Supplement to

NONPARAMETRIC INFERENCE BASED ON CONDITIONAL MOMENT INEQUALITIES

By

Donald W. K. Andrews and Xiaoxia Shi

December 2011
Revised February 2013

COWLES FOUNDATION DISCUSSION PAPER NO. 1840R

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

http://cowles.econ.yale.edu/

This Appendix is not to be published.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

Appendix

to

Nonparametric Inference Based on 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: February 2013

5 Outline

This Appendix provides proofs of Theorems N1 and N2 of Andrews and Shi (2013a) "Nonparametric inference based on conditional moment inequalities," referred to hereafter as ASN. In fact, the results given here cover a much broader class of test statistics than is considered in ASN.

This Appendix is organized as follows. Section 6 defines the class of Cramér-von Mises (CvM) test statistics that are considered. This class includes the statistics that are considered in ASN. Section 7 introduces generalized moment selection (GMS) and plug-in asymptotic (PA) critical values, confidence sets (CS's), and tests. Section 8 establishes the correct asymptotic size of GMS and PA CS's. Theorem N1 of ASN is a corollary to Theorem AN1, which is given in Section 8. Section 9 establishes that GMS and PA 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 N2 of ASN is a corollary to Theorem AN2, which is given in Section 9. Section 10 shows that GMS and PA tests have nontrivial power against some, but not all, $(nb^{d_z})^{1/2}$-local alternatives. Section 11 derives local power results for the KS and CvM tests that cover the case where the DGP does not depend on n and the moment inequalities are binding only on a measure-zero set of X_i. It uses these results to compare the asymptotic power of the KS and CvM tests (in terms of rates of convergence) with that of the CLR test in a simple moment inequality model.

Section 12 provides proofs of the results given in this Appendix. Section 13 provides some additional simulation results to those given in that paper.

We let AS1 and AS2 abbreviate Andrews and Shi (2013b) and Andrews and Shi (2013c), respectively.

6 General Form of the Test Statistic

6.1 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 and $Z_i = z_0$ into equivalent conditional moment inequalities/equalities given only $Z_i = z_0$ by choosing appropriate weighting functions of X_i, i.e., X_i-instruments. Then, we construct a test statistic based on kernel averages of the instrumented moment conditions over Z_i values that lie in a neighborhood of z_0.

The instrumented conditional moment conditions given $Z_i = z_0$ are of the form:

$$\begin{aligned} &E_{F_0}(m_j\left(W_i,\theta_0\right)g_j\left(X_i\right)|Z_i = z_0) \geq 0 \text{ for } j = 1,...,p \text{ and} \\ &E_{F_0}(m_j\left(W_i,\theta_0\right)g_j\left(X_i\right)|Z_i = z_0) = 0 \text{ for } j = p+1,...,k, \text{ for } g = (g_1,...,g_k)' \in \mathcal{G}, \end{aligned} \tag{6.1}$$

where $g = (g_1, ..., g_k)'$ are instruments that depend 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 (6.1) is

$$\Theta_{F_0}(\mathcal{G}) = \{\theta \in \Theta : (6.1) \text{ holds with } \theta \text{ in place of } \theta_0\}. \tag{6.2}$$

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 ASN. Section 6.3 provides conditions for this equality and shows that the instruments defined in (3.6) of ASN satisfy the conditions. Additional sets $\mathcal{G}$ are given in AS1 and AS2.

We construct test statistics based on (6.1). The sample moment functions are

$$\begin{aligned} \overline{m}_n(\theta,g) &= n^{-1}\sum_{i=1}^{n} m(W_i,\theta,g,b) \text{ for } g \in \mathcal{G}, \text{ where} \\ m(W_i,\theta,g,b) &= b^{-d_z/2}K_b(Z_i)m(W_i,\theta,g), \\ K_b(Z_i) &= K\left(\frac{Z_i - z_0}{b}\right), \\ m(W_i,\theta,g) &= \begin{pmatrix} m_1(W_i,\theta)g_1(X_i) \\ m_2(W_i,\theta)g_2(X_i) \\ \vdots \\ m_k(W_i,\theta)g_k(X_i) \end{pmatrix} \text{ for } g \in \mathcal{G}, \end{aligned} \tag{6.3}$$

$b > 0$ is a scalar bandwidth parameter for which $b \to 0$ as $n \to \infty$, and $K(x)$ is a kernel function. The definition of $\overline{m}_n(\theta, g)$ in (6.3) is the same as the definition of $\overline{m}_n(\theta, g)$ in AS1 except for the multiplicand $b^{-d_z/2}K_b(Z_i)$ in $m(W_i, \theta, g, b)$.

For notational simplicity, we omit the dependence of $\overline{m}_n(\theta, g)$ (and various other quantities below) on b.

Note that the normalization $b^{-d_z/2}$ that appears in $m(W_i, \theta, g, b)$ yields $m(W_i, \theta, g, b)$

to have a variance matrix that is $O(1)$, but not $o(1)$. In fact, under the conditions given below, $Var_F(m(W_i, \theta, g, b)) \rightarrow Var_F(m(W_i, \theta, g)|Z_i = z_0)f(z_0)$ as $n \rightarrow \infty$ under $(\theta, F) \in \mathcal{F}$.

If the sample average $\overline{m}_n(\theta, g)$ is divided by the scalar $n^{-1}\sum_{i=1}^n b^{-d_z/2}K_b(Z_i)$ it becomes the Nadaraya-Watson nonparametric kernel estimator of $E(m(W_i, \theta, g)|Z_i = z_0)$. We omit this divisor because doing so simplifies the statistic and has no effect on the test defined below.[23]

We assume the bandwidth b and kernel function $K(x)$ satisfy:

Assumption B. (a) $b = o(n^{-1/(4+d_z)})$ and (b) $nb^{d_z} \rightarrow \infty$ as $n \rightarrow \infty$.

Assumption K. (a) $\int K(z)dz = 1$, (b) $\int zK(z)dz = 0_{d_z}$, (c) $K(z) = 0\ \forall z \notin [-1, 1]^{d_z}$, (d) $K(z) \geq 0\ \forall z \in R^{d_z}$, and (e) $\sup_{z \in R^{d_z}} K(z) < \infty$.

Assumptions B and K are standard assumptions in the nonparametric density and regression literature. When Assumption B is applied to a nonparametric regression or density estimator, part (a) implies that the bias of the estimator goes to zero faster than the variance (and is the weakest condition for which this holds) and part (b) implies that the estimator is asymptotically normal (because it implies that b goes to zero sufficiently slowly that a Lindeberg condition holds).

The sample variance-covariance matrix of $n^{1/2}\overline{m}_n(\theta, g)$ is

$$\widehat{\Sigma}_n(\theta, g) = n^{-1}\sum_{i=1}^n (m(W_i, \theta, g, b) - \overline{m}_n(\theta, g))(m(W_i, \theta, g, b) - \overline{m}_n(\theta, g))'. \tag{6.4}$$

The matrix $\widehat{\Sigma}_n(\theta, g)$ may be singular or nearly singular with non-negligible probability for some $g \in \mathcal{G}$. This is undesirable because the inverse of $\widehat{\Sigma}_n(\theta, g)$ needs to be consistent for its population counterpart uniformly over $g \in \mathcal{G}$ for the test statistics considered below. In consequence, we employ a modification of $\widehat{\Sigma}_n(\theta, g)$, denoted $\overline{\Sigma}_n(\theta, g)$, such that $\det(\overline{\Sigma}_n(\theta, g))$ is bounded away from zero:

$$\overline{\Sigma}_n(\theta, g) = \widehat{\Sigma}_n(\theta, g) + \varepsilon \cdot Diag(\widehat{\Sigma}_n(\theta, 1_k)) \text{ for } g \in \mathcal{G} \tag{6.5}$$

for some fixed $\varepsilon > 0$. In the simulations in Section 4 of ASN, we use $\varepsilon = 5/100$. By design, $\overline{\Sigma}_n(\theta, g)$ is a linear combination of two scale equivariant functions and hence

[23] This holds because division by $n^{-1}\sum_{i=1}^n b^{-d_z/2}K_b(Z_i)$ rescales the test statistic and critical value identically and in consequence the rescaling cancels out.

is scale equivariant.[24] This yields a test statistic that is invariant to rescaling of the moment functions $m(W_i, \theta)$, which is an important property.

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

$$T_n(\theta) = \int S(n^{1/2}\overline{m}_n(\theta, g), \overline{\Sigma}_n(\theta, g))dQ(g), \tag{6.6}$$

where S is a non-negative function, Q is a weight function (i.e., probability measure) on $\mathcal{G}$, and the integral is over $\mathcal{G}$. The functions S and Q are discussed in Sections 6.2 and 6.4 below, respectively.

The Kolmogorov-Smirnov-type (KS) statistic is

$$T_n(\theta) = \sup_{g \in \mathcal{G}} S(n^{1/2}\overline{m}_n(\theta, g), \overline{\Sigma}_n(\theta, g)). \tag{6.7}$$

For brevity, the discussion in this Appendix focusses on CvM statistics and all results stated, except those in Section 11, 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.

6.2 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. Let $\mathcal{S} = \{(m, \Sigma) : m \in R^p_{[+\infty]} \times R^v, \ \Sigma \in \mathcal{W}\}$. Let $R_+ = \{x \in R : x \geq 0\}$.

We consider functions S that satisfy the following conditions.

Assumption S1. $\forall\, (m, \Sigma) \in \mathcal{S}$,

(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.

[24] That is, multiplying the moment functions $m(W_i, \theta)$ by a diagonal matrix, D, changes $\overline{\Sigma}_n(\theta, g)$ into $D\overline{\Sigma}_n(\theta, g)D$.

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

Assumption S2. $S(m, \Sigma)$ is uniformly continuous in the sense that, for all $m_0 \in R^k$ and all pd Σ_0, $\sup_{\mu \in R^p_+ \times \{0\}^v} |S(m + \mu, \Sigma) - S(m_0 + \mu, \Sigma_0)| \to 0$ as $(m, \Sigma) \to (m_0, \Sigma_0)$.[25]

The following two assumptions are used only to establish the power properties of tests.

Assumption S3. $S(m, \Sigma) > 0$ if and only if $m_j < 0$ for some $j = 1, ..., p$ or $m_j \neq 0$ for some $j = p + 1, ..., k$, 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 scalars $a > 0$, $m \in R^k$, and $\Sigma \in \mathcal{W}$.

The functions S_1, S_2, and S_3 in (3.9) of ASN satisfy Assumptions S1-S4 by Lemma 1 of AS1.

6.3 X-Instruments

The collection of instruments $\mathcal{G}$ needs to satisfy the following condition in order for the conditional moments $\{E_F(m(W_i, \theta, g)|Z_i = z_0) : g \in \mathcal{G}\}$ to incorporate the same information as the conditional moments $\{E_F(m(W_i, \theta)|X_i = x, Z_i = z_0) : x \in R^{d_x}\}$.

For any $\theta \in \Theta$ and any distribution F with $E_F(||m(W_i, \theta)||\ |Z_i = z_0) < \infty$, let

$$\begin{aligned} \mathcal{X}_F(\theta) = \{x \in R^{d_x} : & E_F(m_j(W_i, \theta)\,|X_i = x, Z_i = z_0) < 0 \text{ for some } j \leq p \text{ or} \\ & E_F(m_j(W_i, \theta)\,|X_i = x, Z_i = z_0) \neq 0 \text{ for some } j = p + 1, ..., k\}. \end{aligned} \quad (6.8)$$

Assumption NCI. For any $\theta \in \Theta$ and distribution F for which $E_F(||m(W_i, \theta)||\ |Z_i = z_0) < \infty$ and $P_F(X_i \in \mathcal{X}_F(\theta)|Z_i = z_0) > 0$, there exists some $g \in \mathcal{G}$ such that

$$\begin{aligned} & E_F(m_j(W_i, \theta)g_j(X_i)|Z_i = z_0) < 0 \text{ for some } j \leq p \text{ or} \\ & E_F(m_j(W_i, \theta)g_j(X_i)|Z_i = z_0) \neq 0 \text{ for some } j = p + 1, ..., k. \end{aligned}$$

Note that NCI abbreviates "nonparametrically conditionally identified." The following Lemma indicates the importance of Assumption NCI.

[25] Assumption S2 is equivalent to the same condition with μ vectors whose elements exceed $-\eta_1$ for some $\eta_1 < \infty$. This is used in the proofs below.

Lemma AN1. *Assumption* NCI *implies that* $\Theta_F(\mathcal{G}) = \Theta_F$ *for all* F *with* $\sup_{\theta\in\Theta} E_F(||m(W_i,\theta)|| \, |Z_i = z_0) < \infty$.

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

The collection $\mathcal{G}$ also must satisfy the following "manageability" condition. This condition regulates the complexity of $\mathcal{G}$. It ensures that $\{n^{1/2}(\overline{m}_n(\theta,g) - E_{F_n}\overline{m}_n(\theta,g)) : g \in \mathcal{G}\}$ 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. The manageability condition is from Pollard (1990) and is defined and explained in Appendix E of AS2.

Assumption NM. (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 $n \geq 1$.[26,27]

Lemma 3 of AS1 establishes Assumptions NCI and NM for $\mathcal{G}_{c\text{-}cube}$ defined in (3.6) of ASN.[28]

6.4 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 9 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 $\mathcal{G}_{c\text{-}cube}$.

[26] The set of distributions $\mathcal{F}_+$ is defined just prior to (3.17) of ASN.

[27] The asymptotic results given in the paper hold with Assumption NM replaced by any alternative assumption that is sufficient to obtain the requisite empirical process results given in Lemma AN4 below.

[28] Lemma 3 of AS1 and Lemma B2 of AS2 also establish Assumptions NCI and NM of this Appendix for the collections $\mathcal{G}_{box}$, $\mathcal{G}_{B-spline}$, $\mathcal{G}_{box,dd}$, and $\mathcal{G}_{c/d}$ defined there. The proof is the same as in AS2 for Assumptions CI and M with conditioning on $Z_i = z_0$ added throughout.

Weight Function Q for $\mathcal{G}_{c\text{-}cube}$. There is a one-to-one mapping $\Pi_{c\text{-}cube} : \mathcal{G}_{c\text{-}cube} \to AR = \{(a,r) : a \in \{1,...,2r\}^{d_x} \text{ and } r = r_0, r_0+1, ...\}$. Let Q_{AR} be a probability measure on AR. One can take $Q = \Pi_{c\text{-}cube}^{-1} Q_{AR}$. 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

$$T_n(\theta) = \sum_{r=r_0}^{\infty} w(r) \sum_{a \in \{1,...,2r\}^{d_x}} (2r)^{-d_x} S(n^{1/2}\overline{m}_n(\theta, g_{a,r}), \overline{\Sigma}_n(\theta, g_{a,r})), \tag{6.9}$$

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

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

6.5 Computation of Sums, Integrals, and Suprema

The test statistic $T_n(\theta)$ given in (6.9) 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 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

$$\overline{T}_{n,s_n}(\theta) = \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S(n^{1/2}\overline{m}_n(\theta, g_\ell), \overline{\Sigma}_n(\theta, g_\ell)), \tag{6.10}$$

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 (6.9), 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 ASN are of this form when $r_{1,n} < \infty$.

It can 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 and most of the results under (certain) $(nb^{d_z})^{-1/2}$-local alternatives. For brevity we do so here only for the truncated statistics defined in ASN and for the results stated in Theorems N1 and N2 of ASN, see the discussion following the proofs of Theorems AN1 and AN4 in Section 12.2.4 and Comment 2 following Theorem AN2 in Section 9. For other approximate statistics and for power under (certain) $(nb^{d_z})^{-1/2}$-local alternatives, the method of proof is analogous to that used in Section 15.1 of Appendix D of AS2 to prove such results stated in Section 13.1 of Appendix B of AS2 for the tests considered in AS1 and AS2.

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 results for KS tests hold with $\mathcal{G}_n$ in place of $\mathcal{G}$ (although some asymptotic local power results require $\mathcal{G}_n \uparrow \mathcal{G}$ at a sufficiently fast rate). For results of this sort for the tests considered in AS1 and AS2, see Section 13.1 of Appendix B of AS2 and Section 15.1 of Appendix D of AS2.

7 GMS and Plug-in Asymptotic Confidence Sets

7.1 GMS Critical Values

In this section, we define GMS critical values and CS's.

It is shown in Theorem AN4 in Section 12.2.2 that when θ is in the identified set the "uniform asymptotic distribution" of $T_n(\theta)$ is the distribution of $T(h_n)$, where $h_n = (h_{1,n}, h_2)$, $h_{1,n}(\cdot)$ is a function from $\mathcal{G}$ 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 $h_2(\cdot,\cdot)$ is a $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$. For $h = (h_1, h_2)$, define

$$T(h) = \int S(\nu_{h_2}(g) + h_1(g), h_2(g,g) + \varepsilon I_k) dQ(g), \tag{7.1}$$

where

$$\{\nu_{h_2}(g) : g \in \mathcal{G}\} \tag{7.2}$$

is a mean zero R^k-valued Gaussian process with covariance kernel $h_2(\cdot,\cdot)$ on $\mathcal{G} \times \mathcal{G}$, $h_1(\cdot)$

is a function from $\mathcal{G}$ to $R^p_{[+\infty]} \times \{0\}^v$, and ε is as in the definition of $\overline{\Sigma}_n(\theta, g)$ in (6.5).[29] The definition of $T(h)$ in (7.1) applies to CvM test statistics. For the KS test statistic, one replaces $\int ... \, dQ(g)$ by $\sup_{g \in \mathcal{G}} ...$.

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

$$c_0(h, 1-\alpha) \; (= c_0(h_1, h_2, 1-\alpha)) \tag{7.3}$$

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 $\mathcal{G}$ that is constructed to be less than or equal to $h_{1,n}(g)$ for all $g \in \mathcal{G}$ 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 $m = (m_I', m_{II}')'$, the latter property yields a test whose asymptotic level is less than or equal to the nominal level α. (It is arbitrarily close to α for certain $(\theta, F) \in \mathcal{F}$.) The quantities $\widehat{h}_{2,n}(\theta)$ and $\varphi_n(\theta)$ are defined below.

The nominal $1 - \alpha$ GMS critical value is defined to be

$$c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha) = c_0(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) + \eta, \tag{7.4}$$

where $\eta > 0$ is an arbitrarily small positive constant, e.g., 10^{-6}. A nominal $1 - \alpha$ GMS CS is given by

$$CS_n = \{\theta \in \Theta : T_n(\theta) \le c_{n,1-\alpha}(\theta)\}. \tag{7.5}$$

with the critical value $c_{n,1-\alpha}(\theta)$ equal to $c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$.[30]

Next, we define the asymptotic covariance kernel, $\{h_{2,F}(\theta, g, g^*) : g, g^* \in \mathcal{G}\}$, of

[29] The sample paths of $\nu_{h_2}(\cdot)$ are concentrated on the set $U^k_\rho(\mathcal{G})$ of bounded uniformly ρ-continuous R^k-valued functions on $\mathcal{G}$, where ρ is defined in Appendix A of AS2.

[30] 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.

