# Supplemental Appendix for: The Wild Bootstrap with a "Small" Number of "Large" Clusters

Ivan A. Canay
Department of Economics
Northwestern University
iacanay@northwestern.edu

Andres Santos
Department of Economics
U.C.L.A.
andres@econ.ucla.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

November 5, 2019

**Abstract**

This document provides additional results for the authors' paper "The Wild Bootstrap with a "Small" Number of "Large" Clusters". It includes the proofs of auxiliary lemmas, additional details for Remark 2.3, and a generalization of the main results to non-linear models and non-linear hypotheses.

KEYWORDS: Wild bootstrap, Clustered Data, Randomization Tests.
JEL CLASSIFICATION CODES: C12, C15, C23.

# S.1 Auxiliary Lemmas

**Lemma S.1.1.** *Let Assumptions 2.1 and 2.2 hold, $\hat{\Omega}^{-}_{\tilde{Z},n}$ denote the pseudo-inverse of $\hat{\Omega}_{\tilde{Z},n}$, and set $\bar{a} \equiv \sum_{j\in J}\xi_j a_j$ and $U_{n,j} \equiv \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\epsilon_{i,j}$. If $c'\beta = \lambda$, then for any $(g_1,\ldots,g_q) = g \in \mathbf{G}$*

$$\hat{\sigma}_n^2 = c'\hat{\Omega}^{-}_{\tilde{Z},n}\sum_{j\in J}\Big(U_{n,j} - \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}U_{n,\tilde{j}}\Big)\Big(U_{n,j} - \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}U_{n,\tilde{j}}\Big)'\hat{\Omega}^{-}_{\tilde{Z},n}c + o_P(1)$$

$$(\hat{\sigma}_n^*(g))^2 = c'\hat{\Omega}^{-}_{\tilde{Z},n}\sum_{j\in J}\Big(g_jU_{n,j} - \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}g_{\tilde{j}}U_{n,\tilde{j}}\Big)\Big(g_jU_{n,j} - \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}g_{\tilde{j}}U_{n,\tilde{j}}\Big)'\hat{\Omega}^{-}_{\tilde{Z},n}c + o_P(1)\ .$$

PROOF: Recall that $(\hat{\beta}_n', \hat{\gamma}_n')'$ denotes the least squares estimator of $(\beta', \gamma')'$ in (1) and denote the corresponding residuals by $\hat{\epsilon}_{i,j} \equiv (Y_{i,j} - Z_{i,j}'\hat{\beta}_n - W_{i,j}'\hat{\gamma}_n)$. Since $\sqrt{n}(\hat{\beta}_n - \beta)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma)$ are bounded in probability by Assumption 2.1, Lemma S.1.2 and the definition of $U_{n,j}$ yield

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j} &= \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\epsilon_{i,j} - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}Z_{i,j}'\sqrt{n}(\hat{\beta}_n - \beta) - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W_{i,j}'\sqrt{n}(\hat{\gamma}_n - \gamma)\\ &= U_{n,j} - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}_{i,j}'\sqrt{n}(\hat{\beta}_n - \beta) + o_P(1)\ . \end{aligned} \tag{S.1}$$

Next, note that $\hat{\Omega}_{\tilde{Z},n}$ is invertible with probability tending to one by Assumption 2.2(iii). Since $\hat{\Omega}^{-}_{\tilde{Z},n} = \hat{\Omega}^{-1}_{\tilde{Z},n}$ when $\hat{\Omega}_{\tilde{Z},n}$ is invertible, we obtain from Assumptions 2.2(ii)-(iii) that

$$\begin{aligned}&\frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}_{i,j}'\sqrt{n}(\hat{\beta}_n - \beta)\\ &\quad= \frac{n_j}{n}\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}_{i,j}'\hat{\Omega}^{-}_{\tilde{Z},n}\frac{1}{\sqrt{n}}\sum_{\tilde{j}\in J}\sum_{k\in I_{n,\tilde{j}}}\tilde{Z}_{k,\tilde{j}}\epsilon_{k,\tilde{j}} + o_P(1) = \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}U_{n,\tilde{j}} + o_P(1)\ . \end{aligned}\tag{S.2}$$

Therefore, (S.1), (S.2), and the continuous mapping theorem yield

$$\begin{aligned}\hat{V}_n &= \sum_{j\in J}\Big(\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}\Big)\Big(\frac{1}{\sqrt{n}}\sum_{k\in I_{n,j}}\tilde{Z}_{k,j}'\hat{\epsilon}_{k,j}\Big)\\ &= \sum_{j\in J}\Big(U_{n,j} - \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}U_{n,\tilde{j}}\Big)\Big(U_{n,j} - \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}U_{n,\tilde{j}}\Big)' + o_P(1)\ . \end{aligned}\tag{S.3}$$

The first part of the lemma thus follows by the definition of $\hat{\sigma}_n^2$ in (15).

For the second claim of the lemma, note that when $c'\beta = \lambda$, it follows from Assumption 2.1 and Amemiya (1985, Eq. (1.4.5)) that $\sqrt{n}(\hat{\beta}_n^{\mathrm{r}} - \beta)$ and $\sqrt{n}(\hat{\gamma}_n^{\mathrm{r}} - \gamma)$ are bounded in probability. Together with Assumption 2.1 such result in turn also implies that $\sqrt{n}(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathrm{r}})$ and $\sqrt{n}(\hat{\gamma}_n^*(g) - \hat{\gamma}_n^{\mathrm{r}})$ are bounded in probability for all $g \in \mathbf{G}$. Next, recall that the residuals from the bootstrap regression in (4) equal $\hat{\epsilon}^*_{i,j}(g) = g_j\hat{\epsilon}^{\mathrm{r}}_{i,j} - Z_{i,j}'(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathrm{r}}) - W_{i,j}'(\hat{\gamma}_n^*(g) - \hat{\gamma}_n^{\mathrm{r}})$ for all $(g_1,\ldots,g_q) = g \in \mathbf{G}$.

