A Simple Uniformly Valid Test for Inequalities*

Gregory Cox† Xiaoxia Shi‡

July 16, 2019

Abstract

We propose a new test for inequalities that is simple and uniformly valid. The test compares the likelihood ratio statistic to a chi-squared critical value, where the degrees of freedom is the rank of the active inequalities. This test requires no tuning parameters or simulations, and therefore is computationally fast, even with many inequalities. Further, it does not require an estimate of the number of binding or close-to-binding inequalities. To show that this test is uniformly valid, we establish a new bound on the probability of translations of cones under the multivariate normal distribution that may be of independent interest. The leading application of our test is inference in moment inequality models. We also consider testing affine inequalities in the multivariate normal model and testing nonlinear inequalities in general asymptotically normal models.

Keywords: Inequality Testing, Likelihood Ratio, Moment Inequalities, Partial Identification, Uniform Inference.

*We acknowledge helpful feedback from Donald Andrews, Xiaohong Chen, Matthew Shum, Jörg Stoye, the participants of the 2nd Econometrics Jamboree at the UC Berkeley, and econometrics seminars at Columbia University, the UCSD, the UCLA, and the University of Wisconsin at Madison.

†Department of Economics, National University of Singapore (ecsgfc@nus.edu.sg)

‡Department of Economics, University of Wisconsin-Madison (xshi@ssc.wisc.edu)

1 Introduction

This paper considers testing inequalities in three settings of interest. The first setting is testing affine inequalities on the mean of a multivariate normal random vector. The second setting is inference on parameters defined by moment inequalities. The third setting is testing nonlinear inequalities in general asymptotically normal models. In all settings, a new test is proposed—the conditional chi-squared test—that is simple and uniformly valid.

The first setting, the multivariate normal model, is both classical, going back at least to Chernoff (1954), and current, attracting renewed interest because it is the limiting experiment for inference on a parameter defined by moment inequalities.[1] We propose a new test that compares the likelihood ratio statistic, T, to a chi-squared critical value with the degrees of freedom being the rank of the active inequalities, where an *active* inequality is one that holds with equality at the restricted estimator in sample. Active inequalities are the sample counterpart of binding inequalities.[2] We call this test a conditional chi-squared (CC) test because the critical value is calculated conditional on the set of active inequalities.

We show that the CC test is valid. When all the inequalities are binding, this follows from the geometry of the problem and a characterization of the distribution of T as a mixture of chi-squared distributions, as derived by Kudo (1963). In this case, the distribution of T conditional on the set of active inequalities is chi-squared with degrees of freedom equal to the rank of the active inequalities. The CC test controls size by mimicking this characteristic of the distribution of T. When some inequalities are slack (that is, not binding), the distribution of T is more complicated. However, we can still control size by bounding the conditional rejection probabilities using a new bound on the probability of translations of cones under the multivariate normal distribution. In addition, the CC test is not conservative when at least one equality is tested.

The classical test in this setting, as proposed by many papers including Kudo (1963) and Wolak (1987), compares T to the $1-\alpha$ quantile of the least favorable distribution. The least favorable distribution is the mixture of chi-squared distributions that occurs when all the inequalities bind. Relative to the classical test, the CC

[1] For an overview of the classical literature on testing inequalities, see Silvapulle and Sen (2004).

[2] In the moment inequality literature, an inequality is *binding* if it holds with equality at the population value.

test redirects power toward parameter values that violate few inequalities and away from parameter values that violate many (or all) inequalities.[3] This power redirection allows the CC test to reduce, in the presence of very slack inequalities, to the version of the CC test that depends only on the not-very-slack inequalities. We call this property Irrelevance of Distant Inequalities (IDI). Consequently, the CC test is unaffected by adding very slack inequalities. In contrast, adding a very slack inequality increases the critical value and thereby reduces the power of the classical test.

The second setting, moment inequality models, represent a large class of partially identified models that have seen many recent applications.[4] In this setting, we construct a confidence set for the true value of the parameter by inverting a test for the moment inequalities evaluated at given parameter values. The test that we propose is a version of the CC test that compares a (quasi) likelihood ratio (QLR) statistic to a conditional chi-squared critical value. In this case, the restricted estimator is the projection of the estimated moments onto the inequalities. The test is asymptotically uniformly valid because the experiment converges to a multivariate normal limiting experiment with affine inequalities—the first setting. Asymptotically, the inequalities that are relevant are binding or close-to-binding.[5] In this case, the IDI property of the CC test is very useful: any inequality that is not close-to-binding has no effect on the test asymptotically.

Alternative tests in the moment inequality literature achieve the IDI property via a data-dependent first step that determines which inequalities are close-to-binding. Andrews and Soares (2010), hereafter AS, propose a first step that selects a subset of the moments by comparing each standardized moment to a tuning parameter. AS impose assumptions on the rate of divergence of the tuning parameter in order to select the close-to-binding inequalities. Andrews and Barwick (2012), hereafter AB, propose a recommendation for a fixed value of the tuning parameter in AS and determine size-correction. Romano et al. (2014), hereafter RSW, propose a first-step confidence set for the slackness of each inequality. The CC test achieves the IDI property without a data-dependent first step, and is therefore a more practical method

[3] Perlman and Wu (2006) argue for the appropriateness of redirecting power in this way. The tests they propose are not based on a conditional chi-squared critical value. The Perlman and Wu (2006) tests are not simpler than the classical test, and it is not clear whether they are valid.

[4] For an overview of the literature on partially identified models, see Canay and Shaikh (2017).

[5] An inequality is *close-to-binding* if the slackness drifts to zero with the sample size at the $n^{-1/2}$ rate or faster.

for testing moment inequalities. Simulations show that the CC test has comparable size and weighted average power to the tests in AB and RSW.

Our test in the second setting is related to Rosen (2008) in the same way that our test in the first setting is to Wolak (1987). Namely, we use the same test statistic as Rosen (2008) but propose a critical value that depends on the active inequalities, which then leads to a test with the IDI property. For other aspects of the literature on moment inequality models, we refer the reader to the survey paper by Canay and Shaikh (2017).

The third setting, testing nonlinear inequalities in general asymptotically normal models, is widely applicable. We propose a version of the CC test that counts the number of binding inequalities. We show uniform asymptotic validity by appealing to the multivariate normal model as the limiting experiment. This problem has been considered by Kodde and Palm (1986), and they propose a version of the classical Kudo (1963) test. The CC test is the first in this setting to have the IDI property.

The remainder of this paper proceeds as follows. Section 2 covers the multivariate normal model with affine inequalities. Section 3 covers moment inequality models. Section 4 covers general asymptotically normal models with nonlinear inequalities. Section 5 reports the simulation results. Section 6 concludes. An appendix contains the proofs.

2 The Multivariate Normal Model

In this section, we consider testing affine inequalities on the mean of a multivariate normal random vector. This setting is classical and has been considered, for example by Kudo (1963) and Wolak (1987), among others.

Without loss of generality, let there be a single observation from the multivariate normal distribution: $X \sim N(\mu, I)$, where I is the identity matrix.[6] Consider testing the null hypothesis, $H_0 : \mu \in C$, where C is a convex set defined by affine inequalities. Specifically, let $C = \{\mu \in \mathbb{R}^{d_X} | A\mu \leq b\}$, where A is a $d_A \times d_X$ matrix, $b \in \mathbb{R}^{d_A}$, and the inequality is interpreted element by element. Denote the rows of A by a_j', for $j = 1, ..., d_A$. Note that we impose no assumptions on A or d_A, so that the number of inequalities, d_A, can be arbitrarily large, and A can be low rank or sparse.

[6]Assuming $Var(X) = I$ is no stronger than assuming a known nonsingular $Var(X)$ because we can then premultiply X by a full-rank square matrix to orthonormalize it.

Let $P_C X$ denote the projection of X onto C, which is also the restricted estimator of μ. We test H_0 using the likelihood ratio statistic, which is $T = \|X - P_C X\|^2$ in the multivariate normal model. When $A\mu = b$, Kudo (1963) shows that the distribution of T is a mixture of χ^2 distributions with varying degrees of freedom. The weights in the mixture depend on the shape of C, that is, the angles at which the inequalities meet. When $A\mu \neq b$, the distribution of T is more complicated.

We propose a new test for H_0 in this setting. Let J denote a subset of $\{1, ..., d_A\}$. Write $J(X)$ to denote the set of indices for the inequalities that are active. That is, indices for which $a_j' P_C X = b_j$. Let A_J denote the matrix formed by the rows of A corresponding to indices in J. Let $r(A_J)$ denote the rank of A_J. We propose testing H_0 by comparing $T = \|X - P_C X\|^2$ to the $1 - \alpha$ quantile of the χ^2 distribution with $r(A_{J(X)})$ degrees of freedom, denoted by $\chi^2_{r(A_{J(X)}),1-\alpha}$. Notice that the number of degrees of freedom, $r(A_{J(X)})$, is random because we condition on which inequalities are active. For this reason, we refer to this test as a conditional chi-squared (CC) test. Note that this test does not depend on a priori knowledge of the shape of C, and yet still mimics the mixture distribution.

The following simple example illustrates the CC test.

Example 1. *Consider the null hypothesis,* $H_0 : \mu_1 \leq 0$ *and* $\mu_2 \leq 0$. *In this case,* C *is simply the third quadrant of the plane. Suppose we observe* $(X_1, X_2) \sim N(\mu, I_2)$. *We can write the projection of* $X = (X_1, X_2)$ *onto* C *in closed form as*

$$P_C X = (X_1 \mathbb{1}\{X_1 \leq 0\}, X_2 \mathbb{1}\{X_2 \leq 0\})'.$$

This projection sets to zero any value of X_1 *or* X_2 *that is positive. The squared magnitude of this projection is the likelihood ratio statistic,*

$$T = \|X - P_C X\|^2 = X_1^2 \mathbb{1}\{X_1 > 0\} + X_2^2 \mathbb{1}\{X_2 > 0\}.$$

The indices for the active inequalities are determined by the value of X*:*

$$J(X) = \begin{cases} \{1\} & \text{if } X_1 \geq 0, X_2 < 0 \\ \{2\} & \text{if } X_1 < 0, X_2 \geq 0 \\ \{1,2\} & \text{if } X_1 \geq 0, X_2 \geq 0 \\ \emptyset & \text{if } X_1 < 0, X_2 < 0 \end{cases}.$$

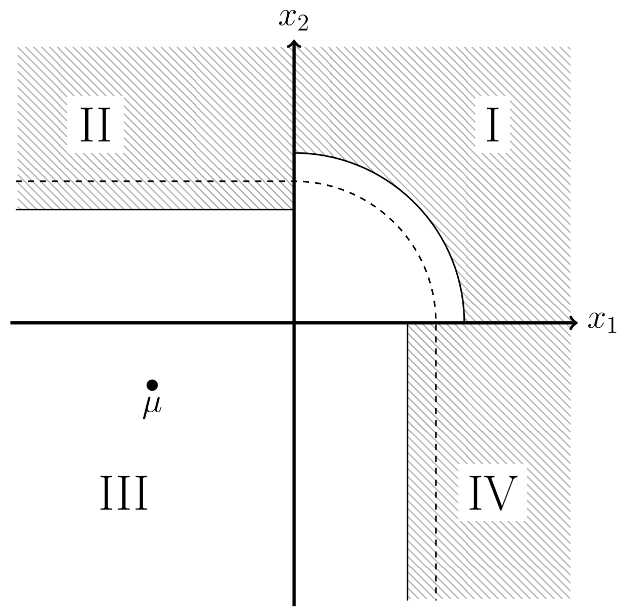

Figure 1: Geometric representation of the CC test (shaded) and the classical Kudo (1963) test (dashed line) in Example 1.

The rejection region for the CC test is illustrated by the shaded region in Figure 1. In quadrant I, both values of X_j are positive, so the CC test rejects if T is larger than $\chi^2_{2,1-\alpha}$. In quadrants II and IV, only one value of X_j is positive, so the CC test rejects if T is larger than $\chi^2_{1,1-\alpha}$. In quadrant III, the CC test never rejects. The dashed line in Figure 1 is discussed below.

Returning to the general setup, we note that the CC test possesses the IDI property (Irrelevance of Distant Inequalities). When an inequality is very slack, the probability that it affects the CC test, either through the likelihood ratio statistic or through the active inequalities, is equal to the probability that it is active, which goes to zero as the degree of slackness diverges. This makes precise the sense in which the CC test possesses the IDI property.

Several papers, including Kudo (1963) and Wolak (1987), propose a classical test for H_0 in this model based on the $1-\alpha$ quantile of the least favorable distribution of T, which is a mixture of chi-squared distributions. Figure 1 illustrates this test in Example 1 with a dashed line. Relative to the classical test, the CC test redirects power toward parameters that violate few inequalities, and away from parameters that violate many (or all) inequalities. This power redirection enables the CC test to have the IDI property.

The following theorem states that the CC test controls size. It also gives conditions under which the test is not conservative.

Theorem 1. (a) *For every* $\mu \in C$*,*

$$\Pr{}_{\mu}(\|X - P_C X\|^2 > \chi^2_{r(A_{J(X)}),1-\alpha}) \leq \alpha(1 - \Pr{}_{\mu}(r(A_{J(X)}) = 0)) \leq \alpha.$$

(b) *If* $A\mu = b$*, then*

$$\Pr{}_{\mu}(\|X - P_C X\|^2 > \chi^2_{r(A_{J(X)}),1-\alpha}) = \alpha(1 - \Pr{}_{\mu}(r(A_{J(X)}) = 0)).$$

Remarks:

1. Part (a) shows that the CC test controls size. Part (b) can be used to show, under some conditions, that there exists a $\mu \in C$ such that the rejection probability is equal to α. This would show that the CC test is not conservative.

2. In order to determine how conservative the CC test is, we evaluate $\Pr_{\mu}(r(A_{J(X)}) = 0)$, which indicates the probability that none of the inequalities are active, at the least favorable $\mu \in C$. A simple case with $\Pr_{\mu}(r(A_{J(X)}) = 0) = 0$ is when at least one equality is being tested. That is, if there exist $j \neq j'$ such that $a_j = -a_{j'} \neq \mathbf{0}$ and $b_j = -b_{j'}$. Otherwise, without an equality, $\Pr_{\mu}(r(A_{J(X)}) = 0)$ usually diminishes exponentially in the number of inequalities.

3. In simple cases with one or two inequalities, including Example 1, it is easy to think of adjustments to the CC test that eliminate the conservativeness.[7] More challenging is to find adjustments that scale well as the number of inequalities increases. Here, we note the availability of an adjustment to the CC test that seems to scale well. Consider a version of the CC test that adjusts α using knowledge of $\Pr_{\mu}(r(A_{J(X)}) = 0)$. That is, the test uses the $1-\beta$ quantile of the chi-squared distribution, where $\beta = \inf_{\mu \in C} \alpha \times [1 - \Pr_{\mu}(r(A_{J(X)}) = 0)]^{-1}$. This can be calculated by simulation. For simplicity, we focus on the basic version of the CC test in this paper, while noting that this adjustment is available.

[7] Indeed, the case of one inequality in one dimension already has a well-known solution: the one-sided t-test is uniformly most powerful. In this case, the adjusted CC test (described in the rest of Remark 3) reduces to the one-sided t-test.

The proof of Theorem 1 relies on (1) a partition of $\mathbb{R}^{d_X}$ that characterizes which inequalities are active, and (2) a property of probabilities of cones under a translation. These are explained in the following two lemmas.

We define some notation for the partition. For any $J \subseteq \{1, ..., d_A\}$, let $J^c = \{1, ..., d_A\} \backslash J$, and let $C_J = \{x \in C : \forall j \in J, a_j' x = b_j, \text{ and } \forall j \in J^c, a_j' x < b_j\}$. Then C_J forms a partition of C. Also let $V_J = \{\sum_{j \in J} v_j a_j : v_j \in \mathbb{R}, v_j \geq 0\}$, and let $K_J = C_J + V_J$.[8] The following lemma shows that K_J forms a partition that characterizes which inequalities are active.

Lemma 1. (a) *If $X \in K_J$, then $X - P_C X \in V_J$ and $P_C X \in C_J$.*

(b) *The set of all K_J for $J \subseteq \{1, ..., d_A\}$ is a partition of $\mathbb{R}^{d_X}$.*

(c) *For every $J \subseteq \{1, ..., d_A\}$, $X \in K_J$ iff $J = J(X)$.*

Next, we define some notation for translations of cones. Let V denote an arbitrary cone in $\mathbb{R}^r$ for a positive integer r.[9] Let V^* denote the polar cone. That is, $V^* = \{\gamma \in \mathbb{R}^r | \langle y, \gamma \rangle \leq 0 \text{ for all } y \in V\}$. For any $\gamma \in V^*$, let $Y \sim N(\gamma, I_r)$. The following lemma provides a property of probabilities of cones under a translation.

Lemma 2. *For every $\gamma \in V^*$, $Pr_\gamma(\|Y\|^2 > \chi^2_{r,1-\alpha} | Y \in V) \leq \alpha$.*

Lemma 2 states that the probability that a random vector, Y, belongs to the tail of its distribution, conditional on belonging to the cone, V, is less than or equal to α, where the tail is any point outside a sphere of radius $\sqrt{\chi^2_{r,1-\alpha}}$. The key assumption is that the mean of Y must belong to the polar cone, V^*, which translates the distribution away from the cone, V. When $\gamma = 0$, this lemma holds with equality because the tail of the χ^2_r distribution has mass exactly α.

We conclude this section by illustrating the proof of Theorem 1 with Example 1.

Example 1, Continued. *Returning to Example 1, we separate the rejection event into the four quadrants of the plane. Notice that*

$$\begin{aligned} \Pr_\mu(T > \chi^2_{r(A_J),1-\alpha}) = & \Pr_\mu(X_1^2 + X_2^2 > \chi^2_{2,1-\alpha} | X \in I) \Pr_\mu(X \in I) \\ & + \Pr_\mu(X_2^2 > \chi^2_{1,1-\alpha} | X \in II) \Pr_\mu(X \in II) \\ & + \Pr_\mu(0 > \chi^2_{0,1-\alpha} | X \in III) \Pr_\mu(X \in III) \end{aligned}$$

[8] When $J = \emptyset$, then $V_J = \{\mathbf{0}_{d_X}\}$.

