Jackknife Standard Errors for Clustered Regression: Online Appendix

Bruce E. Hansen
University of Wisconsin

This version: June 2025

Abstract

This appendix provides extra material on two topics. First, we provide an asymptotic theory, which shows that statistics constructed with the cluster jackknife variance estimator are asymptotic normal. Second, we provide details concerning our implementation of the wild bootstrap in the numerical simulation.

1 Asymptotic Theory

The model is the linear regression with clustered errors

$$\boldsymbol{Y}_g = \boldsymbol{X}_g \beta + \boldsymbol{e}_g. \tag{1}$$

The finite sample theory in the main text treated the regressors as fixed. For our asymptotic theory we instead assume that $(\boldsymbol{Y}_g, \boldsymbol{X}_g)$ are jointly random. Define the unconditional covariance matrix components

$$\begin{aligned}
\boldsymbol{Q}_n &= \frac{1}{n} \sum_{g=1}^{G} \mathbb{E}\left[\boldsymbol{X}_g' \boldsymbol{X}_g\right] \\
\boldsymbol{\Omega}_n &= \frac{1}{n} \sum_{g=1}^{G} \mathbb{E}\left[\boldsymbol{X}_g' \boldsymbol{e}_g \boldsymbol{e}_g' \boldsymbol{X}_g\right] \\
\boldsymbol{V}_n &= \boldsymbol{Q}_n^{-1} \boldsymbol{\Omega}_n \boldsymbol{Q}_n^{-1}.
\end{aligned}$$

An important feature of cluster asymptotic theory is that we allow the possibility of non-standard rates of convergence, which can arise due to within-cluster dependence and non-homogeneous cluster sizes. Thus, we allow the possibility that the covariance matrices $\boldsymbol{\Omega}_n$ and $\boldsymbol{V}_n$, or some sub-components of these matrices, increase with n, rather than converge to constant matrices as they do under non-clustered i.i.d. sampling.

We use the following conditions, which correspond to Theorem 9 of Hansen and Lee (1999), which estalished results for test statistics constructed with the CRVE_1 variance estimator. Let (Y_{ig}, X_{ig}) denote a single observation. Let $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ denote the smallest and largest eigenvalue of a Hermitian matrix $\boldsymbol{A}$, let $\|\boldsymbol{a}\| = \left(\boldsymbol{a}'\boldsymbol{a}\right)^{1/2}$ denote the Euclidean norm for a vector $\boldsymbol{a}$, and let $\|\boldsymbol{A}\| = \left(\lambda_{\max}\left(\boldsymbol{A}'\boldsymbol{A}\right)\right)^{1/2}$ denote the spectral norm of a matrix $\boldsymbol{A}$.

Assumption 1 *The clusters $(\boldsymbol{Y}_g, \boldsymbol{X}_g)$ are mutually independent across g. For some $2 \leq r < s < \infty$, $C < \infty$, and $\delta > 0$,*

1. $\mathbb{E}\left[\boldsymbol{X}_g'\boldsymbol{e}_g\right] = 0$
2. $\max\limits_{g \leq G} \dfrac{n_g^2}{n} \longrightarrow 0$
3. $n^{-1}\left(\sum_{g=1}^{G} n_g^r\right)^{2/r} \leq C$
4. $\mathbb{E}\left|Y_{ig}\right|^{2s} \leq C$
5. $\mathbb{E}\left\|X_{ig}\right\|^{2s} \leq C$
6. $\lambda_{\min}\left(\boldsymbol{Q}_n\right) \geq \delta$
7. $\lambda_{\min}\left(\boldsymbol{\Omega}_n\right) \geq \delta$.