Therefore, we are able to conclude for any $g \in \mathbf{G}$ and $j \in J$ that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \hat{\epsilon}^*_{i,j}(g) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} g_j \hat{\epsilon}^{\mathrm{r}}_{i,j} - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} Z'_{i,j} \sqrt{n}(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n) - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} W'_{i,j} \sqrt{n}(\hat{\gamma}^*_n(g) - \hat{\gamma}^{\mathrm{r}}_n) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} g_j \hat{\epsilon}^{\mathrm{r}}_{i,j} - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \sqrt{n}(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n) + o_P(1) \ , \qquad \text{(S.4)}
\end{aligned}
$$

where in the final equality we employed Lemma S.1.2. Next, recall $\hat{\epsilon}^{\mathrm{r}}_{i,j} \equiv \epsilon_{i,j} - Z'_{i,j}(\hat{\beta}^{\mathrm{r}}_n - \beta) - W'_{i,j}(\hat{\gamma}^{\mathrm{r}}_n - \gamma)$ and note

$$
\begin{aligned}
c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} g_j \hat{\epsilon}^{\mathrm{r}}_{i,j} &= c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} g_j (\epsilon_{i,j} - Z'_{i,j}\sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta) - W'_{i,j}\sqrt{n}(\hat{\gamma}^{\mathrm{r}}_n - \gamma)) \\
&= c'\hat{\Omega}^-_{\tilde{Z},n} g_j U_{n,j} - c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{n} \sum_{i \in I_{n,j}} g_j \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta) + o_P(1) \ , \qquad \text{(S.5)}
\end{aligned}
$$

where the second equality follows from Lemma S.1.2 and $\hat{\Omega}^-_{\tilde{Z},n}$, $\sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta)$, and $\sqrt{n}(\hat{\gamma}^r_n - \gamma)$ being bounded in probability. Moreover, Assumptions 2.2(ii)-(iii) imply

$$
c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{n} \sum_{i \in I_{n,j}} g_j \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta) = c'\Omega^{-1}_{\tilde{Z}} \frac{g_j \xi_j a_j}{\bar{a}} \Omega_{\tilde{Z}} \sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta) + o_P(1) = o_P(1) \ , \qquad \text{(S.6)}
$$

where the final result follows from $c'\hat{\beta}^{\mathrm{r}} = \lambda$ by construction and $c'\beta = \lambda$ by hypothesis. Next, we note that since $\hat{\Omega}^-_{\tilde{Z},n} = \hat{\Omega}^{-1}_{\tilde{Z},n}$ whenever $\hat{\Omega}_{\tilde{Z},n}$ is invertible, and $\hat{\Omega}_{\tilde{Z},n}$ is invertible with probability tending to one by Assumption 2.2(iii), we can conclude that

$$
\begin{aligned}
c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \sqrt{n}(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n) &= c'\hat{\Omega}^-_{\tilde{Z},n} \frac{n_j}{n} \frac{1}{n_j} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \hat{\Omega}^-_{\tilde{Z},n} \sum_{\tilde{j} \in J} \frac{1}{\sqrt{n}} \sum_{k \in I_{n,\tilde{j}}} \tilde{Z}_{k,j} g_{\tilde{j}} \hat{\epsilon}^{\mathrm{r}}_{k,\tilde{j}} + o_P(1) \\
&= c'\hat{\Omega}^-_{\tilde{Z},n} \frac{\xi_j a_j}{\bar{a}} \sum_{\tilde{j} \in J} g_{\tilde{j}} U_{n,\tilde{j}} + o_P(1) \ , \qquad \text{(S.7)}
\end{aligned}
$$

where in the final equality we applied (S.5), (S.6), and $\bar{a} \equiv \sum_{j \in J} \xi_j a_j$. The second part of the lemma then follows from the definition of $(\hat{\sigma}^*_n(g))^2$ in (16) and results (S.4)-(S.7). ■

**Lemma S.1.2.** *Let Assumptions 2.1(ii) and 2.2(iv) hold. It follows that for any $j \in J$ we have*

$$
\frac{1}{n_j} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} W'_{i,j} = o_P(1) \quad \text{and} \quad \frac{1}{n_j} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} Z'_{i,j} = \frac{1}{n_j} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} + o_P(1) \ .
$$

PROOF: Let $\| \cdot \|_F$ denote the Frobenius matrix norm, which recall equals $\|M\|^2_F \equiv \text{trace}\{M'M\}$ for any matrix $M$. By the definition of $\tilde{Z}_{i,j}$ in (8), $\sum_{i \in I_{n,j}} (Z_{i,j} - (\hat{\Pi}^{\mathrm{c}}_{n,j})' W_{i,j}) W'_{i,j} = 0$ by definition of $\hat{\Pi}^{c}_{n,j}$ (see (9)), and the triangle inequality applied to $\|\cdot\|_F$, we then obtain

$$\|\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\|_F = \|\frac{1}{n_j}\sum_{i\in I_{n,j}}(Z_{i,j}-\hat{\Pi}'_n W_{i,j})W'_{i,j}\|_F$$
$$= \|\frac{1}{n_j}\sum_{i\in I_{n,j}}(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F \le \frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F~. \quad \text{(S.8)}$$

