Jackknife Standard Errors for Clustered Regression

Bruce E. Hansen
University of Wisconsin

December 2023

This paper

- Studies variance estimation and confidence interval coverage
  - Clustered regression model
- Bias
  - Standard estimators have arbitrarily large bias
  - Jackknife estimator is conservative – never downward biased
- Confidence intervals
  - Standard intervals have coverage rates arbitrarily close to 0
  - Jackknife interval has coverage bounded by Cauchy distribution
- Theory holds under minimal assumptions, allowing arbitrary cluster sizes, regressor leverage, within-cluster correlation, heteroskedasticity, regression with a single treated cluster, fixed effects, and delete-cluster invertibility failures
- Adjusted critical values
  - Data-based d.o.f. can approximately control size

Model: Clustered Regression

- Cluster-level stacked observations $(\mathbf{Y}_g, \mathbf{X}_g)$
  - $\mathbf{Y}_g$ is $n_g \times 1$
  - $\mathbf{X}_g$ is $n_g \times k$,
  - G clusters
- Regression model

$$
\begin{aligned}
\mathbf{Y}_g &= \mathbf{X}_g \beta + \mathbf{e}_g \\
\mathbb{E}\left[\mathbf{e}_g\right] &= 0
\end{aligned}
$$

- Fixed regressors (conditional)
- Cross-section regression when $n_g = 1$
- Least squares estimation

Cluster Covariance Matrices

- $\mathbb{E}\left[\mathbf{e}_g \mathbf{e}_g'\right] = \mathbf{\Sigma}_g, \qquad n_g \times n_g$
  - Varies by cluster g
  - Can depend on regressors.
  - Allows unconditional and conditional heteroskedasticity.
  - Includes heteroskedastic regression in non-clustered $(n_g = 1)$ case.
- Popular model in contemporary econometrics.
  - $\mathbf{\Sigma}_g$ is treated as unknown and unstructured.
  - Covers both clustered and panel settings.

Least Squares Estimation Variance

- $\mathbf{V} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\sum_{g=1}^{G} \mathbf{X}_g' \Sigma_g \mathbf{X}_g\right) \left(\mathbf{X}'\mathbf{X}\right)^{-1}$
- Cluster-Robust Variance Estimator (CRVE)

$$\widehat{\mathbf{V}}_1 = \frac{G\left(n-1\right)}{\left(G-1\right)\left(n-k\right)} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\sum_{g=1}^{G} \mathbf{X}_g' \widehat{\mathbf{e}}_g \widehat{\mathbf{e}}_g' \mathbf{X}_g\right) \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

- Liang and Zeger (1986), Arellano (1987), Stata
- In the absence of clustering, $\widehat{\mathbf{V}}_1$ simplies to HC_1.
- CRVE_1 and HC_1 dominate empirical practice.

CRVE$_2$

- Bell and McCaffrey (2002), Imbens and Kolesar (2016)
- $\widehat{\mathbf{V}}_2 = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\sum_{g=1}^{G} \mathbf{X}_g' \mathbf{M}_g^{+1/2} \widehat{\mathbf{e}}_g \widehat{\mathbf{e}}_g' \mathbf{M}_g^{+1/2} \mathbf{X}_g \right) \left(\mathbf{X}'\mathbf{X}\right)^{-1}$
  - $\mathbf{M}_g = \mathbf{I}_{n_g} - \mathbf{X}_g \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}_g'$
  - $\mathbf{A}^+ =$ Moore-Penrose pseudoinverse
- Motivation: Unbiased when errors are i.i.d. and $\mathbf{M}_g$ are invertible
- Pseudoinverse allows $\widehat{\mathbf{V}}_2$ to be calculated under invertibility failures
  - Kolesar (2022)
- Implemented in Stata 18

Jackknife Variance Estimator

- Delete-one-cluster estimators

$$\widehat{\beta}_{-g} = \left(\sum_{j \neq g} \mathbf{X}_j' \mathbf{X}_j \right)^{-1} \left(\sum_{j \neq g} \mathbf{X}_j' \mathbf{Y}_j \right)$$