Assumption 1.1 states that the model is a linear projection. Assumption 1.2 and 1.3 regulate the cluster sizes n_g. The assumptions allow n_g to be heterogeneous and increase with n, but do not allow any individual cluster to dominate the full sample asymptotically. Assumption 1.2 specifies that the largest cluster size must increase at a slower rate than the square root of the total sample size. Assumption 1.3 is non-intuitive, but is an additional restriction on the allowable heterogeneity in the cluster sample sizes. The parameter r involves a trade-off with the moment conditions of Assumptions 1.4-1.5. Assumption 1.3 is less restrictive for large r, and more restrictive for small r. (At $r = 2$ it requires the cluster sizes n_g to be bounded. At $r = \infty$ it states $n^{-1}\max\limits_{g \leq G} n_g^2 \leq C$, which is implied by Assumption 1.3 so is redundant.) Assumptions 1.4-1.5 are moment bounds, where the number of required finite moments is $2s$ for some $s > r$. For bounded observations we can set $r = s = \infty$, eliminating the need for Assumption 1.3. The least restrictive moment condition sets $r = 2$, which requires just over four finite moments (as is conventional for regression asymptotic theory) but requires that the cluster sizes n_g are bounded. Assumptions 1.6-1.7 state that the covariance matrix components $\boldsymbol{Q}_n$ and $\boldsymbol{\Omega}_n$ are uniformly full rank.

We now establish that the linear functions of the least squares estimator $\widehat{\beta}$ are asymptotically normal when standardized by the jackknife covariance matrix estimator.

Theorem 1 *Take model (1) under Assumption 1. Let $\widehat{\beta}$ be the least squares estimator of β, and let $\widehat{\boldsymbol{V}}_5$ be the jackknife variance estimator (equation (10) of the text). For any sequence of $k \times q$ full rank matrices $\boldsymbol{R}_n$,*

$$\left(\boldsymbol{R}_n'\widehat{\boldsymbol{V}}_5\boldsymbol{R}_n\right)^{-1/2}\boldsymbol{R}_n'\left(\widehat{\beta} - \beta\right) \underset{d}{\longrightarrow} \mathrm{N}\left(0, \boldsymbol{I}_q\right) \tag{2}$$

as $n \to \infty$. For the case $q = 1$ this implies

$$\frac{\widehat{\theta} - \theta}{\widehat{v}_5} \xrightarrow[d]{} \mathrm{N}(0,1). \tag{3}$$

For the Satterthwaite adjustmnent coefficients a and K (equations (35) and (36) in the text) calculated with $\boldsymbol{\Sigma}_g^0 = \boldsymbol{I}_{n_g}$,

$$a \xrightarrow[p]{} 1 \tag{4}$$

$$K \xrightarrow[p]{} \infty. \tag{5}$$

For the Satterthwaite confidence interval and p-value (equastions (30) and (31) in the text)

$$\mathbb{P}\left[\theta \in \widetilde{C}_5\right] \longrightarrow 1 - \alpha \tag{6}$$

$$p \xrightarrow[d]{} U[0,1], \tag{7}$$

the latter under $\theta = \theta_0$.

Equation (3) shows that the jackknife t-ratios have standard asymptotic normal distributions under the same conditions as for CRVE t-ratios. Equation (2) is a multivariate generalization, showing that sets of coefficient estimates are asymptotically normal, and immediately implies that jackknife Wald statistics have standard asymptotic chi-square distributions.

Equations (4)-(7) describe the properties of the default Satterthwaite approximations under the same conditions. Equation (4) shows that the scale adjustment converges in probability to one, and (5) shows that the Satterthwaite degree-of-freedom K diverges to infinity, meaning that asymptotically the adjustment becomes negligible, and adjusted inference reduces to conventional inference. Equations (6)-(7) show that this implies that the Satterthwaite inference procedures produce correct inferences. Equation (6) shows that the recommended default Satterthwaite confidence interval has asymptotically correct coverage for any regression model satisfying Assumption 1. Similarly, equation (7) shows that the recommend default Satterthwaite p-values have asymptotically correct $U[0,1]$ null distributions.