Moreover, applying a second triangle inequality and the properties of the trace we get

$$\frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F = \frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}\|\times\|W'_{i,j}W_{i,j}\|$$
$$\le \Big\{\frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}\|^2\Big\}^{1/2}\times\Big\{\frac{1}{n_j}\sum_{i\in I_{n,j}}\|W_{i,j}\|^2\Big\}^{1/2} = o_P(1)~, \quad \text{(S.9)}$$

where the inequality follows from the Cauchy-Schwarz inequality, and the final result by Assumption 2.1(ii) and 2.2(iv). Since $\hat{\Pi}_n$ is bounded in probability by Assumption 2.1(ii) and

$$\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}Z'_{i,j} = \frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j} + \frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\hat{\Pi}_n \quad \text{(S.10)}$$

by (8), the second part of the lemma follows. ■

## S.2 Further Details for Remark 2.3

Consider a differences-in-differences application in which, for simplicity, we assume there are only two time periods. Treatment is assigned in the second time period, and for each individual $i$ in group $j$ we let $Y_{i,j}$ denote an outcome of interest, $T_{i,j}\in\{1,2\}$ be the time period at which $Y_{i,j}$ was observed, and $Z_{i,j}\in\{0,1\}$ indicate treatment status. In the canonical differences-in-differences model (Angrist and Pischke, 2008), these variables are assumed to be related by

$$Y_{i,j} = I\{T_{i,j}=2\}\delta + \sum_{\tilde{j}\in J} I\{\tilde{j}=j\}\zeta_{\tilde{j}} + Z_{i,j}\beta + \epsilon_{i,j}~,$$

which we may accommodate in our framework by letting $W_{i,j}$ be cluster-level fixed effects and $I\{T_{i,j}=2\}$. Typically, the groups are such that treatment status is common among all $i\in I_{n,j}$ with $T_{i,j}=2$. This structure implies that $J$ can be partitioned into sets $J(0)$ and $J(1)$ such that $Z_{i,j}=I\{T_{i,j}=2, j\in J(1)\}$. In order to examine the content of Assumptions 2.2(iii)-(iv) in this setting, define

$$\tau \equiv \frac{\sum_{j\in J(1)} n_j(1)p_j}{\sum_{j\in J} n_j(1)p_j}~, \quad \text{(S.11)}$$

where $n_j(t) \equiv \sum_{i \in I_{n,j}} I\{T_{i,j} = t\}$ and $p_j \equiv n_j(2)/n_j$. By direct calculation, it is then possible to verify that $(\hat{\Pi}_n^c)' W_{i,j} = Z_{i,j}$, while

$$\hat{\Pi}_n' W_{i,j} = \begin{cases} -p_j \tau & \text{if } T_{i,j} = 1 \text{ and } j \in J(0) \\ (1-\tau) p_j \tau & \text{if } T_{i,j} = 1 \text{ and } j \in J(1) \\ (1-p_j)\tau & \text{if } T_{i,j} = 2 \text{ and } j \in J(0) \\ \tau + (1-\tau) p_j & \text{if } T_{i,j} = 2 \text{ and } j \in J(1) \end{cases} , \tag{S.12}$$

which implies Assumption 2.2(iv) is violated. On the other hand, these derivations also imply that it may be possible to satisfy Assumption 2.2(iii) by clustering more coarsely. In particular, if we instead group elements of $J$ into larger clusters $\{S_k : k \in K\}$ $(K < q)$ such that

$$\frac{\sum_{j \in J(1) \cap S_k} n_j(1) p_j}{\sum_{j \in S_k} n_j(1) p_j}$$

converges to $\tau$, then Assumption 2.2(iv) is satisfied. In this way, Assumption 2.2(iv) thereby requires the clusters to be "balanced" in the proportion of treated units.

## S.3 A General Result

In this section, we present a result that generalizes Theorem 3.3 and, as explained below, permits us to establish qualitatively similar results for nonlinear null hypotheses and nonlinear models. In what follows, there is no longer a need to distinguish between $Y_{i,j}$, $W_{i,j}$, and $Z_{i,j}$, so we denote by $X_{i,j} \in \mathbf{R}^{d_x}$ the observed data corresponding to the $i$th unit in the $j$th cluster. We consider tests that reject for large values of a test statistic $T_n^{\mathtt{F}}$, whose limiting behavior we will assume below is the same as the limiting behavior of $F(S_n)$, where $S_n$ is the cluster-level "scores" given by

$$S_n \equiv \left\{ \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \psi(X_{i,j}) : j \in J \right\}$$

and $F : \mathbf{R}^q \to \mathbf{R}$ is a known, continuous function. Here, $\psi : \mathbf{R}^{d_x} \to \mathbf{R}^{d_\psi}$ is an unknown function that may depend on the distribution of the data, so, in order to describe a critical value with which to compare $T_n^{\mathtt{F}}$, we assume that there are estimators $\hat{\psi}_n$ of $\psi$ and define

$$\hat{S}_n \equiv \left\{ \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \hat{\psi}_n(X_{i,j}) : j \in J \right\} .$$

Using this notation, the critical value we employ is obtained through the following construction:

**Step 1**: Let $\mathbf{G} = \{-1, 1\}^q$ and for any $g = (g_1, \ldots, g_q) \in \mathbf{G}$ define

$$g\hat{S}_n \equiv \left\{ \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} g_j \hat{\psi}_n(X_{i,j}) : j \in J \right\} .$$

**Step 2**: Compute the $1-\alpha$ quantile of $\{F(g\hat{S}_n)\}_{g\in\mathbf{G}}$, denoted by

$$\hat{c}_n^{\mathtt{F}}(1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{F(g\hat{S}_n) \le u\} \ge 1-\alpha\right\} .$$

Below we develop properties of the test $\phi_n^{\mathtt{F}}$ that rejects whenever $T_n^{\mathtt{F}}$ exceeds $\hat{c}_n^{\mathtt{F}}(1-\alpha)\}$, i.e.,

$$\phi_n^{\mathtt{F}} \equiv I\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\} .$$

In the context of the linear model studied in the main paper, under appropriate choices of $F$, $\psi$, and $\hat{\psi}_n$, the test $\phi_n^{\mathrm{F}}$ is in fact numerically equivalent to the test $\phi_n$ defined in (6). More generally, however, the test $\phi_n^{\mathrm{F}}$ can be interpreted as relying on the "score" bootstrap studied by Kline and Santos (2012). In particular, note that $\hat{c}_n^{\mathrm{F}}(1-\alpha)$ may alternatively be written as

$$\inf\left\{u \in \mathbf{R} : P\left\{F\left(\sum_{j\in J}\frac{\omega_j}{\sqrt{n}}\sum_{i\in I_{n,j}}\hat{\psi}_n(X_{i,j})\right) \le u|X^{(n)}\right\} \ge 1-\alpha\right\} \tag{S.13}$$

where $X^{(n)}$ denotes the data and $\{\omega_j\}_{j=1}^q$ are i.i.d. Rademacher random variables independent of $X^{(n)}$. Whenever $|\mathbf{G}|$ is large, one may therefore approximate $\hat{c}_n^{\mathrm{F}}(1-\alpha)$ by simulating (S.13).

Our analysis will require the following high-level assumption:

**Assumption S.3.1.** *The following statements hold:*

*(i) The test statistic $T_n^{\mathrm{F}}$ satisfies*

$$T_n^{\mathrm{F}} = F(S_n) + o_P(1) .$$

*(ii) The estimator $\hat{\psi}_n$ satisfies*

$$\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\hat{\psi}_n(X_{i,j}) = \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\psi(X_{i,j}) + o_P(1)$$

*for all $j \in J$.*

*(iii) There exists a collection of independent random variables $\{\mathcal{Z}_j\}_{j\in J}$, where $\mathcal{Z}_j \in \mathbf{R}^{d_\psi}$ and $\mathcal{Z}_j \sim N(0,\Sigma_j)$, such that*

$$\left\{\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\psi(X_{i,j}) : j \in J\right\} \xrightarrow{d} \{\mathcal{Z}_j : j \in J\} .$$

*(iv) For any $g \in \mathbf{G}$ and $\tilde{g} \in \mathbf{G}$,*

$$P\{F(\{g_j\mathcal{Z}_j : j \in J\}) = F(\{\tilde{g}_j\mathcal{Z}_j : j \in J\})\} \in \{0,1\} .$$

*(v) There is an integer $\kappa$ such that $|A(g)| = \kappa$ for any $g \in \mathbf{G}$, where*

$$A(g) \equiv \{\tilde{g} \in \mathbf{G} : P\{F(\{g_j\mathcal{Z}_j : j \in J\}) = F(\{\tilde{g}_j\mathcal{Z}_j : j \in J\})\} = 1\} .$$

Assumption S.3.1(i) formalizes the aforementioned requirement that the limiting behavior of $T_n^{\mathtt{F}}$ is the same as the limiting behavior of $F(S_n)$. Assumption S.3.1(ii) encodes homogeneity restrictions qualitatively similar to those in Assumption 2.2; see our discussion of nonlinear restrictions and GMM below. Assumption S.3.1(iii) essentially requires that the dependence within clusters be weak enough to permit application of a suitable central limit theorem to the cluster "scores." Finally, Assumptions S.3.1(iv)-(v) are typically satisfied with $\kappa = 2$ for two-sided tests and $\kappa = 1$ for one-sided tests. By allowing for other values of $\kappa$, however, we can also accommodate settings in which $n_j/n \to 0$ for some $j$ or $\Sigma_j$ in Assumption S.3.1(iii) is positive semi-definite.

We are now prepared to state our result about the properties of $\phi_n^{\mathtt{F}}$. While we are agnostic about the exact form of the null hypothesis, we emphasize that we only expect Assumption S.3.1 to hold under the null hypothesis, so the following result should be interpreted as a statement about the limiting rejection probability of $\phi_n^{\mathtt{F}}$ under the null hypothesis, whatever it may be.

