Supplemental Material to
INFERENCE BASED ON MANY CONDITIONAL MOMENT INEQUALITIES

By

Donald W. K. Andrews and Xiaoxia Shi

July 2015
Revised April 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2010R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles. yale.edu/

This Appendix is not to be published. It will be made available on the web.

Appendix

to

Inference Based on Many Conditional Moment Inequalities

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Xiaoxia Shi
Department of Economics
University of Wisconsin, Madison

June 2010
Revised: April 26, 2016

A Outline

This Appendix provides proofs of Theorems 5.1 and 6.1 of Andrews and Shi (2010) "Inference Based on Many Conditional Moment Inequalities," referred to hereafter as ASM. In fact, the results given here cover a much broader class of test statistics than is considered in ASM. We let AS1 abbreviate Andrews and Shi (2013a) and AS2 abbreviate Andrews and Shi (2013b).

This Appendix is organized as follows. Section B defines the class of test statistics that are considered. This class includes the statistics that are considered in ASM. Section B also provides the definition of manageability that is used in Assumption PS2. Section C introduces the critical values, the confidence sets (CS's), and the tests. Section D establishes the correct asymptotic size of the CS's. Theorem 5.1 of ASM is a corollary to Lemmas D.1 and D.2, which are given in Section D. Section E establishes that the CS's contain fixed parameter values outside the identified set with probability that goes to zero. Equivalently, the tests upon which the CS's are constructed are shown to be consistent tests. Theorem 6.1 of ASM is a corollary to Theorem E.1, which is given in Section E. Section F provides proofs of Lemma 7.1-8.2 of ASM, which verify Assumptions PS1, PS2, SIG1, and SIG2 in the examples given in ASM. Section G provides additional Monte Carlo simulation results for the two simulation examples considered in ASM. These results are designed to analyze the robustness of the tests and CS's to the tuning parameters that are used.

B General Form of the Test Statistic

B.1 Test Statistic

Here we define the general form of the test statistic $T_n(\theta)$ that is used to construct a CS. We transform the conditional moment inequalities/equalities given X_i into equivalent unconditional moment inequalities/equalities by choosing appropriate weighting functions of X_i, i.e., X_i instruments. Then, we construct a test statistic based on the instrumented moment conditions.

The instrumented moment conditions are of the form:

$$
\begin{aligned}
&E_{F_0}[m_j(W_i,\theta_0,\tau)\,g_j(X_i)] \geq 0 \text{ for } j = 1,...,p \text{ and} \\
&E_{F_0}[m_j(W_i,\theta_0,\tau)\,g_j(X_i)] = 0 \text{ for } j = p+1,...,k, \text{ for } g = (g_1,...,g_k)' \in \mathcal{G} \text{ and } \tau \in \mathcal{T},
\end{aligned} \tag{B.1}
$$

where θ_0 and F_0 are the true parameter and distribution, respectively, g is the instrument vector that depends on the conditioning variables X_i, and $\mathcal{G}$ is a collection of instruments. Typically $\mathcal{G}$ contains an infinite number of elements.

The identified set $\Theta_{F_0}(\mathcal{G})$ of the model defined by (B.1) is

$$
\Theta_{F_0}(\mathcal{G}) := \{\theta \in \Theta : \text{(B.1) holds with } \theta \text{ in place of } \theta_0\}. \tag{B.2}
$$

The collection $\mathcal{G}$ is chosen so that $\Theta_{F_0}(\mathcal{G}) = \Theta_{F_0}$, where Θ_{F_0} is the identified set based on the conditional moment inequalities and equalities defined in (2.2) of ASM. Section B.4 provides conditions for this equality and shows that the instruments defined in (3.6) of ASM satisfy the conditions. Additional sets $\mathcal{G}$ are given in AS1 and AS2.

We construct test statistics based on (B.1). The sample moment functions are defined in (3.2) in ASM. The sample variance-covariance matrix of $n^{1/2}\overline{m}_n(\theta,\tau,g)$ is defined in (3.3) in ASM. The matrix $\widehat{\Sigma}_n(\theta,\tau,g)$ may be singular with non-negligible probability for some $g \in \mathcal{G}$. This is undesirable because the inverse of $\widehat{\Sigma}_n(\theta,\tau,g)$ needs to be consistent for its population counterpart uniformly over $g \in \mathcal{G}$ for the test statistics considered below. Thus, we employ a modification of $\widehat{\Sigma}_n(\theta,\tau,g)$, denoted by $\overline{\Sigma}_n(\theta,\tau,g)$ and defined in (3.4) in ASM, such that the smallest eigenvalue of $\overline{\Sigma}_n(\theta,\tau,g)$ is bounded away from zero.

The test statistic $T_n(\theta)$ is either a Cramér-von-Mises-type (CvM) or a Kolmogorov-Smirnov-type (KS) statistic. The CvM statistic is

$$
T_n(\theta) := \sup_{\tau \in \mathcal{T}} \int_{\mathcal{G}} S(n^{1/2}\overline{m}_n(\theta,\tau,g), \overline{\Sigma}_n(\theta,\tau,g))dQ(g), \tag{B.3}
$$

where S is a non-negative function and Q is a weight function (i.e., probability measure) on $\mathcal{G}$. The functions S and Q are discussed in Sections B.2 and B.5 below, respectively.

The Kolmogorov-Smirnov-type (KS) statistic is

$$
T_n(\theta) := \sup_{\tau \in \mathcal{T}} \sup_{g \in \mathcal{G}} S(n^{1/2}\overline{m}_n(\theta,\tau,g), \overline{\Sigma}_n(\theta,\tau,g)). \tag{B.4}
$$

For brevity, the discussion in this Appendix focusses on CvM statistics and all results stated concern CvM statistics. Similar results hold for KS statistics. Such results can be established by extending the results given in Section 13.1 of Appendix B of AS2 and proved in Section 15.1 of Appendix D of AS2.

B.2 S Function Assumptions

Let $m_I := (m_1, ..., m_p)'$ and $m_{II} := (m_{p+1}, ..., m_k)'$. Let Δ be the set of $k \times k$ positive-definite diagonal matrices. Let $\mathcal{W}$ be the set of $k \times k$ positive definite matrices.

Assumption S1. $\forall (m, \Sigma) \in \{(m, \Sigma) : m \in (-\infty, \infty]^p \times R^v, \Sigma \in \mathcal{W}\}$,

(a) $S(Dm, D\Sigma D) = S(m, \Sigma)\ \forall D \in \Delta$,

(b) $S(m_I, m_{II}, \Sigma)$ is non-increasing in each element of m_I,

(c) $S(m, \Sigma) \geq 0$,

(d) S is continuous, and

(e) $S(m, \Sigma + \Sigma_1) \leq S(m, \Sigma)$ for all $k \times k$ positive semi-definite matrices Σ_1.

It is worth pointing out that Assumption S1(d) requires S to be continuous in m at all points m in the extended vector space $(-\infty, \infty]^p \times R^v$, not only for points in R^{p+v}.

Let $\mathcal{M}$ denote a bounded subset of R^k. Let $\mathcal{W}_{cpt}$ denote a compact subset of $\mathcal{W}$.

Assumption S2. $S(m, \Sigma)$ is uniformly continuous in the sense that

$$\lim_{\delta \downarrow 0} \sup_{\mu \in R_+^p \times \{0\}^v} \sup_{\substack{m, m^* \in \mathcal{M} \\ \|m - m^*\| \leq \delta}} \sup_{\substack{\Sigma, \Sigma^* \in \mathcal{W}_{cpt} \\ \|\Sigma - \Sigma^*\| \leq \delta}} |S(m + \mu, \Sigma) - S(m^* + \mu, \Sigma^*)| = 0.[15]$$

Assumption S3. $S(m, \Sigma) > 0$ if and only if $m_j < 0$ for some $j = 1, ..., k$, where $m = (m_1, ..., m_k)'$ and $\Sigma \in \mathcal{W}$.

Assumption S4. For some $\chi > 0$, $S(am, \Sigma) = a^\chi S(m, \Sigma)$ for all scalar $a > 0$, $m \in R^k$, and $\Sigma \in \mathcal{W}$.

[15] It is important that the supremum is only over μ vectors with non-negative elements μ_j for $j \leq p$. Without this restriction on the μ vectors, Assumption S2 would not hold for typical S functions of interest. Also note that Assumption S2 here is Assumption S2$'$, rather than Assumption S2, in AS1. Although Assumption S2 in AS1 is seemingly weaker than Assumption S2$'$, the former implies the latter, i.e. the two assumptions are equivalent. The equivalence can be established by adapting the proof of the well-known result that continuous functions defined on compact sets are uniformly continuous.

It is shown in Lemma 1 of AS1 that the functions S_1-S_3 in (3.9) satisfy Assumptions S1-S4. The function S_4 also does by similar arguments.

B.3 Definition of Manageability

Here we introduce the concept of manageability from Pollard (1990) that is used in Assumption PS2 in ASM and Assumption M that is introduced in the following section. This condition is used to regulate the complexity of $\mathcal{T} \times \mathcal{G}$. It ensures that $\{n^{1/2}(\overline{m}_n(\theta,\tau,g) - E_{F_n}\overline{m}_n(\theta,\tau,g)) : (\tau,g) \in \mathcal{T} \times \mathcal{G}\}$ satisfies a functional central limit theorem (FCLT) under drifting sequences of distributions $\{F_n : n \geq 1\}$. The latter is utilized in the proof of the uniform coverage probability results for the CS's. See Pollard (1990) and Appendix E of AS2 for more about manageability.

Definition (Pollard, 1990, Definition 3.3). The *packing number* $D(\xi,\rho,V)$ for a subset V of a metric space $(\mathcal{V},\rho)$ is defined as the largest b for which there exist points $v^{(1)},...,v^{(b)}$ in V such that $\rho(v^{(s)},v^{(s')}) > \xi$ for all $s \neq s'$. The *covering number* $N(\xi,\rho,V)$ is defined to be the smallest number of closed balls with ρ-radius ξ whose union covers V.

It is easy to see that $N(\xi,\rho,V) \leq D(\xi,\rho,V) \leq N(\xi/2,\rho,V)$.

Let $(\Omega,\mathcal{F},\mathbf{P})$ be the underlying probability space equipped with probability distribution $\mathbf{P}$. Let $\{f_{n,i}(\cdot,\tau) : \Omega \to R : \tau \in \mathcal{T}, i \leq n, n \geq 1\}$ be a triangular array of random processes. Let

$$\mathcal{F}_{n,\omega} := \{(f_{n,1}(\omega,\tau),...,f_{n,n}(\omega,\tau))' : \tau \in \mathcal{T}\}. \tag{B.5}$$

Because $\mathcal{F}_{n,\omega} \subset R^n$, we use the Euclidean metric $\|\cdot\|$ on this space. For simplicity, we omit the metric argument in the packing number function, i.e., we write $D(\xi,V)$ in place of $D(\xi,\|\cdot\|,V)$ when $V \subset \mathcal{F}_{n,\omega}$.

Let $\odot$ denote the element-by-element product. For example for $a,b \in R^n$, $a \odot b = (a_1b_1,...,a_nb_n)'$. Let *envelope functions* of a triangular array of processes $\{f_{n,i}(\omega,\tau) : \tau \in \mathcal{T}, i \leq n, n \geq 1\}$ be an array of functions $\{F_n(\omega) = (F_{n,1}(\omega),...,F_{n,n}(\omega))' : n \geq 1\}$ such that $|f_{n,i}(\omega,\tau)| \leq F_{n,i}(\omega)$ $\forall i \leq n, n \geq 1, \tau \in \mathcal{T}, \omega \in \Omega$.

Definition (Pollard, 1990, Definition 7.9). A triangular array of processes $\{f_{n,i}(\omega,\tau) : \tau \in \mathcal{T}, i \leq n, n \geq 1\}$ is said to be *manageable* with respect to the envelopes $\{F_n(\omega) : n \geq 1\}$ if there exists a deterministic real function λ on $(0,1]$ for which (i) $\int_0^1 \sqrt{\log \lambda(\xi)}d\xi < \infty$ and

(ii) $D(\xi \| \alpha \odot F_n(\omega)\|, \alpha \odot \mathcal{F}_{n,\omega}) \leq \lambda(\xi)$ for $0 < \xi \leq 1$, all $\omega \in \Omega$, all n-vectors α of nonnegative weights, and all $n \geq 1$.

B.4 X Instruments

The collection of instruments $\mathcal{G}$ needs to satisfy the following condition in order for the unconditional moments $\{E_F[m(W_i, \theta, \tau, g)] : (\tau, g) \in \mathcal{T} \times \mathcal{G}\}$ to incorporate the same information as the conditional moments $\{E_F[m(W_i, \theta, \tau)|X_i = x] : x \in R^{d_x}\}$.

For any $\theta \in \Theta$ and any distribution F with $E_F[\|m(W_i, \theta, \tau)\|] < \infty$, $\forall \tau \in \mathcal{T}$, let $\mathcal{X}_F(\theta, \tau)$ be defined as in (6.2) in ASM.

Assumption CI. For any $\theta \in \Theta$ and distribution F for which $E_F[\|m(W_i, \theta, \tau)\|] < \infty$, $\forall \tau \in \mathcal{T}$, if $P_F(X_i \in \mathcal{X}_F(\theta, \tau_*)) > 0$ for some $\tau_* \in \mathcal{T}$, then there exists some $g \in \mathcal{G}$ such that

$$E_F[m_j(W_i, \theta, \tau_*)g_j(X_i)] < 0 \text{ for some } j \leq p \text{ or}$$
$$E_F[m_j(W_i, \theta, \tau_*)g_j(X_i)] \neq 0 \text{ for some } j > p.$$

Note that CI abbreviates "conditionally identified." The following Lemma indicates the importance of Assumption CI. The proof of the lemma is the same as the proof of Lemma 2 in AS1, which is given in AS2, and in consequence, is omitted.

Lemma B.1 *Assumption* CI *implies that* $\Theta_F(\mathcal{G}) = \Theta_F$ *for all* F *with* $\sup_{\theta \in \Theta} E_F[\|m(W_i, \theta, \tau)\|] < \infty$.

Collections $\mathcal{G}$ that satisfy Assumption CI contain non-negative functions whose supports are cubes, boxes, or other sets which are arbitrarily small.

The collection $\mathcal{G}$ also must satisfy the following "manageability" condition.

Assumption M. (a) $0 \leq g_j(x) \leq G$ $\forall x \in R^{d_x}, \forall j \leq k, \forall g \in \mathcal{G}$, for some constant $G < \infty$, and

(b) the processes $\{g_j(X_{n,i}) : g \in \mathcal{G},\ i \leq n, n \geq 1\}$ are manageable with respect to the constant function G for $j = 1, ..., k$, where $\{X_{n,i} : i \leq n, n \geq 1\}$ is a row-wise i.i.d. triangular array with $X_{n,i} \sim F_{X,n}$ and $F_{X,n}$ is the distribution of $X_{n,i}$ under F_n for some $(\theta_n, F_n) \in \mathcal{F}_+$ for $n \geq 1$.[16]

[16]The asymptotic results given in the paper hold with Assumption M replaced by any alternative assumption that is sufficient to obtain the requisite empirical process results given in Lemma D.2 below.

