WORKING PAPER · NO. 2020-134

Inference for Large-Scale Linear Systems with Known Coefficients

Zheng Fang, Andres Santos, Azeem M. Shaikh, and Alexander Torgovitsky

SEPTEMBER 2020

Becker Friedman Institute | Big Data Initiative

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Inference for Large-Scale Linear Systems with Known Coefficients*

Zheng Fang
Department of Economics
Texas A&M University
zfang@tamu.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

September 17, 2020

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. 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 we describe 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 but estimable parameter, and the coefficients of the linear system are known in that A is a $p \times d$ known matrix.

As we discuss in detail in Section 2, the described hypothesis testing problem plays a central role in a surprisingly varied array of empirical settings. Tests of (1), for instance, are useful 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 (2019). Within the treatment effects literature, tests of (1) arise naturally when conducting inference on partially identified causal 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. In addition, we show that for a class of linear programming problems the null hypothesis that the linear program is feasible may be mapped into (1) – an observation that 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). See Remark 3.1 for details.

A common feature of the empirical studies that motivate our analysis is that the dimensions of $x \in \mathbf{R}^d$ and/or $\beta(P) \in \mathbf{R}^p$ are often quite high – e.g., in Nevo et al. (2016) the dimensions p and d are both in excess of 5000. We therefore focus on developing an inference procedure that remains computationally feasible in high-dimensional settings and asymptotically valid under favorable conditions on the relationship between the dimensions of A and the sample size n. To this end, we first obtain a novel geometric characterization of the null hypothesis that is the cornerstone of our approach to inference. Formally, we show that the null hypothesis in (1) is true if and only if

$\beta(P)$ belongs to the range of A and all angles between an estimable parameter and a known set in $\mathbf{R}^d$ are obtuse. This geometric result further provides, to the best of our knowledge, a new characterization of the feasibility of a linear program distinct from, but closely related to, Farkas' lemma that may be of independent interest.

Guided by our geometric characterization of the null hypothesis and our desire for computational and statistical reliability, we propose a test statistic that may be computed through linear programming. While the test statistic is not pivotal, we obtain a suitable critical value by relying on a bootstrap procedure that similarly only requires solving one linear program per bootstrap iteration. Besides delivering computational tractability, the linear programming structure present 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 for our test 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 the sample size, 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. In related work, Andrews et al. (2019) exploit a conditioning argument to develop methods for sub-vector inference in certain conditional moment inequality models. We show in Section 4.3.2 that we may use their insight in the same way to adapt our methodology to conduct inference for the same types problems they consider. Our analysis is also conceptually related to work on sub-vector inference in models involving moment inequalities or shape restrictions; see, among others, Romano and Shaikh (2008), Bugni et al. (2017), Kaido et al. (2019), Gandhi et al. (2019), Chernozhukov et al. (2015), Zhu (2019), and Fang and Seo (2019). 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.

The remainder of the paper is organized as follows. By way of motivation, we first discuss in Section 2 applications in which the null hypothesis in (1) naturally arises. In Sections 3 and 4, we establish our geometric characterization of the null hypothesis and the asymptotic validity of our test. Our simulation studies are contained in Section 5. Proofs and a guide to computation are contained in the Appendix. An R package for implementing our test is available at `https://github.com/conroylau/lpinfer`.

2 Applications

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

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 $h \in \{1, \ldots, H\}$ types of consumers that select among plans $k \in \{1, \ldots, K\}$. Each plan is characterized by a fee F_k, speed s_k, usage allowance $\bar{C}_k$, and overage price p_k. At day t, a consumer of type h with plan k has utility over usage c_t and numeraire y_t given by

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

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 dynamic problem faced by a type h consumer with plan k is then

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

where total wealth I is assumed to be large enough not to 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 $\theta_h \equiv (\zeta_h, \kappa_{1h}, \kappa_{2h}, \mu_h, \sigma_h)$. Therefore, for any function m of Z we obtain the moment restrictions

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

where E_P and E_{θ_h} denote expectations under the distribution P of Z and under θ_h respectively, while x_h is the unknown proportion of each type in the population. After specifying $H = 16807$ different types, Nevo et al. (2016) estimate $x \equiv (x_1, \ldots, x_H)$ by GMM while imposing the constraints that x be a probability measure. The authors then conduct inference on counterfactual demand, which for a known function a equals

$$\sum_{h=1}^{H} a(\theta_h)x_h,$$

by employing the constrained GMM estimator for x and the block bootstrap. We note, however, that the results in Fang and Santos (2018) imply the bootstrap is *inconsistent* for this problem. In contrast, the results in the present paper enable us to conduct asymptotically valid inference on counterfactual demand. For instance, by setting

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

we may obtain a confidence region for counterfactual demand through test inversion (in γ) of the null hypothesis in (1) – here, the final two constraints in (3) impose that probabilities add up to one and the hypothesized value for counterfactual demand. Other applications of the approach in Nevo et al. (2016) to inference in dynamic programs include Blundell et al. (2018) and Illanes and Padi (2019). ■

Example 2.2. (Treatment Effects). Kline and Walters (2016) examine the Head Start Impact Study (HSIS) in which participants where randomly assigned an offer to attend a Head Start school. Each participant can attend a Head Start school (h), other schools (c), or receive home care (n). We let $W \in \{0, 1\}$ denote whether an offer is made, $D(w) \in \{h, c, n\}$ denote potential treatment status, and $Y(d)$ denote test scores given treatment status $d \in \{h, c, n\}$. Under the assumption that a Head Start offer increases the utility of attending a Head Start school but leaves the utility of other programs unchanged, Kline and Walters (2016) partition participants into five groups that are determined by the values of $(D(0), D(1))$. We denote group membership by

$$C \in \{nh, ch, nn, cc, hh\}, \tag{4}$$

where, e.g., $C = nh$ corresponds to $(D(0), D(1)) = (n, h)$. Employing this structure, Kline and Walters (2016) show the local average treatment effect (LATE) identified by HSIS suffices for estimating the benefit cost ratio of a Head Start expansion. The impact of alternative policies, however, depends on partially identified parameters such as

$$\text{LATE}_{nh} \equiv E[Y(h) - Y(n)|C = nh]. \tag{5}$$

To estimate such partially identified parameters, Kline and Walters (2016) rely on a parametric selection model that delivers identification. In contrast, the results in this paper enable us to construct confidence regions for parameters such as LATE_{nh} within the nonparametric framework of Imbens and Angrist (1994). To this end note that, for any function m, the arguments in Imbens and Rubin (1997) imply

$$E_P[m(Y)1\{D=d\}|W=0]-E_P[m(Y)1\{D=d\}|W=1] \\ = \begin{cases} E[m(Y(d))1\{C=dh\}] & \text{if } d\in\{n,c\} \\ E[m(Y(h))1\{C\in\{nh,ch\}\}] & \text{if } d=h \end{cases} \quad (6)$$

while the null hypothesis that LATE_{nh} equals a hypothesized value γ is equivalent to

$$E[(Y(h)-Y(n))1\{C=nh\}]-\gamma P(C=hn)=0. \quad (7)$$

Provided the support of test scores is finite, results (6) and (7) imply that the null hypothesis that there exist a distribution of $(Y(n),Y(h),Y(d),C)$ satisfying (6) and $\text{LATE}_{nh}=\gamma$ is a special case of (1). As in Example 2.1, we may also obtain an asymptotically valid confidence region for LATE_{nh} through test inversion (in γ). Other examples of (1) arising in the treatment effects literature include Balke and Pearl (1994, 1997), Laffërs (2019), Machado et al. (2019), Kamat (2019), and 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

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

where (S_1,S_2) are possibly dependent random variables representing duration until death due to cancer and cardio-vascular disease, D is independent of (S_1,S_2) and denotes the implementation of the war on cancer, and (α,β) are unknown parameters. The observed variables are (T,I,D) where $T=t_k$ if $t_k\le T^*<t_{k+1}$ 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 that there exist known finite sets $\mathcal{S}(\alpha,\beta)$ and $\mathcal{S}_{k,i,d}(\alpha,\beta)\subseteq\mathcal{S}(\alpha,\beta)$ such that (α,β) belongs to the identified set if and only if there is a distribution $f(\cdot,\cdot)$ on $\mathcal{S}(\alpha,\beta)$ satisfying

$$\sum_{(s_1,s_2)\in\mathcal{S}_{k,i,d}(\alpha,\beta)} f(s_1,s_2)=P(T=t_k,I=i|D=d), \\ \sum_{(s_1,s_2)\in\mathcal{S}(\alpha,\beta)} f(s_1,s_2)=1, \text{ and } f(s_1,s_2)\ge 0 \text{ for all } (s_1,s_2)\in\mathcal{S}(\alpha,\beta), \quad (8)$$

where the first equality must hold for all $1\le k\le M$, $i\in\{1,2\}$, and $d\in\{0,1\}$. It follows from (8) 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). We note that, in a similar fashsion, 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 (Cover California), Tebaldi et al. (2019) model the observed plan choice Y by a consumer according to

$$Y \equiv \arg\max_{1 \le j \le J} V_j - \mathsf{p}_j,$$

where J denotes the number of available plans, $V = (V_1, \ldots, V_J)$ is an unobserved vector of valuations, and $\mathsf{p} \equiv (\mathsf{p}_1, \ldots, \mathsf{p}_J)$ denotes post-subsidy prices. Within the regulatory framework of Cover California, post-subsidy prices satisfy $\mathsf{p} = \pi(C)$ for some known function π and C a (discrete-valued) vector of individual characteristics that include age and county of residence. By decomposing C into subvectors (W, S) and assuming V is independent of S conditional on W, Tebaldi et al. (2019) then obtain

$$P(Y = j | C = c) = \int_{\mathcal{V}_j(\pi(c))} f_{V|W}(v|w) dv$$

for $f_{V|W}$ the density of V conditional on W and $\mathcal{V}_j(\mathsf{p}) \equiv \{v : v_j - \mathsf{p}_j \ge v_k - \mathsf{p}_k \text{ for all } k\}$. The authors further show there is a finite partition $\mathbb{V}$ of the support of V satisfying

$$P(Y = j | C = c) = \sum_{\mathcal{V} \in \mathbb{V} : \mathcal{V} \subseteq \mathcal{V}_j(\pi(x))} \int_{\mathcal{V}} f_{V|W}(v|w) dv \quad (9)$$

and such that the identified set for counterfactuals, such as the change in consumer surplus due to a change in subsidies, is characterized by functionals with the structure

$$\sum_{\mathcal{V} \in \mathbb{V} : \mathcal{V} \subseteq \mathcal{V}^\star} a(\mathcal{V}) \int_{\mathcal{V}} f_{V|W}(v|w) dv \quad (10)$$

for known function $a : \mathbb{V} \to \mathbf{R}$ and set $\mathcal{V}^\star$. Arguing as in Example 2.1, it then follows from (9) and (10) that confidence regions for the desired counterfactuals may be obtained through test inversion of hypotheses as in (1). Similar arguments allow us to apply our results to related discrete choice models such as the dynamic potential outcomes framework employed by Torgovitsky (2019) to measure state dependence. ■

Example 2.5. (Revealed Preferences). Building on McFadden and Richter (1990), Kitamura and Stoye (2018) develop a nonparametric specification test for random utility model by showing the null hypothesis has the structure in (1). We note, however, that the arguments showing the asymptotic validity of their test rely on a key restriction on the matrix A: Namely, that $(a_1 - a_0)'(a_2 - a_0) \ge 0$ for any distinct column vectors (a_0, a_1, a_2) of A. While such restriction on A is automatically satisfied in the random utility framework that motivates the analysis in Kitamura and Stoye (2018) and related work (Manski, 2014; Deb et al., 2017; Lazzati et al., 2018), we observe that it can fail in our previously discussed examples. ■

3 Geometry of the Null Hypothesis

In this section, we obtain a geometric characterization of the null hypothesis that guides the construction of our test in Section 4. To this end, we first introduce some additional notation that will prove useful throughout the rest of our analysis.

In what follows, we denote by $\mathbf{R}^k$ the Euclidean space of dimension k and reserve the use of p and d to denote the dimensions of the matrix A. For any two column vectors $(v_1, \ldots, v_k)' \equiv v$ and $(u_1, \ldots, u_k)' \equiv u$ in $\mathbf{R}^k$, we denote their inner product by $\langle v, u\rangle \equiv \sum_{i=1}^k v_i u_i$. The space $\mathbf{R}^k$ can be equipped with the norms $\|\cdot\|_q$ given by

$$\|v\|_q \equiv \{\sum_{i=1}^{k} |v_i|^q\}^{\frac{1}{q}}$$

for any $1 \leq q \leq \infty$, where as usual $\|\cdot\|_\infty$ is understood to equal $\|v\|_\infty \equiv \max_{1\leq i\leq k} |v_i|$. In addition, for any $k \times k$ matrix M, the norm $\|\cdot\|_q$ on $\mathbf{R}^k$ induces an operator norm

$$\|M\|_{o,q} \equiv \sup_{\|v\|_q \leq 1} \|Mv\|_q$$

on M; e.g., $\|M\|_{o,2}$ is the largest singular value of M, and $\|M\|_{o,\infty}$ is the maximum $\|\cdot\|_1$ norm of the rows of M. While the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ play a crucial role in our statistical analysis, our geometric analysis relies more heavily on the norm $\|\cdot\|_2$. In particular, for any closed convex set $C \subseteq \mathbf{R}^k$, we rely on the properties of the $\|\cdot\|_2$-metric projection operator $\Pi_C : \mathbf{R}^k \to C$, which for any vector $v \in \mathbf{R}^k$ is defined pointwise as

$$\Pi_C(v) \equiv \arg\min_{c\in C} \|v - c\|_2;$$

i.e., $\Pi_C(v)$ denotes the unique closest (under $\|\cdot\|_2$) element in C to the vector v. Finally, it will also be helpful to view the $p \times d$ matrix A as a linear map $A : \mathbf{R}^d \to \mathbf{R}^p$. The range $R \subseteq \mathbf{R}^p$ and null space $N \subseteq \mathbf{R}^d$ of A are defined as

$$R \equiv \{b \in \mathbf{R}^p : b = Ax \text{ for some } x \in \mathbf{R}^d\}$$
$$N \equiv \{x \in \mathbf{R}^d : Ax = 0\}.$$

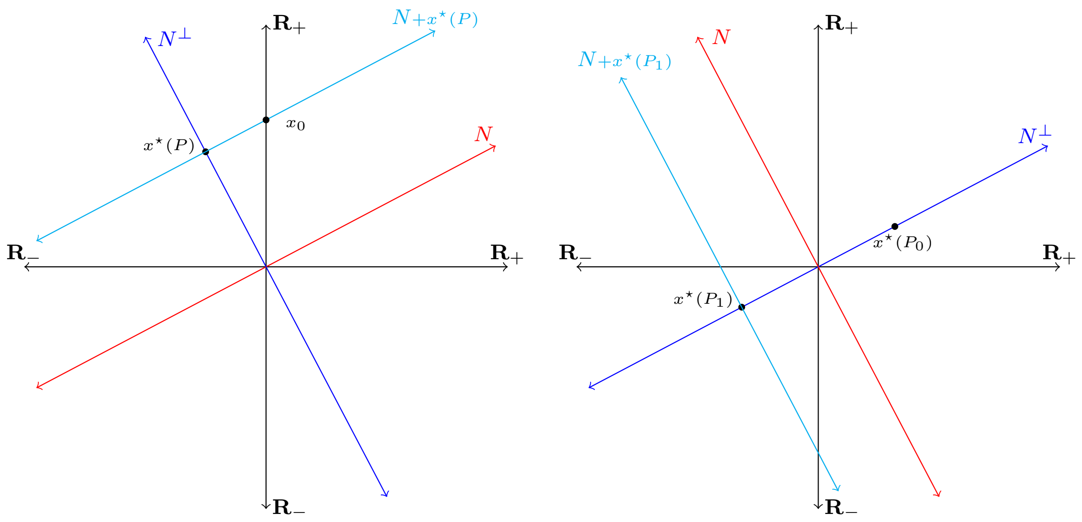
The null space N of A induces a decomposition of $\mathbf{R}^d$ through its orthocomplement

$$N^\perp \equiv \{y \in \mathbf{R}^d : \langle y, x\rangle = 0 \text{ for all } x \in N\};$$

i.e., any vector $x \in \mathbf{R}^d$ can be written as $x = \Pi_N(x) + \Pi_{N^\perp}(x)$ with $\langle \Pi_N(x), \Pi_{N^\perp}(x)\rangle = 0$. For succinctness, we denote such a decomposition of $\mathbf{R}^d$ as $\mathbf{R}^d = N \oplus N^\perp$.

Our first result is a well known consequence of the decomposition $\mathbf{R}^d = N \oplus N^\perp$, but we state it formally due to its importance in our derivations.

Figure 1: Illustration of when requirement (ii) in (11) is satisfied. Left panel: N and $N^{\perp}$ are such that requirement (ii) holds regardless of $x^{\star}(P)$. Right panel: N and $N^{\perp}$ are such that requirement (ii) holds if and only if $x^{\star}(P) \in \mathbf{R}^2_+$.



Lemma 3.1. *For any $\beta(P) \in \mathbf{R}^p$ there exists a unique $x^{\star}(P) \in N^{\perp}$ satisfying*

$$\Pi_R(\beta(P)) = A(x^{\star}(P)).$$

We note, in particular, that if $P \in \mathbf{P}_0$, then $\beta(P)$ must belong to the range of A and as a result $\Pi_R(\beta(P)) = \beta(P)$. Thus, for $P \in \mathbf{P}_0$, Lemma 3.1 implies that there exists a unique $x^{\star}(P) \in N^{\perp}$ satisfying $\beta(P) = A(x^{\star}(P))$. While $x^{\star}(P)$ is the unique solution in $N^{\perp}$, there may nonetheless exist multiple solution in $\mathbf{R}^d$. In fact, Lemma 3.1 and the decomposition $\mathbf{R}^d = N \oplus N^{\perp}$ imply that, provided $\beta(P) \in R$, we have

$$\{x \in \mathbf{R}^d : Ax = \beta(P)\} = x^{\star}(P) + N.$$

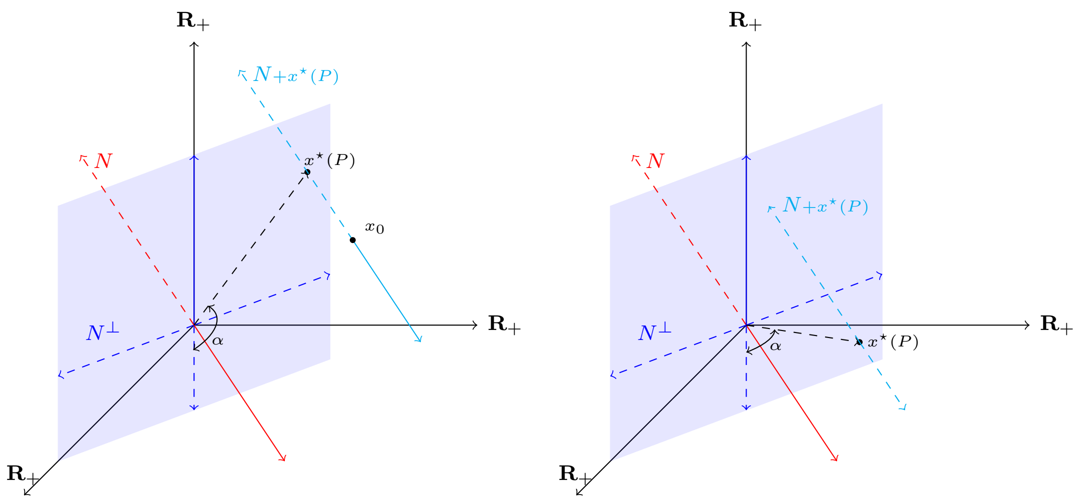
These observations allow us to characterize the null hypothesis in terms of two properties:

$$\text{(i) } \beta(P) \in R \qquad \text{(ii) } \{x^{\star}(P) + N\} \cap \mathbf{R}^d_+ \neq \emptyset; \tag{11}$$

i.e., (i) ensures *some* solution to the equation $Ax = \beta(P)$ exists, while (ii) ensures a *positive* solution $x_0 \in \mathbf{R}^d_+$ exists. Importantly, we note that these two conditions depend on P only through two identified objects: $\beta(P) \in \mathbf{R}^p$ and $x^{\star}(P) \in \mathbf{R}^d$.

Figure 1 illustrates these concepts in the simplest informative setting of $p = 1$ and $d = 2$, in which case N and $N^{\perp}$ are of dimension one and correspond to a rotation of the coordinate axes. Focusing on developing intuition for requirement (ii) in (11) we suppose that $\beta(P) \in R$ so that $A(x^{\star}(P)) = \beta(P)$. The left panel of Figure 1 displays a setting in which condition (ii) holds and an $x_0 \in \mathbf{R}^2_+$ satisfying $Ax_0 = A(x^{\star}(P)) = \beta(P)$

Figure 2: Illustration of Theorem 3.1 with $N = \{x \in \mathbf{R}^3 : x = (\lambda, \lambda, 0)'$ some $\lambda \in \mathbf{R}\}$ and $N^{\perp} \cap \mathbf{R}^3_- = \{x \in \mathbf{R}^3 : x = (0, 0, \lambda)'$ for some $\lambda \leq 0\}$. α denotes angle between $x^{\star}(P)$ and $N^{\perp} \cap \mathbf{R}^3_-$. Left panel: requirement (ii) in (11) holds and α is obtuse. Right panel: requirement (ii) in (11) fails and α is acute.



may be found even though $x^{\star}(P) \notin \mathbf{R}^2_+$. In fact, in the left panel of Figure 1, N and $N^{\perp}$ are such that requirement (ii) in (11) holds regardless of the value of $x^{\star}(P)$ (and hence regardless of P). In contrast, the right panel of Figure 1 displays a scenario in which N and $N^{\perp}$ are such that whether requirement (ii) is satisfied or not depends on $x^{\star}(P)$; e.g., (ii) holds for $x^{\star}(P_0)$ and fails for $x^{\star}(P_1)$. In fact, in the right panel of Figure 1, condition (ii) is satisfied if and only if $x^{\star}(P) \in \mathbf{R}^2_+$.

The preceding discussion highlights that whether condition (ii) in (11) is satisfied can depend delicately on the orientation of N and $N^{\perp}$ in $\mathbf{R}^d$ and the position of $x^{\star}(P)$ in $N^{\perp}$. Our next result, provides a tractable geometric characterization of this relationship.

Theorem 3.1. *For any $\beta(P) \in \mathbf{R}^p$ there exists an $x_0 \in \mathbf{R}^d_+$ satisfying $Ax_0 = \beta(P)$ if and only if $\beta(P) \in R$ and $\langle s, x^{\star}(P)\rangle \leq 0$ for all $s \in N^{\perp} \cap \mathbf{R}^d_-$.*

Theorem 3.1 establishes that the null hypothesis holds if and only if $\beta(P) \in R$ and the angle between $x^{\star}(P)$ and any vector $s \in N^{\perp} \cap \mathbf{R}^d_-$ is obtuse. It is straightforward to verify this relation is indeed present in Figure 1. The content of Theorem 3.1, however, is better appreciated in $\mathbf{R}^3$. Figure 2 illustrates a setting in which $N^{\perp} \cap \mathbf{R}^3_- = \{x \in \mathbf{R}^3 : x = (0, 0, \lambda)'$ for some $\lambda \leq 0\}$. In this case, condition (ii) in (11) holds if and only if the third coordinate of $x^{\star}(P)$ is (weakly) positive, which is equivalent to the angle between $x^{\star}(P)$ and $N^{\perp} \cap \mathbf{R}^3_-$ being obtuse. The left panel of Figure 2 depicts a setting in which the angle is obtuse, and an $x_0 \in \mathbf{R}^3_+$ satisfying $Ax_0 = A(x^{\star}(P))$ may be found even though $x^{\star}(P) \notin \mathbf{R}^3_+$. In contrast, in the left panel of Figure 2, the angle is acute and requirement (ii) in (11) fails to hold.

Remark 3.1. A finite-dimensional linear program can be written in the standard form

$$\min_{x \in \mathbf{R}^d} \langle c, x \rangle \text{ s.t. } Ax = \beta \text{ and } x \geq 0 \tag{12}$$

for some $c \in \mathbf{R}^d$, $\beta \in \mathbf{R}^p$, and $p \times d$ matrix A; see, e.g., Luenberger and Ye (1984). Theorem 3.1 thus provides a characterization of the feasibility of a linear program that is distinct from, but closely related to, Farkas' Lemma and may be of independent interest. We further observe that (12) implies that our results enable us to conduct inference on the value of a linear program whose standard form is such that A and c (as in (12)) are known while β potentially depends on the distribution of the data. This connection was implicitly employed in our discussion of many of the examples in Section 2, where we mapped the original linear programming formulations employed by the papers cited therein into the hypothesis testing problem in (1). ■

4 The Test

Theorem 3.1 provides us with the basis for constructing a variety of tests of the null hypothesis of interest. We next develop one such test, paying special attention to ensuring that it be computationally feasible in high-dimensional problems.

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

$$A^\dagger b \equiv \arg\min_{x \in \mathbf{R}^d} \|x\|_2^2 \text{ s.t. } x \in \arg\min_{\tilde{x} \in \mathbf{R}^d} \|A\tilde{x} - b\|_2^2;$$

i.e., $A^\dagger b$ is the minimum norm solution to minimizing $\|Ax - b\|_2$ over x. Importantly, we note that $A^\dagger b$ is well defined even if there is no solution to the equation $Ax = b$ ($b \notin R$) or the solution is not unique ($d > p$). For our purposes, it is also useful to note that $A^\dagger b$ is the unique element in $N^\perp$ satisfying $A(A^\dagger b) = \Pi_R(b)$, and we may thus interpret $A^\dagger$ as a linear map from $\mathbf{R}^p$ onto $N^\perp$; see Luenberger (1969). Despite its implicit definition, there exist multiple fast algorithms for computing $A^\dagger$. In Appendix S.3, we also provide a numerically equivalent reformulation of our test that avoids computing $A^\dagger$.

In order to build our test statistic, we will assume that there is a suitable 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

$$x^\star(P) = A^\dagger \beta(P) \tag{13}$$

for any $P \in \mathbf{P}_0$, which suggests a sample analogue estimator for $x^\star(P)$. However, while in our leading applications $d \geq p$, it is important to note that the existence of a solution to the equation $Ax = \beta(P)$ locally overidentifies the model when $d < p$ in the sense of Chen and Santos (2018). As a result, the sample analogue estimator for $x^\star(P)$ based on (13) may not be efficient when $d < p$, and we therefore instead set

$$\hat{x}_n^\star = A^\dagger \hat{C}_n \hat{\beta}_n \tag{14}$$

as an estimator for $x^\star(P)$. Here, $\hat{C}_n$ is a $p \times p$ matrix satisfying $\hat{C}_n \beta(P) = \beta(P)$ whenever $P \in \mathbf{P}_0$. For instance, the sample analogue estimator based on (13) corresponds to setting $\hat{C}_n = I_p$ for I_p the $p \times p$ identity matrix. More generally, it is straightforward to show that the specification in (14) also accommodates a variety of minimum distance estimators, which may be preferable to employing $\hat{C}_n = I_p$ when $p > d$.

The estimators $\hat{\beta}_n$ and $\hat{x}_n^\star$ readily allow us to devise a test based on the characterization of the null hypothesis obtained in Theorem 3.1. 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

$$\langle A^\dagger s, x^\star(P)\rangle \leq 0 \text{ for all } s \in \mathbf{R}^p \text{ s.t. } A^\dagger s \leq 0 \text{ (in } \mathbf{R}^d). \tag{15}$$

Thus, with the goal of detecting a violation of condition (15), we introduce the statistic

$$\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle = \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, A^\dagger \hat{C}_n \hat{\beta}_n\rangle \tag{16}$$

where

$$\hat{\mathcal{V}}_n^{\mathrm{i}} \equiv \{s \in \mathbf{R}^p : A^\dagger s \leq 0 \text{ and } \|\hat{\Omega}_n^{\mathrm{i}}(AA')^\dagger s\|_1 \leq 1\}. \tag{17}$$

Here, $\hat{\Omega}_n^{\mathrm{i}}$ 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., (15)). The inclusion of a $\|\cdot\|_1$-norm constraint in $\hat{\mathcal{V}}_n^{\mathrm{i}}$ in (17) ensures the statistic in (16) is not infinite with positive probability. The introduction of the matrix $\hat{\Omega}_n^{\mathrm{i}}$ in (17) grants us an important degree of flexibility in the family of test statistics we examine. In particular, we note that choosing $\hat{\Omega}_n^{\mathrm{i}}$ suitably ensures that the statistic in (16) is scale invariant.

By Theorem 3.1, in addition to (15), any $P \in \mathbf{P}_0$ must satisfy $\beta(P) \in R$. With the goal of detecting a violation of this second requirement, we introduce the statistic

$$\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{e}}} \sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}_n^\star\rangle = \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{e}}} \sqrt{n}\langle s, (I_p - AA^\dagger \hat{C}_n)\hat{\beta}_n\rangle \tag{18}$$

where

$$\hat{\mathcal{V}}_n^{\mathrm{e}} \equiv \{s \in \mathbf{R}^p : \|\hat{\Omega}_n^{\mathrm{e}} s\|_1 \leq 1\}.$$

Here, $\hat{\Omega}_n^{\mathrm{e}}$ 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 $\hat{\Omega}_n^{\text{e}} = I_p$, then by Hölder's inequality (18) equals $\|\hat{\beta}_n - A\hat{x}_n^{\star}\|_{\infty}$. As in (17), introducing $\hat{\Omega}_n^{\text{e}}$ enables us to ensure that the statistic in (18) is scale invariant if so desired. We also observe that in applications in which $d \geq p$ and A is full rank, the requirement $\beta(P) \in R$ is automatically satisfied due to $R = \mathbf{R}^p$ and (18) is identically zero due to $\hat{C}_n = I_p$.

As a test statistic T_n, we simply employ the maximum of (16) and (18); i.e., we set

$$T_n \equiv \max\{\sup_{s \in \hat{\mathcal{V}}_n^{\text{e}}} \sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}_n^{\star}\rangle, \sup_{s \in \hat{\mathcal{V}}_n^{\text{i}}} \sqrt{n}\langle A^{\dagger}s, \hat{x}_n^{\star}\rangle\}, \tag{19}$$

which we note can be computed through linear programming. We do not consider weighting the statistics (16) and (18) when taking the maximum in (19) because weighting them is numerically equivalent to scaling the matrices $\hat{\Omega}_n^{\text{i}}$ and $\hat{\Omega}_n^{\text{e}}$. A variety of alternative test statistics can of course be motivated by Theorem 3.1; some of which may be preferable to T_n in certain applications. A couple of remarks are therefore in order as to why our concern for computational reliability in high-dimensional models has led us to employing T_n. First, we avoided directly studentizing the inequalities in (16) in order to avoid a non-convex optimization problem. Instead, scale-invariance can be ensured by choosing $\hat{\Omega}_n^{\text{i}}$ suitably. Second, we avoided directly studentizing $(\hat{\beta}_n - A\hat{x}_n^{\star})$ in (18) because the asymptotic variance matrix of $(\hat{\beta}_n - A\hat{x}_n^{\star})$ is often rank deficient due to $(I_p - AA^{\dagger}\hat{C}_n)$ being a projection matrix. Third, an alternative norm, say $\|\cdot\|_2$, could be employed in the definitions of $\hat{\mathcal{V}}_n^{\text{i}}$ and $\hat{\mathcal{V}}_n^{\text{e}}$ in (17). At least in our experience, however, linear programs scale better than quadratic programs. In addition, employing $\|\cdot\|_1$-norm constraints implies distributional approximations to T_n can be obtained using coupling arguments with respect to $\|\cdot\|_{\infty}$, which are available under weaker conditions than coupling arguments with respect to, say, $\|\cdot\|_2$.