2 Proof

We start with some preliminary results. Define $\widehat{\boldsymbol{Q}}_n = \frac{1}{n}\boldsymbol{X}'\boldsymbol{X}$ and recall the definition $\boldsymbol{M}_g = \boldsymbol{I}_{n_g} - \boldsymbol{X}_g\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_g'$. We first establish that

$$\left\|\widehat{\boldsymbol{Q}}_n^{-1}\right\| \le O_p(1) \tag{8}$$

$$\frac{1}{n}\max_{g\le G}\left\|\boldsymbol{X}_g\right\|^2 \le o_p(1) \tag{9}$$

$$\max_{g\le G}\left\|\boldsymbol{I}_{n_g} - \boldsymbol{M}_g\right\| = \max_{g\le G}\left\|\boldsymbol{X}_g\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_g'\right\| = o_p(1) \tag{10}$$

$$\max_{g\le G}\left\|\boldsymbol{M}_g^{-1} - \boldsymbol{I}_{n_g}\right\| = o_p(1). \tag{11}$$

By the Schwarz matrix inequality

$$\left\|\boldsymbol{Q}_n^{-1/2}\widehat{\boldsymbol{Q}}_n\boldsymbol{Q}_n^{-1/2} - \boldsymbol{I}_k\right\| \le \left\|\boldsymbol{Q}_n^{-1}\right\|\left\|\widehat{\boldsymbol{Q}}_n - \boldsymbol{Q}_n\right\| \le o_p(1).$$

The final inequality holds because Assumption 1.6 implies $\left\|\boldsymbol{Q}_n^{-1}\right\| \le \delta^{-1} < \infty$ and Theorem 1 of Hansen and Lee (2019) established $\left\|\widehat{\boldsymbol{Q}}_n - \boldsymbol{Q}_n\right\| = o_p(1)$. Equivalently, $\boldsymbol{Q}_n^{-1/2}\widehat{\boldsymbol{Q}}_n\boldsymbol{Q}_n^{-1/2} \underset{p}{\longrightarrow} \boldsymbol{I}_k$. By the continuous mapping theorem we deduce $\boldsymbol{Q}_n^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{Q}_n^{1/2} \underset{p}{\longrightarrow} \boldsymbol{I}_k$. Using the triangle inequality, the Schwarz matrix inequality, and $\left\|\boldsymbol{Q}_n^{-1}\right\| \le \delta^{-1}$,

$$\begin{aligned}
\left\|\widehat{\boldsymbol{Q}}_n^{-1}\right\| &= \left\|\widehat{\boldsymbol{Q}}_n^{-1} - \boldsymbol{Q}_n^{-1} + \boldsymbol{Q}_n^{-1}\right\| \\
&\le \left\|\widehat{\boldsymbol{Q}}_n^{-1} - \boldsymbol{Q}_n^{-1}\right\| + \left\|\boldsymbol{Q}_n^{-1}\right\| \\
&\le \left\|\boldsymbol{Q}_n^{-1}\right\|\left\|\boldsymbol{Q}_n^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{Q}_n^{1/2} - \boldsymbol{I}_k\right\| + \left\|\boldsymbol{Q}_n^{-1}\right\| \\
&\le O_p(1).
\end{aligned}$$

This is (8).

Since the spectral norm is less than the Frobenius norm and $n_g \le \sqrt{n}$ for n sufficiently large by Assumption 1.2,

$$\frac{1}{n}\left\|\boldsymbol{X}_g\right\|^2 \le \frac{1}{n}\sum_{i=1}^{n_g}\left\|X_{ig}\right\|^2 \le \frac{1}{n^{1/2}}\left(\frac{1}{n_g}\sum_{i=1}^{n_g}\left\|X_{ig}\right\|^2\right). \tag{12}$$

Assumption 1.5 implies that $\left\|X_{ig}\right\|^{2r}$ is uniformly integrable. Lemma 1 of Hansen and Lee (2019) shows that this implies that $\left(n_g^{-1}\sum_{i=1}^{n_g}\left\|X_{ig}\right\|^2\right)^r$ is uniformly integrable. Theorem 9.7 of Hansen (2022) shows that this implies $\max_{g\le G}\left(n_g^{-1}\sum_{i=1}^{n_g}\left\|X_{ig}\right\|^2\right) = o_p\left(G^{1/r}\right)$. We find that (12) is uniformly bounded by $o_p\left(n^{-1/2}G^{1/r}\right) \le o_p(1)$, since $G \le n$ and $r \ge 2$. This establishes (9).

