Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure

By

Donald W.K. Andrews and Panle Jia

September 2008
Revised August 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1676R

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

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

Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure

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

Panle Jia
Department of Economics
MIT

June 2008
Revised: May 2011

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant SES-0751517. The authors thank Steve Berry for numerous discussions and comments. The authors also thank the co-editor, three referees, and Michael Wolf for very helpful comments.

Abstract

This paper is concerned with tests and confidence intervals for parameters that are not necessarily identified and are defined by moment inequalities. In the literature, different test statistics, critical value methods, and implementation methods (i.e., the asymptotic distribution versus the bootstrap) have been proposed. In this paper, we compare these methods. We provide a recommended test statistic, moment selection critical value method, and implementation method. We provide data-dependent procedures for choosing the key moment selection tuning parameter κ and a size-correction factor η.

Keywords: Asymptotic size, asymptotic power, bootstrap, confidence set, generalized moment selection, moment inequalities, partial identification, refined moment selection, test, unidentified parameter.

JEL Classification Numbers: C12, C15.

1 Introduction

This paper considers inference in moment inequality models with parameters that need not be identified. We focus on confidence sets for the true parameter, as opposed to the identified set. We construct confidence sets (CS's) by inverting Anderson-Rubin-type test statistics. We consider a class of such statistics and a class of generalized moment selection (GMS) critical values. This approach follows Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007) (CHT), Andrews and Guggenberger (2009) (AG), Andrews and Soares (2010) (AS), and other papers.

GMS and subsampling tests and CS's are the only methods in the literature that apply to arbitrary moment functions and have been shown to have correct asymptotic size in a uniform sense, see AG, AS, and Romano and Shaikh (2008). AS and Bugni (2010) show that GMS tests dominate subsampling tests in terms of asymptotic size and power properties. In addition, in our experience based on simulation results, subsampling tests often are substantially under-sized in finite samples in moment inequality testing problems. Hence, we focus on GMS critical values.

GMS tests and CS's depend on a test statistic function S, a critical value function φ, and a tuning parameter κ. In this paper we determine a combination that performs well in terms of size and power and can be recommended for general use. To do so, we consider asymptotics in which κ equals a *finite* constant plus $o_p(1)$, rather than asymptotics in which $\kappa \to \infty$ as $n \to \infty$, as has been considered elsewhere in the literature.[1]

We find that an *adjusted* Gaussian quasi-likelihood ratio (AQLR) test statistic combined with a "t-test moment selection" critical value performs very well in terms of asymptotic average power compared to other choices considered in the literature.[2] We develop data-dependent methods of selecting κ and a size-correction factor η and show that they yield very good asymptotic and finite-sample size and power. We provide a table that makes them easy to implement in practice for up to ten moment inequalities.

We show that with i.i.d. observations bootstrap critical values out-perform those based on the asymptotic-distribution in terms of finite-sample size. We also show that the bootstrap version of the AQLR test performs similarly in terms of null rejection probabilities and power to an analogous test based on the empirical likelihood ratio (ELR) statistic. The AQLR-based test is noticeably faster to compute than the ELR-

[1] The theoretical arguments mentioned in the preceding paragraph rely on $\kappa \to \infty$ asymptotics.

[2] The "adjustment" in the AQLR test statistic is designed to handle singular asymptotic correlation matrices of the sample moment functions.

based test and avoids computational convergence problems that can arise with the ELR statistic when the correlation matrix of the moment conditions is singular.

The asymptotic results of the paper apply to i.i.d. and time series data and to moment functions that are based on preliminary estimators of point-identified parameters.

In short, the contribution of this paper relative to the literature is to compare moment inequality tests, determine a recommended test, and provide data-dependent tuning parameters.

The remainder of the paper is organized as follows. Section 2 introduces the model and describes the recommended confidence set and test. Section 3 defines the different test statistics and critical values that are compared in the paper. Section 4 provides the numerical comparisons of the tests based on asymptotic average power. Section 5 describes how the recommended data-dependent tuning parameter $\widehat{\kappa}$ and size-correction factor $\widehat{\eta}$ are determined and provides numerical results assessing their performance. Section 6 gives finite-sample results.

Andrews and Jia (2008) (AJ2) provides Supplemental Material that includes: (i) the asymptotic results that are utilized in this paper, (ii) details concerning the numerical results given here, and (iii) additional numerical results.

Let $R_+ = \{x \in R : x \geq 0\}$, $R_{++} = \{x \in R : x > 0\}$, $R_{+,\infty} = R_+ \cup \{+\infty\}$, $K^p = K \times ... \times K$ (with p copies) for any set K, and $0_p = (0, ..., 0)' \in R^p$.

2 Model and Recommended Confidence Set

The moment inequality model is specified as follows. The true value θ_0 ($\in \Theta \subset R^d$) is assumed to satisfy the moment conditions:

$$E_{F_0} m_j(W_i, \theta_0) \geq 0 \text{ for } j = 1, ..., p, \tag{2.1}$$

where $\{m_j(\cdot, \theta) : j = 1, ..., p\}$ are known real-valued moment functions and $\{W_i : i \geq 1\}$ are i.i.d. or stationary random vectors with joint distribution F_0. The observed sample is $\{W_i : i \leq n\}$. The true value θ_0 is not necessarily identified. The results also apply when the moment functions in (2.1) depend on a parameter τ, i.e., when they are of the form $\{m_j(W_i, \theta, \tau) : j \leq p\}$, and a preliminary consistent and asymptotically normal estimator $\widehat{\tau}_n(\theta_0)$ of τ exists, see AJ2. In addition, the asymptotic results in AJ2 allow for moment equalities as well as moment inequalities.

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

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta)\}, \tag{2.2}$$

where $c_n(\theta)$ is a data-dependent critical value.[3]

We now describe the recommended test statistic and critical value. The justifications for these recommendations are described below and are given in detail in AJ2. The recommended test statistic is an adjusted quasi-likelihood ratio (AQLR) statistic, $T_{AQLR,n}(\theta)$, that is a function of the sample moment conditions, $n^{1/2}\overline{m}_n(\theta)$, and an estimator of their asymptotic variance, $\widehat{\Sigma}_n(\theta)$:

$$\begin{aligned} T_{AQLR,n}(\theta) &= S_{2A}(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)) \\ &= \inf_{t \in R^p_{+,\infty}} (n^{1/2}\overline{m}_n(\theta) - t)'\widetilde{\Sigma}_n^{-1}(\theta)(n^{1/2}\overline{m}_n(\theta) - t), \text{ where} \\ \overline{m}_n(\theta) &= (\overline{m}_{n,1}(\theta), ..., \overline{m}_{n,p}(\theta))', \ \overline{m}_{n,j}(\theta) = n^{-1}\sum_{i=1}^n m_j(W_i, \theta) \text{ for } j \leq p, \\ \widetilde{\Sigma}_n(\theta) &= \widehat{\Sigma}_n(\theta) + \max\{\varepsilon - \det(\widehat{\Omega}_n(\theta)), 0\}\widehat{D}_n(\theta), \ \varepsilon = .012, \\ \widehat{D}_n(\theta) &= Diag(\widehat{\Sigma}_n(\theta)), \ \widehat{\Omega}_n(\theta) = \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta)\widehat{D}_n^{-1/2}(\theta), \end{aligned} \tag{2.3}$$

and $Diag(\Sigma)$ denotes the diagonal matrix based on the matrix Σ.[4] Note that the weight matrix $\widetilde{\Sigma}_n(\theta)$ depends only on $\widehat{\Sigma}_n(\theta)$ and hence $T_{AQLR,n}(\theta)$ can be written as a function of $(\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta))$. The function $S_{2A}(\cdot)$ is an adjusted version of the QLR function $S_2(\cdot)$ that appears in Section 3. The adjustment is designed to handle singular variance matrices. Specifically, the matrix $\widetilde{\Sigma}_n(\theta)$ equals the asymptotic variance matrix estimator $\widehat{\Sigma}_n(\theta)$ with an adjustment that ensures that $\widetilde{\Sigma}_n(\theta)$ is always nonsingular and is invariant to scale changes in the moment functions. The matrix $\widehat{\Omega}_n(\theta)$ is the correlation matrix that corresponds to $\widehat{\Sigma}_n(\theta)$.

[3]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 $<$.

[4]The constant $\varepsilon = .012$ was determined numerically based on an average asymptotic power criterion. See Section 6.2 of AJ2 for details.

When the observations are i.i.d. and no parameter τ appears, we take

$$\begin{aligned}\widehat{\Sigma}_n(\theta) &= n^{-1}\sum_{i=1}^{n}(m(W_i,\theta)-\overline{m}_n(\theta))(m(W_i,\theta)-\overline{m}_n(\theta))', \text{ where}\\ m(W_i,\theta) &= (m_1(W_i,\theta),...,m_p(W_i,\theta))'. \end{aligned} \tag{2.4}$$

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 AJ2. For example, with dependent observations, a heteroskedasticity and autocorrelation consistent (HAC) estimator may be required.

The test statistic $T_{AQLR,n}(\theta)$ is computed using a quadratic programming algorithm. Such algorithms are built into GAUSS and Matlab. They are very fast even when p is large. For example, to compute the AQLR test statistic 100,000 times takes 2.6, 2.9, and 4.7 seconds when $p = 2$, 4, and 10, respectively, using GAUSS on a PC with a 3.4 GHz processor.[5]

A moment selection critical value that utilizes a data-dependent tuning parameter $\widehat{\kappa}$ and size-correction factor $\widehat{\eta}$ is referred to as a *refined moment selection* (RMS) critical value. Our recommended RMS critical value is

$$c_n(\theta) = c_n(\theta,\widehat{\kappa}) + \widehat{\eta}, \tag{2.5}$$

where $c_n(\theta,\widehat{\kappa})$ is the $1-\alpha$ quantile of a bootstrap (or "asymptotic normal") distribution of a moment selection version of $T_{AQLR,n}(\theta)$ and $\widehat{\eta}$ is a data-dependent size-correction factor. For i.i.d. data, we recommend using a nonparametric bootstrap version of $c_n(\theta,\widehat{\kappa})$. For dependent data, either a block bootstrap or an asymptotic normal version can be applied.

We now describe the bootstrap version of $c_n(\theta,\widehat{\kappa})$. Let $\{W^*_{i,r} : i \leq n\}$ for $r = 1,...,R$ denote R bootstrap samples of size n (i.i.d. across samples), such as nonparametric i.i.d. bootstrap samples in an i.i.d. scenario or block bootstrap samples in a time series scenario, where R is large. Define the bootstrap variance matrix estimator $\widehat{\Sigma}^*_{n,r}(\theta)$ as $\widehat{\Sigma}_n(\theta)$ is defined (e.g., as in (2.4) in the i.i.d. case) with $\{W^*_{i,r} : i \leq n\}$ in place of $\{W_i : i \leq n\}$ throughout.[6] The p-vectors of re-centered bootstrap sample moments and

[5] In our experience, the GAUSS 9.0 quadratic programming procedure "qprog" is much faster than the Matlab 7 procedure "quadprog."

[6] Note that when a preliminary consistent estimator of a parameter τ appears, the bootstrap moment conditions need to be based on a bootstrap estimator of this preliminary estimator. In such cases, the

$p \times p$ bootstrap weight matrices for $r = 1, ..., R$ are defined by

$$\begin{aligned} m^*_{n,r}(\theta) &= n^{1/2}\left(\overline{m}^*_{n,r}(\theta) - \overline{m}_n(\theta)\right) \text{ and} \\ \widetilde{\Sigma}^*_{n,r}(\theta) &= \widehat{\Sigma}^*_{n,r}(\theta) + \max\{\varepsilon - \det(\widehat{\Omega}^*_{n,r}(\theta)), 0\}\widehat{D}^*_{n,r}(\theta), \text{ where } \varepsilon = .012, \\ \widehat{D}^*_{n,r}(\theta) &= Diag(\widehat{\Sigma}^*_{n,r}(\theta)), \text{ and } \widehat{\Omega}^*_{n,r}(\theta) = \widehat{D}^*_{n,r}(\theta)^{-1/2}\widehat{\Sigma}^*_{n,r}(\theta)\widehat{D}^*_{n,r}(\theta)^{-1/2}. \end{aligned} \tag{2.6}$$

The idea behind the RMS critical value is to compute the critical value using only those moment inequalities that have a noticeable effect on the asymptotic null distribution of the test statistic. Note that moment inequalities that have large positive population means have little or no effect on the asymptotic null distribution. Our preferred RMS procedure employs element-by-element t-tests of the null hypothesis that the mean of $\overline{m}_{n,j}(\theta)$ is zero versus the alternative that it is positive for $j = 1, ..., p$. The j-th moment inequality is selected if

$$\frac{n^{1/2}\overline{m}_{n,j}(\theta)}{\widehat{\sigma}_{n,j}(\theta)} \leq \widehat{\kappa}, \tag{2.7}$$

where $\widehat{\sigma}^2_{n,j}(\theta)$ is the (j, j) element of $\widehat{\Sigma}_n(\theta)$ for $j = 1, ..., p$ and $\widehat{\kappa}$ is a data-dependent tuning parameter (defined in (2.10) below) that plays the role of a critical value in selecting the moment inequalities. Let $\widehat{p}$ denote the number of selected moment inequalities.