We next state a set of assumptions that will enable us to obtain a distributional approximation to T_n. Unless otherwise stated, all quantities are allowed to depend on n, though we leave such dependence implicit to avoid notational clutter.

Assumption 4.1. *For* $\text{j} \in \{\text{e}, \text{i}\}$*: (i)* $\hat{\Omega}_n^{\text{j}}$ *is symmetric; (ii) There is a symmetric matrix* $\Omega^{\text{j}}(P)$ *satisfying* $\|(\Omega^{\text{j}}(P))^{\dagger}(\hat{\Omega}_n^{\text{j}} - \Omega^{\text{j}}(P))\|_{o,\infty} = O_P(a_n/\sqrt{\log(1+p)})$ *uniformly in* $P \in \mathbf{P}$*; (iii)* $\text{range}\{\hat{\Omega}_n^{\text{j}}\} = \text{range}\{\Omega^{\text{j}}(P)\}$ *with probability tending to one uniformly in* $P \in \mathbf{P}$.

Assumption 4.2. *(i)* $\{Z_i\}_{i=1}^n$ *are i.i.d. with* $Z_i \in \mathbf{Z}$ *distributed according to* $P \in \mathbf{P}$*; (ii)* $\hat{x}_n^{\star} = A^{\dagger}\hat{C}_n\hat{\beta}_n$ *for some* $p \times p$ *matrix* $\hat{C}_n$ *satisfying* $\hat{C}_n\beta(P) = \beta(P)$ *for all* $P \in \mathbf{P}_0$*; (iii) There are* $\psi^{\text{i}}(\cdot, P) : \mathbf{Z} \to \mathbf{R}^p$ *and* $\psi^{\text{e}}(\cdot, P) : \mathbf{Z} \to \mathbf{R}^p$ *satisfying uniformly in* $P \in \mathbf{P}$

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

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

Assumption 4.3. *For* $\Sigma^{\rm j}(P) \equiv E_P[\psi^{\rm j}(Z,P)\psi^{\rm j}(Z,P)']$*: (i)* $E_P[\psi^{\rm j}(Z,P)] = 0$ *for all* $P \in \mathbf{P}$ *and* $\rm j \in \{e, i\}$*; (ii) The eigenvalues of* $(\Omega^{\rm j}(P))^{\dagger}\Sigma^{\rm j}(P)(\Omega^{\rm j}(P))^{\dagger}$ *are bounded in* $\rm j \in \{e,i\}$*, n, and* $P \in \mathbf{P}$*; (iii)* $\Psi(z,P) \equiv \|(\Omega^{\rm e}(P))^{\dagger}\psi^{\rm e}(z,P)\|_{\infty} \vee \|(\Omega^{\rm i}(P))^{\dagger}\psi^{\rm i}(z,P)\|_{\infty}$ *satisfies* $\sup_{P\in\mathbf{P}} \|\Psi(\cdot,P)\|_{P,3} \leq M_{3,\Psi} < \infty$ *with* $M_{3,\Psi} \geq 1$.

Assumption 4.4. *For* $\rm j \in \{e,i\}$*: (i)* $\psi^{\rm j}(Z,P) \in \text{range}\{\Omega^{\rm j}(P)\}$ *P-almost surely for all* $P \in \mathbf{P}$*; (ii)* $(I_p - AA^{\dagger}\hat{C}_n)\{\hat{\beta}_n - \beta(P)\} \in \text{range}\{\Sigma^{\rm e}(P)\}$ *and* $AA^{\dagger}\hat{C}_n\{\hat{\beta}_n - \beta(P)\} \in \text{range}\{\Sigma^{\rm i}(P)\}$ *with probability tending to one uniformly in* $P \in \mathbf{P}$.

Because $AA^{\dagger}\hat{C}_n$ is a projection matrix, the relevant asymptotic covariance matrices are often singular. In order to allow $\hat{\Omega}^{\rm i}_n$ and $\hat{\Omega}^{\rm e}_n$ to be sample standard deviation matrices, Assumption 4.1 therefore does not assume invertibility. Instead, Assumption 4.1(ii) requires a suitable form of consistency, and its rate is denoted by $a_n/\sqrt{\log(1+p)}$. Typically a_n will be of order $p/\sqrt{n}$ (up to logs). When $\hat{\Omega}^{\rm e}_n$ and $\hat{\Omega}^{\rm i}_n$ are sample standard deviation matrices, Assumption 4.1(iii) is easily verified due to rank deficiency resulting from the presence of projection matrices. Alternatively, we note that if we employ invertible (e.g., diagonal) weights $\hat{\Omega}^{\rm e}_n$ and $\hat{\Omega}^{\rm i}_n$, then Assumption 4.1(iii) is immediate. Assumptions 4.2(i)–(ii) formalize previously discussed conditions, while Assumption 4.2(iii) requires our estimators to be asymptotically linear with influence functions whose moments are disciplined by Assumption 4.3. Finally, Assumption 4.4(i), together with Assumption 4.1(iii), restricts the manner in which invertibility of $\hat{\Omega}^{\rm e}_n$ and $\hat{\Omega}^{\rm i}_n$ may fail – this condition is again easily verified if we employ invertible weights or sample standard deviation matrices. Similarly, Assumption 4.4(ii) ensures that the support of our estimators be contained in the support of their Gaussian approximations.

Before establishing our distributional approximation to T_n, we need to introduce a final piece of notation. We define the population analogues to $\hat{\mathcal{V}}^{\rm e}_n$ and $\hat{\mathcal{V}}^{\rm i}_n$ as

$$\mathcal{V}^{\rm e}(P) \equiv \{s \in \mathbf{R}^p : \|\Omega^{\rm e}(P)s\|_1 \leq 1\}$$
$$\mathcal{V}^{\rm i}(P) \equiv \{s \in \mathbf{R}^p : A^{\dagger}s \leq 0 \text{ and } \|\Omega^{\rm i}(P)(AA')^{\dagger}s\|_1 \leq 1\}, \tag{20}$$

and for $\psi^{\rm e}(Z,P)$ and $\psi^{\rm i}(Z,P)$ the influence functions in Assumption 4.2(iii) we set $\psi(Z,P) \equiv (\psi^{\rm e}(Z,P)', \psi^{\rm i}(Z,P)')'$ and denote the corresponding asymptotic variance by

$$\Sigma(P) \equiv E_P[\psi(Z,P)\psi(Z,P)'], \tag{21}$$

which note has dimension $2p \times 2p$. For notational simplicity we also define the rate

$$r_n \equiv M_{3,\Psi}(\frac{p^2\log^5(1+p)}{n})^{1/6} + a_n. \tag{22}$$

Our next theorem derives a distributional approximations 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, 4.4 hold, and* $r_n = o(1)$*. Then, there is* $(\mathbb{G}_n^{\rm e}(P)', \mathbb{G}_n^{\rm i}(P)')' \equiv \mathbb{G}_n(P) \sim N(0, \Sigma(P))$ *such that uniformly in* $P \in \mathbf{P}_0$ *we have*

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

4.2 The Critical Value

In order to obtain a suitable critical value, we will assume the availability of "bootstrap" estimates $(\hat{\mathbb{G}}_n^{\rm e\prime}, \hat{\mathbb{G}}_n^{\rm i\prime})'$ whose law conditional on the data $\{Z_i\}_{i=1}^n$ provides a consistent estimate for the joint distribution of $(\mathbb{G}_n^{\rm e}(P)', \mathbb{G}_n^{\rm i}(P)')'$. Given such estimates, we may follow a number of approaches for obtaining critical values; see, e.g., Section 4.3.1. Below we focus on a specific procedure that has favorable power properties in our 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

$$f(s, P) \equiv \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle = \sqrt{n}\langle A^\dagger s, x^\star(P)\rangle, \tag{23}$$

where the second equality follows from $A^\dagger\beta(P) = x^\star(P)$ for all $P \in \mathbf{P}_0$. While $f(\cdot, P)$ cannot be consistently estimated, we may nonetheless construct a suitable upper bound for it. To this end, it is useful to note that in applications some coordinates of $\beta(P)$ may equal a known value for all $P \in \mathbf{P}_0$; see, e.g., Examples 2.1-2.5. We therefore decompose $\beta(P) = (\beta_{\rm u}(P)', \beta_{\rm k}')'$ where $\beta_{\rm k}$ is a known constant for all $P \in \mathbf{P}_0$, and similarly decompose any $b \in \mathbf{R}^p$ into subvectors of comformable dimensions $b = (b_{\rm u}', b_{\rm k})'$. Employing these definitions, we introduce a restricted estimator $\hat\beta_n^{\rm r}$ for $\beta(P)$ by setting

$$\hat\beta_n^{\rm r} \in \arg\min_{b=(b_{\rm u}', b_{\rm k}')'} \sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star - A^\dagger b\rangle \text{ s.t. } b_{\rm k} = \beta_{\rm k},\ Ax = b \text{ for some } x \geq 0, \tag{24}$$

which may be computed through linear programming; see Appendix S.3 for details. Since $f(s, P) \leq 0$ for all $s \in \hat{\mathcal{V}}_n^{\rm i}$ 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 $s \in \hat{\mathcal{V}}_n^{\rm i}$. We thus set

$$\hat{\mathbb{U}}_n(s) \equiv \lambda_n\sqrt{n}\langle A^\dagger s, A^\dagger\hat\beta_n^{\rm r}\rangle, \tag{25}$$

which can be shown to be a suitable estimator for the upper bound $\lambda_n f(s, P)$ provided $\lambda_n \downarrow 0$ – we discuss choices of λ_n in Section 5. As a result, the function $\hat{\mathbb{U}}_n$ provides us with an asymptotic upper bound for the nuisance function $f(\cdot, P)$ on the set $\hat{\mathcal{V}}_n^{\rm i}$. In addition, 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 $s \in \hat{\mathcal{V}}_n^{\rm i}$ and (ii) There is a $b \in \mathbf{R}^p$ satisfying $Ax = b$ for some $x \geq 0$ such that $\hat{\mathbb{U}}_n(s) = \langle A^\dagger s, A^\dagger b\rangle$ for all $s \in \mathbf{R}^p$; see also our discussion in Section 4.3.1. ■

Step 2. Second, we note that the asymptotic approximation obtained in Theorem 4.1

is increasing (in a first-order stochastic dominance sense) in the value of the nuisance function under the pointwise partial order – e.g., if $f(s,P) \geq f(s,P')$ for all $s \in \mathcal{V}^{\mathrm{i}}(P)$, then the distribution of T_n under P first order stochastically dominates the distribution of T_n under P'. Hence, given the upper bound $\hat{\mathbb{U}}_n$ defined in Step 1, the preceding discussion suggests that, for a nominal level α test, we may employ the quantile

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

as a critical value for T_n. We note that computing $\hat{c}_n(1-\alpha)$ requires solving one linear program per bootstrap replication. ■

Given the above definitions, we finally define our test $\phi_n \in \{0,1\}$ to equal

$$\phi_n \equiv 1\{T_n > \hat{c}_n(1-\alpha)\};$$

i.e., we reject the null hypothesis whenever T_n exceeds $\hat{c}_n(1-\alpha)$. In order to establish the asymptotic validity of this test, we impose an additional assumption that enables us to derive the asymptotic properties of the bootstrap estimates $(\hat{\mathbb{G}}_n^{\mathrm{e}\prime}, \hat{\mathbb{G}}_n^{\mathrm{i}\prime})'$.

Assumption 4.5. *(i) There are exchangeable $\{W_{i,n}\}_{i=1}^n$ independent of $\{Z_i\}_{i=1}^n$ with*

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

uniformly in $P \in \mathbf{P}$ for $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$; (ii) For some $a,b>0$, $P(|W_{1,n} - E[W_{1,n}]| > t) \leq 2\exp\{-\frac{t^2}{b+at}\}$ 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) $\sup_{P\in\mathbf{P}} \|\Psi^2(\cdot,P)\|_{P,q} \leq M_{q,\Psi^2} < \infty$ for some $q \in (1,+\infty]$; (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 4.5(i) accommodates a variety of resampling schemes, such as the nonparametric, Bayesian, score, or weighted bootstrap, by simply requiring that $(\hat{\mathbb{G}}_n^{\mathrm{i}\prime}, \hat{\mathbb{G}}_n^{\mathrm{e}\prime})'$ be asymptotically equivalent to an exchangeable bootstrap estimate of the distribution of the (scaled) sample mean of $\psi(\cdot,P)$. In parallel to Assumption 4.2(iii), we note that Assumption 4.5(i) is a linearization assumption on our bootstrap estimates that is automatically satisfied whenever $(\hat{\mathbb{G}}_n^{\mathrm{i}\prime}, \hat{\mathbb{G}}_n^{\mathrm{e}\prime})'$ is linear in the data. Assumptions 4.5(ii)–(iii) impose moment and scale restrictions on the exchangeable bootstrap weights $\{W_{i,n}\}_{i=1}^n$, and are satisfied by commonly used resampling schemes – e.g., the nonparametric and Bayesian bootstrap, and the score or weighted bootstrap under appropriate choices of weights. Assumption 4.5(iv) potentially strengthens the moment restrictions in Assumption 4.3(iii) (if $q > 3/2$) and is imposed to sharpen our estimates of the coupling rate for the bootstrap statistics. Finally, Assumption 4.5(v) is a bootstrap analogue of the previously imposed Assumption 4.4(ii).

The introduced assumptions suffice for establishing that the law of $(\hat{\mathbb{G}}_n^{\mathrm{i}\prime}, \hat{\mathbb{G}}_n^{\mathrm{e}\prime})'$ conditional on the data is a suitable estimator for the distribution of $(\mathbb{G}_n^{\mathrm{e}}(P)', \mathbb{G}_n^{\mathrm{i}}(P)')'$ uniformly in $P \in \mathbf{P}$. Formally, we establish that $(\hat{\mathbb{G}}_n^{\mathrm{e}\prime}, \hat{\mathbb{G}}_n^{\mathrm{i}\prime})$ can be coupled (under $\|\cdot\|_\infty$) to a copy of $(\mathbb{G}_n^{\mathrm{e}}(P)', \mathbb{G}_n^{\mathrm{i}}(P)')'$ that is independent of the data 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 S.2.5 in the Supplemental Appendix. In particular, under appropriate moment restrictions, the bootstrap is consistent provided $p^2(\log^5(p) \vee \log(n))/n = o(1)$. To the best of our knowledge the stated consistency of the exchangeable bootstrap as p grows with n is a novel result that might be of independent interest.

Before establishing that the asymptotic size of our proposed test is at most its nominal level α, we need to introduce a final piece of notation. First, we note that the asymptotic approximation obtained in Theorem 4.1 contains the optimal value of two linear programs. The solution to these programs can be shown to belong to the sets

$$\mathcal{E}^{\mathrm{e}}(P) \equiv \{s \in \mathbf{R}^p : s \text{ is an extreme point of } \Omega^{\mathrm{e}}(P)\mathcal{V}^{\mathrm{e}}(P)\}$$
$$\mathcal{E}^{\mathrm{i}}(P) \equiv \{s \in \mathbf{R}^p : s \text{ is an extreme point of } (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)\}$$

almost surely. We note that while both sets are finite, the cardinality of $\mathcal{E}^{\mathrm{i}}(P)$ can grow exponentially in p, which results in coupling rates obtained via the high-dimensional central limit theorem to be suboptimal (Chernozhukov et al., 2017). In addition, it will be helpful to denote the standard deviations induced by $\mathbb{G}_n^{\mathrm{e}}(P)$ and $\mathbb{G}_n^{\mathrm{i}}(P)$ by

$$\sigma^{\mathrm{e}}(s,P) \equiv \{E_P[(\langle s, (\Omega^{\mathrm{e}}(P))^{\dagger}\mathbb{G}_n^{\mathrm{e}}(P)\rangle)^2]\}^{1/2}$$
$$\sigma^{\mathrm{i}}(s,P) \equiv \{E_P[(\langle \Omega^{\mathrm{i}}(P)s, (\Omega^{\mathrm{i}}(P))^{\dagger}\mathbb{G}_n^{\mathrm{i}}(P)\rangle)^2]\}^{1/2}$$

and denote their upper and (restricted) lower bounds over the set $\mathcal{E}^{\mathrm{e}}(P) \cup \mathcal{E}^{\mathrm{i}}(P)$ by

$$\bar{\sigma}(P) \equiv \sup_{s \in \mathcal{E}^{\mathrm{e}}(P)} \sigma^{\mathrm{e}}(s,P) \vee \sup_{s\in\mathcal{E}^{\mathrm{i}}(P)} \sigma^{\mathrm{i}}(s,P)$$
$$\underline{\sigma}(P) \equiv \inf_{s\in\mathcal{E}^{\mathrm{e}}(P):\sigma^{\mathrm{e}}(s,P)>0} \sigma^{\mathrm{e}}(s,P) \wedge \inf_{s\in\mathcal{E}^{\mathrm{i}}(P):\sigma^{\mathrm{i}}(s,P)>0} \sigma^{\mathrm{i}}(s,P),$$

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

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

Finally, for notational convenience 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, 4.2, 4.3, 4.4, 4.5 hold, and* $\alpha \in (0, 0.5)$. *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 it follows that*

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

In order to leverage our asymptotic approximations to establish the asymptotic validity of our test, Theorem 4.2 imposes an additional rate condition that constrains how p can grow with n. This rate condition depends on the matrix A and the weighting matrices $\Omega^{\mathrm{j}}(P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$. 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 $\Omega^{\mathrm{j}}(P)$ 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 choice of A. Finally, we note that if, as in much of the literature, one considers the case in which p does not grow with n, then Remark 4.1 implies that Theorem 4.2 holds under Assumptions 4.1–4.5 and the requirement $\lambda_n = o(1)$.

Remark 4.1. Whenever $\Omega^{\mathrm{j}}(P)$ is chosen to be 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 $\mathrm{m}(P)$. For instance, under such choice of $\Omega^{\mathrm{j}}(P)$, it is straightforward to show that

$$\max_{s \in \mathcal{E}^{\mathrm{i}}(P)} \sigma^{\mathrm{i}}(s, P) \leq \sup_{s : \|s\|_1 \leq 1} \{E_P[(\langle s, (\Omega^{\mathrm{i}}(P))^{\dagger} \mathbb{G}_n^{\mathrm{i}}(P) \rangle)^2]\}^{1/2} \leq 1$$

by employing the eigen-decomposition of $\Omega^{\mathrm{i}}(P)$ and, moreover, that $\bar{\sigma}(P) \leq 1$. Similarly, if $\sigma^{\mathrm{i}}(s, P) > 0$ for some $s \in \mathcal{E}^{\mathrm{i}}(P)$, then closely related arguments yield

$$\begin{aligned} &\min_{s \in \mathcal{E}^{\mathrm{i}}(P) : \sigma^{\mathrm{i}}(s,P) > 0} \sigma^{\mathrm{i}}(s, P) \\ &\qquad \geq \inf_{s : \|\Omega^{\mathrm{i}}(P)s\|_1 = 1} \{E_P[(\langle \Omega^{\mathrm{i}}(P)s, (\Omega^{\mathrm{i}}(P))^{\dagger} \mathbb{G}_n^{\mathrm{i}}(P) \rangle)^2]\}^{1/2} \geq \inf_{s : \|s\|_1 = 1} \|s\|_2 = \frac{1}{\sqrt{p}} \end{aligned} \tag{26}$$

and, moreover, that $\underline{\sigma}(P) \geq 1/\sqrt{p}$. Finally, by Hölder's and a maximal inequality it is possible to show that $\mathrm{m}(P) \lesssim \sqrt{\log(p)}$. We emphasize, however, that the universal (in A) bound in (26) can be quite conservative for a specific choice of A. For example, applying our procedure for testing whether a vector of means is non-positive corresponds to setting $A = I_p$, in which case $\underline{\sigma}(P) = 1$. ■

4.3 Extensions

We next discuss extensions to our results. For conciseness, we omit a formal analysis, but note they follow by similar arguments to those employed in proving Theorem 4.2.

4.3.1 Two Stage Critical Value

In Theorem 4.2, we focused on a particular choice of critical value due to its favorable power properties in our simulations. It is important to note, however, that other approaches are also available. For instance, an alternative critical value may be obtained by proceeding in a manner that is similar in spirit to the approach pursued by Romano et al. (2014) and Bai et al. (2019) for testing whether a finite-dimensional vector of populations means is nonnegative. Specifically, for some pre-specified $\gamma \in (0,1)$, let

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

and, in place of $\hat{\mathbb{U}}_n$ as introduced in (25), define the upper bound $\tilde{\mathbb{U}}_n$ to be given by

$$\tilde{\mathbb{U}}_n(s) \equiv \min\{\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle + \hat{c}_n^{(1)}(1-\gamma), 0\}.$$

The function $\tilde{\mathbb{U}}_n : \mathbf{R}^p \to \mathbf{R}$ may be interpreted as an upper confidence region for $f(\cdot, P)$ (as in (23)) with uniform (in $P \in \mathbf{P}_0$) asymptotic coverage probability $1-\gamma$. For a nominal level α test, we may then compare the test statistic T_n to the critical value

$$\hat{c}_n^{(2)}(1-\alpha+\gamma) \equiv$$
$$\inf\{u : P(\max\{\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle, \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}}\rangle + \hat{\mathbb{U}}_n(s)\} \leq u | \{Z_i\}_{i=1}^n) \geq 1-\alpha+\gamma\}.$$

Here, the $1-\alpha+\gamma$ quantile is employed instead of the $1-\alpha$ quantile in order to account for the possibility that $f(s,P) > \tilde{\mathbb{U}}_n(s)$ for some $s \in \hat{\mathcal{V}}_n^{\mathrm{i}}$. The asymptotic validity of the resulting test can be established under the same conditions imposed in Theorem 4.2. An appealing feature of the described approach is that it does not require selecting a "bandwidth" λ_n. However, we find in simulations that the power of the resulting test is lower than that of the test ϕ_n. Intuitively, this is due to $\tilde{\mathbb{U}}_n$ not satisfying $\tilde{\mathbb{U}}_n(s) = \langle A^\dagger s, A^\dagger b\rangle$ for some $b \in \mathbf{R}^p$ such that $Ax = b$ with $x \geq 0$. As a result, the upper bound $\tilde{\mathbb{U}}_n$ does not reflect the full structure of the null hypothesis.

4.3.2 Alternative Sampling Frameworks

While we have focused on i.i.d. settings for simplicity, we note that extensions to other asymptotic frameworks are conceptually straightforward. One interesting such extension is to the problem of sub-vector inference in a class of models defined by conditional moment inequalities. In particular, we may follow an insight due to Andrews et al. (2019), who note that in an empirically relevant class of models defined by conditional moment inequalities, the parameter of interest π is known to satisfy

$$E_P[G(D,\pi) - M(W,\pi)\delta|W] \leq 0 \text{ for some } \delta \in \mathbf{R}^{d_\delta} \tag{27}$$

where $G(D,\pi) \in \mathbf{R}^p$, $M(W,\pi)$ is a $p \times d_\delta$ matrix, and both are known functions of (D,W,π). Andrews et al. (2019) observe that the structure of these models is such that testing whether a specified value π_0 satisfies (27) is facilitated by conditioning on $\{W_i\}_{i=1}^n$. As we next argue, their important insight carries over to our framework.

For any $\delta \in \mathbf{R}^{d_\delta}$, let $\delta^+ \equiv \delta \vee 0$ and $\delta_- \equiv -(\delta \wedge 0)$, where $\vee$ and $\wedge$ denote coordinate-wise maximums and minimums – e.g., δ^+ and δ^- are the "positive" and "negative" parts of δ. We then observe that if π_0 satisfies (27), then it follows that

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

Hence, by setting P to denote the distribution of $\{D_i\}_{i=1}^n$ conditional on $\{W_i\}_{i=1}^n$, we may test the null hypothesis that π_0 satisfies (27) by letting $\beta(P)$ and A equal

$$\beta(P) \equiv \frac{1}{n}\sum_{i=1}^n E_P[G(D,\pi_0)|W_i] \qquad A \equiv [\frac{1}{n}\sum_{i=1}^n M(W_i,\pi_0),\ \ -\frac{1}{n}\sum_{i=1}^n M(W_i,\pi_0),\ \ -I_p]$$

and testing whether $\beta(P) = Ax$ for some $x \geq 0$ – note that, crucially, the matrix A does not depend on P due to the conditioning on $\{W_i\}_{i=1}^n$. By letting $\hat{\beta}_n$ equal

$$\hat{\beta}_n \equiv \frac{1}{n}\sum_{i=1}^n G(D_i,\pi_0),$$

our test remains largely the same as in Theorem 4.2, with the exception that the "bootstrap" estimates $(\hat{\mathbb{G}}_n^{\mathrm{e}\prime}, \hat{\mathbb{G}}_n^{\mathrm{i}\prime})'$ must be suitably consistent for the law of

$$((\Omega^{\mathrm{e}}(P))^\dagger(I_p - AA^\dagger\hat{C}_n)\sqrt{n}\{\hat{\beta}_n - \beta(P)\}', (\Omega^{\mathrm{i}}(P))^\dagger AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}')'$$

conditional on $\{W_i\}_{i=1}^n$ (instead of unconditionally, as in Theorem 4.2).

5 Simulations with a Mixed Logit Model

5.1 Model

Example 2.1 is one example of a class of mixture models considered by Fox et al. (2011). A simpler example with the same structure is a static, binary choice logit with random coefficients. In this model, a consumer makes a binary decision $Y \in \{0,1\}$ according to

$$Y = 1\{C_0 + C_1 W - U \geq 0\}, \tag{28}$$

where W is an observed random variable which we will think of as the price the consumer faces for buying a good ($Y = 1$), and (C_0, C_1, U) are latent random variables. The random coefficients, $V \equiv (C_0, C_1)$, represent the consumer's overall preference for buying the good (C_0) as well and their price sensitivity (C_1). The unobservable U is assumed to follow a standard logistic distribution, independently of (V, W).

A consumer of type $v = (c_0, c_1)$ facing price w buys the good with probability

$$P(Y = 1|W = w, V = v) = \frac{1}{1 + \exp(-c_0 - c_1 w)} \equiv \ell(w, v). \tag{29}$$

Bajari et al. (2007) and Fox et al. (2011) approximate the distribution of V using a discrete distribution with known support points $(v_1, \ldots, v_d)$ and unknown respective probabilities $x \equiv (x_1, \ldots, x_d)$. They also assume that V is independent of W, which is a natural baseline case under which random coefficient models are often studied (Ichimura and Thompson, 1998; Gautier and Kitamura, 2013). Under this assumption, (29) can be aggregated into a conditional moment equality in terms of observables:

$$P(Y = 1|W = w) = \sum_{j=1}^{d} x_j \ell(w, v_j). \tag{30}$$

A natural quantity of interest in this model is the elasticity of purchase probability with respect to price. For a consumer of type $v = (c_0, c_1)$ facing price $\bar{w}$, this equals

$$\epsilon(v, \bar{w}) \equiv \left(\frac{\partial}{\partial w} \ell(v, w) \Big|_{w=\bar{w}} \right) \frac{\bar{w}}{\ell(v, \bar{w})} = c_1 \bar{w}(1 - \ell(v, \bar{w})).$$

The cumulative distribution function (c.d.f.) of this elasticity, denoted $F_\epsilon(\cdot|\bar{w})$, satisfies

$$F_\epsilon(t|\bar{w}) \equiv P(\epsilon(V, \bar{w}) \leq t) = \sum_{j=1}^{d} 1\{\epsilon(v_j, \bar{w}) \leq t\} x_j \equiv a(t, \bar{w})' x, \tag{31}$$

where $a(t, \bar{w}) \equiv (a_1(t, \bar{w}), \ldots, a_d(t, \bar{w}))'$ with $a_j(t, \bar{w}) \equiv 1\{\epsilon(v_j, \bar{w}) \leq t\}$. We take the c.d.f. $F_\epsilon(\cdot|\bar{w})$ as our parameter of interest in the discussion ahead.

5.2 Data Generating Processes

We consider data generated from a class of mixed logit models parameterized as follows. The support of W is set to be either $\{0, 1, 2, 3\}$ or $\{0, .2, .4, \ldots, 3\}$, so that it has either 4 or 16 points respectively. In either case its distribution is uniform over these points.

The (known) support of V_0 is generating by taking a Sobol sequence of length $\sqrt{d}$ and rescaling it to lie in $[0, .5]$. Similarly, the support of V_1 is a Sobol sequence of length $\sqrt{d}$ rescaled to $[-3, 0]$. The joint distribution of V is taken to be uniform over the product of the two marginal supports, so that it has d support points in total.

5.3 Identification

Fox et al. (2012) provide identification results for the distribution of random coefficients in multinomial mixed logit model. The binary mixed logit model discussed in the previous section is a special case of these models. However, their conditions require W to be continuously distributed. When W is discretely distributed, one might naturally expect to find that the distributions of V and thus of $\epsilon(V, \bar{w})$ are only partially identified.