Lemma 3 of AS1 establishes Assumptions CI and M for $\mathcal{G}_{\text{c-cube}}$ defined in (3.6) of ASM.[17]

B.5 Weight Function Q

The weight function Q can be any probability measure on $\mathcal{G}$ whose support is $\mathcal{G}$. This support condition is needed to ensure that no functions $g \in \mathcal{G}$, which might have set-identifying power, are "ignored" by the test statistic $T_n(\theta)$. Without such a condition, a CS based on $T_n(\theta)$ would not necessarily shrink to the identified set as $n \to \infty$. Section E below introduces the support condition formally and shows that the probability measure Q considered here satisfies it.

We now give an example of a weight function Q on $\mathcal{G}_{\text{c-cube}}$.

Weight Function Q for $\mathcal{G}_{\text{c-cube}}$. There is a one-to-one mapping $\Pi_{\text{c-cube}} : \mathcal{G}_{\text{c-cube}} \to AR := \{(a,r) : a \in \{1,...,2r\}^{d_x} \text{ and } r = r_0, r_0+1,...\}$. Let Q_{AR} be a probability measure on AR. One can take $Q = \Pi_{\text{c-cube}}^{-1} Q_{AR}$. A natural choice of measure Q_{AR} is uniform on $a \in \{1,...,2r\}^{d_x}$ conditional on r combined with a distribution for r that has some probability mass function $\{w(r) : r = r_0, r_0+1,...\}$. This yields the test statistic

$$\sup_{\tau \in \mathcal{T}} \sum_{r=r_0}^{\infty} w(r) \sum_{a \in \{1,...,2r\}^{d_x}} (2r)^{-d_x} S(n^{1/2}\overline{m}_n(\theta,\tau,g_{a,r}), \overline{\Sigma}_n(\theta,\tau,g_{a,r})), \tag{B.6}$$

where $g_{a,r}(x) := 1(x \in C_{a,r}) \cdot 1_k$ for $C_{a,r} \in \mathcal{C}_{\text{c-cube}}$.

The weight function Q_{AR} with $w(r) := (r^2+100)^{-1}$ is used in the test statistics in ASM, see (3.7).

B.6 Computation of Sums, Integrals, and Suprema

The test statistic $T_n(\theta)$ given in (B.6) involves an infinite sum. A collection $\mathcal{G}$ with an uncountable number of functions g yields a test statistic $T_n(\theta)$ that is an integral with respect to Q. This infinite sum or integral can be approximated by truncation, simulation, or quasi-Monte Carlo (QMC) methods. If $\mathcal{G}$ is countable, let $\{g_1,...,g_{s_n}\}$ denote the first s_n functions g that appear in the infinite sum that defines $T_n(\theta)$. Alternatively, let $\{g_1,...,g_{s_n}\}$ be s_n i.i.d. functions drawn from $\mathcal{G}$ according to the distribution Q. Or, let $\{g_1,...,g_{s_n}\}$ be

[17] Lemma 3 of AS1 and Lemma B2 of AS2 also establish Assumptions CI and M of this Appendix for the collections $\mathcal{G}_{\text{box}}$, $\mathcal{G}_{\text{B-spline}}$, $\mathcal{G}_{\text{box,dd}}$, and $\mathcal{G}_{\text{c/d}}$ defined there.

the first s_n terms in a QMC approximation of the integral with respect to (wrt) Q. Then, an approximate test statistic obtained by truncation, simulation, or QMC methods is

$$\overline{T}_{n,s_n}(\theta) := \sup_{\tau \in \mathcal{T}} \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S(n^{1/2}\overline{m}_n(\theta, \tau, g_\ell), \overline{\Sigma}_n(\theta, \tau, g_\ell)), \tag{B.7}$$

where $w_{Q,n}(\ell) := Q(\{g_\ell\})$ when an infinite sum is truncated, $w_{Q,n}(\ell) := s_n^{-1}$ when $\{g_1, ..., g_{s_n}\}$ are i.i.d. draws from $\mathcal{G}$ according to Q, and $w_{Q,n}(\ell)$ is a suitable weight when a QMC method is used. For example, in (B.6), the outer sum can be truncated at $r_{1,n}$, in which case, $s_n := \sum_{r=r_0}^{r_{1,n}} (2r)^{d_x}$ and $w_{Q,n}(\ell) := w(r)(2r)^{-d_x}$ for ℓ such that g_ℓ corresponds to $g_{a,r}$ for some a. The test statistics in (3.7) of ASM are of this form when $r_{1,n} < \infty$.[18]

It can be shown that truncation at s_n, simulation based on s_n simulation repetitions, or QMC approximation based on s_n terms, where $s_n \to \infty$ as $n \to \infty$, is sufficient to maintain the asymptotic validity of the tests and CS's as well as the asymptotic power results under fixed alternatives.

The KS form of the test statistic requires the computation of a supremum over $g \in \mathcal{G}$. For computational ease, this can be replaced by a supremum over $g \in \mathcal{G}_n$, where $\mathcal{G}_n \uparrow \mathcal{G}$ as $n \to \infty$, in the test statistic and in the definition of the critical value (defined below). The same asymptotic size results and asymptotic power results under fixed alternatives hold for KS tests with $\mathcal{G}_n$ in place of $\mathcal{G}$. For results of this sort for the tests considered in AS1 and AS2, see Section 13.1 of Appendix B in AS2 and Section 15.1 of Appendix D in AS2.

C GMS Confidence Sets

C.1 Bootstrap GMS Critical Values

It is shown in Theorem D.3 in Section D.3.1 below that when θ is in the identified set the "uniform asymptotic distribution" of $T_n(\theta)$ is the distribution of $T(h_n)$, where $T(h)$ is defined below, $h_n := (h_{1,n}, h_2)$, $h_{1,n}(\cdot)$ is a function from $\mathcal{T} \times \mathcal{G}$ to $R^p_{[+\infty]} \times \{0\}^v$ that depends on the slackness of the moment inequalities and on n, where $R_{[+\infty]} := R \cup \{+\infty\}$, and $h_2(\tau, g, \tau^\dagger, g^\dagger)$ is a $k \times k$ matrix-valued covariance kernel on $(\mathcal{T} \times \mathcal{G})^2$.

[18] Typically, the supremum over τ is obtained through smooth optimization techniques and there is no need to approximate $\mathcal{T}$ by a finite set. However, when smooth optimization is not applicable, we can also approximate $\mathcal{T}$ with a finite subset in the same way as approximating $\mathcal{G}$ by a finite subset.

For $h := (h_1, h_2)$, define

$$T(h) := \sup_{\tau \in \mathcal{T}} \int_{\mathcal{G}} S(\nu_{h_2}(\tau, g) + h_1(\tau, g), h_2^{\varepsilon}(\tau, g)) dQ(g), \tag{C.1}$$

where $h_2^{\varepsilon}(\tau, g) = h_2(\tau, g, \tau, g) + \varepsilon I_k$, and

$$\{\nu_{h_2}(\tau, g) : (\tau, g) \in \mathcal{T} \times \mathcal{G}\} \tag{C.2}$$

is a mean zero R^k-valued Gaussian process with covariance kernel $h_2(\cdot, \cdot)$ on $(\mathcal{T} \times \mathcal{G})^2$, $h_1(\cdot)$ is a function from $\mathcal{T} \times \mathcal{G}$ to $R^p_{[+\infty]} \times \{0\}^v$, and ε is as in the definition of $\overline{\Sigma}_n(\theta, \tau, g)$ in (3.4).[19] The definition of $T(h)$ in (C.1) applies to CvM test statistics. For the KS test statistic, one replaces $\int_{\mathcal{G}} \cdots \, dQ(g)$ by $\sup_{g \in \mathcal{G}} \cdots$.

We are interested in tests of nominal level α and CS's of nominal level $1 - \alpha$. Let

$$c_0(h, 1 - \alpha) \ (= c_0(h_1, h_2, 1 - \alpha)) \tag{C.3}$$

denote the $1 - \alpha$ quantile of $T(h)$. If $h_n := (h_{1,n}, h_2)$ was known, we would use $c_0(h_n, 1 - \alpha)$ as the critical value for the test statistic $T_n(\theta)$. However, h_n is not known and $h_{1,n}$ cannot be consistently estimated. In consequence, we replace h_2 in $c_0(h_{1,n}, h_2, 1 - \alpha)$ by a uniformly consistent estimator $\widehat{h}_{2,n}(\theta)$ $(:= \widehat{h}_{2,n}(\theta, \cdot, \cdot))$ of the covariance kernel h_2 and we replace $h_{1,n}$ by a data-dependent GMS function $\varphi_n(\theta)$ $(:= \varphi_n(\theta, \cdot))$ on $\mathcal{T} \times \mathcal{G}$ (defined in Section C.2 below) that is constructed to be less than or equal to $h_{1,n}(\tau, g)$ for all $(\tau, g) \in \mathcal{T} \times \mathcal{G}$ with probability that goes to one as $n \to \infty$. Because $S(m, \Sigma)$ is non-increasing in m_I by Assumption S1(b), where $m := (m_I', m_{II}')'$ and $m_I \in R^p$, the latter property yields a test with asymptotic level less than or equal to the nominal level α. The quantities $\widehat{h}_{2,n}(\theta)$ and $\varphi_n(\theta)$ are defined below.

Using $\widehat{h}_{2,n}(\theta)$ and $\varphi_n(\theta)$, in principle, one can obtain an approximation of $c_0(h_1, h_2, 1 - \alpha)$ using $c_0(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1 - \alpha)$. However, computing $c_0(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1 - \alpha)$ in practice is not easy because it involves the simulation of the Gaussian process $\{\nu_{\widehat{h}_2(\theta)}(\tau, g) : \mathcal{T} \times \mathcal{G}\}$. Although we approximate $\mathcal{G}$ by a finite set in ASM, we may not always do so for $\mathcal{T}$. Even when we also use a finite approximation for $\mathcal{T}$, the number of pairs (τ, g) under consideration often is large. That creates difficulty for simulating the Gaussian process.

[19] The sample paths of $\nu_{h_2}(\cdot, \cdot)$ are concentrated on the set $U^k_{\rho_{h_2}}(\mathcal{T} \times \mathcal{G})$ of bounded uniformly ρ_{h_2}-continuous R^k-valued functions on $\mathcal{T} \times \mathcal{G}$, where ρ_{h_2} is the pseudo-metric on $\mathcal{T} \times \mathcal{G}$ defined by $\rho^2_{h_2}(\iota, \iota^{\dagger}) := tr(h_2(\iota, \iota) - h_2(\iota, \iota^{\dagger}) - h_2(\iota^{\dagger}, \iota) + h_2(\iota^{\dagger}, \iota^{\dagger}))$, where $\iota := (\tau, g)$ and $\iota^{\dagger} := (\tau^{\dagger}, g^{\dagger})$.

Thus, we recommend using a bootstrap version of the critical value instead.

The bootstrap GMS critical value is[20]

$$c^*(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha) := c^*_0(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha+\eta) + \eta, \tag{C.4}$$

where $c^*_0(h, 1-\alpha)$ is the $1-\alpha$ conditional quantile of $T^*(h)$ and $T^*(h)$ is defined as in (C.1) but with $\{\nu_{h_2}(\tau, g) : (\tau, g) \in \mathcal{T} \times \mathcal{G}\}$ replaced by the bootstrap empirical process $\{\nu^*_n(\theta, \tau, g) : (\tau, g) \in \mathcal{T} \times \mathcal{G}\}$. The bootstrap empirical process is defined to be

$$\nu^*_n(\theta, \tau, g) := n^{-1/2} \widehat{D}_n(\theta)^{-1/2} \sum_{i=1}^n (m(W^*_i, \theta, \tau, g) - \overline{m}_n(\theta, \tau, g)), \tag{C.5}$$

where $\{W^*_i : i \leq n\}$ is an i.i.d. bootstrap sample drawn from the empirical distribution of $\{W_i : i \leq n\}$ and $\widehat{D}_n(\theta)$ is defined in (C.10). The function $\widehat{h}^*_{2,n}(\theta, \tau, g, \tau^\dagger, g^\dagger)$ is defined as

$$\widehat{h}^*_{2,n}(\theta, \tau, g, \tau^\dagger, g^\dagger) = \widehat{D}_n^{-1/2}(\theta) \widehat{\Sigma}^*_n(\theta, \tau, g, \tau^\dagger, g^\dagger) \widehat{D}_n^{-1/2}(\theta), \text{where} \tag{C.6}$$

$\widehat{D}_n(\theta)$ is defined in (C.10) below, and

$$\widehat{\Sigma}^*_n(\theta, \tau, g, \tau^\dagger, g^\dagger) : = n^{-1} \sum_{i=1}^n \left(m(W^*_i, \theta, \tau, g) - \overline{m}^*_n(\theta, \tau, g)\right) \left(m(W^*_i, \theta, \tau^\dagger, g^\dagger) - \overline{m}^*_n(\theta, \tau^\dagger, g^\dagger)\right)', \tag{C.7}$$

and $\overline{m}^*_n(\theta, \tau, g) = n^{-1} \sum_{i=1}^n m(W^*_i, \theta, \tau, g)$. Note that we do not recompute $\widehat{D}_n(\theta)$ for the bootstrap samples, which simplifies the theoretical derivations below. Also note that the variance-covariance kernel $\widehat{h}^*_{2,n}(\theta, \tau, g, \tau^\dagger, g^\dagger)$ only enters $c(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha)$ via indices $(\tau, g, \tau^\dagger, g^\dagger)$ such that $(\tau, g) = (\tau^\dagger, g^\dagger)$.

The nominal level $1-\alpha$ GMS CS is given by

$$CS_n := \{\theta \in \Theta : T_n(\theta) \leq c^*_{n,1-\alpha}(\theta)\}, \tag{C.8}$$

[20]The constant η is an *infinitesimal uniformity factor* (IUF) that is employed to circumvent problems that arise due to the presence of the infinite-dimensional nuisance parameter $h_{1,n}$ that affects the distribution of the test statistic in both small and large samples. The IUF obviates the need for complicated and difficult-to-verify uniform continuity and strict monotonicity conditions on the large sample distribution functions of the test statistic.

where the critical value $c^*_{n,1-\alpha}(\theta)$ abbreviates $c^*(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha)$.

When the test statistic, $\overline{T}_{n,s_n}(\theta)$, is a truncated sum, simulated integral, or a QMC quantity, a bootstrap approximate-GMS critical value can be employed. It is defined analogously to the bootstrap GMS critical value but with $T^*(h)$ replaced by $T^*_{s_n}(h)$, where $T^*_{s_n}(h)$ has the same definition as $T^*(h)$ except that a truncated sum, simulated integral, or QMC quantity appears in place of the integral with respect to Q, as in Section B.6. The same functions $\{g_1, ..., g_{s_n}\}$ are used in all bootstrap critical value calculations as in the test statistic $\overline{T}_{n,s_n}(\theta)$.