For $r = 1, ..., R$, let $m^*_{n,r}(\theta, \widehat{p})$ denote the $\widehat{p}$-sub-vector of $m^*_{n,r}(\theta)$ that includes the $\widehat{p}$ selected moment inequalities.[7,8] Analogously, let $\widehat{\Sigma}^*_{n,r}(\theta, \widehat{p})$ denote the $(\widehat{p}\times\widehat{p})$-sub-matrix of $\widehat{\Sigma}^*_{n,r}(\theta)$ that consists of the $\widehat{p}$ selected moment inequalities. The bootstrap quantity $c_n(\theta, \widehat{\kappa})$ is the $1 - \alpha$ sample quantile of

$$\{S_{2A}(m^*_{n,r}(\theta, \widehat{p}), \widehat{\Sigma}^*_{n,r}(\theta, \widehat{p})) : r = 1, ..., R\}, \tag{2.8}$$

where $S_{2A}(\cdot, \cdot)$ is defined as in (2.3) but with p replaced by $\widehat{p}$.

An "asymptotic normal" version of $c_n(\theta, \widehat{\kappa})$ is obtained by replacing the bootstrap quantities $m^*_{n,r}(\theta, \widehat{p})$ and $\widehat{\Sigma}^*_{n,r}(\theta, \widehat{p})$ in (2.8) by $\widehat{\Sigma}^{1/2}_n(\theta, \widehat{p})Z^*_r$ and $\widehat{\Sigma}_n(\theta, \widehat{p})$, respectively, where $\widehat{\Sigma}_n(\theta, \widehat{p})$ denotes the $(\widehat{p} \times \widehat{p})$-sub-matrix of $\widehat{\Sigma}_n(\theta)$ that consists of the $\widehat{p}$ selected

asymptotic normal version of the critical value may be much quicker to compute.

[7]Note that $m^*_{n,r}(\theta, \widehat{p})$ depends not only on the number of moments selected, $\widehat{p}$, but which moments are selected. For simplicity, this is suppressed in the notation.

[8]By definition, $\widehat{p} \geq 1$, i.e., at least one moment must be selected. For specificity, $m^*_{n,r}(\theta, \widehat{p})$ equals the last element of $m^*_{n,r}(\theta)$ if no moments are selected via (2.7).

moment inequalities, $Z_r^* \sim i.i.d.\ N(0_{\widehat{p}}, I_{\widehat{p}})$ for $r = 1, ..., R$, and $\{Z_r^* : r = 1, ..., R\}$ are independent of $\{W_i : i \leq n\}$ conditional on $\widehat{p}$.

The tuning parameter $\widehat{\kappa}$ in (2.7) and the size-correction factor $\widehat{\eta}$ in (2.5) depend on the estimator $\widehat{\Omega}_n(\theta)$ of the asymptotic correlation matrix $\Omega(\theta)$ of $n^{1/2}\overline{m}_n(\theta)$. In particular, they depend on $\widehat{\Omega}_n(\theta)$ through a $[-1,1]$-valued function $\delta(\widehat{\Omega}_n(\theta))$ that is a measure of the amount of negative dependence in the correlation matrix $\widehat{\Omega}_n(\theta)$. We define

$$\delta(\Omega) = \text{smallest off-diagonal element of } \Omega, \tag{2.9}$$

where Ω is a $p \times p$ correlation matrix. The moment selection tuning parameter $\widehat{\kappa}$ and the size-correction factor $\widehat{\eta}$ are defined by

$$\widehat{\kappa} = \kappa(\widehat{\delta}_n(\theta)) \text{ and } \widehat{\eta} = \eta_1(\widehat{\delta}_n(\theta)) + \eta_2(p), \text{ where } \widehat{\delta}_n(\theta) = \delta(\widehat{\Omega}_n(\theta)). \tag{2.10}$$

Table I provides values of $\kappa(\delta)$, $\eta_1(\delta)$, and $\eta_2(p)$ for $\delta \in [-1,1]$ and $p \in \{2, 3, ..., 10\}$ for tests with level $\alpha = .05$ and CS's with level $1 - \alpha = .95$. AJ2 provides simulated values of the mean and standard deviation of the asymptotic distribution of $c_n(\theta, \widehat{\kappa})$. These results, combined with the values of $\eta_1(\delta)$ and $\eta_2(p)$ in Table I, show that the size-correction factor $\widehat{\eta}$ typically is small compared to $c_n(\theta, \widehat{\kappa})$, but not negligible.[9]

Computation of the $\eta_2(p)$ values given in Table I by simulation is not easy because it requires computing the (asymptotic) maximum null rejection probability (MNRP) over a large number of null mean vectors μ and correlation matrices Ω. For this reason, we only provide $\eta_2(p)$ values for $p \leq 10$. For the correlation matrices, we consider both a fixed grid and randomly generated matrices. For the null mean vectors $\mu \in R^p_{+,\infty}$, computation of the $\eta_2(p)$ values is carried out initially for mean vectors that consist only of $0's$ and $\infty's$. Then, the differences are computed between the values obtained by maximization over such μ vectors and the values obtained by maximization over μ vectors that lie in (i) a fixed full grid, (ii) two partial grids, and (iii) 1,000 or 100,000 randomly generated μ vectors (depending on the variance matrix). The differences are found to be .0000 in most cases and small ($\leq .0018$) in all cases, see AJ2 for details. These results indicate, although do not establish unequivocally, that the maxima over $\mu \in R^p_{+,\infty}$ are obtained at μ vectors that consist only of $0's$ and $\infty's$.

[9] For example, for $p = 10$, $\Omega = I_{10}$, five moment inequalities binding, and five moment inequalities completely slack, 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.

In sum, the preferred RMS critical value, $c_n(\theta)$, and CS are computed using the following steps. One computes (i) $\widehat{\Omega}_n(\theta)$ defined in (2.4), (ii) $\widehat{\delta}_n(\theta) =$ smallest off-diagonal element of $\widehat{\Omega}_n(\theta)$, (iii) $\widehat{\kappa} = \kappa(\widehat{\delta}_n(\theta))$ using Table I, (iv) $\widehat{\eta} = \eta_1(\widehat{\delta}_n(\theta)) + \eta_2(p)$ using Table I, (v) the vector of selected moments using (2.7), (vi) the selected bootstrap sample moments, correlation matrices, and weight matrices $\{(m^*_{n,r}(\theta,\widehat{p}), \widehat{\Sigma}^*_{n,r}(\theta,\widehat{p}), \widetilde{\Sigma}^*_{n,r}(\theta,\widehat{p})) : r = 1, ..., R\}$, defined in (2.6) with the non-selected moment inequalities omitted, (vii) $c_n(\theta,\widehat{\kappa})$, which is the .95 sample quantile of $\{S_{2A}(m^*_{n,r}(\theta,\widehat{p}),\ \widehat{\Sigma}^*_{n,r}(\theta,\widehat{p})) : r = 1, ..., R\}$ (for a test of level .05 and a CS of level .95) and (viii) $c_n(\theta) = c_n(\theta,\widehat{\kappa}) + \widehat{\eta}$. The preferred RMS confidence set is computed by determining all the values θ for which the null hypothesis that θ is the true value is not rejected. For the asymptotic normal version of the recommended RMS critical value, in step (vi) one computes the selected sub-vector and sub-matrix of $\widehat{\Sigma}^{1/2}_n(\theta,\widehat{p})Z^*_r$ and $\widehat{\Sigma}_n(\theta,\widehat{p})$, defined in the paragraph following (2.8), and in step (vii) one computes the .95 sample quantile with these quantities in place of $m^*_{n,r}(\theta,\widehat{p})$ and $\widehat{\Sigma}^*_{n,r}(\theta,\widehat{p})$, respectively.

To compute the recommended bootstrap RMS test using $R =$10,000 simulation repetitions takes 1.3, 1.5, and 2.7 seconds when $p = 2$, 4, and 10, respectively, and $n = 250$ using GAUSS on a PC with a 3.4 GHz processor. For the "asymptotic normal" version, the times are .20, .25, and .45 seconds.

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), then a useful approach is to map out the general features of the CS using the "asymptotic normal" version of the MMM/t-Test/κ=2.35 test, which is extremely fast to compute, and then switch to the bootstrap version of the recommended RMS test to find the boundaries of the CS more precisely.[10]

[10] The "asymptotic normal" version of the MMM/t-Test/κ=2.35 test is defined just as the recommended RMS test is defined but with $(S_1, \kappa = 2.35, \eta = 0)$ in place of $(S_{2A}, \widehat{\kappa}, \widehat{\eta})$, respectively, where S_1 is defined in (3.2), and with the bootstrap replaced by the normal asymptotic distribution. The bootstrap version of this test is much slower to compute than the asymptotic normal version and, hence, we do not recommend that it is used for this purpose. The computation times for the "asymptotic normal" version of the MMM/t-Test/κ=2.35 test are .007, .014, and .03 seconds when $p = 2, 4$, and 10, respectively.

3 Test Statistics and Critical Values

We now describe the justification for the recommended RMS test. Details are given in AJ2. The test statistics $T_n(\theta)$ that we consider are of the form

$$T_n(\theta) = S(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)), \tag{3.1}$$

where S is a real function on $(R \cup \{+\infty\})^p \times \mathcal{V}$ and $\mathcal{V}$ is the space of $p \times p$ variance matrices. The leading examples of S are the AQLR function S_{2A} defined above, the QLR function S_2, which is the same as S_{2A} in (2.3) but with $\varepsilon = 0$ (and hence $\widetilde{\Sigma}_n(\theta) = \widehat{\Sigma}_n(\theta)$), the modified method of moments (MMM) function S_1, and the SumMax function S_3:

$$S_1(m, \Sigma) = \sum_{j=1}^{p} [m_j/\sigma_j]_-^2 \text{ and } S_3(m, \Sigma) = \sum_{j=1}^{p_1} [m_{(j)}/\sigma_{(j)}]_-^2, \tag{3.2}$$

where $[x]_- = \min\{x, 0\}$, $m = (m_1, ..., m_p)'$, σ_j^2 is the jth diagonal element of Σ, $[m_{(j)}/\sigma_{(j)}]_-^2$ denotes the jth largest value among $\{[m_\ell/\sigma_\ell]_-^2 : \ell = 1, ..., p\}$, and $p_1 < p$ is some specified integer.[11,12,13] The MMM statistic S_1 has been used by Pakes, Porter, Ho, and Ishii (2004), CHT, Fan and Park (2007), Romano and Shaikh (2008), AG, AS, and Bugni (2010); the (unadjusted) QLR statistic has been used by AG, AS, and Rosen (2008); and the Max and SumMax statistics S_3 have been used by AG, AS, and Azeem Shaikh.[14]

We consider the class of GMS critical values discussed in AS. They rely on a tuning parameter κ and moment selection functions $\varphi_j : (R \cup \{+\infty\})^p \times \Psi \to R_+$ for $j \leq p$, where Ψ is the set of all $p \times p$ correlation matrices. The leading examples of φ_j are

$$\varphi_j^{(1)}(\xi, \Omega) = \begin{cases} 0 & \text{if } \xi_j \leq 1 \\ \infty & \text{if } \xi_j > 1, \end{cases} \quad \varphi_j^{(2)}(\xi, \Omega) = [\kappa(\xi_j - 1)]_+, \; \varphi_j^{(3)}(\xi, \Omega) = [\xi_j]_+,$$
$$\varphi_j^{(4)}(\xi, \Omega) = \kappa\xi_j 1(\xi_j > 1), \text{ and } \varphi_j^{(0)}(\xi, \Omega) = 0 \tag{3.3}$$

for $j \leq p$, where $[x]_+ = \max\{x, 0\}$, $\xi = (\xi_1, ..., \xi_p)'$, Ω is a $p \times p$ correlation matrix,

[11] When constructing a CS, a natural choice for p_1 is the dimension d of θ, see below.

[12] With the functions S_1, S_{2A}, and S_3, there is no restriction on the parameter space for the variance matrix Σ of the moment conditions—Σ can be singular.

[13] 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). This is not desirable in terms of the power of the resulting test.

[14] Personal communication.

and κ in $\varphi_j^{(2)}$ and $\varphi_j^{(4)}$ is the tuning parameter κ. Let $\varphi(\xi, \Omega) = (\varphi_1(\xi, \Omega), ..., \varphi_p(\xi, \Omega))'$ (for any $\varphi_j(\xi, \Omega)$ as in (3.3)). CHT, AS, and Bugni (2010) consider the function $\varphi^{(1)}$; Canay (2010) considers $\varphi^{(2)}$; AS considers $\varphi^{(3)}$; and Fan and Park (2007) use a non-scale-invariant version of $\varphi^{(4)}$. The function $\varphi^{(1)}$ generates the recommended "moment selection t-test" procedure of (2.7), see AJ2 for details. The function $\varphi^{(0)}$ generates a critical value based on the least-favorable distribution evaluated at an estimator of the true variance matrix Σ. It only depends on the data through the estimation of Σ. It is referred to as the "plug-in" asymptotic (PA) critical value. (No value κ is needed for this critical value.) Another φ function is the modified moment selection criterion (MMSC) $\varphi^{(5)}$ function introduced in AS. It is computationally more expensive than the functions $\varphi^{(1)}$-$\varphi^{(4)}$ considered above, but uses all of the information in the p-vector of moment conditions to decide which moments to select. It is a one-sided version of the information-criterion-based moment selection criterion considered in Andrews (1999). For brevity, we do not define $\varphi^{(5)}$ here, but we consider it below.