Using the Schwarz matrix inequality, (8), and (9),

$$\max_{g\le G}\left\|\boldsymbol{I}_{n_g} - \boldsymbol{M}_g\right\| = \max_{g\le G}\left\|\boldsymbol{X}_g\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_g'\right\| \le \left\|\widehat{\boldsymbol{Q}}_n^{-1}\right\|\frac{1}{n}\max_{g\le G}\left\|\boldsymbol{X}_g\right\|^2 \le o_p(1).$$

This is (10).

As $\boldsymbol{I}_{n_g} - \boldsymbol{M}_g$ is positive semi-definite,

$$\max_{g\leq G} \left\| \boldsymbol{I}_{n_g} - \boldsymbol{M}_g \right\| = \max_{g\leq G} \lambda_{\max}\left(\boldsymbol{I}_{n_g} - \boldsymbol{M}_g\right) = 1 - \min_{g\leq G} \lambda_{\min}(\boldsymbol{M}_g). \tag{13}$$

Thus (10) implies that $\min_{g\leq G} \lambda_{\min}(\boldsymbol{M}_g) \underset{p}{\longrightarrow} 1$. One implication is that the matrices $\boldsymbol{M}_g$ are asymptotically invertible, which means that the regresssion is clusterwise invertible. For the remainder of the proof we assume that the sample size is sufficiently large so that this holds.

Equation (13) also implies that

$$\begin{aligned}
\max_{g\leq G} \left\| \boldsymbol{M}_g^{-1} - \boldsymbol{I}_{n_g} \right\| &= \max_{g\leq G} \lambda_{\max}(\boldsymbol{M}_g^{-1} - \boldsymbol{I}_{n_g}) \\
&= \max_{g\leq G} \lambda_{\max}(\boldsymbol{M}_g^{-1}) - 1 \\
&= \frac{1}{\min_{g\leq G} \lambda_{\min}(\boldsymbol{M}_g)} - 1 \\
&\underset{p}{\longrightarrow} 0
\end{aligned}$$

This is (11).

Hansen and Lee (2019, Theorem 9) proved (2) for the CRVE variance estimator under Assumption 1. We establish (2) for the jackknife variance estimator by showing that the replacement of the variance estimators is asymptotically negligible.

It will be useful to define the central component of the CRVE estimator

$$\widehat{\boldsymbol{\Omega}}_n = \frac{1}{n} \sum_{g=1}^{G} \boldsymbol{X}_g' \widehat{\boldsymbol{e}}_g \widehat{\boldsymbol{e}}_g' \boldsymbol{X}_g.$$

We define the analog for the jackknife estimator:

$$\widetilde{\boldsymbol{\Omega}}_n = \frac{1}{n} \sum_{g=1}^{G} \boldsymbol{X}_g' \widehat{\boldsymbol{e}}_{-g} \widehat{\boldsymbol{e}}_{-g}' \boldsymbol{X}_g = \frac{1}{n} \sum_{g=1}^{G} \boldsymbol{X}_g' \boldsymbol{M}_g^{-1} \widehat{\boldsymbol{e}}_g \widehat{\boldsymbol{e}}_g' \boldsymbol{M}_g^{-1} \boldsymbol{X}_g.$$

The second equality holds because MacKinnon, Nielsen, and Webb (2023) established that under clusterwise invertibility, $\widehat{\boldsymbol{e}}_{-g} = \boldsymbol{M}_g^{-1} \widehat{\boldsymbol{e}}_g$.

