$
- $\hat{V}_{2way} = \hat{V}_1 + \hat{V}_2 - \hat{V}_{1\cap 2}$
- S.e. are the square root of the diagonal elements of $\hat{V}_{2way}$

Problems

- Perfectly multicollinear regressors
- $\hat{V}_{2way}[\hat{\beta}]$ not guaranteed to be positive semi-definite

Few clusters

Wald t-statistics

- $w = \frac{\hat{\beta}-\beta_0}{s_{\hat{\beta}}}$
- If $G \rightarrow \infty$, then $w \sim N(0,1)$ under $H_0 : \beta = \beta_0$
- Finite G: unknown distribution of w. Common to use $w \sim T(G-1)$

Few clusters

Wald t-statistics

- $w = \frac{\hat{\beta}-\beta_0}{s_{\hat{\beta}}}$
- If $G \to \infty$, then $w \sim N(0,1)$ under $H_0 : \beta = \beta_0$
- Finite G: unknown distribution of w. Common to use $w \sim T(G-1)$

Problem with Few Clusters

- Despite reasonable precision in estimating β, $\hat{V}_{clu}(\hat{\beta})$ can be downwards-biased
  - OLS leads to "overfitting", with estimated residuals too close to zero to be compared with true errors
- Use wrong critical values from $N(0,1)$ or $T(G-1)$
  - estimate of $\hat{B}$ leads to over-rejection

How few is "few"?

No clear-cut definition of "few". Current consensus:

- Balanced clusters: from less than 20 to less than 50 clusters
- Unbalanced clusters: even more clusters

Solutions

Independent homoskedastic normally distributed errors

- Unbiased variance matrix of error variance $s^2 = \frac{\hat{u}'\hat{u}}{N-K}$
- $w \sim T(N-K)$

Solutions

Independent homoskedastic normally distributed errors

- Unbiased variance matrix of error variance $s^2 = \frac{\hat{u}'\hat{u}}{N-K}$
- $w \sim T(N-K)$

Independent heteroskedastic normally distributed errors

- White's estimate is no longer unbiased
- No way to obtain exact T distribution result for w
- Partial solutions
  - Bias-Corrected CRVE
  - Cluster Bootstrap
  - Improved Critical values using a T-distribution

Solution 1: Bias-Corrected CRVE

Need corrected residuals: $E\left[\hat{u}_g \hat{u}'_g\right] \neq E\left[u_g u'_g\right]$

- $\tilde{u}_g = \sqrt{\frac{G}{G-1}}\,\hat{u}_g$
- $\tilde{u}_g = \left[I_{N_g} - H_{gg}\right]^{-1/2} \hat{u}_g$, where $H_{gg} = X_g \left(X'X\right)^{-1} X'_g$
- $\tilde{u}_g = \sqrt{\frac{G-1}{G}} \left[I_{N_g} - H_{gg}\right]^{-1} \hat{u}_g$

Reduce but not eliminate over-rejection when there are few clusters

Solution 2: Cluster Bootstrap

Idea

- True size Wald test is $0.05 + O(G^{-1})$
- Appropriate bootstrap with asymptotic refinement can obtain true size of $0.05 + O(G^{-3/2})$

Asymptotic refinement

- Achieved bootstrapping a statistics that is asymptotically pivotal: asymptotic distribution does not depend on any unknown parameter
- $\hat{\beta}$ not a.p.
- Wald t-statistics is a.p.

Solution 2: Percentile-t Bootstrap

Obtain B (999) draws, $w_1^*, ..., w_B^*$ from the distribution of the Wald t-statistics as follows:

1. Obtain G clusters $\{(y_1^*, X_1^*), ..., (y_G^*, X_G^*)\}$ by one cluster method (e.g. pairs cluster resampling)
2. Compute OLS estimate $\hat{\beta}_b^*$
3. Calculate the Wald test statistics $w_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{s_{\hat{\beta}_b^*}}$

Cameron, Gelbach, and Miller (2008): Not eliminate test over-rejection

Solution 2: Wild Cluster Bootstrap

Idea

- Estimate main model imposing null hypothesis that you wish to test to give estimate of $\tilde{\beta}_{H_0}$

Example: test statistical significance of one OLS regressor.