We explore this conjecture computationally. Letting $\text{supp}(W)$ denote the support of W, we may then express the identified set for the distribution of V as being equal to

$$\mathbb{X}^\star(P) \equiv \{x \in \mathbf{R}^d_+ : \sum_{j=1}^{d} x_j = 1, \ \sum_{j=1}^{d} x_j \ell(w, v_j) = P(Y = 1|W = w) \text{ for all } w \in \text{supp}(W)\}.$$

In addition, for any $t \in \mathbf{R}$, we denote the identified set for $F_\epsilon(t|\bar{w})$ by $\mathbb{A}^\star(t, \bar{w}|P)$, which simply equals the projection of $\mathbb{X}^\star(P)$ under the linear map introduced in (31):

$$\mathbb{A}^\star(t, \bar{w}|P) \equiv \left\{a(t, \bar{w})'x : x \in \mathbb{X}^\star(P)\right\}.$$

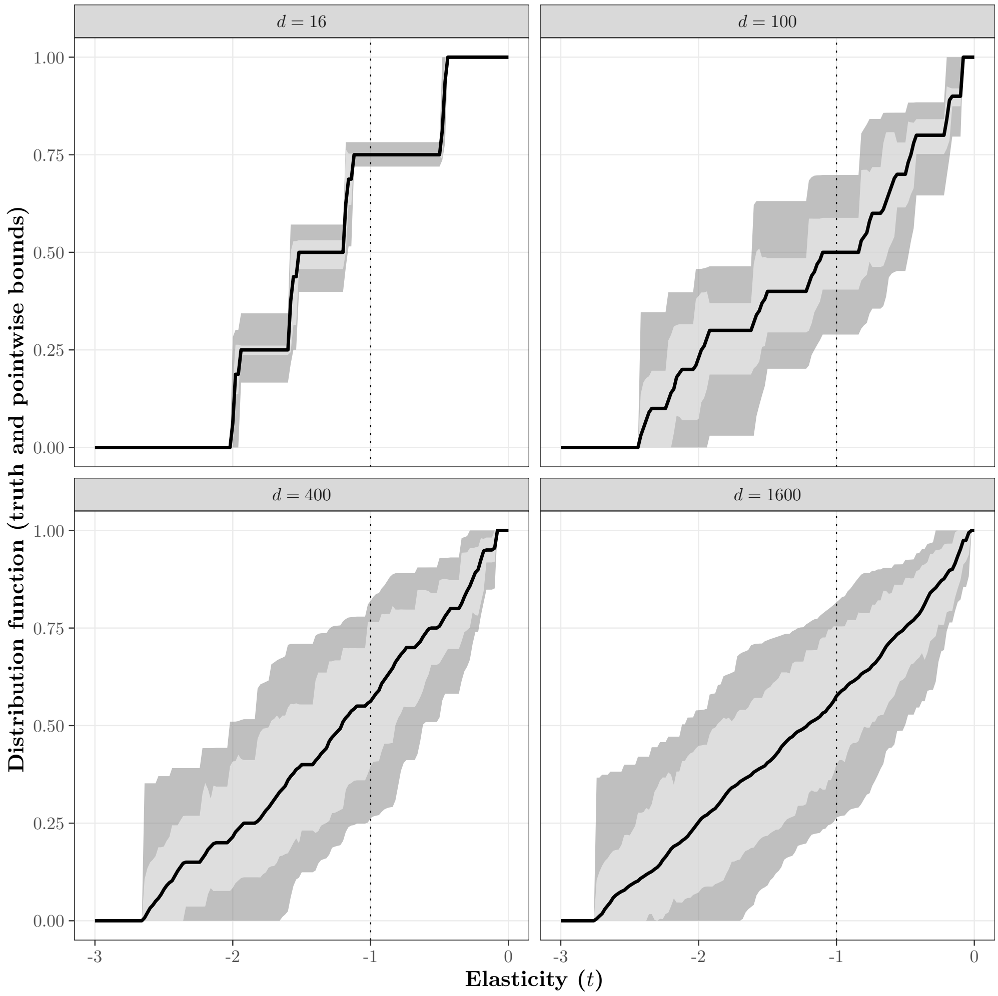
Since $\mathbb{X}^\star(P)$ is a system of linear equalities and inequalities, and $x \mapsto a(t, \bar{w})'x$ is scalar-valued and linear, $\mathbb{A}^\star(t, \bar{w}|P)$ is a closed interval (see, e.g. Mogstad et al., 2018, for a similar argument). The left endpoint of this interval is the solution to the linear program

$$\begin{aligned} \min_{x \in \mathbf{R}^d_+} a(t, \bar{w})'x \quad \text{s.t.} \quad & \sum_{j=1}^{d} x_j = 1 \\ \text{and} \quad & \sum_{j=1}^{d} x_j \ell(w, v_j) = P(Y = 1|W = w) \text{ for all } w \in \text{supp}(W), \end{aligned} \tag{32}$$

and the right endpoint is the solution to its maximization counterpart.

Figure 3 depicts $\mathbb{A}^\star(t, \bar{w}|P)$ as a function of t for $\bar{w} = 1$. The outer and inner bands depict the identified set when the support of W has four and sixteen points, respectively, while the solid line indicates the distribution under the actual data generating process. The identified sets are non-trivial and widen with the number of support points for the unobservable, V, as indexed by d. For $d = 16$, the bounds when W has sixteen support points are narrow, but numerically distinct from a point. This is because the system

Figure 3: Bounds on the distribution of price elasticity $F_\epsilon(t|1)$



These plots are based on the data generating processes described in Section 5.2. The solid black line is the actual value of $F_\varepsilon(t|1)$. The lighter color is $\mathbb{A}^\star(t, 1|P)$ when the support of W has sixteen points. The darker color is the same set when the support of W has only four points. The dotted vertical is the value $t = -1$ used in the Monte Carlo simulations in Section 5.5.

of moment equations that defines $\mathbb{X}^\star(P)$, while known to be nonsingular in principle, is sufficiently close to singular to matter for floating point arithmetic. That is, there are many values of x that satisfy these equations up to machine precision.

5.4 Test Implementation

As in Example 2.1, we may employ our results to test whether a hypothesized $\gamma \in \mathbf{R}$ belongs to the identified set for $F_\epsilon(t|\bar{w})$. Indexing the support of W to have $p - 2$

elements, we may then map such hypothesis into (1) by setting

$$\beta(P) = \begin{pmatrix} P(Y=1|W=w_1) \\ \vdots \\ P(Y=1|W=w_{p-2}) \\ 1 \\ \gamma \end{pmatrix} \qquad A = \begin{pmatrix} \ell(w_1, v_1) & \cdots & \ell(w_1, v_d) \\ \vdots & \vdots & \vdots \\ \ell(w_{p-2}, v_1) & \cdots & \ell(w_{p-2}, v_d) \\ 1 & \cdots & 1 \\ a_1(t, \bar{w}) & \cdots & a_d(t, \bar{w}) \end{pmatrix}. \tag{33}$$

We take $\hat{\beta}_n \equiv (\hat{\beta}_{\mathrm{u},n}, 1, \gamma)' \in \mathbf{R}^p$, where $\hat{\beta}_{\mathrm{u},n}$ is the sample analogue of the first $p-2$ (unknown) components of the vector $\beta(P)$. For designs with $d \geq p$, we set $\hat{x}_n^\star = A^\dagger \hat{\beta}_n$. When $d < p$ we instead set $\hat{x}_n^\star$ to be the unique solution to the minimization

$$\min_{x \in \mathbf{R}^d} \left(\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}} x\right)' \hat{\Xi}_n^{-1} \left(\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}} x\right) \quad \text{s.t.} \quad \sum_{j=1}^d x = 1 \quad \text{and} \quad a(t, \bar{w})' x = \gamma, \tag{34}$$

where A_{u} corresponds to the first $p-2$ rows of A and $\hat{\Xi}_n$ is the sample analogue estimator of asymptotic variance matrix of $\hat{\beta}_{\mathrm{u},n}$. As weighting matrices, we let $\hat{\Omega}_n^{\mathrm{e}}$ be the sample standard deviation matrix of $\hat{\beta}_n$, and $\hat{\Omega}_n^{\mathrm{i}}$ be the sample standard deviation of $A\hat{x}_n^\star$ computed from 250 draws of the nonparametric bootstrap.

We explore two rules for selecting λ_n. To motivate them, we note that an important theoretical restriction on λ_n is that, uniformly in $P \in \mathbf{P}_0$, it satisfy

$$\lambda_n \sqrt{n} \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger A(\hat{x}_n^\star - x^\star(P)) \rangle = o_P(1); \tag{35}$$

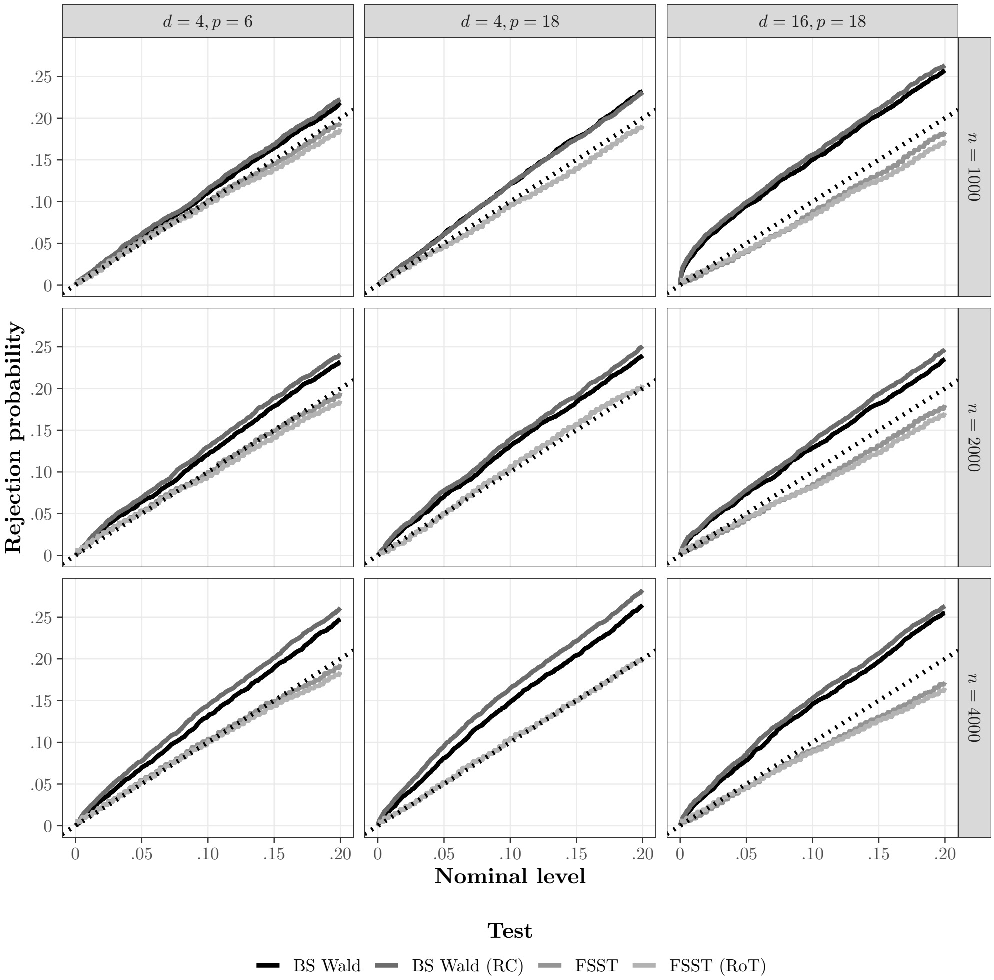
see Lemma S.2.1. Employing our coupling $A(\hat{x}_n^\star - x^\star(P)) \approx \mathbb{G}_n^{\mathrm{i}}(P)$, Hölder's and Markov's inequalities, and $E_P[\|(\Omega_n^{\mathrm{i}}(P))^\dagger \mathbb{G}_n^{\mathrm{i}}(P)\|_\infty] \leq \sqrt{2\log(e \vee p)}$ due to $\Omega_n^{\mathrm{i}}(P)$ being the standard deviation matrix suggests selecting λ_n to satisfy $\lambda_n \sqrt{\log(e \vee p)} = o(1)$ – here $a \vee b \equiv \max\{a, b\}$. For a concrete choice of λ_n, we rely on the law of iterated logarithm and let $\lambda_n^{\mathrm{r}} = 1/\sqrt{\log(e \vee p)\log(e \vee \log(e \vee n))}$. As an alternative to λ_n^{r}, we employ the bootstrap to approximate the law of (35). In particular, for some $\delta_n \downarrow 0$ we let $\lambda_n^{\mathrm{b}} \equiv \min\{1, \hat{\tau}_n(1-\delta_n)\}$ where $\hat{\tau}_n(1-\delta_n)$ denotes the $1-\delta_n$ quantile of

$$\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}} \rangle \tag{36}$$

conditional on the data. For a concrete choice of δ_n we let $\delta_n = 1/\sqrt{\log(e \vee \log(e \vee n))}$.

In Appendix S.3, we describe the computation of our test in more detail. In particular, we show how to reformulate the entire sequence into a series of linear programming problems. We also suggest reformulations that improve the stability of these linear programs while also making it unnecessary to compute $A^\dagger$. An R package for implementing our test is available at https://github.com/conroylau/lpinfer.

Figure 4: Null rejection probabilities for (nearly) point-identified designs



The dotted line is the 45 degree line. "FSST" refers to the test developed in this paper with λ_n^{b}, whereas "FSST (RoT)" uses the rule of thumb choice λ_n^{r}. "BS Wald" corresponds to a Wald test using bootstrap estimates of the standard errors. "BS Wald (RC)" is the same procedure but with standard errors based on bootstrapping with a re-centered GMM criterion. The null hypothesis is that $F_\epsilon(-1|1)$ is equal to its true value. In the case of $d = 16, p = 18$, which is set identified but with a very narrow identified set, we test the null hypothesis that $F_\epsilon(-1|1)$ is equal to the midpoint of the identified set.

5.5 Monte Carlo Simulations

We start by examining the null rejection probabilities of our testing procedure by setting γ to be the lower bound of the population identified set computed via (32) with $t = -1$ and $\bar{w} = 1$. In unreported simulations we found setting γ to be the upper bound of the identified set yielded similar results. We consider sample sizes of $n = 1000$, 2000 and 4000 for each of the data generating processes discussed in Section 5.2. All results are based on 5000 Monte Carlo replications and 250 nonparametric bootstrap draws.

Table 1: Null rejection probabilities for a nominal 0.05 test

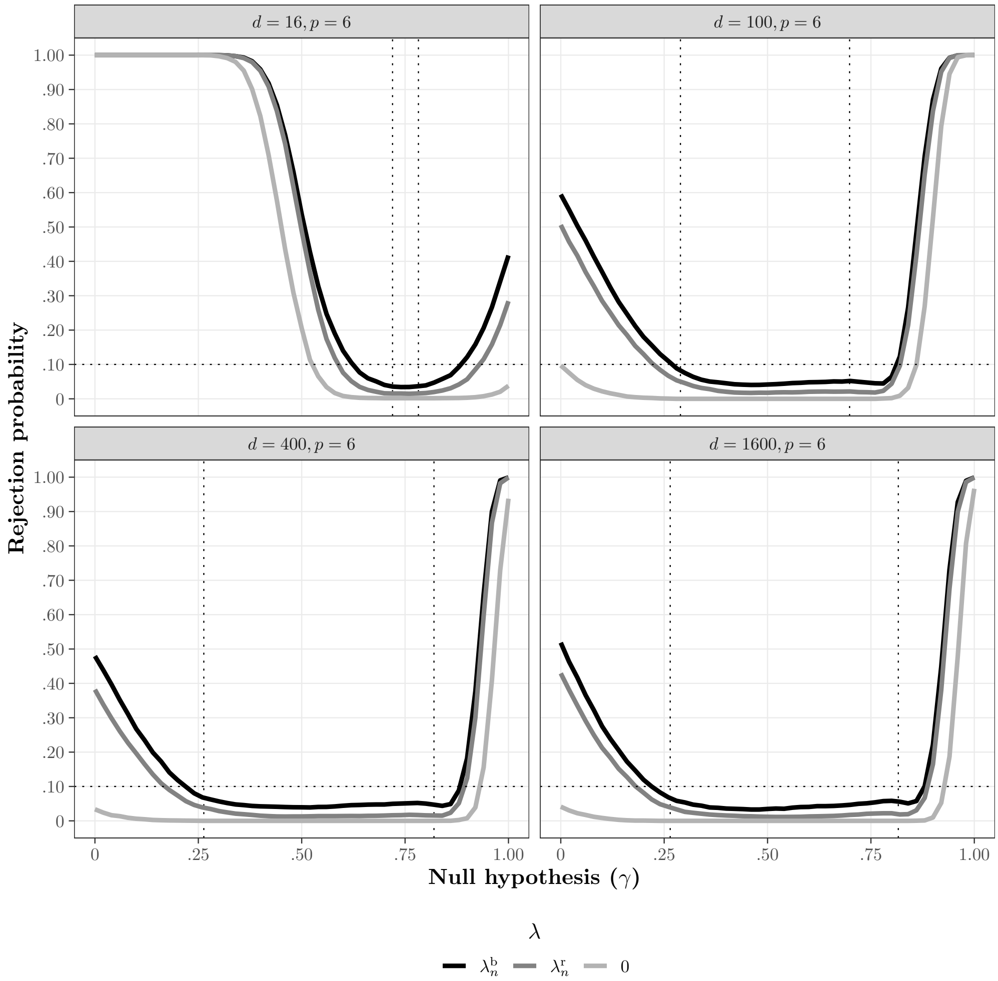
n	p	Test	d = 4	16	100	400	1600
1000	6	BS Wald	.058	–	–	–	–
		BS Wald (RC)	.061	–	–	–	–
		FSST	.055	.014	.032	.021	.023
		FSST (RoT)	.052	.009	.013	.007	.008
	18	BS Wald	.061	.093	–	–	–
		BS Wald (RC)	.061	.099	–	–	–
		FSST	.046	.041	.017	.014	.013
		FSST (RoT)	.046	.042	.014	.011	.009
2000	6	BS Wald	.064	–	–	–	–
		BS Wald (RC)	.069	–	–	–	–
		FSST	.053	.012	.040	.029	.031
		FSST (RoT)	.052	.004	.022	.012	.016
	18	BS Wald	.070	.074	–	–	–
		BS Wald (RC)	.078	.079	–	–	–
		FSST	.050	.044	.019	.014	.017
		FSST (RoT)	.050	.046	.013	.011	.012
4000	6	BS Wald	.070	–	–	–	–
		BS Wald (RC)	.077	–	–	–	–
		FSST	.055	.019	.042	.031	.030
		FSST (RoT)	.053	.003	.027	.019	.017
	18	BS Wald	.081	.080	–	–	–
		BS Wald (RC)	.095	.087	–	–	–
		FSST	.052	.046	.026	.022	.024
		FSST (RoT)	.052	.047	.020	.015	.017

The test abbreviations are described in the notes for Figure 4. The null hypothesis is that $F_\epsilon(-1|1)$ is equal to the lower bound of the population identified set.

We first consider the designs in which $p - 2 \geq d$ so that $F_\epsilon(-1|1)$ is (nearly) point identified. In this case, one might alternatively consider estimating probability weights x_0 satisfying the moment restrictions in (30) by constrained GMM, and then conducting inference on $F_\epsilon(-1|1)$ using a bootstrapped Wald test. For example, this is the approach that appears to have been taken by Nevo et al. (2016) in the related setting discussed in Example 2.1. However, the non-negativity constraints on x_0 imply that the bootstrap will generally not be consistent in this case (Fang and Santos, 2018).

We demonstrate this point in Figure 4 with plots of the actual and nominal level for both our procedure (FSST) and for the bootstrapped Wald test based on constrained GMM. The latter exhibits significant size distortions. For example the GMM test with nominal level 5% rejects nearly 10% of the time when $d = 16, p = 18$ and $n = 1{,}000$, and a nominal level %10 test rejects 15% of the time when $d = 4, p = 18$, and $n = 4{,}000$. Re-centering the GMM criterion before conducting this test (e.g. Hall and Horowitz, 1996) leads to even greater over-rejection. In contrast, FSST has nearly equal nominal

Figure 5: Power curves for FSST nominal 0.10 test, $n = 2000$

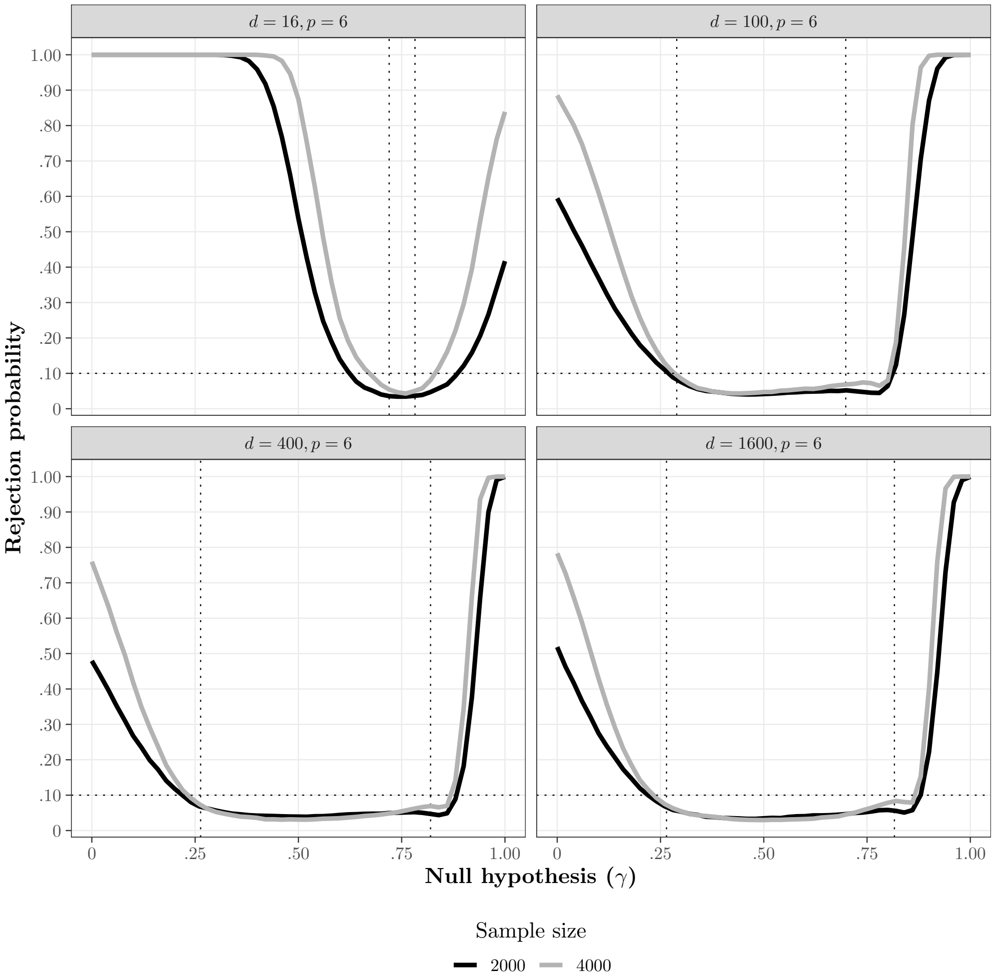


The vertical dotted lines indicate the lower and upper bounds of the population identified set. The horizontal dotted line indicates the nominal level (0.10).

and actual levels when using either λ_n^{b} or the rule-of-thumb (RoT) choice, λ_n^{r}.

In Table 1, we report empirical rejection rates for our procedure using all of the designs, including the partially identified ones. Our approach has null rejection probabilities approximately no greater than the nominal level across all different data generating processes and sample sizes, even with d as high as 1600. Figure 5 illustrates the impact that λ_n has on the power of the test. Both λ_n^{b} and λ_n^{r} provide considerable power gains over the conservative choice of $\lambda_n = 0$. Figure 6 shows how power increases for the choice $\lambda_n = \lambda_n^{\mathrm{b}}$ as the sample size increases from 2000 to 4000.

Figure 6: Power curves for FSST nominal 0.10 test, $\lambda_n = \lambda_n^{\mathrm{b}}$



The vertical dotted lines indicate the lower and upper bounds of the population identified set. The horizontal dotted line indicates the nominal level (0.10).

References

Andrews, I., Roth, J. and Pakes, A. (2019). Inference for linear conditional moment inequalities. Tech. rep., National Bureau of Economic Research.

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

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

Bajari, P., Fox, J. T. and Ryan, S. P. (2007). Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients. *American Economic Review*, **97** 459–463.

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.

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

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.

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. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, **45** 2309–2352.

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

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

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

Fang, Z. and Seo, J. (2019). A general framework for inference on shape restrictions. *arXiv preprint arXiv:1910.07689*.

Fox, J. T., il Kim, K., Ryan, S. P. and Bajari, P. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, **166** 204–212.

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.*

Gautier, E. and Kitamura, Y. (2013). Nonparametric Estimation in Random Coefficients Binary Choice Models. *Econometrica*, **81** 581–607.

Hall, P. and Horowitz, J. L. (1996). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica*, **64** 891–916.

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.

Ichimura, H. and Thompson, T. (1998). Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *Journal of Econometrics*, **86** 269–295.

Illanes, G. and Padi, M. (2019). Competition, asymmetric information, and the annuity puzzle: Evidence from a government-run exchange in chile. *Center for Retirement Research at Boston College*.

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.

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

Kline, P. and Walters, C. (2019). Audits as evidence: Experiments, ensembles, and enforcement. *arXiv preprint arXiv:1907.06622*.

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.

Lafférs, 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, New York.

Luenberger, D. G. and Ye, Y. (1984). *Linear and nonlinear programming*, vol. 2. Springer.

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.

Mogstad, M., Santos, A. and Torgovitsky, A. (2018). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica*, **86** 1589–1619.

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

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.

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

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

Zhu, Y. (2019). Inference in non-parametric/semi-parametric moment equality models with shape restrictions. *Quantitative Economics (forthcoming)*.

Inference for Large-Scale Linear Systems with Known Coefficients: Supplemental Appendix

Zheng Fang
Department of Economics
Texas A&M University

Andres Santos
Department of Economics
UCLA

Azeem M. Shaikh
Department of Economics
University of Chicago

Alexander Torgovitsky
Department of Economics
University of Chicago

September 17, 2020

This Supplemental Appendix to "Inference for Large-Scale Systems of Linear Inequalities" is organized as follows. Appendix S.1 contains the proofs of the theoretical results in Section 3. The proof of Theorems 4.1, 4.2, and required auxiliary results, is contained in Appendix S.2. Throughout, we employ the following notation:

$a \lesssim b$	$a \leq Mb$ for some constant M that is universal in the proof.
$C^{\perp}$	For any set $C \subseteq \mathbf{R}^k$, $C^{\perp} \equiv \{x \in \mathbf{R}^k : \langle x, y \rangle = 0$ for all $y \in C\}$.
N	The null space of the map $A : \mathbf{R}^d \to \mathbf{R}^p$.
R	The range of the map $A : \mathbf{R}^d \to \mathbf{R}^p$.
Π_C	For any closed convex $C \subseteq \mathbf{R}^k$, $\Pi_C y \equiv \arg\min_{x \in C} \|y - x\|_2$.
$x^{\star}(P)$	The unique element of $N^{\perp}$ solving $\Pi_R(\beta(P)) = A(x^{\star}(P))$.

S.1 Results for Section 3

Proof of Lemma 3.1: First note that by definition of R, there exists a $x(P) \in \mathbf{R}^d$ such that $\Pi_R(\beta(P)) = A(x(P))$. Hence, since $\mathbf{R}^d = N \oplus N^{\perp}$ by Theorem 3.4.1 in Luenberger

*We thank Denis Chetverikov, Patrick Kline, and Adriana Lleras-Muney for helpful comments. 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.