Examining the proof of Theorem 9 of Hansen and Lee (2019), the key is equation (89) in their supplemental appendix:

$$\left\| \boldsymbol{\Omega}_n^{-1/2} \left(\widehat{\boldsymbol{\Omega}}_n - \boldsymbol{I}_k \right) \boldsymbol{\Omega}_n^{-1/2} \right\| \underset{p}{\longrightarrow} 0. \tag{14}$$

The analog needed for the jackknife estimator is to demonstrate that

$$\left\| \boldsymbol{\Omega}_n^{-1/2} \left(\widetilde{\boldsymbol{\Omega}}_n - \boldsymbol{I}_k \right) \boldsymbol{\Omega}_n^{-1/2} \right\| \underset{p}{\longrightarrow} 0. \tag{15}$$

Our goal is therefore to demonstrate (15), which is sufficient to establish (2).

Using the triangle inequality

$$
\begin{aligned}
\left\|\boldsymbol{\Omega}_n^{-1/2}\left(\widetilde{\boldsymbol{\Omega}}_n-\boldsymbol{I}_k\right)\boldsymbol{\Omega}_n^{-1/2}\right\| &\leq \left\|\boldsymbol{\Omega}_n^{-1/2}\left(\widetilde{\boldsymbol{\Omega}}_n-\widehat{\boldsymbol{\Omega}}_n\right)\boldsymbol{\Omega}_n^{-1/2}\right\|+\left\|\boldsymbol{\Omega}_n^{-1/2}\left(\widehat{\boldsymbol{\Omega}}_n-\boldsymbol{I}_k\right)\boldsymbol{\Omega}_n^{-1/2}\right\| \\
&= \left\|\boldsymbol{\Omega}_n^{-1/2}\left(\widetilde{\boldsymbol{\Omega}}_n-\widehat{\boldsymbol{\Omega}}_n\right)\boldsymbol{\Omega}_n^{-1/2}\right\|+o_p(1)
\end{aligned}
$$

where the final equality is (14). It is therefore sufficient to show that

$$
\left\|\boldsymbol{\Omega}_n^{-1/2}\left(\widetilde{\boldsymbol{\Omega}}_n-\widehat{\boldsymbol{\Omega}}_n\right)\boldsymbol{\Omega}_n^{-1/2}\right\| \underset{p}{\longrightarrow} 0. \tag{16}
$$

Define $\boldsymbol{P}_g=\boldsymbol{M}_g^{-1}-\boldsymbol{I}_{n_g}$, which satisfies $\max_{g\leq G}\left\|\boldsymbol{P}_g\right\|=o_p(1)$ by (11). Using the Triangle and Schwarz matrix inequalities,

$$
\begin{aligned}
\left\|\boldsymbol{\Omega}_n^{-1/2}\left(\widetilde{\boldsymbol{\Omega}}_n-\widehat{\boldsymbol{\Omega}}_n\right)\boldsymbol{\Omega}_n^{-1/2}\right\| &= \left\|\boldsymbol{\Omega}_n^{-1/2}\left(\frac{1}{n}\sum_{g=1}^{G}\boldsymbol{X}_g'\boldsymbol{P}_g\widehat{\boldsymbol{e}}_g\widehat{\boldsymbol{e}}_g'\boldsymbol{P}_g\boldsymbol{X}_g+\frac{1}{n}\sum_{g=1}^{G}\boldsymbol{X}_g'\boldsymbol{P}_g\widehat{\boldsymbol{e}}_g\widehat{\boldsymbol{e}}_g'\boldsymbol{X}_g+\frac{1}{n}\sum_{g=1}^{G}\boldsymbol{X}_g'\widehat{\boldsymbol{e}}_g\widehat{\boldsymbol{e}}_g'\boldsymbol{P}_g\boldsymbol{X}_g\right)\boldsymbol{\Omega}_n^{-1/2}\right\| \\
&\leq \left\|\frac{1}{n}\sum_{g=1}^{G}\boldsymbol{\Omega}_n^{-1/2}\boldsymbol{X}_g'\boldsymbol{P}_g\widehat{\boldsymbol{e}}_g\widehat{\boldsymbol{e}}_g'\boldsymbol{P}_g\boldsymbol{X}_g\boldsymbol{\Omega}_n^{-1/2}\right\|+2\left\|\frac{1}{n}\sum_{g=1}^{G}\boldsymbol{\Omega}_n^{-1/2}\boldsymbol{X}_g'\boldsymbol{P}_g\widehat{\boldsymbol{e}}_g\widehat{\boldsymbol{e}}_g'\boldsymbol{X}_g\boldsymbol{\Omega}_n^{-1/2}\right\| \\
&\leq \frac{1}{n}\sum_{g=1}^{G}\left\|\boldsymbol{\Omega}_n^{-1/2}\boldsymbol{X}_g'\widehat{\boldsymbol{e}}_g\right\|^2 o_p(1) \\
&\leq o_p(1).
\end{aligned}
$$