- Regress y_{ig} on all components of x_{ig} except regressor with coefficient 0
- Obtain residuals $\tilde{u}_{ig} = y_{ig} - x'_{ig}\tilde{\beta}_{H_0}$

Solution 2: Wild Cluster Bootstrap, cont'd

Implementation

1. Obtain b^{th} resample
    1. Obtain $\tilde{u}_{ig} = y_{ig} - x'_{ig}\tilde{\beta}_{H_0}$
    2. Randomly assign cluster g the weight $d_g = \begin{cases} -1 & \text{w.p. } 0.5 \\ 1 & \text{w.p. } 0.5 \end{cases}$
    3. Generate new pseudo-residuals $u^*_{ig} = d_g \times \tilde{u}_{ig}$ and new outcome variables $y^*_{ig} = x'_{ig}\tilde{\beta}_{H_0} + u^*_{ig}$
2. Compute OLS estimate $\hat{\beta}^*_b$
3. Calculate the Wald test statistics $w^*_b = \frac{\hat{\beta}^*_b - \hat{\beta}}{s_{\hat{\beta}^*_b}}$

Essentially

- Replacing y_g in each resample with $y^*_g = X\tilde{\beta}_{H_0} + \tilde{u}_g$ or $y^*_g = X\tilde{\beta}_{H_0} - \tilde{u}_g$
- Obtain 2^G unique values of $w^*_1, ..., w^*_B$

Solution 2: Wild Cluster Bootstrap, cont'd

$w_1^*, ..., w_B^*$ cannot be used directly to form critical values for a C.I.

- Used to provide a p-value for testing a hypothesis
- To form C.I. need to invert sequence of tests over a range of candidate H_0. 95 % C.I. is set of β_0 for which $p \geq 0.05$

Limitation

- G=5 → 16 possible values of $w_1^*, ..., w_B^*$

Solution 2: Examine distribution of bootstrapped values

Examination

- Summary statistics
- Sample size
- Largest and smallest value
- Histogram
- Missing values

Examples

- If one cluster is an outlier: histogram might have a big mass that sits separately
- Dummy variable in RHS: obtain missing values for w^* because of low variation in treatment in some samples
- Huge $\hat{\beta}^*$ might be caused by bootstrapping drawing nearly multicollinear samples

Solution 3: Improved critical values using a T-distribution

At a minimum

- Use T(G-1) instead of N(0,1)
- For models with L regressors invariant within cluster: use T(G-L)

Solution 3: Data-determined Degrees of Freedom

- Distribution w approximated by $T(v^*)$
- v^* such that first two moments of $v^*\left(s^2_{\hat{\beta}}/\sigma^2_{\hat{\beta}}\right)$ equal the first two moments of the $\chi(v^*)$

$$v^* = \frac{\left(\sum_{j=1}^{G} \lambda_j\right)^2}{\left(\sum_{j=1}^{G} \lambda_j^2\right)} \tag{2}$$

  - λ_j are eigenvalues of $G'\Omega G$
  - G has g^{th} column $(I_N - H)'_g A_g X_g \left(X'X\right)^{-1} e_k$
  - $H = X\left(X'X\right)^{-1} X'$
  - $A_g = \left(I_{N_g} - H_{gg}\right)^{-1/2}$
  - e_k is a vector of zeros aside from 1 in the k^{th} position if $\hat{\beta} = \hat{\beta}_k$

Solution 3: Approximate w using effective number of clusters

- Distribution w approximated by $T(G^*)$
- δ measures cluster heterogeneity

$$G^* = \frac{G}{1+\delta} \tag{3}$$

- $\delta = \frac{1}{G}\sum_{g=1}^{G}\left[\frac{(\gamma_g-\bar{\gamma})^2}{\bar{\gamma}^2}\right]$
- $\gamma_g = e_k'\left(X'X\right)^{-1} X_g \Omega_g X_g \left(X'X\right)^{-1} e_k$

Empirical example

- Individual-level cross section data
- Random "policy" variable generated: equal to one for one-half of states and zero for other half
- Placebo treatment should be statistically insignificant in 95 % of tests performed at significance 0.05
- Montecarlo
  - 1000 replications
  - Generate a dataset by sampling with replacement states
  - Draw states from 3 % sample of individuals within each state. Average of 40 observations per cluster
- Explore effect of the number of clusters