[9] A cone is a set, V, such that for all $v \in V$ and for all $\lambda \geq 0$, $\lambda v \in V$.

$$+ \Pr_\mu(X_1^2 > \chi^2_{1,1-\alpha} | X \in IV) \Pr_\mu(X \in IV).$$

For each quadrant, the conditional rejection probability is less than or equal to α *when* μ *belongs to the null hypothesis. In the notation of Lemma 1, each quadrant can be written as the sum of a cone,* V_J*, and a subset of the third quadrant,* C_J*. For quadrant III,* $J = \emptyset$*,* $V_J = \{\mathbf{0}_2\}$*,* $C_J = \{(x_1, x_2) : x_1 < 0, x_2 < 0\}$*, and the conditional rejection probability is zero. For quadrant II,* $J = \{2\}$*,* $V_J = \{\mathbf{0}_1\} \times \mathbb{R}_+$ *and* $C_J = \{(x_1, x_2) : x_1 < 0, x_2 = 0\}$*, where* $\mathbb{R}_+$ *denotes the nonnegative real numbers. For quadrant IV,* $J = \{1\}$*,* $V_J = \mathbb{R}_+ \times \{\mathbf{0}_1\}$ *and* $C_J = \{(x_1, x_2) : x_1 = 0, x_2 < 0\}$*. In these two cases, the conditional rejection probability is less than or equal to* α *because* $X_j \sim N(\mu_j, 1)$*. By Lemma 2, we have*

$$\Pr_{\mu_j}(X_j^2 > \chi^2_{1,1-\alpha} | X_j \in V_J) \leq \alpha.$$

In this one-dimensional case, this follows from the monotone hazard property of the univariate normal distribution:

$$\Pr_{\mu_j}(X_j^2 > \chi^2_{1,1-\alpha} | X_j \geq 0) = \frac{1 - \Phi(-\mu_j + \sqrt{\chi^2_{1,1-\alpha}})}{1 - \Phi(-\mu_j)} \begin{cases} = \alpha & \text{if } \mu_j = 0 \\ < \alpha & \text{if } \mu_j < 0 \end{cases}.$$

For quadrant I, $J = \{1, 2\}$*,* $V_J = \mathbb{R}^2_+$*, and* $C_J = \{\mathbf{0}_2\}$*. In this case, the conditional rejection probability is less than or equal to* α *because, by Lemma 2, we have*

$$\Pr_\mu(X_1^2 + X_2^2 > \chi^2_{2,1-\alpha} | X \in V_J) \leq \alpha,$$

with equality if $\mu = \mathbf{0}$*. Thus, Lemma 2 can be seen as a multivariate generalization of the monotone hazard property of the univariate normal distribution.*

3 Moment Inequality Models

In this section, we consider constructing a confidence set for parameters defined by moment inequalities, as in AS and AB, among others.

Let Θ be a parameter space for the unknown parameter θ. Let $A(\cdot)$ be a $d_A \times d_m$ matrix-valued function of the parameter, and let $b(\cdot)$ be a $d_A \times 1$ vector-valued function of the parameter. The moment inequality model is defined by the following inequalities:

$$A(\theta_0)E_{F_0}\bar{m}_n(\theta_0) \leq b(\theta_0), \tag{1}$$

where $\bar{m}_n(\theta) = n^{-1}\sum_{i=1}^n m(W_i,\theta)$, $m(\cdot,\theta)$ is a d_m-dimensional moment function of $\theta \in \Theta$, and $\{W_i : i \geq 1\}$ are the data with joint distribution, F_0. The moment inequality model identifies the true parameter value θ_0 up to the identified set,

$$\Theta_0(F_0) = \{\theta \in \Theta : A(\theta)E_{F_0}\bar{m}_n(\theta) \leq b(\theta)\}. \tag{2}$$

This specification of a moment inequality model includes the more familiar specification found, for example, in AS, by taking $b(\theta) = 0$ and

$$A(\theta) = \begin{pmatrix} -I_p & \mathbf{0}_{p\times v} \\ \mathbf{0}_{v\times p} & -I_v \\ \mathbf{0}_{v\times p} & I_v \end{pmatrix}, \tag{3}$$

where $d_A = p + 2v$, the first p moments are inequalities, and the last v moments are equalities. The generalization to equation (1) is useful because, below, we assume the asymptotic variance matrix of $\bar{m}_n(\theta_0)$ is nonsingular. The generalization allows us to incorporate cases where the asymptotic variance matrix of the inequalities is singular, as long as the inequalities can be written as a linear function of a core set of moments, whose asymptotic variance is nonsingular. An example where this is useful is given in Section 5.2 below.

Let $\widehat{\Sigma}_n(\theta)$ be an estimator of the asymptotic variance matrix of $\bar{m}_n(\theta)$. In the case the data are independent, we simply take

$$\widehat{\Sigma}_n(\theta) = n^{-1}\sum_{i=1}^n (m(W_i,\theta) - \bar{m}_n(\theta))(m(W_i,\theta) - \bar{m}_n(\theta))'. \tag{4}$$

Let $\mathcal{F}$ be a set of distributions of the data that satisfies the following assumption. Notice that part (b) of the following assumption is only used to show that the asymptotic minimum coverage probability of the confidence set proposed below is *equal to* $1 - \alpha$.

Assumption 1. (a) *For every sequence,* $\{F_n \in \mathcal{F}\}_{n=1}^\infty$*, for every sequence,* $\{\theta_n \in \Theta_0(F_n)\}_{n=1}^\infty$*, and for every subsequence,* n_m*, there exists a further subsequence,* n_q*,* *there exists a sequence of positive definite $d_m \times d_m$ matrices, $\{D_q\}$, and there exists a positive definite correlation matrix, Ω, such that under the sequence $\{F_{n_q}\}_{q=1}^{\infty}$,*

$$\sqrt{n_q} D_q^{-1/2} (\bar{m}_{n_q}(\theta_{n_q}) - E_{F_{n_q}} \bar{m}_{n_q}(\theta_{n_q})) \rightarrow_d N(\mathbf{0}, \Omega), \tag{5}$$

and

$$\| D_q^{-1/2} \widehat{\Sigma}_{n_q}(\theta_{n_q}) D_q^{-1/2} - \Omega \| \rightarrow_p 0. \tag{6}$$

(b) *There exist sequences $\{F_n \in \mathcal{F}\}$ and $\{\theta_n \in \Theta_0(F_n)\}$ such that for any subsequence, n_m, there exists a further subsequence, n_q, along which D_q is defined in Assumption* 1(a)*, there exists a pair of indices $\ell, \ell' \in \{1, \ldots, d_A\}$, there exists a sequence of positive definite diagonal $d_A \times d_A$ matrices, $\{\Lambda_q\}$, and there exists a finite matrix A_0 such that*

$$a_\ell(\theta_{n_q}) = -a_{\ell'}(\theta_{n_q}) \neq \mathbf{0} \text{ and } b_\ell(\theta_{n_q}) = -b_{\ell'}(\theta_{n_q}) \tag{7}$$

for all q,

$$\Lambda_q A(\theta_{n_q}) D_q^{1/2} \rightarrow A_0, \tag{8}$$

where $r(A_{J,0}) = r(A_J(\theta_{n_q}))$ for all q and all $J \subseteq \{1, \ldots, d_A\}$, and

$$\sqrt{n_q} \Lambda_q (b(\theta_{n_q}) - A(\theta_{n_q}) E_{F_{n_q}} \bar{m}_{n_q}(\theta_{n_q})) \rightarrow \bar{h}, \tag{9}$$

for some $\bar{h} \in \{0, +\infty\}^{d_A}$, where the convergence holds element by element.

Remarks:

1. There are several ways to impose lower level conditions on $\mathcal{F}$ to guarantee that Assumption 1(a) holds. For example, in the case of the simple moment inequality model (with $A(\theta)$ given in (3)) with i.i.d. data, we can assume that $\mathcal{F}$ is the collection of all F's such that

   (i) $\{W_i : i \geq 1\}$ are i.i.d. under F

   (ii) $\sigma_{F,j}^2(\theta) := Var_F(m_j(W_i, \theta)) \in (0, \infty)$ for all $\theta \in \Theta_0(F), j = 1, \ldots, k$

   (iii) $|Corr_F(m(W_i, \theta))| > c$ for all $\theta \in \Theta_0(F)$

   (iv) $E_F |m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq M$ for $j = 1, \ldots, p + v$,

   where $|\cdot|$ denotes the determinant in (iii), and where $c, \delta,$ and M are fixed positive constants not dependent on F or θ. See Andrews and Guggenberger (2009) or AS for other examples.

2. The subsequencing component of Assumption 1 is a technicality used to show uniformity. It is similar to other subsequencing conditions found, for example, in Andrews and Guggenberger (2009) or Andrews et al. (2019).

3. The matrix D_q typically is the diagonal matrix of variances of the elements of $\sqrt{n_q}\bar{m}_{n_q}(\theta_{n_q})$. We allow each diagonal element to go to zero (or infinity) at different rates, to incorporate the cases where different moments are on different scales or where different moments involve time series processes of different stationarity status. Andrews and Guggenberger (2009), AS, and Andrews et al. (2019) also use a diagonal normalizing matrix for this purpose.

   Moreover, the matrix D_q can be non-diagonal, which is useful when the asymptotic variance matrix of $\sqrt{n_q}(\bar{m}_{n_q}(\theta_{n_q}) - E_{F_n}\bar{m}_{n_q}(\theta_{n_q}))$ is singular but a certain rotation of the vector with proper scaling has a non-singular asymptotic variance matrix.

4. Assumption 1(b) states additional conditions that are sufficient for the asymptotic minimum coverage probability of our proposed confidence set below to be equal to $1-\alpha$. It is most easily satisfied when there is an $F \in \mathcal{F}$ under which the data are stationary and the long-run variance matrix of $\{m(W_i,\theta) : i \geq 1\}$ for some $\theta \in \Theta_0(F)$ is non-singular. In this case, D_q can be a constant diagonal matrix equaling the diagonal of the long-run variance matrix of $\{m(W_i,\theta) : i \geq 1\}$, and $\Lambda_q = I_{d_A}$. Then (8) and (9) are satisfied trivially, and (7) is satisfied as long as at least one equality is included.

We propose a confidence set for θ that is defined through test inversion. Specifically, let

$$T_n(\theta_0) = \inf_{t:A(\theta_0)t \leq b(\theta_0)} n(\bar{m}_n(\theta_0) - t)'\widehat{\Sigma}_n^{-1}(\theta_0)(\bar{m}_n(\theta_0) - t) \qquad (10)$$

denote the statistic for testing the hypothesis $H_0 : \theta = \theta_0$. This statistic is the quasi-likelihood ratio statistic used in Rosen (2008) and AS, among others.

We propose a data-dependent critical value that is based on the rank of $A(\theta_0)$ for those inequalities that are active in finite sample. Note that the minimization problem in (10) has a unique solution because the constraint set is convex and the criterion function is strictly convex. This solution is the restricted estimator for the moments, which we denote by $\hat{t}_n(\theta_0)$. Let $a'_\ell(\theta_0)$ denote the ℓth row of $A(\theta_0)$ and let $b_\ell(\theta_0)$ denote the ℓth element of $b(\theta_0)$ for $\ell = 1, 2, \ldots, d_A$. Let

$$\widehat{J}(\theta_0) = \{\ell \in \{1, 2, \ldots, d_A\} : a'_\ell(\theta_0)\hat{t}_n(\theta_0) = b_\ell(\theta_0)\}, \tag{11}$$

which is the set of indices for the active inequalities. As before, let $r(A_J)$ denote the rank of A_J. Let the critical value be

$$c_n(\theta_0, 1-\alpha) = \chi^2_{r(A_{\widehat{J}(\theta_0)}(\theta_0)),1-\alpha}. \tag{12}$$

For any $\alpha \in (0, 1)$, let the nominal $1 - \alpha$ confidence set be

$$CS_n(1-\alpha) = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta, 1-\alpha)\}.$$

Theorem 2. (a) *Under Assumption* 1(a),

$$\liminf_{n\to\infty} \inf_{F_0\in\mathcal{F}} \inf_{\theta\in\Theta_0(F_0)} \Pr{}_{F_0}(\theta \in CS_n(1-\alpha)) \geq 1-\alpha.$$

(b) *Under Assumption* 1(a) *and* (b),

$$\liminf_{n\to\infty} \inf_{F_0\in\mathcal{F}} \inf_{\theta\in\Theta_0(F_0)} \Pr{}_{F_0}(\theta \in CS_n(1-\alpha)) = 1-\alpha.$$

Remarks:

1. Part (a) shows that the confidence set has asymptotic coverage probability greater than or equal to the nominal coverage probability. The proof of part (a) appears complicated, but the general structure of the proof is quite simple, appealing to Theorem 1(a) as the limit experiment. The almost sure representation theorem is used to show that the event that any given subset, J, of the inequalities are active in sample converges to the event that those J inequalities are active in the limit.

2. Part (b) shows that the confidence set has asymptotic minimum coverage probability equal to the nominal coverage probability when at least one equality is imposed on the moments.

3. Notice that no assumptions are placed on $A(\theta)$ for Theorem 2. It can be low-rank or any submatrix of $A(\theta)$ can be local to singular as θ varies. This is achieved by an extra step in the proof that adds inequalities that are redundant in the finite sample but are relevant in the limit (see Lemma 3, in the Appendix).

4 Testing Nonlinear Inequalities

In this section, we consider testing nonlinear inequality restrictions about a parameter, μ, as considered in Kodde and Palm (1986) and Wolak (1991). Unlike Kodde and Palm (1986) or Wolak (1991), our test uses a conditional chi-squared critical value, rather than a fixed least favorable critical value or its approximation.

Let $\mu \in M \subseteq \mathbb{R}^{d_m}$ be an unknown parameter vector, where M is compact, and let $h : M^o \to \mathbb{R}^{d_h}$ be a known function defined on an open set M^o containing M. We assume that $h(\mu) = (h_1(\mu), h_2(\mu))$ with dimensions d_{h1} and d_{h2} satisfying $d_{h1} + d_{h2} = d_h$. We also assume that $h(\mu)$ is continuously differentiable with derivative, $H(\mu) = \frac{\partial}{\partial \mu'} h(\mu)$, that has full rank, d_h, for all $\mu \in M$. We consider a hypothesis that includes nonlinear inequality restrictions, and may also include some nonlinear equality restrictions:

$$H_0 : h_1(\mu) \leq \mathbf{0} \text{ and } h_2(\mu) = \mathbf{0}.$$

This is tested against the alternative hypothesis $H_1 : h_1(\mu) \nleq \mathbf{0}$ or $h_2(\mu) \neq \mathbf{0}$.

We assume that there is an estimator $\hat{\mu}_n$ of μ, whose distribution is indexed by $\mu \in M$ and $F \in \mathcal{F}(\mu)$, where $\mathcal{F}(\mu)$ is a possibly infinite dimensional space for every $\mu \in M$. Below we assume that $\hat{\mu}_n$ is $\sqrt{n}$-consistent and asymptotically normal with asymptotic variance matrix, Σ^μ. We also assume there is an estimator, $\widehat{\Sigma}^\mu_n$ for Σ^μ.

Assumption 2. *For every sequence $\mu_n \in M$ such that $h_1(\mu_n) \leq 0$ and $h_2(\mu_n) = 0$, for every sequence $F_n \in \mathcal{F}(\mu_n)$, and for every subsequence, n_m, there exists a further subsequence, n_q, and a positive definite matrix, Σ^μ, such that*

$$\sqrt{n_q}(\hat{\mu}_{n_q} - \mu_{n_q}) \to_d N(0, \Sigma^\mu) \text{ and } \widehat{\Sigma}^\mu_{n_q} \to_p \Sigma^\mu$$

as $q \to \infty$.

Remark: Assumption 2 is only slightly stronger than a more common convergence in distribution assumption, requiring convergence along arbitrary subsequences of drifting sequences of parameters. This additional subsequencing component is used for asymptotic size control in a uniform sense, as in Remark 2 after Assumption 1.

We use the following quasi-likelihood ratio statistic:

$$T_n = \inf_{t \leq 0} n(h(\hat{\mu}_n) - (t, 0))'(H(\hat{\mu}_n)\widehat{\Sigma}_n^{\mu} H(\hat{\mu}_n)')^{-1}(h(\hat{\mu}_n) - (t, 0)). \tag{13}$$

Let $\hat{t}_n$ denote the solution to the minimization problem, and let $\widehat{J} = \{\ell \in \{1, \ldots, d_{h1}\} : \hat{t}_{\ell,n} = 0\}$, where $\hat{t}_{\ell,n}$ is the ℓth element of $\hat{t}_n$. For any subset, J, of $\{1, \ldots, d_{h1}\}$, let $|J|$ denote the number of elements of J. We use the critical value $\chi^2_{|\widehat{J}|+d_{h2}, 1-\alpha}$. Thus, the CC test rejects H_0 if and only if $T_n > \chi^2_{|\widehat{J}|+d_{h2}, 1-\alpha}$.

Theorem 3. (a) *Under Assumption 2,*

$$\limsup_{n \to \infty} \sup_{\{\mu \in M : h_1(\mu) \leq 0, h_2(\mu) = 0\}} \sup_{F \in \mathcal{F}(\mu)} \Pr{}_{F,\mu}(T_n > \chi^2_{|\widehat{J}|+d_{h2}, 1-\alpha}) \leq \alpha.$$

(b) *In addition, if $d_{h2} > 0$, then*

$$\limsup_{n \to \infty} \sup_{\{\mu \in M : h_1(\mu) \leq 0, h_2(\mu) = 0\}} \sup_{F \in \mathcal{F}(\mu)} \Pr{}_{F,\mu}(T_n > \chi^2_{|\widehat{J}|+d_{h2}, 1-\alpha}) = \alpha.$$

Remark: Part (a) shows that the CC test asymptotically controls size in a uniform sense for H_0. Part (b) shows that, as long as H_0 hypothesizes at least one equality, the asymptotic size is equal to α.

5 Monte Carlo Simulation

In this section, we report Monte Carlo simulation results to compare the CC test to existing tests. Specifically, we compare to two tests in AB: AQLR/Bt, which is their recommended test and calculates the critical value based on the bootstrap, and AQLR/Nm, which is the same as AQLR/Bt except that the critical value is based on simulations for the asymptotic distribution rather than the bootstrap. We denote the AQLR/Bt test by AB/Bt and the AQLR/Nm test by AB/Nm. We also compare to the two-step procedure of RSW.[10]

[10] We choose the AB tests and RSW's two-step test for comparison because both have the IDI property like ours and because the former tests are tuning parameter-free like ours (in the sense that

We make the comparison in two Monte Carlo settings, the first being the generic moment inequality design from AB, and the second being a version of the entry game example popular in the moment inequality literature. Overall, we find that the CC test is comparable to AB/Bt, AB/Nm, and RSW in terms of size and power and meanwhile is much faster computationally.

