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

April 23, 2018

Abstract

This paper studies the properties of the wild bootstrap-based test proposed in Cameron et al. (2008) in settings with clustered data. Cameron et al. (2008) provide simulations that suggest this test works well even in settings with as few as five clusters, but existing theoretical analyses of its properties all rely on an asymptotic framework in which the number of clusters is "large." In contrast to these analyses, we employ an asymptotic framework in which the number of clusters is "small," but the number of observations per cluster is "large." In this framework, we provide conditions under which the limiting rejection probability of an un-Studentized version of the test does not exceed the nominal level. Importantly, these conditions require, among other things, certain homogeneity restrictions on the distribution of covariates. We further establish that the limiting rejection probability of a Studentized version of the test does not exceed the nominal level by more than an amount that decreases exponentially with the number of clusters. We study the relevance of our theoretical results for finite samples via a simulation study.

Keywords: Wild bootstrap, Clustered Data, Randomization Tests.

*We thank Colin Cameron, Patrick Kline, Simon Lee, Magne Mogstad and Ulrich Mueller for helpful comments. The research of the first author was supported by National Science Foundation Grant SES-1530534. The research of the third author was supported by National Science Foundation Grants DMS-1308260, SES-1227091, and SES-1530661. We thank Max Tabord-Meehan for excellent research assistance.

1 Introduction

It is common in the empirical analysis of clustered data to be agnostic about the dependence structure within a cluster (Wooldridge, 2003; Bertrand et al., 2004). The robustness afforded by such agnosticism, however, may unfortunately result in many commonly used inferential methods behaving poorly in applications where the number of clusters is "small" (Donald and Lang, 2007). In response to this concern, Cameron et al. (2008) introduced a procedure based on the wild bootstrap of Liu (1988) and found in simulations that it led to tests that behaved remarkably well even in settings with as few as five clusters. This procedure is sometimes referred to as the "cluster" wild bootstrap, but we henceforth refer to it more compactly as the wild bootstrap. Due at least in part to these simulations, the wild bootstrap has emerged as arguably the most popular method for conducting inference in settings with few clusters. Recent examples of its use as either the leading inferential method or as a robustness check for conclusions drawn under other procedures include Acemoglu et al. (2011), Giuliano and Spilimbergo (2014), Kosfeld and Rustagi (2015), and Meng et al. (2015). The number of clusters in these empirical applications ranges from as few as five to as many as nineteen.

The use of the wild bootstrap in applications with such a small number of clusters contrasts sharply with existing analyses of its theoretical properties, which, to the best of our knowledge, all employ an asymptotic framework where the number of clusters tends to infinity. See, for example, Carter et al. (2017), Djogbenou et al. (2017), and MacKinnon et al. (2017). In this paper, we address this discrepancy by studying its properties in an asymptotic framework in which the number of clusters is fixed, but the number of observations per cluster tends to infinity. In this way, our asymptotic framework captures a setting in which the number of clusters is "small," but the number of observations per cluster is "large."

Our formal results concern the use of the wild bootstrap to test hypotheses about a linear combination of the coefficients in a linear regression model with clustered data. For this testing problem, we first provide conditions under which using the wild bootstrap with an un-Studentized test statistic leads to a test that has limiting rejection probability under the null hypothesis no greater than the nominal level. Our results require, among other things, certain homogeneity restrictions on the distribution of covariates. These homogeneity conditions are satisfied in particular if the distribution of covariates is the same across clusters, but, as explained in Section 2.1, are also satisfied in other circumstances. Importantly, when the regressors consist of cluster-level fixed effects and a single, scalar covariate, these conditions are immediately satisfied for hypotheses about the coefficient on the single, scalar covariate. In this way, our results help explain the remarkable behavior of the wild bootstrap in some simulation studies that feature a single, scalar covariate as well as the poor behavior of the wild bootstrap in simulation studies that violate our homogeneity requirements; see, for example, Ibragimov and Müller (2016) and Section 4 below.

Establishing the properties of a wild bootstrap-based test in an asymptotic framework in which the number of clusters is fixed requires fundamentally different arguments than those employed when the number of clusters diverges to infinity. Importantly, when the number of clusters is fixed, the wild bootstrap distribution is no longer a consistent estimator for the asymptotic distribution of the test statistic and hence "standard" arguments do not apply. Our analysis instead relies on a resemblance of the wild bootstrap-based test to a randomization test based on the group of sign changes with some key differences that, as explained in Section 3, prevent the use of existing results on the large-sample properties of randomization tests, including those in Canay et al. (2017). Despite these differences, we are able to show under our assumptions that the limiting rejection probability of the wild bootstrap-based test equals that of a suitable level-α randomization test.

We emphasize, however, that the asymptotic equivalence described above is delicate in that it relies crucially on the specific implementation of the wild bootstrap recommended by Cameron et al. (2008), which uses Rademacher weights and the restricted least squares estimator. Furthermore, it does not extend to the case where we Studentize the test statistic in the usual way. In that setting, our analysis only establishes that the test that employs a Studentized test statistic has limiting rejection probability under the null hypothesis that does not exceed the nominal level by more than a quantity that decreases exponentially with the number of clusters. In particular, when the number of clusters is eight (or more), this quantity is no greater than approximately 0.008.

This paper is part of a growing literature studying inference in settings where the number of clusters is "small," but the number of observations per cluster is "large." Ibragimov and Müller (2010) and Canay et al. (2017), for instance, develop procedures based on the cluster-level estimators of the coefficients. Importantly, these approaches do not require the homogeneity assumption on the distribution of covariates described above. Canay et al. (2017) is related to our theoretical analysis in that it also exploits a connection with randomization tests, but, as mentioned previously, the results in Canay et al. (2017) are not applicable to our setting. Bester et al. (2011) derives the asymptotic distribution of the full-sample estimator of the coefficients under assumptions similar to our own. Finally, there is a large literature studying the properties of variations of the wild bootstrap, including, in addition to some of the aforementioned references, Webb (2013) and MacKinnon and Webb (2014).

The remainder of the paper is organized as follows. In Section 2, we formally introduce the test we propose to study and the assumptions that will underlie our analysis. Our main results are contained in Section 3. In Section 4, we examine the relevance of our asymptotic analysis for finite samples via a simulation study. Section 5 briefly concludes. The proofs of all results can be found in the Appendix.

2 Setup

We index clusters by $j \in J \equiv \{1, \dots, q\}$ and units in the jth cluster by $i \in I_{n,j} \equiv \{1, \dots, n_j\}$. The observed data consists of an outcome of interest, $Y_{i,j}$, and two random vectors, $W_{i,j} \in \mathbf{R}^{d_w}$ and $Z_{i,j} \in \mathbf{R}^{d_z}$, that are related through the equation

$$Y_{i,j} = W_{i,j}'\gamma + Z_{i,j}'\beta + \epsilon_{i,j} \ , \tag{1}$$

where $\gamma \in \mathbf{R}^{d_w}$ and $\beta \in \mathbf{R}^{d_z}$ are unknown parameters and our requirements on $\epsilon_{i,j}$ are explained below in Section 2.1. Our goal is to test

$$H_0 : c'\beta = \lambda \quad \text{vs.} \quad H_1 : c'\beta \neq \lambda \ , \tag{2}$$

for given values of $c \in \mathbf{R}^{d_z}$ and $\lambda \in \mathbf{R}$, at level $\alpha \in (0,1)$. In this testing problem, γ is a nuisance parameter, such as the coefficient on a constant or the coefficients on cluster-level fixed effects. An important special case of this framework is a test of the null hypothesis that a particular component of β equals a given value. While we do not develop it further in this paper, our results extend straightforwardly to testing null hypotheses concerning multiple linear combinations of β simultaneously.

In order to test (2), we first consider tests that reject for large values of the statistic

$$T_n \equiv |\sqrt{n}(c'\hat{\beta}_n - \lambda)| \ , \tag{3}$$

where $\hat{\gamma}_n$ and $\hat{\beta}_n$ are the ordinary least squares estimator of γ and β in (1). We also consider tests that reject for large values of a Studentized version of T_n, but postpone a more detailed description of such tests to Section 3.2. For a critical value with which to compare T_n, we employ a version of the one proposed by Cameron et al. (2008). Specifically, we obtain a critical value through the following construction:

Step 1: Compute $\hat{\gamma}_n^{\mathrm{r}}$ and $\hat{\beta}_n^{\mathrm{r}}$, the restricted least squares estimators of γ and β in (1) obtained under the constraint that $c'\beta = \lambda$. Note that $c'\hat{\beta}_n^{\mathrm{r}} = \lambda$ by construction. ■

Step 2: Let $\mathbf{G} = \{-1, 1\}^q$ and for any $g = (g_1, \dots, g_q) \in \mathbf{G}$ define

$$Y_{i,j}^*(g) \equiv W_{i,j}'\hat{\gamma}_n^{\mathrm{r}} + Z_{i,j}'\hat{\beta}_n^{\mathrm{r}} + g_j\hat{\epsilon}_{i,j}^{\mathrm{r}} \ , \tag{4}$$

where $\hat{\epsilon}_{i,j}^{\mathrm{r}} = Y_{i,j} - W_{i,j}'\hat{\gamma}_n^{\mathrm{r}} - Z_{i,j}'\hat{\beta}_n^{\mathrm{r}}$. For each $g = (g_1, \dots, g_q) \in \mathbf{G}$ then compute $\hat{\gamma}_n^*(g)$ and $\hat{\beta}_n^*(g)$, the ordinary least squares estimators of γ and β in (1) obtained using $Y_{i,j}^*(g)$ in place of $Y_{i,j}$. ■

Step 3: Compute the $1-\alpha$ quantile of $\{|c'\sqrt{n}(\hat{\beta}^*_n(g)-\hat{\beta}^{\rm r}_n)|\}_{g\in\mathbf{G}}$, denoted by

$$\hat{c}_n(1-\alpha)\equiv\inf\left\{u\in\mathbf{R}:\frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}}I\{|c'\sqrt{n}(\hat{\beta}^*_n(g)-\hat{\beta}^{\rm r}_n)|\le u\}\ge 1-\alpha\right\}~, \quad (5)$$

where $I\{A\}$ equals one whenever the event A is true and equals zero otherwise. ■

In what follows, we study the test ϕ_n of (2) that rejects whenever T_n exceeds the critical value $\hat{c}_n(1-\alpha)$, i.e.,

$$\phi_n\equiv I\{T_n>\hat{c}_n(1-\alpha)\}~. \quad (6)$$

It is worth noting that the critical value $\hat{c}_n(1-\alpha)$ defined in (5) may also be written as

$$\inf\{u\in\mathbf{R}:P\{|c'\sqrt{n}(\hat{\beta}^*_n(\omega)-\hat{\beta}^{\rm r}_n)|\le u|X^{(n)}\}\ge 1-\alpha\}~,$$

where $X^{(n)}$ denotes the full sample of observed data and $\omega\sim\text{Unif}(\mathbf{G})$ independently of $X^{(n)}$. This way of writing $\hat{c}_n(1-\alpha)$ coincides with the existing literature on the wild bootstrap that sets the cluster weights $\omega=(\omega_1,\ldots,\omega_q)$ to be i.i.d. Rademacher random variables – i.e., ω_j equals ± 1 with equal probability. Furthermore, it suggests a natural way of approximating $\hat{c}_n(1-\alpha)$ using simulation, which may be helpful when $|\mathbf{G}|$ is large.