The fact $n^{-1}\sum_{g=1}^{G}\left\|\boldsymbol{\Omega}_n^{-1/2}\boldsymbol{X}_g'\widehat{\boldsymbol{e}}_g\right\|^2=O_p(1)$ follows implicitly from Hansen and Lee's proof of (14). This is (16), which completes the proof (2).

Theorem 7 in the main text established (4) and (5) under (8)-(9).

We next establish (6). Given the definition of $\widetilde{C}_5$ and results (3), (4), and (5),

$$
\begin{aligned}
\mathbb{P}\left[\theta\in\widetilde{C}_5\right] &= \mathbb{P}\left[\left|\frac{\widehat{\theta}-\theta}{\widehat{v}_5}\right|\leq\frac{t_K^{1-\alpha/2}}{a}\right] \\
&\to \mathbb{P}\left[|\mathrm{N}(0,1)|\leq t_\infty^{1-\alpha/2}\right] \\
&= 1-\alpha
\end{aligned}
$$

which is (6).

Similarly, given the definition of the p-value p,

$$
p=2\left(1-G\left(a\left|\frac{\widehat{\theta}-\theta}{\widehat{v}_5}\right|,K\right)\right)\to 2\left(1-\Phi\left(|\mathrm{N}(0,1)|\right)\right)\sim U[0,1]
$$

which is (7).

This completes the proof. ■

3 Wild Bootstrap

We describe here the details of our implementation of the wild bootstrap with jackknife standard errors, as used in the numerical simulation. Our implementation is modeled on MacKinnon (2023) who describes a fast wild bootstrap implementation with CRVE_1 standard errors.

It is useful to describe the algorithm first for a fixed value of the parameter θ, and then discuss how this is used to (separately) calculate p-values and confidence intervals. Define $v_0^2 = R'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R$ and $\boldsymbol{Z}_g = \boldsymbol{X}_g\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R$. Let $\widetilde{\beta} = \widehat{\beta} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R\widehat{\theta}/v_0^2$ denote the constrained least squares estimator subject to $\widetilde{\theta} = R'\widetilde{\beta} = 0$. Let $\widetilde{\boldsymbol{e}}_g = \boldsymbol{Y}_g - \boldsymbol{X}_g\widetilde{\beta} = \widehat{\boldsymbol{e}}_g + \boldsymbol{Z}_g\widehat{\theta}/v_0^2$ denote the associated cluster-level residual. For given θ, let $\widetilde{\beta}(\theta) = \widetilde{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R\theta/v_0^2$ denote the constrained least squares estimator subject to $\widetilde{\theta}(\theta) = R'\widetilde{\beta} = \theta$. Let $\widetilde{\boldsymbol{e}}_g(\theta) = \boldsymbol{Y}_g - \boldsymbol{X}_g\widetilde{\beta}(\theta) = \widetilde{\boldsymbol{e}}_g - \boldsymbol{Z}_g\theta/v_0^2$ denote the associated cluster-level residual.

