Econometrica Supplementary Material

SUPPLEMENT TO “INFERENCE FOR PARAMETERS DEFINED BY MOMENT INEQUALITIES: A RECOMMENDED MOMENT SELECTION PROCEDURE”

(*Econometrica*, Vol. 80, No. 6, November 2012, 2805–2826)

BY DONALD W. K. ANDREWS AND PANLE JIA BARWICK[1]

CONTENTS

[1]Andrews gratefully acknowledges the research support of the National Science Foundation via Grants SES-0751517 and SES-1058376. The authors thank Steve Berry for numerous discussions and comments. The authors also thank a co-editor, three referees, Michael Wolf, and Giovanni Campiani for very helpful comments.

 DOI: 10.3982/ECTA8166

S1. INTRODUCTION

THIS IS A SUPPLEMENT to the paper by Andrews and Barwick. The latter paper is referred to hereafter as AB1. The contents of this supplement are summarized as follows.

Sections S2–S5 provide the asymptotic results upon which AB1 is based.

Section S2 specifies the model considered, which allows for both moment inequalities and equalities (whereas AB1 considers only moment inequalities).

Section S3 defines the class of test statistics that are considered.

Section S4 defines in detail the class of refined moment selection (RMS) critical values that are introduced in AB1, gives the basic idea behind RMS critical values, defines data-dependent tuning parameters $\widehat{\kappa}$ and data-dependent size-correction factors $\widehat{\eta}$, and discusses plug-in asymptotic (PA) critical values.

Section S5 establishes that RMS CS's have correct asymptotic size (defined in a uniform sense), derives the asymptotic power of RMS tests against local alternatives, discusses an asymptotic average power criterion for comparing RMS tests, and discusses the unidimensional asymptotic power envelope.

Section S6 provides numerical results supplemental to those reported in AB1. Section S6.1 contains additional results that assess the performance of the data-dependent method for choosing $\widehat{\kappa}$ and $\widehat{\eta}$ for the AQLR/t-test/κauto test. Section S6.2 discusses the determination of the recommended adjustment constant $\varepsilon = .012$ for the recommended AQLR test statistic. Section S6.3 considers the case where the sample moments have a singular asymptotic correlation matrix. It provides comparisons of several tests based on their asymptotic average power, finite-sample maximum null rejection probabilities (MNRP's), and finite-sample average power. Section S6.4 provides a table of the κ values that maximize asymptotic average power (i.e., the best κ values), which are used in the construction of Table II. Section S6.4 also provides a table of the asymptotic MNRP's (which are used for size-correction) of the RMS tests that appear in Table II when no size-correction factor is employed (i.e., $\eta = 0$). Section S6.5 is similar to Section 4, which compares the asymptotic power of various RMS tests, except that it considers 19 correlation matrices Ω (rather than 3) but fewer tests. Section S6.6 compares several generalized moment selection (GMS) and RMS tests, where the GMS tests are based on non-data-dependent tuning parameters κ and no size-correction factors η. Section S6.7 gives asymptotic MNRP and power results for some tests that are not considered in AB1. Section S6.8 discusses the relative computation times of the asymptotic normal and bootstrap versions of the AQLR/t-test/κauto and MMM/t-test/$\kappa = 2.35$ tests. Section S6.9 provides information on the magnitude of the (random) RMS critical values for the recommended AQLR/t-test/κauto test.

Section S7 provides details concerning the numerical results reported in AB1 and in Section S6 herein. Section S7.1 provides the μ vectors used in AB1 (which define the alternatives over which asymptotic and finite-sample average power is computed). Section S7.2 describes some details concerning the assessment of the properties of the automatic method of choosing κ. Section S7.3 discusses the determination and computation of the asymptotic power envelope. Section S7.4 discusses the computation of the κ values that maximize asymptotic average power that are reported in Table II. Sections S7.5 and S7.6 describe the numerical computation of $\eta_2(p)$, which is part of the recommended size-correction function $\eta(\cdot)$. Section S7.6 also describes how the maximum over μ vectors in the null is computed for the finite-sample results.

Section S8 describes the GAUSS computer programs that were used to compute the numerical results.

Section S9 gives an alternative parametrization of the moment inequality/equality model to that given in AB1 (that is conducive to the calculation of the uniform asymptotic properties of CS's and tests) and provides proofs of the results given in Section S5.

Throughout, we use the following notation. Let $R_+ = \{x \in R : x \geq 0\}$, $R_{++} = \{x \in R : x > 0\}$, $R_{+,\infty} = R_+ \cup \{+\infty\}$, $R_{[+\infty]} = R \cup \{+\infty\}$, $R_{[\pm\infty]} = R \cup \{\pm\infty\}$, $K^p = K \times \cdots \times K$ (with p copies) for any set K, and $\infty^p = (+\infty, \ldots, +\infty)'$ (with p copies). All limits are as $n \to \infty$ unless specified otherwise. Let d.f. abbreviate distribution function, p.d. abbreviate positive definite, $\mathrm{cl}(\Psi)$ denote the closure of a set Ψ, and 0_v denote a v-vector of zeros.

S2. MOMENT INEQUALITY/EQUALITY MODEL

For brevity, the model considered in AB1 only allows for moment inequalities. Here we consider a more general model that allows for both inequalities and equalities. The moment inequality/equality model is as follows. The true value θ_0 ($\in \Theta \subset R^d$) is assumed to satisfy the moment conditions

$$\begin{aligned} E_{F_0} m_j(W_i, \theta_0) &\geq 0 \quad \text{for} \quad j = 1, \ldots, p \\ &= 0 \quad \text{for} \quad j = p+1, \ldots, p+v, \end{aligned} \tag{S2.1}$$

where $\{m_j(\cdot, \theta) : j = 1, \ldots, k\}$ are known real-valued moment functions, $k = p + v$, and $\{W_i : i \geq 1\}$ are i.i.d. or stationary random vectors with joint distribution F_0. Either p or v may be zero. The observed sample is $\{W_i : i \leq n\}$. The true value θ_0 is not necessarily point identified.

We are interested in tests and confidence sets (CS's) for the true value θ_0.

Generic values of the parameters are denoted (θ, F). For the case of i.i.d. observations, the parameter space $\mathcal{F}$ for (θ, F) is the set of all (θ, F) that satisfy

$$\text{(S2.2)} \quad \begin{aligned} &\text{(i)} && \theta \in \Theta, \\ &\text{(ii)} && E_F m_j(W_i, \theta) \geq 0 \quad \text{for} \quad j = 1, \dots, p, \\ &\text{(iii)} && E_F m_j(W_i, \theta) = 0 \quad \text{for} \quad j = p+1, \dots, k, \\ &\text{(iv)} && \{W_i : i \geq 1\} \quad \text{are i.i.d. under } F, \\ &\text{(v)} && \sigma^2_{F,j}(\theta) = \text{Var}_F\big(m_j(W_i, \theta)\big) > 0, \\ &\text{(vi)} && \text{Corr}_F\big(m(W_i, \theta)\big) \in \Psi, \\ &\text{(vii)} && E_F\big|m_j(W_i, \theta)/\sigma_{F,j}(\theta)\big|^{2+\delta} \leq M \quad \text{for} \quad j = 1, \dots, k, \end{aligned}$$

where $\text{Var}_F(\cdot)$ and $\text{Corr}_F(\cdot)$ denote variance and correlation matrices, respectively, when F is the true distribution, Ψ is the parameter space for $k \times k$ correlation matrices specified at the end of Section S3, and $M < \infty$ and $\delta > 0$ are constants.

The asymptotic results apply to the case of dependent observations. We specify $\mathcal{F}$ for dependent observations in Section S9. The asymptotic results also apply when the moment functions in (S2.1) depend on a parameter τ, that is, when they are of the form $\{m_j(W_i, \theta, \tau) : j \leq k\}$, and a preliminary consistent and asymptotically normal estimator $\widehat{\tau}_n(\theta_0)$ of τ exists (where θ_0 is the true value of θ). The existence of such an estimator requires that τ is point identified given θ_0. In this case, the sample moment functions take the form $\overline{m}_{n,j}(\theta) = \overline{m}_{n,j}(\theta, \widehat{\tau}_n(\theta))$ $(= n^{-1}\sum_{i=1}^n m_j(W_i, \theta, \widehat{\tau}_n(\theta)))$. The asymptotic variance of $n^{1/2}\overline{m}_{n,j}(\theta)$ typically is affected by the estimation of τ and is defined accordingly. Nevertheless, all of the asymptotic results given below hold in this case using the definition of $\mathcal{F}$ given in Section S9 below with the definitions of $m_j(W_i, \theta)$ and $\overline{m}_{n,j}(\theta)$ changed suitably, as described there.

We consider a confidence set obtained by inverting a test. The test is based on a test statistic $T_n(\theta_0)$ for testing $H_0 : \theta = \theta_0$. The nominal level $1 - \alpha$ CS for θ is

$$\text{(S2.3)} \qquad \text{CS}_n = \big\{\theta \in \Theta : T_n(\theta) \leq c_n(\theta)\big\},$$

where $c_n(\theta)$ is a data-dependent critical value.[2] In other words, the confidence set includes all parameter values θ for which one does not reject the null hypothesis that θ is the true value.

[2]When θ is in the interior of the identified set, it may be the case that $T_n(\theta) = 0$ and $c_n(\theta) = 0$. In consequence, it is important that the inequality in the definition of CS_n is $\leq$, not $<$.

S3. TEST STATISTICS

In this section, we define the test statistics $T_n(\theta)$ that we consider. The statistic $T_n(\theta)$ is of the form

$$\text{(S3.1)} \quad T_n(\theta) = S\big(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)\big), \quad \text{where}$$
$$\overline{m}_n(\theta) = \big(\overline{m}_{n,1}(\theta), \dots, \overline{m}_{n,k}(\theta)\big)',$$
$$\overline{m}_{n,j}(\theta) = n^{-1}\sum_{i=1}^{n} m_j(W_i, \theta) \quad \text{for} \quad j \le k,$$

$\widehat{\Sigma}_n(\theta)$ is a $k \times k$ variance matrix estimator defined below, S is a real function on $(R^p_{[+\infty]} \times R^v) \times \mathcal{V}_{k\times k}$, and $\mathcal{V}_{k\times k}$ is the space of $k \times k$ variance matrices. (The set $R^p_{[+\infty]} \times R^v$ contains k-vectors whose first p elements are either real or $+\infty$ and whose last v elements are real.)

The estimator $\widehat{\Sigma}_n(\theta)$ is an estimator of the asymptotic variance matrix of the sample moments $n^{1/2}\overline{m}_n(\theta)$. When the observations are i.i.d. and no parameter τ appears,

$$\text{(S3.2)} \quad \widehat{\Sigma}_n(\theta) = n^{-1}\sum_{i=1}^{n}\big(m(W_i, \theta) - \overline{m}_n(\theta)\big)\big(m(W_i, \theta) - \overline{m}_n(\theta)\big)', \quad \text{where}$$
$$m(W_i, \theta) = \big(m_1(W_i, \theta), \dots, m_k(W_i, \theta)\big)'.$$

The correlation matrix $\widehat{\Omega}_n(\theta)$ that corresponds to $\widehat{\Sigma}_n(\theta)$ is defined by

$$\text{(S3.3)} \quad \widehat{\Omega}_n(\theta) = \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta)\widehat{D}_n^{-1/2}(\theta), \quad \text{where} \quad \widehat{D}_n(\theta) = \mathrm{Diag}\big(\widehat{\Sigma}_n(\theta)\big)$$

and $\mathrm{Diag}(\Sigma)$ denotes the diagonal matrix based on the matrix Σ.

With temporally dependent observations or when a preliminary estimator of a parameter τ appears, a different definition of $\widehat{\Sigma}_n(\theta)$ often is required; see Section S9. For example, with dependent observations, a heteroskedasticity and autocorrelation consistent (HAC) estimator may be required.

We now define the leading examples of the test statistic function S. The first is the modified method of moments (MMM) test function S_1 defined by

$$\text{(S3.4)} \quad S_1(m, \Sigma) = \sum_{j=1}^{p}[m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v}(m_j/\sigma_j)^2, \quad \text{where}$$
$$[x]_- = \begin{cases} x, & \text{if } x < 0, \\ 0, & \text{if } x \ge 0, \end{cases} \qquad m = (m_1, \dots, m_k)',$$

and σ_j^2 is the jth diagonal element of Σ. AB1 lists papers in the literature that consider this test statistic and the other subsequent test statistics.[3]

The second function S is the quasi-likelihood ratio (QLR) test function S_2 defined by

$$S_2(m,\Sigma)=\inf_{t=(t_1,0_v):t_1\in R^p_{+,\infty}}(m-t)'\Sigma^{-1}(m-t). \tag{S3.5}$$

The origin of the QLR S function is as follows. Suppose one replaces m in (S3.5) by a data vector $X\in R^k$ that has a known $k\times k$ variance matrix Σ. Then the resulting QLR statistic is the likelihood ratio statistic for the model with $X\sim N(\mu,\Sigma)$, $\mu=(\mu_1',\mu_2')'\in R^p\times R^v=R^k$, the null hypothesis $H_0^*:\mu_1\ge 0_p$ and $\mu_2=0_v$, and the alternative hypothesis $H_1^*:\mu_1\not\ge 0_p$ and/or $\mu_2\ne 0_v$. The QLR statistic has been considered in many papers on tests of inequality constraints; for example, see Kudo (1963) and Silvapulle and Sen (2005, Sec. 3.8). In the moment inequality literature, it has been considered by Rosen (2008), Andrews and Guggenberger (2009) (AG), and Andrews and Soares (2010) (AS).

Note that under the null and local alternative hypotheses, GEL test statistics behave asymptotically (to the first order) the same as the QLR statistic $T_n(\theta)$ based on S_2 (see Sections 8.1 and 10.3 in AG, Section 10.1 in AS, and Canay (2010)). Although GEL statistics are not of the form given in (S3.1), the results of the present paper (viz., Theorems 1 and 3 below) hold for such statistics under the assumptions given in AG provided the class of moment condition correlation matrices consists of matrices whose determinants are bounded away from zero.

Next we consider an adjusted QLR (AQLR) test function denoted S_{2A}, which is the recommended S function in AB1. It has the property that its weight matrix (whose inverse appears in the quadratic form) is nonsingular even if the estimator of the asymptotic variance matrix of the moment conditions is singular. It is defined by

$$S_{2A}(m,\Sigma)=\inf_{t=(t_1,0_v):t_1\in R^p_{+,\infty}}(m-t)'\widetilde{\Sigma}_\Sigma^{-1}(m-t),\quad\text{where} \tag{S3.6}$$
$$\widetilde{\Sigma}_\Sigma=\Sigma+\max\{\varepsilon-\det(\Omega_\Sigma),0\}D_\Sigma,$$
$$D_\Sigma=\mathrm{Diag}(\Sigma),\quad \Omega_\Sigma=D_\Sigma^{-1/2}\Sigma D_\Sigma^{-1/2},\quad\text{and}\quad \varepsilon>0.$$

Note that the adjustment to the matrix Σ is designed so that $\widetilde{\Sigma}_\Sigma$ is equivariant to scale changes in the moment functions. Based on the results in Section S6.3, the recommended choice of ε for S_{2A} is $\varepsilon=.012$.

[3]Several papers in the literature use a variant of S_1 that is not invariant to rescaling of the moment functions (i.e., with $\sigma_j=1$ for all j), which is not desirable in terms of the power of the resulting test.

The function S_3 is a function that directs power against alternatives with p_1 ($< p$) moment inequalities violated. The test function S_3 is defined by

$$(S3.7)\qquad S_3(m,\Sigma)=\sum_{j=1}^{p_1}[m_{(j)}/\sigma_{(j)}]_-^2+\sum_{j=p+1}^{p+v}(m_j/\sigma_j)^2,$$

where $[m_{(j)}/\sigma_{(j)}]_-^2$ denotes the jth largest value among $\{[m_\ell/\sigma_\ell]_-^2 : \ell=1,\dots,p\}$ and $p_1 < p$ is some specified integer.[4,5]

The asymptotic results given in Section S5 hold for all functions S that satisfy the following assumption.

ASSUMPTION S: (a) *$S(m,\Sigma)=S(Dm,D\Sigma D)$ for all $m\in R^k$, $\Sigma\in R^{k\times k}$, and p.d. diagonal $D\in R^{k\times k}$.*

(b) *$S(m,\Omega)\ge 0$ for all $m\in R^k$ and $\Omega\in\Psi$.*

(c) *$S(m,\Omega)$ is continuous at all $m\in R^p_{[+\infty]}\times R^v$ and $\Omega\in\Psi$.*[6]

(d) *$S(m,\Omega)>0$ if and only if $m_j<0$ for some $j=1,\dots,p$ or $m_j\neq 0$ for some $j=p+1,\dots,k$, where $m=(m_1,\dots,m_k)'$ and $\Omega\in\Psi$.*

(e) *For all $\ell\in R^p_{[+\infty]}\times R^v$, all $\Omega\in\Psi$, and $Z\sim N(0_k,\Omega)$, the d.f. of $S(Z+\ell,\Omega)$ at x is* (i) *continuous for $x>0$ and* (ii) *unless $v=0$ and $\ell=\infty^p$, strictly increasing for $x>0$.*

In Assumption S, the set Ψ is as in condition (vi) of (S2.2) when the observations are i.i.d. and no preliminary estimator of a parameter τ appears. Otherwise, Ψ is the parameter space for the correlation matrix of the asymptotic distribution of $n^{1/2}\overline{m}_n(\theta)$ under (θ,F), denoted $\mathrm{AsyCorr}_F(n^{1/2}\overline{m}_n(\theta))$.[7]

The functions S_1, S_{2A}, and S_3 satisfy Assumption S for any choice of Ψ. The function S_2 satisfies Assumption S provided the determinants of the correlation matrices in Ψ are bounded away from zero.[8]

[4]When constructing a CS, a natural choice for p_1 is the dimension d of θ; see Section S5.3 below.

[5]With the functions S_1, S_{2A}, and S_3, the parameter space Ψ for the correlation matrices in Assumption S and in condition (vi) of (S2.2) can be any nonempty subset of the set Ψ_1 of all $k\times k$ correlation matrices. With the function S_2, the asymptotic results below require that the correlation matrices in Ψ have determinants bounded away from zero because Σ^{-1} appears in the definition of S_2.

[6]Let $B\subset R^w$. We say that a real function G on $R^p_{[+\infty]}\times B$ is continuous at $x\in R^p_{[+\infty]}\times B$ if $y\to x$ for $y\in R^p_{[+\infty]}\times B$ implies that $G(y)\to G(x)$. In Assumption S(c), $S(m,\Omega)$ is viewed as a function with domain Ψ_1.

[7]More specifically, for dependent observations or when a preliminary estimator of a parameter τ appears, Ψ is as in condition (v) of (S9.2) in Section S9.

[8]For the functions S_1–S_3, see Lemma 1 of AG for a proof that Assumptions S(a)–(d) hold and see AS for a proof that Assumption S(e) holds. The proof for S_{2A} is the same as that for S_2 with $\widetilde{\Sigma}_\Sigma$ in place of Σ. By construction, $\widetilde{\Sigma}_\Sigma$ has a determinant that is bounded away from zero even if the latter property fails for Σ.

S4. REFINED MOMENT SELECTION

This section is concerned with critical values for use with the test statistics introduced in Section S3. We proceed in four steps. First, we explain the idea behind moment selection critical values and discuss a tuning parameter $\widehat{\kappa}$ that determines the extent of the moment selection. Second, we introduce a function φ that helps one to select "relevant" moment inequalities. Third, we define the RMS critical value. Last, we specify a size-correction factor $\widehat{\eta}$ that delivers correct asymptotic size even when $\widehat{\kappa}$ does not diverge to infinity. Because the CS's defined in (S2.3) are obtained by inverting tests, we discuss both tests and CS's below.

S4.1. *Basic Idea and Tuning Parameter* $\widehat{\kappa}$

The idea behind *generalized moment selection* and *refined moment selection* is to use the data to determine whether a given moment inequality is satisfied and is far from being an equality. If so, one takes the critical value to be smaller than it would be if all moment inequalities were binding—both under the null and under the alternative.

Under a suitable sequence of null distributions $\{F_n : n \geq 1\}$, the asymptotic null distribution of $T_n(\theta)$ is the distribution of

$$S\big(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0\big), \quad \text{where} \quad Z^* \sim N(0_k, I_k), \tag{S4.1}$$

$h_1 \in R^p_{+,\infty}$, Ω_0 is a $k \times k$ correlation matrix, and both h_1 and Ω_0 typically depend on the true value of θ. The correlation matrix Ω_0 can be consistently estimated. But the $1/n^{1/2}$ local asymptotic mean parameter h_1 cannot be (uniformly) consistently estimated.[9]

A moment selection critical value is the $1-\alpha$ quantile of a data-dependent version of the asymptotic null distribution, $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$, that replaces Ω_0 with a consistent estimator and replaces h_1 with a p-vector in $R^p_{+,\infty}$ whose value depends on a measure of the slackness of the moment inequalities. The measure of slackness is

$$\xi_n(\theta) = \widehat{\kappa}^{-1} n^{1/2} \widehat{D}_n^{-1/2}(\theta) \overline{m}_n(\theta) \in R^k, \tag{S4.2}$$

where $\widehat{\kappa}$ is a tuning parameter (that may depend on θ). For a generalized moment selection (GMS) critical value (as in AS), $\{\widehat{\kappa} = \kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \to \infty$, such as $\kappa_n = (\ln n)^{1/2}$

[9]The asymptotic distribution of the test statistic $T_n(\theta)$ is a discontinuous function of the expected values of the moment inequality functions. This is not a feature of its finite-sample distribution. For this reason, sequences of distributions $\{F_n : n \geq 1\}$ in which these expected values may drift to zero—rather than a fixed distribution F—need to be considered. See Andrews and Guggenberger (2009) for details.

The local parameter h_1 cannot be estimated consistently because doing so requires an estimator of the mean $h_1/n^{1/2}$ that is consistent at rate $o_p(n^{-1/2})$, which is not possible.

or $\kappa_n = (2\ln\ln n)^{1/2}$. In contrast, for an RMS critical value, $\widehat{\kappa}$ does not go to infinity as $n \to \infty$ and is data-dependent.

Data dependence of $\widehat{\kappa}$ is obtained by taking $\widehat{\kappa}$ to depend on $\widehat{\Omega}_n(\theta)$,

$$\widehat{\kappa} = \kappa\big(\widehat{\Omega}_n(\theta)\big), \tag{S4.3}$$

where $\kappa(\cdot)$ is a function from Ψ to R_{++}. A suitable choice of function $\kappa(\cdot)$ improves the power properties of the RMS procedure because the asymptotic power of the test depends on the probability limit of $\widehat{\kappa}$ through $\Omega(\theta)$.

We assume that $\kappa(\Omega)$ satisfies the following assumption.

ASSUMPTION κ: (a) $\kappa(\Omega)$ *is continuous at all* $\Omega \in \Psi$. (b) $\kappa(\Omega) > 0$ *for all* $\Omega \in \Psi$.[10]

S4.2. *Moment Selection Function* φ

Next, we discuss the moment selection function φ that determines how nonbinding moment inequalities are detected. Let $\xi_{n,j}(\theta)$, $h_{1,j}$, and $[\Omega_0^{1/2}Z^*]_j$ denote the jth elements of $\xi_n(\theta)$, h_1, and $\Omega_0^{1/2}Z^*$, respectively, for $j = 1, \ldots, p$. When $\xi_{n,j}(\theta)$ is zero or close to zero, this indicates that $h_{1,j}$ is zero or fairly close to zero and the desired replacement of $h_{1,j}$ in $S(\Omega_0^{1/2}Z^* + (h_1, 0_v), \Omega_0)$ is 0. On the other hand, when $\xi_{n,j}(\theta)$ is large, this indicates $h_{1,j}$ is large and the desired replacement of $h_{1,j}$ in $S(\Omega_0^{1/2}Z^* + (h_1, 0_v), \Omega_0)$ is ∞ or some large value.

We replace $h_{1,j}$ in $S(\Omega_0^{1/2}Z^* + (h_1, 0_v), \Omega_0)$ by $\varphi_j(\xi_n(\theta), \widehat{\Omega}_n(\theta))$ for $j = 1, \ldots, p$, where $\varphi_j : (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi \to R_{[\pm\infty]}$ is a function that is chosen to deliver the properties described above. The leading choices for the function φ_j are

$$\varphi_j^{(1)}(\xi, \Omega) = \begin{cases} 0, & \text{if } \xi_j \le 1, \\ \infty, & \text{if } \xi_j > 1, \end{cases} \qquad \varphi_j^{(2)}(\xi, \Omega) = \big[\kappa(\Omega)(\xi_j - 1)\big]_+, \tag{S4.4}$$

$$\varphi_j^{(3)}(\xi, \Omega) = [\xi_j]_+, \quad \varphi_j^{(4)}(\xi, \Omega) = \begin{cases} 0, & \text{if } \xi_j \le 1, \\ \kappa(\Omega)\xi_j, & \text{if } \xi_j > 1, \end{cases} \quad \text{and}$$

$$\varphi_j^{(0)}(\xi, \Omega) = 0$$

for $j = 1, \ldots, p$, where $[x]_+ = \max\{x, 0\}$, and $\kappa(\Omega)$ in $\varphi_j^{(2)}$ and $\varphi_j^{(4)}$ is the same tuning parameter function that appears in (S4.3). Let $\varphi^{(r)}(\xi, \Omega) =$

[10]For simplicity, the recommended function $\kappa(\Omega) = \kappa(\delta(\Omega))$ given in AB1 is constant on intervals of $\delta(\Omega)$ values and has jumps from one interval to the next. Hence, it does not satisfy Assumption κ. However, the function $\kappa(\delta)$ in Table I can be replaced by a continuous linearly interpolated function whose value at the left-hand point in each interval of δ equals the value given in Table I. Such a function satisfies Assumption κ. Numerical calculations show that the grid of δ values in Table I is sufficiently fine that the finite-sample and asymptotic properties of the recommended RMS test are not sensitive to whether the $\kappa(\delta)$ function is linearly interpolated.

$(\varphi_1^{(r)}(\xi,\Omega),\dots,\varphi_p^{(r)}(\xi,\Omega),0,\dots,0)' \in R^p_{[\pm\infty]} \times \{0\}^v$ for $r=1,\dots,4$. Chernozhukov, Hong, and Tamer (2007), AS, and Bugni (2010) consider the function $\varphi^{(1)}$, Hansen (2005) and Canay (2010) consider $\varphi^{(2)}$, AS consider $\varphi^{(3)}$, and Fan and Park (2007) consider $\varphi^{(4)}$.[11]

The function $\varphi^{(1)}$ generates a "moment selection t-test" procedure, which is the recommended φ function in AB1. Note that $\xi_{n,j}(\theta_0)\le 1$ is equivalent to the condition $n^{1/2}\overline{m}_{n,j}(\theta)/\widehat{\sigma}_{n,j}(\theta)\le\widehat{\kappa}$ in AB1.

The functions $\varphi^{(2)}-\varphi^{(4)}$ exhibit less steep rates of increase than $\varphi^{(1)}$ as functions of ξ_j for $j=1,\dots,p$.

For the asymptotic results given below, the only condition needed on the φ_j functions is that they are continuous on a set that has probability 1 under a certain distribution.

ASSUMPTION φ: *For all* $j=1,\dots,p$, *all* $\beta\in R^p_{[+\infty]}\times R^v$, *and all* $\Omega\in\Psi$, $\varphi_j(\xi,\Omega)$ *is continuous at* (ξ,Ω) *for all* $(\xi',0_v')'$ *in a set* $\Xi(\beta,\Omega)\subset R^p_{[+\infty]}\times R^v$ *for which* $P(\kappa^{-1}(\Omega)[\Omega^{1/2}Z^*+\beta]\in\Xi(\beta,\Omega))=1$, *where* $Z^*\sim N(0_k,I_k)$.

The functions φ_j in (S4.4) all satisfy Assumption φ.