(1969), it follows that $x(P) = \Pi_N(x(P)) + \Pi_{N^\perp}(x(P))$ and we set $x^\star(P) = \Pi_{N^\perp}(x(P))$. Since $A(\Pi_N(x(P))) = 0$ by definition of N, we then obtain that

$$\Pi_R(\beta(P)) = A(x(P)) = A(\Pi_{N^\perp}x(P) + \Pi_N x(P)) = A(x^\star(P)).$$

To see $x^\star(P)$ is the unique element in $N^\perp$ satisfying $\Pi_R(\beta(P)) = A(x^\star(P))$, let $\tilde{x}(P) \in N^\perp$ be any element satisfying $A(\tilde{x}(P)) = \Pi_R(\beta(P)) = A(x^\star(P))$. Since $A(\tilde{x}(P) - x^\star(P)) = 0$, it then follows that $\tilde{x}(P) - x^\star(P) \in N$. However, we also have $\tilde{x}(P) - x^\star(P) \in N^\perp$ since $\tilde{x}(P), x^\star(P) \in N^\perp$ and $N^\perp$ is a vector subspace of $\mathbf{R}^d$. Thus, we obtain $x^\star(P) \cap \tilde{x}(P) \in N \cap N^\perp$, and since $N \cap N^\perp = \{0\}$ we can conclude $\tilde{x}(P) = x^\star(P)$, which establishes $x^\star(P)$ is indeed unique. ■

Proof of Theorem 3.1: Fix any $\beta(P) \in \mathbf{R}^p$ and recall $\Pi_R(\beta(P))$ denotes its projection onto R (under $\|\cdot\|_2$). Next note that by Farkas' Lemma (see, e.g., Corollary 5.85 in Aliprantis and Border (2006)) it follows that the statement

$$\Pi_R(\beta(P)) = A\tilde{x} \text{ for some } \tilde{x} \geq 0 \quad \text{(S.1)}$$

holds if and only if there *does not* exist a $y \in \mathbf{R}^p$ satisfying the following inequalities:

$$A'y \leq 0 \text{ (in } \mathbf{R}^d) \text{ and } \langle y, \Pi_R(\beta(P))\rangle > 0. \quad \text{(S.2)}$$

In particular, there being no $y \in \mathbf{R}^p$ satisfying (S.1) is equivalent to the statement

$$\langle y, \Pi_R(\beta(P))\rangle \leq 0 \text{ for all } y \in \mathbf{R}^p \text{ such that } A'y \leq 0 \text{ (in } \mathbf{R}^d). \quad \text{(S.3)}$$

Next note Lemma 3.1 implies that there is a unique $x^\star(P) \in N^\perp$ such that $\Pi_R(\beta(P)) = Ax^\star(P)$. Therefore, $\langle y, Ax^\star(P)\rangle = \langle A'y, x^\star(P)\rangle$ implies (S.3) is equivalent to

$$\langle A'y, x^\star(P)\rangle \leq 0 \text{ for all } y \in \mathbf{R}^p \text{ such that } A'y \leq 0 \text{ (in } \mathbf{R}^d). \quad \text{(S.4)}$$

Moreover, since $\{A'y : y \in \mathbf{R}^p \text{ and } A'y \leq 0\} = \text{range}\{A'\} \cap \mathbf{R}^d_-$, (S.4) is equivalent to

$$\langle s, x^\star(P)\rangle \leq 0 \text{ for all } s \in \text{range}\{A'\} \cap \mathbf{R}^d_- \quad \text{(S.5)}$$

Theorem 6.7.3 in Luenberger (1969) implies that $\text{range}\{A'\} = N^\perp$. Since $\text{range}\{A'\}$ is closed, we have that condition (S.5) is satisfied if and only if the following holds:

$$\langle s, x^\star(P)\rangle \leq 0 \text{ for all } s \in N^\perp \cap \mathbf{R}^d_-. \quad \text{(S.6)}$$

In summary, we have shown that statement (S.1) is satisfied if and only if condition (S.6) holds. Since in addition $\beta(P) \in R$ if and only if $\beta(P) = \Pi_R(\beta(P))$, the claim of the theorem follows. ■

S.2 Results for Section 4

Proof of Theorem 4.1: First note that by Lemma S.2.4 there exists a Gaussian vector $(\mathbb{G}_n^{\rm e}(P)', \mathbb{G}_n^{\rm i}(P)')' \equiv \mathbb{G}_n(P) \in \mathbf{R}^{2p}$ with $\mathbb{G}_n(P) \sim N(0, \Sigma(P))$ satisfying

$$\|(\Omega^{\rm e}(P))^\dagger\{(I_p - AA^\dagger\hat{C}_n)\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \mathbb{G}_n^{\rm e}(P)\}\|_\infty = O_P(r_n)$$
$$\|(\Omega^{\rm i}(P))^\dagger\{AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \mathbb{G}_n^{\rm i}(P)\}\|_\infty = O_P(r_n) \qquad \text{(S.7)}$$

uniformly in $P \in \mathbf{P}$. Further note that Assumption 4.4(i) implies range$\{\Sigma^{\rm j}(P)\} \subseteq$ range$\{\Omega^{\rm j}(P)\}$ for $\rm j \in \{e, i\}$ and $P \in \mathbf{P}$. Therefore, Assumption 4.4(ii) yields

$$(I_p - AA^\dagger\hat{C}_n)\sqrt{n}\{\hat{\beta}_n - \beta(P)\} \in \text{range}\{\Omega^{\rm e}(P)\}$$
$$AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\} \in \text{range}\{\Omega^{\rm i}(P)\} \qquad \text{(S.8)}$$

with probability tending to one uniformly in $P \in \mathbf{P}$. Next, note that $AA^\dagger s = s$ whenever $s \in R$ and Theorem 3.1 imply $(I_p - AA^\dagger)\beta(P) = 0$ for all $P \in \mathbf{P}_0$. Therefore, $\hat{x}_n^\star = A^\dagger\hat{C}_n\hat{\beta}_n$ and $\hat{C}_n\beta(P) = \beta(P)$ for all $P \in \mathbf{P}_0$ by Assumption 4.2(ii) yield that

$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm e}} \sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}_n^\star\rangle$$
$$= \sup_{s\in\hat{\mathcal{V}}_n^{\rm e}} \langle s, (I_p - AA^\dagger\hat{C}_n)\sqrt{n}\hat{\beta}_n\rangle = \sup_{s\in\hat{\mathcal{V}}_n^{\rm e}} \langle s, (I_p - AA^\dagger\hat{C}_n)\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle \qquad \text{(S.9)}$$

for all $P \in \mathbf{P}_0$. Similarly, employing that $A^\dagger AA^\dagger = A^\dagger$ (see, e.g., Proposition 6.11.1(5) in Luenberger (1969)) together with $A\hat{x}_n^\star = AA^\dagger\hat{C}_n\hat{\beta}_n$ and $\hat{C}_n\beta(P) = \beta(P)$ for all $P \in \mathbf{P}_0$ by Assumption 4.2(ii), allows us to conclude that for all $P \in \mathbf{P}_0$

$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle = \sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} \sqrt{n}\langle A^\dagger s, A^\dagger\hat{C}_n\hat{\beta}_n\rangle$$
$$= \sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} \langle A^\dagger s, A^\dagger AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle. \qquad \text{(S.10)}$$

Moreover, if $P \in \mathbf{P}_0$, then $\sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle \leq 0$ for all s satisfying $A^\dagger s \leq 0$ by Theorem 3.1, $A^\dagger s \in N^\perp \cap \mathbf{R}^d_-$ whenever $A^\dagger s \leq 0$, and $x^\star(P) = A^\dagger\beta(P)$. Hence, $r_n = o(1)$, results (S.7), (S.8), (S.9) and (S.10) and Theorem S.2.1 applied with $\hat{\mathbb{W}}_n^{\rm e}(P) = (I_p - AA^\dagger\hat{C}_n)\sqrt{n}\{\hat{\beta}_n - \beta(P)\}$, $\hat{\mathbb{W}}_n^{\rm i}(P) = AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}$, $\hat{f}_n(s, P) = \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle$, $\mathbf{Q} = \mathbf{P}_0$, and $\omega_n = r_n$ together with $a_n + r_n = O(r_n)$ imply uniformly in $P \in \mathbf{P}_0$ that

$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm e}} \sqrt{n}\langle s, \hat{\beta}_n - A\hat{x}_n^\star\rangle = \sup_{s\in\mathcal{V}^{\rm e}(P)} \langle s, \mathbb{G}_n^{\rm e}(P)\rangle + O_P(r_n) \qquad \text{(S.11)}$$
$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle = \sup_{s\in\mathcal{V}^{\rm i}(P)} \langle A^\dagger s, A^\dagger\mathbb{G}_n^{\rm i}(P)\rangle + \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle + O_P(r_n), \qquad \text{(S.12)}$$

from which the claim of the theorem follows. ■

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

$$\mathbb{M}_n(s,P) \equiv \langle A^\dagger s, A^\dagger \mathbb{G}_n^{\rm i}(P)\rangle \qquad \mathbb{U}_n(s,P) \equiv \sqrt{n}\langle A^\dagger s, A^\dagger \beta(P)\rangle \tag{S.13}$$

$$\mathbb{A}_n^{\rm e}(s,P) \equiv \langle s, (\Omega^{\rm e}(P))^\dagger \mathbb{G}_n^{\rm e}(P)\rangle \qquad \mathbb{A}_n^{\rm i}(s,P) \equiv \langle s, \mathbb{G}_n^{\rm i}(P) + \sqrt{n}\beta(P)\rangle. \tag{S.14}$$

We also 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

$$\sup_{P\in\mathbf{P}} \frac{\mathrm{m}(P) + \bar{\sigma}(P) z_{\tau_n}}{\underline{\sigma}^2(P)} = o(\ell_n^{-1}), \tag{S.15}$$

which is feasible by hypothesis. Also note that since $\eta > 0.5$, there is $\epsilon > 0$ such that $\eta - \epsilon > 0.5$ and for $z_{\eta-\epsilon}$ the $\eta - \epsilon$ quantile of a standard normal random variable, let

$$E_{1n}(P) \equiv \{\hat{c}_n(\eta) \geq (\underline{\sigma}(P) z_{\eta-\epsilon})/2\} \tag{S.16}$$

$$E_{2n}(P) \equiv \{\mathbb{U}_n(s,P) \leq \hat{\mathbb{U}}_n(s) + \ell_n \text{ for all } s \in \hat{\mathcal{V}}_n^{\rm i}\}. \tag{S.17}$$

Next note that $0 \in \hat{\mathcal{V}}_n^{\rm e}$ and $0 \in \hat{\mathcal{V}}_n^{\rm i}$ together yield that $\hat{c}_n(\eta) \geq 0$. Therefore, $\phi_n = 1$ implies $T_n > 0$, which together with Lemma S.2.3 implies that the conclusion of the theorem is immediate on the set $\mathbf{D}_0 \equiv \{P \in \mathbf{P}_0 : \sigma^{\rm j}(s,P) = 0 \text{ for all } s \in \mathcal{E}^{\rm j}(P) \text{ and all } {\rm j} \in \{{\rm e},{\rm i}\}\}$. We therefore assume without loss of generality that for all $P \in \mathbf{P}_0$, $\sigma^{\rm j}(s,P) > 0$ for some $s \in \mathcal{E}^{\rm j}(P)$ and some ${\rm j} \in \{{\rm e},{\rm i}\}$. Next, we also observe that since $\phi_n = 1$ implies $T_n > 0$, Lemma S.2.2 allows us to conclude that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\phi_n = 1) = \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(T_n > \hat{c}_n(\eta);\ E_{1n}(P)). \tag{S.18}$$

Further observe that, for ${\rm j} \in \{{\rm e},{\rm i}\}$, $\mathbb{G}_n^{\rm j}(P) \in \text{range}\{\Sigma^{\rm j}(P)\} \subseteq \text{range}\{\Omega^{\rm j}(P)\}$ almost surely by Theorem 3.6.1 in Bogachev (1998) and Assumption 4.4(i). Hence, it follows that $\Omega^{\rm j}(P)(\Omega^{\rm j}(P))^\dagger \mathbb{G}_n^{\rm j}(P) = \mathbb{G}_n^{\rm j}(P)$ almost surely for ${\rm j} \in \{{\rm e},{\rm i}\}$, which together with Hölder's inequality, Assumption 4.1(ii), the definitions of $\mathcal{V}^{\rm e}(P)$ and $\mathcal{V}^{\rm i}(P)$, and $\mathbb{U}_n(s,P) \leq 0$ for $s \in \mathcal{V}^{\rm i}(P)$ and $P \in \mathbf{P}_0$ by Theorem 3.1 imply that almost surely

$$\sup_{s\in\mathcal{V}^{\rm e}(P)} \langle s, \mathbb{G}_n^{\rm e}(P)\rangle = \sup_{s\in\mathcal{V}^{\rm e}(P)} \langle \Omega^{\rm e}(P)s, (\Omega^{\rm e}(P))^\dagger \mathbb{G}_n^{\rm e}(P)\rangle < \infty$$

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

Thus, by Theorem 4.1 and Lemmas S.2.12, S.2.13 we obtain uniformly in $P \in \mathbf{P}_0$ that

$$T_n = \max_{s\in\mathcal{E}^{\rm e}(P)} \mathbb{A}_n^{\rm e}(s,P) \vee \max_{s\in\mathcal{E}^{\rm i}(P)} \mathbb{A}_n^{\rm i}(s,P) + O_P(r_n). \tag{S.19}$$

For any $\tau \in (0,1)$ and $\mathbb{M}_n(s,P)$ as in (S.13), we next let $c_n^{(1)}(\tau,P)$ denote the τ^{th} quantile

$$c_n^{(1)}(\tau,P) \equiv \inf\{u : P(\sup_{s\in\mathcal{V}^{\mathrm{i}}(P)} \mathbb{M}_n(s,P) \le u) \ge \tau\}. \tag{S.20}$$

Employing $c_n^{(1)}(\tau,P)$ we further define a "truncated" subset $\mathcal{E}^{\mathrm{i},\tau}(P) \subseteq \mathcal{E}^{\mathrm{i}}(P)$ by

$$\mathcal{E}^{\mathrm{i},\tau}(P) \equiv \{s \in \mathcal{E}^{\mathrm{i}}(P) : -\langle s, \sqrt{n}\beta(P)\rangle \le c_n^{(1)}(\tau,P)\}. \tag{S.21}$$

Next note that $0 \in \mathcal{V}^{\mathrm{i}}(P)$ satisfying $\mathbb{M}_n(0,P) = 0$ implies $\sup_{s\in\mathcal{V}^{\mathrm{i}}(P)} \mathbb{M}_n(s,P)$ is nonnegative almost surely and therefore $c_n^{(1)}(\tau,P) \ge 0$. Since in addition $0 \in \mathcal{E}^{\mathrm{i}}(P)$ by Lemma S.2.13, it follows $0 \in \mathcal{E}^{\mathrm{i},\tau}(P)$ and therefore we obtain that

$$\begin{aligned} P(\max_{s\in\mathcal{E}^{\mathrm{i}}(P)} \mathbb{A}_n^{\mathrm{i}}(s,P) = \max_{s\in\mathcal{E}^{\mathrm{i},\tau}(P)} \mathbb{A}_n^{\mathrm{i}}(s,P)) \\ \ge P(\max_{s\in\mathcal{E}^{\mathrm{i}}(P)\setminus\mathcal{E}^{\mathrm{i},\tau}(P)} \mathbb{A}_n^{\mathrm{i}}(s,P) \le 0) \ge P(\sup_{s\in\mathcal{V}^{\mathrm{i}}(P)} \mathbb{M}_n(s,P) \le c_n^{(1)}(\tau,P)) \ge \tau, \end{aligned}$$

where the second and final inequalities hold by definitions (S.13) and (S.20), and $\mathcal{E}^{\mathrm{i}}(P) \subseteq (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$. Next define the sets $\mathcal{C}_n(\mathrm{j},P)$ according to the relation

$$\mathcal{C}_n(\mathrm{j},P) \equiv \begin{cases} \mathcal{E}^{\mathrm{e}}(P) & \text{if } \mathrm{j} = \mathrm{e} \\ \mathcal{E}^{\mathrm{i},\tau_n}(P) & \text{if } \mathrm{j} = \mathrm{i} \end{cases}.$$

Given these definitions, we then obtain from results (S.15), (S.18), and (S.19) that

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\phi_n = 1) \\ &\le \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\max_{s\in\mathcal{E}^{\mathrm{e}}(P)} \mathbb{A}_n^{\mathrm{e}}(s,P) \vee \max_{s\in\mathcal{E}^{\mathrm{i},\tau_n}(P)} \mathbb{A}_n^{\mathrm{i}}(s,P) > \hat{c}_n(\eta) - \ell_n;\ E_{1n}(P)) \\ &= \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\max_{\mathrm{j}\in\{\mathrm{e},\mathrm{i}\}} \max_{s\in\mathcal{C}_n(\mathrm{j},P)} \mathbb{A}_n^{\mathrm{j}}(s,P) > \hat{c}_n(\eta) - \ell_n;\ E_{1n}(P)) \end{aligned} \tag{S.22}$$

due to $\tau_n \uparrow 1$ and $r_n = o(\ell_n)$ by construction. Further define the set $\mathcal{A}_n(P)$ to equal

$$\mathcal{A}_n(P) \equiv \{(\mathrm{j},s) : \mathrm{j} \in \{\mathrm{e},\mathrm{i}\},\ s \in \mathcal{C}_n(\mathrm{j},P),\ \sigma^{\mathrm{j}}(s,P) > 0\},$$

and note that, for n sufficiently large, $\inf_{P\in\mathbf{P}}(\underline{\sigma}(P)z_{\eta-\epsilon}) - 2\ell_n > 0$ by requirement (S.15), in which case the event $E_{1n}(P)$ implies $\hat{c}_n(\eta) - \ell_n > 0$. Hence, since for all $P \in \mathbf{P}_0$ we have $E[\mathbb{A}_n^{\mathrm{e}}(s,P)] = 0$ for all $s \in \mathcal{E}^{\mathrm{e}}(P)$ and $E[\mathbb{A}_n^{\mathrm{i}}(s,P)] \le 0$ for all $s \in \mathcal{E}^{\mathrm{i},\tau_n}(P)$ due to $\langle (AA')^{\dagger}s, \beta(P)\rangle \le 0$ for all $s \in \mathcal{V}^{\mathrm{i}}(P)$ by Theorem 3.1, we can conclude from result (S.22) that the claim of the theorem is immediate if $\mathcal{A}_n(P) = \emptyset$. Therefore, assuming without loss of generality that $\mathcal{A}_n(P) \neq \emptyset$ we obtain from the same observations that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\phi_n = 1) \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) > \hat{c}_n(\eta) - \ell_n)$$
$$= \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) > \hat{c}_n(\eta) - \ell_n;\ E_{2n}(P)), \quad \text{(S.23)}$$

where the final inequality holds for $E_{2n}(P)$ as defined in (S.17) by Lemma S.2.1.

For any $P \in \mathbf{P}_0$, it follows that under $E_{2n}(P)$, $\hat{c}_n(\eta)$ is P-almost surely bounded from below by the conditional on $\{Z_i\}_{i=1}^n$ η quantile of the random variable

$$\max\{\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle, \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}}\rangle + \mathbb{U}_n(s,P)\} - \ell_n. \quad \text{(S.24)}$$

Moreover, by Theorem S.2.5 there is a Gaussian vector $(\mathbb{G}_n^{\mathrm{e}\star}(P)', \mathbb{G}_n^{\mathrm{i}\star}(P)')' \equiv \mathbb{G}_n^\star(P)$ with $\mathbb{G}_n^\star(P) \sim N(0, \Sigma(P))$, independent of $\{Z_i\}_{i=1}^n$, and satisfying

$$\|(\Omega^{\mathrm{e}}(P))^\dagger\{\hat{\mathbb{G}}_n^{\mathrm{e}} - \mathbb{G}_n^{\mathrm{e}\star}(P)\}\|_\infty \vee \|(\Omega^{\mathrm{i}}(P))^\dagger\{\hat{\mathbb{G}}_n^{\mathrm{i}} - \mathbb{G}_n^{\mathrm{i}\star}(P)\}\|_\infty = O_P(b_n)$$

uniformly in $P \in \mathbf{P}$. Since $r_n = o(1)$ implies $a_n = o(1)$, we may then apply Theorem S.2.1 with $\hat{\mathbb{W}}_n = \hat{\mathbb{G}}_n$, $\mathbb{W}_n(P) = \mathbb{G}_n^\star(P)$, and $\hat{f}_n(s,P) = \mathbb{U}_n(s,P)$ to conclude that

$$\max\{\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle, \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}}\rangle + \mathbb{U}_n(s,P)\}$$
$$= \max\{\sup_{s\in\mathcal{V}^{\mathrm{e}}(P)} \langle s, \mathbb{G}_n^{\mathrm{e}\star}(P)\rangle, \sup_{s\in\mathcal{V}^{\mathrm{i}}(P)} \langle A^\dagger s, A^\dagger \mathbb{G}_n^{\mathrm{i}\star}(P)\rangle + \mathbb{U}_n(s,P)\} + O_P(b_n)$$
$$= \max\{\max_{s\in\mathcal{E}^{\mathrm{e}}(P)} \langle s, (\Omega^{\mathrm{e}}(P))^\dagger \mathbb{G}_n^{\mathrm{e}\star}(P)\rangle, \max_{s\in\mathcal{E}^{\mathrm{i}}(P)} \langle s, \mathbb{G}_n^{\mathrm{i}\star}(P) + \sqrt{n}\beta(P)\rangle\} + O_P(b_n) \quad \text{(S.25)}$$

uniformly in $P \in \mathbf{P}_0$, and where the second equality follows by arguing as in (S.19). Therefore, defining $c_n^{(2)}(\eta, P)$ to be the following η quantile

$$c_n^{(2)}(\eta, P) \equiv \inf\{u : P(\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) \leq u) \geq \eta\},$$

we obtain from $E_{2n}(P)$ implying that $\hat{c}_n(\eta)$ is P-almost surely bounded from below by the conditional on $\{Z_i\}_{i=1}^n$ η quantile of (S.24) for any $P \in \mathbf{P}_0$, results (S.23) and (S.25), $\mathbb{G}_n(P)$ 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

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\phi_n = 1) \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) > c_n^{(2)}(\eta_n, P) - 3\ell_n) \quad \text{(S.26)}$$

for some sequence η_n satisfying $\eta_n \uparrow \eta$.

To conclude, for any $(\mathrm{j}, s) \in \mathcal{A}_n(P)$ we define the random variable $\mathbb{N}((\mathrm{j}, s), P)$ by

$$\mathbb{N}((\mathrm{j}, s), P) \equiv \frac{\mathbb{A}_n^{\mathrm{j}}(s, P) - c_n^{(2)}(\eta_n, P)}{\sigma^{\mathrm{j}}(s, P)} + \frac{c_n^{(1)}(\tau_n, P) + 0 \vee c_n^{(2)}(\eta_n, P)}{\underline{\sigma}(P)}.$$

Then note that $E[\mathbb{N}((\mathrm{j}, s), P)] \geq 0$ for any $(\mathrm{j}, s) \in \mathcal{A}_n(P)$, by definition of $\mathcal{E}^{\mathrm{i},\tau_n}(P)$, $c_n^{(1)}(\eta_n, P) \geq 0$, and $\sigma^{\mathrm{j}}(s, P) \geq \underline{\sigma}(P)$ for all $(\mathrm{j}, s) \in \mathcal{A}_n(P)$. Thus, since in addition $\mathrm{Var}\{\mathbb{N}((\mathrm{j}, s), P)\} = 1$ for any $(\mathrm{j}, s) \in \mathcal{A}_n(P)$ and $\mathcal{A}_n(P)$ is finite due to $\mathcal{E}^{\mathrm{e}}(P)$ and $\mathcal{E}^{\mathrm{i}}(P)$ being finite by Corollary 19.1.1 in Rockafellar (1970), Lemma S.2.11 implies

$$\begin{aligned}
P(|\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s, P) - c_n^{(2)}(\eta_n, P)| \leq 3\ell_n) \\
&\leq P(|\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{N}((\mathrm{j}, s), P) - \frac{c_n^{(1)}(\tau_n, P) + 0 \vee c_n^{(2)}(\eta_n, P)}{\underline{\sigma}(P)}| \leq \frac{3\ell_n}{\underline{\sigma}(P)}) \\
&\leq \frac{12\ell_n}{\underline{\sigma}(P)} \max\{\mathrm{med}\{\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{N}((\mathrm{j}, s), P)\}, 1\} \qquad \text{(S.27)}
\end{aligned}$$

for any $P \in \mathbf{P}_0$. Next note the definition of $\mathbb{N}((\mathrm{j}, s), P)$, $\Omega^{\mathrm{j}}(P)(\Omega^{\mathrm{j}}(P))^\dagger \mathbb{G}_n^{\mathrm{j}}(P) = \mathbb{G}_n^{\mathrm{j}}(P)$ for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$, $\mathcal{E}^{\mathrm{e}}(P) \subset \Omega^{\mathrm{e}}(P)\mathcal{V}^{\mathrm{e}}(P)$, and $\mathcal{E}^{\mathrm{i},\tau_n}(P) \subseteq (AA')^\dagger \mathcal{V}^{\mathrm{i}}(P)$ imply that

$$\begin{aligned}
&\mathrm{med}\{\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{N}((\mathrm{j}, s), P)\} \\
&\leq \frac{1}{\underline{\sigma}(P)}\{\mathrm{med}\{\sup_{s\in\mathcal{V}^{\mathrm{e}}(P)} \langle s, \mathbb{G}_n^{\mathrm{e}}(P)\rangle \vee \sup_{s\in\mathcal{V}^{\mathrm{i}}(P)} \langle A^\dagger s, A^\dagger \mathbb{G}_n^{\mathrm{i}}(P)\rangle\} + c_n^{(1)}(\tau_n, P) + |c_n^{(2)}(\eta_n, P)|\} \\
&= \frac{\mathrm{m(P)}}{\underline{\sigma}(P)} + \frac{c_n^{(1)}(\tau_n, P) + |c_n^{(2)}(\eta_n, P)|}{\underline{\sigma}(P)} \qquad \text{(S.28)}
\end{aligned}$$

for all $P \in \mathbf{P}_0$ and n. Furthermore, by Borell's inequality (see, for example, the corollary in pg. 82 of Davydov et al. (1998)) we also have the bound

$$c_n^{(1)}(\tau_n, P) \leq \mathrm{m}(P) + z_{\tau_n} \bar{\sigma}(P) \qquad \text{(S.29)}$$

for all $P \in \mathbf{P}$ and n sufficiently large due to $\tau_n \uparrow 1$. Since $P \in \mathbf{P}_0$ implies $\langle s, \beta(P)\rangle \leq 0$ for any $s \in \mathcal{E}^{\mathrm{i},\tau_n}(P) \subset (AA')^\dagger \mathcal{V}^{\mathrm{i}}(P)$ by Theorem 3.1, Borell's inequality yields

$$c_n^{(2)}(\eta_n, P) \leq \mathrm{m}(P) + \bar{\sigma}(P) z_{\eta_n} \qquad \text{(S.30)}$$

for n sufficiently large by $\eta_n \uparrow \eta > 1/2$ and definition of $\mathrm{m}(P)$. Also, $\eta_n > 1/2$ for n sufficiently large and $0 \geq \langle s, \sqrt{n}\beta(P)\rangle \geq -c_n^{(1)}(\tau_n, P)$ for all $s \in \mathcal{E}^{\mathrm{i},\tau_n}(P)$ by (S.21) imply

$$\begin{aligned}
c_n^{(2)}(\eta_n, P) &\geq \mathrm{med}\{\max_{s\in\mathcal{E}^{\mathrm{e}}(P):\sigma^{\mathrm{e}}(s,P)>0} \mathbb{A}_n^{\mathrm{e}}(s, P) \vee \max_{s\in\mathcal{E}^{\mathrm{i},\tau_n}(P):\sigma^{\mathrm{i}}(s,P)>0} \langle s, \mathbb{G}_n^{\mathrm{i}}(P)\rangle\} - c_n^{(1)}(\tau_n, P) \\
&\geq -c_n^{(1)}(\tau_n, P), \qquad \text{(S.31)}
\end{aligned}$$

where in the last inequality we employed that $E[\mathbb{A}_n^{\mathrm{e}}(s, P)] = 0$ for all $s \in \mathcal{E}^{\mathrm{e}}(P)$ and

$E[\langle s, \mathbb{G}_n^{\mathrm{i}}(P)\rangle] = 0$ for all $s \in \mathcal{E}^{\mathrm{i}}(P)$ imply $\mathrm{med}\{\mathbb{A}_n^{\mathrm{e}}(s,P)\} \geq 0$ for any $s \in \mathcal{E}^{\mathrm{e}}(P)$ and $\mathrm{med}\{\langle s, \mathbb{G}_n^{\mathrm{i}}(P)\rangle\} \geq 0$ for any $s \in \mathcal{E}^{\mathrm{i}}(P)$. Therefore, results (S.27), (S.28), (S.29), (S.30), (S.31), $\tau_n \uparrow 1$ implying $z_{\tau_n} \uparrow \infty$, and ℓ_n satisfying restriction (S.15) finally yield that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(|\max_{(\mathrm{j},s)\in\mathcal{A}_n(P)} \mathbb{A}_n^{\mathrm{j}}(s,P) - c_n^{(2)}(\eta_n, P)| \leq 3\ell_n) \lesssim \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} \frac{\ell_n(\mathrm{m}(P) + z_{\tau_n}\bar{\sigma}(P))}{\underline{\sigma}^2(P)} = 0. \quad \text{(S.32)}$$

Thus, (S.26) and (S.32) together with the definition of $c_n^{(2)}(\eta_n, P)$ and $\eta_n \uparrow \eta$ imply

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

Since $\eta = 1 - \alpha$, the claim of the theorem therefore follows. ■

Lemma S.2.1. *Let Assumptions 4.1, 4.2, 4.3, 4.4(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*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \{\sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle - \hat{\mathbb{U}}_n(s)\} \leq \ell_n) = 1.$$

Proof: First note that Theorem 3.1 implies that $\langle A^\dagger s, A^\dagger\beta(P)\rangle \leq 0$ for all $s \in \hat{\mathcal{V}}_n^{\mathrm{i}}$ and $P \in \mathbf{P}_0$. Therefore, the definition of $\hat{\mathbb{U}}_n(s)$ and $\lambda_n \in [0,1]$ allow us to conclude that

$$\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle - \hat{\mathbb{U}}_n(s) \leq \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \lambda_n\sqrt{n}\langle A^\dagger s, A^\dagger\{\beta(P) - \hat{\beta}_n^{\mathrm{r}}\}\rangle$$
$$\leq \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \lambda_n\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star - A^\dagger\hat{\beta}_n^{\mathrm{r}}\rangle + \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \lambda_n\sqrt{n}\langle A^\dagger s, A^\dagger\beta(P) - \hat{x}_n^\star\rangle. \quad \text{(S.33)}$$

Moreover, the definition of $\hat{\beta}_n^{\mathrm{r}}$ in (24), $\hat{x}_n^\star \equiv A^\dagger\hat{C}_n\hat{\beta}_n$ with $\hat{C}_n\beta(P) = \beta(P)$ for any $P \in \mathbf{P}_0$ by Assumption 4.2(ii), $\beta(P) \in R$ for any $P \in \mathbf{P}_0$, and (S.33) yield

$$\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, A^\dagger\beta(P)\rangle - \hat{\mathbb{U}}_n(s) \leq \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} 2\lambda_n|\langle A^\dagger s, \sqrt{n}\{\hat{x}_n^\star - A^\dagger\beta(P)\}\rangle|$$
$$= \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} 2\lambda_n|\langle A^\dagger s, A^\dagger AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle|. \quad \text{(S.34)}$$

By applying Theorem S.2.1 twice, once with $\hat{\mathbb{W}}_n^{\mathrm{i}}(P) = AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}$ and $\hat{\mathbb{W}}_n^{\mathrm{e}}(P) = \mathbb{G}_n^{\mathrm{e}}(P)$, and once with $\hat{\mathbb{W}}_n^{\mathrm{i}}(P) = AA^\dagger\hat{C}_n\sqrt{n}\{\beta(P) - \hat{\beta}_n\}$ and $\hat{\mathbb{W}}_n^{\mathrm{e}}(P) = -\mathbb{G}_n^{\mathrm{e}}(P)$, and in both cases setting $\hat{f}_n(s,P) = 0$ for all $s \in \mathbf{R}^p$, we obtain from Lemma

S.2.4 and $(-\mathbb{G}_n^{\rm e}(P)', -\mathbb{G}_n^{\rm i}(P)')' \sim N(0, \Sigma(P))$ that uniformly in $P \in \mathbf{P}_0$ we have

$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} \langle A^\dagger s, A^\dagger A A^\dagger \hat{C}_n \sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle = \sup_{s\in\mathcal{V}^{\rm i}(P)} \langle A^\dagger s, A^\dagger \mathbb{G}_n^{\rm i}(P)\rangle + O_P(r_n)$$
$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} \langle A^\dagger s, A^\dagger A A^\dagger \hat{C}_n \sqrt{n}\{\beta(P) - \hat{\beta}_n\}\rangle = \sup_{s\in\mathcal{V}^{\rm i}(P)} \langle A^\dagger s, A^\dagger(-\mathbb{G}_n^{\rm i}(P))\rangle + O_P(r_n). \quad \text{(S.35)}$$

Thus, since $\Omega^{\rm i}(P)(\Omega^{\rm i}(P))^\dagger \mathbb{G}_n^{\rm i}(P) = \mathbb{G}_n^{\rm i}(P)$ almost surely due to $\mathbb{G}_n^{\rm i}(P) \in \text{range}\{\Sigma^{\rm i}(P)\} \subseteq \text{range}\{\Omega^{\rm i}(P)\}$ almost surely by Theorem 3.6.1 in Bogachev (1998) and Assumption 4.4(i), we obtain from results (S.34), (S.35), and Hölder's inequality that

$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} 2\lambda_n|\langle A^\dagger s, A^\dagger A A^\dagger \hat{C}_n \sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle| = \sup_{s\in\mathcal{V}^{\rm i}(P)} 2\lambda_n|\langle A^\dagger s, A^\dagger \mathbb{G}_n^{\rm i}(P)\rangle| + O_P(\lambda_n r_n)$$
$$\leq 2\lambda_n\|(\Omega^{\rm i}(P))^\dagger \mathbb{G}_n^{\rm i}(P)\|_\infty + O_P(\lambda_n r_n) = O_P(\lambda_n\sqrt{\log(1+p)}) \quad \text{(S.36)}$$

uniformly in $P \in \mathbf{P}_0$, and where the final equality follows from $r_n = o(1)$, Markov's inequality, and $\sup_{P\in\mathbf{P}} E_P[\|(\Omega^{\rm i}(P))^\dagger \mathbb{G}_n^{\rm i}(P)\|_\infty] \lesssim \sqrt{\log(1+p)}$ by Lemma S.2.8 and Assumption 4.3(ii). The claim of the Lemma then follows from results (S.34), (S.36), and $\lambda_n\sqrt{\log(1+p)} = o(\ell_n)$ by hypothesis. ■

Lemma S.2.2. *Let Assumptions 4.1, 4.2(i)(ii), 4.3, 4.4, 4.5 hold, $\eta \in (0.5, 1)$, $0 < \epsilon < \eta - 0.5$, and z_η be the η quantile of $N(0,1)$. If $r_n \vee b_n = o(1)$ and $\sup_{P\in\mathbf{P}}(\mathrm{m}(P) + \bar{\sigma}(P))/\underline{\sigma}^2(P) = o(r_n^{-1} \wedge b_n^{-1})$, then for each $P \in \mathbf{P}_0$ there are $\{E_n(P)\}$ with*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\{Z_i\}_{i=1}^n \in E_n(P)) = 1 \quad \text{(S.37)}$$

and on $E_n(P)$ it holds that $\hat{c}_n(\eta) \geq (\underline{\sigma}(P) z_{\eta-\epsilon})/2$ whenever $T_n > 0$.