For a GMS critical value as in AS, $\{\kappa = \kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \to \infty$, such as $\kappa_n = (\ln n)^{1/2}$. 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 $\widehat{\Omega}_n(\theta)$: $\widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta))$,where $\kappa(\cdot)$ is an R_{++}-valued function. We justify RMS critical values using asymptotics in which κ equals a *finite* constant plus $o_p(1)$, rather than asymptotics in which $\kappa \to \infty$ as $n \to \infty$. This differs from the asymptotics in other papers in the moment inequality literature.

There are four reasons for using finite-κ asymptotics. First, they provide better approximations because κ is finite, not infinite, in any given application. Second, for any given (S, φ), they allow one to compute a best κ value in terms of asymptotic average power, which in turn allows one to compare different (S, φ) functions (each evaluated at its own best κ value) in terms of asymptotic average power. One cannot determine a best κ value in terms of asymptotic average power when $\kappa \to \infty$ because asymptotic power is always higher if κ is smaller, asymptotic size does not depend on κ, and finite-sample size is worse if κ smaller.[15] Third, for the recommended (S, φ) functions, the finite-κ asymptotic formula for the best κ value lets one determine a data-dependent κ value that is approximately optimal in terms of asymptotic average power. Fourth, finite-κ

[15]This does not imply that one cannot size-correct a test and then consider the $\kappa \to \infty$ asymptotic properties of such a test. Rather, the point is that $\kappa \to \infty$ asymptotics do not allow one to determine a suitable formula for size correction for the reason given.

asymptotics permit one to compute size-correction factors that depend on κ, which is a primary determinant of a test's finite-sample size. In contrast, if $\kappa \to \infty$ the asymptotic properties of tests under the null hypothesis do not depend on κ. Even the higher-order errors in null rejection probabilities do not depend on κ, see Bugni (2010). Thus, with $\kappa \to \infty$ asymptotics, the determination of a desirable size-correction factor based on κ is not possible.

For brevity, the finite-κ asymptotic results are given in AJ2. These results include uniform asymptotic size and $n^{-1/2}$-local power results. We use these results to compare different (S, φ) functions below and to develop recommended $\widehat{\kappa}$ and $\widehat{\eta}$ values.

For $Z^* \sim N(0_p, I_p)$ and $\beta \in (R \cup \{+\infty\})^p$, let $q_S(\beta, \Omega)$ denote the $1-\alpha$ quantile of $S(\Omega^{1/2}Z^* + \beta, \Omega)$. For constants $\kappa > 0$ and $\eta \geq 0$, define

$$\begin{aligned} & AsyPow(\mu, \Omega, S, \varphi, \kappa, \eta) \\ = & P\left(S(\Omega^{1/2}Z^* + \mu, \Omega) > q_S\left(\varphi(\kappa^{-1}[\Omega^{1/2}Z^* + \mu], \Omega), \Omega\right) + \eta\right), \end{aligned} \tag{3.4}$$

where $\mu \in R^p$ and $\Omega \in \Psi$. The asymptotic power of an RMS test of the null hypothesis that the true value is θ, based on (S, φ) with data-dependent $\widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta))$, and $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$, is shown in AJ2 to be $AsyPow(\mu, \Omega(\theta), S, \varphi, \kappa(\Omega(\theta)), \eta(\Omega(\theta)))$, where μ is a p-vector whose elements depend on the limits (as $n \to \infty$) of the normalized population means of the p moment inequalities and $\Omega(\theta)$ is the population correlation matrix of the moment functions evaluated at the null value θ.

We compare the power of different RMS tests by comparing their asymptotic average power for a chosen set $\mathcal{M}_p(\Omega)$ of alternative parameter vectors $\mu \in R^p$ for a given correlation matrix Ω. The asymptotic average power of the RMS test based on $(S, \varphi, \kappa, \eta)$ for constants $\kappa > 0$ and $\eta \geq 0$ is

$$|\mathcal{M}_p(\Omega)|^{-1} \sum_{\mu \in \mathcal{M}_p(\Omega)} AsyPow(\mu, \Omega, S, \varphi, \kappa, \eta), \tag{3.5}$$

where $|\mathcal{M}_p(\Omega)|$ denotes the number of elements in $\mathcal{M}_p(\Omega)$.

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. In consequence, we specify the sets $\mathcal{M}_p(\Omega)$ considered below to be ones for which most vectors μ have half or more elements positive (since positive elements correspond to non-binding inequalities), which is suitable for the typical case in which $p \geq 2d$.

To compare (S,φ) functions based on asymptotic $\mathcal{M}_p(\Omega)$-average power requires choices of functions $(\kappa(\cdot),\eta(\cdot))$. We use the functions $\kappa^*(\Omega)$ and $\eta^*(\Omega)$ that are optimal in terms of maximizing asymptotic $\mathcal{M}_p(\Omega)$-average power. These are determined numerically, see AJ2 for details. Given Ω, $\kappa^*(\Omega)$, and $\eta^*(\Omega)$, we compare (S,φ) functions by comparing their values of the quantity in (3.5) evaluated at $\kappa=\kappa^*(\Omega)$, and $\eta=\eta^*(\Omega)$.

Once we have determined a recommended (S,φ), we determine data-dependent values $\widehat{\kappa}$ and $\widehat{\eta}$ that are suitable for use with this (S,φ) combination.

Note that generalized empirical likelihood (GEL) test statistics, including the empirical likelihood ratio (ELR) statistic, behave the same asymptotically (to the first order) as the (unadjusted) QLR statistic $T_n(\theta)$ based on S_2 under the null and local alternative hypotheses for *nonsingular* correlation matrices of the moment conditions. See Sections 8.1 and 10.3 of AG, Section 10.1 of AS, and Canay (2010). In consequence, although GEL statistics are not of the form given in (3.1), the asymptotic results of the present paper, given in AJ2, hold for such statistics under the assumptions given in AG for classes of moment condition correlation matrices whose determinants are bounded away from zero. Hence, in the latter case, the recommended $\widehat{\kappa}$ and $\widehat{\eta}$ values given in Table I can be used with GEL statistics. However, an advantage of the AQLR statistic in comparison to GEL statistics is that its asymptotic properties are known and well-behaved whether or not the moment condition correlation matrix is singular. There are also substantial computational reasons to prefer the AQLR statistic to GEL statistics such as ELR, see Section 6 below.

4 Asymptotic Average Power Comparisons

In the numerical work reported here, we focus on results for $p=2,4$, and 10. For each value of p, we consider three correlation matrices Ω: Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}. The matrix Ω_{Zero} equals I_p for $p=2,4$, and 10. The matrices Ω_{Neg} and Ω_{Pos} are Toeplitz matrices with correlations on the diagonals (as they go away from the main diagonal) given by the following: For $p=2$: $\rho=-.9$ for Ω_{Neg} and $\rho=.5$ for Ω_{Pos}. For $p=4$: $\rho=(-.9,.7,-.5)$ for Ω_{Neg} and $\rho=(.9,.7,.5)$ for Ω_{Pos}. For $p=10$: $\rho=(-.9,.8,-.7,.6,-.5,.4,-.3,.2,-.1)$

for Ω_{Neg} and $\rho = (.9, .8, .7, .6, .5, ..., .5)$ for Ω_{Pos}.

For $p = 2$, the set of μ vectors $\mathcal{M}_2(\Omega)$ for which asymptotic average power is computed includes seven elements: $\mathcal{M}_2(\Omega) = \{(-\mu_1, 0), (-\mu_2, 1), (-\mu_3, 2), (-\mu_4, 3), (-\mu_5, 4), (-\mu_6, 7), (-\mu_7, -\mu_7)\}$, where μ_j depends on Ω and is such that the power envelope is .75 at each element of $\mathcal{M}_2(\Omega)$. Consistent with the discussion in Section 3, most elements of $\mathcal{M}_2(\Omega)$ have less than two negative elements. The positive elements of the μ vectors are chosen to cover a reasonable range of the parameter space. For brevity, the values of μ_j in $\mathcal{M}_2(\Omega)$ and the sets $\mathcal{M}_p(\Omega)$ for $p = 4, 10$ are given in AJ2. The elements of $\mathcal{M}_p(\Omega)$ for $p = 4, 10$ are selected such that the power envelope is .80 and .85, respectively, at each element of the set.

In AJ2 we also provide results for two singular Ω matrices and 19 nonsingular Ω matrices (for each p) that cover a grid of $\delta(\Omega)$ values from -1.0 to 1.0. The qualitative results reported here are found to apply as well to the broader range of Ω matrices. Some special features of the results based on the singular variance matrices are commented on below.

We compare tests based on the following functions: $(S, \varphi) =$ (MMM, PA), (MMM, t-Test), (Max, PA), (Max, t-Test), (SumMax, PA), (SumMax, t-Test), (AQLR, PA), (AQLR, t-Test), (AQLR, $\varphi^{(3)}$), (AQLR, $\varphi^{(4)}$), and (AQLR, MMSC).[16] We also consider the "pure ELR" test, for which Canay (2010) establishes a large deviation asymptotic optimality result. This test rejects the null when the ELR statistic exceeds a fixed constant (that is the same for all Ω).[17] The reason for reporting results for this test is to show that these asymptotic optimality results do not provide theoretical grounds for favoring the ELR test or ELR test statistic over other tests or test statistics.

For each test, Table II reports the asymptotic average power given the κ value that maximizes asymptotic average power for the test, denoted κ=Best. The best κ values are determined numerically using grid search, see AJ2 for details. For all tests and $p = 2, 4, 10$, the best κ values are decreasing from Ω_{Neg} to Ω_{Zero} to Ω_{Pos}. For example,

[16]The statistics MMM, AQLR, Max, and SumMax use the functions S_1, S_2, S_3 with $p_1 = 1$, and S_3 with $p_1 = 2$, respectively. The PA, t-Test, and MMSC critical values use the functions $\varphi^{(0)}$, $\varphi^{(1)}$, and $\varphi^{(5)}$, respectively.

[17]The level .05 pure ELR asymptotic critical value is determined numerically by calculating the constant for which the maximum asymptotic null rejection probability of the ELR statistic over all mean vectors in the null hypothesis and over all positive definite correlation matrices Ω is .05. See AJ2 for details. The critical values are found to be 5.07, 7.99, and 16.2 for $p = 2$, 4, and 10, respectively. These critical values yield asymptotic null rejection probabilities of .05 when Ω contains elements that are close to -1.0.

for the AQLR/t-Test test, the best κ values for $(\Omega_{Neg}, \Omega_{Zero}, \Omega_{Pos})$ are $(2.5, 1.4, .6)$ for $p = 10$, $(2.5, 1.4, .8)$ for $p = 4$, and $(2.6, 1.7, .6)$ for $p = 2$.

The asymptotic power results are size-corrected.[18,19] The critical values, size-correction factors, and power results are each calculated using $40,000$ simulation repetitions, except where stated otherwise, which yields a simulation standard error of .0011 for the power results.

Table II shows that the MMM/PA test has very low asymptotic power compared to the AQLR/t-Test/κBest test (which is shown in boldface) especially for $p = 4, 10$. Similarly, the Max/PA and SumMax/PA tests have low power. The AQLR/PA test has better power than the other PA tests, but it is still very low compared to the AQLR/t-Test/κBest test.

Table II also shows that the MMM/t-Test/κBest test has equal asymptotic average power to the AQLR/t-Test/κBest test for Ω_{Zero} and only slightly lower power for Ω_{Pos}. But, it has substantially lower power for Ω_{Neg}. For example, for $p = 10$, the comparison is .18 versus .55. The Max/t-Test/κBest test has noticeably lower average power than the AQLR/t-Test/κBest test for Ω_{Neg}, slightly lower power for Ω_{Zero}, and essentially equal power for Ω_{Pos}. It is strongly dominated in terms of average power. The SumMax/t-Test/κBest test also is strongly dominated by the AQLR/t-Test/κBest test in terms of asymptotic average power. The power differences between these two tests are especially large for Ω_{Neg}. For example, for $p = 10$ and Ω_{Neg}, their powers are .20 and .55, respectively.

Next we compare tests that use the AQLR test statistic but different critical values—due to the use of different functions φ. The AQLR/$\varphi^{(2)}$/κBest test is essentially dominated by the AQLR/t-Test/κBest, although the differences are not large. The AQLR /$\varphi^{(3)}$/κBest test has noticeably lower asymptotic average power than the AQLR/t-Test/κBest test for Ω_{Neg}, somewhat lower power for Ω_{Zero}, and equal power for Ω_{Pos}. The differences increase with p.

The AQLR/$\varphi^{(4)}$/κBest test has almost the same asymptotic average power as the AQLR/t-Test/κBest test for Ω_{Zero} and Ω_{Pos} and slightly lower power for Ω_{Neg}. This

[18] Size-correction here is done for the fixed known value of Ω. It is not based on the least-favorable Ω matrix because the results are asymptotic and Ω can be estimated consistently.

[19] The maximum null rejection probability calculations used in the size correction were calculated using μ vectors that consist of $0's$ and $\infty's$. Then, additional calculations were carried out to determine whether the maximum over $\mu \in R^p_{+,\infty}$ is attained at such a μ vector in each case. No evidence was found to suggest otherwise. See Section 7 of the Supplemental Material for details.

is because the $\varphi^{(4)}$ and $\varphi^{(1)}$ functions are similar. The AQLR/MMSC/κBest test and AQLR/t-Test/κBest tests have quite similar power. Nevertheless, the AQLR/MMSC/ κBest test is not the recommended test for reasons given below. We experimented with several smooth versions of the $\varphi^{(1)}$ critical value function in conjunction with the AQLR statistic. We were not able to find any that improved upon the asymptotic average power of the AQLR/t-Test/κBest test. Some were inferior. All such tests have substantial disadvantages relative to the AQLR/t-test in terms of the computational ease of determining suitable data-dependent κ and η values.