Next, we define the asymptotic covariance kernel, $\{h_{2,F}(\theta,\tau,g,\tau^\dagger,g^\dagger) : (\tau,g),(\tau^\dagger,g^\dagger) \in \mathcal{T} \times \mathcal{G}\}$, of $n^{1/2}(\overline{m}_n(\theta,\tau,g) - E_F\overline{m}_n(\theta,\tau,g))$ after normalization via a diagonal matrix $D_F^{-1/2}(\theta)$. Define

$$\begin{aligned} h_{2,F}(\theta,\tau,g,\tau^\dagger,g^\dagger) &:= D_F^{-1/2}(\theta)\Sigma_F(\theta,\tau,g,\tau^\dagger,g^\dagger)D_F^{-1/2}(\theta), \text{ where} \\ \Sigma_F(\theta,\tau,g,\tau^\dagger,g^\dagger) &:= Cov_F(m(W_i,\theta,\tau,g), m(W_i,\theta,\tau^\dagger,g^\dagger)'), \\ D_F(\theta) &:= Diag(\sigma^2_{F,1}(\theta),\ldots,\sigma^2_{F,k}(\theta)), \end{aligned} \tag{C.9}$$

and $\sigma^2_{F,j}(\theta)$ is introduced above Assumption PS1.

Correspondingly, the sample covariance kernel $\widehat{h}_{2,n}(\theta)$ $(= \widehat{h}_{2,n}(\theta,\cdot,\cdot))$, which is an estimator of $h_{2,F}(\theta,\tau,g,\tau^\dagger,g^\dagger)$, is defined by

$$\begin{aligned} \widehat{h}_{2,n}(\theta,\tau,g,\tau^\dagger,g^\dagger) &:= \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta,\tau,g,\tau^\dagger,g^\dagger)\widehat{D}_n^{-1/2}(\theta), \text{ where} \\ \widehat{\Sigma}_n(\theta,\tau,g,\tau^\dagger,g^\dagger) &:= n^{-1}\sum_{i=1}^n \left(m(W_i,\theta,\tau,g) - \overline{m}_n(\theta,\tau,g)\right)\left(m(W_i,\theta,\tau^\dagger,g^\dagger) - \overline{m}_n(\theta,\tau^\dagger,g^\dagger)\right)', \\ \widehat{D}_n(\theta) &:= Diag(\widehat{\sigma}^2_{n,1}(\theta),\ldots,\widehat{\sigma}^2_{n,k}(\theta)), \end{aligned} \tag{C.10}$$

and $\widehat{\sigma}^2_{n,j}(\theta)$ is a consistent estimator of $\sigma^2_{F,j}(\theta)$ introduced below (3.4).

Note that $\widehat{\Sigma}_n(\theta,\tau,g)$, defined in (3.3), equals $\widehat{\Sigma}_n(\theta,\tau,g,\tau,g)$.

C.2 Definition of $\varphi_{\mathbf{n}}(\theta)$

Next, we define $\varphi_n(\theta)$. As discussed above, $\varphi_n(\theta)$ is constructed such that $\varphi_n(\theta,\tau,g) \leq h_{1,n}(\tau,g)$ $\forall(\tau,g) \in \mathcal{T} \times \mathcal{G}$ with probability that goes to one as $n \to \infty$ uniformly over

$(\theta, F) \in \mathcal{F}$. Let

$$\xi_n(\theta, \tau, g) := \kappa_n^{-1} n^{1/2} \overline{D}_n^{-1/2}(\theta, \tau, g) \overline{m}_n(\theta, \tau, g), \text{ where } \overline{D}_n(\theta, \tau, g) := Diag(\overline{\Sigma}_n(\theta, \tau, g)), \tag{C.11}$$

$\overline{\Sigma}_n(\theta, \tau, g)$ is defined in (3.4), and $\{\kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \to \infty$. The jth element of $\xi_n(\theta, \tau, g)$, denoted by $\xi_{n,j}(\theta, \tau, g)$, measures the slackness of the moment inequality $E_F m_j(W_i, \theta, \tau, g) \geq 0$ for $j = 1, ..., p$.

Define $\varphi_n(\theta, \tau, g) := (\varphi_{n,1}(\theta, \tau, g), ..., \varphi_{n,p}(\theta, \tau, g), 0, ..., 0)' \in R^k$ via, for $j \leq p$,

$$\begin{aligned} \varphi_{n,j}(\theta, \tau, g) &:= \overline{h}_{2,n,j}^{-1/2}(\theta, \tau, g) B_n 1(\xi_{n,j}(\theta, \tau, g) > 1), \\ \overline{h}_{2,n}(\theta, \tau, g) &:= \widehat{D}_n^{-1/2}(\theta) \overline{\Sigma}_n(\theta, \tau, g) \widehat{D}_n^{-1/2}(\theta), \text{ and} \\ \overline{h}_{2,n,j}(\theta, \tau, g) &:= [\overline{h}_{2,n}(\theta, \tau, g)]_{jj}. \end{aligned} \tag{C.12}$$

We assume:

Assumption GMS1.(a) $\overline{\varphi}_n(\theta, \tau, g)$ satisfies (C.12) and $\{B_n : n \geq 1\}$ is a nondecreasing sequence of positive constants, and

(b) $\kappa_n \to \infty$ and $B_n/\kappa_n \to 0$ as $n \to \infty$.

In ASM and Andrews and Shi (2014), we use $\kappa_n = (0.3 \ln(n))^{1/2}$ and $B_n = (0.4 \ln(n)/ \ln \ln(n))^{1/2}$, which satisfy Assumption GMS1.

The multiplicand $\overline{h}_{2,n,j}^{-1/2}(\theta, \tau, g)$ in (C.12) is an "ε-adjusted" standard deviation estimator for the jth normalized sample moment based on g (see (3.4) for the ε-adjustment in $\overline{\Sigma}_n(\theta, \tau, g)$). It provides a suitable scaling for $\varphi_n(\theta, \tau, g)$.

D Asymptotic Size

In this section, we show that the bootstrap GMS CS's have correct uniform asymptotic coverage probabilities, i.e., correct asymptotic size.

D.1 Notation

First, define

$$\begin{aligned}
h_{1,n,F}(\theta,\tau,g) &:= n^{1/2}D_F^{-1/2}(\theta)E_F m(W_i,\theta,\tau,g),\\
h_{n,F}(\theta,\tau,g,\tau^\dagger,g^\dagger) &:= (h_{1,n,F}(\theta,\tau,g), h_{2,F}(\theta,\tau,g,\tau^\dagger,g^\dagger)),\\
\widehat{h}_{2,n,F}(\theta,\tau,g,\tau^\dagger,g^\dagger) &:= D_F^{-1/2}(\theta)\widehat{\Sigma}_n(\theta,\tau,g,\tau^\dagger,g^\dagger)D_F^{-1/2}(\theta),\\
\overline{h}_{2,n,F}(\theta,\tau,g) &:= \widehat{h}_{2,n,F}(\theta,\tau,g,\tau,g) + \varepsilon D_F^{-1/2}(\theta)\widehat{D}_n(\theta)D_F^{-1/2}(\theta)\\
&= D_F^{-1/2}(\theta)\overline{\Sigma}_n(\theta,\tau,g)D_F^{-1/2}(\theta), \text{ and} \qquad \text{(D.1)}\\
\nu_{n,F}(\theta,\tau,g) &:= n^{-1/2}\sum_{i=1}^n D_F^{-1/2}(\theta)[m(W_i,\theta,\tau,g) - E_F m(W_i,\theta,\tau,g)],
\end{aligned}$$

where $m(W_i,\theta,\tau,g)$, $\widehat{\Sigma}_n(\theta,\tau,g,\tau^\dagger,g^\dagger)$, $\overline{\Sigma}_n(\theta,\tau,g)$, $D_F(\theta)$, and $\widehat{D}_n(\theta)$ are defined in (3.2), (3.3), (3.4), and (5.1) of ASM, and (C.10), respectively.

Below we write $T_n(\theta)$ as a function of the quantities in (D.1). As defined, (i) $h_{1,n,F}(\theta,\tau,g)$ is the k-vector of normalized means of the moment functions for $(\tau,g) \in \mathcal{T}\times\mathcal{G}$, which measures the slackness of the population moment conditions under (θ,F), and it has the very useful feature that it is non-negative when $(\theta,F)\in\mathcal{F}$ by (2.1) of ASM, (ii) $h_{n,F}(\theta,\tau,g,\tau^\dagger,g^\dagger)$ contains the approximation to the normalized means of $D_F^{-1/2}(\theta)m(W_i,\theta,\tau,g)$ and the covariances of $D_F^{-1/2}(\theta)m(W_i,\theta,\tau,g)$ and $D_F^{-1/2}(\theta)m(W_i,\theta,\tau^\dagger,g^\dagger)$, (iii) $\widehat{h}_{2,n,F}(\theta,\tau,g,\tau^\dagger,g^\dagger)$ and $\overline{h}_{2,n,F}(\theta,\tau,g)$ are hybrid quantities—part population, part sample—based on the matrices $\widehat{\Sigma}_n(\theta,\tau,g,\tau^\dagger,g^\dagger)$ and $\overline{\Sigma}_n(\theta,\tau,g)$, respectively, and (iv) $\nu_{n,F}(\theta,\tau,g)$ is the sample average of the moment functions $D_F^{-1/2}(\theta)m(W_i,\theta,\tau,g)$ normalized to have mean zero and variance that is $O(1)$, but not $o(1)$. Note that $\nu_{n,F}(\theta,\cdot,\cdot)$ is an empirical process indexed by $(\tau,g)\in\mathcal{T}\times\mathcal{G}$ with covariance kernel given by $h_{2,F}(\theta,\tau,g,\tau^\dagger,g^\dagger)$.

The normalized sample moments $n^{1/2}\overline{m}_n(\theta,\tau,g)$ can be written as

$$n^{1/2}\overline{m}_n(\theta,\tau,g) = D_F^{1/2}(\theta)(\nu_{n,F}(\theta,\tau,g) + h_{1,n,F}(\theta,\tau,g)). \qquad \text{(D.2)}$$

The test statistic $T_n(\theta)$, defined in (B.3), can be written as

$$T_n(\theta) = \sup_{\tau\in\mathcal{T}}\int_{\mathcal{G}} S(\nu_{n,F}(\theta,\tau,g) + h_{1,n,F}(\theta,\tau,g), \overline{h}_{2,n,F}(\theta,\tau,g))dQ(g). \qquad \text{(D.3)}$$

Note the close resemblance between $T_n(\theta)$ and $T(h)$ (defined in (C.1)).

Let $\mathcal{H}_1$ denote the set of all functions from $\mathcal{T} \times \mathcal{G}$ to $R^p_{[+\infty]} \times \{0\}^v$.

For notational simplicity, for any function of the form $r_F(\theta, \tau, g)$ for $(\tau, g) \in \mathcal{T} \times \mathcal{G}$, let $r_F(\theta)$ denote the function $r_F(\theta, \cdot, \cdot)$ on $\mathcal{T} \times \mathcal{G}$. Correspondingly, for any function of the form $r_F(\theta, \tau, g, \tau^\dagger, g^\dagger)$ for $(\tau, g), (\tau^\dagger, g^\dagger) \in \mathcal{T} \times \mathcal{G}$, let $r_F(\theta)$ denote the function $r_F(\theta, \cdot, \cdot, \cdot, \cdot)$ on $(\mathcal{T} \times \mathcal{G})^2$. Thus, $h_{2,F}(\theta)$ abbreviates the asymptotic covariance kernel $\{h_{2,F}(\theta, \tau, g, \tau^\dagger, g^\dagger) : (\tau, g), (\tau^\dagger, g^\dagger) \in \mathcal{T} \times \mathcal{G}\}$ defined in (C.9). Define

$$\mathcal{H}_2 := \{h_{2,F}(\theta) : (\theta, F) \in \mathcal{F}\}, \tag{D.4}$$

where, as defined at the end of Section 2, $\mathcal{F}$ is the subset of $\mathcal{F}_+$ that satisfies Assumption PS3. On the space of $k \times k$ matrix-valued covariance kernels on $(\mathcal{T} \times \mathcal{G})^2$, which is a superset of $\mathcal{H}_2$, we use the uniform metric d defined by

$$d(h_2^{(1)}, h_2^{(2)}) := \sup_{(\tau,g),(\tau^\dagger,g^\dagger) \in \mathcal{T} \times \mathcal{G}} \|h_2^{(1)}(\tau, g, \tau^\dagger, g^\dagger) - h_2^{(2)}(\tau, g, \tau^\dagger, g^\dagger)\|. \tag{D.5}$$

Let $\Rightarrow$ denote weak convergence. Let $\{a_n\}$ denote a subsequence of n. Let $\rho_{h_2}(\theta)$ be the intrinsic pseudometric on $\mathcal{T} \times \mathcal{G}$ for the tight Gaussian process $\nu_{h_2}(\theta)$ with variance-covariance kernel h_2:

$$\begin{aligned} &\rho_{h_2}(\tau, g, \tau^\dagger, g^\dagger) \\ &:= tr\left(h_2(\tau, g, \tau, g) - h_2(\tau, g, \tau^\dagger, g^\dagger) - h_2(\tau^\dagger, g^\dagger, \tau, g) + h_2(\tau^\dagger, g^\dagger, \tau^\dagger, g^\dagger)\right). \end{aligned} \tag{D.6}$$

D.2 Proof of Theorem 5.1

Theorem 5.1 of ASM is a result of two lemmas. The two lemmas together imply the uniform validity of the GMS CS over $\mathcal{F}$ under Assumptions M, S1,, S2, andGMS1.

The first lemma below establishes the uniform asymptotic size under two high-level assumptions (given below). The second lemma verifies these two assumptions under Assumptions M, S1, and S2.

Assumption PS4. For any subsequence $\{a_n\}$ of $\{n\}$ and any sequence $\{(\theta_{a_n}, F_{a_n}) \in \mathcal{F}_+ : n \geq 1\}$ for which

$$\lim_{n\to\infty} d(h_{2,F_{a_n}}(\theta_{a_n}), h_2) = 0 \tag{D.7}$$

for some $k \times k$ matrix-valued covariance kernel $h_2(\tau, g, \tau^\dagger, g^\dagger)$ on $(\mathcal{T} \times \mathcal{G})^2$, we have

(i) $\nu_{a_n,F_{a_n}}(\theta_{a_n}) \Rightarrow \nu_{h_2}(\cdot)$ and

(ii) $d(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}), h_2) \to_p 0$ as $n \to \infty$, where $\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n})$ is defined in (D.1).

Assumption PS5. For any subsequence $\{a_n\}$ of $\{n\}$ and any sequence $\{(\theta_{a_n}, F_{a_n}) \in \mathcal{F}_+ : n \geq 1\}$, conditional on any sample path ω for which

$$\lim_{n\to\infty} d(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n})(\omega), h_2) = 0, \tag{D.8}$$

for some $k \times k$ matrix-valued covariance kernel $h_2(\tau, g, \tau^\dagger, g^\dagger)$ on $(\mathcal{T} \times \mathcal{G})^2$, we have (i) $\nu^*_{a_n}(\theta) \Rightarrow \nu_{h_2}$ and (ii) $d(\widehat{h}^*_{2,a_n}(\theta_{a_n}), h_2) \to_p 0$.