The functions $\varphi^{(r)}$ for $r=1,\dots,4$ all exhibit "element-by-element" determination of which moments to "select" because $\varphi_j^{(r)}(\xi,\Omega)$ depends only on (ξ,Ω) through ξ_j. This has significant computational advantages because $\varphi_j^{(r)}(\xi_n(\theta),\widehat{\Omega}_n(\theta))$ is very easy to compute. On the other hand, when $\widehat{\Omega}_n(\theta)$ is nondiagonal, the whole vector $\xi_n(\theta)$ contains information about the magnitude of the population mean of $\overline{m}_n(\theta)$. The function $\varphi^{(5)}$ that is introduced in AS and defined below exploits this information. It is related to the information-criterion-based moment selection criteria (MSC) considered in Andrews (1999) for a different moment selection problem. We refer to $\varphi^{(5)}$ as the modified MSC (MMSC) φ function. It is computationally more expensive than the functions $\varphi^{(1)}$–$\varphi^{(4)}$ considered above.

Define $c=(c_1,\dots,c_k)'$ to be a selection k-vector of 0's and 1's. If $c_j=1$, the jth moment condition is selected; if $c_j=0$, it is not selected. The moment equality functions are always selected, so $c_j=1$ for $j=p+1,\dots,k$. Let $|c|=\sum_{j=1}^k c_j$. For $\xi\in R^p_{[+\infty]}\times R^v_{[\pm\infty]}$, define $c\cdot\xi=(c_1\xi_1,\dots,c_k\xi_k)'\in R^p_{[+\infty]}\times R^v_{[\pm\infty]}$, where $c_j\xi_j=0$ if $c_j=0$ and $\xi_j=\infty$. Let $\mathcal{C}$ denote the parameter space for the selection vectors, for example, $\mathcal{C}=\{0,1\}^p\times\{1\}^v$. Let $\zeta(\cdot)$ be a strictly increasing real function on R_+. Given $(\xi,\Omega)\in(R^p_{[+\infty]}\times R^v_{[\pm\infty]})\times\Psi$, the selection vector $c(\xi,\Omega)\in\mathcal{C}$ that is chosen is the vector in $\mathcal{C}$ that minimizes the MMSC defined by

$$S(-c\cdot\xi,\Omega)-\zeta\big(|c|\big). \tag{S4.5}$$

[11]The function used by Fan and Park (2007) is not exactly equal to $\varphi_j^{(4)}$. Let $\widehat{\sigma}_{n,j}(\theta)$ denote the (j,j) element of $\widehat{\Sigma}_n(\theta)$. The function Fan and Park (2007) use is $\varphi_j^{(4)}(\xi,\Omega)$ with "if $\xi_j\le 1$" replaced by "if $\widehat{\sigma}_{n,j}(\theta)\xi_j\le 1$," and likewise for $>$ in place of $<$. This yields a non-scale-invariant φ_j function, which is not desirable, so we define $\varphi_j^{(4)}$ as is.

The minus sign that appears in the first argument of the S function ensures that a large *positive* value of ξ_j yields a large value of $S(-c\cdot\xi,\Omega)$ when $c_j=1$, as desired. Since $\zeta(\cdot)$ is increasing, $-\zeta(|c|)$ is a bonus term that rewards inclusion of more moments. For $j=1,\dots,p$, define

$$\varphi_j^{(5)}(\xi,\Omega)=\begin{cases}0, & \text{if } c_j(\xi,\Omega)=1,\\ \infty, & \text{if } c_j(\xi,\Omega)=0.\end{cases} \tag{S4.6}$$

The MMSC is analogous to the Bayesian information criterion (BIC) and the Hannan–Quinn information criterion (HQIC) when $\zeta(x)=x$, $\kappa_n=(\log n)^{1/2}$ for BIC, and $\kappa_n=(Q\ln\ln n)^{1/2}$ for some $Q\geq 2$ for HQIC; see AS. Some calculations show that when $\widehat{\Omega}_n(\theta)$ is diagonal, $S=S_1$, S_2, or S_{2A}, and $\zeta(x)=x$, the function $\varphi^{(5)}$ reduces to $\varphi^{(1)}$.

S4.3. *RMS Critical Value $c_n(\theta)$*

The (asymptotic normal) RMS critical value is equal to the $1-\alpha$ quantile of $S(\Omega^{1/2}Z^*+\beta,\Omega)$ evaluated at $\beta=\varphi(\xi_n(\theta),\widehat{\Omega}_n(\theta))$ and $\Omega=\widehat{\Omega}_n(\theta)$ plus a size-correction factor $\widehat{\eta}$. More specifically, given a choice of function

$$\varphi(\xi,\Omega)=\big(\varphi_1(\xi,\Omega),\dots,\varphi_p(\xi,\Omega),0,\dots,0\big)'\in R^p_{[+\infty]}\times\{0\}^v, \tag{S4.7}$$

the replacement for the k-vector $(h_1,0_v)$ in $S(\Omega_0^{1/2}Z^*+(h_1,0_v),\Omega_0)$ is given by

$$\varphi\big(\xi_n(\theta),\widehat{\Omega}_n(\theta)\big). \tag{S4.8}$$

For $Z^*\sim N(0_k,I_k)$ (independent of $\{W_i:i\geq 1\}$) and $\beta\in R^k_{[+\infty]}$, let $q_S(\beta,\Omega)$ denote the $1-\alpha$ quantile of

$$S\big(\Omega^{1/2}Z^*+\beta,\Omega\big). \tag{S4.9}$$

One can compute $q_S(\beta,\Omega)$ by simulating R i.i.d. random variables $\{Z_r^*:r=1,\dots,R\}$ with $Z_r^*\sim N(0_k,I_k)$ and taking $q_S(\beta,\Omega)$ to be the $1-\alpha$ sample quantile of $\{S(\Omega^{1/2}Z_r^*+\beta,\Omega):r=1,\dots,R\}$, where R is large.

The nominal $1-\alpha$ (asymptotic normal) RMS critical value is

$$c_n(\theta)=q_S\Big(\varphi\big(\xi_n(\theta),\widehat{\Omega}_n(\theta)\big),\widehat{\Omega}_n(\theta)\Big)+\eta\big(\widehat{\Omega}_n(\theta)\big), \tag{S4.10}$$

where $\widehat{\eta}=\eta(\widehat{\Omega}_n(\theta))$ is a size-correction factor that is specified in Section S4.4.

The bootstrap RMS critical value is obtained by replacing $q_S(\varphi(\xi_n(\theta),\widehat{\Omega}_n(\theta)),\widehat{\Omega}_n(\theta))$ in (S4.10) by $q_S^*(\varphi(\xi_n(\theta),\widehat{\Omega}_n(\theta)))$, where $q_S^*(\beta)$ is the $1-\alpha$ quantile of $S(\widehat{D}^*_{n,r}(\theta)^{-1/2}m^*_{n,r}(\theta)+\beta,\widehat{\Omega}^*_{n,r}(\theta))$ for $\beta\in R^k_{[+\infty]}$, and $m^*_{n,r}(\theta)$,

$\widehat{D}^*_{n,r}(\theta)$, and $\widehat{\Omega}^*_{n,r}(\theta)$ are bootstrap quantities defined in AB1. The quantity $q^*_S(\varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)))$ can be computed by taking the $1-\alpha$ sample quantile of $\{S(\widehat{D}^*_{n,r}(\theta)^{-1/2}m^*_{n,r}(\theta) + \varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)), \widehat{\Omega}^*_{n,r}(\theta)) : r = 1, \ldots, R\}$.

For the recommended RMS critical value defined in AB1, the asymptotic normal critical value is of the form in (S4.10) with $S = S_{2A}$, $\varphi = \varphi^{(1)}$, and $\eta(\Omega) = \eta_1(\delta(\Omega)) + \eta_2(p)$. The bootstrap critical value uses $q^*_{S_{2A}}(\cdot)$ in place of $q_{S_{2A}}(\cdot, \widehat{\Omega}_n(\theta))$.

S4.4. *Size-Correction Factor* $\widehat{\eta}$

We now discuss the size-correction factor $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$. Such a factor is necessary to deliver correct asymptotic size under asymptotics in which $\widehat{\kappa}$ does not diverge to infinity. This factor can be viewed as giving an asymptotic size refinement to a GMS critical value.

As noted above, we show in the proofs (see Section S9) that under a suitable sequence of true parameters and distributions $\{(\theta_n, F_n) : n \geq 1\}$, $T_n(\theta_n) \to_d S(\Omega^{1/2}Z^* + (h_1, 0_v), \Omega)$ for some $(h_1, \Omega) \in R^p_{+,\infty} \times \Psi$. Furthermore, we show that under such a sequence, the asymptotic coverage probability (CP) of an RMS CS based on a data-dependent tuning parameter $\widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta))$ and a fixed size-correction constant η is

$$\text{(S4.11)}\quad \mathrm{CP}(h_1, \Omega, \eta) = P\Big[S\big(\Omega^{1/2}Z^* + (h_1, 0_v), \Omega\big) \leq q_S\big(\varphi\big(\kappa^{-1}(\Omega)\big[\Omega^{1/2}Z^* + (h_1, 0_v)\big], \Omega\big), \Omega\big) + \eta\Big],$$

where $Z^* \sim N(0_k, I_k)$. (Correspondingly, the null rejection probability of an RMS test with fixed η for testing $H_0 : \theta = \theta_0$ is $1 - \mathrm{CP}(h_1, \Omega, \eta)$.)

We let Δ ($\subset R^p_{+,\infty} \times \Psi$) denote the set of all (h_1, Ω) values that can arise given the model specification $\mathcal{F}$. More precisely, Δ is defined as follows. Let the normalized mean vector and asymptotic correlation matrix of the sample moment functions be denoted by

$$\text{(S4.12)}\quad \gamma_1(\theta, F) = \mathrm{Diag}^{-1/2}\big(\mathrm{AsyVar}_F\big(n^{1/2}\overline{m}_n(\theta)\big)\big)E_F m(W_i, \theta) \geq 0_p,$$
$$\Omega(\theta, F) = \mathrm{AsyCorr}_F\big(n^{1/2}\overline{m}_n(\theta)\big),$$

where $\mathrm{AsyVar}_F(n^{1/2}\overline{m}_n(\theta))$ and $\mathrm{AsyCorr}_F(n^{1/2}\overline{m}_n(\theta))$ denote the variance and correlation matrices, respectively, of the asymptotic distribution of $n^{1/2}\overline{m}_n(\theta)$ when the true parameter is θ and the true distribution is F.[12] Then Δ is defined

[12]For dependent observations and when a preliminary estimator of a parameter τ appears, the parameter space $\mathcal{F}$ of (θ, F) is defined in Section S9.1 such that both $\mathrm{AsyVar}_F(n^{1/2}\overline{m}_n(\theta))$ and $\mathrm{AsyCorr}_F(n^{1/2}\overline{m}_n(\theta))$ exist. These limits equal $\mathrm{Var}_F(m(W_i, \theta))$ and $\mathrm{Corr}_F(m(W_i, \theta))$, respectively, in the case of i.i.d. observations with no preliminary estimator of a parameter τ.

by

(S4.13) $\Delta = \big\{(h_1, \Omega) \in R^p_{+,\infty} \times \mathrm{cl}(\Psi) : \exists$ a subsequence $\{w_n\}$ of $\{n\}$ and a sequence $\big\{(\theta_{w_n}, F_{w_n}) \in \mathcal{F} : n \geq 1\big\}$ with $\gamma_1(\theta_{w_n}, F_{w_n}) \geq 0_p$ and $\Omega(\theta_{w_n}, F_{w_n}) \in \Psi$ for which $w_n^{1/2}\gamma_1(\theta_{w_n}, F_{w_n}) \to h_1$, $\Omega(\theta_{w_n}, F_{w_n}) \to \Omega$, and $\theta_{w_n} \to \theta_*$ for some θ_* in $\mathrm{cl}(\Theta)\big\}$.

Our primary focus is on the standard case in which

(S4.14) $\Delta = R^p_{+,\infty} \times \mathrm{cl}(\Psi)$.

This arises when there are no restrictions on the moment functions beyond the inequality/equality restrictions, and h_1 and Ω are variation-free. Our asymptotic results cover the general case in (S4.13) in which Δ may be restricted, as well as the standard case in (S4.14).

To determine the asymptotic size of an RMS test or CS, it suffices to have $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ satisfy the following assumption

ASSUMPTION η1: *$\eta(\Omega)$ is continuous at all $\Omega \in \Psi$.*[13]

However, for an RMS CS to have asymptotic size greater than or equal to $1-\alpha$, $\eta(\cdot)$ must be chosen to satisfy the first condition that follows. If it also satisfies the second, stronger, condition, then its asymptotic size equals $1-\alpha$. Let $\mathrm{CP}(h_1, \Omega, \eta(\Omega)-) = \lim_{x \downarrow 0} \mathrm{CP}(h_1, \Omega, \eta(\Omega) - x)$.

ASSUMPTION η2: $\inf_{(h_1,\Omega)\in\Delta} \mathrm{CP}(h_1, \Omega, \eta(\Omega)-) \geq 1-\alpha$.

ASSUMPTION η3: (a) $\inf_{(h_1,\Omega)\in\Delta} \mathrm{CP}(h_1, \Omega, \eta(\Omega)) = 1-\alpha$.
(b) $\inf_{(h_1,\Omega)\in\Delta} \mathrm{CP}(h_1, \Omega, \eta(\Omega)-) = \inf_{(h_1,\Omega)\in\Delta} \mathrm{CP}(h_1, \Omega, \eta(\Omega))$.

Assumption η3(b) is a continuity condition that is not restrictive. The left-hand side (l.h.s.) quantity inside the probability in (S4.11) has a d.f. that is continuous and strictly increasing for positive values. The corresponding right-hand side (r.h.s.) quantity is positive. These two quantities are quite different nonlinear functions of the same underlying normal random vector. Hence, they are equal with probability 0, which implies that Assumption η3(b) holds.

The function $\eta(\Omega)$ depends on S, φ, and the tuning parameter function $\kappa(\Omega)$. For notational simplicity, we suppress this dependence. Functions $\eta(\cdot)$ that satisfy Assumption η2 and/or η3 are not uniquely defined. The smallest

[13] An analogous comment to that in footnote 10 also applies to the recommended function $\eta(\cdot)$ given in AB1 and Assumption η1.

function that satisfies Assumption $\eta3$(a), denoted $\eta^*(\Omega)$, exists and is defined as follows. For each $\Omega \in \Psi$, define $\eta^*(\Omega)$ to be the smallest value[14] η for which

$$\inf_{h_1:(h_1,\Omega)\in\Delta} \mathrm{CP}(h_1, \Omega, \eta) = 1 - \alpha. \tag{S4.15}$$

When Δ satisfies (S4.14), the infimum is over $h_1 \in R^p_{+,\infty}$. For purposes of minimizing the probability of false coverage of the CS (or, equivalently, maximizing the power of the tests on which the CS is based), it is desirable to take $\eta(\Omega)$ as close to $\eta^*(\Omega)$ as possible subject to $\eta(\Omega) \geq \eta^*(\Omega)$. For computational tractability and storability, however, it is convenient to use a function $\eta(\cdot)$ that is simpler than $\eta^*(\Omega)$, for example, a function that depends on Ω only through a scalar function of Ω, as with the recommended RMS critical value described in AB1.[15]

S4.5. *Plug-in Asymptotic Critical Values*

We now discuss CS's based on a plug-in asymptotic (PA) critical value. The least favorable asymptotic null distributions of the statistic $T_n(\theta)$ are those for which the moment inequalities hold as equalities. These distributions depend on the correlation matrix Ω of the moment functions. PA critical values are determined by the least favorable asymptotic null distribution for given Ω evaluated at a consistent estimator of Ω. Such critical values have been considered in the literature on multivariate one-sided tests; see Silvapulle and Sen (2005) for references. AG and AS consider them in the context of the moment inequality literature. Rosen (2008) considers variations of PA critical values that make adjustments in the case where it is known that if one moment inequality holds as an equality, then another cannot.[16]

The PA critical value is

$$q_S\big(0_k, \widehat{\Omega}_n(\theta)\big). \tag{S4.16}$$

The PA critical value can be viewed as a special case of an RMS critical value with $\varphi_j(\xi, \Omega) = 0$ for all $j = 1, \ldots, k$ and $\eta(\widehat{\Omega}_n(\theta)) = 0$. This implies that the asymptotic results stated below for RMS CS's and tests also apply to PA CS's and tests.

S5. ASYMPTOTIC RESULTS

This section provides asymptotic results for RMS CS's and tests. It establishes that RMS CS's have correct asymptotic size (defined in a uniform sense),

[14] A smallest value exists because $\mathrm{CP}(h_1, \Omega, \eta)$ is right-continuous in η.

[15] Note that even if $\eta(\Omega) \neq \eta^*(\Omega)$, Assumption $\eta3$(a) still can hold.

[16] This method delivers corrrect asymptotic size in a uniform sense only if when one moment inequality holds as an equality, then the other is strictly bounded away from zero.

derives the asymptotic power of RMS tests against local alternatives, discusses an asymptotic average power criterion for comparing RMS tests, and discusses the unidimensional asymptotic power envelope.

S5.1. *Asymptotic Size*

The exact and asymptotic confidence sizes of an RMS CS are

$$\text{(S5.1)}\qquad \text{ExCS}_n = \inf_{(\theta,F)\in\mathcal{F}} P_F\big(T_n(\theta)\le c_n(\theta)\big) \quad \text{and}$$
$$\text{AsyCS} = \liminf_{n\to\infty} \text{ExCS}_n,$$

respectively. The definition of AsyCS takes the "inf" before the "lim." This builds uniformity over (θ, F) into the definition of AsyCS. Uniformity is required for the asymptotic size to give a good approximation to the finite-sample size of a CS.

Theorems 1 and 3 below apply to i.i.d. observations, in which case $\mathcal{F}$ is defined in (S2.2). They also apply to stationary temporally dependent observations and to cases in which the moment functions depend on a preliminary consistent estimator of a parameter τ, in which cases $\mathcal{F}$ is defined in Section S9.

THEOREM 1: *Suppose Assumptions* S, κ, φ, *and* η1 *hold*, *and* $0<\alpha<1$. *Then the nominal level* $1-\alpha$ *RMS CS based on* S, φ, $\widehat{\kappa}=\kappa(\widehat{\Omega}_n(\theta))$, *and* $\widehat{\eta}=\eta(\widehat{\Omega}_n(\theta))$ *satisfies the following statements*:

(a) $\text{AsyCS}\in[\inf_{(h_1,\Omega)\in\Delta}\text{CP}(h_1,\Omega,\eta(\Omega)-), \inf_{(h_1,\Omega)\in\Delta}\text{CP}(h_1,\Omega,\eta(\Omega))]$.

(b) $\text{AsyCS}\ge 1-\alpha$ *provided Assumption* η2 *holds*.

(c) $\text{AsyCS}= 1-\alpha$ *provided Assumption* η3 *holds*.

COMMENTS: (i) Theorem 1(b) shows that an RMS CS based on a size-correction factor $\widehat{\eta}=\eta(\widehat{\Omega}_n(\theta))$ that satisfies Assumption η2 is asymptotically valid in a uniform sense under asymptotics that do not require $\widehat{\kappa}\to\infty$ as $n\to\infty$. In contrast, the GMS CS introduced in AS requires $\widehat{\kappa}\to\infty$ as $n\to\infty$.

(ii) Theorem 1 holds even if there are restrictions such that one moment inequality cannot hold as an equality if another moment inequality does. Rosen (2008) discussed models in which restrictions of this sort arise.

(iii) Theorem 1 applies to moment conditions based on weak instruments (because the tests considered are of an Anderson–Rubin form).

(iv) Define the asymptotic size of an RMS test of $H_0:\theta=\theta_0$ by

$$\text{(S5.2)}\qquad \text{AsySz}(\theta_0)=\liminf_{n\to\infty}\sup_{(\theta,F)\in\mathcal{F}:\theta=\theta_0} P_F\big(T_n(\theta_0)>c_n(\theta_0)\big).$$

The proof of Theorem 1 shows that under the assumptions in Theorem 1, (a) $\text{AsySz}(\theta_0)\in[1-\inf_{(h_1,\Omega)\in\Delta_0}\text{CP}(h_1,\Omega,\eta(\Omega)), 1-\inf_{(h_1,\Omega)\in\Delta_0}\text{CP}(h_1,\Omega,\eta(\Omega)-)]$, where Δ_0 is defined as Δ is defined in (S4.14) or as in (S4.13) but with the sequence $\{\theta_{w_n} : n \geq 1\}$ replaced by the constant θ_0, (b) $\text{AsySz}(\theta_0) \leq \alpha$ provided Assumption $\eta 2$ holds, and (c) $\text{AsySz}(\theta_0) = \alpha$ provided Assumption $\eta 3$ holds, where Δ in Assumptions $\eta 2$ and $\eta 3$ is replaced by Δ_0. The primary case of interest is when $\Delta_0 = R^p_{+,\infty} \times \text{cl}(\Psi)$, which occurs when there are no restrictions on the moment functions beyond the inequality/equality restrictions, and h_1 and Ω are variation free.

(v) The proofs of Theorem 1 and all other results stated here are provided in Section S9.

S5.2. *Asymptotic Power*

In this section, we compute the asymptotic power of RMS tests against $1/n^{1/2}$ local alternatives. These results have immediate consequences for the length or volume of a CS based on these tests, because the power of a test for a point that is not the true value is the probability that the CS does not include that point. (See Pratt (1961) for an equation that links CS volume and probabilities of false coverage.) We use these results to define tuning parameters $\kappa = \kappa(\Omega)$ and size-correction factors $\eta = \eta(\Omega)$ that maximize average power for a selected set of alternative parameter values. We also use the results to compare different choices of test function S and moment selection function φ in terms of asymptotic average power.

For given θ_0, we consider tests of

$$\text{(S5.3)} \qquad H_0 : E_F m_j(W_i, \theta_0) \geq 0 \quad \text{for} \quad j = 1, \ldots, p, \\ = 0 \quad \text{for} \quad j = p+1, \ldots, k,$$

where F denotes the true distribution of the data. (More precisely, by this we mean H_0: the true $(\theta, F) \in \mathcal{F}$ satisfies $\theta = \theta_0$.) The alternative is $H_1 : H_0$ does not hold.

Let

$$\text{(S5.4)} \qquad \sigma^2_{F,j}(\theta) = \text{AsyVar}_F\big(n^{1/2}\overline{m}_{n,j}(\theta)\big) \quad \text{for} \quad j = 1, \ldots, k, \\ D(\theta, F) = \text{Diag}\big\{\sigma^2_{F,1}(\theta), \ldots, \sigma^2_{F,k}(\theta)\big\}, \\ \Omega(\theta, F) = \text{AsyCorr}_F\big(n^{1/2}\overline{m}_n(\theta)\big).$$

Note that this definition of $\sigma^2_{F,j}(\theta)$ reduces to that given in (S2.2) when the observations are i.i.d. Let $\widehat{\sigma}^2_{n,j}(\theta)$ denote the (j, j) element of $\widehat{\Sigma}_n(\theta)$ for $j = 1, \ldots, k$.

We now introduce the $1/n^{1/2}$ local alternatives. The first two assumptions are the same as in AS. The third assumption is a high-level assumption that allows for dependent observations and sample moment functions that may depend on a preliminary estimator $\widehat{\tau}_n(\theta)$. It is shown to hold automatically with i.i.d. observations when there is no preliminary estimator of a parameter τ.

ASSUMPTION LA1: *The true parameters* $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ *satisfy the following statements*:

(a) $\theta_n = \theta_0 - \lambda n^{-1/2}(1 + o(1))$ *for some* $\lambda \in R^d$ *and* $F_n \to F_0$ *for some* $(\theta_0, F_0) \in \mathcal{F}$.

(b) $n^{1/2} E_{F_n} m_j(W_i, \theta_n)/\sigma_{F_n,j}(\theta_n) \to h_{1,j}$ *for some* $h_{1,j} \in R_{+,\infty}$ *for* $j = 1, \dots, p$.

(c) $\sup_{n \geq 1} E_{F_n} |m_j(W_i, \theta_0)/\sigma_{F_n,j}(\theta_0)|^{2+\delta} < \infty$ *for* $j = 1, \dots, k$ *for some* $\delta > 0$.

ASSUMPTION LA2: *The* $k \times d$ *matrix* $\Pi(\theta, F) = (\partial/\partial\theta')[D^{-1/2}(\theta, F) E_F m(W_i, \theta)]$ *exists and is continuous in* (θ, F) *for all* (θ, F) *in a neighborhood of* (θ_0, F_0).[17]

ASSUMPTION LA3: *The true parameters* $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ *satisfy the following statements*:

(a) $A_n^0 = (A_{n,1}^0, \dots, A_{n,k}^0)' \to_d Z \sim N(0_k, \Omega_0)$ *as* $n \to \infty$, *where* $A_{n,j}^0 = n^{1/2}(\overline{m}_{n,j}(\theta_0) - E_{F_n} m_j(W_i, \theta_0))/\sigma_{F_n,j}(\theta_0)$.

(b) $\widehat{\sigma}_{n,j}(\theta_0)/\sigma_{F_n,j}(\theta_0) \to_p 1$ *as* $n \to \infty$ *for* $j = 1, \dots, k$.

(c) $\widehat{D}_n^{-1/2}(\theta_0)\widehat{\Sigma}_n(\theta_0)\widehat{D}_n^{-1/2}(\theta_0) \to_p \Omega_0$ *as* $n \to \infty$.

When the observations are i.i.d. for each $(\theta, \Omega) \in \mathcal{F}$, Assumption LA3 holds automatically as shown in the following lemma.

ASSUMPTION LA3*: (a) *For each* $n \geq 1$, *the observations* $\{W_i : i \leq n\}$ *are i.i.d. under* $(\theta_n, F_n) \in \mathcal{F}$, (b) $\widehat{\Sigma}_n(\theta)$ *is defined by* (S3.2), *and* (c) *no preliminary estimator of a parameter* τ *appears in the sample moment functions*.

LEMMA 2: *Assumptions* LA1 *and* LA3* *imply Assumption* LA3.

The asymptotic distribution of $T_n(\theta_0)$ under local alternatives depends on the limit of the normalized moment inequality functions when evaluated at the null value θ_0. Under Assumptions LA1 and LA2, it can be shown that

$$\text{(S5.5)} \quad \lim_{n\to\infty} n^{1/2} D^{-1/2}(\theta_0, F_n) E_{F_n} m(W_i, \theta_0) = \mu = (h_1, 0_v) + \Pi_0 \lambda \in R^p_{[+\infty]} \times R^v, \quad \text{where} \quad h_1 = (h_{1,1}, \dots, h_{1,p})' \quad \text{and} \quad \Pi_0 = \Pi(\theta_0, F_0).$$

By definition, if $h_{1,j} = \infty$, then $h_{1,j} + x = \infty$ for any $x \in R$. Let $\Pi_{0,j}$ denote the jth row of Π_0 written as a column d-vector for $j = 1, \dots, k$. Note that $(h_1, 0_v) + \Pi_0\lambda \in R^p_{[+\infty]} \times R^v$. Let $\mu = (\mu_1, \dots, \mu_k)'$. The true distribution F_n is

[17]When a preliminary estimator of a parameter τ appears in the sample moment functions, then in Assumptions LA1 and LA2 and (S5.5), $m_j(W_i, \theta)$ and $m(W_i, \theta)$ are defined to be $m_j(W_i, \theta, \tau_0)$ and $m(W_i, \theta, \tau_0)$, respectively, where τ_0 denotes the true value of the parameter τ under the true distribution F.

in the alternative, not the null (for n large) when $\mu_j = h_{1,j} + \Pi'_{0,j}\lambda < 0$ for some $j = 1, \ldots, p$ or $\Pi'_{0,j}\lambda \neq 0$ for some $j = p+1, \ldots, k$.