Proof: First note that by Lemma S.2.5 there is a Gaussian vector $(\mathbb{G}_n^{\rm e\star}(P)', \mathbb{G}_n^{\rm i\star}(P)')' \equiv \mathbb{G}_n^\star(P) \sim N(0, \Sigma(P))$ that is independent of $\{Z_i\}_{i=1}^n$ and satisfies

$$\|(\Omega^{\rm e}(P))^\dagger\{\hat{\mathbb{G}}_n^{\rm e} - \mathbb{G}_n^{\rm e\star}(P)\}\|_\infty \vee \|(\Omega^{\rm i}(P))^\dagger\{\hat{\mathbb{G}}_n^{\rm i} - \mathbb{G}_n^{\rm i\star}(P)\}\|_\infty = O_P(b_n)$$

uniformly in $P \in \mathbf{P}$. Further define $\hat{\mathbb{L}}_n \in \mathbf{R}$ and $\mathbb{L}_n^\star(P) \in \mathbf{R}$ to be given by

$$\hat{\mathbb{L}}_n \equiv \max\{\sup_{s\in\hat{\mathcal{V}}_n^{\rm e}} \langle s, \hat{\mathbb{G}}_n^{\rm e}\rangle, \sup_{s\in\hat{\mathcal{V}}_n^{\rm i}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\rm i}\rangle + \hat{\mathbb{U}}_n(s)\} \quad \text{(S.38)}$$
$$\mathbb{L}_n^\star(P) \equiv \max\{\sup_{s\in\mathcal{V}^{\rm e}(P)} \langle s, \mathbb{G}_n^{\rm e\star}(P)\rangle, \sup_{s\in\mathcal{V}^{\rm i}(P)} \langle A^\dagger s, A^\dagger \mathbb{G}_n^{\rm i\star}(P)\rangle + \hat{\mathbb{U}}_n(s)\}, \quad \text{(S.39)}$$

and note that since $\langle A^\dagger s, A^\dagger \hat{\beta}_n^{\rm r}\rangle \leq 0$ for all $s \in \mathbf{R}^p$ such that $A^\dagger s \leq 0$ by Theorem 3.1, it follows from Lemma S.2.5, Assumptions 4.4(i) and 4.5(v), and Theorem S.2.1 applied with $\mathbb{W}_n(P) = \mathbb{G}_n^\star(P)$, $\hat{\mathbb{W}}_n = \hat{\mathbb{G}}_n$, and $\hat{f}_n(\cdot, P) = \hat{\mathbb{U}}_n(\cdot)$ that uniformly in $P \in \mathbf{P}$

$$\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}\star} \rangle = \sup_{s \in \mathcal{V}^{\mathrm{e}}(P)} \langle s, \mathbb{G}_n^{\mathrm{e}\star}(P) \rangle + O_P(b_n) \tag{S.40}$$

$$\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{G}}_n^{\mathrm{i}\star} \rangle + \hat{\mathbb{U}}_n(s) = \sup_{s \in \mathcal{V}^{\mathrm{i}}(P)} \langle A^\dagger s, A^\dagger \mathbb{G}_n^{\mathrm{i}\star}(P) \rangle + \hat{\mathbb{U}}_n(s) + O_P(b_n). \tag{S.41}$$

We establish the lemma by studying three separate cases.

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

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

and note that $z_{\eta-\epsilon} > 0$ due to $\eta - \epsilon > 0.5$, and therefore result (S.40), Markov's inequality, and $b_n \times \sup_{P \in \mathbf{P}} 1/\underline{\sigma}(P) = o(1)$ by hypothesis, imply that

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}_0^{\mathrm{e}}} P(\{Z_i\}_{i=1}^n \in E_n(P)) = 1.$$

Then note that whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ the triangle inequality allows us to conclude

$$\begin{aligned} &P(\sup_{s \in \mathcal{V}^{\mathrm{e}}(P)} \langle s, \mathbb{G}_n^{\mathrm{e}\star}(P) \rangle \leq \hat{c}_n(\eta) + \frac{\underline{\sigma}(P) z_{\eta-\epsilon}}{2} | \{Z_i\}_{i=1}^n) \\ &\geq P(\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}\star} \rangle \leq \hat{c}_n(\eta) | \{Z_i\}_{i=1}^n) - \epsilon \geq P(\hat{\mathbb{L}}_n \leq \hat{c}_n(\eta) | \{Z_i\}_{i=1}^n) - \epsilon \geq \eta - \epsilon \end{aligned} \tag{S.42}$$

where the second inequality follows from (S.38), while the final inequality holds by definition of $\hat{c}_n(\eta)$. Also note that $\mathbb{G}_n^{\mathrm{e}\star}(P) \sim N(0, \Sigma^{\mathrm{e}}(P))$, Theorem 3.6.1 in Bogachev (1998), and Assumption 4.4(i) imply $\mathbb{G}_n^{\mathrm{e}\star}(P) = \Omega^{\mathrm{e}}(P)(\Omega^{\mathrm{e}}(P))^\dagger \mathbb{G}_n^{\mathrm{e}\star}(P)$ almost surely. Therefore, by symmetry of $\Omega^{\mathrm{e}}(P)$ we can conclude that almost surely

$$\begin{aligned} \sup_{s \in \mathcal{V}^{\mathrm{e}}(P)} \langle s, \mathbb{G}_n^{\mathrm{e}\star}(P) \rangle &= \sup_{s \in \mathcal{V}^{\mathrm{e}}(P)} \langle \Omega^{\mathrm{e}}(P) s, (\Omega^{\mathrm{e}}(P))^\dagger \mathbb{G}_n^{\mathrm{e}\star}(P) \rangle \\ &= \max_{s \in \mathcal{E}^{\mathrm{e}}(P)} \langle s, (\Omega^{\mathrm{e}}(P))^\dagger \mathbb{G}_n^{\mathrm{e}\star}(P) \rangle, \end{aligned}$$

where the second equality holds by Lemma S.2.12 and the supremum being finite by Hölder's inequality. Hence, the (unconditional) distribution of $\sup_{s \in \mathcal{V}^{\mathrm{e}}(P)} \langle s, \mathbb{G}_n^{\mathrm{e}\star}(P) \rangle$ first order stochastically dominates the distribution $N(0, \underline{\sigma}^2(P))$ whenever $P \in \mathbf{P}_0^{\mathrm{e}}$ by definition of $\underline{\sigma}(P)$. In particular, $\mathbb{G}_n^{\mathrm{e}\star}(P)$ being independent of $\{Z_i\}_{i=1}^n$ and result (S.42) imply that whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ and $P \in \mathbf{P}_0^{\mathrm{e}}$ we must have

$$\hat{c}_n(\eta) + \frac{\underline{\sigma}(P) z_{\eta-\epsilon}}{2} \geq \underline{\sigma}(P) z_{\eta-\epsilon},$$

which establishes the claim of the lemma for the subset $\mathbf{P}_0^{\mathrm{e}} \subseteq \mathbf{P}_0$.

Case II: Suppose $P \in \mathbf{P}_0^{\mathrm{i}} \equiv \{P \in \mathbf{P}_0 : \sigma^{\mathrm{i}}(s,P) > 0$ for some $s \in \mathcal{E}^{\mathrm{i}}(P)$ and $\sigma^{\mathrm{e}}(s,P) = 0$ for all $s \in \mathcal{E}^{\mathrm{e}}(P)\}$, and define the event $E_n(P) \equiv \bigcap_{j=1}^4 E_{j,n}(P)$, where

$$\begin{aligned}
E_{1n}(P) &\equiv \{\hat{\mathcal{V}}_n^{\mathrm{i}} \subseteq 2\mathcal{V}^{\mathrm{i}}(P)\} \\
E_{2n}(P) &\equiv \{AA^\dagger \hat{C}_n\{\hat{\beta}_n - \beta(P)\} \in \mathrm{range}\{\Sigma^{\mathrm{i}}(P)\}\} \\
E_{3n}(P) &\equiv \{P(|\hat{\mathbb{L}}_n - \mathbb{L}_n^\star(P)| > (\underline{\sigma}(P) z_{\eta-\epsilon})/2 | \{Z_i\}_{i=1}^n) \leq \epsilon\} \\
E_{4n}(P) &\equiv \{T_n = \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, \hat{x}_n^\star \rangle\}.
\end{aligned}$$

Next note that $\Omega^{\mathrm{i}}(P)(\hat{\Omega}_n^{\mathrm{i}})^\dagger \hat{\Omega}_n^{\mathrm{i}} = \Omega^{\mathrm{i}}(P)$ with probability tending to one uniformly in $P \in \mathbf{P}$ by Assumption 4.1(iii), Lemma S.2.10, and symmetry of $\hat{\Omega}_n^{\mathrm{i}}$ and $\Omega^{\mathrm{i}}(P)$. Since $\hat{\Omega}_n^{\mathrm{i}}(\hat{\Omega}_n^{\mathrm{i}})^\dagger \hat{\Omega}_n^{\mathrm{i}} = \hat{\Omega}_n^{\mathrm{i}}$ by Proposition 6.11.1(6) in Luenberger (1969), we obtain from the definition of $\hat{\mathcal{V}}_n^{\mathrm{i}}$ that with probability tending to one uniformly in $P \in \mathbf{P}$

$$\begin{aligned}
\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \|\Omega^{\mathrm{i}}(P)(AA')^\dagger s\|_1 &\leq 1 + \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \|(\hat{\Omega}_n^{\mathrm{i}} - \Omega^{\mathrm{i}}(P))(\hat{\Omega}_n^{\mathrm{i}})^\dagger \hat{\Omega}_n^{\mathrm{i}}(AA')^\dagger s\|_1 \\
&\leq 1 + \|(\hat{\Omega}_n^{\mathrm{i}} - \Omega^{\mathrm{i}}(P))(\hat{\Omega}_n^{\mathrm{i}})^\dagger\|_{o,1} = 1 + \|(\hat{\Omega}_n^{\mathrm{i}})^\dagger(\hat{\Omega}_n^{\mathrm{i}} - \Omega^{\mathrm{i}}(P))\|_{o,\infty}, \quad \text{(S.43)}
\end{aligned}$$

where the final equality follows from Assumptions 4.1(i)(ii) and Theorem 6.5.1 in Luenberger (1969). Therefore, (S.43) and Lemma S.2.6 (for $E_{1n}(P)$), Assumption 4.4(ii) (for $E_{2n}(P)$), results (S.40) and (S.41) together with $\eta - \epsilon > 0.5$, Markov's inequality and $b_n \times \sup_{P \in \mathbf{P}} 1/\underline{\sigma}(P) = o(1)$ (for $E_{3n}(P)$), and Lemma S.2.3 (for $E_{4n}(P)$), yield

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}_0^{\mathrm{i}}} P(\{Z_i\}_{i=1}^n \in E_n(P)) = 1.$$

Next note that if $\{Z_i\}_{i=1}^n \in E_n(P)$ then the event $E_{1n}(P)$ allows us to conclude

$$T_n = \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star \rangle \leq \sup_{s \in \mathcal{V}^{\mathrm{i}}(P)} 2\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star \rangle. \quad \text{(S.44)}$$

Furthermore, since $A^\dagger AA^\dagger = A^\dagger$ by Proposition 6.11.1(5) in Luenberger (1969), Assumption 4.2(ii), $AA^\dagger \beta(P) = \beta(P)$ whenever $P \in \mathbf{P}_0$ due to $\beta(P) \in R$ by Theorem 3.1, symmetry of $\Omega^{\mathrm{i}}(P)$, and $AA^\dagger \hat{C}_n \sqrt{n}\{\hat{\beta}_n - \beta(P)\} \in \mathrm{range}\{\Omega^{\mathrm{i}}(P)\}$ whenever $\{Z_i\}_{i=1}^n \in E_{2n}(P)$ due to $\mathrm{range}\{\Sigma^{\mathrm{i}}(P)\} \subseteq \mathrm{range}\{\Omega^{\mathrm{i}}(P)\}$ by Assumption 4.4(i) imply

$$\begin{aligned}
\sqrt{n}\langle A^\dagger s, \hat{x}_n^\star \rangle &= \langle A^\dagger s, A^\dagger AA^\dagger \hat{C}_n \sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle A^\dagger s, A^\dagger \beta(P)\rangle \\
&= \langle \Omega^{\mathrm{i}}(P)(AA')^\dagger s, (\Omega^{\mathrm{i}}(P))^\dagger AA^\dagger \hat{C}_n \sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle (AA')^\dagger s, \beta(P)\rangle \quad \text{(S.45)}
\end{aligned}$$

for any $s \in \mathcal{V}^{\mathrm{i}}(P)$ whenever $\{Z_i\}_{i=1}^n \in E_n(P)$. Further note that since $\langle A^\dagger s, A^\dagger \beta(P)\rangle \leq 0$ whenever $P \in \mathbf{P}_0$ and $s \in \mathcal{V}^{\mathrm{i}}(P)$ by Theorem 3.1, Hölder's inequality implying (S.45) is bounded in $s \in \mathcal{V}^{\mathrm{i}}(P)$ together with Lemmas S.2.12 and S.2.13 implies that

$$\sup_{s \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)} \langle \Omega^{\mathrm{i}}(P)s, (\Omega^{\mathrm{i}}(P))^{\dagger} AA^{\dagger}\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle s, \beta(P)\rangle$$
$$= \max_{s \in \mathcal{E}^{\mathrm{i}}(P)} \langle \Omega^{\mathrm{i}}(P)s, (\Omega^{\mathrm{i}}(P))^{\dagger} AA^{\dagger}\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle s, \beta(P)\rangle. \quad \text{(S.46)}$$

Hence, results (S.44), (S.45), and (S.46) together establish that the set $\mathcal{S}^{\mathrm{i}}(P)$ given by

$$\mathcal{S}^{\mathrm{i}}(P) \equiv \{s \in \mathcal{E}^{\mathrm{i}}(P) : \langle \Omega^{\mathrm{i}}(P)s, (\Omega^{\mathrm{i}}(P))^{\dagger} AA^{\dagger}\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\} + \sqrt{n}\langle s, \beta(P)\rangle > 0\}$$

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 $\sqrt{n}\langle s, \beta(P)\rangle \leq 0$ for all $s \in \mathcal{S}^{\mathrm{i}}(P)$ due to $\mathcal{S}^{\mathrm{i}}(P) \subseteq \mathcal{E}^{\mathrm{i}}(P) \subset (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$, $P \in \mathbf{P}_0$, and Theorem 3.1, it follows that whenever $\mathcal{S}^{\mathrm{i}}(P) \neq \emptyset$ we must have

$$\langle \Omega^{\mathrm{i}}(P)s, (\Omega^{\mathrm{i}}(P))^{\dagger} AA^{\dagger}\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle > 0 \quad \text{(S.47)}$$

for all $s \in \mathcal{S}^{\mathrm{i}}(P)$. Also note that if $\{Z_i\}_{i=1}^n \in E_n(P) \subseteq E_{2n}(P)$, then range$\{\Sigma^{\mathrm{i}}(P)\}$ equaling the support of $\mathbb{G}_n^{\mathrm{i}}(P)$ by Theorem 3.6.1 in Bogachev (1998) implies that $\sigma^{\mathrm{i}}(s, P) > 0$ for any s satisfying (S.47). Thus, we have so far shown that if $P \in \mathbf{P}_0^{\mathrm{i}}$, then

$$\mathcal{S}^{\mathrm{i}}(P) \neq \emptyset \text{ and } \sigma^{\mathrm{i}}(s, P) > 0 \text{ for all } s \in \mathcal{S}^{\mathrm{i}}(P) \quad \text{(S.48)}$$

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

$$\max_{s \in \mathcal{S}^{\mathrm{i}}(P)} \langle s, AA^{\dagger}\hat{\beta}_n^{\mathrm{r}}\rangle = 0 \quad \text{(S.49)}$$

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

$$0 \geq \sup_{s \in \mathcal{V}^{\mathrm{i}}(P)} \langle A^{\dagger}s, A^{\dagger}\hat{\beta}_n^{\mathrm{r}}\rangle = \sup_{s \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)} \langle s, AA^{\dagger}\hat{\beta}_n^{\mathrm{r}}\rangle = \max_{s \in \mathcal{E}^{\mathrm{i}}(P)} \langle s, AA^{\dagger}\hat{\beta}_n^{\mathrm{r}}\rangle, \quad \text{(S.50)}$$

where the first equality follows from $A^{\dagger}AA^{\dagger} = A^{\dagger}$ by Proposition 6.11.1(5) in Luenberger (1969) and the second equality from Lemmas S.2.12 and S.2.13. Furthermore, since $AA^{\dagger}\hat{C}_n\beta(P) = \beta(P)$ due to $\hat{C}_n\beta(P) = \beta(P)$ by Assumption 4.2(ii) and $\beta(P) \in R$, we obtain from symmetry of $\Omega^{\mathrm{i}}(P)$ and $AA^{\dagger}\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\} \in$ range$\{\Omega^{\mathrm{i}}(P)\}$ whenever $\{Z_i\}_{i=1}^n \in E_{2n}(P)$ due to range$\{\Sigma^{\mathrm{i}}(P)\} \subseteq$ range$\{\Omega^{\mathrm{i}}(P)\}$ by Assumption 4.4(i), that

$$\max_{s \in \mathcal{E}^{\mathrm{i}}(P)\setminus\mathcal{S}^{\mathrm{i}}(P)} \langle s, AA^{\dagger}\hat{C}_n\hat{\beta}_n\rangle = \max_{s \in \mathcal{E}^{\mathrm{i}}(P)\setminus\mathcal{S}^{\mathrm{i}}(P)} \langle s, AA^{\dagger}\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle s, \beta(P)\rangle$$
$$= \max_{s \in \mathcal{E}^{\mathrm{i}}(P)\setminus\mathcal{S}^{\mathrm{i}}(P)} \langle \Omega^{\mathrm{i}}(P)s, (\Omega^{\mathrm{i}}(P))^{\dagger} AA^{\dagger}\hat{C}_n\sqrt{n}\{\hat{\beta}_n - \beta(P)\}\rangle + \sqrt{n}\langle s, \beta(P)\rangle \leq 0, \quad \text{(S.51)}$$

where the inequality follows by definition of $\mathcal{S}^{\mathrm{i}}(P)$. Thus, if we suppose by way of contradiction that (S.49) fails to hold, then (S.50), (S.51), $\mathcal{S}^{\mathrm{i}}(P) \subseteq \mathcal{E}^{\mathrm{i}}(P)$, and $\mathcal{E}^{\mathrm{i}}(P)$

being finite, imply there exists a $\gamma^\star \in (0,1)$ (depending on $\hat{\beta}_n$ and $\hat{\beta}_n^{\mathrm{r}}$) such that

$$0 \geq \max_{s\in\mathcal{E}^{\mathrm{i}}(P)} \langle s, AA^\dagger\{(1-\gamma^\star)\hat{\beta}_n^{\mathrm{r}}+\gamma^\star \hat{C}_n\hat{\beta}_n\}\rangle = \sup_{s\in(AA')^\dagger\mathcal{V}^{\mathrm{i}}(P)} \langle s, AA^\dagger\{(1-\gamma^\star)\hat{\beta}_n^{\mathrm{r}}+\gamma^\star \hat{C}_n\hat{\beta}_n\}\rangle$$

$$= \sup_{s\in\mathcal{V}^{\mathrm{i}}(P)} \langle A^\dagger s, A^\dagger\{(1-\gamma^\star)\hat{\beta}_n^{\mathrm{r}}+\gamma^\star AA^\dagger \hat{C}_n\hat{\beta}_n\}\rangle \quad \text{(S.52)}$$

where the equalities follow from Lemmas S.2.12 and S.2.13, and again employing $A^\dagger AA^\dagger = A^\dagger$ by Proposition 6.11.1(5) in Luenberger (1969). However, by construction $\hat{\beta}_n^{\mathrm{r}} \in R$ and $AA^\dagger\hat{C}_n\hat{\beta}_n \in R$, and therefore result (S.52) and Theorem 3.1 imply that

$$(1-\gamma^\star)\hat{\beta}_n^{\mathrm{r}} + \gamma^\star AA^\dagger\hat{C}_n\hat{\beta}_n = Ax \text{ for some } x \geq 0.$$

Moreover, note if $T_n > 0$, then $\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}}\langle A^\dagger s, \hat{x}_n^\star\rangle > 0$ by definition and as a result $T_n > 0$ implies $\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}}\langle A^\dagger s, \hat{x}_n^\star - A^\dagger\hat{\beta}_n^{\mathrm{r}}\rangle > 0$ due to $\langle A^\dagger s, A^\dagger\hat{\beta}_n^{\mathrm{r}}\rangle \leq 0$ for all $s \in \hat{\mathcal{V}}_n$ by Theorem 3.1. Hence, if $T_n > 0$, then $\hat{x}_n^\star = A^\dagger\hat{C}_n\hat{\beta}_n$, $A^\dagger AA^\dagger = A^\dagger$, and $\gamma^\star \in (0,1)$ yield

$$\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, \hat{x}_n^\star - A^\dagger\{(1-\gamma^\star)\hat{\beta}_n^{\mathrm{r}} + \gamma^\star AA^\dagger\hat{C}_n\hat{\beta}_n\}\rangle$$

$$= (1-\gamma^\star)\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}}\langle A^\dagger s, \hat{x}_n^\star - A^\dagger\hat{\beta}_n^{\mathrm{r}}\rangle < \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}}\langle A^\dagger s, \hat{x}_n^\star - A^\dagger\hat{\beta}_n^{\mathrm{r}}\rangle,$$

which is impossible by definition of $\hat{\beta}_n^{\mathrm{r}}$. We thus obtain that if $\{Z_i\}_{i=1}^n \in E_n(P)$ and $T_n > 0$, then result (S.49) must hold.

To conclude, note that results (S.48) and (S.49) imply there is a $\hat{s}_n \in \mathcal{V}^{\mathrm{i}}(P)$ depending only on P and $\{Z_i\}_{i=1}^n$ such that $(AA')^\dagger\hat{s}_n \in \mathcal{E}^{\mathrm{i}}(P)$, $\sigma((AA')^\dagger\hat{s}_n, P) > 0$, and $0 = \lambda_n\langle A^\dagger\hat{s}_n, A^\dagger\hat{\beta}_n^{\mathrm{r}}\rangle \equiv \hat{\mathbb{U}}_n(\hat{s}_n) = 0$ whenever $\{Z_i\}_{i=1}^n \in E_n(P)$ and $T_n > 0$. Therefore, the definitions of $\hat{\mathbb{L}}_n$, $\mathbb{L}_n^\star(P)$, and $\hat{c}_n(\eta)$ together with $\{Z_i\}_{i=1}^n \in E_n(P) \subseteq E_{3n}(P)$ yield

$$P(\langle A^\dagger\hat{s}_n, A^\dagger\mathbb{G}_n^{\mathrm{i}\star}(P)\rangle \leq \hat{c}_n(\eta) + \frac{\underline{\sigma}(P)z_{\eta-\epsilon}}{2}|\{Z_i\}_{i=1}^n)$$
$$\geq P(\mathbb{L}_n^\star(P) \leq \hat{c}_n(\eta) + \frac{\underline{\sigma}(P)z_{\eta-\epsilon}}{2}|\{Z_i\}_{i=1}^n)$$
$$\geq P(\hat{\mathbb{L}}_n \leq \hat{c}_n(\eta)|\{X_i\}_{i=1}^n) - \epsilon$$
$$\geq \eta - \epsilon \quad \text{(S.53)}$$

whenever $P \in \mathbf{P}_0^{\mathrm{i}}$, $\{Z_i\}_{i=1}^n \in E_n(P)$, and $T_n > 0$. Furthermore, since $\mathbb{G}_n^{\mathrm{i}\star}(P) \in \mathrm{range}\{\Omega^{\mathrm{i}}(P)\}$ by Assumption 4.4(i) and Theorem 3.6.1 in Bogachev (1998), we have

$$\langle A^\dagger\hat{s}_n, A^\dagger\mathbb{G}_n^{\mathrm{i}\star}(P)\rangle = \langle\Omega^{\mathrm{i}}(P)(AA')^\dagger\hat{s}_n, (\Omega^{\mathrm{i}}(P))^\dagger\mathbb{G}_n^{\mathrm{i}\star}(P)\rangle \quad \text{(S.54)}$$

almost surely. Hence, $\mathbb{G}_n^{\mathrm{i}\star}(P)$ being independent of $\{Z_i\}_{i=1}^n$ implies $\langle A^\dagger\hat{s}_n, A^\dagger\mathbb{G}_n^{\mathrm{i}\star}(P)\rangle \sim N(0, (\sigma^{\mathrm{i}}((AA')^\dagger\hat{s}_n, P))^2)$ conditional on $\{Z_i\}_{i=1}^n$. Since (S.54) and $\sigma^{\mathrm{i}}((AA')^\dagger\hat{s}_n, P) >$

0 imply that the distribution of $\langle A^\dagger \hat{s}_n, A^\dagger \mathbb{G}_n^{\mathrm{i}\star}(P)\rangle$ conditional on $\{Z_i\}_{i=1}^n$ first order stochastically dominates $N(0, \underline{\sigma}(P))$ random variable, (S.53) yields

$$\hat{c}_n(\eta) + \frac{\underline{\sigma}(P) z_{\eta-\epsilon}}{2} \geq \underline{\sigma}(P) z_{\eta-\epsilon},$$

which establishes the claim of the lemma for the subset $\mathbf{P}_0^{\mathrm{i}}$.

Case III: For the final case, suppose $P \in \mathbf{P}_0^{\mathrm{d}} \equiv \{P \in \mathbf{P}_0 : \sigma^{\mathrm{j}}(s, P) = 0 \text{ for all } s \in \mathcal{E}^{\mathrm{j}}(P) \text{ and } \mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}\}$. Then, by Lemma S.2.3 we may set $E_n(P) \equiv \{T_n = 0\}$ and the claim of the lemma for the subset $\mathbf{P}_0^{\mathrm{d}}$ follows. ■

Theorem S.2.1. *Let Assumptions 4.1, 4.3(ii), and 4.4(i) hold with $a_n = o(1)$, set $\Sigma(P) \equiv E_P[\psi(X, P)\psi(X, P)']$, and suppose $(\hat{\mathbb{W}}_n^{\mathrm{e}}(P)', \hat{\mathbb{W}}_n^{\mathrm{i}}(P)')' \equiv \hat{\mathbb{W}}_n(P) \in \mathbf{R}^{2p}$ satisfies*

$$\|(\Omega^{\mathrm{e}}(P))^\dagger \{\hat{\mathbb{W}}_n^{\mathrm{e}}(P) - \mathbb{W}_n^{\mathrm{e}}(P)\}\|_\infty \vee \|(\Omega^{\mathrm{i}}(P))^\dagger \{\hat{\mathbb{W}}_n^{\mathrm{i}}(P) - \mathbb{W}_n^{\mathrm{i}}(P)\}\|_\infty = O_P(\omega_n) \quad \text{(S.55)}$$

for $\omega_n > 0$, $\mathbb{W}_n(P) \equiv (\mathbb{W}_n^{\mathrm{e}}(P)', \mathbb{W}_n^{\mathrm{i}}(P)')' \sim N(0, \Sigma(P))$, and $\hat{\mathbb{W}}_n^{\mathrm{e}}(P) \in \mathrm{range}\{\Omega^{\mathrm{e}}(P)\}$ and $\hat{\mathbb{W}}_n^{\mathrm{i}}(P) \in \mathrm{range}\{\Omega^{\mathrm{i}}(P)\}$ with probability tending to one uniformly in $P \in \mathbf{P}$. Then, for any $\mathbf{Q} \subseteq \mathbf{P}$ and possibly random function $\hat{f}_n(\cdot, P) : \mathbf{R}^p \to \mathbf{R}$ satisfying

$$\gamma \hat{f}_n(s, P) \leq \hat{f}_n(\gamma s, P) \leq 0 \quad \text{(S.56)}$$

for all s with $A^\dagger s \leq 0$, $\gamma \in [0, 1]$, and $P \in \mathbf{Q}$, it follows uniformly in $P \in \mathbf{Q}$ that

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

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

Proof: We establish only the second claim of the theorem, noting that the first claim follows from slightly simpler but largely identical arguments. First note that since $\Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \hat{\mathbb{W}}_n^{\mathrm{i}}(P) = \hat{\mathbb{W}}_n^{\mathrm{i}}(P)$ whenever $\hat{\mathbb{W}}_n^{\mathrm{i}}(P) \in \mathrm{range}\{\Omega^{\mathrm{i}}(P)\}$, it follows that

$$\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \hat{\mathbb{W}}_n^{\mathrm{i}}(P)\rangle + \hat{f}_n(s, P) = \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \hat{\mathbb{W}}_n^{\mathrm{i}}(P)\rangle + \hat{f}_n(s, P) \quad \text{(S.57)}$$

with probability tending to one uniformly in $P \in \mathbf{P}$. Further note that Lemma S.2.10 and Assumption 4.1(iii) imply $\hat{\Omega}_n^{\mathrm{i}}(\hat{\Omega}_n^{\mathrm{i}})^\dagger \Omega^{\mathrm{i}}(P) = \Omega^{\mathrm{i}}(P)$ with probability tending to one uniformly in $P \in \mathbf{P}$. Thus, since $\hat{\Omega}_n^{\mathrm{i}}$ and $\Omega^{\mathrm{i}}(P)$ are symmetric by Assumption 4.1(i)(ii), it follows that $\Omega^{\mathrm{i}}(P) = \Omega^{\mathrm{i}}(P)(\hat{\Omega}_n^{\mathrm{i}})^\dagger \hat{\Omega}_n^{\mathrm{i}}$ with probability tending to one uniformly in $P \in \mathbf{P}$. Employing the triangle inequality, the definition of $\hat{\mathcal{V}}_n^{\mathrm{i}}$, and $\hat{\Omega}_n^{\mathrm{i}}(\hat{\Omega}_n^{\mathrm{i}})^\dagger \hat{\Omega}_n^{\mathrm{i}} = \hat{\Omega}_n^{\mathrm{i}}$

by Proposition 6.11.1(6) in Luenberger (1969) we can conclude that

$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}}\|\Omega^{\rm i}(P)(AA')^\dagger s\|_1 \le 1+\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}}\|(\hat{\Omega}_n^{\rm i}-\Omega^{\rm i}(P))(AA')^\dagger s\|_1$$
$$= 1+\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}}\|(\hat{\Omega}_n^{\rm i}-\Omega^{\rm i}(P))(\hat{\Omega}_n^{\rm i})^\dagger\hat{\Omega}_n^{\rm i}(AA')^\dagger s\|_1 \le 1+\|(\hat{\Omega}_n^{\rm i}-\Omega^{\rm i}(P))(\hat{\Omega}_n^{\rm i})^\dagger\|_{o,1} \quad \text{(S.58)}$$

with probability tending to one uniformly in $P\in\mathbf{P}$. Further note that Theorem 6.5.1 in Luenberger (1969), symmetry of $\hat{\Omega}_n^{\rm i}$ and $\Omega^{\rm i}(P)$, and Lemma S.2.6 imply

$$\|(\hat{\Omega}_n^{\rm i}-\Omega^{\rm i}(P))(\hat{\Omega}_n^{\rm i})^\dagger\|_{o,1} = \|(\hat{\Omega}_n^{\rm i})^\dagger(\hat{\Omega}_n^{\rm i}-\Omega^{\rm i}(P))\|_{o,\infty} = O_P(\frac{a_n}{\sqrt{\log(1+p)}}) \quad \text{(S.59)}$$

uniformly in $P\in\mathbf{P}$. Next, note that since $\Omega^{\rm i}(P)(A^\dagger)'A^\dagger = \Omega^{\rm i}(P)(AA')^\dagger$ (see, e.g., Seber (2008) pg. 139), Hölder's inequality, and results (S.58) and (S.59) yield