- Jackknife variance estimators (Tukey, 1958)

$$\begin{aligned} \widehat{\mathbf{V}}_3 &= \frac{G-1}{G} \sum_{g=1}^{G} \left(\widehat{\beta}_{-g} - \overline{\beta} \right) \left(\widehat{\beta}_{-g} - \overline{\beta} \right)' \\ \overline{\beta} &= \frac{1}{G} \sum_{g=1}^{G} \widehat{\beta}_{-g} \\ \widehat{\mathbf{V}}_4 &= \frac{G-1}{G} \sum_{g=1}^{G} \left(\widehat{\beta}_{-g} - \widehat{\beta} \right) \left(\widehat{\beta}_{-g} - \widehat{\beta} \right)' \end{aligned}$$

- Cluster version of HC_3 (MacKinnon and White, 1985)

Invertibility Failures

- Jackknife variance undefined if there is a cluster g for which $\mathbf{X}'\mathbf{X} - \mathbf{X}_g'\mathbf{X}_g$ is noninvertible
  - Identical to context where $\mathbf{M}_g$ is noninvertible
- Occurs when $\mathbf{X}$ includes
  - Fixed effects
  - Treatment dummy with a single treated cluster
  - One component of $\mathbf{X}_g$ non-zero only for a single cluster
  - Saturated regressions with sparse cell proportions
- Commonplace in empirical applications
- Common Solution (e.g. Stata)
  - Drop noninvertible clusters
  - Similar to dropping noninvertible designs in bootstrap
  - We adopt this definition for $\widehat{\mathbf{V}}_3$ & $\widehat{\mathbf{V}}_4$: noninvertible clusters dropped
  - Properties previously unexplored

Recommended Jackknife Estimator

- Generalized delete-one-cluster estimators

$$\widetilde{\beta}_{-g} = \left(\sum_{j \neq g} \mathbf{X}_j' \mathbf{X}_j \right)^{+} \left(\sum_{j \neq g} \mathbf{X}_j' \mathbf{Y}_j \right)$$

$$\widehat{\mathbf{V}}_5 = \sum_{g=1}^{G} \left(\widetilde{\beta}_{-g} - \widehat{\beta} \right) \left(\widetilde{\beta}_{-g} - \widehat{\beta} \right)'$$

- Differs from $\widehat{\mathbf{V}}_3$ and $\widehat{\mathbf{V}}_4$ in three respects, leading to $\widehat{\mathbf{V}}_5 > \widehat{\mathbf{V}}_4 > \widehat{\mathbf{V}}_3$
  - $\widehat{\mathbf{V}}_5$ does not drop noninvertible clusters
  - $\widehat{\mathbf{V}}_5$ is centered at $\widehat{\beta}$
  - $\widehat{\mathbf{V}}_5$ does not have degree-of-freedom correction $(G-1)/G$
- Properties are explored in this paper.

Variance Estimation Bias

- Classical variance estimators are unbiased under classical assumptions.
- What we now show
  - Proposed jackknife estimator is never downward biased.
  - Existing estimators can have arbitrarily large bias.
- We focus on downward bias as this is the issue which causes undercoverage of confidence intervals and oversized tests.

Jackknife Variance Estimator is Conservative

Theorem 1: In a linear regression with full rank $\mathbf{X}$

$$\mathbb{E}\left[\widehat{\mathbf{V}}_5\right] \geq \mathbf{V}$$

- This holds in any clustered regression setting
- Minimal assumptions
- Allows clusterwise noninvertibility (e.g. fixed effects)

Worst Case Bias

- Focus on individual coefficients $\theta = R'\beta$ and their standard errors $\widehat{v}_j$
- Sets of Models for fixed G and k
  - $\mathcal{F}$ is the class of all regressor and covariance matrices $(\mathbf{X}, \mathbf{\Sigma})$ such that $\mathbf{X}$ is full rank, $\mathbf{\Sigma}$ has finite elements, and $\nu^2 > 0$.
  - $\mathcal{F}^* \subset \mathcal{F}$ is the subset where $\mathbf{X}$ satisfies clusterwise invertibility.
  - $\mathcal{F}_0 \subset \mathcal{F}$ and $\mathcal{F}_0^* \subset \mathcal{F}^*$ are the subsets where $\mathbf{\Sigma} = \mathbf{I}_n\sigma^2$.