For constants $\kappa > 0$ and $\eta \geq 0$, define

$$\text{(S5.6)} \quad \begin{aligned} &\mathrm{AsyPow}(\mu, \Omega, S, \varphi, \kappa, \eta) \\ &\quad = P\big(S\big(\Omega^{1/2}Z^* + \mu, \Omega\big) > q_S\big(\varphi\big(\kappa^{-1}\big[\Omega^{1/2}Z^* + \mu\big], \Omega\big), \Omega\big) + \eta\big), \\ &\mathrm{AsyPow}^-(\mu, \Omega, S, \varphi, \kappa, \eta) = \lim_{x \downarrow 0} \mathrm{AsyPow}(\mu, \Omega, S, \varphi, \kappa, \eta - x), \end{aligned}$$

where $Z^* \sim N(0_k, I_k)$, $\mu \in R^k$, $\Omega \in \Psi$, $\kappa \in R_{++}$, and the functions S, φ, and q_S are as defined in Section S3, (S4.4) or (S4.6), and (S4.9), respectively.[18] Typically, $\mathrm{AsyPow}(\mu, \Omega, S, \varphi, \kappa, \eta) = \mathrm{AsyPow}^-(\mu, \Omega, S, \varphi, \kappa, \eta)$ because the l.h.s. quantity in the probability in (S5.6) is a nonlinear function of a normal random vector that has a continuous and strictly increasing d.f. (unless $v = 0$ and $\mu = \infty^p$, which cannot hold under the alternative hypothesis) and the r.h.s. quantity in the probability in (S5.6) is a quite different nonlinear function of the same normal random vector.

For a sequence of constants $\{\zeta_n : n \geq 1\}$, let $\zeta_n \to [\zeta_{1,\infty}, \zeta_{2,\infty}]$ denote that $\zeta_{1,\infty} \leq \liminf_{n\to\infty} \zeta_n \leq \limsup_{n\to\infty} \zeta_n \leq \zeta_{2,\infty}$.

THEOREM 3: *Under Assumptions* S, κ, φ, η1 *and* LA1–LA3, *the RMS test based on* $S, \varphi, \widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta))$, *and* $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ *satisfies*

$$\begin{aligned} &P_{F_n}\big(T_n(\theta_0) > c_n(\theta_0)\big) \\ &\quad \to \big[\mathrm{AsyPow}\big(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0)\big), \\ &\qquad \mathrm{AsyPow}^-\big(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0)\big)\big], \end{aligned}$$

where $\mu = (h_1, 0_v) + \Pi_0\lambda$.

COMMENTS 1: (i) Theorem 3 provides the $1/n^{1/2}$ local alternative power function of RMS and PA tests. Typically, $\mathrm{AsyPow}(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0)) = \mathrm{AsyPow}^-(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0))$ and the asymptotic local power function is unique for any given (μ, Ω_0).

(ii) The results of Theorem 3 hold under the null and alternative hypotheses.

(iii) For moment conditions based on weak instruments, the results of Theorem 3 still hold. But with weak instruments, RMS and PA tests have power less than or equal to α against $1/n^{1/2}$ local alternatives because $\Pi'_{0,j}\lambda = 0$ for all $j = 1, \ldots, k$.

[18]For some functions φ, such as $\varphi^{(1)}$ and $\varphi^{(4)}$, $\kappa = 0$ is permissible because $\lim_{\kappa\downarrow 0} \varphi(\kappa^{-1}[\Omega^{1/2}Z + \mu], \Omega)$ is well defined. For example, for $\varphi^{(1)}$ and $x \in R$, $\lim_{\kappa\downarrow 0} \varphi(\kappa^{-1}x, \Omega) = 0$ if $x \leq 0$ and $\lim_{\kappa\downarrow 0} \varphi(\kappa^{-1}x, \Omega) = \infty$ if $x > 0$.

S5.3. *Average Power*

RMS tests depend on S, φ, $\kappa(\Omega)$, and $\eta(\Omega)$. We compare the power of RMS tests by comparing their asymptotic average power for a chosen set $\mathcal{M}_k(\Omega)$ of alternative parameter vectors $\mu \in R^k$ for $\Omega \in \Psi$.[19] Let $|\mathcal{M}_k(\Omega)|$ denote the number of elements in $\mathcal{M}_k(\Omega)$. The asymptotic average power of the RMS test based on $(S, \varphi, \kappa, \eta)$ for constants $\kappa > 0$ and $\eta \geq 0$ is

$$\text{(S5.7)} \quad \left|\mathcal{M}_k(\Omega)\right|^{-1} \sum_{\mu \in \mathcal{M}_k(\Omega)} \text{AsyPow}(\mu, \Omega, S, \varphi, \kappa, \eta).$$

We are interested in comparing the (S, φ) functions defined in (S3.4)–(S3.7), (S4.4), and (S4.6) in terms of asymptotic $\mathcal{M}_k(\Omega)$ average power. To do so requires choices of functions $(\kappa(\cdot), \eta(\cdot))$ for each (S, φ). We use the tuning and size-correction functions $\kappa^*(\Omega)$ and $\eta^*(\Omega)$ that are optimal in terms of asymptotic $\mathcal{M}_k(\Omega)$ average power. They are defined as follows. Given Ω and $\kappa > 0$, let $\eta^*(\Omega, \kappa)$ be defined as in (S4.15) with $\Delta = R^p_{+,\infty} \times \text{cl}(\Omega)$ and tuning parameter $\kappa > 0$. The optimal tuning parameter $\kappa^*(\Omega)$ maximizes (S5.7) with η replaced by $\eta^*(\Omega, \kappa)$ over $\kappa > 0$. The optimal size-correction factor then is $\eta^*(\Omega) = \eta^*(\Omega, \kappa^*(\Omega))$ and the test based on $(\kappa^*(\Omega), \eta^*(\Omega))$ has asymptotic size α. (Obviously, $\kappa^*(\cdot)$ and $\eta^*(\cdot)$ depend on (S, φ).)

Given $\eta^*(\Omega)$ and $\kappa^*(\Omega)$, we compare (S, φ) functions by comparing their values of

$$\text{(S5.8)} \quad \left|\mathcal{M}_k(\Omega)\right|^{-1} \sum_{\mu \in \mathcal{M}_k(\Omega)} \text{AsyPow}\big(\mu, \Omega, S, \varphi, \kappa^*(\Omega), \eta^*(\Omega)\big),$$

which depend on Ω.

We are interested in constructing tests that yield CS's that are as small as possible. The boundary of a CS, like the boundary of the identified set, is determined at any given point by the moment inequalities that are binding at that point. The number of binding moment inequalities at a point depends on the dimension, d, of the parameter θ. Typically, the boundary of a confidence set is determined by d (or fewer) moment inequalities; that is, at most d moment inequalities are binding and at least $p - d$ are slack; see Figure 1. (Note that the axes in Figure 1 are θ_1 and θ_2.) In consequence, we specify the sets $\mathcal{M}_k(\Omega)$ considered below to be ones for which most vectors μ have half or more elements positive (since positive elements correspond to nonbinding inequalities), which is suitable for the typical case in which $p \geq 2d$.

[19]As indicated, we allow this set to depend on Ω. The reason is that the power of any test and the asymptotic power envelope depend on Ω. Hence, it is natural to vary the magnitude of $\|\mu\|$ for $\mu \in \mathcal{M}_k(\Omega)$ as Ω varies.

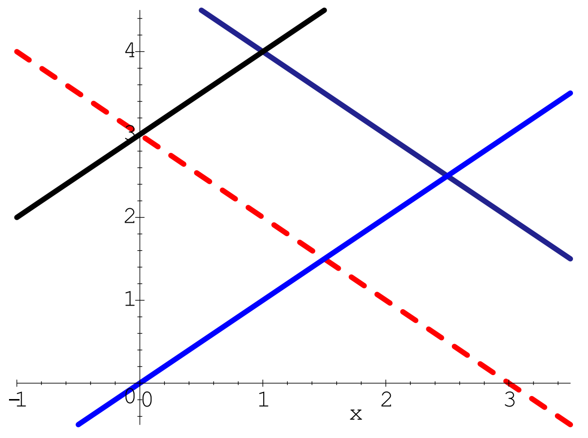

FIGURE 1.—Confidence set for a parameter $\theta \in R^d$ for $d = 2$ based on $p = 4$ moment inequalities.

S5.4. *Asymptotic Power Envelope*

To assess the power performance of RMS tests in an absolute sense, it is of interest to compare their asymptotic power to the asymptotic power envelope. For details on the determination and computation of the latter, see Section S7.3.

We note that the asymptotic power envelope is a "unidirectional" envelope. One does not expect a test that is designed to perform well for multidirectional alternatives to be on, or close to, the unidirectional envelope. This is analogous to the fact that the power of a standard F-test for a p-dimensional restriction with an unrestricted alternative hypothesis in a normal linear regression model is not close to the unidimensional power envelope. For example, for $p = 2, 4, 10$, when the asymptotic power envelope is .75, .80, and .85, respectively, the F test has power .65, .60, and .49, respectively.[20] Clearly, the larger is p, the greater is the difference between the power of a test designed for p-directional alternatives and the unidirectional power envelope.

S6. NUMERICAL RESULTS

This section gives supplemental numerical results to those given in AB1.

[20]These asymptotic power results are obtained by some simple calculations based on the distribution function of the noncentral χ^2 distribution with $p = 1, 2, 4, 10$ degrees of freedom, where the noncentral χ^2 distribution with $p = 1$ degrees of freedom is used for the power envelope calculations.

Section S6.1 describes how the approximately optimal $\kappa(\cdot)$ and $\eta(\cdot)$ functions given in Table I are determined and provides numerical results concerning their properties.[21]

Section S6.2 discusses the determination of the recommended adjustment constant $\varepsilon = .012$ for the recommended AQLR test statistic, which is based on the S_{2A} function.[22]

Section S6.3 considers the case where the sample moments have a singular asymptotic correlation matrix. It provides comparisons of several tests based on their asymptotic average power, finite-sample maximum null rejection probabilities (MNRP's), and finite-sample average power. It also defines the empirical likelihood ratio (ELR) statistic, discusses its computation, and defines the bootstrap employed with the ELR test.

Section S6.4 provides a table of the κ values that maximize asymptotic average power for various tests. These are the κ values that yield the asymptotic power reported in Table II. Section S6.4 also provides a table that is analogous to Table II but reports asymptotic MNRP's rather than asymptotic power.

Section S6.5 provides results that supplement those of AB1 by comparing (S, φ) functions for a larger number of Ω matrices. These are results based on the best κ values in terms of asymptotic average power.

Section S6.6 provides asymptotic MNRP and power comparisons (based on fixed κ asymptotics) of several GMS tests and the recommended RMS test, which is the AQLR/t-test/κauto test.

Section S6.7 provides additional asymptotic MNRP and power results for some GMS and RMS tests that are not considered explicitly in AB1.

Section S6.8 provides comparative computation times for tests based on the AQLR and MMM test statistics, and the "asymptotic normal" and bootstrap versions of the t-test (i.e., $\varphi^{(1)}$) moment selection critical values.[23]

S6.1. *Approximately Optimal* $\kappa(\Omega)$ *and* $\eta(\Omega)$ *Functions*

S6.1.1. *Definitions of* $\kappa(\Omega)$ *and* $\eta(\Omega)$

Here, we describe how the recommended $\kappa(\Omega)$ and $\eta(\Omega)$ functions defined in AB1 are determined. These functions are for use with the recommended AQLR/t-test test.

First, for $p = 2$ and given $\rho \in (-1, 1)$, where ρ denotes the correlation that appears in Ω, we compute numerically the values of κ that maximize the

[21]These functions determine the data-dependent tuning parameter $\widehat{\kappa}$ and size-correction factor $\widehat{\eta}$.

[22]The constant $\varepsilon > 0$ ensures that the matrix $\widetilde{\Sigma}_n(\theta)$, whose inverse appears in the AQLR statistic, is nonsingular even if the estimator $\widehat{\Sigma}_n(\theta)$ of the asymptotic variance of the sample moment conditions is singular.

[23]Note that Section S7.6 provides additional numerical results concerning the computation of $\eta_2(p)$.

asymptotic average (size-corrected) power of the nominal .05 AQLR/t-test test over a fine grid of 31 κ values. We do this for each ρ in a fine grid of 43 values.[24] Because the power results are size-corrected, a by-product of determining the best κ value for each ρ value is the size-correction value η that yields asymptotically correct size for each ρ.[25]

Second, by a combination of intuition and the analysis of numerical results, we postulate that for $p \geq 3$, the optimal function $\kappa^*(\Omega)$ defined in Section S5.3 is well approximated by a function that depends on Ω only through the $[-1, 1]$-valued function $\delta(\Omega) =$ smallest off-diagonal element of Ω.

The explanation for this is as follows: (i) Given Ω, the value $\kappa^*(\Omega)$ that yields maximum asymptotic average power is such that the size-correction value $\eta^*(\Omega)$ is not very large. (This is established numerically for a variety of p and Ω.) The reason is that the larger is $\eta^*(\Omega)$, the closer is the test to the PA test and the lower is the power of the test for μ vectors that have less than p negative elements. (ii) The size-correction value $\eta^*(\Omega)$ is small if the rejection probability at the least favorable null vector μ is close to α when using the size-correction factor $\eta(\Omega) = 0$. (This is self-evident.) (iii) We postulate that null vectors μ that have two elements equal to zero and the rest equal to infinity are nearly least favorable null vectors. If true, then the size of the AQLR/t-test test depends on the two-dimensional submatrices of Ω that are the correlation matrices that correspond to the cases where only two moment conditions appear. (iv) The size of a test for given κ and $p = 2$ is decreasing in the correlation ρ. In consequence, the least favorable two-dimensional submatrix of Ω is the one with the smallest correlation. Hence, the value of κ that makes the size of the test equal to α for a small value of η is (approximately) a function of Ω through $\delta(\Omega)$. Note that this is just a heuristic explanation. It is not intended to be a proof.

Next, because $\delta(\Omega)$ corresponds to a particular 2×2 submatrix of Ω with correlation δ $(= \delta(\Omega))$, we take $\kappa(\Omega)$ to be the value that maximizes asymptotic average power when $p = 2$ and $\rho = \delta$, as specified in Table I and described in the second paragraph of this section.[26] We take $\eta(\Omega)$ to be the value determined by $p = 2$ and δ, that is, $\eta_1(\delta)$ in Table I, but allow for an adjustment that

[24]The grid of 31 κ values is $\{0, .2, .4, .6, .8, 1.0, 1.1, 1.2, \ldots, 2.9, 3.0, 3.2, \ldots, 3.8, 4.2\}$. The grid of 43 ρ values is $\{.99, .975, .95, .90, .85, \ldots, -.90, -.95, -.975, -.99\}$. The results are based on 40,000 critical-value repetitions and 40,000 size and power repetitions. Size-corrrection is done for the given value of ρ, not uniformly over $\rho \in [-1, 1]$, because ρ can be consistently estimated and hence is known asymptotically.

[25]The asymptotic size of the QLR/t-test for given κ is found numerically to be decreasing in ρ for $\rho \in [-1, 1]$. Hence, for $\rho \in [a_1, a_2)$, we take η to be the size-correction value that yields correct asymptotic size for $\rho = a_1$. See Section S7.5 for a discussion of how the maximum null rejection probability over $\mu \geq 0$ is calculated.

[26]For $\rho \in [-.8, 1.0]$, we use the κ values that maximize average asymptotic power for $p = 2$ as the automatic κ values. For $\rho \in [-1.0, -.8)$, however, we use somewhat larger κ values than those that maximize average power. The reason is that numerical results show that the best κ values (in terms of power) for $\rho \in [-1.0, -.85]$ (and $p = 2$) are somewhat smaller than for $\rho = -.80$.

depends on p (viz., $\eta_2(p)$) that is defined to guarantee that the test has correct asymptotic significance level (up to numerical error).[27] In particular, $\eta_1(\delta) \in R$ is defined to be such that

$$\inf_{h_1 \in R^2_{+,\infty}} \mathrm{CP}\big(h_1, \Omega_\delta, \eta_1(\delta)\big) = 1 - \alpha, \tag{S6.1}$$

where Ω_δ is the 2×2 correlation matrix with correlation δ (and $\kappa(\Omega)$ that appears in the definition of $\mathrm{CP}(h_1, \Omega, \eta)$ in (S4.11) is as just defined). The numerical calculation of $\eta_1(\delta)$ is described above in the second paragraph of this section. Next, $\eta_2(p) \in R$ is defined to be such that

$$\inf_{h_1 \in R^p_{+,\infty}, \Omega \in \Psi} \mathrm{CP}\big(h_1, \Omega, \eta_1\big(\delta(\Omega)\big) + \eta_2(p)\big) = 1 - \alpha, \tag{S6.2}$$

where $\kappa(\Omega)$ and $\eta_1(\delta(\Omega))$ are defined as described above. The numerical calculation of $\eta_2(p)$ is described in Section S7.5.

S6.1.2. *Automatic κ Power Assessment*

We now discuss numerical evaluations of how well the proposed method does in approximating the best κ, namely, $\kappa^*(\Omega)$. Three groups of results are provided and each group considers $p = 2, 4, 10$. The first group consists of the three Ω matrices considered in AB1 and the results are given by comparing the rows of Table II labeled AQLR/t-test/κbest and AQLR/t-test/κauto. The second group consists of a fixed set of 19 Ω matrices (defined in Section S7.2) chosen such that $\delta(\Omega)$ takes values on a grid in $[-.99, .99]$. The third group consists of 500 randomly generated Ω matrices for $p = 2, 4$ and 250 randomly generated Ω matrices for $p = 10$. See Section S7.2 for details concerning their distributions.

For the second group of results, the asymptotic power results are size-corrected and are based on (40,000, 40,000, 40,000) critical-value, size-correction, and power simulation repetitions for $p = 2$ and 4. For $p = 10$, they are based on (1000, 1000, 1000) repetitions. Average power is computed

Thus, there is a small deviation from the feature that the best κ value is monotone decreasing in ρ. When using the κ values for $p = 2$ with $p = 4, 10$, numerical results show that imposing monotonicity of κ in ρ yields better results for $p = 4$ in the sense that a smaller value $\eta_2(p)$ is needed for size correction (which leads to higher power over the entire range of δ values). For this reason, we define $\kappa(\delta)$ in Table I to take values for $\delta \in [-1.0, -.80)$ that are slightly larger than the power maximizing values. The resultant loss in power for $p = 2$ is small, being around .01 for $\delta \in [-1.0, -.80)$.

[27] One could define $\eta(\Omega)$ to depend separately on $\delta(\Omega)$ and p, say $\eta(\Omega) = \overline{\eta}(\delta(\Omega), p)$ for some function $\overline{\eta}$. This would yield a much more complicated function $\eta(\Omega)$ than the function $\eta(\Omega) = \eta_1(\delta(\Omega)) + \eta_2(p)$ that we use. Numerical results indicate that more complicated functions $\overline{\eta}$ are not needed. The simple function that we use works quite well.

TABLE S-I

ASYMPTOTIC POWER DIFFERENCES BETWEEN AQLR/t-TEST/κAUTO AND AQLR/t-TEST/κBEST TESTS FOR NOMINAL LEVEL .05 SIZE-CORRECTED TESTS[a]

	δ									
	−.99	−.975	−.95	−.9	−.8	−.7	−.6	−.5	−.4	−.2
$p = 2$	.022	.017	.009	.002	.000	.000	.000	.001	.000	.000
$p = 4$	.011	.007	.007	.009	.001	.001	.002	.003	.003	.001
$p = 10$	.004	.006	.004	.006	.004	.006	.012	.009	.006	.007
	.0	.2	.4	.6	.8	.9	.95	.975	.99	
$p = 2$	.001	.000	.000	.000	.000	.000	.000	.000	.000	
$p = 4$	.001	.001	.001	.000	.000	.000	.000	.000	.000	
$p = 10$	.002	.008	.002	.000	.000	.000	.000	.000	.000	

[a] κ = Auto denotes the data dependent-method of choosing κ described in (2.9)–(2.10) of AB1. κ = Best denotes the κ value that maximizes asymptotic average power.

for μ vectors that consist of linear combinations of the μ vectors defined in Section S7.1; see Section S7.2 for definitions of the linear combinations.

For all three groups, we assess the proposed method of selecting κ, referred to as the κauto method, by comparing the asymptotic average power of the κauto test with the corresponding κbest test, whose κ value is determined numerically to maximize asymptotic average power.

The results for the 19 Ω matrices are given in Table S-I. These results show that the κauto method works very well. There is very little difference between the asymptotic average power of the AQLR/t-test/κauto and AQLR/t-test/κbest tests. Only in 3 cases out of 57 is a difference of .010 or more detected.

The results for the randomly generated Ω matrices are similarly good for the κauto method. For $p = 2$, across the 500 Ω matrices, the average power differences have average equal to .0010, standard deviation equal to .0032, and range equal to [.000, .022]. For $p = 4$, across the 500 Ω matrices, the average power difference is .0012, the standard deviation is .0016, and the range is [.000, .010]. For $p = 10$, across the 250 Ω matrices, the average power differences have average equal to .0183, standard deviation equal to .0069, and range equal to [.000, .037].

In conclusion, the κauto method performs very well in terms of selecting κ values that maximize the asymptotic average power.

S6.2. *AQLR Statistic and Choice of ε*

There exist moment inequality models of practical importance in which the asymptotic variance matrix of the sample moment conditions is necessarily singular. For example, this occurs in the missing data example in Imbens and

Manski (2004) when the probability p of observing a variable is 0 or 1. It also occurs in simple entry models; for example, see Canay (2010).[28]

To handle models of this sort, AB1 introduces the AQLR statistic, which is based on the S_{2A} function. The AQLR statistic is designed so that the determinant of the random $k \times k$ matrix $\widetilde{\Sigma}_n(\theta)$ that enters the quadratic form in S_{2A} is at least as large as ε. Hence, if $\varepsilon > 0$, there is no difficulty in inverting $\widetilde{\Sigma}_n(\theta)$, $\widetilde{\Sigma}_n^{-1}(\theta)$ converges in probability to the inverse of the probability limit of $\widetilde{\Sigma}_n(\theta)$, and the asymptotic results of this paper hold even if the asymptotic variance matrix of the sample moment conditions is singular.

AB1 gives a recommended value of $\varepsilon = .012$. It is determined as follows. We simulate the asymptotic average power of the AQLR/t-test/κauto test as a function of ε for certain singular correlation matrices for $p = 2, 4$, and 10. For $p = 2$, Ω is singular only if the correlation ρ is $+1$ or -1. When $\rho = +1$ or close to $+1$, we find that the performance of the AQLR/t-test/κauto test (under the null and the alternative) is not sensitive to ε, provided ε is not too large. Even taking $\varepsilon = 0$ and using the Moore–Penrose inverse, the performance of the test is the same as when ε is positive. Similar results are obtained for $p = 4, 10$ when the correlation is positive and close to or equal to 1.

In consequence, we focus on cases with perfect negative correlation. For $p = 2$, we consider the correlation matrix $\Omega_{\mathrm{Sg,Neg}}$ with correlation $\rho = -1$. For $p = 4$, we consider the Toeplitz correlation matrix $\Omega_{\mathrm{Sg,Neg}}$ with $\rho = (-1, 1, -1)$, where ρ indexes the correlations on the diagonals of $\Omega_{\mathrm{Sg,Neg}}$ (as one moves away from the main diagonal). For $p = 10$, we consider the Toeplitz correlation matrix $\Omega_{\mathrm{Sg,Neg}}$ with $\rho = (-1, 1, -1, \dots, 1, -1)$.

For each value of p, we find that there is a sharp discontinuity in the asymptotic average power of the AQLR/t-test/κauto test as a function of ε at the point $\varepsilon = 0$ and no discontinuity in its asymptotic null rejection probabilities. (When $\varepsilon = 0$, the AQLR test is defined using the Moore–Penrose inverse of $\Omega_{\mathrm{Sg,Neg}}$.) Also, for all values of $\varepsilon > 0$, the asymptotic average power of the AQLR/t-test/κauto test is not very sensitive to the value of ε provided $\varepsilon > 0$, but power decreases when ε is made large enough. Based on these observations, we take the recommended value of ε to be the largest value that has asymptotic average power within .001 of the maximum asymptotic average power over $\varepsilon \in [10^{-6}, 1]$ for $p = 2$. As shown in Table S-II, this value is $\varepsilon = .012$. Table S-II gives the asymptotic average power of the AQLR/t-test/κauto test as a function of ε for $p = 2, 4, 10$. Asymptotic average power is computed for the vectors μ in $\mathcal{M}_p(\Omega_{\mathrm{Neg}})$, which is defined in Section S7.1. Table S-II is based on (40,000, 40,000, 40,000) critical-value, size-correction, and power simulation repetitions, respectively. Table S-II shows that the choice $\varepsilon = .012$ also works

[28]In the missing data model, even the variance submatrix consisting of the binding moment inequalities is singular when $p = 1$. In the entry model, the variance submatrix consisting of the binding moment inequalities is singular when the profit of one firm is not effected by the entry of the other firm or vice versa or both, which are cases of practical interest.

TABLE S-II

ASYMPTOTIC AVERAGE POWER OF THE AQLR/t-TEST/κAUTO TEST AS A FUNCTION OF THE ADJUSTMENT CONSTANT ε FOR $p = 2$, 4, AND 10[a]

			$p = 2$ & $\Omega_{\text{Sg,Neg}}$					
ε	.0	.000001	.00001	.0001	.001	.005	.010	.011
Avg Asy Power	.5616	.8752	.8752	.8752	.8751	.8749	.8745	.8744
ε	.0120	.0121	.0125	.013	.015	.02	.05	
Avg Asy Power	.8744	.8742	.8701	.8676	.8603	.8486	.8265	
			$p = 4$ & $\Omega_{\text{Sg,Neg}}$					
ε	.0	.0001	.001	.005	.01	.012	.02	
Avg Asy Power	.3905	.9401	.9400	.9398	.9396	.9395	.9392	
			$p = 10$ & $\Omega_{\text{Sg,Neg}}$					
ε	.0	.0001	.001	.005	.01	.012	.02	
Avg Asy Power	.2903	.9718	.9718	.9717	.9715	.9715	.9713	

[a] κ = Auto denotes the data dependent-method of choosing κ described in (2.9)–(2.10) of AB1.

well for $p = 4, 10$. For $p = 4$, the choice of $\varepsilon = .012$ yields asymptotic average power that is within .0006 of the maximum over different ε values. For $p = 10$, it is within .0003 of the maximum.

We note that the discontinuity at $\varepsilon = 0$ of the asymptotic average power of the AQLR/t-test/κauto test also is found in finite samples when perfect negative correlation is present; see Table S-V below. However, somewhat surprisingly, no discontinuity at $\varepsilon = 0$ is found for the null rejection probabilities, either asymptotic or finite sample, of the AQLR/t-test/κauto test when perfect negative (or positive) correlation is present; see Table S-IV below. (The AQLR/t-test/κauto test with $\varepsilon = 0$ equals the MP-QLR/t-test/κauto test.)

S6.3. *Singular Variance Matrices*

In this section, we present results that are similar to those in Tables II and III except that they are based on singular matrices $\Omega_{\text{Sg,Neg}}$ and $\Omega_{\text{Sg,Pos}}$, rather than the nonsingular matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}. As noted in Section S6.2, singular and near singular matrices arise in a number of moment inequality models of practical importance.

The matrices $\Omega_{\text{Sg,Neg}}$ for $p = 2, 4, 10$ are the same matrices that are considered in Section S6.2. The matrices $\Omega_{\text{Sg,Pos}}$ for $p = 2, 4, 10$ are correlation matrices with all elements equal to 1.