$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}}|\langle A^\dagger s, A^\dagger\Omega^{\rm i}(P)(\Omega^{\rm i}(P))^\dagger(\hat{\mathbb{W}}_n^{\rm i}(P)-\mathbb{W}_n^{\rm i}(P))\rangle|$$
$$\le (1+O_P(\frac{a_n}{\sqrt{\log(1+p)}}))\|(\Omega^{\rm i}(P))^\dagger(\hat{\mathbb{W}}_n^{\rm i}(P)-\mathbb{W}_n^{\rm i}(P))\|_\infty = O_P(\omega_n) \quad \text{(S.60)}$$

uniformly in $P\in\mathbf{P}$, and where the final equality follows from $a_n=o(1)$ by assumption. Therefore, combining results (S.57) and (S.60) we conclude that uniformly in $P\in\mathbf{P}$

$$\sup_{s\in\hat{\mathcal{V}}_n^{\rm i}}\langle A^\dagger s, A^\dagger\hat{\mathbb{W}}_n^{\rm i}(P)\rangle + \hat{f}_n(s,P)$$
$$= \sup_{s\in\hat{\mathcal{V}}_n^{\rm i}}\langle A^\dagger s, A^\dagger\Omega^{\rm i}(P)(\Omega^{\rm i}(P))^\dagger\mathbb{W}_n^{\rm i}(P)\rangle + \hat{f}_n(s,P)+O_P(\omega_n). \quad \text{(S.61)}$$

We next aim to replace $\hat{\mathcal{V}}_n^{\rm i}$ with $\mathcal{V}^{\rm i}(P)$ in (S.61). To this end, let $\hat{s}_n\in\hat{\mathcal{V}}_n^{\rm i}$ satisfy

$$\langle A^\dagger\hat{s}_n, A^\dagger\Omega^{\rm i}(P)(\Omega^{\rm i}(P))^\dagger\mathbb{W}_n^{\rm i}(P)\rangle + \hat{f}_n(\hat{s}_n,P)$$
$$= \sup_{s\in\hat{\mathcal{V}}_n^{\rm i}}\langle A^\dagger s, A^\dagger\Omega^{\rm i}(P)(\Omega^{\rm i}(P))^\dagger\mathbb{W}_n^{\rm i}(P)\rangle + \hat{f}_n(s,P)+O(\omega_n), \quad \text{(S.62)}$$

where note $\hat{s}_n$ is random and (S.62) is meant to hold surely. Set $\bar{s}_n\equiv\gamma_n\hat{s}_n$ with

$$\gamma_n \equiv (\|\Omega^{\rm i}(P)(AA')^\dagger\hat{s}_n\|_1\vee 1)^{-1}\in[0,1], \quad \text{(S.63)}$$

and note that since $\gamma_n\le 1$, result (S.58) and $\hat{s}_n\in\hat{\mathcal{V}}_n^{\rm i}$ allow us to conclude that

$$0\le 1-\gamma_n\le 1-(1+\|(\hat{\Omega}_n^{\rm i}-\Omega^{\rm i}(P))(\hat{\Omega}_n^{\rm i})^\dagger\|_{o,1})^{-1} \quad \text{(S.64)}$$

with probability tending to one uniformly in $P \in \mathbf{P}$. Hence, (S.59) and (S.64) yield

$$0 \le 1 - \gamma_n \le O_P(\frac{a_n}{\sqrt{\log(1+p)}}) \tag{S.65}$$

uniformly $P \in \mathbf{P}$ due to $a_n = o(1)$. Next, we note $A^\dagger \hat{s}_n \le 0$ since $\hat{s}_n \in \hat{\mathcal{V}}_n^{\mathrm{i}}$ and therefore $A^\dagger \bar{s}_n = \gamma_n A^\dagger \hat{s}_n \le 0$ because $\gamma_n \ge 0$. Since $\bar{s}_n = \gamma_n \hat{s}_n$ and (S.63) further imply

$$\|\Omega^{\mathrm{i}}(P)(AA')^\dagger \bar{s}_n\|_1 = (\|\Omega^{\mathrm{i}}(P)(AA')^\dagger \hat{s}_n\|_1 \vee 1)^{-1} \|\Omega^{\mathrm{i}}(P)(AA')^\dagger \hat{s}_n\|_1 \le 1, \tag{S.66}$$

it follows that $\bar{s}_n \in \mathcal{V}^{\mathrm{i}}(P)$. Moreover, $\hat{s}_n - \bar{s}_n = (1-\gamma_n)\hat{s}_n$, $\gamma_n \hat{f}_n(\hat{s}_n, P) \le \hat{f}_n(\gamma_n \hat{s}_n, P)$ and $\hat{f}_n(\hat{s}_n, P) \le 0$ for all $P \in \mathbf{Q}$ by (S.56), and Hölder's inequality yield

$$\begin{aligned} \langle A^\dagger(\hat{s}_n - \bar{s}_n), & A^\dagger \Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)\rangle + \hat{f}_n(\hat{s}_n, P) - \hat{f}_n(\bar{s}_n, P) \\ & \le (1-\gamma_n)\{\langle A^\dagger \hat{s}_n, A^\dagger \Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)\rangle + \hat{f}_n(\hat{s}_n, P)\} \\ & \le (1-\gamma_n)\{\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \|\Omega^{\mathrm{i}}(P)(AA')^\dagger s\|_1\} \|(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)\|_\infty. \end{aligned}$$

In particular, since $\sup_{P \in \mathbf{P}} E_P[\|(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)\|_\infty] \lesssim \sqrt{\log(1+p)}$ by Lemma S.2.8 and Assumption 4.3(ii), Markov's inequality, results (S.58), (S.59), (S.62), and (S.65), and $\bar{s}_n \in \mathcal{V}^{\mathrm{i}}(P)$ allow us to conclude that uniformly in $P \in \mathbf{Q}$ we have

$$\begin{aligned} \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} & \langle A^\dagger s, A^\dagger \Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)\rangle + \hat{f}_n(s, P) \\ & \le \sup_{s \in \mathcal{V}^{\mathrm{i}}(P)} \langle A^\dagger s, A^\dagger \Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)\rangle + \hat{f}_n(s, P) + O_P(\omega_n + a_n) \end{aligned} \tag{S.67}$$

uniformly in $P \in \mathbf{Q}$. The reverse inequality to (S.67) can be established by very similar arguments, and therefore we can conclude that uniformly in $P \in \mathbf{Q}$ we have

$$\begin{aligned} \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} & \langle A^\dagger s, A^\dagger \Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)\rangle + \hat{f}_n(s, P) \\ & = \sup_{s \in \mathcal{V}^{\mathrm{i}}(P)} \langle A^\dagger s, A^\dagger \Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)\rangle + \hat{f}_n(s, P) + O_P(\omega_n + a_n). \end{aligned} \tag{S.68}$$

Finally, note $\mathbb{W}_n^{\mathrm{i}}(P)$ almost surely belongs to the range of $\Sigma^{\mathrm{i}}(P) : \mathbf{R}^p \to \mathbf{R}^p$ by Theorem 3.6.1 in Bogachev (1998). Hence, since Assumption 4.4(i) implies $\Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \Sigma^{\mathrm{i}}(P)$ it follows that $\mathbb{W}_n^{\mathrm{i}}(P) = \Omega^{\mathrm{i}}(P)(\Omega^{\mathrm{i}}(P))^\dagger \mathbb{W}_n^{\mathrm{i}}(P)$ P-almost surely. The second claim of the theorem thus follows from (S.61), and (S.68). ■

Lemma S.2.3. *Let Assumptions 4.1, 4.2(ii), 4.4, 4.5(v) hold, $a_n = o(1)$, and for $\mathrm{j} \in \{\mathrm{e}, \mathrm{i}\}$ set $\mathbf{D}_0^{\mathrm{j}} \equiv \{P \in \mathbf{P}_0 : \sigma^{\mathrm{j}}(s, P) = 0$ for all $s \in \mathcal{E}^{\mathrm{j}}(P)\}$. Then:*

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

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

Proof: First note that Theorem 3.6.1 in Bogachev (1998) and Assumption 4.4(i) imply $\mathbb{G}_n^{\mathrm{e}}(P)\in\mathrm{range}\{\Sigma^{\mathrm{e}}(P)\}\subseteq\mathrm{range}\{\Omega^{\mathrm{e}}(P)\}$ almost surely. Hence, $\Omega^{\mathrm{e}}(P)(\Omega^{\mathrm{e}}(P))^\dagger\mathbb{G}_n^{\mathrm{e}}(P) = \mathbb{G}_n^{\mathrm{e}}(P)$ almost surely and symmetry of $\Omega^{\mathrm{e}}(P)$ imply for any $P\in\mathbf{D}_0^{\mathrm{e}}$ that

$$\begin{aligned}\sup_{s\in\mathcal{V}^{\mathrm{e}}(P)}|\langle s,\mathbb{G}_n^{\mathrm{e}}(P)\rangle| &= \sup_{s\in\mathcal{V}^{\mathrm{e}}(P)}|\langle\Omega^{\mathrm{e}}(P)s,(\Omega^{\mathrm{e}}(P))^\dagger\mathbb{G}_n^{\mathrm{e}}(P)\rangle|\\ &= \max_{s\in\mathcal{E}^{\mathrm{e}}(P)}|\langle s,(\Omega^{\mathrm{e}}(P))^\dagger\mathbb{G}_n^{\mathrm{e}}(P)\rangle| = 0,\end{aligned}\quad\text{(S.69)}$$

where the second equality follows from Hölder's inequality implying the supremum is finite and Lemma S.2.12. Also note that $\hat{\Omega}_n^{\mathrm{e}}(\hat{\Omega}_n^{\mathrm{e}})^\dagger\Omega^{\mathrm{e}}(P) = \Omega^{\mathrm{e}}(P)$ with probability tending to one uniformly in $P\in\mathbf{P}$ by Assumption 4.1(iii) and Lemma S.2.10. Thus, by symmetry of $\hat{\Omega}_n^{\mathrm{e}}$ and $\Omega^{\mathrm{e}}(P)$ we obtain that $\Omega^{\mathrm{e}}(P) = \Omega^{\mathrm{e}}(P)(\hat{\Omega}_n^{\mathrm{e}})^\dagger\hat{\Omega}_n^{\mathrm{e}}$, which together with the triangle inequality, definition of $\hat{\mathcal{V}}_n^{\mathrm{e}}$, and $\hat{\Omega}_n^{\mathrm{e}}(\hat{\Omega}_n^{\mathrm{e}})^\dagger\hat{\Omega}_n^{\mathrm{e}} = \hat{\Omega}_n^{\mathrm{e}}$ by Proposition 6.11.1(6) in Luenberger (1969) imply with probability tending to one uniformly in $P\in\mathbf{P}$ that

$$\begin{aligned}\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}}\|\Omega^{\mathrm{e}}(P)s\|_1 &\leq 1+\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}}\|(\hat{\Omega}_n^{\mathrm{e}}-\Omega^{\mathrm{e}}(P))s\|_1\\ &= 1+\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}}\|(\hat{\Omega}_n^{\mathrm{e}}-\Omega^{\mathrm{e}}(P))(\hat{\Omega}_n^{\mathrm{e}})^\dagger\hat{\Omega}_n^{\mathrm{e}}s\|_1 \leq 1+\|(\hat{\Omega}_n^{\mathrm{e}})^\dagger(\hat{\Omega}_n^{\mathrm{e}}-\Omega^{\mathrm{e}}(P))\|_{o,\infty},\end{aligned}$$

where the final inequality follows from Theorem 6.5.1 in Luenberger (1969). Therefore, Lemma S.2.6 and $a_n = o(1)$ imply that $\hat{\mathcal{V}}_n^{\mathrm{e}}\subseteq 2\mathcal{V}^{\mathrm{e}}(P)$ with probability tending to one uniformly in $P\in\mathbf{P}$. We can thus conclude from $0\in\hat{\mathcal{V}}_n^{\mathrm{e}}$, result (S.69), Assumption 4.4(ii), and the support of $\mathbb{G}_n^{\mathrm{e}}(P)$ being equal to the range of $\Sigma^{\mathrm{e}}(P)$ by Theorem 3.6.1 in Bogachev (1998) that with probability tending to one uniformly in $P\in\mathbf{D}_0^{\mathrm{e}}$

$$0\leq\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}}|\sqrt{n}\langle s,\hat{\beta}_n - A\hat{x}_n^\star\rangle| \leq \sup_{s\in 2\mathcal{V}^{\mathrm{e}}(P)}|\langle s,(I_p - AA^\dagger\hat{C}_n)\sqrt{n}\{\hat{\beta}_n-\beta(P)\}\rangle| = 0.\quad\text{(S.70)}$$

Moreover, identical arguments but relying on Assumption 4.5(v) instead of 4.4(i) yield

$$0\leq\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}}|\langle s,\hat{\mathbb{G}}_n^{\mathrm{e}}\rangle| \leq \sup_{s\in 2\mathcal{V}^{\mathrm{e}}(P)}|\langle s,\hat{\mathbb{G}}_n^{\mathrm{e}}\rangle| = 0\quad\text{(S.71)}$$

with probability tending to one uniformly in $P\in\mathbf{D}_0^{\mathrm{e}}$. The first claim of the lemma therefore follows from results (S.70) and (S.71).

For the second claim of the lemma, we note that identical arguments to those employed for the first claim readily establish that $\hat{\mathcal{V}}_n^{\mathrm{i}}\subseteq 2\mathcal{V}^{\mathrm{i}}(P)$ and

$$\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}}|\langle A^\dagger s, A^\dagger AA^\dagger\hat{C}_n\sqrt{n}\{\hat{\beta}_n-\beta(P)\}\rangle| = \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}}|\langle A^\dagger s, A^\dagger\hat{\mathbb{G}}_n^{\mathrm{i}}\rangle| = 0\quad\text{(S.72)}$$

with probability tending to one uniformly in $P \in \mathbf{D}_0^{\mathrm{i}}$. Further note that since $A^\dagger A A^\dagger = A^\dagger$ by Proposition 6.11.1(5) in Luenberger (1969), it follows that $A^\dagger A A^\dagger \hat{C}_n \hat{\beta}_n = \hat{x}_n^\star$ due to $\hat{x}_n^\star \equiv A^\dagger \hat{C}_n \hat{\beta}_n$ by Assumption 4.2(ii) and therefore (S.72) yields

$$\sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} |\langle A^\dagger s, \hat{x}_n^\star - A^\dagger \beta(P)\}\rangle| = 0 \tag{S.73}$$

with probability tending to one uniformly in $P \in \mathbf{D}_0^{\mathrm{i}}$. Therefore, since $\langle A^\dagger s, A^\dagger \beta(P)\rangle \leq 0$ for any $P \in \mathbf{P}_0$ and $s \in \mathcal{V}^{\mathrm{i}}(P)$ by Theorem 3.1, we obtain from $0 \in \hat{\mathcal{V}}_n^{\mathrm{i}}$ and (S.73) that

$$0 \leq \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle \leq \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} |\langle A^\dagger s, \hat{x}_n^\star - A^\dagger \beta(P)\}\rangle| + \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger \beta(P)\rangle = 0$$

with probability tending to one uniformly in $P \in \mathbf{D}_0^{\mathrm{i}}$. ■

Lemma S.2.4. *Set $\Sigma(P) \equiv E_P[\psi(X,P)\psi(X,P)']$ and $r_n \equiv a_n + M_{3,\Psi} p^{1/3}(\log(1+p))^{5/6}/n^{1/6}$. If Assumptions 4.2(i)(iii), 4.3, 4.4(i) hold, and $r_n = o(1)$, then there exists a Gaussian $(\mathbb{G}_n^{\mathrm{e}}(P)', \mathbb{G}_n^{\mathrm{i}}(P)')' \equiv \mathbb{G}_n(P) \sim N(0, \Sigma(P))$ satisfying uniformly in $P \in \mathbf{P}$:*

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

Proof: We first set $\tilde{\psi}(Z,P) \equiv (((\Omega^{\mathrm{e}}(P))^\dagger \psi^{\mathrm{e}}(Z,P))', ((\Omega^{\mathrm{i}}(P))^\dagger \psi^{\mathrm{i}}(Z,P))')' \in \mathbf{R}^{2p}$, define

$$\tilde{\Sigma}(P) \equiv E_P[\tilde{\psi}(Z,P)\tilde{\psi}(Z,P)'], \tag{S.74}$$

and let $S_n(P) \in \mathbf{R}^{2p}$ be normally distributed with mean zero and variance $\tilde{\Sigma}(P)/n$. Next observe that since $\|a\|_2^2 \leq 2p\|a\|_\infty^2$ for any $a \in \mathbf{R}^{2p}$ we can conclude that

$$E_P[\|S_n(P)\|_2^2 \|S_n(P)\|_\infty] \leq 2pE_P[\|S_n(P)\|_\infty^3] \lesssim p(\frac{\sqrt{\log(1+p)}}{\sqrt{n}})^3, \tag{S.75}$$

where the second inequality follows from Lemma S.2.8 and Assumption 4.3(ii). Moreover, by similar arguments, Assumption 4.3(iii), and result (S.75) we can conclude

$$\begin{aligned} n\{E_P[\|\frac{\tilde{\psi}(Z,P)}{\sqrt{n}}\|_2^2 \|\frac{\tilde{\psi}(Z,P)}{\sqrt{n}}\|_\infty] + E_P[\|S_n(P)\|_2^2 \|S_n(P)\|_\infty]\} \\ \lesssim n\{\frac{p}{n^{3/2}} E_P[\Psi^3(Z,P)] + E_P[\|S_n(P)\|_2^2 \|S_n(P)\|_\infty]\} \\ \lesssim \frac{p}{\sqrt{n}}\{M_{3,\Psi}^3 + (\log(1+p))^{3/2}\}. \end{aligned} \tag{S.76}$$

Setting $\mathbb{Z} \sim N(0, I_{2p})$, we then obtain by Assumptions 4.2(i), 4.3(i), Lemma 39 in Belloni et al. (2019), and (S.76) that for any $\delta > 0$ there is a $\tilde{\mathbb{G}}_n(P) \sim N(0, \tilde{\Sigma}(P))$ such that

$$P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}(Z_i, P) - \tilde{\mathbb{G}}_n(P)\|_\infty > \delta)$$
$$\lesssim \min_{t\geq 0}\{P(\|\mathbb{Z}\|_\infty > t) + \frac{t^2}{\delta^3}\frac{p}{\sqrt{n}}\{M_{3,\Psi}^3 + (\log(1+p))^{3/2}\}$$
$$\lesssim \min_{t\geq 0}\{\exp\{-\frac{t^2}{8\log(1+p)}\} + \frac{t^2}{\delta^3}\frac{pM_{3,\Psi}^3(\log(1+p))^{3/2}}{\sqrt{n}}\}, \tag{S.77}$$

where the final inequality follows from Proposition A.2.1 in van der Vaart and Wellner (1996), $E[\|\mathbb{Z}\|_\infty^2] \lesssim \log(1+p)$ by Lemma S.2.8, and we employed that $M_{3,\Psi} \geq 1$ by Assumption 4.3(iii) in order to simplify the bound. Thus, by setting $t = K\sqrt{\log(1+p)}$ and $\delta^3 = K^3 pM_{3,\Psi}^3(\log(1+p))^{5/2}/\sqrt{n}$ in (S.77) for any $K > 0$ yields

$$\lim_{K\uparrow\infty}\limsup_{n\to\infty}\sup_{P\in\mathbf{P}} P(\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}(Z_i, P) - \tilde{\mathbb{G}}_n(P)\|_\infty > K\frac{M_{3,\Psi}p^{1/3}(\log(1+p))^{5/6}}{n^{1/6}})$$
$$\lesssim \lim_{K\uparrow\infty}\{\exp\{-\frac{K^2}{8}\} + \frac{1}{K}\} = 0. \tag{S.78}$$

Since $r_n \equiv M_{3,\Psi}p^{1/3}(\log(1+p))^{5/6}/n^{1/6} + a_n$, result (S.78), Assumption 4.2(iii), writing $\tilde{\mathbb{G}}_n(P) \equiv (\tilde{\mathbb{G}}_n^{\mathrm{e}}(P)', \tilde{\mathbb{G}}_n^{\mathrm{i}}(P)')'$, and the triangle inequality imply that uniformly in $P \in \mathbf{P}$

$$\|(\Omega^{\mathrm{e}}(P))^\dagger(I_p - AA^\dagger)\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \tilde{\mathbb{G}}_n^{\mathrm{e}}(P)\|_\infty = O_P(r_n)$$
$$\|(\Omega^{\mathrm{i}}(P))^\dagger AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \tilde{\mathbb{G}}_n^{\mathrm{i}}(P)\|_\infty = O_P(r_n). \tag{S.79}$$

To conclude, note that for $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$, $\tilde{\mathbb{G}}_n^{\mathrm{j}}(P) \sim N(0, (\Omega^{\mathrm{j}}(P))^\dagger\Sigma^{\mathrm{j}}(P)(\Omega^{\mathrm{j}}(P))^\dagger)$ and therefore Theorem 3.6.1 in Bogachev (1998) implies that $\tilde{\mathbb{G}}_n^{\mathrm{j}}(P)$ belongs to the range of the map $(\Omega^{\mathrm{j}}(P))^\dagger\Sigma^{\mathrm{j}}(P)(\Omega^{\mathrm{j}}(P))^\dagger : \mathbf{R}^p \to \mathbf{R}^p$ almost surely. Thus, since for $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$ we have $(\Omega^{\mathrm{j}}(P))^\dagger\Omega^{\mathrm{j}}(P)(\Omega^{\mathrm{j}}(P))^\dagger = (\Omega^{\mathrm{j}}(P))^\dagger$ it follows that $(\Omega^{\mathrm{j}}(P))^\dagger\Omega^{\mathrm{j}}(P)\tilde{\mathbb{G}}_n^{\mathrm{j}}(P) = \tilde{\mathbb{G}}_n^{\mathrm{j}}(P)$ almost surely. Hence, setting $\mathbb{G}_n^{\mathrm{j}}(P) = \Omega^{\mathrm{j}}(P)\tilde{\mathbb{G}}_n^{\mathrm{j}}(P)$ for $\mathrm{j} \in \{\mathrm{e},\mathrm{i}\}$ we can conclude

$$\|(\Omega^{\mathrm{e}}(P))^\dagger\{(I_p - AA^\dagger)\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \mathbb{G}_n^{\mathrm{e}}(P)\}\|_\infty = O_P(r_n)$$
$$\|(\Omega^{\mathrm{i}}(P))^\dagger\{AA^\dagger\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \mathbb{G}_n^{\mathrm{i}}(P)\}\|_\infty = O_P(r_n) \tag{S.80}$$

uniformly in $P \in \mathbf{P}$ by result (S.79). Since $\tilde{\mathbb{G}}_n(P) \sim N(0, \tilde{\Sigma}(P))$, and Assumption 4.4(i) implies $\Omega^{\mathrm{j}}(P)(\Omega^{\mathrm{j}}(P))^\dagger\psi^{\mathrm{j}}(Z, P) = \psi^{\mathrm{j}}(Z, P)$ P-almost surely, we can conclude from the definition of $\tilde{\Sigma}(P)$ in (S.74) and $\mathbb{G}_n(P) = ((\Omega^{\mathrm{e}}(P)\tilde{\mathbb{G}}_n^{\mathrm{e}}(P))', (\Omega^{\mathrm{i}}(P)\tilde{\mathbb{G}}_n^{\mathrm{i}}(P))')'$ that $\mathbb{G}_n(P) \sim N(0, \Sigma(P))$ and thus the claim of the lemma follows. ■

Lemma S.2.5. *Let Assumptions 4.2(i), 4.3, 4.4(i), 4.5(i)-(iv) hold, and define*

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

If $b_n = o(1)$, then there is a Gaussian vector $(\mathbb{G}_n^{\mathrm{e}\star}(P)', \mathbb{G}_n^{\mathrm{i}\star}(P)') \equiv \mathbb{G}_n^{\star}(P) \sim N(0, \Sigma(P))$ independent of $\{Z_i\}_{i=1}^n$ and satisfying uniformly in $P \in \mathbf{P}$:

$$\|(\Omega^{\mathrm{e}}(P))^{\dagger}\{\hat{\mathbb{G}}_n^{\mathrm{e}} - \mathbb{G}_n^{\mathrm{e}\star}(P)\}\|_\infty \vee \|(\Omega^{\mathrm{i}}(P))^{\dagger}\{\hat{\mathbb{G}}_n^{\mathrm{i}} - \mathbb{G}_n^{\mathrm{i}\star}(P)\}\|_\infty = O_P(b_n).$$

Proof: For ease of exposition we divide the proof into multiple steps. In the arguments that follow, we let $\varphi(Z, P) \equiv (\varphi^{\mathrm{e}}(Z, P)', \varphi^{\mathrm{i}}(Z, P)')' \in \mathbf{R}^{2p}$, where

$$\varphi^{\mathrm{e}}(Z, P) \equiv (\Omega^{\mathrm{e}}(P))^{\dagger}\psi^{\mathrm{e}}(Z, P) \qquad \varphi^{\mathrm{i}}(Z, P) \equiv (\Omega^{\mathrm{i}}(P))^{\dagger}\psi^{\mathrm{i}}(Z, P). \tag{S.81}$$

Step 1: (Distributional Representation). Let $\{U_i\}_{i=1}^{\infty}$ be an i.i.d. sequence independent of $\{Z_i, W_{i,n}\}_{i=1}^n$ with U_i uniformly distributed on $(0, 1]$. We further set $(U_{(1),n}, \ldots U_{(n),n})$ to denote the order statistics of $\{U_i\}_{i=1}^n$ and $R_{i,n}$ to denote the rank of each U_i (i.e., $U_i = U_{(R_{i,n}),n}$). By Lemma 13.1(iv) in van der Vaart (1999), it then follows that the vector $R_n \equiv (R_{1,n}, \ldots, R_{n,n})$ is uniformly distributed on the set of all $n!$ permutations of $\{1, \ldots, n\}$ and hence by Assumption 4.5(i) we can conclude that

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

where $\stackrel{d}{=}$ denotes equality in distribution and $\bar{W}_n \equiv \sum_{i=1}^n W_{i,n}/n$.