2.1 Assumptions

We next introduce the assumptions that will underlie our analysis of the properties of the test ϕ_n defined in (6) as well as its Studentized counterpart. In order to state these assumptions formally, we require some additional notation. In particular, it is useful to introduce a $d_w\times d_z$-dimensional matrix $\hat{\Pi}_n$ satisfying the orthogonality conditions

$$\sum_{j\in J}\sum_{i\in I_{n,j}}(Z_{i,j}-\hat{\Pi}'_nW_{i,j})W'_{i,j}=0~. \quad (7)$$

Our assumptions will guarantee that, with probability tending to one, $\hat{\Pi}_n$ is the unique $d_w\times d_z$ matrix satisfying (7). Thus, $\hat{\Pi}_n$ corresponds to the coefficients of a least squares projection of $Z_{i,j}$ on $W_{i,j}$. The "residuals" from this projection,

$$\tilde{Z}_{i,j}\equiv Z_{i,j}-\hat{\Pi}'_nW_{i,j}~, \quad (8)$$

will play an important role in our analysis as well. Finally, for every $j \in J$, let $\hat{\Pi}^{c}_{n,j}$ be a $d_w \times d_z$-dimensional matrix satisfying the orthogonality conditions

$$\sum_{i \in I_{n,j}} (Z_{i,j} - (\hat{\Pi}^{c}_{n,j})' W_{i,j}) W'_{i,j} = 0 \ . \tag{9}$$

Because the restrictions in (9) involve only data from cluster j, there may be multiple matrices $\hat{\Pi}^{c}_{n,j}$ satisfying (9) even asymptotically. Non-uniqueness occurs, for instance, when $W_{i,j}$ includes cluster-level fixed effects. For our purposes, however, we only require that for each $j \in J$ the quantities $(\hat{\Pi}^{c}_{n,j})' W_{i,j}$ with $i \in I_{n,j}$ are uniquely defined, which is satisfied by construction.

Using this notation, we may now introduce our assumptions. Before doing so, we note that all limits are understood to be as $n \to \infty$ and it is assumed that $n_j \to \infty$ as $n \to \infty$. Importantly, the number of clusters, q, is fixed in our asymptotic framework.

Assumption 2.1. *The following statements hold:*

(i) The quantity

$$\frac{1}{\sqrt{n}} \sum_{j \in J} \sum_{i \in I_{n,j}} \begin{pmatrix} W_{i,j} \epsilon_{i,j} \\ Z_{i,j} \epsilon_{i,j} \end{pmatrix}$$

converges in distribution.

(ii) The quantity

$$\frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \begin{pmatrix} W_{i,j} W'_{i,j} & W_{i,j} Z'_{i,j} \\ Z_{i,j} W'_{i,j} & Z_{i,j} Z'_{i,j} \end{pmatrix}$$

converges in probability to a positive-definite matrix.

Assumption 2.1 imposes sufficient conditions to ensure that the ordinary least squares estimators of γ and β in (1) are well behaved. It further implies that the least squares estimators of γ and β in (1) subject to the restriction that $c'\beta = \gamma$ are well behaved under the null hypothesis in (2). Assumption 2.1 in addition guarantees $\hat{\Pi}_n$ converges in probability to a well-defined limit. The requirements of Assumption 2.1 are satisfied, for example, whenever the within-cluster dependence is sufficiently weak to permit application of suitable laws of large numbers and central limit theorems.

Whereas Assumption 2.1 governs the asymptotic properties of the restricted and unrestricted least squares estimators, our next assumption imposes additional conditions that are employed in our analysis of the wild bootstrap.

Assumption 2.2. *The following statements hold:*

(i) There exists a collection of independent random variables $\{\mathcal{Z}_j\}_{j \in J}$, *where* $\mathcal{Z}_j \in \mathbf{R}^{d_z}$

and $\mathcal{Z}_j \sim N(0, \Sigma_j)$ with Σ_j positive definite for all $j \in J$, such that

$$\left\{ \frac{1}{\sqrt{n_j}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \epsilon_{i,j} : j \in J \right\} \xrightarrow{d} \{\mathcal{Z}_j : j \in J\} \ .$$

(ii) For each $j \in J$, $n_j/n \to \xi_j > 0$.

(iii) For each $j \in J$,

$$\frac{1}{n_j} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \xrightarrow{P} a_j \Omega_{\tilde{Z}} \ ,$$

where $a_j > 0$ and $\Omega_{\tilde{Z}}$ is positive definite.

(iv) For each $j \in J$,

$$\frac{1}{n_j} \sum_{i \in I_{n,j}} \| W'_{i,j} (\hat{\Pi}_n - \hat{\Pi}^c_{n,j}) \|^2 \xrightarrow{P} 0 \ .$$

The distributional convergence in Assumption 2.2(i) is satisfied, for example, whenever the within-cluster dependence is sufficiently weak to permit application of a suitable central limit theorem and the data are independent across clusters or, as explained in Bester et al. (2011), the "boundaries" of the clusters are "small." The additional requirement that $\mathcal{Z}_j$ have full rank covariance matrices requires that $Z_{i,j}$ can not be expressed as a linear combination of $W_{i,j}$ within each cluster. Assumption 2.2(ii) governs the relative sizes of the clusters. It permits clusters to have different sizes, but not dramatically so. Assumptions 2.2(iii)-(iv) are the main "homogeneity" assumptions required for our analysis of the wild bootstrap. These two assumptions are satisfied, for example, whenever the distributions of $(W'_{i,j}, Z'_{i,j})'$ are the same across clusters, but may also hold when that is not the case. For example, if $Z_{i,j}$ is a scalar, then Assumption 2.2(iii) reduces to the weak requirement that the average of $\tilde{Z}^2_{i,j}$ within each cluster converges in probability to a non-zero constant. Similarly, if $W_{i,j}$ includes only cluster-level fixed effects, then Assumption 2.2(iv) is trivially satisfied; see Example 2.1. In contrast, Assumption 2.2 is violated by the simulation design in Ibragimov and Müller (2016), in which the size of the wild bootstrap-based test exceeds its nominal level. Finally, we note that under additional conditions it is possible to test Assumptions 2.2(iii)-(iv) directly.

It is worth noting that our assumptions can accommodate additive "common shocks" η_j to each cluster by defining them to be the coefficients on cluster-level fixed effects rather than part of $\epsilon_{i,j}$. This viewpoint is helpful in seeing that Models 1 and 2 in our simulation study in Section 4 satisfy our assumptions. As the following example illustrates, even when such "common shocks" are not present, it may be desirable to include cluster-level fixed effects.

Example 2.1. (*Cluster-Level Fixed Effects*) In certain applications, adding additional regressors $W_{i,j}$ can aid in verifying Assumptions 2.2(iii)-(iv). In order to gain an appreciation for this possibility, suppose that

$$Y_{i,j} = \gamma + Z'_{i,j}\beta + \epsilon_{i,j}$$

with $\gamma \in \mathbf{R}$, $E[\epsilon_{i,j}] = 0$ and $E[Z_{i,j}\epsilon_{i,j}] = 0$. If the researcher specifies that $W_{i,j}$ is simply a constant, then Assumption 2.2(iv) demands that the cluster-level sample means of $Z_{i,j}$ all tend in probability to the same constant, while Assumption 2.2(iii) implies the cluster-level sample covariance matrices of $Z_{i,j}$ all tend in probability to the same, positive-definite matrix up to scale. On the other hand, if the researcher specifies that $W_{i,j}$ includes only cluster-level fixed effects, then Assumption 2.2(iv) is immediately satisfied, while Assumption 2.2(iii) is again satisfied whenever the cluster-level sample covariance matrices of $Z_{i,j}$ all tend in probability to the same, positive-definite matrix up to scale. ■

We conclude this section with a brief illustration of the content of our assumptions in a canonical differences-in-differences setting (Angrist and Pischke, 2008).

Example 2.2. (*Differences-in-Differences*) 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. 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

$$\lambda \equiv \frac{\sum_{j \in J(1)} n_j(1) p_j}{\sum_{j \in J} n_j(1) p_j} \ , \tag{10}$$

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 $(\hat{\Pi}^{c}_n)' W_{i,j} = Z_{i,j}$, while

$$\hat{\Pi}'_n W_{i,j} = \begin{cases} -p_j\lambda & \text{if } T_{i,j} = 1 \text{ and } j \in J(0) \\ (1-\lambda)p_j\lambda & \text{if } T_{i,j} = 1 \text{ and } j \in J(1) \\ (1-p_j)\lambda & \text{if } T_{i,j} = 2 \text{ and } j \in J(0) \\ \lambda + (1-\lambda)p_j & \text{if } T_{i,j} = 2 \text{ and } j \in J(1) \end{cases} \ , \tag{11}$$

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 λ, 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. ■

3 Asymptotic Properties

In this section, we first analyze the properties of the test ϕ_n defined in (6) under Assumptions 2.1 and 2.2. We then proceed to analyze the properties of a Studentized version of this test under the same assumptions.

3.1 Main Result

The following theorem establishes that the test ϕ_n has limiting rejection probability under the null hypothesis that does not exceed the nominal level α. It further establishes a lower bound on the limiting rejection probability of the test under the null hypothesis.

Theorem 3.1. *If Assumptions 2.1 and 2.2 hold and $c'\beta = \lambda$, then*

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

To gain some intuition into the conclusion of Theorem 3.1, it is important to note that the wild bootstrap does not re-sample the regressors. As a result, differences in T_n and its bootstrap counterpart are exclusively due to differences in the "scores." Formally, $T_n = F_n(s_n)$ for some function $F_n \colon \mathbf{R}^q \to \mathbf{R}$ and

$$s_n \equiv \left\{ \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \epsilon_{i,j} : j \in J \right\} \tag{12}$$

denoting the cluster "scores," while, for any $g \in \mathbf{G}$, $|\sqrt{n} c'(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n)| = F_n(g \hat{s}_n)$ where

$$\hat{s}_n \equiv \left\{ \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \hat{\epsilon}^{\mathrm{r}}_{i,j} : j \in J \right\} \tag{13}$$

denotes the cluster "bootstrap scores" and the notation ga is shorthand for $(g_1 a_1, \ldots, g_q a_q)$ for any $a \in \mathbf{R}^q$. This observation implies that the test ϕ_n defined in (6) rejects if and only if

$$\sum_{g\in\mathbf{G}} I\{F_n(s_n) > F_n(g\hat{s}_n)\} > \lceil|\mathbf{G}|(1-\alpha)\rceil \ , \tag{14}$$

where, for any $x \in \mathbf{R}$, $\lceil x\rceil$ represents the smallest integer larger than x. The characterization of ϕ_n in (14) reveals a resemblance to a randomization test, but also highlights an important difference: the action g is applied to a different statistic (i.e., $\hat{s}_n$) than the one defining the full-sample test statistic (i.e., s_n). This distinction prevents the application of results in Canay et al. (2017). In fact, s_n and $\hat{s}_n$ do not even tend in distribution to the same limit.

