The Wild Bootstrap with a "Small" Number of "Large" Clusters*

Ivan A. Canay[†] Andres Santos[‡] Azeem M. Shaikh[§]

August 8, 2019

Abstract

This paper studies the properties of the wild bootstrap-based test proposed in Cameron et al. (2008) for testing hypotheses about the coefficients in a linear regression model 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 an unstudentized version of the test is valid in the sense that it has limiting rejection probability under the null hypothesis that does not exceed the nominal level. Importantly, these conditions require, among other things, certain homogeneity restrictions on the distribution of covariates. In contrast, we establish that a studentized version of the test may only over-reject the null hypothesis by a "small" amount in the sense that it has limiting rejection probability under the null hypothesis that does not exceed the nominal level by more than an amount that decreases exponentially with the number of clusters. We obtain results qualitatively similar to those for the studentized version of the test for closely related "score" bootstrap-based tests, which permit testing hypotheses about parameters in nonlinear models. We illustrate the relevance of our theoretical results for applied work via a simulation study and empirical application.

Keywords: Wild bootstrap, Clustered Data, Randomization Tests.

*We thank Colin Cameron, Patrick Kline, Simon Lee, James MacKinnon, 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 and Yong Cai for excellent research assistance.

[†]Northwestern University, Department of Economics. iacanay@northwestern.edu

[‡]U.C.L.A., Department of Economics. andres@econ.ucla.edu

[§]University of Chicago, Department of Economics. amshaikh@uchicago.edu

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. (2019), and MacKinnon et al. (2019). 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 main 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 unstudentized test statistic leads to a test that is valid in the sense that it 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. While our conditions are not necessary, we believe our results also help shed some light on the poor behavior of the wild bootstrap in simulation studies that violate our homogeneity requirements; see, e.g., 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 may only over-reject the null hypothesis by a "small" amount in the sense that it 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.

The arguments used in establishing these properties for the studentized wild bootstrap-based test permit us to establish qualitatively similar results for wild bootstrap-based tests of nonlinear null hypotheses and closely related "score" bootstrap-based tests in nonlinear models. In particular, under conditions that include suitable "homogeneity" restrictions, we show that the limiting rejection probability of these tests under the null hypothesis does not exceed the nominal level by more than an amount that decreases exponentially with the number of clusters. We defer a formal statement of these results to Appendix C, but briefly discuss "score" bootstrap-based tests of linear null hypotheses in the generalized method of moments (GMM) framework of Hansen (1982) in the main text. Due to the differences with the wild bootstrap-based tests described previously, our discussion focuses on implementation and the homogeneity requirements needed in our formal result.

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 restriction described above. Canay et al. (2017) is related to our theoretical analysis in that it also employs 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 (2017).

The remainder of the paper is organized as follows. In Section 2, we formally introduce the test we study and the assumptions that will underlie our analysis. Our theoretical results are contained in Section 3. In Sections 4 and 5 , we illustrate the relevance of our asymptotic analysis for applied work via a simulation study and empirical application. We conclude in Section 6 with a summary of the main implications of our results for empirical work. The proofs of all results and a number of extensions can be found in the Appendix.

2 Setup

We index clusters by $j \in J \equiv \{1, \ldots, q\}$ and units in the jth cluster by $i \in I_{n,j} \equiv \{1, \ldots, n_j\}$. 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

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

where $\beta \in \mathbf{R}^{d_z}$ and $\gamma \in \mathbf{R}^{d_w}$ are unknown parameters and our requirements on $\epsilon_{i,j}$ are explained below in Section 2.1. In what follows, we consider β to be the parameter of primary interest and view γ as a nuisance parameter. For example, in the context of a randomized controlled trial, $Z_{i,j}$ may be an indicator for treatment status and $W_{i,j}$ may be a vector of "controls" such as additional unit-level characteristics or cluster-level fixed effects. Our hypothesis of interest therefore concerns only β. Specifically, we aim 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 $\alpha \in (0, 1)$. An important special case of this framework is a test of the null hypothesis that a particular component of β equals a given value.

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

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

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

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

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

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

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

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

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

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

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

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

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

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 obtained from linearly regressing $Z_{i,j}$ on $W_{i,j}$ employing the entire sample. The residuals from this regression,

$$\tilde{Z}_{i,j} \equiv Z_{i,j} - \hat{\Pi}_n' W_{i,j} \ , \tag{8}$$

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

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

Because the restrictions in (9) involve only data from cluster j, there may be multiple matrices $\hat{\Pi}^{\rm c}_{n,j}$ 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 $(\hat{\Pi}^{\rm c}_{n,j})' W_{i,j}$ with $i \in I_{n,j}$, i.e., the fitted values obtained from a linear regression of $Z_{i,j}$ on $W_{i,j}$ using only data from cluster 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 for all $j \in J$ 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

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

converges in distribution.

(ii) The quantity

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

converges in probability to a positive-definite matrix.

Assumption 2.1 imposes sufficient conditions to ensure that the ordinary least squares estimators of β and γ in (1) are well behaved. It further implies that the least squares estimators of β and γ subject to the restriction that $c'\beta = \lambda$ 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 and there is no perfect colinearity in $(Z'_{i,j}, W'_{i,j})'$.

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 $j \in J$,

$$\frac{1}{n_j} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \xrightarrow{P} a_j \Omega_{\tilde{Z}} \ , \tag{10}$$

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

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

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

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 $(Z'_{i,j}, W'_{i,j})'$ 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) by, for example, comparing the sample second moments matrices of $(Z'_{i,j}, W'_{i,j})'$ across clusters.

We conclude with three examples that illustrate the content of our assumptions.

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). For example, suppose that

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

with $E[\epsilon_{i,j}] = 0$, 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 also note that including cluster-level fixed effects is important for accommodating the model in Moulton (1986), where the error term is assumed to be of the form $v_j + \epsilon_{i,j}$. ■

Example 2.2. (*Cluster-Level Parameter Heterogeneity*) It is common in empirical work to consider models in which the parameters vary across clusters. As a stylized example, let

$$Y_{i,j} = \gamma + Z_{i,j}\beta_j + \eta_{i,j} \ , \tag{11}$$

where $Z_{i,j} \in \mathbf{R}$, $E[\eta_{i,j}] = 0$, and $E[Z_{i,j}\eta_{i,j}] = 0$. For β equal to a suitable weighted average of the β_j, we may write (11) in the form of (1) by setting $\epsilon_{i,j} = Z_{i,j}(\beta_j - \beta) + \eta_{i,j}$. By doing so, we see that unless $\beta_j = \beta$ for all $j \in J$, Assumption 2.2(i) is violated, as it requires that

$$\frac{1}{\sqrt{n_j}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\epsilon_{i,j} = \frac{1}{\sqrt{n_j}} \sum_{i \in I_{n,j}} (Z_{i,j} - \bar{Z}_n)(Z_{i,j}(\beta_j - \beta) + \eta_{i,j})$$

converge in distribution for all $j \in J$. A direct application of other methods that are valid with a "small" number of "large" clusters, such as Ibragimov and Müller (2010, 2016), and Canay et al. (2017), for this problem would also require that $\beta_j = \beta$ for all $j \in J$. We emphasize, however, that these methods would not require such an assumption for inference about $(\beta_j : j \in J)$. ■

Example 2.3. (*Differences-in-Differences*) It is difficult to satisfy our Assumptions 2.2(iii)-(iv) in settings where $Z_{i,j}$ is constant within cluster, i.e., $Z_{i,j}$ does not vary with $i \in I_{n,j}$. A popular setting in which this occurs and the wild bootstrap is commonly employed is differences-in-differences where treatment status is assigned at the level of the cluster. We illustrate this point in Appendix B with a stylized differences-in-differences example. ■

3 Main Results

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 and discuss extensions to non-linear models and hypotheses.

3.1 Unstudentized Test

Our first result shows that the unstudentized wild bootstrap-based test ϕ_n is indeed valid in the sense that its limiting rejection probability under the null hypothesis is no greater than the nominal level α. In addition we show the test is not too conservative by establishing a lower bound on its limiting rejection probability under the null hypothesis.

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

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

In the proof of Theorem 3.1, 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 wild bootstrap variants that construct 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 β and γ instead of the restricted estimators $\hat{\beta}^{\mathrm{r}}_n$ and $\hat{\gamma}^{\mathrm{r}}_n$. Notably, the use of the restricted estimators and Rademacher weights has been encouraged by Davidson and MacKinnon (1999), Cameron et al. (2008), and Davidson and Flachaire (2008).

While we focus on the ordinary least square setting of Section 2, we emphasize the conclusion of Theorem 3.1 can be easily 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 straightforward 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.

We next examine the power of the wild bootstrap-based test against $n^{-1/2}$-local alternatives. To this end, suppose

$$Y_{i,j} = Z'_{i,j}\beta_n + W'_{i,j}\gamma_n + \epsilon_{i,j}$$

with β_n satisfying $c'\beta_n = \lambda + \delta/\sqrt{n}$. Below, we denote by $P_{\delta,n}$ the distribution of the data in order to emphasize its dependence on both n and the local parameter δ. Our next result shows that the limiting rejection probability of ϕ_n along such sequences of local alternatives exceeds the nominal level (at least for sufficiently large values of $|\delta|$). While we do not present it as a part of the result, the proof in fact provides a lower bound on the limiting rejection probability of ϕ_n along such sequences of local alternatives for any value of δ. In addition to Assumptions 2.1 and 2.2, we impose that $\lceil |\mathbf{G}|(1-\alpha)\rceil < |\mathbf{G}| - 1$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x, in order to ensure that the critical value is not simply equal to the largest possible value of $|\sqrt{n}c'(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n)|$. This requirement will always be satisfied unless either α or q is too small.

Theorem 3.2. *If Assumptions 2.1 and 2.2 hold under $\{P_{\delta,n}\}$ and $\lceil |\mathbf{G}|(1-\alpha)\rceil < |\mathbf{G}| - 1$, then*

$$\lim_{|\delta|\to\infty} \liminf_{n\to\infty} P_{\delta,n}\{T_n > \hat{c}_n(1-\alpha)\} = 1 \ .$$

Remark 3.1. In order to appreciate why Theorem 3.1 does not follow from results in Canay et al. (2017), note that $T_n = F_n(s_n)$ for some function $F_n\colon \mathbf{R}^q \to \mathbf{R}$ and

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

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

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

and $ga = (g_1a_1, \ldots, g_qa_q)$ for any $a \in \mathbf{R}^q$. These observations and the definition of ϕ_n in (6) reveals a resemblance to a randomization test, but also highlights an important difference: the critical value is computed by applying g to a different statistic (i.e., $\hat{s}_n$) than the one defining the test statistic (i.e., s_n). This distinction prevents the application of results in Canay et al. (2017), as s_n and $\hat{s}_n$ do not even converge in distribution to the same limit. ■

Remark 3.2. 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 Studentized Test

We now analyze 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

$$\hat{\Omega}_{\tilde{Z},n} \equiv \frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \ , \tag{14}$$

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

$$\hat{\sigma}^2_n \equiv c' \hat{\Omega}^{-1}_{\tilde{Z},n} \hat{V}_n \hat{\Omega}^{-1}_{\tilde{Z},n} c \ . \tag{15}$$

Next, for any $g \in \mathbf{G} \equiv \{-1, 1\}^q$, recall that $(\hat{\beta}^*_n(g)', \hat{\gamma}^*_n(g)')'$ denotes the unconstrained ordinary least squares estimator of $(\beta', \gamma')'$ obtained from regressing $Y^*_{i,j}(g)$ (as defined in (4)) on $Z_{i,j}$ and $W_{i,j}$. 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 $\hat{\epsilon}^*_{i,j}(g) = Y^*_{i,j}(g) - Z'_{i,j}\hat{\beta}^*_n(g) - W'_{i,j}\hat{\gamma}^*_n(g)$, as the wild bootstrap-analogue to $\hat{V}_n$, and

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

to be the wild bootstrap-analogue to $\hat{\sigma}^2_n$. Notice that since the regressors are not re-sampled 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

$$\hat{c}^{\mathsf{s}}_n(1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\left\{|\sqrt{n}\frac{c'(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n)}{\hat{\sigma}^*_n(g)}| \leq u\right\} \geq 1-\alpha\right\} . \tag{17}$$

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.,

$$\phi^{\mathsf{s}}_n \equiv I\{T_n/\hat{\sigma}_n > \hat{c}^{\mathsf{s}}_n(1-\alpha)\} . \tag{18}$$

Our next result bounds the limiting rejection probability of ϕ^{s}_n under the null hypothesis.

Theorem 3.3. *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 3.3 indicates that studentizing the test-statistic T_n may lead to the test over-rejecting the null hypothesis in the sense that the limiting rejection probability of the test exceeds its nominal level, but by a "small" amount that decreases exponentially with the number of clusters. 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 level-$(\alpha + 2^{1-q})$ randomization test, from which the theorem follows. This implies, for example, that in applications with eight or more clusters, the limiting amount by which the test over-rejects the null hypothesis will be no greater than 0.008. These results also imply that it is possible to "size correct" the test simply by replacing α with $\alpha - 2^{1-q}$.

It is important to emphasize that there are compelling reasons for studentizing T_n in an asymptotic framework in which the number of clusters tends to infinity. In such a setting, the asymptotic distribution of $T_n/\hat{\sigma}_n$ is pivotal, while that of T_n is not. As a result, the analysis in Djogbenou et al. (2019) implies that the rejection probability of $\phi_n^{\mathtt{s}}$ under the null hypothesis converges to the nominal level α at a faster rate than the rejection probability of ϕ_n under the null hypothesis. Combined with Theorem 3.3, these results suggest that it may be preferable to employ the studentized test $\phi_n^{\mathtt{s}}$ unless the number of clusters q is sufficiently small for the difference between the upper bound in Theorem 3.3 and α to be of concern for the application at hand.

3.3 Discussion of Extensions

The arguments used in establishing Theorem 3.3 can be used to establish qualitatively similar results in a variety of other settings, such as tests of nonlinear null hypotheses and tests in nonlinear models, under suitable homogeneity requirements. We reserve the statement of formal results to Appendix C, but briefly discuss in this section tests of linear null hypotheses in a GMM framework. Given that there are no natural "residuals" in this framework, we do not employ the wild bootstrap to obtain a critical value. Instead, we rely on a specific variant of the "score" bootstrap as studied by Kline and Santos (2012). Our discussion therefore emphasizes computation of the critical value and the homogeneity assumptions needed in our formal result.

Denote by $X_{i,j} \in \mathbf{R}^{d_x}$ the observed data corresponding to ith unit in the jth cluster. Let

$$\hat{\beta}_n \equiv \arg\min_{b \in \mathbf{R}^{d_\beta}} \left(\frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} m(X_{i,j}, b) \right)' \hat{\Sigma}_n \left(\frac{1}{n} \sum_{j \in J} \sum_{i \in I_{n,j}} m(X_{i,j}, b) \right) , \qquad (19)$$

where $m(X_{i,j}, \cdot) : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_m}$ is a moment function and $\hat{\Sigma}_n$ is a $d_m \times d_m$ weighting matrix. Under suitable conditions, $\hat{\beta}_n$ is consistent for its estimand, which we denote by β. As in

Section 3.1, we consider testing

$$H_0 : c'\beta = \lambda \quad \text{vs.} \quad H_1 : c'\beta \neq \lambda \ , \tag{20}$$

at level $\alpha \in (0,1)$ by employing the test statistic $T_n^{\text{gmm}} \equiv |\sqrt{n}(c'\hat{\beta}_n - \lambda)|$. The critical value with which we compare T_n^{gmm} is computed as follows:

Step 1: Compute $\hat{\beta}_n^{\text{r}}$, the restricted GMM estimator obtained by minimizing the criterion in (19) under the constraint $c'b = \lambda$. Note that $c'\hat{\beta}_n^{\text{r}} = \lambda$ by construction.

Step 2: For any $b \in \mathbf{R}^{d_\beta}$, let $\hat{\Gamma}_n(b) \equiv (\hat{D}_n(b)'\hat{\Sigma}_n\hat{D}_n(b))^{-1}\hat{D}_n(b)\hat{\Sigma}_n$, where we define

$$\hat{D}_n(b) \equiv \frac{1}{n}\sum_{j\in J}\sum_{i\in I_{n,j}} \nabla m(X_{i,j}, b) \tag{21}$$

for $\nabla m(X_{i,j}, b)$ the Jacobian of $m(X_{i,j}, \cdot) : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_m}$ evaluated at b. For $\mathbf{G} = \{-1,1\}^q$ and writing an element $g \in \mathbf{G}$ as $g = (g_1, \ldots, g_q)$, we set as our critical value

$$\hat{c}_n^{\text{gmm}}(1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\left\{\sum_{j\in J}\frac{g_j}{\sqrt{n}}\sum_{i\in I_{n,j}} c'\hat{\Gamma}_n(\hat{\beta}_n^{\text{r}})m(X_{i,j}, \hat{\beta}_n^{\text{r}})\right\} \geq 1-\alpha\right\} \ .$$

We then obtain a test of (19) by rejecting whenever T_n^{gmm} is larger than $\hat{c}_n^{\text{gmm}}(1-\alpha)$, i.e.,

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

It is instructive to examine how ϕ_n^{gmm} simplifies in the context of Section 3.1. To this end, suppose $W_{i,j}$ is empty in (1), and set $X_{i,j} = (Y_{i,j}, Z_{i,j}')'$ and $m(X_{i,j}, b) = (Y_{i,j} - Z_{i,j}'b)Z_{i,j}$. It is straightforward to show that in this case

$$\tilde{Z}_{i,j} = Z_{i,j}, \quad m(X_{i,j}, \hat{\beta}_n^{\text{r}}) = \hat{\epsilon}_{i,j}^{\text{r}} Z_{i,j}, \quad \text{and} \quad \hat{D}_n(\hat{\beta}_n^{\text{r}}) = \hat{\Omega}_{\tilde{Z},n} \ .$$

As a result, the test ϕ_n^{gmm} is numerically equivalent to the test ϕ_n defined in (6). In this sense, ϕ_n^{gmm} may be viewed as a natural generalization of ϕ_n to the GMM setting. Moreover, the observation that $\hat{D}_n(\hat{\beta}_n^{\text{r}}) = \hat{\Omega}_{\tilde{Z},n}$ suggests that the appropriate generalization of the "homogeneity" requirement imposed in Assumption 2.2(iii) is to require for all $j \in J$ that

$$\frac{1}{n_j} \sum_{i \in I_{n,j}} \nabla m(X_{i,j}, \beta) \xrightarrow{P} a_j D(\beta) \tag{22}$$

for some $a_j > 0$ and $d_m \times d_\beta$ matrix $D(\beta)$ independent of $j \in J$. Indeed, in Appendix C, we show that under conditions including (22), the test ϕ_n^{gmm} has limiting rejection probability under the null hypothesis that is bounded by $\alpha + 2^{1-q}$. We thus find that nonlinearities, similar to studentiziation, may cause ϕ_n^{gmm} to over-reject by a "small" amount, in the sense that its limiting rejection probability under the null hypothesis exceeds the nominal level by an amount that decreases exponentially with q.

4 Simulation Study

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

$$Y_{i,j} = \gamma + Z_{i,j}'\beta + \sigma(Z_{i,j})(\eta_j + \epsilon_{i,j}) \ , \tag{23}$$

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:

Unstud: Corresponds to the unstudentized test studied in Theorem 3.1.

Stud: Corresponds to the studentized test studied in Theorem 3.3.

ET-US: Corresponds to the equi-tailed analog of the unstudentized test. This test rejects when the unstudentized 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).

ET-S: Corresponds to the equi-tailed 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 (15) and (17) respectively.

Each of the tests may be implemented with or without fixed effects (see Example 2.1), and with Rademacher weights or the alternative weighting scheme described 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 3.3.

We organize our discussion of the results by test.

Unstud: As expected in light of Theorem 3.1 and Example 2.1, Table 1 shows the unstudentized 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 rejec-

	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
	Unstud	6.48	9.90	9.34	9.42	9.24	14.48	13.80	12.48	15.40	14.42	13.06	12.16
Model 1	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
$n = 50$	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
	Unstud	9.02	5.96	9.70	9.98	10.58	15.84	15.60	15.42	14.26	13.62	13.78	13.72
Model 2	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
$n = 50$	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
	Unstud	7.24	9.72	9.46	10.16	10.54	15.48	14.32	14.24	15.58	14.78	13.48	12.88
Model 1	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
$n = 300$	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
	Unstud	6.96	9.68	9.74	10.12	12.30	17.74	16.20	15.26	15.50	14.86	14.08	13.34
Model 2	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
$n = 300$	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 $\alpha = 10\%$.

tion 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 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 unstudentized 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

	Test	Rade - with Fixed effects q=4	5	6	8	Rade - without Fixed effects q=4	5	6	8	Mammen - with Fixed effects q=4	5	6	8
Model 1 $n = 50$	unstud	19.80	33.14	39.34	42.28	20.42	34.94	39.54	40.74	35.46	37.86	40.84	42.50
	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
	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
Model 2 $n = 50$	unstud	13.34	20.28	20.04	18.88	15.56	25.16	23.38	21.58	22.68	22.28	20.94	20.34
	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
	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
Model 1 $n = 300$	unstud	22.22	39.20	42.46	48.32	21.80	39.72	40.84	44.80	38.30	42.10	43.38	48.08
	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
	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
Model 2 $n = 300$	unstud	15.60	23.98	24.72	20.86	17.46	27.72	26.92	22.88	24.58	23.98	24.52	21.08
	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
	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 $\alpha = 10\%$.

	Test	Rade - with Fixed effects q=4	5	6	8	Rade - without Fixed effects q=4	5	6	8
Model 3 $n = 50$	unstud	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 $n = 50$	unstud	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 $n = 300$	unstud	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
	ET-US	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 $n = 300$	unstud	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 $\alpha = 10\%$.

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

		Rade - with Fixed effects				Rade - without Fixed effects			
		q				q			
	Test	4	5	6	8	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 $\alpha = 12.5\%$.

is larger than two.

Stud: The studentized test studied in Theorem 3.3 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 cluster 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 unstudentized test, except when $q = 4$ where the studentized test exhibits higher power.

Theorem 3.3 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 3.3 can be conservative.

ET-US/ET-S: The equi-tailed versions of the unstudentized 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 unstudentized 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 Unstud and Stud. These differences in turn translate into lower power under the alternative hypothesis; see Table 2.

5 Empirical Application

In their investigation into the causes of the Chinese Great Famine between 1958 and 1960, Meng et al. (2015) study the relationship between province-level mortality and agricultural productivity during both famine years and non-famine years. To this end, in their baseline specification, Meng et al. (2015) estimate by ordinary least squares the equation

$$Y_{j,t+1} = Z^{(1)}_{j,t}\beta_1 + Z^{(2)}_{j,t}\beta_2 + W'_{j,t}\gamma + \epsilon_{j,t} \tag{24}$$

using data from 19 provinces between 1953 and 1982, where

$$\begin{aligned} Y_{j,t+1} &= \log(\text{number of deaths in province } j \text{ during year } t+1) \\ Z^{(1)}_{j,t} &= \log(\text{predicted grain production in province } j \text{ during year } t) \\ Z^{(2)}_{j,t} &= Z^{(1)}_{j,t} \times I\{t \text{ is a famine year}\} \end{aligned}$$

and $W_{j,t}$ is vector of year-level fixed effects and other covariates. We henceforth refer to this as Analysis #1. As robustness checks, Meng et al. (2015) additionally consider the following:

Analysis #2: Repeating Analysis #1 using only data between 1953 and 1965.

Analysis #3: Repeating Analysis #1 using four additional provinces.

Analysis #4: Repeating Analysis #2 using four additional provinces.

Analysis #5: Repeating Analysis #1 using actual rather than predicted grain production.

Analysis #6: Repeating Analysis #2 using actual rather than predicted grain production.

The results of these six analyses can be found in Table 2 of Meng et al. (2015). Among other things, for each analysis, Meng et al. (2015) report the ordinary least squares estimate of β_1 as well as its heteroskedasticity-consistent standard errors, and the ordinary least squares estimate of $\beta_1 + \beta_2$ as well as a p-value for testing the null hypothesis that $\beta_1 + \beta_2 = 0$ computed using heteroskedasticity-consistent standard errors. In unreported results, they write in footnote 33 that conclusions computed using the wild bootstrap are similar.

In Table 5, we consider for each of these six analyses different ways of testing the null hypotheses that $\beta_1 = 0$ and $\beta_1 + \beta_2 = 0$. For each analysis and for each null hypothe-

Analysis	H_0	FE	Coef	T_n	$T_n/\hat{\sigma}_n$	Wild p-value	Wild S. p-value	Cluster p-value	Robust p-value
#1	$\beta_1 = 0$	No	0.148	3.532	3.195	0.019	0.029	0.005	0.000
		Yes	0.141	3.363	2.899	0.026	0.028	0.010	0.000
	$\beta_1 + \beta_2 = 0$	No	0.141	3.371	2.368	0.054	0.061	0.029	0.001
		Yes	0.145	3.470	2.937	0.046	0.081	0.009	0.001
#2	$\beta_1 = 0$	No	0.103	1.614	2.473	0.041	0.047	0.024	0.013
		Yes	0.088	1.374	1.900	0.037	0.052	0.074	0.023
	$\beta_1 + \beta_2 = 0$	No	0.098	1.533	1.829	0.070	0.072	0.084	0.025
		Yes	0.050	0.790	0.893	0.321	0.353	0.383	0.270
#3	$\beta_1 = 0$	No	0.156	4.097	3.877	0.013	0.014	0.001	0.000
		Yes	0.140	3.676	3.182	0.027	0.027	0.004	0.001
	$\beta_1 + \beta_2 = 0$	No	0.115	3.023	3.140	0.049	0.029	0.005	0.007
		Yes	0.174	4.577	4.245	0.017	0.032	0.000	0.000
#4	$\beta_1 = 0$	No	0.120	2.071	3.245	0.029	0.026	0.004	0.005
		Yes	0.084	1.445	1.818	0.082	0.080	0.083	0.047
	$\beta_1 + \beta_2 = 0$	No	0.094	1.628	2.576	0.056	0.030	0.017	0.033
		Yes	0.057	0.975	1.010	0.297	0.281	0.323	0.248
#5	$\beta_1 = 0$	No	0.137	3.262	3.885	0.015	0.008	0.001	0.000
		Yes	0.135	3.227	3.322	0.015	0.011	0.004	0.000
	$\beta_1 + \beta_2 = 0$	No	0.113	2.689	1.784	0.168	0.141	0.091	0.004
		Yes	0.024	0.576	0.394	0.803	0.692	0.699	0.739
#6	$\beta_1 = 0$	No	0.090	1.419	3.215	0.031	0.021	0.005	0.015
		Yes	0.087	1.371	2.380	0.012	0.011	0.029	0.008
	$\beta_1 + \beta_2 = 0$	No	0.089	1.402	1.528	0.160	0.171	0.144	0.045
		Yes	-0.124	1.943	1.303	0.227	0.180	0.209	0.340

Table 5: Results for model (24) for each of the six analyses in Table 2 of Meng et al. (2015). 'Coef' denotes the estimated value of β_1 or $\beta_1 + \beta_2$. T_n denotes the corresponding value of the statistic in (3). $T_n/\hat{\sigma}_n$ denotes the corresponding value of the Studentized statistic in (18). 'Wild p-value' is the corresponding p-value using the un-Studentized wild bootstrap. 'Wild S. p-value' is the corresponding p-value using the Studentized wild bootstrap. 'Cluster p-value' is the corresponding p-value using cluster-robust standard errors. 'Robust p-value' is the corresponding p-value using heteroskedasticity-consistent standard errors

sis, we report the ordinary least squares estimate of the quantity of interest; the value of the unstudentized test statistic T_n defined in (3); the value of the studentized test statistic $T_n/\hat{\sigma}_n$, where $\hat{\sigma}_n^2$ is defined in (15); the wild bootstrap p-value corresponding to T_n; the wild bootstrap p-value corresponding to $T_n/\hat{\sigma}_n$; the p-value computed using cluster-robust standard errors; and, finally, the p-value computed using heteroskedasticity-consistent standard errors. We also repeat each of these exercises after adding cluster-level fixed effects.

Our results permit the following observations:

1. The inclusion or exclusion of cluster-level fixed effects may have a significant impact on the wild bootstrap p-values (both unstudentized and studentized). For an extreme example

of this phenomenon, see the p-values for testing the null hypothesis that $\beta_1 + \beta_2 = 0$ in Analyses #2 and #4, where, the wild bootstrap p-values with cluster-level fixed effects are far above any conventional significance level whereas those without cluster-level fixed effects are quite small. We note that in light of our discussion in Example 2.1 we would expect the results with cluster-level fixed effects included to be more reliable.

2. The unstudentized wild bootstrap p-values may be both smaller or larger than the studentized wild bootstrap p-values. Importantly, in some cases, these differences may be meaningful in that they may lead tests based on these p-values to reach different conclusions. In order to illustrate this point, see the p-values for testing the null hypothesis that $\beta_1 + \beta_2 = 0$ in Analyses #1 and #4. Given that in this application $2^{1-q} \leq 2^{-18}$, Theorem 3.3 and the benefits of studentizing as the number of clusters diverges to infinity (Djogbenou et al., 2019) suggest that test based on the studentized wild bootstrap p-values are preferable to those based on unstudentized wild bootstrap p-values in this application.

3. The wild bootstrap p-values (both unstudentized and studentized) may be both smaller or larger than the p-values computed using cluster-robust standard errors. As in our preceding point, in some cases these differences may be meaningful in that they may lead tests based on these p-values to reach different conclusions. In order to illustrate this point, see the p-values for testing the null hypothesis that $\beta_1 = 0$ in Analyses #2 and #3. Since p-values based on cluster-robust standard errors are only theoretically justified in a framework where the number of clusters tend to infinity, our analysis suggests that in this setting it is preferable to employ wild bootstrap-based p-values.

Recall that both Theorems 3.1 and 3.3 rely on the homogeneity requirements described in Assumption 2.2(iii). We therefore conclude our empirical application with a brief examination of the plausibility of this assumption in this example. We pursue this exercise only in the context of Analysis #1, i.e., using predicted versus actual grain production and using data on 19 provinces between 1953 and 1982. To this end, we compute below the matrix on the left-hand side of (10) for several different provinces. If Assumption 2.2(iii) held, then we would expect these matrices to be approximately proportional to one another. This property does not appear to hold in this application. To see this, consider the values of these matrices for Beijing (corresponding to $j = 1$) and Tianjin (corresponding to $j = 2$):

$$\Omega_{1,n} = \begin{pmatrix} 0.302 & 0.066 \\ 0.066 & 0.987 \end{pmatrix} \quad \text{and} \quad \Omega_{2,n} = \begin{pmatrix} 0.228 & 0.021 \\ 0.021 & 0.012 \end{pmatrix} .$$

The lower diagonal elements of these matrices differ by a factor of > 80, whereas the other elements differ by a factor that is at least an order of magnitude smaller. Similar results hold for other pairs of provinces and other analyses. These observations suggest that Assumption 2.2(iii) does not hold in this application. In light of the simulation study in Section 4, we may therefore wish to be cautious when applying the wild bootstrap in this setting.

6 Recommendations for Empirical Practice

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. Our results have a number of important implications for applied work that we summarize below:

- Wild bootstrap-based tests can be valid even if the number of clusters is "small." This conclusion, however, applies to a specific variant of the wild bootstrap-based test proposed in Cameron et al. (2008). Practitioners should, in particular, use Rademacher weights and avoid other weights such those in Mammen (1993) in such settings. Practitioners should also avoid reporting wild bootstrap-based standard errors because t-tests based on such standard errors are not asymptotically valid in an asymptotic framework in which the number of clusters is fixed.

- The studentized version of the wild bootstrap-based test has a limiting rejection probability that exceeds the nominal level by an amount of at most 2^{1-q}. In an asymptotic framework in which the number clusters diverges to infinity, however, the studentized test exhibits advantages over its unstudentized counterpart. Therefore, we recommend employing studentized wild bootstrap-based test unless the number of clusters is sufficiently small for the factor 2^{1-q} to be of concern.

- Our results rely on certain "homogeneity" assumptions on the distribution of covariates across clusters. These "homogeneity" requirements can sometimes be weakened by including cluster-level fixed effects. Whenever the number of clusters is small and the "homogeneity" assumptions are implausible, however, we recommend instead employing an inference procedure that does not rely on these types of "homogeneity" conditions, such as those developed in Canay et al. (2017).

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? *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. *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 a t-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. (2019). Asymptotic theory and wild bootstrap inference with clustered errors. *Journal of Econometrics*. Forthcoming.

Donald, S. G. and Lang, K. (2007). Inference with difference-in-differences and other panel data. *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.

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

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.

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

Kosfeld, M. and Rustagi, D. (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in Ethiopia. *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. *Annals of Statistics*, **16** 1696–1708.

MacKinnon, J. G., Nielsen, M. Ø. and Webb, M. D. (2019). Bootstrap and asymptotic inference with multiway clustering. Tech. rep., Working Paper 1415, Economics Department, Queen's University.

MacKinnon, J. G. and Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, **32** 233–354.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *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. *Review of Economic Studies*, **82** 1568–1611.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics*, **32** 385–397.

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.

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. *American Economic Review*, **93** 133–138.

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 $\hat{\epsilon}^{\mathrm{r}}_{i,j} \equiv (Y_{i,j} - Z_{i,j}'\hat{\beta}^{\mathrm{r}}_n - W_{i,j}'\hat{\gamma}^{\mathrm{r}}_n)$, 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

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

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

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

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, i.e., $\iota \equiv (1,\ldots,1) \in \mathbf{R}^q$, 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 $\hat{\epsilon}^{\mathrm{r}}_{i,j} = (Y_{i,j} - Z'_{i,j}\hat{\beta}^{\mathrm{r}}_n - W'_{i,j}\hat{\gamma}^{\mathrm{r}}_n)$ that whenever $E_n = 1$ we obtain

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

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 $c'\hat{\beta}^{\mathrm{r}}_n = \lambda$ 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

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

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

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

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

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

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 yields

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

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

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

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 $\epsilon > 0$

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

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

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

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

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

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 non-singular by Assumption 2.2(iii). Finally, noting that $T(\iota S) = T(-\iota S) = T(S)$, we can conclude $T(S) > T^{(k^*)}(S|\mathbf{G} \setminus \{\pm\iota\})$ if and only if $T(S) > T^{(k^*)}(S|\mathbf{G})$, which together with (A-10) and (A-18) yields

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

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., Lehmann and Romano, 2005, Theorem 15.2.1). 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 $\alpha - \frac{1}{2^{q-1}} \le 0$, in which case the result trivially follows. Assume $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil \le |\mathbf{G}| - 2$ and note that

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

where the first inequality follows from result (A-7), the second inequality follows from Portmanteau's theorem (see, e.g., van der Vaart and Wellner, 1996, Theorem 1.3.4(iii)), the third inequality holds because $P\{T^{(\mathbf{z}+2)}(S|\mathbf{G}) > T^{(\mathbf{z})}(S|\mathbf{G})\} = 1$ for any integer $\mathbf{z} \le |\mathbf{G}| - 2$ by (A-1) and Assumption 2.2(i)-(ii), 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). Thus, the lower bound holds and the theorem follows. ■

Proof of Theorem 3.2: Throughout the proof, all convergence in distribution and probability statements are understood to be along the sequence $\{P_{\delta,n}\}$. Following the notation in the proof of Theorem 3.1, we first let $\mathbb{S} \equiv \mathbf{R}^{d_z \times d_z} \times \bigotimes_{j\in J} \mathbf{R}^{d_z}$ and write an element of $s \in \mathbb{S}$ 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 $j \in J$. We then define the map $T : \mathbb{S} \to \mathbf{R}$ to be given by

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

for any $s \in \mathbb{S}$ such that s_1 is invertible, and set $T(s) = 0$ whenever s_1 is not invertible. We again identify any $(g_1, \ldots, g_q) = g \in \mathbf{G} = \{-1, 1\}^q$ with an action $s \in \mathbb{S}$ defined by $gs = (s_1, \{g_j s_{2,j} : j \in J\})$. We finally define $E_n \in \mathbf{R}$ and $S_n \in \mathbb{S}$ to equal

$$\begin{aligned}
E_n &\equiv I\{\hat{\Omega}_{\tilde{Z},n} \text{ is invertible}\} \\
S_n &\equiv \Big(\hat{\Omega}_{\tilde{Z},n}, \Big\{ \sum_{i\in I_{n,j}} \frac{\tilde{Z}_{i,j}\epsilon_{i,j}}{\sqrt{n}} + \frac{\tilde{Z}_{i,j}\tilde{Z}_{i,j}'}{n}\sqrt{n}(\beta_n - \hat{\beta}_n^{\mathrm{r}})\Big\}\Big) \ ,
\end{aligned}$$

where

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

Since $c'\hat{\beta}_n^{\mathrm{r}} = \lambda$, the Frisch-Waugh-Lovell theorem implies, whenever $E_n = 1$, that

$$\begin{aligned} |\sqrt{n}(c'\hat{\beta}_n - \lambda)| &= |\sqrt{n}c'(\hat{\beta}_n - \beta_n) + \sqrt{n}c'(\beta_n - \hat{\beta}_n^{\mathrm{r}})| \\ &= \Big|c'\hat{\Omega}_{\tilde{Z},n}^{-1}\sum_{j\in J}\sum_{i\in I_{n,j}} \frac{\tilde{Z}_{i,j}\epsilon_{i,j}}{\sqrt{n}} + \sqrt{n}c'(\beta_n - \hat{\beta}_n^{\mathrm{r}})\Big| \\ &= \Big|c'\hat{\Omega}_{\tilde{Z},n}^{-1}\sum_{j\in J}\sum_{i\in I_{n,j}} \Big(\frac{\tilde{Z}_{i,j}\epsilon_{i,j}}{\sqrt{n}} + \frac{\tilde{Z}_{i,j}\tilde{Z}_{i,j}'}{n}\sqrt{n}(\beta_n - \hat{\beta}_n^{\mathrm{r}})\Big)\Big| \\ &= T(S_n) \ , \end{aligned} \quad \text{(A-21)}$$

where the final equality follows from the definition of $T : \mathbb{S} \to \mathbf{R}$. Also note that Amemiya (1985, Eq. (1.4.5)), Assumption 2.1, and $\sqrt{n}c'(\beta_n - \lambda) = \delta$ imply that $\sqrt{n}(\hat{\beta}_n^{\mathrm{r}} - \beta_n) = O_P(1)$ and $\sqrt{n}(\hat{\gamma}_n^{\mathrm{r}} - \gamma_n) = O_P(1)$. Therefore, manipulations similar to those in (A-21), Lemma A.2 and $n_j/n \to \xi_j > 0$ by Assumption 2.2(ii) imply, whenever $E_n = 1$, that for any $g \in \mathbf{G}$

$$\begin{aligned} |\sqrt{n}c'(\hat{\beta}_n^*(g) - \hat{\beta}_n^{\mathrm{r}})| &= \Big|c'\hat{\Omega}_{\tilde{Z},n}^{-1}\sum_{j\in J}\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}} g_j\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}^{\mathrm{r}}\Big| \\ &= \Big|c'\hat{\Omega}_{\tilde{Z},n}^{-1}\sum_{j\in J}\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}} g_j\Big(\tilde{Z}_{i,j}Z_{i,j}'(\beta_n - \hat{\beta}_n^{\mathrm{r}}) + \tilde{Z}_{i,j}W_{i,j}'(\gamma_n - \hat{\gamma}_n^{\mathrm{r}}) + \tilde{Z}_{i,j}\epsilon_{i,j}\Big)\Big| \\ &= \Big|c'\hat{\Omega}_{\tilde{Z},n}^{-1}\sum_{j\in J}\sum_{i\in I_{n,j}} g_j\Big(\frac{\tilde{Z}_{i,j}\epsilon_{i,j}}{\sqrt{n}} + \frac{\tilde{Z}_{i,j}\tilde{Z}_{i,j}'}{n}\sqrt{n}(\beta_n - \hat{\beta}_n^{\mathrm{r}})\Big)\Big| + o_P(1) \ . \end{aligned}$$

We next study the asymptotic behavior of $T(gS_n)$. To this end, we first note that Amemiya (1985, Eq. (1.4.5)) and the partitioned inverse formula imply, whenever $E_n = 1$, that

$$\hat{\beta}_n^{\mathrm{r}} = \hat{\beta}_n - \hat{\Omega}_{\tilde{Z},n}^{-1}c\frac{c'\hat{\beta}_n - \lambda}{c'\hat{\Omega}_{\tilde{Z},n}^{-1}c} = \hat{\beta}_n - \hat{\Omega}_{\tilde{Z},n}^{-1}c\left(\frac{c'(\hat{\beta}_n - \beta_n)}{c'\hat{\Omega}_{\tilde{Z},n}^{-1}c} + \frac{c'\beta_n - \lambda}{c'\hat{\Omega}_{\tilde{Z},n}^{-1}c}\right) \ . \quad \text{(A-22)}$$

Therefore, employing that $\sqrt{n}(c'\beta_n - \lambda) = \delta$ by hypothesis, we conclude whenever $E_n = 1$ that

$$\sum_{i\in I_{n,j}} \frac{\tilde{Z}_{i,j}\tilde{Z}_{i,j}'}{n}\sqrt{n}(\beta_n - \hat{\beta}_n^{\mathrm{r}}) = \sum_{i\in I_{n,j}} \frac{\tilde{Z}_{i,j}\tilde{Z}_{i,j}'}{n}\Big\{\Big(I_{d_z} - \hat{\Omega}_{\tilde{Z},n}^{-1}\frac{cc'}{c'\hat{\Omega}_{\tilde{Z},n}^{-1}c}\Big)\sqrt{n}(\beta_n - \hat{\beta}_n) + \frac{\hat{\Omega}_{\tilde{Z},n}^{-1}c}{c'\hat{\Omega}_{\tilde{Z},n}c}\delta\Big\} \ , \quad \text{(A-23)}$$

where I_{d_z} denotes the $d_z \times d_z$ identity matrix. Since Assumptions 2.2(ii)-(iii) imply $\hat{\Omega}_{\tilde{Z},n} \xrightarrow{P} \bar{a}\Omega_{\tilde{Z}}$ 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, it follows $E_n = 1$ with probability tending to one. Hence, results (A-22), (A-23), and Assumptions 2.2(ii)-(iii) yield

$$\limsup_{n\to\infty} P_{\delta,n}\Big\{\Big|\sqrt{n}c'(\hat{\beta}^*_n(g)-\hat{\beta}^{\rm r}_n) - c'\hat{\Omega}^{-1}_{\tilde{Z},n}\sum_{j\in J}\sum_{i\in I_{n,j}} g_j\Big(\frac{\tilde{Z}_{i,j}\epsilon_{i,j}}{\sqrt{n}} + c\frac{\xi_j a_j\delta}{c'\hat{\Omega}^{-1}_{\tilde{Z},n}c}\Big)\Big| > \epsilon;\ E_n = 1\Big\} = 0\ . \tag{A-24}$$

In particular, results (A-21) and (A-24), $\hat{\Omega}_{\tilde{Z},n} \xrightarrow{P} \bar{a}\Omega_{\tilde{Z}}$, and Assumption 2.2(i) establish that

$$(T_n, \{\sqrt{n}c'(\hat{\beta}^*_n(g)-\hat{\beta}^{\rm r}_n) : g\in\mathbf{G}\}) \xrightarrow{d} (T(S_\delta), \{T(gS_\delta) : g\in\mathbf{G}\})$$

where

$$S_\delta \equiv \Big(\bar{a}\Omega_{\tilde{Z}}, \Big\{\sqrt{\xi_j}\mathcal{Z}_j + c\frac{\bar{a}\xi_j a_j\delta}{c'\Omega^{-1}_{\tilde{Z}}c} : j\in J\Big\}\Big)\ .$$

By definition of $\hat{c}_n(1-\alpha)$ and Portmanteau's theorem (see, e.g., van der Vaart and Wellner, 1996, Theorem 1.3.4(ii)), it then follows that

$$\liminf_{n\to\infty} P_{\delta,n}\{T_n > \hat{c}_n(1-\alpha)\} \geq P\Big\{T(S_\delta) > \inf\Big\{u\in\mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{T(gS_\delta)\leq u\} \geq 1-\alpha\Big\}\Big\}\ . \tag{A-25}$$

To conclude the proof, we denote the ordered values of $\{T(gs) : g\in\mathbf{G}\}$ according to

$$T^{(1)}(s|\mathbf{G}) \leq \cdots \leq T^{(|\mathbf{G}|)}(s|\mathbf{G})\ .$$

Then observe that since $\lceil |\mathbf{G}|(1-\alpha)\rceil < |\mathbf{G}|-1$ by hypothesis, result (A-25) implies that

$$\liminf_{|\delta|\to\infty}\liminf_{n\to\infty} P_{\delta,n}\Big\{T_n > \hat{c}_n(1-\alpha)\Big\} \geq \liminf_{|\delta|\to\infty} P\Big\{T(S_\delta) = T^{(|\mathbf{G}|)}(S_\delta|\mathbf{G})\Big\}\ .$$

Let $\iota = (1,\cdots,1)\in\mathbf{R}^q$, and note that since $T(\iota S) = T(-\iota S)$, the triangle inequality yields

$$\begin{aligned} &P\{T(S_\delta) = T^{(|\mathbf{G}|)}(S_\delta|\mathbf{G})\} \\ &\quad\geq P\Big\{\Big|\sum_{j\in J}(\frac{\sqrt{\xi_j}}{\bar{a}}c'\Omega^{-1}_{\tilde{Z}}\mathcal{Z}_j + \xi_j a_j\delta)\Big| \geq \max_{g\in\mathbf{G}\setminus\{\pm\iota\}}\Big|\sum_{j\in J} g_j(\frac{\sqrt{\xi_j}}{\bar{a}}c'\Omega^{-1}_{\tilde{Z}}\mathcal{Z}_j + \xi_j a_j\delta)\Big|\Big\} \\ &\quad\geq P\Big\{|\delta|\Big(\sum_{j\in J}\xi_j a_j - \max_{g\in\mathbf{G}\setminus\{\pm\iota\}}|\sum_{j\in J}\xi_j a_j g_j|\Big) \geq 2\sum_{j\in J}\Big|\frac{\sqrt{\xi_j}}{\bar{a}}c'\Omega^{-1}_{\tilde{Z}}\mathcal{Z}_j\Big|\Big\}\ . \end{aligned}$$

Since $a_j\xi_j > 0$ for all $1\leq j\leq J$ and every $g\in\mathbf{G}\setminus\{\pm\iota\}$ must have at least one coordinate equal to 1 and at least one coordinate equal to -1, it follows that

$$\sum_{j\in J}\xi_j a_j - \max_{g\in\mathbf{G}\setminus\{\pm\iota\}}\Big|\sum_{j\in J}\xi_j a_j g_j\Big| > 0\ .$$

Hence, since $\sum_{j\in J}|\sqrt{\xi_j}c'\Omega_{\tilde Z}^{-1}\mathcal{Z}_j| = O_P(1)$ by Assumption 2.2(i), we finally obtain that

$$\liminf_{|\delta|\to\infty}\liminf_{n\to\infty} P_{\delta,n}\{T_n > \hat c_n(1-\alpha)\}$$
$$\geq \liminf_{|\delta|\to\infty} P\Big\{|\delta|\Big(\sum_{j\in J}\xi_j a_j - \max_{g\in\mathbf{G}\setminus\{\pm\iota\}}\Big|\sum_{j\in J}\xi_j a_j g_j\Big|\Big) \geq 2\sum_{j\in J}\Big|\frac{\sqrt{\xi_j}}{\bar a}c'\Omega_{\tilde Z}^{-1}\mathcal{Z}_j\Big|\Big\} = 1 ,$$

which establishes the claim of the theorem. ■

Proof of Theorem 3.3: 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

$$T(s) \equiv |c'(s_1)^{-1}(\sum_{j\in J}s_{2,j}) - \lambda| \tag{A-26}$$
$$W(s) \equiv \Big(c'(s_1)^{-1}\sum_{j\in J}\Big(s_{2,j} - \frac{\xi_j a_j}{\bar a}\sum_{\tilde j\in J}s_{2,\tilde j}\Big)\Big(s_{2,j} - \frac{\xi_j a_j}{\bar a}\sum_{\tilde j\in J}s_{2,\tilde j}\Big)'(s_1)^{-1}c\Big)^{1/2} , \tag{A-27}$$

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

$$A_n \equiv I\{\hat\Omega_{\tilde Z,n}\text{ is invertible, } \hat\sigma_n > 0,\text{ and } \hat\sigma_n^*(g) > 0 \text{ for all } g\in\mathbf{G}\} \tag{A-28}$$
$$S_n \equiv \Big(\hat\Omega_{\tilde Z,n},\Big\{\frac{1}{\sqrt n}\sum_{i\in I_{n,j}}\tilde Z_{i,j}\epsilon_{i,j} : j\in J\Big\}\Big) \tag{A-29}$$

where recall $\hat\Omega_{\tilde Z,n}$ was defined in (14) 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

$$\Big\{\frac{\sqrt{n_j}}{\sqrt n}\frac{1}{\sqrt{n_j}}\sum_{i\in I_{n,j}}\tilde Z_{i,j}\epsilon_{i,j} : j\in J\Big\} \xrightarrow{d} \{\sqrt{\xi_j}\mathcal{Z}_j : j\in J\} . \tag{A-30}$$

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-30) together with the definition of S_n in (A-29), Assumption 2.2(ii)-(iii), and the continuous mapping theorem then allows us to establish

$$S_n \xrightarrow{d} S \equiv \Big(\bar a\Omega_{\tilde Z},\{\sqrt{\xi_j}\mathcal{Z}_j : j\in J\}\Big) , \tag{A-31}$$

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

$$\hat{\sigma}_n = W(S_n) + o_P(1) \qquad \hat{\sigma}^*_n(g) = W(gS_n) + o_P(1) \tag{A-32}$$

due to the definition of $W \colon \mathbb{S} \to \mathbf{R}$ in (A-27) and Lemma A.1. Moreover, $\hat{\Omega}_{\tilde{Z},n}$ being invertible with probability tending to one additionally allows us to conclude that

$$\begin{aligned}\liminf_{n\to\infty} P\{A_n = 1\} &= \liminf_{n\to\infty} P\{\hat{\sigma}_n > 0 \text{ and } \hat{\sigma}^*_n(g) > 0 \text{ for all } g \in \mathbf{G}\} \\ &\geq P\{W(gS) > 0 \text{ for all } g \in \mathbf{G}\} = 1 \ , \end{aligned} \tag{A-33}$$

where the inequality in (A-33) holds by (A-31), (A-32), the continuous mapping theorem, and Portmanteau's Theorem (see, e.g., van der Vaart and Wellner, 1996, Theorem 1.3.4(ii)). In turn, the final equality in (A-33) 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 $\hat{\epsilon}^{\mathrm{r}}_{i,j} = (Y_{i,j} - Z'_{i,j}\hat{\beta}^{\mathrm{r}}_n - W'_{i,j}\hat{\gamma}^{\mathrm{r}}_n)$ and note that whenever $A_n = 1$ we obtain

$$\begin{aligned}\sqrt{n}c'(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n) &= c'\hat{\Omega}^{-1}_{\tilde{Z},n}\frac{1}{\sqrt{n}}\sum_{j\in J}\sum_{i\in I_{n,j}} g_j\tilde{Z}_{i,j}\hat{\epsilon}^{\mathrm{r}}_{i,j} \\ &= c'\hat{\Omega}^{-1}_{\tilde{Z},n}\frac{1}{\sqrt{n}}\sum_{j\in J}\sum_{i\in I_{n,j}} g_j\tilde{Z}_{i,j}(\epsilon_{i,j} - Z'_{i,j}(\hat{\beta}^{\mathrm{r}}_n - \beta) - W'_{i,j}(\hat{\gamma}^{\mathrm{r}}_n - \gamma)) \ .\end{aligned} \tag{A-34}$$

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

$$\limsup_{n\to\infty} P\Big\{|c'\hat{\Omega}^{-1}_{\tilde{Z},n}\sum_{j\in J}\frac{g_j}{n}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\sqrt{n}(\hat{\gamma}^{\mathrm{r}}_n - \gamma)| > \epsilon;\ A_n = 1\Big\} = 0 \tag{A-35}$$

for any $\epsilon > 0$. Similarly, since $\sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta)$ 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 $\epsilon > 0$

$$\begin{aligned}&\limsup_{n\to\infty} P\Big\{\Big|c'\hat{\Omega}^{-1}_{\tilde{Z},n}\sum_{j\in J}\frac{n_j}{n}\frac{1}{n_j}g_j\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}Z'_{i,j}\sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta)\Big| > \epsilon;\ A_n = 1\Big\} \\ &\quad = \limsup_{n\to\infty} P\Big\{\Big|c'\hat{\Omega}^{-1}_{\tilde{Z},n}\sum_{j\in J}\frac{n_j}{n}\frac{1}{n_j}g_j\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j}\sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta)\Big| > \epsilon;\ A_n = 1\Big\} \\ &\quad = \limsup_{n\to\infty} P\Big\{\Big|c'\Omega^{-1}_{\tilde{Z}}\sum_{j\in J}\frac{\xi_j a_j g_j}{\bar{a}}\Omega_{\tilde{Z}}\sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta)\Big| > \epsilon;\ A_n = 1\Big\} = 0 \ .\end{aligned} \tag{A-36}$$

It follows from results (A-32)-(A-36) together with $T(S_n) = T_n$ whenever $\hat{\Omega}_{\tilde{Z},n}$ is invertible, that

$$\begin{aligned}&((|\sqrt{n}(c'\hat{\beta}_n - \lambda)|, \hat{\sigma}_n), \{(|c'\sqrt{n}(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n)|, \hat{\sigma}^*_n(g)) : g \in \mathbf{G}\}) \\ &\qquad = ((T(S_n), W(S_n)), \{(T(gS_n), W(gS_n)) : g \in \mathbf{G}\}) + o_P(1) \ .\end{aligned} \tag{A-37}$$

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 $P\{W(gS) > 0$ for all $g \in \mathbf{G}\} = 1$ as argued in (A-33). 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

$$\mathrm{t}^{(1)}(s|\mathbf{G}) \leq \ldots \leq \mathrm{t}^{(|\mathbf{G}|)}(s|\mathbf{G}) \ . \tag{A-38}$$

Next, we observe that result (A-33) and a set inclusion inequality allow us to conclude that

$$\limsup_{n\to\infty} P\Big\{\frac{T_n}{\hat{\sigma}_n} > \hat{c}_n^{\mathsf{s}}(1-\alpha)\Big\} \leq \limsup_{n\to\infty} P\Big\{\frac{T_n}{\hat{\sigma}_n} \geq \hat{c}_n^{\mathsf{s}}(1-\alpha);\ A_n = 1\Big\}$$
$$\leq P\Big\{\mathrm{t}(S) \geq \inf\Big\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{\mathrm{t}(gS) \leq u\} \geq 1-\alpha\Big\}\Big\} \ , \tag{A-39}$$

where the final inequality follows by results (A-31), (A-37), and the continuous mapping and Portmanteau theorems (see, e.g., van der Vaart and Wellner, 1996, Theorem 1.3.4(iii)). Therefore, setting $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil$, we can then obtain from result (A-39) that

$$\limsup_{n\to\infty} P\Big\{\frac{T_n}{\hat{\sigma}_n} > \hat{c}_n^{\mathsf{s}}(1-\alpha)\Big\}$$
$$\leq P\{\mathrm{t}(S) > \mathrm{t}^{(k^*)}(S)\} + P\{\mathrm{t}(S) = \mathrm{t}^{(k^*)}(S)\} \leq \alpha + P\{\mathrm{t}(S) = \mathrm{t}^{(k^*)}(S)\} \ , \tag{A-40}$$

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., Lehmann and Romano, 2005, Theorem 15.2.1). Moreover, applying Lehmann and Romano (2005, Theorem 15.2.2) yields

$$P\{\mathrm{t}(S) = \mathrm{t}^{(k^*)}(S)\}$$
$$= E[P\{\mathrm{t}(S) = \mathrm{t}^{k^*}(S)|S \in \{gS : g \in \mathbf{G}\}\}] = E\left[\frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{\mathrm{t}(gS) = \mathrm{t}^{(k^*)}(S)\}\right] \ . \tag{A-41}$$

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

$$P\{\mathrm{t}(gS) = \mathrm{t}(\tilde{g}S)\} = 0 \tag{A-42}$$

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,

$$\frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{\mathrm{t}(gS) = \mathrm{t}^{(k^*)}(S)\} = \frac{1}{|\mathbf{G}|} \times 2 = \frac{1}{2^{q-1}} \tag{A-43}$$

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-40) and (A-43). 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* $(g_1, \ldots, g_q) = g \in \mathbf{G}$

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

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 A.2 and the definition of $U_{n,j}$ yield

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\hat{\epsilon}_{i,j} &= \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\epsilon_{i,j} - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} Z'_{i,j}\sqrt{n}(\hat{\beta}_n - \beta) - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} W'_{i,j}\sqrt{n}(\hat{\gamma}_n - \gamma) \\ &= U_{n,j} - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\tilde{Z}'_{i,j}\sqrt{n}(\hat{\beta}_n - \beta) + o_P(1) \ . \qquad \text{(A-44)}\end{aligned}$$

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

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

Therefore, (A-44), (A-45), and the continuous mapping theorem yield

$$\begin{aligned}\hat{V}_n &= \sum_{j \in J} \Big(\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j}\hat{\epsilon}_{i,j}\Big)\Big(\frac{1}{\sqrt{n}} \sum_{k \in I_{n,j}} \tilde{Z}'_{k,j}\hat{\epsilon}_{k,j}\Big) \\ &= \sum_{j \in J} \Big(U_{n,j} - \frac{\xi_j a_j}{\bar{a}} \sum_{\tilde{j} \in J} U_{n,\tilde{j}}\Big)\Big(U_{n,j} - \frac{\xi_j a_j}{\bar{a}} \sum_{\tilde{j} \in J} U_{n,\tilde{j}}\Big)' + o_P(1) \ . \qquad \text{(A-46)}\end{aligned}$$

The first part of the lemma thus follows by the definition of $\hat{\sigma}^2_n$ 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 $\sqrt{n}(\hat{\beta}^{\rm r}_n - \beta)$ and $\sqrt{n}(\hat{\gamma}^{\rm r}_n - \gamma)$ are bounded in probability. Together with Assumption 2.1 such result in turn also implies that $\sqrt{n}(\hat{\beta}^*_n(g) - \hat{\beta}^{\rm r}_n)$ and $\sqrt{n}(\hat{\gamma}^*_n(g) - \hat{\gamma}^{\rm r}_n)$ are bounded in probability for all $g \in \mathbf{G}$. Next, recall that the residuals from the bootstrap regression in (4) equal $\hat{\epsilon}^*_{i,j}(g) = g_j\hat{\epsilon}^{\rm r}_{i,j} - Z'_{i,j}(\hat{\beta}^*_n(g) - \hat{\beta}^{\rm r}_n) - W'_{i,j}(\hat{\gamma}^*_n(g) - \hat{\gamma}^{\rm r}_n)$ 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

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \hat{\epsilon}^*_{i,j}(g) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} g_j \hat{\epsilon}^{\mathrm{r}}_{i,j} - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} Z'_{i,j} \sqrt{n}(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n) - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} W'_{i,j} \sqrt{n}(\hat{\gamma}^*_n(g) - \hat{\gamma}^{\mathrm{r}}_n) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} g_j \hat{\epsilon}^{\mathrm{r}}_{i,j} - \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \sqrt{n}(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n) + o_P(1) \ , \qquad \text{(A-47)}
\end{aligned}
$$

where in the final equality we employed Lemma A.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

$$
\begin{aligned}
c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{\sqrt{n}} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} g_j \hat{\epsilon}^{\mathrm{r}}_{i,j} &= c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} g_j (\epsilon_{i,j} - Z'_{i,j}\sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta) - W'_{i,j}\sqrt{n}(\hat{\gamma}^{\mathrm{r}}_n - \gamma)) \\
&= c'\hat{\Omega}^-_{\tilde{Z},n} g_j U_{n,j} - c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{n} \sum_{i \in I_{n,j}} g_j \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta) + o_P(1) \ , \quad \text{(A-48)}
\end{aligned}
$$

where the second equality follows from Lemma A.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

$$
c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{n} \sum_{i \in I_{n,j}} g_j \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta) = c'\Omega^{-1}_{\tilde{Z}} \frac{g_j \xi_j a_j}{\bar{a}} \Omega_{\tilde{Z}} \sqrt{n}(\hat{\beta}^{\mathrm{r}}_n - \beta) + o_P(1) = o_P(1) \ , \qquad \text{(A-49)}
$$

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

$$
\begin{aligned}
c'\hat{\Omega}^-_{\tilde{Z},n} \frac{1}{n} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \sqrt{n}(\hat{\beta}^*_n(g) - \hat{\beta}^{\mathrm{r}}_n) &= c'\hat{\Omega}^-_{\tilde{Z},n} \frac{n_j}{n} \frac{1}{n_j} \sum_{i \in I_{n,j}} \tilde{Z}_{i,j} \tilde{Z}'_{i,j} \hat{\Omega}^-_{\tilde{Z},n} \sum_{\tilde{j} \in J} \frac{1}{\sqrt{n}} \sum_{k \in I_{n,\tilde{j}}} \tilde{Z}_{k,j} g_{\tilde{j}} \hat{\epsilon}^{\mathrm{r}}_{k,\tilde{j}} + o_P(1) \\
&= c'\hat{\Omega}^-_{\tilde{Z},n} \frac{\xi_j a_j}{\bar{a}} \sum_{\tilde{j} \in J} g_{\tilde{j}} U_{n,\tilde{j}} + o_P(1) \ , \qquad \text{(A-50)}
\end{aligned}
$$

where in the final equality we applied (A-48), (A-49), 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 (A-47)-(A-50). ■

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

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

PROOF: Let $\|\cdot\|_F$ denote the Frobenius matrix norm, which recall equals $\|M\|_F^2 \equiv \text{trace}\{M'M\}$ 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

$$\|\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\|_F = \|\frac{1}{n_j}\sum_{i\in I_{n,j}}(Z_{i,j}-\hat{\Pi}'_n W_{i,j})W'_{i,j}\|_F$$
$$= \|\frac{1}{n_j}\sum_{i\in I_{n,j}}(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F \le \frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F \ . \quad \text{(A-51)}$$

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

$$\frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}W'_{i,j}\|_F = \frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}\|\times\|W'_{i,j}W_{i,j}\|$$
$$\le \left\{\frac{1}{n_j}\sum_{i\in I_{n,j}}\|(\hat{\Pi}^{c}_{n,j}-\hat{\Pi}_n)'W_{i,j}\|^2\right\}^{1/2}\times\left\{\frac{1}{n_j}\sum_{i\in I_{n,j}}\|W_{i,j}\|^2\right\}^{1/2} = o_P(1) \ , \quad \text{(A-52)}$$

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

$$\frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}Z'_{i,j} = \frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}\tilde{Z}'_{i,j} + \frac{1}{n_j}\sum_{i\in I_{n,j}}\tilde{Z}_{i,j}W'_{i,j}\hat{\Pi}_n \quad \text{(A-53)}$$

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

B 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

$$\tau \equiv \frac{\sum_{j\in J(1)} n_j(1)p_j}{\sum_{j\in J} n_j(1)p_j} \ , \quad \text{(B-54)}$$

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

$$\hat{\Pi}_n' W_{i,j} = \begin{cases} -p_j \tau & \text{if } T_{i,j} = 1 \text{ and } j \in J(0) \\ (1-\tau) p_j \tau & \text{if } T_{i,j} = 1 \text{ and } j \in J(1) \\ (1-p_j)\tau & \text{if } T_{i,j} = 2 \text{ and } j \in J(0) \\ \tau + (1-\tau) p_j & \text{if } T_{i,j} = 2 \text{ and } j \in J(1) \end{cases} , \tag{B-55}$$

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 τ, 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.

C 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 ith unit in the jth 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 ψ 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

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

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

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

In the context of the linear model studied in the main paper, under appropriate choices of F, ψ, and $\hat{\psi}_n$, the test ϕ_n^{F} is in fact numerically equivalent to the test ϕ_n defined in (6). More generally, however, the test ϕ_n^{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

$$\inf\left\{u \in \mathbf{R} : P\left\{F\left(\sum_{j\in J}\frac{\omega_j}{\sqrt{n}}\sum_{i\in I_{n,j}}\hat{\psi}_n(X_{i,j})\right) \le u|X^{(n)}\right\} \ge 1-\alpha\right\} \tag{C-56}$$

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 (C-56).

Our analysis will require the following high-level assumption:

Assumption C.1. *The following statements hold:*

(i) The test statistic T_n^{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 κ *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 C.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 C.1(ii) encodes homogeneity restrictions qualitatively similar to those in Assumption 2.2; see our discussion of nonlinear restrictions and GMM below. Assumption C.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 C.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 κ, however, we can also accommodate settings in which $n_j/n \to 0$ for some j or Σ_j in Assumption C.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 C.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 C.1. *If Assumption C.1 holds, then*

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

Proof of Theorem C.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 C.1(ii)-(iii) and the continuous mapping theorem imply

$$(F(S_n), \{F(g\hat{S}_n) : g \in \mathbf{G}\}) \xrightarrow{d} (F(S), \{F(gS) : g \in \mathbf{G}\}) \ . \tag{C-57}$$

Hence, by Assumption C.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

$$\begin{aligned} \limsup_{n\to\infty} P\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\} &\leq \limsup_{n\to\infty} P\{T_n^{\mathtt{F}} \geq \hat{c}_n^{\mathtt{F}}(1-\alpha)\} \\ &\leq P\left\{F(S) \geq \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{F(gS) \leq u\} \geq 1-\alpha\right\}\right\} \ . \end{aligned} \tag{C-58}$$

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

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

Setting $k^* \equiv \lceil |\mathbf{G}|(1-\alpha)\rceil$, we then obtain from (C-58) and Assumption C.1(iii) that

$$\begin{aligned} \limsup_{n\to\infty} P\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\} &\leq P\{F(S) > F^{(k^*)}(S|\mathbf{G})\} + P\{F(S) = F^{(k^*)}(S|\mathbf{G})\} \\ &\leq \alpha + P\{F(S) = F^{(k^*)}(S|\mathbf{G})\} \ , \end{aligned} \tag{C-59}$$

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

$$P\{F(S) = F^{(k^*)}(S|\mathbf{G})\} = E[P\{F(S) = F^{(k^*)}(S|\mathbf{G})|S \in \{gS\}_{g\in\mathbf{G}}\}$$
$$= E\left[\frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{F(gS) = F^{(k^*)}(gS|\mathbf{G})\}\right] = \frac{\kappa}{2^q} \ , \quad \text{(C-60)}$$

where the final equality follows from Assumptions C.1(iv)-(v). The claim of the upper bound in the theorem therefore follows from results (C-59) and (C-60).

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 (C-57) and the Portmanteau Theorem, see, e.g., Theorem 1.3.4(ii) in van der Vaart and Wellner (1996) yield

$$\liminf_{n\to\infty} P\{T_n^{\mathtt{F}} > \hat{c}_n^{\mathtt{F}}(1-\alpha)\} \geq P\{F(S) > F^{(k^*)}(S|\mathbf{G})\} \geq P\{F(S) \geq F^{(k^*+\kappa)}(S|\mathbf{G})\} \ , \quad \text{(C-61)}$$

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 C.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

$$P\{F(S) \geq F^{(k^*+\kappa)}(S|\mathbf{G})\} \geq \alpha - \frac{\kappa}{2^q} \ . \quad \text{(C-62)}$$

Thus, the lower bound holds by (C-61) and (C-62), and the claim of the theorem follows. ■

C.1 Applications of the General Result

Below, we apply Theorem C.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).

C.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 β, consider testing

$$H_0 : h(\beta) = 0 \quad \text{vs.} \quad H_1 : h(\beta) \neq 0 \ . \quad \text{(C-63)}$$

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

$$\hat{c}_n^{\mathtt{F}}(1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\Big\{\|\nabla h(\hat{\beta}_n^{\mathrm{r}})\sum_{j\in J}\frac{g_j}{\sqrt{n}}\sum_{i\in I_{n,j}}\hat{\Omega}_{\tilde{Z},n}^{-1}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}^{\mathrm{r}}\|^2 \leq u\Big\} \geq 1-\alpha\right\} \ , \quad \text{(C-64)}$$

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

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

under the null hypothesis.

Theorem C.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 β, then*

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

Sketch of Proof: Theorem C.2 follows from an application of Theorem C.1. To map ϕ_n^{F} into the context of Theorem C.1, we let $X_{i,j} = (Y_{i,j}, Z_{i,j}', W_{i,j}')'$ and define

$$\psi(X_{i,j}) = \nabla h(\beta)(\bar{a}\Omega_{\tilde{Z}})^{-1}\tilde{Z}_{i,j}\epsilon_{i,j} \ , \tag{C-65}$$

where recall $\bar{a} = \sum_{j\in J} a_j\xi_j$. It then follows by standard arguments and $\hat{\Omega}_{\tilde{Z},n} \xrightarrow{P} \bar{a}\Omega_{\tilde{Z}}$ by Assumptions 2.2(ii)-(iii), that T_n^{F} satisfies Assumption C.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 (C-65). Moreover, by setting

$$\hat{\psi}_n(X_{i,j}) = \nabla h(\hat{\beta}_n^{\mathrm{r}})\hat{\Omega}_{\tilde{Z},n}^{-1}\tilde{Z}_{i,j}\hat{\epsilon}_{i,j}^{\mathrm{r}} \ , \tag{C-66}$$

we verify the critical value in (C-64) has the exact structure required by Theorem C.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 C.1(ii) is satisfied. Finally, Assumptions C.1(iii)-(v) hold immediately with $\kappa = 2$ by Assumptions 2.2(i)-(ii). ■

Remark C.1. In this application it is also natural to consider employing the critical value

$$\hat{c}_n^{\mathrm{F}}(1-\alpha) \equiv \inf\left\{u \in \mathbf{R} : \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} I\{\|\sqrt{n}h(\hat{\beta}_n^*(g))\|^2 \leq u\} \geq 1-\alpha\right\} \tag{C-67}$$

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 $\bar{\beta}_n(g) \xrightarrow{P} \hat{\beta}_n^{\mathrm{r}}$. Hence, the continuity of the the Jacobian ∇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 $\hat{c}_n^{\mathrm{F}}(1-\alpha)$ as in (C-64) and $\tilde{c}_n^{\mathrm{F}}(1-\alpha)$ as in (C-67). Inspecting the proof of Theorem C.1 (see, in particular, (C-57), (C-58), and (C-61)), then reveals the conclusion of Theorem C.1 continues to apply if we employ $\tilde{c}_n^{\mathrm{F}}(1-\alpha)$ in place of $\hat{c}_n^{\mathrm{F}}(1-\alpha)$; i.e.

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

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

C.1.2 Generalized Method of Moments

In this section, we apply Theorem C.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

$$\hat{\beta}_n \equiv \arg\min_b \left(\frac{1}{n}\sum_{j \in J}\sum_{i \in I_{n,j}} m(X_{i,j}, b)\right)' \hat{\Sigma}_n \left(\frac{1}{n}\sum_{j \in J}\sum_{i \in I_{n,j}} m(X_{i,j}, b)\right) \ , \tag{C-68}$$

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 β. For $h : \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_h}$ with $d_\beta \leq d_h$ and h continuously differentiable at β, we consider testing

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

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

$$\hat{D}_n(b) \equiv \frac{1}{n}\sum_{j \in J}\sum_{i \in I_{n,j}} \nabla m(X_{i,j}, b) \tag{C-69}$$

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 $\hat{\beta}_n^{\mathrm{r}}$ the GMM estimator computed subject to the restriction $h(\hat{\beta}_n^{\mathrm{r}}) = 0$,

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

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 C.1 to establish properties of $\phi_n^{\mathtt{F}}$, we impose the following assumption:

Assumption C.2. *The following statements hold:*

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

(ii) There are full rank matrices Σ *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* Σ_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 C.3. *If Assumption C.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 C.3 follows from an application of Theorem C.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

$$\psi(X_{i,j}) = \nabla h(\beta)(D(\beta)'\Sigma D(\beta))^{-1} D(\beta)'\Sigma m(X_{i,j}, \beta) . \tag{C-70}$$

Assumption C.2(iii), continuity of $||\cdot||^2$, and the continuous mapping theorem imply Assumption C.1(i). Assumption C.1(iii) follows from C.2(iv) with

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

Assumptions C.1(iv) and C.1(v) are then immediate with $\kappa = 2$. We are then left with Assumption C.1(ii). By the mean value theorem and the definition of $\hat{\psi}_n(X_{i,j})$, we obtain

$$\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}\hat{\psi}_n(X_{i,j}) = \nabla h(\hat{\beta}_n^{\rm r})(\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\hat{D}_n(\hat{\beta}_n^{\rm r}))^{-1}\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\frac{1}{\sqrt{n}}\sum_{i\in I_{n,j}}m(X_{i,j},\beta)$$
$$+\nabla h(\hat{\beta}_n^{\rm r})(\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\hat{D}_n(\hat{\beta}_n^{\rm r}))^{-1}\hat{D}_n(\hat{\beta}_n^{\rm r})'\hat{\Sigma}_n\frac{1}{n}\sum_{i\in I_{n,j}}\nabla m(X_{i,j},\bar{\beta}_n)\sqrt{n}(\hat{\beta}_n^{\rm r}-\beta) \ , \quad \text{(C-71)}$$

where $\bar{\beta}_n$ lies between $\hat{\beta}_n^{\rm r}$ and β. Assumptions C.2(i), C.2(ii), and C.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 C.2(i) and C.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 β by Assumptions C.2(i)-(iii). This completes the argument. ■