$n^{1/2}\overline{m}_n(\theta, g)$ after normalization via a diagonal matrix $D_F^{-1/2}(\theta, z_0)$. Define[31]

$$\begin{aligned} h_{2,F}(\theta, g, g^*) &= D_F^{-1/2}(\theta, z_0)\Sigma_F(\theta, g, g^*, z_0)D_F^{-1/2}(\theta, z_0), \text{ where} \\ \Sigma_F(\theta, g, g^*, z) &= E_F(m(W_i, \theta, g)m(W_i, \theta, g^*)'|Z_i = z)f(z) \text{ and} \\ D_F(\theta, z) &= Diag(\Sigma_F(\theta, 1_k, 1_k, z)) \; (= Diag(E_F(m(W_i, \theta)m(W_i, \theta)'|Z_i = z)f(z))). \end{aligned} \tag{7.6}$$

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

$$\begin{aligned} \widehat{h}_{2,n}(\theta, g, g^*) &= \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta, g, g^*)\widehat{D}_n^{-1/2}(\theta), \text{ where} \\ \widehat{\Sigma}_n(\theta, g, g^*) &= n^{-1}\sum_{i=1}^n (m(W_i, \theta, g, b) - \overline{m}_n(\theta, g))\,(m(W_i, \theta, g^*, b) - \overline{m}_n(\theta, g^*))' \text{ and} \\ \widehat{D}_n(\theta) &= Diag(\widehat{\Sigma}_n(\theta, 1_k, 1_k)). \end{aligned} \tag{7.7}$$

Note that $\widehat{\Sigma}_n(\theta, g)$, defined in (6.4), equals $\widehat{\Sigma}_n(\theta, g, g)$ and $\widehat{\Sigma}_n(\theta, 1_k, 1_k)$ is the sample variance-covariance matrix of $\{m(W_i, \theta) : n \geq 1\}$.

The quantity $\varphi_n(\theta)$ is defined in Section 7.4 below.

7.2 GMS Critical Values for Approximate Test Statistics

When the test statistic is approximated via a truncated sum, simulated integral, or QMC quantity, as discussed in Section 6.5, the statistic $T(h)$ in Section 7.1 is replaced by

$$\overline{T}_{s_n}(h) = \sum_{\ell=1}^{s_n} w_{Q,n}(\ell)S(\nu_{h_2}(g_\ell) + h_1(g_\ell), h_2(g_\ell, g_\ell) + \varepsilon I_k), \tag{7.8}$$

where $\{g_\ell : \ell = 1, ..., s_n\}$ are the same functions $\{g_1, ..., g_{s_n}\}$ that appear in the approximate statistic $\overline{T}_{n,s_n}(\theta)$. We call the critical value obtained using $\overline{T}_{s_n}(h)$ an approximate GMS (A-GMS) critical value.

Let $c_{0,s_n}(h, 1-\alpha)$ denote the $1-\alpha$ quantile of $\overline{T}_{s_n}(h)$ for fixed $\{g_1, ..., g_{s_n}\}$. The

[31] Note that $D_F(\theta, z) = Diag(\sigma^2_{F,1}(\theta, z), ..., \sigma^2_{F,k}(\theta, z))$, where $\sigma^2_{F,j}(\theta, z) = E_F(m_j^2(W_i, \theta)|Z_i = z)f(z)$. Also note that the means, $E_F m(W_i, \theta, g)$, $E_F m(W_i, \theta, g^*)$, and $E_F m(W_i, \theta)$, are not subtracted off in the definitions of $\Sigma_F(\theta, g, g^*, z)$ and $D_F(\theta, z)$. The reason is that the population means of the sample-size n quantities based on $m(W_i, \theta, g, b)$ are smaller than the second moments by an order of magnitude and, hence, are asymptotically negligible. See Lemmas AN6 and AN7 below.

A-GMS critical value is defined to be

$$c_{s_n}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha) = c_{0,s_n}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) + \eta. \qquad (7.9)$$

This critical value is a quantile that can be computed by simulation as follows. Let $\{\overline{T}_{s_n,\tau}(h) : \tau = 1, ..., \tau_{reps}\}$ be τ_{reps} i.i.d. random variables each with the same distribution as $\overline{T}_{s_n}(h)$ and each with the same functions $\{g_1, ..., g_{s_n}\}$, where $h = (h_1, h_2)$ is evaluated at $(\varphi_n(\theta), \widehat{h}_{2,n}(\theta))$. Then, the A-GMS critical value, $c_{s_n}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$, is the $1-\alpha+\eta$ sample quantile of $\{\overline{T}_{s_n,\tau}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta)) : \tau = 1, ..., \tau_{reps}\}$ plus η for very small $\eta > 0$ and large τ_{reps}.

7.3 Bootstrap GMS Critical Values

Bootstrap versions of the GMS critical value in (7.4) and the A-GMS critical value in (7.9) can be employed. The bootstrap GMS critical value is

$$c^*(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha) = c^*_0(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha+\eta) + \eta, \qquad (7.10)$$

where $c^*_0(h, 1-\alpha)$ is the $1-\alpha$ quantile of $T^*(h)$ and $T^*(h)$ is defined as in (7.1) but with $\{\nu_{h_2}(g) : g \in \mathcal{G}\}$ and $\widehat{h}_{2,n}(\theta)$ replaced by the bootstrap empirical process $\{\nu^*_n(g) : g \in \mathcal{G}\}$ and the bootstrap covariance kernel $\widehat{h}^*_{2,n}(\theta)$, respectively. By definition, $\nu^*_n(g) = n^{-1/2}\sum_{i=1}^n (m(W^*_i, \theta, g, b) - \overline{m}_n(\theta, g))$, where $\{W^*_i : i \leq n\}$ is an i.i.d. bootstrap sample drawn from the empirical distribution of $\{W_i : i \leq n\}$. Also, $\widehat{h}^*_{2,n}(\theta, g, g^*)$, $\widehat{\Sigma}^*_n(\theta, g, g^*)$, and $\widehat{D}^*_n(\theta)$ are defined as in (7.7) with W^*_i in place of W_i. Note that $\widehat{h}^*_{2,n}(\theta, g, g^*)$ only enters $c(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha)$ via functions (g, g^*) such that $g = g^*$.

When the test statistic, $\overline{T}_{n,s_n}(\theta)$, is a truncated sum, simulated integral, or a QMC quantity, a bootstrap A-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 7.2. The same functions $\{g_1, ..., g_{s_n}\}$ are used in all bootstrap critical value calculations as in the test statistic $\overline{T}_{n,s_n}(\theta)$.

7.4 Definition of $\boldsymbol{\varphi}_{\mathbf{n}}(\boldsymbol{\theta})$

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

$$\xi_n(\theta, g) = \kappa_n^{-1} n^{1/2} \overline{D}_n^{-1/2}(\theta, g) \overline{m}_n(\theta, g), \text{ where } \overline{D}_n(\theta, g) = Diag(\overline{\Sigma}_n(\theta, g)), \quad (7.11)$$

$\overline{\Sigma}_n(\theta, g)$ is defined in (6.5), and $\{\kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \to \infty$. The jth element of $\xi_n(\theta, g)$, denoted $\xi_{n,j}(\theta, g)$, measures the slackness of the moment inequality $E_F m_j(W_i, \theta, g) \geq 0$ for $j = 1, ..., p$.

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

$$\begin{aligned} &\varphi_{n,j}(\theta, g) = \overline{h}_{2,n,j}(\theta, g)^{1/2} B_n 1(\xi_{n,j}(\theta, g) > 1), \\ &\overline{h}_{2,n}(\theta, g) = \widehat{D}_n^{-1/2}(\theta) \overline{\Sigma}_n(\theta, g) \widehat{D}_n^{-1/2}(\theta), \text{ and } \overline{h}_{2,n,j}(\theta, g) = [\overline{h}_{2,n}(\theta, g)]_{jj}. \end{aligned} \quad (7.12)$$

We assume:

Assumption GMS1. (a) $\varphi_n(\theta, g)$ satisfies (7.12), where $\{B_n : n \geq 1\}$ is a non-decreasing sequence of positive constants, and

(b) for some $\zeta > 1$, $\kappa_n - \zeta B_n \to \infty$ as $n \to \infty$.

The constants $\{B_n : n \geq 1\}$ in Assumption GMS1 need not diverge to infinity for the GMS CS to have asymptotic size greater than or equal to $1 - \alpha$. However, for the GMS CS not to be asymptotically conservative, B_n must diverge to ∞, see Assumption GMS2(b) below. In ASN, 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 $\overline{h}_{2,n,j}(\theta, g)^{1/2}$ in (7.12) is an "ε-adjusted" standard deviation estimator for the jth normalized sample moment based on g. It provides a suitable scaling for $\varphi_n(\theta, g)$.

The following assumption is not needed for GMS CS's to have uniform asymptotic coverage probability greater than or equal to $1 - \alpha$. It is used, however, to show that GMS CS's are not asymptotically conservative. For $(\theta, F) \in \mathcal{F}$ and $j = 1, ..., k$, define $h_{1,\infty,F}(\theta) = \{h_{1,\infty,F}(\theta, g) : g \in \mathcal{G}\}$ to have jth element equal to ∞ if $E_F m_{F,j}(\theta, X_i, z_0) \times g_j(X_i) > 0$ and $j \leq p$ and 0 otherwise, where $m_{F,j}(\theta, x, z)$ denotes the jth element of $m_F(\theta, x, z)$. Let $h_{\infty,F}(\theta) = (h_{1,\infty,F}(\theta), h_{2,F}(\theta))$, where $h_{2,F}(\theta) = \{h_{2,F}(\theta, g, g^*) : (g, g^*) \in \mathcal{G} \times \mathcal{G}\}$.

Assumption GMS2. (a) For some $(\theta_c, F_c) \in \mathcal{F}$, the distribution function of $T(h_{\infty,F_c}(\theta_c))$ is continuous and strictly increasing at its $1-\alpha$ quantile plus δ, viz., $c_0(h_{\infty,F_c}(\theta_c), 1-\alpha) + \delta$, for all $\delta > 0$ sufficiently small and $\delta = 0$,

(b) $B_n \to \infty$ as $n \to \infty$, and

(c) $(nb^{d_z})^{1/2}/\kappa_n \to \infty$ as $n \to \infty$.

Assumption GMS2(a) is not restrictive. For example, it holds for typical choices of S and Q, such as S_1 and S_3 and Q as in ASN, for any (θ_c, F_c) for which $Q(\{g \in \mathcal{G} : h_{1,\infty,F_c}(\theta_c, g) = 0\}) > 0$. This is established in Lemma B3 in Section 13.3 of AS2. Assumption GMS2(c) is satisfied by typical choices of κ_n, such as $\kappa_n = (0.3 \ln n)^{1/2}$, because the bandwidth b should always be taken such that $b^{d_z} \geq cn^{-1+\delta}$ for some $c, \delta > 0$.

7.5 "Plug-in Asymptotic" Confidence Sets

Next, for comparative purposes, we define plug-in asymptotic (PA) critical values. Subsampling critical values also can be considered, see Appendix B of AS2 for details. We strongly recommend GMS critical values over PA and subsampling critical values for the same reasons as given in AS1 plus the fact that the finite-sample simulations in Section 4 show better performance by GMS critical values than PA and subsampling critical values.

PA critical values are obtained from the asymptotic null distribution that arises when all conditional moment inequalities hold as equalities a.s. The PA critical value is

$$c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha) = c_0(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) + \eta, \tag{7.13}$$

where $0_{\mathcal{G}}$ denotes the R^k-valued function on $\mathcal{G}$ that is identically $(0, ..., 0)' \in R^k$, and $\widehat{h}_{2,n}(\theta)$ is defined in (7.7). The nominal $1-\alpha$ PA CS is given by (7.5) with the critical value $c_{n,1-\alpha}(\theta)$ equal to $c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha)$.

Bootstrap PA, A-PA, and bootstrap A-PA critical values are defined analogously to their GMS counterparts in Sections 7.2 and 7.3.

8 Asymptotic Size

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

For simplicity, let $h_{2,F}(\theta)$ abbreviate the asymptotic covariance kernel $\{h_{2,F}(\theta, g, g^*) : g, g^* \in \mathcal{G}\}$ defined in (7.6). Define

$$\mathcal{H}_2 = \{h_{2,F}(\theta) : (\theta, F) \in \mathcal{F}\}. \tag{8.1}$$

On the space of $k \times k$-matrix-valued covariance kernels on $\mathcal{G} \times \mathcal{G}$, which is a superset of $\mathcal{H}_2$, we use the uniform metric d defined by

$$d(h_2^{(1)}, h_2^{(2)}) = \sup_{g,g^* \in \mathcal{G}} ||h_2^{(1)}(g, g^*) - h_2^{(2)}(g, g^*)||. \tag{8.2}$$

The following Theorem gives uniform asymptotic coverage probability results for GMS and PA CS's.

Theorem AN1. *Suppose Assumptions* B, K, NM, S1, *and* S2 *hold and Assumption* GMS1 *also holds when considering* GMS CS's. *Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$, GMS *and* PA *confidence sets* CS_n *satisfy*

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

(b) *if Assumption* GMS2 *also holds and* $h_{2,F_c}(\theta_c) \in \mathcal{H}_{2,cpt}$ *(for* $(\theta_c, F_c) \in \mathcal{F}$ *as in Assumption* GMS2), *then the* GMS *confidence set satisfies*

$$\lim_{\eta\to 0} \liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F(\theta \in CS_n) = 1 - \alpha,$$

where η *is as in the definition of* $c(h, 1-\alpha)$.

Comments. 1. Theorem AN1(a) shows that GMS and PA CS's have correct uniform asymptotic size over compact sets of covariance kernels. Theorem AN1(b) shows that GMS CS's are at most infinitesimally conservative asymptotically (i.e., their asymptotic size is infinitessimally close to their nominal size). The uniformity results hold whether the moment conditions involve "weak" or "strong" instrumental variables X_i.

2. As in AS1, an analogue of Theorem AN1(b) holds for PA CS's if Assumption GMS2(a) holds and $E_{F_c}(m_j(W_i, \theta_c)|X_i, Z_i = z_0) = 0$ a.s. for $j \leq p$ (i.e., if the conditional moment inequalities hold as equalities a.s.) under some $(\theta_c, F_c) \in \mathcal{F}$. However, the latter condition is restrictive—it fails in many applications.

3. Theorem N1 of ASN for the case $r_{1,n} = \infty$ is proved by verifying the conditions of Theorem AN1 (that is, by showing that Assumptions B, K, NM, S1, S2, and GMS1 hold for b, K, and S defined as in ASN).[32] Assumption B holds by the definition of b following (3.2) of ASN. Assumption K holds for the Epanechnikov kernel $K(x) = 0.75 \max\{1 - x^2, 0\}$ employed in (3.2) of ASN. The functions S_1, S_2, and S_3 in (3.9) of ASN satisfy Assumptions S1-S4 by Lemma 1 of AS1. Lemma 3 of AS1 establishes Assumptions NCI and NM for $\mathcal{G}_{c\text{-}cube}$ defined in (3.6) of ASN. Assumption GMS1 holds immediately for κ_n and B_n defined in (3.10) and (3.11) of ASN, respectively. Assumptions GMS2(b) and (c) hold by the definitions of b, κ_n, and B_n of ASN. Assumption GMS2(a) holds for the functions S_1 and S_3 by Lemma B3 given in Section 13.3 in Appendix B of AS2. For the function S_2, part (b) of Theorem N1 is stated to hold in Comment 2 following Theorem N1 only if Assumption GMS2(a) is assumed to hold. (That is, we do not have a proof that this Assumption GMS2(a) necessarily holds with the function S_2. But, it seems that it should hold in most models.)

4. Theorem N1 of ASN 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 of Theorems AN1 and AN4 (where Theorem AN4 given in Section 12.2 is used in the proof of Theorem AN1), for details see Section 12.2.4 following the proofs of Theorems AN1 and AN4.

9 Power Against Fixed Alternatives

We now show that the power of GMS and PA tests converges to one as $n \to \infty$ for all fixed alternatives (for which the moment functions have $4 + \delta$ moments finite). Thus, both tests are consistent tests. This implies that for any fixed distribution F_0 and any parameter value θ_* *not* in the identified set Θ_{F_0}, the GMS and PA CS's do not include θ_* with probability approaching one. In this sense, GMS and PA CS's based on $T_n(\theta)$ fully exploit the conditional moment inequalities and equalities. CS's based on a finite number of unconditional moment inequalities and equalities do not have this property.

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

The null hypothesis is

$$H_0 : E_{F_0}(m_j(W_i, \theta_*)|X_i, Z_i = z_0) \geq 0 \text{ a.s. } [F_{X,0}] \text{ for } j = 1, ..., p \text{ and}$$
$$E_{F_0}(m_j(W_i, \theta_*)|X_i, Z_i = z_0) = 0 \text{ a.s. } [F_{X,0}] \text{ for } j = p+1, ..., k, \qquad (9.1)$$

where θ_* denotes the null parameter value and F_0 denotes the fixed true distribution of the data. The alternative hypothesis is $H_1 : H_0$ does not hold. The following assumption specifies the properties of fixed alternatives (FA).

Let $\mathcal{F}_+$ denote all (θ, F) that satisfy Assumptions PS1-PS3 that define $\mathcal{F}$ except Assumptions PS1(c) and PS1(d) (which impose the conditional moment inequalities and equalities). As defined, $\mathcal{F} \subset \mathcal{F}_+$. Note that $\mathcal{F}_+$ includes (θ, F) pairs for which θ lies outside of the identified set Θ_F as well as all values in the identified set.

Assumption NFA. The value $\theta_* \in \Theta$ and the true distribution F_0 satisfy: (a) $P_{F_0}(X_i \in \mathcal{X}_{F_0}(\theta_*)|Z_i = z_0) > 0$, where $\mathcal{X}_{F_0}(\theta_*)$ is defined in (6.8), and (b) $(\theta_*, F_0) \in \mathcal{F}_+$.

Assumption NFA(a) states that violations of the conditional moment inequalities or equalities occur for the null parameter θ_* for X_i values in some set with positive conditional probability given $Z_i = z_0$ under F_0. Thus, under Assumption NFA(a), the moment conditions specified in (9.1) do not hold.

For $g \in \mathcal{G}$, define

$$m_j^*(g) = E_{F_0}(m_j(W_i, \theta_*)g_j(X_i)|Z_i = z_0)f(z_0)/\sigma_{F_0,j}(\theta_*, z_0) \text{ and}$$
$$\beta(g) = \max\{-m_1^*(g), ..., -m_p^*(g), |m_{p+1}^*(g)|, ..., |m_k^*(g)|\}. \qquad (9.2)$$

Under Assumptions NFA(a) and NCI, $\beta(g_0) > 0$ for some $g_0 \in \mathcal{G}$.

For a test based on $T_n(\theta)$ to have power against all fixed alternatives, the weighting function Q cannot "ignore" any elements $g \in \mathcal{G}$, because such elements may have identifying power for the identified set. This requirement is captured in the following assumption.

Let $F_{X,0}$ denote the distribution of X_i under F_0. Define the pseudo-metric ρ_X on $\mathcal{G}$ by

$$\rho_X(g, g^*) = (E_{F_{X,0}}||g(X_i) - g^*(X_i)||^2)^{1/2} \text{ for } g, g^* \in \mathcal{G}. \qquad (9.3)$$

Let $\mathcal{B}_{\rho_X}(g, \delta)$ denote an open ρ_X-ball in $\mathcal{G}$ centered at g with radius δ.