Montecarlo exercise

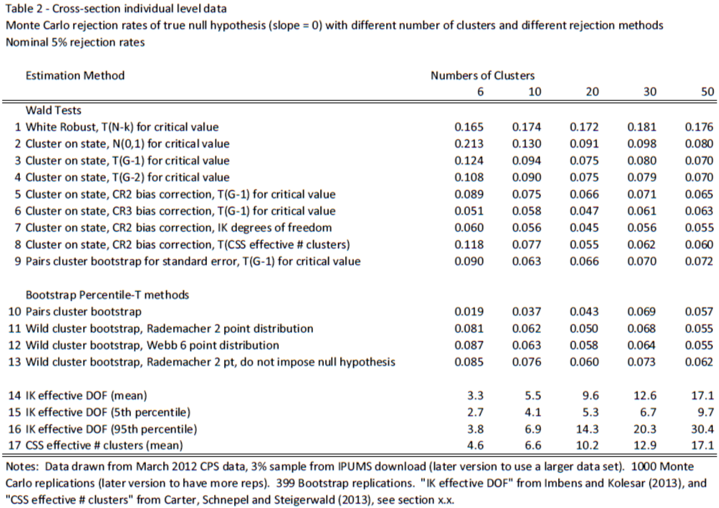

Table 2 - Cross-section individual level data

Monte Carlo rejection rates of true null hypothesis (slope = 0) with different number of clusters and different rejection methods

Nominal 5% rejection rates

	Estimation Method	Numbers of Clusters				
		6	10	20	30	50
	Wald Tests					
1	White Robust, T(N-k) for critical value	0.165	0.174	0.172	0.181	0.176
2	Cluster on state, N(0,1) for critical value	0.213	0.130	0.091	0.098	0.080
3	Cluster on state, T(G-1) for critical value	0.124	0.094	0.075	0.080	0.070
4	Cluster on state, T(G-2) for critical value	0.108	0.090	0.075	0.079	0.070
5	Cluster on state, CR2 bias correction, T(G-1) for critical value	0.089	0.075	0.066	0.071	0.065
6	Cluster on state, CR3 bias correction, T(G-1) for critical value	0.051	0.058	0.047	0.061	0.063
7	Cluster on state, CR2 bias correction, IK degrees of freedom	0.060	0.056	0.045	0.056	0.055
8	Cluster on state, CR2 bias correction, T(CSS effective # clusters)	0.118	0.077	0.055	0.062	0.060
9	Pairs cluster bootstrap for standard error, T(G-1) for critical value	0.090	0.063	0.066	0.070	0.072
	Bootstrap Percentile-T methods					
10	Pairs cluster bootstrap	0.019	0.037	0.043	0.069	0.057
11	Wild cluster bootstrap, Rademacher 2 point distribution	0.081	0.062	0.050	0.068	0.055
12	Wild cluster bootstrap, Webb 6 point distribution	0.087	0.063	0.058	0.064	0.055
13	Wild cluster bootstrap, Rademacher 2 pt, do not impose null hypothesis	0.085	0.076	0.060	0.073	0.062
14	IK effective DOF (mean)	3.3	5.5	9.6	12.6	17.1
15	IK effective DOF (5th percentile)	2.7	4.1	5.3	6.7	9.7
16	IK effective DOF (95th percentile)	3.8	6.9	14.3	20.3	30.4
17	CSS effective # clusters (mean)	4.6	6.6	10.2	12.9	17.1

Notes: Data drawn from March 2012 CPS data, 3% sample from IPUMS download (later version to use a larger data set). 1000 Monte Carlo replications (later version to have more reps). 399 Bootstrap replications. "IK effective DOF" from Imbens and Kolesar (2013), and "CSS effective # clusters" from Carter, Schnepel and Steigerwald (2013), see section x.x.

Montecarlo exercise

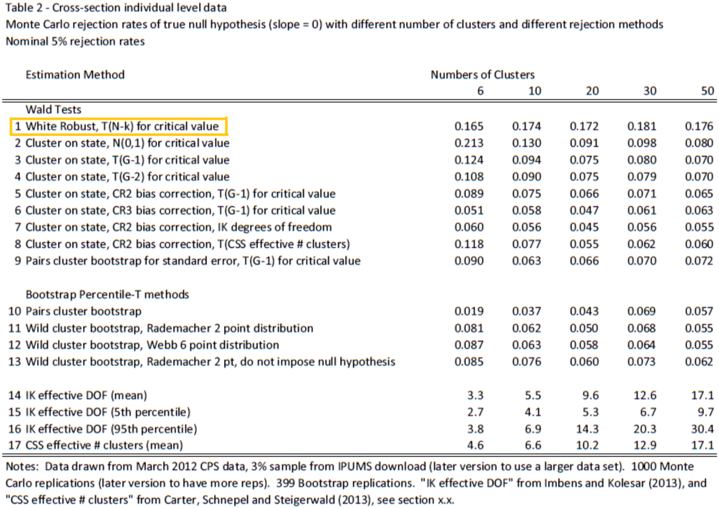

Table 2 - Cross-section individual level data
Monte Carlo rejection rates of true null hypothesis (slope = 0) with different number of clusters and different rejection methods
Nominal 5% rejection rates