S6.3.1. *Asymptotic Power Comparisons*

Table S-III provides asymptotic average power comparisons of MMM, Max, AQLR, and MP-QLR test statistics combined with PA, t-test/κbest, and t-test/κauto critical values. Note that MP-QLR statistics are QLR statistics that

TABLE S-III

ASYMPTOTIC POWER COMPARISONS (SIZE-CORRECTED) FOR SINGULAR VARIANCE MATRICES[a]

Statistic	Critical Value	Tuning Param. κ	$p=10$		$p=4$		$p=2$	
			$\Omega_{\mathrm{Sg,Neg}}$	$\Omega_{\mathrm{Sg,Pos}}$	$\Omega_{\mathrm{Sg,Neg}}$	$\Omega_{\mathrm{Sg,Pos}}$	$\Omega_{\mathrm{Sg,Neg}}$	$\Omega_{\mathrm{Sg,Pos}}$
MMM	PA	—	.03	.27	.17	.40	.48	.51
MMM	t-test	Best	.15	.79	.31	.77	.52	.73
Max	PA	—	.28	.81	.36	.78	.48	.73
Max	t-test	Best	.28	.82	.38	.78	.52	.73
AQLR	PA	—	.96	.81	.92	.78	.85	.73
AQLR	***t*-test**	**Best**	.98	.82	.95	.78	.89	.73
AQLR	t-test	Auto	.97	.82	.94	.78	.87	.73
MP-QLR	PA	—	.29	.81	.39	.78	.56	.73
MP-QLR	t-test	Best	.29	.82	.39	.78	.56	.73
MP-QLR	t-test	Auto	.29	.82	.39	.78	.56	.73

[a] κ = Auto denotes the data dependent-method of choosing κ described in (2.9)–(2.10) of AB1. κ = Best denotes the κ value that maximizes asymptotic average power.

use the Moore–Penrose inverse of the singular matrix $\Omega_{\mathrm{Sg,Neg}}$ or $\Omega_{\mathrm{Sg,Pos}}$ as the weight matrix of the quadratic form. The power results are size-corrected, as in Table II. Average power is computed for the vectors μ in $\mathcal{M}_p(\Omega_{\mathrm{Neg}})$ when $\Omega = \Omega_{\mathrm{Sg,Neg}}$ and for the μ vectors in $\mathcal{M}_p(\Omega_{\mathrm{Pos}})$ when $\Omega = \Omega_{\mathrm{Sg,Pos}}$, where $\mathcal{M}_p(\Omega_{\mathrm{Neg}})$ and $\mathcal{M}_p(\Omega_{\mathrm{Pos}})$ are defined in Section S7.1. The results in Table S-III for $p = 2$, 4, and 10 are based on (40,000, 40,000, 40,000) critical-value, size-correction, and power simulation repetitions, respectively.

Table S-III shows that the AQLR/t-test/κauto test dominates the tests based on the MMM and Max statistics in terms of asymptotic average power. The differences in power are quite large for $\Omega_{\mathrm{Sg,Neg}}$ and small for $\Omega_{\mathrm{Sg,Pos}}$ (at least when the t-test/κbest critical values are used for the MMM and Max tests). In fact, the superiority of the AQLR/t-test/κauto test over the MMM and Max tests for $\Omega_{\mathrm{Sg,Neg}}$ is larger than it is for Ω_{Neg}; see Table II.

Table S-III shows that the AQLR/t-test/κauto test has vastly superior asymptotic average power compared to that of the MP-QLR/t-test/κauto test for $\Omega_{\mathrm{Sg,Neg}}$ and has the same power for $\Omega_{\mathrm{Sg,Pos}}$. Hence, it is clear that the adjustment made to the QLR statistic is beneficial.

Table S-III also shows that the data-dependent method of choosing κ and η works well with the singular matrices $\Omega_{\mathrm{Sg,Neg}}$ and $\Omega_{\mathrm{Sg,Pos}}$. The difference in asymptotic average power between the κbest and κauto versions of the AQLR/t-test test is .00 in three cases, .01 in two cases, and .02 in one case.

S6.3.2. *Finite-Sample MNRP and Power Comparisons*

Next we consider the finite-sample properties of the asymptotic normal and bootstrap versions of the AQLR/t-test/κauto and MP-QLR/t-test/κauto tests

TABLE S-IV

FINITE-SAMPLE MAXIMUM NULL REJECTION PROBABILITIES FOR SINGULAR VARIANCE MATRICES FOR NOMINAL .05 TESTS

			$p = 10$		$p = 4$		$p = 2$	
Test	Distrib.	n	$\Omega_{\text{Sg,Neg}}$	$\Omega_{\text{Sg,Pos}}$	$\Omega_{\text{Sg,Neg}}$	$\Omega_{\text{Sg,Pos}}$	$\Omega_{\text{Sg,Neg}}$	$\Omega_{\text{Sg,Pos}}$
AQLR/Nm	$N(0,1)$	100	.061	.038	.053	.045	.065	.053
AQLR/Bt			.050	.045	.048	.045	.051	.052
MP-QLR/Nm	$N(0,1)$	100	.044	.038	.050	.045	.049	.053
MP-QLR/Bt			.036	.045	.043	.045	.052	.052
AQLR/Nm	χ^2_3	100	.071	.043	.052	.050	.060	.066
AQLR/Bt			.045	.043	.048	.042	.050	.055
MP-QLR/Nm	χ^2_3	100	.071	.043	.050	.050	.045	.066
MP-QLR/Bt			.044	.043	.042	.042	.051	.055

with the singular matrices $\Omega_{\text{Sg,Neg}}$ and $\Omega_{\text{Sg,Pos}}$. The results are analogous to those given in Table III but with different Ω matrices and fewer distributions considered. We provide results for sample size $n = 100$. We consider the same numbers of moment inequalities $p = 2$, 4, and 10. We take the mean zero variance I_p random vector $Z^\dagger = \text{Var}^{-1/2}(m(W_i, \theta))(m(W_i, \theta) - Em(W_i, \theta))$ to be i.i.d. across elements and consider two distributions for the elements: standard normal (i.e., $N(0, 1)$) and chi-squared with 3 degrees of freedom, χ^2_3. The latter distribution is centered and scaled to have mean zero and variance 1. Average power is computed for the vectors μ in $\mathcal{M}_p(\Omega_{\text{Neg}})$ when $\Omega = \Omega_{\text{Sg,Neg}}$ and for the μ vectors in $\mathcal{M}_p(\Omega_{\text{Pos}})$ when $\Omega = \Omega_{\text{Sg,Pos}}$. The average power results are size-corrected based on the true Ω matrix. We use (3000, 3000, 3000) critical-value, size-correction, and rejection-probability repetitions for $p = 2$ and 4. We use (1000, 1000, 1000) repetitions for results for $p = 10$.

Table S-IV gives the finite-sample maximum null rejection probabilities (MNRP's) of the tests. There is very little difference in the MNRP's of the AQLR and MP-QLR versions of the tests. For both versions, the bootstrap and asymptotic normal implementation methods perform similarly and quite well. The bootstrap is slightly better overall. For the bootstrap version of the AQLR/t-test/κauto test, the MNRP's lie in the range [.042, .055]. An interesting feature of the results is that there is no overrejection by the asymptotic normal version of the AQLR/t-test/κauto test with Ω_{Neg}, χ^2_3 distribution, and $p = 4, 10$, whereas substantial overrejection is reported in Table III in the same scenario except with Ω_{Neg} in place of $\Omega_{\text{Sg,Neg}}$.

We conclude that the bootstrap version of the AQLR/t-test/κauto test, which is the recommended test, works very well in terms of MNRP's with singular variance matrices.

Table S-V reports the finite-sample average power results with the singular matrices $\Omega_{\text{Sg,Neg}}$ and $\Omega_{\text{Sg,Pos}}$. The AQLR-based tests all outperform the MP-

TABLE S-V

FINITE-SAMPLE (SIZE-CORRECTED) AVERAGE POWER FOR SINGULAR VARIANCE MATRICES FOR NOMINAL .05 TESTS

			$p=10$		$p=4$		$p=2$	
Test	Distrib.	n	$\Omega_{\text{Sg,Neg}}$	$\Omega_{\text{Sg,Pos}}$	$\Omega_{\text{Sg,Neg}}$	$\Omega_{\text{Sg,Pos}}$	$\Omega_{\text{Sg,Neg}}$	$\Omega_{\text{Sg,Pos}}$
AQLR/PA	$N(0,1)$	100	.97	.79	.92	.77	.85	.73
AQLR/Nm			.96	.78	.93	.77	.85	.72
AQLR/Bt			.97	.78	.93	.78	.86	.71
MP-QLR/PA	$N(0,1)$	100	.31	.79	.40	.77	.54	.73
MP-QLR/Nm			.29	.78	.39	.77	.55	.72
MP-QLR/Bt			.29	.78	.39	.78	.54	.71
AQLR/PA	χ_3^2	100	.97	.78	.92	.75	.85	.72
AQLR/Nm			.96	.78	.94	.74	.85	.66
AQLR/Bt			.97	.78	.94	.74	.86	.65
MP-QLR/PA	χ_3^2	100	.31	.78	.41	.76	.56	.72
MP-QLR/Nm			.29	.78	.40	.74	.57	.67
MP-QLR/Bt			.29	.78	.39	.74	.56	.65

QLR-based tests by a wide margin for $\Omega_{\text{Sg,Neg}}$ and perform essentially the same for $\Omega_{\text{Sg,Pos}}$. For example, for $p=10$, $\Omega_{\text{Sg,Neg}}$, and the $N(0,1)$ distribution, the power difference is .97 to .29 for the recommended AQLR/t-test/κauto test compared to the MP-QLR/t-test/κauto test for the bootstrap versions of these tests.

For all tests considered, the bootstrap and asymptotic normal implementations of the tests perform quite similarly. This is consistent with the MNRP results in Table S-IV. For all tests, the results for the normal and χ_3^2 distributions are quite similar. This also is consistent with the MNRP results in Table S-IV, but differs from the results in Table III.

Based on Table S-V, we conclude that the bootstrap version of the AQLR/t-test/κauto test, which is the recommended test, works very well in terms of finite-sample average power with singular variance matrices.

S6.3.3. *ELR Test With Singular Correlation Matrix*

In this section, we define the empirical likelihood ratio (ELR) statistic for the case where no equality constraints appear (i.e., $v=0$), describe the method used to compute the ELR statistic, and compare the finite-sample properties of the bootstrap versions of the ELR/t-test/κauto and AQLR/t-test/κauto tests with the singular matrices $\Omega_{\text{Sg,Neg}}$ and $\Omega_{\text{Sg,Pos}}$.

When $v=0$, the ELR statistic can be written as

$$T_n^{\text{ELR}}(\theta) = \max_{\lambda=(\lambda_1,\dots,\lambda_p)':\lambda_\ell\le 0,\forall \ell\le p} 2\sum_{i=1}^n \bigl(1+\lambda' m(W_i,\theta)\bigr); \tag{S6.3}$$

see Canay (2010). This expression is easier to compute than an equivalent expression given in Canay (2010) and AG, so we use it in the numerical work.

The constrained optimization (CO) module of GAUSS was used to compute the ELR statistic. We found that it was necessary to do a careful analysis of the optimization algorithm used. Arbitrarily selecting a preprogrammed generic optimization algorithm and presuming that it will give accurate and timely results is not a wise procedure whether the correlation matrix is nonsingular or singular.

The CO module contains five algorithms: BFGS, DFP, NR, scaled BFGS, and scaled DFP; four line search methods: step length $=1$, cubic or quadratic step, step halving, and Brent's method; and two gradient/Hessian computation methods: numerical and analytical. We investigated the properties of each of these methods with nonsingular and singular correlation matrices in many different combinations before selecting one to use. For nonsingular correlation matrices, scaled BFGS and scaled DFP had substantial convergence and accuracy problems regardless of the line search method and gradient/Hessian method employed. DFP often had similar convergence problems. BFGS and NR worked well in terms of giving accurate results with line search method 1 and 2, and numerical derivatives. BFGS did not work well in terms of accuracy with analytic gradient/Hessian. NR worked well in terms of accuracy and convergence properties with line search methods 1 and 2, and with numerical and analytic gradient/Hessian. NR was fastest with line search 1 and analytic gradient/Hessian, which is the method we employed to compute the results given in Table III for nonsingular correlation matrices.

For singular variance matrices, all methods in CO had convergence problems when $p=4$ and $p=10$. This is because with a singular correlation matrix, the Hessian of the empirical likelihood objective function is singular almost surely. For $p=2$, NR with line search 1 and analytic gradient/Hessian worked well. In consequence, we only report results for singular correlation matrices for $p=2$. We provide results for the matrices $\Omega_{\mathrm{Sg,Neg}}$ and $\Omega_{\mathrm{Sg,Pos}}$ defined above. We use (5000, 5000) critical-value and rejection-probability repetitions under the null and the alternative.

The bootstrap version of the ELR/t-test/κauto is based on bootstrap samples that are recentered by the average of the observations from the original sample; that is, the original sample is $\{W_1,\dots,W_n\}$, the bootstrap sample $\{W_1^*,\dots,W_n^*\}$ is n i.i.d. draws from the empirical distribution of the original sample, and the recentered bootstrap sample is $\{W_1^*-\overline{W}_n,\dots,W_n^*-\overline{W}_n\}$, where $\overline{W}_n=n^{-1}\sum_{i=1}^n W_i\in R^p$.

For $p=2$, Table S-VI shows that the performance of the ELR/t-test/κauto/Bt and AQLR/t-test/κauto/Bt tests is essentially the same in terms of MNRP's and average power. Hence, the most important distinction between the two tests is the speed and reliability of their computation. The AQLR test has a substantial advantage in these dimensions, especially when the correlation matrix is singular.

TABLE S-VI

FINITE-SAMPLE MAXIMUM NULL REJECTION PROBABILITIES AND (SIZE-CORRECTED) AVERAGE POWER FOR SINGULAR VARIANCE MATRICES FOR NOMINAL .05 TESTS

			$p=2$	
Test	Distrib.	H_0/H_1	$\Omega_{\text{Sg,Neg}}$	$\Omega_{\text{Sg,Pos}}$
AQLR/Bt	$N(0,1)$	H_0	.053	.051
ELR/Bt	$N(0,1)$	H_0	.054	.051
AQLR/Bt	t_3	H_0	.055	.055
ELR/Bt	t_3	H_0	.048	.053
AQLR/Bt	χ^2_3	H_0	.052	.052
ELR/Bt	χ^2_3	H_0	.053	.052
AQLR/Bt	$N(0,1)$	H_1	.86	.72
ELR/Bt	$N(0,1)$	H_1	.86	.72
AQLR/Bt	t_3	H_1	.86	.74
ELR/Bt	t_3	H_1	.87	.73
AQLR/Bt	χ^2_3	H_1	.86	.65
ELR/Bt	χ^2_3	H_1	.86	.65

S6.4. *κ Values That Maximize Asymptotic Average Power*

The κ values that maximize asymptotic average power, that is, the best κ values, which are used in the construction of Table II, are given in Table S-VII.

Table S-VIII gives the asymptotic maximum null rejection probabilities (where the maximum is over all mean vectors in the null hypothesis for a fixed correlation matrix Ω) of the RMS tests that appear in Table II and are based on the $\kappa =$ best tuning parameter and no size-correction factor (i.e., $\eta = 0$). The results show that the κ value that maximizes asymptotic average power also has quite good asymptotic size properties even with $\eta = 0$, with the exceptions of the AQLR/$\varphi^{(2)}$, AQLR/$\varphi^{(3)}$, AQLR/$\varphi^{(4)}$, and AQLR/MMSC tests.

S6.5. *Comparison of (S, φ) Functions: 19 Ω Matrices*

Here we compare the MMM/t-test/κbest, AQLR/t-test/κbest, AQLR/t-test/κauto, and AQLR/MMSC/κbest tests. This section is quite similar to Section 4 of AB1 except that 19 Ω matrices are considered here, rather than 3, and fewer tests are considered.[29] The 19 Ω matrices are the same as those considered in Table S-I and are defined in Section S7.2.

The qualitative results reported in AB1 are found in Table S-IX to apply as well to the broader range of Ω matrices that are considered.

[29]For the AQLR/MMSC/κbest test, we report only results for $p = 2, 4$ because the results for $p = 10$ are very time consuming.

TABLE S-VII

κ VALUES THAT MAXIMIZE (SIZE-CORRECTED) ASYMPTOTIC AVERAGE POWER[a]

Statistic	Critical Value	$p=10$			$p=4$			$p=2$		
		Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
MMM	t-test	2.5	1.4	.4	2.5	1.4	.2	2.5	1.7	.6
Max	t-test	2.4	1.4	.6	2.5	1.5	.8	2.5	1.8	.6
SumMax	t-test	2.3	1.3	.4	2.5	1.6	.4	2.5	1.7	.6
AQLR	t-test	2.5	1.4	.6	2.5	1.4	.8	2.6	1.7	.6
AQLR	$\varphi^{(2)}$	2.1[b]	.6[b]	.0[b]	2.4[c]	1.0[d]	.2[d]	2.0[d]	1.2[d]	.2[d]
AQLR	$\varphi^{(3)}$	12.5[b]	2.3[b]	1.1[b]	9.0[c]	2.8[d]	1.4[d]	10.0[d]	1.4[d]	1.2[d]
AQLR	$\varphi^{(4)}$	2.7[b]	1.4[b]	.2[b]	2.5[c]	1.4[d]	.4[d]	2.2[d]	1.9[d]	.2[d]
AQLR	MMSC	5.3[b]	1.1[b]	.2[b]	5.7	1.4	.8	2.8	1.7	.6

[a] Unless otherwise noted, all cases are based on (40,000, 40,000, 40,000) critical-value, size-correction, and rejection-probability repetitions.

[b] Results are based on (1000, 1000, 1000) repetitions.

[c] Results are based on (2000, 2000, 2000) repetitions.

[d] Results are based on (5000, 5000, 5000) repetitions.

S6.6. *Comparison of RMS and GMS Procedures*

In this section, we provide asymptotic MNRP and power comparisons (based on fixed κ asymptotics) of several GMS tests and the recommended RMS test, which is the AQLR/t-test/κauto test.

We consider GMS tests based on $(S, \varphi) = $ (MMM, t-test), (AQLR, t-test), and (AQLR, MMSC). The GMS tests depend on a tuning parameter κ $(= \kappa_n)$

TABLE S-VIII

COMPARISONS OF ASYMPTOTIC MAXIMUM NULL REJECTION PROBABILITIES[a]

Statistic	Critical Value	Tuning Param. κ	$p=10$			$p=4$			$p=2$		
			Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
MMM	t-test	Best	.059	.061	.054	.054	.058	.058	.054	.053	.051
Max	t-test	Best	.056	.057	.052	.053	.055	.052	.054	.052	.052
SumMax	t-test	Best	.060	.060	.054	.054	.055	.056	.054	.053	.051
AQLR	$\varphi^{(2)}$	Best	.092[b]	.102[b]	.066[b]	.064[c]	.057[d]	.052[d]	.062[d]	.059[d]	.054[d]
AQLR	$\varphi^{(3)}$	Best	.113[b]	.111[b]	.066[b]	.098[c]	.063[d]	.052[d]	.072[d]	.068[d]	.055[d]
AQLR	$\varphi^{(4)}$	Best	.088[b]	.089[b]	.066[b]	.066[c]	.057[d]	.052[d]	.062[d]	.058[d]	.056[d]
AQLR	t-test	Best	.058	.061	.051	.053	.058	.051	.053	.053	.051
AQLR	MMSC	Best	.088[b]	.097[b]	.066[b]	.055	.058	.051	.052	.053	.051

[a] These results use κ = Best and $\eta = 0$. Unless otherwise noted, results are based on (40,000, 40,000) critical-value and rejection-probability repetitions.

[b] Results are based on (1000, 1000) repetitions.

[c] Results are based on (2000, 2000) repetitions.

[d] Results are based on (5000, 5000) repetitions.

TABLE S-IX

ASYMPTOTIC POWER COMPARISONS (SIZE-CORRECTED) FOR 19 Ω MATRICES[a]

(a) $p = 10$

			$\delta(\Omega)$									
Statist.	Crit. Val.	κ	−.99	−.975	−.95	−.9	−.8	−.7	−.6	−.5	−.4	−.2
MMM	t-test	κbest	.16	.16	.17	.18	.20	.23	.28	.34	.42	.57
AQLR	t-test	κbest	.96	.94	.76	.55	.47	.48	.50	.52	.55	.61
AQLR	t-test	κauto	.96	.94	.76	.55	.47	.47	.49	.51	.54	.60
Power	Envelope	—	.98	.98	.94	.85	.74	.73	.74	.75	.77	.81
			0.0	.2	.4	.6	.8	.9	.95	.975	.99	
MMM	t-test	κbest	.67	.36	.50	.85	.82	.81	.80	.80	.79	
AQLR	t-test	κbest	.67	.37	.50	.85	.83	.83	.82	.82	.82	
AQLR	t-test	κauto	.67	.36	.50	.85	.83	.83	.82	.82	.82	
Power	Envelope	—	.85	.47	.59	.89	.85	.83	.82	.82	.82	

(b) $p = 4$

			$\delta(\Omega)$									
Statist.	Crit. Val.	κ	−.99	−.975	−.95	−.9	−.8	−.7	−.6	−.5	−.4	−.2
MMM	t-test	κbest	.30	.30	.30	.31	.34	.37	.42	.48	.53	.62
AQLR	t-test	κbest	.93	.87	.74	.60	.53	.53	.55	.57	.59	.64
AQLR	t-test	κauto	.92	.87	.73	.59	.53	.53	.54	.56	.59	.64
AQLR	MMSC	κbest	.93	.88	.75	.63	.55	.54	.55	.57	.60	.64
Power	Envelope	—	.95	.94	.87	.80	.70	.70	.70	.72	.73	.77
			0.0	.2	.4	.6	.8	.9	.95	.975	.99	
MMM	t-test	κbest	.69	.45	.58	.79	.79	.78	.77	.77	.77	
AQLR	t-test	κbest	.69	.46	.59	.80	.79	.78	.78	.78	.78	
AQLR	t-test	κauto	.69	.46	.59	.80	.79	.78	.78	.78	.78	
AQLR	MMSC	κbest	.69	.46	.59	.80	.79	.78	.78	.78	.78	
Power	Envelope	—	.80	.54	.66	.83	.81	.79	.79	.78	.78	

(*Continues*)

that does not depend on Ω. We consider the values $\kappa = 2.35$ and $\kappa = 1.87$. The former corresponds to the Bayesian information criterion (BIC) choice $\kappa_n = (\ln n)^{1/2}$ for $n = 250$ and the latter corresponds to the law of the iterated logarithm (LIL) choice $\kappa_n = (2 \ln \ln n)^{1/2}$ for $n = 300$. Note that the BIC choice yields $\kappa_n \in [2.15, 2.63]$ for $n \in [100, 1000]$ and the LIL choice yields $\kappa_n \in [1.75, 1.97]$ for $n \in [100, 1000]$.

Tables S-X and S-XI provide the asymptotic MNRP and power results, respectively, for $p = 2, 4, 10$ and $\Omega = \Omega_{\text{Neg}}, \Omega_{\text{Zero}}, \Omega_{\text{Pos}}$. The critical values are

TABLE S-IX—*Continued*

(c) $p = 2$

			$\delta(\Omega)$									
Statist.	Crit. Val.	κ	−.99	−.975	−.95	−.9	−.8	−.7	−.6	−.5	−.4	−.2
MMM	t-test	κbest	.52	.52	.51	.51	.52	.54	.57	.59	.62	.66
AQLR	t-test	κbest	.86	.83	.76	.65	.60	.59	.60	.61	.62	.66
AQLR	t-test	κauto	.84	.81	.76	.65	.60	.59	.60	.61	.62	.66
AQLR	MMSC	κbest	.86	.83	.76	.65	.60	.59	.60	.61	.62	.66
Power	Envelope	—	.88	.86	.83	.75	.70	.69	.69	.70	.70	.73
			0.0	.2	.4	.6	.8	.9	.95	.975	.99	
MMM	t-test	κbest	.69	.59	.66	.72	.73	.73	.73	.73	.73	
AQLR	t-test	κbest	.69	.59	.66	.73	.73	.73	.74	.73	.73	
AQLR	t-test	κauto	.69	.59	.66	.73	.73	.73	.74	.73	.73	
AQLR	MMSC	κbest	.69	.59	.66	.73	.73	.73	.74	.73	.73	
Power	Envelope	—	.75	.63	.70	.75	.74	.74	.74	.73	.73	

[a] $\kappa =$ best denotes the κ value that maximizes asymptotic average power. The results are based on (40,000, 40,000, 40,000) critical-value, size-correction, and rejection-probability repetitions for $p = 2$, 4, and 10.

obtained using 40,000 simulation repetitions, and both the MNRP and power results are obtained using 40,000 repetitions, which yields a simulation standard error of .0011.[30] The power results are size-corrected.

TABLE S-X
ASYMPTOTIC MNRP COMPARISONS OF NOMINAL .05 GMS TESTS AND THE RECOMMENDED RMS TEST

			$p = 10$			$p = 4$			$p = 2$		
Statistic	Critical Value	Tuning Param. κ	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
MMM	t-test	2.35	.061	.054	.052	.056	.052	.052	.055	.051	.050
MMM	t-test	1.87	.073	.056	.052	.070	.054	.052	.065	.052	.050
AQLR	t-test	2.35	.060	.054	.050	.056	.052	.051	.056	.051	.050
AQLR	t-test	1.87	.076	.056	.050	.073	.054	.051	.075	.052	.050
AQLR	MMSC	2.35	.148[a]	.081[a]	.064[a]	.111	.052	.051	.057	.051	.050
AQLR	MMSC	1.87	.173[a]	.082[a]	.064[a]	.119	.054	.051	.075	.052	.050
AQLR	t-test	Auto	.044	.046	.038	.047	.049	.047	.051	.051	.050

[a] These results are based on (1000, 1000) critical-value and rejection-probability repetitions. All other results are based on (40,000, 40,000) repetitions.

[30] This is true except for the AQLR/MMSC tests with $p = 10$, which are based on (1000, 1000) critical-value and rejection-probability repetitions.

TABLE S-XI

ASYMPTOTIC POWER COMPARISONS (SIZE-CORRECTED) FOR NOMINAL .05 GMS TESTS AND THE RECOMMENDED RMS TEST

Statistic	Critical Value	Tuning Param. κ	$p=10$ Ω_{Neg}	$p=10$ Ω_{Zero}	$p=10$ Ω_{Pos}	$p=4$ Ω_{Neg}	$p=4$ Ω_{Zero}	$p=4$ Ω_{Pos}	$p=2$ Ω_{Neg}	$p=2$ Ω_{Zero}	$p=2$ Ω_{Pos}
MMM	t-test	2.35	.18	.64	.68	.31	.68	.67	.51	.68	.68
MMM	t-test	1.87	.16	.66	.71	.28	.69	.70	.48	.69	.69
AQLR	t-test	2.35	.55	.64	.79	.60	.68	.76	.64	.68	.70
AQLR	t-test	1.87	.52	.66	.80	.56	.69	.77	.59	.69	.71
AQLR	MMSC	2.35	.46[a]	.60[a]	.74[a]	.56	.68	.75	.64	.68	.70
AQLR	MMSC	1.87	.44[a]	.63[a]	.76[a]	.54	.69	.76	.59	.69	.71
AQLR	t-test	Auto	.55	.67	.82	.59	.69	.78	.65	.69	.73
Power	Envelope	—	.85	.85	.85	.80	.80	.80	.75	.75	.75

[a]These results are based on (1000, 1000, 1000) critical-value, size-correction, and rejection-probability repetitions. All other results are based on (40,000, 40,000, 40,000) repetitions.

Table S-X shows that the GMS tests, AQLR/t-test, and MMM/t-test with $\kappa=1.87$ have asymptotic MNRP that is close to .050 for Ω_{Pos}, is slightly above .050 for Ω_{Zero}, and is noticeably above .050 for Ω_{Neg}. For example, for Ω_{Neg}, the AQLR/t-test/$\kappa = 1.87$ test has MNRP .075, .073, and .076 for $p = 2, 4,$ and 10, respectively. These tests with $\kappa = 2.35$ have asymptotic MNRP that is closer to .050 than when $\kappa = 1.87$. There is still some overrejection with Ω_{Neg}, but it is noticeably smaller. For example, for Ω_{Neg}, the AQLR/t-test/$\kappa = 2.35$ test has MNRP .056, .056, and .060 for $p = 2, 4,$ and 10, respectively.

The AQLR/MMSC test shows substantial overrejection whenever $p = 10$ or $\Omega = \Omega_{\text{Neg}}$ for both $\kappa = 1.87$ and 2.35. For example, the MNRP for the AQLR/MMSC/$\kappa = 2.35$ test is .148 for Ω_{Neg}.