In conclusion, we find that the best (S, φ) choices in terms of asymptotic average power (based on κ=Best) are: AQLR/t-Test and AQLR/MMSC, followed closely by AQLR/$\varphi^{(2)}$ and AQLR/$\varphi^{(4)}$. Each of these tests out-performs the PA tests by a wide margin in terms of asymptotic power.

The AQLR/MMSC test has the following drawbacks: (i) its computation time is very high when p is large, such as $p = 10$, because the test statistic must be computed for all 2^p possible combinations of selected moment vectors and (ii) the best κ value varies widely with Ω and p, which makes it quite difficult to specify a data-dependent κ value that performs well. Similarly, the AQLR/$\varphi^{(2)}$ and AQLR/$\varphi^{(4)}$ tests have substantial computational drawbacks for determining a data-dependent κ values, see AJ2 for details.

Based on the power results discussed above and on the computational factors, we take the AQLR/t-Test to be the recommended test and we develop data-dependent $\widehat{\kappa}$ and $\widehat{\eta}$ for this test.

The last row of Table II gives the asymptotic power envelope, which is a "uni-directional" envelope, see AJ2 for details. One does not expect a test that is designed to perform well for multi-directional alternatives to be on, or close to, the uni-directional envelope. In fact, it is surprising how close the AQLR/t-Test/κBest test is to the power envelope when $\Omega = \Omega_{Pos}$. As expected, the larger is p the greater is the difference between the power of a test designed for p-directional alternatives and the uni-directional power envelope.

When the sample correlation matrix is singular, the QLR test statistic can be defined using the Moore-Penrose generalized inverse in the definition of the weighting matrix. Let MP-QLR denote this statistic. For the case of singular correlation matrices, AJ2 provides asymptotic power comparisons of the AQLR/t-Test/κBest test, the MP-QLR/t-Test/κBest test, and several other tests.

The results show that the AQLR/t-Test/κBest test has vastly superior asymptotic average power to that of the MP-QLR/t-Test/κBest test (e.g., .98 versus .29 when $p = 10$) when the correlation matrix exhibits perfect negative correlation and the same power when only perfect positive correlation is present. Hence, it is clear that the adjustment made to the QLR statistic is beneficial. The results also show that the AQLR/t-Test/κBest test strongly dominates tests based on the MMM and Max statistics in terms of asymptotic average power with singular correlation matrices.

Finally, results for the "pure ELR" test show that it has very poor asymptotic power properties.[20] For example, for $p = 10$, its power ranges 1/3 to 1/7 that of the AQLR/t-Test/κBest test (and of the feasible AQLR/t-Test/κAuto test, which is the recommended test of Section 2). The poor power properties of this "asymptotically optimal" test imply that the (generalized Neyman-Pearson) large deviations asymptotic optimality criterion is not a suitable criterion in this context.[21]

Note that the poor power of the "pure ELR" test does not imply that the ELR test statistic is a poor choice of test statistic. When combined with a good critical value, such as the data-dependent critical value recommended in this paper or a similar critical value, it yields a test with very good power. The point is that the large deviations asymptotic optimality result does not provide convincing evidence in favor of the ELR statistic.

5 Approximately Optimal $\kappa(\Omega)$ and $\eta(\Omega)$ Functions

Next, we describe how the recommended $\kappa(\Omega)$ and $\eta(\Omega)$ functions for the AQLR/t-Test test, defined in Section 2 and referred to, are determined.

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 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. Because the power results

[20] The power of the pure ELR test and AQLR/t-Test/κAuto test, which is the recommended test of Section 2, in the nine cases considered in Table II are: for $p = 10$, (.19, .55), (.17, .67), and (.12, .82); for $p = 4$, (.44, .59), (.42, .69), (.39, .78); and for $p = 2$, (.57, .65), (.55, .69), and (.54, .73). See Table S-XIII of AJ2.

[21] In our view, the large-deviation asymptotic optimality criterion is not appropriate when comparing tests with substantially different asymptotic properties under non-large deviations. In particular, this criterion is questionable when the alternative hypothesis is multi-dimensional because it implies that a test can be "optimal" against alternatives in all directions, which is incompatible with the finite sample and local asymptotic behavior of tests in most contexts.

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.

Second, by a combination of intuition and the analysis of numerical results, we postulate that for $p \geq 3$ the optimal function $\kappa^*(\Omega)$ is well approximated by a function that depends on Ω only through the $[-1, 1]$-valued function $\delta(\Omega)$ defined in (2.9).

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 larger is the fraction, $\eta^*(\Omega)/(c_n(\theta, \kappa^*(\Omega)) + \eta^*(\Omega))$ of the critical value that does not depend on the data (for Ω known), the closer is the critical value to the PA critical value that does not depend on the data at all (for known Ω), and the lower is the power of the test for μ vectors that have less than p elements negative and some elements strictly positive. (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.[22] If true, then the size of the AQLR/t-Test test depends on the two-dimensional sub-matrices of Ω that are the correlation matrices for 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 sub-matrix 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 $\delta(\Omega)$ defined in (2.9). (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. We take $\eta(\Omega)$ to be the value determined by $p = 2$ and δ, i.e., $\eta_1(\delta)$ in (2.10) and Table I, but allow for an adjustment that depends on p, viz., $\eta_2(p)$, that is defined to guarantee that the test has correct asymptotic significance level (up to numerical error).

[22] The reason for this postulation is that a test with given κ has larger null rejection probability the more negative are the correlations between the moments. A variance matrix of dimension three by three or greater has restrictions on its correlations imposed by the positive semi-definiteness property. If all of the correlations are equal, they cannot be arbitarily close to -1. In constrast, with a two-dimensional variance matrix, the correlation can be arbitarily close to -1.

See AJ2 for details.

We refer to the proposed method of selecting $\kappa(\Omega)$ and $\eta(\Omega)$, described in Section 2, as the κAuto method. We examine numerically how well the κAuto method does in approximating the best κ, viz., $\kappa^*(\Omega)$.[23] We provide four groups of results and consider $p = 2, 4, 10$ for each group. The first group consists of the three Ω matrices considered in Table II. The rows of Table II for the AQLR/t-Test/κBest and AQLR/t-Test/κAuto tests show that the κAuto method works very well. It has the same asymptotic average power as the AQLR/t-Test/κBest test for all p and Ω values except one case where the difference is just .01.

The second group consists of a set of 19 Ω matrices for which $\delta(\Omega)$ takes values on a grid in $[-.99, .99]$. In 53 of the 57 (=3×19) cases, the difference in asymptotic average power of the AQLR/t-Test/κBest and AQLR/t-Test/κAuto tests is less than .01.

The third group consists of two singular Ω matrices. One with perfect negative and positive correlations and the other with perfect positive correlations. The AQLR/t-Test/κAuto test has the same asymptotic average power as the AQLR/t-Test/κBest test for 3 (p, Ω) combinations, power that is lower by .01 for 2 combinations, and power that is lower by .02 for one combination.

The fourth group consists 500 randomly generated Ω matrices for $p = 2, 4$ and 250 randomly generated Ω matrices for $p = 10$. For $p = 2$, across the 500 Ω matrices, the asymptotic 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.

6 Finite-Sample Results

The recommended RMS test, AQLR/t-Test/κAuto, can be implemented in finite samples via the "asymptotic normal" and the bootstrap versions of the t-Test/κAuto critical value. Here we determine which of these two methods performs better in finite samples. We also compare these tests to the bootstrap version of the ELR/t-Test/κAuto

[23]For brevity, details of the numerical results are given in AJ2.

test, which has the same first-order asymptotic properties as the AQLR-based tests (for correlation matrices whose determinants are bounded away from zero by $\varepsilon = .012$ or more). See Sec. 6.3.3 of AJ2 for the definition of the ELR statistic and details of its computation.

In short, we find that the bootstrap version (denoted Bt in Table III) of the AQLR/t-Test/κAuto test performs better than the asymptotic normal version (denoted Nm) in terms of the closeness of its null rejection probabilities to its nominal level and similarly on average in terms of its power. The AQLR bootstrap test also performs slightly better than the ELR bootstrap test in terms of power, is noticeably superior in terms of computation time, and is essentially the same (up to simulation error) in terms of null rejection probabilities. In addition, the AQLR bootstrap test is found to perform quite well in an absolute sense. Its null rejection probabilities are close to its nominal level and the difference between its finite-sample and asymptotic power is relatively small.

We provide results for sample size $n = 100$. We consider the same correlation matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos} as above and the same numbers of moment inequalities $p = 2$, 4, and 10. We take the mean zero variance I_p random vector $Z^\dagger = Var^{-1/2}(m(W_i, \theta))(m(W_i, \theta) - Em(W_i, \theta))$ to be i.i.d. across elements and consider three distributions for the elements: standard normal (i.e., N(0, 1)), t_3, and chi-squared with three degrees of freedom χ^2_3. All of these distributions are centered and scaled to have mean zero and variance one. The power results are "size-corrected" based on the true Ω matrix. For $p = 2$, 4, and 10, we use 5000, 3000, and 1000 critical value and rejection probability repetitions, respectively, for the results under the null and under the alternative.[24]

We note that the finite-sample testing problem for *any* moment inequality model fits into the framework above for some correlation matrix Ω and some distribution of $Z^\dagger$. Hence, the finite-sample results given here provide a level of generality that usually is lacking with finite-sample simulation results.

The upper part of Table III provides the finite-sample maximum null rejection probabilities (MNRP's) of the nominal .05 normal and bootstrap versions of the AQLR/t-Test/κAuto test as well the bootstrap version of the ELR/t-Test/κAuto test. The MNRP is the maximum rejection probability over mean vectors μ in the null hypothesis for a given correlation matrix Ω and a given distribution of $Z^\dagger$. The lower part of Table

[24] The binding constraint on the number of simulation repetitions is the ELR test, see below for details.

III provides MNRP-corrected finite-sample average power for the same three tests. The average power results are for the same mean vectors μ in the alternative hypothesis as considered above for asymptotic power.

Table III shows that the AQLR/t-Test/κAuto bootstrap test performs well with MNRP's in the range of [.043, .066]. In contrast, the AQLR normal test over-rejects somewhat for some Ω matrices with the normal and t_3 distributions for which its MNRP's are in the range of $[.045, .092]$. With the skewed distribution, χ^2_3, the AQLR normal test over-rejects the null hypothesis substantially with its MNRP's being in the range $[.068, .153]$. The fact that over-rejection is largest for a skewed distribution is not surprising because the first term in the Edgeworth expansion of a sample average is a skewness term and the statistics considered here are simple functions of sample averages.

The ELR bootstrap test performs similarly to the AQLR bootstrap test in terms of null rejection probabilities. Its average amount of over-rejection over the 27 cases is .012, whereas it is .005 for the AQLR bootstrap test.

For the N(0, 1), t, and χ^2_3 distributions, Table III shows that the AQLR bootstrap test has finite-sample average power compared to the AQLR normal test that is similar, inferior, and superior, respectively.

The ELR bootstrap test performs similarly to the AQLR bootstrap test in terms of power. Computation of the ELR/t-Test/κAuto bootstrap test using $R =$10,000 simulation repetitions takes 9.5, 11.8, 16.2, 31.9, 59.0, and 182.7 seconds when $p = 2$, 4, 10, 20, 30, and 50, respectively, and $n = 250$ using GAUSS on a PC with a 3.4 GHz processor. This is slower than the AQLR/t-Test/κAuto bootstrap test (see Section 2) by a factor of 4.4 to 7.7.

AJ2 reports additional finite-sample results for the case of singular correlation matrices. The results for the AQLR/t-Test/κAuto test show that the bootstrap version performs better than the normal version in terms of MNRP's but similarly in terms of average power. Both tests perform well in an absolute sense. The bootstrap version of the MP-QLR/t-Test/κAuto test also is found to have good MNRP's. However, its finite-sample average power is much inferior to that of the AQLR/t-Test/κAuto bootstrap test—quite similar to the asymptotic power differences.

For the ELR/t-Test/κAuto bootstrap test, results for singular correlation matrices are reported in AJ2 only for the case of $p = 2$. The reason is that with a singular correlation matrix, the Hessian of the empirical likelihood objective function is singular a.s., which causes difficulties for standard derivative-based optimization algorithms when computing the ELR test statistic. With $p = 4$ and $p = 10$, the constrained optimization algorithm in GAUSS exhibits convergence problems and computation times are prohibitively large. For $p = 2$, the ELR bootstrap test's performance is essentially the same as that of the AQLR bootstrap test in terms of MNRP's and power.

In conclusion, we find that the AQLR/t-test/κAuto bootstrap test, which is the recommended test, performs well in an absolute sense with both nonsingular and singular variance matrices and out-performs the other tests considered in terms of asymptotic and finite-sample MNRP's or power, computational time, and/or computational stability.

References

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

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.

Andrews, D. W. K. and P. Jia (2008): "Supplemental Material to 'Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure'," unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica,* 78, 119-157.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica,* 78, 735-753.

Canay, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics,* 156, 408-425.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica,* 75, 1243-1284.