Step 2: (Couple to i.i.d.). We next define $\tau_n : [0, 1] \to \{W_{i,n} - \bar{W}_n\}_{i=1}^n$ to be given by

$$\tau_n(u) \equiv \inf\{c : \frac{1}{n}\sum_{i=1}^n 1\{W_{i,n} - \bar{W}_n \leq c\} \geq u\};$$

i.e., τ_n is the empirical quantile function of the sample $\{W_{i,n} - \bar{W}_n\}_{i=1}^n$. In addition, set

$$S_n(P) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n (W_{R_i,n} - \bar{W}_n)\varphi(Z_i, P)$$
$$L_n(P) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n (\varphi(Z_i, P) - \bar{\varphi}_n(P))\tau_n(U_i)$$

where $\bar{\varphi}_n(P) \equiv \sum_{i=1}^n \varphi(Z_i, P)/n$. Letting $S_{j,n}(P)$ and $L_{j,n}(P)$ denote the j^{th} coordinates of $S_n(P)$ and $L_n(P)$ respectively, we then observe that Theorem 3.1 in Hájek (1961) (see in particular equation (3.11) in page 512) allows us to conclude that

$$E[(S_{j,n}(P) - L_{j,n}(P))^2|\{Z_i, W_{i,n}\}_{i=1}^n] \lesssim \mathrm{Var}\{L_{j,n}(P)|\{Z_i, W_{i,n}\}_{i=1}^n\}\frac{\max_{1\leq i\leq n}|W_{i,n} - \bar{W}_n|}{(\sum_{i=1}^n (W_{i,n} - \bar{W}_n)^2)^{1/2}}. \tag{S.82}$$

In order to study the properties of $L_n(P)$ it is convenient to define $\xi_{i,n}(P)$ to equal

$$\xi_{i,n}(P) \equiv (\varphi(Z_i, P) - \bar{\varphi}_n(P))\frac{\tau_n(U_i)}{\sqrt{n}}. \tag{S.83}$$

Then note that since $\{U_i\}_{i=1}^n$ are i.i.d. uniform on $(0,1]$ and independent of $\{Z_i, W_{i,n}\}_{i=1}^n$, and τ_n is the empirical quantile function of $\{W_{i,n} - \bar{W}_n\}_{i=1}^n$ it follows that

$$E[\xi_{i,n}(P)|\{Z_i, W_{i,n}\}_{i=1}^n] = \frac{1}{\sqrt{n}}(\varphi(Z_i, P) - \bar{\varphi}_n(P))(\frac{1}{n}\sum_{i=1}^n W_{i,n} - \bar{W}_n) = 0$$

$$E[\xi_{i,n}(P)\xi_{i,n}(P)'|\{Z_i, W_{i,n}\}_{i=1}^n] = \frac{\hat{\sigma}_n^2}{n}(\varphi(Z_i, P) - \bar{\varphi}_n(P))(\varphi(Z_i, P) - \bar{\varphi}_n(P))', \tag{S.84}$$

where $\hat{\sigma}_n^2 \equiv \sum_{i=1}^n (W_{i,n} - \bar{W}_n)^2/n$. Hence, since the variables $\{\xi_{i,n}(P)\}_{i=1}^n$ are independent conditional on $\{Z_i, W_{i,n}\}_{i=1}^n$ it follows from $L_n(P) = \sum_{i=1}^n \xi_{i,n}(P)$ that

$$\text{Var}\{L_{j,n}(P)|\{Z_i, W_{i,n}\}_{i=1}^n\} = \frac{\hat{\sigma}_n^2}{n}\sum_{i=1}^n (\varphi_j(Z_i, P) - \bar{\varphi}_{j,n}(P))^2, \tag{S.85}$$

where $\varphi_j(Z_i, P)$ and $\bar{\varphi}_{j,n}(P)$ denote the j^{th} coordinates of $\varphi(Z_i, P)$ and $\bar{\varphi}_n(P)$ respectively. Thus, since for any random variable $(V_1, \ldots, V_{2p}) \equiv V \in \mathbf{R}^{2p}$ Jensen's inequality implies $E[\|V\|_\infty] \leq \sqrt{2p}\max_{1\leq j\leq 2p}(E[V_j^2])^{1/2}$, results (S.82) and (S.85) yield

$$\begin{aligned} &E[\|S_n(P) - L_n(P)\|_\infty|\{Z_i, W_{i,n}\}_{i=1}^n] \\ &\qquad \lesssim \sqrt{p}\max_{1\leq j\leq 2p}(\frac{\hat{\sigma}_n}{n^{3/2}}\sum_{i=1}^n (\varphi_j(Z_i, P) - \bar{\varphi}_{j,n}(P))^2)^{1/2}(\max_{1\leq i\leq n}|W_{i,n} - \bar{W}_n|)^{1/2}. \end{aligned} \tag{S.86}$$

Next, we note that the definition of $\varphi(z, P)$ implies that $\Psi(z, P)$, as introduced in Assumption 4.3(iii), satisfies $\Psi(z, P) = \|\varphi(z, P)\|_\infty$. Hence, for $M_{3,\Psi}$ as introduced in Assumption 4.3(iii), Markov and Jensen's inequalities imply for any $C > 0$ that

$$\begin{aligned} &\sup_{P\in\mathbf{P}} P(|\frac{1}{n}\sum_{i=1}^n \Psi^2(Z_i, P)| > CM_{3,\Psi}^2) \\ &\qquad \leq \frac{1}{CM_{3,\Psi}^2}\sup_{P\in\mathbf{P}} E_P[|\frac{1}{n}\sum_{i=1}^n \Psi^2(Z_i, P)|] \leq \frac{1}{CM_{3,\Psi}^2}\sup_{P\in\mathbf{P}}\|\Psi(\cdot, P)\|_{P,2}^2 \leq \frac{1}{C}. \end{aligned}$$

Thus, using that $\Psi(z, P) = \|\varphi(z, P)\|_\infty$ we conclude uniformly in $P \in \mathbf{P}$ that

$$\max_{1\leq j\leq 2p}\frac{1}{n}\sum_{i=1}^n (\varphi_j(Z_i, P) - \bar{\varphi}_{j,n}(P))^2 \leq \frac{1}{n}\sum_{i=1}^n \Psi^2(Z_i, P) = O_P(M_{3,\Psi}^2). \tag{S.87}$$

Moreover, by the triangle inequality, Assumption 4.5(ii), Lemma 2.2.10 in van der Vaart and Wellner (1996), and $E[|V|] \leq \|V\|_{\psi_1}$ for any random variable V and $\|\cdot\|_{\psi_1}$ the Orlicz norm based on $\psi_1 = e^x - 1$, we can conclude that

$$E[\max_{1\leq i\leq n} |W_{i,n} - \bar{W}_n|] \leq E[\max_{1\leq i\leq n} |W_{i,n} - E[W_{1,n}]|] + E[|\bar{W}_n - E[W_{1,n}]|]$$
$$\lesssim \log(1+n) + E[|W_{1,n}|]. \quad \text{(S.88)}$$

Thus, $\hat{\sigma}_n^2 \xrightarrow{P} 1$ by Assumption 4.5(iii), $E[|W_{1,n}|]$ being uniformly bounded in n by Jensen's inequality and Assumption 4.5(iii), and results (S.86), (S.87), (S.88) yield

$$E[\|S_n(P) - L_n(P)\|_\infty | \{Z_i, W_{i,n}\}_{i=1}^n] = O_P(\frac{\sqrt{p\log(1+n)}M_{3,\Psi}}{n^{1/4}})$$

uniformly in $P \in \mathbf{P}$. By Fubini's theorem and Markov's inequality we may therefore conclude that unconditionally (on $\{Z_i, W_{i,n}\}_{i=1}^n$) and uniformly in $P \in \mathbf{P}$ we have

$$\|S_n(P) - L_n(P)\|_\infty = O_P(\frac{\sqrt{p\log(1+n)}M_{3,\Psi}}{n^{1/4}}).$$

Step 3: (Couple to Gaussian). We next proceed by coupling $L_n(P)$ to a (conditionally) Gaussian vector. To this end, recall the definition of $\xi_{i,n}(P)$ in (S.83) and let

$$\bar{G}_{i,n}(P) \sim N(0, \mathrm{Var}\{\xi_{i,n}(P) | \{Z_i, W_{i,n}\}_{i=1}^n\})$$

and $\{\bar{G}_{i,n}(P)\}_{i=1}^n$ be mutually independent conditional on $\{Z_i, W_{i,n}\}_{i=1}^n$. Then note that $\|a\|_2^2 \leq 2p\|a\|_\infty^2$ for any $a \in \mathbf{R}^{2p}$, Lemma S.2.8, and result (S.84) imply

$$\sum_{i=1}^n E[\|\bar{G}_{i,n}(P)\|_2^2 \|\bar{G}_{i,n}(P)\|_\infty | \{Z_i, W_{i,n}\}_{i=1}^n] \leq 2p \sum_{i=1}^n E[\|\bar{G}_{i,n}(P)\|_\infty^3 | \{Z_i, W_{i,n}\}_{i=1}^n]$$
$$\lesssim p\log^{3/2}(1+p)\frac{\hat{\sigma}_n^3}{n^{3/2}} \sum_{i=1}^n \|\varphi(Z_i, P) - \bar{\varphi}_n(P)\|_\infty^{3/2}. \quad \text{(S.89)}$$

Similarly, employing the definition of $\xi_{i,n}(P)$, $\{U_i\}_{i=1}^n$ being independent of $\{Z_i, W_{i,n}\}_{i=1}^n$, and τ_n being the empirical quantile function of $\{W_{i,n} - \bar{W}_n\}_{i=1}^n$, we obtain that

$$\sum_{i=1}^n E[\|\xi_{i,n}(P)\|_2^2 \|\xi_{i,n}(P)\|_\infty | \{Z_i, W_{i,n}\}_{i=1}^n]$$
$$\leq \frac{2p}{\sqrt{n}} (\frac{1}{n}\sum_{i=1}^n \|\varphi(Z_i, P) - \bar{\varphi}_n(P)\|_\infty^3)(\frac{1}{n}\sum_{i=1}^n |W_{i,n} - \bar{W}_n|^3). \quad \text{(S.90)}$$

Therefore, results (S.89) and (S.90), $\Psi(Z_i, P) = \|\varphi(Z_i, P)\|_\infty$, and multiple applications of the triangle and Jensen's inequalities yield the upper bound

$$\sum_{i=1}^{n} E[\|\bar{G}_{i,n}(P)\|_2^2\|\bar{G}_{i,n}(P)\|_\infty + \|\xi_{i,n}(P)\|_2^2\|\xi_{i,n}(P)\|_\infty|\{Z_i, W_{i,n}\}_{i=1}^n]$$
$$\lesssim \frac{p\log^{3/2}(1+p)}{\sqrt{n}}(\frac{1}{n}\sum_{i=1}^{n}|W_{i,n}|^3)(\frac{1}{n}\sum_{i=1}^{n}\{\Psi^3(Z_i,P)+\Psi^{3/2}(Z_i,P)\}) \equiv B_n(P), \quad \text{(S.91)}$$

where the final equality is definitional. Next, let $\mathcal{B}$ denote the Borel σ-field on $\mathbf{R}^{2p}$ and for any $A \in \mathcal{B}$ and $\epsilon > 0$ set $A^\epsilon \equiv \{a \in \mathbf{R}^{2p} : \inf_{\tilde{a}\in A}\|a - \tilde{a}\|_\infty \leq \epsilon\}$ – i.e., A^ϵ is an $\|\cdot\|_\infty$-enlargement of A. Strassen's Theorem (see Theorem 10.3.1 in Pollard (2002)), Lemma 39 in Belloni et al. (2019), and result (S.91) then establish for any $\delta > 0$ that

$$\sup_{A\in\mathcal{B}}\{P(L_n(P)\in A|\{Z_i,W_{i,n}\}_{i=1}^n) - P(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{G}_{i,n}(P)\in A^{3\delta}|\{Z_i,W_{i,n}\}_{i=1}^n)\}$$
$$\lesssim \min_{t\geq 0}(2P(\|\mathbb{Z}\|_\infty > t) + \frac{B_n(P)}{\delta^3}t^2) \quad \text{(S.92)}$$

where $\mathbb{Z} \in \mathbf{R}^{2p}$ is distributed according to $\mathbb{Z} \sim N(0, I_{2p})$. Furthermore, Proposition A.2.1 in van der Vaart and Wellner (1996) and Lemma S.2.8 imply for some $C < \infty$

$$\sup_{P\in\mathbf{P}} E_P[\min_{t\geq 0}(2P(\|\mathbb{Z}\|_\infty > t) + \frac{B_n(P)}{\delta^3}t^2)]$$
$$\lesssim \min_{t\geq 0}(\exp\{-\frac{t^2}{C\log(1+p)}\} + \sup_{P\in\mathbf{P}} E_P[B_n(P)]\frac{t^2}{\delta^3})$$
$$\lesssim \min_{t\geq 0}(\exp\{-\frac{t^2}{C\log(1+p)}\} + \frac{p\log^{3/2}(1+p)M_{3,\Psi}^3}{\sqrt{n}}\frac{t^2}{\delta^3}), \quad \text{(S.93)}$$

where the final inequality follows from (S.91), $E[|W_{i,n}|^3]$ being bounded uniformly in n by Assumption 4.5(iii), Jensen's inequality, $\sup_{P\in\mathbf{P}}\|\Psi\|_{P,3} \leq M_{3,\Psi}$ with $M_{3,\Psi} \geq 1$ by Assumption 4.3(iii), and $\{W_{i,n}\}_{i=1}^n$ being independent of $\{Z_i\}_{i=1}^n$ by Assumption 4.5(i). Hence, for any $K > 0$, $p\log^{5/2}(1+p)M_{3,\Psi}^3/\sqrt{n} \leq b_n^3$, (S.92), and (S.93) imply

$$\sup_{P\in\mathbf{P}} E_P[\sup_{A\in\mathcal{B}} E_P[1\{L_n(P)\in A\} - 1\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{G}_{i,n}(P)\in A^{3Kb_n}\}|\{Z_i,W_{i,n}\}_{i=1}^n]]$$
$$\lesssim \min_{t\geq 0}(\exp\{-\frac{t^2}{C\log(1+p)}\} + \frac{t^2}{K^3\log(1+p)}) \leq \exp\{-\frac{K^2}{C}\} + \frac{1}{K}, \quad \text{(S.94)}$$

where the final inequality follows by setting $t = K\sqrt{\log(1+p)}$. Theorem 4 in Monrad and Philipp (1991) and result (S.94) then imply that there exists a $\bar{\mathbb{G}}_n(P)$ such that

$$\|L_n(P) - \bar{\mathbb{G}}_n(P)\|_\infty = O_P(b_n)$$

uniformly in $P \in \mathbf{P}$, and its distribution conditional on $\{Z_i, W_{i,n}\}_{i=1}^n$ is given by

$$\bar{\mathbb{G}}_n(P) \sim N(0, \sum_{i=1}^n \text{Var}\{\xi_{i,n}(P)|\{Z_i, W_{i,n}\}_{i=1}^n\}). \tag{S.95}$$

Step 4: (Remove Dependence). We next couple $\bar{\mathbb{G}}_n(P)$ to a Gaussian vector $\tilde{\mathbb{G}}_n^\star(P)$ that is independent of $\{Z_i, W_{i,n}\}_{i=1}^n$. To this end, we first note result (S.84) implies

$$\hat{\Lambda}_n(P) \equiv \sum_{i=1}^n \text{Var}\{\xi_{i,n}(P)|\{Z_i, W_{i,n}\}_{i=1}^n\} = \frac{\hat{\sigma}_n^2}{n}\sum_{i=1}^n (\varphi(Z_i, P)\varphi(Z_i, P)' - \bar{\varphi}_n(P)\bar{\varphi}_n(P)').$$

Moreover, $E_P[\varphi(Z, P)] = 0$ and $\sup_{P\in\mathbf{P}} \max_{1\le j\le 2p} \|\varphi_j(\cdot, P)\|_{P,2}$ being bounded in n by definition of φ and Assumptions 4.3(i)(ii), and $\|aa'\|_{o,2} = \|a\|_2^2$ for any $a \in \mathbf{R}^{2p}$ imply

$$\sup_{P\in\mathbf{P}} E_P[\|\bar{\varphi}_n(P)\bar{\varphi}_n(P)'\|_{o,2}] = \sup_{P\in\mathbf{P}} E_P[\|\bar{\varphi}_n(P)\|_2^2]$$
$$= \sup_{P\in\mathbf{P}} \sum_{j=1}^{2p} E_P[(\frac{1}{n}\sum_{i=1}^n \varphi_j(Z_i, P))^2] \lesssim \frac{p}{n}. \tag{S.96}$$

Also, $\|\varphi(Z_i, P)\|_2^2 \le 2p\Psi^2(Z_i, P)$, Assumption 4.5(iv), and Jensen's inequality imply

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

Setting $\Lambda(P) \equiv E_P[\varphi(Z, P)\varphi(Z, P)']$, we then note that $b_n = o(1)$, Lemma S.2.9, $\|\Lambda(P)\|_{o,2}$ being uniformly bounded in n and $P \in \mathbf{P}$ by Assumption 4.3(ii) and definition of $\varphi(Z, P)$, and Markov's inequality allow us to conclude that

$$\|\frac{1}{n}\sum_{i=1}^n \varphi(Z_i, P)\varphi(Z_i, P)' - \Lambda(P)\|_{o,2} = O_P(\{\frac{p\log(1+p)n^{1/q}M_{q,\Psi^2}}{n}\}^{1/2}) \tag{S.97}$$

uniformly in $P \in \mathbf{P}$. Therefore, the triangle inequality, (S.96), (S.97), $\|\Lambda(P)\|_{o,2}$ being bounded in n and $P \in \mathbf{P}$ and Assumption 4.5(iii) yield

$$\|\hat{\Lambda}_n(P) - \Lambda(P)\|_{o,2} \le |\hat{\sigma}_n^2 - 1|\|\Lambda(P)\|_{o,2} + O_P(\{\frac{p\log(1+p)n^{1/q}M_{q,\Psi^2}}{n}\}^{1/2})$$
$$= O_P(\{\frac{p\log(1+p)n^{1/q}M_{q,\Psi^2}}{n}\}^{1/2})$$

uniformly in $P \in \mathbf{P}$. Hence, since the distribution of $\bar{\mathbb{G}}_n(P)$ conditional on $\{Z_i, W_{i,n}\}_{i=1}^n$ equals (S.95), we may apply Lemma S.2.7 with $V_n = \{Z_i, W_{i,n}\}_{i=1}^n$ to conclude that there exists a $\tilde{\mathbb{G}}_n^\star(P) \sim N(0,\Lambda(P))$ independent of $\{Z_i, W_{i,n}\}_{i=1}^n$ with

$$\|\bar{\mathbb{G}}_n(P) - \tilde{\mathbb{G}}_n^\star(P)\|_\infty = O_P((\frac{p\log^3(1+p)n^{1/q}M_{q,\Psi^2}}{n})^{1/4})$$

uniformly in $P \in \mathbf{P}$.

Step 5: (Couple $\hat{\mathbb{G}}_n$). Combining Steps 2, 3, and 4, we obtain that there exists a Gaussian vector $\tilde{\mathbb{G}}_n^\star(P)$ 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, in particular, $\tilde{\mathbb{G}}_n^\star(P)$ 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 $(\breve{\mathbb{G}}_n^{e\star}(P)', \breve{\mathbb{G}}_n^{i\star}(P)')' \equiv \breve{\mathbb{G}}_n^\star(P) \sim N(0,\Lambda(P))$ independent of $\{Z_i\}_{i=1}^n$ and such that

$$\|\frac{1}{\sqrt{n}}\sum_{i=1}^n (W_{i,n} - \bar{W}_n)\varphi(Z_i,P) - \breve{\mathbb{G}}_n^\star(P)\|_\infty = O_P(b_n) \tag{S.98}$$

uniformly in $P \in \mathbf{P}$. To conclude, set $\mathbb{G}_n^{j\star}(P) \equiv \Omega^j(P)\breve{\mathbb{G}}_n^{j\star}(P)$ for $j \in \{e,i\}$ and $\mathbb{G}_n^\star(P) \equiv (\mathbb{G}_n^{e\star}(P)', \mathbb{G}_n^{i\star}(P)')'$. Then note that since $\Omega^j(P)(\Omega^j(P))^\dagger\psi^j(Z,P) = \psi^j(Z,P)$ P-almost surely for $j \in \{e,i\}$ by Assumption 4.4(i), it follows from $\Lambda(P) \equiv E_P[\varphi(Z,P)\varphi(Z,P)']$ and the definition of $\varphi(Z,P)$ that $\mathbb{G}_n^\star(P) \sim N(0,\Sigma(P))$ as desired. Furthermore, since $\breve{\mathbb{G}}_n^\star(P)$ belongs to the range of $\Lambda(P)$ almost surely by Theorem 3.6.1 in Bogachev (1998), it follows that $\breve{\mathbb{G}}_n^{j\star}(P) = (\Omega^j(P))^\dagger\Omega^j(P)\breve{\mathbb{G}}_n^{j\star}(P) = (\Omega^j(P))^\dagger\mathbb{G}_n^{j\star}(P)$ for $j \in \{e,i\}$. The lemma thus follows from (S.98), the definition of $\varphi(Z,P)$, and Assumption 4.5(i). ■

Lemma S.2.6. *If Assumption 4.1 holds and $a_n/\sqrt{\log(1+p)} = o(1)$, then $\|(\hat{\Omega}_n^e)^\dagger(\hat{\Omega}_n^e - \Omega^e(P))\|_{o,\infty} \vee \|(\hat{\Omega}_n^i)^\dagger(\hat{\Omega}_n^i - \Omega^i(P))\|_{o,\infty} = O_P(a_n/\sqrt{\log(1+p)})$ uniformly in $P \in \mathbf{P}$.*

Proof: First note Assumption 4.1 and Lemma S.2.10 imply $\Omega^e(P)(\Omega^e(P))^\dagger\hat{\Omega}_n^e = \hat{\Omega}_n^e$ and $(\hat{\Omega}_n^e)^\dagger\hat{\Omega}_n^e(\Omega^e(P))^\dagger = (\Omega^e(P))^\dagger$ with probability tending to one uniformly in $P \in \mathbf{P}$. Since $\Omega^e(P)(\Omega^e(P))^\dagger\Omega^e(P) = \Omega^e(P)$ by Proposition 6.11.1(6) in Luenberger (1969), we thus obtain, with probability tending to one uniformly in $P \in \mathbf{P}$, that

$$\begin{aligned}
&\|(\hat{\Omega}_n^e)^\dagger(\hat{\Omega}_n^e - \Omega^e(P))\|_{o,\infty} \\
&\quad = \|(\hat{\Omega}_n^e)^\dagger\Omega^e(P)(\Omega^e(P))^\dagger(\hat{\Omega}_n^e - \Omega^e(P))\|_{o,\infty} \\
&\quad \le \|(\hat{\Omega}_n^e)^\dagger(\Omega^e(P) - \hat{\Omega}_n^e)\|_{o,\infty} \times o_P(1) + \|(\hat{\Omega}_n^e)^\dagger\hat{\Omega}_n^e(\Omega^e(P))^\dagger(\hat{\Omega}_n^e - \Omega^e(P))\|_{o,\infty} \\
&\quad = \|(\hat{\Omega}_n^e)^\dagger(\Omega^e(P) - \hat{\Omega}_n^e)\|_{o,\infty} \times o_P(1) + \|(\Omega^e(P))^\dagger(\hat{\Omega}_n^e - \Omega^e(P))\|_{o,\infty},
\end{aligned} \tag{S.99}$$

where the inequality holds due to $\|(\Omega^e(P))^\dagger(\hat{\Omega}_n^e - \Omega^e(P))\|_{o,\infty} = O_P(a_n/\sqrt{1+\log(p)})$ uniformly in $P \in \mathbf{P}$ by Assumption 4.1(ii) and $a_n/\sqrt{1+\log(p)} = o(1)$ by hypothesis. Since $\|(\Omega^e(P))^\dagger(\hat{\Omega}_n^e - \Omega^e(P))\|_{o,\infty} = O_P(a_n/\sqrt{1+\log(p)})$ uniformly in $P \in \mathbf{P}$ by Assumption 4.1(ii), result (S.99) implies $\|(\hat{\Omega}_n^e)^\dagger(\hat{\Omega}_n^e - \Omega^e(P))\|_{o,\infty} = O_P(a_n/\sqrt{\log(1+p)}$

uniformly in $P \in \mathbf{P}$. The claim $\|(\hat{\Omega}^{\mathrm{i}}_n)^{\dagger}(\hat{\Omega}^{\mathrm{i}}_n - \Omega^{\mathrm{i}}(P))\|_{o,\infty} = O_P(a_n/\sqrt{\log(1+p)})$ uniformly in $P \in \mathbf{P}$ can be established by identical arguments. ■

Lemma S.2.7. *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 Gaussian $\mathbb{G}^{\star}_n(P) \sim N(0, \Sigma_n(P))$ independent of V_n and satisfying $\|\bar{\mathbb{G}}_n(P) - \mathbb{G}^{\star}_n(P)\|_{\infty} = O_P(\sqrt{\log(1+d_n)\delta_n})$ 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 corresponding eigenvalues of $\hat{\Sigma}_n(P)$. Further letting $\mathcal{N}_{d_n}$ be independent of $(V_n, \bar{\mathbb{G}}_n(P))$ and distributed according to $\mathcal{N}_{d_n} \sim N(0, I_{d_n})$, we then define $\mathbb{Z}_n(P) \in \mathbf{R}^{d_n}$ to be given by

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

Since $\mathcal{N}_{d_n}$ is independent of V_n and $\bar{\mathbb{G}}_n(P)$ is Gaussian conditional on V_n it follows that $\mathbb{Z}_n(P)$ is Gaussian conditional on V_n as well. Moreover, we have

$$E[\mathbb{Z}_n(P)\mathbb{Z}_n(P)'|V_n] = \sum_{j=1}^{d_n} \hat{\nu}_j(P)\hat{\nu}_j(P)' = I_{d_n},$$

by direct calculation, and hence we conclude that $\mathbb{Z}_n(P) \sim N(0, I_{d_n})$ and is independent of V_n. Next, we note that Theorem 3.6.1 in Bogachev (1998) implies that $\bar{\mathbb{G}}_n(P)$ belongs to the range of $\hat{\Sigma}_n(P) : \mathbf{R}^{d_n} \to \mathbf{R}^{d_n}$ almost surely. Thus, since $\{\hat{\nu}_j(P) : \hat{\lambda}_j(P) \neq 0\}$ is an orthonormal basis for the range of $\hat{\Sigma}_n(P)$, we obtain that almost surely

$$\hat{\Sigma}_n^{1/2}(P)\mathbb{Z}_n(P) = \sum_{j:\hat{\lambda}_j(P)\neq 0} \hat{\nu}_j(P)(\hat{\nu}_j(P)'\bar{\mathbb{G}}_n(P)) = \bar{\mathbb{G}}_n(P). \tag{S.100}$$

Employing that $\mathbb{Z}_n(P)$ is independent of V_n, we then define the desired $\mathbb{G}^{\star}_n(P)$ by

$$\mathbb{G}^{\star}_n(P) \equiv \Sigma_n^{1/2}(P)\mathbb{Z}_n(P). \tag{S.101}$$

Next, set $\hat{\Delta}_n(P) \equiv \hat{\Sigma}_n^{1/2}(P) - \Sigma_n^{1/2}(P)$ and let $\hat{\Delta}_n^{(j,k)}(P)$ denote its (j,k) entry. Note (S.100), (S.101), Lemma S.2.8, and $\sup_{\|v\|_2=1}\langle v, a\rangle = \|a\|_2$ for any vector $a \in \mathbf{R}^{d_n}$ yield

$$\begin{aligned} E[\|\bar{\mathbb{G}}_n(P) - \mathbb{G}^{\star}_n(P)\|_{\infty}|V_n] &\lesssim \sqrt{\log(1+d_n)} \max_{1\le j\le d_n} (\sum_{k=1}^{d_n}(\hat{\Delta}_n^{(j,k)}(P))^2)^{1/2} \\ &= \sqrt{\log(1+d_n)} \sup_{\|v\|_2=1} \|\hat{\Delta}_n(P)v\|_{\infty} \le \sqrt{\log(1+d_n)}\|\hat{\Delta}_n(P)\|_{o,2}, \end{aligned} \tag{S.102}$$

where $\|\hat{\Delta}_n(P)\|_{o,2}$ denotes the operator norm of $\hat{\Delta}_n(P) : \mathbf{R}^{d_n} \to \mathbf{R}^{d_n}$ when $\mathbf{R}^{d_n}$ is endowed with the norm $\|\cdot\|_2$, and the final inequality follows from $\|\cdot\|_\infty \leq \|\cdot\|_2$. Moreover, Theorem X.1.1 in Bhatia (1997) further implies that

$$\|\hat{\Delta}_n(P)\|_{o,2}^2 \leq \|\hat{\Sigma}_n(P) - \Sigma(P)\|_{o,2} = O_P(\delta_n), \quad \text{(S.103)}$$

where the equality holds uniformly in $P \in \mathbf{P}$ by hypothesis. Therefore, Fubini's theorem, Markov's inequality, and result (S.102) allow us to conclude for any $C > 0$ that

$$\sup_{P\in\mathbf{P}} P(\|\bar{\mathbb{G}}_n(P) - \mathbb{G}_n^\star(P)\|_\infty > C^2\sqrt{\log(1+d_n)\delta_n} \text{ and } \|\hat{\Delta}_n(P)\|_{o,2} \leq C\sqrt{\delta_n})$$
$$\leq \sup_{P\in\mathbf{P}} E_P[\frac{\|\hat{\Delta}_n(P)\|_{o,2}}{C^2\sqrt{\delta_n}} 1\{\|\hat{\Delta}_n(P)\|_{o,2} \leq C\sqrt{\delta_n}\}] \leq \frac{1}{C}. \quad \text{(S.104)}$$

The claim of the lemma then follows from results (S.103) and (S.104). ■

Lemma S.2.8. *Let $Z = (Z_1, \ldots, Z_p) \in \mathbf{R}^p$ be jointly Gaussian with $E[Z_j] = 0$ and $E[Z_j^2] \leq \sigma^2$ for all $1 \leq j \leq p$. Then, there is a universal $K < \infty$ such that for any $q \geq 1$*

$$E[\|Z\|_\infty^q] \leq (\frac{q!\sqrt{\log(1+p)}\sigma K}{\sqrt{\log(2)}})^q.$$

Proof: The result is well known and stated here for completeness and ease of reference. Define the function $\psi_2 : \mathbf{R} \to \mathbf{R}$ to equal $\psi_2(u) = \exp\{u^2\} - 1$ for any $u \in \mathbf{R}$ and recall that for any random variable $V \in \mathbf{R}$ its Orlicz norm $\|V\|_{\psi_2}$ is given by

$$\|V\|_{\psi_2} \equiv \inf\{C > 0 : E[\psi(\frac{|V|}{C})] \leq 1\}.$$

Further note that for any $q \geq 1$ and random variable V we have $(E[|V|^q])^{1/q} \leq q!\|V\|_{\psi_2}/\sqrt{\log(2)}$; see, e.g., van der Vaart and Wellner (1996) pg. 95. Hence, Lemmas 2.2.1 and 2.2.2 in van der Vaart and Wellner (1996) imply that there exist finite constants K_0 and K_1 such that for all $q \geq 1$ it follows that

$$E[\|Z\|_\infty^q] \leq (\frac{q!}{\sqrt{\log(2)}})^q \| \max_{1\leq j\leq p} |Z_j| \|_{\psi_2}^q$$
$$\leq (\frac{q!}{\sqrt{\log(2)}})^q \{K_0\sqrt{\log(1+p)} \max_{1\leq j\leq p} \|Z_j\|_{\psi_2}\}^q \leq (\frac{q!\sqrt{\log(1+p)}\sigma K_1}{\sqrt{\log(2)}})^q,$$

for all $q \geq 1$. The claim of the lemma therefore follows. ■

Lemma S.2.9. *Let $\{V_i\}_{i=1}^n$ be an i.i.d. sample with $V_i \in \mathbf{R}^k$ and $\Sigma \equiv E[VV']$. Then:*

$$E[\|\frac{1}{n}\sum_{i=1}^n V_iV_i' - \Sigma\|_{o,2}] \leq \max\{\|\Sigma\|_{o,2}^{1/2}\delta, \delta^2\},$$

where $\delta \equiv D\sqrt{E[\max_{1\leq i\leq n} \|V_i\|_2^2]\log(1+k)/n}$ for some universal constant D.

Proof: This is essentially Theorem E.1 in Kato (2013) if $k \geq 2$. Suppose $k = 1$. Then by Lemma 2.3.1 in van der Vaart and Wellner (1996) it follows that

$$E[\|\frac{1}{n}\sum_{i=1}^{n} V_i V_i' - \Sigma\|_{o,2}] \leq 2E[|\frac{1}{n}\sum_{i=1}^{n} \epsilon_i V_i^2|], \tag{S.105}$$

where $\{\epsilon_i\}_{i=1}^n$ are i.i.d. Rademacher random variables that are independent of $\{V_i\}_{i=1}^n$. For $\|\cdot\|_{\psi_2}$ the Orlicz norm induced by $\psi_2(u) = \exp\{u^2\} - 1$, it then follows from $E[|U|] \leq \|U\|_{\psi_2}/\sqrt{\log(2)}$ for any random variable $U \in \mathbf{R}$ (see, e.g., van der Vaart and Wellner (1996) pg. 95) and Lemma 2.2.7 in van der Vaart and Wellner (1996) that

$$\begin{aligned} E[|\frac{1}{n}\sum_{i=1}^{n} \epsilon_i V_i^2|] &= E[E[|\frac{1}{n}\sum_{i=1}^{n} \epsilon_i V_i^2||\{V_i\}_{i=1}^n]] \\ &\leq \frac{\sqrt{6}}{\sqrt{\log(2)}} E[\{\sum_{i=1}^{n}(\frac{V_i^2}{n})^2\}^{1/2}] \leq \frac{\sqrt{6}}{\sqrt{\log(2)}} E[\max_{1\leq i\leq n}|V_i|\{\sum_{i=1}^{n}(\frac{V_i}{n})^2\}^{1/2}]. \end{aligned} \tag{S.106}$$

Therefore, the Cauchy-Schwarz's inequality and result (S.106) allow us to conclude that

$$E[|\frac{1}{n}\sum_{i=1}^{n} \epsilon_i V_i^2|] \leq \frac{\sqrt{6}}{\sqrt{\log(2)}}\{E[\max_{1\leq i\leq n}|V_i|^2]\}^{1/2}\{\frac{E[V^2]}{n}\}^{1/2},$$

which together with (S.105) establishes the claim of the lemma. ■

Lemma S.2.10. *Let Ω_1 and Ω_2 be $k \times k$ symmetric matrices such that* $\text{range}\{\Omega_1\} = \text{range}\{\Omega_2\}$. *It then follows that $\Omega_2\Omega_2^\dagger\Omega_1 = \Omega_1$ and $\Omega_2^\dagger\Omega_2\Omega_1^\dagger = \Omega_1^\dagger$.*

Proof: For any $k \times k$ matrix M, let $R(M) \subseteq \mathbf{R}^k$ and $N(M) \subseteq \mathbf{R}^k$ denote its range and null space. Also recall that any vector subspace $V \subseteq \mathbf{R}^k$ we set $V^\perp \equiv \{s \in \mathbf{R}^k : \langle s, v\rangle = 0 \text{ for all } v \in V\}$. In order to establish the first claim of the lemma, let $s_1 \in \mathbf{R}^k$ be arbitrary and observe that since $R(\Omega_1) = R(\Omega_2)$ it follows that there exists an $s_2 \in \mathbf{R}^k$ such that $\Omega_1 s_1 = \Omega_2 s_2$. Therefore, Proposition 6.11.1(6) in Luenberger (1969) yields

$$\Omega_2\Omega_2^\dagger\Omega_1 s_1 = \Omega_2\Omega_2^\dagger\Omega_2 s_2 = \Omega_2 s_2 = \Omega_1 s_1.$$

Hence, since $s_1 \in \mathbf{R}^k$ was arbitrary, it follows that $\Omega_2\Omega_2^\dagger\Omega_1 = \Omega_1$.