Lemma D.1 *Suppose Assumptions* PS4, PS5, S1, S2, *and* SIG1 *hold, and Assumption* GMS1 *holds when considering* GMS *critical values. Then, for any compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$*, the* GMS *CS satisfies:*

$$\liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F(\theta \in CS_n) \geq 1 - \alpha.$$

Lemma D.2 *Suppose Assumptions* M, S1, *and* S2 *hold. Then,*

(a) *Assumption* PS4 *holds and*

(b) *Assumption* PS5 *holds.*

Comments. 1. Lemma D.1(a) shows that GMS CS has correct uniform asymptotic size. The uniformity results hold whether the moment conditions involve "weak" or "strong" IV's X_i.

2. Theorem 5.1 of ASM for the case $r_{1,n} = \infty$ is proved by verifying the conditions of Lemma D.2 (that is, by showing that Assumptions M, S1, S2, and GMS1 hold for the $\mathcal{G}_{\text{c-cube}}$ set and the S functions considered in ASM).[21] The functions S_1, S_2, and S_3 in (3.9) of ASM satisfy Assumptions S1 and S2 by Lemma 1 of AS1 and the function S_4 of ASM satisfy Assumptions S1 and S2 by similar arguments. Lemma 3 of AS1 establishes Assumption M for $\mathcal{G}_{\text{c-cube}}$ defined in (3.6) of ASM. Assumption GMS1 holds immediately for κ_n and B_n used in (4.1) and (4.2) of ASM, respectively. Theorem 5.1 of ASM holds for $r_{1,n}$ such that $r_{1,n} < \infty$ and $r_{1,n} \to \infty$ as $n \to \infty$ by minor alterations to the proofs.

[21] The quantity $r_{1,n}$ is the test statistic truncation value that appears in (3.7) of ASM. It satisfies either $r_{1,n} = \infty$ for all $n \geq 1$ or $r_{1,n} < \infty$ and $r_{1,n} \to \infty$ as $n \to \infty$.

D.3 Proof of Lemma D.1

D.3.1 Theorem D.3

The following Theorem provides a uniform asymptotic distributional result for the test statistic $T_n(\theta)$. It is an analogue of Theorem 1 in AS1. It is used in the proof of Lemma D.1.

Theorem D.3 *Suppose Assumptions* PS4, S1, S2, *and* SIG1 *hold. Then, for all compact subsets $\mathcal{H}_{2,cpt}$ of $\mathcal{H}_2$, for all constants $x_{h_{n,F}(\theta)} \in R$ that may depend on (θ, F) and n through $h_{n,F}(\theta)$, and all $\delta > 0$, we have*

(a) $\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} \left[P_F(T_n(\theta) > x_{h_{n,F}(\theta)}) - P\left(T(h_{n,F}(\theta)) + \delta > x_{h_{n,F}(\theta)}\right)\right] \leq 0$ *and*

(b) $\liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} \left[P_F(T_n(\theta) > x_{h_{n,F}(\theta)}) - P\left(T(h_{n,F}(\theta)) - \delta > x_{h_{n,F}(\theta)}\right)\right] \geq 0,$

where $T(h)$ is the function defined in (C.1).

Proof of Theorem D.3. Theorem D.3 is similar to Theorem 1 in AS1. The proof of the latter theorem goes through with the following modifications:

(i) Redefine $SubSeq(h_2)$ to be the set of subsequences $\{(\theta_{a_n}, F_{a_n}) \in \mathcal{F} : n \geq 1\}$ where $\{a_n\}$ is a subsequence of $\{n\}$, such that (D.7) holds.

(ii) Replace $\int \cdots dQ(g)$ by $\sup_{\tau\in\mathcal{T}} \int_{\mathcal{G}} \cdots dQ(g)$. In other instances where g and $\mathcal{G}$ appear, replace g with (τ, g) and $\mathcal{G}$ with $\mathcal{T} \times \mathcal{G}$.

(iii) Replace "by Lemma A1" with "by Assumption PS4."

(iv) Change the paragraph at the bottom of p. 6 of AS2 to the following:

"Given this and Assumption SIG1, by the almost sure representation theorem, e.g., see Pollard (1990, Thm. 9.4), there exists a probability space and random quantities $\tilde{\nu}_{a_n}(\cdot)$, $\tilde{h}_{2,a_n}(\cdot)$, $\tilde{V}_{a_n}$, and $\tilde{\nu}_0(\cdot)$ defined on it such that (i) $(\tilde{\nu}_{a_n}(\cdot), \tilde{h}_{2,a_n}(\cdot), \tilde{V}_{a_n})$ has the same distribution as $(\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot), \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, \cdot), D_{F_{a_n}}^{-1/2}(\theta_{a_n})\widehat{D}_{a_n}(\theta_{a_n})D_{F_{a_n}}^{-1/2}(\theta_{a_n}))$, (ii) $(\tilde{\nu}_0(\cdot))$ has the same distribution as $\nu_{h_{2,0}}(\cdot)$, and

$$\text{(iii)} \quad \sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}} \left\| \begin{pmatrix} \tilde{\nu}_{a_n}(\tau, g) \\ \tilde{h}_{2,a_n}(\tau, g) \\ vec(\tilde{V}_{a_n}) \end{pmatrix} - \begin{pmatrix} \tilde{\nu}_0(\tau, g) \\ h_{2,0}(\tau, g) \\ vec(I_k) \end{pmatrix} \right\| \to 0 \text{ as } n \to \infty, \text{ a.s.} \qquad \text{(D.9)}$$

(v) Replace $Diag(\tilde{h}_{2,a_n}(1_k))$ by $\tilde{V}_{a_n}$.

With the above modifications, the proof of Theorem 1 in AS2 up to the proof of (12.7) of AS2 goes through. The proof of (12.7) in AS2, which relies on a dominated convergence argument, does not go through because the test statistic considered in this paper is not of the pure CvM type, and thus, $\tilde{T}_{a_n}$ and $\tilde{T}_{a_n,0}$ are not integrals with respect to (τ, g).

We change the proof of (12.7) in AS2 to the following.

As in the proof of (12.7) in AS2, we fix a sample path ω at which $(\tilde{\nu}_{a_n}(\tau, g), \tilde{h}_{2,a_n}(\tau, g))(\omega)$ converges to $(\tilde{\nu}_0(\tau, g), h_{2,0}(\tau, g))(\omega)$ uniformly over $(\tau, g) \in \mathcal{T} \times \mathcal{G}$ as $n \to \infty$ and $\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}} \|\tilde{\nu}_0(\tau, g)(\omega)\| < \infty$. Let $\tilde{\Omega}$ be the collection of such sample paths. By (D.9), $P(\tilde{\Omega}) = 1$. For a fixed $\omega \in \tilde{\Omega}$, by Assumption S2, we have

$$\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}} \sup_{\mu\in[0,\infty)^p\times\{0\}^v} |S(\tilde{\nu}_{a_n}(\tau, g)(\omega) + \mu, \tilde{h}^{\varepsilon}_{2,a_n}(\tau, g)(\omega)) - S(\tilde{\nu}_0(\tau, g)(\omega) + \mu, h^{\varepsilon}_{2,0}(\tau, g))| \to 0, \tag{D.10}$$

as $n \to \infty$, where $\tilde{h}^{\varepsilon}_{2,n}(\tau, g) := \tilde{h}_{2,n}(\tau, g) + \varepsilon \tilde{V}_{a_n}$, and $h^{\varepsilon}_{2,0}(\tau, g) := h_{2,0}(\tau, g) + \varepsilon I_k$. Thus, for every $\omega \in \tilde{\Omega}$,

$$\begin{aligned} |\tilde{T}_{a_n}(\omega) - \tilde{T}_{a_n,0}(\omega)| \le \sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}} & |S(\tilde{\nu}_{a_n}(\tau, g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, \tau, g), \tilde{h}^{\varepsilon}_{2,a_n}(\tau, g)(\omega)) - \\ & - S(\tilde{\nu}_0(\tau, g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, \tau, g), h^{\varepsilon}_{2,0}(\tau, g))| \\ & \to 0 \text{ as } n \to \infty. \end{aligned} \tag{D.11}$$

This verifies (12.7) in AS2. $\square$

D.3.2 Proof of Lemma D.1

Lemma D.1 is similar to Theorem 2(a) of AS1 and we modify the proof of the latter in AS2 to fit the context of Lemma D.1. In addition to notational changes, a substantial modification is needed because Theorem 2 of AS1 does not cover bootstrap critical values.

Specifically, the proof of Theorem 2(a) in AS2 with the following modifications provides the proof of Lemma D.1.

(i) Replace all references to "Assumption M" of AS1 by references to "Assumption PS4" stated above and Assumptions S1 and S2 of AS1 by Assumptions S1 and S2 stated above. Replace $\int \cdots dQ(g)$ by $\sup_{\tau\in\mathcal{T}} \int_{\mathcal{G}} \cdots dQ(g)$. In other instances where g and $\mathcal{G}$ appear, replace g with (τ, g) and $\mathcal{G}$ with $\mathcal{T} \times \mathcal{G}$. Let $\widehat{D}_{a_n}(\theta_{a_n})$ be defined as in (C.10) above, rather than as in AS1 and AS2.

(ii) Replace references to "Theorem 1(a)" of AS1 with references to "Theorem D.3(a)" stated above.

(iii) Redefine $SubSeq(h_2)$ to be the set of subsequences $\{(\theta_{a_n}, F_{a_n}) \in \mathcal{F} : n \geq 1\}$ for which (D.7) holds, where $\{a_n\}$ is a subsequence of $\{n\}$.

(iv) Replace references to "Lemma A1" of AS2 to references to "Assumption PS4" stated above.

(v) In both the statement and the proof of Lemma A3 in AS2, replace $c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$ with $c_0(\varphi_n(\theta), h_{2,F}(\theta), 1-\alpha)$, and $c(h_{1,n,F}(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$ with $c_0(h_{1,n,F}(\theta), h_{2,F}(\theta), 1-\alpha)$. The proof of Lemma A3 given in AS2 goes through with the following changes:

In the 6th and 7th last lines of the proof of Lemma A3 in AS2, delete "$\varepsilon^{-1/2} h_{2,0,j}^{-1/2}(1_k, 1_k)(1 + o_p(1)) =$", and change "by Lemma A1(b) and (5.2)" to "by Assumption SIG1 and (D.1)."

(vi) Replace Lemma A4 in AS2 with Lemma D.4 given immediately below. The proof of the Lemma D.4 given below is self-contained and does not rely on an analogue of Lemma A5 of AS2.

No other changes are needed in the proof of Theorem 2(a) in AS2. □

The following lemma is used in the proof of Lemma D.1 given immediately above.

Lemma D.4 *Suppose Assumptions* PS4, PS5, S1, S2, *and* GMS1 *hold. Then, for all* $\delta \in (0, \eta)$, *where* $\eta > 0$ *is defined in* (C.4),

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\left(c^*(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha) < c_0(\varphi_n(\theta), h_{2,F}(\theta), 1-\alpha) + \delta\right) = 0.$$

Prove of Lemma D.4. The result of the Lemma is equivalent to

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F(c_0^*(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha+\eta) + \eta < c_0(\varphi_n(\theta), h_{2,F}(\theta), 1-\alpha) + \delta) = 0. \tag{D.12}$$

By considering a sequence $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ that is within $\zeta_n \to 0$ of the supremum in the above display for all $n \geq 1$, it suffices to show that

$$\lim_{n\to\infty} P_{F_n}(c_0^*(\varphi_n(\theta_n), \widehat{h}^*_{2,n}(\theta_n), 1-\alpha+\eta) + \eta < c_0(\varphi_n(\theta_n), h_{2,F_n}(\theta_n), 1-\alpha) + \delta) = 0. \tag{D.13}$$

Given any subsequence $\{u_n\}$ of $\{n\}$, there exists a further subsequence $\{w_n\}$ such that $d(h_{2,F_{w_n}}(\theta_{w_n}), h_{2,0}) \to 0$ as $n \to \infty$ for some matrix-valued covariance function $h_{2,0}$ by the compactness of $\mathcal{H}_{2,cpt}$. It suffices to show that (D.13) holds with w_n in place of n.

By Assumption PS4(ii), $d(h_{2,F_{w_n}}(\theta_{w_n}), h_{2,0}) \to 0$ implies that $d(\widehat{h}_{2,w_n,F_{w_n}}(\theta_{w_n}), h_{2,0}) \to_p 0$, which then implies

$$d(\widehat{h}_{2,w_n}(\theta_{w_n}), h_{2,0}) \to_p 0, \tag{D.14}$$

where $\widehat{h}_{2,n}(\theta)$ and $\widehat{h}_{2,n,F}(\theta)$ are defined in (C.10) and (D.1), respectively. Then, by a general convergence in probability result, given any subsequence of $\{w_n\}$ there exists a further subsequence $\{a_n\}$ such that

$$d(\widehat{h}_{2,a_n}(\theta_{a_n}), h_{2,0}) \to 0 \text{ a.s.} \tag{D.15}$$

Hence, it suffices to show (D.13) with a_n in place of n. Let $\bar{\Omega}$ be the set of sample paths ω such that $d(\widehat{h}_{2,a_n}(\theta_{a_n})(\omega), h_2) \to 0$. The above display implies that $P(\bar{\Omega}) = 1$.

Consider an arbitrary sample path $\omega \in \bar{\Omega}$. Below we show that for all constants $x_n \in R$ (possibly dependent on ω) and all $\xi > 0$,

$$\begin{aligned}\limsup_{n\to\infty} \Big[&P\left(T^*(\varphi_{a_n}(\theta_{a_n}), \widehat{h}^*_{2,a_n}(\theta_{a_n})) \le x_{a_n}|\omega\right) - \\ &P\left(T(\varphi_{a_n}(\theta_{a_n}), h_{2,F_{a_n}}(\theta_{a_n})) \le x_{a_n} + \xi|\omega\right)\Big] \le 0,\end{aligned} \tag{D.16}$$

where in the first line $P(\cdot|\omega)$ denotes bootstrap probability conditional on the original sample path ω, in the second line $P(\cdot|\omega)$ denotes $\nu_{h_{2,F_{a_n}}(\theta_{a_n})}(\cdot)$ probability conditional on the original sample path ω, and $\nu_{h_{2,F_{a_n}}(\theta_{a_n})}(\cdot)$ is the Gaussian process defined in (C.2) with $h_2 = h_{2,F_{a_n}}(\theta_{a_n})$, which is taken to be independent of the original sample $\{W_i : i \le n\}$ and, hence, is independent of $\varphi_{a_n}(\theta_{a_n})$.