Estimation Method	Numbers of Clusters				
	6	10	20	30	50
Wald Tests					
1 White Robust, T(N-k) for critical value	0.165	0.174	0.172	0.181	0.176
2 Cluster on state, N(0,1) for critical value	0.213	0.130	0.091	0.098	0.080
3 Cluster on state, T(G-1) for critical value	0.124	0.094	0.075	0.080	0.070
4 Cluster on state, T(G-2) for critical value	0.108	0.090	0.075	0.079	0.070
5 Cluster on state, CR2 bias correction, T(G-1) for critical value	0.089	0.075	0.066	0.071	0.065
6 Cluster on state, CR3 bias correction, T(G-1) for critical value	0.051	0.058	0.047	0.061	0.063
7 Cluster on state, CR2 bias correction, IK degrees of freedom	0.060	0.056	0.045	0.056	0.055
8 Cluster on state, CR2 bias correction, T(CSS effective # clusters)	0.118	0.077	0.055	0.062	0.060
9 Pairs cluster bootstrap for standard error, T(G-1) for critical value	0.090	0.063	0.066	0.070	0.072
Bootstrap Percentile-T methods					
10 Pairs cluster bootstrap	0.019	0.037	0.043	0.069	0.057
11 Wild cluster bootstrap, Rademacher 2 point distribution	0.081	0.062	0.050	0.068	0.055
12 Wild cluster bootstrap, Webb 6 point distribution	0.087	0.063	0.058	0.064	0.055
13 Wild cluster bootstrap, Rademacher 2 pt, do not impose null hypothesis	0.085	0.076	0.060	0.073	0.062
14 IK effective DOF (mean)	3.3	5.5	9.6	12.6	17.1
15 IK effective DOF (5th percentile)	2.7	4.1	5.3	6.7	9.7
16 IK effective DOF (95th percentile)	3.8	6.9	14.3	20.3	30.4
17 CSS effective # clusters (mean)	4.6	6.6	10.2	12.9	17.1

Notes: Data drawn from March 2012 CPS data, 3% sample from IPUMS download (later version to use a larger data set). 1000 Monte Carlo replications (later version to have more reps). 399 Bootstrap replications. "IK effective DOF" from Imbens and Kolesar (2013), and "CSS effective # clusters" from Carter, Schnepel and Steigerwald (2013), see section x.x.

G=50, ignoring clustering leads to great over-rejection but less variable

Montecarlo exercise

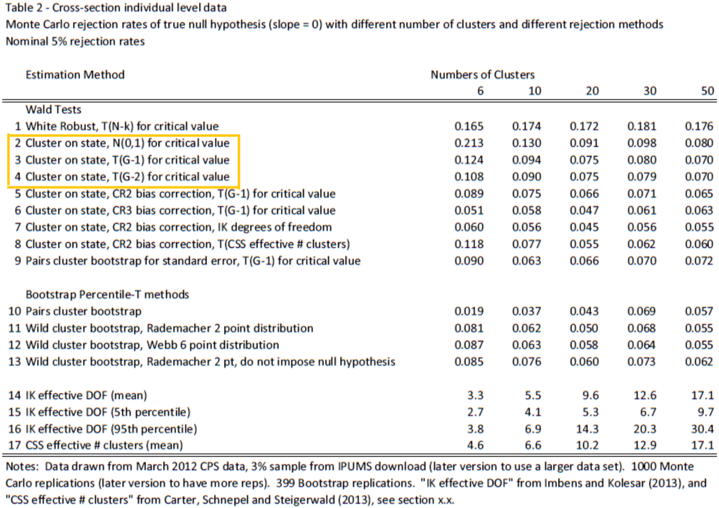

Table 2 - Cross-section individual level data
Monte Carlo rejection rates of true null hypothesis (slope = 0) with different number of clusters and different rejection methods
Nominal 5% rejection rates