5.1 Generic Moment Inequality Design

AB introduce a Monte Carlo design that is sufficiently general to be used for comparing different testing procedures. We use this exact Monte Carlo design. In this subsection, we run a horse race of maximum null rejection probabilities (MNRP) and size-corrected average power of our test, AB/Bt, and AB/Nm.

We briefly describe the Monte Carlo design here and refer the readers to Section 6 of AB for further details. Consider the moment inequality model

$$E[W_i - \theta] \geq \mathbf{0},$$

and the null hypothesis $H_0 : \theta = \mathbf{0}$, where W_i is a p-dimensional random vector. Let the data $\{W_i\}_{i=1}^n$ be i.i.d. with sample size n. Let $W_i = \Omega^{1/2} Z^\dagger + \mu$, where Ω is a correlation matrix, μ is a mean-vector, and $Z^\dagger$ is a vector of independent random variables with zero mean and unit variance. Three choices of distribution for the elements of $Z^\dagger$ are considered: $N(0, 1)$ to represent a best-case scenario for asymptotic approximation, t_3 to represent fat-tailed distributions, and χ^2_3 to represent skewed distributions. Three choices of Ω are considered: Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}, indicating negative, zero, and positive correlation among the moments, respectively. The exact numerical specifications of these matrices for different p's are in Section 4 of AB and Section S7.1 of the Supplemental Material of AB. Also, three choices of p, 2, 4, and 10, are considered.

For each combination of the distribution of $Z^\dagger$, Ω and p, we follow AB exactly to compute the MNRP and the size-corrected average power. Specifically, we compute the rejection probability under a set of μ values that satisfies $\mu \geq \mathbf{0}$ (so that the moment inequalities hold), and take the maximum over this set of μ values to be the MNRP. The set of μ values considered are $\{0, \infty\}^p$. This may differ from the

AB propose and use an optimal choice of the AS tuning parameter), and the latter test is insensitive to reasonable choices of its tuning parameter.

Table 1: Finite Sample Maximum Null Rejection Probabilities and Size-Corrected Average Power of Nominal 5% Tests

			$p=10$			$p=4$			$p=2$		
Test	Distribution	H_0/H_1	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
CC	$N(0,1)$	H_0	.074	.069	.043	.058	.049	.034	.048	.039	.035
AB/Bt	$N(0,1)$	H_0	.061	.062	.058	.053	.056	.049	.054	.053	.052
AB/Nm	$N(0,1)$	H_0	.088	.092	.057	.065	.062	.049	.056	.058	.053
CC	t_3	H_0	.067	.063	.043	.053	.048	.037	.046	.041	.032
AB/Bt	t_3	H_0	.043	.055	.055	.051	.058	.052	.057	.055	.056
AB/Nm	t_3	H_0	.059	.067	.045	.050	.049	.047	.053	.047	.046
CC	χ_3^2	H_0	.132	.145	.068	.086	.101	.060	.070	.073	.062
AB/Bt	χ_3^2	H_0	.062	.066	.057	.050	.055	.050	.054	.053	.056
AB/Nm	χ_3^2	H_0	.136	.153	.068	.093	.101	.062	.085	.087	.080
CC	$N(0,1)$	H_1	.54	.57	.73	.59	.63	.68	.59	.61	.63
AB/Bt	$N(0,1)$	H_1	.46	.62	.77	.54	.64	.76	.63	.68	.71
AB/Nm	$N(0,1)$	H_1	.45	.59	.78	.54	.63	.76	.63	.68	.71
CC	t_3	H_1	.64	.66	.77	.67	.70	.73	.66	.68	.69
AB/Bt	t_3	H_1	.56	.67	.79	.61	.71	.78	.67	.72	.71
AB/Nm	t_3	H_1	.58	.69	.84	.66	.76	.81	.70	.76	.72
CC	χ_3^2	H_1	.42	.42	.66	.52	.50	.64	.54	.54	.59
AB/Bt	χ_3^2	H_1	.43	.51	.72	.53	.57	.70	.57	.59	.62
AB/Nm	χ_3^2	H_1	.37	.42	.72	.48	.53	.71	.56	.57	.61

Note: CC denotes the conditional chi-squared test, AB/Bt denotes the adjusted quasi-likelihood ratio (AQLR) test with bootstrap critical value in AB, and AB/Nm denotes the AQLR test with asymptotic normality-based critical value. The rejection probabilities under H_0 are the maximum null rejection probabilities, and those under H_1 are the average power across the alternative points specified in AB. The numbers for the AB tests are from Table III of AB, which use 5000, 3000, 1000 for both critical value and rejection probability repetitions, respectively for $p = 2, 4, 10$. The numbers for the CC test are computed using 5000 rejection probability repetitions for all p.

true maximum of null rejection probabilities over all nonnegative μ's. For power comparison, we compute the average power for a multitude of μ values that do not satisfy $\mu \geq \mathbf{0}$. These μ values are given in Section 4 of AB and Section S7.1 of the Supplemental Material of AB. The power is size-corrected by adding a positive number to the critical value where the positive number is set to make the size-corrected MNRP equal to the nominal level.

The results are reported in Table 1. Since we follow the design and the procedure to obtain MNRP and size-corrected average power of AB exactly, it is not necessary to re-do their simulations. Thus, the results for AB/Bt and AB/Nm tests are from Table III of AB. The results for our tests are based on 5000 Monte Carlo repetitions.

The first panel of the table shows the MNRP of the CC test and the two versions of the AB adjusted quasi-likelihood ratio tests. As we can see, our test has good

Table 2: Finite Sample Maximum Null Rejection Probabilities and Size-Corrected Average Power of Nominal 5% CC Test

			$p = 10$			$p = 4$			$p = 2$		
n	Distribution	H_0/H_1	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
100	$N(0,1)$	H_0	.074	.069	.043	.058	.049	.034	.048	.039	.035
500	$N(0,1)$	H_0	.058	.054	.042	.051	.043	.034	.046	.035	.034
2000	$N(0,1)$	H_0	.055	.056	.042	.056	.046	.032	.046	.038	.033
100	t_3	H_0	.067	.063	.043	.053	.048	.037	.046	.041	.032
500	t_3	H_0	.051	.051	.044	.051	.046	.033	.048	.039	.035
2000	t_3	H_0	.055	.053	.039	.051	.048	.035	.048	.038	.034
100	χ^2_3	H_0	.132	.145	.068	.086	.101	.060	.070	.073	.062
500	χ^2_3	H_0	.074	.079	.052	.062	.066	.044	.054	.049	.043
2000	χ^2_3	H_0	.061	.066	.048	.057	.057	.039	.049	.042	.036
100	$N(0,1)$	H_1	.54	.57	.73	.59	.63	.68	.59	.61	.63
500	$N(0,1)$	H_1	.58	.63	.73	.60	.62	.67	.58	.61	.62
2000	$N(0,1)$	H_1	.58	.60	.72	.57	.62	.67	.58	.61	.62
100	t_3	H_1	.64	.66	.77	.67	.70	.73	.66	.68	.69
500	t_3	H_1	.64	.66	.75	.63	.66	.70	.62	.64	.65
2000	t_3	H_1	.61	.63	.74	.62	.65	.70	.61	.64	.65
100	χ^2_3	H_1	.42	.42	.66	.52	.50	.64	.54	.54	.59
500	χ^2_3	H_1	.54	.56	.72	.57	.59	.68	.58	.61	.62
2000	χ^2_3	H_1	.57	.57	.73	.59	.60	.68	.60	.62	.63

Note: The rejection probabilities under H_0 are the maximum null rejection probabilities, and those under H_1 are the average power across the alternative points specified in AB. We use 5000 Monte Carlo repetitions for all p.

size-control unless the data have the highly skewed χ^2_3 distribution. This may be due to the fact that this test is based on asymptotic normality and the asymptotic normal approximation can be poor when the sample size is small ($n = 100$) and the underlying distribution is highly skewed. In Table 2, we verify that the over-rejection of the CC test disappears as the sample size increases.

Interestingly, the over-rejection is slightly worse for the AB/Nm test, which is also based on the asymptotic normal approximation. The AB/Bt test has good size control across all cases.[11]

Another thing to note is that the CC test appears conservative when the moments

[11] A bootstrap version of the CC test can be computed by replacing the $\chi^2_{r(A_J(\theta_0)),1-\alpha}$ critical value by the conditional $1-\alpha$ quantile of $n\|P^n_J \widehat{\Sigma}^{*,-1/2}_n(\theta_0)(\bar{m}^*_n(\theta_0) - \bar{m}_n(\theta_0))\|^2$, where P^n_J is the projection matrix onto the space spanned by the rows of $A_J(\theta_0)\widehat{\Sigma}^{1/2}_n$, and $\bar{m}^*_n(\theta_0)$ and $\widehat{\Sigma}^*_n(\theta_0)$ are the bootstrap counterparts of $\bar{m}_n(\theta_0)$ and $\widehat{\Sigma}_n(\theta_0)$, respectively. Similar to the AB/Bt test, the bootstrap version of the CC test can reduce the over-rejection in the presence of highly skewed distributions, but with an increased computational cost. For this reason, we recommend the χ^2-based test unless the user has reason to suspect a high degree of skewness in their moments.

are positively correlated (that is, in the Ω_{Pos} cases). This is consistent with Theorem 1 because, when moments are positively correlated, it is more likely none of the inequalities are active in finite sample. Theorem 1 suggests the CC test should be less conservative when the probability that none of the inequalities are active is lower. This occurs when the moments are negatively correlated or the number of inequalities increases, and this is what we observe in Table 1.

The second panel of the table shows the finite sample size-corrected average power of the CC test against that of AB/Bt and AB/Nm. It is encouraging to see that our simple method has comparable power with the more intricate AB/Bt and AB/Nm tests. In the cases with $p = 4$ and 10, negatively correlated moments, and symmetric errors ($N(0, 1)$ or t_3), the CC test has noticeably better average size-corrected power than both AB/Bt and AB/Nm.

The RSW test is not included in Table 1, but because RSW already compare their test and AB/Bt in their paper using the same design, we can draw a comparison between RSW and the CC test indirectly. RSW's results show that the average size-corrected power of AB/Bt is slightly higher than RSW in all cases, although RSW can have higher power against some of the individual alternatives included in the average power calculation. Their results and ours together are sufficient to show that RSW and the CC test do not dominate each other in terms of either average size-corrected power as defined in AB or size-corrected power against individual alternatives. Their results also show that the RSW test has similar size performance (at $n = 100$) as AB/Bt which is similar to the CC test when the error is normal or Student-t.

5.2 Entry Game

In this subsection, we report the finite sample performance of the CC test in the stylized entry game considered in Shi and Shum (2015). An entry game similar to this is used widely in the moment inequality literature for illustration purposes. We include this model for two reasons: first, it includes moment equalities and thus is useful for illustrating Theorem 2(b); second, it can be written as the moment inequality model in equation (1) with nonzero $b(\theta)$, and thus differs from the Monte Carlo design in the previous subsection.

The game considered is a two-firm entry game with complete information. Player j, $j = 1, 2$ enters the market if the profit of entering exceeds 0: $y_j = 1\{\pi_j \geq 0\}$.

The profit is modeled as $\pi_j = a_j + \delta_j y_{-j} + \varepsilon_j$, where a_j is the expected monopoly profit, δ_j is the competition effect assumed to be nonpositive, and $(\varepsilon_1, \varepsilon_2)$ follows an independent standard normal distribution.

Consider a simple model with no covariates and let Y denote the equilibrium outcome which could take one of four values $\{(1,1),(1,0),(0,1),(0,0)\}$, with the first number in the pair indicating the entry status of the first player and the second number indicating the entry status of the second player. We assume that the agents play a pure strategy Nash equilibrium and the econometrician does not know the equilibrium selection mechanism when there are multiple equilibria. Then the model implies the following moment inequalities/equalities:

$$\begin{aligned} E[1\{Y=(0,0)\}] &= g_{00}(a,\delta) \\ E[1\{Y=(1,1)\}] &= g_{11}(a,\delta) \\ E[1\{Y=(1,0)\}] &\leq g_{10}(a,\delta) \\ E[1\{Y=(1,0)\}] &\geq 1 - g_{00}(a,\delta) - g_{11}(a,\delta) - g_{01}(a,\delta), \end{aligned} \tag{14}$$

where

$$\begin{aligned} g_{00}(a,\delta) &= (1-\Phi(a_1))(1-\Phi(a_2)) \\ g_{11}(a,\delta) &= \Phi(a_1+\delta_1)\Phi(a_2+\delta_2) \\ g_{10}(a,\delta) &= \Phi(a_1)(1-\Phi(a_2+\delta_2)) \\ g_{01}(a,\delta) &= \Phi(a_2)(1-\Phi(a_1+\delta_1)). \end{aligned} \tag{15}$$

This model fits into the framework of (1) with

$$\begin{aligned} \bar{m}_n(\theta) &= n^{-1}\sum_{i=1}^{n} \begin{pmatrix} 1(Y_i=(0,0)) \\ 1(Y_i=(1,1)) \\ 1(Y_i=(1,0)) \end{pmatrix} \\ A(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix} \text{ and} \\ b(\theta) &= \begin{pmatrix} g_{00}(a,\delta) \\ -g_{00}(a,\delta) \\ g_{11}(a,\delta) \\ -g_{11}(a,\delta) \\ g_{10}(a,\delta) \\ g_{00}(a,\delta)+g_{11}(a,\delta)+g_{01}(a,\delta)-1 \end{pmatrix}. \end{aligned} \tag{16}$$

Table 3: Finite Sample Rejection Probabilities of Nominal 5% Tests

Test	θ_{null}	$n = 100$	$n = 250$	$n = 500$
CC	$\theta_0 = (0.5, 0.5, -0.25, -0.25)'$	.0742	.0568	.0492
AB/Bt	$\theta_0 = (0.5, 0.5, -0.25, -0.25)'$	.0250	.0450	.0370
AB/Nm	$\theta_0 = (0.5, 0.5, -0.25, -0.25)'$	.0580	.0550	.0360
RSW	$\theta_0 = (0.5, 0.5, -0.25, -0.25)'$	.0272	.0448	.0428
CC	$\theta_1 - \theta_0 = (-0.13, 0.13, 0, 0)'$	.2106	.4402	.6904
AB/Bt	$\theta_1 - \theta_0 = (-0.13, 0.13, 0, 0)'$	.1140	.3450	.6540
AB/Nm	$\theta_1 - \theta_0 = (-0.13, 0.13, 0, 0)'$	.1840	.3880	.6730
RSW	$\theta_1 - \theta_0 = (-0.13, 0.13, 0, 0)'$	.1170	.3634	.6736
CC	$\theta_2 - \theta_0 = (0, -0.2, -0.21, 0.25)'$	.1960	.3284	.5652
AB/Bt	$\theta_2 - \theta_0 = (0, -0.2, -0.21, 0.25)'$	.0950	.2750	.5120
AB/Nm	$\theta_2 - \theta_0 = (0, -0.2, -0.21, 0.25)'$	.1840	.3140	.5310
RSW	$\theta_2 - \theta_0 = (0, -0.2, -0.21, 0.25)'$	.1136	.2840	.5278

Note: CC denotes the conditional chi-squared test, AB/Bt denotes the adjusted quasi-likelihood ratio (AQLR) test with bootstrap critical value in AB, AB/Nm denotes the AQLR test with asymptotic normality-based critical value, RSW denotes the Bonferroni-corrected test in Romano, Shaikh, and Wolf (2014). All tests use 5000 rejection probability repetitions. The AB/Bt, AB/Nm, and RSW tests use 999 repetitions for their simulated critical values.

A confidence set for the parameter $\theta = (a_1, a_2, \delta_1, \delta_2)'$ can be constructed by inverting a test for the moment inequalities/equalities in (14) at each given value of θ. In this exercise, we compute the rejection probabilities of the CC test, the AB tests, and the RSW test for the null hypothesis: $H_0 : \theta = \theta_{null}$ for several θ_{null} values.

We consider an i.i.d. sample with three choices of sample sizes: $n = 100, 250$, and 500. We generate the data according to the entry game described as above, and let the equilibrium selection rule maximize the joint profit of the two firms when there are multiple equilibria. We set the true value of the parameters to be $\theta_0 = (0.5, 0.5, -0.25, -0.25)$.[12] The null parameter values tested are θ_0, $\theta_1 = (0.37, 0.63, -0.25, -0.25)$, and $\theta_2 = (0.50, 0.30, -0.46, 0.00)'$. The two false null values are chosen to be outside the identified set, both being about .14 away from their corresponding closest points in the identified set in terms of Euclidean distance.[13]

Table 3 reports the results. As we can see from the table, the CC test has moderate

[12]Other choices of the true parameter value and null parameter values yield qualitatively similar results.

[13]The identified set of θ has an empty interior (see Shi and Shum (2015)). Thus, θ_0 is on the boundary of the identified set.

over-rejection at the true parameter value at $n = 100$ but the over-rejection disappears as the sample size increases. The AB tests and the RSW test seem to have under-rejection in most cases, especially AB/Bt. On the other hand, at the false null values, the CC test has noticeably better (higher) rejection rates than the AB tests and the RSW test.

The CC test is a lot faster to compute than either of the three existing tests. For example, on a 2010 Dell Precision workstation with 12G RAM and 3.33GHZ Intel Xeon X5650 CPU running Matlab 2018b, at $n = 500$, the CC test (repeated 5000 times for all three null values) finishes in less than 0.01 hours, while the AB/Nm test uses 4.28 hours. The AB/Bt test (4.77 hours) is slightly more time consuming than AB/Nm, while the RSW (5.8 hours) is even more so possibly because it involves two bootstrapping components rather than one.[14] This is because the CC test requires only one calculation of the quasi likelihood ratio statistic, while the AB and RSW tests require calculating this statistic and its bootstrapped or simulated version many times. Thus, the computational improvement of the CC test is on the same order as the number of simulations or bootstrap draws the AB or RSW tests require. Such a computational improvement is especially useful for inverting a test to construct a confidence set for θ.