For each bootstrap draw we simulate a $G\times 1$ vector $\boldsymbol{\phi} \sim \mathrm{N}(0, \boldsymbol{I}_G)$ with gth element ϕ_g. The bootstrap dependent variable equals $\boldsymbol{Y}_g^* = \widetilde{\boldsymbol{e}}_g(\theta)\phi_g$. The bootstrap version of $\widehat{\theta}$ is

$$\widehat{\theta}^* = R'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y}^*\right) = \sum_{g=1}^{G}\boldsymbol{Z}_g'\widetilde{\boldsymbol{e}}_g(\theta)\phi_g = \boldsymbol{a}_0'\boldsymbol{\phi} - \boldsymbol{a}_1'\boldsymbol{\phi}\theta$$

where $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$ are $G\times 1$ with gth elements $a_{0g} = \boldsymbol{Z}_g'\widetilde{\boldsymbol{e}}_g$ and $a_{1g} = \boldsymbol{Z}_g'\boldsymbol{Z}_g/v_0^2$, respectively. The bootstrap version of $\widehat{\theta}_{-g} - \widehat{\theta}$ is

$$\begin{aligned}
\widehat{\theta}_{-g}^* - \widehat{\theta}^* &= R'\left(\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}_g'\boldsymbol{X}_g\right)^{-}\left(\sum_{h\neq g}\boldsymbol{X}_h'\boldsymbol{Y}_h^*\right) - \widehat{\theta}^* \\
&= R'\left(\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}_g'\boldsymbol{X}_g\right)^{-}\left(\sum_{h\neq g}\boldsymbol{X}_h'\widetilde{\boldsymbol{e}}_h(\theta)\phi_h\right) - \left(\boldsymbol{a}_0'\boldsymbol{\phi} - \boldsymbol{a}_1'\boldsymbol{\phi}\theta\right) \\
&= \left(\boldsymbol{d}_{0g} - \boldsymbol{a}_0\right)'\boldsymbol{\phi} - \left(\boldsymbol{d}_{1g} - \boldsymbol{a}_1\right)'\boldsymbol{\phi}\theta
\end{aligned}$$

where $\boldsymbol{d}_{0g}$ and $\boldsymbol{d}_{1g}$ are $G\times 1$. Their gth elements are 0, and for $h \neq g$ their hth elements are $R'\left(\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}_g'\boldsymbol{X}_g\right)^{-}\boldsymbol{X}_h'\widetilde{\boldsymbol{e}}_h$ and $R'\left(\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{X}_g'\boldsymbol{X}_g\right)^{-}\boldsymbol{X}_h'\boldsymbol{Z}_h/v_0^2$, respectively. Stacking, we obtain the $G\times 1$ vector $\widehat{\theta}_-^* - \widehat{\theta}^* = \boldsymbol{C}_0'\boldsymbol{\phi} - \boldsymbol{C}_1'\boldsymbol{\phi}\theta$ where $\boldsymbol{C}_0$ and $\boldsymbol{C}_1$ are $G\times G$ with gth column $\boldsymbol{d}_{0g} - \boldsymbol{a}_0$ and $\boldsymbol{d}_{1g} - \boldsymbol{a}_1$, respectively.

The bootstrap version of $\widehat{v}_5^2$ is

$$\widehat{v}_5^{2*} = \sum_{g=1}^{G}\left(\widehat{\theta}_{-g}^* - \widehat{\theta}^*\right)^2 = \boldsymbol{\phi}'\boldsymbol{C}_0\boldsymbol{C}_0'\boldsymbol{\phi} - 2\boldsymbol{\phi}'\boldsymbol{C}_0\boldsymbol{C}_1'\boldsymbol{\phi}\theta + \boldsymbol{\phi}'\boldsymbol{C}_1\boldsymbol{C}_1'\boldsymbol{\phi}\theta^2.$$

The bootstrap version of the squared t-statistic $\left(\widehat{\theta} - \theta\right)^2/\widehat{v}_5^2$ is

$$T^* = \frac{\widehat{\theta}^{*2}}{\widehat{v}_5^{2*}} = \frac{\left(\boldsymbol{a}_0'\boldsymbol{\phi} - \boldsymbol{a}_1'\boldsymbol{\phi}\theta\right)^2}{\boldsymbol{\phi}'\boldsymbol{C}_0\boldsymbol{C}_0'\boldsymbol{\phi} - 2\boldsymbol{\phi}'\boldsymbol{C}_0\boldsymbol{C}_1'\boldsymbol{\phi}\theta + \boldsymbol{\phi}'\boldsymbol{C}_1\boldsymbol{C}_1'\boldsymbol{\phi}\theta^2}.$$