	Estimation Method	Numbers of Clusters				
		6	10	20	30	50
	Wald Tests					
1	White Robust, T(N-k) for critical value	0.165	0.174	0.172	0.181	0.176
2	Cluster on state, N(0,1) for critical value	0.213	0.130	0.091	0.098	0.080
3	Cluster on state, T(G-1) for critical value	0.124	0.094	0.075	0.080	0.070
4	Cluster on state, T(G-2) for critical value	0.108	0.090	0.075	0.079	0.070
5	Cluster on state, CR2 bias correction, T(G-1) for critical value	0.089	0.075	0.066	0.071	0.065
6	Cluster on state, CR3 bias correction, T(G-1) for critical value	0.051	0.058	0.047	0.061	0.063
7	Cluster on state, CR2 bias correction, IK degrees of freedom	0.060	0.056	0.045	0.056	0.055
8	Cluster on state, CR2 bias correction, T(CSS effective # clusters)	0.118	0.077	0.055	0.062	0.060
9	Pairs cluster bootstrap for standard error, T(G-1) for critical value	0.090	0.063	0.066	0.070	0.072
	Bootstrap Percentile-T methods					
10	Pairs cluster bootstrap	0.019	0.037	0.043	0.069	0.057
11	Wild cluster bootstrap, Rademacher 2 point distribution	0.081	0.062	0.050	0.068	0.055
12	Wild cluster bootstrap, Webb 6 point distribution	0.087	0.063	0.058	0.064	0.055
13	Wild cluster bootstrap, Rademacher 2 pt, do not impose null hypothesis	0.085	0.076	0.060	0.073	0.062
14	IK effective DOF (mean)	3.3	5.5	9.6	12.6	17.1
15	IK effective DOF (5th percentile)	2.7	4.1	5.3	6.7	9.7
16	IK effective DOF (95th percentile)	3.8	6.9	14.3	20.3	30.4
17	CSS effective # clusters (mean)	4.6	6.6	10.2	12.9	17.1

Notes: Data drawn from March 2012 CPS data, 3% sample from IPUMS download (later version to use a larger data set). 1000 Monte Carlo replications (later version to have more reps). 399 Bootstrap replications. "IK effective DOF" from Imbens and Kolesar (2013), and "CSS effective # clusters" from Carter, Schnepel and Steigerwald (2013), see section x.x.

Use different critical values: improvement but stil over-rejection

Montecarlo exercise

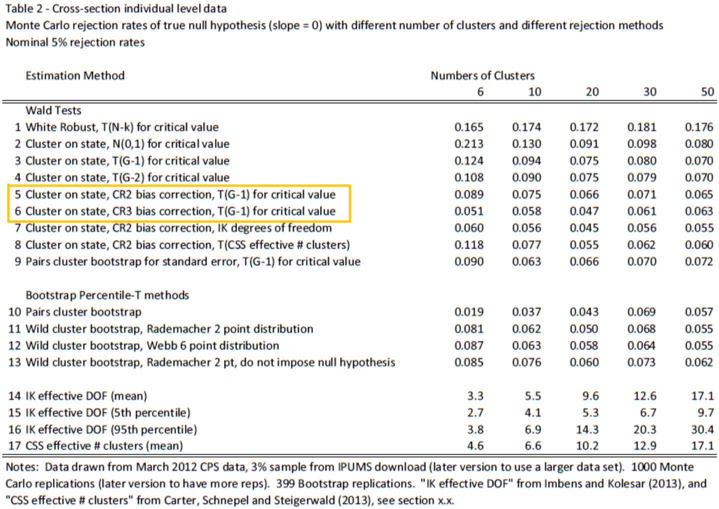

Table 2 - Cross-section individual level data

Monte Carlo rejection rates of true null hypothesis (slope = 0) with different number of clusters and different rejection methods