The interval $(0, \eta - \delta)$ is non-empty because $\delta \in (0, \eta)$ by assumption. Using (D.16), we obtain, for all $\xi \in (0, \eta - \delta)$,

$$\begin{aligned}&\limsup_{n\to\infty} P\left(T^*(\varphi_{a_n}(\theta_{a_n}), \widehat{h}^*_{2,a_n}(\theta_{a_n})) \le c_0(\varphi_{a_n}(\theta_{a_n}), h_{2,F_{a_n}}(\theta_{a_n}), 1-\alpha) + \delta - \eta|\omega\right) \\ &\le \limsup_{n\to\infty} P\left(T(\varphi_{a_n}(\theta_{a_n}), h_{2,F_{a_n}}(\theta_{a_n})) \le c_0(\varphi_{a_n}(\theta_{a_n}), h_{2,F_{a_n}}(\theta_{a_n}), 1-\alpha) + \delta - \eta + \xi|\omega\right) \\ &\le 1 - \alpha,\end{aligned} \tag{D.17}$$

where the second inequality holds because $\delta - \eta + \xi < 0$ for $\xi \in (0, \eta - \delta)$. For any df F with $1-\alpha+\eta$ quantile denoted by $q_{1-\alpha+\eta}$, we have $F(q_{1-\alpha+\eta}) \geq 1-\alpha+\eta$. Hence, if $F(x) < 1-\alpha+\eta$, then $x < q_{1-\alpha+\eta}$. Combining this with the result in (D.17) implies that given any $\delta \in (0, \eta)$, for n sufficiently large,

$$c_0(\varphi_{a_n}(\theta_{a_n})(\omega), h_{2,F_{a_n}}(\theta_{a_n}), 1-\alpha) + \delta - \eta < c_0^*(\varphi_{a_n}(\theta_{a_n}), h_{2,F_{a_n}}(\theta_{a_n}), 1-\alpha+\eta)(\omega), \quad \text{(D.18)}$$

where the indexing by ω denotes that the result holds for fixed $\omega \in \bar{\Omega}$. Because (D.18) holds for all $\omega \in \bar{\Omega}$ and $P(\bar{\Omega}) = 1$, the bounded convergence theorem applies and establishes (D.13).

It remains to prove the result in (D.16). This result follows from an analogous argument to that used to prove Theorem D.3(b). Note the common structure of the original sample and bootstrap sample test statistics:

$$\begin{aligned} T_n(\theta_n) &= \sup_{\tau \in \mathcal{T}} \int S(\nu_{n,F_n}(\theta_n, \tau, g) + h_{1,n,F_n}(\theta_n, \tau, g), \overline{h}_{2,n,F_n}(\theta_n, \tau, g)) dQ(g), \\ T_n^*(\varphi_n(\theta_n), \widehat{h}_{2,n}^*(\theta_n)) &= \sup_{\tau \in \mathcal{T}} \int S(\nu_n^*(\theta_n, \tau, g) + \varphi_n(\theta_n, \tau, g), \overline{h}_{2,n}^*(\theta_n, \tau, g)) dQ(g), \text{ where} \\ \overline{h}_{2,n}^*(\theta, \tau, g) &:= \widehat{h}_{2,n}^*(\theta, \tau, g) + \varepsilon I_k, \end{aligned}$$

$\nu_{n,F}$, $h_{1,n,F}$, and $\overline{h}_{2,n,F}$ are defined in (D.1), $\varphi_n(\theta)$ is defined in (C.12), and $\widehat{h}_{2,n}^*(\theta)$ is defined following (C.5) using (C.10) with W_i^* in place of W_i.

The result of Theorem D.3(b) with $T_n(\theta_n)$ replaced by $T_n^*(\varphi_n(\theta_n), \widehat{h}_{2,n}^*(\theta_n))$, with $T(h_{n,F}(\theta))$ replaced by $T(\varphi_n(\theta_n), h_{2,F_n}(\theta_n))$, and with δ replaced by ξ, when applied to the subsequence $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\}$ is the result of (D.16). The result in (D.16) follows by the same argument as that for Theorem D.3(b) with $\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot)$ replaced by $\nu_{a_n}^*(\theta_{a_n}, \cdot)(\omega)$, where $\nu_{a_n}^*(\theta_{a_n}, \cdot)(\omega)$ denotes the bootstrap empirical process given the sample path ω of the original sample, with $\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, \cdot, \cdot)$ replaced by $\widehat{h}_{2,a_n}^*(\theta_{a_n}, \cdot, \cdot)(\omega)$, and with Assumption PS4 replaced by Assumption PS5, which guarantees that $\nu_{a_n}^*(\theta_{a_n})(\omega) \Rightarrow \nu_{h_{2,0}}$ and $d(\widehat{h}_{2,a_n}^*(\theta_{a_n})(\omega), h_{2,0}) \to_p 0$. $\square$

D.4 Proof of Lemma D.2

The verification of Assumption PS4 is the same as the proof of Lemma A1 given in Appendix E of AS2 except with some notation changes and with Lemma D.5 below replacing Lemma E1(a) in AS2 in the proof. (Lemma A1 of AS2 is stated in Appendix A of AS2.) The verification of Assumption PS5 is the same as that of Assumption PS4 except that all arguments are conditional on the sample path ω (specified in Assumption PS5). Details are omitted for brevity.

Lemma D.5 *Let $(\Omega, \mathbb{F}, P)$ be a probability space and let ω denote a generic element in Ω. Suppose that the row-wise i.i.d. triangular arrays of random processes $\{\phi_{n,i}(\omega, g) : g \in \mathcal{G}, i \leq n, n \geq 1\}$ and $\{c_{n,i}(\omega, \tau) : \tau \in \mathcal{T}, i \leq n, n \geq 1\}$ are manageable with respect to the envelopes $\{F_n(\omega) : \Omega \to R^n : n \geq 1\}$ and $\{C_n(\omega) : \Omega \to R^n : n \geq 1\}$, respectively. Then, $\{\phi_{n,i}(\omega, g)c_{n,i}(\omega, \tau) : (\tau, g) \in \mathcal{T} \times \mathcal{G}, i \leq n, n \geq 1\}$ is manageable with respect to the envelopes $\{F_n(\omega) \odot C_n(\omega) : n \geq 1\}$, where $\odot$ stands for the coordinate-wise product.*

Proof of Lemma D.5. For a positive number ξ and a Euclidean space G, the packing number $D(\xi, G)$ is defined in Section B.3. For each $\omega \in \Omega$ and each $n \geq 1$, let $\mathcal{F}_{n,\omega} := \{(\phi_{n,1}(\omega, g), ..., \phi_{n,n}(\omega, g))' : g \in \mathcal{G}\}$, and let $\mathcal{C}_{n,\omega} := \{(c_{n,1}(\omega, \tau), ..., c_{n,n}(\omega, \tau))' : \tau \in \mathcal{T}\}$. Let $\lambda_\phi(\varepsilon)$ and $\lambda_c(\varepsilon)$ be the deterministic functions that (i) bound from above $D(\varepsilon||\alpha \odot F_n(\omega)||, \alpha \odot \mathcal{F}_{n,\omega})$ and $D(\varepsilon||\alpha \odot C_n(\omega)||, \alpha \odot \mathcal{C}_{n,\omega})$, respectively, for an arbitrary nonnegative n-vector α, and (ii) satisfy $\int_0^1 \sqrt{\log \lambda_\phi(x)}dx < \infty$ and $\int_0^1 \sqrt{\log \lambda_c(x)}dx < \infty$. Such functions exist by the assumed manageability of the triangular arrays of random processes in the lemma.

For an arbitrary $\varepsilon > 0$, construct a bound for $D(\varepsilon||\alpha \odot F_n(\omega) \odot C_n(\omega)||, \alpha \odot \mathcal{F}_{n,\omega} \odot \mathcal{C}_{n,\omega})$ as follows:

$$
\begin{aligned}
& D(\varepsilon||\alpha \odot F_n(\omega) \odot C_n(\omega)||, \alpha \odot \mathcal{F}_{n,\omega} \odot \mathcal{C}_{n,\omega}) \\
\leq\ & D((\varepsilon/4)||\alpha \odot F_n(\omega) \odot C_n(\omega)||, \alpha \odot F_n(\omega) \odot \mathcal{C}_{n,\omega}) \\
& \times D((\varepsilon/4)||\alpha \odot F_n(\omega) \odot C_n(\omega)||, \alpha \odot C_n(\omega) \odot \mathcal{F}_{n,\omega}) \\
\leq\ & \sup_{\alpha^* \in R_+^n} D((\varepsilon/4)||\alpha^* \odot C_n(\omega)||, \alpha^* \odot \mathcal{C}_{n,\omega}) \sup_{\alpha^* \in R_+^n} D((\varepsilon/4)||\alpha^* \odot F_n(\omega)||, \alpha^* \odot \mathcal{F}_{n,\omega}) \\
\leq\ & \lambda_\phi(\varepsilon/4)\lambda_c(\varepsilon/4),
\end{aligned} \tag{D.19}
$$

where the first inequality holds by the displayed equation following (5.2) in Pollard (1990), the second inequality holds because $\alpha \odot F_n(\omega), \alpha \odot C_n(\omega) \in R^n_+$, and the last inequality holds by the definitions of $\lambda_\phi(\varepsilon)$ and $\lambda_c(\varepsilon)$.

Then, the manageability of $\{\phi_{n,i}(\omega, g)c_{n,i}(\omega, g) : (\tau, g) \in \mathcal{T} \times \mathcal{G}, i \leq n, n \geq 1\}$ with respect to the envelopes $\{F_n(\omega) \odot C_n(\omega) : n \geq 1\}$ is proved by the following calculations:

$$\begin{aligned}\int_0^1 \sqrt{\log(\lambda_\phi(x/4)\lambda_c(x/4))}dx &\leq \int_0^1 \sqrt{\log \lambda_\phi(x/4)}dx + \int_0^1 \sqrt{\log \lambda_c(x/4)}dx \\ &= 4\int_0^{1/4} \sqrt{\log \lambda_\phi(y)}dy + 4\int_0^{1/4} \sqrt{\log \lambda_c(y)}dy \\ &\leq 4\int_0^1 \sqrt{\log \lambda_\phi(y)}dy + 4\int_0^1 \sqrt{\log \lambda_c(y)}dy \\ &< \infty, \end{aligned} \tag{D.20}$$

where the first inequality holds by the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b > 0$, the equality holds by a change of variables, the second inequality holds because the integrands are nonnegative on $(1/4, 1]$, and the last inequality holds by the definitions of $\lambda_\phi(\varepsilon)$ and $\lambda_c(\varepsilon)$. $\square$

E Power Against Fixed Alternatives

We now show that the power of the GMS test converges to one as $n \to \infty$ for all fixed alternatives. Thus, the test is a consistent test.

Recall that the null hypothesis is

$$\begin{aligned} H_0 : E_{F_0}[m_j(W_i, \theta_*, \tau)|X_i] &\geq 0 \text{ a.s. } [F_{X,0}] \text{ for } j = 1, ..., p \text{ and} \\ E_{F_0}[m_j(W_i, \theta_*, \tau)|X_i] &= 0 \text{ a.s. } [F_{X,0}] \text{ for } j = p+1, ..., k, \ \forall \tau \in \mathcal{T}, \end{aligned} \tag{E.1}$$

where θ_* denotes the null parameter value and F_0 denotes the fixed true distribution of the data. The alternative is that H_0 does not hold. Assumption MFA of ASM specifies the properties of fixed alternatives. For convenience, we restate this assumption here. Recall that $\mathcal{X}_F(\theta, \tau)$, defined in (6.2), is the set of points $x \in R^{d_x}$ such that under F there is a violation of some conditional moment inequality or equality, evaluated at (θ, τ), conditional on $X_i = x$.

Assumption MFA. The value $\theta_* \in \Theta$ and the true distribution F_0 satisfy: (a) for some

$\tau_* \in \mathcal{T}$, $P_{F_0}(X_i \in \mathcal{X}_{F_0}(\theta_*, \tau_*)) > 0$ and (b) $(\theta_*, F_0) \in \mathcal{F}_+$.

The following assumption requires the measure Q on $\mathcal{G}$ to have full support. For each $(\theta, F, \tau) \in \mathcal{F}_+ \times \mathcal{T}$, define a pseudometric on $\mathcal{G}$: $d_{(\theta,,F,\tau)}(g, g^\dagger) = \|E_F[m(W_i, \theta, \tau)(g(X_i) - g^\dagger(X_i))]\|$ for $g, g^\dagger \in \mathcal{G}$. Let $\mathcal{B}_{d_{(\theta,F,\tau)}}(g_0, \delta) = \{g \in \mathcal{G} : d_{(\theta,F,\tau)}(g, g_0) \leq \delta\}$.

Assumption MQ. The support of Q under $d_{(\theta,F,\tau)}$ is $\mathcal{G}$ for all $(\theta, F, \tau) \in \mathcal{F}_+ \times \mathcal{T}$. That is, $\forall(\theta, F, \tau) \in \mathcal{F}_+ \times \mathcal{T}$, $\forall\delta > 0$, and $\forall g_0 \in \mathcal{G}$, $Q(B_{d(\theta,F,\tau)}(g_0, \delta)) > 0$.

The following theorem shows that the GMS test is consistent against all fixed alternatives defined in Assumption MFA.

Theorem E.1 *Suppose Assumptions* PS4, PS5, MFA, CI, MQ, S1, S3, S4, *and* SIG2 *hold. Then,*

(a) $\lim_{n\to\infty} P_{F_0}(T_n(\theta_*) > c^*(\varphi_n(\theta_*), \widehat{h}^*_{2,n}(\theta_*), 1-\alpha) = 1$, *and*

(b) $\lim_{n\to\infty} P_{F_0}(T_n(\theta_*) > c^*(0^k_{\mathcal{T}\times\mathcal{G}}, \widehat{h}^*_{2,n}(\theta_*), 1-\alpha) = 1$.

Comments. 1. Theorem 6.1 of ASM for the case $r_{1,n} = \infty$ is proved by verifying that the conditions of Theorem E.1 (except Assumption MFA) hold for the $\mathcal{G}_{\text{c-cube}}$ set, the S functions, and the measure Q_{AR} defined as in ASM. (See Section B.5 for the definition of Q_{AR} with weight function $w(r) := (r^2 + 100)^{-1}$.) Assumption CI holds for $\mathcal{G}_{\text{c-cube}}$ defined in (3.6) of ASM by Lemma 3 of AS1. Assumption MQ holds for $\mathcal{G}_{\text{c-cube}}$ and Q_{AR} because $\mathcal{G}_{\text{c-cube}}$ is countable and Q_{AR} has a probability mass function that is positive at each element in $\mathcal{G}_{\text{c-cube}}$. Assumptions S1-S4 hold for the functions S_1, S_2, and S_3 defined in (3.9) of ASM by Lemma 1 of AS1, and for S_4 in (3.9) by similar arguments. Assumptions PS4 and PS5 hold by Lemma D.2 provided Assumption M holds. Assumption M holds for $\mathcal{G}_{\text{c-cube}}$ by Lemma 3 of AS1. (Note that Assumption M with F_0 in place of F_n in part (b) holds because $\mathcal{C}_{\text{c-cube}}$ is a Vapnik-Cervonenkis class of sets.)

2. Theorem 6.1 of ASM holds for $r_{1,n}$ such that $r_{1,n} < \infty$ and $r_{1,n} \to \infty$ as $n \to \infty$ by making some alterations to the proof of Theorem E.1. The alterations required are the same as those described for A-CvM tests in the proof of Theorem B2 in Appendix D of AS2.[22]

[22] The proof of Theorem B2 describes alterations to the proof of Theorem 3 of AS1, which is given in Appendix C of AS2, to accommodate A-CvM tests based on truncation, simulation, or quasi-Monte Carlo computation and KS tests. Theorem 3 of AS1 establishes that the tests in AS1 have asymptotic power equal to one for fixed alternative distributions.