Assumption Q. The support of Q under the pseudo-metric ρ_X is $\mathcal{G}$. That is, for all

$\delta > 0$, $Q(\mathcal{B}_{\rho_X}(g, \delta)) > 0$ for all $g \in \mathcal{G}$.

Assumption Q holds for Q_{AR} and $\mathcal{G}_{c\text{-}cube}$ defined above and in ASN because $\mathcal{G}_{c\text{-}cube}$ is countable and Q_{AR} has a probability mass function that is positive at each element in $\mathcal{G}_{c\text{-}cube}$. Appendix B of AS2 verifies Assumption Q for four other choices of Q and $\mathcal{G}$.

The following Theorem shows that GMS and PA tests are consistent against all fixed alternatives.

Theorem AN2. *Suppose Assumptions* B, K, NFA, NCI, Q, S1, S3, *and* S4 *hold and* Assumption NM *holds with* F_0 *in place of* F_n *in* Assumption NM(b). 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) $\lim_{n\to\infty} P_{F_0}(T_n(\theta_*) > c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha)) = 1$.

Comments. 1. Theorem N2 of ASN for the case $r_{1,n} = \infty$ is proved by verifying that the conditions of Theorem AN2 (except Assumption NFA) hold for b, K, S, and $\mathcal{G}_{c\text{-}cube}$ defined as in ASN. By Comment 3 to Theorem AN1, Assumptions B, K, S1, S3, and S4 hold. Assumption NCI holds for $\mathcal{G}_{c\text{-}cube}$ as defined in (3.6) of ASN by Lemma 3 of AS1. As noted above, Assumption Q holds for $\mathcal{G}_{c\text{-}cube}$ and Q_{AR}. Assumption NM holds for $\mathcal{G}_{c\text{-}cube}$ *with* F_0 *in place of* F_n in part (b) because $\mathcal{C}_{c\text{-}cube}$ is a Vapnik-Cervonenkis class of sets. (For more details, see Lemma 3 of AS1 for the verification of Assumption NM under F_n.)

2. Theorem N2 of ASN 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 AN2. The alterations required are the same as those described for A-CvM tests in the proof of Theorem B2 in Appendix D of AS2.[33]

10 Power Against $(\mathbf{nb^{d_z}})^{-1/2}$-Local Alternatives

In this section, we show that GMS and PA tests have power against certain, but not all, $(nb^{d_z})^{-1/2}$-local alternatives.

We show that a GMS test has asymptotic power that is greater than or equal to that of a PA test (based on the same test statistic) under all alternatives with strict inequality in certain scenarios.

[33] 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.

For given $\theta_{n,*} \in \Theta$ for $n \geq 1$, we consider tests of

$$H_0 : E_{F_n}(m_j(W_i, \theta_{n,*})|Z_i = z_0) \geq 0 \text{ for } j = 1, ..., p,$$
$$E_{F_n}(m_j(W_i, \theta_{n,*})|Z_i = z_0) = 0 \text{ for } j = p+1, ..., k, \quad (10.1)$$

and $(\theta_{n,*}, F_n) \in \mathcal{F}$, where F_n denotes the true distribution of the data. The null values $\theta_{n,*}$ are allowed to drift with n or be fixed for all n. Drifting $\theta_{n,*}$ values are of interest because they allow one to consider the case of a fixed identified set, say Θ_0, and to derive the asymptotic probability that parameter values $\theta_{n,*}$ that are not in the identified set, but drift toward it at rate $n^{-1/2}$, are excluded from a GMS or PA CS. In this scenario, the sequence of true distributions are ones that yield Θ_0 to be the identified set, i.e., $F_n \in \mathcal{F}_0 = \{F : \Theta_F = \Theta_0\}$.

The true parameters and distributions are denoted (θ_n, F_n). We consider the Kolmogorov-Smirnov metric on the space of distributions F.

Let $f_n(z)$ denote the density of Z_i wrt μ_{Leb} under F_n.

The $(nb^{d_z})^{-1/2}$-local alternatives are defined as follows.

Assumption NLA1. The true parameters and distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ and the null parameters $\{\theta_{n,*} : n \geq 1\}$ satisfy:

(a) $\theta_{n,*} = \theta_n + \lambda(nb^{d_z})^{-1/2}(1 + o(1))$ for some $\lambda \in R^{d_\theta}$, $\theta_{n,*} \in \Theta$, $\theta_{n,*} \to \theta_0$, and $F_n \to F_0$ for some $(\theta_0, F_0) \in \mathcal{F}$,

(b) $(nb^{d_z})^{1/2} E_{F_n}(m_j(W_i, \theta_n, g)|Z_i = z_0) f_n(z_0)/\sigma_{F_n,j}(\theta_n, z_0) \to h_{1,j}(g)$ for some $h_{1,j}(g) \in R_{+,\infty}$ for $j = 1, ..., p$ and all $g \in \mathcal{G}$,

(c) $d(h_{2,F_n}(\theta_n), h_{2,F_0}(\theta_0)) \to 0$ and $d(h_{2,F_n}(\theta_{n,*}), h_{2,F_0}(\theta_0)) \to 0$ as $n \to \infty$ (where d is defined in (8.2)), and

(d) $(\theta_n, F_n) \in \mathcal{F}_+$ for all $n \geq 1$.

Assumption NLA2. The $k \times d$ matrix $\Pi_F(\theta, g) = (\partial/\partial\theta')[D_F^{-1/2}(\theta, z_0) E_F(m(W_i, \theta, g) |Z_i = z_0) f(z_0)]$ exists and is continuous in (θ, F) for all (θ, F) in a neighborhood of (θ_0, F_0) for all $g \in \mathcal{G}$.

For notational simplicity, we let h_2 abbreviate $h_{2,F_0}(\theta_0)$ throughout this section. Assumption NLA1(a) states that the true values $\{\theta_n : n \geq 1\}$ are $(nb^{d_z})^{-1/2}$-local to the null values $\{\theta_{n,*} : n \geq 1\}$. Assumption NLA1(b) specifies the asymptotic behavior of the (normalized) moment inequality functions when evaluated at the true values $\{\theta_n : n \geq 1\}$. Under the true values, these (normalized) moment inequality functions are non-negative. Assumption NLA1(c) specifies the asymptotic behavior of the covariance kernels $\{h_{2,F_n}(\theta_n, \cdot, \cdot) : n \geq 1\}$ and $\{h_{2,F_n}(\theta_{n,*}, \cdot, \cdot) : n \geq 1\}$. Assumption NLA2 is a smoothness condition on the normalized expected conditional moment functions given $Z_i = z_0$. Given the smoothing properties of the expectation operator, this condition is not restrictive.

Under Assumptions NLA1 and NLA2, we show that the moment inequality functions evaluated at the null values $\{\theta_{n,*} : n \geq 1\}$ satisfy:

$$\begin{aligned} &\lim_{n\to\infty} n^{1/2} D_{F_n}^{-1/2}(\theta_{n,*}, b) E_{F_n} m(W_i, \theta_{n,*}, g, b) = h_1(g) + \Pi_0(g)\lambda \in R^k, \text{ where} \\ &h_1(g) = (h_{1,1}(g), ..., h_{1,p}(g), 0, ..., 0)' \in R^k, \ \Pi_0(g) = \Pi_{F_0}(\theta_0, g), \text{ and} \\ &D_F(\theta, b) = Diag(Var_F(b^{-d_z/2} K_b(Z_i) m(W_i, \theta))). \end{aligned} \tag{10.2}$$

If $h_{1,j}(g) = \infty$, then by definition $h_{1,j}(g) + y = \infty$ for any $y \in R$. We have $h_1(g) + \Pi_0(g)\lambda \in R^p_{[+\infty]} \times R^v$. Let $\Pi_{0,j}(g)$ denote the jth row of $\Pi_0(g)$ written as a column d_θ-vector for $j = 1, ..., k$.

The null hypothesis, defined in (10.1), does not hold (at least for n large) when the following assumption holds.

Assumption LA3. For some $g \in \mathcal{G}$, $h_{1,j}(g) + \Pi_{0,j}(g)'\lambda < 0$ for some $j = 1, ..., p$ or $\Pi_{0,j}(g)'\lambda \neq 0$ for some $j = p + 1, ..., k$.

Under the following assumption, if $\lambda = \beta\lambda_0$ for some $\beta > 0$ and some $\lambda_0 \in R^{d_\theta}$, then the power of GMS and PA tests against the perturbation λ is arbitrarily close to one for β arbitrarily large:

Assumption LA3′. $Q(\{g \in \mathcal{G} : h_{1,j}(g) < \infty$ and $\Pi_{0,j}(g)'\lambda_0 < 0$ for some $j = 1, ..., p$ or $\Pi_{0,j}(g)'\lambda_0 \neq 0$ for some $j = p + 1, ..., k\}) > 0$.

Assumption LA3′ requires that either (i) the moment equalities detect violations of the null hypothesis for a set of g functions with positive Q measure or (ii) the moment inequalities are not too far from being binding, i.e., $h_{1,j}(g) < \infty$, and the perturbation λ_0 occurs in a direction that yields moment inequality violations for a set of g functions with positive Q measure.

Assumption LA3 is employed with the KS test. It is weaker than Assumption LA3′, which is employed for the CvM test. If Assumption LA3 holds with $\lambda = \beta\lambda_0$ (and some other assumptions), then the power of KS-GMS and KS-PA tests against the perturbation λ is arbitrarily close to one for β arbitrarily large. For brevity, we do not prove this here. The proof is analogous to the proof of such results for the KS tests considered in AS1 and AS2, see Section 13.1 of Appendix B and Section 15.1 of Appendix D of AS2.

Assumptions LA3 and LA3$'$ can fail to hold even when the null hypothesis is violated. This typically happens if the true parameter/true distribution is fixed, i.e., $(\theta_n, F_n) = (\theta_0, F_0) \in \mathcal{F}$ for all n in Assumption NLA1(a), the null hypothesis parameter $\theta_{n,*}$ drifts with n as in Assumption NLA1(a), and $P_{F_0}(X_i \in \mathcal{X}_{zero}|Z_i = z_0) = 0$, where $\mathcal{X}_{zero} = \{x \in R^{d_x} : E_{F_0}(m(W_i, \theta_0)|X_i = x, Z_i = z_0) = 0\}$. In such cases, typically $h_{1,j}(g) = \infty$ $\forall g \in \mathcal{G}$ (because the conditional moment inequalities are non-binding with probability one), Assumptions LA3 and LA3 $'$ fail, and Theorem AN3 below shows that GMS and PA tests have trivial asymptotic power against these $(nb^{d_z})^{-1/2}$-local alternatives. See Section 11 for local power results that apply when Assumption LA3 or LA3$'$ fail to hold.

The asymptotic distribution of $T_n(\theta_{n,*})$ under $(nb^{d_z})^{-1/2}$-local alternatives is shown to be $J_{h,\lambda}$. By definition, $J_{h,\lambda}$ is the distribution of

$$T(h_1 + \Pi_0\lambda, h_2) = \int S(\nu_{h_2}(g) + h_1(g) + \Pi_0(g)\lambda, h_2(g) + \varepsilon I_k)dQ(g), \tag{10.3}$$

where $h = (h_1, h_2)$, Π_0 denotes $\Pi_0(\cdot)$, and $\nu_{h_2}(\cdot)$ is a mean zero Gaussian process with covariance kernel $h_2 = h_{2,F_0}(\theta_0)$. For notational simplicity, the dependence of $J_{h,\lambda}$ on Π_0 is suppressed.

Next, we introduce two assumptions, viz., Assumptions NLA4 and LA5, that are used only for GMS tests in the context of local alternatives. The asymptotic analogues of $\overline{\Sigma}_n(\theta, g)$ and its diagonal matrix are

$$\overline{\Sigma}_F(\theta, g, z_0) = \Sigma_F(\theta, g, g, z_0) + \varepsilon\Sigma_F(\theta, 1_k, 1_k, z_0) \text{ and } \overline{D}_F(\theta, g, z_0) = Diag(\overline{\Sigma}_F(\theta, g, z_0)), \tag{10.4}$$

where $\Sigma_F(\theta, g, g, z_0)$ is defined in (7.6).

Assumption NLA4. $\kappa_n^{-1}(nb^{d_z})^{1/2}\overline{D}_{F_n}^{-1/2}(\theta_n, g, z_0)E_{F_n}(m(W_i, \theta_n, g)|Z_i = z_0)f(z_0) \to \pi_1(g)$, where $\pi_1(g) = (\pi_{1,1}(g), ..., \pi_{1,k}(g))'$, for some $\pi_{1,j}(g) \in R_{+,\infty}$ for $j = 1, ..., p$, $\pi_{1,j}(g) = 0$ for $j = p+1, ..., k$, and all $g \in \mathcal{G}$.

In Assumption NLA4 the functions are evaluated at the true value θ_n, not at the null value $\theta_{n,*}$, and $(\theta_n, F_n) \in \mathcal{F}$. In consequence, the moment functions in Assumption NLA4 satisfy the moment inequalities and $\pi_{1,j}(g) \geq 0$ for all $j = 1, ..., p$ and $g \in \mathcal{G}$. Note that

$0 \le \pi_{1,j}(g) \le h_{1,j}(g)$ for all $j = 1, ..., p$ and all $g \in \mathcal{G}$ (by Assumption NLA1(b) and $\kappa_n \to \infty$.)

Let $c_0(\varphi(\pi_1), h_2, 1-\alpha)$ denote the $1-\alpha$ quantile of

$$\begin{aligned} T(\varphi(\pi_1), h_2) &= \int S(\nu_{h_2}(g) + \varphi(\pi_1(g)), h_2(g) + \varepsilon I_k) dQ(g), \text{ where} \\ \varphi(\pi_1(g)) &= (\varphi(\pi_{1,1}(g)), ..., \varphi(\pi_{1,p}(g)), 0, ..., 0)' \in R^k \text{ and} \\ \varphi(x) &= 0 \text{ if } x \le 1 \text{ and } \varphi(x) = \infty \text{ if } x > 1. \end{aligned} \tag{10.5}$$

Let $\varphi(\pi_1)$ denote $\varphi(\pi_1(\cdot))$. The probability limit of the GMS critical value $c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$ is shown below to be $c(\varphi(\pi_1), h_2, 1-\alpha) = c_0(\varphi(\pi_1), h_2, 1-\alpha+\eta) + \eta$.

Assumption LA5. (a) $Q(\mathcal{G}_\varphi) = 1$, where $\mathcal{G}_\varphi = \{g \in \mathcal{G} : \pi_{1,j}(g) \neq 1 \text{ for } j = 1, ..., p\}$, and

(b) the distribution function (df) of $T(\varphi(\pi_1), h_2)$ is continuous and strictly increasing at $x = c(\varphi(\pi_1), h_2, 1-\alpha)$, where $h_2 = h_{2,F_0}(\theta_0)$.

The value 1 that appears in $\mathcal{G}_\varphi$ in Assumption LA5(a) is the discontinuity point of φ. Assumption LA5(a) implies that the $(nb^{d_z})^{-1/2}$-local power formulae given below do not apply to certain "discontinuity vectors" $\pi_1(\cdot)$, but this is not particularly restrictive.[34] Assumption LA5(b) typically holds because of the absolute continuity of the Gaussian random variables $\nu_{h_2}(g)$ that enter $T(\varphi(\pi_1), h_2)$.[35]

The following assumption is used only for PA tests.

Assumption LA6. The df of $T(0_\mathcal{G}, h_2)$ is continuous and strictly increasing at $x = c(0_\mathcal{G}, h_2, 1-\alpha)$, where $h_2 = h_{2,F_0}(\theta_0)$.

The probability limit of the PA critical value is shown to be $c(0_\mathcal{G}, h_2, 1-\alpha) = c_0(0_\mathcal{G}, h_2, 1-\alpha+\eta) + \eta$, where $c_0(0_\mathcal{G}, h_2, 1-\alpha)$ denotes the $1-\alpha$ quantile of $J_{(0_\mathcal{G}, h_2), 0_{d_\theta}}$.

Theorem AN3. *Under Assumptions* B, K, NM, S1, S2, *and* NLA1-NLA2,

[34] Assumption LA5(a) is not particularly restrictive because in cases where it fails, one can obtain lower and upper bounds on the local asymptotic power of GMS tests by replacing $c(\varphi(\pi_1), h_2, 1-\alpha)$ by $c(\varphi(\pi_1-), h_2, 1-\alpha)$ and $c(\varphi(\pi_1+), h_2, 1-\alpha)$, respectively, in Theorem AN3(a). By definition, $\varphi(\pi_1-) = \varphi(\pi_1(\cdot)-)$ and $\varphi(\pi_1(g)-)$ is the limit from the left of $\varphi(x)$ at $x = \pi_1(g)$. Likewise $\varphi(\pi_1+) = \varphi(\pi_1(\cdot)+)$ and $\varphi(\pi_1(g)+)$ is the limit from the right of $\varphi(x)$ at $x = \pi_1(g)$.

[35] If Assumption LA5(b) fails, one can obtain lower and upper bounds on the local asymptotic power of GMS tests by replacing $J_{h,\lambda}(c(\varphi(\pi_1), h_2, 1-\alpha))$ by $J_{h,\lambda}(c(\varphi(\pi_1), h_2, 1-\alpha)+)$ and $J_{h,\lambda}(c(\varphi(\pi_1), h_2, 1-\alpha)-)$, respectively, in Theorem AN3(a), where the latter are the limits from the left and right, respectively, of $J_{h,\lambda}(x)$ at $x = c(\varphi(\pi_1), h_2, 1-\alpha)$.

(a) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(\varphi_n(\theta_{n,*}), \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)) = 1 - J_{h,\lambda}(c(\varphi(\pi_1), h_2, 1-\alpha))$ *provided Assumptions* GMS1, NLA4, *and* LA5 *also hold,*

(b) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)) = 1 - J_{h,\lambda}(c(0_{\mathcal{G}}, h_2, 1-\alpha))$ *provided Assumption* LA6 *also holds, and*

(c) $\lim_{\beta\to\infty}[1 - J_{h,\beta\lambda_0}(c(\varphi(\pi_1), h_2, 1-\alpha))] = \lim_{\beta\to\infty}[1 - J_{h,\beta\lambda_0}(c(0_{\mathcal{G}}, h_2, 1-\alpha))] = 1$ *provided Assumptions* LA3$'$, S3, *and* S4 *hold.*

Comments. 1. Theorems AN3(a) and AN3(b) provide the $(nb^{d_z})^{-1/2}$-local alternative power functions of the GMS and PA tests, respectively. Theorem AN3(c) shows that the asymptotic power of GMS and PA tests is arbitrarily close to one if the $(nb^{d_z})^{-1/2}$-local alternative parameter $\lambda = \beta\lambda_0$ is sufficiently large in the sense that its scale β is large.