6 Conclusion

This paper first considers testing a statistical hypothesis defined by affine inequalities in a multivariate normal model. We propose a new test that compares the likelihood ratio statistic to a chi-squared critical value, where the number of degrees of freedom is the rank of the active inequalities. We show that this test controls size in the multivariate normal model. We also propose a version of the test for moment inequality models and for testing nonlinear inequalities in general asymptotically normal models. We show that these versions of the test asymptotically control size in a uniform sense. These tests provide a simple and practical way to test inequalities or to construct a confidence set for a parameter defined by moment inequalities.

[14] In this example, we are able to and already do use the tabulated optimal tuning parameter and size-correction parameter values given in AB, which drastically reduces computational time for the AB tests.

A Proofs

Proof of Theorem 1

We first prove part (a). Fix $\mu \in C$. First, we use lemma 1, parts (b) and (c) to get

$$\Pr(||X - P_C X||^2 > \chi^2_{r(A_{J(X)}),1-\alpha}) = \sum_{J \subseteq \{1,\ldots,d_A\}} \Pr(X \in K_J \text{ and } ||X - P_C X||^2 > \chi^2_{r(A_J),1-\alpha}). \tag{17}$$

For each J, we consider the span of V_J as a subspace of $\mathbb{R}^{d_X}$. Let P_J denote the projection onto $\text{span}(V_J)$, and M_J denote the projection onto its orthogonal complement. We note that, given J, there exists a $\kappa_J \in \text{span}(V_J)$ such that for every $z \in C_J$, $P_J z = \kappa_J$. This follows because for two $z_1, z_2 \in C_J$, and for any $v \in \text{span}(V_J)$, $\langle z_1 - z_2, v \rangle = 0$, which implies that $z_1 - z_2 \perp \text{span}(V_J)$, so that $P_J(z_1 - z_2) = \mathbf{0}_{d_X}$.

We write $P_J X = P_J(X - P_C X) + P_J P_C X = X - P_C X + \kappa_J$, where the second equality follows by lemma 1(a) and the above discussion. We also write $M_J X = X - P_J X = P_C X - \kappa_J$. Therefore,

$$\begin{aligned}
&\Pr(X \in K_J \text{ and } ||X - P_C X||^2 > \chi^2_{r(A_J),1-\alpha}) \\
=& \Pr(M_J X + \kappa_J \in C_J, P_J X - \kappa_J \in V_J, \text{ and } ||P_J X - \kappa_J||^2 > \chi^2_{r(A_J),1-\alpha}) \\
=& \Pr(M_J X + \kappa_J \in C_J) \times \Pr(P_J X - \kappa_J \in V_J \text{ and } ||P_J X - \kappa_J||^2 > \chi^2_{r(A_J),1-\alpha}),
\end{aligned} \tag{18}$$

where the first equality uses the facts that $X = P_J X + M_J X$ and $K_J = C_J + V_J$, and the second equality follows from the fact that $P_J X$ is independent of $M_J X$.

For J such that $r(A_J) = 0$, we have $\mathit{span}(V_J) = \{\mathbf{0}_{d_X}\}$. Thus, $P_J X = \kappa_J = \mathbf{0}_{d_X}$. Then by (18), we have

$$\Pr(X \in K_J \text{ and } ||X - P_C X||^2 > \chi^2_{r(A_J),1-\alpha}) = 0. \tag{19}$$

For J such that $r(A_J) > 0$, we would like to apply Lemma 2 to the second probability in (18). We first define a linear isometry from $\text{span}(V_J)$ to $\mathbb{R}^{r(A_J)}$. Let B_J be a $d_X \times r(A_J)$ matrix whose columns form a basis for $\text{span}(V_J)$. Then $P_J X = B_J(B_J' B_J)^{-1} B_J' X$. The projection matrix $B_J(B_J' B_J)^{-1} B_J'$ is idempotent with rank $r(A_J)$, and thus there exists a $d_X \times r(A_J)$ matrix with orthonormal columns, Q_J, such that $Q_J Q_J' = B_J(B_J' B_J)^{-1} B_J'$. The linear isometry from $\text{span}(V_J)$ to $\mathbb{R}^{r(A_J)}$ is

$Q_J(X) = Q_J'X$. This is an isometry because for any $v_1, v_2 \in \text{span}(V_J)$,

$$\begin{aligned}
\|v_1 - v_2\|^2 &= (v_1 - v_2)'(v_1 - v_2) \\
&= (v_1 - v_2)'(P_J(v_1 - v_2)) \\
&= (v_1 - v_2)'Q_J Q_J'(v_1 - v_2) \\
&= \|Q_J(v_1) - Q_J(v_2)\|^2, \qquad (20)
\end{aligned}$$

where the second equality holds because $v_1, v_2 \in \text{span}(V_J)$.

Now let $Q_J'V_J = \{Q_J'v : v \in V_J\}$. Then $P_JX - \kappa_J \in V_J$ if and only if $Q_J'(P_JX - \kappa_J) \in Q_J'V_J$ because an isometry is bijective. Therefore, we have

$$\begin{aligned}
&\Pr(P_JX - \kappa_J \in V_J \text{ and } ||P_JX - \kappa_J||^2 > \chi^2_{r(A_J),1-\alpha}) \\
&= \Pr(Q_J'(P_JX - \kappa_J) \in Q_J'V_J \text{ and } ||Q_J'(P_JX - \kappa_J)||^2 > \chi^2_{r(A_J),1-\alpha}). \qquad (21)
\end{aligned}$$

Since $X \sim N(\mu, I)$, we have

$$Q_J'(P_JX - \kappa_J) \sim N(Q_J'(P_J\mu - \kappa_J), Q_J'IQ_J) = N(Q_J'(P_J\mu - \kappa_J), I). \qquad (22)$$

Next, note that $Q_J'V_J$ is a cone in $\mathbb{R}^{r(A_J)}$. The random vector $Q_J'(P_JX - \kappa_J) \sim N(\gamma, I)$ where $\gamma = Q_J'(P_J\mu - \kappa_J)$. The vector γ is in the polar cone because, for all $\tilde{y} \in Q_J'V_J$, there exists a $y = \sum_{j \in J} v_j a_j \in V_J$ such that $\tilde{y} = Q_J'y$, and thus

$$\begin{aligned}
\langle \gamma, \tilde{y} \rangle &= \langle Q_J'(P_J\mu - \kappa_J), Q_J'y \rangle \\
&= \langle P_J\mu - \kappa_J, y \rangle \\
&= \langle (\mu - M_J\mu - P_Jz), y \rangle \\
&= \langle (\mu - M_J\mu - z + M_Jz), y \rangle \\
&= \langle (\mu - z), y \rangle \\
&= \sum_{j \in J} v_j(\langle \mu, a_j \rangle - \langle z, a_j \rangle) \\
&\le 0, \qquad (23)
\end{aligned}$$

where z is any element[15] of C_J so that $\kappa_J = P_Jz$, the second equality holds because

[15]If C_J is empty, so that no such z exists, then $\Pr(X \in K_J) = 0$, so that the following inequalities hold trivially.

$\langle Q'_J(P_J\mu - \kappa_J), Q'_J y\rangle = y'Q_J Q'_J(P_J\mu - \kappa_J) = y'P_J(P_J\mu - \kappa_J) = y'(P_J\mu - \kappa_J)$, and the inequality follows because $\langle z, a_j\rangle = b_j \geq \langle \mu, a_j\rangle$, using the facts that $z \in C_J$ and $\mu \in C$.

Therefore, we can apply Lemma 2 to get that

$$\begin{aligned}
&\Pr(P_J X - \kappa_J \in V_J \text{ and } ||P_J X - \kappa_J||^2 > \chi^2_{r(A_J),1-\alpha}\}) \\
&= Pr\left(Q'_J(P_J X - \kappa_J) \in Q'_J V_J \text{ and } ||Q'_J(P_J X - \kappa_J)||^2 > \chi^2_{r(A_J),1-\alpha}\right) \\
&\leq \alpha \Pr(Q'_J(P_J X - \kappa_J) \in Q'_J V_J) \\
&= \alpha \Pr(P_J X - \kappa_J \in V_J). \qquad (24)
\end{aligned}$$

Plugging this and (19) into equation (17) for every J, we get that

$$\begin{aligned}
&\sum_{J\subseteq\{1,\ldots,d_A\}} \Pr(X \in K_J \text{ and } ||X - P_C X||^2 > \chi^2_{r(A_J),1-\alpha}) \\
&\leq \sum_{J\subseteq\{1,\ldots,d_A\}: r(A_J)>0} \alpha \Pr(M_J X + \kappa_J \in C_J) \times \Pr(P_J X - \kappa_J \in V_J), \\
&= \alpha \times \left(\sum_{J\subseteq\{1,\ldots,d_A\}: r(A_{J(X)})>0} \Pr(X \in K_J)\right) \\
&= \alpha(1 - \Pr(r(A_{J(X)}) = 0)) \leq \alpha,
\end{aligned}$$

where the last equality uses the fact that K_J form a partition of $\mathbb{R}^{d_X}$.

Next, we prove part (b). We start with equation (17). We can exclude those J for which $\Pr(X \in K_J) = 0$:

$$\begin{aligned}
&\Pr(||X - P_C X||^2 > \chi^2_{r(A_{J(X)}),1-\alpha}) \\
&= \sum_{J\subseteq\{1,\ldots,d_A\} \text{ s.t. } \Pr(X\in K_J)>0} \Pr(X \in K_J \text{ and } ||X - P_C X||^2 > \chi^2_{r(A_J),1-\alpha}). \qquad (25)
\end{aligned}$$

As before, for each J, we define $\kappa_J \in \text{span}(V_J)$ to be such that for every $z \in C_J$, $P_J z = \kappa_J$. We write $P_J X = X - P_C X + \kappa_J$ and $M_J X = P_C X - \kappa_J$. Then (18) holds. We show that $\kappa_J = P_J\mu$. We note that since $\Pr(X \in K_J) > 0$, it must be the case that C_J is nonempty. Let $z \in C_J$ and for every $\lambda \in [0, 1]$ let $\mu_\lambda = \lambda z + (1 - \lambda)\mu$.

Recall that in part (b), we have $A\mu = b$. Thus, for each $\lambda \in (0,1]$

$$
\begin{aligned}
a_j'\mu_\lambda &= \lambda a_j' z + (1-\lambda)a_j'\mu = \lambda b_j + (1-\lambda)b_j = b_j \text{ for } j \in J, \text{ and} \\
a_j'\mu_\lambda &= \lambda a_j' z + (1-\lambda)a_j'\mu < \lambda b_j + (1-\lambda)b_j = b_j \text{ for } j \in J^c.
\end{aligned}
$$

This implies that $\mu_\lambda \in C_J$, and hence, for every $\lambda \in (0,1]$, $P_J\mu_\lambda = \kappa_J$. Take $\lambda \to 0$ and by the continuity of the projection, $\kappa_J = P_J\mu$.

Next, we apply the isometry to $P_JX - \kappa_J = P_J(X-\mu)$ in the second probability of equation (18). Then,

$$
\begin{aligned}
&\Pr(P_JX - \kappa_J \in V_J \text{ and } ||P_JX - \kappa_J||^2 > \chi^2_{r(A_J),1-\alpha}) \\
=&\Pr(Q_J'P_J(X-\mu) \in Q_J'V_J \text{ and } ||Q_J'P_J(X-\mu)||^2 > \chi^2_{r(A_J),1-\alpha}) \\
=&Pr\left(\frac{Q_J'P_J(X-\mu)}{||Q_J'P_J(X-\mu)||} \in Q_J'V_J \text{ and } ||Q_J'P_J(X-\mu)||^2 > \chi^2_{r(A_J),1-\alpha}\right), \qquad (26)
\end{aligned}
$$

where the second equality follows from the fact that $Q_J'V_J$ is a cone.

We notice that $Q_J'P_J(X-\mu) \sim N(0,I)$, where the dimension is $r(A_J)$. By a property of the multivariate standard normal distributions, the magnitude of the vector, $||Q_J'P_J(X-\mu)||$ is independent of the direction, $\frac{Q_J'P_J(X-\mu)}{||Q_J'P_J(X-\mu)||}$. Therefore, whenever $r(A_J) > 0$,

$$
\begin{aligned}
&Pr\left(\frac{Q_J'P_J(X-\mu)}{||Q_J'P_J(X-\mu)||} \in Q_J'V_J \text{ and } ||Q_J'P_J(X-\mu)||^2 > \chi^2_{r(A_J),1-\alpha}\right) \\
=&Pr\left(\frac{Q_J'P_J(X-\mu)}{||Q_J'P_J(X-\mu)||} \in Q_J'V_J\right) \times Pr\left(||Q_J'P_J(X-\mu)||^2 > \chi^2_{r(A_J),1-\alpha}\right) \\
=&Pr\left(\frac{Q_J'P_J(X-\mu)}{||Q_J'P_J(X-\mu)||} \in Q_J'V_J\right) \times \alpha \\
=&Pr\left(Q_J'P_J(X-\mu) \in Q_J'V_J\right) \times \alpha \\
=&Pr\left(P_J(X-\mu) \in V_J\right) \times \alpha, \qquad (27)
\end{aligned}
$$

where the second equality follows from the fact that $||Q_J'P_J(X-\mu)||^2 \sim \chi^2_{r(A_J)}$, the third equality uses the fact that $Q_J'V_J$ is a cone, and the fourth equality uses the isometry. When $r(A_J) = 0$, the probability is equal to zero.

Plugging this into equation (25), we get that

$$
\Pr(||X - P_CX||^2 > \chi^2_{r(A_{J(X)}),1-\alpha})
$$

$$
\begin{aligned}
&= \sum_{J\subseteq\{1,\dots,d_A\} \text{ s.t. } \Pr(X\in K_J)>0 \text{ and } r(A_J)>0} \alpha \Pr(M_J X + \kappa_J \in C_J) \times \Pr(P_J X - \kappa_J \in V_J), \\
&= \sum_{J\subseteq\{1,\dots,d_A\} \text{ s.t. } \Pr(X\in K_J)>0 \text{ and } r(A_J)>0} \alpha \Pr(X \in K_J) \\
&= \alpha \times \left(\sum_{J\subseteq\{1,\dots,d_A\} \text{ s.t. } r(A_J)>0} \Pr(X \in K_J) \right) \\
&= \alpha \times (1 - \Pr(r(A_{J(X)}) = 0)),
\end{aligned}
$$

where the first equality follows from equations (18), (19), (26), and (27) and the last equality uses the fact that K_J form a partition of $\mathbb{R}^{d_X}$. □

Proof of Theorem 2

We first prove part (a). Let $\{\theta_n, F_n\}_{n=1}^{\infty}$ be an arbitrary sequence satisfying $F_n \in \mathcal{F}$ and $\theta_n \in \Theta_0(F_n)$ for all n. Let $\{n_m\}$ be an arbitrary subsequence of $\{n\}$. It is sufficient to show that there exists a further subsequence, $\{n_q\}$, such that as $q \to \infty$,

$$
\liminf_{q\to\infty} \Pr{}_{F_{n_q}}(T_{n_q}(\theta_{n_q}) \leq c_{n_q}(\theta_{n_q}, 1-\alpha)) \geq 1 - \alpha. \tag{28}
$$

Fix an arbitrary subsequence, $\{n_m\}$. By Assumption 1(a), there exists a further subsequence, $\{n_q\}$, a sequence of positive definite matrices, D_q, and a positive definite correlation matrix, Ω_0, such that[16]

$$
\sqrt{n_q} D_q^{-1/2}(\bar{m}_{n_q}(\theta_{n_q}) - E_{F_{n_q}} \bar{m}_{n_q}(\theta_{n_q})) \to_d Y \sim N(0, \Omega_0), \text{ and} \tag{29}
$$

$$
D_q^{-1/2} \widehat{\Sigma}_{n_q}(\theta_{n_q}) D_q^{-1/2} \to_p \Omega_0. \tag{30}
$$

We introduce some simplified notation. Let $\widehat{\Omega}_q = D_q^{-1/2} \widehat{\Sigma}_{n_q}(\theta_{n_q}) D_q^{-1/2}$, $X = \Omega_0^{-1/2} Y \sim N(0, I)$, $Y_q = \sqrt{n_q} D_q^{-1/2}(\bar{m}_{n_q}(\theta_{n_q}) - E_{F_{n_q}} \bar{m}_{n_q}(\theta_{n_q}))$, and $X_q = \widehat{\Omega}_q^{-1/2} Y_q$. Equations (29) and (30) imply that

$$
X_q \to_d X \sim N(0, I), \text{ and} \tag{31}
$$

$$
\widehat{\Omega}_q \to_p \Omega_0. \tag{32}
$$

The remainder of the proof proceeds in four steps. (A) In the first step, the

[16] For notational simplicity, we denote all further subsequences by $\{n_q\}$

problem defined in (10) is transformed to include additional inequalities. (B) In the second step, notation is defined for partitioning $\mathbb{R}^{d_m}$ according to Lemma 1, for both finite q and the limit. (C) In the third step, the almost sure representation theorem is invoked on the convergence in (31) and (32). (D) In the final step, we show that (almost surely) the event $T_{n_q}(\theta_{n_q}) \leq c_{n_q}(\theta_{n_q}, 1-\alpha)$ eventually implies a limiting event based on X and Ω_0. This limiting event has probability greater than or equal to $1-\alpha$ from Theorem 1.

(A) Consider the sequence of matrices $A(\theta_{n_q})D_q^{1/2}$. For each q, let Λ_q denote a $d_A \times d_A$ diagonal matrix with positive entries on the diagonal such that each row of $\Lambda_q A(\theta_{n_q})D_q^{1/2}$ is either zero or belongs to the unit circle. Such a Λ_q always exists by taking the diagonal element to be the inverse of the magnitude of the corresponding row of $A(\theta_{n_q})D_q$, if it is nonzero, and one otherwise. Let $g_q = \sqrt{n_q}\Lambda_q(b(\theta_{n_q}) - A(\theta_{n_q})E_{F_{n_q}}\bar{m}_{n_q}(\theta_{n_q}))$. With this notation, we can write

$$T_{n_q}(\theta_{n_q}) = \inf_{y:\Lambda_q A(\theta_{n_q})D_q^{1/2}y \leq g_q} (Y_q - y)'\widehat{\Omega}_q^{-1}(Y_q - y), \tag{33}$$

which adds and subtracts $E_{F_{n_q}}\bar{m}_{n_q}(\theta_{n_q})$ in the objective and applies the change of variables, $y = \sqrt{n_q}D_q^{-1/2}(t - E_{F_{n_q}}\bar{m}_{n_q}(\theta_{n_q}))$.