Proof of Theorem E.1. Because $\varphi_n(\theta_*) \geq 0$ and $c^*(\cdot, \widehat{h}^*_{2,n}(\theta_*), 1-\alpha)$ is weakly decreasing by definition, we have that part (a) is implied by part (b). Therefore, it suffices to show part (b) only.

Let

$$A(\theta_*) := \sup_{\tau \in \mathcal{T}} \int_{\mathcal{G}} S(E_{F_0}[m(W_i, \theta_*, \tau) g(X_i)], \overline{\Sigma}_{F_0}(\theta_*, \tau, g) dQ(g). \tag{E.2}$$

First, we show that

$$|n^{-\chi/2} T_n(\theta_*) - A(\theta_*)| \to_p 0. \tag{E.3}$$

For any $\delta > 0$,

$$\begin{aligned}
&P_{F_0}(|n^{-\chi/2} T_n(\theta_*) - A(\theta_*)| > \delta) \\
\leq\ & P_{F_0}\left(\sup_{\substack{\mu \in [0,\infty) \\ (\tau,g) \in \mathcal{T} \times \mathcal{G}}} \left| S(n^{-1/2} \nu_{n,F_0}(\theta_*, \tau, g) + \mu, \widehat{h}^{\varepsilon}_{2,n,F_0}(\theta_*, \tau, g)) - S(\mu, h^{\varepsilon}_{2,F_0}(\theta_*, \tau, g)) \right| > \delta \right) \\
\leq\ & P_{F_0}\left(\sup_{(\tau,g) \in \mathcal{T} \times \mathcal{G}} ||n^{-1/2} \nu_{n,F_0}(\theta_*, \tau, g)|| + \sup_{(\tau,g) \in \mathcal{T} \times \mathcal{G}} ||\widehat{h}^{\varepsilon}_{2,n,F_0}(\theta_*, \tau, g) - h^{\varepsilon}_{2,F_0}(\theta_*, \tau, g)|| > \xi_\delta \right) \\
\to\ & 0 \text{ as } n \to \infty,
\end{aligned} \tag{E.4}$$

where $\widehat{h}^{\varepsilon}_{2,n,F_0}(\theta, \tau, g) := \widehat{h}_{2,n,F_0}(\theta, \tau, g) + \varepsilon D^{-1/2}_{F_0}(\theta_n) \widehat{D}_0(\theta_n) D^{-1/2}_{F_0}(\theta_n)$, $h^{\varepsilon}_{2,F_0}(\theta_*, \tau, g) := h_{2,F_0}(\theta_*, \tau, g) + \varepsilon I_k$, the first inequality uses Assumption S4, and the second inequality holds for some $\xi_\delta > 0$ by Assumptions PS4, S2 and SIG2. This establishes (E.3).

Next we show that $A(\theta_*) > 0$. By Assumption MFA, there exists a $\tau_* \in \mathcal{T}$ and either a $j_0 \leq p$ such that $P_{F_0}(E_{F_0}[m_{j_0}(W_i, \theta_*, \tau_*)|X_i] < 0) > 0$ or a $j_0 > p$ such that $P_{F_0}(E_{F_0}[m_{j_0}(W_i, \theta_*, \tau_*)|X_i] \neq 0) > 0$. Without loss of generality, assume that $j_0 \leq p$. By Assumption CI, there is a $g_* \in \mathcal{G}$ such that $E_{F_0}[m_{j_0}(W_i, \theta_*, \tau_*) g_{*j_0}(X_i)] < 0$, where $g_{*j_0}(X_i)$ denotes the j_0th element of $g_*(X_i)$.

Because $E_{F_0}[m_{j_0}(W_i, \theta_*, \tau_*) g_{*j_0}(X_i)]$ is continuous in g with respect to the pseudo-metric $d_{(\theta_*, F_0, \tau_*)}$, there exists a $\delta > 0$ such that $\forall g \in \mathcal{B}_{d(\theta_*, F_0, \tau_*)}(g_*, \delta)$, $E_{F_0}[m_{j_0}(W_i, \theta_*, \tau_*) g_{j_0}(X_i)]$ has the same sign as $E_{F_0}[m_{j_0}(W_i, \theta_*, \tau_*) g_{*j_0}(X_i)]$, i.e., $E_{F_0}[m_{j_0}(W_i, \theta_*, \tau_*) g_{j_0}(X_i)] < 0$, $\forall g \in \mathcal{B}_{d(\theta_*, F_0, \tau_*)}(g_*, \delta)$. By Assumption MQ, $Q(\mathcal{B}_{d(\theta_*, F_0, \tau_*)}(g_*, \delta)) > 0$. Therefore,

$$A(\theta_*) \geq \int_{\mathcal{B}_{d(\theta_*, F_0, \tau_*)}(g_*, \delta)} S(E_{F_0}[m(W_i, \theta_*, \tau_*) g(X_i)], \overline{\Sigma}_{F_0}(\theta_*, \tau_*, g) dQ(g) > 0, \tag{E.5}$$

where the second inequality holds by Assumption S3 and $Q(\mathcal{B}_{d(\theta_*,F_0,\tau_*)}(g_*,\delta)) > 0$.

Analogous arguments to those used to establish (14.34) of AS2 show

$$c^*(0^k_{\mathcal{T}\times\mathcal{G}}, \widehat{h}^*_{2,n}(\theta_*), 1-\alpha) = O_p(1). \tag{E.6}$$

Equations (E.3), (E.5), and (E.6) give

$$\begin{aligned} & P_{F_0}(T_n(\theta_*) > c^*(0^k_{\mathcal{T}\times\mathcal{G}}, \widehat{h}^*_{2,n}(\theta_*), 1-\alpha)) \\ = \ & P_{F_0}(n^{-\chi/2}T_n(\theta_*) > n^{-\chi/2}c^*(0^k_{\mathcal{T}\times\mathcal{G}}, \widehat{h}^*_{2,n}(\theta_*), 1-\alpha)) \\ = \ & P_{F_0}(A(\theta_*) + o_p(1) > o_p(1)) \\ \to \ & 1 \text{ as } n \to \infty, \end{aligned} \tag{E.7}$$

which establishes part (b). □

F Proofs of Results Concerning the Examples

Proof of Lemma 7.1. Assumption PS1(a) holds because $\Theta = \{0\}$. Assumptions PS1(b) holds by the condition given in the lemma. Assumption PS1(c) holds because $\sigma^2_{F,1}(0) = 1$. Assumption PS1(d) holds because $|1\{Y_2 \leq \tau\} - 1\{Y_1 \leq \tau\}| \leq 1$. Assumption PS1(e) holds because

$$E_F M^{2+\delta}(W) = 1/\sigma^{2+\delta}_F(0) = 1. \tag{F.1}$$

Next, we verify Assumption PS2. For $j = 1, 2$, consider the set $\mathcal{M}_{n,j,y_j} := \{(-1\{y_{j,i} \leq \tau\})^n_{i=1} \in R^n : \tau \in \mathcal{T}\}$ for an arbitrary realization $\{y_{j,i} : i \leq n\}$ of the random vector $\{Y_{j,i} : i \leq n\}$. The set has pseudo-dimension (defined on p. 15 of Pollard (1990)) at most one by Lemma 4.4 of Pollard (1990). Then, by Corollary 4.10 of Pollard (1990), there exist constants $c_1 \geq 1$ and $c_2 > 0$ (not depending on j, n, ε, or $\{y_{j,i} : i \leq n\}$) such that

$$D(\varepsilon||\alpha||, \alpha \odot \mathcal{M}_{n,j,y_j}) \leq c_1\varepsilon^{-c_2} \tag{F.2}$$

for $\varepsilon \in (0,1]$, every rescaling vector $\alpha \in R^n_+$, and $j = 1, 2$. In consequence, by the stability of the L_2 packing numbers (see Pollard (1990, p. 22)), we have

$$
\begin{aligned}
& D(2\varepsilon||\alpha||, (\alpha \odot \mathcal{M}_{n,1,y_1}) \oplus (\alpha \odot \mathcal{M}_{n,2,y_2})) \\
\leq\ & D(\varepsilon||\alpha||, \alpha \odot \mathcal{M}_{n,1,y_1}) D(\varepsilon||\alpha||, \alpha \odot \mathcal{M}_{n,2,y_2}) \\
\leq\ & c_1^2 \varepsilon^{-2c_2}, \qquad \text{(F.3)}
\end{aligned}
$$

where $A \oplus B = \{a + b : a \in A, b \in B\}$ for any two sets $A, B \subset R^n$.

Now consider the set $\mathcal{M}_{n,y_1,y_2} := \{(1\{y_{2,i} \leq \tau\} - 1\{y_{1,i} \leq \tau\})_{i=1}^n \in R^n : \tau \in \mathcal{T}\}$. By definition, $\alpha \odot \mathcal{M}_{n,y_1,y_2} \subset (\alpha \odot \mathcal{M}_{n,1,y_1}) \oplus (\alpha \odot \mathcal{M}_{n,2,y_2})$. Thus,

$$
D(2\varepsilon||\alpha||, \alpha \odot \mathcal{M}_{n,y_1,y_2}) \leq c_1^2 \varepsilon^{-2c_2}. \qquad \text{(F.4)}
$$

Lastly, because c_1 and c_2 do not depend on n or $\{(Y_{1i}, Y_{2i}) : i \leq n\}$, the manageability of $\{1\{Y_{2,i} \leq \tau\} - 1\{Y_{1,i} \leq \tau\} : \tau \in \mathcal{T}, i \leq n, n \geq 1\}$ holds by the following calculations:

$$
\int_0^1 \sqrt{\log(c_1^2 \varepsilon^{-2c_2})} d\varepsilon = \int_{\sqrt{\log(A)}}^{\infty} (2A^{1/W}/W) x^2 e^{-x^2/W} dx < \infty, \qquad \text{(F.5)}
$$

where $A := c_1^2$, $W := 2c_2$, $\log(A) \geq 0$ because $c_1 \geq 1$, and the equality holds by change of variables with $x = \sqrt{\log(A\varepsilon^{-W})}$ or, equivalently, $\varepsilon = A^{1/W} e^{-x^2/W}$, which yields $d\varepsilon = (2A^{1/W}/W) x e^{-x^2/W} dx$. This completes the proof. $\square$

Proof of Lemma 7.2. We prove part (a) first. Assumption PS1(a) holds because $\Theta = \{0\}$. Assumptions PS1(b) and PS1(c) hold by conditions (i) and (ii) of the lemma, respectively. Assumption PS1(d) holds because

$$
\begin{aligned}
& |(\tau - Y_2)^{s-1} 1\{Y_2 \leq \tau\} - (\tau - Y_1)^{s-1}) 1\{Y_1 \leq \tau\}| \\
\leq\ & (\tau - Y_2)^{s-1} 1\{Y_2 \leq \tau\} + (\tau - Y_1)^{s-1}) 1\{Y_1 \leq \tau\} \\
\leq\ & (B - Y_2)^{s-1} + (B - Y_1)^{s-1}. \qquad \text{(F.6)}
\end{aligned}
$$

Assumption PS1(e) holds because

$$
M(W) \leq 2(B - (-B))^{s-1} / \sigma_{F,1}(0) \leq 2^s B^{s-1} / \underline{\sigma}. \qquad \text{(F.7)}
$$

Next, we verify Assumption PS2. Consider the set $\mathcal{M}_{n,1,y_1} := \{(-(\tau - y_{1,i})^{s-1}1\{y_{1,i} \leq \tau\})_{i=1}^n \in R^n : \tau \in \mathcal{T}\}$ for an arbitrary realization $\{y_{1,i} : i \leq n\}$ of the random vector $\{Y_{1,i} : i \leq n\}$. First, we show that this set has pseudo-dimension (as defined in Pollard (1990, p. 15)) at most one. Suppose not. Then, there exists a vector $x = (x_1, x_2)' \in R^2$ and a pair (i, i') such that $\{(-(\tau - y_{1,i})^{s-1}1\{y_{1,i} \leq \tau\}, (-(\tau - y_{1,i'})^{s-1}1\{y_{1,i'} \leq \tau\}) : \tau \in \mathcal{T}\}$ surrounds x.[23] Thus, there exists $\tau_1, \tau_2 \in \mathcal{T}$ such that

$$\begin{aligned}
(\tau_1 - y_{1,i})^{s-1}1\{y_{1,i} \leq \tau_1\} &> x_1, \\
(\tau_1 - y_{1,i'})^{s-1}1\{y_{1,i'} \leq \tau_1\} &< x_2, \\
(\tau_2 - y_{1,i})^{s-1}1\{y_{1,i} \leq \tau_2\} &< x_1, \text{ and} \\
(\tau_2 - y_{1,i'})^{s-1}1\{y_{1,i'} \leq \tau_2\} &> x_2. \qquad \text{(F.8)}
\end{aligned}$$

This yields

$$\begin{aligned}
(\tau_1 - y_{1,i})^{s-1}1\{y_{1,i} \leq \tau_1\} &> (\tau_2 - y_{1,i})^{s-1}1\{y_{1,i} \leq \tau_2\} \text{ and} \\
(\tau_1 - y_{1,i'})^{s-1}1\{y_{1,i'} \leq \tau_1\} &< (\tau_2 - y_{1,i'})^{s-1}1\{y_{1,i'} \leq \tau_2\}. \qquad \text{(F.9)}
\end{aligned}$$

Due to the monotonicity of the function $G_s(y, \tau) := (\tau - y)^{s-1}1\{y \leq \tau\}$ in τ for any y, the first inequality in the equation above implies that $\tau_1 > \tau_2$, and the second inequality implies that $\tau_1 < \tau_2$, which is a contradiction. Therefore, $\mathcal{M}_{n,1,y_1}$ has pseudo-dimension at most one.

The remainder of the proof of part (a) is the same as the corresponding part of the proof of Lemma 7.1 and, hence, for brevity, is omitted.