2. We have $c(\varphi(\pi_1), h_2, 1-\alpha) \leq c(0_{\mathcal{G}}, h_2, 1-\alpha)$ (because $\varphi(\pi_1(g)) \geq 0$ for all $g \in \mathcal{G}$ and $S(m, \Sigma)$ is non-increasing in m_I by Assumption S1(b), where $m = (m_I', m_{II}')'$). Hence, the asymptotic local power of a GMS test is greater than or equal to that of a PA test. Strict inequality holds whenever $\pi_1(\cdot)$ is such that $Q(\{g \in \mathcal{G} : \varphi(\pi_1(g)) > 0\}) > 0$. The latter typically occurs whenever the conditional moment inequality $E_{F_n}(m_j(W_i, \theta_{n,*}) |X_i, Z_i = z_0)$ for some $j = 1, ..., p$ is bounded away from zero as $n \to \infty$ with positive X_i probability.

3. The results of Theorem AN3 hold under the null hypothesis as well as under the alternative. The results under the null quantify the degree of asymptotic non-similarity of the GMS and PA tests.

4. Suppose the assumptions of Theorem AN3 hold and each distribution F_n generates the same identified set, call it $\Theta_0 = \Theta_{F_n}$ $\forall n \geq 1$. Then, Theorem AN3(a) implies that the asymptotic probability that a GMS CS includes, $\theta_{n,*}$, which lies within $O((nb^{d_z})^{-1/2})$ of the identified set, is $J_{h,\lambda}(c(\varphi(\pi_1), h_2, 1-\alpha))$. If $\lambda = \beta\lambda_0$ and Assumptions LA3$'$, S3, and S4 also hold, then $\theta_{n,*}$ is not in Θ_0 (at least for β large) and the asymptotic probability that a GMS or PA CS includes $\theta_{n,*}$ is arbitrarily close to zero for β arbitrarily large by Theorem AN3(c). Analogous results hold for PA CS's.

11 Asymptotic Local Power and Comparisons with the CLR Test

In this Section, we derive local power results for the KS and CvM tests that cover the case where the DGP does not depend on n and the moment inequalities are binding only on a measure-zero set of X_i. The results of this section only yield the rates of convergence (of the null hypothesis parameter values to the true parameter value) for which the tests have non-trivial asymptotic power. In contrast, the results of Section 10 yield asymptotic distributions from which actual power approximations can be obtained.

Next, we compare the asymptotic power of the KS and CvM tests (in terms of rates of convergence) with that of the CLR test in a simple moment inequality model. We find that the KS and CvM tests have higher power than the CLR test for more flat conditional moment functions and lower power for more curved conditional moment functions.

11.1 Power Against $\mathbf{a_n}$-Local Alternatives

Here we study the asymptotic local power of the KS and CvM tests under conditions that allow for a fixed true DGP's with non-flat conditional moment functions, as well as DGP's that depend on n. The results are stated under high-level assumptions (specifically, Assumptions NLA7 and NLA7$'$ below). These assumptions are verified for a simple moment inequality model in Section 11.2 below.

For a sequence of positive constants $\{a_n : n \geq 1\}$ such that $a_n \to 0$, define a sequence of a_n-local alternatives suitable for the KS test as follows.

Assumption NLA7. The true parameters and distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ and the null parameters $\{\theta_{n,*} : n \geq 1\}$ satisfy $(\theta_{n,*}, F_n) \in \mathcal{F}_+$, and

(a) $\theta_{n,*} = \theta_n + a_n$, $\theta_n \to \theta_0$, and $F_n \to F_0$ for some $(\theta_0, F_0) \in \mathcal{F}$,

(b) $d(h_{2,F_n}(\theta_{n,*}), h_{2,F_0}(\theta_0)) \to 0$, and

(c) for some sequence $\{g_n \in \mathcal{G} : n \geq 1\}$, we have $\lim_{n\to\infty}(nb^{d_z})^{1/2} D_{F_n}^{-1/2}(\theta_{n,*}, z_0)$ $\times E_{F_n} m_{F_n}(\theta_{n,*}, X_i, z_0) g_n(X_i) \to h_1 \in [-\infty, \infty]^k$, where $h_{1,j} = -\infty$ for some $j \leq p$ or $|h_{1,j}| = \infty$ for some $j > p$ and $h_{1,j}$ denotes the jth element of h_1.

In Assumption NLA7, $\mathcal{F}_+$ is defined in the paragraph following (3.16). In Assumption NLA7(b), d is the uniform metric defined in (8.2).

The following assumption defines the sequence of a_n-local alternatives suitable for the CvM test.

Assumption NLA7$'$. The true parameters and distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ and the null parameters $\{\theta_{n,*} : n \geq 1\}$ satisfy $(\theta_{n,*}, F_n) \in \mathcal{F}_+$, Assumptions NLA7(a) and NLA7(b) hold and

(c) for some sequence $\{G_n \subseteq \mathcal{G} : n \geq 1\}$, $Q(G_n)^{1/\chi}(nb^{d_z})^{1/2} \min_{g \in G_n} \beta_{F_n}(\theta_{n,*}, g) \to \infty$, where $\beta_{F_n}(\theta_{n,*}, g)$ is defined as $\beta(g)$ is defined in (9.2) with F_0 and θ_* replaced by F_n and $\theta_{n,*}$ respectively, and χ is the degree of homogeneity in Assumption S4.

The following theorem shows that the KS and CvM tests have power that approaches one under the sequences defined in Assumptions NLA7 and NLA7′, respectively.

Theorem AN4. *Suppose Assumptions* B, K, NM, *and* S1-S4 *hold. In addition, suppose Assumption* NLA7 *holds when the* KS *statistic* (*defined in* (6.6)) *is used and Assumption* NLA7′ *holds when the* CvM *statistic* (*defined in* (6.7)) *is used. Then,*

(a) $\lim_{n\to\infty} \Pr_{F_n}(T_n(\theta_{n,*}) > B) = 1$ *for any constant* $B > 0$,

(b) $\lim_{n\to\infty} \Pr_{F_n}\left(T_n(\theta_{n,*}) > c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)\right) = 1,$ *and*

(c) $\lim_{n\to\infty} \Pr_{F_n}\left(T_n(\theta_{n,*}) > c(\varphi_n(\theta_{n,*}), \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)\right) = 1.$

Comments. 1. Theorem AN4(a) shows that the test statistic $T_n(\theta_{n,*})$ diverges to infinity in probability under the sequence of local alternatives. Theorem AN4(b) and AN4(c) show that the tests employing the PA and GMS critical values, respectively, reject the null hypothesis with probability that goes to one as $n \to \infty$.

2. The proof of Theorem AN4 is given in Section 12.3 below.

11.2 A Non-flat Bound Example

Now we verify Assumptions NLA7 and NLA7′ for a simple moment inequality example. To maximize clarity, we consider a fixed true parameter: $(\theta_n, F_n) = (\theta_0, F_0)$ for all n.

Example. We consider the following moment inequality model with $k = p = 1$ and $d_x = d_z = 1$:

$$E_{F_0}(Y_i - \theta_0 | X_i = x, Z_i = z_0) \geq 0, \text{ a.s. } [F_{X,0}], \tag{11.6}$$

where $z_0 = 0$. The identified set for θ_0 is $(-\infty, \bar{\theta}]$, where $\bar{\theta} = \min_x E_{F_0}(Y_i | X_i = x, Z_i = 0)$.

We consider the fixed distribution F_0 under which $X_i, Z_i \sim$Unif$([-1/2, 1/2])$, X_i and Z_i are independent, $E_{F_0}(Y_i | X_i = x, Z_i = 0)$ is uniquely minimized at $x = 0$, $E(Y_i | X_i = x, Z_i = z_0) = c|x|^\pi$ for some $c, \pi > 0$ and x in a neighborhood of 0, and $\inf_x \text{Var}(Y_i | X_i = x, Z_i = 0) > 0$. For this F_0, $\bar{\theta} = 0$, $m_{F_0}(\theta, x, z_0) = c|x|^\pi - \theta$, and $\inf_\theta D_{F_0}(\theta, z_0) > 0$. We consider the true parameter $\theta_0 = \bar{\theta} \equiv 0$.

To be consistent with the rest of the paper, we consider an S function with $\chi = 2$, $\mathcal{G} = \mathcal{G}_{c-cube}$, and Q as defined in Section 6.4 with $w(r) = (r^2 + 100)^{-1}$.

For this example, the following Theorem, combined with Theorem AN4, characterizes the sequences of null parameters that the KS and CvM tests reject with probability approaching one as $n \to \infty$.

Theorem AN5. (a) *Assumption* NLA7 *is satisfied for the example in* (11.6) *with null hypothesis parameter values* $\theta_{n,*} = \theta_0 + a_n$ *if*

$$a_n(nb)^{\pi/(2\pi+2)} \to \infty. \tag{11.7}$$

(b) *Assumption* NLA7′ *is satisfied for the example in* (11.6) *with null hypothesis parameter values* $\theta_{n,*} = \theta_0 + a_n$ *if*

$$a_n(nb)^{\pi/(2\pi+5)} \to \infty. \tag{11.8}$$

Proof of Theorem AN5. We verify only part (c) of Assumptions NLA7 and NLA7′ because parts (a) and (b) of these assumptions are straightforward to verify.

To verify Assumption NLA7(c), let r_n^* be the smallest integer such that $(2r_n^*)^{-1} \leq (\theta_{n,*}/c)^{1/\pi}$. For n large enough, we have $(2r_n^*)^{-1} > (\theta_{n,*}/c)^{1/\pi}/2$ (because otherwise $(2(r_n^* - 1))^{-1} \leq (\theta_{n,*}/c)^{1/\pi}$). Let $g_n(x) = 1(x \in (0, (2r_n^*)^{-1}])$. Then, $g_n \in \mathcal{G}_{c-cube}$ and

$$\begin{aligned} E_{F_0} m_{F_0}(\theta_{n,*}, X_i, z_0) g_n(X_i) &= \int_0^{(2r_n^*)^{-1}} (c \cdot x^\pi - \theta_{n,*}) dx \\ \leq \int_0^{(\theta_{n,*}/c)^{1/\pi}/2} (c \cdot x^\pi - \theta_{n,*}) dx &= -\frac{2^\pi(\pi+1) - 1}{2^{\pi+1}(\pi+1)} c^{-1/\pi} \theta_{n,*}^{1+1/\pi}, \end{aligned} \tag{11.9}$$

where the first inequality holds because $c \cdot x^\pi - \theta_{n,*} < 0$ on the integral range and the second equality holds by direct calculation. Given (11.7) and (11.9), additional elementary algebra shows that Assumption NLA7(c) holds.

To verify Assumption NLA7′(c), consider $G_n = \{g_n\}$ with g_n defined as in the KS case above. Then,

$$\begin{aligned} Q(G_n) &= (2r_n^*)^{-1} \cdot ((r_n^*)^2 + 100)^{-1} \\ &\geq 2^{-1}(\theta_{n,*}/c)^{1/\pi} \cdot ((\theta_{n,*}/c)^{-2/\pi} + 100)^{-1} \\ &= 2^{-1}(\theta_{n,*}/c)^{3/\pi}(1 + o(1)). \end{aligned} \tag{11.10}$$

Thus, $Q(G_n)^{1/2}(nb)^{1/2}\beta_{F_0}(\theta_{n,*}, g_n) \geq \bar{C} \cdot (1 + o(1))(nb)^{1/2}\theta_{n,*}^{(2\pi+5)/(2\pi)}$ for some constant $\bar{C} > 0$ (because $\beta_{F_0}(\theta_{n,*}, g_n) = E_{F_0} m_{F_0}(\theta_{n,*}, X_i, z_0) g_n(X_i) \geq \frac{2^{\pi}(\pi+1)-1}{2^{\pi+1}(\pi+1)} c^{-1/\pi}\theta_{n,*}^{1+1/\pi}$ in the present case). This, (11.8), and some additional elementary algebra show Assumption NLA7′(c) holds. □

11.3 Power Comparisons with the CLR Test

In this subsection, we continue the example above and compare the local power properties of the KS and CvM tests to that of the CLR test.

The CLR test is based on nonparametric estimation of $E(Y|X = x, X = z_0)$. Suppose that the uniform convergence rate of the nonparametric estimator of this conditional expectation is γ_n, where $\gamma_n \to \infty$. Then, by Theorems 1-3 of CLR, the CLR test has power approaching one as $n \to \infty$ if $\theta_{n,*} - \theta_0$ converges to zero slower than γ_n^{-1}, that is,

$$\gamma_n(\theta_{n,*} - \theta_0) \to \infty. \tag{11.11}$$

The relative power properties of KS, CvM, and CLR tests is obtained by comparing (11.7), (11.8) and (11.11). Specifically, the KS tests have better asymptotic local power than the CLR test if $(nb)^{-\pi/(2\pi+2)}\gamma_n \to 0$. The opposite is true if $\gamma_n^{-1}(nb)^{\pi/(2\pi+2)} \to 0$. The CvM test has better asymptotic local power than the CLR test if $(nb)^{-\pi/(2\pi+5)}\gamma_n \to 0$. The opposite is true if $\gamma_n^{-1}(nb)^{\pi/(2\pi+5)} \to 0$.

The conditions above translate into thresholds for π, above which the KS and CvM tests have better asymptotic local power than the CLR test, and below which the opposite is true. For the KS test versus the CLR test, the π threshold implied by the above conditions is

$$\pi^* = \frac{2\log\gamma_n}{\log nb - 2\log\gamma_n}, \tag{11.12}$$

which solves $(nb)^{-\pi^*/(2\pi^*+2)}\gamma_n = 1$. For the CvM test versus the CLR test, the π threshold is

$$\pi^* = \frac{5\log\gamma_n}{\log nb - 2\log\gamma_n}. \tag{11.13}$$

By design, π controls the flatness of the curve $E(Y_i|X_i = x, Z_i = z_0)$ at its bottom, with a larger π yielding a flatter curve. The above analysis shows that the KS and CvM tests have higher asymptotic local power than the CLR test for flatter (but not necessarily completely flat) bound curves, while the CLR test has higher power for more curved bound curves.

Next, we calculate the threshold for π for the b chosen in this paper and the γ_n implied by the recommended tuning parameters in CLR. In this paper, we choose $b \approx n^{-2/7}$ and thus $nb \approx n^{5/7}$, where $c_n \approx d_n$ means $c_n = O(d_n)$ and $d_n = O(c_n)$. In CLR, for the local linear version of their test, the recommended bandwidth is $h \approx n^{-1/6}n^{1/10}n^{-1/7} = n^{-44/210}$. This implies a pointwise convergence rate for the local linear bound estimator of $n^{1/2}h \approx n^{1/2}n^{-44/210} = n^{61/210}$. The uniform convergence rate should be slightly slower, giving $\gamma_n = o(n^{61/210})$. Thus, the π threshold for the KS test versus the CLR test is

$$\frac{61/210}{5/14 - 61/210} = \frac{61}{14} \doteq 4.4. \tag{11.14}$$

The π threshold for the CvM test versus the CLR test is

$$\frac{5 \times 61/210}{5/7 - 2 \times 61/210} = \frac{305}{28} \doteq 10.9. \tag{11.15}$$

Finally, we note that the analysis in this section only compares the asymptotic local power of the tests under a fixed true (θ_0, F_0). It does not necessarily have a direct implication for the relative power of the tests for any given finite sample size n when the bound curve is not completely flat. In fact, the Monte Carlo experiments in this paper and in AS1 show that the CvM tests have higher finite-sample power than the CLR test for bound curves that are not as flat as $c|x|^{10}$. Such finite-sample behavior can be explained by the asymptotic local power results under drifting sequences of true DGP's given in Section 10. We believe these provide better finite-sample approximations than the results of this section.

12 Proofs

12.1 Proof of Lemma AN1

Proof of Lemma AN1. We have: $\theta \notin \Theta_F(\mathcal{G})$ implies that $E_F(m_j(W_i, \theta)g_j(X_i)|Z_i = z_0) < 0$ for some $j \leq p$ or $E_F(m_j(W_i, \theta)g_j(X_i)|Z_i = z_0) \neq 0$ for some $j = p+1, ..., k$. By the law of iterated expectations and $g_j(x) \geq 0$ for all $x \in R^{d_x}$ and $j \leq p$, this implies that $P_F(X_i \in \mathcal{X}_F(\theta)|Z_i = z_0) > 0$ and, hence, $\theta \notin \Theta_F$.

On the other hand, $\theta \notin \Theta_F$ implies that $P_F(X_i \in \mathcal{X}_F(\theta)|Z_i = z_0) > 0$ and the latter implies that $\theta \notin \Theta_F(\mathcal{G})$ by Assumption NCI. $\square$

12.2 Proof of Theorem AN1

In this section, we prove Theorem AN1. We start by introducing some notation. Next, we establish Theorem AN4, which is used in the proof of Theorem AN1. To prove Theorem AN4 we use Lemmas AN2-AN4. The proofs of the latter use Lemmas AN5-AN7.

12.2.1 Notation

First, we define sample-size n population analogues of the asymptotic covariance kernels that are defined in (7.6). We make their dependence on $b = b_n$ explicit. Let[36]

$$\begin{aligned} h_{2,F}(\theta, g, g^*, b) &= D_F^{-1/2}(\theta, b)\Sigma_F(\theta, g, g^*, b)D_F^{-1/2}(\theta, b) \\ &= Cov_F\left(D_F^{-1/2}(\theta, b)m(W_i, \theta, g, b), D_F^{-1/2}(\theta, b)m(W_i, \theta, g^*, b)\right), \\ \Sigma_F(\theta, g, g^*, b) &= Cov_F(m(W_i, \theta, g, b), m(W_i, \theta, g^*, b)), \text{ and} \qquad (12.1) \\ D_F(\theta, b) &= Diag(\Sigma_F(\theta, 1_k, 1_k, b)) \ (= Diag(Var_F(b^{-d_z/2}K_b(Z_i)m(W_i, \theta)))). \end{aligned}$$

Let $h_{2,F}(\theta, b)$ abbreviate the sample-size n covariance kernel $\{h_{2,F}(\theta, g, g^*, b) : g, g^* \in \mathcal{G}\}$ of $n^{1/2}\overline{m}_n(\theta, g)$, which depends on n through b.