Theorem 2: Variance estimators can be arbitrarily biased

$$
\begin{aligned}
\inf_{(\mathbf{X},\boldsymbol{\Sigma})\in\mathcal{F}_0^*} \frac{\mathbb{E}\left[\widehat{v}_1^2\right]}{v^2} &= 0 \\
\inf_{(\mathbf{X},\boldsymbol{\Sigma})\in\mathcal{F}^*} \frac{\mathbb{E}\left[\widehat{v}_2^2\right]}{v^2} &= 0 \\
\inf_{(\mathbf{X},\boldsymbol{\Sigma})\in\mathcal{F}^*} \frac{\mathbb{E}\left[\widehat{v}_3^2\right]}{v^2} &= \left(\frac{G-1}{G}\right)^2 < 1 \\
\inf_{(\mathbf{X},\boldsymbol{\Sigma})\in\mathcal{F}^*} \frac{\mathbb{E}\left[\widehat{v}_4^2\right]}{v^2} &= \frac{G-1}{G} < 1 \\
\inf_{(\mathbf{X},\boldsymbol{\Sigma})\in\mathcal{F}_0} \frac{\mathbb{E}\left[\widehat{v}_3^2\right]}{v^2} &= 0 \\
\inf_{(\mathbf{X},\boldsymbol{\Sigma})\in\mathcal{F}_0} \frac{\mathbb{E}\left[\widehat{v}_4^2\right]}{v^2} &= 0
\end{aligned}
$$

Summary of Theorem 2

- CRVE_1 standard error has arbitrary large downward bias (*fully downward biased*) even under i.i.d. errors and clusterwise invertibility. Severe bias can arise from extreme regressor leverage, including unbalanced cluster sizes.
- CRVE_2 standard error has full downward bias when we allow general covariance matrices (heteroskedasticity and correlation).
- Conventional jackknife standard errors $\widehat{v}_3$ and $\widehat{v}_4$ are downward biased under clusterwise invertibility.
- Conventional jackknife standard errors $\widehat{v}_3$ and $\widehat{v}_4$ are fully downward biased under clusterwise noninvertibility.
    - This shows that the "solution" of deleting noninvertible clusters is a non-solution.
- Recommended standard errors $\widehat{v}_5$ are never downward biased.
    - Allows heteroskedasticity, correlation, clusterwise noninvertibility.

Confidence Intervals

Given a critical value c, a confidence interval for θ is

$$\widehat{C}_j(c) = \widehat{\theta} \pm c\widehat{v}_j$$

Theorem 3: If $\mathbf{e}_g \sim \mathrm{N}(0, \mathbf{\Sigma}_g)$, then for any $1 \leq c < \infty$

$$\inf_{(\mathbf{X},\mathbf{\Sigma})\in\mathcal{F}} \mathbb{P}\left[\theta \in \widehat{C}_5(c)\right] \geq F(c; 1, 1)$$

where $F(x; k_1, k_2)$ is the F distribution.

Interpretation: The jackknife confidence interval has coverage bounded by the Cauchy distribution.

Implication: Size distortion (using the jackknife) is bounded.