The bootstrap repeats this calculation for B independent draws of the random vector $\boldsymbol{\phi}$. As described by MacKinnon (2023), it is computationally efficient to calculate the vectors and matrices $\boldsymbol{a}_0$, $\boldsymbol{a}_1$, $\boldsymbol{C}_0$, and $\boldsymbol{C}_1$ before making the draws $\boldsymbol{\phi}$ and calculating the bootstrap statistics T^*. This way, calculation of T^* only involves a small number of basic matrix operations. It is also useful to observe that given the statistics $\boldsymbol{a}_0'\boldsymbol{\phi}$, $\boldsymbol{a}_1'\boldsymbol{\phi}$, $\boldsymbol{\phi}'\boldsymbol{C}_0\boldsymbol{C}_0'\boldsymbol{\phi}$, $\boldsymbol{\phi}'\boldsymbol{C}_0\boldsymbol{C}_1'\boldsymbol{\phi}$, and $\boldsymbol{\phi}'\boldsymbol{C}_1\boldsymbol{C}_1'\boldsymbol{\phi}$ the bootstrap statistic T^* is a simple function of θ (a ratio of quadratics). This is a useful insight for confidence interval construction, where T^* will need to be iteratively re-calculated for many values of θ.

For any θ, the bootstrap $1-\alpha$ critical value $c(\theta)$ is the $1-\alpha$ quantile of the empirical distribution of the bootstrap statistics T^* across the bootstrap draws $\boldsymbol{\phi}$.

A hypothesis $\theta = \theta_0$ is accepted if $\left(\widehat{\theta}-\theta_0\right)^2/\widehat{v}_5^2 \le c(\theta_0)$ and rejected otherwise. This is how we calculate the coverage probabilities in the simulation.

The $1-\alpha$ level wild bootstrap confidence interval for θ is the set of values which are accepted by the bootstrap test; equivalently, the set of θ which satisfy $\left(\widehat{\theta}-\theta\right)^2/\widehat{v}_5^2 \le c(\theta)$. Let θ_j be the set of solutions to $\left(\widehat{\theta}-\theta\right)^2/\widehat{v}_5^2 = c(\theta)$. There are at least two solutions, satisfying $\theta_1 \le \widehat{\theta} \le \theta_2$, but is possible that there are more. We define the confidence interval as $[\theta_L, \theta_U]$ with $\theta_L = \min \theta_j$ and $\theta_U = \max_j \theta$.

For the confidence interval length results presented in the simulation we calculate the confidence interval endpoints as follows. We take the endpoints of the Satterthwaite intervals as initial values, thus $\theta_L = \widehat{\theta} - t_K^{1-\alpha/2}\widehat{v}_5/a$ and $\theta_U = \widehat{\theta} + t_K^{1-\alpha/2}\widehat{v}_5/a$, then search for solutions to $\left(\widehat{\theta}-\theta\right)^2/\widehat{v}_5^2 = c(\theta)$ using a local grid search to find values which bracket the solution, and then apply the bisection algorithm. This locates the unique endpoint solution when it exists, and otherwise locates the solution closest to the Satterthwaite interval.

References

[1] Hansen, Bruce E. (2022): *Econometrics*, Princeton University Press.

[2] Hansen, Bruce E. and Seojeong Lee (2019): "Asymptotic theory for clustered samples," *Journal of Econometrics*, 210, 268-290.

[3] MacKinnon, James G. (2023): "Fast cluster bootstrap methods for linear regression models," *Econometrics and Statistics*, 26, 52-71.

[4] MacKinnon, James G., Morten Orregaard Nielsen, and Matthew D. Webb (2023): "Fast and reliable jackknife and bootstrap methods for cluster-robust inference," *Journal of Applied Econometrics*, 38, 671-694.