Fan, Y. and S. Park (2007): "Confidence Sets for Some Partially Identified Parameters," unpublished manuscript, Department of Economics, Vanderbilt University.

Imbens, G. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica,* 72, 1845-1857.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2004): "Applications of Moment Inequalities," unpublished working paper, Department of Economics, Harvard University.

Romano, J. P. and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Inference and Planning (Special Issue in Honor of T. W. Anderson),* 2786-2807.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics,* 146, 107-117.

Table I. Moment Selection Tuning Parameters $\kappa(\delta)$ and Size-Correction Factors $\eta_1(\delta)$ and $\eta_2(p)$ for $\alpha = .05$[1]

δ	$\kappa(\delta)$	$\eta_1(\delta)$	δ	$\kappa(\delta)$	$\eta_1(\delta)$	δ	$\kappa(\delta)$	$\eta_1(\delta)$
$[-1, -.975)$	2.9	.025	$[-.30, -.25)$	2.1	.111	$[.45, .50)$	0.8	.023
$[-.975, -.95)$	2.9	.026	$[-.25, -.20)$	2.1	.082	$[.50, .55)$	0.6	.033
$[-.95, -.90)$	2.9	.021	$[-.20, -.15)$	2.0	.083	$[.55, .60)$	0.6	.013
$[-.90, -.85)$	2.8	.027	$[-.15, -.10)$	2.0	.074	$[.60, .65)$	0.4	.016
$[-.85, -.80)$	2.7	.062	$[-.10, -.05)$	1.9	.082	$[.65, .70)$	0.4	.000
$[-.80, -.75)$	2.6	.104	$[-.05, .00)$	1.8	.075	$[.70, .75)$	0.2	.003
$[-.75, -.70)$	2.6	.103	$[.00, .05)$	1.5	.114	$[.75, .80)$	0.0	.002
$[-.70, -.65)$	2.5	.131	$[.05, .10)$	1.4	.112	$[.80, .85)$	0.0	.000
$[-.65, -.60)$	2.5	.122	$[.10, .15)$	1.4	.083	$[.85, .90)$	0.0	.000
$[-.60, -.55)$	2.5	.113	$[.15, .20)$	1.3	.089	$[.90, .95)$	0.0	.000
$[-.55, -.50)$	2.5	.104	$[.20, .25)$	1.3	.058	$[.95, .975)$	0.0	.000
$[-.50, -.45)$	2.4	.124	$[.25, .30)$	1.2	.055	$[.975, .99)$	0.0	.000
$[-.45, -.40)$	2.2	.158	$[.30, .35)$	1.1	.044	$[.99, 1.0]$	0.0	.000
$[-.40, -.35)$	2.2	.133	$[.35, .40)$	1.0	.040			
$[-.35, -.30)$	2.1	.138	$[.40, .45)$	0.8	.051			

p	2	3	4	5	6	7	8	9	10
$\eta_2(p)$	.00	.15	.17	.24	.31	.33	.37	.45	.50

[1]The values in Table I are obtained by simulating asymptotic formulae using 40,000 critical-value and 40,000 rejection-probability simulation repetitions, see AJ2 for details.

Table II. Asymptotic Average Power Comparisons (Size-Corrected): MMM, Max, SumMax, & AQLR Statistics, & PA, t-Test, $\varphi^{(2)}$, $\varphi^{(3)}$, $\varphi^{(4)}$, & MMSC Critical Values with κ=Best[1]

	Crit.	Tuning	$p=10$			$p=4$			$p=2$		
Stat.	Val.	Par. κ	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
MMM	PA	-	.04	.36	.34	.20	.53	.45	.48	.62	.59
MMM	t-Test	Best	.18	.67	.79	.31	.69	.76	.51	.69	.72
Max	PA	-	.19	.44	.70	.30	.57	.71	.48	.64	.66
Max	t-Test	Best	.25	.58	.82	.35	.66	.78	.51	.69	.72
SumMax	PA	-	.10	.43	.62	.20	.55	.60	.48	.62	.59
SumMax	t-Test	Best	.20	.65	.81	.31	.69	.77	.51	.69	.72
AQLR	PA	-	.35	.36	.69	.46	.53	.70	.58	.69	.65
AQLR	***t*-Test**	**Best**	**.55**	**.67**	**.82**	**.60**	**.69**	**.78**	**.65**	**.69**	**.73**
AQLR	t-Test	Auto	.55	.67	.82	.59	.69	.78	.65	.69	.73
AQLR	$\varphi^{(2)}$	Best	.51†	.65†	.81†	.60$^{\diamond}$	.69*	.78*	.66*	.69*	.72*
AQLR	$\varphi^{(3)}$	Best	.43†	.63†	.81†	.55$^{\diamond}$	.68*	.78*	.61*	.69*	.72*
AQLR	$\varphi^{(4)}$	Best	.51†	.65†	.81†	.60$^{\diamond}$	.70*	.78*	.66*	.69*	.72*
AQLR	MMSC	Best	.56†	.66†	.81†	.63	.69	.78	.65	.69	.73
Power	Envelope	-	.85	.85	.85	.80	.80	.80	.75	.75	.75

[1] κ=Best denotes the κ value that maximizes asymptotic average power. All cases not marked with a *, $^\diamond$, or † are based on (40,000, 40,000, 40,000) critical-value, size-correction, and power repetitions, respectively.

*Results are based on (5000, 5000, 5000) repetitions.

$^\diamond$Results are based on (2000, 2000, 2000) repetitions.

†Results are based on (1000, 1000, 1000) repetitions.

Table III. Finite-Sample Maximum Null Rejection Probabilities (MNRP's) and ("Size-Corrected") Average Power of the Nominal .05 AQLR/t-Test/κAuto Test with Normal (AQLR/Nm) and Bootstrap-Based (AQLR/Bt) Critical Values and ELR/t-Test/κAuto Test with Bootstrap-Based (ELR/Bt) Critical Values

			$p=10$			$p=4$			$p=2$		
Test	Dist	H_0/H_1	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
AQLR/Nm	N(0,1)	H_0	.088	.092	.057	.065	.062	.049	.056	.058	.053
AQLR/Bt	N(0,1)	H_0	.061	.062	.058	.053	.056	.049	.054	.053	.052
ELR/Bt	N(0,1)	H_0	.075	.076	.073	.059	.065	.054	.055	.058	.053
AQLR/Nm	t_3	H_0	.059	.067	.045	.050	.049	.047	.053	.047	.046
AQLR/Bt	t_3	H_0	.043	.055	.055	.051	.058	.052	.057	.055	.056
ELR/Bt	t_3	H_0	.059	.072	.072	.056	.067	.058	.057	.057	.055
AQLR/Nm	χ^2_3	H_0	.136	.153	.068	.093	.101	.062	.085	.087	.080
AQLR/Bt	χ^2_3	H_0	.062	.066	.057	.050	.055	.050	.054	.053	.056
ELR/Bt	χ^2_3	H_0	.068	.077	.065	.054	.061	.054	.053	.054	.055
AQLR/Nm	N(0,1)	H_1	.45	.59	.78	.54	.63	.76	.63	.68	.71
AQLR/Bt	N(0,1)	H_1	.46	.62	.77	.54	.64	.76	.63	.68	.71
ELR/Bt	N(0,1)	H_1	.49	.61	.76	.56	.64	.75	.63	.68	.71
AQLR/Nm	t_3	H_1	.58	.69	.84	.66	.76	.81	.70	.76	.72
AQLR/Bt	t_3	H_1	.56	.67	.79	.61	.71	.78	.67	.72	.71
ELR/Bt	t_3	H_1	.55	.62	.76	.61	.67	.76	.64	.68	.71
AQLR/Nm	χ^2_3	H_1	.37	.42	.72	.48	.53	.71	.56	.57	.61
AQLR/Bt	χ^2_3	H_1	.43	.51	.72	.53	.57	.70	.57	.59	.62
ELR/Bt	χ^2_3	H_1	.41	.47	.70	.53	.56	.70	.56	.59	.62

Supplemental Material to "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure"

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

Panle Jia
Department of Economics
MIT

June 2008
Revised: May 2011

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant SES-0751517. The authors thank Steve Berry for numerous discussions and comments and the co-editor and three referees for comments.

Contents

1 Introduction

This paper contains Supplemental Material to the paper Andrews and Jia (2008), which we refer to hereafter as AJ1.

The contents of this paper are summarized as follows.

Sections 2-5 provide the asymptotic results upon which AJ1 is based.

Section 2 specifies the model considered, which allows for both moment inequalities and equalities (whereas AJ1 only considers moment inequalities).

Section 3 defines the class of test statistics that are considered.

Section 4 defines in detail the class of refined moment selection (RMS) critical values that are introduced in AJ1, 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 5 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 uni-dimensional asymptotic power envelope.

Section 6 provides supplemental numerical results to those reported in AJ1. Section 6.1 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 6.2 discusses the determination of the recommended adjustment constant $\varepsilon = .012$ for the recommended AQLR test statistic. Section 6.3 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 6.4 provides tables of the κ values that maximize asymptotic average power (i.e., the best κ values), which are used in the construction of Table II of AJ1 and of the asymptotic MNRP's (which are used for "size-correction") of the RMS tests that appear in Table II of AJ1 (which reports asymptotic power) when no size-correction factor is employed, i.e., $\eta = 0$. Section 6.5 is similar to Section 4 of AJ1, which compares the asymptotic power of various RMS tests, except that it considers 19 correlation matrices Ω (rather than three) but fewer tests. Section 6.6 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 6.7 gives asymptotic MNRP and power results for some tests that are not considered in AJ1. Section 6.8 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 6.9 provides information on the magnitude of the (random) RMS critical values for the recommended AQLR/t-Test/κAuto test.

Section 7 provides details concerning the numerical results reported in AJ1 and in Section 6 of this paper. Section 7.1 provides the μ vectors used in AJ1 (which define the alternatives over which asymptotic and finite-sample average power is computed). Section 7.2 describes some details concerning the assessment of the properties of the automatic method of choosing κ. Section 7.3 discusses the determination and computation of the asymptotic power envelope. Section 7.4 discusses the computation of the κ values that maximize asymptotic average power that are reported in Table II of AJ1. Sections 7.5 and 7.6 describe the numerical computation of $\eta_2(p)$, which is part of the recommended size-correction function $\eta(\cdot)$.

Section 8 describes the GAUSS computer programs that were used to compute the numerical results.

Section 9 gives an alternative parametrization of the moment inequality/equality model to that given in AJ1 (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 5.

Throughout, we use the following notation. Let $R_+ = \{x \in R : x \geq 0\}$, $R_{++} = \{x \in R : x > 0\}$, $R_{+,\infty} = R_+ \cup \{+\infty\}$, $R_{[+\infty]} = R \cup \{+\infty\}$, $R_{[\pm\infty]} = R \cup \{\pm\infty\}$, $K^p = K \times ... \times K$ (with p copies) for any set K, $\infty^p = (+\infty, ..., +\infty)'$ (with p copies).

All limits are as $n \to \infty$ unless specified otherwise. Let "df" abbreviate "distribution function," "pd" abbreviate "positive definite," $cl(\Psi)$ denote the closure of a set Ψ, and 0_v denote a v-vector of zeros.

2 Moment Inequality/Equality Model

For brevity, the model considered in AJ1 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:

$$\begin{aligned} &E_{F_0} m_j(W_i, \theta_0) \geq 0 \text{ for } j = 1, ..., p \text{ and} \\ &E_{F_0} m_j(W_i, \theta_0) = 0 \text{ for } j = p+1, ..., p+v, \end{aligned} \tag{2.1}$$

where $\{m_j(\cdot, \theta) : j = 1, ..., k\}$ 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 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:

$$\begin{aligned} &\text{(i) } \theta \in \Theta, \text{ (ii) } E_F m_j(W_i, \theta) \geq 0 \text{ for } j = 1, ..., p, \text{ (iii) } E_F m_j(W_i, \theta) = 0 \\ &\text{for } j = p+1, ..., k, \text{ (iv) } \{W_i : i \geq 1\} \text{ are i.i.d. under } F, \\ &\text{(v) } \sigma^2_{F,j}(\theta) = Var_F(m_j(W_i, \theta)) > 0, \text{ (vi) } Corr_F(m(W_i, \theta)) \in \Psi, \text{ and} \\ &\text{(vii) } E_F |m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq M \text{ for } j = 1, ..., k, \end{aligned} \tag{2.2}$$

where $Var_F(\cdot)$ and $Corr_F(\cdot)$ 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 3, 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 9 below. The asymptotic results also apply when the moment functions in (2.1) depend on a parameter τ, i.e., 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 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 9 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

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta)\}, \tag{2.3}$$

where $c_n(\theta)$ is a data-dependent critical value.[1] 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.