In the proof of Theorem 3.1 in the Appendix, we show under Assumptions 2.1 and 2.2 that the limiting rejection probability of ϕ_n equals that of a level-α randomization test, from which the conclusion of the theorem follows immediately. Despite the resemblance described above, relating the limiting rejection probability of ϕ_n to that of a level-α randomization test is delicate. In fact, the conclusion of Theorem 3.1 is not robust to variants of ϕ_n that construct "bootstrap" outcomes $Y^*_{i,j}(g)$ in other ways, such as the weighting schemes in Mammen (1993) and Webb (2013). We explore this in our simulation study in Section 4. The conclusion of Theorem 3.1 is also not robust to the use of the ordinary least squares estimators of γ and β instead of the restricted estimators $\hat{\gamma}^{\mathrm{r}}_n$ and $\hat{\beta}^{\mathrm{r}}_n$. Notably, the use of the restricted estimators is encouraged by Davidson and MacKinnon (1999) and Cameron et al. (2008).

Remark 3.1. The proof of Theorem 3.1 differs considerably from the existing literature on the properties of ϕ_n in asymptotic frameworks where the number of clusters is "large." In particular, those analyses all proceed by first deriving the limit in distribution of T_n and then establishing that $\hat{c}_n(1-\alpha)$ tends in probability to the appropriate quantile of this limiting distribution. In our asymptotic framework, in contrast, the bootstrap distribution is not a consistent estimator for the limiting distribution of T_n and $\hat{c}_n(1-\alpha)$ need not even settle down. ■

Remark 3.2. The conclusion of Theorem 3.1 can be extended to linear models with endogeneity. In particular, one may consider the test obtained by replacing the ordinary least squares estimator and the least squares estimator restricted to satisfy $c'\beta = \lambda$ with instrumental variable counterparts. Under assumptions that parallel Assumptions 2.1 and 2.2, it is possible to show using arguments similar to those in the proof of Theorem 3.1 that the conclusion of Theorem 3.1 holds for the test obtained in this way. ■

Remark 3.3. For testing certain null hypotheses, it is possible to provide conditions under which wild bootstrap-based tests are valid in finite samples. In particular, suppose that $W_{i,j}$ is empty and the goal is to test a null hypothesis that specifies all values of β. For such a problem, $\hat{\epsilon}^{\mathrm{r}}_{i,j} = \epsilon_{i,j}$ and as a result the wild bootstrap-based test is numerically equivalent to a randomization test. Using this observation, it is then straightforward to provide conditions under which a wild bootstrap-based test of such null hypotheses is level α in finite samples. For example, sufficient conditions are that $\{(\epsilon_{i,j}, Z_{i,j}) : i \in I_{n,j}\}$ be independent across clusters and

$$\{\epsilon_{i,j} : i \in I_{n,j}\}|\{Z_{i,j} : i \in I_{n,j}\} \stackrel{d}{=} \{-\epsilon_{i,j} : i \in I_{n,j}\}|\{Z_{i,j} : i \in I_{n,j}\}$$

for all $j \in J$. Davidson and Flachaire (2008) present related results under independence between $\epsilon_{i,j}$ and $Z_{i.j}$. In contrast, because we are focused on tests of (2), which only specify the value of a linear combination of the coefficients in (1), wild bootstrap-based tests are not guaranteed finite-sample validity even under such strong conditions. ■

3.2 Studentization

We now analyze the limiting rejection probability under the null hypothesis of a Studentized version of ϕ_n. Before proceeding, we require some additional notation in order to define formally the variance estimators that we employ. To this end, let

$$\hat{\Omega}_{\tilde{Z},n} \equiv \frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} , \tag{15}$$

where $\tilde{Z}_{i,j}$ is defined as in (8). For $\hat{\gamma}_n$ and $\hat{\beta}_n$ the ordinary least squares estimators of γ and β in (1) and $\hat{\epsilon}_{i,j} \equiv Y_{i,j} - W'_{i,j}\hat{\gamma}_n - Z'_{i,j}\hat{\beta}_n$, define

$$\hat{V}_n \equiv \frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \sum_{k \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{k,j} \hat{\epsilon}_{i,j} \hat{\epsilon}_{k,j} .$$

Using this notation, we define our Studentized test statistic to be $T_n/\hat{\sigma}_n$, where

$$\hat{\sigma}_n^2 \equiv c' \hat{\Omega}^{-1}_{\tilde{Z},n} \hat{V}_n \hat{\Omega}^{-1}_{\tilde{Z},n} c . \tag{16}$$

Next, for any $g \in \mathbf{G} \equiv \{-1, 1\}^q$, recall that $(\hat{\gamma}_n^*(g)', \hat{\beta}_n^*(g)')'$ denotes the unconstrained ordinary least squares estimator of $(\gamma', \beta')'$ obtained from regressing $Y^*_{i,j}(g)$ (as defined in (4)) on $W_{i,j}$ and $Z_{i,j}$. We therefore define the $d_z \times d_z$ covariance matrix

$$\hat{V}_n^*(g) \equiv \frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \sum_{k \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{k,j} \hat{\epsilon}^*_{i,j}(g) \hat{\epsilon}^*_{k,j}(g) ,$$

with $\hat{\epsilon}^*_{i,j}(g) = Y^*_{i,j}(g) - W'_{i,j}\hat{\gamma}^*_n(g) - Z'_{i,j}\hat{\beta}^*_n(g)$, as the wild bootstrap-analogue to $\hat{V}_n$, and

$$\hat{\sigma}^*_n(g)^2 \equiv c' \hat{\Omega}^{-1}_{\tilde{Z},n} \hat{V}^*_n(g) \hat{\Omega}^{-1}_{\tilde{Z},n} c \tag{17}$$

to be the wild bootstrap-analogue to $\hat{\sigma}^2_n$. Notice that since the regressors are not resampled when implementing the wild bootstrap, the matrix $\hat{\Omega}_{\tilde{Z},n}$ is employed in computing both $\hat{\sigma}_n$ and $\hat{\sigma}^*_n(g)$. Finally, we set as our critical value

$$\hat{c}^{\mathsf{s}}_n(1-\alpha) \equiv \inf \left\{ u \in \mathbf{R} : \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} I \left\{ |\sqrt{n} \frac{c'(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n)}{\hat{\sigma}^*_n(g)}| \leq u \right\} \geq 1 - \alpha \right\} . \qquad (18)$$

As in Section 2, we can employ simulation to approximate $\hat{c}^{\mathsf{s}}_n(1-\alpha)$ by generating q-dimensional vectors of i.i.d. Rademacher random variables independently of the data.

Using this notation, the Studentized version of ϕ_n that we consider is the test ϕ^{s}_n of (2) that rejects whenever $T_n/\hat{\sigma}_n$ exceeds the critical value $\hat{c}^{\mathsf{s}}_n(1-\alpha)$, i.e.,

$$\phi^{\mathsf{s}}_n \equiv I\{T_n/\hat{\sigma}_n > \hat{c}^{\mathsf{s}}_n(1-\alpha)\} . \qquad (19)$$

The following theorem studies the limiting rejection probability of this test under the null hypothesis.

Theorem 3.2. *If Assumptions 2.1 and 2.2 hold and $c'\beta = \lambda$, then*

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

Theorem 3.2 indicates that Studentizing the test-statistic T_n may lead to the limiting rejection proability of the test exceeding its nominal level, but by an amount no greater than 2^{1-q}, where q denotes the number of clusters. As explained further in Remark 3.4 below, the reason for this possible over-rejection is that Studentizing T_n results in a test whose limiting rejection probability no longer equals that of a level-α randomization test. Its limiting rejection probability, however, can still be bounded by that of a modified randomization test that rejects the null hypothesis whenever the p-value is *weakly* smaller than α instead of *strictly* smaller than α. This modified randomization test has rejection probability under the null hypothesis bounded above by $\alpha + 2^{1-q}$, from which the conclusion of the theorem follows. This implies, for example, that in applications with eight or more clusters, the amount by which the limiting rejection probabilty under the null hypothesis exceeds the nominal level will be no greater than 0.008. Of course, these results also imply that it is possible to "size correct" the test simply by replacing α with $\alpha - 2^{1-q}$.

Remark 3.4. Recall from the discussion in Section 3 that ϕ_n may be written as in (14). In a similar way, ϕ^{s}_n defined in (19) can be shown to reject if and only if

$$\sum_{g \in \mathbf{G}} I\{F^{\mathsf{s}}_n(t_n) > F^{\mathsf{s}}_n(g\hat{t}_n)\} > \lceil |\mathbf{G}|(1-\alpha) \rceil , \qquad (20)$$

for a function F^{s}_n and suitable statistics t_n and $\hat{t}_n$. In contrast to the situation with ϕ_n, however, it is possible that $F^{\mathsf{s}}_n(t_n) > F^{\mathsf{s}}_n(g\hat{t}_n)$ when $g = \pm(1, \ldots, 1) \in \mathbf{G}$. As a result, a test that rejects if and only if (20) occurs may differ even asymptotically from a test that follows the same decision rule but employs $F_n^{\mathsf{s}}(gt_n)$ in place of $F_n^{\mathsf{s}}(g\hat{t}_n)$. This subtle distinction underlies the differences in the conclusions of Theorems 3.1 and 3.2. ■

4 Simulation Study

In this section, we illustrate the results in Section 3 with a simulation study. In all cases, data is generated as

$$Y_{i,j} = \gamma + Z'_{i,j}\beta + \sigma(Z_{i,j})(\eta_j + \epsilon_{i,j}) \ , \tag{21}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, q$, where η_j, $Z_{i,j}$, $\sigma(Z_{i,j})$ and $\epsilon_{i,j}$ are specified as follows.

Model 1: We set $\gamma = 1$; $d_z = 1$; $Z_{i,j} = A_j + \zeta_{i,j}$ where $A_j \perp\!\!\!\perp \zeta_{i,j}$, $A_j \sim N(0,1)$, $\zeta_{i,j} \sim N(0,1)$; $\sigma(Z_{i,j}) = Z_{i,j}^2$; and $\eta_j \perp\!\!\!\perp \epsilon_{i,j}$ with $\eta_j \sim N(0,1)$ and $\epsilon_{i,j} \sim N(0,1)$.

Model 2: As in Model 1, but we set $Z_{i,j} = \sqrt{j}(A_j + \zeta_{i,j})$.

Model 3: As in Model 1, but $d_z = 3$; $\beta = (\beta_1, 1, 1)$; $Z_{i,j} = A_j + \zeta_{i,j}$ with $A_j \sim N(0, \mathbb{I}_3)$ and $\zeta_{i,j} \sim N(0, \Sigma_j)$, where $\mathbb{I}_3$ is a 3×3 identity matrix and Σ_j, $j = 1, \ldots, q$, is randomly generated following Marsaglia and Olkin (1984).

Model 4: As in Model 1, but $d_z = 2$, $Z_{i,j} \sim N(\mu_1, \Sigma_1)$ for $j > q/2$ and $Z_{i,j} \sim N(\mu_2, \Sigma_2)$ for $j \leq q/2$, where $\mu_1 = (-4, -2)$, $\mu_2 = (2, 4)$, $\Sigma_1 = \mathbb{I}_2$,

$$\Sigma_2 = \begin{bmatrix} 10 & 0.8 \\ 0.8 & 1 \end{bmatrix} ,$$

$\sigma(Z_{i,j}) = (Z_{1,i,j} + Z_{2,i,j})^2$, and $\beta = (\beta_1, 2)$.