The recommended RMS test has asymptotic MNRP that is close to its nominal level .050. For Ω_{Neg}, it has MNRP .051, .047, and .044 for $p = 2, 4,$ and 10, respectively.

Based on Table S-X, we conclude that some GMS tests have moderate to large problems of overrejection asymptotically under fixed κ asymptotics for some Ω matrices. However, some GMS tests with $\kappa = 2.35$ perform fairly well and overreject by a relatively small amount. The recommended RMS test performs well. It shows no sign of overrejection.

Next, we discuss the asymptotic power results given in Table S-XI. Table S-XI shows that the GMS tests given by MMM/t-test with $\kappa = 2.35$ and $\kappa = 1.87$ have quite low power compared to the recommended RMS test (i.e., the AQLR/t-test/κauto test) for Ω_{Neg} and noticeably lower power for Ω_{Pos}. For Ω_{Neg}, the powers of the MMM/t-test tests are decreasing in p rather quickly.

The GMS tests AQLR/t-test/$\kappa = 2.35$ and AQLR/t-test/$\kappa = 1.87$ have power that is similar to that of the recommended RMS test, but lower on average. The

GMS tests AQLR/MMSC/$\kappa = 2.35$ and AQLR/MMSC/$\kappa = 1.87$ have lower power than the corresponding t-test versions, especially for $p = 10$.

We conclude that (i) the best GMS test in terms of asymptotic MNRP and power is the AQLR/t-test/$\kappa = 2.35$, (ii) the recommended RMS test performs similarly to this GMS test, but has slightly higher power on average and does not overreject under the null hypothesis, and (iii) the recommended RMS test outperforms the other GMS tests considered by a noticeable margin in terms of asymptotic MNRP and/or power.

S6.7. *Additional Asymptotic MNRP and Power Results*

Table S-XII reports asymptotic MNRP results for some tests that are not considered in AB1 or above. Table S-XIII does likewise for asymptotic power.

The critical values for the pure ELR test are based on a constant critical value that does not depend on Ω (i.e., it is least favorable over Ω). It is approximated by taking the maximum critical value for the AQLR/PA test over 43 Ω matrices.[31] (Each of these PA critical values is computed using all null mean vectors μ that consist of 0's and ∞'s.) The critical values are found to be 5.07, 7.99, and 16.2 for $p = 2$, 4, and 10, respectively.

S6.8. *Comparative Computation Times*

As reported in the paper, to compute the recommended bootstrap RMS test (i.e., AQLR/t-test/κauto/Bt) using 10,000 bootstrap repetitions takes .34, .39, and .86 seconds when $p = 2$, 4, and 10, respectively, and $n = 250$ using a PC with a 3.2-GHz processor. For the asymptotic normal version of the recommended RMS test, i.e., AQLR/t-test/κauto/Nm, using 10,000 critical value simulations, the times are .08, .09, and .16, seconds, respectively.

In contrast, to compute the bootstrap version of the MMM/t-test/$\kappa = 2.35$ test using 10,000 bootstrap repetitions takes .19, .24, and .60 seconds when $p = 2$, 4, and 10, respectively, and $n = 250$. For the asymptotic normal version of the MMM/t-test/$\kappa = 2.35$ test, the times are .003, .004, and .009 seconds, respectively. Note that the computation times are not affected by whether κ is taken to be κauto or $\kappa = 2.35$. The difference between the results in the previous paragraph and this paragraph is due to the different statistics used: AQLR and MMM.

The results indicate that the bootstrap version of the MMM-based test is between 1.4 and 1.8 times faster than the corresponding bootstrap version of the AQLR-based test. On the other hand, the asymptotic normal version of the MMM-based test is very much faster (from 17 to 30 times) than the asymptotic

[31] For any given value of $\delta = \delta(\Omega)$, these 43 matrices are defined just as the 19 Toeplitz matrices are defined in Section S7.2. The $\delta(\Omega)$ values considered are the 43 values specified by the endpoints for δ in Table I, but including $-.99$ and excluding -1.0 and 1.0.

TABLE S-XII

ASYMPTOTIC MNRP COMPARISONS OF NOMINAL .05 TESTS WITH $\eta = 0$[a]

Statistic	Critical Value	Tuning Param. κ	$p=10$ Ω_{Neg}	$p=10$ Ω_{Zero}	$p=10$ Ω_{Pos}	$p=4$ Ω_{Neg}	$p=4$ Ω_{Zero}	$p=4$ Ω_{Pos}	$p=2$ Ω_{Neg}	$p=2$ Ω_{Zero}	$p=2$ Ω_{Pos}
MMM	PA	—	.052	.048	.046	.051	.050	.050	.053	.050	.049
AQLR	PA	—	.048	.048	.047	.050	.050	.051	.051	.050	.049
ELR	Const.	—	.021	.010	.000	.048	.025	.006	.047	.031	.025
MMM	t-test	Best	.059	.061	.054	.054	.058	.058	.054	.053	.051
MMM	t-test	2.35	.061	.054	.052	.056	.052	.052	.055	.051	.050
MMM	t-test	1.87	.073	.056	.052	.070	.054	.052	.065	.052	.050
Max	PA	—	.051	.049	.047	.051	.051	.051	.053	.050	.050
Max	t-test	Best	.056	.057	.052	.053	.055	.052	.054	.052	.052
Max	t-test	2.35	.056	.053	.051	.054	.052	.052	.055	.051	.050
Max	t-test	1.87	.066	.054	.051	.065	.053	.052	.065	.052	.050
SumMax	PA	—	.051	.047	.047	.051	.050	.051	.053	.050	.049
SumMax	t-test	Best	.060	.060	.054	.054	.055	.056	.054	.053	.051
SumMax	t-test	2.35	.059	.054	.052	.056	.052	.052	.055	.051	.050
SumMax	t-test	1.87	.071	.056	.052	.070	.053	.052	.065	.052	.050
AQLR	$\varphi^{(2)}$	Best	.092[b]	.102[b]	.066[b]	.064[c]	.057[d]	.052[d]	.062[d]	.059[d]	.054[d]
AQLR	$\varphi^{(2)}$	2.35	.090[b]	.081[b]	.065[b]	.058[c]	.057[d]	.052[d]	.062[d]	.056[d]	.053[d]
AQLR	$\varphi^{(2)}$	1.87	.098[b]	.081[b]	.065[b]	.066[c]	.057[d]	.052[d]	.062[d]	.056[d]	.053[d]
AQLR	$\varphi^{(3)}$	Best	.113[b]	.111[b]	.066[b]	.098[c]	.063[d]	.052[d]	.072[d]	.068[d]	.055[d]
AQLR	$\varphi^{(3)}$	2.35	.245[b]	.111[b]	.065[b]	.153[c]	.065[d]	.052[d]	.118[d]	.062[d]	.054[d]
AQLR	$\varphi^{(3)}$	1.87	.262[b]	.114[b]	.065[b]	.162[c]	.068[d]	.052[d]	.127[d]	.065[d]	.054[d]
AQLR	$\varphi^{(4)}$	Best	.088[b]	.089[b]	.066[b]	.066[c]	.057[d]	.052[d]	.062[d]	.058[d]	.056[d]
AQLR	$\varphi^{(4)}$	2.35	.092[b]	.081[b]	.065[b]	.062[c]	.057[d]	.052[d]	.062[d]	.056[d]	.053[d]
AQLR	$\varphi^{(4)}$	1.87	.105[b]	.082[b]	.065[b]	.077[c]	.057[d]	.052[d]	.074[d]	.058[d]	.053[d]
AQLR	t-test	Best	.058	.061	.051	.053	.058	.051	.053	.053	.051
AQLR	t-test	2.35	.060	.054	.050	.056	.052	.051	.056	.051	.050
AQLR	t-test	1.87	.076	.056	.050	.073	.054	.051	.075	.052	.050
AQLR	t-test	Auto	.044	.046	.038	.047	.049	.047	.051	.051	.050
AQLR	MMSC	Best	.088[b]	.097[b]	.066[b]	.055	.058	.051	.052	.053	.051
AQLR	MMSC	2.35	.148[b]	.081[b]	.064[b]	.111	.052	.051	.057	.051	.050
AQLR	MMSC	1.87	.173[b]	.082[b]	.064[b]	.119	.054	.051	.075	.052	.050

[a] κ = Best denotes the κ value that maximizes asymptotic average power. Unless stated otherwise, results are based on (40,000, 40,000) critical-value and rejection-probability repetitions.

[b] Results are based on (1000, 1000) repetitions.

[c] Results are based on (2000, 2000) repetitions.

[d] Results are based on (5000, 5000) repetitions.

normal version of the AQLR-based test. (This is because generation of the bootstrap samples dominates the computation time for the bootstrap version of the MMM-based test.)

When constructing a CS, if the computation time is burdensome (because one needs to carry out many tests with different values of θ as the null value),

TABLE S-XIII

ASYMPTOTIC POWER COMPARISONS (SIZE-CORRECTED) OF NOMINAL .05 TESTS[a]

Statistic	Critical Value	Tuning Param. κ	$p=10$ Ω_{Neg}	$p=10$ Ω_{Zero}	$p=10$ Ω_{Pos}	$p=4$ Ω_{Neg}	$p=4$ Ω_{Zero}	$p=4$ Ω_{Pos}	$p=2$ Ω_{Neg}	$p=2$ Ω_{Zero}	$p=2$ Ω_{Pos}
MMM	PA	—	.04	.36	.34	.20	.53	.45	.48	.62	.59
AQLR	PA	—	.35	.36	.69	.45	.53	.70	.58	.62	.65
ELR	Const.	—	.19	.17	.12	.44	.42	.39	.57	.55	.54
MMM	t-test	Best	.18	.67	.79	.31	.69	.76	.51	.69	.72
MMM	t-test	2.35	.18	.64	.68	.31	.68	.67	.51	.68	.68
MMM	t-test	1.87	.16	.66	.71	.28	.69	.70	.48	.69	.69
Max	PA	—	.19	.44	.70	.30	.57	.71	.48	.64	.66
Max	t-test	Best	.25	.58	.82	.35	.66	.78	.51	.69	.72
Max	t-test	2.35	.24	.57	.80	.35	.65	.76	.51	.68	.71
Max	t-test	1.87	.23	.58	.80	.33	.66	.77	.48	.69	.71
SumMax	PA	—	.10	.43	.62	.20	.55	.60	.48	.62	.59
SumMax	t-test	Best	.20	.65	.81	.31	.69	.77	.51	.69	.72
SumMax	t-test	2.35	.20	.62	.76	.31	.68	.72	.51	.68	.68
SumMax	t-test	1.87	.19	.64	.78	.28	.69	.73	.48	.69	.69
AQLR	$\varphi^{(2)}$	Best	.51[b]	.65[b]	.81[b]	.60[c]	.69[d]	.78[d]	.66[d]	.69[d]	.72[d]
AQLR	$\varphi^{(2)}$	2.35	.50[b]	.58[b]	.77[b]	.60[c]	.65[d]	.75[d]	.64[d]	.68[d]	.70[d]
AQLR	$\varphi^{(2)}$	1.87	.50[b]	.60[b]	.78[b]	.60[c]	.66[d]	.76[d]	.64[d]	.68[d]	.70[d]
AQLR	$\varphi^{(3)}$	Best	.43[b]	.63[b]	.81[b]	.55[c]	.68[d]	.78[d]	.61[d]	.69[d]	.72[d]
AQLR	$\varphi^{(3)}$	2.35	.36[b]	.63[b]	.80[b]	.52[c]	.68[d]	.77[d]	.59[d]	.68[d]	.72[d]
AQLR	$\varphi^{(3)}$	1.87	.36[b]	.63[b]	.81[b]	.52[c]	.68[d]	.77[d]	.59[d]	.69[d]	.72[d]
AQLR	$\varphi^{(4)}$	Best	.51[b]	.65[b]	.81[b]	.60[c]	.70[d]	.78[d]	.66[d]	.69[d]	.72[d]
AQLR	$\varphi^{(4)}$	2.35	.51[b]	.60[b]	.78[b]	.60[c]	.66[d]	.75[d]	.66[d]	.69[d]	.70[d]
AQLR	$\varphi^{(4)}$	1.87	.51[b]	.63[b]	.79[b]	.58[c]	.68[d]	.76[d]	.61[d]	.69[d]	.71[d]
AQLR	t-test	Best	.55	.67	.82	.60	.69	.78	.65	.69	.73
AQLR	t-test	2.35	.55	.64	.79	.60	.68	.76	.51	.68	.68
AQLR	t-test	1.87	.52	.66	.80	.56	.69	.77	.48	.69	.69
AQLR	t-test	Auto	.55	.67	.82	.59	.69	.78	.65	.69	.73
AQLR	MMSC	Best	.56[b]	.66[b]	.81[b]	.63	.69	.78	.65	.69	.73
AQLR	MMSC	2.35	.46[b]	.60[b]	.74[b]	.56	.68	.75	.64	.68	.70
AQLR	MMSC	1.87	.44[b]	.63[b]	.76[b]	.54	.69	.76	.59	.69	.71
Power	Envelope	—	.85	.85	.85	.80	.80	.80	.75	.75	.75

[a] κ = Best denotes the κ value that is best in terms of asymptotic average power. Unless stated otherwise, results are based on (40,000, 40,000, 40,000) critical-value, size-correction, and rejection-probability repetitions.

[b] Results are based on (1000, 1000, 1000) repetitions.

[c] Results are based on (2000, 2000, 2000) repetitions.

[d] Results are based on (5000, 5000, 5000) repetitions.

then the results above suggest that a useful approach is to map out the general features of the CS using the asymptotic normal version of the MMM/t-test/$\kappa = 2.35$ test, which is very fast to compute, and then switch to the bootstrap version of the AQLR/t-test/κauto test to find the boundaries of the CS more precisely.

Computation of the ELR/t-test/κauto bootstrap test using 10,000 bootstrap repetitions takes 3.1, 3.8, and 5.6 seconds when $p = 2$, 4, and 10, respectively, and $n = 250$. This is slower than the AQLR/t-test/κauto bootstrap test by factors of 9.3, 9.8, and 6.6.

S6.9. *Magnitude of RMS Critical Values*

Table S-XIV provides information on the magnitude of the recommended RMS critical value for the AQLR/t-test/κauto test when the size-correction factor $\widehat{\eta}$ is not included. (Recall that the RMS critical value equals $c_n(\theta, \widehat{\kappa}) + \widehat{\eta}$.) Specifically, Table S-XIV provides simulated values of the mean and standard deviation of the asymptotic distribution of the data-dependent quantile $c_n(\theta, \widehat{\kappa}) = q_{S_{2A}}(\varphi^{(1)}(\xi_n(\theta), \widehat{\Omega}_n(\theta)), \widehat{\Omega}_n(\theta))$ in various scenarios. The mean values in Table S-XIV can be compared with the values of the components $\eta_1(\delta)$ and $\eta_2(p)$ (given in Table I) of the size-correction factor $\widehat{\eta}$ ($= \eta_1(\widehat{\delta}_n(\theta)) + \eta_2(p)$) to see how large the quantile $c_n(\theta, \widehat{\kappa})$ is (on average) compared to the size-correction factor $\widehat{\eta}$.

The asymptotic distribution of $c_n(\theta, \widehat{\kappa})$ depends on h_1 and Ω. Table S-XIV considers the same three correlation matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos} as considered elsewhere in AB1 and above; see AB1 for their definitions. Table S-XIV considers h_1 vectors that consist of 0's and ∞'s. (Other h_1 vectors are of interest, but for brevity we do not consider them here.) When an element of h_1 equals ∞, the corresponding moment inequality is far from binding

TABLE S-XIV

MEAN AND STANDARD DEVIATION (SD) OF THE ASYMPTOTIC DISTRIBUTION OF THE DATA-DEPENDENT RMS CRITICAL VALUES EXCLUDING THE SIZE-CORRECTION FACTOR $\widehat{\eta}$[a]

	Ω_{Neg}		Ω_{Zero}		Ω_{Pos}	
Number of Zero's in h_1	Mean $c_n(\theta, \widehat{\kappa})$	SD $c_n(\theta, \widehat{\kappa})$	Mean $c_n(\theta, \widehat{\kappa})$	SD $c_n(\theta, \widehat{\kappa})$	Mean $c_n(\theta, \widehat{\kappa})$	SD $c_n(\theta, \widehat{\kappa})$
1	2.7	.00	2.7	.00	2.7	.00
2	5.0	.13	4.1	.53	3.5	.55
3	6.2	.11	5.2	.52	4.1	.68
4	7.5	.11	6.2	.54	4.5	.76
5	8.7	.13	7.2	.57	5.0	.82
6	9.8	.14	8.1	.59	5.3	.86
7	10.9	.16	8.9	.57	5.6	.89
8	11.9	.16	9.7	.63	5.9	.90
9	12.9	.17	10.6	.66	6.1	.92
10	13.8	.17	11.4	.68	6.3	.94

[a] Results are based on 40,000 simulation repetitions.

and the moment selection procedure detects this with probability 1 asymptotically and does not include this moment when computing $c_n(\theta, \widehat{\kappa})$. When an element of h_1 equals 0, the corresponding moment inequality is binding and the moment selection procedure includes this moment with high probability but not with probability 1, even asymptotically. (It is for this reason that $c_n(\theta, \widehat{\kappa})$ is random asymptotically.) In consequence, the asymptotic distribution depends on h_1 through the number of zeros in h_1 and through the submatrix of Ω that corresponds to the zeros in h_1. The matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos} are defined such that for any value of p, the submatrix of Ω of dimension equal to the number of zeros in h_1 is the same (provided $p \geq$ number of zeros in h_1). In consequence, the results of Table S-XIV hold for any value of p. For example, if $p = 20$, $\Omega = \Omega_{\text{Neg}}$, and the number of zeros in h_1 is 5, one obtains the same mean and standard deviation of the asymptotic distribution of $c_n(\theta, \widehat{\kappa})$ as when $p = 15$, $\Omega = \Omega_{\text{Neg}}$, and the number of zeros in h_1 is 5.

The results of Table S-XIV, combined with the magnitudes of the size-correction factors given in Table I, show that the size-correction factor $\widehat{\eta} = \eta_1(\widehat{\delta}_n(\theta)) + \eta_2(p)$ typically is small compared to $c_n(\theta, \widehat{\kappa})$, but is not negligible. For example, for $p = 10$, $\Omega = \Omega_{\text{Zero}} = I_{10}$, and $h_1 = (0, 0, 0, 0, 0, \infty, \infty, \infty, \infty, \infty)'$ (which corresponds to five moment inequalities being binding and five being very far from binding), the mean and standard deviation of the asymptotic distribution of $c_n(\theta, \widehat{\kappa})$ are 7.2 and .57, respectively, whereas the size-correction factor is .614.

S7. DETAILS CONCERNING THE NUMERICAL RESULTS

This section contains (i) the definition of the μ vectors used in AB1 (which define the alternatives over which asymptotic and finite-sample average power is computed), (ii) a description of some details concerning the assessment of the properties of the automatic method of choosing κ, (iii) a discussion of the determination and computation of the asymptotic power envelope, (iv) a discussion of the computation of the κ values that maximize asymptotic average power that are reported in Table II, (v) a description of the numerical computation of $\eta_2(p)$, which is part of the recommended size-correction function $\eta(\cdot)$, and (vi) a brief description of the computation of the finite-sample MNRP's.

S7.1. *μ Vectors*

For $p = 2$, the μ vectors considered are

$$\text{(S7.1)} \quad \mathcal{M}_2(I_2) = \{(-2.309, 0), (-2.309, 1), (-2.309, 2), (-2.309, 3), (-2.309, 4), (-2.309, 7), (-1.6263, -1.6263)\},$$

$$\mathcal{M}_2(\Omega_{\text{Neg}}) = \{(-1.001, 0), (-1.804, 1), (-2.303, 2), (-2.309, 3), (-2.309, 4), (-2.309, 7), (-0.5165, -0.5165)\},$$
$$\mathcal{M}_2(\Omega_{\text{Pos}}) = \mathcal{M}_2(I_2) \quad \text{except the last vector is} \quad (-2.0040, -2.0040).$$

The power envelope at each of these μ vectors is .750.

For $p = 4$, the μ vectors in $\mathcal{M}_4(I_4)$ are defined by

$$\begin{aligned} \text{(S7.2)} \quad & \mathcal{M}_4(\Omega) \\ & = \{(-\mu_1, -\mu_1, 1, 1), (-\mu_2, -\mu_2, 2, 2), (-\mu_3, -\mu_3, 3, 3), \\ & (-\mu_4, -\mu_4, 4, 4), (-\mu_5, -\mu_5, 7, 7), (-\mu_6, -\mu_6, 1, 7), \\ & (-\mu_7, -\mu_7, 2, 7), (-\mu_8, -\mu_8, 3, 7), (-\mu_9, -\mu_9, 4, 7), \\ & (-\mu_{10}, 1, 1, 1), (-\mu_{11}, 2, 2, 2), (-\mu_{12}, 3, 3, 3), (-\mu_{13}, 4, 4, 4), \\ & (-\mu_{14}, 7, 7, 7), (-\mu_{15}, 1, 1, 7), (-\mu_{16}, 2, 2, 7), (-\mu_{17}, 3, 3, 7), \\ & (-\mu_{18}, 4, 4, 7), (-\mu_{19}, -\mu_{19}, 0, 0), (-\mu_{20}, 0, 0, 0), \\ & (-\mu_{21}, 25, 25, 25), (-\mu_{22}, -\mu_{22}, 25, 25), (-\mu_{23}, -\mu_{23}, -\mu_{23}, 25), \\ & (-\mu_{24}, -\mu_{24}, -\mu_{24}, -\mu_{24})\}, \end{aligned}$$

and $\mu_j = 1.7388$ for $j = 1, \dots, 9, 19, 22$, $\mu_j = 2.4705$ for $j = 10, \dots, 18, 20, 21$, $\mu_{23} = 1.4242$, and $\mu_{24} = 1.2350$.

For $p = 4$, the μ vectors in $\mathcal{M}_4(\Omega_{\text{Neg}})$ are defined by (S7.2), and $\mu_1 = 0.5505$, $\mu_j = 0.5526$ for $j = 2, \dots, 5$, $\mu_6 = 0.5505$, $\mu_j = 0.5526$ for $j = 7, 8, 9$, $\mu_{10} = 1.8814$, $\mu_{11} = 2.4283$, $\mu_j = 2.4705$ for $j = 12, 13, 14, 17, 18, 21$, $\mu_{15} = 1.8814$, $\mu_{16} = 2.4283$, $\mu_{19} = 0.3176$, $\mu_{20} = 0.8624$, $\mu_{22} = 0.5526$, $\mu_{23} = 0.2607$, and $\mu_{24} = 0.1756$.

For $p = 4$, the μ vectors in $\mathcal{M}_4(\Omega_{\text{Pos}})$ are defined by (S7.2), and $\mu_j = 2.4047$ for $j = 1, \dots, 9, 19, 22$, $\mu_j = 2.4705$ for $j = 10, \dots, 18, 20, 21$, $\mu_{23} = 2.2628$, and $\mu_{24} = 2.1293$.

For $p = 4$, the power envelope at each of the μ vectors is .800.

For $p = k = 10$, $\mathcal{M}_{10}(\Omega)$ includes 40 vectors:

$$\begin{aligned} \text{(S7.3)} \quad & \mathcal{M}_{10}(\Omega) \\ & = \{(-\mu_1, -\mu_1, 1, \dots, 1), (-\mu_2, -\mu_2, 2, \dots, 2), \\ & (-\mu_3, -\mu_3, 3, \dots, 3), (-\mu_4, -\mu_4, 4, \dots, 4), \\ & (-\mu_5, -\mu_5, 7, \dots, 7), (-\mu_6, -\mu_6, 1, 1, 1, 7, \dots, 7), \\ & (-\mu_7, -\mu_7, 2, 2, 2, 7, \dots, 7), (-\mu_8, -\mu_8, 3, 3, 3, 7, \dots, 7), \\ & (-\mu_9, -\mu_9, 4, 4, 4, 7, \dots, 7), \\ & (-\mu_{10}, -\mu_{10}, -\mu_{10}, -\mu_{10}, 1, \dots, 1), \end{aligned}$$

$$
\begin{aligned}
&(-\mu_{11}, -\mu_{11}, -\mu_{11}, -\mu_{11}, 2, \ldots, 2),\\
&(-\mu_{12}, -\mu_{12}, -\mu_{12}, -\mu_{12}, 3, \ldots, 3),\\
&(-\mu_{13}, -\mu_{13}, -\mu_{13}, -\mu_{13}, 4, \ldots, 4),\\
&(-\mu_{14}, -\mu_{14}, -\mu_{14}, -\mu_{14}, 7, \ldots, 7),\\
&(-\mu_{15}, -\mu_{15}, -\mu_{15}, -\mu_{15}, 1, 1, 1, 7, 7, 7),\\
&(-\mu_{16}, -\mu_{16}, -\mu_{16}, -\mu_{16}, 2, 2, 2, 7, 7, 7),\\
&(-\mu_{17}, -\mu_{17}, -\mu_{17}, -\mu_{17}, 3, 3, 3, 7, 7, 7),\\
&(-\mu_{18}, -\mu_{18}, -\mu_{18}, -\mu_{18}, 4, 4, 4, 7, 7, 7),\\
&(-\mu_{19}, 1, \ldots, 1), (-\mu_{20}, 2, \ldots, 2), (-\mu_{21}, 3, \ldots, 3),\\
&(-\mu_{22}, 4, \ldots, 4), (-\mu_{23}, 7, \ldots, 7), (-\mu_{24}, 1, 1, 1, 7, \ldots, 7),\\
&(-\mu_{25}, 2, 2, 2, 7, \ldots, 7), (-\mu_{26}, 3, 3, 3, 7, \ldots, 7),\\
&(-\mu_{27}, 4, 4, 4, 7, \ldots, 7), (-\mu_{28}, -\mu_{28}, 0, \ldots, 0),\\
&(-\mu_{29}, -\mu_{29}, -\mu_{29}, -\mu_{29}, 0, \ldots, 0), (-\mu_{30}, 0, \ldots, 0),\\
&(-\mu_{31}, 25, \ldots, 25), (-\mu_{32}, -\mu_{32}, 25, \ldots, 25),\\
&(-\mu_{33}, -\mu_{33}, -\mu_{33}, 25, \ldots, 25),\\
&(-\mu_{34}, -\mu_{34}, -\mu_{34}, -\mu_{34}, 25, \ldots, 25),\\
&(-\mu_{35}, -\mu_{35}, -\mu_{35}, -\mu_{35}, -\mu_{35}, 25, \ldots, 25),\\
&(-\mu_{36}, \ldots, -\mu_{36}, 25, 25, 25, 25), (-\mu_{37}, \ldots, -\mu_{37}, 25, 25, 25),\\
&(-\mu_{38}, \ldots, -\mu_{38}, 25, 25), (-\mu_{39}, \ldots, -\mu_{39}, 25),\\
&(-\mu_{40}, \ldots, -\mu_{40})\big\}.
\end{aligned}
$$

For $p = 10$, the μ vectors in $\mathcal{M}_{10}(I_{10})$ are defined by (S7.3), and $\mu_j = 1.8927$ for $j = 1, \ldots, 9, 28, 32$, $\mu_j = 1.3360$ for $j = 10, \ldots, 18, 29, 34$, $\mu_j = 2.6817$ for $j = 19, \ldots, 27, 30, 31$, $\mu_{33} = 1.5463$, $\mu_{35} = 1.1963$, $\mu_{36} = 1.0893$, $\mu_{37} = 1.0099$, $\mu_{38} = 0.9465$, $\mu_{39} = 0.8882$, and $\mu_{40} = 0.8440$.