3 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

$$\begin{aligned} &T_n(\theta) = S(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)), \text{ where} \\ &\overline{m}_n(\theta) = (\overline{m}_{n,1}(\theta), ..., \overline{m}_{n,k}(\theta))', \ \overline{m}_{n,j}(\theta) = n^{-1}\sum_{i=1}^n m_j(W_i, \theta) \text{ for } j \leq k, \end{aligned} \tag{3.1}$$

$\widehat{\Sigma}_n(\theta)$ 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

[1] 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 $<$.

moments $n^{1/2}\overline{m}_n(\theta)$. When the observations are i.i.d. and no parameter τ appears,

$$\begin{aligned}\widehat{\Sigma}_n(\theta) &= n^{-1}\sum_{i=1}^{n}(m(W_i,\theta)-\overline{m}_n(\theta))(m(W_i,\theta)-\overline{m}_n(\theta))', \text{ where} \\ m(W_i,\theta) &= (m_1(W_i,\theta),...,m_p(W_i,\theta))'. \end{aligned} \tag{3.2}$$

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

$$\widehat{\Omega}_n(\theta) = \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta)\widehat{D}_n^{-1/2}(\theta), \text{ where } \widehat{D}_n(\theta) = Diag(\widehat{\Sigma}_n(\theta)) \tag{3.3}$$

and $Diag(\Sigma)$ 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 9. 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

$$\begin{aligned} S_1(m,\Sigma) &= \sum_{j=1}^{p}[m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v}(m_j/\sigma_j)^2, \text{ where} \\ [x]_- &= \begin{cases} x & \text{if } x < 0 \\ 0 & \text{if } x \geq 0, \end{cases} \quad m = (m_1,...,m_k)', \end{aligned} \tag{3.4}$$

and σ_j^2 is the jth diagonal element of Σ. AJ1 lists papers in the literature that consider this test statistic and the other test statistics below.[2]

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

$$S_2(m,\Sigma) = \inf_{t=(t_1,0_v):t_1\in R^p_{+,\infty}} (m-t)'\Sigma^{-1}(m-t). \tag{3.5}$$

The origin of the QLR S function is as follows. Suppose one replaces m in (3.5) 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 $H_0^*: \mu_1 \geq 0_p$ & $\mu_2 = 0_v$ and the alternative

[2] 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.

hypothesis $H_1^* : \mu_1 \ngeq 0_p$ &/or $\mu_2 \neq 0_v$. The QLR statistic has been considered in many papers on tests of inequality constraints, e.g., see Kudo (1963) and Silvapulle and Sen (2005, Sec. 3.8). In the moment inequality literature, it has been considered by Andrews and Guggenberger (2009) (AG), Andrews and Soares (2010) (AS), and Rosen (2008).

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 (3.1), 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 have determinants bounded away from zero.

Next we consider an adjusted QLR (AQLR) test function denoted S_{2A}, which is the recommended S function in AJ1. 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

$$\begin{aligned} S_{2A}(m,\Sigma) &= \inf_{t=(t_1,0_v):t_1\in R^p_{+,\infty}} (m-t)'\widetilde{\Sigma}_\Sigma^{-1}(m-t), \text{ where} \\ \widetilde{\Sigma}_\Sigma &= \Sigma + \max\{\varepsilon - \det(\widetilde{\Omega}_\Sigma), 0\} \widetilde{D}_\Sigma, \\ \widetilde{D}_\Sigma &= Diag(\widetilde{\Sigma}_\Sigma),\ \widetilde{\Omega}_\Sigma = \widetilde{D}_\Sigma^{-1/2}\widetilde{\Sigma}_\Sigma \widetilde{D}_\Sigma^{-1/2}, \text{ and } \varepsilon > 0. \end{aligned} \tag{3.6}$$

Note that the adjustment to the matrix Σ is designed so that $\widetilde{\Sigma}$ is invariant to scale changes in the moment functions. Based on the results in Section 6.3, the recommended choice of ε for S_{2A} is $\varepsilon = .012$.

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

$$S_3(m,\Sigma) = \sum_{j=1}^{p_1} [m_{(j)}/\sigma_{(j)}]_-^2 + \sum_{j=p+1}^{p+v} (m_j/\sigma_j)^2, \tag{3.7}$$

where $[m_{(j)}/\sigma_{(j)}]_-^2$ denotes the jth largest value among $\{[m_\ell/\sigma_\ell]_-^2 : \ell = 1, ..., p\}$ and $p_1 < p$ is some specified integer.[3,4]

[3] When constructing a CS, a natural choice for p_1 is the dimension d of θ, see Section 5.3 below.

[4] 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 (2.2) can be any non-empty 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

The asymptotic results given in Section 5 below 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 pd diagonal $D \in R^{k\times k}$.

(b) $S(m,\Omega) \geq 0$ 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 $\Omega \in \Psi$.[5]

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

(e) For all $\ell \in R^p_{[+\infty]} \times R^v$, all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the df 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 (2.2) 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 $AsyCorr_F(n^{1/2}\overline{m}_n(\theta))$.[6]

The functions S_1, S_{2A}, and S_3 satisfy Assumption S. The function S_2 satisfies Assumption S provided the determinants of the correlation matrices in Ψ are bounded away from zero.[7]

4 Refined Moment Selection

This section is concerned with critical values for use with the test statistics introduced in Section 3. 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. Lastly, we specify a size-correction factor $\widehat{\eta}$ that delivers correct asymptotic size even when $\widehat{\kappa}$ does

Ψ have determinants bounded away from zero because Σ^{-1} appears in the definition of S_2.

[5] 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.

[6] More specifically, for dependent observations or when a preliminary estimator of a parameter τ appears, Ψ is as in condition (v) of (9.2) in Section 9.

[7] For the functions S_1-S_3, see Lemma 1 of AG for a proof that Assumptions S(a)-S(d) hold and 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.

not diverge to infinity. Because the CS's defined in (2.3) are obtained by inverting tests, we discuss both tests and CS's below.

4.1 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

$$S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0), \text{ where } Z^* \sim N(0_k, I_k), \tag{4.1}$$

$h_1 \in R^p_{+,\infty}$, 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.[8]

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 by 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

$$\xi_n(\theta) = \widehat{\kappa}^{-1} n^{1/2} \widehat{D}_n^{-1/2}(\theta) \overline{m}_n(\theta) \in R^k, \tag{4.2}$$

where $\widehat{\kappa}$ is a tuning parameter. 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 $\kappa_n = (\ln n)^{1/2}$ 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.

[8] 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.

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

$$\widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta)), \tag{4.3}$$

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:

Assumption $\boldsymbol{\kappa}$. (a) $\kappa(\Omega)$ is continuous at all $\Omega \in \Psi$. (b) $\kappa(\Omega) > 0$ for all $\Omega \in \Psi$.[9]

4.2 Moment Selection Function φ

Next, we discuss the moment selection function φ that determines how non-binding 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 $j = 1, ..., p$. 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 $j = 1, ..., p$, 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

$$\varphi_j^{(1)}(\xi, \Omega) = \begin{cases} 0 & \text{if } \xi_j \leq 1 \\ \infty & \text{if } \xi_j > 1, \end{cases} \quad \varphi_j^{(2)}(\xi, \Omega) = [\kappa(\Omega)(\xi_j - 1)]_+,$$
$$\varphi_j^{(3)}(\xi, \Omega) = [\xi_j]_+, \text{ and } \varphi_j^{(4)}(\xi, \Omega) = \begin{cases} 0 & \text{if } \xi_j \leq 1 \\ \kappa(\Omega)\xi_j & \text{if } \xi_j > 1 \end{cases} \tag{4.4}$$

for $j = 1, ..., p$, where $[x]_+ = \max\{x, 0\}$ and $\kappa(\Omega)$ in $\varphi_j^{(2)}$ and $\varphi_j^{(4)}$ is the same tuning

[9] For simplicity, the recommended function $\kappa(\Omega) = \kappa(\delta(\Omega))$ given in AJ1 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 of AJ1 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 or not.

parameter function that appears in (4.3). Let $\varphi^{(r)}(\xi,\Omega) = (\varphi_1^{(r)}(\xi,\Omega),...,\varphi_p^{(r)}(\xi,\Omega), 0,...,0)' \in R^p_{[\pm\infty]} \times \{0\}^v$ for $r = 1,...,4$. Chernozhukov, Hong, and Tamer (2007), AS, and Bugni (2010) consider the function $\varphi^{(1)}$; Canay (2010) considers $\varphi^{(2)}$; AS considers $\varphi^{(3)}$; and Fan and Park (2007) consider $\varphi^{(4)}$.[10]

The function $\varphi^{(1)}$ generates a "moment selection t-test" procedure, which is the recommended φ function in AJ1. Note that $\xi_{n,j}(\theta_0) \leq 1$ is equivalent to the condition $n^{1/2}\overline{m}_{n,j}(\theta)/\widehat{\sigma}_{n,j}(\theta) \leq \widehat{\kappa}$ in AJ1.

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

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 one under a certain distribution:

Assumption φ. For all $j = 1,...,p$, 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 (4.4) all satisfy Assumption φ.

The functions $\varphi^{(r)}$ for $r = 1,...,4$ all exhibit "element by element" determination of which moments to "select" because $\varphi_j^{(r)}(\xi,\Omega)$ only depends 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 non-diagonal, 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 $c = (c_1,...,c_k)'$ 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 $j = p+1,...,k$. Let $|c| = \sum_{j=1}^k c_j$. For $\xi \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, define $c \cdot \xi = (c_1\xi_1,...,c_k\xi_k)' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, 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, e.g., $\mathcal{C} = \{0,1\}^p \times \{1\}^v$. Let $\zeta(\cdot)$

[10] 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 $\xi_j \leq 1$" replaced by "if $\widehat{\sigma}_{n,j}(\theta)\xi_j \leq 1$," 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.

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

$$S(-c \cdot \xi, \Omega) - \zeta(|c|). \tag{4.5}$$

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 $j = 1, ..., p$, define

$$\varphi_j^{(5)}(\xi, \Omega) = \begin{cases} 0 & \text{if } c_j(\xi, \Omega) = 1 \\ \infty & \text{if } c_j(\xi, \Omega) = 0. \end{cases} \tag{4.6}$$

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 $\varphi^{(1)}$.

4.3 RMS Critical Value $\mathbf{c_n}(\boldsymbol{\theta})$

The (asymptotic normal) RMS critical value is equal to the $1 - \alpha$ quantile of $S\left(\Omega^{1/2} Z^* + \beta, \Omega\right)$ 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

$$\varphi(\xi, \Omega) = (\varphi_1(\xi, \Omega), ..., \varphi_p(\xi, \Omega), 0, ..., 0)' \in R^p_{[+\infty]} \times \{0\}^v, \tag{4.7}$$

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

$$\varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)). \tag{4.8}$$

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

$$S\left(\Omega^{1/2} Z^* + \beta, \Omega\right). \tag{4.9}$$

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

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

$$c_n(\theta) = q_S\left(\varphi\Big(\xi_n(\theta), \widehat{\Omega}_n(\theta)\Big), \widehat{\Omega}_n(\theta)\right) + \eta(\widehat{\Omega}_n(\theta)), \tag{4.10}$$

where $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ is a size-correction factor that is specified in Section 4.4 below.

The bootstrap RMS critical value is obtained by replacing $q_S(\varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)), \widehat{\Omega}_n(\theta))$ in (4.10) 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 $\beta \in R_{[+\infty]}^k$ and $m_{n,r}^*(\theta)$, $\widehat{D}_{n,r}^*(\theta)$, and $\widehat{\Omega}_{n,r}^*(\theta)$ are bootstrap quantities defined in AJ1. The quantity $q_S^*(\varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)))$ can be computed by taking the $1-\alpha$ sample quantile of $\{S(\widehat{D}_{n,r}^*(\theta)^{-1/2} m_{n,r}^*(\theta) + \varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)), \widehat{\Omega}_{n,r}^*(\theta)) : r = 1, ..., R\}$.

For the recommended RMS critical value defined in AJ1, the asymptotic normal critical value is of the form in (4.10) 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 $q_{S_{2A}}(\cdot, \widehat{\Omega}_n(\theta))$.

4.4 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 viewed as giving an asymptotic size refinement to a GMS critical value.

As noted above, we show in the proofs (see Section 9) 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 $(h_1, \Omega) \in R_{+,\infty}^p \times \Psi$. Furthermore, we show that under such a sequence the asymptotic coverage probability 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

$$\begin{aligned} CP(h_1, \Omega, \eta) = P\big(& S\left(\Omega^{1/2} Z^* + (h_1, 0_v), \Omega\right) \leq \\ & q_S\left(\varphi\left(\kappa^{-1}(\Omega)[\Omega^{1/2} Z^* + (h_1, 0_v)], \Omega\right), \Omega\right) + \eta\big), \end{aligned} \tag{4.11}$$

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 $1 - CP(h_1, \Omega, \eta)$.)

We let Δ ($\subset R_{+,\infty}^p \times \Psi$) 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

$$\gamma_1(\theta, F) = Diag^{-1/2}\left(AsyVar_F\left(n^{1/2}\overline{m}_n(\theta)\right)\right) E_F m(W_i, \theta) \geq 0_p \text{ and}$$
$$\Omega(\theta, F) = AsyCorr_F\left(n^{1/2}\overline{m}_n(\theta)\right), \quad (4.12)$$

where $AsyVar_F(n^{1/2}\overline{m}_n(\theta))$ and $AsyCorr_F(n^{1/2}\overline{m}_n(\theta))$ 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.[11] Then, Δ is defined by

$$\begin{aligned}\Delta = \{&(h_1, \Omega) \in R^p_{+,\infty} \times cl(\Psi) : \exists \text{ a subsequence } \{w_n\} \text{ of } \{n\} \text{ and}\\ &\text{a sequence } \{(\theta_{w_n}, F_{w_n}) \in \mathcal{F} : n \geq 1\} \text{ with } \gamma_1(\theta_{w_n}, F_{w_n}) \geq 0_p \text{ and}\\ &\Omega(\theta_{w_n}, F_{w_n}) \in \Psi \text{ for which } w_n^{1/2}\gamma_1(\theta_{w_n}, F_{w_n}) \to h_1,\ \Omega(\theta_{w_n}, F_{w_n}) \to \Omega,\\ &\text{and } \theta_{w_n} \to \theta_* \text{ for some } \theta_* \text{ in } cl(\Theta)\}. \end{aligned} \quad (4.13)$$

Our primary focus is on the standard case in which

$$\Delta = R^p_{+,\infty} \times cl(\Psi). \quad (4.14)$$

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 (4.13) in which Δ may be restricted, as well as the standard case in (4.14).