For each of the above specifications, we test the null hypothesis $H_0 : \beta_1 = 1$ against the unrestricted alternative at level $\alpha = 10\%$. We further consider different values of (n, q) with $n \in \{50, 300\}$ and $q \in \{4, 5, 6, 8\}$ as well as both $\beta_1 = 1$ (i.e., under the null hypothesis) and $\beta_1 = 0$ (i.e., under the alternative hypothesis).

The results of our simulations are presented in Tables 1–4 below. Rejection probabilities are computed using 5000 replications. Rows are labeled in the following way:

un-Stud: Corresponds to the un-Studentized test studied in Theorem 3.1.

Stud: Corresponds to the Studentized test studied in Theorem 3.2.

ET-uS: Corresponds to the equi-tail analog of the un-Studentized test. This test rejects when the un-Studentized test statistic $T_n = \sqrt{n}(c'\hat{\beta}_n - \lambda)$ is either below $\hat{c}_n(\alpha/2)$ or above $\hat{c}_n(1-\alpha/2)$, where $\hat{c}_n(1-\alpha)$ is defined in (5).

		Rade - with Fixed effects				Rade - without Fixed effects				Mammen - with Fixed effects			
		q				q				q			
	Test	4	5	6	8	4	5	6	8	4	5	6	8
Model 1 $n = 50$	un-Stud	6.48	9.90	9.34	9.42	9.24	14.48	13.80	12.48	15.40	14.42	13.06	12.16
	Stud	7.36	10.42	9.54	9.76	7.74	10.80	10.04	9.86	6.10	6.26	5.16	4.58
	ET-uS	1.48	7.40	9.64	9.26	1.50	11.42	14.00	12.16	2.32	3.14	3.30	4.74
	ET-S	4.24	8.64	9.90	9.52	3.08	8.34	10.32	9.46	24.98	25.72	24.32	22.04
Model 2 $n = 50$	un-Stud	9.02	5.96	9.70	9.98	10.58	15.84	15.60	15.42	14.26	13.62	13.78	13.72
	Stud	9.44	7.74	9.72	10.08	8.18	10.38	10.06	11.04	5.56	5.92	4.60	4.10
	ET-uS	6.68	1.58	9.88	9.72	1.34	12.44	15.68	15.00	1.16	1.54	2.22	3.58
	ET-S	7.60	4.02	10.34	9.88	2.48	8.30	10.24	10.80	26.86	25.42	25.26	25.40
Model 1 $n = 300$	un-Stud	7.24	9.72	9.46	10.16	10.54	15.48	14.32	14.24	15.58	14.78	13.48	12.88
	Stud	8.42	10.22	9.64	10.16	8.62	11.24	10.42	10.86	6.62	6.88	5.30	4.58
	ET-uS	2.10	7.14	9.66	9.84	1.10	12.00	14.42	13.82	1.82	2.66	3.62	4.70
	ET-S	4.18	8.12	10.12	9.92	2.80	8.78	10.74	10.56	26.06	25.08	24.38	24.14
Model 2 $n = 300$	un-Stud	6.96	9.68	9.74	10.12	12.30	17.74	16.20	15.26	15.50	14.86	14.08	13.34
	Stud	8.26	10.16	9.86	10.16	8.88	10.96	10.28	10.66	6.64	6.18	4.80	4.34
	ET-uS	2.00	7.26	10.00	9.96	1.30	13.60	16.24	14.74	0.98	1.80	2.36	3.40
	ET-S	4.36	8.16	10.42	9.88	3.02	8.00	10.44	10.40	27.14	26.80	26.66	25.42

Table 1: Rejection probability under the null hypothesis $\beta_1 = 1$ with $\alpha = 10\%$.

ET-S: Corresponds to the equi-tail analog of the Studentized test. This test rejects when the Studentized test statistic $T_n/\hat{\sigma}_n$ is either below $\hat{c}^{\mathsf{s}}_n(\alpha/2)$ or above $\hat{c}^{\mathsf{s}}_n(1-\alpha/2)$, where $\hat{\sigma}_n$ and $\hat{c}^{\mathsf{s}}_n(1-\alpha)$ are defined in (16) and (18) respectively.

Each of the tests may be implemented with or without fixed effects (see Example 2.1), and with Rademacher weights or alternative weighting schemes as in Mammen (1993).

Tables 1 and 2 display the results for Models 1 and 2 under the null and alternative hypotheses respectively. These two models satisfy Assumptions 2.2(iii)–(iv) when the regression includes cluster-level fixed effects but not when only a constant term is included; see Example 2.1. Table 3 displays the results for Models 3 and 4 under the null hypothesis. These two models violate Assumptions 2.2(iii)–(iv) and are included to explore sensitivity to violations of these conditions. Finally, Table 4 displays results for Model 1 with $\alpha = 12.5\%$ to study the possible over-rejection under the null hypothesis of the Studentized test, as described in Theorem 3.2.

We organize our discussion of the results by test.

un-Stud: As expected in light of Theorem 3.1 and Example 2.1, Table 1 shows the un-Studentized test has rejection probability under the null hypothesis very close to the nominal level when the regression includes cluster-level fixed effects and the number of clusters is larger than four. When $q = 4$, however, the test is conservative in the sense that the rejection probability under the null hypothesis may be strictly below its nominal level. In fact, when $\alpha = 5\%$ (not reported), the test rarely rejects when $q = 4$ and is somewhat conservative for $q = 5$. Table 1 also illustrates the importance of including cluster-level fixed effects in the

	Test	Rade - with Fixed effects				Rade - without Fixed effects				Mammen - with Fixed effects			
		q				q				q			
		4	5	6	8	4	5	6	8	4	5	6	8
	un-Stud	19.80	33.14	39.34	42.28	20.42	34.94	39.54	40.74	35.46	37.86	40.84	42.50
Model 1	Stud	22.44	33.72	39.22	42.40	20.76	31.84	34.94	35.90	18.08	18.68	20.78	28.88
$n = 50$	ET-uS	5.64	28.80	39.70	41.62	4.60	30.32	39.90	40.16	10.14	15.84	22.06	29.26
	ET-S	11.08	30.10	39.76	41.72	9.58	28.40	35.66	35.44	51.16	51.94	54.50	55.76
	un-Stud	13.34	20.28	20.04	18.88	15.56	25.16	23.38	21.58	22.68	22.28	20.94	20.34
Model 2	Stud	16.00	20.66	19.66	18.40	13.94	19.24	17.86	16.68	12.42	11.74	10.12	10.50
$n = 50$	ET-uS	3.88	17.56	20.32	18.58	3.00	21.68	23.50	21.08	3.02	4.58	5.74	6.88
	ET-S	8.86	18.50	20.08	18.18	6.26	16.50	18.24	16.34	37.70	36.42	35.40	33.26
	un-Stud	22.22	39.20	42.46	48.32	21.80	39.72	40.84	44.80	38.30	42.10	43.38	48.08
Model 1	Stud	25.26	40.04	42.64	48.26	22.68	36.18	37.02	39.58	19.90	22.30	22.08	34.52
$n = 300$	ET-uS	6.12	33.78	42.88	47.80	4.70	34.16	41.14	44.20	11.80	20.16	25.78	35.68
	ET-S	11.98	35.82	43.26	47.90	10.70	31.94	37.62	39.20	54.10	55.86	56.40	59.96
	un-Stud	15.60	23.98	24.72	20.86	17.46	27.72	26.92	22.88	24.58	23.98	24.52	21.08
Model 2	Stud	17.90	24.24	24.72	20.64	15.70	21.30	20.72	17.80	14.40	13.10	13.16	12.90
$n = 300$	ET-uS	4.88	20.44	25.06	20.40	3.22	23.60	27.16	22.28	3.66	5.52	7.38	8.06
	ET-S	9.36	21.50	25.24	20.30	6.78	18.46	21.00	17.46	42.04	39.88	39.32	34.92

Table 2: Rejection probability under the alternative hypothesis $\beta_1 = 0$ with $\alpha = 10\%$.

regression: when the test does not employ cluster-level fixed effects, the rejection probability often exceeds the nominal level. In addition, Table 1 shows that the Rademacher weights play an important role in our results, and may not extended to other weighting schemes such as those proposed by Mammen (1993). Indeed, the rejection probability under the null hypothesis exceeds the nominal level for all values of q and n when we use these alternative weights; see the last four columns in Tables 1 and 2. We therefore do not consider these alternative weights in Tables 3 and 4.

Models 3 and 4 are heterogeneous, in the sense that Assumption 2.2(iii) is always violated and Assumption 2.2(iv) is violated if cluster-level fixed effects are not included. Table 3 shows that the rejection probability of the un-Studentized test under the null hypothesis exceeds the nominal level in nearly all specifications, including those employing cluster-level fixed effects. These results highlight the importance of Assumptions 2.2(iii)–(iv) for our results and for the reliability of the wild bootstrap when the number of clusters is small. Our findings are consistent with our theoretical results in Section 3 and simulations in Ibragimov and Müller (2016), who find that the wild bootstrap may have rejection probability under the null hypothesis greater than the nominal level whenever the dimension of the regressors is larger than two.

Stud: The Studentized test studied in Theorem 3.2 has rejection probability under the null hypothesis very close to the nominal level in Table 1 across the different specifications. Remarkably, this test seems to be less sensitive to whether clus-

	Test	Rade - with Fixed effects q 4	5	6	8	Rade - without Fixed effects q 4	5	6	8
Model 3 $n = 50$	un-Stud	11.58	13.90	13.32	13.24	26.68	37.16	32.38	26.12
	Stud	11.14	12.74	11.94	11.44	19.98	18.62	14.54	12.66
	ET-uS	5.62	10.82	12.78	12.92	8.66	31.40	33.18	25.62
	ET-S	7.06	10.24	11.34	11.38	13.52	16.08	15.10	12.46
Model 4 $n = 50$	un-Stud	12.96	17.70	16.30	12.96	12.44	22.64	18.00	14.22
	Stud	13.00	16.34	14.62	10.88	15.24	22.68	17.22	12.84
	ET-uS	5.52	14.68	16.56	12.72	3.60	19.08	18.20	14.02
	ET-S	7.62	14.30	15.10	10.76	9.60	20.70	17.66	12.74
Model 3 $n = 300$	un-Stud	12.26	15.10	13.52	12.66	30.10	39.08	33.26	26.06
	Stud	12.32	13.52	11.40	10.96	22.00	19.38	15.44	12.96
	ETuNS	5.88	12.20	14.14	12.38	14.20	32.34	16.14	12.74
	ET-S	8.20	11.86	11.94	10.74	17.80	16.70	13.00	11.98
Model 4 $n = 300$	un-Stud	13.54	17.18	15.94	12.84	14.72	24.38	17.56	13.78
	Stud	13.40	15.78	14.94	11.72	17.12	25.10	17.66	12.58
	ET-uS	5.60	13.98	16.36	12.68	4.32	19.66	17.80	13.60
	ET-S	7.88	13.38	15.46	11.56	10.42	22.16	18.14	12.36

Table 3: Rejection probability under the null hypothesis $\beta_1 = 1$ with $\alpha = 10\%$.

	Test	Rade - with Fixed effects q 4	5	6	8	Rade - without Fixed effects q 4	5	6	8
Model 1 - $n = 50$	Stud	14.76	14.26	12.96	11.26	16.60	15.28	13.80	12.42
Model 1 - $n = 300$	Stud	14.56	13.54	13.10	11.76	16.30	14.34	13.94	12.10

Table 4: Rejection probability under the null hypothesis $\beta_1 = 1$ with $\alpha = 12.5\%$.

ter level fixed effects are included in the regression or not. Nonetheless, when cluster-level fixed effects are included the rejection probability under the null hypothesis is closer to the nominal level of $\alpha = 10\%$. In the heterogeneous models of Table 3, however, the rejection probability of the Studentized test under the null hypothesis exceeds the nominal level in many of the specifications, especially when $q < 8$. Here, the inclusion of cluster-level fixed effects attenuates the amount of over-rejection. Finally, Table 2 shows that the rejection probability under the alternative hypothesis is similar to that of the un-Studentized test, except when $q = 4$ where the Studentized test exhibits higher power.

Theorem 3.2 establishes that the asymptotic size of the Studentized test does not exceed its nominal level by more than 2^{1-q}. Table 4 examines this conclusion by considering Studentized tests with nominal level $\alpha = 12.5\%$. Our simulation results shows that the rejection probability under the null hypothesis indeed exceeds the nominal level, but by an amount that is in fact smaller than 2^{1-q}. This conclusion suggests that the upper bound in Theorem 3.2 can be conservative.

ET-uS/ET-S: The equi-tailed versions of the un-Studentized and Studentized tests behave similar to their symmetric counterparts when q is not too small.

When $q \geq 6$, the rejection probability under the null and alternative hypotheses are very close to those of the un-Studentized and Studentized tests; see Tables 1-3. When $q < 6$, however, the equi-tailed versions of these tests have rejection probability under the null hypothesis below those of un-Stud and Stud. These differences in turn translate into lower power under the alternative hypothesis; see Table 2.

5 Concluding remarks

This paper has studied the properties of the wild bootstrap-based test proposed in Cameron et al. (2008) for use in settings with clustered data. In contrast to previous analyses of this test, we employ an asymptotic framework in which the number of clusters is "small," but the number of observations per cluster is "large," which coincides with the types of settings in which it is frequently being used. Our analysis highlights the importance of certain homogeneity assumptions on the distribution of covariates in ensuring that the test behaves well under the null hypothesis when there are few clusters. The practical relevance of these conditions in finite samples is confirmed via a small simulation study. It follows that when these conditions are implausible and there are few clusters, researchers may wish to consider methods that do not impose such restrictions, such as Ibragimov and Müller (2010) and Canay et al. (2017).

A Proof of Theorems

Proof of Theorem 3.1: We first introduce notation that will help streamline our argument. Let $\mathbb{S} \equiv \mathbf{R}^{d_z \times d_z} \times \bigotimes_{j \in J} \mathbf{R}^{d_z}$ and write any $s \in \mathbb{S}$ as $s = (s_1, \{s_{2,j} : j \in J\})$ where $s_1 \in \mathbf{R}^{d_z \times d_z}$ is a (real) $d_z \times d_z$ matrix, and $s_{2,j} \in \mathbf{R}^{d_z}$ for all $j \in J$. Further let $T \colon \mathbb{S} \to \mathbf{R}$ satisfy

$$T(s) \equiv |c'(s_1)^{-1}(\sum_{j \in J} s_{2,j})| \tag{A-1}$$

for any $s \in \mathbb{S}$ such that s_1 is invertible, and let $T(s) = 0$ whenever s_1 is not invertible. We also identify any $(g_1, \ldots, g_q) = g \in \mathbf{G} = \{-1, 1\}^q$ with an action on $s \in \mathbb{S}$ given by $gs = (s_1, \{g_j s_{2,j} : j \in J\})$. For any $s \in \mathbb{S}$ and $\mathbf{G}' \subseteq \mathbf{G}$, denote the ordered values of $\{T(gs) : g \in \mathbf{G}'\}$ by

$$T^{(1)}(s|\mathbf{G}') \leq \cdots \leq T^{(|\mathbf{G}'|)}(s|\mathbf{G}').$$

Next, let $(\hat{\gamma}_n', \hat{\beta}_n')'$ be the least squares estimators of $(\gamma', \beta')'$ in (1) and recall that $\hat{\epsilon}^{\mathrm{r}}_{i,j} \equiv (Y_{i,j} - W_{i,j}'\hat{\gamma}^{\mathrm{r}}_n - Z_{i,j}'\hat{\beta}^{\mathrm{r}}_n)$, where $(\hat{\gamma}^{\mathrm{r}\prime}_n, \hat{\beta}^{\mathrm{r}\prime}_n)'$ are the constrained least squares estimators of the same parameters restricted to satisfy $c'\hat{\beta}^{\mathrm{r}}_n = \lambda$. By the Frisch-Waugh-Lovell theorem, $\hat{\beta}_n$ can be obtained by regressing $Y_{i,j}$ on $\tilde{Z}_{i,j}$, where $\tilde{Z}_{i,j}$ is the residual from the projection of $Z_{i,j}$ on $W_{i,j}$ defined in (8). Using this notation we can define the statistics $S_n, S_n^* \in \mathbb{S}$ to be given by

$$S_n \equiv \Big(\hat{\Omega}_{\tilde{Z},n}, \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\epsilon_{i,j} : j \in J\}\Big) \tag{A-2}$$

$$S_n^* \equiv \Big(\hat{\Omega}_{\tilde{Z},n}, \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\hat{\epsilon}^{\mathrm{r}}_{i,j} : j \in J\}\Big), \tag{A-3}$$

where

$$\hat{\Omega}_{\tilde{Z},n} \equiv \frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\tilde{Z}_{i,j}'. \tag{A-4}$$

Next, let E_n denote the event $E_n \equiv I\{\hat{\Omega}_{\tilde{Z},n}$ is invertible$\}$, and note that whenever $E_n = 1$ and $c'\beta = \lambda$, the Frisch-Waugh-Lovell theorem implies that

$$|\sqrt{n}(c'\hat{\beta}_n - \lambda)| = |\sqrt{n}c'(\hat{\beta}_n - \beta)| = |c'\hat{\Omega}^{-1}_{\tilde{Z},n} \sum_{j \in J} \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\epsilon_{i,j}| = T(S_n). \tag{A-5}$$

Moreover, by identical arguments it also follows that for any action $g \in \mathbf{G}$ we similarly have

$$|\sqrt{n}c'(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n)| = |c'\hat{\Omega}^{-1}_{\tilde{Z},n} \sum_{j \in J} \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} g_j \tilde{Z}_{i,j}\hat{\epsilon}^{\mathrm{r}}_{i,j}| = T(gS_n^*) \tag{A-6}$$

whenever $E_n = 1$. Therefore, for any $x \in \mathbf{R}$ letting $\lceil x \rceil$ denote the smallest integer larger than x and $k^* \equiv \lceil |\mathbf{G}|(1-\alpha) \rceil$, we obtain from (A-5) and (A-6) that

$$I\{T_n > \hat{c}_n(1-\alpha);\ E_n = 1\} = I\{T(S_n) > T^{(k^*)}(S_n^*|\mathbf{G});\ E_n = 1\}. \tag{A-7}$$

In addition, it follows from Assumptions 2.2(ii)-(iii) that $\hat{\Omega}_{\tilde{Z},n} \xrightarrow{P} \bar{a}\Omega_{\tilde{Z}}$, where $\bar{a} \equiv \sum_{j \in J} \xi_j a_j > 0$

and $\Omega_{\tilde{Z}}$ is a $d_z \times d_z$ invertible matrix. Hence, we may conclude that

$$\liminf_{n\to\infty} P\{E_n = 1\} = 1. \tag{A-8}$$

Further let $\iota \in \mathbf{G}$ correspond to the identity action, e.g. $\iota \equiv (1, \ldots, 1) \in \mathbf{R}^q$, and similarly define $-\iota \equiv (-1, \ldots, -1) \in \mathbf{R}^q$. Then note that since $T(-\iota S_n^*) = T(\iota S_n^*)$, we can conclude from (A-3) and $\hat{\epsilon}^{\mathrm{r}}_{i,j} = (Y_{i,j} - W_{i,j}'\hat{\gamma}_n^{\mathrm{r}} - Z_{i,j}'\hat{\beta}_n^{\mathrm{r}})$ that whenever $E_n = 1$ we obtain

$$\begin{aligned} T(-\iota S_n^*) = T(\iota S_n^*) &= \Big|c'\hat{\Omega}_{\tilde{Z},n}^{-1}\sum_{j\in J}\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}(Y_{i,j} - W_{i,j}'\hat{\gamma}_n^{\mathrm{r}} - Z_{i,j}'\hat{\beta}_n^{\mathrm{r}})\Big| \\ &= \Big|c'\hat{\Omega}_{\tilde{Z},n}^{-1}\sum_{j\in J}\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}(Y_{i,j} - \tilde{Z}_{i,j}'\hat{\beta}_n^{\mathrm{r}})\Big| = |\sqrt{n}c'(\hat{\beta}_n - \hat{\beta}_n^{\mathrm{r}})| = T(S_n), \end{aligned} \tag{A-9}$$

where the third equality follows from $\sum_{j\in J}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W_{i,j}' = 0$ due to $\tilde{Z}_{i,j} \equiv (Z_{i,j} - \hat{\Pi}_n' W_{i,j})$ and the definition of $\hat{\Pi}_n$ (see (7)). In turn, the fourth equality in (A-9) follows from (A-4) and the Frisch-Waugh-Lovell theorem as in (A-5), while the final result in (A-9) is implied by $c'\hat{\beta}_n^{\mathrm{r}} = \lambda$ and (A-5). In particular, (A-9) implies that if $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil > |\mathbf{G}| - 2$, then $I\{T(S_n) > T^{(k^*)}(S_n^*|\mathbf{G}); E_n = 1\} = 0$, which establishes the upper bound in Theorem 3.1 due to (A-7) and (A-8). We therefore assume that $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil \le |\mathbf{G}| - 2$, in which case

$$\begin{aligned} \limsup_{n\to\infty} E[\phi_n] &= \limsup_{n\to\infty} P\{T(S_n) > T^{(k^*)}(S_n^*|\mathbf{G});\ E_n = 1\} \\ &= \limsup_{n\to\infty} P\{T(S_n) > T^{(k^*)}(S_n^*|\mathbf{G}\setminus\{\pm\iota\});\ E_n = 1\} \\ &\le \limsup_{n\to\infty} P\{T(S_n) \ge T^{(k^*)}(S_n^*|\mathbf{G}\setminus\{\pm\iota\});\ E_n = 1\}, \end{aligned} \tag{A-10}$$

where the first equality follows from (A-7) and (A-8), the second equality is implied by (A-9) and $k^* \le |\mathbf{G}| - 2$, and the final inequality follows by set inclusion.

To examine the right hand side of (A-10), we first note that Assumptions 2.2(i)-(ii) and the continuous mapping theorem imply that

$$\Big\{\frac{\sqrt{n_j}}{\sqrt{n}}\frac{1}{\sqrt{n_j}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\epsilon_{i,j} : j\in J\Big\} \xrightarrow{L} \{\sqrt{\xi_j}\mathcal{Z}_j : j \in J\}. \tag{A-11}$$

Since $\xi_j > 0$ for all $j \in J$ by Assumption 2.1(ii), and the variables $\{\mathcal{Z}_j : j \in J\}$ have full rank covariance matrices by Assumption 2.1(i), it follows that $\{\sqrt{\xi_j}\mathcal{Z}_j : j \in J\}$ have full rank covariance matrices as well. Combining (A-11) together with the definition of S_n in (A-2) and the previously shown result $\hat{\Omega}_{\tilde{Z},n} \xrightarrow{P} \bar{a}\Omega_{\tilde{Z}}$ then allows us to establish

$$S_n \xrightarrow{L} S \equiv \Big(\bar{a}\Omega_{\tilde{Z}}, \{\sqrt{\xi_j}\mathcal{Z}_j : j \in J\}\Big). \tag{A-12}$$

We further note that whenever $E_n = 1$, the definition of S_n and S_n^* in (A-2) and (A-3), together with the triangle inequality, yield for every $g \in \mathbf{G}$ an upper bound of the form

$$|T(gS_n) - T(gS_n^*)| \leq |c'\hat{\Omega}_{\tilde{Z},n}^{-1} \sum_{j\in J} \frac{n_j}{n}\frac{1}{n_j} \sum_{i\in I_{n,j}} g_j \tilde{Z}_{i,j} Z'_{i,j}\sqrt{n}(\beta - \hat{\beta}_n^{\mathrm{r}})| \\ + |c'\hat{\Omega}_{\tilde{Z},n}^{-1} \sum_{j\in J} \frac{n_j}{n}\frac{1}{n_j} \sum_{i\in I_{n,j}} g_j \tilde{Z}_{i,j} W'_{i,j}\sqrt{n}(\gamma - \hat{\gamma}_n^{\mathrm{r}})|. \quad \text{(A-13)}$$

In what follows, we aim to employ (A-13) to establish that $T(gS_n) = T(gS_n^*) + o_P(1)$. To this end, note that whenever $c'\beta = \lambda$ 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. Thus, Lemma A.2 implies that

$$\limsup_{n\to\infty} P\{|c'\hat{\Omega}_{\tilde{Z},n}^{-1} \sum_{j\in J} \frac{n_j}{n}\frac{1}{n_j} \sum_{i\in I_{n,j}} g_j \tilde{Z}_{i,j} W'_{i,j}\sqrt{n}(\gamma - \hat{\gamma}_n^{\mathrm{r}})| > \epsilon;\ E_n = 1\} = 0 \quad \text{(A-14)}$$

for any $\epsilon > 0$. Moreover, Lemma A.2 and Assumptions 2.2(ii)-(iii) establish for any $\epsilon > 0$ that

$$\limsup_{n\to\infty} P\{|c'\hat{\Omega}_{\tilde{Z},n}^{-1} \sum_{j\in J} \frac{n_j}{n}\frac{1}{n_j} \sum_{i\in I_{n,j}} g_j \tilde{Z}_{i,j} Z'_{i,j}\sqrt{n}(\beta - \hat{\beta}_n^{\mathrm{r}})| > \epsilon;\ E_n = 1\} \\ = \limsup_{n\to\infty} P\{|c'\hat{\Omega}_{\tilde{Z},n}^{-1} \sum_{j\in J} \frac{n_j}{n}\frac{1}{n_j} \sum_{i\in I_{n,j}} g_j \tilde{Z}_{i,j} \tilde{Z}'_{i,j}\sqrt{n}(\beta - \hat{\beta}_n^{\mathrm{r}})| > \epsilon;\ E_n = 1\} \\ = \limsup_{n\to\infty} P\{|c'\Omega_{\tilde{Z}}^{-1} \sum_{j\in J} \frac{\xi_j g_j a_j}{\bar{a}} \Omega_{\tilde{Z}}\sqrt{n}(\beta - \hat{\beta}_n^{\mathrm{r}})| > \epsilon;\ E_n = 1\}, \quad \text{(A-15)}$$

where recall $\bar{a} \equiv \sum_{j\in J} \xi_j a_j$. Hence, if $c'\beta = \lambda$, then (A-15) and $c'\hat{\beta}_n^{\mathrm{r}} = \lambda$ yield for any $\epsilon > 0$

$$\limsup_{n\to\infty} P\{|c'\hat{\Omega}_{\tilde{Z},n}^{-1} \sum_{j\in J} \frac{n_j}{n}\frac{1}{n_j} \sum_{i\in I_{n,j}} g_j \tilde{Z}_{i,j} Z'_{i,j}\sqrt{n}(\beta - \hat{\beta}_n^{\mathrm{r}})| > \epsilon;\ E_n = 1\} \\ = \limsup_{n\to\infty} P\{|\sum_{j\in J} \frac{\xi_j g_j a_j}{\bar{a}} \sqrt{n}(c'\beta - c'\hat{\beta}_n^{\mathrm{r}})| > \epsilon; E_n = 1\} = 0. \quad \text{(A-16)}$$

Since we had defined $T(s) = 0$ for any $s = (s_1, \{s_{2,j} : j \in J\})$ whenever s_1 is not invertible, it follows that $T(gS_n^*) = T(gS_n)$ whenever $E_n = 0$. Therefore, results (A-13), (A-14), and (A-16) imply $T(gS_n^*) = T(gS_n) + o_P(1)$ for any $g \in \mathbf{G}$. We thus obtain from result (A-12) that

$$(T(S_n), \{T(gS_n^*) : g \in \mathbf{G}\}) \xrightarrow{L} (T(S), \{T(gS) : g \in \mathbf{G}\}) \quad \text{(A-17)}$$

due to the continuous mapping theorem. Moreover, since $E_n \xrightarrow{P} 1$ by result (A-8), it follows that $(T(S_n), E_n, \{T(gS_n^*) : g \in \mathbf{G}\})$ converge jointly as well. Hence, Portmanteau's theorem, see e.g. Theorem 1.3.4(iii) in van der Vaart and Wellner (1996), implies

$$\limsup_{n\to\infty} P\{T(S_n) \geq T^{(k^*)}(S_n^*|\mathbf{G} \setminus \{\pm\iota\});\ E_n = 1\} \\ \leq P\{T(S) \geq T^{(k^*)}(gS|\mathbf{G} \setminus \{\pm\iota\})\} = P\{T(S) > T^{(k^*)}(gS|\mathbf{G} \setminus \{\pm\iota\})\}, \quad \text{(A-18)}$$

where in the equality we exploited that $P\{T(S) = T(gS)\} = 0$ for all $g \in \mathbf{G} \setminus \{\pm\iota\}$ since the covariance matrix of $\mathcal{Z}_j$ is full rank for all $j \in J$ and $\Omega_{\tilde{Z}}$ is nonsingular by Assumption 2.2(iii). Finally, noting that $T(\iota S) = T(-\iota S) = T(S)$, we can conclude $T(S) > T^{(k^*)}(gS|\mathbf{G} \setminus \{\pm\iota\})$ if and only if $T(S) > T^{(k^*)}(gS|\mathbf{G})$, which together with (A-10) and (A-18) yields

$$\limsup_{n\to\infty} E[\phi_n] \leq P\{T(S) > T^{(k^*)}(gS|\mathbf{G} \setminus \{\pm\iota\})\} = P\{T(S) > T^{(k^*)}(gS|\mathbf{G})\} \leq \alpha, \tag{A-19}$$

where the final inequality follows by $gS \stackrel{d}{=} S$ for all $g \in \mathbf{G}$ and the properties of randomization tests, see e.g. Theorem 15.2.1 in Lehmann and Romano (2005). This completes the proof of the upper bound in the statement of the Theorem.

For the lower bound, first note that $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil > |\mathbf{G}| - 2$ implies that $\alpha - \frac{1}{2^{q-1}} \leq 0$, in which case the result trivially follows. Assume $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil \leq |\mathbf{G}| - 2$ and note that

$$\begin{aligned}
\limsup_{n\to\infty} E[\phi_n] &\geq \liminf_{n\to\infty} P\{T(S_n) > T^{(k^*+1)}(S_n^*|\mathbf{G});\ E_n = 1\} \\
&\geq P\{T(S) > T^{(k^*+1)}(gS|\mathbf{G})\} \\
&\geq P\{T(S) > T^{(k^*+2)}(gS|\mathbf{G})\} + \tau P\{T(S) = T^{(k^*+2)}(gS|\mathbf{G})\} \\
&= \alpha - \frac{1}{2^{q-1}},
\end{aligned} \tag{A-20}$$

where the first inequality follows from result (A-7) and $T^{(k^*+1)}(gS|\mathbf{G}) \geq T^{(k^*)}(gS|\mathbf{G})$, the second inequality follows from Portmanteau's theorem, see e.g. Theorem 1.3.4(iii) in van der Vaart and Wellner (1996), the third inequality holds for any $\tau \in [0,1]$ due to $T^{(k^*+2)}(gS|\mathbf{G}) \geq T^{(k^*+1)}(gS|\mathbf{G})$, and the last equality follows from noticing that $k^*+2 = \lceil |\mathbf{G}|((1-\alpha)+2/|\mathbf{G}|)\rceil = \lceil |\mathbf{G}|(1-\alpha')\rceil$ with $\alpha' = \alpha - \frac{1}{2^{q-1}}$ and the properties of randomization tests (see, e.g., Lehmann and Romano, 2005, Theorem 15.2.1) together with setting τ equal to

$$\tau = \frac{|\mathbf{G}|\alpha' - M^+(S)}{M^0(S)},$$

where

$$\begin{aligned}
M^+(S) &= |\{1 \leq j \leq |\mathbf{G}| : T^{(j)}(gS|\mathbf{G}) > T^{(k^*+2)}(gS|\mathbf{G})\}| \\
M^0(S) &= |\{1 \leq j \leq |\mathbf{G}| : T^{(j)}(gS|\mathbf{G}) = T^{(k^*+2)}(gS|\mathbf{G})\}|.
\end{aligned}$$

Thus, the lower bound holds and the claim of the Theorem follows. ■

Proof of Theorem 3.2: The proof follows similar arguments as those employed in establishing Theorem 3.1, and thus we keep exposition more concise. We again start by introducing notation that will streamline our arguments. Let $\mathbb{S} \equiv \mathbf{R}^{d_z \times d_z} \times \bigotimes_{j\in J} \mathbf{R}^{d_z}$ and write an element $s \in \mathbb{S}$ by $s = (s_1, \{s_{2,j} : j \in J\})$ where $s_1 \in \mathbf{R}^{d_z \times d_z}$ is a (real) $d_z \times d_z$ matrix, and $s_{2,j} \in \mathbf{R}^{d_z}$ for any $j \in J$. Further define the functions $T\colon \mathbb{S} \to \mathbf{R}$ and $W\colon \mathbb{S} \to \mathbf{R}$ to be pointwise given by

$$T(s) \equiv |c'(s_1)^{-1}(\sum_{j\in J} s_{2,j}) - \lambda| \tag{A-21}$$

$$W(s) \equiv \Big(c'(s_1)^{-1} \sum_{j\in J}\Big(s_{2,j} - \frac{\xi_j a_j}{\bar{a}} \sum_{\tilde{j}\in J} s_{2,\tilde{j}}\Big)\Big(s_{2,j} - \frac{\xi_j a_j}{\bar{a}} \sum_{\tilde{j}\in J} s_{2,\tilde{j}}\Big)'(s_1)^{-1}c\Big)^{1/2}, \tag{A-22}$$

for any $s \in \mathbb{S}$ such that s_1 is invertible, and set $T(s) = 0$ and $W(s) = 1$ whenever s_1 is not invertible. We further identify any $(g_1, \ldots, g_q) = g \in \mathbf{G} = \{-1,1\}^q$ with an action on $s \in \mathbb{S}$

defined by $gs = (s_1, \{g_j s_{2,j} : j \in J\})$. Finally, we set $A_n \in \mathbf{R}$ and $S_n \in \mathbb{S}$ to equal

$$A_n \equiv I\{\hat{\Omega}_{\tilde{Z},n} \text{ is invertible, } \hat{\sigma}_n > 0, \text{ and } \hat{\sigma}_n^*(g) > 0 \text{ for all } g \in \mathbf{G}\} \tag{A-23}$$

$$S_n \equiv \Big(\hat{\Omega}_{\tilde{Z},n}, \{\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\epsilon_{i,j} : j \in J\}\Big) \tag{A-24}$$

where recall $\hat{\Omega}_{\tilde{Z},n}$ was defined in (15) and $\tilde{Z}_{i,j}$ was defined in (8).

First, note that by Assumptions 2.2(i)-(ii) and the continuous mapping theorem we obtain

$$\Big\{\frac{\sqrt{n_j}}{\sqrt{n}} \frac{1}{\sqrt{n_j}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\epsilon_{i,j} : j \in J\Big\} \xrightarrow{L} \{\sqrt{\xi_j}\mathcal{Z}_j : j \in J\}. \tag{A-25}$$

Since $\xi_j > 0$ for all $j \in J$ by Assumption 2.2(ii), and the variables $\{\mathcal{Z}_j : j \in J\}$ have full rank covariance matrices by Assumption 2.2(i), it follows that $\{\sqrt{\xi_j}\mathcal{Z}_j : j \in J\}$ have full rank covariance matrices as well. Combining (A-25) together with the definition of S_n in (A-24), Assumption 2.2(ii)-(iii), and the continuous mapping theorem then allows us to establish

$$S_n \xrightarrow{L} S \equiv \Big(\bar{a}\Omega_{\tilde{Z}}, \{\sqrt{\xi_j}\mathcal{Z}_j : j \in J\}\Big), \tag{A-26}$$

where $\bar{a} \equiv \sum_{j \in J} \xi_j a_j > 0$. Since $\Omega_{\tilde{Z}}$ is invertible by Assumption 2.2(iii) and $\bar{a} > 0$, it follows that $\hat{\Omega}_{\tilde{Z},n}$ is invertible with probability tending to one. Hence, we can conclude that

$$\hat{\sigma}_n = W(S_n) + o_P(1) \qquad\qquad \hat{\sigma}_n^*(g) = W(gS_n) + o_P(1) \tag{A-27}$$

due to the definition of $W\colon \mathbb{S} \to \mathbf{R}$ in (A-22) and Lemma A.1. Moreover, $\hat{\Omega}_{\tilde{Z},n}$ being invertible with probability tending to one additionally allows us to conclude that

$$\begin{aligned}\liminf_{n\to\infty} P\{A_n = 1\} &= \liminf_{n\to\infty} P\{\hat{\sigma}_n > 0 \text{ and } \hat{\sigma}_n^*(g) > 0 \text{ for all } g \in \mathbf{G}\} \\ &\geq P\{W(gS) > 0 \text{ for all } g \in \mathbf{G}\} = 1, \end{aligned} \tag{A-28}$$

where the inequality in (A-28) holds by (A-26), (A-27), the continuous mapping theorem, and Portmanteau's Theorem; see, e.g., Theorem 1.3.4(ii) in van der Vaart and Wellner (1996). In turn, the final equality in (A-28) follows from $\{\sqrt{\xi_j}\mathcal{Z}_j : j \in J\}$ being independent and continuously distributed with covariance matrices that are full rank.

Next, recall that $\hat{\epsilon}^{\mathrm{r}}_{i,j} = (Y_{i,j} - W'_{i,j}\hat{\gamma}^{\mathrm{r}}_n - Z'_{i,j}\hat{\beta}^{\mathrm{r}}_n)$ and note that whenever $A_n = 1$ we obtain

$$\begin{aligned}\sqrt{n}c'(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n) &= c'\hat{\Omega}^{-1}_{\tilde{Z},n}\frac{1}{\sqrt{n}}\sum_{j \in J}\sum_{i \in I_{n,j}} g_j\tilde{Z}_{i,j}\hat{\epsilon}^{\mathrm{r}}_{i,j} \\ &= c'\hat{\Omega}^{-1}_{\tilde{Z},n}\frac{1}{\sqrt{n}}\sum_{j \in J}\sum_{i \in I_{n,j}} g_j\tilde{Z}_{i,j}(\epsilon_{i,j} - W'_{i,j}(\hat{\gamma}^{\mathrm{r}}_n - \gamma) - Z'_{i,j}(\hat{\beta}^{\mathrm{r}}_n - \beta)).\end{aligned} \tag{A-29}$$

Further note that $c'\beta = \lambda$, Assumption 2.1, and Amemiya (1985, Eq. (1.4.5)) together imply that $\sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta)$ and $\sqrt{n}(\hat{\gamma}^{\mathrm{r}}_n - \gamma)$ are bounded in probability. Therefore, Lemma A.2 implies

$$\limsup_{n\to\infty} P\{|c'\hat{\Omega}^{-1}_{\tilde{Z},n}\sum_{j \in J}\frac{g_j}{n}\sum_{i \in I_{n,j}} \tilde{Z}_{i,j}W'_{i,j}\sqrt{n}(\hat{\gamma}^{\mathrm{r}}_n - \gamma)| > \epsilon;\ A_n = 1\} = 0 \tag{A-30}$$

for any $\epsilon > 0$. Similarly, since $\sqrt{n}(\hat\beta_n^{\rm r} - \beta)$ is bounded in probability and $\Omega_{\tilde Z}$ is invertible by Assumption 2.2(iii), Lemma A.2 together with Assumptions 2.2(ii)-(iii) imply for any $\epsilon > 0$

$$\begin{aligned}
&\limsup_{n\to\infty} P\{|c'\hat\Omega_{\tilde Z,n}^{-1}\sum_{j\in J}\frac{n_j}{n}\frac{1}{n_j}g_j\sum_{i\in I_{n,j}}\tilde Z_{i,j}Z_{i,j}'\sqrt{n}(\hat\beta_n^{\rm r}-\beta)| > \epsilon;\ A_n = 1\} \\
&\quad = \limsup_{n\to\infty} P\{|c'\hat\Omega_{\tilde Z,n}^{-1}\sum_{j\in J}\frac{n_j}{n}\frac{1}{n_j}g_j\sum_{i\in I_{n,j}}\tilde Z_{i,j}\tilde Z_{i,j}'\sqrt{n}(\hat\beta_n^{\rm r}-\beta)| > \epsilon;\ A_n = 1\} \\
&\quad = \limsup_{n\to\infty} P\{|c'\Omega_{\tilde Z}^{-1}\sum_{j\in J}\frac{\xi_j a_j g_j}{\bar a}\sum_{i\in I_{n,j}}\Omega_{\tilde Z}\sqrt{n}(\hat\beta_n^{\rm r}-\beta)| > \epsilon;\ A_n = 1\}. \qquad \text{(A-31)}
\end{aligned}$$

It follows from results (A-27)-(A-31) together with $T(S_n) = T_n$ whenever $\hat\Omega_{\tilde Z,n}$ is invertible, that

$$\begin{aligned}
&((|\sqrt{n}(c'\hat\beta_n - \lambda)|, \hat\sigma_n), \{(|c'\sqrt{n}(\hat\beta_n^*(g) - \hat\beta_n^{\rm r})|, \hat\sigma_n^*(g)) : g \in \mathbf{G}\}) \\
&\qquad = ((T(S_n), W(S_n)), \{(T(gS_n), W(gS_n)) : g \in \mathbf{G}\}) + o_P(1). \qquad \text{(A-32)}
\end{aligned}$$

To conclude, we define a function $\mathrm{t}\colon \mathbb{S} \to \mathbf{R}$ to be given by $\mathrm{t}(s) = T(s)/W(s)$. Then note that, for any $g \in \mathbf{G}$, gS assigns probability one to the continuity points of $\mathrm{t} : \mathbb{S} \to \mathbf{R}$ since $\Omega_{\tilde Z}$ is invertible and $P\{W(gS) > 0 \text{ for all } g \in \mathbf{G}\} = 1$ as argued in (A-28). In what follows, for any $s \in \mathbb{S}$ it will prove helpful to employ the ordered values of $\{\mathrm{t}(gs) : g \in \mathbf{G}\}$, which we denote by

$$\mathrm{t}^{(1)}(s|\mathbf{G}) \le \ldots \le \mathrm{t}^{(|\mathbf{G}|)}(s|\mathbf{G}). \qquad \text{(A-33)}$$

Next, we observe that result (A-28) and a set inclusion inequality allow us to conclude that

$$\begin{aligned}
\limsup_{n\to\infty} P\Big\{\frac{T_n}{\hat\sigma_n} > \hat c_n^{\rm s}(1-\alpha)\Big\} &\le \limsup_{n\to\infty} P\Big\{\frac{T_n}{\hat\sigma_n} \ge \hat c_n^{\rm s}(1-\alpha);\ A_n = 1\Big\} \\
&\le P\left\{\mathrm{t}(S) \ge \inf\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{\mathrm{t}(gS) \le u\} \ge 1-\alpha\}\right\}, \qquad \text{(A-34)}
\end{aligned}$$

where the final inequality follows by results (A-26), (A-32), and the continuous mapping and Portmanteau theorems; see, e.g., Theorem 1.3.4(iii) in van der Vaart and Wellner (1996). Therefore, setting $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil$, we can then obtain from result (A-34) that

$$\begin{aligned}
&\limsup_{n\to\infty} P\Big\{\frac{T_n}{\hat\sigma_n} > \hat c_n^{\rm s}(1-\alpha)\Big\} \\
&\qquad \le P\{\mathrm{t}(S) > \mathrm{t}^{(k^*)}(S)\} + P\{\mathrm{t}(S) = \mathrm{t}^{(k^*)}(S)\} \le \alpha + P\{\mathrm{t}(S) = \mathrm{t}^{(k^*)}(S)\}, \qquad \text{(A-35)}
\end{aligned}$$

where in the final inequality we exploited 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

$$\begin{aligned}
&P\{\mathrm{t}(S) = \mathrm{t}^{(k^*)}(S)\} \\
&\qquad = E[P\{\mathrm{t}(S) = \mathrm{t}^{k^*}(S)|S \in \{gS\}_{g\in\mathbf{G}}\}] = E[\frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{\mathrm{t}(gS) = \mathrm{t}^{(k^*)}(S)\}]. \qquad \text{(A-36)}
\end{aligned}$$

For any $g = (g_1, \ldots, g_q) \in \mathbf{G}$ then let $-g = (-g_1, \ldots, -g_q) \in \mathbf{G}$ and note that $\mathrm{t}(gS) = \mathrm{t}(-gS)$ with probability one. However, if $\tilde{g}, g \in \mathbf{G}$ are such that $\tilde{g} \notin \{g, -g\}$, then

$$P\{\mathrm{t}(gS) = \mathrm{t}(\tilde{g}S)\} = 0 \tag{A-37}$$

since, by Assumption 2.2, $S = (\bar{a}\Omega_{\tilde{Z}}, \{\sqrt{\xi_j}\mathcal{Z}_j : j \in J\})$ is such that $\Omega_{\tilde{Z}}$ is invertible, $\xi_j > 0$ for all $j \in J$, and $\{\mathcal{Z}_j : j \in J\}$ are independent with full rank covariance matrices. Hence,

$$\frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{\mathrm{t}(gS) = \mathrm{t}^{(k^*)}(S)\} = \frac{1}{|\mathbf{G}|} \times 2 = \frac{1}{2^{q-1}} \tag{A-38}$$

with probability one, and where in the final equality we exploited that $|\mathbf{G}| = 2^q$. The claim of the upper bound in the Theorem therefore follows from results (A-35) and (A-38). Finally, the lower bound follows from similar arguments to those in (A-20) and so we omit them here. ■

Lemma A.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 $(g_1, \ldots, g_q) = g \in \mathbf{G}$*

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

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

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, we can conclude from Lemma A.2 and the definition of $U_{n,j}$ that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j} &= \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\epsilon_{i,j} - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}Z'_{i,j}\sqrt{n}(\hat{\beta}_n - \beta) - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\sqrt{n}(\hat{\gamma}_n - \gamma)\\ &= U_{n,j} - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\sqrt{n}(\hat{\beta}_n - \beta) + o_P(1).\end{aligned} \tag{A-39}$$

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

$$\begin{aligned}&\frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\sqrt{n}(\hat{\beta}_n - \beta)\\ &\quad = \frac{n_j}{n}\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\hat{\Omega}^{-}_{\tilde{Z},n}\frac{1}{\sqrt{n}}\sum_{\tilde{j}\in J}\sum_{k\in I_{n,\tilde{j}}}\tilde{Z}_{k,\tilde{j}}\epsilon_{k,\tilde{j}} + o_P(1) = \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}U_{n,\tilde{j}} + o_P(1).\end{aligned} \tag{A-40}$$

Therefore, (A-39), (A-40), and the continuous mapping theorem yield

$$\hat{V}_n = \sum_{j\in J}\Big(\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}\Big)\Big(\frac{1}{\sqrt{n}}\sum_{k\in I_{n,j}}\tilde{Z}'_{k,j}\hat{\epsilon}_{k,j}\Big)$$
$$= \sum_{j\in J}\Big(U_{n,j} - \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}U_{n,\tilde{j}}\Big)\Big(U_{n,j} - \frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}U_{n,\tilde{j}}\Big)' + o_P(1). \quad \text{(A-41)}$$

The first part of the lemma thus follows by the definition of $\hat{\sigma}_n^2$ in (16).

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 $\hat{\epsilon}_{i,j}^*(g) = g_j\hat{\epsilon}_{i,j}^{\mathrm{r}} - Z'_{i,j}(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathrm{r}}) - W'_{i,j}(\hat{\gamma}_n^*(g) - \hat{\gamma}_n^{\mathrm{r}})$ 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

$$\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}^*(g)$$
$$= \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}g_j\hat{\epsilon}_{i,j}^{\mathrm{r}} - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}Z'_{i,j}\sqrt{n}(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathrm{r}}) - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\sqrt{n}(\hat{\gamma}_n^*(g) - \hat{\gamma}_n^{\mathrm{r}})$$
$$= \frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}g_j\hat{\epsilon}_{i,j}^{\mathrm{r}} - \frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\sqrt{n}(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathrm{r}}) + o_P(1), \quad \text{(A-42)}$$

where in the final equality we employed Lemma A.2. Next, recall $\hat{\epsilon}_{i,j}^{\mathrm{r}} \equiv \epsilon_{i,j} - Z'_{i,j}(\hat{\beta}_n^{\mathrm{r}} - \beta) - W'_{i,j}(\hat{\gamma}_n^{\mathrm{r}} - \gamma)$ and note

$$c'\hat{\Omega}_{\tilde{Z},n}^{-}\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}g_j\hat{\epsilon}_{i,j}^{\mathrm{r}} = c'\hat{\Omega}_{\tilde{Z},n}^{-}\frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}g_j(\epsilon_{i,j} - Z'_{i,j}\sqrt{n}(\hat{\beta}_n^{\mathrm{r}} - \beta) - W'_{i,j}\sqrt{n}(\hat{\gamma}_n^{\mathrm{r}} - \gamma))$$
$$= c'\hat{\Omega}_{\tilde{Z},n}^{-}g_jU_{n,j} - c'\hat{\Omega}_{\tilde{Z},n}^{-}\frac{1}{n}\sum_{i\in I_{n,j}}g_j\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\sqrt{n}(\hat{\beta}_n^{\mathrm{r}} - \beta) + o_P(1), \quad \text{(A-43)}$$

where the second equality follows from Lemma A.2 and $\hat{\Omega}_{\tilde{Z},n}^{-}$, $\sqrt{n}(\hat{\beta}_n^{\mathrm{r}} - \beta)$, and $\sqrt{n}(\hat{\gamma}_n^{r} - \gamma)$ being bounded in probability. Moreover, Assumptions 2.2(ii)-(iii) imply

$$c'\hat{\Omega}_{\tilde{Z},n}^{-}\frac{1}{n}\sum_{i\in I_{n,j}}g_j\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\sqrt{n}(\hat{\beta}_n^{\mathrm{r}} - \beta) = c'\Omega_{\tilde{Z}}^{-1}\frac{g_j\xi_j a_j}{\bar{a}}\Omega_{\tilde{Z}}\sqrt{n}(\hat{\beta}_n^{\mathrm{r}} - \beta) + o_P(1) = o_P(1), \quad \text{(A-44)}$$

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 $\hat{\Omega}_{\tilde{Z},n}^{-} = \hat{\Omega}_{\tilde{Z},n}^{-1}$ 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

$$\begin{aligned}
c'\hat{\Omega}^{-}_{\tilde{Z},n}\frac{1}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\sqrt{n}(\hat{\beta}^{*}_{n}(g)-\hat{\beta}^{\mathrm{r}}_{n}) \\
&= c'\hat{\Omega}^{-}_{\tilde{Z},n}\frac{n_j}{n}\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\hat{\Omega}^{-}_{\tilde{Z},n}\sum_{\tilde{j}\in J}\frac{1}{\sqrt{n}}\sum_{k\in I_{n,\tilde{j}}}\tilde{Z}_{k,j}g_{\tilde{j}}\hat{\epsilon}^{\mathrm{r}}_{k,\tilde{j}}+o_P(1) \\
&= c'\hat{\Omega}^{-}_{\tilde{Z},n}\frac{\xi_j a_j}{\bar{a}}\sum_{\tilde{j}\in J}g_{\tilde{j}}U_{n,\tilde{j}}+o_P(1), \qquad \text{(A-45)}
\end{aligned}$$

where in the final equality we applied (A-43), (A-44), 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 (17) and results (A-42)-(A-45). ■

Lemma A.2. *Let Assumptions 2.1(ii) and 2.2(iv) hold. It follows that for any $j\in J$ we have*

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

Proof: Let $\|\cdot\|_F$ denote the Frobenius matrix norm, which recall equals $\|M\|^2_F\equiv\mathrm{trace}\{M'M\}$ 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 $\hat{\Pi}^{\mathrm{c}}_{n,j}$ (see (9)), and the triangle inequality applied to $\|\cdot\|_F$, we then obtain

$$\begin{aligned}
\|\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\|_F &= \|\frac{1}{n_j}\sum_{i\in I_{n,j}}(Z_{i,j}-\hat{\Pi}'_nW_{i,j})W'_{i,j}\|_F \\
&= \|\frac{1}{n_j}\sum_{i\in I_{n,j}}(\hat{\Pi}^{\mathrm{c}}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F \le \frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{\mathrm{c}}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F. \qquad \text{(A-46)}
\end{aligned}$$

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

$$\begin{aligned}
\frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{\mathrm{c}}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F &= \frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{\mathrm{c}}_{n,j}-\hat{\Pi}_n)'W_{i,j}\|\times\|W'_{i,j}W_{i,j}\| \\
&\le \{\frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{\mathrm{c}}_{n,j}-\hat{\Pi}_n)'W_{i,j}\|^2\}^{1/2}\times\{\frac{1}{n_j}\sum_{i\in I_{n,j}}\|W_{i,j}\|^2\}^{1/2}=o_P(1), \qquad \text{(A-47)}
\end{aligned}$$

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

$$\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}Z'_{i,j}=\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j}+\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\hat{\Pi}_n \qquad \text{(A-48)}$$

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

References

Acemoglu, D., Cantoni, D., Johnson, S. and Robinson, J. A. (2011). The consequences of radical reform: The french revolution. *American Economic Review*, **101** 3286–3307.

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.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, **119** 249–275.

Bester, C. A., Conley, T. G. and Hansen, C. B. (2011). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, **165** 137–151.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, **90** 414–427.

Canay, I. A., Romano, J. P. and Shaikh, A. M. (2017). Randomization tests under an approximate symmetry assumption. *Econometrica*, **85** 1013–1030.

Carter, A. V., Schnepel, K. T. and Steigerwald, D. G. (2017). Asymptotic behavior of at-test robust to cluster heterogeneity. *Review of Economics and Statistics*, **99** 698–709.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, **146** 162–169.

Davidson, R. and MacKinnon, J. G. (1999). The size distortion of bootstrap tests. *Econometric theory*, **15** 361–376.

Djogbenou, A. A., MacKinnon, J. G. and Nielsen, M. O. (2017). Validity of wild bootstrap inference with clustered errors. Tech. rep., Queen's Economics Department Working Paper.

Donald, S. G. and Lang, K. (2007). Inference with difference-in-differences and other panel data. *The review of Economics and Statistics*, **89** 221–233.

Giuliano, P. and Spilimbergo, A. (2014). Growing up in a recession. *The Review of Economic Studies*, **81** 787–817.

Ibragimov, R. and Müller, U. K. (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics*, **28** 453–468.

Ibragimov, R. and Müller, U. K. (2016). Inference with few heterogeneous clusters. *Review of Economics and Statistics*, **98** 83–96.

Kosfeld, M. and Rustagi, D. (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in ethiopia. *The American Economic Review*, **105** 747–783.

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

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. *The Annals of Statistics*, **16** 1696–1708.

MacKinnon, J. G., Nielsen, M. Ø. and Webb, M. D. (2017). Bootstrap and asymptotic inference with multiway clustering. Tech. rep.

MacKinnon, J. G. and Webb, M. D. (2014). Wild bootstrap inference for wildly different cluster sizes. Tech. rep., Queen's Economics Department Working Paper.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics* 255–285.

Marsaglia, G. and Olkin, I. (1984). Generating correlation matrices. *SIAM Journal on Scientific and Statistical Computing*, **5** 470–475.

Meng, X., Qian, N. and Yared, P. (2015). The institutional causes of china's great famine, 1959–1961. *The Review of Economic Studies*, **82** 1568–1611.

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

Webb, M. D. (2013). Reworking wild bootstrap based inference for clustered errors. Tech. rep., Queen's Economics Department Working Paper.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *The American Economic Review*, **93** 133–138.