For $p = 10$, the μ vectors in $\mathcal{M}_{10}(\Omega_{\text{Neg}})$ are defined by (S7.3), and $\mu_j = 0.6016$ for $j = 1, \ldots, 9$, $\mu_j = 0.3475$ for $j = 10, \ldots, 18$, $\mu_{19} = 1.9847$, $\mu_{20} = 2.5835$, $\mu_j = 2.6817$ for $j = 21, 22, 23, 26, 27, 31$, $\mu_{24} = 1.9847$, $\mu_{25} = 2.5835$, $\mu_{28} = 0.5341$, $\mu_{29} = 0.3322$, $\mu_{30} = 1.1551$, $\mu_{32} = 0.6016$, $\mu_{33} = 0.4195$, $\mu_{34} = 0.3475$, $\mu_{35} = 0.2985$, $\mu_{36} = 0.2674$, $\mu_{37} = 0.2430$, $\mu_{38} = 0.2254$, $\mu_{39} = 0.2106$, and $\mu_{40} = 0.1993$.

For $p = 10$, the μ vectors in $\mathcal{M}_{10}(\Omega_{\text{Pos}})$ are defined by (S7.3), and $\mu_j = 2.6227$ for $j = 1, \ldots, 9$, $\mu_j = 2.4676$ for $j = 10, \ldots, 18$, $\mu_j = 2.6817$ for $j = 19, \ldots, 27$, $\mu_{28} = 2.6227$, $\mu_{29} = 2.4676$, $\mu_{30} = 2.6817$, $\mu_{31} = 2.6817$, $\mu_{32} =$

2.6227, $\mu_{33} = 2.5401$, $\mu_{34} = 2.4676$, $\mu_{35} = 2.4005$, $\mu_{36} = 2.3140$, $\mu_{37} = 2.2846$, $\mu_{38} = 2.2565$, $\mu_{39} = 2.2343$, and $\mu_{40} = 2.2066$.

For $p = 10$, the power envelope at each of the μ vectors is .850.

S7.2. *Automatic κ Power Assessment Details*

The 19 matrices Ω that are considered in Table S-I in Section S6.1.2 are Toeplitz matrices with elements on the diagonals given by the $(p-1)$ vectors ρ defined as follows. For $p = 2$, ρ takes the values for δ specified in Table S-I. For $p = 4, 10$, if $\delta \geq 0$, $\rho = (\delta, \ldots, \delta)$. For $p = 4$, if $\delta = -.99$, $\rho = (-.99, .97, -.95)$; if $\delta = -.975$, $\rho = (-.975, .94, -.90)$; if $\delta = -.95$, $\rho = (-.95, .9, -.8)$; and if $-.9 \leq \delta < 0$, $\rho = (\delta/(-.9)) \times (-.9, .7, -.5)$. For $p = 10$, if $\delta = -.99$, $\rho = (-.99, .97, -.95, .93, -.91, .89, -.87, .85, -.83)$; if $\delta = -.975$, $\rho = (-.975, .94, -.90, .86, -.82, .78, -.76, .74, -.72)$; if $\delta = -.95$, $\rho = (-.95, .9, -.8, .7, -.6, .5, -.4, .3, -.2)$; and if $-.9 \leq \delta < 0$, $\rho = (\delta/(-.9)) \times (-.9, .8, -.7, .6, -.5, .4, -.3, .2, -.1)$.

The randomly generated Ω matrices discussed in AB1 (that are used to assess the performance of the automatic κ method) have the following distributions. For $p = 2, 4$, and 10, the Ω matrices are i.i.d. with $\Omega = \mathrm{Diag}^{-1/2}(BB')BB'\,\mathrm{Diag}^{-1/2}(BB')$, where B is a $p \times p$ matrix with independent $N(2.5, 4)$ elements. For $p = 2, 4$, 500 Ω matrices are used. For $p = 10$, 250 Ω matrices are used.

The set of alternative hypothesis mean vectors μ, denoted $\mathcal{M}_p(\Omega)$ (used when assessing the asymptotic average power properties of the automatic κ method for Ω matrices that do not equal Ω_{Neg}, Ω_{Zero}, or Ω_{Pos}) contain linear combinations of μ vectors in $\mathcal{M}_p(\Omega_{\mathrm{Neg}})$, $\mathcal{M}_p(\Omega_{\mathrm{Zero}})$, and $\mathcal{M}_p(\Omega_{\mathrm{Pos}})$. Specifically, for a given matrix Ω, $\mathcal{M}_p(\Omega)$ is defined by (i) $\mathcal{M}_p(\Omega) = \mathcal{M}_p(\Omega_{\mathrm{Neg}})$ if $\delta(\Omega) \in [-1.0, -.90]$, (ii) if $\delta(\Omega) \in [-.9, 0]$, $\mathcal{M}_p(\Omega) = \{\mu : \mu = (1 + \delta/.9)\mu_{\mathrm{Zero},j} - (\delta/.9)\mu_{\mathrm{Neg},j}$ for $j = 1, \ldots, J_p\}$, where $\mu_{\mathrm{Zero},j}$ denotes the jth element of $\mathcal{M}_p(\Omega_{\mathrm{Zero}})$ and analogously for $\mathcal{M}_p(\Omega_{\mathrm{Neg}})$ and $\mathcal{M}_p(\Omega_{\mathrm{Pos}})$, and J_p denotes the numbers of elements in $\mathcal{M}_p(\Omega_{\mathrm{Zero}})$, (iii) if $\delta(\Omega) \in [0, .5]$, $\mathcal{M}_p(\Omega) = \{\mu : \mu = (1 - \delta/.5)\mu_{\mathrm{Zero},j} + (\delta/.5)\mu_{\mathrm{Pos},j}$ for $j = 1, \ldots, J_p\}$, and (iv) if $\delta(\Omega) \in [0.5, 1.0]$, $\mathcal{M}_p(\Omega) = \mathcal{M}_p(\Omega_{\mathrm{Pos}})$.

S7.3. *Asymptotic Power Envelope*

We obtain an upper bound on the asymptotic power envelope by considering the simple-versus-simple likelihood ratio (SSLR) test for the desired alternative distribution and some selected null distribution, with the critical value chosen so that the test has the desired asymptotic null rejection rate α at the specified null distribution. This method of obtaining an upper bound on a power envelope also has been exploited in different contexts by Andrews, Moreira, and Stock (2008) and Müller and Watson (2008). If the specified null distribution is such that the SSLR test has maximum rejection probability equal to

α over all null distributions, then the specified null distribution is least favorable and the SSLR test actually provides the asymptotic power envelope at the alternative distribution considered.

We assume that one observes $(n^{1/2}\overline{m}_n(\theta_0), \Sigma)$ and the null hypothesis H_0 is defined as in (S5.3). The simple alternative is $H_1 : F = F_n$, where F_n is a $n^{1/2}$ local alternative with asymptotic mean vector μ_{Alt}. Asymptotically, the distribution of $n^{1/2}\overline{m}_n(\theta_0)$ under the alternative is $N(\mu_{\text{Alt}}, \Sigma)$. We take the specified asymptotic null distribution to be $N(\mu_{\text{Null}}, \Sigma)$, where μ_{Null} is defined to minimize $(\mu - \mu_{\text{Alt}})'\Sigma^{-1}(\mu - \mu_{\text{Alt}})$ over $\mu \in R^p_{[+\infty]}$. In the numerical results reported below, we find that this choice of null distribution is least favorable. Thus, the upper bound on the asymptotic power envelope, up to numerical accuracy (based on 40,000 simulation repetitions), is the asymptotic power envelope.

S7.4. *Computation of the κ Values That Maximize Asymptotic Average Power*

Here we discuss the computation of the κ values that maximize asymptotic average power. These best κ values are used in the asymptotic power comparisons given in Table II. For all of the RMS tests in Table II, the best κ values are determined by grid search to an accuracy of .2. On a subset of cases this is found to be sufficiently small that the asymptotic average power is within .01 of the maximum based on a finer grid. The grid of κ values used for the t-test critical values and each test statistic considered are subsets of $\{.0, .2, \ldots, 3.6, 3.8, 4.2\}$ with lower and upper bounds on the elements of each subset being determined (by previous computations) to include the best κ value. For all of the test statistics considered, the average power values are well behaved as a function of κ, there is no difficulty in finding the best κ value, and the best κ value is within the interior of the range considered. To ensure the latter, for the AQLR/MMSC test, the following alternative grids are used in special cases: for $p = 4$ and Ω_{Neg}, $\{4.9, 5.1, \ldots, 6.5\}$; for $p = 10$ and Ω_{Neg}, $\{4.1, 4.4, \ldots, 6.5\}$. For the AQLR/$\varphi^{(3)}$ test, the following alternative grids are used in special cases: for $p = 2$ and Ω_{Neg}, $\{5.0, 5.5, \ldots, 10.5\}$; for $p = 4$ and Ω_{Neg}, $\{3.5, 4.0, \ldots, 10.5\}$; for $p = 10$ and Ω_{Neg}, $\{11.5, 12.0, \ldots, 14.0\}$.

S7.5. *Numerical Computation of $\eta_2(p)$*

The size-correction factor $\eta_2(p)$ is determined as follows. Let p and Ω be given. For given (h_1, Ω), we compute the .95 sample quantile of

$$\begin{aligned}(\text{S7.4}) \qquad &\big\{S_{2A}\big(\Omega^{1/2}Z_r + (h_1, 0_v), \Omega\big) \\ &\quad - q_{S_{2A}}\big(\varphi^{(1)}\big(\kappa^{-1}(\Omega)\big[\Omega^{1/2}Z_r + (h_1, 0_v)\big], \Omega\big), \Omega\big) \\ &\quad - \eta_1\big(\delta(\Omega)\big) : r = 1, \ldots, R\big\},\end{aligned}$$

where $Z_r \sim$ i.i.d. $N(0_k, I_k)$ for $r = 1, \dots, R$, where $R = 40{,}000$. Call the sample quantile $\eta_{h_1,\Omega}$. Up to simulation error, $\eta_{h_1,\Omega}$ is the smallest value that satisfies

$$\text{(S7.5)} \qquad \mathrm{CP}\big(h_1, \Omega, \eta_1\big(\delta(\Omega)\big) + \eta_{h_1,\Omega}\big) = 1 - \alpha.$$

The same simulated random variables $\{Z_r : r = 1, \dots, R\}$ are used for all (h_1, Ω) considered. The critical value $q_{S_{2A}}(\varphi^{(1)}(\kappa^{-1}(\Omega)[\Omega^{1/2} Z_r + (h_1, 0_v)], \Omega), \Omega)$ in (S7.4) is obtained by simulation for each r. (The number of simulation repetitions employed is R here too and the same random numbers are used for each r.)

Let $\mathcal{E}_1$ denote the set of all p vectors whose elements are 0's and ∞'s. By considering a variety of subcases, we find that size is (essentially) attained for $\mu \in \mathcal{E}_1$; see Section S7.6.[32] Thus, to obtain good numerical approximations, it suffices to restrict attention to maximization of $\eta_{h_1,\Omega}$ over $\mathcal{E}_1$, rather than over $R^p_{+,\infty}$. In addition, we approximate the maximization of $\eta_{h_1,\Omega}$ over the parameter space Ψ for Ω to a maximization of a finite set $\Psi^* \subset \Psi$. Given this, $\eta_2(p) \in R$ is defined to be

$$\text{(S7.6)} \qquad \sup_{h_1 \in \mathcal{E}_1, \Omega \in \Psi^*} \eta_{h_1,\Omega}.$$

For $p \leq 10$, the set Ψ^* is a set of correlation matrices that includes (i) 43 Toeplitz matrices Ω that are such that $\delta(\Omega)$ takes values in a grid between $-.99$ and $.99$,[33] and (ii) 500 randomly generated matrices Ω that are generated by $\Omega = \mathrm{Corr}(V)$, where $V = BB'$ and B is a $p \times p$ matrix with i.i.d. $N(0, 1)$ elements. As the number of randomly generated matrices Ω goes to infinity, the maximum of $\eta_{h_1,\Omega}$ over Ψ^* approaches the maximum of $\eta_{h_1,\Omega}$ over Ψ. Since the same underlying random variables $\{Z_r : r = 1, \dots, R\}$ are used for each (h_1, Ω) considered, an empirical process central limit theorem (CLT) guarantees that as R and the number of random matrices Ω considered go to infinity, the calculated critical values converge to the desired value $\eta_2(p)$ that satisfies

$$\text{(S7.7)} \qquad \inf_{h_1 \in \mathcal{E}_1, \Omega \in \Psi} \mathrm{CP}\big(h_1, \Omega, \eta_1\big(\delta(\Omega)\big) + \eta_2(p)\big) = 1 - \alpha.$$

S7.6. *Maximization Over μ Vectors in the Null Hypothesis*

In this section, we report the results of calculations that assess the impact of using the restricted set of null mean vectors $\mathcal{E}_1$, rather than all of $R^p_{+,\infty}$ when

[32]In the numerical results, we use 25 in place of ∞, but there is no sensitivity to this choice. Results for 15 and 35 give identical results because when the mean is sufficiently large, say 15, 25, or 35, the probability of observing a sample mean that is negative is so close to zero that the precise value of the mean does not affect the rejection probabilities.

[33]For any given value of $\delta = \delta(\Omega)$, these 43 matrices are defined just as the 19 Toeplitz matrices are defined in Section S7.2. The $\delta(\Omega)$ values considered are the 43 values specified by the endpoints for δ in Table I, but including $-.99$ and excluding -1.0 and 1.0.

computing (i) $\eta_2(p)$, (ii) the asymptotic MNRP's for tests that employ the asymptotically best κ values ($\kappa =$ best), and (iii) the finite-sample results of AB1 and those reported above.

S7.6.1. *Computation of* $\eta_2(p)$

Here we assess the impact of using $\mathcal{E}_1$, rather than all of $R^p_{+,\infty}$ when computing $\eta_2(p)$. First, for the AQLR/t-test/κauto test, we compute the difference between the asymptotic MNRP when the maximum is over μ vectors in $\mathcal{E}_1$ with the asymptotic MNRP when the maximum is over several larger sets of μ vectors. The larger sets include (i) three different grids of fixed μ vectors, which are described in the following subsection, and (ii) 1000 randomly generated μ vectors plus $\mathcal{E}_1$.[34] These results are for the 43 fixed Toeplitz variance matrices that are described in Section S7.5. The results are given in Table S-XV.

Second, for 260 randomly generated variance matrices, we compute the differences in asymptotic MNRP when the maximum is over $\mathcal{E}_1$ and when the maximum is over 1000 randomly generated μ vectors (with the same distribution as in the previous paragraph) plus $\mathcal{E}_1$.[35] These results are given in Table S-XVI.

Third, we report results for the variance matrix, Ω_{LF_1}, that is found to be least favorable (LF) over the 43 fixed Toeplitz variance matrices used in the compu-

TABLE S-XV

MAXIMUM DIFFERENCES IN NOMINAL .05 ASYMPTOTIC MNRP'S DUE TO DIFFERENT SETS OF MEAN VECTORS μ USED IN THE COMPUTATIONS WITH 43 TOEPLITZ VARIANCE MATRICES[a]

p	$\mathcal{E}_1$ versus Full Grid Plus $\mathcal{E}_1$	$\mathcal{E}_1$ versus Large Partial Grid Plus $\mathcal{E}_1$	$\mathcal{E}_1$ versus Small Partial Grid Plus $\mathcal{E}_1$	$\mathcal{E}_1$ versus 1000 Random μ Plus $\mathcal{E}_1$
2	.0001	.0005	.0005	.0004
3	.0005	.0000	.0000	.0005
4	.0003	.0000	.0000	.0005
5	.0000	.0000	.0000	.0000
6	.0000	.0000	.0000	.0000
7	.0000	.0000	.0000	.0000
8	.0000	.0000	.0000	.0000
9	—	.0000	.0000	.0000
10	—	.0000	.0000	.0000

[a]The maximum is over the 43 Toeplitz variance matrices.

[34] The random μ vectors have elements that are i.i.d. with probability .5 of equalling 0 and probability .5 of being uniform on [0, 8].

[35] The variance matrices are generated via $V = BB'$, where B is a $p \times p$ matrix with i.i.d. $N(0, 1)$ elements.

TABLE S-XVI

MAXIMUM DIFFERENCES IN NOMINAL .05 ASYMPTOTIC MNRP'S DUE TO DIFFERENT SETS OF MEAN VECTORS μ USED IN THE COMPUTATIONS WITH 260 RANDOM VARIANCE MATRICES[a]

p	$\mathcal{E}_1$ versus 1000 Random μ Vectors Plus $\mathcal{E}_1$
3	.0000
4	.0000
5	.0000
6	.0000
7	.0000
8	.0025
9	.0026
10	.0024

[a]The maximum is over the 260 variance matrices.

tation of $\eta_2(p)$ for $p = 3, \ldots, 10$.[36] We also report results for the variance matrix, Ω_{LF_2}, that is found to be least favorable (LF) over the 500 randomly generated variance matrices used in the computation of $\eta_2(p)$ for $p = 3, \ldots, 10$.[37] For these two variance matrices and $p = 3, \ldots, 10$, we report the differences in asymptotic MNRP when the maximum is over $\mathcal{E}_1$ and when the maximum is over 100,000 randomly generated μ vectors (with the same distribution as above) plus $\mathcal{E}_1$. The results are given in Table S-XVII.

Fourth, in Table S-XVIII, we report the effect of potential inaccuracy in $\eta_2(p)$ on the asymptotic MNRP's of the AQLR/t-test/κauto test.

All results are based on 40,000 simulation repetitions for the critical-value calculations and the rejection probabilities.

Definitions of the Grids of μ Vectors. The three sets of fixed grids of μ vectors considered are (i) a full grid, (ii) a large partial grid, and (iii) a small partial grid. The partial grids are considered because a finer mesh can be used with these grids than with a full grid. A full grid is not computable for $p = 9$ and 10 because there are too many μ vectors. The grids are defined as follows.

- *Full grid of μ vectors*. This set of μ vectors consists of p vectors whose elements (i) all come from a vector, GridVec, of dimension #grid and (ii) contain

[36]That is, Ω_{LF_1} is the matrix that yields the largest MNRP over the 43 matrices when the MNRP is computed using all μ vectors with 0's and ∞'s, and $\eta_2(p)$ is set equal to 0. This matrix is found to be I_p for seven of the eight values of p and within .0001 of being LF for the other case. So, for simplicity, we take $\Omega_{\mathrm{LF}_1} = I_p$ for $p = 3, \ldots, 10$.

[37]That is, Ω_{LF_2} is the matrix that yields the largest MNRP over the 500 random matrices used to compute $\eta_2(p)$ when the MNRP is computed using all μ vectors with 0's and ∞'s and $\eta_2(p)$ is set equal to 0.

TABLE S-XVII

DIFFERENCES IN NOMINAL .05 ASYMPTOTIC MNRP'S DUE TO DIFFERENT SETS OF MEAN VECTORS μ USED IN THE COMPUTATIONS

	$\mathcal{E}_1$ versus 100,000 Random μ Vectors Plus $\mathcal{E}_1$	
p	Difference for $\Omega = \Omega_{\text{LF}_1}$	Difference for $\Omega = \Omega_{\text{LF}_2}$
3	.0000	.0000
4	.0000	.0000
5	.0000	.0000
6	.0000	.0000
7	.0000	.0000
8	.0000	.0000
9	.0000	.0000
10	.0000	.0000

at least one zero. The number of such vectors is $(\#\text{grid})^p - (\#\text{grid} - 1)^p$, where #grid is the number of elements in GridVec. The GridVec vectors used with the full grid are, for $p = 2, 3$, $\#\text{grid} = 24$ and GridVec $= \{0, .05, .1, .2, .3, .5, .75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7, 8, 9, 10, 15, 20\}$; for $p = 4$, $\#\text{grid} = 18$ and GridVec $= \{0, .25, .5, .75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 9, 10\}$; for $p = 5$, $\#\text{grid} = 8$ and GridVec $= \{0, .5, 1, 1.5, 2, 2.5, 3, 4\}$; for $p = 6$, $\#\text{grid} = 5$ and GridVec $= \{0, 1, 2, 3, 4\}$; for $p = 7$, $\#\text{grid} = 4$ and GridVec $= \{0, 1, 2.5, 4\}$; and for $p = 8$, $\#\text{grid} = 3$ and GridVec $= \{0, 2.5, 3.5\}$.

- *Large partial grid of μ vectors*. This set of μ vectors consists of p vectors whose elements (i) all come from a vector, GridVec, of dimension #grid, (ii) are nondecreasing, and (iii) contain at least one zero. For example, if $p = 4$ and GridVec $= \{0, 1, 2, 3, 4\}$, then $\#\text{grid} = 5$ and the μ vectors are of the form

TABLE S-XVIII

DIFFERENCES IN MNRP'S WHEN $\eta_2(p)$ IS INCREASED OR DECREASED BY 25% OR 50%

p	Ω	+25%	−25%	+50%	−50%
3	Ω_{Zero}	.0009	.0006	.0019	.0017
4	Ω_{Neg}	.0013	.0012	.0022	.0022
4	Ω_{Zero}	.0011	.0011	.0014	.0026
4	Ω_{Pos}	.0010	.0010	.0014	.0023
6	Ω_{Zero}	.0012	.0016	.0025	.0036
8	Ω_{Zero}	.0018	.0018	.0033	.0041
10	Ω_{Neg}	.0022	.0022	.0042	.0044
10	Ω_{Zero}	.0020	.0030	.0039	.0052
10	Ω_{Pos}	.0024	.0030	.0046	.0054

$(0, 0, 0, 0), \ldots, (0, 0, 2, 3), (0, 0, 2, 4), (0, 0, 3, 4), \ldots, (0, 4, 4, 4)$. The number of such vectors does not have a simple closed form expression.

The GridVec vectors used with the large partial grid are, for $p = 2, 3$, and 4, #grid = 24 and GridVec = $\{0, .05, .1, .2, .3, .5, .75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7, 8, 9, 10, 15, 20\}$; for $p = 5$, #grid = 11 and GridVec = $\{0, .5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7\}$; for $p = 6$, #grid = 8 and GridVec = $\{0, 1, 2, 3, 4, 5, 6, 7\}$; for $p = 7$, #grid = 7 and GridVec = $\{0, 1, 2, 3, 4, 5, 6\}$; for $p = 8$, #grid = 6 and GridVec = $\{0, 1, 2, 4, 5, 6\}$; for $p = 9$, #grid = 5 and GridVec = $\{0, 1, 2, 4, 6\}$; and for $p = 10$, #grid = 4 and GridVec = $\{0, 2, 4, 6\}$.

• *Small partial grid of μ vectors*. This set of μ vectors consists of p vectors whose elements (i) all come from a vector, GridVec, of dimension #grid, (ii) take only two different values, (iii) are nondecreasing, and (iv) contain at least one zero (to guarantee that the vector is on the boundary of the null hypothesis). For example, if $p = 4$ and GridVec = $\{0, 1, 2, 3, 4\}$, then #grid = 5 and the μ vectors are of the form $(0, 0, 0, 0), (0, 0, 0, 1), \ldots, (0, 0, 3, 3), (0, 0, 4, 4), (0, 1, 1, 1), \ldots, (0, 4, 4, 4)$. The number of such vectors is $(p - 1) * (\#\text{grid} - 1) + 1$.

The GridVec vector used with the small partial grid is $\forall p = 2, \ldots, 10$, #grid = 24 and GridVec = $\{0, .05, .1, .2, .3, .5, .75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7, 8, 9, 10, 15, 20\}$.

MNRP Difference Results. Tables S-XV, S-XVI, and S-XVII provide the results. Table S-XV shows that the differences in asymptotic MNRP's of the AQLR/t-test/κauto test from maximizing over $\mathcal{E}_1$ versus the full grid is .0005 or less. The differences in MNRP's from maximizing over $\mathcal{E}_1$ versus the large and small partial grids are very small, being .0000 in all cases with $p \geq 3$ and .0005 for $p = 2$. Table S-XVI shows that the difference in MNRP's from maximizing over $\mathcal{E}_1$ versus 1000 random μ vectors and 260 random Ω matrices is .0000 for $p \leq 7$ and always .0026 or less.

For computation of the $\eta_2(p)$ values, what is most relevant is the difference between the MNRP over $\mathcal{E}_1$ and $R^p_{+,\infty}$ evaluated at the least favorable variance matrix. In consequence, Table S-XVII reports the differences for the two LF matrices Ω_{LF_1} and Ω_{LF_2} defined above. These results are based on 100,000 randomly generated μ vectors. In all 16 cases considered, the differences are .0000.[38]

In sum, extensive simulations fail to find a noticeable effect of restricting the MNRP calculations for the AQLR/t-test/κauto test to μ vectors in $\mathcal{E}_1$ compared to calculations based on broader sets of μ vectors in $R^p_{+,\infty}$.

[38]One might wonder why the simulated differences are not small but positive, due to simulation error, even if the true differences are zero. We believe the reason is due to the high positive correlation between the two statistics whose difference is being computed. Given high positive correlation, the simulation error is small.

Potential Effects of Inaccuracy in $\eta_2(p)$. Next, we report the potential effects of inaccuracy in the calculation of $\eta_2(p)$. Table S-XVIII provides the differences in MNRP's when $\eta_2(p)$ is given by the value in Table I compared to when it is increased or decreased by 25% or 50%. These results answer the question, "How much would the asymptotic MNRP's change if the $\eta_2(p)$ values in Table I are inaccurate by as much as 25% or 50%?" The results are based on (40,000, 40,000) critical-value and null rejection-probability repetitions.

Table S-XVIII shows that even relatively large percentage changes in $\eta_2(p)$ have fairly small effects on the MNRP's. With a change of ±50%, the difference in MNRP is .0054 or less in all cases.

S7.6.2. *Computation of MNRP's for Tests Based on Best* κ *Values*

Table II reports asymptotic power comparisons for tests using (infeasible) critical values that employ the asymptotically best κ values ($\kappa =$ best). The MNRP's for these tests and the size correction that is based on the MNRP's are computed using all mean vectors μ in $\mathcal{E}_1$. In this section, we report numerical results designed to see whether the restriction to $\mathcal{E}_1$, rather than $R^p_{+,\infty}$, affects the results. We compute asymptotic MNRP differences of the types reported in Tables S-XV and S-XVI, but for tests other than the AQLR/t-test/κauto test. We compute results for a subset of the cases considered in Tables S-XV and S-XVI.[39] (Unlike the results reported in these tables, only the three variance matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos} that appear in Table II are considered here.)

We discuss the computationally fast and slow tests separately. The computationally fast tests are the MMM, Max, SumMax, and AQLR test statistics combined with the t-test/κbest critical values. The slow tests are the AQLR test statistic combined with the $\varphi^{(2)}/\kappa$best, $\varphi^{(3)}/\kappa$best, and $\varphi^{(4)}/\kappa$best critical values. The AQLR statistic combined with the MMSC critical value is discussed separately.

For the fast tests and the AQLR/MMSC/κbest test, we compute results for all of the cases in Tables S-XV and S-XVI for $p = 2$, 4, and 10, and Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}. For the slow tests, we compute results for the full grid for $p = 2$ and 4 and for 1000 random μ vectors for $p = 10$.

For the fast tests, the number of simulations used is (40,000, 40,000, 40,000) for the critical values, size correction, and rejection probabilities, respectively, in all cases considered. For the slow tests, (10,000, 10,000, 10,000) repetitions are used for $p = 2$, (1000, 1000, 1000) are used for $p = 4$, and (2000, 2000, 2000) repetitions are used for $p = 10$. (More repetitions are used here for $p = 10$ than $p = 4$ because fewer μ vectors are considered.) For

[39]Even if it was the case that considering $\mathcal{E}_1$, rather than $R^p_{+,\infty}$, affects the results for the κbest tests, the comparisons in Table II are still meaningful because they provide an upper bound on the size-corrected power of the κbest tests. Hence, comparisons between the recommended AQLR/t-test/κauto test and the various infeasible κbest tests in Table II are still quite informative. In any event, the numerical results given below indicate that there is not a significant effect.

the AQLR/MMSC/κbest test, (40,000, 40,000, 40,000) repetitions are used for $p = 2$ and 4 and (10,000, 10,000, 10,000) repetitions are used for $p = 10$.

The results are easy to state, so no table is provided. In all cases but 5 out of 192, the difference between the MNRP computed over $\mathcal{E}_1$ and over the larger set is found to be .0000. The five exceptions are the following. For the AQLR/$\varphi^{(j)}$/κbest for $j = 2, 3, 4$ with $p = 4$, Ω_{Pos}, and the full grid, the differences obtained are .0040, .0030, and .0030, respectively. For the AQLR/MMSC/κbest test with $p = 4$ and Ω_{Neg} using 1000 random μ and the full grid, the differences are .0034 and .0037, respectively.