Nominal 5% rejection rates

	Estimation Method	Numbers of Clusters				
		6	10	20	30	50
	Wald Tests					
1	White Robust, T(N-k) for critical value	0.165	0.174	0.172	0.181	0.176
2	Cluster on state, N(0,1) for critical value	0.213	0.130	0.091	0.098	0.080
3	Cluster on state, T(G-1) for critical value	0.124	0.094	0.075	0.080	0.070
4	Cluster on state, T(G-2) for critical value	0.108	0.090	0.075	0.079	0.070
5	Cluster on state, CR2 bias correction, T(G-1) for critical value	0.089	0.075	0.066	0.071	0.065
6	Cluster on state, CR3 bias correction, T(G-1) for critical value	0.051	0.058	0.047	0.061	0.063
7	Cluster on state, CR2 bias correction, IK degrees of freedom	0.060	0.056	0.045	0.056	0.055
8	Cluster on state, CR2 bias correction, T(CSS effective # clusters)	0.118	0.077	0.055	0.062	0.060
9	Pairs cluster bootstrap for standard error, T(G-1) for critical value	0.090	0.063	0.066	0.070	0.072
	Bootstrap Percentile-T methods					
10	Pairs cluster bootstrap	0.019	0.037	0.043	0.069	0.057
11	Wild cluster bootstrap, Rademacher 2 point distribution	0.081	0.062	0.050	0.068	0.055
12	Wild cluster bootstrap, Webb 6 point distribution	0.087	0.063	0.058	0.064	0.055
13	Wild cluster bootstrap, Rademacher 2 pt, do not impose null hypothesis	0.085	0.076	0.060	0.073	0.062
14	IK effective DOF (mean)	3.3	5.5	9.6	12.6	17.1
15	IK effective DOF (5th percentile)	2.7	4.1	5.3	6.7	9.7
16	IK effective DOF (95th percentile)	3.8	6.9	14.3	20.3	30.4
17	CSS effective # clusters (mean)	4.6	6.6	10.2	12.9	17.1

Notes: Data drawn from March 2012 CPS data, 3% sample from IPUMS download (later version to use a larger data set). 1000 Monte Carlo replications (later version to have more reps). 399 Bootstrap replications. "IK effective DOF" from Imbens and Kolesar (2013), and "CSS effective # clusters" from Carter, Schnepel and Steigerwald (2013), see section x.x.

Residual bias adjustment: perform best with fewer clusters

Montecarlo exercise

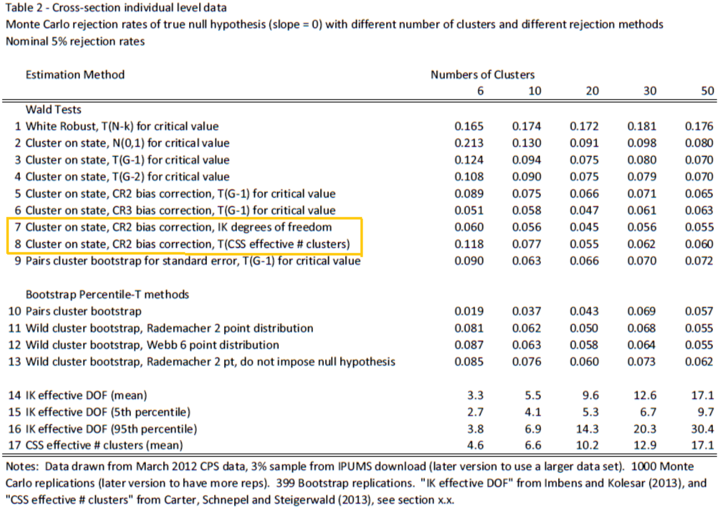

Table 2 - Cross-section individual level data

Monte Carlo rejection rates of true null hypothesis (slope = 0) with different number of clusters and different rejection methods

Nominal 5% rejection rates

	Estimation Method	Numbers of Clusters				
		6	10	20	30	50
	Wald Tests					
1	White Robust, T(N-k) for critical value	0.165	0.174	0.172	0.181	0.176
2	Cluster on state, N(0,1) for critical value	0.213	0.130	0.091	0.098	0.080
3	Cluster on state, T(G-1) for critical value	0.124	0.094	0.075	0.080	0.070
4	Cluster on state, T(G-2) for critical value	0.108	0.090	0.075	0.079	0.070
5	Cluster on state, CR2 bias correction, T(G-1) for critical value	0.089	0.075	0.066	0.071	0.065
6	Cluster on state, CR3 bias correction, T(G-1) for critical value	0.051	0.058	0.047	0.061	0.063
7	Cluster on state, CR2 bias correction, IK degrees of freedom	0.060	0.056	0.045	0.056	0.055
8	Cluster on state, CR2 bias correction, T(CSS effective # clusters)	0.118	0.077	0.055	0.062	0.060
9	Pairs cluster bootstrap for standard error, T(G-1) for critical value	0.090	0.063	0.066	0.070	0.072
	Bootstrap Percentile-T methods					
10	Pairs cluster bootstrap	0.019	0.037	0.043	0.069	0.057
11	Wild cluster bootstrap, Rademacher 2 point distribution	0.081	0.062	0.050	0.068	0.055
12	Wild cluster bootstrap, Webb 6 point distribution	0.087	0.063	0.058	0.064	0.055
13	Wild cluster bootstrap, Rademacher 2 pt, do not impose null hypothesis	0.085	0.076	0.060	0.073	0.062
14	IK effective DOF (mean)	3.3	5.5	9.6	12.6	17.1
15	IK effective DOF (5th percentile)	2.7	4.1	5.3	6.7	9.7
16	IK effective DOF (95th percentile)	3.8	6.9	14.3	20.3	30.4
17	CSS effective # clusters (mean)	4.6	6.6	10.2	12.9	17.1