In order to establish the second claim of the lemma, first note that $R(M^\dagger) = N(M)^\perp$ for any $k \times k$ matrix M. Thus, since for $j \in \{1, 2\}$ we have $N(\Omega_j)^\perp = R(\Omega_j)$ due to $\Omega_j' = \Omega_j$ and Theorem 6.7.3(2) in Luenberger (1969), we can conclude that

$$R(\Omega_2^\dagger) = N(\Omega_2)^\perp = R(\Omega_2) = R(\Omega_1) = N^\perp(\Omega_1) = R(\Omega_1^\dagger),$$

where the third equality holds by assumption. Letting $s_1 \in \mathbf{R}^k$ be arbitrary, it then follows that there exists an $s_2 \in \mathbf{R}^k$ for which $\Omega_1^\dagger s_1 = \Omega_2^\dagger s_2$, and thus

$$\Omega_2^\dagger \Omega_2 \Omega_1^\dagger s_1 = \Omega_2^\dagger \Omega_2 \Omega_2^\dagger s_2 = \Omega_2^\dagger s_2 = \Omega_1^\dagger s_1,$$

where the second equality holds by Proposition 6.11.1(5) in Luenberger (1969). Since $s_1 \in \mathbf{R}^k$ was arbitrary, it follows that $\Omega_2^\dagger \Omega_2 \Omega_1^\dagger = \Omega_1^\dagger$. ■

Lemma S.2.11. *Let* $(\mathbb{Z}_1, \ldots, \mathbb{Z}_d)' \equiv \mathbb{Z} \in \mathbf{R}^d$ *be Gaussian with* $E[\mathbb{Z}_j] \geq 0$, $\text{Var}\{\mathbb{Z}_j\} = \sigma^2 > 0$ *for all* $1 \leq j \leq d$, *and define* $\mathbb{S} \equiv \max_{1 \leq j \leq d} \mathbb{Z}_j$ *and* $\text{m} \equiv \text{med}\{\mathbb{S}\}$. *Then, the distribution of* $\mathbb{S}$ *is absolutely continuous and its density is bounded on* $\mathbf{R}$ *by*

$$\frac{2}{\sigma} \max\{\frac{\text{m}}{\sigma}, 1\}.$$

Proof: The result immediately follows from results in Chapter 11 of Davydov et al. (1998). First, let F denote the c.d.f. of $\mathbb{S}$ and note that Theorem 11.2 in Davydov et al. (1998) implies that F is absolutely continuous with density F' satisfying

$$F'(r) = q(r) \exp\{-\frac{r^2}{2\sigma^2}\}, \tag{S.107}$$

where $q : \mathbf{R} \to \mathbf{R}_+$ is a nondecreasing function. Moreover, we can further conclude that

$$q(r) \int_r^\infty \exp\{-\frac{u^2}{2\sigma^2}\} du \leq \int_r^\infty q(u) \exp\{-\frac{u^2}{2\sigma^2}\} du = P(\mathbb{S} \geq r) \leq 1, \tag{S.108}$$

where the first inequality follows from $q : \mathbf{R} \to \mathbf{R}_+$ being nondecreasing and the equality follows from (S.107). Setting Φ and Φ' to denote the c.d.f. and density of a standard normal random variable respectively, then note that we may write

$$\int_r^\infty \exp\{-\frac{u^2}{2\sigma^2}\} du = \sqrt{2\pi} \int_r^\infty \Phi'(u/\sigma) du = \sqrt{2\pi}\sigma(1 - \Phi(r/\sigma)). \tag{S.109}$$

Therefore, we can combine results (S.107), (S.108), and (S.109) to obtain the bound

$$F'(r) \leq \frac{\exp\{-r^2/2\sigma^2\}}{\sqrt{2\pi}\sigma(1 - \Phi(r/\sigma))} = \frac{\Phi'(r/\sigma)}{\sigma(1 - \Phi(r/\sigma))} \leq \frac{2}{\sigma} \max\{\frac{r}{\sigma}, 1\}, \tag{S.110}$$

where the final result follows from 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)).

Next note that for any $\eta > 0$, the definitions of $\mathbb{S}$ and m, and the distribution of $\mathbb{S}$ first order stochastically dominating that of $\mathbb{Z}_j$ for any $1 \leq j \leq d$ imply that

$$P(\mathbb{S} \leq \text{m} + \eta) \geq P(\mathbb{S} \leq \max_{1 \leq j \leq p} \text{med}\{\mathbb{Z}_j\} + \eta) \geq P(\max_{1 \leq j \leq d} (\mathbb{Z}_j - E[\mathbb{Z}_j]) \leq \eta) > 0, \tag{S.111}$$

where the final inequality follows from $E[\mathbb{Z}]$ belonging to the support of $\mathbb{Z}$. Theorem

11.2 in Davydov et al. (1998) thus implies $q : \mathbf{R} \to \mathbf{R}_+$ is continuous at any $r > \mathrm{m}$, which together with (S.107) and the first fundamental theorem of calculus establishes F is in fact differentiable at any $r > \mathrm{m}$ with derivative given by F'. Setting $\Gamma \equiv \Phi^{-1} \circ F$, then observe $F = \Phi \circ \Gamma$ and hence at any $r > \mathrm{m}$ we obtain

$$F'(r) = \Phi'(\Gamma(r))\Gamma'(r) \tag{S.112}$$

for Γ' the derivative of Γ. However, note that 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 $[0, \infty)$ and $\Gamma(r) \in [0, \infty)$ for any $r > \mathrm{m}$. In particular, (S.112) implies that F' is decreasing on $(\mathrm{m}, +\infty)$ which together with (S.110) yields

$$\sup_{r\in(\mathrm{m},+\infty)} F'(r) = \limsup_{r\downarrow \mathrm{m}} F'(r) \leq \limsup_{r\downarrow \mathrm{m}} \frac{\Phi'(r/\sigma)}{\sigma(1-\Phi(r/\sigma))} = \frac{\Phi'(\mathrm{m}/\sigma)}{\sigma(1-\Phi(\mathrm{m}/\sigma))}. \tag{S.113}$$

Since result (S.110) implies $F'(r)$ is bounded by $2\max\{\mathrm{m}/\sigma, 1\}/\sigma$ on $(-\infty, \mathrm{m}]$ and result (S.113) implies the same bound applies on $(\mathrm{m}, +\infty)$, the claim of the lemma follows. ■

Lemma S.2.12. *Let $C \subseteq \mathbf{R}^k$ be a nonempty, closed, 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: The claim that $\mathcal{E} \neq \emptyset$ follows from Corollary 18.5.3 in Rockafellar (1970). Moreover, for $\mathcal{D}$ the set of extreme directions of C, Corollary 19.1.1 in Rockafellar (1970) implies both $\mathcal{E}$ and $\mathcal{D}$ are finite. Thus, writing $\mathcal{E} = \{a_j\}_{j=1}^m$ and $\mathcal{D} \equiv \{a_j\}_{j=m+1}^n$ (with $n = m$ when $\mathcal{D} = \emptyset$), Theorem 18.5 in Rockafellar (1970) yields the representation

$$C \equiv \{c \in \mathbf{R}^k : c = \sum_{j=1}^n a_j\lambda_j \text{ s.t. } \sum_{j=1}^m \lambda_j = 1 \text{ and } \lambda_j \geq 0 \text{ for all } 1 \leq j \leq n\}. \tag{S.114}$$

Next note that if $\sup_{c\in C}\langle c, y\rangle$ is finite, then Corollary 5.3.7 in Borwein and Lewis (2010) implies that the supremum is attained. Hence, by result (S.114) we obtain

$$\begin{aligned}\sup_{c\in C}\langle c, y\rangle &= \max_{\{\lambda_j\}_{j=1}^n} \langle y, \sum_{j=1}^n \lambda_j a_j\rangle \text{ s.t. } \sum_{j=1}^m \lambda_j = 1,\ \lambda_j \geq 0 \text{ for } 1 \leq j \leq n \\ &= \max_{\{\lambda_j\}_{j=1}^m} \langle y, \sum_{j=1}^m \lambda_j a_j\rangle \text{ s.t. } \sum_{j=1}^m \lambda_j = 1,\ \lambda_j \geq 0 \text{ for } 1 \leq j \leq m,\end{aligned} \tag{S.115}$$

where the second equality follows due to $\sup_{c\in C}\langle c, y\rangle$ being finite implying we must have $\langle y, a_j\rangle \leq 0$ for all $m+1 \leq j \leq n$. Since $\mathcal{E} = \{a_j\}_{j=1}^m$ and the maximization in (S.115) is solved by setting $\lambda_{j^\star} = 1$ for some $1 \leq j^\star \leq m$, the claim of the lemma follows. ■

Lemma S.2.13. *Let $\mathcal{V}^{\mathrm{i}}(P)$ be as defined in (20). Then, the set $(AA')^\dagger\mathcal{V}^{\mathrm{i}}(P)$ is nonempty, closed, polyehdral, contains no lines, and zero is one of its extreme points.*

Proof: First note that $0 \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ and therefore $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ is non-empty.

To show $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ is closed, suppose $\{v_j\}_{j=1}^{\infty} \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ and $\|v_j - v^{\star}\|_2 = o(1)$ for some $v^{\star} \in \mathbf{R}^p$. Since $v_j \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ it follows that there is an $s_j \in \mathcal{V}^{\mathrm{i}}(P)$ such that $v_j = (AA')^{\dagger}s_j$. Next, let $\tilde{s}_j \equiv AA^{\dagger}s_j$ and note that

$$(AA')^{\dagger}\tilde{s}_j = (AA')^{\dagger}AA^{\dagger}s_j = (A')^{\dagger}A^{\dagger}s_j = (AA')^{\dagger}s_j \tag{S.116}$$

since $(AA')^{\dagger}A = (A')^{\dagger}$ by Proposition 6.11.1(8) in Luenberger (1969) and $(A')^{\dagger}A^{\dagger} = (AA')^{\dagger}$ (see, e.g., Seber (2008) pg. 139). Moreover, note $A^{\dagger}\tilde{s}_j = A^{\dagger}AA^{\dagger}s_j = A^{\dagger}s_j$ by Proposition 6.11.1(5) in Luenberger (1969), while (S.116) implies $\|\Omega^{\mathrm{i}}(P)(AA')^{\dagger}\tilde{s}_j\|_1 = \|\Omega^{\mathrm{i}}(P)(AA')^{\dagger}s_j\|_1$. Hence, if $s_j \in \mathcal{V}^{\mathrm{i}}(P)$, then $\tilde{s}_j \in \mathcal{V}^{\mathrm{i}}(P)$, and by (S.116) we have $(AA')^{\dagger}\tilde{s}_j = v_j$. Furthermore, by construction $\tilde{s}_j \in R$ and hence $(AA')(AA')^{\dagger}\tilde{s}_j = \tilde{s}_j$, which together with $(AA')^{\dagger}\tilde{s}_j = v_j$ implies $\tilde{s}_j = AA'v_j$. By continuity, it then follows from $\|v_j - v^{\star}\|_2 = o(1)$ that $\|\tilde{s}_j - s^{\star}\|_2 = o(1)$ for $s^{\star} = AA'v^{\star}$ and thus $s^{\star} \in \mathcal{V}^{\mathrm{i}}(P)$ due to $\mathcal{V}^{\mathrm{i}}(P)$ being closed. Furthermore, $v_j = (AA')^{\dagger}\tilde{s}_j$ yields

$$\|v^{\star} - (AA')^{\dagger}s^{\star}\|_2 \le \lim_{n\to\infty} \|v_j - v^{\star}\|_2 + \|(AA')^{\dagger}(\tilde{s}_j - s^{\star})\|_2 = 0 \tag{S.117}$$

due to $\|v_j - v^{\star}\|_2 = o(1)$ and $\|\tilde{s}_j - s^{\star}\|_2 = o(1)$. Since, as argued, $s^{\star} \in \mathcal{V}^{\mathrm{i}}(P)$, we can conclude that $v^{\star} \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ and hence that $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ is closed as desired.

The fact that $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ is polyhedral is immediate from definition of $\mathcal{V}^{\mathrm{i}}(P)$, and thus we next show $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ contains no lines. To this end, suppose $v \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$, which implies $v = (AA')^{\dagger}s$ for some $s \in \mathcal{V}^{\mathrm{i}}(P)$. Since $A'(AA')^{\dagger} = A^{\dagger}$ by Proposition 6.11.1(9) in Luenberger (1969), we are able to conclude that

$$A'v = A'(AA')^{\dagger}s = A^{\dagger}s \le 0 \tag{S.118}$$

due to $s \in \mathcal{V}^{\mathrm{i}}(P)$. Similarly, if $-v \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$, then we must have $A'(-v) \le 0$ and thus $-v, v \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ imply that $A'v = 0$. However, for $N(A')^{\perp}$ the orthocomplement to the null space of A', note that $v = (AA')^{\dagger}s = (A')^{\dagger}A^{\dagger}s$ implies that

$$v \in N(A')^{\perp}. \tag{S.119}$$

Since $v \in N(A')^{\perp}$ and $A'v = 0$ imply $v = 0$, it follows that if $-v, v \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$, then $v = 0$ and hence $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ contains no lines as claimed.

Finally, to see that zero is an extreme point of $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ suppose that $0 = \lambda v_1 + (1-\lambda)v_2$ for some $v_1, v_2 \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ and $\lambda \in (0,1)$. By result (S.118) holding for any $v \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$, $\lambda \in (0,1)$, and $0 = A'0 = A'(\lambda v_1 + (1-\lambda)v_2)$, it then follows that $A'v_1 = A'v_2 = 0$. Therefore, result (S.119) holding for any $v \in (AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$ implies $v_1 = v_2 = 0$, which verifies that zero is indeed an extreme point of $(AA')^{\dagger}\mathcal{V}^{\mathrm{i}}(P)$. ■

S.3 Computational Details

In this appendix, we provide details on how we compute our test statistic, T_n, defined in (19), the restricted estimator $\hat{\beta}_n^{\mathrm{r}}$, defined in (24), and obtain a critical value. One computational theme that we found important in our simulations is that the pseudoinverse $A^\dagger$ can be poorly conditioned. As we show below, however, it is possible to implement our procedure without ever needing to compute $A^\dagger$ explicitly.

First, we need to select a specific estimator $\hat{x}_n^\star$. In the mixed logit simulation in Section 5, the parameter $\beta(P)$ can be decomposed into $\beta(P) = (\beta_{\mathrm{u}}(P)', \beta_{\mathrm{k}}')'$, where $\beta_{\mathrm{u}}(P) \in \mathbf{R}^{p_{\mathrm{u}}}$ and $\beta_{\mathrm{k}} \in \mathbf{R}^{p_{\mathrm{k}}}$ is a known constant for all $P \in \mathbf{P}_0$. Similarly, we decompose any $b \in \mathbf{R}^p$ into $b = (b_{\mathrm{u}}', b_{\mathrm{k}}')'$ with $b_{\mathrm{u}} \in \mathbf{R}^{p_{\mathrm{u}}}$ and $b_{\mathrm{k}} \in \mathbf{R}^{p_{\mathrm{k}}}$, and partition the matrix A into the corresponding submatrices A_{u} (dimension $p_{\mathrm{u}} \times k$) and A_{k} (dimension $p_{\mathrm{k}} \times k$). In our simulations, we then set $\hat{x}_n^\star$ to be a solution to the quadratic program

$$\min_{x \in \mathbf{R}^d} \left(\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}} x\right)' \hat{\Xi}_n^{-1} \left(\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}} x\right) \quad \text{s.t.} \quad A_{\mathrm{k}} x = \beta_{\mathrm{k}}, \tag{S.120}$$

where $\hat{\beta}_n = (\hat{\beta}_{\mathrm{u},n}', \beta_{\mathrm{k}})'$ and $\hat{\Xi}_n$ is an estimate of the asymptotic variance matrix of $\hat{\beta}_{\mathrm{u},n}$. While the solution to (S.120) may not be unique, we note that any two minimizers x_1 and x_2 of (S.120) must satisfy $Ax_1 = Ax_2$. Since in our reformulations below $\hat{x}_n^\star$ only enters through $A\hat{x}_n^\star$, the specific choice of minimizer in (S.120) is immaterial.

Throughout, we let $\hat{\Omega}_n^{\mathrm{e}}$ be the sample standard deviation matrix of the entire vector $\hat{\beta}_n$. Note that, since $\hat{\beta}_n = (\hat{\beta}_{\mathrm{u},n}', \beta_{\mathrm{k}})'$ and β_{k} is non-stochastic, $\hat{\Omega}_n^{\mathrm{e}}$ has the form

$$\hat{\Omega}_n^{\mathrm{e}} = \begin{bmatrix} \hat{\Xi}_n^{1/2} & 0 \\ 0 & 0 \end{bmatrix}. \tag{S.121}$$

We further let $\hat{\Omega}_n^{\mathrm{i}}$ be the sample standard deviation of $A\hat{x}_n^\star$, although this choice of studentization plays no special computational role in what follows.

Next, consider the first component of T_n (see (19)), which we reproduce here as

$$T_n^{\mathrm{e}} \equiv \sup_{s \in \hat{\mathcal{V}}_n^{\mathrm{e}}} \sqrt{n} \langle s, \hat{\beta}_n - A\hat{x}_n^\star \rangle \quad \text{where} \quad \hat{\mathcal{V}}_n^{\mathrm{e}} \equiv \{s \in \mathbf{R}^p : \|\hat{\Omega}_n^{\mathrm{e}} s\|_1 \le 1\}. \tag{S.122}$$

As in the main text, the superscript "e" alludes to the relation to the "equality" condition in Theorem 3.1 – i.e. this statistic is designed to detect violations of the requirement that $\beta(P)$ is an element of the range of A. As noted in the main text, $\hat{\beta}_n = A\hat{x}_n^\star$ and hence $T_n^{\mathrm{e}} = 0$ whenever A is full rank and $d \ge p$. In other cases, we use the fact that $\hat{x}_n^\star$, as the solution to (S.120), must satisfy $A_{\mathrm{k}} \hat{x}_n^\star = \beta_{\mathrm{k}}$, and that our choice of $\hat{\Omega}_n^{\mathrm{e}}$ in (S.121) has $\hat{\Xi}_n^{1/2}$ as its upper left block. From these observations, we deduce that

$$\begin{aligned} T_n^{\mathrm{e}} &= \sup_{s_{\mathrm{u}}\in\mathbf{R}^{p_{\mathrm{u}}}} \sqrt{n}\langle s_{\mathrm{u}}, \hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}}\hat{x}_n^\star\rangle \quad \text{s.t.} \quad \|\hat{\Xi}_n^{1/2}s_{\mathrm{u}}\|_1 \le 1 \\ &= \|\sqrt{n}\hat{\Xi}_n^{1/2}\left(\hat{\beta}_{\mathrm{u},n} - A_{\mathrm{u}}\hat{x}_n^\star\right)\|_\infty. \end{aligned} \tag{S.123}$$

Thus, T_n^{e} can be computed by simply taking the maximum of a vector of length p_{u}.

The second component of T_n, defined in (19), is reproduced here as

$$T_n^{\mathrm{i}} \equiv \sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star\rangle \quad \text{where} \quad \hat{\mathcal{V}}_n^{\mathrm{i}} \equiv \{s\in\mathbf{R}^p : A^\dagger s \le 0 \text{ and } \|\hat{\Omega}_n^{\mathrm{i}}(AA')^\dagger s\|_1 \le 1\}, \tag{S.124}$$

where the superscript "i" alludes to the relation to the "inequality condition in Theorem 3.1 – i.e. this statistic is designed to detect violations of the requirement that a positive solution to $Ax = \beta(P)$ exists. To compute T_n^{i} without explicitly using $A^\dagger$, we first note

$$A^\dagger = A'(AA')^\dagger, \tag{S.125}$$

see, e.g., Proposition 6.11.1(9) in Luenberger (1969). Then, we observe that

$$\text{range}\{(AA')^\dagger\} = \text{null}\{AA'\}^\perp = \text{range}\{AA'\} = \text{range}\{A\}. \tag{S.126}$$

The first equality in (S.126) is a property of pseudoinverses, see, e.g., Luenberger (1969, pg. 164). The second equality is a standard result in linear algebra, see, e.g., Theorem 6.6.1 in Luenberger (1969). This result is also used in the third equality, which uses the following logic: if $t = As$ for some $s\in\mathbf{R}^p$, then also $t = As_1$, where $s_1 \in \text{null}\{A\}^\perp = \text{range}\{A'\}$ is determined from the orthogonal decomposition $s = s_0 + s_1$ with $s_0 \in \text{null}\{A\}$, and hence $t \in \text{range}\{AA'\}$ implying $\text{range}\{A\} \subseteq \text{range}\{AA'\}$. Since trivially $\text{range}\{AA'\} \subseteq \text{range}\{A\}$ the third equality follows. We thus obtain that

$$\begin{aligned} T_n^{\mathrm{i}} &= \sup_{s\in\mathbf{R}^p} \sqrt{n}\langle A'(AA')^\dagger, \hat{x}_n^\star\rangle \quad \text{s.t.} \quad A'(AA')^\dagger s \le 0 \quad \text{and} \quad \|\hat{\Omega}_n^{\mathrm{i}}(AA')^\dagger s\|_1 \le 1, \\ &= \sup_{x\in\mathbf{R}^d} \sqrt{n}\langle A'Ax, \hat{x}_n^\star\rangle \quad \text{s.t.} \quad A'Ax \le 0 \quad \text{and} \quad \|\hat{\Omega}_n^{\mathrm{i}}Ax\|_1 \le 1, \\ &= \sup_{x\in\mathbf{R}^d, s\in\mathbf{R}^p} \sqrt{n}\langle s, A\hat{x}_n^\star\rangle \quad \text{s.t.} \quad Ax = s, \quad A's \le 0 \quad \text{and} \quad \|\hat{\Omega}_n^{\mathrm{i}}s\|_1 \le 1, \end{aligned} \tag{S.127}$$

where the first equality follows from (S.125), the second from (S.126), and in the third we substituted $s = Ax$. The final program in (S.127) can be written explicitly as a linear program by introducing non-negative slack variables, so that

$$T_n^{\mathrm{i}} = \sup_{x\in\mathbf{R}^d, s\in\mathbf{R}^p, \phi^+\in\mathbf{R}^p_+, \phi^-\in\mathbf{R}^p_+} \sqrt{n}\langle s, A\hat{x}_n^\star\rangle$$

$$\text{s.t.} \quad Ax = s,\ A's \le 0,\ \langle \mathbf{1}_p, \phi^+\rangle + \langle \mathbf{1}_p, \phi^-\rangle \le 1,\ \phi^+ - \phi^- = \hat{\Omega}_n^{\mathrm{i}} s, \quad \text{(S.128)}$$

where $\mathbf{1}_p \in \mathbf{R}^p$ is the vector with all coordinates equal to one. Note that if $d \ge p$ and A has full rank, then the constraint $Ax = s$ is redundant since Ax ranges across all of $\mathbf{R}^p$ as x varies across $\mathbf{R}^d$. In these cases, the constraint $Ax = s$ together with the variable x can be entirely removed from the linear program in (S.128). Taking the maximum of (S.123) and (S.128) yields our test statistic T_n.

Turning to our bootstrap procedure, we first show how to solve (24) to find $\hat{\beta}_n^{\mathrm{r}}$. The optimization problem to solve is here reproduced as:

$$\min_{x\in\mathbf{R}^d_+, b=(b_{\mathrm{u}}', b_{\mathrm{k}}')'} \left[\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \sqrt{n}\langle A^\dagger s, \hat{x}_n^\star - A^\dagger b\rangle\right] \text{ s.t. } b_{\mathrm{k}} = \beta_{\mathrm{k}},\ Ax = b. \quad \text{(S.129)}$$

We first observe that the inner problem has the same structure as (S.124), but with $\hat{x}_n^\star$ replaced by $\hat{x}_n^\star - A^\dagger b$, where b is a fixed variable of optimization from the outer problem. Applying the same logic employed in (S.127) to this inner problem yields

$$\sup_{x\in\mathbf{R}^d} \sqrt{n}\langle A'Ax, \hat{x}_n^\star - A^\dagger b\rangle \quad \text{s.t.} \quad A'Ax \le 0 \quad \text{and} \quad \|\hat{\Omega}_n^{\mathrm{i}} Ax\|_1 \le 1. \quad \text{(S.130)}$$

Introducing slack variables as in (S.128) turns (S.130) into a linear program. The dual of the resulting linear program can be shown to be given by

$$\inf_{\phi_1\in\mathbf{R}_+, \phi_p\in\mathbf{R}^p, \phi_d\in\mathbf{R}^d_+} \phi_1 \quad \text{s.t.} \quad \mathbf{1}_p\phi_1 - \phi_p \ge 0,\ \mathbf{1}_p\phi_1 + \phi_p \ge 0$$

$$- A'\hat{\Omega}_n^{\mathrm{i}}\phi_p + A'A\phi_d = \sqrt{n}A'A(\hat{x}_n^\star - A^\dagger b). \quad \text{(S.131)}$$

Next, let $V \equiv \mathrm{range}\{AA'\}$ and note that since $A^\dagger = A'(AA')^\dagger$ by Proposition 6.11.1(8) in Luenberger (1969), it follows that $A'AA^\dagger b = A'AA'(AA')^\dagger b = A'\Pi_V b$. However, by (S.126), $V \equiv \mathrm{range}\{AA'\} = \mathrm{range}\{A\} = \mathrm{null}\{A'\}^\perp$, where the final equality follows by Theorem 6.6.1 in Luenberger (1969). Hence, $A'\Pi_V b = A'b$ and (S.131) equals

$$\inf_{\phi_1\in\mathbf{R}_+, \phi_p\in\mathbf{R}^p, \phi_d\in\mathbf{R}^d_+} \phi_1 \quad \text{s.t.} \quad \mathbf{1}_p\phi_1 - \phi_p \ge 0,\ \mathbf{1}_p\phi_1 + \phi_p \ge 0$$

$$- A'\hat{\Omega}_n^{\mathrm{i}}\phi_p + A'A\phi_d = \sqrt{n}A'(A\hat{x}_n^\star - b). \quad \text{(S.132)}$$

Substituting (S.132) back into the inner problem in (S.129) then yields a single linear program that determines $\hat{\beta}_n^{\mathrm{r}}$. Given $\hat{\beta}_n^{\mathrm{r}}$ it is then straightforward to compute our bootstrap statistic. For instance, in the simulations in Section 5, we let

$$\hat{\mathbb{G}}_n^{\mathrm{e}} = \sqrt{n}\{(\hat{\beta}_{b,n} - A\hat{x}_{b,n}^\star) - (\hat{\beta}_n - A\hat{x}_n^\star)\}$$
$$\hat{\mathbb{G}}_n^{\mathrm{i}} = \sqrt{n}A(\hat{x}_{b,n}^\star - \hat{x}_n^\star)$$

where $\hat{\beta}_{b,n}$ and $\hat{x}_{b,n}^\star$ are nonparametric bootstrap analogues to $\hat{\beta}_n$ and $\hat{x}_n^\star$. Arguing as in result (S.123) it is then straightforward to show that

$$\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{e}}} \langle s, \hat{\mathbb{G}}_n^{\mathrm{e}}\rangle = \|\sqrt{n}\hat{\Xi}_n^{1/2}\hat{\mathbb{G}}_n^{\mathrm{e}}\|_\infty \tag{S.133}$$

In analogy to (S.123), we note that (S.133) equals zero whenever A is full rank and $d \geq p$. Next, we may employ the same arguments as in (S.127) and (S.128) and noting $AA^\dagger\hat{\mathbb{G}}_n^{\mathrm{i}} = \hat{\mathbb{G}}_n^{\mathrm{i}}$ due to $AA^\dagger A = A$ by Proposition 6.11.1(6) in Luenberger (1969) to obtain

$$\begin{aligned}
&\sup_{s\in\hat{\mathcal{V}}_n^{\mathrm{i}}} \langle A^\dagger s, A^\dagger(\hat{\mathbb{G}}_n^{\mathrm{i}} + \sqrt{n}\lambda_n\hat{\beta}_n^{\mathrm{r}})\rangle \\
&\qquad = \sup_{x\in\mathbf{R}^d, s\in\mathbf{R}^p, \phi^+\in\mathbf{R}_+^p, \phi^-\in\mathbf{R}_+^p} \langle s, \hat{\mathbb{G}}_n^{\mathrm{i}} + \sqrt{n}\lambda_n\hat{\beta}_n^{\mathrm{r}}\rangle \\
&\qquad\quad \text{s.t. } Ax = s,\ A's \leq 0,\ \langle \mathbf{1}_p, \phi^+\rangle + \langle \mathbf{1}_p, \phi^-\rangle \leq 1,\ \phi^+ - \phi^- = \hat{\Omega}_n^{\mathrm{i}} s.
\end{aligned} \tag{S.134}$$

As in (S.128), we note that if A is full rank and $d \geq p$, then the constraint $Ax = s$ and the variable x may be dropped from the linear program in (S.134). The critical value is then obtained by computing the $1-\alpha$ quantile of the maximum of (S.133) and (S.134) across bootstrap iterations. Finally, we note the same arguments also shows that the problem (36) used to determine λ_n^{b} is equivalent to (S.128) with $A\hat{x}_n^\star$ replaced by $\hat{\mathbb{G}}_n^{\mathrm{i}}$.

References

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

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, New York.

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

Borwein, J. and Lewis, A. S. (2010). *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media.

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

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

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.

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

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

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

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.

Pollard, D. (2002). *A user's guide to measure theoretic probability*, vol. 8. Cambridge University Press.

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

Seber, G. A. (2008). *A matrix handbook for statisticians*, vol. 15. John Wiley & Sons.

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

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