Next, define

$$\begin{aligned} h_{1,n,F}(\theta, g, b) &= n^{1/2}D_F^{-1/2}(\theta, b)E_F m(W_i, \theta, g, b), \\ h_{1,n,F}^{\dagger}(\theta, g, b) &= (nb^{d_z})^{1/2}D_F^{-1/2}(\theta, b)E_F(m_F(\theta, X_i, z_0) \odot g(X_i)), \\ h_{n,F}^{\dagger}(\theta, g, g^*, b) &= (h_{1,n,F}^{\dagger}(\theta, g, b), h_{2,F}(\theta, g, g^*, b)), \\ \widehat{h}_{2,n,F}(\theta, g, g^*, b) &= D_F^{-1/2}(\theta, b)\widehat{\Sigma}_n(\theta, g, g^*)D_F^{-1/2}(\theta, b), \\ \overline{h}_{2,n,F}(\theta, g, b) &= \widehat{h}_{2,n,F}(\theta, g, g, b) + \varepsilon\widehat{h}_{2,n,F}(\theta, 1_k, 1_k, b) \\ &= D_F^{-1/2}(\theta, b)\overline{\Sigma}_n(\theta, g)D_F^{-1/2}(\theta, b), \text{ and} \qquad (12.2) \\ \nu_{n,F}(\theta, g, b) &= n^{-1/2}\sum_{i=1}^{n} D_F^{-1/2}(\theta, b)[m(W_i, \theta, g, b) - E_F m(W_i, \theta, g, b)], \end{aligned}$$

[36]For simplicity, there is some abuse of notation in the definitions in (12.1) because $h_{2,F}(\theta, g, g^*, b)$ has a different definition than $h_{2,F}(\theta, g, g^*, z_0)$ in (7.6), but the only difference in the notation is b versus z_0. The same is true for $\Sigma_F(\theta, g, g^*, b)$ and $D_F(\theta, b)$ versus $\Sigma_F(\theta, g, g^*, z_0)$ and $D_F(\theta, z_0)$.

where $m_F(\theta, x, z)$, $m(W_i, \theta, g, b)$, $\overline{\Sigma}_n(\theta, g)$, and $\widehat{\Sigma}_n(\theta, g, g^*)$ are defined in (2.15) of ASN, (6.3), (6.5), and (7.7), respectively, and $A \odot B$ denotes the direct (i.e., element by element) product of two matrices or vectors, A and B, with the same dimensions. Below we write $T_n(\theta)$ as a function of the quantities in (12.2). As defined, (i) $h_{1,n,F}(\theta, g, b)$ is the k-vector of normalized means of the moment functions $D_F^{-1/2}(\theta, b)m(W_i, \theta, g, b)$ for $g \in \mathcal{G}$, which measure the slackness of the population moment conditions under (θ, F), (ii) $h^\dagger_{1,n,F}(\theta, g, b)$ is an approximation to $h_{1,n,F}(\theta, g, b)$ that has the very useful feature that it is non-negative when $(\theta, F) \in \mathcal{F}$ because $m_F(\theta, X_i, z_0) \geq 0$ a.s. by (2.15) of ASN and Assumptions PS1(c) and (d) stated in ASN, (ii) $h^\dagger_{n,F}(\theta, g, g^*, b)$ contains the approximation to the normalized means of $D_F^{-1/2}(\theta, b)m(W_i, \theta, g, b)$ and the covariances of $D_F^{-1/2}(\theta, b)$ $m(W_i, \theta, g, b)$ and $D_F^{-1/2}(\theta, b)m(W_i, \theta, g^*, b)$, (iii) $\widehat{h}_{2,n,F}(\theta, g, g^*, b)$ and $\overline{h}_{2,n,F}(\theta, g, b)$ are hybrid quantities—part population, part sample—based on $\widehat{\Sigma}_n(\theta, g, g^*)$ and $\overline{\Sigma}_n(\theta, g)$, respectively, and (iv) $\nu_{n,F}(\theta, g, b)$ is the sample average of $D_F^{-1/2}(\theta, b)m(W_i, \theta, g, b)$ normalized to have mean zero and variance that is $O(1)$ but not $o(1)$. Note that $\nu_{n,F}(\theta, \cdot, b)$ is an empirical process indexed by $g \in \mathcal{G}$ with covariance kernel given by $h_{2,F}(\theta, g, g^*, b)$.

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

$$n^{1/2}\overline{m}_n(\theta, g) = D_F^{1/2}(\theta, b)(\nu_{n,F}(\theta, g, b) + h_{1,n,F}(\theta, g, b)). \tag{12.3}$$

The test statistic $T_n(\theta)$, defined in (6.6), can be written as

$$T_n(\theta) = \int S(\nu_{n,F}(\theta, g, b) + h_{1,n,F}(\theta, g, b), \overline{h}_{2,n,F}(\theta, g, b))dQ(g). \tag{12.4}$$

Note the close resemblance between $T_n(\theta)$ and $T(h)$ (defined in (7.1)).

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

For notational simplicity, for any function of the form $r_F(\theta, g, b)$ for $g \in \mathcal{G}$, let $r_F(\theta, b)$ denote the function $r_F(\theta, \cdot, b)$ on $\mathcal{G}$. Correspondingly, for any function of the form $r_F(\theta, g, g^*, b)$ for $g, g^* \in \mathcal{G}$, let $r_F(\theta, b)$ denote the function $r_F(\theta, \cdot, \cdot, b)$ on $\mathcal{G}^2$.

12.2.2 Theorem AN4

The following Theorem provides a uniform asymptotic distributional result for the test statistic $T_n(\theta)$. It is an analogue of Theorem 1 of AS1. It used in the proof of Theorem AN1.

Theorem AN4. *Suppose Assumptions* B, K, NM, S1, *and* S2 *hold. Then, for every* *compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$*, all constants* $x_{h^\dagger_{n,F}(\theta,b)} \in R$ *that may depend on* (θ, F) *and* n *through* $h^\dagger_{n,F}(\theta, b)$*, and all* $\delta > 0$*, we have*

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

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

where $T(h) = \int S(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)dQ(g)$*,* $\nu_{h_2}(\cdot)$ *is the Gaussian process defined in* (7.2)*, and* $h^\dagger_{n,F}(\theta, b) = h^\dagger_{n,F}(\theta, \cdot, \cdot, b)$ *is defined in* (12.2)*.*

Comments. 1. Theorem AN4 gives a uniform asymptotic approximation to the distribution function of $T_n(\theta)$. Uniformity holds without any restrictions on the true normalized mean (i.e., moment inequality slackness) functions $\{h_{1,n,F_n}(\theta_n, b) : n \geq 1\}$. In particular, Theorem AN4 does not require $\{h_{1,n,F_n}(\theta_n, b) : n \geq 1\}$ to converge as $n \to \infty$ or to belong to a compact set. The Theorem does not require that $T_n(\theta)$ has a unique asymptotic distribution under any sequence $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$.

2. The supremum and infimum in Theorem AN4 are over compact sets of asymptotic covariance kernels $\mathcal{H}_{2,cpt}$, rather than the parameter spaces $\mathcal{H}_2$ of covariance kernels. This is not particularly problematic because the potential asymptotic size problems that arise in moment inequality models are due to the pointwise discontinuity of the asymptotic distribution of the test statistic as a function of the means of the moment inequality functions, not as a function of the covariances between different moment inequalities.

12.2.3 Lemmas AN2-AN4

The proof of Theorem AN4 uses the following three Lemmas. The first Lemma is a key result that establishes that the finite-sample covariance kernel $h_{2,F}(\theta, b)$ converges to the asymptotic covariance kernel $h_{2,F}(\theta)$ in the sup norm d uniformly over $(\theta, F) \in \mathcal{F}_+$.

Lemma AN2. *Suppose Assumptions* B, K, *and* NM *hold. Then,*

$$\text{(a)} \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g,g^*\in\mathcal{G}} \|\Sigma_F(\theta,g,g^*,b) - \Sigma_F(\theta,g,g^*,z_0)\| \to 0,$$

$$\text{(b)} \sup_{(\theta,F)\in\mathcal{F}_+} \left\|D_F^{-1}(\theta,z_0)D_F(\theta,b) - I_k\right\| \to 0, \text{ and}$$

$$\text{(c)} \sup_{(\theta,F)\in\mathcal{F}_+} d(h_{2,F}(\theta,b), h_{2,F}(\theta)) \to 0.$$

Comment. Lemma AN2 is a key ingredient in the proof of Lemma AN4, which in turn is used in the proofs of Theorems AN4 and AN1. See Comment 3 to Lemma AN4 for a description of how Lemma AN2 is employed.

The next Lemma shows that the bias due to taking averages over values z ($\neq z_0$) for which the conditional moment inequalities in (2.1) of ASN do *not* hold is negligible asymptotically.

Lemma AN3. *Suppose Assumptions* B, K, *and* NM *hold. Then,*

$$\limsup_{n\to\infty} \sup_{(\theta,F)\in\mathcal{F}} \sup_{g\in\mathcal{G}} ||h_{1,n,F}(\theta,g,b) - h^{\dagger}_{1,n,F}(\theta,g,b)|| \to 0.$$

Comment. For Lemma AN3 to hold, a key feature of the definition of $h^{\dagger}_{1,n,F}(\theta,g,b)$, given in (12.2), is that the normalization is by $D_F^{-1/2}(\theta,b)$ (not $D_F^{-1/2}(\theta,z_0)$), which is the same normalization as in $h_{1,n,F}(\theta,g,b)$.

The next Lemma is analogous to Lemma A1 of AS2. It is used in the proofs of Theorems AN4 and AN1-AN3. It establishes a functional CLT and uniform LLN for certain independent non-identically distributed empirical processes as well as uniform convergence of the estimator of the covariance kernel.

Let $\mathcal{H}_{2,+} = \{h_{2,F}(\theta) : (\theta,F) \in \mathcal{F}_+\}$. By definition, $\mathcal{H}_{2,+}$ is a set of $k\times k$-matrix-valued covariance kernels on $\mathcal{G}\times\mathcal{G}$ that includes $\mathcal{H}_2$.

Definition SubSeq(h₂). For $h_2 \in \mathcal{H}_{2,+}$, $SubSeq(h_2)$ is the set of subsequences $\{(\theta_{a_n}, F_{a_n}) \in \mathcal{F}_+ : n \geq 1\}$, where $\{a_n : n \geq 1\}$ is some subsequence of $\{n\}$, for which

$$\text{(i)} \lim_{n\to\infty} \sup_{g,g^*\in\mathcal{G}} ||h_{2,F_{a_n}}(\theta_{a_n},g,g^*) - h_2(g,g^*)|| = 0$$

and (ii) $\{W_i : i \geq 1\}$ are i.i.d. under F_{a_n}.

Note that the definition of $SubSeq(h_2)$ here differs from the definition of $SubSeq(h_2)$ in AS2 because (i) the summands of the sample averages are $m(W_i, \theta, g, b) = b^{-d_z/2} K_b(Z_i)$ $m(W_i, \theta, g)$, rather than $m(W_i, \theta, g)$, and $\{m(W_i, \theta, g, b) m(W_i, \theta, g^*, b)' : n \geq 1\}$ is not uniformly integrable, which complicates the proof of Lemma AN4(b) below, (ii) $SubSeq$ (h_2) requires $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}_+$, and (iii) $SubSeq(h_2)$ does not impose any conditions related to Assumption NM. The latter are imposed separately in the results below.

The sample paths of the Gaussian process $\nu_{h_2}(\cdot)$, which is defined in (7.2) and appears in the following Lemma, are bounded and uniformly ρ-continuous a.s. The pseudo-metric ρ on $\mathcal{G}$ is a pseudo-metric commonly used in the empirical process literature:

$$\rho^2(g, g^*) = tr\left(h_2(g, g) - h_2(g, g^*) - h_2(g^*, g) + h_2(g^*, g^*)\right). \tag{12.5}$$

For $h_2(\cdot, \cdot) = h_{2,F}(\theta, \cdot, \cdot)$, where $(\theta, F) \in \mathcal{F}$, this metric can be written equivalently as

$$\begin{aligned} \rho^2(g, g^*) &= E_F ||D_F^{-1/2}(\theta)[\widetilde{m}(W_i, \theta, g) - \widetilde{m}(W_i, \theta, g^*)]||^2, \text{ where} \\ \widetilde{m}(W_i, \theta, g) &= m(W_i, \theta, g) - E_F m(W_i, \theta, g). \end{aligned} \tag{12.6}$$

Lemma AN4. *Suppose Assumptions* B *and* NM *hold. For any subsequence* $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$ with $h_2 \in \mathcal{H}_{2,+}$,
(a) $\nu_{a_n, F_{a_n}}(\theta_{a_n}, \cdot, b_{a_n}) \Rightarrow \nu_{h_2}(\cdot)$ *as* $n \to \infty$ (*as processes indexed by* $g \in \mathcal{G}$), *and*
(b) $\sup_{g, g^* \in \mathcal{G}} ||\widehat{h}_{2, a_n, F_{a_n}}(\theta_{a_n}, g, g^*, b_{a_n}) - h_2(g, g^*)|| \to_p 0$ *as* $n \to \infty$.

Comments. 1. To obtain uniform asymptotic coverage probability results for CS's, Lemma AN4 is applied with $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ and $h_2 \in \mathcal{H}_2$. To obtain power results under fixed and local alternatives, Lemma AN4 is applied with $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}_+ \backslash \mathcal{F}$ for all $n \geq 1$ and $h_2 \in \mathcal{H}_{2,+}$.

2. Assumption PS3(d) stated in ASN only needs to hold with an exponent $2 + \delta$ for some $\delta > 0$, rather than 4, for Lemma AN4(a) to hold. For Lemma AN4(b), which gives consistency of the estimator of the covariance kernel, the exponent 4 is needed to control the variance of the covariance estimator.

3. The proof of Lemma AN4(a) is an extension of the proof of Lemma A1 of AS2 (which is given in Appendix E of AS2). The proof of Lemma AN4(b) is different from that of Lemma A1 of AS2 because the summands $m(W_i, \theta, g, b)$ are not uniformly integrable, so a standard uniform law of large numbers cannot be employed. Rather, an empirical process maximal inequality is utilized.

4. To prove Theorem AN4, we adjust the proof of Theorem 1 of AS1. The proof of Theorem 1 of AS1 uses a subsequence argument to reduce a uniform result over $(\theta, F) \in \mathcal{F}$ for which $h_{2,F}(\theta) \in \mathcal{H}_{2,cpt}$ as $n \to \infty$ to a result for a subsequence $\{(\theta_{a_n}, F_{a_n}) \in \mathcal{F} : n \geq 1\}$ for which the covariance kernels $\{h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) : n \geq 1\}$ satisfy $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ for some limit $h_{2,0} \in \mathcal{H}_2$.

In AS1 and AS2, the covariance kernel $h_{2,F}(\theta)$ of $\nu_n(\theta, \cdot)$ is a normalized sum of terms $m(W_i, \theta, g)$ and does not depend on n. Hence, the sample-size n and the asymptotic covariance kernels are the same. In contrast, in this paper, the covariance kernel $h_{2,F}(\theta, b)$ of $\nu_{n,F}(\theta, \cdot, b)$ is a normalized sum of terms $m(W_i, \theta, g, b)$ and it depends on n through b. Here, the subsequence of covariance kernels $\{h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) : n \geq 1\}$ (that arises from the subsequence argument in AS2) is a subsequence of asymptotic kernels. We use Lemma AN2(c) to show that if $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$, then the sample-size a_n covariance kernel $h_{2,F_{a_n}}(\theta_{a_n}, b_{a_n})$ satisfies $d(h_{2,F_{a_n}}(\theta_{a_n}, b_{a_n}), h_{2,0}) \to 0$ as $n \to \infty$. This holds because

$$\begin{aligned} & d(h_{2,F_{a_n}}(\theta_{a_n}, b_{a_n}), h_{2,0}) \\ \leq\ & d(h_{2,F_{a_n}}(\theta_{a_n}, b_{a_n}), h_{2,F_{a_n}}(\theta_{a_n})) + d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \\ \leq\ & \sup_{(\theta,F)\in\mathcal{F}} d(h_{2,F}(\theta, b_{a_n}), h_{2,F}(\theta)) + d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \\ \to\ & 0, \end{aligned} \tag{12.7}$$

where the first inequality holds by the triangle inequality and the convergence holds by Lemma AN2(c). The convergence result in (12.7) is the condition that is needed to obtain the weak convergence of the empirical process $\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot, b_{a_n})$ in Lemma AN4(a).

12.2.4 Proofs of Theorems AN4 and AN1

Proof of Theorem AN4. We alter the proof of Theorem 1 of AS1 to prove Theorem AN4. The statements of Theorem 1 of AS1 and Theorem AN4 differ because $h_{n,F}(\theta)$ appears in the former result, whereas $h^\dagger_{n,F}(\theta, b)$ appears in the latter. The proof of Theorem 1 of AS1 is given in AS2. Throughout this proof, $x_{h_{a_n},F_{a_n}(\theta_{a_n})}$ is replaced by $x_{h^\dagger_{a_n},F_{a_n}(\theta_{a_n},b_{a_n})}$. By Lemma AN3, for the sequence $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\}$ that appears in the proof in AS2, we have

$$\sup_{g\in\mathcal{G}} ||h_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n}) - h^{\dagger}_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n})|| \to 0. \tag{12.8}$$

We define $\widetilde{T}_{a_n}$ as in (12.5) of AS2, but change the definition of $\widetilde{T}_{a_n,0}$ to

$$\widetilde{T}_{a_n,0} = \int S\left(\tilde{\nu}_0(g) + h^{\dagger}_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n}), h^{\varepsilon}_{2,0}(g)\right) dQ(g). \tag{12.9}$$

By construction, $\widetilde{T}_{a_n,0}$ has the same distribution as $T(h^{\dagger}_{a_n,F_{a_n}}(\theta_{a_n}, b_{a_n}))$ for all $n \geq 1$. With this change in the definition of $\widetilde{T}_{a_n,0}$, we need to show that (12.7) of AS2 holds. The rest of the proof of Theorem 1 given in AS2 goes through without any changes.

We change the proof of (12.7) by replacing $\tilde{\nu}_{a_n}(g)(\omega)$ by

$$\tilde{\nu}_{a_n}(g)(\omega) + h^{\dagger}_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n}) - h_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n}) \tag{12.10}$$

in (12.10), (12.12), and (12.13) of AS2. The quantity in (12.10) converges to $\tilde{\nu}_0(g)(\omega)$ for all $\omega \in \widetilde{\Omega}$ using (12.8) above. Given the (12.10) replacement, (12.11) of AS2 holds with $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n})$ replaced by $h^{\dagger}_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n})$ in the first summand on the lhs. In addition, $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n})$ is replaced by $h^{\dagger}_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n})$ in the second summand on the lhs of (12.11) due to the new definition of $\widetilde{T}_{a_n,0}$ given in (12.9). With the above changes, the first line of (12.14) of AS2 holds with $\tilde{\nu}_{a_n}(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n})$ replaced by $\tilde{\nu}_{a_n}(g)(\omega) + h^{\dagger}_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n})$. In consequence, the second inequality of (12.14) of AS2 holds because $h^{\dagger}_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n}) \geq 0$ (since $m_F(\theta, X_i, z_0) \geq 0$ a.s. by (2.15) of ASN and Assumption PS1(c) and (d) of ASN). The remainder of the proof of (12.7) of AS2 goes through without any changes. $\square$

Proof of Theorem AN1. We adjust the proof of Theorem 2(a) in AS1 to prove part (a) of Theorem AN1. The proof of Theorem 2(a) of AS1 is given by the combination of Lemmas A2-A5 stated in Appendix A of AS2. Hence, we need to establish analogues of these Lemmas that hold in the context of this paper.

In the analogue of Lemma A2, the quantity $c_0(h_{n,F}(\theta), 1-\alpha)$ is replaced by $c_0(h^{\dagger}_{n,F}(\theta, b), 1-\alpha)$ because the latter is the $1-\alpha$ quantile of the distribution of $T(h^{\dagger}_{n,F}(\theta, b))$, which depends on $h^{\dagger}_{n,F}(\theta, b)$, not $h_{n,F}(\theta)$. Given this change, the proof of Lemma A2 of AS2 goes through making use of Theorem AN4 in place of Theorem 1 of AS1. Note that the quantity $x_{h_{n,F}(\theta)}$ that appears in Theorem 1 of AS1 and in the proof of Lemma A2 of AS2 is changed to $x_{h^\dagger_{n,F}(\theta,b)}$ in Theorem AN4 because we take $x_{h^\dagger_{n,F}(\theta,b)} = c_0(h^\dagger_{n,F}(\theta,b), 1-\alpha)+\delta$ in the proof of the analogue of Lemma A2.