To prove part (b), consider an arbitrary sequence $\{F_n : n \geq 1\}$ such that $(0, F_n) \in \mathcal{F}_+$

[23] As defined in Pollard (1990, p. 15), a set $A \subset R^2$ *surrounds* x if there exists points $a, b, c, d \in A$, where $a = (a_1, a_2)'$ etc., such that $a_1 > x_1$, $a_2 > x_2$, $b_1 > x_1$, $b_2 < x_2$, $c_1 < x_1$, $c_2 > x_2$, $d_1 < x_1$, and $d_2 < x_2$.

for all n. Under this sequence, we have for any $\zeta > 0$ and $j = 1, 2$,

$$
\begin{aligned}
\Pr_{F_n}(|\overline{Y}_{j,n} - E_{F_n}(Y_j)| > \zeta) &\leq \frac{E_{F_n}(\overline{Y}_{j,n} - E_{F_n}(Y_j))^2}{\zeta^2} \\
&= \frac{E_{F_n}(Y_j - E_{F_n}(Y_j))^2}{n\zeta^2} \\
&\leq \frac{E_{F_n}(Y_j^2)}{n\zeta} \\
&\leq B^2/(n\zeta) \to 0 \text{ as } n \to \infty,
\end{aligned} \tag{F.10}
$$

where the last inequality holds because the support of Y_j is contained in $\mathcal{T}$ and $\mathcal{T}$ is contained in $[-B, B]$ by condition (iii) of the lemma. Similarly, we have under the sequence $\{F_n : n \geq 1\}$,

$$
n^{-1}\sum_{i=1}^{n}(Y_{j,i} - E_{F_n}(Y_j))^{2(s-1)} - E_{F_n}(Y_j - E_{F_n}(Y_j))^{2(s-1)} \to_p 0, \tag{F.11}
$$

for $j = 1, 2$. Therefore, we have

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(Y_{j,i} - \overline{Y}_{j,n})^{2(s-1)} - E_{F_n}(Y_j - E_{F_n}(Y_j))^{2(s-1)} \\
&= n^{-1}\sum_{i=1}^{n}(Y_{j,i} - E_{F_n}(Y_j))^{2(s-1)} - E_{F_n}(Y_j - E_{F_n}(Y_j))^{2(s-1)} \\
&\quad + \sum_{b=0}^{2(s-1)-1}\binom{2(s-1)}{b}\left[n^{-1}\sum_{i=1}^{n}(Y_{j,i} - E_{F_n}(Y_j))^b\right](E_{F_n}(Y_j) - \overline{Y}_{j,n})^{2s-2-b}, \\
&= o_p(1) \\
&\quad + \sum_{b=0}^{2(s-1)-1}\binom{2(s-1)}{b}\left[n^{-1}\sum_{i=1}^{n}(Y_{j,i} - E_{F_n}(Y_j))^b\right](E_{F_n}(Y_j) - \overline{Y}_{j,n})^{2s-2-b}, \\
&= o_p(1),
\end{aligned} \tag{F.12}
$$

where the second equality holds by (F.11), and the last equality holds by (F.10) and the boundedness of Y_j.

Therefore,

$$\begin{aligned}|\widehat{\sigma}^2_{n,1}(0) - \sigma^2_{F_n,1}(0)|/\sigma^2_{F_n,1}(0) &\leq \underline{\sigma}^{-2}|\widehat{\sigma}^2_{n,1}(0) - \sigma^2_{F_n,1}(0)| \\ &\leq \underline{\sigma}^{-2}\sum_{j=1}^{2}\left|n^{-1}\sum_{i=1}^{n}(Y_{j,i} - \overline{Y}_{j,n})^{2(s-1)} - E_{F_n}(Y_j - E_{F_n}(Y_j))^{2(s-1)}\right| \\ &\rightarrow_p 0, \text{ as } n \rightarrow \infty. \end{aligned} \tag{F.13}$$

Because this holds for an arbitrary sequence $\{F_n : n \geq 1\}$ such that $(0, F_n) \in \mathcal{F}_+$, it establishes both Assumptions SIG1 and SIG2. Thus, part (b) holds. $\square$

Proof of Lemma 8.1. We show that parts (a) and (b) are equivalent by solving a linear programming problem. We show that parts (b) and (c) are equivalent by employing the convex polyhedral cone representation of linear inequalities developed in Gale (1951).

First, we show the equivalence between parts (a) and (b).

For a set $A \subset R^{d_\beta}$, let A^c denote the complement of A in R^{d_β}. By basic set operations, the statement in part (a) is equivalent to

$$\cap_{j=1}^{m} H(c_j)^c \subset H(\overline{c})^c. \tag{F.14}$$

Because m is finite and $H(c)^c$ is an open set for any $c \in R^{d_\beta}\backslash\{0\}$, (F.14) is equivalent to

$$\cap_{j=1}^{m} cl(H(c_j)^c) \subset cl(H(\overline{c})^c). \tag{F.15}$$

Note that $cl(H(c)^c) = \{b \in R^{d_\beta} : b'c \leq 0\}$ and (F.15) is equivalent to the redundancy of the inequality restriction $b'\overline{c} \leq 0$ on b relative to the system of linear inequalities $b'c_j \leq 0$ for $j = 1, ..., m$. In turn, the latter is equivalent to the statement that $V = 0$, where

$$V := \max_{b \in R^{d_\beta}} b'\overline{c} \text{ subject to } b'c_j \leq 0 \text{ for } j = 1, ..., m \text{ and } b'\overline{c} \leq 1. \tag{F.16}$$

Now we solve the linear programming problem in (F.16) using the Lagrange multiplier method. It is well known that

$$V = \min_{\lambda_j \geq 0 : j=1,...,m+1} \max_{b \in R^{d_\beta}} \left(b'\overline{c} - \sum_{j=1}^{m}\lambda_j b'c_j - \lambda_{m+1}(b'\overline{c} - 1)\right). \tag{F.17}$$

Because the maximization over b is unconstrained, for any $\lambda_1, ..., \lambda_{m+1} \geq 0$ such that $(1 - \lambda_{m+1})\overline{c} - \sum_{j=1}^m \lambda_j c_j \neq 0$, the maximum is infinite. But, $V \leq 1$ by the inequality $b'\overline{c} \leq 1$ in (F.16). Thus, the optimal $\lambda_1, ..., \lambda_{m+1}$ must satisfy

$$(1 - \lambda_{m+1})\overline{c} - \sum_{j=1}^{m} \lambda_j c_j = 0. \tag{F.18}$$

This implies that

$$V = \min_{\lambda_j \geq 0: j=1,...,m+1} \lambda_{m+1} \text{ subject to } (1 - \lambda_{m+1})\overline{c} - \sum_{j=1}^{m} \lambda_j c_j = 0. \tag{F.19}$$

Now we show that $V = 0$ iff there exist $\lambda_1, ..., \lambda_m \geq 0$ such that

$$\overline{c} = \sum_{j=1}^{m} \lambda_j c_j, \tag{F.20}$$

which establishes the equivalence between parts (a) and (b). Suppose that there exists $\lambda_1, ..., \lambda_m \geq 0$ such that (F.20) holds, then $V \leq \min\{\lambda_{m+1} \geq 0 : \lambda_{m+1}\overline{c} = 0^{d_\beta}\} = 0$ by (F.19). However, by (F.16) (with $b = 0^{d_\beta}$), $V \geq 0$. Thus, $V = 0$. Conversely, suppose that $V = 0$, then there exists $\lambda_1, ..., \lambda_m \geq 0$ such that $(1-0)\overline{c} - \sum_{j=1}^m \lambda_j c_j = 0^{d_\beta}$ by (F.19), which implies (F.20).

Next, we establish the equivalence between parts (b) and (c). Using the terminology in Gale (1951), let $P(c_1, ..., c_m)$ denote the convex polyhedral cone generated by the vectors $(c_1, ..., c_m)$. That is,

$$P(c_1, ..., c_m) := \left\{ c \in R^{d_\beta} : c = \sum_{j=1}^{m} \lambda_j c_j \text{ for some } \lambda_1, ..., \lambda_m \geq 0 \right\}.$$

Then, part (b) is equivalent to $\overline{c} \in P(c_1, ..., c_m)$.

If $rk([c_1, ..., c_m]) = d_\beta$, then by Weyl's Theorem (see Theorem 1 of Gale (1951)), $P(c_1, ..., c_m)$ is an intersection of at most $\binom{m}{d_\beta - 1}$ half-spaces or, in other words, there exist $b^1, ..., b^N \in R^{d_\beta}$, where $N := \binom{m}{d_\beta - 1}$, such that $P(c_1, ..., c_m) = \{c : [b^1, \cdots, b^N]'c \geq 0\}$. Then, the equivalence between parts (b) and (c) is established with $B(c_1, ..., c_m) := [b^1, \cdots,$ $b^N, 0^{d_\beta \times (M-N)}]$ for an arbitrary $[b^1, \cdots, b^N]$ that satisfies $P(c_1, ..., c_m) = \{c : [b^1, \cdots, b^N]'c \geq$

$0\}$.

If $rk([c_1, ..., c_m]) < d_\beta$, let $L(c_1, ..., c_m)$ be the linear subspace spanned by $c_1, ..., c_m$. Let the dimension of this linear subspace be d_L. Applying Weyl's Theorem on $L(c_1, ..., c_m)$, we have that there exist $b^1, ..., b^{N_1} \in R^{d_\beta}$, where $N_1 := \binom{m}{d_L - 1}$, such that $P(c_1, ..., c_m) = \{c : [b^1, \cdots, b^{N_1}]'c \geq 0\} \cap L(c_1, ..., c_m)$. Moreover, by the property of linear subspaces, there exist $b^{N_1+1}, ..., b^{N_2} \in R^{d_\beta}$, where $N_2 := N_1 + d_\beta - d_L$, such that $L(c_1, ..., c_m) = \{c : [b^{N_1+1}, \cdots, b^{N_2}]'c = 0\}$. Therefore,

$$P(c_1, ..., c_m) = \{c : [b^1, \cdots, b^{N_2}, -b^{N_1+1}, \cdots, -b^{N_2}]'c \geq 0\}. \tag{F.21}$$

Then, the equivalence between parts (b) and (c) holds with $B(c_1, ..., c_m) := [b^1, \cdots, b^{N_2}, -b^{N_1+1}, \cdots, -b^{N_2}, 0^{d_\beta \times (M - 2N_2 + N_1)}]$ for arbitrary $b^1, ..., b^{N_2}$ that satisfy (F.21). $\square$

Proof of Lemma 8.2. Assumption PS1(a) holds because $\theta \in \Theta$. Assumptions PS1(b) holds by the i.i.d. condition in the lemma. Assumption PS1(c) holds by $\sigma^2_{F,1}(\theta) = 1$. Assumption PS1(d) holds because $|F_\beta(\mathcal{S}, \theta) - 1\{S(Y_1, Y_2, X_1) \subset \mathcal{S}\}| \leq 1$. Assumption PS1(e) holds because

$$E_F M^{2+\delta}(W) = 1/\sigma^{2+\delta}_{F,1}(0) = 1. \tag{F.22}$$

Next, we verify Assumption PS2. Consider the set $\mathcal{M}_n := \{(F_\beta(\mathcal{S}(\tau), \theta) - 1\{(y_{1i} - 1/2)B(\tau)'(1, x'_{1i}, y'_{2i})' \geq 0\})_{i=1}^n \in R^n : \tau = (c_1, \ldots, c_m), c_1, \ldots, c_m \in R^{d_\beta} \backslash \{0^{d_\beta}\}\}$ for an arbitrary realization $\{(y_{1,i}, y'_{2,i}, x'_{1,i})' : i \leq n\}$ of $\{(Y_{1,i}, Y'_{2,i}, X'_{1,i})' : i \leq n\}$. The set has pseudo-dimension at most M by Lemma 4.4 of Pollard (1990). Then, by Corollary 4.10 of Pollard (1990), there exist constants $c_1 \geq 1$ and $c_2 > 0$ (not depending on n, $\{(y_{1,i}, y'_{2,i}, x'_{1,i})' : i \leq n\}$, or ε) such that

$$D(\varepsilon \|\alpha\|, \alpha \odot \mathcal{M}_n) \leq c_1 \varepsilon^{-c_2} \tag{F.23}$$

for all $0 < \varepsilon \leq 1$ and every rescaling vector $\alpha \in R^n_+$. In consequence, the manageability of $\{(F_\beta(\mathcal{S}(\tau), \theta) - 1\{Y_{1,i} - 1/2)B(\tau)'(1, X'_{1,i}, Y'_{2,i})' \geq 0\})_{i=1}^n \in R^n : \tau = (c_1, \ldots, c_m), c_1, \ldots, c_m \in R^{d_\beta} \backslash \{0^{d_\beta}\}\}$ follows from the calculations in (F.5) with $A := c_1$ and $W := c_2$, which completes the proof. $\square$

Proof of Lemma 9.1 Assumptions PS1(a)-(c) and (e) hold by assumption. Next, we show

Assumption PS1(d) holds. We have

$$|h(Q_\theta(w,v),u)| = |\sup_{q\in Q_\theta(w,v)} q'u| \le \sup_{q\in Q_\theta(w,v)} \|q\|\|u\| \le \sup_{q\in Q_\theta(w,v)} \|q\| \le M(w)/2, \qquad \text{(F.24)}$$

where the first inequality holds by the Cauchy-Schwarz inequality, the second inequality holds because u satisfies $\|u\| \le 1$, and the last inequality holds by condition (iii) of the lemma. Assumption PS1(d) follows from the following calculations:

$$\begin{aligned}\left|\int h(Q_\theta(W,v),u)dF_{V|X}(v,X;\theta) - u'q(X)\right| &\le \int |h(Q_\theta(W,v),u)|dF_{V|X}(v,X;\theta) + |u'q(X)| \\ &\le M(W)/2 + |u'q(X)| \\ &\le M(W)/2 + \|u\|\|q(X)\| \\ &\le M(W), \qquad \text{(F.25)}\end{aligned}$$

where the second inequality holds by (F.24), the third inequality holds by the Cauchy-Schwarz inequality and the last inequality holds by $\|u\| \le 1$ and condition (iv) of the lemma.

Now, we show that Assumption PS2 holds. Let $m(W,\theta,u) := \int h(Q_\theta(W,v),u) dF_{V|X}(v,X;\theta) - u'q(X)$. Consider an arbitrary sequence (θ_n, F_n) that satisfy the conditions in the lemma. Arguments similar to those for Assumption PS1(d) above show that $m(W,\theta,u)$ is Lipschitz continuous in u with Lipschitz constant $M(W)$ for all θ. Given the Lipschitz continuity, for any nonnegative n-vector $\alpha := (\alpha_1,\ldots,\alpha_n)$, any $u_1 \in R^d$, $u_2 \in R^d$ such that $\|u_1\| \le 1$ and $\|u_2\| \le 1$, and any n realizations $(w_1,\ldots,w_n)$ of W (under F_n), we have

$$\sum_{i=1}^n (\alpha_i m(w_i,\theta_n,u_1) - \alpha_i m(w_i,\theta_n,u_2))^2 \le \left(\sum_{i=1}^n (\alpha_i M(w_i))^2\right) \|u_1 - u_2\|^2. \qquad \text{(F.26)}$$

Let $\mathcal{F}_{n(w_1,\ldots,w_n)} = \{(m(w_i,\theta_n,u)_{i=1}^n : u \in R^d, \|u\| \le 1\}$ and let $\vec{M}_n(w_1,\ldots,w_n) = (M(w_1),\ldots, M(w_n))'$. Then, (F.26) implies that, for all $\xi \in (0,1]$,

$$D(\xi\|\alpha \odot \vec{M}_n(w_1,\ldots,w_n)\|, \alpha \odot \mathcal{F}_{n(w_1,\ldots,w_n)}) \le D(\xi, \{u \in R^d : \|u\| \le 1\}) \le C/\xi^d, \qquad \text{(F.27)}$$

for some constant $C < \infty$. Assumption PS2 holds because $\int_0^1 \sqrt{\log C - d\log\xi}d\xi = 2C^{1/d}d^{-1}$

$\int_0^\infty x^2 e^{-x^2/d} dx < \infty$. □

Proof of Lemma 9.2. We prove part (a) first. Assumptions PS1(a)-(e) hold by assumption. Now we verify Assumption PS2.