Notes: Data drawn from March 2012 CPS data, 3% sample from IPUMS download (later version to use a larger data set). 1000 Monte Carlo replications (later version to have more reps). 399 Bootstrap replications. "IK effective DOF" from Imbens and Kolesar (2013), and "CSS effective # clusters" from Carter, Schnepel and Steigerwald (2013), see section x.x.

Data-determined df: improvement with more clusters

Montecarlo exercise

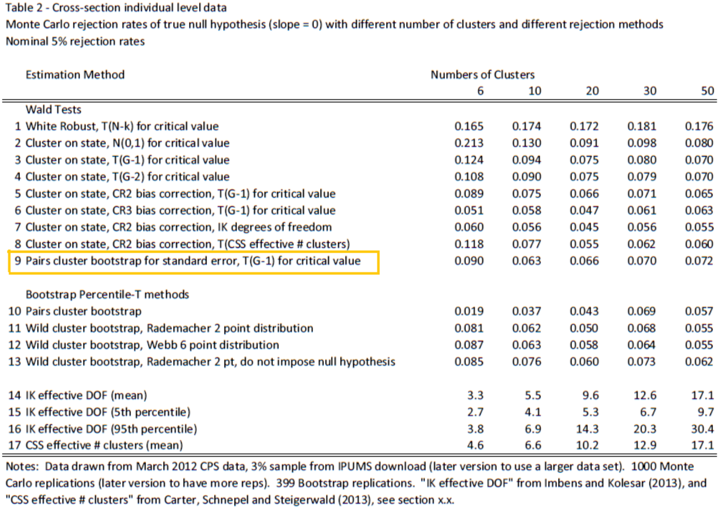

Table 2 - Cross-section individual level data
Monte Carlo rejection rates of true null hypothesis (slope = 0) with different number of clusters and different rejection methods
Nominal 5% rejection rates

Estimation Method	Numbers of Clusters				
	6	10	20	30	50
Wald Tests					
1 White Robust, T(N-k) for critical value	0.165	0.174	0.172	0.181	0.176
2 Cluster on state, N(0,1) for critical value	0.213	0.130	0.091	0.098	0.080
3 Cluster on state, T(G-1) for critical value	0.124	0.094	0.075	0.080	0.070
4 Cluster on state, T(G-2) for critical value	0.108	0.090	0.075	0.079	0.070
5 Cluster on state, CR2 bias correction, T(G-1) for critical value	0.089	0.075	0.066	0.071	0.065
6 Cluster on state, CR3 bias correction, T(G-1) for critical value	0.051	0.058	0.047	0.061	0.063
7 Cluster on state, CR2 bias correction, IK degrees of freedom	0.060	0.056	0.045	0.056	0.055
8 Cluster on state, CR2 bias correction, T(CSS effective # clusters)	0.118	0.077	0.055	0.062	0.060
9 Pairs cluster bootstrap for standard error, T(G-1) for critical value	0.090	0.063	0.066	0.070	0.072
Bootstrap Percentile-T methods					
10 Pairs cluster bootstrap	0.019	0.037	0.043	0.069	0.057
11 Wild cluster bootstrap, Rademacher 2 point distribution	0.081	0.062	0.050	0.068	0.055
12 Wild cluster bootstrap, Webb 6 point distribution	0.087	0.063	0.058	0.064	0.055
13 Wild cluster bootstrap, Rademacher 2 pt, do not impose null hypothesis	0.085	0.076	0.060	0.073	0.062
14 IK effective DOF (mean)	3.3	5.5	9.6	12.6	17.1
15 IK effective DOF (5th percentile)	2.7	4.1	5.3	6.7	9.7
16 IK effective DOF (95th percentile)	3.8	6.9	14.3	20.3	30.4
17 CSS effective # clusters (mean)	4.6	6.6	10.2	12.9	17.1

Notes: Data drawn from March 2012 CPS data, 3% sample from IPUMS download (later version to use a larger data set). 1000 Monte Carlo replications (later version to have more reps). 399 Bootstrap replications. "IK effective DOF" from Imbens and Kolesar (2013), and "CSS effective # clusters" from Carter, Schnepel and Steigerwald (2013), see section x.x.