Uncoverage of Standard Intervals

Theorem 4: If $\mathbf{e}_g \sim \mathrm{N}(0, \mathbf{\Sigma}_g)$, for any $0 \leq c < \infty$

$$
\begin{aligned}
\inf_{\mathbf{X} \in \mathcal{F}_0^*} \mathbb{P}\left[\theta \in \widehat{C}_1(c)\right] &= 0 \\
\inf_{(\mathbf{X}, \mathbf{\Sigma}) \in \mathcal{F}^*} \mathbb{P}\left[\theta \in \widehat{C}_2(c)\right] &= 0 \\
\inf_{(\mathbf{X}, \mathbf{\Sigma}) \in \mathcal{F}_0} \mathbb{P}\left[\theta \in \widehat{C}_3(c)\right] &= 0 \\
\inf_{(\mathbf{X}, \mathbf{\Sigma}) \in \mathcal{F}_0} \mathbb{P}\left[\theta \in \widehat{C}_4(c)\right] &= 0
\end{aligned}
$$

Interpretation: Intervals can have coverage arbitrarily close to zero.

Improved Inference

- In highly leverage settings t-ratios are non-normally distributed.
- Can we improve on t_{G-1} critical values?
- Bell and McCaffrey (2002) suggested an adjustment.
  - They made this suggestion for HC_2 and CRVE_2.
  - Endorsed by Imbens and Kolesar (2016), coded in Stata 18.
  - We extend this suggestion to the jackknife.

Approximate F Inference

Theorem 5: If $\mathbf{e}_g \sim \mathrm{N}\left(0, \mathbf{\Sigma}_g\right)$,

$$\mathbb{P}\left[\theta \in \widehat{C}_5(c)\right] \gtrapprox F\left(a^2 x; 1, K\right)$$

where a and K are functions of the regressors $\mathbf{X}$ and covariance matrix $\mathbf{\Sigma}$.

- The bound uses a Satterthwaite (1946) approximation.
- **Interpretation**: The jackknife t-ratio is approximately t with a non-standard scale a and d.o.f. K.

Unknown Constants

- The constants a and K depend on the unknown variance matrices $\mathbf{\Sigma}_g$.
  - If they were known, then a and K could be calculated
  - But they are unknown
- Bell-McCaffrey suggest using the reference model $\mathbf{\Sigma}_g = \mathbf{I}_{n_g}\sigma^2$.
- **Theorem 5**: When $\mathbf{\Sigma}_g = \mathbf{I}_{n_g}\sigma^2$, computationally convenient [but lengthy] algebraic expressions for a and K are given in the paper.
- They are functions only of $\mathbf{X}$ and selector vector R.

Adjusted Confidence Intervals

Our recommend adjusted $100(1-\alpha)\%$ confidence interval for θ is

$$\widetilde{C}_5 = \widehat{\theta} \pm \frac{t_K^{1-\alpha/2}\widehat{v}_5}{a}.$$

where

- $t_K^{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the student t distribution with K degrees of freedom.
- $\widehat{v}_5$ is our recommended jackknife standard error.
- R code posted, Stata code forthcoming.

Simulation Evidence

- Baseline model $\mathbf{Y}_g = \alpha + \mathbf{X}_g \beta + \mathbf{e}_g$
- Design 1
  - $\mathbf{X}_g \sim \mathrm{N}(0, \mathbf{I}_g + \mathbf{1}_g \mathbf{1}_g')$
  - $\mathbf{e}_g \sim \mathrm{N}(0, \mathbf{I}_g + \mathbf{1}_g \mathbf{1}_g' + \mathbf{h}_g \mathbf{h}_g')$ where $\mathbf{h}_g' \mathbf{1}_g = 0$
  - $n_g = 10$ for all g
- Design 2: $\mathbf{X}_g \sim \frac{\sqrt{2}}{\exp(2)} \exp\left(\mathrm{N}(0, \mathbf{I}_g + \mathbf{1}_g \mathbf{1}_g')\right)$
- Design 3: n_g heterogeneous
- Designs 4, 5, and 6: $\mathbf{e}_g \sim \mathrm{N}(0, \mathbf{I}_g + \mathbf{X}_g \mathbf{X}_g')$
- 95% Confidence intervals for β
- Number of clusters $G = \{6, 12, 40, 100\}$
- Simulation replications: $20,000$