We can apply Lemma 3 to $\Lambda_q A(\theta_{n_q})D_q^{1/2}$ and g_q to get a further subsequence, n_q, a sequence of matrices, B_q, a sequence of vectors, h_q, matrices A_0 and B_0, and vectors g_0 and h_0, satisfying conditions (a)-(d) of Lemma 3. Let

$$\bar{A}_q = \begin{bmatrix} \Lambda_q A(\theta_{n_q})D_q^{1/2} \\ B_q \end{bmatrix} \text{ and } \bar{h}_q = \begin{bmatrix} g_q \\ h_q \end{bmatrix},$$

and similarly for $\bar{A}_0$ and $\bar{h}_0$. Let $d_{\bar{A}} = d_A + d_B$. We have that

$$T_{n_q}(\theta_{n_q}) = \inf_{y:\bar{A}_q y \leq \bar{h}_q} (Y_q - y)'\widehat{\Omega}_q^{-1}(Y_q - y) \tag{34}$$

$$= \inf_{z:\bar{A}_q\widehat{\Omega}_q^{1/2}z \leq \bar{h}_q} (X_q - z)'(X_q - z), \tag{35}$$

where the first equation follows from condition (b) of Lemma 3 and the second equation follows from the change of variables $z = \widehat{\Omega}_q^{-1/2}y$.

Equation (35) has changed the problem by adding additional inequalities. We verify that the rank of the binding inequalities is unchanged. For any positive definite matrix, Ω, let $J_q(x,\Omega)$ be the set of indices for the binding inequalities in the problem:

$$\inf_{y:\Lambda_q A(\theta_{n_q}) D_q^{1/2} y \leq g_q} (x-y)'\Omega^{-1}(x-y).$$

Recall that $\widehat{J}(\theta_{n_q})$ is the set of binding inequalities for the problem defined in (10), which is equal to $J_q(Y_q, \widehat{\Omega}_q)$ by a change of variables. Similarly, let $\bar{J}_q(x,\Omega)$ be the set of binding inequalities in the problem:

$$\inf_{z:\bar{A}_q \Omega^{1/2} z \leq \bar{h}_q} (x-z)'(x-z). \tag{36}$$

Also let $z_q^*(x,\Omega)$ denote the unique minimizer. We have that for any $y \in \mathbb{R}^{d_m}$ and for any positive definite Ω,

$$r(A_{J_q(y,\Omega)}(\theta_{n_q})) = r([\Lambda_q A(\theta_{n_q}) D_q^{1/2}]_{J_q(y,\Omega)}) = r([\bar{A}_q]_{\bar{J}_q(\Omega^{-1/2}y,\Omega)}) = r([\bar{A}_q \Omega^{1/2}]_{\bar{J}_q(\Omega^{-1/2}y,\Omega)}), \tag{37}$$

where the first equality follows because Λ_q is diagonal with positive entries on the diagonal and D_q is positive definite, the second equality follows by condition (c) of Lemma 3, and the final equality follows from the fact that Ω is positive definite.

Before proceeding to the next step, we simplify the rank calculation by taking a further subsequence. Notice that for each $J \subseteq \{1,...,d_{\bar{A}}\}$, $r([\bar{A}_q]_J) \in \{1,...,d_m\}$. We can denote it by r_J^q, and then take a subsequence, n_q, so that for all J, r_J^q does not depend on q. Similarly, we define $r_J^\infty = r([\bar{A}_0]_J)$. Note that by the convergence of $\bar{A}_q$ to $\bar{A}_0$, $r_J^q \geq r_J^\infty$ for all J.

(B) For any positive definite $d_m \times d_m$ matrix, Ω, and for every $J \subseteq \{1,...,d_{\bar{A}}\}$, let

$$\begin{aligned}
A^q(\Omega) &= \bar{A}_q \Omega^{1/2} \\
a_\ell^{q\prime}(\Omega) &= \ell^{\text{th}} \text{ row of } A^q(\Omega) \\
C^q(\Omega) &= \{x \in \mathbb{R}^{d_m} : a_\ell^{q\prime}(\Omega) x \leq \bar{h}_{\ell,q} \text{ for all } \ell = 1,...,d_{\bar{A}}\} \\
C_J^q(\Omega) &= \{x \in C^q : a_\ell^{q\prime}(\Omega) x = \bar{h}_{\ell,q} \text{ for all } \ell \in J \text{ and } a_\ell^{q\prime}(\Omega) x < \bar{h}_{\ell,q} \text{ for all } \ell \in J^c\} \\
V_J^q(\Omega) &= \left\{ \sum_{\ell \in J} v_\ell a_\ell^q(\Omega) : v_\ell \in \mathbb{R}, v_\ell \geq 0 \right\}, \text{ and} \\
K_J^q(\Omega) &= C_J^q(\Omega) + V_J^q(\Omega).
\end{aligned} \tag{38}$$

Furthermore, for every $J \subseteq \{1, ..., d_{\bar{A}}\}$, let $P_J^q(\Omega)$ denote the projection onto $\text{span}(V_J^q(\Omega))$, and let $M_J^q(\Omega)$ denote its orthogonal projection. There exists a $\kappa_J^q(\Omega) \in \text{span}(V_J^q(\Omega))$ such that for every $x \in C_J^q(\Omega)$, $P_J^q(\Omega)x = \kappa_J^q(\Omega)$. This follows because for two $x_1, x_2 \in C_J^q(\Omega)$, and for any $v \in \text{span}(V_J^q(\Omega))$, $v'(x_1 - x_2) = 0$, which implies that $P_J^q(\Omega)(x_1 - x_2) = 0$.

For every given Ω, we can apply Lemma 1 to the objects defined in (38). This implies that

(a) if $x \in K_J^q(\Omega)$ then $x - z_q^*(x, \Omega) \in V_J^q(\Omega)$ and $z_q^*(x, \Omega) \in C_J^q(\Omega)$,

(b) the sets $K_J^q(\Omega)$ for all $J \subseteq \{1, \ldots, d_{\bar{A}}\}$ form a partition of $\mathbb{R}^{d_m}$, and

(c) for each $J \subseteq \{1, \ldots, d_{\bar{A}}\}$, we have

$$x \in K_J^q(\Omega) \text{ iff } J = \bar{J}_q(x, \Omega). \tag{39}$$

These properties imply that, for all $x \in K_J^q(\Omega)$, we can write

$$P_J^q(\Omega)x = P_J^q(\Omega)(x - z_q^*(x, \Omega)) + P_J^q(\Omega)z_q^*(x, \Omega) = x - z_q^*(x, \Omega) + \kappa_J^q(\Omega), \tag{40}$$

where the second equality follows by (a) and the definition of $\kappa_J^q(\Omega)$. Then, we can also write $M_J^q(\Omega)x = x - P_J^q(\Omega)x = z_q^*(x, \Omega) - \kappa_J^q(\Omega)$.

We define similar notation for the limiting objects. Let $J^\infty = \{\ell \in \{1, ..., d_{\bar{A}}\} : \bar{h}_{\ell,0} < \infty\}$. These are the indices for the inequalities that are "close-to-binding." For any positive definite matrix, Ω, let $A^\infty(\Omega)$ denote the matrix formed by the rows of $\bar{A}_0\Omega^{1/2}$ associated with the indices in J^∞. For notational simplicity, we refer to the rows of $A^\infty(\Omega)$ using $\ell \in J^\infty$ even though the matrix $A^\infty(\Omega)$ has been compressed.

Let

$$\begin{aligned}
&a_\ell^{\infty\prime}(\Omega) = \ell^{\text{th}} \text{ row of } A^\infty(\Omega) \text{ for } \ell \in J^\infty \\
&C^\infty(\Omega) = \{x \in \mathbb{R}^{d_m} : a_\ell^\infty(\Omega)'x \leq \bar{h}_{\ell,0} \text{ for all } \ell \in J^\infty\} \\
&C_J^\infty(\Omega) = \{x \in C^\infty(\Omega) : a_\ell^\infty(\Omega)'x = \bar{h}_{\ell,0} \text{ for all } \ell \in J \text{ and } a_\ell^\infty(\Omega)'x < \bar{h}_{\ell,0} \text{ for all } \ell \in J^\infty \backslash J\} \\
&V_J^\infty(\Omega) = \left\{ \sum_{\ell \in J} v_\ell a_\ell^\infty(\Omega) : v_\ell \in \mathbb{R}, v_\ell \geq 0 \right\}, \text{ and} \\
&K_J^\infty(\Omega) = C_J^\infty(\Omega) + V_J^\infty(\Omega).
\end{aligned} \tag{41}$$

Furthermore, for every $J \subseteq J^\infty$, let $P_J^\infty(\Omega)$ denote the projection onto $\text{span}(V_J^\infty(\Omega))$. There exists a $\kappa_J^\infty(\Omega) \in \text{span}(V_J^\infty(\Omega))$ such that for every $x \in C_J^\infty(\Omega)$,

$$P_J^\infty(\Omega)x = \kappa_J^\infty(\Omega). \tag{42}$$

This follows because for two $x_1, x_2 \in C_J^\infty(\Omega)$, and for any $v \in \text{span}(V_J^\infty(\Omega))$, $v'(x_1 - x_2) = 0$, which implies that $P_J^\infty(\Omega)(x_1 - x_2) = 0$.

Let h^∞ denote the vector formed from the elements of $\bar{h}_0$ that are finite. Let $J^\infty(x, \Omega)$ be the indices for the binding inequalities in the problem:

$$\inf_{z: A^\infty(\Omega) z \leq h^\infty} (x - z)'(x - z).$$

Also let $z_\infty^*(x, \Omega)$ denote the unique minimizer. We can apply Lemma 1 to the objects defined in (41). This implies that

(a^∞) if $x \in K_J^\infty(\Omega)$ then $x - z_\infty^*(x, \Omega) \in V_J^\infty(\Omega)$ and $z_\infty^*(x, \Omega) \in C_J^\infty(\Omega)$,

(b^∞) the set of all $K_J^\infty(\Omega)$ form a partition of $\mathbb{R}^{d_m}$, and

(c^∞) $x \in K_J^\infty(\Omega)$ iff $J = J^\infty(x, \Omega)$.

Before proceeding to the next step, consider $M_J^q(\Omega_0)$, which is a sequence of projection matrices in $\mathbb{R}^{d_m}$ onto a space of dimension $d_m - r_J^q$. Since the space of such matrices is compact, we can find a subsequence, n_q, such that for all $J \subseteq \{1, ..., d_{\bar{A}}\}$, $M_J^q(\Omega_0) \to M_J^N$, where M_J^N is a projection matrix onto a subspace, N_J, of dimension $d_m - r_J^q$.[17] Furthermore, for any sequence of positive definite matrices such that $\Omega_q \to \Omega_0$, we have $M_J^q(\Omega_q) \to M_J^N$. This follows because, if we let E_q denote a $d_m \times r_J^q$ matrix whose columns form an orthonormal basis for $\text{span}(V_J^q(\Omega_0))$ (which is the range of $P_J^q(\Omega_0)$) then for any positive definite matrix, Ω, the columns of $\Omega^{1/2}\Omega_0^{-1/2}E_q$ form a basis for $\text{span}(V_J^q(\Omega))$, which implies that

$$\begin{aligned} M_J^q(\Omega_q) &= I_{d_m} - \Omega_q^{1/2}\Omega_0^{-1/2}E_q(E_q'\Omega_0^{-1/2}\Omega_q\Omega_0^{-1/2}E_q)^{-1}E_q'\Omega_0^{-1/2}\Omega_q^{1/2} \\ &= I_{d_m} - E_q(E_q'E_q)^{-1}E_q' + o(1) = M_J^q(\Omega_0) + o(1). \end{aligned}$$

[17] Recall that r_J^q does not depend on q due to the construction of the subsequence $\{n_q\}$.

(C) Next, we invoke the almost sure representation theorem on the convergence in (31) and (32).[18] Then, we can treat the convergence in (31) and (32) as holding almost surely.[19]

We now construct an event, $\mathcal{X} \subseteq \mathbb{R}^{d_m}$, such that $Pr(X \in \mathcal{X}) = 1$. For every $L \subseteq J^\infty$, let

$$V_{L+}^\infty = \{x \in V_L^\infty | \forall L' \subseteq L, \text{ if } r_{L'}^q < r_L^\infty \text{ then } M_{L'}^N x \neq 0\}. \tag{43}$$

For each $L \subseteq J^\infty$ such that $r_L^\infty > 0$, let

$$\mathcal{X}_L = \{x \in K_L^\infty : P_L^\infty x - \kappa_L^\infty \in V_{L+}^\infty \text{ and } (P_L^\infty x - \kappa_L^\infty)'(P_L^\infty x - \kappa_L^\infty) \neq \chi^2_{r_L^\infty, 1-\alpha}\}. \tag{44}$$

Since $r_L^\infty > 0$, $P_L^\infty X \sim N(0, P_L^\infty)$, which is absolutely continuous on span(V_L^∞), and therefore the probability that $P_L^\infty X - \kappa_L^\infty$ lies in any one of the finitely many subspaces, $null(M_{L'}^N) = \{x \in \mathbb{R}^{d_m} : M_{L'}^N x = 0\}$, each with dimension $r_{L'}^q < r_L^\infty$, is zero. Also, $(P_L^\infty X - \kappa_L^\infty)'(P_L^\infty X - \kappa_L^\infty)$ is absolutely continuous because it can be written as the sum of $r(A_L^\infty)$ squared normal random variables. Therefore,

$$\Pr(P_L^\infty X - \kappa_L^\infty \in V_L^\infty / V_{L+}^\infty \text{ or } (P_L^\infty X - \kappa_L^\infty)'(P_L^\infty X - \kappa_L^\infty) = \chi^2_{r(A_L^\infty), 1-\alpha}) = 0. \tag{45}$$

For $L \subseteq J^\infty$ such that $r(A_L^\infty) = 0$, let $\mathcal{X}_L = K_L^\infty$. Then, let $\mathcal{X} = \cup_{L \subseteq J^\infty} \mathcal{X}_L$. Therefore, by property (b^∞) and equation (45), $\Pr(X \in \mathcal{X}) = 1$.

(D) We consider the set of all sequences such that $x_q \to x^\infty \in \mathcal{X}$ and $\Omega_q \to \Omega_0$. By the definition of $\mathcal{X}$, these sequences occur with probability one. Below we show that for each sequence,

$$\mathbb{1}\{\|x_q - z_q^*(x_q, \Omega_q)\|^2 \leq \chi^2_{r^q_{\bar{J}(x_q, \Omega_q)}}\} \geq \mathbb{1}\{\|x^\infty - z_\infty^*(x^\infty, \Omega_0)\|^2 \leq \chi^2_{r^\infty_{J^\infty(x^\infty, \Omega_0)}}\} \tag{46}$$

eventually. Notice that by (35) and (36), the probability of the left hand side is equal to $\Pr_{F_{n_q}}(T_{n_q}(\theta_{n_q}) \leq c_{n_q}(\theta_{n_q}, 1-\alpha))$. If (46) holds, then by the bounded convergence theorem,

$$\liminf_{q \to \infty} \Pr_{F_{n_q}}(T_{n_q}(\theta_{n_q}) \leq c_{n_q}(\theta_{n_q}, 1-\alpha)) \geq \Pr\left(\|X - z_\infty^*(X, \Omega_0)\|^2 \leq \chi^2_{r^\infty_{J^\infty(X, \Omega_0)}}\right). \tag{47}$$

[18]See van der Vaart and Wellner (1996), Theorem 1.10.3, for the a.s. representation theorem.

[19]This can be formalized by defining random variables, $\tilde{X}_q$, $\tilde{X}$, and $\tilde{\Omega}_q$, satisfying $\tilde{X}_q =_d X_q$, $\tilde{X} =_d X$, $\tilde{\Omega}_q =_d \widehat{\Omega}_q$, $\tilde{X}_q \to_{a.s.} \tilde{X}$, and $\tilde{\Omega}_q \to_{a.s.} \Omega_0$.

Also,

$$\Pr\left(\|X - z_\infty^*(X,\Omega_0)\|^2 \leq \chi^2_{r^\infty_{J^\infty(X,\Omega_0)}}\right) \geq 1-\alpha \tag{48}$$

by Theorem 1(a), because $z_\infty^*(X,\Omega_0) = P_{C^\infty}X$, where P_{C^∞} is the projection of X onto $C^\infty = \{z \in \mathbb{R}^{d_m} : A^\infty(\Omega_0)z \leq h^\infty\}$. Together, (47) and (48) imply (28) for the given subsequence, n_q.

To finish the proof of part (a), we prove (46). Fix a sequence, $x_q \to x_\infty \in \mathcal{X}$ and $\Omega_q \to \Omega_0$. For the rest of this step, consider $A^\infty(\Omega)$, $P_J^\infty(\Omega)$, $\kappa_J^\infty(\Omega)$, and the objects defined in (41) and let the objects without the argument (Ω) denote the objects evaluated at Ω_0. For example, $A^\infty = A^\infty(\Omega_0)$. Similarly, consider the objects defined in (38) and let the objects without the argument (Ω) denote the objects evaluated at Ω_q.

Let L^∞ be the subset of J^∞ for which $x_\infty \in K^\infty_{L^\infty}$. We show that

$$1 = \sum_{L \subseteq \{1,\ldots,d_{\bar{A}}\}} \mathbb{1}\{x_q \in K_L^q\} = \sum_{L \subseteq L^\infty : r_L^q \geq r^\infty_{L^\infty}} \mathbb{1}\{x_q \in K_L^q\} \tag{49}$$

eventually. Property (b) above implies that the first equality holds at every q. Thus it is sufficient to show the second equality. For the second equality, it is sufficient to show that, for all $L \notin \{L \subseteq L^\infty : r_L^q \geq r^\infty_{L^\infty}\}$, $x_q \notin K_L^q$ eventually. Specifically, we consider three cases: (I) $L \not\subseteq J^\infty$, (II) $L \subseteq J^\infty$ but $L \not\subseteq L^\infty$, (III) $L \subseteq L^\infty$ but $r_L^q < r^\infty_{L^\infty}$. By Lemma 7, $z_q^*(x_q,\Omega_q) \to z_\infty^*(x_\infty,\Omega_0)$, where the conditions of the lemma are satisfied by condition (d) from Lemma 3 and the fact that $\bar{h}_q \geq 0$ for all q.