In the statement of the analogue of Lemma A3 of AS2, $c(h_{1,n,F}(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$ is replaced by $c(h^\dagger_{1,n,F}(\theta,b), \widehat{h}_{2,n}(\theta), 1-\alpha)$. To prove the analogue of Lemma A3 of AS2, we use the property of the sequence $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\}$ constructed there (that $d(h_{2,a_n,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$) and Lemma AN2(c) to show that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_{2,0})$. In the rest of the proof, we make the following changes: $h_{1,a_n,F_{a_n}}(\theta_{a_n})$ is replaced by $h^\dagger_{1,a_n,F_{a_n}}(\theta_{a_n}, b_{a_n})$ in (12.16) and (12.17), but not in (12.22), and $h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g, b_{a_n})$ is replaced by $h^\dagger_{1,a_n,F_{a_n},j}(\theta_{a_n}, g, b_{a_n})$ in the first three lines of (12.23), and in the second appearance of $h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g, b_{a_n})$ in the fourth, fifth, and seventh lines of (12.23). In addition, the empirical process and other finite-sample quantities depend on b_{a_n} in the proof. The second equality of (12.23) holds because $h^\dagger_{1,a_n,F_{a_n},j}(\theta_{a_n}, g, b_{a_n}) \geq 0$ (because $m_F(\theta, X_i, z_0) \geq 0$ a.s. by (2.15) of ASN and Assumptions PS1(c) and (d) of ASN). The equality in (12.23) holds by the argument given plus the result of Lemma AN3, which implies that $h^\dagger_{1,a_n,F_{a_n},j}(\theta_{a_n}, g, b_{a_n}) = h_{1,a_n,F_{a_n},j}(\theta_{a_n}, g, b_{a_n}) + o(1)$ uniformly over $g \in \mathcal{G}$.

In the statement of the analogue of Lemma A4 of AS2, $h_{1,n,F}(\theta)$ is replaced by $h^\dagger_{1,n,F}(\theta, b)$ twice. In the proof of the analogue of Lemma A4 of AS2, we use Lemma AN2(c) to show that the sequence $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\}$ constructed there is in $SubSeq(h_{2,0})$ (as in the proof of the analogue of Lemma A3). The rest of the proof of the analogue of Lemma A4 goes through with the only changes being that $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g)$ is replaced by $h^\dagger_{1,a_n,F_{a_n}}(\theta_{a_n}, g, b_{a_n})$ throughout and $h_{2,F_{a_n}}(\theta_{a_n}, g)$ depends on b_{a_n}.

The proof of the analogue of Lemma A5 of AS2 goes through without any changes.

Given that the analogues of Lemmas A1-A5 of AS2 hold, the proof of Theorem AN1(a) is complete.

Next, we prove part (b) of Theorem AN1. To do so, we adjust the proof of Theorem 2(b) of AS1, which is given in Appendix C of AS2. The proof of Theorem 2(b) in AS2 goes through as is with the following two changes. First, Lemma AN4 is used in place of Lemma A1 of AS2. Second, (14.16) of AS2 is replaced by the following:

$$\begin{aligned}\kappa_n^{-1} h_{1,n,F_c}(\theta_c, g) &= \kappa_n^{-1} h^\dagger_{1,n,F_c}(\theta_c, g) + o(\kappa_n^{-1}) \\ &= (nb^{d_z})^{1/2} \kappa_n^{-1} D_F^{-1/2}(\theta, b) E_F(m_F(\theta, X_i, z_0) \odot g(X_i)) + o(\kappa_n^{-1}) \\ &\to h_{1,\infty,F_c}(\theta_c, g), \qquad (12.11)\end{aligned}$$

where the first equality holds by Lemma AN3, the second equality holds by the definition of $h^{\dagger}_{1,n,F_c}(\theta_c, g)$ in (12.2), and the convergence holds because (i) the diagonal elements of the diagonal matrix $D_F^{-1/2}(\theta, b)$ are bounded away from zero by Lemma AN2(b) and Assumption PS3(a) of ASN, (ii) $(nb^{d_z})^{1/2}\kappa_n^{-1} \to \infty$ by Assumption GMS2(c), (iii) by the definition of $h_{1,\infty,F_c}(\theta_c, g)$ (given just before Assumption GMS2 in Section 7.4), the jth element of $h_{1,\infty,F_c}(\theta_c, g)$ equals 0 if $E_F m_{F,j}(\theta, X_i, z_0) g_j(X_i) = 0$ and equals ∞ if $E_F m_{F,j}(\theta, X_i, z_0) g_j(X_i) > 0$, and (iv) $\kappa_n^{-1} \to 0$ by Assumption GMS1(b). This completes the proof of Theorem AN1(b). $\square$

Theorem N1 of ASN 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 of Theorems AN1 and AN4 (where Theorem AN4 given in Section 12.2 is used in the proof of Theorem AN1).[37] The alterations to the proof of Theorem AN4 (given above) involve changing the definition of $\widetilde{T}_{a_n,0}$ in (12.9) so that its integrand (which is just a summand in the present case because $\mathcal{G}$ is countable, Q is a measure on $\mathcal{G}$, and the integral reduces to a sum for the test statistic in (3.7) of ASN) is non-zero only for $r \leq r_{1,n}$. The definition of $\widetilde{T}_{a_n}$ needs to be changed likewise. With these changes the bounded convergence theorem argument, as in the proof of Theorem 1 of AS1 given in Appendix A of AS2, goes through.[38] Lemmas AN2-AN4, which are used in the proof of Theorem AN4, do not require any changes. The proof of Theorem AN1(a) is based on analogues of Lemmas A2-A4 in Appendix A of AS2 (as well as Theorem AN4). Again one only needs to truncate the integrals (which reduce to sums because $\mathcal{G}$ is countable) to terms with $r \leq r_{1,n}$ wherever the integrals appear in the proofs. The proof of Theorem AN1(b) is based on the proof of Theorem 2(b) given in Appendix C of AS2. In this case as well, one only needs to truncate the integrals (which reduce to sums because $\mathcal{G}$ is countable) to terms with $r \leq r_{1,n}$ wherever the integrals appear in the proofs—specifically, in (14.11), (14.14), (14.20), and (14.23). The bounded convergence theorem argument given in the proof of Theorem 2(b) to obtain (14.20) and (14.23) goes through with these changes.[39]

[37]Note that the truncated test statistic in ANS is of the form in (6.10) of Section 6.5.

[38]The alterations needed here are simpler than those in the proof of Thm. B1 in Appendix B of AS2, which considers approximate CvM tests, because ANS deals only with CvM tests based on a countable set $\mathcal{G}_{c\text{-}cube}$ and in consequence the bounded convergence argument goes through.

[39]Note that Comment 2 to Theorem B1 in Appendix B of AS2 which says "Theorem 2(b) is not given here because the proof of Theorem 2(b) does not go through with KS or A-CvM test statistics" only applies to simulated and quasi-Monte Carlo A-CvM test statistics. With truncated sums, as in ASN, the proof does go through.

12.2.5 Lemmas AN5-AN7 and Proofs of Lemmas AN2-AN4

The proof of Lemma AN2 uses the following three Lemmas.

Let $A \odot B$ denote the direct (i.e., element-by-element) product of two matrices A and B with the same dimensions.

Lemma AN5. *Suppose Assumption* NM *holds. Then, for all* $g, g^* \in \mathcal{G}$ *and* $(\theta, F) \in \mathcal{F}_+$,

$$\Sigma_F(\theta, g, g^*, z_0) = E_F \Sigma_F(\theta, X_i, z_0) \odot (g(X_i) g^*(X_i)'),$$

where $\Sigma_F(\theta, x, z)$ *and* $\Sigma_F(\theta, g, g^*, z)$ *are defined in* (2.15) *of* ASN *and* (7.6), *respectively.*

Lemma AN6. *Suppose Assumptions* B, K, *and* NM *hold. Then,*

$$\sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g\in\mathcal{G}} ||b^{-d_z/2} E_F K_b(Z_i) m(W_i, \theta, g)|| = O(b^{d_z/2}) = o(1).$$

Lemma AN7. *Suppose Assumptions* B, K, *and* NM *hold. Then,*

$$\begin{aligned}&\sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g,g^*\in\mathcal{G}} ||b^{-d_z} E_F K_b^2(Z_i) m(W_i, \theta, g) m(W_i, \theta, g^*)' \\ &- E_F \Sigma_F(\theta, X_i, z_0) \odot (g(X_i) g^*(X_i)')|| \to 0.\end{aligned}$$

Proof of Lemma AN2. Using the definitions in (7.6) and (12.1), part (a) is established as follows. We have

$$\begin{aligned}\Sigma_F(\theta, g, g^*, b) &= Cov_F(b^{-d_z/2} K_b(Z_i) m(W_i, \theta, g), b^{-d_z/2} K_b(Z_i) m(W_i, \theta, g^*)) \\ &= b^{-d_z} E_F K_b^2(Z_i) m(W_i, \theta, g) m(W_i, \theta, g^*)' \\ &\quad - b^{-d_z/2} E_F K_b(Z_i) m(W_i, \theta, g) \cdot b^{-d_z/2} E_F K_b(Z_i) m(W_i, \theta, g^*)' \\ &= E_F[\Sigma_F(\theta, X_i, z_0) \odot (g(X_i) g^*(X_i)')] + o(1) \\ &= \Sigma_F(\theta, g, g^*, z_0) + o(1), \qquad (12.12)\end{aligned}$$

where the $o(1)$ term holds uniformly over $g, g^* \in \mathcal{G}$ and $(\theta, F) \in \mathcal{F}_+$, the third equality holds by Lemmas AN6 and AN7, and the fourth equality holds by Lemma AN5.

Part (b) follows from part (a) by taking $g = g^* = 1_k$ because $D_F(\theta, b) = Diag(\Sigma_F(\theta,$

$1_k, 1_k, b))$, $D_F(\theta, z_0) = Diag(\Sigma_F(\theta, 1_k, 1_k, z_0))$, and $\sup_{(\theta,F)\in\mathcal{F}_+} \left\| D_F^{-1}(\theta, z_0) \right\| < \infty$ by Assumption PS3(a) of ASN.

Part (c) follows from parts (a) and (b) because

$$\begin{aligned} h_{2,F}(\theta, g, g^*, b) &= \left[D_F^{-1/2}(\theta, b) D_F^{1/2}(\theta, z_0)\right] \left[D_F^{-1/2}(\theta, z_0)\Sigma_F(\theta, g, g^*, b) D_F^{-1/2}(\theta, z_0)\right] \\ &\quad \times \left[D_F^{1/2}(\theta, z_0) D_F^{-1/2}(\theta, b)\right], \\ h_{2,F}(\theta, g, g^*, z_0) &= D_F^{-1/2}(\theta, z_0)\Sigma_F(\theta, g, g^*, z_0) D_F^{-1/2}(\theta, z_0), \end{aligned} \tag{12.13}$$

and $\sup_{(\theta,F)\in\mathcal{F}_+} \left\| D_F^{-1/2}(\theta, z_0) \right\| < \infty$. $\square$

Proof of Lemma AN3. For notational simplicity, suppose $m_F(\theta, x, z)$ (defined in (2.15) of ASN to equal $E_F(m(W_i, \theta)|X_i = x, Z_i = z) f(z|x))$ is a scalar. This is without loss of generality (wlog) because we could argue element by element. By a two-term Taylor expansion of $m_F(\theta, x, z_0 + bz^*)$ around $z^* = 0$, we have

$$\begin{aligned} &\sup_{(\theta,F)\in\mathcal{F}_+} \left| \int K(z^*) \left[m_F(\theta, x, z_0 + bz^*) - m_F(\theta, x, z_0)\right] dz^* \right| \\ &= \sup_{(\theta,F)\in\mathcal{F}_+} \left| b \int z^{*\prime} K(z^*) \, dz^* \frac{\partial}{\partial z} m_F(\theta, x, z_0) + \frac{b^2}{2} \int K(z^*) z^{*\prime} \frac{\partial^2}{\partial z \partial z'} m_F(\theta, x, \widetilde{z}) z^* dz^* \right| \\ &\leq b^2 \sup_{z\in[-1,1]^{d_z}} |K(z)| \cdot \sup_{(\theta,F)\in\mathcal{F}} \sup_{z\in\mathcal{Z}_0} \left\| \frac{\partial^2}{\partial z \partial z'} m_F(\theta, x, z) \right\| \cdot \left| \int_{[-1,1]^{d_z}} z^{*\prime} z^* dz^* \right| \\ &= b^2 L_m(x) C \end{aligned} \tag{12.14}$$

for some $C < \infty$, where the Taylor expansion is valid by Assumption PS3(b) of ASN, $\widetilde{z}$ is some intermediate point that is in $\mathcal{Z}_0$ for b sufficiently small, the inequality uses Assumption K(c), the last equality uses Assumptions K(d) and K(e), and $L_m(x)$ is defined in Assumption PS3(b) of ASN.

Using (12.14), we have: for all $(\theta, F) \in \mathcal{F}$ and $g \in \mathcal{G}$,

$$
\begin{aligned}
& |E_F m(W_i, \theta, g, b) - b^{d_z/2} E_F m_F(\theta, X_i, z_0) g(X_i)| \\
= & |b^{-d_z/2} E_F K_b(Z_i) m(W_i, \theta) g(X_i) - b^{d_z/2} E_F m_F(\theta, X_i, z_0) g(X_i)| \\
= & \left| \int \left(\int b^{-d_z/2} K\left(\frac{z - z_0}{b}\right) m_F(\theta, x, z) dz - b^{d_z/2} m_F(\theta, x, z_0) \right) g(x) f(x) d\mu_X(x) \right| \\
= & b^{d_z/2} \left| \int \left(\int \left[K(z^*) m_F(\theta, x, z_0 + bz^*) - K(z^*) m_F(\theta, x, z_0) \right] dz^* \right) g(x) f(x) d\mu_X(x) \right| \\
\leq & b^{d_z/2} \int b^2 L_m(x) C G f(x) d\mu_X(x) \\
\leq & b^{2+d_z/2} C G C_2, \qquad (12.15)
\end{aligned}
$$

where $CGC_2 < \infty$, the first equality holds by the definition of $m(W_i, \theta, g, b)$, the second equality uses iterated expectations with conditioning on (X_i, Z_i) and the definition of $m_F(\theta, x, z)$, the third equality holds by change of variables with $z^* = (z - z_0)/b$, the first inequality holds by (12.14) and Assumption NM(a), and the second inequality holds by Assumption PS3(b) of ASN.

By Assumption B(a), $n^{1/2} O(b^{2+d_z/2}) = o(1)$. This and (12.15) give

$$
\sup_{(\theta,F)\in\mathcal{F}} \sup_{g\in\mathcal{G}} |n^{1/2} E_F m(W_i, \theta, g, b) - (nb^{d_z})^{1/2} E_F m_F(\theta, X_i, z_0) g(X_i)| = o(1). \qquad (12.16)
$$

Equations (12.14)-(12.16) also hold with $D_F^{-1/2}(\theta, b)$ multiplying each quantity inside the absolute values (using Lemma AN2(b) and Assumption PS3(a) of ASN). Equation (12.16) (with the multiplicand $D_F^{-1/2}(\theta, b)$ added inside the absolute values) and the definitions of $h_{1,n,F}(\theta, g, b)$ and $h_{1,n,F}^{\dagger}(\theta, g, b)$ give

$$
\begin{aligned}
& \sup_{(\theta,F)\in\mathcal{F}} \sup_{g\in\mathcal{G}} |h_{1,n,F}(\theta, g, b) - h_{1,n,F}^{\dagger}(\theta, g, b)| \\
= & \sup_{(\theta,F)\in\mathcal{F}} \sup_{g\in\mathcal{G}} |n^{1/2} D_F^{-1/2}(\theta, b) E_F m(W_i, \theta, g, b) \\
& - (nb^{d_z})^{1/2} D_F^{-1/2}(\theta, b) E_F m_F(\theta, X_i, z_0) g(X_i)| \\
= & o(1). \qquad (12.17)
\end{aligned}
$$

This completes the proof of Lemma AN3. □

Proof of Lemma AN4. The proof of part (a) follows the same argument as used to prove Lemma A1(a) of AS2 using Lemmas E1-E3 in Appendix E of AS2. Lemmas E1 and E2 hold without change.

The results of Lemma E3 of AS2 hold for $SubSeq(h_2)$ as defined here with $h_2 \in \mathcal{H}_{2,+}$ and with $m(W_{n,i}(\omega), \theta_n, g)$ and $D_{F_n}^{-1/2}(\theta_n)$ replaced by $m(W_{n,i}(\omega), \theta_n, g, b)$ and $D_{F_n}^{-1/2}(\theta_n, b)$, respectively, in (16.4) of AS2. Lemma E3 of AS2 is proved by verifying conditions (i)-(v) of Theorem 10.6 of Pollard (1990). The proof in the present context requires some adjustments.

In the verification of (i), $m(W_{n,i}(\omega), \theta_n, g)$ and $\sigma_{F_n,j}(\theta_n)$ are replaced by $m(W_{n,i}(\omega), \theta_n, g, b)$ and the (j, j) element of $D_{F_n}^{1/2}(\theta_n, b)$ in (16.35)-(16.36) of AS2.

In the verification of (ii), $D_{F_n}(\theta_n)$ and $\Sigma_{F_n}(\theta_n, g, g^*)$ are replaced by $D_{F_n}(\theta_n, b)$ and $\Sigma_{F_n}(\theta_n, g, g^*, b)$ in (16.37) of AS2. Then, condition (i) of $SubSeq(h_2)$ plus Lemma AN2(c) deliver the desired convergence. Lemma AN2(c) is required in the proof in the current case, but not in AS2, because the finite-sample covariance kernel of the empirical process depends on b in the present case.

In the verification of (iii), one can ignore the $\sigma_{F_n,j}^{-1}(\theta_n)$ and $G(X_i)$ multiplicands in (16.38) of AS2 because Lemma AN2(b) and Assumption PS3(a) of ASN imply that $\sigma_{F_n,j}^{-1}(\theta_n)$ is uniformly bounded over $(\theta, F) \in \mathcal{F}_+$ and $n \geq 1$ and Assumption NM(a) implies that $G(X_i) = G < \infty$. Then, Lemma AN2(a) gives the desired result.