**Theorem S.3.1.** *If Assumption S.3.1 holds, then*

$$\alpha - \frac{\kappa}{2^q} \le \liminf_{n\to\infty} P\left\{T_n^{\mathrm{F}} > \hat{c}_n^{\mathrm{F}}(1-\alpha)\right\} \le \limsup_{n\to\infty} P\left\{T_n^{\mathrm{F}} > \hat{c}_n^{\mathrm{F}}(1-\alpha)\right\} \le \alpha + \frac{\kappa}{2^q} \ .$$

Proof of Theorem S.3.1: The proof follows arguments similar to those employed in establishing Theorem 3.1. We again start by introducing notation that will streamline our arguments. Let $\mathbb{S} \equiv \bigotimes_{j\in J} \mathbf{R}^{d_\psi}$ and write an element of $s \in \mathbb{S}$ by $\{s_j : j \in J\}$. We further identify any $(g_1, \ldots, g_q) = g \in \mathbf{G}$ with an action on $s \in \mathbb{S}$ by $gs = \{g_j s_j : j \in J\}$. Since $F$ is continuous by hypothesis, note that Assumptions S.3.1(ii)-(iii) and the continuous mapping theorem imply

$$(F(S_n), \{F(g\hat{S}_n) : g \in \mathbf{G}\}) \xrightarrow{d} (F(S), \{F(gS) : g \in \mathbf{G}\}) \ . \tag{S.14}$$

Hence, by Assumption S.3.1(i), a set inclusion restriction, and the Portmanteau theorem (see, e.g., Theorem 1.3.4(iii) in van der Vaart and Wellner (1996)), we obtain

$$\begin{aligned}\limsup_{n\to\infty} P\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\} &\le \limsup_{n\to\infty} P\{T_n^{\mathtt{F}} \ge \hat{c}_n^{\mathtt{F}}(1-\alpha)\} \\ &\le P\left\{F(S) \ge \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{F(gS) \le u\} \ge 1-\alpha\right\}\right\} \ .\end{aligned} \tag{S.15}$$

In what follows, for any $s \in \mathbb{S}$, we denote the ordered values of $\{F(gs) : g \in \mathbf{G}\}$ according to

$$F^{(1)}(s|\mathbf{G}) \le \cdots \le F^{(|\mathbf{G}|)}(s|\mathbf{G}) \ .$$

Setting $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil$, we then obtain from (S.15) and Assumption S.3.1(iii) that

$$\begin{aligned}\limsup_{n\to\infty} P\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\} &\le P\{F(S) > F^{(k^*)}(S|\mathbf{G})\} + P\{F(S) = F^{(k^*)}(S|\mathbf{G})\} \\ &\le \alpha + P\{F(S) = F^{(k^*)}(S|\mathbf{G})\} \ ,\end{aligned} \tag{S.16}$$

where in the final inequality we employed that $gS \stackrel{d}{=} S$ for all $g \in \mathbf{G}$ and the basic properties of randomization tests; see, e.g., Theorem 15.2.1 in Lehmann and Romano (2005). Moreover, applying

Theorem 15.2.2 in Lehmann and Romano (2005) yields

$$P\{F(S) = F^{(k^*)}(S|\mathbf{G})\} = E[P\{F(S) = F^{(k^*)}(S|\mathbf{G})|S \in \{gS\}_{g\in\mathbf{G}}\}$$
$$= E\left[\frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{F(gS) = F^{(k^*)}(gS|\mathbf{G})\}\right] = \frac{\kappa}{2^q} , \quad \text{(S.17)}$$

where the final equality follows from Assumptions S.3.1(iv)-(v). The claim of the upper bound in the theorem therefore follows from results (S.16) and (S.17).

For the lower bound, note that $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil > |\mathbf{G}| + \kappa$ implies $\alpha - \kappa/|\mathbf{G}| \leq 0$, in which case the lower bound is immediate. Assume $k^* \leq |\mathbf{G}| - \kappa$ and note that result (S.14) and the Portmanteau Theorem, see, e.g., Theorem 1.3.4(ii) in van der Vaart and Wellner (1996) yield

$$\liminf_{n\to\infty} P\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\} \geq P\{F(S) > F^{(k^*)}(S|\mathbf{G})\} \geq P\{F(S) \geq F^{(k^*+\kappa)}(S|\mathbf{G})\} , \quad \text{(S.18)}$$

where the last inequality holds because $P\{F^{(\mathtt{z}+\kappa)}(S|\mathbf{G}) > F^{(\mathtt{z})}(S|\mathbf{G})\} = 1$ for any integer $\mathtt{z} \leq |\mathbf{G}| - \kappa$ by Assumptions S.3.1(iv)-(v). Next note $k^* + \kappa = \lceil |\mathbf{G}|((1-\alpha) + \kappa/|\mathbf{G}|)\rceil = \lceil |\mathbf{G}|(1-\alpha')\rceil$ with $\alpha' = \alpha - \kappa/2^q$ and so the properties of randomization tests (see Lehmann and Romano, 2005, Theorem 15.2.1) imply

$$P\{F(S) \geq F^{(k^*+\kappa)}(S|\mathbf{G})\} \geq \alpha - \frac{\kappa}{2^q} . \quad \text{(S.19)}$$

Thus, the lower bound holds by (S.18) and (S.19), and the claim of the theorem follows. ■

## S.3.1 Applications of the General Result

Below, we apply Theorem S.3.1 to establish results qualitatively similar to Theorem 3.3 for tests of nonlinear null hypotheses in both the linear model of Section 2 and the GMM framework of Hansen (1982).