Let (θ_n, F_n) be an arbitrary sequence that satisfies all the conditions of the lemma. Consider n arbitrary realizations $(w_1, \ldots, w_n)$ of W under F_n for arbitrary $n \geq 1$. By (9.2) and conditions (iii) and (iv) of the lemma, $\{M(w_i) : i \leq n, n \geq 1\}$ are envelopes for the triangular array of processes $\{m(w_i, \theta_n, \tau) : i \leq n, n \geq 1, \tau = 1, 2, \ldots\}$. Let

$$\mathcal{F}_{n(w_1,\ldots,w_n)} = \{(m(w_1, \theta_n, \tau), \ldots, m(w_n, \theta_n, \tau))' : \tau = 1, 2, \ldots\}. \tag{F.28}$$

Let $\vec{M}_n(w_1, \ldots, w_n) = (M(w_1), \ldots, M(w_n))'$. Then, for any $\xi \in (0, 1]$ and any nonnegative n-vector α, $D(\xi \|\alpha \odot \vec{M}_n(w_1, \ldots, w_n)\|, \alpha \odot \mathcal{F}_{n(w_1,\ldots,w_n)}) \leq \lambda_{\mathcal{T}}(\xi)$ because $\alpha \odot (m(w_1, \theta_n, \tau), \ldots,$ $m(w_n, \theta_n, \tau))'$ belongs to the $\xi \|\alpha \odot \vec{M}_n(w_1, \ldots, w_n)\|$-neighborhood of 0^n for all $\tau \geq \lambda_{\mathcal{T}}(\xi)$. The latter holds because, for all $\tau \geq \lambda_{\mathcal{T}}(\xi)$,

$$\begin{aligned}
\sum_{i=1}^n (\alpha_i m(w_i, \theta, \tau))^2 &= w_{\mathcal{T}}^2(\tau) \sum_{i=1}^n (\alpha_i \widetilde{m}(w_i, \theta, \tau))^2 \\
&\leq w_{\mathcal{T}}^2(\tau) \sum_{i=1}^n (\alpha_i M(w_i))^2 \\
&\leq \xi^2 \|\alpha \odot \vec{M}_n(w_1, \ldots, w_n)\|^2,
\end{aligned} \tag{F.29}$$

where the last inequality holds because $\tau \geq \lambda_{\mathcal{T}}(\xi)$ iff $w_{\mathcal{T}}(\tau) \leq \xi$ (because $\lambda_{\mathcal{T}}(\xi)$ is the inverse function of the decreasing function $w_{\mathcal{T}}(\xi)$). By assumption, $\int_0^1 \sqrt{\log(\lambda_{\mathcal{T}}(\xi))} d\xi < \infty$. Hence, Assumption PS2 holds.

The proof of part (b) is similar to that of Lemma 7.2(b) and is omitted for brevity. □

G Additional Simulation Results

In this section we report additional Monte Carlo simulation results that investigate the sensitivity of the performance of the MCMI tests to different choices of tuning parameters.

Tables 3-6 report the null and alternative hypothesis rejection probabilities of the CvM/GMS and KS/GMS tests in the first-order stochastic dominance example studied in Section 7.2 for various choices of the tuning parameters. The first two tables cover null data generating processes for which first-order stochastic dominance holds. The last two tables cover alternative data generating processes for which first-order stochastic dominance does not hold. These data generating processes are the same as those considered in Tables 1 and 2.

In each table, we report the rejection probabilities of the nominal .05 level CvM/GMS and KS/GMS tests for a base case specification of the tuning parameters and 12 variants of the base case. The base case sets $r_{1,n} = 3$, $N_\tau = 25$, $\kappa_n = (0.3\ln(n))^{1/2}$, $B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}$, $\varepsilon = 0.01$, and Weight Constant $= 100$, where the weight constant is the constant added to r^2 in (3.7). The larger the weight constant is, the more weight is given to smaller hypercubes relative to the larger cubes. In each variant of the base case, one and only one tuning parameter is changed in order to isolate the effect of that tuning parameter on the performance of the tests. The power results in Tables 5 and 6 are size-corrected based on the data generating process for Table 3 under which the tests have asymptotic null rejection probabilities equal to .05.

As one can see from the tables, the base case specification performs well compared to the other specifications. Overall, the sensitivity to the choice of tuning parameters is not large for the range of tuning parameter values that is considered.

Table 3: Sensitivity to Tuning Parameters for First-Order Stochastic Dominance Tests – Null 1: $(c_1, c_2, c_3, c_4) = (0, 0, 0.85, 0.6)$

	CvM/GMS	KS/GMS
Base case	.057	.064
$r_{1,n} = 2$	.056	.054
$r_{1,n} = 4$	.059	.055
$N_\tau = 20$	.061	.074
$N_\tau = 30$	.062	.068
$\kappa_n = (0.3\ln(n))^{1/2}/2$	.070	.077
$\kappa_n = 2(0.3\ln(n))^{1/2}$	.052	.055
$B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}/2$	.057	.064
$B_n = 2(0.4\ln(n)/\ln(\ln(n)))^{1/2}$	.057	.064
$\varepsilon = 0.005$	.056	.065
$\varepsilon = 0.02$	.058	.056
Weight Constant = 50	.056	.067
Weight Constant = 200	.057	.064

Note: Sample size $n = 250$. The critical values use 1000 repetitions. Nominal size = .05. Base case specifications: $r_{1,n} = 3$, $N_\tau = 25$, $\kappa_n = (0.3\ln(n))^{1/2}$, $B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}$, $\varepsilon = 0.01$, Weight Constant = 100.

Table 4: Sensitivity to Tuning Parameters for First-Order Stochastic Dominance Tests – Null 2: $(c_1, c_2, c_3, c_4) = (0.15, 0, 0.85, 0.6)$

	CvM/GMS	KS/GMS
Base case	.014	.019
$r_{1,n} = 2$	.016	.018
$r_{1,n} = 4$	.014	.019
$N_\tau = 20$	.015	.015
$N_\tau = 30$	.018	.019
$\kappa_n = (0.3\ln(n))^{1/2}/2$	.026	.029
$\kappa_n = 2(0.3\ln(n))^{1/2}$	.010	.011
$B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}/2$	.013	.019
$B_n = 2(0.4\ln(n)/\ln(\ln(n)))^{1/2}$	.016	.019
$\varepsilon = 0.005$	.014	.018
$\varepsilon = 0.02$	.014	.017
Weight Constant = 50	.014	.019
Weight Constant = 200	.014	.019

Note: Sample size $n = 250$. The critical values use 1000 repetitions. Nominal size = .05. Base case specifications: $r_{1,n} = 3$, $N_\tau = 25$, $\kappa_n = (0.3\ln(n))^{1/2}$, $B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}$, $\varepsilon = 0.01$, Weight Constant = 100.

Table 5: Sensitivity to Tuning Parameters for First-Order Stochastic Dominance Tests – Alternative 1: $(c_1, c_2, c_3, c_4) = (-0.25, 0.2, 0.85, 0.6)$

	CvM/GMS	KS/GMS
Base case	.505	.379
$r_{1,n} = 2$	.513	.411
$r_{1,n} = 4$	.509	.367
$N_\tau = 20$	.486	.359
$N_\tau = 30$	.470	.405
$\kappa_n = (0.3\ln(n))^{1/2}/2$	.493	.363
$\kappa_n = 2(0.3\ln(n))^{1/2}$	.507	.384
$B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}/2$	.508	.379
$B_n = 2(0.4\ln(n)/\ln(\ln(n)))^{1/2}$	.506	.379
$\varepsilon = 0.005$	.505	.381
$\varepsilon = 0.02$	.504	.392
Weight Constant = 50	.505	.379
Weight Constant = 200	.506	.379

Note: Sample size $n = 250$. The critical values use 1000 repetitions. Nominal size = .05. Base case specifications: $r_{1,n} = 3$, $N_\tau = 25$, $\kappa_n = (0.3\ln(n))^{1/2}$, $B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}$, $\varepsilon = 0.01$, Weight Constant = 100.

Table 6: Sensitivity to Tuning Parameters for First-Order Stochastic Dominance Tests – Alternative 2: $(c_1, c_2, c_3, c_4) = (0.35, 0, 0.85, 0.23)$

	CvM/GMS	KS/GMS
Base case	.581	.295
$r_{1,n} = 2$	.628	.397
$r_{1,n} = 4$	.609	.246
$N_\tau = 20$	.560	.275
$N_\tau = 30$	.539	.309
$\kappa_n = (0.3\ln(n))^{1/2}/2$	.590	.318
$\kappa_n = 2(0.3\ln(n))^{1/2}$	.463	.159
$B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}/2$	.532	.258
$B_n = 2(0.4\ln(n)/\ln(\ln(n)))^{1/2}$	.589	.301
$\varepsilon = 0.005$	.615	.302
$\varepsilon = 0.02$	.529	.268
Weight Constant = 50	.580	.295
Weight Constant = 200	.582	.295

Note: Sample size $n = 250$. The critical values use 1000 repetitions. Nominal size = .05. Base case specifications: $r_{1,n} = 3$, $N_\tau = 25$, $\kappa_n = (0.3\ln(n))^{1/2}$, $B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}$, $\varepsilon = 0.01$, Weight Constant = 100.

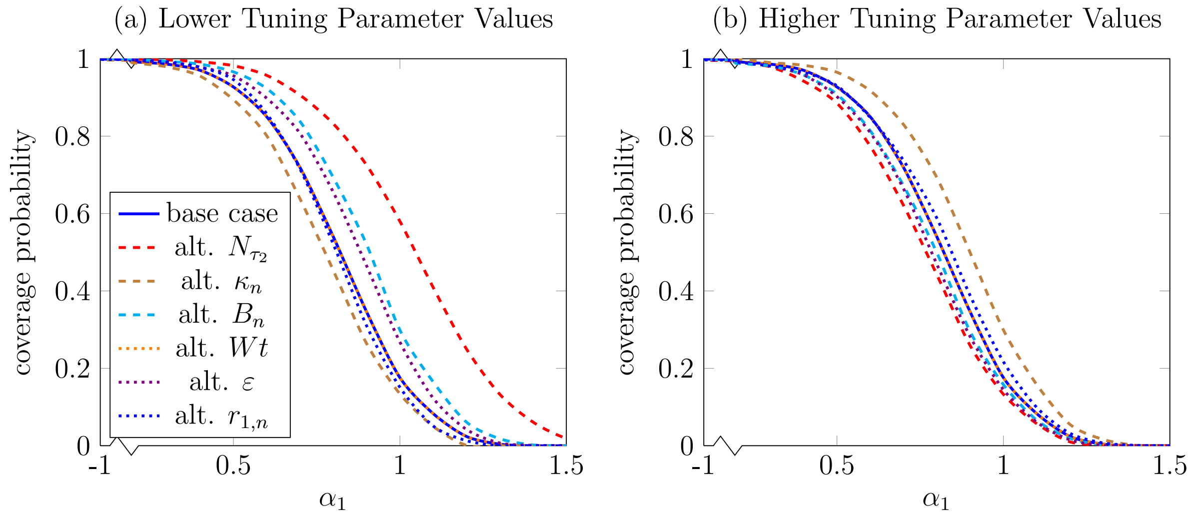

Figure 3: Sensitivity to Tuning Parameters for the KS/GMS CS Applied on the IV Random-Coefficients Binary-Outcome Model. (Nominal size = .95, $n = 250$, $(\alpha_0, \gamma_0, \gamma_1) = (0, -1, 1)$, and $(\alpha_0, \alpha_1, \gamma_0, \gamma_1)$ is in the identified set if and only if $\alpha_1 \leq -0.8274$.)

In the random-coefficient binary-outcome example studied in Section 8.2, the KS/GMS CS is the best performing CS. For this reason, we perform sensitivity analysis on this CS. Figure 3 reports the coverage probabilities of the KS/GMS CS under various choices of the tuning parameters. The base case values for $r_{1,n}$, κ_n, B_n, ε, and the Weight Constant are the same as they are in the stochastic dominance example. The base case value for N_{τ_2} is 15, which results in 120 grid points on $\mathcal{T}$. In the sensitivity analysis, we alter one and only one tuning parameter each time, and recompute the coverage probabilities.

Graph (a) of Figure 3 depicts the coverage probabilities when the tuning parameters are altered to lower values. The lower values are: $r_{1,n} = 2$, $N_{\tau_2} = 10$, $\kappa_n = (0.3\ln(n))^{1/2}/2$, $B_n = (0.4\ln(n)/\ln(\ln(n)))^{1/2}/2$, $\varepsilon = 0.005$, and Weight Constant = 50. Graph (b) depicts the coverage probabilities when they are altered to higher values. The higher values are: $r_{1,n} = 4$, $N_{\tau_2} = 20$, $\kappa_n = 2(0.3\ln(n))^{1/2}$, $B_n = 2(0.4\ln(n)/\ln(\ln(n)))^{1/2}$, $\varepsilon = 0.02$, and Weight Constant = 200.

Recall that the coverage probabilities are computed for fixed values of $(\alpha_0, \alpha_1, \gamma_0, \gamma_1)$. The fixed values considered are $(0, \alpha_1, -1, 1)$ for α_1 running from -1 (its true value), to -0.8274 (right boundary of the identified set at $(\alpha_0, \gamma_0, \gamma_1) = (0, -1, 1)$), and finally to 1.5. The coverage probabilities when $\alpha_1 \leq -0.8274$ should ideally be 0.95 or higher, and those when $\alpha_1 > -0.8274$ are false coverage probabilities, and should ideally be lower than 0.95 and get lower as α_1 moves to the right.

As one can see, the base case specification performs well relative to the other specifications. The coverage probabilities are generally insensitive to the tuning parameter changes considered, except when N_{τ_2} is changed. Coarser τ grids result in noticeably higher (worse) false coverage probabilities.

The sensitivity analysis sheds some light on the choice of the grid points for approximating $\mathcal{T}$. In both examples, we find that the size of the test or the coverate probability of the CS for points in the identified set is insensitive to $\mathcal{T}$, but the power or false coverage probability is noticeably worse when an insufficient number of grid points is used. Based on this, we recommend increasing the number of grid points until the test statistic value stabilizes, or increasing it to the maximum that the computational resources of the user allows.

References

Andrews D. W. K., Shi, X., 2013a. Inference based on conditional moment inequalities. Econometrica 81, 609–666.

Andrews D. W. K., Shi, X., 2013b. Supplemental material to 'Inference based on conditional moment inequalities.' Available at Econometrica Supplemental Material 81, http://www.econometricsociety.org/ecta/supmat/9370_Proofs.pdf.

Andrews D. W. K., Shi, X., 2014. Nonparametric inference based on conditional moment inequalities. Journal of Econometrics 179, 31–45. Appendix of proofs available as Cowles Foundation Discussion Paper No. 1840RR, Yale University, 2011.

Chernozhukov, V., D. Chetverikov, and K. Kato, 2014. Testing many moment inequalities. Department of Economics, MIT. Unpublished manuscript.

Gale, D., 1951. Convex polyhedral cones and linear inequalities. Chapter XVII in Activity Analysis of Production and Allocation, ed. by T. C. Koopmans, pp. 287–297. Wiley, New York.

Pollard, D., 1990. Empirical Process Theory and Application. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. II. Institute of Mathematical Statistics.