Confidence Interval Methods

1. $\widehat{\beta} \pm t_{G-1}^{.975} \widehat{v}_1$
2. $\widehat{\beta} \pm t_{G-1}^{.975} \widehat{v}_2$
3. $\widehat{\beta} \pm t_{G-1}^{.975} \widehat{v}_3$
4. $\widehat{\beta} \pm t_{G-1}^{.975} \widehat{v}_4$
5. $\widehat{\beta} \pm t_{G-1}^{.975} \widehat{v}_5$
6. $\widehat{\beta} \pm t_{G-1}^{.975} \widehat{v}_6$, where $\widehat{v}_6$ is from nonparametric pairs cluster bootstrap
7. Bell-McCaffrey: $\widehat{\beta} \pm t_{K}^{.975} \widehat{v}_2$
8. Adjusted t_K : $\widehat{\beta} \pm t_{K}^{.975} \widehat{v}_5 / a$
9. Nonparametric pairs cluster bootstrap symmetric percentile-t using $\widehat{v}_1$
10. Nonparametric pairs cluster bootstrap symmetric percentile-t using $\widehat{v}_5$
11. Wild cluster bootstrap symmetric percentile-t using $\widehat{v}_1$
12. Wild cluster bootstrap symmetric percentile-t using $\widehat{v}_5$

	Conventional t_{G-1}						Adjusted		Boot		Wild	
s.e.	$\widehat{v}_1$	$\widehat{v}_2$	$\widehat{v}_3$	$\widehat{v}_4$	$\widehat{v}_5$	$\widehat{v}_6$	$\widehat{v}_2$	$\widehat{v}_5$	$\widehat{v}_1$	$\widehat{v}_5$	$\widehat{v}_1$	$\widehat{v}_5$
$G = 6$												
D1	.91	.93	.95	.95	**.96**	.94	.95	**.96**	.96	.96	.94	.94
D2	.85	.91	.95	.96	**.97**	.98	.99	**.99**	.93	.96	.97	.96
D3	.83	.90	.95	.95	**.96**	.99	.99	**.99**	.93	.97	.96	.95
D4	.89	.91	.93	.93	**.95**	.90	.94	**.94**	.95	.96	.95	.95
D5	.60	.72	.87	.87	**.89**	.91	.90	**.93**	.80	.92	.85	.91
D6	.57	.70	.87	.87	**.89**	.92	.90	**.93**	.78	.92	.86	.91
$G = 40$												
D1	.94	.94	.95	.95	**.95**	.94	.95	**.95**	.95	.95	.95	.95
D2	.86	.90	.93	.93	**.93**	.96	.97	**.98**	.89	.92	.95	.96
D3	.80	.88	.94	.94	**.94**	.95	.98	**.98**	.90	.95	.94	.96
D4	.93	.93	.94	.94	**.94**	.93	.94	**.94**	.95	.95	.95	.95
D5	.71	.79	.88	.88	**.89**	.84	.93	**.95**	.90	.94	.94	.94
D6	.63	.74	.87	.87	**.87**	.83	.92	**.95**	.86	.93	.91	.93

Model with Noninvertibility

- $\mathbf{Y}_g = \alpha + \mathbf{X}_g \beta + \mathbf{D}_g \gamma + \mathbf{e}_g$
- $\mathbf{D}_g$ is cluster-level dummy variable
- Confidence intervals for β