(I) Let $L \not\subseteq J^\infty$. Then, there exists a $\ell \in L$ such that $\bar{h}_{\ell,q} \to \infty$. Then $a_\ell^{q\prime} z_q^*(x_q,\Omega_q) < \bar{h}_{\ell,q}$ eventually because $z_q^*(x_q,\Omega_q) \to z_\infty^*(x_\infty,\Omega_0)$. This implies that $z_q^*(x_q,\Omega_q) \notin C_L^q$, and therefore by (a), $x_q \notin K_L^q$ eventually.

(II) Let $L \subseteq J^\infty$ but $L \not\subseteq L^\infty$. Then, there exists a $\ell \in L$ such that $a_\ell^{\infty\prime} z_\infty^*(x_\infty,\Omega_0) < \bar{h}_{\ell,0}$. By the fact that $a_\ell^{q\prime} z_q^*(x_q,\Omega_q) \to a_\ell^{\infty\prime} z_\infty^*(x_\infty,\Omega_0)$ and $\bar{h}_{\ell,q} \to \bar{h}_{\ell,0}$, we have that $a_\ell^{q\prime} z_q^*(x_q,\Omega_q) < \bar{h}_{\ell,q}$ eventually. This implies that $z_q^*(x_q,\Omega_q) \notin C_L^q$, and therefore by property (a) above, $x_q \notin K_L^q$ eventually.

(III) Let $L \subseteq L^\infty$ such that $r_L^q < r^\infty_{L^\infty}$. This case is impossible if $r^\infty_{L^\infty} = 0$. Thus we only need to consider $r^\infty_{L^\infty} > 0$. Note that $x_\infty - z_\infty^*(x_\infty,\Omega_0) = P^\infty_{L^\infty} x_\infty - \kappa^\infty_{L^\infty}$ by property (a^∞) above. Also, by the definition of $\mathcal{X}$ we have $x_\infty \in \mathcal{X}_{L^\infty}$, which implies that $x_\infty - z_\infty^*(x_\infty,\Omega_0) \in V^\infty_{L^\infty +}$, which in turn means that $M_L^N(x_\infty - z_\infty^*(x_\infty,\Omega_0)) \neq 0$. By the convergence that $M_L^q(x_q - z_q^*(x_q,\Omega_q)) \to M_L^N(x_\infty - z_\infty^*(x_\infty,\Omega_0))$, we have that

$M_L^q(x_q - z_q^*(x_q, \Omega_q)) \neq 0$ eventually. However, if $x_q \in K_L^q$, then by property (a) above, $x_q - z_q^*(x_q, \Omega_q) \in V_L^q$, which implies that $M_L^q(x_q - z_q^*(x_q, \Omega_q)) = 0$. This means that $x_q \notin K_L^q$ eventually. Therefore, (49) holds eventually.

We now verify (46). Notice that

$$
\begin{aligned}
&\mathbb{1}\{\|x_q - z_q^*(x_q, \Omega_q)\|^2 \leq \chi^2_{r^q_{\bar{J}(x_q,\Omega_q)},1-\alpha}\} \\
=&\sum_{J \subseteq \{1,\dots,d_{\bar{A}}\}} \mathbb{1}\{\|x_q - z_q^*(x_q, \Omega_q)\|^2 \leq \chi^2_{r^q_{\bar{J}(x_q,\Omega_q)},1-\alpha}\}\mathbb{1}\{x_q \in K_J^q\} \\
=&\sum_{L \subseteq L^\infty : r_L^q \geq r_{L^\infty}^\infty} \mathbb{1}\{\|x_q - z_q^*(x_q, \Omega_q)\|^2 \leq \chi^2_{r^q_{\bar{J}(x_q,\Omega_q)},1-\alpha}\}\mathbb{1}\{x_q \in K_L^q\} \\
=&\sum_{L \subseteq L^\infty : r_L^q \geq r_{L^\infty}^\infty} \mathbb{1}\{\|x_q - z_q^*(x_q, \Omega_q)\|^2 \leq \chi^2_{r^q_L,1-\alpha}\}\mathbb{1}\{x_q \in K_L^q\} \\
\geq&\sum_{L \subseteq L^\infty : r_L^q \geq r_{L^\infty}^\infty} \mathbb{1}\{\|x_q - z_q^*(x_q, \Omega_q)\|^2 \leq \chi^2_{r^\infty_{L^\infty},1-\alpha}\}\mathbb{1}\{x_q \in K_L^q\} &(50)\\
=&\mathbb{1}\{\|x_\infty - z_\infty^*(x_\infty, \Omega_0)\|^2 \leq \chi^2_{r^\infty_{L^\infty},1-\alpha}\} \sum_{L \subseteq L^\infty : r_L^q \geq r_{L^\infty}^\infty} \mathbb{1}\{x_q \in K_L^q\} &(51)\\
=&\mathbb{1}\{\|x_\infty - z_\infty^*(x_\infty, \Omega_0)\|^2 \leq \chi^2_{r^\infty_{L^\infty},1-\alpha}\}\mathbb{1}\{x_\infty \in K_{L^\infty}^\infty\} \\
=&\sum_{J \subseteq J^\infty} \mathbb{1}\{\|x_\infty - z_\infty^*(x_\infty, \Omega_0)\|^2 \leq \chi^2_{r^\infty_J,1-\alpha}\}\mathbb{1}\{x_\infty \in K_J^\infty\} \\
=&\sum_{J \subseteq J^\infty} \mathbb{1}\{\|x_\infty - z_\infty^*(x_\infty, \Omega_0)\|^2 \leq \chi^2_{r^\infty_{J^\infty(x_\infty,\Omega_0)},1-\alpha}\}\mathbb{1}\{x_\infty \in K_J^\infty\} \\
=&\mathbb{1}\{\|x_\infty - z_\infty^*(x_\infty, \Omega_0)\|^2 \leq \chi^2_{r^\infty_{J^\infty(x_\infty,\Omega_0)},1-\alpha}\}, &(52)
\end{aligned}
$$

where: the first equality follows from property (b); the second equality follows from (49); the third equality follows from property (c); the inequality follows because $r_L^q \geq r_{L^\infty}^\infty$; the fourth equality must hold eventually because, when $r_{L^\infty}^\infty > 0$,

$$
\|x_q - z_q^*(x_q, \Omega_q)\|^2 \to \|x_\infty - z_\infty^*(x_\infty, \Omega_0)\|^2 = \|P_{L^\infty}^\infty x_\infty - \kappa_{L^\infty}^\infty\|^2 \neq \chi^2_{r^\infty_{L^\infty},1-\alpha}
$$

(the equality follows from property (a^∞) above because $x_\infty \in K_{L^\infty}^\infty$, and the inequality follows because $x \in \mathcal{X}_{L^\infty}$) and when $r_{L^\infty}^\infty = 0$, $A_{L^\infty}^\infty = 0$, we have $A_{L^\infty}^q = 0$ eventually (because each row of A^∞ either belongs to the unit circle or is zero), and therefore, $x_q \in K_L^q$ for $L \subseteq K^\infty$ implies $x_q - z_q^*(x_q, \Omega_q) = 0$ eventually; the fifth equality follows from (49) and $x_\infty \in K_{L^\infty}^\infty$; the sixth equality follows because all the terms with $J \neq L^\infty$ are zero; the seventh equality follows from (c$^\infty$); and the final equality follows from (a$^\infty$). This verifies (46), proving part (a).

Next, we prove part (b). Let $\{(\theta_n, F_n)\}$ be the sequence specified in Assumption 1(b). It is sufficient to show that for every subsequence, n_m, there exists a further subsequence, n_q, such that

$$\lim_{q\to\infty} \Pr{}_{F_{n_q}}(T_{n_q}(\theta_{n_q}) \leq c_{n_q}(\theta_{n_q}, 1-\alpha)) = 1-\alpha. \tag{53}$$

Let n_m be an arbitrary subsequence. The proof follows that of part (a) with the following changes.

(i) The existence of Λ_q satisfying (33) now follows from Assumption 1(b).

(ii) The existence of B_q is no longer needed, so we can take $\bar{A}_q = \Lambda_q A(\theta_{n_q}) D_q^{1/2}$ and $\bar{h}_q = g_q$. Then, $\bar{A}_q \to A_0$ and $\bar{h}_q \to \bar{h}$ follows from (8) and (9). Notice that without B_q, (37) still holds without appealing to Lemma 3.

(iii) We show that (46) is satisfied with equality by the arguments in part (a) with modifications (iv)-(vi) below. In this case, the bounded convergence implies that equality holds in (47). Also, an appeal to Theorem 1(b) implies that equality holds in (48), where the conditions of Theorem 1(b) are satisfied because $A^\infty\mu = h^\infty$ since $\mu = 0$ and $h^\infty = 0$, and $\Pr(A_{0,J^\infty(X,\Omega_0)} = 0) = 0$ since $a_\ell^\infty = -a_{\ell'}^\infty \neq 0$ (so at least one of those inequalities is always active in the limit).

(iv) The conditions of Lemma 7 are satisfied, not by Lemma 3, but by Lemma 8, where the condition is satisfied by Assumption 1(b).

(v) The inequality in (50) holds with equality because $r_L^q = r_L^\infty \leq r_{L^\infty}^\infty$ for all L by Assumption 1(b).

(vi) Equality (51) holds when $r_{L^\infty}^\infty = 0$ because by Assumption 1(b), $A_{L^\infty}^q = 0$ eventually. No change is needed when $r_{L^\infty}^\infty > 0$.

Combining these changes with the proof of part (a) proves (53), and therefore, part (b). □

Proof of Theorem 3

The proof of Theorem 3 follows from the same argument as the proof Theorem 2 with the following changes: there is no dependence on θ; the distribution is indexed by both $\mu \in M$ and $F \in \mathcal{F}(\mu)$; $\bar{m}_n(\theta)$ is replaced by $h(\hat{\mu}_n)$; $E_F\bar{m}_n(\theta)$ is replaced by

$h(\mu)$; $b(\theta)$ is replaced by $0_{d_{h1}+2d_{h2}}$; and

$$A(\theta) = \begin{bmatrix} I_{d_{h1}} & 0 \\ 0 & I_{d_{h2}} \\ 0 & -I_{d_{h2}} \end{bmatrix}. \tag{54}$$

The identified set is given by $M_0 = \{\mu \in M : h_1(\mu) \leq 0, h_2(\mu) = 0\}$. We let $\widehat{\Sigma}_n(\theta) = H(\hat{\mu}_n)\widehat{\Sigma}_n^{\mu} H(\hat{\mu}_n)'$, which does not depend on θ.

We verify a modification of Assumption 1(a) to account for the above changes. Fix a sequence, μ_n, F_n, such that $\mu_n \in M_0$, and fix a subsequence, n_m. By Assumption 2, there exists a further subsequence, n_q, and a positive definite matrix, Σ^{μ}, such that $\sqrt{n_q}(\hat{\mu}_{n_q} - \mu_{n_q}) \to N(0, \Sigma^{\mu})$ and $\widehat{\Sigma}_n^{\mu} \to_p \Sigma^{\mu}$. Furthermore, by compactness of M, there exists a further subsequence, n_q, such that $\mu_{n_q} \to \mu_0 \in M$. Let $\Sigma = H(\mu_0)\Sigma^{\mu} H(\mu_0)'$. Then, (5) is satisfied with $D_q = I$ by the delta method (note that Ω needs not be a correlation matrix in this case), and (6) is satisfied by the continuous differentiability of $h(\mu)$ and the consistency of $\hat{\mu}_{n_q}$ for μ_0.

We also note that $r(A_{\widehat{\mathcal{J}}(\theta_0)}(\theta_0)) = |\widehat{J}| + d_{h2}$ because of the definition of $\widehat{J}$ and the fact that $A(\theta)$ takes the form of (54). Therefore,

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{\{\mu \in M : h_1(\mu) \leq 0, h_2(\mu) = 0\}} \sup_{F \in \mathcal{F}(\mu)} \Pr{}_{F,\mu}(T_n > \chi^2_{|\widehat{J}|+d_{h2}, 1-\alpha}) \\ =& 1 - \liminf_{n\to\infty} \inf_{\{\mu \in M : h_1(\mu) \leq 0, h_2(\mu) = 0\}} \inf_{F \in \mathcal{F}(\mu)} \Pr{}_{F,\mu}(T_n \leq \chi^2_{|\widehat{J}|+d_{h2}, 1-\alpha}) \\ \leq& 1 - (1-\alpha) = \alpha, \end{aligned} \tag{55}$$

where the inequality follows from the modified proof of Theorem 2(a).

For part (b), we verify a modification to Assumption 1(b). Under the null hypothesis there exists a $\mu_0 \in M_0$ and an $F_0 \in \mathcal{F}(\mu_0)$. We take our sequence to be the constant sequence, (μ_0, F_0), for all n. Then, for every subsequence, n_m, there exists a further subsequence, n_q, satisfying Assumption 1(a) with $D_q = I$. Condition (7) is satisfied with $\ell = d_{h1} + 1$ and $\ell' = d_{h1} + d_{h2} + 1$ by (54) and the fact that $d_{h2} > 0$. Condition (8) is satisfied with $\Lambda_q = I$ because the left hand side does not depend on q. Also, condition (9) is satisfied because every element of $\Lambda_q(b(\theta_{n_q}) - A(\theta_{n_q})E_{F_{n_q}}\bar{m}_{n_q}(\theta_{n_q}))$, which is just equal to $-A(\theta)h(\mu_0)$, does not depend on q and is nonnegative. Therefore, when multiplied by $\sqrt{n_q}$, each element either diverges to $+\infty$ or converges to zero. Therefore, by the modified proof of Theorem 2(b), the inequality in (55) holds with equality, and part (b) of Theorem 3 holds. $\square$

Proof of Lemma 1

(a) By assumption, $X \in K_J = C_J + V_J$. So, we write $X = X_1 + X_2$, where $X_1 \in C_J$ and $X_2 \in V_J$. Then, $P_C X_1 = X_1$ because $X_1 \in C$ already. We show that $P_C X = X_1$. By a property of projection onto convex sets, it is necessary and sufficient that for all $y \in C$, we have $\langle X - X_1, y - X_1 \rangle \leq 0$.[20] This follows because $X_2 = \sum_{j \in J} v_j a_j$ with $v_j \geq 0$, so

$$\langle X_2, y - X_1 \rangle = \sum_{j \in J} v_j (\langle a_j, y \rangle - \langle a_j, X_1 \rangle) \leq 0,$$

where the inequality uses the fact that $y \in C$, so $a_j' y \leq b_j$ and $X_1 \in C_J$, so $a_j' X_1 = b_j$. Combining these, we get that $P_C X = X_1 \in C_J$ and $X - P_C X = X - X_1 = X_2 \in V_J$.

(b) We first show that every X belongs to some K_J. For every X, $P_C X \in C$, so there exists a J such that $P_C X \in C_J$.

By the inner-product property of projection, we know that for all $y \in C$, $\langle y - P_C X, X - P_C X \rangle \leq 0$. Using this fact, let $z \perp \text{span}(V_J)$. Then, there exists a $\epsilon > 0$ such that $P_C X + \epsilon z$ and $P_C X - \epsilon z$ both belong to C.[21] Then, $\langle \epsilon z, X - P_C X \rangle \leq 0$ and $\langle -\epsilon z, X - P_C X \rangle \leq 0$. These two inequalities imply that $\langle z, X - P_C X \rangle = 0$. Thus, $X - P_C X$ is orthogonal to all vectors, z, which are orthogonal to $\text{span}(V_J)$. This implies that $X - P_C X \in \text{span}(V_J)$.

If $X - P_C X \notin V_J$, then by the separating hyperplane theorem,[22] there exists a direction, $c \in \text{span}(V_J)$ such that $\langle c, X - P_C X \rangle > 0$ and $\langle c, a_j \rangle < 0$ for all $j \in J$. We consider $P_C X + \epsilon c$. We show that for ϵ sufficiently small, (1) $P_C X + \epsilon c \in C$, and (2) $\langle X - P_C X, \epsilon c \rangle > 0$.

(1) For $j \in J$, $\langle P_C X + \epsilon c, a_j \rangle = b_j + \epsilon \langle c, a_j \rangle < b_j$, where the equality follows because $P_C X \in C_J$ and the inequality follows from the definition of c. For

[20] See Section 3.12 in Luenberger (1969). Hereafter, call this property of projection onto a convex set the "inner-product property."

[21] This uses the slackness of the inequalities in the definition of C_J.

[22] See Section 11 of Rockafellar (1970) or Section 5.12 in Luenberger (1969).

$j \in J^c$, $\langle P_C X + \epsilon c, a_j\rangle = \langle P_C X, a_j\rangle + \epsilon\langle c, a_j\rangle$, which is less than b_j for ϵ sufficiently small because $\langle P_C X, a_j\rangle < b_j$.

(2) $\langle X - P_C X, \epsilon c\rangle = \epsilon\langle X - P_C X, c\rangle > 0$ by the definition of c.

This contradicts the inner-product property of projection onto a convex set, and therefore $X - P_C X \in V_J$, and $X \in K_J$.

We next show that no X belongs to two distinct K_J. If $X \in K_J$ and $K_{J'}$, then, by part (a), $P_C X \in C_J$ and $P_C X \in C_{J'}$. But this is a contradiction because the projection onto a convex set is unique, and the C_J form a partition of C.

(c) If $X \in K_J$, then $P_C X \in C_J$, so all the inequalities in J are binding. If $X \notin K_J$, then X is in a different $K_{J'}$, for some $J' \neq J$, by part (b). Thus, $J \neq J(X) = J'$.

□

Proof of Lemma 2

It is without loss of generality to assume $\gamma = -\lambda e_1$ for some $\lambda \geq 0$ because the inequality to be proved is invariant to a rotation of Y.