### S.3.1.1 Nonlinear Null Hypotheses

Recall the setup introduced in Section 2, including Assumptions 2.1 and 2.2. For $h : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_h}$ with $d_h \leq d_\beta$ and $h$ continuously differentiable at $\beta$, consider testing

$$H_0 : h(\beta) = 0 \quad \text{vs.} \quad H_1 : h(\beta) \neq 0 . \quad \text{(S.20)}$$

We employ $T_n^{\mathtt{F}} = \|\sqrt{n}h(\hat{\beta}_n)\|^2$, where $\|\cdot\|$ is the Euclidean norm, as our test statistic. For our critical value, we use

$$\hat{c}_n^{\mathtt{F}}(1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\Big\{\|\nabla h(\hat{\beta}_n^{\mathrm{r}})\sum_{j\in J}\frac{g_j}{\sqrt{n}}\sum_{i\in I_{n,j}} \hat{\Omega}_{\tilde{Z},n}^{-1}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}^{\mathrm{r}}\|^2 \leq u\Big\} \geq 1-\alpha\right\} , \quad \text{(S.21)}$$

where $\nabla h(\hat{\beta}_n^{\mathrm{r}})$ denotes the Jacobian of $h : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_h}$, and $(\hat{\gamma}_n^{\mathrm{r}}, \hat{\beta}_n^{\mathrm{r}})$ are understood to be computed subject to the restriction that $h(\beta) = 0$ rather than $c'\beta = \lambda$. The following theorem bounds the limiting rejection probability of the test

$$\phi_n^{\mathtt{F}} \equiv I\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\}$$

under the null hypothesis.

**Theorem S.3.2.** *If Assumptions 2.1 and 2.2 hold and $h(\beta) = 0$ for $h : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_h}$ with $d_h \leq d_\beta$ and $h$ continuously differentiable at $\beta$, then*

$$\alpha - \frac{1}{2^{q-1}} \leq \liminf_{n\to\infty} P\left\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\right\} \leq \limsup_{n\to\infty} P\left\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\right\} \leq \alpha + \frac{1}{2^{q-1}} \ .$$

Sketch of Proof: Theorem S.3.2 follows from an application of Theorem S.3.1. To map $\phi_n^{\mathtt{F}}$ into the context of Theorem S.3.1, we let $X_{i,j} = (Y_{i,j}, Z'_{i,j}, W'_{i,j})'$ and define

$$\psi(X_{i,j}) = \nabla h(\beta)(\bar{a}\Omega_{\tilde{Z}})^{-1}\tilde{Z}_{i,j}\epsilon_{i,j} \ , \tag{S.22}$$

where recall $\bar{a} = \sum_{j\in J} a_j\xi_j$. It then follows by standard arguments and $\hat{\Omega}_{\tilde{Z},n} \stackrel{P}{\to} \bar{a}\Omega_{\tilde{Z}}$ by Assumptions 2.2(ii)-(iii), that $T_n^{\mathtt{F}}$ satisfies Assumption S.3.1(i) with $F : \mathbf{R}^q \to \mathbf{R}$ given by $F(c) = \|\sum_{j\in J} c_j\|^2$ for any $c = (c_1, \ldots, c_q)$ and $\psi(X_{i,j})$ as in (S.22). Moreover, by setting

$$\hat{\psi}_n(X_{i,j}) = \nabla h(\hat{\beta}_n^{\mathrm{r}})\hat{\Omega}_{\tilde{Z},n}^{-1}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}^{\mathrm{r}} \ , \tag{S.23}$$

we verify the critical value in (S.21) has the exact structure required by Theorem S.3.1. Further note that arguments similar to those leading to (A-37) in the proof of Theorem 3.3 yield

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\hat{\psi}_n(X_{i,j}) &= \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\nabla h(\beta)(\bar{a}\Omega_{\tilde{Z}})^{-1}\tilde{Z}_{i,j}(\epsilon_{i,j} + \tilde{Z}'_{i,j}(\beta - \hat{\beta}_n^{\mathrm{r}})) + o_P(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\psi(X_{i,j}) + \nabla h(\beta)\frac{a_j\xi_j}{\bar{a}}\sqrt{n}(\beta - \hat{\beta}_n^{\mathrm{r}}) + o_P(1) = \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\psi(X_{i,j}) + o_P(1) \ ,\end{aligned}$$

where the second equality follows from Assumption 2.2(iii), and the final equality follows from $\nabla h(\beta)\sqrt{n}(\beta - \hat{\beta}_n^{\mathrm{r}}) = o_P(1)$ due to $h(\hat{\beta}_n^{\mathrm{r}}) = h(\beta) = 0$. Hence, Assumption S.3.1(ii) is satisfied. Finally, Assumptions S.3.1(iii)-(v) hold immediately with $\kappa = 2$ by Assumptions 2.2(i)-(ii). ■

**Remark S.3.1.** In this application it is also natural to consider employing the critical value

$$\tilde{c}_n^{\mathtt{F}}(1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{\|\sqrt{n}h(\hat{\beta}_n^*(g))\|^2 \leq u\} \geq 1-\alpha\right\} \tag{S.24}$$

where, again, $\hat{\beta}_n^*(g)$ is understood to be computed as in Section 2 but by using $(\hat{\gamma}_n^{\mathrm{r}}, \hat{\beta}_n^{\mathrm{r}})$ corresponding to the restriction $h(\beta) = 0$ rather than $c'\beta = \lambda$. By the mean value theorem we then obtain

$$\sqrt{n}h(\hat{\beta}_n^*(g)) = \nabla h(\bar{\beta}_n(g))\sqrt{n}(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathrm{r}}) = \sum_{j\in J}\frac{g_j}{\sqrt{n}}\sum_{i\in I_{n,j}}\nabla h(\bar{\beta}_n(g))\hat{\Omega}_{\tilde{Z},n}^{-1}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}^{\mathrm{r}}$$

for some $\bar{\beta}_n(g)$ satisfying $\bar{\beta}_n(g) \xrightarrow{P} \hat{\beta}_n^{\rm r}$. Hence, the continuity of the the Jacobian $\nabla h$ implies that

$$\sqrt{n} h(\hat{\beta}_n^*(g)) = \sum_{j \in J} \frac{g_j}{\sqrt{n}} \sum_{i \in I_{n,j}} \hat{\psi}_n(X_{i,j}) + o_P(1) ,$$

which reveals a close relation between $\hat{c}_n^{\rm F}(1-\alpha)$ as in (S.21) and $\tilde{c}_n^{\rm F}(1-\alpha)$ as in (S.24). Inspecting the proof of Theorem S.3.1 (see, in particular, (S.14), (S.15), and (S.18)), then reveals the conclusion of Theorem S.3.1 continues to apply if we employ $\tilde{c}_n^{\rm F}(1-\alpha)$ in place of $\hat{c}_n^{\rm F}(1-\alpha)$; i.e.