Condition (iv) is the Lindeberg condition. In the verification of (iv), one can ignore the $\sigma_{F_n,j}^{-1}(\theta_n)$ and $G(X_i)$ multiplicands in (16.39) of AS2 for the same reasons as above. The required condition reduces to: for all $\xi > 0$, some $\delta > 0$, and all $j \leq k$,

$$\begin{aligned} &A_n = \sum_{i=1}^{n} E_{F_n} m_{n,j}^2(W_i, \theta_n, b) 1(|m_{n,j}(W_i, \theta_n, b)| > \xi) \to 0, \text{ where} \\ &m_{n,j}(W_i, \theta, b) = n^{-1/2} b^{-d_z/2} K_b(Z_i) m_j(W_i, \theta). \end{aligned} \tag{12.18}$$

We have

$$
\begin{aligned}
A_n &\leq nE_{F_n}|m_{n,j}(W_i,\theta_n,b)|^{2+\delta}/\xi^\delta \\
&= n^{-\delta/2}b^{-\delta d_z/2}\left(b^{-d_z}E_{F_n}K_b^{2+\delta}(Z_i)|m_j(W_i,\theta_n)|^{2+\delta}/\xi^\delta\right) \\
&= (nb^{d_z})^{-\delta/2}\left(b^{-d_z}\int K^{2+\delta}\left(\frac{z-z_0}{b}\right)E_{F_n}(|m_j(W_i,\theta_n)|^{2+\delta}|Z_i=z)f_n(z)dz/\xi^\delta\right) \\
&= (nb^{d_z})^{-\delta/2}\left(\int K^{2+\delta}(z^*)E_{F_n}(|m_j(W_i,\theta_n)|^{2+\delta}|Z_i=z_0+bz^*)f_n(z_0+bz^*)dz^*/\xi^\delta\right) \\
&\leq (nb^{d_z})^{-\delta/2}\left(C_5^*\int K^{2+\delta}(z^*)dz^*/\xi^\delta\right) \\
&\to 0 \qquad (12.19)
\end{aligned}
$$

for some constant $C_5^* < \infty$, where the first inequality holds using identical distributions, the first equality holds by algebra, the second equality holds by iterated expectations, the third equality holds by change of variables with $z^* = (z-z_0)/b$, the second inequality holds for b sufficiently small that $z_0 + bz^* \in \mathcal{Z}_0$ by Assumption PS3(e) of ASN, and the convergence holds by Assumptions B(b), K(c), and K(e).

In the verification of (v), $D_{F_n}(\theta_n)$ and $m(W_i,\theta_n,g)$ are replaced by $D_{F_n}(\theta_n,b)$ and $m(W_i,\theta_n,g,b)$ in (16.40) of Section 16.6 in Appendix E of AS2 and the convergence holds by condition (i) of $SubSeq(h_2)$ plus Lemma AN2(c). This completes the changes needed in the proof of Lemma E3 of AS2.

Given that the results of Lemma E3 of AS2 hold for $SubSeq(h_2)$ as defined here, the proof of Lemma A1(a) in AS2 establishes Lemma AN4(a) with only minor changes. In particular, $D_{F_n}(\theta_n)$ is replaced by $D_{F_n}(\theta_n,b)$ in (16.8) of AS2 and the second and last equalities in (16.8) of AS2 hold by (16.40) of AS2 with the changes described in the previous paragraph. This completes the proof of part (a) of Lemma AN4.

Now, we prove part (b) of the Lemma. The multiplicand $D_F^{-1/2}(\theta,b)$, which appears in $\widehat{h}_{2,n,F}(\theta,g,g^*,b)$, equals $D_F^{-1/2}(\theta,z_0)+o(1)$ uniformly over $(\theta,F)\in\mathcal{F}$ by Lemma AN2(b) and $\sup_{(\theta,F)\in\mathcal{F}}||D_F^{-1/2}(\theta,z_0)||<\infty$ by Assumption PS3(a) of ASN. Hence, one can ignore the $D_F^{-1/2}(\theta,b)$ multiplicand when verifying part (b) of the Lemma. Doing so transforms $\widehat{h}_{2,n,F}(\theta,g,g^*,b)$ into $\widehat{\Sigma}_n(\theta,g,g^*)$.

Part of the proof of part (b) is similar to the proof of Lemma A1(b) of AS2. As in AS2, for notational simplicity, we establish results for the sequence $\{n\}$, rather than the subsequence $\{a_n : n \geq 1\}$. Two terms appear in the rhs of (16.16) of AS2. The second term can be shown to be $o_p(1)$. The argument is as follows. The second term (ignoring the $D_F^{-1/2}(\theta, b)$ multiplicand) is the following quantity multiplied by its transpose:

$$n^{-1} \sum_{i=1}^{n} m(W_i, \theta, g, b) = n^{-1} \sum_{i=1}^{n} b^{-d_z/2} K_b(Z_i) m_j(W_i, \theta_n) g(X_i). \tag{12.20}$$

This quantity has mean that is $o_p(1)$ by Lemma AN6. The difference between this quantity and its mean is $o_p(1)$ by Lemma E2 of AS2. The conditions of Lemma E2 are verified by the argument given in (16.18)-(16.22) of AS2 with (16.21), which verifies an $L^{1+\eta}$-boundedness condition, replaced by L^2-boundedness of $b^{-d_z/2} K_b(Z_i) m_j(W_i, \theta_n) g(X_i)$, which holds by Lemma AN7.

The first term appearing in (16.16) of AS2 (ignoring the $D_F^{-1/2}(\theta, b)$ multiplicand) is

$$Q_n(g, g^*) = n^{-1} \sum_{i=1}^{n} m(W_i, \theta, g, b) m(W_i, \theta, g^*, b)'. \tag{12.21}$$

To complete the proof of part (b), we need to show that the supremum over $(g, g^*) \in \mathcal{G}^2$ of $Q_n(g, g^*)$ minus its expectation is $o_p(1)$ under $\{(\theta_n, F_n) : n \geq 1\}$. This cannot be done using the uniform law of large numbers given in Lemma E2 of AS2, as is done in the proof of Lemma A1(b) in AS2, because the summands do not satisfy an $L^{1+\eta}$-boundedness condition when $m(W_i, \theta, g)$ is replaced by $m(W_i, \theta, g, b)$.

In fact, the summands of $Q_n(g, g^*)$ do not even satisfy a uniform integrability condition, as the following calculations show. For simplicity, suppose $m(W_i, \theta)$ is a scalar and is independent of Z_i. Let $m_{n,i}(b)$ and $m_{n,i}$ denote $m(W_i, \theta_n, g, b)$ and $m(W_i, \theta_n, g)$, respectively. We have: for $L < \infty$,

$$\begin{aligned} & E_{F_n} m_{n,i}^2(b) 1(m_{n,i}^2(b) > L) \\ = \; & E_{F_n} b^{-d_z} K_b^2(Z_i) m_{n,i}^2 1(b^{-d_z} K_b^2(Z_i) m_{n,i}^2 > L) \\ = \; & E_{F_n} \cdot E_{F_n}(b^{-d_z} K_b^2(Z_i) m_{n,i}^2 1(b^{-d_z} K_b^2(Z_i) m_{n,i}^2 > L) | Z_i) \end{aligned}$$

$$\begin{aligned} = & \int b^{-d_z} K^2\left(\frac{z - z_0}{b}\right) E_{F_n}\left(m_{n,i}^2 1\left(b^{-d_z} K^2\left(\frac{z - z_0}{b}\right) m_{n,i}^2 > L\right) | Z_i = z\right) f_n(z) dz \\ = & \int K^2(z^*) E_{F_n}(m_{n,i}^2 1(K^2(z^*) m_{n,i}^2 > L b^{d_z}) | Z_i = z_0 + b z^*) f_n(z_0 + b z^*) dz^*, \end{aligned} \tag{12.22}$$

where the second equality holds by iterated expectations and the fourth equality holds by change of variables with $z^* = (z - z_0)/b$. The $\limsup_{n\to\infty}$ of the rhs in (12.22) is not small for L large because $b^{d_z} \to 0$. Hence, uniform integrability fails.

Instead, we show that

$$\sup_{g,g^*\in\mathcal{G}} |Q_n(g,g^*) - E_{F_n} Q_n(g,g^*)| \to_p 0 \qquad (12.23)$$

under $\{(\theta_n, F_n) : n \geq 1\}$ by using the maximal inequality (7.10) of Pollard (1990, p. 38) for manageable processes, which is applicable by Assumption NM(b) and Lemma E1 of AS2. For notational simplicity, suppose $m(W_i, \theta, g, b)$ is a scalar. (This is wlog because we can argue element by element.) The maximal inequality says that

$$E_{F_n} \sup_{g,g^*\in\mathcal{G}} |Q_n(g,g^*) - E_{F_n} Q_n(g,g^*)| \leq n^{-1} C E_{F_n} ||F_n^*|| \leq n^{-1} C (E_{F_n} ||F_n^*||^2)^{1/2}, \quad (12.24)$$

where C is some finite constant and F_n^* (using Pollard's notation) is an n-vector of envelope functions that satisfies $F_n^* = (F_{n,1}^*, ..., F_{n,n}^*)'$, $||F_n^*||^2 = \sum_{i=1}^n F_{n,i}^{*2}$, and

$$F_{n,i}^* = b^{-d_z} K_b^2(Z_i) ||m(W_i, \theta_n)||^2 G^2 \geq \sup_{g,g^*\in\mathcal{G}} ||m(W_i, \theta_n, g, b) m(W_i, \theta_n, g^*, b)||. \quad (12.25)$$

We have

$$\begin{aligned}
& n^{-1} (E_{F_n} ||F_n^*||^2)^{1/2} \\
= \; & n^{-1/2} (E_{F_n} F_{n,1}^{*2})^{1/2} \\
= \; & n^{-1/2} G^2 (E_{F_n} b^{-2d_z} K_b^4(Z_i) ||m(W_i, \theta_n)||^4)^{1/2} \\
= \; & (nb^{d_z})^{-1/2} G^2 \left(\int b^{-d_z} K^4 \left(\frac{z - z_0}{b} \right) E_{F_n} (||m(W_i, \theta_n)||^4 | Z_i = z) f_n(z) dz \right)^{1/2} \\
= \; & (nb^{d_z})^{-1/2} G^2 \left(\int K^4(z^*) E_{F_n} (||m(W_i, \theta_n)||^4 | Z_i = z_0 + bz^*) f_n(z_0 + bz^*) dz^* \right)^{1/2} \\
\leq \; & (nb^{d_z})^{-1/2} G^2 \left(\int K^4(z^*) dz^* \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{z\in\mathcal{Z}_0} E_F (||m(W_i, \theta)||^4 | Z_i = z) f(z) \right)^{1/2} \\
\to \; & 0, \qquad (12.26)
\end{aligned}$$

where the first equality holds by identical distributions for $i = 1, ..., n$ under F_n, the second equality holds using Assumption NM(a), the third equality holds by iterated expectations, the fourth equality holds by change of variables with $z^* = (z - z_0)/b$, the inequality holds for b sufficiently small using Assumption K(c), and the convergence holds by Assumptions B(b) and K(c)-(e) and Assumption PS3(e) of ASN. This completes the proof of part (b) of the Lemma. □

12.2.6 Proofs of Lemmas AN5-AN7

Proof of Lemma AN5. Using Assumptions PS2(a)-(d) of ASN (which hold for $(\theta, F) \in \mathcal{F}_+$), we have

$$\begin{aligned}\Sigma_F(\theta, g, g^*, z) &= E_F(m(W_i, \theta, g)m(W_i, \theta, g^*)'|Z_i = z)f(z) \\ &= \int\int m(y, x, z, \theta, g)m(y, x, z, \theta, g^*)'f(y, x|z)d\mu_Y(y)d\mu_X(x)f(z) \\ &= \int\int m(y, x, z, \theta, g)m(y, x, z, \theta, g^*)'f(y, x, z)d\mu_Y(y)d\mu_X(x). \quad (12.27)\end{aligned}$$

In addition, we have

$$\begin{aligned}&E_F[\Sigma_F(\theta, X_i, z) \odot (g(X_i)g^*(X_i)')] \\ &= \int [\Sigma_F(\theta, x, z) \odot (g(x)g^*(x)')]f(x)d\mu_X(x) \\ &= \int \left[\int m(y, x, z, \theta)m(y, x, z, \theta)'f(y|x, z)d\mu_Y(y)f(z|x) \odot (g(x)g^*(x)')\right] f(x)d\mu_X(x) \\ &= \int\int m(y, x, z, \theta, g)m(y, x, z, \theta, g^*)'f(y, x, z)d\mu_Y(y)d\mu_X(x), \quad (12.28)\end{aligned}$$

where the last equality uses $m(w, \theta, g) = m(w, \theta) \odot g(x)$ for $w = (y, x, z)'$. □

Proof of Lemma AN6. Define

$$m_F(\theta, g, z) = E_F(m(W_i, \theta, g)|Z_i = z)f(z). \quad (12.29)$$

We have

$$
\begin{aligned}
& \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g\in\mathcal{G}} ||b^{-d_z/2} E_F K_b(Z_i) m(W_i,\theta,g)|| \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g\in\mathcal{G}} ||b^{-d_z/2} \int K_b(z) m_F(\theta,g,z)dz|| \\
\leq & \ b^{-d_z/2} \int K\left(\frac{z-z_0}{b}\right) \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g\in\mathcal{G}} ||m_F(\theta,g,z)||dz \\
= & \ b^{d_z/2} \int K(z^*) \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{g\in\mathcal{G}} ||m_F(\theta,g,z_0+bz^*)||dz^* \\
\leq & \ b^{d_z/2} \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{z\in\mathcal{Z}_0} \sup_{g\in\mathcal{G}} ||m_F(\theta,g,z)|| \\
\to & \ 0,
\end{aligned} \tag{12.30}
$$

where the first equality holds by iterated expectations conditioning on Z_i using Assumption PS2(a) of ASN, the second equality holds by change of variables with $z^* = (z-z_0)/b$, the second inequality holds using Assumption K(a), and the convergence holds by Assumption B(a) and the result:

$$
\sup_{(\theta,F)\in\mathcal{F}_+, z\in\mathcal{Z}_0, g\in\mathcal{G}} ||m_F(\theta,g,z)|| < \infty. \tag{12.31}
$$

Equation (12.31) is established as follows. We have

$$
\begin{aligned}
m_F(\theta,g,z) &= E_F[E_F(m(W_i,\theta,g)|X_i,Z_i=z)]f(z) \\
&= \int E_F(m(W_i,\theta)|X_i=x,Z_i=z)g(x)f(x|z)d\mu_X(x)f(z) \\
&= \int m_F(\theta,x,z)g(x)f(x,z)d\mu_X(x),
\end{aligned} \tag{12.32}
$$

where the second equality uses Assumption PS2(e) of ASN. Hence, we obtain

$$
\begin{aligned}
& \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{z\in\mathcal{Z}_0} \sup_{g\in\mathcal{G}} ||m_F(\theta,g,z)|| \\
\leq & \ G \sup_{(\theta,F)\in\mathcal{F}_+} \sup_{z\in\mathcal{Z}_0} \int ||m_F(\theta,x,z)||f(x,z)d\mu_X(x) < \infty,
\end{aligned} \tag{12.33}
$$

where the first inequality holds by Assumption NM(a) and the second inequality holds by Assumption PS3(c). $\square$

Proof of Lemma AN7. For notational simplicity, we suppose $m(W_i,\theta,g)$ is a scalar. (This is wlog because we could argue element by element.) For all $g,g^*\in\mathcal{G}$, we have

$$\begin{aligned}
& J_b(g,g^*) \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} |b^{-d_z}E_FK_b^2(Z_i)m(W_i,\theta,g)m(W_i,\theta,g^*)-E_F[\Sigma_F(\theta,X_i,z_0)\odot(g(X_i)g^*(X_i))]| \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} |b^{-d_z}E_FK_b^2(Z_i)m^2(W_i,\theta)g(X_i)g^*(X_i)-E_F\Sigma_F(\theta,X_i,z_0)g(X_i)g^*(X_i)| \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} \left|\int\left(\int b^{-d_z}K^2\left(\frac{z-z_0}{b}\right)\Sigma_F(\theta,x,z)dz-\Sigma_F(\theta,x,z_0)\right)g(x)g^*(x)f(x)d\mu_X(x)\right| \\
= & \sup_{(\theta,F)\in\mathcal{F}_+} \left|\int\left(\int\left[K^2(z^*)\Sigma_F(\theta,x,z_0+bz^*)-K^2(z^*)\Sigma_F(\theta,x,z_0)\right]dz^*\right)\right. \\
& \left.\times g(x)g^*(x)f(x)d\mu_X(x)\right|, \qquad (12.34)
\end{aligned}$$

where the first equality defines $J_b(g,g^*)$, the second equality holds by the definition of $m(W_i,\theta,g)$, the third equality uses iterated expectations with conditioning on (X_i,Z_i) and Assumptions PS2(b) and (c) of ASN, and the fourth equality holds by change of variables with $z^*=(z-z_0)/b$.

Using (12.34), we have

$$\begin{aligned}
\sup_{g,g^*\in\mathcal{G}} J_b(g,g^*) &\leq G\sup_{(\theta,F)\in\mathcal{F}_+}\int\left(\int K^2(z^*)L_\Sigma(x)b||z^*||dz^*\right)f(x)d\mu_X(x) \\
&\leq bGC\sup_{(\theta,F)\in\mathcal{F}_+}\int L_\Sigma(x)f(x)d\mu_X(x) \\
&\to 0, \qquad (12.35)
\end{aligned}$$

where the first inequality holds by Assumption PS3(d) of ASN and Assumption NM(a), the second inequality holds for some $C<\infty$ by Assumptions K(c) and K(e), and the convergence holds by Assumptions B(a) and PS3(d) of ASN. □

12.3 Proofs of Theorems AN2-AN4

Proof of Theorem AN2. Theorem AN2 is analogous to Theorem 3 of AS1. The proof of Theorem 3 of AS1 that is given in Section 14.2 in Appendix C of AS2 goes through with a few changes in the present context. First, $E_{F_0}(\cdot)$ is replaced by $E_{F_0}(\cdot|Z_i=z_0)$ in $m^*(g)$ and elsewhere. Second, $n^{1/2}\beta(g_0)$ is replaced throughout by $(nb^{d_z})^{1/2}\beta(g_0)$. Third, Assumption NFA(a) is used in place of Assumption FA(a) to obtain the inequality in (14.28) of AS2. Fourth, the proof uses Lemma AN4, which employs Assumptions NFA(b) and NFA(c), in place of Lemma A1 of AS2.

Fifth, the second equality of (14.33) of AS2 does not hold. It relies on $n^{-1/2}h_{1,n,F_0}(\theta_*, g) = m^*(g)$, which in the present context is replaced by $(nb^{d_z})^{-1/2}h_{1,n,F_0}(\theta_*, g, b) = m^*(g)$, which does not hold. However, we have

$$\begin{aligned}(nb^{d_z})^{-1/2}h_{1,n,F_0}(\theta_*, g, b) &= D_{F_0}^{-1/2}(\theta_*, b)b^{-d_z/2}E_{F_0}m(W_i, \theta_*, g, b)\\ &= D_{F_0}^{-1/2}(\theta_*, z_0)E_{F_0}m(\theta_*, X_i, z_0)g(X_i) + O(b^2)\\ &= D_{F_0}^{-1/2}(\theta_*, z_0)E_{F_0}(m(W_i, \theta_*, g)|Z_i = z_0)f(z_0) + O(b^2)\\ &= m^*(g) + o(1), \qquad (12.36)\end{aligned}$$

where the second equality holds by Lemma AN2(b) and (12.15) (which holds for $(\theta_*, F_0) \in \mathcal{F}_+$), the third equality holds by the same argument as in the proof of Lemma AN5 with $m(y, x, z, \theta, g)m(y, x, z, \theta, g^*)'$ replaced by $m(y, x, z, \theta, g)$ throughout, and the fourth equality holds by the definition of $m^*(g)$ and Assumption B(a).