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

Assumption $\boldsymbol{\eta}$1. $\eta(\Omega)$ is continuous at all $\Omega \in \Psi$.[12]

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 $CP(h_1, \Omega, \eta(\Omega)-) = \lim_{x \downarrow 0} CP(h_1, \Omega, \eta(\Omega) - x)$.

[11] For dependent observations and when a preliminary estimator of a paramter τ appears, the parameter space $\mathcal{F}$ of (θ, F) is defined in Section 9.1 such that both $AsyVar_F(n^{1/2}\overline{m}_n(\theta))$ and $AsyCorr_F(n^{1/2}\overline{m}_n(\theta))$ exist. These limits equal $Var_F(m(W_i, \theta)))$ and $Corr_F(m(W_i, \theta)))$, respectively, in the case of i.i.d. observations with no preliminary estimator of a parameter τ.

[12] An analogous comment to that in footnote 9 also applies to the recommended function $\eta(\cdot)$ given in AJ1 and Assumption η1.

Assumption $\boldsymbol{\eta}$2. $\inf_{(h_1,\Omega)\in\Delta} CP(h_1,\Omega,\eta(\Omega)-) \geq 1-\alpha$.

Assumption $\boldsymbol{\eta}$3. (a) $\inf_{(h_1,\Omega)\in\Delta} CP(h_1,\Omega,\eta(\Omega)) = 1-\alpha$.

(b) $\inf_{(h_1,\Omega)\in\Delta} CP(h_1,\Omega,\eta(\Omega)-) = \inf_{(h_1,\Omega)\in\Delta} CP(h_1,\Omega,\eta(\Omega))$.

Assumption η3(b) is a continuity condition that is not restrictive. The left-hand side (lhs) quantity inside the probability in (4.11) has a df that is continuous and strictly increasing for positive values. The corresponding right-hand side (rhs) quantity is positive. These two quantities are quite different nonlinear functions of the same underlying normal random vector. Hence, they are equal with probability zero, 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 Assumptions η2 and/or η3 are not uniquely defined. The smallest function that satisfies Assumption η3(a), denoted $\eta^*(\Omega)$, exists and is defined as follows. For each $\Omega \in \Psi$, define $\eta^*(\Omega)$ to be the smallest value η for which

$$\inf_{h_1:(h_1,\Omega)\in\Delta} CP(h_1,\Omega,\eta) = 1-\alpha.^{13} \tag{4.15}$$

When Δ satisfies (4.14), 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 upon 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 $\eta^*(\Omega)$, e.g., a function that depends on Ω only through a scalar function of Ω, as with the recommended RMS critical value described in AJ1.[14]

4.5 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

[13] A smallest value exists because $CP(h_1,\Omega,\eta)$ is right continuous in η.

[14] Note that even if $\eta(\Omega) \neq \eta^*(\Omega)$, Assumption η3(a) still can hold.

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.[15]

The PA critical value is

$$q_S(0_k, \widehat{\Omega}_n(\theta)). \tag{4.16}$$

The PA critical value can be viewed as a special case of an RMS critical value with $\varphi_j(\xi, \Omega) = 0$ for all $j = 1, ..., k$ 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.

5 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), derives the asymptotic power of RMS tests against local alternatives, discusses an asymptotic average power criterion for comparing RMS tests, and discusses the uni-dimensional asymptotic power envelope.

5.1 Asymptotic Size

The exact and asymptotic confidence sizes of an RMS CS are

$$ExCS_n = \inf_{(\theta,F)\in\mathcal{F}} P_F(T_n(\theta) \le c_n(\theta)) \text{ and } AsyCS = \liminf_{n\to\infty} ExCS_n, \tag{5.1}$$

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 (2.2). 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 9 below.

[15] 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.

Theorem 1 *Suppose Assumptions* S, κ, φ, *and* $\eta 1$ *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*

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

(b) $AsyCS \geq 1 - \alpha$ *provided Assumption* $\eta 2$ *holds, and*

(c) $AsyCS = 1 - \alpha$ *provided Assumption* $\eta 3$ *holds.*

Comments. 1. Theorem 1(b) shows that an RMS CS based on a size-correction factor $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ that satisfies Assumption $\eta 2$ 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 $n \to \infty$.

2. 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) discusses models in which restrictions of this sort arise.

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

4. Define the asymptotic size of an RMS test of $H_0 : \theta = \theta_0$ by

$$AsySz(\theta_0) = \limsup_{n\to\infty} \sup_{(\theta,F)\in\mathcal{F}:\theta=\theta_0} P_F(T_n(\theta_0) > c_n(\theta_0)). \tag{5.2}$$

The proof of Theorem 1 shows that under the assumptions in Theorem 1, (a) $AsySz(\theta_0) \in [1 - \inf_{(h_1,\Omega)\in\Delta_0} CP(h_1, \Omega, \eta(\Omega)), 1 - \inf_{(h_1,\Omega)\in\Delta_0} CP(h_1, \Omega, \eta(\Omega)-)]$, where Δ_0 is defined as Δ is defined in (4.14) or as in (4.13) but with the sequence $\{\theta_{w_n} : n \geq 1\}$ replaced by the constant θ_0, (b) $AsySz(\theta_0) \leq \alpha$ provided Assumption $\eta 2$ holds, and (c) $AsySz(\theta_0) = \alpha$ 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 $\Delta_0 = R^p_{+,\infty} \times cl(\Psi)$, which occurs when there are no restrictions on the moment functions beyond the inequality/equality restrictions and h_1 and Ω are variation free.

5. The proofs of Theorem 1 and all other results stated here are provided in Section 9.

5.2 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

$$\begin{aligned} H_0 : E_F m_j(W_i, \theta_0) &\geq 0 \text{ for } j = 1, ..., p \text{ and} \\ E_F m_j(W_i, \theta_0) &= 0 \text{ for } j = p+1, ..., k, \end{aligned} \tag{5.3}$$

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

$$\begin{aligned} \sigma^2_{F,j}(\theta) &= AsyVar_F(n^{1/2}\overline{m}_{n,j}(\theta)) \text{ for } j = 1, ..., p, \\ D(\theta, F) &= Diag\{\sigma^2_{F,1}(\theta), ..., \sigma^2_{F,k}(\theta)\}, \text{ and} \\ \Omega(\theta, F) &= AsyCorr_F(n^{1/2}\overline{m}_n(\theta)). \end{aligned} \tag{5.4}$$

Note that this definition of $\sigma^2_{F,j}(\theta)$ reduces to that given in (2.2) when the observations are i.i.d. Let $\widehat{\sigma}^2_{n,j}(\theta)$ is the (j, j) element of $\widehat{\Sigma}_n(\theta)$ for $j = 1, ..., k$.

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:

(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 $j = 1, ..., p$, and

(c) $\sup_{n \geq 1} E_{F_n} |m_j(W_i, \theta_0)/\sigma_{F_n,j}(\theta_0)|^{2+\delta} < \infty$ for $j = 1, ..., k$ 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).[16]

[16] When a preliminary estimator of a parameter τ appears in the sample moment functions, then in Assumptions LA1 and LA2 and (5.5), $m_j(W_i, \theta)$ and $m(W_i, \theta)$ are defined to be $m_j(W_i, \theta, \tau_0)$ and

Assumption LA3. The true parameters $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ satisfy:

(a) $A_n^0 = (A_{n,1}^0, ..., A_{n,k}^0)' \to_d Z \sim N(0_k, \Omega_0)$ 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 $j = 1, ..., k$, and

(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 (3.2), 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

$$\lim_{n\to\infty} n^{1/2} D^{-1/2}(\theta_0, F_n) E_{F_n} m(W_i, \theta_0) = \mu = (h_1, 0_v) + \Pi_0 \lambda \in R^p_{[+\infty]} \times R^v, \text{ where}$$
$$h_1 = (h_{1,1}, ..., h_{1,p})' \text{ and } \Pi_0 = \Pi(\theta_0, F_0). \tag{5.5}$$

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 $j = 1, ..., k$. Note that $(h_1, 0_v) + \Pi_0\lambda \in R^p_{[+\infty]} \times R^v$. Let $\mu = (\mu_1, ..., \mu_k)'$. The true distribution F_n is in the alternative, not the null (for n large) when $\mu_j = h_{1,j} + \Pi_{0,j}'\lambda < 0$ for some $j = 1, ..., p$ or $\Pi_{0,j}'\lambda \neq 0$ for some $j = p+1, ..., k$.

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

$$\begin{aligned} &AsyPow(\mu, \Omega, S, \varphi, \kappa, \eta) \\ &= P\left(S(\Omega^{1/2}Z^* + \mu, \Omega) > q_S\left(\varphi(\kappa^{-1}[\Omega^{1/2}Z^* + \mu], \Omega), \Omega\right) + \eta\right) \text{ and} \\ &AsyPow^-(\mu, \Omega_0, S, \varphi, \kappa, \eta) = \lim_{x \downarrow 0} AsyPow(\mu, \Omega_0, S, \varphi, \kappa, \eta - x), \end{aligned} \tag{5.6}$$

where $Z^* \sim N(0_k, I_k)$, $\mu \in R^k$, $\Omega \in \Psi$, $\kappa \in R_{++}$, the functions S, φ, and q_S are as defined

$m(W_i, \theta, \tau_0)$, respectively, where τ_0 denotes the true value of the parameter τ under the true distribution F.

in Section 3, (4.4) or (4.6), and (4.9), respectively.[17] Typically, $AsyPow(\mu, \Omega, S, \varphi, \kappa, \eta) = AsyPow^{-}(\mu, \Omega, S, \varphi, \kappa, \eta)$ because the lhs quantity in the probability in (5.6) is a nonlinear function of a normal random vector that has a continuous and strictly increasing df (unless $v = 0$ and $\mu = \infty^p$, which cannot hold under the alternative hypothesis) and the rhs quantity in the probability in (5.6) 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, $\eta 1$, *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*

$$P_{F_n}(T_n(\theta_0) > c_n(\theta_0)) \\ \to [AsyPow(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0)), AsyPow^{-}(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0))],$$

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

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

2. The results of Theorem 3 hold under the null and alternative hypotheses.

3. 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 $j = 1, ..., k$.

5.3 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 $\Omega \in \Psi$.[18] Let $|\mathcal{M}_k(\Omega)|$ denote the number of elements

[17] 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 $x > 0$.

[18] 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.

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

$$|\mathcal{M}_k(\Omega)|^{-1} \sum_{\mu \in \mathcal{M}_k(\Omega)} AsyPow(\mu, \Omega, S, \varphi, \kappa, \eta). \tag{5.7}$$

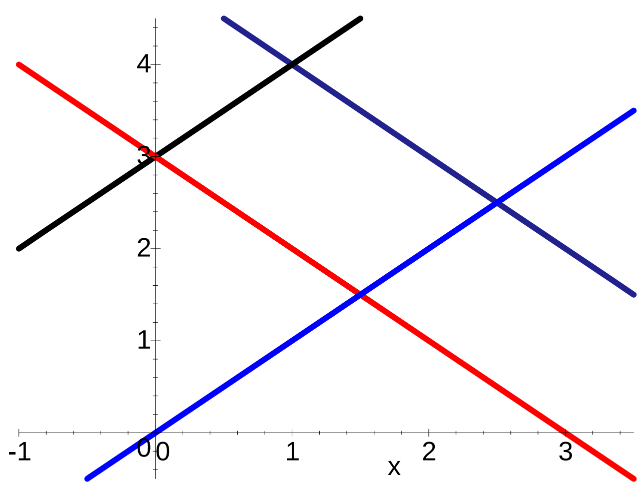
We are interested in comparing the (S, φ) functions defined in (3.4)-(3.7), (4.4), and (4.6) in terms of asymptotic $\mathcal{M}_k(\Omega)$-average power. To do so requires choices of functions $(\kappa(\cdot), \eta(\cdot))$ for each (S, φ). We use the tuning and size-correction functions $\kappa^*(\Omega)$ 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 (4.15) with $\Delta = R^p_{+,\infty} \times cl(\Omega)$ and tuning parameter $\kappa > 0$. The optimal tuning parameter $\kappa^*(\Omega)$ maximizes (5.7) 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

$$|\mathcal{M}_k(\Omega)|^{-1} \sum_{\mu \in \mathcal{M}_k(\Omega)} AsyPow(\mu, \Omega, S, \varphi, \kappa^*(\Omega), \eta^*(\Omega)), \tag{5.8}$$

which depend on Ω.

Figure 1. Confidence Set for a Parameter $\theta \in R^d$ for $d = 2$ Based on $p = 4$ Moment Inequalities



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. 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 non-binding inequalities), which is suitable for the typical case in which $p \geq 2d$.

5.4 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 7 below.