$$\alpha - \frac{1}{2^{q-1}} \leq \liminf_{n \to \infty} P\left\{T_n^{\rm F} > \tilde{c}_n^{\rm F}(1-\alpha)\right\} \leq \limsup_{n \to \infty} P\left\{T_n^{\rm F} > \tilde{c}_n^{\rm F}(1-\alpha)\right\} \leq \alpha + \frac{1}{2^{q-1}} .$$

We note that if $h$ is linear, then $\hat{c}_n^{\rm F}(1-\alpha)$ and $\tilde{c}_n^{\rm F}(1-\alpha)$ are numerically equivalent and the upper bound on the limiting rejection probability can be shown to equal $\alpha$ (instead of $\alpha + 1/2^{q-1}$). ■

#### S.3.1.2 Generalized Method of Moments

In this section, we apply Theorem S.3.1 to study the properties of "score" bootstrap-based tests of nonlinear null hypotheses in a GMM setting with a "small" number of "large" clusters. As mentioned previously, the reason for relying on the "score" bootstrap instead of the wild bootstrap stems from there being no natural "residuals" in this setting.

To this end, let

$$\hat{\beta}_n \equiv \arg\min_b \left(\frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} m(X_{i,j}, b)\right)' \hat{\Sigma}_n \left(\frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} m(X_{i,j}, b)\right) , \qquad \text{(S.25)}$$

where $m(X_{i,j}, \cdot) : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_m}$ is a moment function $\hat{\Sigma}_n$ is a $d_m \times d_m$ weighting matrix. Under suitable conditions, see, e.g., Newey and McFadden (1994), $\hat{\beta}_n$ is consistent for its estimand, which we denote by $\beta$. For $h : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_h}$ with $d_\beta \leq d_h$ and $h$ continuously differentiable at $\beta$, we consider testing

$$H_0 : h(\beta) = 0 \quad \text{vs.} \quad H_1 : h(\beta) \neq 0 .$$

We again employ $T_n^{\tt F} = \|\sqrt{n} h(\hat{\beta}_n)\|^2$, where $\|\cdot\|$ is the Euclidean norm, as our test statistic. In order to describe a critical value with which to compare $T_n^{\tt F}$, define, for any $b \in \mathbf{R}^{d_\beta}$, the matrix

$$\hat{D}_n(b) \equiv \frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \nabla m(X_{i,j}, b) \qquad \text{(S.26)}$$

where $\nabla m(X_{i,j}, b)$ denotes the Jacobian of $m(X_{i,j}, \cdot) : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_m}$ at $b$. Further define, for $\hat{\beta}_n^{\rm r}$ the GMM estimator computed subject to the restriction $h(\hat{\beta}_n^{\rm r}) = 0$,

$$\hat{\psi}_n(X_{i,j}) = \nabla h(\hat{\beta}_n^{\rm r})(\hat{D}_n(\hat{\beta}_n^{\rm r})' \hat{\Sigma}_n \hat{D}_n(\hat{\beta}_n^{\rm r}))^{-1} \hat{D}_n(\hat{\beta}_n^{\rm r})' \hat{\Sigma}_n m(X_{i,j}, \hat{\beta}_n^{\rm r}) .$$

Using this notation, our critical value is given by

$$\hat{c}_n^{\mathrm{F}}(1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\Big\{\|\sum_{j\in J}\frac{g_j}{\sqrt{n}}\sum_{i\in I_{n,j}}\hat{\psi}_n(X_{i,j})\|^2 \le u\Big\} \ge 1-\alpha\right\} .$$

The test we study is therefore given by

$$\phi_n^{\mathtt{F}} \equiv I\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\} .$$

In order to apply Theorem S.3.1 to establish properties of $\phi_n^{\mathtt{F}}$, we impose the following assumption:

**Assumption S.3.2.** *The following statements hold:*

*(i)* $h : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_h}$ *is continuously differentiable at* $\beta$*.*

*(ii) There are full rank matrices* $\Sigma$ *and* $D(\beta)$ *such that* $\hat{\Sigma}_n \xrightarrow{P} \Sigma$ *and* $\hat{D}_n(b_n) \xrightarrow{P} D(\beta)$ *for any random variable* $b_n \in \mathbf{R}^{d_\beta}$ *satisfying* $b_n \xrightarrow{P} \beta$*.*

*(iii) The restricted and unrestricted estimators satisfy* $\sqrt{n}(\hat{\beta}_n^{\mathrm{r}} - \beta) = O_P(1)$ *and*

$$\sqrt{n}h(\hat{\beta}_n) = \nabla h(\beta)(D(\beta)'\Sigma D(\beta))^{-1}D(\beta)'\Sigma\frac{1}{\sqrt{n}}\sum_{j\in J}\sum_{i\in I_{n,j}} m(X_{i,j},\beta) + o_P(1) .$$