Using (12.36), the second equality of (14.33) of AS2 holds with $m^*(g)/\beta(g_0)$ replaced by $m^*(g)/\beta(g_0) + o(1)$.

These are the only changes needed to the proof of Theorem 3 of AS1. □

Proof of Theorem AN3. Theorem AN3 is analogous to Theorem 4 of AS1. First, we give an analogue of (14.37) in the proof of Theorem 4 of AS1 given in Section 14.3 of Appendix C in AS2. We have

$$\begin{aligned}&h_{1,n,F_n}(\theta_{n,*}, g, b)\\ &= n^{1/2}D_{F_n}^{-1/2}(\theta_{n,*}, b)E_{F_n}m(W_i, \theta_{n,*}, g, b)\\ &= (nb^{d_z})^{1/2}(I_k + o(1))D_{F_n}^{-1/2}(\theta_{n,*}, z_0)E_{F_n}m(\theta_{n,*}, X_i, z_0)g(X_i) + o(1) \qquad (12.37)\\ &= (nb^{d_z})^{1/2}(I_k + o(1))D_{F_n}^{-1/2}(\theta_{n,*}, z_0)E_{F_n}(m(W_i, \theta_{n,*}, g)|Z_i = z_0)f_n(z_0) + o(1),\end{aligned}$$

where the first equality holds by (12.2), the second equality holds by Lemma AN2(b) and (12.15) because $n^{1/2}b^{2+d_z/2} \to 0$ if $b = o(n^{-1/(4+d_z)})$, and the third equality holds by the same argument as in the proof of Lemma AN5 above.

Next, by element-by-element mean-value expansions about θ_n, we have

$$
\begin{aligned}
& D_{F_n}^{-1/2}(\theta_{n,*}, z_0) E_{F_n}(m(W_i, \theta_{n,*}, g)|Z_i = z_0) f_n(z_0) \\
= \ & D_{F_n}^{-1/2}(\theta_n, z_0) E_{F_n}(m(W_i, \theta_n, g)|Z_i = z_0) f_n(z_0) \\
& + \Pi_{F_n}(\theta_{n,g}, g)(\theta_{n,*} - \theta_n),
\end{aligned} \tag{12.38}
$$

using Assumption NLA2, where $\theta_{n,g}$ may differ across rows of $\Pi_{F_n}(\theta_{n,g}, g)$, $\theta_{n,g}$ lies between $\theta_{n,*}$ and θ_n, and $\theta_{n,g} \to \theta_0$.

Combining (12.37) and (12.38) gives the analogue of (14.37) of AS2:

$$
\begin{aligned}
& h_{1,n,F_n}(\theta_{n,*}, g, b) \\
= \ & (nb^{d_z})^{1/2}(I_k + o(1)) D_{F_n}^{-1/2}(\theta_n, z_0) E_{F_n}(m(W_i, \theta_n, g)|Z_i = z_0) f_n(z_0) \\
& + (I_k + o(1)) \Pi_{F_n}(\theta_{n,g}, g)(nb^{d_z})^{1/2}(\theta_{n,*} - \theta_n) \\
\to \ & h_1(g) + \Pi_0(g)\lambda,
\end{aligned} \tag{12.39}
$$

where $h_1(g)$ and $\Pi_0(g)$ are defined in (10.2) and the convergence uses Assumptions NLA1(a), NLA1(b), and NLA2.

Now, the proof of Theorem AN3 is similar to the proof of Theorem 4 of AS1 given in AS2 with the following changes:

(i) $\{(\theta_{n,*}, F_n) \in \mathcal{F} : n \geq 1\} \in SubSeq(h_2)$, where $h_2 = h_{2,F_0}(\theta_0) \in \mathcal{H}_{2,+}$ by Assumptions NLA1(a) and NLA1(c)-(e),

(ii) part (i) and Assumptions B and MN imply that the results of Lemma AN4 hold under $\{(\theta_{n,*}, F_n) \in \mathcal{F} : n \geq 1\}$ and these results are used in place of Lemma A1 of AS2,

(iii) equation (14.38) of AS2 is replaced by

$$
\begin{aligned}
& \kappa_n^{-1} \overline{D}_{F_n}^{-1/2}(\theta_{n,*}, g, b) D_{F_n}^{1/2}(\theta_{n,*}, b) h_{1,n,F_n}(\theta_{n,*}, g, b) \\
= \ & (I_k + o(1)) \kappa_n^{-1} (nb^{d_z})^{1/2} \overline{D}_{F_n}^{-1/2}(\theta_n, g, z_0) E_{F_n}(m(W_i, \theta_n, g)|Z_i = z_0) f_n(z_0) \\
& + \kappa_n^{-1} \overline{D}_{F_0}^{-1/2}(\theta_0, g, z_0) D_{F_0}^{1/2}(\theta_0, z_0)(I_k + o(1)) \Pi_{F_n}(\theta_{n,g}, g)(nb^{d_z})^{1/2}(\theta_{n,*} - \theta_n)] \\
= \ & \pi_1(g) + o(1),
\end{aligned} \tag{12.40}
$$

where the first equality holds by the equality in (12.39) and Lemma AN2(b) and the second equality holds because (a) the first term on the rhs of the first equality is $\pi_1(g) + o(1)$ by Assumption NLA4 and (b) the second term on the rhs of the first equality is $o(1)$ by the convergence of the second term in (12.39) plus $\kappa_n^{-1} \to 0$, and

(iv) in the verification of (14.23) in part (ix) of the proof of Theorem 4 of AS1 given in Section 14.3 of Appendix C in AS2, (12.39) is used in place of (14.37) of AS2. This completes the proof. □

Proof of Theorem AN4. First we establish part (a) of the Theorem for the KS statistic. For the KS statistic defined in (6.7), we have

$$\begin{aligned} T_n(\theta_{n,*}) &\geq S(n^{1/2}\overline{m}_n(\theta_{n,*}, g_n), \overline{\Sigma}_n(\theta_{n,*}, g_n)) \\ &= S\left(\nu_{n,F_n}(\theta_{n,*}, g_n, b) + h_{1,n,F_n}(\theta_{n,*}, g_n, b), \overline{h}_{2,n,F_n}(\theta_{n,*}, g_n, b)\right) \\ &\geq S(\nu_{n,F_n}(\theta_{n,*}, g_n, b) + h_{1,n,F_n}(\theta_{n,*}, g_n, b), C \cdot I_k) \text{ w.p.a.1.} \end{aligned} \tag{12.41}$$

for some constant C sufficiently large, where g_n is as in Assumption NLA7(c) and "w.p.a.1." abbreviates "with probability that approaches one as $n \to \infty$." The second inequality holds w.p.a.1 using Assumption S1(e) because $\|\overline{h}_{2,n,F_n}(\theta_{n,*}, g_n, b) - (h_{2,F_n}(\theta_{n,*}, g_n, g_n) + \varepsilon I_k)\| \to_p 0$ (by (12.2) and Lemma AN2(c)) and $C \cdot I_k - (h_{2,F_n}(\theta_{n,*}, g_n, g_n) + \varepsilon I_k)$ is positive definite w.p.a.1 (as required in Assumption S1(e)) because the largest eigenvalue of $h_{2,F_n}(\theta_{n,*}, g_n, g_n)$ is bounded (because it is a correlation matrix divided by a diagonal matrix with diagonal elements that are bounded away from zero).

By Lemmas AN2(c) and AN4(a) and Assumption NLA7(b),

$$\nu_{n,F_n}(\theta_{n,*}, g_n, b) = O_p(1). \tag{12.42}$$

Also observe that

$$\begin{aligned} & h_{1,n,F_n}(\theta_{n,*}, g_n, b) \\ = \; & n^{1/2}(I_k + o(1))D_{F_n}^{-1/2}(\theta_{n,*}, z_0)E_{F_n}m(W_i, \theta_{n,*}, g_n, b) \\ = \; & (I_k + o(1))D_{F_n}^{-1/2}(\theta_{n,*}, z_0)(nb^{d_z})^{1/2}E_{F_n}m_{F_n}(\theta_{n,*}, X_i, z_0)g_n(X_i) + o(1) \\ \to \; & h_1, \end{aligned} \tag{12.43}$$

where the first equality holds by Lemma AN2(b), the second equality holds by (12.16) (with $\mathcal{F}$ in (12.16) and (12.15) replaced by $\mathcal{F}_+$, which does not invalidate either), and the convergence holds by Assumption NLA7(c).

Equations (12.41)-(12.43) along with Assumptions S3 and S4 imply part (a) for the KS statistic.

Now we establish part (a) of the Theorem for the CvM statistic. For the CvM statistic defined in (6.6), by similar arguments as those for (12.41), we have

$$T_n(\theta_{n,*}) \geq \int_{G_n} S(\nu_{n,F_n}(\theta_{n,*}, g, b) + h_{1,n,F_n}(\theta_{n,*}, g, b), C \cdot I_k) dQ(g). \tag{12.44}$$

By Lemmas AN2(c) and AN4(a) and Assumption NLA7(b), $\sup_{g \in G_n} \|\nu_{n,F_n}(\theta_{n,*}, g, b)\| = O_p(1)$. Then, for any $\delta > 0$, there exists $C_\delta \in (0, \infty)$ large enough such that

$$\liminf_{n\to\infty} \Pr{}_{F_n} \left(\sup_{g \in G_n} \|\nu_{n,F_n}(\theta_{n,*}, g, b)\| \leq C_\delta \right) > 1 - \delta. \tag{12.45}$$

Thus, for any $\delta > 0$,

$$\begin{aligned} & 1 - \delta \\ & \leq \liminf_{n\to\infty} \Pr{}_{F_n} \left(T_n(\theta_{n,*}) \geq Q(G_n) \inf_{\|\nu\| \leq C_\delta} \inf_{g \in G_n} S\left(\nu + h_{1,n,F_n}(\theta_{n,*}, g, b), C \cdot I_k\right) \right) \\ & = \liminf_{n\to\infty} \Pr{}_{F_n} \left(T_n(\theta_{n,*}) \geq Q(G_n) S(\nu_n + h_{1,n,F_n}(\theta_{n,*}, g_n, b), C \cdot I_k) + o(1) \right), \end{aligned} \tag{12.46}$$

where the equality holds for some $\nu_n \in [-C_\delta, C_\delta]^k$ and $g_n \in G_n$ that approximately achieves the infima. Next, we have

$$\begin{aligned} & Q(G_n) S(\nu_n + h_{1,n,F_n}(\theta_{n,*}, g_n, b), C \cdot I_k) \\ = \; & Q(G_n)(nb^{d_z})^{\chi/2} \beta^{\chi}_{F_n}(\theta_{n,*}, g_n) S\left(\frac{\nu_n + h_{1,n,F_n}(\theta_{n,*}, g_n, b)}{(nb^{d_z})^{1/2} \beta_{F_n}(\theta_{n,*}, g_n)}, C \cdot I_k \right) \\ \to \; & \infty, \end{aligned} \tag{12.47}$$

where the equality holds by Assumption S4 and the convergence holds because $Q(G_n) \times (nb^{d_z})^{\chi/2} \beta^{\chi}_{F_n}(\theta_{n,*}, g_n) \to \infty$ (by Assumption NLA7$'$(c)) and the quantity $S\left(\frac{\nu_n + h_{1,n,F_n}(\theta_{n,*}, g_n, b)}{(nb^{d_z})^{1/2} \beta_{F_n}(\theta_{n,*}, g_n)}, C \cdot I_k \right)$ is bounded away from zero. The latter holds using Assumption S3 and the fact that at least one element of the vector $\frac{\nu_n + h_{1,n,F_n}(\theta_{n,*}, g_n, b)}{(nb^{d_z})^{1/2} \beta_{F_n}(\theta_{n,*}, g_n)}$ is bounded away from the corresponding elements of the vectors in the set $[0, \infty]^p \times 0^v$ (which holds by $\nu_n / (nb^{d_z})^{1/2} \beta_{F_n}(\theta_{n,*}, g_n) \to 0$, the first three lines of (12.43), and the definition of $\beta_F(\theta, g)$).

Combining (12.46) and (12.47), for any $\delta > 0$ and any $B < \infty$, we have $\liminf_{n\to\infty} \Pr_{F_n}(T_n(\theta_{n,*}) > B) \geq 1 - \delta$. This completes the proof of part (a) for the CvM statistic.

Part (b) is implied by part (a) and $c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}), 1 - \alpha) = O_p(1)$ (which holds by an argument that is analogous to that used to prove (14.34) of AS2).

Part (c) is implied by part (b) and $c(\varphi_n(\theta_{n,*}), \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)) \leq c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)$. $\square$

13 Additional Simulation Results

In this section, we provide some additional simulation results. Tables A1 and A2 report the robustness results for the CvM/Max and KS/Max test statistics in the kinked and the peaked bound cases, respectively, for the quantile selection model. As in Tables I-III, the results in Tables A1 and A2 are for the lower endpoints of the identified intervals. Tables A3 and A4 report the robustness results for the CvM and KS test statistics in the kinked and tilted bound cases, respectively, for the conditional treatment effect model.

Both Tables A1 and A2 show that there is little sensitivity to r_1, ε, the GMS tuning parameters, and the kernel bandwidth in terms of coverage probabilities. There is some sensitivity in terms of the FCP's. The FCP decreases (gets better) with the sample size for the KS/MAX-GMS/Asy pair and is stable for the CvM/Max-GMS/Asy pair. The FCP is smaller (better) with (κ_n, B_n) halved and bigger with (κ_n, B_n) doubled.

There is quite a bit sensitivity to the kernel bandwidth. With both the kinked and the peaked bound, doubling the bandwidth reduces the FCP's for tests with the KS/Max statistics. The same is true with the kinked bound and the CvM/Max statistic. However, with the peaked bound, both doubling and halving the bandwidth increases the FCP's.

Tables A1 and A2 show that 0.50 CI's cover the true value with probability noticeably higher than 0.50. This indicates that the lower boundary point of the 0.50 CI as an estimator for the lower end point of the identified set is not median unbiased, but does not have an inward bias which has been a concern in the literature.

Table A1. Nonparametric Quantile Selection Model with Kinked Bound: Variations on the Base Case

		(a) Coverage Probabilities		(b) False Cov Probs (CPcor)	
	Statistic:	CvM/Max	KS/Max	CvM/Max	KS/Max
Case	Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Base Case: ($n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7}$)		.989	.987	.49	.57
$n = 100$		.988	.991	.48	.59
$n = 500$		.989	.991	.45	.54
$r_1 = 2$		.988	.987	.50	.53
$r_1 = 4$		.990	.989	.48	.60
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.991	.987	.49	.55
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.993	.991	.56	.61
$\varepsilon = 1/100$		.989	.987	.47	.57
$b = 0.5b^0 n^{-2/7}$		.986	.987	.69	.77
$b = 2b^0 n^{-2/7}$		.997	.995	.35	.45
$\alpha = .5$		.771	.739	.05	.06
$\alpha = .5$ & $n = 500$		.787	.753	.05	.06

Table A2. Nonparametric Quantile Selection Model with Peaked Bound: Variations on the Base Case

	Statistic:	(a) Coverage Probabilities CvM/Max	KS/Max	(b) False Cov Probs (CPcor) CvM/Max	KS/Max
Case	Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Base Case: $(n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7})$		.991	.991	.49	.53
$n = 100$		.989	.990	.56	.65
$n = 500$		.994	.995	.50	.45
$r_1 = 2$		.990	.990	.51	.50
$r_1 = 4$		.992	.991	.48	.58
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.992	.990	.47	.52
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.994	.994	.54	.56
$\varepsilon = 1/100$		.991	.991	.47	.53
$b = 0.5b^0 n^{-2/7}$		.988	.989	.62	.70
$b = 2b^0 n^{-2/7}$		.997	.996	.53	.47
$\alpha = .5$		.803	.761	.04	.05
$\alpha = .5$ & $n = 500$		.836	.795	.04	.04

Tables A3 and A4 show the sensitivity results for the nonparametric conditional treatment effect model with kinked bound and tilted bound, respectively.

Table A3 shows that, with the kinked bound, the test has NRP's smaller than 0.05 for all the test configurations and sample sizes that we experimented with. This is expected because with the kinked bound, the conditional moment inequality is only binding at a measure-zero set of the instrumental variable and Assumption GMS2 is not likely to hold. The ARP's are relatively stable as we vary r_1, decrease ε or decrease (κ_n, B_n). Doubling (κ_n, B_n) makes the ARP's smaller (worse). Both doubling and halving the kernel bandwidth reduces ARP's noticeably.

Table A3. Nonparametric Conditional Treatment Effect Model with Kinked Bound: Variations on the Base Case

Case	Statistic: / Crit Val:	(a) Null Rejection Probabilities CvM GMS/Asy	KS GMS/Asy	(b) Rej Probs under H_1 (NRP-corrected) CvM GMS/Asy	KS GMS/Asy
Base Case: $(n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7})$		.000	.000	.52	.49
$n = 100$		.000	.000	.65	.55
$n = 500$		.000	.000	.33	.40
$r_1 = 2$		.000	.000	.52	.53
$r_1 = 4$		.000	.000	.51	.45
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.000	.000	.52	.52
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.000	.000	.44	.42
$\varepsilon = 1/100$		.000	.000	.52	.44
$b = 0.5b^0 n^{-2/7}$		.000	.000	.38	.30
$b = 2b^0 n^{-2/7}$		.000	.000	.34	.43

Table A4 shows a new aspect of the sensitivity analysis. The NRP for the CvM test in the base case is somewhat bigger than 0.05. Halving the bandwidth reduces NRP's to below 0.05. while doubling the bandwidth increases the NRP's to disastrous level. This is expected because with the tilted bound the unconditional moment formed using the kernel functions has negative expectation for any fixed bandwidth. The negative expectation converges to zero as the bandwidth converges to zero. Thus, letting b converge to zero is central to the theoretical validity of our method. Using a large b deviates from the asymptotic theory.

The ARP's in Table A4 are reasonably stable across different configurations and sample sizes, except that they are somewhat sensitive to the kernel bandwidth.

Table A4. Nonparametric Conditional Treatment Effect Model with Tilted Bound: Variations on the Base Case

		(a) Null Rejection Probabilities		(b) Rej Probs under H_1 (NRP-corrected)	
	Statistic:	CvM	KS	CvM	KS
Case	Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Base Case: ($n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7}$)		.072	.047	.53	.36
$n = 100$		.085	.042	.49	.34
$n = 500$		.072	.050	.53	.40
$r_1 = 2$		.074	.059	.52	.38
$r_1 = 4$		.069	.036	.53	.32
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.081	.054	.50	.35
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.066	.045	.53	.36
$\varepsilon = 1/100$		.071	.040	.52	.31
$b = 0.5 b^0 n^{-2/7}$		.044	.023	.29	.14
$b = 2 b^0 n^{-2/7}$		.467	.313	.69	.57

References

Andrews, D. W. K., Shi, X., 2013a. Nonparametric inference based on conditional moment inequalities. Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K., Shi, X., 2013b. Inference based on conditional moment inequalities. Econometrica 81, forthcoming.

Andrews, D. W. K., Shi, X., 2013c. Supplemental Material to "Inference based on conditional moment inequalities" Econometrica 81, forthcoming, available on the Econometric Society website.