First we change variables to spherical coordinates. Essentially, this reduces the problem from any cone, V, to a ray. Calculate

$$\begin{aligned}
&\alpha Pr_\gamma(Y \in V) - Pr_\gamma(\{Y \in V \text{ and } ||Y||^2 > \chi^2_{n,1-\alpha}\}) \\
&= \int \mathbb{1}\{Y \in V\}(\alpha - \mathbb{1}\{||Y||^2 > \chi^2_{r,1-\alpha}\})\phi(Y + \lambda e_1)dY \\
&= \int\int (\alpha - \mathbb{1}\{\rho^2 > \chi^2_{r,1-\alpha}\})\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\lambda\cos(\phi_1))^2}d\rho \mathbb{1}\{Y/||Y|| \in V\}c(\phi,\lambda)d\phi_1\cdots d\phi_{n-1},
\end{aligned}$$

where the last equality is a conversion to spherical coordinates: $\rho = ||Y||$, ϕ_1, ϕ_2, ..., ϕ_{r-1} are angles that satisfy $Y_1 = \rho\cos(\phi_1)$, $Y_2 = \rho\sin(\phi_1)\cos(\phi_2)$, ..., $Y_{r-1} = \rho\sin(\phi_1)\cdots\sin(\phi_{r-2})\cos(\phi(r-1))$, and $Y_n = \rho\sin(\phi_1)\cdots\sin(\phi_{r-2})\sin(\phi_{r-1})$, $c(\phi,\lambda) = \frac{\Gamma(r/2)}{2\pi^{r/2}}e^{-\frac{1}{2}\lambda^2(1-\cos^2(\phi_1))}\prod_{i=1}^{r-2}\sin^{r-1-i}(\phi_i) \geq 0$ (which mostly comes from the Jacobian of the transformation, but also includes a constant), and $\beta = \frac{1}{2^{r/2-1}\Gamma(r/2)}$. Because any $Y \in V$ has $Y_1 \geq 0$[23], ϕ_1 only ranges from 0 to $\pi/2$. ϕ_i for $i = 2, ..., n-2$ ranges from 0 to π, while ϕ_{n-1} ranges from 0 to 2π.

[23]This follows because $-e_1$ is in the polar cone, so $\langle Y, -e_1\rangle \leq 0$.

We focus on the inner integral, which is along a ray defined by ϕ with distribution that is very similar to a χ distribution, but with an extra noncentrality parameter, $\lambda \cos(\phi_1) \geq 0$. Let $\bar{\lambda} = \lambda \cos(\phi_1)$, and we show that for every $\bar{\lambda} > 0$, the inner integral is nonnegative. This is sufficient because the weights in the outer integral are all nonnegative.

For every $\bar{\lambda} \geq 0$, let

$$f(\bar{\lambda}) = \int_0^\infty (\alpha - \mathbb{1}\{\rho^2 > \chi^2_{r,1-\alpha}\})\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}dr.$$

Let $f'(\bar{\lambda})$ denote the derivative of f. We show that (1) $f(0) \geq 0$ and (2) for all $\bar{\lambda} \geq 0$, $f'(\bar{\lambda}) \geq -\left(\sqrt{\chi^2_{r,1-\alpha}} + \bar{\lambda}\right) f(\bar{\lambda})$. Together, these two properties imply that $f(\bar{\lambda}) \geq 0$ because, if not, then there exists a $\bar{\lambda} > 0$ such that $f(\bar{\lambda}) < 0$. Then, by the mean value theorem, there exists a $\tilde{\lambda} \in (0, \bar{\lambda})$ such that $f(\tilde{\lambda}) < 0$ and $f'(\tilde{\lambda}) < 0$, which contradicts property (2).

Property (1) holds because, for $\bar{\lambda} = 0$, ρ is a χ distribution with r degrees of freedom, so the probability that $\rho^2 > \chi^2_{r,1-\alpha}$ is α. To show that property (2) holds, we evaluate:

$$\begin{aligned}
f'(\bar{\lambda}) &= \frac{d}{d\bar{\lambda}} \int (\alpha - \mathbb{1}\{\rho^2 > \chi^2_{r,1-\alpha}\})\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}d\rho \\
&= \int (\alpha - \mathbb{1}\{\rho^2 > \chi^2_{r,1-\alpha}\})\beta\rho^{r-1}\frac{d}{d\bar{\lambda}}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}d\rho \\
&= -\int (\rho+\bar{\lambda})(\alpha - \mathbb{1}\{\rho^2 > \chi^2_{r,1-\alpha}\})\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}d\rho \\
&= -\int (\rho+\bar{\lambda})\alpha\mathbb{1}\{\rho^2 \leq \chi^2_{r,1-\alpha}\}\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}d\rho \\
&\quad + \int (\rho+\bar{\lambda})(1-\alpha)\mathbb{1}\{\rho^2 > \chi^2_{r,1-\alpha}\}\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}d\rho \\
&\geq -\int \left(\sqrt{\chi^2_{r,1-\alpha}} + \bar{\lambda}\right)\alpha\mathbb{1}\{\rho^2 \leq \chi^2_{r,1-\alpha}\}\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}d\rho \\
&\quad + \int \left(\sqrt{\chi^2_{r,1-\alpha}} + \bar{\lambda}\right)(1-\alpha)\mathbb{1}\{\rho^2 > \chi^2_{r,1-\alpha}\}\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}d\rho \\
&= -\left(\sqrt{\chi^2_{r,1-\alpha}} + \bar{\lambda}\right)\int (\alpha - \mathbb{1}\{\rho^2 > \chi^2_{r,1-\alpha}\})\beta\rho^{r-1}e^{-\frac{1}{2}(\rho+\bar{\lambda})^2}d\rho \\
&= -\left(\sqrt{\chi^2_{r,1-\alpha}} + \bar{\lambda}\right) f(\bar{\lambda}),
\end{aligned}$$

where the second equality follows by dominated convergence and the inequality follows from the events $\{\rho^2 > \chi^2_{r,1-\alpha}\}$ and $\{\rho^2 \leq \chi^2_{r,1-\alpha}\}$. $\square$

Lemmas 3–6

For any matrix, A, and for any vector, g, let $C(A, g) = \{x \in \mathbb{R}^{d_x} : Ax \leq g\}$.

Lemma 3. *Let A_n be a sequence of $d_A \times d_x$ matrices such that each row is either zero or belongs to the unit circle. Let g_n be a sequence of nonnegative d_A-vectors. Then, there exists a subsequence, n_q, a sequence of $d_B \times d_x$ matrices, B_q, and a sequence of nonnegative d_B-vectors h_q such that the following hold.*[24]

(a) *$A_{n_q} \to A_0$, $B_q \to B_0$, $g_{n_q} \to g_0$, and $h_q \to h_0$ (some of the elements of g_0 and h_0 may be $+\infty$, in which case the convergence/divergence occurs elementwise).*

(b) *$C(A_{n_q}, g_{n_q}) \subseteq C(B_q, h_q)$ for all q.*

(c) *If we let $J^A_q(x) = \{j \in \{1, ..., d_A\} : a'_{j,n_q}x = g_{j,n_q}\}$ and let $J^B_q(x) = \{j \in \{1, ..., d_B\} : b'_{j,q}x = h_{j,q}\}$, then for all $x \in C(A_{n_q}, g_{n_q})$ and for all q, $r(\{a_{j,n_q}\}_{j \in J^A_q(x)}) = r(\{a_{j,n_q}\}_{j \in J^A_q(x)} \cup \{b_{j,q}\}_{j \in J^B_q(x)})$.*

(d) *$C(A_{n_q}, g_{n_q}) \cap C(B_q, h_q) \to C(A_0, g_0) \cap C(B_0, h_0)$ pointwise, which means that*

  (i) *for every sequence $x_q \in C(A_{n_q}, g_{n_q}) \cap C(B_q, h_q)$ such that $x_q \to x_0$, $x_0 \in C(A_0, g_0) \cap C(B_0, h_0)$, and*

  (ii) *for every $x_0 \in C(A_0, g_0) \cap C(B_0, h_0)$, there exists a sequence, $x_q \in C(A_{n_q}, g_{n_q}) \cap C(B_q, h_q)$ such that $x_q \to x_0$.*

Proof of Lemma 3. Before proving the lemma, we note that for any subsequence, n_q such that $A_{n_q} \to A_0$ and $g_{n_q} \to g_0$, and for any $B_q \to B_0$ and $h_q \to h_0$, condition (d)(i) is satisfied. Specifically, let x_q denote a sequence that belongs to $C(A_{n_q}, g_{n_q}) \cap C(B_q, h_q)$ for all q, and such that $x_q \to x_0$. Then

$$a'_{j,0}x_0 = \lim_{q\to\infty} a'_{j,n_q}x_q \leq \lim_{q\to\infty} g_{j,n_q} = g_{j,0}.$$

Also, by the convergence of h_q, we have that

$$b'_{j,0}x_0 = \lim_{q\to\infty} b'_{j,q}x_q \leq \lim_{q\to\infty} h_{j,q} = h_{j,0}.$$

[24]We use $b_{j,q}$ to denote the transpose of the j^{th} row of B_q, and similarly for a_{j,n_q}, $a_{j,0}$, and $b_{j,0}$.

Therefore, $x_0 \in C(A_0, g_0) \cap C(B_0, h_0)$.

We also note that for any q, B_q, and h_q satisfying (b), condition (c) must also be satisfied. If not, then there exists a q, an $x \in C(A_{n_q}, g_{n_q})$ and a $j' \in J_q^B(x)$ such that $b_{j',q}$ cannot be written as a linear combination of a_{j,n_q} for $j \in J_q^A(x)$. This implies that there exists a v such that $b'_{j',q}v > 0$ and $v \perp a_{j,n_q}$ for all $j \in J_q^A(x)$. But then, $x + \alpha v \in C(A_{n_q}, g_{n_q})$ for sufficiently small α, at the same time that $b'_{j',q}(x + \alpha v) > h_q$. This contradicts the fact that $C(A_{n_q}, g_{n_q}) \subseteq C(B_q, h_q)$. Therefore, (c) holds.

We now prove the lemma by finding a subsequence, n_q, and sequences $\{B_q\}$ and $\{h_q\}$ that satisfy conditions (a), (b), and (d)(ii). We first consider A_n and g_n. By the compactness of the unit circle, let n_q be a subsequence so that A_{n_q} converges to some A_0. Also suppose g_{n_q} converges along the subsequence to some vector $g_0 \in (\mathbb{R}_+ \cup \{+\infty\})^{d_A}$.

Let J_A^+ denote the subset of $\{1, ..., d_A\}$ for which $g_{j,0} > 0$, and let J_A^0 denote the subset for which $g_{j,0} = 0$. Consider $A_{J_A^0,0}$, which defines a cone in $\mathbb{R}^{d_x}$: $C(A_{J_A^0,0}, 0) = \{x \in \mathbb{R}^{d_x} : A_{J_A^0,0}x \leq 0\}$. Let S denote the smallest linear subspace of $\mathbb{R}^{d_x}$ that contains this cone. (For any two subspaces containing the cone, the intersection also does. Therefore, the smallest subspace is well-defined.) Let the dimension of S be d_S. Let J_A^S be the subset of J_A^0 for which $a_{j,0} \perp S$ for all $j \in J_A^S$. Let $J_A^N = \{1, ..., d_A\} \backslash J_A^S$.

Next, we define sequences B_q and h_q that satisfy conditions (a), (b), and (d)(ii) by induction on the dimension of S. If $d_S = 0$, then no B_q or h_q is required. Condition (a) is satisfied by the above choice of the subsequence. Condition (b) is satisfied because $C(B_q, h_q) = \mathbb{R}^{d_x}$ for all q. Condition (d)(ii) is satisfied because $C(A_0, g_0) = \{0\}$, and then we can take $x_q = 0$ for all q, which belongs to $C(A_{n_q}, g_{n_q})$ and converges to $x_0 = 0 \in C(A_0, g_0)$.

If $d_S > 0$, then suppose that the conclusion of Lemma 3 holds for all values of the dimension of S less than d_S. Let $C_q = C(A_{J_A^S,n_q}, g_{J_A^S,n_q})$. Let C_q^S be the projection of C_q onto S. That is, $C_q^S = \{P_S x : x \in C_q\}$, where P_S denotes the projection onto S and $M_S = I - P_S$. The fact that C_q is a polyhedral set (defined by finitely many affine inequalities) implies by Theorem 19.3 in Rockafellar (1970) that C_q^S is also a polyhedral set. Therefore, there exists a $d_{B_1} \times d_x$ matrix of unit vectors in S, B_q^1 and a vector h_q^1 such that $C_q^S = \{y \in S : B_q^1 y \leq h_q^1\}$. We note that C_q^S contains zero, so $h_q^1 \geq 0$. Let n_q be a further subsequence so that $B_q^1 \to B_0^1$ and $h_q^1 \to h_0^1$, where some of the elements of h_0^1 may be $+\infty$, in which case the convergence holds elementwise. We note that this construction satisfies conditions

(a) and (b) because $C(A_{n_q}, g_{n_q}) \subseteq C_q \subseteq C(B_q^1, h_q^1)$ for all q, where the second subset holds because $B_q^1 x = B_q^1 M_S x + B_q^1 P_S x = B_q^1 P_S x \leq h_q^1$ for all $x \in C_q$ because the rows of B_q^1 belong to S and $P_S x \in C_q^S$.

Let J_B^+ denote the set of $j \in \{1, ..., d_{B_1}\}$ for which $h_{j,0}^1 > 0$, and let J_B^0 denote the set for which $h_{j,0}^1 = 0$, where $h_{j,0}^1$ is the jth element of h_0^1. Consider $B_{J_B^0,0}^1$ and $A_{J_A^0,0}$, which together define a cone in S: $\{x \in S : B_{J_B^0,0}^1 x \leq 0 \text{ and } A_{J_A^0,0} x \leq 0\}$. As before, let $S^\dagger$ denote the smallest linear subspace of S that contains this cone. Let $J_B^{S^\dagger}$ denote the set of all $j \in J_B^0$ for which $b_{j,0}^1 \perp S^\dagger$. Also let $J_A^{S^\dagger}$ denote the set of all $j \in J_A^0$ for which $a_{j,0} \perp S^\dagger$. Let the dimension of $S^\dagger$ be $d_{S^\dagger}$.

If $d_{S^\dagger} < d_S$, then the result follows by the induction assumption. In particular, if we let

$$\tilde{A}_q = \begin{bmatrix} A_{n_q} \\ B_q^1 \end{bmatrix} \text{ and } \tilde{g}_q = \begin{bmatrix} g_{n_q} \\ h_q^1 \end{bmatrix},$$

then the subspace, $\tilde{S}$, defined to be the smallest linear subspace containing $C(\tilde{A}_0, \tilde{g}_0)$, is equal to $S^\dagger$. Therefore, there exists a further subsequence, n_q, and another matrix of inequalities, B_q^2 and h_q^2 such that: (a) $B_q^2 \to B_0^2$ and $h_q^2 \to h_0^2$, (b) $C(\tilde{A}_q, \tilde{g}_q) \subseteq C(B_q^2, h_q^2)$ for all q along the subsequence, and (d)(ii) $C(\tilde{A}_q, \tilde{g}_q) \cap C(B_q^2, h_q^2) \to C(\tilde{A}_0, \tilde{g}_0) \cap C(B_0^2, h_0^2)$ pointwise. It is easy to see that these conditions imply conditions (a), (b), and (d)(ii) for the original A_n and g_n along this subsequence, with

$$B_q = \begin{bmatrix} B_q^1 \\ B_q^2 \end{bmatrix} \text{ and } h_q = \begin{bmatrix} h_q^1 \\ h_q^2 \end{bmatrix},$$

using the fact that $C(\tilde{A}_q, \tilde{g}_q) = C(A_{n_q}, g_{n_q}) \cap C(B_q^1, h_q^1)$.

Therefore, we only need to show condition (d)(ii) in the case that $d_{S^\dagger} = d_S$. In this case, $S = S^\dagger$, and so $J_B^{S^\dagger} = \emptyset$ and $J_A^{S^\dagger} = J_A^S$. Fix $x_0 \in C(A_0, g_0) \cap C(B_0^1, h_0^1)$. We show that for every $\epsilon > 0$ there exists a Q such that for all $q \geq Q$ there exists a $y_q \in C(A_{n_q}, g_{n_q}) \cap C(B_q^1, h_q^1)$ such that $\|y_q - x_0\| \leq 2\epsilon$. If true, then this can be used to construct a sequence, $y_q \to x_0$, satisfying condition (d)(ii).

Fix $\epsilon > 0$. By Lemma 4, there exists a point, $\tilde{x}$, in S that satisfies ${b_{j,0}^1}' \tilde{x} < h_{j,0}^1$ for all $j \in \{1, ..., d_{B_1}\}$, and $a_{j,0}' \tilde{x} < g_{j,0}$ for all $j \in J_A^N$. There exists a $\lambda \in (0,1)$ small enough that $x^\dagger = \lambda \tilde{x} + (1 - \lambda) x_0 \in \bar{B}(x_0, \epsilon)$, where $\bar{B}(x_0, \epsilon)$ denotes the closed ball of radius ϵ around x_0. Note that $x^\dagger$ satisfies $a_{j,0}' x^\dagger < g_{j,0}$ for all $j \in J_A^N$ and ${b_{j,0}^1}' x^\dagger < h_{j,0}$ for all $j \in \{1, ..., d_{B_1}\}$. Therefore, there exists a $\delta \in (0, \epsilon)$ and a Q such that for all

$q \geq Q$, and for all $x \in \bar{B}(x^\dagger, \delta)$, ${b_{j,q}^1}' x < h_{j,q}^1$ for all $j \in \{1, ..., d_{B_1}\}$, and $a_{j,n_q}' x < g_{j,n_q}$ for all $j \in J_A^N$. Notice that, for all $q \geq Q$, $x^\dagger \in C_q^S = \{y \in S : B_q^1 y \leq h_q^1\}$, which means that there exists a $y_q \in C_q$ such that $x^\dagger = P_S y_q$. By Lemma 5 applied to $K = \{x^\dagger\}$ (where the condition is satisfied because, by Lemma 6, $S = \{x \in \mathbb{R}^{d_x} : A_{J^S,0} x \leq 0\}$), there exists a larger Q such that for all $q \geq Q$, $y_q \in \bar{B}(x^\dagger, \delta)$. Therefore, $\|y_q - x_0\| \leq 2\epsilon$. □

Lemma 4. *Let A be a $d_A \times d_x$ matrix. Let g be nonnegative. Let J^+ denote the subset of $\{1, ..., d_A\}$ such that $g_j > 0$, and let J^0 denote the subset of $\{1, ..., d_A\}$ such that $g_j = 0$. Let S denote the smallest linear subspace containing $C(A_{J^0}, 0) = \{x \in \mathbb{R}^{d_x} : A_{J^0} x \leq 0\}$. Let J^S be the subset of J^0 for which $A_{J^S} \perp S$. Let $J^N = \{1, ..., d_A\} \backslash J^S$. There exists a $\tilde{x} \in S$ such that $a_j' \tilde{x} < g_j$ for all $j \in J^N$.*

Proof of Lemma 4. First, let $M > \max_{j \in J^+} \|a_j\|$, and let $\epsilon \in (0, \min_{j \in J^+}\{g_j\}/M)$. Then, for all $\tilde{x} \in \bar{B}(0, \epsilon)$, $a_j' \tilde{x} < g_j$ for all $j \in J^+$. Also, for every $j \in J^N \cap J^0$, $\{x \in S : a_j' x = 0\}$ defines a subspace of S. We note that for all $j \in J^N \cap J^0$, $\{x \in S : a_j' x = 0\}$ is a proper subset of S, because else j would belong to J^S. By the definition of S, $S \cap C(A_{J^N \cap J^0}, 0)$ is not contained within any of these subspaces. In particular, for each $j \in J^N \cap J^0$, we can find a $\tilde{x}_j$ and a neighborhood, N_j, (relatively open in S) that belongs to $S \cap C(A_{J^N \cap J^0, 0}, 0) \backslash \{x \in S : a_j' x = 0\}$. Indeed, we can consider $j \in J^N \cap J^0$ sequentially, and define each neighborhood to be a subset of the previous one. Therefore, the final $\tilde{x}_j$ must belong to $S \cap C(A_{J^N \cap J^0, 0}, 0)$ and satisfy $a_j' \tilde{x} < 0$ for all $j \in J^N \cap J^0$. Take $\tilde{x} = \lambda \tilde{x}_j$, where $\lambda > 0$ is small enough that $\tilde{x} \in \bar{B}(0, \epsilon)$. Then, $\tilde{x}$ satisfies $a_j' \tilde{x} < g_j$ for all $j \in J^N$. □

Lemma 5. *Let $A_n \to A_0$ and $g_n \to 0$, where $g_n \geq 0$ for all n. Suppose $S = \{x \in \mathbb{R}^{d_x} : A_0 x \leq 0\}$ is a linear subspace of $\mathbb{R}^{d_x}$. Let $S^\perp$ denote the orthogonal subspace to S in $\mathbb{R}^{d_x}$. Let $P_S x$ denote the projection of $x \in \mathbb{R}^{d_x}$ onto S and let $M_S x$ denote $x - P_S x$. Then, for every $K \subseteq S$, compact, and for every $\varepsilon > 0$, we have*

$$\{x \in C(A_n, g_n) : P_S x \in K, \|M_S x\| \geq \varepsilon\} = \emptyset$$

eventually as $n \to \infty$.