*(iv) There exists a collection of independent random variables* $\{\mathcal{N}_j\}_{j\in J}$*, where* $\mathcal{N}_j \in \mathbf{R}^{d_m}$ *and* $\mathcal{N}_j \sim N(0,\Sigma_j)$ *with* $\Sigma_j$ *positive definite, such that*

$$\left\{\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}} m(X_{i,j},\beta) : j \in J\right\} \xrightarrow{d} \{\mathcal{N}_j : j \in J\} .$$

*(v) For each* $j \in J$ *there is an* $a_j > 0$ *such that*

$$\frac{1}{n}\sum_{i\in I_{n,j}} \nabla m(X_{i,j}, b_n) \xrightarrow{P} a_j D(\beta)$$

*for any random variable* $b_n \in \mathbf{R}^{d_\beta}$ *satisfying* $b_n \xrightarrow{P} \beta$*.*

The following theorem bounds the limiting rejection probability of $\phi_n^{\mathtt{F}}$ under the null hypothesis.

**Theorem S.3.3.** *If Assumption S.3.2 holds and* $h(\beta) = 0$*, then*

$$\alpha - \frac{1}{2^{q-1}} \le \liminf_{n\to\infty} P\{T_n^{\mathrm{F}} > \hat{c}_n^{\mathrm{F}}(1-\alpha)\} \le \limsup_{n\to\infty} P\{T_n^{\mathrm{F}} > \hat{c}_n^{\mathrm{F}}(1-\alpha)\} \le \alpha - \frac{1}{2^{q-1}}$$

Sketch of Proof: Theorem S.3.3 follows from an application of Theorem S.3.1. Let $F : \mathbf{R}^q \to \mathbf{R}$ be given by $F(c) = \|\sum_{j\in J} c_j\|^2$ for any $c = (c_1,\ldots,c_q) \in \mathbf{R}^q$ and set $\psi : \mathbf{R}^{d_x} \to \mathbf{R}^{d_\beta}$ to equal

$$\psi(X_{i,j}) = \nabla h(\beta)(D(\beta)'\Sigma D(\beta))^{-1}D(\beta)'\Sigma m(X_{i,j},\beta) . \tag{S.27}$$

Assumption S.3.2(iii), continuity of $||\cdot||^2$, and the continuous mapping theorem imply Assumption S.3.1(i). Assumption S.3.1(iii) follows from S.3.2(iv) with

$$\mathcal{Z}_j = \nabla h(\beta)(D(\beta)'\Sigma D(\beta))^{-1}D(\beta)'\Sigma\mathcal{N}_j \ .$$

Assumptions S.3.1(iv) and S.3.1(v) are then immediate with $\kappa = 2$. We are then left with Assumption S.3.1(ii). By the mean value theorem and the definition of $\hat{\psi}_n(X_{i,j})$, we obtain

$$\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\hat{\psi}_n(X_{i,j}) = \nabla h(\hat{\beta}_n^{\rm r})(\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\hat{D}_n(\hat{\beta}_n^{\rm r}))^{-1}\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}} m(X_{i,j},\beta)$$
$$+ \nabla h(\hat{\beta}_n^{\rm r})(\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\hat{D}_n(\hat{\beta}_n^{\rm r}))^{-1}\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\frac{1}{n}\sum_{i\in I_{n,j}}\nabla m(X_{i,j},\bar{\beta}_n)\sqrt{n}(\hat{\beta}_n^{\rm r}-\beta) \ , \quad \text{(S.28)}$$

where $\bar{\beta}_n$ lies between $\hat{\beta}_n^{\rm r}$ and $\beta$. Assumptions S.3.2(i), S.3.2(ii), and S.3.2(iv) imply that the first term satisfies

$$\nabla h(\hat{\beta}_n^{\rm r})(\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\hat{D}_n(\hat{\beta}_n^{\rm r}))^{-1}\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}} m(X_{i,j},\beta) = \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\psi(X_{i,j}) + o_P(1) \ .$$

Assumptions S.3.2(i) and S.3.2(ii)-(iv), imply that the second term equals

$$\begin{aligned}\nabla h(\beta)(D(\beta)'\Sigma D(\beta))^{-1}D(\beta)'\Sigma(a_jD(\beta))\sqrt{n}(\hat{\beta}_n^{\rm r}-\beta) + o_P(1) &= a_j\nabla h(\beta)\sqrt{n}(\hat{\beta}_n^{\rm r}-\beta)+o_P(1) \\ &= o_P(1) \ ,\end{aligned}$$

where final equality follows from $0 = h(\hat{\beta}_n^{\rm r}) - h(\beta) = \nabla h(\bar{\beta}_n)\sqrt{n}(\hat{\beta}_n^{\rm r}-\beta) = \nabla h(\beta)\sqrt{n}(\hat{\beta}_n^{\rm r}-\beta) + o_P(1)$ for $\bar{\beta}_n$ between $\hat{\beta}_n^{\rm r}$ and $\beta$ by Assumptions S.3.2(i)-(iii). This completes the argument. ■

# References


Amemiya, T. (1985). *Advanced econometrics*. Harvard university press.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50** pp. 1029–1054.

Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, **1** 23–41.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Verlag.

Newey, W. K. and McFadden, D. (1994). Large-sample estimation and hypothesis testing. *Handbook of Econometrics*, **4** 2111–2245.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag.