Inference for Large-Scale Linear Systems with Known Coefficients*

Zheng Fang
Department of Economics
Emory University
zheng.fang@emory.edu

Andres Santos
Department of Economics
UCLA
andres@econ.ucla.edu

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

Alexander Torgovitsky
Department of Economics
University of Chicago
torgovitsky@uchicago.edu

August 1, 2022

Abstract

This paper considers the problem of testing whether there exists a non-negative solution to a possibly under-determined system of linear equations with known coefficients. This hypothesis testing problem arises naturally in a number of settings, including random coefficient, treatment effect, and discrete choice models, as well as a class of linear programming problems. As a first contribution, we obtain a novel geometric characterization of the null hypothesis in terms of identified parameters satisfying an infinite set of inequality restrictions. Using this characterization, we devise a test that requires solving only linear programs for its implementation, and thus remains computationally feasible in the high-dimensional applications that motivate our analysis. The asymptotic size of the proposed test is shown to equal at most the nominal level uniformly over a large class of distributions that permits the number of linear equations to grow with the sample size.

KEYWORDS: linear programming, linear inequalities, moment inequalities, random coefficients, partial identification, exchangeable bootstrap, uniform inference.

*We thank Denis Chetverikov, Patrick Kline, and Adriana Lleras-Muney for helpful comments. Omkar Katta and Conroy Lau provided outstanding research assistance. The research of the third author was supported by NSF grant SES-1530661. The research of the fourth author was supported by NSF grant SES-1846832.

1 Introduction

Given an independent and identically distributed (i.i.d.) sample $\{Z_i\}_{i=1}^n$ with Z distributed according to $P \in \mathbf{P}$, this paper studies the hypothesis testing problem

$$H_0 : P \in \mathbf{P}_0 \qquad H_1 : P \in \mathbf{P} \setminus \mathbf{P}_0, \tag{1}$$

where $\mathbf{P}$ is a "large" set of distributions satisfying conditions described below and $\mathbf{P}_0 \equiv \{P \in \mathbf{P} : \beta(P) = Ax \text{ for some } x \geq 0\}$. Here, "$x \geq 0$" signifies that all coordinates of $x \in \mathbf{R}^d$ are non-negative, $\beta(P) \in \mathbf{R}^p$ denotes an unknown parameter, and the coefficients of the linear system are known in that A is known.

As we discuss in Section 2, the described hypothesis testing problem plays a central role in a surprisingly varied array of empirical settings. Tests of (1) can be used for obtaining asymptotically valid confidence regions for counterfactual broadband demand in the analysis of Nevo et al. (2016), and for conducting inference on the fraction of employers engaging in discrimination in the audit study of Kline and Walters (2021). Within the treatment effects literature, tests of (1) arise naturally when examining the testable implications of the model proposed by Imbens and Angrist (1994) and when conducting inference on partially identified parameters, such as in the studies by Kline and Walters (2016) and Kamat (2019) of the Head Start program, or the analysis of unemployment state dependence by Torgovitsky (2019). The null hypothesis in (1) has also been shown by Kitamura and Stoye (2018) to play a central role in testing whether a cross-sectional sample is rationalizable by a random utility model; see Manski (2014), Deb et al. (2017), and Lazzati et al. (2018) for related examples.

Tests of the null hypothesis in (1) can also be used to test whether a class of linear programs is feasible or, through test inversion, to obtain confidence regions for the optimal value of these linear programs. More precisely, our results are applicable to any linear programs whose standard form has the structure

$$\min_{x \in \mathbf{R}^d_+} c'x \quad \text{s.t.} \quad Ax = \beta(P)$$

for known A and c. We emphasize that linear (in)equality constraints on x and/or $\beta(P)$ can be incorporated by using slack variables appropriately. This connection to linear programming enables us to conduct inference in the competing risks model of Honoré and Lleras-Muney (2006), the empirical study of the California

Affordable Care Act marketplace by Tebaldi et al. (2019), and the dynamic discrete choice model of Honoré and Tamer (2006).

The null hypothesis in (1) can equivalently be represented as a system of linear inequalities in $\beta(P)$ through, e.g., Fourier-Motzkin elimination. Such a representation would enable us to test (1) by relying on approaches devised by the literature on testing for the validity of moment inequalities; see Canay and Shaikh (2017) for a review. Unfortunately, in the empirical applications that motivate us the dimensions p and, in particular, d are large, making obtaining such a representation computationally infeasible (Kitamura and Stoye, 2018). We proceed instead by obtaining a novel geometric characterization of the null hypothesis that forms the cornerstone of our approach to inference. Specifically, we show that the null hypothesis in (1) holds if and only if: (i) there is an $x \in \mathbf{R}^d$ (not necessarily positive) solving $Ax = \beta(P)$; and (ii) the minimum norm solution to $Ax = \beta(P)$, denoted $x^\star(P)$, forms an obtuse angle with any vector in the intersection of the row space of A and the negative orthant in $\mathbf{R}^d$. Condition (ii) can be represented as a finite number of linear inequalities in $x^\star(P)$, though enumerating such inequalities can again be computationally prohibitive in applications with large p and/or d. We show that such enumeration is unnecessary: One can instead evaluate whether condition (ii) holds by computing the largest inner product between $x^\star(P)$ and the vectors in the intersection of the row space of A and the negative orthant – a task that may be accomplished by solving a linear program.

Our geometric characterization can be employed to construct a variety of different tests. Guided by computational and statistical reliability when p and/or d are large, we focus on a test that can be computed through linear programming. Our test statistic employs a linear program to compute the largest violation of the "inequality" restrictions in our geometric characterization of the null hypothesis. We obtain a critical value through a bootstrap procedure that requires solving one linear program per bootstrap iteration. The resulting test is similar in spirit to generalized moment selection in incorporating information on whether inequalities are "slack" or "close" to binding (Andrews and Soares, 2010). An R package for implementing our test is available at `https://github.com/conroylau/lpinfer`. In Section 4, we additionally describe variants of this test that may be less computationally tractable, but may be attractive in some low-dimensional settings in terms of their power properties. We emphasize, however, that, none of these tests are motivated by any specific optimality criterion.

Besides delivering computational tractability, the linear programming structure in our test enables us to establish the consistency of our asymptotic approximations under the requirement that p^2/n tends to zero (up to logs). Leveraging the consistency of such approximations to establish the asymptotic validity of our test further requires us to verify an anti-concentration condition at a particular quantile (Chernozhukov et al., 2014). We show that the required anti-concentration property indeed holds under a condition that relates the allowed rate of growth of p relative to n to the matrix A. This result enables us to derive a sufficient, but more stringent, condition on the rate of growth of p relative to n that delivers anti-concentration universally in A. Furthermore, if, as in much of the related literature, p is fixed with n, then our results imply that our test is asymptotically valid under "weak" regularity conditions on $\mathbf{P}$.

Our paper is related to important work by Kitamura and Stoye (2018), who study (1) in the context of testing the validity of a random utility model. Their inference procedure, however, relies on conditions on A that can be violated in the broader set of applications that motivate us; see Section 2. Andrews et al. (2019) and Cox and Shi (2019) propose methods for sub-vector inference in certain conditional moment inequality models that can be related to (1). However, applying their tests, which were designed with a different problem in mind, to (1) can require non-trivial theoretical extensions or be computationally challenging – in particular when, as in most of our examples, $\beta(P)$ has non-zero known coordinates and/or d is very large. On the other hand, we show in Section 4.4.2 that an important insight in Andrews et al. (2019) allows us to adapt our methodology to conduct subvector inference in a class of conditional moment inequality models. Our analysis is also conceptually related to work on sub-vector inference in models involving moment inequalities and to a literature on shape restrictions; see, e.g., Romano and Shaikh (2008), Bugni et al. (2017), Kaido et al. (2019), Gandhi et al. (2019), Chernozhukov et al. (2015), Zhu (2020), and Fang and Seo (2021). While these procedures are designed for general problems that do not possess the specific structure in (1), they are, as a result, less computationally tractable and/or rely on more demanding and high-level conditions than the ones we employ.

2 Applications

In order to fix ideas, we first discuss a number of empirical settings in which the hypothesis testing problem described in (1) arises naturally.

Example 2.1. (Dynamic Programming). Building on Fox et al. (2011), Nevo et al. (2016) estimate a model for residential broadband demand in which there are $h \in \{1, \ldots, d\}$ types of consumers that select among plans $k \in \{1, \ldots, K\}$. Each plan has fee F_k, speed s_k, usage allowance $\bar{C}_k$, and overage price p_k. At day t, a type h consumer with plan k has utility over c_t and numeraire y_t equal to

$$u_h(c_t, y_t, v_t; k) = v_t(\frac{c_t^{1-\zeta_h}}{1-\zeta_h}) - c_t(\kappa_{1h} + \frac{\kappa_{2h}}{\log(s_k)}) + y_t,$$

where v_t is an i.i.d. shock following a truncated log-normal distribution with mean μ_h and variance σ_h^2. The problem faced by a type h consumer with plan k is

$$\max_{c_1,\ldots,c_T} \sum_{t=1}^{T} E[u_h(c_t, y_t, v_t; k)] \text{ s.t. } F_k + \mathsf{p}_k \max\{\sum_{t=1}^{T} c_t - \bar{C}_k, 0\} + \sum_{t=1}^{T} y_t \leq I, \quad (2)$$

where the expectation is over v_t and total wealth I is assumed to be large enough to not restrict usage. From (2), it follows that the distribution of observed plan choice and daily usage, denoted by $Z \in \mathbf{R}^{T+1}$, for a consumer of type h is characterized by $\theta_h \equiv (\zeta_h, \kappa_{1h}, \kappa_{2h}, \mu_h, \sigma_h)$. Hence, for any function m of Z we obtain the restriction

$$E_P[m(Z)] = \sum_{h=1}^{d} E_{\theta_h}[m(Z)]x_h,$$

where E_P and E_{θ_h} denote expectations under the distribution P of Z and under θ_h, respectively, and x_h is the unknown proportion of each type in the population. After specifying $d = 16807$ different types and $p = 120000$ moments, Nevo et al. (2016) estimate $x \equiv (x_1, \ldots, x_d)$ by GMM while constraining x to be a probability measure. The authors then employ the constrained GMM estimator for x and the block bootstrap to conduct inference on counterfactual demand, which equals $\sum_{h=1}^{d} a(\theta_h)x_h$ for a known function a. We note, however, that the results in Fang and Santos (2018) imply the bootstrap is *inconsistent* for this problem – the bootstrap fails because the restriction $x \geq 0$ causes the GMM estimator to not be "smooth" function of the moments. In contrast, the results in this paper enable us to conduct asymptotically valid inference. For instance, by setting

$$\beta(P) \equiv \begin{pmatrix} E_P[m(Z)] \\ 1 \\ \gamma \end{pmatrix} \qquad A \equiv \begin{pmatrix} E_{\theta_1}[m(Z)] & \cdots & E_{\theta_d}[m(Z)] \\ 1 & \cdots & 1 \\ a(\theta_1) & \cdots & a(\theta_d) \end{pmatrix} \quad (3)$$

we may obtain a confidence region for counterfactual demand through test inversion (in γ) of the null hypothesis in (1). Other applications of the approach in Nevo et al. (2016) include Blundell et al. (2018) and Illanes and Padi (2019). ■

Example 2.2. (Treatment Effects). Consider the heterogenous treatment effects model of Imbens and Angrist (1994) in which an instrument $W \in \{0,1\}$, potential treatments $(D(0), D(1))$, and potential outcomes $(Y(0), Y(1))$ satisfy

$$(D(0), D(1), Y(0), Y(1)) \perp\!\!\!\perp W \text{ and } D(1) \geq D(0) \text{ a.s.} \tag{4}$$

The requirements in (4) yield testable restrictions on the distributions of observables $(Y, D, W) \equiv (Y(D), D(W), W)$ (Balke and Pearl, 1994; Angrist and Imbens, 1995; Kitagawa, 2015) that may be mapped into (1). Specifically, assuming for simplicity that Y has discrete support $\mathcal{K}$, and letting $j^c \equiv 1-j$, we note (4) yields

$$\begin{aligned}
P(Y \in B, D = j | W = 0) &= \sum_{l \in \{0,1\}: l \geq j} \sum_{(m,k) \in B \times \mathcal{K}} P((Y(j), Y(j^c), D(0), D(1)) = (m, k, j, l)) \\
P(Y \in B, D = j | W = 1) &= \sum_{l \in \{0,1\}, l \leq j} \sum_{(m,k) \in B \times \mathcal{K}} P((Y(j), Y(j^c), D(0), D(1)) = (m, k, l, j)) \\
1 &= \sum_{l,j \in \{0,1\}: l \geq j} \sum_{m,k \in \mathcal{K}} P((Y(0), Y(1), D(0), D(1)) = (m, k, j, l))
\end{aligned}$$

for any set B. These restrictions may be written as $\beta(P) = Ax$ with $x \geq 0$ denoting the distribution of $(Y(0), Y(1), D(0), D(1))$. For K the number of support points of Y, in this problem $d = 3K^2$ and p is as large as $4K + 1$. For instance, in estimating the distribution of compliers in Angrist and Krueger (1991), Imbens and Rubin (1997) let W indicate fourth quarter birth and discretize log weekly earning into 55 bins, yielding $d = 9075$ and $p = 221$. As in Example 2.1, we may also construct confidence intervals for linear functionals of the distribution of $(Y(0), Y(1), D(0), D(1))$ such as the average treatment effect (Balke and Pearl, 1997; Laffers, 2019; Machado et al., 2019; Kamat, 2019; Bai et al., 2020). ■

Example 2.3. (Duration Models). In studying the efficacy of President Nixon's war on cancer, Honoré and Lleras-Muney (2006) employ the competing risks model

$$(T^*, I) = \begin{cases} (\min\{S_1, S_2\}, \arg\min\{S_1, S_2\}) & \text{if } W = 0 \\ (\min\{\alpha S_1, \beta S_2\}, \arg\min\{\alpha S_1, \beta S_2\}) & \text{if } W = 1 \end{cases},$$

where (S_1, S_2) represent duration until death due to cancer and cardio-vascular disease, $W \perp\!\!\!\perp (S_1, S_2)$ indicates the implementation of the war on cancer, and (α, β) are unknown parameters. The observed variables are (T, I, W) where $T = t_k$

if $t_k \leq T^* < t_{k+1}$ for $k = 1, \ldots, M$ and $t_{M+1} = \infty$, reflecting data sources often contain interval observations of duration. While (α, β) is partially identified, Honoré and Lleras-Muney (2006) show there are known finite sets $\mathcal{S}(\alpha, \beta)$ and $\mathcal{S}_{k,i,w}(\alpha, \beta) \subseteq \mathcal{S}(\alpha, \beta)$ such that (α, β) belongs to the identified set if and only if

$$\sum_{(s_1,s_2)\in\mathcal{S}_{k,i,w}(\alpha,\beta)} f(s_1, s_2) = P(T = t_k, I = i|W = w),$$
$$\sum_{(s_1,s_2)\in\mathcal{S}(\alpha,\beta)} f(s_1, s_2) = 1, \text{ and } f(s_1, s_2) \geq 0 \text{ for all } (s_1, s_2) \in \mathcal{S}(\alpha, \beta), \quad (5)$$

for some distribution f on $\mathcal{S}(\alpha, \beta)$, and where the first equality must hold for all $1 \leq k \leq M$, $i \in \{1, 2\}$, and $w \in \{0, 1\}$. In the context of Honoré and Lleras-Muney (2006) empirical analysis, (5) yields $p = 141$ and $d = 4900$. It follows from (5) that testing whether a particular (α, β) belongs to the identified set is a special case of (1). Through test inversion, the results in this paper therefore allow us to construct a confidence region for the identified set that satisfies the coverage requirement proposed by Imbens and Manski (2004). Similarly, our results also apply to the dynamic discrete choice model of Honoré and Tamer (2006). ■

Example 2.4. (Discrete Choice). In their study of demand for health insurance in the California Affordable Care Act marketplace (Covered California), Tebaldi et al. (2019) assume a consumer's utility for plan j equals $V_j - \mathsf{p}_j$, where $V = (V_1, \ldots, V_J)$ is an unobserved vector of valuations and $\mathsf{p} \equiv (\mathsf{p}_1, \ldots, \mathsf{p}_J)$ denotes post-subsidy prices. In Covered California, post-subsidy prices satisfy $\mathsf{p} = \pi(C)$ for some known function π and C a (discrete-valued) vector of individual characteristics. By decomposing C into subvectors (W, S) and assuming V is independent of S conditional on W, Tebaldi et al. (2019) show there is a finite partition $\mathbb{V}$ of $\mathbf{R}^J$ and known sets $\mathcal{V}_j(\mathsf{p})$ such that observed plan choice Y satisfies

$$P(Y = j|C = c) = \sum_{\mathcal{V}\in\mathbb{V}:\mathcal{V}\subseteq\mathcal{V}_j(\pi(c))} \int_{\mathcal{V}} f_{V|W}(v|w)dv \quad (6)$$

for $f_{V|W}$ the density of V conditional on W. Moreover, counterfactuals such as the change in consumer surplus due to a change in subsidies, can be written as

$$\sum_{\mathcal{V}\in\mathbb{V}} a(\mathcal{V}) \int_{\mathcal{V}} f_{V|W}(v|w)dv \quad (7)$$

for known function a. Arguing as in Example 2.1, it then follows from (6) and (7) that confidence regions for the desired counterfactuals may be obtained through test inversion of hypotheses as in (1). In Tebaldi et al. (2019), the corresponding matrix A has dimensions as high as 253×15000. ■

Example 2.5. (Revealed Preferences). Building on McFadden and Richter (1990), Kitamura and Stoye (2018) develop a nonparametric specification test for a random utility model (RUM). In the simplest setting they study, Kitamura and Stoye (2018) suppose there are K goods and for each individual we observe the prices $\mathsf{p} \in \mathbf{R}^K$ they faced, their budget set $\mathcal{B}(\mathsf{p}) \equiv \{y \in \mathbf{R}^K_+ : \mathsf{p}'y = 1\}$, and their chosen consumption bundle $Y \in \mathcal{B}(\mathsf{p})$. Under the assumption that p has discrete support $\{\mathsf{p}_1, \ldots, \mathsf{p}_J\}$, the authors build a finite partition $\mathbb{V}$ of $\bigcup_{j=1}^J \mathcal{B}(\mathsf{p}_j)$ which they use to show the distribution P of (Y, p) is compatible with RUM if and only if $\beta(P) = Ax$ for some $x \geq 0$ – here, each coordinate of $\beta(P)$ equals $P(Y \in \mathcal{V} | \mathsf{p} = \mathsf{p}_j)$ for some $\mathcal{V} \in \mathbb{V}$, each column of A represents a rationalizable non-stochastic demand system, and x represents a vector of probabilities. Kitamura and Stoye (2018) propose a test for (1) and implement it using the U.K. Family Expenditure Survey – an application in which p and d can be as large as 79 and 313440. We note, however, that the arguments establishing the asymptotic validity of their test rely on a key restriction on A: That $(a_1 - a_0)'(a_2 - a_0) \geq 0$ for any distinct column vectors (a_0, a_1, a_2) of A. While this restriction is satisfied in the application that motivates Kitamura and Stoye (2018) and related work (Manski, 2014; Deb et al., 2017; Lazzati et al., 2018), it can fail in our previous examples. ■

3 Geometry of the Null Hypothesis

In this section, we obtain a geometric characterization of the condition that a $\beta \in \mathbf{R}^p$ satisfies $\beta = Ax$ for some $x \geq 0$. This characterization yields an alternative formulation of the null hypothesis that guides the construction of our test.

We view A as a linear map from $\mathbf{R}^d$ to $\mathbf{R}^p$, with range $R \equiv \{b \in \mathbf{R}^p : b = Ax \text{ for some } x \in \mathbf{R}^d\}$ and null space $N \equiv \{x \in \mathbf{R}^d : Ax = 0\}$. It will also be helpful to introduce the orthocomplement to N, which we denote by $N^\perp \equiv \{y \in \mathbf{R}^d : \langle y, x \rangle = 0 \text{ for all } x \in N\}$, where $\langle v, u \rangle \equiv v'u$ for any vectors v, u. The orthocomplement $N^\perp$ satisfies the following well-known property.

Lemma 3.1. *If $\beta \in R$, then there is a unique $x^\star \in N^\perp$ satisfying $\beta = Ax^\star$.*

Provided $\beta \in R$, Lemma 3.1 and the orthogonality of N and $N^\perp$ imply that the set of solutions to $Ax = \beta$ is given by $x^\star + N$. Hence, the restriction that

$\beta = Ax$ for some $x \geq 0$ is equivalent to two conditions: (i) $\beta \in R$ and (ii) $x^\star + N$ intersects the positive orthant, $\mathbf{R}^d_+$; i.e., (i) ensures *some* solution to the equation $Ax = \beta$ exists, while (ii) ensures a (weakly) *positive* solution exists. Our next result shows these requirements are equivalent to a system of linear equalities (in $\mathbf{R}^p$) and inequalities (in $\mathbf{R}^d$) whose validity is computationally easier to verify.

Theorem 3.1. *For any $\beta \in \mathbf{R}^p$ there exists an $x \geq 0$ satisfying $Ax = \beta$ if and only if $\beta \in R$ and $\langle s, x^\star \rangle \leq 0$ for all $s \in N^\perp \cap \mathbf{R}^d_-$.*

The proof of Theorem 3.1 is based on Farkas' Lemma, which can be viewed as an implication of the separating hyperplane theorem. Observe that $Ax = \beta$ for some $x \geq 0$ if and only if there *does not* exist a hyperplane that separates β from the conic hull generated by the columns of A (i.e., $\{Ax : x \geq 0\}$). Theorem 3.1 translates this requirement (in $\mathbf{R}^p$) into the requirement that there does not exist a vector s in the range of A' (equivalently, $N^\perp$) that separates $x^\star$ from the positive orthant (in $\mathbf{R}^d$).

4 The Test

The results in Section 3 imply that the null hypothesis in (1) holds if and only if $\beta(P) \in R$ and the unique element $x^\star(P) \in N^\perp$ solving $\beta(P) = A(x^\star(P))$ satisfies $\langle s, x^\star(P) \rangle \leq 0$ for all $s \in N^\perp \cap \mathbf{R}^d_-$. Based on this characterization, we next develop a test that is computationally feasible when p and/or d are large.

4.1 The Test Statistic

In what follows, we let $A^\dagger$ denote the Moore-Penrose pseudoinverse of A, which is a $d \times p$ matrix implicitly defined for any $b \in \mathbf{R}^p$ through the optimization problem

$$A^\dagger b \equiv \arg\min_{x \in \mathbf{R}^d} \|x\|_2 \text{ s.t. } x \in \arg\min_{\tilde{x} \in \mathbf{R}^d} \|A\tilde{x} - b\|_2;$$

i.e., $A^\dagger b$ is the minimum norm minimizer of $\|Ax - b\|_2$ over x, where $\|v\|_q \equiv (\sum_{i=1}^k |v_i|^q)^{1/q}$ for any vector $v \equiv (v_1, \ldots, v_k)'$ and $1 \leq q \leq \infty$. Importantly, $A^\dagger b$ is defined even if there is no $x \in \mathbf{R}^d$ satisfying $Ax = b$ or the solution is not unique. It is also helpful to note that $A^\dagger$ is a map from $\mathbf{R}^p$ onto $N^\perp$ (Luenberger, 1969).

We assume that there is an estimator $\hat{\beta}_n$ of $\beta(P)$ that is constructed from an i.i.d. sample $\{Z_i\}_{i=1}^n$ with $Z_i \in \mathbf{Z}$ distributed according to $P \in \mathbf{P}$. Since $\beta(P) \in R$ under the null hypothesis, Lemma 3.1 implies $x^\star(P) = A^\dagger \beta(P)$ for any $P \in \mathbf{P}_0$. This observation suggests employing $\hat{x}_n^\star = A^\dagger \hat{\beta}_n$ as an estimator for $x^\star(P)$, which we focus on for ease of exposition. We note, however, that when $d < p$, the analysis in Chen and Santos (2018) implies employing such an estimator may be inefficient. In such instances, it may be preferable to implement our test with alternative minimum distance estimators of $\hat{x}_n^\star$; see Fang et al. (2021) for details.

To devise a test based on Theorem 3.1 we first note that since the range of $A^\dagger$ equals $N^\perp$, the condition $\langle s, x^\star(P)\rangle \leq 0$ for all $s \in N^\perp \cap \mathbf{R}^d_-$ is equivalent to

$$\langle A^\dagger s, x^\star(P)\rangle \leq 0 \text{ for all } s \in \mathbf{R}^p \text{ s.t. } A^\dagger s \leq 0 \text{ (in } \mathbf{R}^d). \tag{8}$$

Therefore, to detect violations of condition (8) we introduce the statistic

$$\sup_{s \in \mathcal{V}^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle \quad \text{where} \quad \mathcal{V}^{\mathrm{i}} \equiv \{s \in \mathbf{R}^p : A^\dagger s \leq 0 \text{ and } \|\Omega^{\mathrm{i}}(AA')^\dagger s\|_1 \leq 1\}. \tag{9}$$

Here, Ω^{i} is a $p \times p$ symmetric matrix and the "i" superscript alludes to the relation to the "inequality" condition in Theorem 3.1 (i.e., (8)). The inclusion of a norm constraint in $\mathcal{V}^{\mathrm{i}}$ ensures the statistic in (9) is not infinite with positive probability. The introduction of Ω^{i} in (9) provides flexibility in the family of test statistics we examine. For ease of exposition, we focus on the case in which Ω^{i} is not stochastic. Our results can be extended to allow Ω^{i} to depend on the data, as is needed to let Ω^{i} be an estimator of the asymptotic standard deviation of $\sqrt{n}A\hat{x}_n^\star$ – a choice that we employ in simulations because it ensures (9) is scale-invariant. We state the additional assumptions required for such an extension at the start of the Appendix and refer the reader to Fang et al. (2021) for the relevant analysis.

By Theorem 3.1, any $P \in \mathbf{P}_0$ must satisfy $\beta(P) \in R$ in addition to (8). To detect violations of this second requirement, we introduce the statistic

$$\sup_{s \in \mathcal{V}^{\mathrm{e}}} \sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}_n^\star\rangle \quad \text{where} \quad \mathcal{V}^{\mathrm{e}} \equiv \{s \in \mathbf{R}^p : \|\Omega^{\mathrm{e}} s\|_1 \leq 1\}. \tag{10}$$

Here, Ω^{e} is a $p \times p$ symmetric matrix and the "e" superscript alludes to the relation to the "equality" condition in Theorem 3.1 (i.e., $\beta(P) \in R$). In particular, note that if $\Omega^{\mathrm{e}} = I_p$, then (10) equals $\sqrt{n}\|\hat{\beta}_n - A\hat{x}_n^\star\|_\infty$. For ease of exposition we again assume Ω^{e} to be non-stochastic. In applications in which $d \geq p$ and A is full rank, the requirement $\beta(P) \in R$ is automatically satisfied and (10) is identically zero.

For our test statistic T_n, we use the maximum of the statistics in (9) and (10):

$$T_n \equiv \max\{\sup_{s\in\mathcal{V}^{\mathrm{e}}} \sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}_n^\star\rangle, \sup_{s\in\mathcal{V}^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle\}, \tag{11}$$

which can be computed through linear programming. A variety of alternative test statistics can of course be motivated by Theorem 3.1. A couple of remarks are therefore in order as to why our interest on high-dimensional applications has led us to employ T_n. Focusing on (9) for conciseness, note that it is a special case of

$$\sup_{s\in\mathbf{R}^p} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle \text{ s.t. } A^\dagger s \leq 0 \text{ and } \omega(s) \leq 1, \tag{12}$$

where ω is a convex weight function satisfying $\omega(s) = \omega(-s)$, $\omega(s) \geq 0$, and $\omega(\gamma s) = \gamma\omega(s)$ for any $\gamma \geq 0$ – e.g., to recover (9) set $\omega(s) = \|\Omega^{\mathrm{i}}(AA')^\dagger s\|_1$. The linearity of the objective and the homogeneity of ω imply that (12) in fact equals

$$\max\{0, \sup_{s\in\mathbf{R}^p} \frac{\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle}{\omega(s)} \text{ s.t. } A^\dagger s \leq 0 \text{ and } \omega(s) > 0\}. \tag{13}$$

Representation (13) shows that (12) implicitly weights each term $\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle$ while remaining computationally tractable – i.e., (12) can be computed by convex programming, while (13) cannot. For instance, if we set $\omega(s) = \|\Omega^{\mathrm{i}}(AA')^\dagger s\|_2$ with Ω^{i} the standard deviation of $\sqrt{n}A\hat{x}_n^\star$, then $\omega(s) = (\mathrm{Var}\{\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle\})^{1/2}$ and by (13) the statistic in (12) implicitly studentizes. In (9), we instead use the weighting $\omega(s) = \|\Omega^{\mathrm{i}}(AA')^\dagger s\|_1$ because: (i) It ensures (9) is a linear program, which scales better than a quadratically-constrained program; and (ii) Using a $\|\cdot\|_1$-constraint allows us to obtain distributional approximations using coupling arguments under $\|\cdot\|_\infty$, which are available under weaker conditions on p than under $\|\cdot\|_2$. Nonetheless, we emphasize that in certain applications, a researcher may prefer to use weighting functions such as $\omega(s) = \|\Omega^{\mathrm{i}}(AA^\dagger)s\|_2$ instead. We expect that, under suitable restrictions, a version of our test that simply replaces $\|\Omega^{\mathrm{i}}(AA')^\dagger s\|_1$ with the desired $\omega(s)$ everywhere will be asymptotically valid.

4.2 The Distribution

We next introduce assumptions that enable us to approximate the distribution of T_n. Unless otherwise stated, all quantities are allowed to depend on n.

Assumption 4.1. *(i) $\{Z_i\}_{i=1}^n$ are i.i.d. with $Z_i \in \mathbf{Z}$ and $Z_i \sim P \in \mathbf{P}$; (ii) There* *are* $\psi^{\mathrm{i}}(\cdot, P) : \mathbf{Z} \to \mathbf{R}^p$ *and* $\psi^{\mathrm{e}}(\cdot, P) : \mathbf{Z} \to \mathbf{R}^p$ *satisfying uniformly in* $P \in \mathbf{P}$

$$\|(\Omega^{\mathrm{e}})^{\dagger}\{(I_p - AA^{\dagger})\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{\mathrm{e}}(Z_i, P)\}\|_{\infty} = O_P(a_n)$$
$$\|(\Omega^{\mathrm{i}})^{\dagger}\{AA^{\dagger}\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{\mathrm{i}}(Z_i, P)\}\|_{\infty} = O_P(a_n).$$

Assumption 4.2. *For* $\Sigma^{\mathrm{j}}(P) \equiv E_P[\psi^{\mathrm{j}}(Z, P)\psi^{\mathrm{j}}(Z, P)']$*: (i)* $E_P[\psi^{\mathrm{j}}(Z, P)] = 0$ *for all* $P \in \mathbf{P}$ *and* $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$*; (ii) The eigenvalues of* $(\Omega^{\mathrm{j}})^{\dagger}\Sigma^{\mathrm{j}}(P)(\Omega^{\mathrm{j}})^{\dagger}$ *are bounded in* $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$*,* n*, and* $P \in \mathbf{P}$*; (iii)* $\Psi(z, P) \equiv \|(\Omega^{\mathrm{e}})^{\dagger}\psi^{\mathrm{e}}(z, P)\|_{\infty} \vee \|(\Omega^{\mathrm{i}})^{\dagger}\psi^{\mathrm{i}}(z, P)\|_{\infty}$ *satisfies* $\sup_{P \in \mathbf{P}} \|\Psi(\cdot, P)\|_{P,3} \leq M_{3,\Psi} < \infty$ *with* $M_{3,\Psi} \geq 1$*; (iv)* $\Omega^{\mathrm{i}}, \Omega^{\mathrm{e}}$ *are symmetric.*

Assumption 4.3. *(i)* $\psi^{\mathrm{j}}(Z, P) \in \mathrm{range}\{\Omega^{\mathrm{j}}\}$ *P-almost surely for all* $P \in \mathbf{P}$ *and* $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$*; (ii)* $(I_p - AA^{\dagger})\{\hat{\beta}_n - \beta(P)\} \in \mathrm{range}\{\Sigma^{\mathrm{e}}(P)\}$ *and* $AA^{\dagger}\{\hat{\beta}_n - \beta(P)\} \in \mathrm{range}\{\Sigma^{\mathrm{i}}(P)\}$ *with probability tending to one uniformly in* $P \in \mathbf{P}$.

Assumption 4.1(ii) requires our estimators to be asymptotically linear with influence functions whose moments are disciplined by Assumption 4.2(i)-(iii). In Assumption 4.2(iv) we additionally impose that the weighting matrices be symmetric. Finally, Assumption 4.3(i) is a mild regularity condition that ensures the approximating distribution is not infinite with positive probability. For similar reasons, Assumption 4.3(ii) ensures that the supports of our estimators are contained in the supports of their Gaussian approximations.

For $\psi^{\mathrm{e}}(Z, P)$ and $\psi^{\mathrm{i}}(Z, P)$ the influence functions in Assumption 4.1(ii) we set $\psi(Z, P) \equiv (\psi^{\mathrm{e}}(Z, P)', \psi^{\mathrm{i}}(Z, P)')'$ and let $\Sigma(P) \equiv E_P[\psi(Z, P)\psi(Z, P)']$. For notational simplicity we also define the rate $r_n \equiv M_{3,\Psi}(p^2 \log^5(1 + p)/n)^{1/6} + a_n$. Our next theorem gives a distributional approximation for T_n that, under appropriate moment conditions, is valid uniformly in $P \in \mathbf{P}_0$ provided $p^2 \log^5(p)/n = o(1)$.

Theorem 4.1. *Let Assumptions 4.1, 4.2, 4.3 hold, and* $r_n = o(1)$*. Then, there is* $(\mathbb{G}_n^{\mathrm{e}}(P)', \mathbb{G}_n^{\mathrm{i}}(P)')' \equiv \mathbb{G}_n(P) \sim N(0, \Sigma(P))$ *such that uniformly in* $P \in \mathbf{P}_0$

$$T_n = \max\{\sup_{s \in \mathcal{V}^{\mathrm{e}}} \langle s, \mathbb{G}_n^{\mathrm{e}}(P)\rangle, \sup_{s \in \mathcal{V}^{\mathrm{i}}} \langle A^{\dagger}s, A^{\dagger}\mathbb{G}_n^{\mathrm{i}}(P)\rangle + \sqrt{n}\langle A^{\dagger}s, A^{\dagger}\beta(P)\rangle\} + O_P(r_n).$$

The asymptotic approximation in Theorem 4.1 depends on linear programs whose solutions must be attained at one of a finite number of extreme points. It follows that T_n is asymptotically equivalent to the maximum of a Gaussian vector – an observation that suggests a connection to the high-dimensional central limit theorem of Chernozhukov et al. (2019). The proof of Theorem 4.1, however, does not rely on Chernozhukov et al. (2019) because the number of extreme points depends on A in a non-transparent way and upper bounds, such as that in Mc-Mullen (1970), are exponential in p. Nonetheless, we note that for certain A and $\hat{\beta}_n$, Chernozhukov et al. (2019) may yield better coupling rates than Theorem 4.1 and allow p to be larger than n. On the other hand, we should not expect such conditions to apply when $\hat{\beta}_n$ is a vector of empirical probabilities as in Examples 2.2-2.5 – a setting we expect to at least require $p/n = o(1)$.

4.3 The Critical Value

To obtain a critical value, we assume the availability of "bootstrap" estimates $(\hat{\mathbb{G}}_n^{\mathrm{e}\prime}, \hat{\mathbb{G}}_n^{\mathrm{i}\prime})'$ for the distribution of $(\mathbb{G}_n^{\mathrm{e}}(P)', \mathbb{G}_n^{\mathrm{i}}(P)')'$. Given such estimates, we may follow a number of approaches for obtaining critical values; see, e.g., Section 4.4.1. Below we focus on an approach that has favorable power properties in simulations.

Step 1. First, we observe that the main challenge in employing Theorem 4.1 for inference is the presence of the nuisance function $f(\cdot, P) : \mathbf{R}^p \to \mathbf{R}$ given by

$$f(s, P) \equiv \sqrt{n}\langle A^\dagger s, A^\dagger \beta(P)\rangle. \tag{14}$$

While $f(\cdot, P)$ cannot be consistently estimated, we can construct a suitable upper bound for it. To this end, we note that in applications some coordinates of $\beta(P)$ may be known; see Section 2. We therefore decompose $\beta(P) = (\beta_{\mathrm{u}}(P)', \beta_{\mathrm{k}}')'$ where β_{k} is a known constant, and similarly decompose any $b \in \mathbf{R}^p$ into subvectors of conformable dimensions $b = (b_{\mathrm{u}}', b_{\mathrm{k}}')'$. We then define the restricted estimator

$$\hat{\beta}_n^{\mathrm{r}} \in \arg \min_{b=(b_{\mathrm{u}}', b_{\mathrm{k}}')'} \sup_{s \in \mathcal{V}^{\mathrm{i}}} |\langle A^\dagger s, \hat{x}_n^\star - A^\dagger b\rangle| \text{ s.t. } b_{\mathrm{k}} = \beta_{\mathrm{k}},\ Ax = b \text{ for some } x \geq 0, \tag{15}$$

which may be computed through linear programming; see Appendix **??**. Since $f(s, P) \leq 0$ for all $s \in \mathcal{V}^{\mathrm{i}}$ and $P \in \mathbf{P}_0$ by Theorem 3.1, it follows that under the null hypothesis $\lambda_n f(s, P) \geq f(s, P)$ for any $\lambda_n \leq 1$ and $s \in \mathcal{V}^{\mathrm{i}}$. We therefore set

$$\hat{\mathbb{U}}_n(s) \equiv \lambda_n \sqrt{n}\langle A^\dagger s, A^\dagger \hat{\beta}_n^{\mathrm{r}}\rangle, \tag{16}$$

which is a consistent estimator for $\lambda_n f(s, P)$ provided $\lambda_n \downarrow 0$ at a suitable rate – we discuss choices of λ_n in Remark 4.2. The upper bound $\hat{\mathbb{U}}_n$ reflects the structure of the null hypothesis in that: (i) $\hat{\mathbb{U}}_n(s) \leq 0$ for all $s \in \mathcal{V}^{\mathrm{i}}$ and (ii) There is a $b \in \mathbf{R}^p$

satisfying $Ax = b$ for some $x \geq 0$ such that $\hat{\mathbb{U}}_n(s) = \sqrt{n}\langle A^\dagger s, A^\dagger b\rangle$. ■

Step 2. The asymptotic approximation from Theorem 4.1 is increasing (in a first-order stochastic dominance sense) in $f(\cdot, P)$ (under the pointwise partial order). Hence, for a nominal level α test we may use the bootstrap quantile

$$\hat{c}_n(1-\alpha) \equiv \inf\{u : P(\max\{\sup_{s\in\mathcal{V}^{\mathrm{e}}}\langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle, \sup_{s\in\mathcal{V}^{\mathrm{i}}}\langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}}\rangle + \hat{\mathbb{U}}_n(s)\} \leq u | \{Z_i\}_{i=1}^n) \geq 1-\alpha\}$$

as a critical value for T_n. Computing $\hat{c}_n(1-\alpha)$ requires solving one linear program per bootstrap replication. We also note that because $0 \in \mathcal{V}^{\mathrm{i}}$, any $s \in \mathcal{V}^{\mathrm{i}}$ for which $\sqrt{n}\langle A^\dagger s, A^\dagger \beta(P)\rangle$ tends to minus infinity plays an asymptotically negligible role in the distributional approximation of Theorem 4.1. Our critical value reflects this structure because $\hat{\mathbb{U}}_n(s)$ and $\lambda_n\sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle$ are asymptotically equivalent, and thus any s for which $\lambda_n\sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle$ tends to minus infinity plays an asymptotic negligible role in determining $\hat{c}_n(1-\alpha)$. E.g., in an asymptotic setting in which P is fixed and $\lambda_n\sqrt{n} \to \infty$, any s satisfying $\langle A^\dagger s, A^\dagger\beta(P)\rangle < 0$ plays a negligible role in both the distribution of T_n and our bootstrap approximation. ■

Given the above definitions, we finally define $\phi_n \equiv 1\{T_n > \hat{c}_n(1-\alpha)\}$ as our test; i.e., we reject the null hypothesis whenever T_n exceeds $\hat{c}_n(1-\alpha)$. To establish the asymptotic validity of this test, we impose one final assumption.

Assumption 4.4. *(i) For exchangeable $\{W_{i,n}\}_{i=1}^n$ independent of $\{Z_i\}_{i=1}^n$ we have*

$$\|(\Omega^{\mathrm{j}}(P))^\dagger\{\hat{\mathbb{G}}_n^{\mathrm{j}} - \frac{1}{\sqrt{n}}\sum_{i=1}^n (W_{i,n} - \bar{W}_n)\psi^{\mathrm{j}}(Z_i, P)\}\|_\infty = O_P(a_n)$$

uniformly in $P \in \mathbf{P}$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$ and $\bar{W}_n \equiv \sum_{i=1}^n W_{i,n}/n$; (ii) For some $a, b > 0$, $P(|W_{1,n} - E[W_{1,n}]| > t) \leq 2\exp\{-t^2/(b + at)\}$ for all $t \in \mathbf{R}_+$ and n; (iii) $|\sum_{i=1}^n (W_{i,n} - \bar{W}_n)^2/n - 1| = O_P(n^{-1/2})$ and $\sup_n E[|W_{1,n}|^3] < \infty$; (iv) For some $q \in (1, +\infty]$, $\sup_{P\in\mathbf{P}} \|\Psi^2(\cdot, P)\|_{P,q} \leq M_{q,\Psi^2} < \infty$; (v) For $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$, $\hat{\mathbb{G}}_n^{\mathrm{j}} \in \mathrm{range}\{\Sigma^{\mathrm{j}}(P)\}$ with probability tending to one uniformly in $P \in \mathbf{P}$.

Assumption 4.4 accommodates a variety of resampling schemes, such as the nonparametric, Bayesian, score, or weighted bootstrap. In parallel to Assumption 4.1(ii), Assumption 4.4(i) imposes a linearization assumption on our bootstrap estimates. Assumptions 4.4(ii)(iii) state restrictions on the bootstrap weights that are satisfied by commonly used resampling schemes. Assumption 4.4(iv) strengthens the moment restrictions in Assumption 4.2(iii) (if $q > 3/2$) and is imposed to sharpen our estimates of the coupling rate for the bootstrap statistics.

Finally, Assumption 4.4(v) is a bootstrap analogue to Assumption 4.3(ii). These conditions suffice for showing that the distribution of $(\hat{\mathbb{G}}_n^{\mathrm{i}\prime}, \hat{\mathbb{G}}_n^{\mathrm{e}\prime})'$ conditional on the data is suitably consistent for the distribution of $(\mathbb{G}_n^{\mathrm{e}}(P)', \mathbb{G}_n^{\mathrm{i}}(P)')'$ at a rate

$$b_n \equiv \frac{\sqrt{p\log(1+n)}M_{3,\Psi}}{n^{1/4}} + (\frac{p\log^{5/2}(1+p)M_{3,\Psi}}{\sqrt{n}})^{1/3} + (\frac{p\log^3(1+p)n^{1/q}M_{q,\Psi^2}}{n})^{1/4} + a_n;$$

see Lemma A.5. In particular, under appropriate moment restrictions, the bootstrap is consistent provided $p^2/n = o(1)$ (up to logs). The consistency of the exchangeable bootstrap when p grows with n is to our knowledge a novel result.

Before establishing the asymptotic validity of our test, we introduce some final pieces of notation. First, we note that the asymptotic approximation obtained in Theorem 4.1 contains two linear programs, whose solutions can be shown to belong to the sets $\mathcal{E}^{\mathrm{e}} \equiv \{s \in \mathbf{R}^p : s \text{ is an extreme point of } \Omega^{\mathrm{e}}\mathcal{V}^{\mathrm{e}}\}$ and $\mathcal{E}^{\mathrm{i}} \equiv \{s \in \mathbf{R}^p : s \text{ is an extreme point of } (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}\}$. We also define $\sigma^{\mathrm{e}}(s,P) \equiv \{E_P[(\langle s, (\Omega^{\mathrm{e}})^{\dagger}\mathbb{G}_n^{\mathrm{e}}(P)\rangle)^2]\}^{1/2}$, $\sigma^{\mathrm{i}}(s,P) \equiv \{E_P[(s, \mathbb{G}_n^{\mathrm{i}}(P)\rangle)^2]\}^{1/2}$, and set

$$\bar{\sigma}(P) \equiv \max_{\mathrm{j}\in\{\mathrm{e},\mathrm{i}\}} \max_{s\in\mathcal{E}^{\mathrm{j}}} \sigma^{\mathrm{j}}(s,P) \qquad \underline{\sigma}(P) \equiv \min_{\mathrm{j}\in\{\mathrm{e},\mathrm{i}\}} \min_{s\in\mathcal{E}^{\mathrm{j}}:\sigma^{\mathrm{j}}(s,P)>0} \sigma^{\mathrm{j}}(s,P)$$

where we let $\underline{\sigma}(P) = +\infty$ if $\sigma^{\mathrm{j}}(s,P) = 0$ for all $s \in \mathcal{E}^{\mathrm{j}}$, $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$. For any random variable $V \in \mathbf{R}$, let $\mathrm{med}\{V\}$ denote its median, and for any $P \in \mathbf{P}$ define

$$\mathrm{m}(P) \equiv \mathrm{med}\{\max\{\sup_{s\in\mathcal{V}^{\mathrm{e}}}\langle s, \mathbb{G}_n^{\mathrm{e}}(P)\rangle, \sup_{s\in\mathcal{V}^{\mathrm{i}}}\langle A^{\dagger}s, A^{\dagger}\mathbb{G}_n^{\mathrm{i}}(P)\rangle\}\}.$$

Lastly, we introduce the sequence $\xi_n \equiv r_n \vee b_n \vee \lambda_n\sqrt{\log(1+p)}$. Our next result establishes the asymptotic validity of the proposed test.

Theorem 4.2. *Let Assumptions 4.1–4.4 hold, $\alpha \in (0, 0.5)$, and $0 \le \lambda_n \le 1$. If ξ_n satisfies $\xi_n = o(1)$ and $\sup_{P\in\mathbf{P}}(\mathrm{m}(P) + \bar{\sigma}(P))/\underline{\sigma}^2(P) = o(\xi_n^{-1})$, then*

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} E_P[\phi_n] \le \alpha. \tag{17}$$

Under additional requirements, it is possible to show (17) holds with equality. For instance, if p is fixed with n and $\sqrt{n}\lambda_n \to \infty$, then the limiting rejection probability of ϕ_n tends to α for any P on the "boundary" of $\mathbf{P}_0$ – a result that, together with Theorem 4.2, implies the asymptotic size of our test equals α. We also note that Theorem 4.2 imposes a rate condition that constrains how p can grow with n. This rate condition depends on A and the weighting matrices Ω^{e}

and Ω^{i}. As we show in Remark 4.1 below, it is possible to obtain universal (in A) bounds for $(\mathrm{m}(P)+\bar{\sigma}(P))/\underline{\sigma}^2(P)$ when setting Ω^{j} to be the standard deviation matrix of $\mathbb{G}_n^{\mathrm{j}}(P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$. While such bounds provide sufficient conditions for the rate requirements in Theorem 4.2, we emphasize that they can be quite conservative for a specific A. Finally, we note that if, as in much of the literature, one considers the case in which p is fixed with n, then Remark 4.1 implies Theorem 4.2 holds under Assumptions 4.1–4.4 and the requirements $\lambda_n \downarrow 0$ and $a_n = o(1)$.

Remark 4.1. Whenever Ω^{j} equals the standard deviation matrix of $\mathbb{G}_n^{\mathrm{j}}(P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$, it is possible to obtain universal (in A) bounds on $\bar{\sigma}(P)$, $\underline{\sigma}(P)$, and $\mathrm{m}(P)$. In particular, under such choice of Ω^{j}, we have $p^{-1/2} \leq \underline{\sigma}(P) \leq \bar{\sigma}(P) \leq 1$ and $\mathrm{m}(P) \lesssim \sqrt{\log(1+p)}$. The resulting universal (in A) bound for $(\mathrm{m}(P) + \bar{\sigma}(P))/\underline{\sigma}^2(P)$, however, may be quite conservative for a specific A. ■

Remark 4.2. In simulations we find two choice of λ_n to perform well. The first is to set Ω^{i} to be an estimate of the standard deviation matrix of $\mathbb{G}_n^{\mathrm{i}}(P)$ and, based on the law of iterated logartihm, set $\lambda_n^{\mathrm{r}} = 1/\sqrt{\log(e \vee p)\log(e \vee \log(e \vee n))}$. The second is to set $\lambda_n^{\mathrm{b}} = \min\{1, 1/\hat{\tau}_n\}$ where $\hat{\tau}_n$ is the $1 - 1/\sqrt{\log(e \vee \log(e \vee n))}$ bootstrap quantile of $\sup_{s \in \mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}} \rangle$, which typically is smaller than λ_n^{r}. ■

4.4 Extensions

4.4.1 Two Stage Critical Value

We have focused on a particular choice of critical value due to its favorable power properties in simulations. An alternative critical value may be obtained by following Romano et al. (2014) and Bai et al. (2019). Specifically, we may set

$$\hat{c}_n^{(1)}(1-\gamma) \equiv \inf\{u : P(\sup_{s \in \mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, -A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}} \rangle \leq u | \{Z_i\}_{i=1}^n) \geq 1 - \gamma\}$$

for some $\gamma \in (0, \alpha)$ and define $\tilde{\mathbb{U}}_n(s) \equiv \min\{\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star \rangle + \hat{c}_n^{(1)}(1-\gamma), 0\}$. The function $\tilde{\mathbb{U}}_n$ is an upper confidence region for $f(\cdot, P)$ (as in (14)) with uniform (in $P \in \mathbf{P}_0$) asymptotic coverage probability $1 - \gamma$. For a nominal level α test, we may then compare T_n to the $1 - \alpha + \gamma$ bootstrap quantile of

$$\max\{\sup_{s \in \mathcal{V}^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}} \rangle, \sup_{s \in \mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}} \rangle + \tilde{\mathbb{U}}_n(s)\}.$$

An appealing feature of the described test is that it does not require selecting λ_n. However, in simulations we find its power is lower than that of ϕ_n. Intuitively, this is due to $\tilde{\mathbb{U}}_n$ not reflecting the structure of the null hypothesis in that it *does not* satisfy $\tilde{\mathbb{U}}_n(s) = \sqrt{n}\langle A^\dagger s, A^\dagger b\rangle$ for some b such that $Ax = b$ with $x \geq 0$.

4.4.2 Alternative Sampling Frameworks

While we have focused on i.i.d. settings, we note that extensions to other asymptotic frameworks can be straightforward. Consider, for example, Andrews et al. (2019), who study a class of models in which the parameter of interest π satisfies

$$E_P[G(D,\pi) - M(V,\pi)\delta|V] \leq 0 \text{ for some } \delta \in \mathbf{R}^{d_\delta} \tag{18}$$

where $G(D,\pi) \in \mathbf{R}^p$, $M(V,\pi)$ is a $p \times d_\delta$ matrix, and both are known functions of (D, V, π). Andrews et al. (2019) observe that testing whether a specified value π_0 satisfies (18) is facilitated by conditioning on $\{V_i\}_{i=1}^n$. Similarly, setting $\delta^+ \equiv \delta \vee 0$ and $\delta_- \equiv -(\delta \wedge 0)$ for any $\delta \in \mathbf{R}^{d_\delta}$, we note that if π_0 satisfies (18), then

$$\frac{1}{n}\sum_{i=1}^n E_P[G(D,\pi_0)|V_i] = \frac{1}{n}\sum_{i=1}^n M(V_i,\pi_0)(\delta^+ - \delta^-) - \Delta \text{ for some } \Delta \in \mathbf{R}^p_+,\ \delta \in \mathbf{R}^{d_\delta},$$

which may be mapped into (1) after conditioning on $\{V_i\}_{i=1}^n$ by setting $\beta(P) \equiv \sum_i E_P[G(D,\pi_0)|V_i]/n$ – note A does not depend on P due to the conditioning on $\{V_i\}_{i=1}^n$. By letting $\hat{\beta}_n \equiv \sum_{i=1}^n G(D_i,\pi_0)/n$, our test remains largely the same, with the exception that $(\hat{\mathbb{G}}_n^{\mathrm{e}\prime}, \hat{\mathbb{G}}_n^{\mathrm{i}\prime})'$ must be consistent for the law of $((\Omega^{\mathrm{e}})^\dagger(I_p - AA^\dagger\hat{C}_n)\sqrt{n}\{\hat{\beta}_n - \beta(P)\}', (\Omega^{\mathrm{i}})^\dagger AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}')'$ conditional on $\{V_i\}_{i=1}^n$.

Appendix

While we assumed Ω^{i} and Ω^{e} to be non-stochastic for ease of exposition, we note Theorems 4.1 and 4.2 can be extended to allow for estimated weights. Writing $\Omega^{\mathrm{i}}(P)$ and $\Omega^{\mathrm{e}}(P)$, such an extension requires that Assumptions 4.2-4.3 hold uniformly in P with Ω^{j} replaced by $\Omega^{\mathrm{j}}(P)$ and that their estimators $\hat{\Omega}_n^{\mathrm{j}}$ satisfy:

Assumption A.1. *For* $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$*: (i)* $\hat{\Omega}_n^{\mathrm{j}}$ *is symmetric; (ii) There is a symmetric matrix* $\Omega^{\mathrm{j}}(P)$ *satisfying* $\|(\Omega^{\mathrm{j}}(P))^\dagger(\hat{\Omega}_n^{\mathrm{j}} - \Omega^{\mathrm{j}}(P))\|_{o,\infty} = O_P(a_n/\sqrt{\log(1+p)})$ *uniformly in* $P \in \mathbf{P}$*; (iii)* $\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P(\mathrm{range}\{\hat{\Omega}_n^{\mathrm{j}}\} = \mathrm{range}\{\Omega^{\mathrm{j}}(P)\}) = 1$.

Proof of Lemma 3.1: Follows from result 3F in Strang (1993) and $N^\perp$ equaling the range of A' by Theorem 6.6.2 in Luenberger (1969). ■

Proof of Theorem 3.1: Fix any $\beta \in R$ and note that by Farkas' Lemma (see, e.g., Corollary 5.85 in Aliprantis and Border (2006)) $\beta = A\tilde{x}$ for some $\tilde{x} \geq 0$ if and only if there *does not* exist a $y \in \mathbf{R}^p$ satisfying $A'y \leq 0$ (in $\mathbf{R}^d$) and $\langle y, \beta\rangle > 0$. In particular, the condition $\beta = A\tilde{x}$ for some $\tilde{x} \geq 0$ is equivalent to

$$\langle y, \beta\rangle \leq 0 \text{ for all } y \in \mathbf{R}^p \text{ such that } A'y \leq 0 \text{ (in } \mathbf{R}^d). \tag{A.1}$$

Next note there is a unique $x^\star \in N^\perp$ satisfying $\beta = Ax^\star$ by Lemma 3.1. Therefore, $\langle y, Ax^\star\rangle = \langle A'y, x^\star\rangle$, $\{A'y : y \in \mathbf{R}^p \text{ and } A'y \leq 0\} = \text{range}\{A'\} \cap \mathbf{R}^d_-$, and $\text{range}\{A'\} = N^\perp$ by Theorem 6.6.3 in Luenberger (1969) imply

$$\langle s, x^\star\rangle \leq 0 \text{ for all } s \in N^\perp \cap \mathbf{R}^d_- \tag{A.2}$$

is equivalent to (A.1). In summary, if $\beta \in R$, then $\beta = A\tilde{x}$ for some $\tilde{x} \geq 0$ if and only if (A.2) holds and hence the theorem follows. ■

Proof of Theorem 4.1: First note that since $\beta(P) \in R$ for all $P \in \mathbf{P}_0$ by Theorem 3.1, we have $AA^\dagger\beta(P) = \beta(P)$ and therefore $\hat{x}^\star_n = A^\dagger\hat{\beta}_n$ yields that

$$\sup_{s\in\mathcal{V}^{\mathrm{e}}} \sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}^\star_n\rangle = \sup_{s\in\mathcal{V}^{\mathrm{e}}}\langle s, (I_p - AA^\dagger)\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle \tag{A.3}$$

for all $P \in \mathbf{P}_0$. Similarly, employing that $A^\dagger AA^\dagger = A^\dagger$ (see Proposition 6.11.1(5) in Luenberger (1969)) together with $\hat{x}^\star_n = A^\dagger\hat{\beta}_n$ implies for all $P \in \mathbf{P}_0$ that

$$\sup_{s\in\mathcal{V}^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, \hat{x}^\star_n\rangle = \sup_{s\in\mathcal{V}^{\mathrm{i}}}\langle A^\dagger s, A^\dagger AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle. \tag{A.4}$$

Next note that Assumption 4.3(i) implies $\text{range}\{\Sigma^{\mathrm{j}}(P)\} \subseteq \text{range}\{\Omega^{\mathrm{j}}\}$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$. Therefore, Assumption 4.3(ii) implies $(I_p - AA^\dagger)\sqrt{n}\{\hat{\beta}_n - \beta(P)\} \in \text{range}\{\Omega^{\mathrm{e}}\}$ and $AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\} \in \text{range}\{\Omega^{\mathrm{i}}\}$ with probability tending to one uniformly in $P \in \mathbf{P}$. Hence, by Lemma A.4 we may apply Lemma A.2 with $\hat{\mathbb{W}}^{\mathrm{e}}_n(P) = (I_p - AA^\dagger)\sqrt{n}\{\hat{\beta}_n - \beta(P)\}$, $\hat{\mathbb{W}}^{\mathrm{i}}_n(P) = AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\}$, and $\hat{f}_n(s, P) = \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle$, which together with results (A.3) and (A.4) establish the claim of the theorem. ■

Proof of Theorem 4.2: For notational simplicity we first set $\eta \equiv 1 - \alpha$ and define

$$\mathbb{M}_n(s, P) \equiv \langle A^\dagger s, A^\dagger\mathbb{G}^{\mathrm{i}}_n(P)\rangle \qquad \mathbb{U}_n(s, P) \equiv \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle \tag{A.5}$$

$$\mathbb{A}^{\mathrm{e}}_n(s, P) \equiv \langle s, (\Omega^{\mathrm{e}})^\dagger\mathbb{G}^{\mathrm{e}}_n(P)\rangle \qquad \mathbb{A}^{\mathrm{i}}_n(s, P) \equiv \langle s, \mathbb{G}^{\mathrm{i}}_n(P) + \sqrt{n}\beta(P)\rangle, \tag{A.6}$$

and set sequences $\ell_n \downarrow 0$ and $\tau_n \uparrow 1$ to satisfy $r_n \vee b_n \vee \lambda_n \sqrt{\log(1+p)} = o(\ell_n)$ and

$$\sup_{P\in\mathbf{P}} (\mathrm{m}(P) + \bar{\sigma}(P) z_{\tau_n})/\underline{\sigma}^2(P) = o(\ell_n^{-1}). \tag{A.7}$$

where z_{τ_n} is the τ_n quantile of $N(0,1)$. Also set $\epsilon > 0$ to satisfy $\eta - \epsilon > 0.5$ and

$$E_n(P) \equiv \{\hat{c}_n(\eta) \geq (\underline{\sigma}(P) z_{\eta-\epsilon})/2,\ \mathbb{U}_n(s,P) \leq \hat{\mathbb{U}}_n(s) + \ell_n \text{ for all } s \in \mathcal{V}^{\mathrm{i}}\}. \tag{A.8}$$

Next, note that $0 \in \mathcal{V}^{\mathrm{e}}$ and $0 \in \mathcal{V}^{\mathrm{i}}$ imply $\hat{c}_n(\eta) \geq 0$. Therefore, $\phi_n = 1$ implies $T_n > 0$, which together with Lemma A.6 implies the theorem is immediate on $\mathbf{D}_0 \equiv \{P \in \mathbf{P}_0 : \sigma^{\mathrm{j}}(s,P) = 0 \text{ for all } s \in \mathcal{E}^{\mathrm{j}} \text{ and all j} \in \{\mathrm{e},\mathrm{i}\}\}$. Hence, without loss of generality, we assume that for all $P \in \mathbf{P}_0$, $\sigma^{\mathrm{j}}(s,P) > 0$ for some $s \in \mathcal{E}^{\mathrm{j}}$ and $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$. Furthermore, since $\phi_n = 1$ implies $T_n > 0$, Lemmas A.1 and A.3 yield

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\phi_n = 1) = \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(T_n > \hat{c}_n(\eta);\ E_n(P)). \tag{A.9}$$

Moreover, for $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$, $\mathbb{G}_n^{\mathrm{j}}(P) \in \mathrm{range}\{\Sigma^{\mathrm{j}}(P)\} \subseteq \mathrm{range}\{\Omega^{\mathrm{j}}\}$ by Theorem 3.6.1 in Bogachev (1998) and Assumption 4.3(i). Hence, $\Omega^{\mathrm{j}}(\Omega^{\mathrm{j}})^{\dagger}\mathbb{G}_n^{\mathrm{j}}(P) = \mathbb{G}_n^{\mathrm{j}}(P)$ for $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$, which together with Hölder's inequality, symmetry of Ω^{j}, the definitions of $\mathcal{V}^{\mathrm{e}}$ and $\mathcal{V}^{\mathrm{i}}$, and $\mathbb{U}_n(s,P) \leq 0$ for $s \in \mathcal{V}^{\mathrm{i}}$ and $P \in \mathbf{P}_0$ by Theorem 3.1 imply

$$\sup_{s\in\mathcal{V}^{\mathrm{e}}} \langle s, \mathbb{G}_n^{\mathrm{e}}(P)\rangle = \sup_{s\in\mathcal{V}^{\mathrm{e}}} \langle \Omega^{\mathrm{e}} s, (\Omega^{\mathrm{e}})^{\dagger}\mathbb{G}_n^{\mathrm{e}}(P)\rangle < \infty$$
$$\sup_{s\in\mathcal{V}^{\mathrm{i}}} \mathbb{M}_n(s,P) + \mathbb{U}_n(s,P) = \sup_{s\in\mathcal{V}^{\mathrm{i}}} \langle \Omega^{\mathrm{i}}(AA')^{\dagger} s, (\Omega^{\mathrm{i}})^{\dagger}\mathbb{G}_n^{\mathrm{i}}(P)\rangle + \mathbb{U}_n(s,P) < \infty.$$

Thus, by Theorem 4.1 and Lemmas A.10, A.11 we obtain uniformly in $P \in \mathbf{P}_0$

$$T_n = \max_{s\in\mathcal{E}^{\mathrm{e}}} \mathbb{A}_n^{\mathrm{e}}(s,P) \vee \max_{s\in\mathcal{E}^{\mathrm{i}}} \mathbb{A}_n^{\mathrm{i}}(s,P) + O_P(r_n). \tag{A.10}$$

Next define $c_n^{(1)}(\tau,P) \equiv \inf\{u : P(\sup_{s\in\mathcal{V}^{\mathrm{i}}} \mathbb{M}_n(s,P) \leq u) \geq \tau\}$ for any $\tau \in (0,1)$ and set $\mathcal{E}^{\mathrm{i},\tau}(P) \equiv \{s \in \mathcal{E}^{\mathrm{i}} : -\langle s, \sqrt{n}\beta(P)\rangle \leq c_n^{(1)}(\tau,P)\}$. Since $0 \in \mathcal{V}^{\mathrm{i}}$ implies $c_n^{(1)}(\tau,P) \geq 0$, Lemma A.11 yields $0 \in \mathcal{E}^{\mathrm{i},\tau}(P)$ and therefore we have

$$\begin{aligned} P(\max_{s\in\mathcal{E}^{\mathrm{i}}} \mathbb{A}_n^{\mathrm{i}}(s,P) &= \max_{s\in\mathcal{E}^{\mathrm{i},\tau}(P)} \mathbb{A}_n^{\mathrm{i}}(s,P)) \\ &\geq P(\max_{s\in\mathcal{E}^{\mathrm{i}}\setminus\mathcal{E}^{\mathrm{i},\tau}(P)} \mathbb{A}_n^{\mathrm{i}}(s,P) \leq 0) \geq P(\sup_{s\in\mathcal{V}^{\mathrm{i}}} \mathbb{M}_n(s,P) \leq c_n^{(1)}(\tau,P)) \geq \tau, \end{aligned} \tag{A.11}$$

where the second and final inequalities hold by definitions of $\mathcal{E}^{\mathrm{i},\tau}(P)$ and $c_n^{(1)}(\tau,P)$, and $\mathcal{E}^{\mathrm{i}} \subseteq (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$. Next define the sets $\mathcal{C}_n(\mathrm{e},P) \equiv \mathcal{E}^{\mathrm{e}}$ and $\mathcal{C}_n(\mathrm{i},P) \equiv \mathcal{E}^{\mathrm{i},\tau_n}(P)$ and note that results (A.9), (A.10), (A.11), $\tau_n \uparrow 1$, and $r_n = o(\ell_n)$ imply that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\phi_n = 1) \le \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\max_{\mathrm{j}\in\{\mathrm{e},\mathrm{i}\}} \max_{s\in\mathcal{C}_n(\mathrm{j},P)} \mathbb{A}_n^{\mathrm{j}}(s,P) > \hat{c}_n(\eta) - \ell_n;\ E_n(P)). \tag{A.12}$$

Defining $\mathcal{A}_n(P) \equiv \{(\mathrm{j}, s) : \mathrm{j} \in \{\mathrm{e},\mathrm{i}\},\ \ s \in \mathcal{C}_n(\mathrm{j}, P),\ \sigma^{\mathrm{j}}(s, P) > 0\}$, then note that, by (A.7), $E_n(P)$ implies $\hat{c}_n(\eta) - \ell_n > 0$ for n sufficiently large. Since $E[\mathbb{A}_n^{\mathrm{e}}(s,P)] = 0$ for all $s \in \mathcal{E}^{\mathrm{e}}$, while $E[\mathbb{A}_n^{\mathrm{i}}(s,P)] \le 0$ for all $s \in \mathcal{E}^{\mathrm{i},\tau_n}(P)$ due to $\langle (AA')^{\dagger}s, \beta(P)\rangle \le 0$ for all $s \in \mathcal{V}^{\mathrm{i}}$ by Theorem 3.1, (A.12) yields the theorem if $\mathcal{A}_n(P) = \emptyset$. Hence, assuming without loss of generality $\mathcal{A}_n(P) \neq \emptyset$, we obtain

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\phi_n = 1) \le \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) > \hat{c}_n(\eta) - \ell_n;\ E_n(P)). \tag{A.13}$$

By Lemma A.5 there is a $(\mathbb{G}_n^{\mathrm{e}\star}(P)', \mathbb{G}_n^{\mathrm{i}\star}(P)')' \equiv \mathbb{G}_n^{\star}(P) \sim N(0, \Sigma(P))$, independent of $\{Z_i\}_{i=1}^n$, and satisfying $\|(\Omega^{\mathrm{j}})^{\dagger}\{\hat{\mathbb{G}}_n^{\mathrm{j}} - \mathbb{G}_n^{\mathrm{j}\star}\}\|_\infty = O_P(b_n)$ uniformly in $P \in \mathbf{P}$ for $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$. Thus, by Lemma A.2 and arguing as in (A.10) we get

$$\begin{aligned}\max\{&\sup_{s\in\mathcal{V}^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle, \sup_{s\in\mathcal{V}^{\mathrm{i}}} \langle A^{\dagger}s, A^{\dagger}\hat{\mathbb{G}}_n^{\mathrm{i}}\rangle + \mathbb{U}_n(s,P)\} \\ &= \max\{\max_{s\in\mathcal{E}^{\mathrm{e}}} \langle s, (\Omega^{\mathrm{e}})^{\dagger}\mathbb{G}_n^{\mathrm{e}\star}(P)\rangle, \max_{s\in\mathcal{E}^{\mathrm{i}}} \langle s, \mathbb{G}_n^{\mathrm{i}\star}(P) + \sqrt{n}\beta(P)\rangle\} + O_P(b_n)\end{aligned} \tag{A.14}$$

uniformly in $P \in \mathbf{P}_0$. Setting $c_n^{(2)}(\eta, P) \equiv \inf\{u : P(\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) \le u) \ge \eta\}$, we then obtain from results (A.13) and (A.14), $E_n(P)$ implying that $\hat{c}_n(\eta)$ is bounded from below by the conditional on $\{Z_i\}_{i=1}^n$ η quantile of

$$\max\{\sup_{s\in\mathcal{V}^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle, \sup_{s\in\mathcal{V}^{\mathrm{i}}} \langle A^{\dagger}s, A^{\dagger}\hat{\mathbb{G}}_n^{\mathrm{i}}\rangle + \mathbb{U}_n(s,P)\} - \ell_n,$$

$\mathbb{G}_n(P)$ and $\mathbb{G}_n^{\star}(P)$ sharing the same distribution, $\mathbb{G}_n^{\star}(P)$ being independent of $\{Z_i\}_{i=1}^n$, Lemma 11 in Chernozhukov et al. (2013), and $b_n = o(\ell_n)$ that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\phi_n = 1) \le \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) > c_n^{(2)}(\eta_n, P) - 3\ell_n) \tag{A.15}$$

for some $\eta_n \uparrow \eta$. Next set $\mathbb{N}((\mathrm{j},s),P) \equiv (\mathbb{A}_n^{\mathrm{j}}(s,P) - c_n^{(2)}(\eta_n,P))/\sigma^{\mathrm{j}}(s,P)$ and $\underline{\mu}_n(P) = -(c_n^{(1)}(\tau_n,P) + |c_n^{(2)}(\tau_n,P)|)/\underline{\sigma}(P)$ and note that by definition of $\mathcal{E}^{\mathrm{i},\tau_n}(P)$ we have $E[\mathbb{N}((\mathrm{j},s),P)] \ge \underline{\mu}_n(P)$. Therefore, Lemma A.9 yields that

$$\begin{aligned}P(|\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) - c_n^{(2)}(\eta_n,P)| \le 3\ell_n) &\le P(|\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{N}((\mathrm{j},s),P)| \le \frac{3\ell_n}{\underline{\sigma}(P)}) \\ &\lesssim (\ell_n/\underline{\sigma}(P))(1 \vee (\mathrm{med}\{\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{N}((\mathrm{j},s),P)\} - \underline{\mu}_n(P))).\end{aligned} \tag{A.16}$$

Moreover, $\Omega^{\mathrm{j}}(\Omega^{\mathrm{j}})^{\dagger}\mathbb{G}_n^{\mathrm{j}}(P)=\mathbb{G}_n^{\mathrm{j}}(P)$, $\mathcal{E}^{\mathrm{e}}\subset\Omega^{\mathrm{e}}\mathcal{V}^{\mathrm{e}}$, and $\mathcal{E}^{\mathrm{i},\tau_n}(P)\subseteq(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$ imply

$$\mathrm{med}\{\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)}\mathbb{N}((\mathrm{j},s),P)\}\leq(\mathrm{m}(P)+|c_n^{(2)}(\eta_n,P)|)/\underline{\sigma}(P) \tag{A.17}$$

for all $P\in\mathbf{P}_0$. Furthermore, since $\tau_n\uparrow 1$, $\eta_n\uparrow\eta>1/2$ and $\langle s,\beta(P)\rangle\leq 0$ for any $s\in\mathcal{E}^{\mathrm{i},\tau_n}(P)\subset(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$ by Theorem 3.1, we obtain by Borell's inequality (see, e.g., the corollary in pg. 82 of Davydov et al. (1998)) that $c_n^{(1)}(\tau_n,P)\vee c_n^{(2)}(\eta_n,P)\leq\mathrm{m}(P)+z_{\tau_n}\bar{\sigma}(P)$ for n sufficiently large. Also note $c_n^{(2)}(\eta_n,P)\geq -c_n^{(1)}(\tau_n,P)$ by definition of $\mathcal{E}^{\mathrm{i},\tau_n}(P)$ and $\eta_n\uparrow\eta>1/2$. Hence, (A.16) and (A.17) yield

$$P(|\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)}\mathbb{A}_n^{\mathrm{j}}(s,P)-c_n^{(2)}(\eta_n,P)|\leq 3\ell_n)\lesssim\frac{\ell_n}{\underline{\sigma}(P)}(1\vee\frac{\mathrm{m}(P)+z_{\tau_n}(P)\bar{\sigma}(P)}{\underline{\sigma}(P)}),$$

which together with (A.7), (A.15), and $\eta_n\uparrow\eta\equiv 1-\alpha$ yield the theorem. ■

Lemma A.1. *Let Assumptions 4.1, 4.2, 4.3(i) hold, $\lambda_n\in[0,1]$, and $r_n=o(1)$. Then, for any sequence ℓ_n satisfying $\lambda_n\sqrt{\log(1+p)}=o(\ell_n)$ it follows that $P(\sup_{s\in\mathcal{V}^{\mathrm{i}}}\{\sqrt{n}\langle A^{\dagger}s,A^{\dagger}\beta(P)\rangle-\hat{\mathbb{U}}_n(s)\}\leq\ell_n)$ tends to one uniformly in $P\in\mathbf{P}_0$.*

Proof: Theorem 3.1 implies $\langle A^{\dagger}s,A^{\dagger}\beta(P)\rangle\leq 0$ for all $s\in\mathcal{V}^{\mathrm{i}},P\in\mathbf{P}_0$, and hence

$$\begin{aligned}\sup_{s\in\mathcal{V}^{\mathrm{i}}}\{\sqrt{n}\langle A^{\dagger}s,A^{\dagger}\beta(P)\rangle-\hat{\mathbb{U}}_n(s)\}&\leq\sup_{s\in\mathcal{V}^{\mathrm{i}}}\lambda_n\sqrt{n}\langle A^{\dagger}s,A^{\dagger}\{\beta(P)-\hat{\beta}_n^{\mathrm{r}}\}\rangle\\&\leq\sup_{s\in\mathcal{V}^{\mathrm{i}}}\lambda_n\sqrt{n}|\langle A^{\dagger}s,\hat{x}_n^{\star}-A^{\dagger}\hat{\beta}_n^{\mathrm{r}}\rangle|+\sup_{s\in\mathcal{V}^{\mathrm{i}}}\lambda_n\sqrt{n}|\langle A^{\dagger}s,A^{\dagger}\beta(P)-\hat{x}_n^{\star}\rangle|.\end{aligned} \tag{A.18}$$

The definition of $\hat{\beta}_n^{\mathrm{r}}$, $\hat{x}_n^{\star}=A^{\dagger}\hat{\beta}_n$, $\beta(P)\in R$ for any $P\in\mathbf{P}_0$, and (A.18) then yield

$$\sup_{s\in\mathcal{V}^{\mathrm{i}}}\{\sqrt{n}\langle A^{\dagger}s,A^{\dagger}\beta(P)\rangle-\hat{\mathbb{U}}_n(s)\}\leq\sup_{s\in\mathcal{V}^{\mathrm{i}}}2\lambda_n|\langle A^{\dagger}s,A^{\dagger}AA^{\dagger}\sqrt{n}\{\hat{\beta}_n-\beta(P)\}\rangle|. \tag{A.19}$$

Applying Lemma A.2 with $(\hat{\mathbb{W}}_n^{\mathrm{i}}(P),\hat{\mathbb{W}}_n^{\mathrm{e}}(P))=\pm(AA^{\dagger}\sqrt{n}\{\hat{\beta}_n-\beta(P)\},\mathbb{G}_n^{\mathrm{e}}(P))$, we then obtain from Lemma A.4 and $\pm(\mathbb{G}_n^{\mathrm{e}}(P)',\mathbb{G}_n^{\mathrm{i}}(P)')'\sim N(0,\Sigma(P))$ that

$$\sup_{s\in\mathcal{V}^{\mathrm{i}}}|\langle A^{\dagger}s,A^{\dagger}AA^{\dagger}\sqrt{n}\{\hat{\beta}_n-\beta(P)\}\rangle|=\sup_{s\in\mathcal{V}^{\mathrm{i}}}|\langle A^{\dagger}s,A^{\dagger}\mathbb{G}_n^{\mathrm{i}}(P)\rangle|+O_P(r_n) \tag{A.20}$$

uniformly in $P\in\mathbf{P}_0$. Since Theorem 3.6.1 in Bogachev (1998) and Assumption 4.3(i) imply $\mathbb{G}_n^{\mathrm{i}}(P)\in\mathrm{range}\{\Sigma^{\mathrm{i}}(P)\}\subseteq\mathrm{range}\{\Omega^{\mathrm{i}}\}$, we have $\Omega^{\mathrm{i}}(\Omega^{\mathrm{i}})^{\dagger}\mathbb{G}_n^{\mathrm{i}}(P)=\mathbb{G}_n^{\mathrm{i}}(P)$. Therefore, (A.19), (A.20), and Hölder's inequality yield, uniformly in $P\in\mathbf{P}_0$, that

$$\begin{aligned}\sup_{s\in\mathcal{V}^{\mathrm{i}}}2\lambda_n|\langle A^{\dagger}s,A^{\dagger}AA^{\dagger}\sqrt{n}\{\hat{\beta}_n-\beta(P)\}\rangle|&=\sup_{s\in\mathcal{V}^{\mathrm{i}}}2\lambda_n|\langle A^{\dagger}s,A^{\dagger}\mathbb{G}_n^{\mathrm{i}}(P)\rangle|+O_P(\lambda_n r_n)\\&\leq 2\lambda_n\|(\Omega^{\mathrm{i}})^{\dagger}\mathbb{G}_n^{\mathrm{i}}(P)\|_{\infty}+O_P(\lambda_n r_n)=O_P(\lambda_n\sqrt{\log(1+p)}),\end{aligned} \tag{A.21}$$

where the final equality follows from Markov's inequality and Lemma A.8 and Assumption 4.2(ii) implying $\sup_{P\in\mathbf{P}} E_P[\|(\Omega^{\rm i})^\dagger\mathbb{G}_n^{\rm i}(P)\|_\infty] \lesssim \sqrt{\log(1+p)}$. The Lemma follows from (A.19), (A.21), and $\lambda_n\sqrt{\log(1+p)} = o(\ell_n)$. ■

Lemma A.2. *Let Assumptions 4.2(ii), 4.2(iv), and 4.3(i) hold, and suppose* $(\hat{\mathbb{W}}_n^{\rm e}(P)', \hat{\mathbb{W}}_n^{\rm i}(P)')' \equiv \hat{\mathbb{W}}_n(P)$ *satisfies* $\|(\Omega^{\rm j})^\dagger\{\hat{\mathbb{W}}_n^{\rm j}(P) - \mathbb{W}_n^{\rm j}(P)\}\|_\infty = O_P(\omega_n)$ *for* ${\rm j}\in\{{\rm e},{\rm i}\}$ *and* $\mathbb{W}_n(P) \equiv (\mathbb{W}_n^{\rm e}(P)', \mathbb{W}_n^{\rm i}(P)')' \sim N(0,\Sigma(P))$. *If* $\hat{\mathbb{W}}_n^{\rm j}(P) \in \operatorname{range}\{\Omega^{\rm j}\}$ *with probability tending to one uniformly in* $P\in\mathbf{P}$ *for* ${\rm j}\in\{{\rm e},{\rm i}\}$, *then for any possibly random function* $\hat{f}_n(\cdot,P):\mathbf{R}^p\to\mathbf{R}$ *it follows uniformly in* $P\in\mathbf{P}$ *that*

$$\sup_{s\in\mathcal{V}^{\rm e}}\langle s, \hat{\mathbb{W}}_n^{\rm e}(P)\rangle = \sup_{s\in\mathcal{V}^{\rm e}}\langle s, \mathbb{W}_n^{\rm e}(P)\rangle + O_P(\omega_n)$$

$$\sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, A^\dagger\hat{\mathbb{W}}_n^{\rm i}(P)\rangle + \hat{f}_n(s,P) = \sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, A^\dagger\mathbb{W}_n^{\rm i}(P)\rangle + \hat{f}_n(s,P) + O_P(\omega_n).$$

Proof: We establish only the second claim noting that the first follows from identical arguments. First note $\Omega^{\rm i}(A^\dagger)'A^\dagger = \Omega^{\rm i}(AA')^\dagger$, Hölder's inequality, definition of $\mathcal{V}^{\rm i}$, and symmetry of $\Omega^{\rm i}$ imply uniformly in $P\in\mathbf{P}$ that

$$\sup_{s\in\mathcal{V}^{\rm i}}|\langle A^\dagger s, A^\dagger\Omega^{\rm i}(\Omega^{\rm i})^\dagger(\hat{\mathbb{W}}_n^{\rm i}(P) - \mathbb{W}_n^{\rm i}(P))\rangle| \le \|(\Omega^{\rm i})^\dagger(\hat{\mathbb{W}}_n^{\rm i}(P) - \mathbb{W}_n^{\rm i}(P))\|_\infty = O_P(\omega_n). \tag{A.22}$$

Hence, since $\Omega^{\rm i}(\Omega^{\rm i})^\dagger\hat{\mathbb{W}}_n^{\rm i}(P) = \hat{\mathbb{W}}_n^{\rm i}(P)$ whenever $\hat{\mathbb{W}}_n^{\rm i}(P)\in\operatorname{range}\{\Omega^{\rm i}\}$, we obtain from $\hat{\mathbb{W}}_n^{\rm i}(P)\in\operatorname{range}\{\Omega^{\rm i}\}$ with probability tending to one and (A.22) that

$$\sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, A^\dagger\hat{\mathbb{W}}_n^{\rm i}(P)\rangle + \hat{f}_n(s,P) = \sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, A^\dagger\Omega^{\rm i}(\Omega^{\rm i})^\dagger\mathbb{W}_n^{\rm i}(P)\rangle + \hat{f}_n(s,P) + O_P(\omega_n) \tag{A.23}$$

uniformly in $P\in\mathbf{P}$. Finally, note $\mathbb{W}_n^{\rm i}(P)\in\operatorname{range}\{\Sigma^{\rm i}(P)\}$ by Theorem 3.6.1 in Bogachev (1998). Since Assumption 4.3(i) implies $\Sigma^{\rm i}(P) = \Omega^{\rm i}(\Omega^{\rm i})^\dagger\Sigma^{\rm i}(P)$ it follows $\mathbb{W}_n^{\rm i}(P) = \Omega^{\rm i}(\Omega^{\rm i})^\dagger\mathbb{W}_n^{\rm i}(P)$, which together with (A.23) establishes the lemma. ■

Lemma A.3. *Let Assumptions 4.1(i), 4.2, 4.3, 4.4 hold,* $\eta\in(0.5,1)$, $\epsilon\in(0,\eta-0.5)$, z_η *be the* η *quantile of* $N(0,1)$, $r_n\vee b_n = o(1)$, $\sup_{P\in\mathbf{P}}(\mathrm{m}(P)+\bar{\sigma}(P))/\underline{\sigma}^2(P) = o(r_n^{-1}\wedge b_n^{-1})$. *Then, then there are* $\{E_n(P)\}$ *with* $\liminf_{n\to\infty}\inf_{P\in\mathbf{P}_0} P(\{Z_i\}_{i=1}^n\in E_n(P)) = 1$ *and on* $E_n(P)$ *it holds that* $\hat{c}_n(\eta) \ge (\underline{\sigma}(P)z_{\eta-\epsilon})/2$ *whenever* $T_n > 0$.

Proof: By Lemma A.5 there are $(\mathbb{G}_n^{{\rm e}\star}(P)', \mathbb{G}_n^{{\rm i}\star}(P)')' \equiv \mathbb{G}_n^\star(P) \sim N(0,\Sigma(P))$ that are independent of $\{Z_i\}_{i=1}^n$ and satisfy $\|(\Omega^{\rm j})^\dagger\{\hat{\mathbb{G}}_n^{\rm j} - \mathbb{G}_n^{{\rm j}\star}(P)\|_\infty = O_P(b_n)$ uniformly in $P\in\mathbf{P}$ for ${\rm j}\in\{{\rm e},{\rm i}\}$. Further define $\hat{\mathbb{L}}_n\in\mathbf{R}$ and $\mathbb{L}_n^\star(P)\in\mathbf{R}$ to be given by

$$\hat{\mathbb{L}}_n \equiv \max\{\sup_{s\in\mathcal{V}^{\rm e}}\langle s,\hat{\mathbb{G}}_n^{\rm e}\rangle, \sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, A^\dagger\hat{\mathbb{G}}_n^{\rm i}\rangle + \hat{\mathbb{U}}_n(s)\} \tag{A.24}$$

$$\mathbb{L}_n^\star(P) \equiv \max\{\sup_{s\in\mathcal{V}^{\rm e}}\langle s, \mathbb{G}_n^{{\rm e}\star}(P)\rangle, \sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, A^\dagger\mathbb{G}_n^{{\rm i}\star}(P)\rangle + \hat{\mathbb{U}}_n(s)\}, \tag{A.25}$$

and note that Assumptions 4.3(i) and 4.4(v) together with Lemma A.2 yield

$$\sup_{s\in\mathcal{V}^{\rm e}}\langle s, \hat{\mathbb{G}}_n^{\rm e}\rangle = \sup_{s\in\mathcal{V}^{\rm e}}\langle s, \mathbb{G}_n^{{\rm e}\star}(P)\rangle + O_P(b_n) \tag{A.26}$$

$$\sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, A^\dagger\hat{\mathbb{G}}_n^{\rm i}\rangle + \hat{\mathbb{U}}_n(s) = \sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, A^\dagger\mathbb{G}_n^{{\rm i}\star}(P)\rangle + \hat{\mathbb{U}}_n(s) + O_P(b_n) \tag{A.27}$$

uniformly in $P \in \mathbf{P}$. We establish the lemma by studying three separate cases.

Case I: Suppose $P \in \mathbf{P}_0^{\rm e} \equiv \{P \in \mathbf{P}_0 : \sigma^{\rm e}(s, P) > 0$ for some $s \in \mathcal{E}^{\rm e}\}$. First set

$$E_n(P) \equiv \{P(|\sup_{s\in\mathcal{V}^{\rm e}}\langle s, \hat{\mathbb{G}}_n^{\rm e}\rangle - \sup_{s\in\mathcal{V}^{\rm e}}\langle s, \mathbb{G}_n^{{\rm e}\star}(P)\rangle| > (\underline{\sigma}(P)z_{\eta-\epsilon})/2|\{Z_i\}_{i=1}^n) \le \epsilon\}$$

and note that (A.26), Markov's inequality, and $b_n \times \sup_{P\in\mathbf{P}} 1/\underline{\sigma}(P) = o(1)$ imply the probability of $E_n(P)$ tends to one uniformly in $P \in \mathbf{P}_0^{\rm e}$. Moreover, whenever $\{Z_i\}_{i=1}^n \in E_n(P)$, $\sup_{s\in\mathcal{V}^{\rm e}}\langle s, \hat{\mathbb{G}}_n^{\rm e}\rangle \le \hat{\mathbb{L}}_n$ and the definition of $\hat{c}_n(\eta)$ imply

$$P(\sup_{s\in\mathcal{V}^{\rm e}}\langle s, \mathbb{G}_n^{{\rm e}\star}(P)\rangle \le \hat{c}_n(\eta) + \underline{\sigma}(P)z_{\eta-\epsilon}/2|\{Z_i\}_{i=1}^n) \ge \eta - \epsilon. \tag{A.28}$$

Also note that $\mathbb{G}_n^{{\rm e}\star}(P) \sim N(0, \Sigma^{\rm e}(P))$, Theorem 3.6.1 in Bogachev (1998), and Assumption 4.3(i) yield $\mathbb{G}_n^{{\rm e}\star}(P) = \Omega^{\rm e}(\Omega^{\rm e})^\dagger\mathbb{G}_n^{{\rm e}\star}(P)$ almost surely. Hence, Lemma A.10 implies $\sup_{s\in\mathcal{V}^{\rm e}}\langle s, \mathbb{G}_n^{{\rm e}\star}(P)\rangle = \max_{s\in\mathcal{E}^{\rm e}}\langle s, (\Omega^{\rm e})^\dagger\mathbb{G}_n^{{\rm e}\star}(P)\rangle$. It follows that the distribution of $\sup_{s\in\mathcal{V}^{\rm e}}\langle s, \mathbb{G}_n^{{\rm e}\star}(P)\rangle$ first order stochastically dominates $N(0, \underline{\sigma}^2(P))$ whenever $P \in \mathbf{P}_0^{\rm e}$. Thus, $\mathbb{G}_n^{{\rm e}\star}(P)$ being independent of $\{Z_i\}_{i=1}^n$ and (A.28) imply that $\hat{c}_n(\eta) + \underline{\sigma}(P)z_{\eta-\epsilon}/2 \ge \underline{\sigma}(P)z_{\eta-\epsilon}$ whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ and $P \in \mathbf{P}_0^{\rm e}$.

Case II: Suppose $P \in \mathbf{P}_0^{\rm i} \equiv \{P \in \mathbf{P}_0 : \sigma^{\rm i}(s, P) > 0$ for some $s \in \mathcal{E}^{\rm i}$ and $\sigma^{\rm e}(s, P) = 0$ for all $s \in \mathcal{E}^{\rm e}\}$, and set $E_n(P) \equiv E_{1,n}(P) \cap E_{2,n}(P)$, where $E_{1n}(P) \equiv \{AA^\dagger\{\hat{\beta}_n - \beta(P)\} \in \mathrm{range}\{\Sigma^{\rm i}(P)\}$, $T_n = \sup_{s\in\mathcal{V}^{\rm i}}\langle A^\dagger s, \hat{x}_n^\star\rangle\}\}$ and $E_{2n}(P) \equiv \{P(|\hat{\mathbb{L}}_n - \mathbb{L}_n^\star(P)| > (\underline{\sigma}(P)z_{\eta-\epsilon})/2|\{Z_i\}_{i=1}^n) \le \epsilon\}$. Note Assumption 4.3(ii) and Lemma A.6 (for $E_{1n}(P)$), and (A.26), (A.27), Markov's inequality, and $\sup_{P\in\mathbf{P}} 1/\underline{\sigma}(P) = o(b_n^{-1})$ (for $E_{2n}(P)$), imply the probability of $E_n(P)$ tends to one uniformly in $P \in \mathbf{P}_0^{\rm i}$.

Since $A^\dagger AA^\dagger = A^\dagger$ by Proposition 6.11.1(5) in Luenberger (1969), $AA^\dagger\beta(P) = \beta(P)$ whenever $P \in \mathbf{P}_0$ and $AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\} \in \mathrm{range}\{\Omega^{\rm i}\}$ whenever $\{Z_i\}_{i=1}^n \in E_{1n}(P)$ by Assumption 4.3(i) imply that if $\{Z_i\}_{i=1}^n \in E_n(P)$ and $s \in \mathcal{V}^{\rm i}$, then

$$\begin{aligned}\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle &= \langle A^\dagger s, A^\dagger AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle \\ &= \langle\Omega^{\rm i}(AA')^\dagger s, (\Omega^{\rm i})^\dagger AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle(AA')^\dagger s, \beta(P)\rangle.\end{aligned} \tag{A.29}$$

Since $\langle A^\dagger s, A^\dagger \beta(P)\rangle \le 0$ for all $P \in \mathbf{P}_0$, $s \in \mathcal{V}^{\mathrm{i}}$ by Theorem 3.1, Hölder's inequality implies (A.29) is bounded above in $s \in \mathcal{V}^{\mathrm{i}}$. Hence, by Lemmas A.10 and A.11

$$\sup_{s\in(AA')^\dagger\mathcal{V}^{\mathrm{i}}} \langle \Omega^{\mathrm{i}} s, (\Omega^{\mathrm{i}})^\dagger AA^\dagger \sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle s, \beta(P)\rangle$$
$$= \max_{s\in\mathcal{E}^{\mathrm{i}}} \langle \Omega^{\mathrm{i}} s, (\Omega^{\mathrm{i}})^\dagger AA^\dagger \sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle s, \beta(P)\rangle. \quad (A.30)$$

Thus, (A.29) and (A.30) imply $\mathcal{S}^{\mathrm{i}}(P) \equiv \{s \in \mathcal{E}^{\mathrm{i}} : \langle \Omega^{\mathrm{i}} s, (\Omega^{\mathrm{i}})^\dagger AA^\dagger \sqrt{n}\{\hat{\beta}_n - \beta(P)\} + \sqrt{n}\langle s, \beta(P)\rangle > 0\}$ is such that $\mathcal{S}^{\mathrm{i}}(P) \neq \emptyset$ whenever $T_n > 0$ and $\{Z_i\}_{i=1}^n \in E_n(P)$. Moreover, since $\sqrt{n}\langle s, \beta(P)\rangle \le 0$ for all $s \in \mathcal{S}^{\mathrm{i}}(P)$ and $P \in \mathbf{P}_0$ by Theorem 3.1, it follows that if $\mathcal{S}^{\mathrm{i}}(P) \neq \emptyset$, then $\langle \Omega^{\mathrm{i}} s, (\Omega^{\mathrm{i}})^\dagger AA^\dagger \sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle > 0$ for all $s \in \mathcal{S}^{\mathrm{i}}(P)$ and thus by Theorem 3.6.1 in Bogachev (1998) we have

$$\mathcal{S}^{\mathrm{i}}(P) \neq \emptyset \text{ and } \sigma^{\mathrm{i}}(s, P) > 0 \text{ for all } s \in \mathcal{S}^{\mathrm{i}}(P) \quad (A.31)$$

whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ and $T_n > 0$. We next aim to show that in addition

$$\max_{s\in\mathcal{S}^{\mathrm{i}}(P)} \langle s, AA^\dagger \hat{\beta}_n^{\mathrm{r}}\rangle = 0 \quad (A.32)$$

whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ and $T_n > 0$. To this end, note Theorem 3.1 yields that

$$0 \ge \sup_{s\in\mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\beta}_n^{\mathrm{r}}\rangle = \sup_{s\in(AA')^\dagger\mathcal{V}^{\mathrm{i}}} \langle s, AA^\dagger \hat{\beta}_n^{\mathrm{r}}\rangle = \max_{s\in\mathcal{E}^{\mathrm{i}}} \langle s, AA^\dagger \hat{\beta}_n^{\mathrm{r}}\rangle, \quad (A.33)$$

due to Lemmas A.10 and A.11 and $A^\dagger AA^\dagger = A^\dagger$. Since $AA^\dagger \beta(P) = \beta(P)$ for all $P \in \mathbf{P}_0$, the symmetry of Ω^{i} and $AA^\dagger \sqrt{n}\{\hat{\beta}_n - \beta(P)\} \in \mathrm{range}\{\Omega^{\mathrm{i}}\}$ whenever $\{Z_i\}_{i=1}^n \in E_{1n}(P)$ due to Assumption 4.3(i), we obtain by definition of $\mathcal{S}^{\mathrm{i}}(P)$

$$\max_{s\in\mathcal{E}^{\mathrm{i}}\setminus\mathcal{S}^{\mathrm{i}}(P)} \langle s, AA^\dagger \hat{\beta}_n\rangle = \max_{s\in\mathcal{E}^{\mathrm{i}}\setminus\mathcal{S}^{\mathrm{i}}(P)} \langle \Omega^{\mathrm{i}} s, (\Omega^{\mathrm{i}})^\dagger AA^\dagger \sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle s, \beta(P)\rangle \le 0. \quad (A.34)$$

Thus, if we suppose that (A.32) fails to hold, then (A.33), (A.34), $\mathcal{S}^{\mathrm{i}}(P) \subseteq \mathcal{E}^{\mathrm{i}}$, and $\mathcal{E}^{\mathrm{i}}$ being finite, imply there is a $\gamma \in (0,1)$ depending on $\hat{\beta}_n^{\mathrm{r}}$ and $\hat{\beta}_n$ satisfying

$$0 \ge \max_{s\in\mathcal{E}^{\mathrm{i}}} \langle s, AA^\dagger\{(1-\gamma)\hat{\beta}_n^{\mathrm{r}} + \gamma\hat{\beta}_n\}\rangle = \sup_{s\in\mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, A^\dagger\{(1-\gamma)\hat{\beta}_n^{\mathrm{r}} + \gamma AA^\dagger \hat{\beta}_n\}\rangle \quad (A.35)$$

where the equality follow from Lemmas A.10, A.11, and $A^\dagger AA^\dagger = A^\dagger$. However, since $\hat{\beta}_n^{\mathrm{r}} \in R$ and $AA^\dagger \hat{\beta}_n \in R$, result (A.35) and Theorem 3.1 imply $(1-\gamma)\hat{\beta}_n^{\mathrm{r}} + \gamma AA^\dagger \hat{\beta}_n = Ax$ for some $x \ge 0$. Moreover, if $T_n > 0$, then $\sup_{s\in\mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, \hat{x}_n^\star\rangle > 0$ whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ and hence $T_n > 0$ implies $\sup_{s\in\mathcal{V}^{\mathrm{i}}} \langle A^\dagger s, \hat{x}_n^\star - A^\dagger \hat{\beta}_n^{\mathrm{r}}\rangle > 0$ due to $\langle A^\dagger s, A^\dagger \hat{\beta}_n^{\mathrm{r}}\rangle \le 0$ for all $s \in \mathcal{V}^{\mathrm{i}}$. In particular, if $T_n > 0$, then $\sup_{s\in\mathcal{V}^{\mathrm{i}}} |\langle A^\dagger s, \hat{x}_n^\star - A^\dagger \hat{\beta}_n^{\mathrm{r}}\rangle| > 0$ and therefore $\hat{x}_n^\star = A^\dagger \hat{\beta}_n$, $A^\dagger AA^\dagger = A^\dagger$, and $\gamma \in (0,1)$ imply

$\sup_{s\in\mathcal{V}^{\mathrm{i}}}|\langle A^\dagger s, \hat{x}_n^\star - A^\dagger\{(1-\gamma)\hat{\beta}_n^{\mathrm{r}} + \gamma AA^\dagger\hat{\beta}_n\}\rangle| < \sup_{s\in\mathcal{V}^{\mathrm{i}}}|\langle A^\dagger s, \hat{x}_n^\star - A^\dagger\hat{\beta}_n^{\mathrm{r}}\rangle|$, which is impossible by definition of $\hat{\beta}_n^{\mathrm{r}}$. Thus, under $E_n(P)$, (A.32) holds when $T_n > 0$.

Results (A.31) and (A.32) imply that, whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ and $T_n > 0$, there is a $\hat{s}_n \in \mathcal{V}^{\mathrm{i}}$ with $(AA')^\dagger\hat{s}_n \in \mathcal{E}^{\mathrm{i}}$, $\sigma((AA')^\dagger\hat{s}_n, P) > 0$, and $0 = \lambda_n\langle A^\dagger\hat{s}_n, A^\dagger\hat{\beta}_n^{\mathrm{r}}\rangle \equiv \hat{\mathbb{U}}_n(\hat{s}_n)$. Hence, the definitions of $\mathbb{L}_n^\star(P)$, $E_n(P)$, and $\hat{c}_n(\eta)$ yield that

$$P(\langle A^\dagger\hat{s}_n, A^\dagger\mathbb{G}_n^{\mathrm{i}\star}(P)\rangle \leq \hat{c}_n(\eta) + \underline{\sigma}(P)z_{\eta-\epsilon}/2|\{Z_i\}_{i=1}^n) \\ \geq P(\mathbb{L}_n^\star(P) \leq \hat{c}_n(\eta) + \underline{\sigma}(P)z_{\eta-\epsilon}/2|\{Z_i\}_{i=1}^n) \geq \eta - \epsilon \quad \text{(A.36)}$$

whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ and $T_n > 0$. Since $\mathbb{G}_n^{\mathrm{i}\star}(P) \in \mathrm{range}\{\Omega^{\mathrm{i}}\}$ by Assumption 4.3(i) and Theorem 3.6.1 in Bogachev (1998), we have $\langle A^\dagger\hat{s}_n, A^\dagger\mathbb{G}_n^{\mathrm{i}\star}(P)\rangle = \langle\Omega^{\mathrm{i}}(AA')^\dagger\hat{s}_n, (\Omega^{\mathrm{i}})^\dagger\mathbb{G}_n^{\mathrm{i}\star}(P)\rangle$ and hence $\langle A^\dagger\hat{s}_n, A^\dagger\mathbb{G}_n^{\mathrm{i}\star}(P)\rangle \sim N(0, (\sigma^{\mathrm{i}}((AA')^\dagger\hat{s}_n, P))^2)$ conditional on $\{Z_i\}_{i=1}^n$. Since $\sigma^{\mathrm{i}}((AA')^\dagger\hat{s}_n, P) > 0$ implies, by definition, that $\sigma^{\mathrm{i}}((AA')^\dagger\hat{s}_n, P) > \underline{\sigma}(P)$, result (A.36) yields $\hat{c}_n(\eta) + \underline{\sigma}(P)z_{\eta-\epsilon}/2 \geq \underline{\sigma}(P)z_{\eta-\epsilon}$.

<u>Case III:</u> Suppose $P \in \mathbf{P}_0^{\mathrm{d}} \equiv \{P \in \mathbf{P}_0 : \sigma^{\mathrm{j}}(s,P) = 0$ for all $s \in \mathcal{E}^{\mathrm{j}}$ and $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}\}$. Then, by Lemma A.6 we may set $E_n(P) \equiv \{T_n = 0\}$ and the lemma follows. ■

Lemma A.4. *If Assumptions 4.1, 4.2, 4.3(i) hold and $r_n = o(1)$, then there exists $(\mathbb{G}_n^{\mathrm{e}}(P)', \mathbb{G}_n^{\mathrm{i}}(P)')' \equiv \mathbb{G}_n(P) \sim N(0, \Sigma(P))$ satisfying uniformly in $P \in \mathbf{P}$:*

$$\|(\Omega^{\mathrm{e}})^\dagger\{(I_p - AA^\dagger)\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \mathbb{G}_n^{\mathrm{e}}(P)\}\|_\infty = O_P(r_n) \quad \text{(A.37)}$$
$$\|(\Omega^{\mathrm{i}})^\dagger\{AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \mathbb{G}_n^{\mathrm{i}}(P)\}\|_\infty = O_P(r_n). \quad \text{(A.38)}$$

Proof: Set $\tilde{\psi}(Z,P) \equiv (((\Omega^{\mathrm{e}})^\dagger\psi^{\mathrm{e}}(Z,P))', ((\Omega^{\mathrm{i}})^\dagger\psi^{\mathrm{i}}(Z,P))')' \in \mathbf{R}^{2p}$, define $\tilde{\Sigma}(P) \equiv E_P[\tilde{\psi}(Z,P)\tilde{\psi}(Z,P)']$, and let $S_n(P) \sim N(0, \tilde{\Sigma}(P)/n)$. Since $\|a\|_2^2 \leq 2p\|a\|_\infty^2$ for any $a \in \mathbf{R}^{2p}$, Lemma A.8 and Assumptions 4.2(ii)(iii) imply that

$$E_P[\|\frac{\tilde{\psi}(Z,P)}{\sqrt{n}}\|_2^2\|\frac{\tilde{\psi}(Z,P)}{\sqrt{n}}\|_\infty + \|S_n(P)\|_2^2\|S_n(P)\|_\infty] \lesssim \frac{p}{n^{3/2}}(M_{3,\Psi}^3 + (\log(1+p))^{3/2}). \quad \text{(A.39)}$$

For $\mathbb{Z} \sim N(0, I_{2p})$, we obtain by Assumptions 4.1(i), 4.2(i), Lemma 38 in Belloni et al. (2019), and (A.39) that for any $\delta > 0$ there is a $\tilde{\mathbb{G}}_n(P) \sim N(0, \tilde{\Sigma}(P))$ with

$$P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{\psi}(Z_i,P) - \tilde{\mathbb{G}}_n(P)\|_\infty > \delta) \lesssim \min_{t\geq 0}\{P(\|\mathbb{Z}\|_\infty > t) + \frac{t^2}{\delta^3}\frac{p}{\sqrt{n}}\{M_{3,\Psi}^3 + (\log(1+p))^{3/2}\} \\ \lesssim \min_{t\geq 0}\{\exp\{-\frac{t^2}{8\log(1+p)}\} + \frac{t^2}{\delta^3}\frac{pM_{3,\Psi}^3(\log(1+p))^{3/2}}{\sqrt{n}}\}, \quad \text{(A.40)}$$

where the final inequality follows from Proposition A.2.1 in van der Vaart and Wellner (1996) and $E[\|\mathbb{Z}\|_\infty^2] \lesssim \log(1+p)$ by Lemma A.8. Setting $t = K\sqrt{\log(1+p)}$,

$\delta^3 = K^3 p M_{3,\Psi}^3 (\log(1+p))^{5/2}/\sqrt{n}$, and taking limits as $K \uparrow \infty$ then yields

$$\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\psi}(Z_i, P) - \tilde{\mathbb{G}}_n(P)\|_\infty = O_P(\frac{M_{3,\Psi} p^{1/3}(\log(1+p))^{5/6}}{n^{1/6}}) \tag{A.41}$$

uniformly in $P \in \mathbf{P}$. Writing $\tilde{\mathbb{G}}_n(P) \equiv (\tilde{\mathbb{G}}_n^{\mathrm{e}}(P)', \tilde{\mathbb{G}}_n^{\mathrm{i}}(P)')'$, then note $\tilde{\mathbb{G}}_n^{\mathrm{j}}(P) \in \mathrm{range}\{(\Omega^{\mathrm{j}})^\dagger \Sigma^{\mathrm{j}}(P)(\Omega^{\mathrm{j}})^\dagger\}$ almost surely by Theorem 3.6.1 in Bogachev (1998) and Assumption 4.2(iv). Since $(\Omega^{\mathrm{j}})^\dagger \Omega^{\mathrm{j}} (\Omega^{\mathrm{j}})^\dagger = (\Omega^{\mathrm{j}})^\dagger$ it follows $(\Omega^{\mathrm{j}})^\dagger \Omega^{\mathrm{j}} \tilde{\mathbb{G}}_n^{\mathrm{j}}(P) = \tilde{\mathbb{G}}_n^{\mathrm{j}}(P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$. Setting $\mathbb{G}_n^{\mathrm{j}}(P) = \Omega^{\mathrm{j}} \tilde{\mathbb{G}}_n^{\mathrm{j}}(P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$ we then obtain from (A.41), Assumption 4.1(ii), and the definition of r_n that (A.37) and (A.38) indeed hold uniformly in $P \in \mathbf{P}$. Moreover, since $\tilde{\mathbb{G}}_n(P) \sim N(0, \tilde{\Sigma}(P))$ and Assumption 4.3(i) implies $\Omega^{\mathrm{j}}(\Omega^{\mathrm{j}})^\dagger \psi^{\mathrm{j}}(Z, P) = \psi^{\mathrm{j}}(Z, P)$, we obtain from the definition of $\tilde{\Sigma}(P)$ and $\mathbb{G}_n(P) = ((\Omega^{\mathrm{e}} \tilde{\mathbb{G}}_n^{\mathrm{e}}(P))', (\Omega^{\mathrm{i}} \tilde{\mathbb{G}}_n^{\mathrm{i}}(P))')'$ that $\mathbb{G}_n(P) \sim N(0, \Sigma(P))$ as desired. ■

Lemma A.5. *Let Assumptions 4.1(i), 4.2, 4.3(i), 4.4(i)-(iv) hold. If $b_n = o(1)$, then there is a $(\mathbb{G}_n^{\mathrm{e}\star}(P)', \mathbb{G}_n^{\mathrm{i}\star}(P)')' \equiv \mathbb{G}_n^\star(P) \sim N(0, \Sigma(P))$ independent of $\{Z_i\}_{i=1}^n$ satisfying $\|(\Omega^{\mathrm{j}})^\dagger\{\hat{\mathbb{G}}_n^{\mathrm{j}} - \mathbb{G}_n^{\mathrm{j}\star}(P)\}\|_\infty = O_P(b_n)$ uniformly in $P \in \mathbf{P}$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$.*

Proof: In what follows, we let $\varphi(Z, P) \equiv (\varphi^{\mathrm{e}}(Z, P)', \varphi^{\mathrm{i}}(Z, P)')' \in \mathbf{R}^{2p}$, where $\varphi^{\mathrm{e}}(Z, P) \equiv (\Omega^{\mathrm{e}})^\dagger \psi^{\mathrm{e}}(Z, P)$ and $\varphi^{\mathrm{i}}(Z, P) \equiv (\Omega^{\mathrm{i}})^\dagger \psi^{\mathrm{i}}(Z, P)$.

Step 1: Let $\{U_i\}_{i=1}^\infty$ be i.i.d., independent of $\{Z_i, W_{i,n}\}_{i=1}^n$, with U_i uniformly distributed on $(0, 1]$, and set $R_{i,n}$ to be the rank of U_i in the sample $\{U_i\}_{i=1}^n$. By Lemma 13.1(iv) in van der Vaart (1999), $R_n \equiv (R_{1,n}, \ldots, R_{n,n})$ is uniformly distributed on the set of all $n!$ permutations of $\{1, \ldots, n\}$. Letting $\stackrel{d}{=}$ denote equality in distribution, we then obtain from Assumption 4.4(i) that

$$(\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_{i,n} - \bar{W}_n)\varphi(Z_i, P), \{Z_i\}_{i=1}^n) \stackrel{d}{=} (\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_{R_{i},n} - \bar{W}_n)\varphi(Z_i, P), \{Z_i\}_{i=1}^n).$$

Step 2: Next set $\tau_n(u) \equiv \inf\{c : n^{-1}\sum_{i=1}^n 1\{W_{i,n} - \bar{W}_n \le c\} \ge u\}$, $\xi_{i,n}(P) \equiv n^{-1/2}(\varphi(Z_i, P) - \bar{\varphi}_n(P))\tau_n(U_i)$ for $\bar{\varphi}_n(P) \equiv \sum_{i=1}^n \varphi(Z_i, P)/n$, and

$$S_n(P) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n (W_{R_i,n} - \bar{W}_n)\varphi(Z_i, P) \quad L_n(P) \equiv \sum_{i=1}^n \xi_{i,n}(P), \tag{A.42}$$

We next couple $L_n(S)$ and $S_n(P)$. Letting A_j denote the j^{th} coordiante of a vector A, then note Theorem 3.1 in Hájek (1961) (see eq. (3.11) in page 512) yields

$$E[(S_{j,n}(P) - L_{j,n}(P))^2 | \{Z_i, W_{i,n}\}_{i=1}^n] \lesssim \mathrm{Var}\{L_{j,n}(P) | \{Z_i, W_{i,n}\}_{i=1}^n\} \frac{\max_{1\le i\le n}|W_{i,n} - \bar{W}_n|}{(\sum_{i=1}^n (W_{i,n} - \bar{W}_n)^2)^{1/2}}. \tag{A.43}$$

Since $\{U_i\}_{i=1}^n$ are i.i.d. uniform on $(0,1]$ we have $E[\tau_n(U_i)|\{Z_i, W_{i,n}\}_{i=1}^n] = 0$, which implies $E[\xi_{i,n}(P)|\{Z_i, W_{i,n}\}_{i=1}^n] = 0$. For $\hat{\sigma}_n^2 \equiv \sum_{i=1}^n (W_{i,n} - \bar{W}_n)^2/n$ we then obtain

$$\mathrm{Var}\{\xi_{i,n}(P)|\{Z_i, W_{i,n}\}_{i=1}^n\} = \frac{\hat{\sigma}_n^2}{n}(\varphi(Z_i, P) - \bar{\varphi}_n(P))(\varphi(Z_i, P) - \bar{\varphi}_n(P))'. \quad (A.44)$$

Noting that $E[\|V\|_\infty] \leq \sqrt{2p}\max_{1\leq j\leq 2p}(E[V_j^2])^{1/2}$ for any $(V_1, \ldots, V_{2p}) \equiv V \in \mathbf{R}^{2p}$ by Jensen's inequality, we then obtain from (A.43), (A.42), and (A.44) that

$$E[\|S_n(P) - L_n(P)\|_\infty|\{Z_i, W_{i,n}\}_{i=1}^n]$$
$$\lesssim \sqrt{p}\max_{1\leq j\leq 2p}(\frac{\hat{\sigma}_n}{n^{3/2}}\sum_{i=1}^n(\varphi_j(Z_i, P) - \bar{\varphi}_{j,n}(P))^2)^{1/2}(\max_{1\leq i\leq n}|W_{i,n} - \bar{W}_n|)^{1/2}. \quad (A.45)$$

Assumptions 4.4(ii)(iii) and Lemma 2.2.10 in van der Vaart and Wellner (1996) imply $E[\max_{1\leq i\leq n}|W_{i,n} - \bar{W}_n|] \lesssim \log(1+n)$. Thus, Assumptions 4.4(iii) and 4.2(iii), (A.45), Fubini's theorem and the triangle and Markov's inequality yield $\|S_n(P) - L_n(P)\|_\infty = O_P(\sqrt{p\log(1+n)}M_{3,\Psi}/n^{1/4})$ uniformly in $P \in \mathbf{P}$.

Step 3: We next couple $L_n(P)$ to a conditionally Gaussian vector. Set $\{\bar{G}_{i,n}(P)\}_{i=1}^n$ to be mutually independent conditional on $\{Z_i, W_{i,n}\}_{i=1}^n$ and satisfying $\bar{G}_{i,n}(P) \sim N(0, \mathrm{Var}\{\xi_{i,n}(P)|\{Z_i, W_{i,n}\}_{i=1}^n\})$. Employing the inequality $\|a\|_2^2 \leq 2p\|a\|_\infty^2$ for any $a \in \mathbf{R}^{2p}$ together with Lemma A.8, result (A.44), and direct calculation yield

$$\sum_{i=1}^n E[\|\bar{G}_{i,n}(P)\|_2^2\|\bar{G}_{i,n}(P)\|_\infty + \|\xi_{i,n}(P)\|_2^2\|\xi_{i,n}(P)\|_\infty|\{Z_i, W_{i,n}\}_{i=1}^n]$$
$$\lesssim \frac{p\log^{\frac{3}{2}}(1+p)}{\sqrt{n}}(\frac{1}{n}\sum_{i=1}^n|W_{i,n}|^3)(\frac{1}{n}\sum_{i=1}^n\{\Psi^3(Z_i, P) + \Psi^{\frac{3}{2}}(Z_i, P)\}) \equiv B_n(P) \quad (A.46)$$

where $\Psi(Z_i, P) = \|\varphi(Z_i, P)\|_\infty$. Let $\mathcal{B}$ denote the Borel σ-field on $\mathbf{R}^{2p}$ and for any $A \in \mathcal{B}$ and $\epsilon > 0$ set $A^\epsilon \equiv \{a \in \mathbf{R}^{2p} : \inf_{\tilde{a}\in A}\|a - \tilde{a}\|_\infty \leq \epsilon\}$. Strassen's Theorem, Lemma 38 in Belloni et al. (2019), and result (A.46) then imply for any $\delta > 0$

$$\sup_{A\in\mathcal{B}}\{P(L_n(P) \in A|\{Z_i, W_{i,n}\}_{i=1}^n) - P(\frac{1}{\sqrt{n}}\sum_{i=1}^n\bar{G}_{i,n}(P) \in A^{3\delta}|\{Z_i, W_{i,n}\}_{i=1}^n)\}$$
$$\lesssim \min_{t\geq 0}(2P(\|\mathbb{Z}\|_\infty > t) + \frac{B_n(P)}{\delta^3}t^2) \quad (A.47)$$

where $\mathbb{Z} \sim N(0, I_{2p})$. Furthermore, Proposition A.2.1 in van der Vaart and Wellner (1996), Lemma A.8, result (A.46), and Assumptions 4.2(iii) and 4.4(i)(iii) imply

$$\sup_{P\in\mathbf{P}}E_P[\min_{t\geq 0}(2P(\|\mathbb{Z}\|_\infty > t) + \frac{B_n(P)}{\delta^3}t^2)]$$

$$\lesssim \min_{t\geq 0}(\exp\{-\frac{t^2}{C\log(1+p)}\} + \frac{p\log^{3/2}(1+p)M_{3,\Psi}^3}{\sqrt{n}}\frac{t^2}{\delta^3}). \quad \text{(A.48)}$$

for some $C < \infty$. Hence, $p\log^{5/2}(1+p)M_{3,\Psi}^3/\sqrt{n} \leq b_n^3$, (A.47), and (A.48) imply

$$\sup_{P\in\mathbf{P}} E_P[\sup_{A\in\mathcal{B}} E_P[1\{L_n(P)\in A\} - 1\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \bar{G}_{i,n}(P)\in A^{3Kb_n}\}|\{Z_i,W_{i,n}\}_{i=1}^n]]$$
$$\lesssim \min_{t\geq 0}(\exp\{-\frac{t^2}{C\log(1+p)}\} + \frac{t^2}{K^3\log(1+p)}) \leq \exp\{-\frac{K^2}{C}\} + \frac{1}{K}, \quad \text{(A.49)}$$

where $K > 0$ is arbitrary and in the final inequality we set $t = K\sqrt{\log(1+p)}$. Theorem 4 in Monrad and Philipp (1991) and (A.49) imply there is a $\bar{\mathbb{G}}_n(P)$ satisfying $\|L_n(P) - \bar{\mathbb{G}}_n(P)\|_\infty = O_P(b_n)$ uniformly in $P\in\mathbf{P}$ and $\bar{\mathbb{G}}_n(P) \sim N(0, \sum_{i=1}^n \mathrm{Var}\{\xi_{i,n}(P)|\{Z_i,W_{i,n}\}_{i=1}^n\})$ conditional on $\{Z_i,W_{i,n}\}_{i=1}^n$.

Step 4: We next couple $\bar{\mathbb{G}}_n(P)$ to a Gaussian $\tilde{\mathbb{G}}_n^\star(P)$ independent of $\{Z_i,W_{i,n}\}_{i=1}^n$. To this end, note that $E_P[\varphi(Z,P)] = 0$ and $\sup_{P\in\mathbf{P}}\max_{1\leq j\leq 2p}\|\varphi_j(\cdot,P)\|_{P,2}$ being bounded in n by Assumptions 4.2(i)(ii), and $\|aa'\|_{o,2} = \|a\|_2^2$ for any $a\in\mathbf{R}^{2p}$ imply

$$\sup_{P\in\mathbf{P}} E_P[\|\bar{\varphi}_n(P)\bar{\varphi}_n(P)'\|_{o,2}] = \sup_{P\in\mathbf{P}} E_P[\|\bar{\varphi}_n(P)\|_2^2] \lesssim \frac{p}{n}. \quad \text{(A.50)}$$

Jensen's inequality, $\|\varphi(Z_i,P)\|_2^2 \leq 2p\Psi^2(Z_i,P)$, and Assumption 4.4(iv) imply

$$\sup_{P\in\mathbf{P}} E_P[\max_{1\leq i\leq n}\|\varphi(Z_i,P)\|_2^2] \lesssim \sup_{P\in\mathbf{P}} p(E_P[\max_{1\leq i\leq n}\Psi^{2q}(Z_i,P)])^{1/q} \leq pn^{1/q}M_{q,\Psi^2}.$$

Setting $\Lambda(P) \equiv E_P[\varphi(Z,P)\varphi(Z,P)']$, we then note that $b_n = o(1)$, Theorem E.1 in Kato (2013), $\|\Lambda(P)\|_{o,2}$ being uniformly bounded in n and $P\in\mathbf{P}$ by Assumption 4.2(ii) and definition of $\varphi(Z,P)$, and Markov's inequality yield that

$$\|\frac{1}{n}\sum_{i=1}^n \varphi(Z_i,P)\varphi(Z_i,P)' - \Lambda(P)\|_{o,2} = O_P(\{\frac{p\log(1+p)n^{1/q}M_{q,\Psi^2}}{n}\}^{1/2}) \quad \text{(A.51)}$$

uniformly in $P\in\mathbf{P}$. Setting $\hat{\Lambda}_n(P) \equiv \sum_{i=1}^n \mathrm{Var}\{\xi_{i,n}(P)|\{Z_i,W_{i,n}\}_{i=1}^n\}$, we then obtain from (A.44), (A.50), (A.51), $\|\Lambda(P)\|_{o,2}$ being bounded in n and $P\in\mathbf{P}$ and Assumption 4.4(iii) imply, uniformly in $P\in\mathbf{P}$, that $\|\hat{\Lambda}_n(P) - \Lambda(P)\|_{o,2} = O_P(\{p\log(1+p)n^{1/q}M_{q,\Psi^2}/n\}^{1/2})$. Since $\bar{\mathbb{G}}_n(P) \sim N(0,\hat{\Lambda}_n(P))$ conditional on $\{Z_i,W_{i,n}\}_{i=1}^n$, applying Lemma A.7 with $V_n = \{Z_i,W_{i,n}\}_{i=1}^n$ implies there is a $\tilde{\mathbb{G}}_n^\star(P) \sim N(0,\Lambda(P))$ independent of $\{Z_i,W_{i,n}\}_{i=1}^n$ satisfying uniformly in $P\in\mathbf{P}$ $\|\bar{\mathbb{G}}_n(P) - \tilde{\mathbb{G}}_n^\star(P)\|_\infty = O_P((p\log^3(1+p)n^{1/q}M_{q,\Psi^2}/n)^{1/4})$.

Step 5: By Steps 2, 3, and 4, there exists a Gaussian $\tilde{\mathbb{G}}_n^\star(P)$ that is independent of $\{Z_i,W_{i,n}\}_{i=1}^n$ and satisfies $\|S_n(P) - \tilde{\mathbb{G}}_n^\star(P)\|_\infty = O_P(b_n)$ uniformly in $P\in\mathbf{P}$.

Since $\tilde{\mathbb{G}}_n^\star(P)$ is independent of $\{Z_i\}_{i=1}^n$, the representation in Step 1 and Lemma 2.11 in Dudley and Philipp (1983) imply that there exists a $(\breve{\mathbb{G}}_n^{\mathrm{e}\star}(P)', \breve{\mathbb{G}}_n^{\mathrm{i}\star}(P)')' \equiv \breve{\mathbb{G}}_n^\star(P) \sim N(0, \Lambda(P))$ independent of $\{Z_i\}_{i=1}^n$ and such that uniformly in $P \in \mathbf{P}$

$$\|\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_{i,n} - \bar{W}_n)\varphi(Z_i, P) - \breve{\mathbb{G}}_n^\star(P)\|_\infty = O_P(b_n). \tag{A.52}$$

To conclude, set $\mathbb{G}_n^{\mathrm{j}\star}(P) \equiv \Omega^{\mathrm{j}}\breve{\mathbb{G}}_n^{\mathrm{j}\star}(P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$ and $\mathbb{G}_n^\star(P) \equiv (\mathbb{G}_n^{\mathrm{e}\star}(P)', \mathbb{G}_n^{\mathrm{i}\star}(P)')'$. Since $\Omega^{\mathrm{j}}(\Omega^{\mathrm{j}})^\dagger\psi^{\mathrm{j}}(Z, P) = \psi^{\mathrm{j}}(Z, P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$ by Assumption 4.3(i), it follows from $\Lambda(P) \equiv E_P[\varphi(Z, P)\varphi(Z, P)']$ and the definition of φ that $\mathbb{G}_n^\star(P) \sim N(0, \Sigma(P))$. Furthermore, $\breve{\mathbb{G}}_n^\star(P) \in \mathrm{range}\{\Lambda(P)\}$ by Theorem 3.6.1 in Bogachev (1998), and hence $\breve{\mathbb{G}}_n^{\mathrm{j}\star}(P) = (\Omega^{\mathrm{j}})^\dagger\Omega^{\mathrm{j}}\breve{\mathbb{G}}_n^{\mathrm{j}\star}(P) = (\Omega^{\mathrm{j}})^\dagger\mathbb{G}_n^{\mathrm{j}\star}(P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$. The lemma follows from (A.52), the definition of $\varphi(Z, P)$, and Assumption 4.4(i). ■

Lemma A.6. *Let Assumptions 4.2(iv), 4.3, 4.4(v) hold, and for* $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$ *define the sets* $\mathbf{D}_0^{\mathrm{j}} \equiv \{P \in \mathbf{P}_0 : \sigma^{\mathrm{j}}(s, P) = 0$ *for all* $s \in \mathcal{E}^{\mathrm{j}}\}$. *Then:*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{D}_0^{\mathrm{e}}} P(\sup_{s\in\mathcal{V}^{\mathrm{e}}} |\sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}_n^\star\rangle| = \sup_{s\in\mathcal{V}^{\mathrm{e}}} |\langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle| = 0) = 1$$

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{D}_0^{\mathrm{i}}} P(\sup_{s\in\mathcal{V}^{\mathrm{i}}} |\langle A^\dagger s, A^\dagger\hat{\mathbb{G}}_n^{\mathrm{i}}\rangle| = \sup_{s\in\mathcal{V}^{\mathrm{i}}} |\langle A^\dagger s, A^\dagger\beta(P) - \hat{x}_n^\star\rangle| = \sup_{s\in\mathcal{V}^{\mathrm{i}}}\langle A^\dagger s, \hat{x}_n^\star\rangle = 0) = 1$$

Proof: Theorem 3.6.1 in Bogachev (1998) and Assumption 4.3(i) imply $\mathbb{G}_n^{\mathrm{e}}(P) \in \mathrm{range}\{\Sigma^{\mathrm{e}}(P)\} \subseteq \mathrm{range}\{\Omega^{\mathrm{e}}\}$ almost surely. Hence, $\Omega^{\mathrm{e}}(\Omega^{\mathrm{e}})^\dagger\mathbb{G}_n^{\mathrm{e}}(P) = \mathbb{G}_n^{\mathrm{e}}(P)$ almost surely and Ω^{e} being symmetric by Assumption 4.2(iv) imply for any $P \in \mathbf{D}_0^{\mathrm{e}}$

$$\sup_{s\in\mathcal{V}^{\mathrm{e}}} |\langle s, \mathbb{G}_n^{\mathrm{e}}(P)\rangle| = \sup_{s\in\mathcal{V}^{\mathrm{e}}} |\langle \Omega^{\mathrm{e}} s, (\Omega^{\mathrm{e}})^\dagger\mathbb{G}_n^{\mathrm{e}}(P)\rangle| = \max_{s\in\mathcal{E}^{\mathrm{e}}} |\langle s, (\Omega^{\mathrm{e}})^\dagger\mathbb{G}_n^{\mathrm{e}}(P)\rangle| = 0, \tag{A.53}$$

where the second equality follows from Lemma A.10. Result (A.53), Assumption 4.3(ii), and the support of $\mathbb{G}_n^{\mathrm{e}}(P)$ being equal to $\mathrm{range}\{\Sigma^{\mathrm{e}}(P)\}$ by Theorem 3.6.1 in Bogachev (1998) imply that with probability tending to one uniformly in $P \in \mathbf{D}_0^{\mathrm{e}}$

$$\sup_{s\in\mathcal{V}^{\mathrm{e}}} |\sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}_n^\star\rangle| = \sup_{s\in\mathcal{V}^{\mathrm{e}}} |\langle s, (I_p - AA^\dagger)\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle| = 0. \tag{A.54}$$

Identical arguments but relying on Assumption 4.4(v) instead of 4.3(i) also yield that $\sup_{s\in\mathcal{V}^{\mathrm{e}}} |\langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle| = 0$ with probability tending to one uniformly in $P \in \mathbf{D}_0^{\mathrm{e}}$.

To establish the second claim, we note that similar arguments and $A^\dagger AA^\dagger\{\hat{\beta}_n - \beta(P)\} = \hat{x}_n^\star - A^\dagger\beta(P)$ by Proposition 6.11.1(5) in Luenberger (1969) imply

$$\sup_{s\in\mathcal{V}^{\mathrm{i}}} |\langle A^\dagger s, \hat{x}_n^\star - A^\dagger\beta(P)\rangle| = \sup_{s\in\mathcal{V}^{\mathrm{i}}} |\langle A^\dagger s, A^\dagger\hat{\mathbb{G}}_n^{\mathrm{i}}\rangle| = 0 \tag{A.55}$$

with probability tending to one uniformly in $P \in \mathbf{D}_0^{\mathrm{i}}$. To conclude, we note that $\langle A^\dagger s, A^\dagger \beta(P)\rangle \leq 0$ for any $P \in \mathbf{P}_0$ and $s \in \mathcal{V}^{\mathrm{i}}$ by Theorem 3.1. Hence, $0 \in \mathcal{V}^{\mathrm{i}}$ and (A.55) imply $0 \leq \sup_{s\in\mathcal{V}^{\mathrm{i}}}\langle A^\dagger s, \hat{x}_n^\star\rangle \leq \sup_{s\in\mathcal{V}^{\mathrm{i}}} |\langle A^\dagger s, \hat{x}_n^\star - A^\dagger\beta(P)\rangle| = 0$ with probability tending to one uniformly in $P \in \mathbf{D}_0^{\mathrm{i}}$ as well. ■

Lemma A.7. *Let $\{V_n\}_{n=1}^\infty$ be random variables with distribution parametrized by $P \in \mathbf{P}$ and $\bar{\mathbb{G}}_n(P) \in \mathbf{R}^{d_n}$ be such that $\bar{\mathbb{G}}_n(P) \sim N(0, \hat{\Sigma}_n(P))$ conditionally on V_n. If there exist non-random matrices $\Sigma_n(P)$ such that $\|\hat{\Sigma}_n(P) - \Sigma_n(P)\|_{o,2} = O_P(\delta_n)$ uniformly in $P \in \mathbf{P}$, then there exists a $\mathbb{G}_n^\star(P) \sim N(0, \Sigma_n(P))$ independent of V_n and satisfying $\|\bar{\mathbb{G}}_n(P) - \mathbb{G}_n^\star(P)\|_\infty = O_P(\sqrt{\log(1+d_n)\delta_n})$ uniformly in $P \in \mathbf{P}$.*

Proof: Let $\{\hat{\nu}_j(P)\}_{j=1}^{d_n}$ and $\{\hat{\lambda}_j(P)\}_{j=1}^{d_n}$ denote the unit length eigenvectors and eigenvalues of $\hat{\Sigma}_n(P)$. For some $\mathcal{N}_{d_n} \sim N(0, I_{d_n})$ independent of $(V_n, \bar{\mathbb{G}}_n(P))$ set

$$\mathbb{Z}_n(P) \equiv \sum_{j:\hat{\lambda}_j(P)\neq 0} \hat{\nu}_j(P)\frac{\hat{\nu}_j(P)'\bar{\mathbb{G}}_n(P)}{\hat{\lambda}_j^{1/2}(P)} + \sum_{j:\hat{\lambda}_j(P)=0} \hat{\nu}_j(P)(\hat{\nu}_j(P)'\mathcal{N}_{d_n}).$$

Since $\mathcal{N}_{d_n}$ is independent of $(V_n, \bar{\mathbb{G}}_n(P))$, it follows $\mathbb{Z}_n(P) \sim N(0, I_{d_n})$ conditionally on V_n and hence is independent of V_n. We set $\mathbb{G}_n^\star(P) \equiv \Sigma_n^{1/2}(P)\mathbb{Z}_n(P)$, which is independent of V_n. By Theorem 3.6.1 in Bogachev (1998), $\bar{\mathbb{G}}_n(P) \in \mathrm{range}\{\hat{\Sigma}_n(P)\}$ and therefore the spectral decomposition of $\hat{\Sigma}^{1/2}$ implies $\hat{\Sigma}_n^{1/2}(P)\mathbb{Z}_n(P) = \bar{\mathbb{G}}_n(P)$. Setting $\hat{\Delta}_n(P) \equiv \hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)$ and $\hat{\Delta}_n^{(j,k)}(P)$ be its (j,k) entry, we obtain from Lemma A.8, $\sup_{\|v\|_2=1}\langle v, a\rangle = \|a\|_2$ for any a, and $\|\cdot\|_\infty \leq \|\cdot\|_2$ that

$$\begin{aligned}E[\|\bar{\mathbb{G}}_n(P) - \mathbb{G}_n^\star(P)\|_\infty | V_n] &\lesssim \sqrt{\log(1+d_n)}\max_{1\leq j\leq d_n}(\sum_{k=1}^{d_n}(\hat{\Delta}_n^{(j,k)}(P))^2)^{1/2} \\ &= \sqrt{\log(1+d_n)}\sup_{\|v\|_2=1}\|\hat{\Delta}_n(P)v\|_\infty \leq \sqrt{\log(1+d_n)}\|\hat{\Delta}_n(P)\|_{o,2}. \quad (A.56)\end{aligned}$$

Hence, Fubini's theorem, Markov's inequality, and (A.56) yield for any $C > 0$

$$\begin{aligned}&\sup_{P\in\mathbf{P}} P(\|\bar{\mathbb{G}}_n(P) - \mathbb{G}_n^\star(P)\|_\infty > C^2\sqrt{\log(1+d_n)\delta_n} \text{ and } \|\hat{\Delta}_n(P)\|_{o,2} \leq C\sqrt{\delta_n}) \\ &\qquad \leq \sup_{P\in\mathbf{P}} E_P[\frac{\|\hat{\Delta}_n(P)\|_{o,2}}{C^2\sqrt{\delta_n}}1\{\|\hat{\Delta}_n(P)\|_{o,2} \leq C\sqrt{\delta_n}\}] \leq \frac{1}{C}. \quad (A.57)\end{aligned}$$

Since $\|\hat{\Delta}_n(P)\|_{o,2}^2 \leq \|\hat{\Sigma}_n(P) - \Sigma(P)\|_{o,2}$ by Theorem X.1.1 in Bhatia (1997), (A.57) and $\|\hat{\Sigma}_n(P) - \Sigma(P)\|_{o,2} = O_P(\delta_n)$ uniformly in $P \in \mathbf{P}$ imply the lemma. ■

Lemma A.8. *Let $Z \equiv (Z_1, \ldots, Z_p) \in \mathbf{R}^p$ be jointly Gaussian with $E[Z_j] = 0$ and $E[Z_j^2] \leq \sigma^2$ for all $1 \leq j \leq p$. It then follows that for any $q \geq 1$ there exists a constant $K_q < \infty$ such that $E[\|Z\|_\infty^q] \leq K_q(\sigma\sqrt{\log(1+p)})^q$.*

Proof: The result is well known and follows from Corollary 2.2.8 and Proposition A.2.4 in van der Vaart and Wellner (1996). ■

Lemma A.9. *Let* $(\mathbb{Z}_1, \ldots, \mathbb{Z}_d)' \in \mathbf{R}^d$ *be Gaussian with* $E[\mathbb{Z}_j] \geq \underline{\mu}$, $\mathrm{Var}\{\mathbb{Z}_j\} = \sigma^2 > 0$ *for all* j*, and set* $\mathbb{S} \equiv \max_{1 \leq j \leq d} \mathbb{Z}_j$ *and* $\mathrm{m} \equiv \mathrm{med}\{\mathbb{S}\}$*. Then, the distribution of* $\mathbb{S}$ *is absolutely continuous with density bounded by* $(2/\sigma)\max\{(\mathrm{m} - (\underline{\mu} \wedge 0))/\sigma, 1\}$*.*

Proof: First assume $\underline{\mu} \geq 0$ and let F and Φ denote the c.d.f. of $\mathbb{S}$ and a standard normal. Theorem 11.2 in Davydov et al. (1998) implies F is absolutely continuous and its density satisfies $F'(r) = q(r)\Phi'(r/\sigma)$ for some nondecreasing q. Moreover,

$$q(r)\sigma(1 - \Phi(r/\sigma)) \leq \int_r^\infty q(u)\Phi'(u/\sigma)du = \int_r^\infty F'(u)du = P(\mathbb{S} \geq r) \leq 1, \quad \text{(A.58)}$$

due to the function q being nondecreasing. Hence, $F'(r) = q(r)\Phi'(r/\sigma)$, result (A.58), and Mill's inequality implying $\Phi'(r)/(1 - \Phi(r)) \leq 2\max\{r, 1\}$ for all $r \in \mathbf{R}$ (see, e.g., pg. 64 in Chernozhukov et al. (2014)) yield the bound

$$F'(r) \leq \Phi'(r/\sigma)/(\sigma(1 - \Phi(r/\sigma))) \leq (2/\sigma)\max\{r/\sigma, 1\}. \quad \text{(A.59)}$$

Next note $P(\mathbb{S} \leq \mathrm{m} + \eta) \geq P(\max_j(\mathbb{Z}_j - E[\mathbb{Z}_j]) \leq \eta) > 0$ for any $\eta > 0$ due to $\mathrm{m} \geq E[\mathbb{Z}_j]$ for all j. As a result, Theorem 11.2 in Davydov et al. (1998) implies q is continuous at any $r > \mathrm{m}$, which together with $F'(r) = q(r)\Phi'(r/\sigma)$ establishes F is differentiable (with derivative F') at any $r > \mathrm{m}$. Setting $\Gamma \equiv \Phi^{-1} \circ F$, then observe $F = \Phi \circ \Gamma$ and hence at any $r > \mathrm{m}$ we obtain $F'(r) = \Phi'(\Gamma(r))\Gamma'(r)$ for Γ' the derivative of Γ. However, Γ' is decreasing since Γ is concave by Proposition 11.3 in Davydov et al. (1998), while $\Phi'(\Gamma(r))$ is decreasing on $[\mathrm{m}, +\infty)$ due to Φ' being decreasing on $[0, \infty)$, $\Gamma(r) \in [0, \infty)$ for any $r > \mathrm{m}$, and Γ being nondecreasing. Therefore, F' is decreasing on $(\mathrm{m}, +\infty)$ which together with (A.59) yields

$$\sup_{r \in (\mathrm{m}, +\infty)} F'(r) = \limsup_{r \downarrow \mathrm{m}} F'(r) \leq \Phi'(\mathrm{m}/\sigma)/(\sigma(1 - \Phi(\mathrm{m}/\sigma)). \quad \text{(A.60)}$$

Since (A.59) implies F' is bounded by $2\max\{\mathrm{m}/\sigma, 1\}/\sigma$ on $(-\infty, \mathrm{m}]$ and (A.60) implies the same bound on $(\mathrm{m}, +\infty)$, the lemma follows when $\underline{\mu} \geq 0$. If $\underline{\mu} < 0$, then set $\mathbb{S} = (\max_j \mathbb{Z}_j - \underline{\mu}) + \underline{\mu}$ and apply our first bound to $(\max_j \mathbb{Z}_j - \underline{\mu})$. ■

Lemma A.10. *Let* $C \subseteq \mathbf{R}^k$ *be a nonempty polyhedral set containing no lines and* $\mathcal{E}$ *denote its extreme points. Then:* $\mathcal{E} \neq \emptyset$ *and for any* $y \in \mathbf{R}^k$ *such that* $\sup_{c \in C}\langle c, y\rangle < \infty$*, it follows that* $\sup_{c \in C}\langle c, y\rangle = \max_{c \in \mathcal{E}}\langle c, y\rangle$*.*

Proof: Follows from Corollary 32.3.4 in Rockafellar (1970). ■

Lemma A.11. *Let $\mathcal{V}^{\mathrm{i}}$ be as defined in (9). Then, $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$ is a nonempty polyhedral set, contains no lines, and zero is one of its extreme points.*

Proof: Since $0 \in \mathcal{V}^{\mathrm{i}}$ and $\mathcal{V}^{\mathrm{i}}$ is polyhedral, Theorem 19.3 in Rockafellar (1970) implies $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$ is nonempty and polyhedral. Next note that $A'(AA')^{\dagger} = A^{\dagger}$ by Proposition 6.11.1(9) in Luenberger (1969), and hence since any $v \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$ satisfies $v = (AA')^{\dagger}s$ for some $s \in \mathcal{V}^{\mathrm{i}}$, we obtain for any $v = (AA')^{\dagger}s$ that

$$A'v = A'(AA')^{\dagger}s = A^{\dagger}s \leq 0. \tag{A.61}$$

For $N(A')^{\perp}$ the orthocomplement to the null space of A', note $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}} \subseteq N(A')^{\perp}$ because $(AA')^{\dagger} = (A')^{\dagger}A^{\dagger}$. Thus, if $\pm v \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$, then (A.61) implies $A'v = 0$ which together with $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}} \subseteq N(A')^{\perp}$ yields $v \in N(A') \cap N(A')^{\perp} = \{0\}$, implying $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$ contains no lines. Similarly, if $0 = \lambda v_1 + (1-\lambda)v_2$ with $v_1, v_2 \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}$ and $\lambda \in (0,1)$, then (A.61) implies $A'v_1 = A'v_2 = 0$ which together with $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}} \subseteq N(A')^{\perp}$ yields $v_1 = v_2 = 0$ implying zero is an extreme point. ■

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

Andrews, D. W. K. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** 119–157.

Andrews, I., Roth, J. and Pakes, A. (2019). Inference for linear conditional moment inequalities. Tech. rep., NBER.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, **90** 431–442.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, **106** 979–1014.

Bai, Y., Santos, A. and Shaikh, A. (2019). A practical method for testing many moment inequalities. *Working Paper. University of Chicago*.

Bai, Y., Shaikh, A. M. and Vytlacil, E. J. (2020). Partial identification of treatment effect rankings with instrumental variables. *Working Paper. University of Chicago*.

Balke, A. and Pearl, J. (1994). Counterfactual probabilities: Computational methods, bounds and applications. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 46–54.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, **92** 1171–1176.

Belloni, A., Chernozhukov, V., Chetverikov, D. and Fernández-Val, I. (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics*.

Bhatia, R. (1997). *Matrix Analysis*. Springer, NY.

Blundell, W., Gowrisankaran, G. and Langer, A. (2018). Escalation of scrutiny: The gains from dynamic enforcement of environmental regulations. Tech. rep., NBER.

Bogachev, V. I. (1998). *Gaussian measures*. 62, American Mathematical Soc.

Bugni, F. A., Canay, I. A. and Shi, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, **8** 1–38.

Canay, I. A. and Shaikh, A. M. (2017). Practical and theoretical advances in inference for partially identified models. *Advances in Economics and Econometrics*, **2** 271–306.

Chen, X. and Santos, A. (2018). Overidentification in regular models. *Econometrica*, **86** 1771–1817.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, **162** 47–70.

Chernozhukov, V., Chetverikov, D., Kato, K. and Koike, Y. (2019). Improved central limit theorem and bootstrap approximations in high dimensions. *arXiv preprint arXiv:1912.10529*.

Chernozhukov, V., Lee, S. and Rosen, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica*, **81** 667–737.

Chernozhukov, V., Newey, W. K. and Santos, A. (2015). Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311*.

Cox, G. and Shi, X. (2019). Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models. *arXiv preprint arXiv:1907.06317*.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local Properties of Distribuions of Stochastic Functionals*. American Mathematical Society, Providence.

Deb, R., Kitamura, Y., Quah, J. K.-H. and Stoye, J. (2017). Revealed price preference: Theory and stochastic testing. *Cowles Foundation Discussion Paper*.

Dudley, R. and Philipp, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **62** 509–552.

Fang, Z. and Santos, A. (2018). Inference on directionally differentiable functions. *The Review of Economic Studies*, **86** 377–412.

Fang, Z., Santos, A., Shaikh, A. M. and Torgovitsky, A. (2021). Inference for large-scale linear systems with known coefficients. *arXiv preprint arxiv:2009.08568v2*.

Fang, Z. and Seo, J. (2021). A projection framework for testing shape restrictions that form convex cones. *Econometrica*, **89** 2439–2458.

Fox, J. T., Kim, K. I., Ryan, S. P. and Bajari, P. (2011). A simple estimator for the distribution of random coefficients. *Quantitative Economics*, **2** 381–418.

Gandhi, A., Lu, Z. and Shi, X. (2019). Estimating demand for differentiated products with zeroes in market share data. *Working Paper. UW-Madison*.

Hájek, J. (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. *The Annals of Mathematical Statistics*, **32** 506–523.

Honoré, B. E. and Lleras-Muney, A. (2006). Bounds in competing risks models and the war on cancer. *Econometrica*, **74** 1675–1698.

Honoré, B. E. and Tamer, E. (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, **74** 611–629.

Illanes, G. and Padi, M. (2019). Competition, asymmetric information, and the annuity puzzle: Evidence from a government-run exchange in chile. *Working Paper*.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. **72** 1845–1857.

Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, **64** 555–574.

Kaido, H., Molinari, F. and Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, **87** 1397–1432.

Kamat, V. (2019). Identification with latent choice sets.

Kato, K. (2013). Quasi-Bayesian analysis of nonparametric instrumental variables models. *The Annals of Statistics*, **41** 2359–2390.

Kitagawa, T. (2015). A test for instrument validity. *Econometrica*, **83** 2043–2063.

Kitamura, Y. and Stoye, J. (2018). Nonparametric analysis of random utility models. *Econometrica*, **86** 1883–1909.

Kline, P. and Walters, C. (2021). Reasonable doubt: Experimental detection of job-level employment discrimination. *Econometrica*, **89** 765–792.

Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, **131** 1795–1848.

Laffërs, L. (2019). Bounding average treatment effects using linear programming. *Empirical Economics*, **57** 727–767.

Lazzati, N., Quah, J. and Shirai, K. (2018). Nonparametric analysis of monotone choice. *Available at SSRN 3301043*.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, NY.

Machado, C., Shaikh, A. M. and Vytlacil, E. J. (2019). Instrumental variables and the sign of the average treatment effect. *Journal of Econometrics*.

Manski, C. F. (2014). Identification of income–leisure preferences and evaluation of income tax policy. *Quantitative Economics*, **5** 145–174.

McFadden, D. and Richter, M. K. (1990). Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO* 161–186.

McMullen, P. (1970). The maximum numbers of faces of a convex polytope. *Mathematika*, **17** 179–184.

Monrad, D. and Philipp, W. (1991). Nearby variables with nearby conditional laws and a strong approximation theorem for hilbert space valued martingales. *Probability Theory and Related Fields*, **88** 381–404.

Nevo, A., Turner, J. L. and Williams, J. W. (2016). Usage-based pricing and demand for residential broadband. *Econometrica*, **84** 411–443.

Rockafellar, R. T. (1970). *Convex analysis*, vol. 28. Princeton university press.

Romano, J. P. and Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference – Special Issue in Honor of Ted Anderson*.

Romano, J. P., Shaikh, A. M. and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica*, **82** 1979–2002.

Strang, G. (1993). *Introduction to linear algebra*. Wellesley-Cambridge Press.

Tebaldi, P., Torgovitsky, A. and Yang, H. (2019). Nonparametric estimates of demand in the california health insurance exchange. Tech. rep., NBER.

Torgovitsky, A. (2019). Nonparametric inference on state dependence in unemployment. *Econometrica*, **87** 1475–1505.

van der Vaart, A. (1999). *Asymptotic Statistics*. Cambridge University Press, NY.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, NY.

Zhu, Y. (2020). Inference in non-parametric/semi-parametric moment equality models with shape restrictions. *Quantitative Economics*, **11** 609–636.