Proof of Lemma 5. Suppose that the conclusion of the lemma is not true. Then there exists a sequence $\{x_n \in C(A_n, g_n)\}$ and a subsequence n_m such that $P_S x_{n_m} \in K$ and $\|M_S x_{n_m}\| \geq \varepsilon$ for all $m \geq 1$. Define the unit vector $x_{n_m}^\perp = M_S x_{n_m} / \|M_S x_{n_m}\|$. Then, by the compactness of K and the unit circle, there exists a further subsequence n_q such that $P_S x_{n_q} \to x^S$ and $x^{\perp}_{n_q} \to x^{\perp}$ for some $x^S \in S$ and $x^{\perp} \in S^{\perp}$ as $q \to \infty$.

Because $x^{\perp} \in S^{\perp}$ and $x^{\perp} \neq 0$, we know that $x^{\perp} \notin S = \{x \in \mathbb{R}^{d_x} : A_0 x \leq 0\}$, and therefore there exists a j such that

$$a'_{j,0} x^{\perp} > 0. \tag{56}$$

Also, since $x^S \in S$, $a'_{j,0} x^S \leq 0$. Since S is a linear subspace, we have $a'_{j,0}(-x^S) \leq 0$ as well. This shows that $a'_{j,0} x^S = 0$ (and more generally, $S = \{x \in \mathbb{R}^{d_x} : A_0 x = 0\}$). These imply that

$$a'_{j,0}(x^S + x^{\perp}) > 0. \tag{57}$$

Now consider

$$\begin{aligned} a'_{j,n_q} x_{n_q} - g_{j,n_q} &= a'_{j,n_q} P_S x_{n_q} + a'_{j,n_q} M_S x_{n_q} - g_{j,n_q} \\ &= o(1) + a'_{j,0} x^S + \|M_S x_{n_q}\|(o(1) + a'_{j,0} x^{\perp}) - o(1) \\ &= o(1) + \|M_S x_{n_q}\|(o(1) + a'_{j,0} x^{\perp}). \end{aligned} \tag{58}$$

By (56), $o(1) + a'_{j,0} x^{\perp} > 0$ eventually. Thus, eventually

$$a'_{j,n_q} x_{n_q} - g_{j,n_q} > 0. \tag{59}$$

This contradicts the definition of the sequence x_n which requires that $x_n \in C(A_n, g_n)$ for all n. □

Lemma 6. *Let A be a matrix. Let S be the smallest linear subspace containing $C = C(A, 0)$. Let $J = \{j : a_j \perp S\}$. Then, $S = C(A_J, 0)$.*

Proof of Lemma 6. First, notice that if $x \in S$, then $x \perp a_j$ for all $j \in J$, and therefore, $A_J x = 0$, so $x \in C(A_J, 0)$.

To go the other way, let $x \in C(A_J, 0)$. Lemma 4 implies that there exists a $\tilde{x} \in S$ such that $a'_j \tilde{x} < 0$ for all $j \in J^c$, where $J^c = \{1, ..., d_A\} \backslash J$. Consider $y = x + M\tilde{x}$ for M large. We note that $A_J y = A_J x + M A_J \tilde{x} \leq 0$ since $x \in C(A_J, 0)$ and $\tilde{x} \in S \subseteq C(A_J, 0)$. We also note that for every $j \in J^c$, $a'_j y = a'_j x + M a'_j \tilde{x} \to -\infty$ as M diverges. Thus, there exists an M large enough that $y \in C(A, 0)$. This implies that

$y \in S$ because $C(A, 0) \subseteq S$. This also implies that $x = y - M\tilde{x} \in S$ because S is a linear subspace. $\square$

Lemmas 7 and 8

The following lemmas are used in the proof of Theorem 2. For any $d_A \times d_x$ real-valued matrix A and vector $h \in \mathbb{R}^{d_A}_{+,\infty} := [0, \infty]^{d_A}$, define

$$z(x, A, h) = \arg \min_{z:Az \leq h} \|x - z\|^2,$$

The lemma considers a sequence of $d_A \times d_x$ real-valued matrices $\{A_n\}_{n=1}^{\infty}$ and a sequence of $d_A \times 1$ vectors $h_n \in \mathbb{R}^{d_A}_{+} := [0, \infty)^{d_A}$ such that, as $n \to \infty$, $A_n \to A_0$ and $h_n \to h_0$ for a $d_A \times d_x$ real-valued matrix A_0 and a vector $h_0 \in \mathbb{R}^{d_A}_{+,\infty}$. Also, let $x_n \in \mathbb{R}^{d_x}$ be a sequence of vectors such that $x_n \to x_0 \in \mathbb{R}^{d_x}$ as $n \to \infty$.

Assumption 3. *For any z_0 such that $A_0 z_0 \leq h_0$, there exists a sequence, $\{z_n^*\}$, such that for large enough n,*

$$A_n z_n^* \leq h_n \text{ and } z_n^* \to z_0 \text{ as } n \to \infty. \tag{60}$$

Lemma 7. *Under Assumption 3, $z(x_n, A_n, h_n) \to z(x_0, A_0, h_0)$.*

Proof of Lemma 7. Assumption 3 implies that there exists a sequence, z_n^*, such that, for large enough n,

$$A_n z_n^* \leq h_n \text{ and } z_n^* \to z(x_0, A_0, h_0) \text{ as } n \to \infty. \tag{61}$$

This implies that

$$\|x_n - z(x_n, A_n, h_n)\|^2 \leq \|x_n - z_n^*\|^2 \to \|x_0 - z(x_0, A_0, h_0)\|^2. \tag{62}$$

Taking lim sup on both sides, we get

$$\limsup{}_{n\to\infty} \|x_n - z(x_n, A_n, h_n)\|^2 \leq \|x_0 - z(x_0, A_0, h_0)\|^2. \tag{63}$$

Now note that $z(x_n, A_n, h_n) = \arg\min_{z:A_n x \leq h_n} \|x - z\|^2$. This sequence of minimizers is necessarily bounded because otherwise (63) cannot hold. Thus for any subsequence $\{n_m\}$ there is a further subsequence $\{n_q\}$ such that $z(x_{n_q}, A_{n_q}, h_{n_q}) \to z_\infty$ for some $z_\infty \in \mathbb{R}^{d_x}$. Since $A_{n_q} z(x_{n_q}, A_{n_q}, h_{n_q}) \leq h_{n_q}$, we have $A_0 z_\infty \leq h_0$. Thus,

$$\lim_{q\to\infty} \|x_{n_q} - z(x_{n_q}, A_{n_q}, h_{n_q})\|^2 = \|x_0 - z_\infty\|^2 \geq \|x_0 - z(x_0, A_0, h_0)\|^2. \tag{64}$$

Since the subsequence is arbitrary, this implies that

$$\text{liminf}_{n\to\infty} \|x_n - z(x_n, A_n, h_n)\|^2 \geq \|x_0 - z(x_0, A_0, h_0)\|^2. \tag{65}$$

Combining (63) and (65), we have $\lim_{n\to\infty} \|x_n - z(x_n, A_n, h_n)\|^2 = \|x_0 - z(x_0, A_0, h_0)\|^2$. This, (64), and the uniqueness of $\arg\min_{z:A_0 z \leq h_0} \|x_0 - z\|^2$ together imply that

$$z(x_n, A_n, h_n) \to z_\infty = z(x_0, A_0, h_0) \text{ as } n \to \infty,$$

proving the lemma. □

Lemma 8. *If for all $J \subseteq \{1, \ldots, d_A\}$, $r(A_{J,n}) = r(A_{J,0})$ for all n, then Assumption 3 holds.*

Proof of Lemma 8. Let $a'_{j,0}$ denote the jth row of A_0 and let $a'_{j,n}$ denote the jth row of A_n. Let $J_0 = \{j = 1, \ldots, d_A : a'_{j,0} z_0 = h_{j,0}\}$. If $J_0 = \emptyset$, then $z_n^* = z_0$ satisfies the requirement by $A_n \to A_0$ and $h_n \to h_0$. If $J_0 \neq \emptyset$ but $r(A_{J_0,0}) = 0$, then $a_{j,0} = \mathbf{0}$ for all $j \in J_0$, which implies that $a_{j,n} = \mathbf{0}$ for all $j \in J_0$ by the rank condition stated in the lemma. Then, we can again let $z_n^* = z_0$ and $a'_{j,n} z_n^* = 0 \leq h_{j,n}$ for all $j \in J_0$. Again, $\{z_n^*\}$ satisfies the requirement due to $A_n \to A_0$ and $h_n \to h_0$.

Now suppose that $r(A_{J_0,0}) > 0$. The key for the next step is to partition J_0 into two subsets J_0^* and J_0^o. We require the partition to satisfy the following conditions:

(i) J_0^* contains $r(A_{J_0,0})$ elements such that $\{a_{j^*,0} : j^* \in J_0^*\}$ has full rank, and for any element in $j^o \in J_0^o$, there exists a unique linear representation $a_{j^o,0} = \sum_{j^* \in J_0^*} w_{j^o,j^*} a_{j^*,0}$, where $w_{j^o,j^*} : j^* \in J_0^*$ are real-valued weights.

(ii) The linear representation satisfies: for any $j^* \in J_0^*$ and $j^o \in J_0^o$ such that $w_{j^o,j^*} \neq 0$, we have $h_{j^*,0} \leq h_{j^o,0}$.

Such a partition always exists. To see why, note that the existence of a partition satisfying (i) is guaranteed by the fact that $r(A_{J_0,0}) = r(\{a_{j,0} : j \in J_0\})$. The number of partitions satisfying (i) is finite because J_0 is a finite set. If we choose the partition to be one that minimizes $\sum_{j^* \in J_0^*} h_{j^*,0}$ among those satisfying (i), then the chosen partition also satisfies (ii).

We note that for all n, $r(A_{J_0^*,n}) = r(A_{J_0})$ implies that for every $j^o \in J_0^o$ and $j^* \in J_0^*$, there exist weights, $w_{j^o,j^*,n}$, such that

$$a_{j^o,n} = \sum_{j^* \in J_0^*} w_{j^o,j^*,n} a_{j^*,n}.$$

Furthermore, we know that if $w_{j^o,j^*,n} \neq 0$, then $w_{j^o,j^*} \neq 0$. This follows because, otherwise, we would have

$$r(A_{\{j^o\} \cup (J_0^* \setminus \{j^*\}),n}) > r(A_{(J_0^* \setminus \{j^*\}),n}) = r(A_{(J_0^* \setminus \{j^*\}),0}) = r(A_{\{j^o\} \cup (J_0^* \setminus \{j^*\}),0}),$$

contradicting the assumed rank condition.

Let $A_{J_0^*,0}$ denote the submatrix of A_0 formed by the rows selected by J_0^*, and let $A_{J_0^*,n}$, $h_{J_0^*,0}$, and $h_{J_0^*,n}$ be defined analogously. Now let D be a $(d_x - |J_0|) \times d_x$ matrix, the rows of which form an orthonormal basis for the orthogonal complement of the space spanned by $\{a_{j,0} : j \in J_0^*\}$. Then the matrix $\begin{pmatrix} A_{J_0^*,0} \\ D \end{pmatrix}$ is invertible, which implies that the matrix $\begin{pmatrix} A_{J_0^*,n} \\ D \end{pmatrix}$ is invertible for large enough n. Let $h^{\wedge}_{J_0^*,n} = \min(h_{J_0^*,n}, h_{J_0^*,0})$ where the minimum is taken element by element. Let

$$z_n^{\dagger} = \begin{pmatrix} A_{J_0^*,n} \\ D \end{pmatrix}^{-1} \begin{pmatrix} h^{\wedge}_{J_0^*,n} \\ Dz_0 \end{pmatrix}.$$

It is easy to verify that

$$z_n^{\dagger} \to \begin{pmatrix} A_{J_0^*,0} \\ D \end{pmatrix}^{-1} \begin{pmatrix} h_{J_0^*,0} \\ Dz_0 \end{pmatrix} = z_0, \text{ and} \tag{66}$$

$$A_{J_0^*,n} z_n^{\dagger} = h^{\wedge}_{J_0^*,n} \leq h_{J_0^*,n}. \tag{67}$$

If $a'_{j,n} z_n^{\dagger} \leq h_{j,n}$ for all $j \in J_0^o$ for large enough n, then (61) holds with $z_n^* = z_n^{\dagger}$ and we are done. Otherwise, let

$$\lambda_n = \begin{cases} \min\left\{1, \min_{j \in J_0^o : h_{j,0} > 0} \frac{h_{j,n}}{a'_{j,n} z_n^{\dagger}}\right\} & \text{if } \{j \in J_0^o : h_{j,0} > 0\} \neq \emptyset \\ 1 & \text{otherwise} \end{cases}.$$

This is well-defined for large enough n since $a'_{j,n} z_n^{\dagger} \to a'_{j,0} z_0 = h_{j,0}$ and thus $a'_{j,n} z_n^{\dagger} \neq 0$

for large enough n. Also, by definition $\lambda_n \leq 1$, and

$$\lambda_n \to \min_{j \in J_0^o : h_{j,0} > 0} \frac{h_{j,0}}{a_{j,0}' z_0} = 1. \tag{68}$$

Now let

$$z_n^* = \lambda_n z_n^\dagger. \tag{69}$$

Then for any $j \in J_0^o$ such that $h_{j,0} > 0$, we have

$$a_{j,n}' z_n^* \leq h_{j,n}. \tag{70}$$

For any $j \in J_0^o$ such that $h_{j,0} = 0$, we have

$$\begin{aligned} a_{j,n}' z_n^* &= \lambda_n \sum_{j^* \in J_0^*} w_{j,j^*,n} a_{j^*,n}' z_n^\dagger \\ &= \lambda_n \sum_{j^* \in J_0^*} w_{j,j^*,n} \min(h_{j^*,n}, h_{j^*,0}) \\ &= 0 \leq h_{j,n}, \end{aligned} \tag{71}$$

where the first equality follows by the definition of the weights, $w_{j,j^*,n}$, the second equality follows from the definition of $z_n^\dagger$, the third equality follows because, if $w_{j,j^*,n} \neq 0$, then $w_{j,j^*} \neq 0$, and therefore $0 \leq \min(h_{j^*,n}, h_{j^*,0}) \leq h_{j^*,0} \leq h_{j,0} = 0$ by property (ii) of the partition.

Equations (66), (68), and (69) together imply that $z_n^* \to z_0$. This also implies that, for all $j \notin J_0$, $a_{j,n}' z_n^* - h_{j,n} \to a_{j,0}' z_0 - h_{j,0} < 0$ and thus, for large enough n,

$$A_{\{1,\dots,d_A\} \setminus J_0, n} z_n^* < h_{\{1,\dots,d_A\} \setminus J_0, n}.$$

This combined with equations (67), $\lambda_n \leq 1$, and (69)-(71) implies that $A_n z_n^* \leq h_n$. Therefore, $\{z_n^*\}$ satisfies the requirement and the lemma is proved. □

References

Andrews, D. and Barwick, P. (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica*, 80:2805–2862.

Andrews, D., Cheng, X., and Guggenberger, P. (2019). Generic results for establishing the asymptotic size of confidence sets and tests. Cowles Foundation Discussion Paper No. 1813.

Andrews, D. and Guggenberger, P. (2009). Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory*, 25:669–709.

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

Canay, I. A. and Shaikh, A. (2017). Practical and theoretical advances for inference in partially identified models. In B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson (Eds) *Advances in Economics and Econometrics*: Volume 2: Eleventh World Congress, (Econometric Society Monographs, pp. 271-306). Cambridge University Press.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25:573–578.

Kodde, D. and Palm, F. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, 54:1243–1248.

Kudo, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, 50:403–418.

Luenberger, D. (1969). *Optimization by Vector Space Methods*. John Wiley & Sons, Inc.

Perlman, M. D. and Wu, L. (2006). Some improved tests for multivariate one-sided hypotheses. *Metrika*, 64:23–39.

Rockafellar, R. (1970). *Convex Analysis*. Princeton University Press.

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

Rosen, A. (2008). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics*, 146:107–117.

Shi, X. and Shum, M. (2015). Simple two-stage inference for a class of partially identified models. *Econometric Theory*, 31:493–520.

Silvapulle, M. and Sen, P. (2004). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley-Interscience.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer.

Wolak, F. (1987). An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association*, 82:782–793.

Wolak, F. (1991). The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica*, 59:981–995.