	Conventional t_{G-1}						Adjusted		Boot		Wild	
s.e.	$\widehat{V}_1$	$\widehat{V}_2$	$\widehat{V}_3$	$\widehat{V}_4$	$\widehat{V}_5$	$\widehat{V}_6$	$\widehat{V}_2$	$\widehat{V}_5$	$\widehat{V}_1$	$\widehat{V}_5$	$\widehat{V}_1$	$\widehat{V}_5$
$G = 6$												
D1	.91	.93	.93	.94	**.96**	.94	.95	**.96**	.96	.97	.94	.94
D2	.85	.91	.92	.93	**.97**	.98	.99	**.99**	.93	.97	.96	.96
D3	.83	.90	.87	.89	**.96**	.99	.99	**.99**	.93	.97	.96	.96
D4	.89	.91	.91	.91	**.95**	.91	.94	**.94**	.95	.95	.95	.95
D5	.60	.72	.80	.82	**.89**	.91	.91	**.94**	.81	.93	.88	.92
D6	.59	.71	.74	.76	**.89**	.93	.91	**.93**	.80	.92	.90	.93
$G = 40$												
D1	.94	.94	.95	.95	**.95**	.94	.95	**.95**	.95	.95	.95	.95
D2	.86	.90	.93	.93	**.93**	.96	.97	**.98**	.89	.92	.95	.96
D3	.82	.88	.91	.91	**.94**	.96	.98	**.98**	.89	.94	.94	.96
D4	.93	.93	.94	.94	**.94**	.93	.94	**.94**	.95	.95	.95	.95
D5	.71	.79	.88	.88	**.89**	.84	.93	**.95**	.90	.94	.94	.93
D6	.65	.75	.79	.79	**.88**	.84	.93	**.95**	.87	.93	.92	.93

Model with Noninvertibility

- $\mathbf{Y}_g = \alpha + \mathbf{X}_g\beta + \mathbf{D}_g\gamma + \mathbf{e}_g$
- $\mathbf{D}_g$ is cluster-level dummy variable
- Confidence intervals for γ

	Conventional t_{G-1}						Adjusted		Boot		Wild	
s.e,	$\widehat{v}_1$	$\widehat{v}_2$	$\widehat{v}_3$	$\widehat{v}_4$	$\widehat{v}_5$	$\widehat{v}_6$	$\widehat{v}_2$	$\widehat{v}_5$	$\widehat{v}_1$	$\widehat{v}_5$	$\widehat{v}_1$	$\widehat{v}_5$
$G=6$												
D1	.64	.66	.68	.68	**1.0**	1.0	.73	**1.0**	1.0	1.0	.98	.99
D2	.64	.66	.66	.67	**1.0**	1.0	.72	**1.0**	1.0	1.0	.99	.99
D3	.60	.64	.68	.69	**1.0**	1.0	.76	**1.0**	1.0	1.0	.94	.95
D4	.71	.75	.76	.77	**1.0**	1.0	.82	**1.0**	1.0	1.0	.95	.96
D5	.66	.70	.72	.73	**1.0**	1.0	.75	**1.0**	1.0	1.0	.99	.99
D6	.71	.77	.79	.80	**1.0**	1.0	.86	**1.0**	1.0	1.0	.96	.96
$G=40$												
D1	.28	.28	.28	.28	**1.0**	.77	.29	**1.0**	1.0	1.0	1.0	1.0
D2	.26	.26	.26	.26	**1.0**	.75	.27	**1.0**	1.0	1.0	1.0	1.0
D3	.33	.38	.43	.43	**1.0**	.74	.43	**1.0**	1.0	1.0	.85	.88
D4	.44	.44	.45	.45	**1.0**	.87	.45	**1.0**	1.0	1.0	1.0	1.0
D5	.35	.38	.43	.43	**1.0**	.79	.39	**1.0**	1.0	1.0	1.0	1.0
D6	.50	.58	.60	.61	**1.0**	.85	.64	**1.0**	1.0	1.0	.92	.93

Summary of Simulations

- For any inference method (conventional, adjusted, pairs bootstrap, wild bootstrap)
  - Coverage is best with jackknife $\widehat{v}_5$
- Conventional jackknife interval with t_{G-1} criticals works reasonably well.
- Conventional standard errors can work poorly.
  - Especially under noninvertibility
- Pairs/wild bootstrap with $\widehat{v}_5$ works well
  - But not better than adjusted intervals
  - Adjusted intervals computationally much simpler!

Summary Recommendation

- In linear regression use jackknife standard errors.
  - Do not discard noninvertible clusters, rather use pseudoinverse
- Use adjusted student t critical values
  - Developing Stata and R code.
- Be wary when regressors are leveraged
  - Saturated dummy variable models
  - log-normal regressors