In conclusion, we do not find evidence that the restriction to the set $\mathcal{E}_1$, rather than $R^p_{+,\infty}$, has a significant effect on the MNRP results for the tests based on $\kappa = \text{best}$ critical values. The evidence against there being such an effect is fairly strong for $p = 2$ and 4 because of the full grid results that are reported. It is less strong for $p = 10$ because a full grid could not be considered due to computational constraints.

S7.6.3. *Computation of Finite-Sample MNRP's*

The finite-sample MNRP's reported in Tables III and S-IV–S-VI are computed for a given covariance matrix Ω by maximizing over the null mean vectors $\mu \in \mathcal{E}_1$, where $\mathcal{E}_1$ denotes the set of all p vectors whose elements are 0's and ∞'s. MNRP-corrected critical values also are computed using $\mathcal{E}_1$. The first justification for using $\mathcal{E}_1$, rather than the larger set $R^p_{+,\infty}$, is the small differences found between $\mathcal{E}_1$ and larger sets of μ vectors in the asymptotic scenario; see Sections S7.6.1 and S7.6.2. The second justification is a finite-sample analysis that is analogous to that in Section S7.6.1 for the recommended tests AQLR/t-test/κauto/Bt and AQLR/t-test/κauto/Nm. We report the results here.

Table S-XIX reports the differences in MNRP's of these tests when they are computed over $\mathcal{E}_1$ compared to when they are computed over a full grid of μ vectors plus $\mathcal{E}_1$. The grid size (#grid) is 24 for $p = 2$ and 10 for $p = 4$. The GridVec's are $\{0, .05, .1, .2, .3, .5, .75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7, 8, 9, 10, 15, 20\}$ for $p = 2$ and $\{0, .25, .5, .75, 1, 1.5, 2, 3, 4, 6\}$ for $p = 4$. The sample size is $n = 250$, as in AB1. The same three matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos} are employed as in AB1. Results are reported for the same three distributions $N(0, 1)$, t_3, and χ^2_3 (all rescaled to have mean 0 and variance 1) as considered in AB1. For $p = 2$, the results use 5000 critical-value simulation repetitions and 5000 null rejection-probability simulations repetitions. For $p = 4$, 1000 and 1000 repetitions, respectively, are used.

Table S-XIX shows that in all of the 36 cases considered the difference in MNRP's is found to be .0000. Hence, these results are consistent with the least favorable null rejection vector being in $\mathcal{E}_1$ for the cases considered.

Next, Table S-XX reports MNRP difference results for $p = 10$. For $p = 10$, it is not possible to compute results for a full grid of μ vectors. Instead, we report results for the same large partial grid, small partial grid, and 1000 randomly generated μ vectors as described in Section S7.6.1. We report results for the

TABLE S-XIX

DIFFERENCES IN NOMINAL .05 FINITE-SAMPLE ($n = 250$) MNRP'S DUE TO DIFFERENT SETS OF NULL MEAN VECTORS μ USED IN THE COMPUTATIONS

p	Test	Distribution	Variance Matrix	$\mathcal{E}_1$ versus Full Grid Plus $\mathcal{E}_1$
2	AQLR/t-test/κauto/Bt	$N(0,1)$	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
		t_3	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
		χ^2_3	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
2	AQLR/t-test/κauto/Nm	$N(0,1)$	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
		t_3	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
		χ^2_3	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
4	AQLR/t-test/κauto/Bt	$N(0,1)$	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
		t_3	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
		χ^2_3	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
4	AQLR/t-test/κauto/Nm	$N(0,1)$	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
		t_3	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000
		χ^2_3	Ω_{Neg}	.0000
			Ω_{Zero}	.0000
			Ω_{Pos}	.0000

TABLE S-XX

DIFFERENCES IN NOMINAL .05 FINITE-SAMPLE ($n = 250$) MNRP'S DUE TO DIFFERENT SETS OF NULL MEAN VECTORS μ USED IN THE COMPUTATIONS

p	Test	Distribution	Variance Matrix	$\mathcal{E}_1$ versus Large Partial Grid Plus $\mathcal{E}_1$	$\mathcal{E}_1$ versus Small Partial Grid Plus $\mathcal{E}_1$	$\mathcal{E}_1$ versus 1000 Random μ Plus $\mathcal{E}_1$
10	AQLR/t-test/	$N(0,1)$	Ω_{Neg}	.0000	.0000	.0000
	κauto/Bt		Ω_{Zero}	.0000	.0000	.0000
			Ω_{Pos}	.0000	.0000	.0000
		t_3	Ω_{Neg}	.0000	.0000	.0000
			Ω_{Zero}	.0000	.0000	.0000
			Ω_{Pos}	.0000	.0000	.0000
		χ^2_3	Ω_{Neg}	.0000	.0000	.0000
			Ω_{Zero}	.0000	.0000	.0000
			Ω_{Pos}	.0000	.0000	.0000
10	AQLR/t-test/	$N(0,1)$	Ω_{Neg}	.0000	.0000	.0000
	κauto/Nm		Ω_{Zero}	.0000	.0000	.0000
			Ω_{Pos}	.0000	.0000	.0000
		t_3	Ω_{Neg}	.0000	.0000	.0000
			Ω_{Zero}	.0000	.0000	.0000
			Ω_{Pos}	.0000	.0000	.0000
		χ^2_3	Ω_{Neg}	.0000	.0000	.0000
			Ω_{Zero}	.0000	.0000	.0000
			Ω_{Pos}	.0000	.0000	.0000

same sample size, variance matrices, and distributions as in Table S-XIX. The results use 1000 critical-value simulation repetitions and 1000 null rejection-probability simulation repetitions.

The results of Table S-XX for $p = 10$ are the same as those in Table S-XIX for $p = 2$ and $p = 4$. In all cases, the difference in MNRP's is .0000. So, with $p = 10$ too, the results are consistent with the least favorable null rejection vector being in $\mathcal{E}_1$ for the cases considered.

S8. COMPUTER PROGRAMS

This section lists the GAUSS computer programs that were used to carry out the numerical results reported in AB1 and above. These programs are available as additional Supplemental Material on the Econometric Society website. Also available on the Econometric Society website is the translation of some of these programs into Matlab.

- rmsprg_final: This program is designed for users who want to carry out a test using the recommended RMS test (or any of several related tests). It was not used to compute any of the numerical results.

• etaprg1_final: This program was used when computing the $\eta_2(p)$ values based on 500 randomly generated variance matrices.

• etaprg2_final: This program was used when computing the $\eta_2(p)$ values based on 43 fixed variance matrices.

• finsamp3_final: This programs was used to compute all of the finite-sample results reported in Tables III, S-IV, S-V, and S-VI.

• kappaprg_final: This program was used for many purposes, including (i) computation of the best ε value for use with the AQLR statistic, as reported in Table S-II; (ii) assessment of how well the choice $\varepsilon = .012$ based on $p = 2$ performs for $p = 4, 10$, as reported in Table S-II; (iii) determination of the best κ values and the corresponding $\eta_1(\delta)$ values for the AQLR/t-test/κauto test for $p = 2$, as reported in Table I; (iv) asymptotic MNRP and power comparisons based on best κ values for a variety of test statistics and the three main variance matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}, as reported in Tables II, S-XII, and S-XIII; (v) determination of the asymptotic MNRP's and power for a variety of tests when $\kappa = 2.35$ and $\kappa = 1.87$ (which are BIC and HQIC values, respectively), as reported in Tables S-X, S-XI, S-XII, and S-XIII; (vi) asymptotic power comparisons for a variety of tests and the power envelope for 19 Ω matrices, as reported in Tables S-I and S-IX; (vii) asymptotic power comparisons for a variety of tests for singular variance matrices, as reported in Table S-III; (viii) determination of the pure/constant ELR critical values for the ELR tests whose MNRP's and power are reported in Tables S-XII and S-XIII; (ix) determination of the asymptotic MNRP's and power for the ELR test with pure/constant critical values, as reported in Tables S-XII and S-XIII; and (x) changes in asymptotic MNRP's when $\eta_2(p)$ is increased or decreased by 25% or 50%, as reported in Table S-XVIII.

• powprg_final: This program was used to compute the difference in average asymptotic power between the AQLR/t-test/κauto and AQLR/t-test/κbest tests for 500 randomly generated Ω matrices, as reported in Table S-I and Section S6.1.2.

• rmsprg_fs_short_final: This program was not used to compute any of the results reported in AB1 or this Supplement. It is a shortened version of finsamp3_final that computes finite-sample results for the main tests of interest: AQLR/t-test/κauto implemented using the asymptotic distribution or the bootstrap and MMM/t-test/$\kappa = 2.35$.

• sizediffprg11_final: This program computes the differences in MNRP's for a variety of tests when the mean vectors μ considered are (i) all vectors consisting of 0's and ∞'s and (ii) these μ vectors plus randomly generated μ vectors, as reported in Table S-XVI and Section S7.6.2.

• sizediffprg22_final: This program computes the differences in asymptotic MNRP's for a variety of tests when the mean vectors μ considered are (i) all vectors consisting of 0's and ∞'s, and (ii) these μ vectors plus a full grid of μ vectors, or a large partial grid of μ vectors, or a small partial grid of μ vectors, as reported in Table S-XV, the first column of results in Table S-XVII, and Section S7.6.2.

• sizediffprg22_LF_final: This program computes the same differences as sizediffprg22_final but for the least favorable variance matrices that were determined when calculating $\eta_2(p)$ using 500 random variance matrices for $p = 3, \ldots, 10$. These results are reported in the last column of Table S-XVII.

• sizediffprg22_finsamp_final: This program computes the differences in finite-sample MNRP's for a variety of tests when the mean vectors μ considered are (i) all vectors consisting of 0's and ∞'s, and (ii) these μ vectors plus a full grid of μ vectors, or a large partial grid of μ vectors, or a small partial grid of μ vectors, as reported in Tables S-XIX and S-XX.

S9. ALTERNATIVE PARAMETRIZATION AND PROOFS

This section provides proofs of the results given in Section S5. In addition, the first subsection gives an alternative parametrization of the moment inequality/equality model to that given in (S2.1). This parametrization is conducive to the calculation of the asymptotic properties of CS's and tests. It was first used in AG. The first subsection also specifies the parameter space for the case of dependent observations and for the case where a preliminary estimator of a parameter τ appears. The second subsection provides proofs of the results stated in the paper.

S9.1. *Alternative Parametrization*

In this section, we specify a one-to-one mapping between the parameters (θ, F) with parameter space $\mathcal{F}$ and a new parameter $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ with corresponding parameter space Γ. The latter parametrization is amenable to establishing the asymptotic uniformity results of Theorem 1.

As stated above, the true value θ_0 ($\in \Theta \subset R^d$) is assumed to satisfy the moment conditions in (S2.1). For the case where the sample moment functions depend on a preliminary estimator $\widehat{\tau}_n(\theta)$ of an identified parameter vector τ with true parameter τ_0, we define $m_j(W_i, \theta) = m_j(W_i, \theta, \tau_0)$, $m(W_i, \theta) = (m_1(W_i, \theta, \tau_0), \ldots, m_k(W_i, \theta, \tau_0))'$, $\overline{m}_{n,j}(\theta) = n^{-1}\sum_{i=1}^{n} m_j(W_i, \theta, \widehat{\tau}_n(\theta))$, and $\overline{m}_n(\theta) = (\overline{m}_{n,1}(\theta), \ldots, \overline{m}_{n,k}(\theta))'$. (Hence, in this case, $\overline{m}_n(\theta) \neq n^{-1} \times \sum_{i=1}^{n} m(W_i, \theta)$.)

We define $\gamma_1 = (\gamma_{1,1}, \ldots, \gamma_{1,p})' \in R_+^p$ by writing the moment inequalities in (S2.1) as moment equalities

$$\text{(S9.1)} \qquad \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0 \quad \text{for} \quad j = 1, \ldots, p,$$

where $\sigma_{F,j}^2(\theta)$ is the variance of the asymptotic distribution of $n^{1/2}\overline{m}_{n,j}(\theta)$ under (θ, F). Also, let $\Omega = \Omega(\theta, F) = \text{AsyCorr}_F(n^{1/2}\overline{m}_n(\theta))$ denote the correlation matrix of the asymptotic distribution of $n^{1/2}\overline{m}_n(\theta)$ under (θ, F). When no preliminary estimator of a parameter τ appears, $\sigma_{F,j}^2(\theta) = \lim_{n\to\infty} \text{Var}_F(n^{1/2} \times \overline{m}_{n,j}(\theta))$ and $\Omega(\theta, F) = \lim_{n\to\infty} \text{Corr}_F(n^{1/2}\overline{m}_n(\theta))$, where $\text{Var}_F(n^{1/2}\overline{m}_{n,j}(\theta))$

and $\mathrm{Corr}_F(n^{1/2}\overline{m}_n(\theta))$ denote the finite-sample variance of $n^{1/2}\overline{m}_{n,j}(\theta)$ and correlation matrix of $n^{1/2}\overline{m}_n(\theta)$ under (θ, F), respectively. Let $\gamma_2 = (\gamma_{2,1}, \gamma_{2,2}) = (\theta, \mathrm{vech}_*(\Omega(\theta, F))) \in R^q$, where $\mathrm{vech}_*(\Omega)$ denotes the vector of elements of Ω that lie below the main diagonal, $q = d + k(k-1)/2$, and $\gamma_3 = F$.

For i.i.d. observations and no preliminary estimator of a parameter τ, the parameter space for γ is defined by $\Gamma = \{\gamma = (\gamma_1, \gamma_2, \gamma_3)$: for some $(\theta, F) \in \mathcal{F}$, where $\mathcal{F}$ is defined in (S2.2), γ_1 satisfies (S9.1), $\gamma_2 = (\theta, \mathrm{vech}_*(\Omega(\theta, F)))$, and $\gamma_3 = F\}$.

For dependent observations and for sample moment functions that depend on a preliminary estimator $\widehat{\tau}_n(\theta)$, we specify the parameter space Γ for the moment inequality model using a set of high-level conditions. To verify the high-level conditions using primitive conditions, one has to specify an estimator $\widehat{\Sigma}_n(\theta)$ of the asymptotic variance matrix $\Sigma(\theta)$ of $n^{1/2}\overline{m}_n(\theta)$. For brevity, we do not do so here. Since there is a one-to-one mapping from γ to (θ, F), Γ also defines the parameter space $\mathcal{F}$ of (θ, F). Let Ψ be a specified set of $k \times k$ correlation matrices. The parameter space Γ is defined to include parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3) = (\gamma_1, (\theta, \gamma_{2,2}), F)$ that satisfy

$$\text{(S9.2)} \quad \text{(i)} \quad \theta \in \Theta,$$

$$\text{(ii)} \quad \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0 \quad \text{for} \quad j = 1, \ldots, p,$$

$$\text{(iii)} \quad E_F m_j(W_i, \theta) = 0 \quad \text{for} \quad j = p+1, \ldots, k,$$

$$\text{(iv)} \quad \sigma_{F,j}^2(\theta) = \mathrm{AsyVar}_F\big(n^{1/2}\overline{m}_{n,j}(\theta)\big) \quad \text{exists and lies in } (0, \infty) \quad \text{for} \quad j = 1, \ldots, k,$$

$$\text{(v)} \quad \mathrm{AsyCorr}_F\big(n^{1/2}\overline{m}_n(\theta)\big) \quad \text{exists and equals } \Omega_{\gamma_{2,2}} \in \Psi,$$

$$\text{(vi)} \quad \{W_i : i \geq 1\} \quad \text{are stationary under } F,$$

where $\gamma_1 = (\gamma_{1,1}, \ldots, \gamma_{1,p})'$ and $\Omega_{\gamma_{2,2}}$ is the $k \times k$ correlation matrix determined by $\gamma_{2,2}$.[40] Furthermore, Γ must be restricted by enough additional conditions such that under any sequence $\{\gamma_{n,h} = (\gamma_{n,h,1}, (\theta_{n,h}, \mathrm{vech}_*(\Omega_{n,h})), F_{n,h}) : n \geq 1\}$ of parameters in Γ that satisfies $n^{1/2}\gamma_{n,h,1} \to h_1$ and $(\theta_{n,h}, \mathrm{vech}_*(\Omega_{n,h})) \to h_2 = (h_{2,1}, h_{2,2})$ for some $h = (h_1, h_2) \in R^p_{+,\infty} \times R^q_{[\pm\infty]}$, we have

$$\text{(S9.3)} \quad \text{(vii)} \quad A_n = (A_{n,1}, \ldots, A_{n,k})' \to_d Z_{h_{2,2}} \sim N(0_k, \Omega_{h_{2,2}}) \quad \text{as} \quad n \to \infty, \quad \text{where}$$

$$A_{n,j} = n^{1/2}\big(\overline{m}_{n,j}(\theta_{n,h}) - E_{F_{n,h}} m_j(W_i, \theta_{n,h})\big) / \sigma_{F_{n,h},j}(\theta_{n,h}),$$

[40] In Andrews and Guggenberger (2009), a strong mixing condition is imposed in condition (vi) of (S9.2). This condition is used to verify Assumption E0 in that paper and is not needed with RMS critical values.

(viii) $\widehat{\sigma}_{n,j}(\theta_{n,h})/\sigma_{F_{n,h},j}(\theta_{n,h}) \to_p 1$ as $n \to \infty$ for $j = 1, \dots, k$,

(ix) $\widehat{D}_n^{-1/2}(\theta_{n,h})\widehat{\Sigma}_n(\theta_{n,h})\widehat{D}_n^{-1/2}(\theta_{n,h}) \to_p \Omega_{h_{2,2}}$ as $n \to \infty$,

(x) conditions (vii)–(ix) hold for all subsequences $\{w_n\}$ in place of $\{n\}$,

where $\Omega_{h_{2,2}}$ is the $k \times k$ correlation matrix for which $\text{vech}_*(\Omega_{h_{2,2}}) = h_{2,2}$, $\widehat{\sigma}_{n,j}^2(\theta) = [\widehat{\Sigma}_n(\theta)]_{jj}$ for $1 \le j \le k$, and $\widehat{D}_n(\theta) = \text{Diag}\{\widehat{\sigma}_{n,1}^2(\theta), \dots, \widehat{\sigma}_{n,k}^2(\theta)\}$ $(= \text{Diag}(\widehat{\Sigma}_n(\theta)))$.[41,42]

For example, for i.i.d. observations, conditions (i)–(vi) in (S2.2) imply conditions (i)–(vi) in (S9.2). Furthermore, conditions (i)–(vi) in (S2.2) plus the definition of $\widehat{\Sigma}_n(\theta)$ in (S3.2) and the additional condition (vii) in (S2.2) imply conditions (vii)–(x) in (S9.3). For a proof, see Lemma 2 of AG.

For dependent observations or when a preliminary estimator of a parameter τ appears, one needs to specify a particular variance estimator $\widehat{\Sigma}_n(\theta)$ before one can specify primitive "additional conditions" beyond conditions (i)–(vi) in (S9.2) that ensure that Γ is such that any sequences $\{\gamma_{w_n,h} : n \ge 1\}$ in Γ satisfy (S9.3). For brevity, we do not do so here.

We now specify the set Δ, defined in (S4.13), in the parametrization introduced above. Define

(S9.4) $$H = \big\{h \in R^p_{[\pm\infty]} \times R^q_{[\pm\infty]} : \exists \text{ subsequence } \{w_n\} \text{ of } \{n\} \text{ and sequence } \{\gamma_{w_n,h} \in \Gamma : n \ge 1\} \text{ for which } w_n^{1/2}\gamma_{w_n,h,1} \to h_1 \text{ and } \gamma_{w_n,h,2} \to h_2\big\}.$$

Then Δ can be written equivalently as

(S9.5) $$\Delta = \big\{(h_1, \Omega_{h_{2,2}}) \in R^p_{+,\infty} \times \text{cl}(\Psi) : h = (h_1, h_{2,1}, h_{2,2}) \in H \text{ for some } h_{2,1} \in \text{cl}(\Theta), \text{ where } h_{2,2} = \text{vech}_*(\Omega_{h_{2,2}})\big\}.$$

In words, Δ is the set of "slackness" parameters h_1 and correlation matrices Ω that correspond to some limit point h in H.

[41]When a preliminary estimator $\widehat{\tau}_n(\theta)$ appears, $A_{n,j}$ can be written equivalently as $n^{1/2}(n^{-1}\sum_{i=1}^n m_j(W_i, \theta_{n,h}, \widehat{\tau}_n(\theta_{n,h})) - E_{F_{n,h}} m_j(W_i, \theta_{n,h}, \tau_0))/\sigma_{F_{n,h},j}(\theta_{n,h})$, which typically is asymptotically normal with an asymptotic variance matrix $\Omega_{h_{2,2}}$ that reflects the fact that τ_0 has been estimated. When a preliminary estimator $\widehat{\tau}_n(\theta)$ appears, $\widehat{\Sigma}_n(\theta)$ needs to be defined to take account of the fact that τ_0 has been estimated. When no preliminary estimator $\widehat{\tau}_n(\theta)$ appears, $A_{n,j}$ can be written equivalently as $n^{1/2}(\overline{m}_{n,j}(\theta_{n,h}) - E_{F_{n,h}}\overline{m}_{n,j}(\theta_{n,h}))/\sigma_{F_{n,h},j}(\theta_{n,h})$.

[42]Condition (x) of (S9.3) requires that conditions (vii)–(ix) must hold under any sequence of parameters $\{\gamma_{w_n,h} : n \ge 1\}$ that satisfies the conditions preceding (S9.3) with n replaced by w_n.

S9.2. *Proofs*

The proof of Theorem 1 uses the following lemmas. Let

(S9.6) $\quad \mathrm{CP}_n(\gamma) = P_\gamma\big(T_n(\theta) \le c_n(\theta)\big).$

As above, for a sequence of constants $\{\zeta_n : n \ge 1\}$, $\zeta_n \to [\zeta_{1,\infty}, \zeta_{2,\infty}]$ denotes that $\zeta_{1,\infty} \le \liminf_{n\to\infty} \zeta_n \le \limsup_{n\to\infty} \zeta_n \le \zeta_{2,\infty}$.

LEMMA 4: *Suppose Assumptions* S, φ, κ, *and* $\eta 1$ *hold. Let* $\{\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3}) : n \ge 1\}$ *be a sequence of points in* Γ *that satisfies* (i) $n^{1/2}\gamma_{n,h,1} \to h_1$ *for some* $h_1 \in R^p_{+,\infty}$ *and* (ii) $\gamma_{n,h,2} \to h_2$ *for some* $h_2 = (h_{2,1}, h_{2,2}) \in R^q_{[\pm\infty]}$. *Let* $h = (h_1, h_2)$ *and let* $\Omega_{h_{2,2}}$ *be the correlation matrix that corresponds to* $h_{2,2}$.

(a) *Then*, $\mathrm{CP}_n(\gamma_{n,h}) \to [\mathrm{CP}(h_1, \Omega_{h_{2,2}}, \eta(\Omega_{h_{2,2}})-), \mathrm{CP}(h_1, \Omega_{h_{2,2}}, \eta(\Omega_{h_{2,2}}))]$.

(b) *Also, for any subsequence* $\{w_n : n \ge 1\}$ *of* $\{n\}$, *the result of part* (a) *holds with* w_n *in place of* n *provided conditions* (i) *and* (ii) *above hold with* w_n *in place of* n.

LEMMA 5: *Suppose Assumptions* S(b)–(e) *hold. Then* $q_S(\beta, \Omega)$ *is continuous on* $(R^p_{[+\infty]} \times R^v) \times \Psi$.

PROOF OF THEOREM 1: First, we prove part (a). Let $\{\gamma^*_n = (\gamma^*_{n,1}, \gamma^*_{n,2}, \gamma^*_{n,3}) \in \Gamma : n \ge 1\}$ be a sequence such that $\liminf_{n\to\infty} \mathrm{CP}_n(\gamma^*_n) = \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} \mathrm{CP}_n(\gamma)$ (= AsyCS). Such a sequence always exists. Let $\{u_n : n \ge 1\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} \mathrm{CP}_{u_n}(\gamma^*_{u_n})$ exists and equals $\liminf_{n\to\infty} \mathrm{CP}_n(\gamma^*_n) = \mathrm{AsyCS}$. Such a subsequence always exists.

Let $\gamma^*_{n,1,j}$ denote the jth component of $\gamma^*_{n,1}$ for $j = 1, \ldots, p$. Either (i) $\limsup_{n\to\infty} u_n^{1/2}\gamma^*_{u_n,1,j} < \infty$ or (ii) $\limsup_{n\to\infty} u_n^{1/2}\gamma^*_{u_n,1,j} = \infty$. If (i) holds, then for some subsequence $\{w_n\}$ of $\{u_n\}$,

(S9.7) $\quad w_n^{1/2}\gamma^*_{w_n,1,j} \to h^*_{1,j} \quad \text{for some } h^*_{1,j} \in R_+.$

If (ii) holds, then for some subsequence $\{w_n\}$ of $\{u_n\}$,

(S9.8) $\quad w_n^{1/2}\gamma^*_{w_n,1,j} \to h^*_{1,j}, \quad \text{where} \quad h^*_{1,j} = \infty.$

In addition, for some subsequence $\{w_n\}$ of $\{u_n\}$,

(S9.9) $\quad \gamma^*_{w_n,2} \to h^*_2 \quad \text{for some } h^*_2 \in \mathrm{cl}(\Gamma_2).$

By taking successive subsequences over the p components of $\gamma^*_{u_n,1}$ and $\gamma^*_{u_n,2}$, we find that there exists a subsequence $\{w_n\}$ of $\{u_n\}$ such that for each $j = 1, \ldots, p$, either (S9.7) or (S9.8) applies and (S9.9) holds. In consequence, (i) $w_n^{1/2}\gamma_{w_n,h,1} \to h^*_1$ for some $h^*_1 \in R^p_{+,\infty}$, (ii) $\gamma_{w_n,h,2} \to h^*_2$ for some $h^*_2 \in R^q_{[\pm\infty]}$,

(iii) $h^* = (h_1^*, h_2^*) \in H$ (for H defined in (S9.4)), and (iv) $\lim_{n\to\infty} \mathrm{CP}_{w_n}(\gamma_{w_n}^*) =$ AsyCS. Hence, by Lemma 4(b),

$$\text{(S9.10)}\quad \mathrm{AsyCS} = \lim_{n\to\infty} \mathrm{CP}_{w_n}\big(\gamma_{w_n}^*\big) \geq \mathrm{CP}\big(h_1^*, \Omega_{h_{2,2}^*}, \eta(\Omega_{h_{2,2}^*})-\big) \geq \inf_{(h_1,\Omega)\in\Delta} \mathrm{CP}\big(h_1, \Omega, \eta(\Omega)-\big),$$

where the second inequality holds because $(h_1^*, \Omega_{h_{2,2}^*}) \in \Delta$ by the definition of Δ in (S9.5).

Next, by the definition of Δ in (S9.5), for each $(h_1, \Omega_{h_{2,2}}) \in \Delta$, there exists a subsequence $\{t_n : n \geq 1\}$ of $\{n\}$ and a sequence of points $\{\gamma_{t_n,h} = (\gamma_{t_n,h,1}, \gamma_{t_n,h,2}, \gamma_{t_n,h,3}) \in \Gamma : n \geq 1\}$ such that conditions (i) and (ii) of Lemma 4 hold with t_n in place of n. Hence,

$$\text{(S9.11)}\quad \mathrm{AsyCS} = \liminf_{n\to\infty} \inf_{(\theta,F)\in\mathcal{F}} P_F\big(T_n(\theta) \leq c_n(\theta)\big) \leq \liminf_{n\to\infty} \mathrm{CP}_{t_n}(\gamma_{t_n,h}) \leq \mathrm{CP}\big(h_1, \Omega_{h_{2,2}}, \eta(\Omega_{h_{2,2}})\big),$$

where the second inequality holds by Lemma 4(b). Since (S9.11) holds for all $(h_1, \Omega_{h_{2,2}}) \in \Delta$, we have

$$\text{(S9.12)}\quad \mathrm{AsyCS} \leq \inf_{(h_1,\Omega)\in\Delta} \mathrm{CP}\big(h_1, \Omega, \eta(\Omega)\big).$$

Combining (S9.10) and (S9.12) establishes part (a) of the theorem.