Bootstrap without asymptotic refinement: as 2-4

Montecarlo exercise

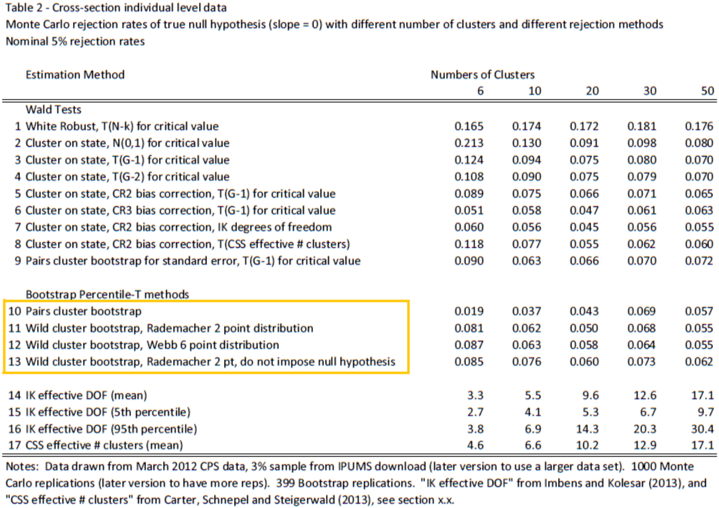

Table 2 - Cross-section individual level data

Monte Carlo rejection rates of true null hypothesis (slope = 0) with different number of clusters and different rejection methods

Nominal 5% rejection rates

	Estimation Method	Numbers of Clusters 6	10	20	30	50
	Wald Tests					
1	White Robust, T(N-k) for critical value	0.165	0.174	0.172	0.181	0.176
2	Cluster on state, N(0,1) for critical value	0.213	0.130	0.091	0.098	0.080
3	Cluster on state, T(G-1) for critical value	0.124	0.094	0.075	0.080	0.070
4	Cluster on state, T(G-2) for critical value	0.108	0.090	0.075	0.079	0.070
5	Cluster on state, CR2 bias correction, T(G-1) for critical value	0.089	0.075	0.066	0.071	0.065
6	Cluster on state, CR3 bias correction, T(G-1) for critical value	0.051	0.058	0.047	0.061	0.063
7	Cluster on state, CR2 bias correction, IK degrees of freedom	0.060	0.056	0.045	0.056	0.055
8	Cluster on state, CR2 bias correction, T(CSS effective # clusters)	0.118	0.077	0.055	0.062	0.060
9	Pairs cluster bootstrap for standard error, T(G-1) for critical value	0.090	0.063	0.066	0.070	0.072
	Bootstrap Percentile-T methods					
10	Pairs cluster bootstrap	0.019	0.037	0.043	0.069	0.057
11	Wild cluster bootstrap, Rademacher 2 point distribution	0.081	0.062	0.050	0.068	0.055
12	Wild cluster bootstrap, Webb 6 point distribution	0.087	0.063	0.058	0.064	0.055
13	Wild cluster bootstrap, Rademacher 2 pt, do not impose null hypothesis	0.085	0.076	0.060	0.073	0.062
14	IK effective DOF (mean)	3.3	5.5	9.6	12.6	17.1
15	IK effective DOF (5th percentile)	2.7	4.1	5.3	6.7	9.7
16	IK effective DOF (95th percentile)	3.8	6.9	14.3	20.3	30.4
17	CSS effective # clusters (mean)	4.6	6.6	10.2	12.9	17.1

Notes: Data drawn from March 2012 CPS data, 3% sample from IPUMS download (later version to use a larger data set). 1000 Monte Carlo replications (later version to have more reps). 399 Bootstrap replications. "IK effective DOF" from Imbens and Kolesar (2013), and "CSS effective # clusters" from Carter, Schnepel and Steigerwald (2013), see section x.x.

Bootstrap with asymptotic refinement: mild over-rejection