Part (b) of the theorem follows from part (a) and Assumption $\eta 2$. Part (c) of the theorem follows from part (a) and Assumption $\eta 3$. *Q.E.D.*

PROOF OF LEMMA 4: For notational simplicity, let Ω_0 denote $\Omega_{h_{2,2}}$. To establish part (a), we show below that

$$\text{(S9.13)}\quad \begin{pmatrix} T_n(\theta_{n,h}) \\ c_n(\theta_{n,h}) \end{pmatrix} \to_d \begin{pmatrix} S\big(Z + (h_1, 0_v), \Omega_0\big) \\ q_S\big(\varphi\big(\kappa^{-1}(\Omega_0)\big[Z + (h_1, 0_v)\big], \Omega_0\big), \Omega_0\big) + \eta(\Omega_0) \end{pmatrix} \quad \text{as} \quad n \to \infty$$

under $\{\gamma_{n,h} : n \geq 1\}$, where $Z \sim N(0_k, \Omega_0)$. Hence, by the definition of convergence in distribution, for every continuity point x of the asymptotic distribution of $T_n(\theta_{n,h}) - c_n(\theta_{n,h})$, we have

$$\text{(S9.14)}\quad P_{\gamma_{n,h}}\big(T_n(\theta_{n,h}) \leq c_n(\theta_{n,h}) + x\big) \to P\Big[S\big(Z + (h_1, 0_v), \Omega_0\big) \leq q_S\big(\varphi\big(\kappa^{-1}(\Omega_0)\big[Z + (h_1, 0_v)\big], \Omega_0\big), \Omega_0\big) + \eta(\Omega_0) + x\Big] = \mathrm{CP}\big(h_1, \Omega_0, \eta(\Omega_0) + x\big).$$

There exist continuity points $x > 0$ and $x < 0$ arbitrarily close to zero. Hence, we have

$$\text{(S9.15)}\quad \begin{aligned}&\limsup_{n\to\infty} P_{\gamma_{n,h}}\big(T_n(\theta_{n,h}) \le c_n(\theta_{n,h})\big)\\ &\quad\le \lim_{x\downarrow 0}\limsup_{n\to\infty} P_{\gamma_{n,h}}\big(T_n(\theta_{n,h}) \le c_n(\theta_{n,h}) + x\big)\\ &\quad= \lim_{x\downarrow 0} \mathrm{CP}\big(h_1, \Omega_0, \eta(\Omega_0) + x\big) = \mathrm{CP}\big(h_1, \Omega_0, \eta(\Omega_0)\big),\end{aligned}$$

where the first equality holds by (S9.14) and the second equality holds because $\mathrm{CP}(h_1, \Omega_0, \eta(\Omega_0) + x)$ is a d.f. and hence is right-continuous. Analogously,

$$\text{(S9.16)}\quad \begin{aligned}\liminf_{n\to\infty} P_{\gamma_{n,h}}\big(T_n(\theta_{n,h}) \le c_n(\theta_{n,h})\big) &\ge \lim_{x\downarrow 0} \mathrm{CP}\big(h_1, \Omega_0, \eta(\Omega_0) - x\big)\\ &= \mathrm{CP}\big(h_1, \Omega_0, \eta(\Omega_0)-\big),\end{aligned}$$

where the equality holds by definition. Equations (S9.15) and (S9.16) combine to establish part (a).

Next, we prove (S9.13). Using Assumption S(a), we have

$$\text{(S9.17)}\quad T_n(\theta) = S\big(\widehat{D}_n^{-1/2}(\theta)n^{1/2}\overline{m}_n(\theta), \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta)\widehat{D}_n^{-1/2}(\theta)\big).$$

For i.i.d. or dependent observations with or without preliminary estimators of identified parameters, (S9.3) holds (using the fact that $\gamma \in \Gamma$ if and only if $(\theta, F) \in \mathcal{F}$ and using Lemma 2 of AG to show that (S9.3) holds for i.i.d. observations). By (S9.3), the jth element of $\widehat{D}_n^{-1/2}(\theta_{n,h})n^{1/2}\overline{m}_n(\theta_{n,h})$ equals $(1 + o_p(1))(A_{n,j} + n^{1/2}\gamma_{n,h,1,j})$, where $\gamma_{n,h,1} = (\gamma_{n,h,1,1}, \ldots, \gamma_{n,h,1,p})'$ and, by definition, $\gamma_{n,h,1,j} = 0$ for $j = p + 1, \ldots, k$. If $h_{1,j} = \infty$ and $j \le p$, where $h_1 = (h_{1,1}, \ldots, h_{1,p})'$, then $A_{n,j} + n^{1/2}\gamma_{n,h,1,j} \to_p \infty$ under $\{\gamma_{n,h} : n \ge 1\}$ by condition (vii) of (S9.3) and the definition of $\{\gamma_{n,h} : n \ge 1\}$. Hence, if any element of h_1 equals ∞, $\widehat{D}_n^{-1/2}(\theta_{n,h})n^{1/2}\overline{m}_n(\theta_{n,h})$ does not converge in distribution (to a proper finite random vector) and the continuous mapping theorem cannot be applied to obtain the asymptotic distribution of the right-hand side of (S9.17) or of the RMS critical value, which is defined by

$$\text{(S9.18)}\quad c_n(\theta) = q_S\big(\varphi\big(\xi_n(\theta), \widehat{\Omega}_n(\theta)\big), \widehat{\Omega}_n(\theta)\big) + \eta\big(\widehat{\Omega}_n(\theta)\big).$$

To circumvent these problems, we consider k-vector-valued functions of $\widehat{D}_n^{-1/2}(\theta_{n,h})n^{1/2}\overline{m}_n(\theta_{n,h})$ and $\xi_n(\theta_{n,h})$ that converge in distribution whether or not some elements of h_1 equal ∞. Then we write the right-hand sides of (S9.17) and (S9.18) as continuous functions of these k-vectors and apply the continuous mapping theorem. Let $G(\cdot)$ be a strictly increasing continuous d.f. on R, such as the standard normal d.f.

For $j \leq k$, we have

$$\text{(S9.19)} \quad \begin{aligned} G_{\kappa,n,j} &= G\big(\xi_{n,j}(\theta_{n,h})\big) \\ &= G\big(\kappa^{-1}\big(\widehat{\Omega}_n(\theta_{n,h})\big)\widehat{\sigma}_{n,j}^{-1}(\theta_{n,h})n^{1/2}\overline{m}_{n,j}(\theta_{n,h})\big) \\ &= G\big(\kappa^{-1}\big(\widehat{\Omega}_n(\theta_{n,h})\big)\widehat{\sigma}_{n,j}^{-1}(\theta_{n,h})\sigma_{F_{n,h},j}(\theta_{n,h})\big[A_{n,j} + n^{1/2}\gamma_{n,h,1,j}\big]\big), \end{aligned}$$

where $A_{n,j}$ is defined in (S9.3) and, by definition, $\gamma_{n,h,1,j} = 0$ for $j = p + 1, \dots, k$.

Let $Z = (Z_1, \dots, Z_k)' \sim N(0_k, \Omega_0)$. Define $h_{1,j} = 0$ for $j = p + 1, \dots, k$. If $j \leq p$ and $h_{1,j} < \infty$ or if $j = p + 1, \dots, k$, then

$$\text{(S9.20)} \quad G_{\kappa,n,j} \to_d G\big(\kappa^{-1}(\Omega_0)[Z_j + h_{1,j}]\big)$$

using (S9.19), conditions (vii) and (viii) of (S9.3) (which yield $A_{n,j} + n^{1/2}\gamma_{n,h,1,j} \to_d Z_j + h_{1,j}$), Assumption κ and condition (ix) of (S9.3) (which yield $\kappa^{-1}(\widehat{\Omega}_n(\theta_{n,h})) \to_p \kappa^{-1}(\Omega_0)$), and the continuous mapping theorem.

If $j \leq p$ and $h_{1,j} = \infty$, then

$$\text{(S9.21)} \quad G_{\kappa,n,j} \to_p 1$$

using (S9.19), $A_{n,j} = O_p(1)$, $\kappa^{-1}(\widehat{\Omega}_n(\theta_{n,h})) \to_p \kappa^{-1}(\Omega_0) > 0$, and $G(x) \to 1$ as $x \to \infty$. The results in (S9.20) and (S9.21) hold jointly and combine to give

$$\text{(S9.22)} \quad \begin{aligned} &G_{\kappa,n} = (G_{\kappa,n,1}, \dots, G_{\kappa,n,k})' \to_d G_{\kappa,\infty}, \quad \text{where} \\ &G_{\kappa,\infty} = \big(G\big(\kappa^{-1}(\Omega_0)[Z_1 + h_{1,1}]\big), \dots, G\big(\kappa^{-1}(\Omega_0)[Z_k + h_{1,k}]\big)\big)' \end{aligned}$$

and $G(Z_{h_{2,2},j} + h_{1,j})$ denotes $G(\infty) = 1$ when $h_{1,j} = \infty$.

Let G^{-1} denote the inverse of G. For $x = (x_1, \dots, x_k)' \in R^p_{[+\infty]} \times R^v$, let $G_{(k)}(x) = (G(x_1), \dots, G(x_k))' \in (0, 1]^p \times (0, 1)^v$. For $z = (z_1, \dots, z_k)' \in (0, 1]^p \times (0, 1)^v$, let $G^{-1}_{(k)}(z) = (G^{-1}(z_1), \dots, G^{-1}(z_k))' \in R^p_{[+\infty]} \times R^v$. Define $\widetilde{q}_S(z, \Omega)$ as

$$\text{(S9.23)} \quad \widetilde{q}_{S,\varphi}(z, \Omega) = q_S\big(\varphi\big(G^{-1}_{(k)}(z), \Omega\big), \Omega\big)$$

for $z \in (0, 1]^p \times (0, 1)^v$ and $\Omega \in \Psi$.

Assumption φ and Lemma 5 imply that $\widetilde{q}_{S,\varphi}(z, \Omega)$ is continuous at (z, Ω) for all $z \in \mathcal{Z}((h_1, 0_v), \Omega_0)$ and $\Omega = \Omega_0$, where

$$\text{(S9.24)} \quad \begin{aligned} &\mathcal{Z}\big((h_1, 0_v), \Omega_0\big) = \big\{z \in (0, 1]^p \times (0, 1)^v : G^{-1}_{(k)}(z) \in \Xi\big((h_1, 0_v), \Omega\big)\big\}, \\ &P\big(G_{\kappa,\infty} \in \mathcal{Z}\big((h_1, 0_v), \Omega_0\big)\big) \\ &\quad = P\big(\kappa^{-1}(\Omega_0)\big[Z + (h_1, 0_v)\big] \in \Xi\big((h_1, 0_v), \Omega_0\big)\big) \\ &\quad = 1, \end{aligned}$$

and $\Xi(\beta,\Omega)$ is defined in Assumption φ.

We now have

$$
\begin{aligned}
\text{(S9.25)}\quad c_n(\theta_{n,h}) &= q_S\big(\varphi\big(\xi_n(\theta_{n,h}),\widehat{\Omega}_n(\theta_{n,h})\big),\widehat{\Omega}_n(\theta_{n,h})\big)+\eta\big(\widehat{\Omega}_n(\theta_{n,h})\big)\\
&= q_S\big(\varphi\big(G^{-1}_{(k)}(G_{\kappa,n}),\widehat{\Omega}_n(\theta_{n,h})\big),\widehat{\Omega}_n(\theta_{n,h})\big)+\eta\big(\widehat{\Omega}_n(\theta_{n,h})\big)\\
&= \widetilde{q}_{S,\varphi}\big(G_{\kappa,n},\widehat{\Omega}_n(\theta_{n,h})\big)+\eta\big(\widehat{\Omega}_n(\theta_{n,h})\big)\\
&\to_d \widetilde{q}_{S,\varphi}(G_{\kappa,\infty},\Omega_0)+\eta(\Omega_0)\\
&= q_S\big(\varphi\big(G^{-1}_{(k)}(G_{\kappa,\infty}),\Omega_0\big),\Omega_0\big)+\eta(\Omega_0)\\
&= q_S\big(\varphi\big(\kappa^{-1}(\Omega_0)\big[Z+(h_1,0_v)\big],\Omega_0\big),\Omega_0\big)+\eta(\Omega_0),
\end{aligned}
$$

where the first equality holds by the definition of $c_n(\theta_{n,h})$, the second equality holds by the definitions of $G_{\kappa,n}$ and $G^{-1}_{(k)}(\cdot)$, the third and fourth equalities hold by the definition of $\widetilde{q}_{S,\varphi}(\cdot,\cdot)$, the convergence holds by (S9.22), condition (ix) of (S9.3), Assumption η1, and the continuous mapping theorem using (S9.24), and the last equality holds by the definitions of $G_{\kappa,\infty}$ and $G^{-1}_{(k)}(\cdot)$ and the definition that if $h_{1,j}=\infty$, then the corresponding element of $Z+(h_1,0_v)$ equals ∞.

We now use an analogous argument to that in (S9.19)–(S9.25) to show that

$$
\text{(S9.26)}\quad T_n(\theta_{n,h})\to_d S\big(Z+(h_1,0_v),\Omega_0\big).
$$

The argument only differs from that given above in that (i) $\kappa(\cdot)$ is replaced by 1 throughout, (ii) the function $q_S(\varphi(m,\Omega),\Omega)$ is replaced by $S(m,\Omega)$, (iii) the function $\widetilde{q}_{S,\varphi}(z,\Omega)=q_S(\varphi(G^{-1}_{(k)}(z),\Omega),\Omega)$ is replaced by $\widetilde{S}(z,\Omega)=S(G^{-1}_{(k)}(z),\Omega)$, and (iv) the continuity argument in the paragraph containing (S9.24) is replaced by the assertion that $\widetilde{S}(z,\Omega)$ is continuous at all $(z,\Omega)\in((0,1]^p\times(0,1)^v)\times\Psi$ by Assumption S(c).

The convergence in (S9.25) and (S9.26) is joint because the two results can be obtained by a single application of the continuous mapping theorem. Hence, the verification of (S9.13) is complete and part (a) is proved.

Next, we prove part (b). By the same argument as above but using condition (x) of (S9.3) in place of conditions (vii)–(ix), the results of (S9.25) and (S9.26) hold with $\{w_n\}$ in place of $\{n\}$ for any subsequence $\{w_n\}$. Hence, (S9.13) and (S9.14) hold with the same changes, which implies that part (b) holds. *Q.E.D.*

PROOF OF LEMMA 5: Given $(\beta_0,\Omega_0)\in(R^p_{[+\infty]}\times R^v)\times\Psi$, we consider three cases: (i) $q_S(\beta_0,\Omega_0)>0$, (ii) $q_S(\beta_0,\Omega_0)=0$ and either $v>0$ or both $v=0$ and $\beta_0\neq\infty^p$, and (iii) $q_S(\beta_0,\Omega_0)=0$, $v=0$, and $\beta_0=\infty^p$.

In case (i), given $\varepsilon>0$, we want to show that if (β,Ω) is sufficiently close to (β_0,Ω_0), then $|q_S(\beta,\Omega)-q_S(\beta_0,\Omega_0)|<\varepsilon$. Let $Z^*\sim N(0_k,I_k)$. By Assumption S(e), the d.f. of $S(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0)$ is strictly increasing at $x = q_S(\beta_0, \Omega_0) > 0$. Hence, for some $\varepsilon_U > 0$,

$$P\big(S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big) \le q_S(\beta_0, \Omega_0) + \varepsilon\big) = 1 - \alpha + \varepsilon_U. \tag{S9.27}$$

The d.f. of $S(\Omega^{1/2}Z^* + \beta, \Omega)$ at $x > 0$ is continuous in (β, Ω) at (β_0, Ω_0) by the bounded convergence theorem because

(S9.28)
(a) $S\big(\Omega^{1/2}Z^* + \beta, \Omega\big) \to S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big)$ a.s.,
(b) $1\big(S\big(\Omega^{1/2}Z^* + \beta, \Omega\big) \le x\big) \to 1\big(S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big) \le x\big)$ a.s. except if $S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big) = x$,
(c) $P\big(S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big) = x\big) = 0$,
(d) the indicator function is bounded,

where (a) holds by Assumption S(c), (b) holds by (a), and (c) holds because the d.f. of $S(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0)$ is continuous at all $x > 0$ by Assumption S(e).

In consequence, for all (β, Ω) sufficiently close to (β_0, Ω_0), we have

$$\begin{aligned} &\big|P\big(S\big(\Omega^{1/2}Z^* + \beta, \Omega\big) \le q_S(\beta_0, \Omega_0) + \varepsilon\big) \\ &\quad - P\big(S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big) \le q_S(\beta_0, \Omega_0) + \varepsilon\big)\big| < \varepsilon_U/2. \end{aligned} \tag{S9.29}$$

Equations (S9.27) and (S9.29) imply that

$$P\big(S\big(\Omega^{1/2}Z^* + \beta, \Omega\big) \le q_S(\beta_0, \Omega_0) + \varepsilon\big) \ge 1 - \alpha + \varepsilon_U/2. \tag{S9.30}$$

The definition of a quantile and (S9.30) imply that

$$q_S(\beta, \Omega) \le q_S(\beta_0, \Omega_0) + \varepsilon. \tag{S9.31}$$

By a completely analogous argument, for (β, Ω) sufficiently close to (β_0, Ω_0), $q_S(\beta, \Omega) \ge q_S(\beta_0, \Omega_0) - \varepsilon$. Hence, $|q_S(\beta, \Omega) - q_S(\beta_0, \Omega_0)| < \varepsilon$ and the proof is complete for case (i).

In case (ii), $P(S(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0) \le 0) \ge 1 - \alpha$ because $q_S(\beta_0, \Omega_0) = 0$. Also, in case (ii), $S(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0)$ has a strictly increasing d.f. for $x > 0$ by Assumption S(e) (because $v = 0$ and $\beta_0 = \infty^p$ does not hold in case (ii)). These results imply that given $\varepsilon > 0$, there exists $\varepsilon_1 > 0$ such that

$$P\big(S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big) \le \varepsilon\big) = 1 - \alpha + \varepsilon_1. \tag{S9.32}$$

Because the d.f. of $S(\Omega^{1/2}Z^* + \beta, \Omega)$ at $\varepsilon > 0$ is continuous in (β, Ω) by (S9.28), for all (β, Ω) sufficiently close to (β_0, Ω_0), we have

$$\big|P\big(S\big(\Omega^{1/2}Z^* + \beta, \Omega\big) \le \varepsilon\big) - P\big(S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big) \le \varepsilon\big)\big| < \varepsilon_1/2. \tag{S9.33}$$

Equations (S9.32) and (S9.33) imply

$$\text{(S9.34)} \quad P\big(S\big(\Omega^{1/2}Z^* + \beta, \Omega\big) \leq \varepsilon\big) \geq 1 - \alpha.$$

This and the definition of a quantile imply that $q_S(\beta, \Omega) \leq \varepsilon$. Since $q_S(\beta, \Omega) \geq 0$ for all (β, Ω) by Assumption S(b), the proof for case (ii) is complete.

In case (iii), $S(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0) = S(\infty^p, \Omega_0) = 0$ a.s. by Assumption S(b) and (d). This and the continuity in (β, Ω) at (β_0, Ω_0) of the d.f. of $S(\Omega^{1/2}Z^* + \beta, \Omega)$ at $x > 0$, which holds by (S9.28), give, for all $x > 0$,

$$\text{(S9.35)} \quad \lim_{(\beta,\Omega)\to(\beta_0,\Omega_0)} P\big(S\big(\Omega^{1/2}Z^* + \beta, \Omega\big) \leq x\big) = P\big(S\big(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0\big) \leq x\big) = 1.$$

Equation (S9.35) implies that given any $x > 0$ for all (β, Ω) sufficiently close to (β_0, Ω_0), the d.f. of $S(\Omega^{1/2}Z^* + \beta, \Omega)$ at $x > 0$ is greater than $1 - \alpha$ and hence $q_S(\beta, \Omega) \leq x$. Since $q_S(\beta, \Omega) \geq 0$ for all (β, Ω) and $x > 0$ is arbitrary, the proof for case (iii) is complete. *Q.E.D.*

PROOF OF LEMMA 2: Assumption LA3(a) holds by the Liapounov triangular array CLT for rowwise i.i.d. random variables with mean zero and variance 1 using Assumptions LA1(a) and (c), and LA3* and the Cramér–Wold device. Assumption LA3(b) and (c) hold by standard arguments using a weak law of large numbers for rowwise i.i.d. random variables with variance 1 using Assumptions LA1(a) and (c), and LA3*. Note that Assumption LA3 does not follow from (S9.3) because in Assumption LA3 the functions are evaluated at θ_0, which is not the true value (unless $\lambda = 0$). *Q.E.D.*

PROOF OF THEOREM 3: The proof follows a similar line of argument to that of Lemma 4(a). We start by showing that under the given assumptions, (S9.13) holds with $(h_1, 0_v)$ replaced by $(h_1, 0_v) + \Pi_0\lambda$. By element-by-element mean-value expansions about $\theta = \theta_n$ and Assumptions LA1 and LA2, we obtain

$$\text{(S9.36)} \quad D^{-1/2}(\theta_0, F_n)E_{F_n}m(W_i, \theta_0) = D^{-1/2}(\theta_n, F_n)E_{F_n}m(W_i, \theta_n) + \Pi\big(\theta_n^*, F_n\big)(\theta_0 - \theta_n),$$
$$n^{1/2}D^{-1/2}(\theta_0, F_n)E_{F_n}m(W_i, \theta_0) \to (h_1, 0_v) + \Pi_0\lambda,$$

where $D(\theta, F) = \mathrm{Diag}\{\sigma^2_{F,1}(\theta), \ldots, \sigma^2_{F,k}(\theta)\}$, θ_n^* may differ across rows of $\Pi(\theta_n^*, F_n)$, θ_n^* lies between θ_0 and θ_n, $\theta_n^* \to \theta_0$, and $\Pi(\theta_n^*, F_n) \to \Pi_0$.

For the same reason as described above following (S9.17), to obtain the asymptotic distribution of $T_n(\theta_0)$ we use the same type of argument as in the proof of Lemma 4(a). Let $G(\cdot)$ be a strictly increasing continuous d.f. on R, such as the standard normal d.f. Using (S9.36), Assumption LA3, and

$\kappa^{-1}(\widehat{\Omega}_n(\theta_0)) \to_p \kappa^{-1}(\Omega(\theta_0))$ (which holds by Assumptions κ and LA3), for $j = 1, \ldots, k$, we have

$$
\begin{aligned}
\text{(S9.37)}\quad G^0_{\kappa,n,j} &= G\big(\kappa^{-1}\big(\widehat{\Omega}_n(\theta_0)\big)\widehat{\sigma}^{-1}_{n,j}(\theta_0)n^{1/2}\overline{m}_{n,j}(\theta_0)\big) \\
&= G\big(\kappa^{-1}\big(\widehat{\Omega}_n(\theta_0)\big)\widehat{\sigma}^{-1}_{n,j}(\theta_0)\sigma_{F_n,j}(\theta_0) \\
&\quad \times \big[A^0_{n,j} + n^{1/2}\sigma^{-1}_{F_n,j}(\theta_0)E_{F_n}m_j(W_i,\theta_0)\big]\big), \\
G^0_{\kappa,n,j} &\to_p 1 \quad \text{if} \quad j \le p \quad \text{and} \quad h_{1,j} = \infty, \\
G^0_{\kappa,n,j} &\to_d G\big(\kappa^{-1}\big(\Omega(\theta_0)\big)\big[Z_j + h_{1,j} + \Pi'_{0,j}\lambda\big]\big) \\
&\quad \text{if} \quad j \le p \quad \text{and} \quad h_{1,j} < \infty, \\
G^0_{\kappa,n,j} &\to_d G\big(\kappa^{-1}\big(\Omega(\theta_0)\big)\big[Z_j + \Pi'_{0,j}\lambda\big]\big) \quad \text{if} \quad j = p+1, \ldots, k, \\
G^0_{\kappa,n} &= \big(G^0_{\kappa,n,1}, \ldots, G^0_{\kappa,n,k}\big) \to_d G^0_{\kappa,\infty} \\
&= \big(G\big(\kappa^{-1}\big(\Omega(\theta_0)\big)\big[Z_1 + h_{1,1} + \Pi'_{0,1}\lambda\big]\big), \ldots, \\
&\quad G\big(\kappa^{-1}\big(\Omega(\theta_0)\big)\big[Z_k + \Pi'_{0,k}\lambda\big]\big)\big)',
\end{aligned}
$$

where $Z = (Z_1, \ldots, Z_k)'$ and $Z_j + h_{1,j} + \Pi'_{0,j}\lambda = \infty$ by definition if $h_{1,j} = \infty$. Now, the same argument as in (S9.23)–(S9.25) of the proof of Lemma 4(a) gives

$$
\text{(S9.38)}\quad c_n(\theta_0) \to_d q_S\big(\varphi\big(\kappa^{-1}(\Omega_0)\big[Z + (h_1, 0_v) + \Pi_0\lambda\big], \Omega_0\big), \Omega_0\big) + \eta(\Omega_0).
$$

The only difference in the proof is that $\mathcal{Z}((h_1, 0_v), \Omega_0)$ and $\Xi((h_1, 0_v), \Omega)$ are replaced by $\mathcal{Z}((h_1, 0_v) + \Pi_0\lambda, \Omega_0)$ and $\Xi((h_1, 0_v) + \Pi_0\lambda, \Omega)$, respectively.

Next, by the same argument as in (S9.26) in the proof of Lemma 4(a), we obtain

$$
\text{(S9.39)}\quad T_n(\theta_0) \to_d S\big(\big[Z + (h_1, 0_v) + \Pi_0\lambda\big], \Omega_0\big).
$$

Furthermore, the convergence in (S9.38) and (S9.39) is joint, which establishes that (S9.13) holds with $(h_1, 0)$ replaced by $(h_1, 0_v) + \Pi_0\lambda$. Finally, given the latter result, the result of the theorem holds by the same argument as in (S9.14)–(S9.16) in the proof of Lemma 4(a) with $(h_1, 0_v)$ replaced by $(h_1, 0_v) + \Pi_0\lambda$ and $\mathrm{CP}(h_1, \Omega_0, \eta(\Omega_0))$ replaced by $\mathrm{AsyPow}(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0))$. *Q.E.D.*

REFERENCES

ANDREWS, D. W. K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–564. [10]

ANDREWS, D. W. K., AND P. GUGGENBERGER (2009): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709. [6,8,56]

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157. [6]

ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2008): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression With Weak Instruments," *Journal of Econometrics*, 146, 241–254. [43]

BUGNI, F. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753. [10]

CANAY, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408–425. [6,10,25,30]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [10]

FAN, Y., AND S. PARK (2007): "Confidence Sets for Some Partially Identified Parameters," Unpublished Manuscript, Department of Economics, Vanderbilt University. [10]

HANSEN, P. (2005): "A Test for Superior Predictive Ability," *Journal of Business & Economic Statistics*, 23, 365–380. [10]

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [24,25]

KUDO, A. (1963): "A Multivariate Analogue of the One-Sided Test," *Biometrika*, 50, 403–418. [6]

MÜLLER, U., AND M. WATSON (2008): "Low-Frequency Robust Cointegration Testing," *Econometrica*, 76, 979–1016. [43]

PRATT, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56, 541–567. [16]

ROSEN, A. M. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117. [6,14,15]

SILVAPULLE, M. J., AND P. K. SEN (2005): *Constrained Statistical Inference*. New York: Wiley. [6,14]

Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse Avenue, New Haven, CT 06520-8281, U.S.A.; donald.andrews@yale.edu

and

Dept. of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142, U.S.A.; pjia@mit.edu.

Manuscript received October, 2008; final revision received March, 2012.