We note that the asymptotic power envelope is a "uni-directional" envelope. One does not expect a test that is designed to perform well for multi-directional alternatives to be on, or close to, the uni-directional 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 uni-dimensional power envelope. For example, for $p = 2, 4, 10$, when the asymptotic power envelope is $.75, .80, .85$, respectively, the F test has power $.65, .60, .49$, respectively.[19] Clearly, the larger is p the greater is the difference between the power of a test designed for p-directional alternatives and the uni-directional power envelope.

6 Numerical Results

This section gives supplemental numerical results to those given in AJ1.

Section 6.1 describes how the approximately optimal $\kappa(\cdot)$ and $\eta(\cdot)$ functions given in Table I of AJ1 are determined and provides numerical results concerning their properties.[20]

[19] 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.

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

Section 6.2 discusses the determination of the recommended adjustment constant $\varepsilon = .012$ for the recommended AQLR test statistic, which is based on the S_{2A} function.[21]

Section 6.3 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 6.4 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 of AJ1. Section 6.4 also provides a table that is analogous to Table II of AJ1 but reports asymptotic MNRP's rather than asymptotic power.

Section 6.5 provides results that supplement those of AJ1 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 6.7 provides additional asymptotic MNRP and power results for some GMS and RMS tests that are not considered explicitly in AJ1.

Section 6.8 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.[22]

6.1 Approximately Optimal $\boldsymbol{\kappa}(\boldsymbol{\Omega})$ and $\boldsymbol{\eta}(\boldsymbol{\Omega})$ Functions

6.1.1 Definitions of $\boldsymbol{\kappa}(\boldsymbol{\Omega})$ and $\boldsymbol{\eta}(\boldsymbol{\Omega})$

Here, we describe how the recommended $\kappa(\Omega)$ and $\eta(\Omega)$ functions defined in AJ1 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 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.[23] Because the power results

[21] 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.

[22] Note that Section 7.6 below provides additional numerical results concerning the computation of $\eta_2(p)$.

[23] The grid of 31 κ values is $\{0, .2, .4, .6, .8, 1.0, 1.1, 1.2, ...,2.9, 3.0, 3.2, ..., 3.8, 4.2\}$. The grid of 43 ρ values is $\{.99, .975, .95, .90, .85, ..., -.90, -.95, -.975, -.99\}$. The results are based on 40,000 critical value

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.[24]

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 5.3 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 elements negative. (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 sub-matrices 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 sub-matrix 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 by 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 of AJ1 and described in the second paragraph of this section.[25] We take $\eta(\Omega)$ to be the value determined by $p = 2$ and δ, i.e., $\eta_1(\delta)$

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.

[24] The asymptotic size of the QLR/tTest 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 7.5 for a discussion of how the maximum null rejection probability over $\mu \geq 0$ was calculated.

[25] 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 the ones that maximize average power. The reason is as follows. 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 $\rho = -.80$. Thus, there is a small deviation from the feature that the best κ value is monotone decreasing in ρ. When using the

in Table I of AJ1, but allow for an adjustment that depends on p, viz., $\eta_2(p)$, that is defined to guarantee that the test has correct asymptotic significance level (up to numerical error).[26] In particular, $\eta_1(\delta) \in R$ is defined to be such that

$$\inf_{h_1 \in R^2_{+,\infty}} CP(h_1, \Omega_\delta, \eta_1(\delta)) = 1 - \alpha, \tag{6.1}$$

where Ω_δ is the 2 by 2 correlation matrix with correlation δ (and $\kappa(\Omega)$ that appears in the definition of $CP(h_1, \Omega, \eta)$ in (4.11) 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

$$\inf_{h_1 \in R^p_{+,\infty}, \Omega \in \Psi} CP(h_1, \Omega, \eta_1(\delta(\Omega)) + \eta_2(p)) = 1 - \alpha, \tag{6.2}$$

where $\kappa(\Omega)$ and $\eta_1(\delta(\Omega))$ are defined as described above. The numerical calculation of $\eta_2(p)$ is described in Section 7.5 below.

6.1.2 Automatic κ Power Assessment

We now discuss numerical evaluations of how well the proposed method does in approximating the best κ, viz., $\kappa^*(\Omega)$. 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 AJ1 and the results are given by comparing the rows of Table II of AJ1 labelled AQLR/t-Test/κBest and AQLR/t-Test/κAuto. The second group consists of a fixed set of 19 Ω matrices (defined in Section 7.2 below) 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 7.2 below for details concerning their distributions.

For the second group of results, the asymptotic power results are size-corrected and

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 $\delta \in [-1.0, -.80)$.

[26] 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.

are based on (40000, 40000, 40000) 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 for μ vectors that consist of linear combinations of the μ vectors defined in Section 7.1 below, see Section 7.2 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 three 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.

Table S-I. Asymptotic Power Differences Between AQLR/t-Test/κAuto and AQLR/t-Test/κBest Tests for Nominal Level .05 Size-Corrected Tests

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

δ	.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

6.2 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, e.g., see Canay (2010).[27]

In order to handle models of this sort, AJ1 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.

AJ1 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 one or equal to one.

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

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 $\Omega_{Sg,Neg}$.) Also, for all values of $\varepsilon > 0$,

[27] In the missing data model, even the variance sub-matrix consisting of the binding moment inequalities is singular when $p = 1$. In the entry model, the variance sub-matrix 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.

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 $\mathcal{M}_p(\Omega_{Neg})$, which is defined in Section 7.1. Table S-II is based on (40000, 40000, 40000) critical-value, size-correction, and power simulation repetitions, respectively. Table S-II shows that the choice $\varepsilon = .012$ also works 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.)

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

	$p = 2$ & $\Omega_{Sg,Neg}$							
ε:	.0	.000,001	.000,01	.000,1	.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	
	$p = 4$ & $\Omega_{Sg,Neg}$							
ε:	.0	.000,1	.001	.005	.01	.012	.02	
Avg Asy Power	.3905	.9401	.9400	.9398	.9396	.9395	.9392	
	$p = 10$ & $\Omega_{Sg,Neg}$							
ε:	.0	.000,1	.001	.005	.01	.012	.02	
Avg Asy Power	.2903	.9718	.9718	.9717	.9715	.9715	.9713	

6.3 Singular Variance Matrices

In this section, we present results that are similar to those in Tables II and III of AJ1 except that they are based on singular matrices $\Omega_{Sg,Neg}$ and $\Omega_{Sg,Pos}$, rather than the nonsingular matrices Ω_{Neg}, $\Omega_{Zero,}$ and Ω_{Pos}. As noted in Section 6.2, singular and near singular matrices arise in a number of moment inequality models of practical importance.

The matrices $\Omega_{Sg,Neg}$ for $p = 2, 4, 10$ are the same matrices that are considered in Section 6.2. The matrices $\Omega_{Sg,Pos}$ for $p = 2, 4, 10$ are correlation matrices with all elements equal to one.

6.3.1 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 use the Moore-Penrose inverse of the singular matrix $\Omega_{Sg,Neg}$ or $\Omega_{Sg,Pos}$ as the weight matrix of the quadratic form. The power results are size corrected, as in Table II of AJ1. Average power is computed for the vectors μ in $\mathcal{M}_p(\Omega_{Neg})$ when $\Omega = \Omega_{Sg,Neg}$ and for the μ vectors in

$\mathcal{M}_p(\Omega_{Pos})$ when $\Omega = \Omega_{Sg,Pos}$, where $\mathcal{M}_p(\Omega_{Neg})$ and $\mathcal{M}_p(\Omega_{Pos})$ are defined in Section 7.1. The results in Table S-III for $p = 2$, 4, and 10 are based on (40000, 40000, 40000) 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 $\Omega_{Sg,Neg}$ and small for $\Omega_{Sg,Pos}$ (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 $\Omega_{Sg,Neg}$ is larger than it is for Ω_{Neg}, see Table II in AJ1.

Table S-III shows that the AQLR/t-Test/κAuto test has vastly superior asymptotic average power to that of the MP-QLR/t-Test/κAuto test for $\Omega_{Sg,Neg}$ and the same power for $\Omega_{Sg,Pos}$. 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 $\Omega_{Sg,Neg}$ and $\Omega_{Sg,Pos}$. 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.

Table S-III. Asymptotic Power Comparisons (Size-Corrected) for Singular Variance Matrices: MMM, Max, SumMax, AQLR, & MP-QLR Statistics, and PA & t-Test Critical Values with κ=Best & κ=Auto

	Crit.	Tuning	$p = 10$		$p = 4$		$p = 2$	
Stat.	Val.	Par. κ	$\Omega_{Sg,Neg}$	$\Omega_{Sg,Pos}$	$\Omega_{Sg,Neg}$	$\Omega_{Sg,Pos}$	$\Omega_{Sg,Neg}$	$\Omega_{Sg,Pos}$
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

6.3.2 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 with the singular matrices $\Omega_{Sg,Neg}$ and $\Omega_{Sg,Pos}$. The results are analogous to those given in Table III of AJ1 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 $Z^\dagger = Var^{-1/2}(m(W_i, \theta))(m(W_i, \theta) - Em(W_i, \theta))$ 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 three degrees of freedom χ_3^2. The latter distribution is centered and scaled to have mean zero and variance one. Average power is computed for the vectors μ in $\mathcal{M}_p(\Omega_{Neg})$ when $\Omega = \Omega_{Sg,Neg}$ and for the μ vectors in $\mathcal{M}_p(\Omega_{Pos})$ when $\Omega = \Omega_{Sg,Pos}$. 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 over-rejection by the asymptotic normal version of the AQLR/t-test/κAuto test with Ω_{Neg}, χ_3^2 distribution, and $p = 4, 10$, whereas substantial over-rejection is reported in Table III of AJ1 in the same scenario except with Ω_{Neg} in place of $\Omega_{Sg,Neg}$.

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 $\Omega_{Sg,Neg}$ and $\Omega_{Sg,Pos}$. The AQLR-based tests all out-perform the MP-QLR-based tests by a wide margin for $\Omega_{Sg,Neg}$ and perform essentially the same for $\Omega_{Sg,Pos}$. For example, for $p = 10$ and $\Omega_{Sg,Neg}$, 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^2_3 distributions are quite similar. This also is consistent with the MNRP results in Table S-IV, but differs from the results in Table III of AJ1.

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.

Table S-IV. Finite-Sample Maximum Null Rejection Probabilities for Singular Variance Matrices of the Nominal .05 AQLR/t-Test/κAuto and MP-QLR/t-Test/κAuto Tests Based on Normal and Bootstrap-Based Critical Values

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

Table S-V. Finite-Sample ("Size-Corrected") Average Power for Singular Variance Matrices of the Nominal .05 AQLR/t-Test/κAuto, MP-QLR/t-Test/κAuto, AQLR/PA, and MP-QLR/PA Tests Based on Normal and Bootstrap-Based Critical Values

Test		Dist	n	$P = 10$		$p = 4$		$p = 2$	
				$\Omega_{Sg,Neg}$	$\Omega_{Sg,Pos}$	$\Omega_{Sg,Neg}$	$\Omega_{Sg,Pos}$	$\Omega_{Sg,Neg}$	$\Omega_{Sg,Pos}$
AQLR	PA	N(0,1)	100	.97	.79	.92	.77	.85	.73
AQLR	Norm			.96	.78	.93	.77	.85	.72
AQLR	Boot			.97	.78	.93	.78	.86	.71
MP-QLR	PA	N(0,1)	100	.31	.79	.40	.77	.54	.73
MP-QLR	Norm			.29	.78	.39	.77	.55	.72
MP-QLR	Boot			.29	.78	.39	.78	.54	.71
AQLR	PA	χ^2_3	100	.97	.78	.92	.75	.85	.72
AQLR	Norm			.96	.78	.94	.74	.85	.66
AQLR	Boot			.97	.78	.94	.74	.86	.65
MP-QLR	PA	χ^2_3	100	.31	.78	.41	.76	.56	.72
MP-QLR	Norm			.29	.78	.40	.74	.57	.67
MP-QLR	Boot			.29	.78	.39	.74	.56	.65

6.3.3 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 $\Omega_{Sg,Neg}$ and $\Omega_{Sg,Pos}$.

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

$$T_n^{ELR}(\theta) = \max_{\lambda=(\lambda_1,\ldots,\lambda_p)':\lambda_\ell \leq 0, \forall \ell \leq p} 2\sum_{i=1}^{n}(1 + \lambda' m(W_i, \theta)), \tag{6.3}$$

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 pre-programmed 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 one and two 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 one and two and with numerical and analytic gradient/Hessian. NR was fastest with line search one and analytic gradient/Hessian, which is the method we employed to compute the results given in Table III of AJ1 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 a.s. For $p = 2$, NR with line search one 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 $\Omega_{Sg,Neg}$ and $\Omega_{Sg,Pos}$ 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 $\{W_1, ..., W_n\}$, the bootstrap sample $\{W_1^*, ..., W_n^*\}$ is n i.i.d. draws from the empirical distribution of the original sample, and the recentered bootstrap sample is $\{W_1^* - \overline{W}_n, ..., W_n^* - \overline{W}_n\}$, 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.

6.4 κ Values That Maximize Asymptotic Average Power

The κ values that maximize asymptotic average power, i.e., the best κ values, which are used in the construction of Table II of AJ1, 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 and a fixed correlation matrix Ω) of the RMS tests that appear in Table II of AJ1 and are based on the κ=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.

Table S-VI. Finite-Sample Maximum Null Rejection Probabilities (MNRP's) and ("Size-Corrected") Average Power for Singular Variance Matrices of the Nominal .05 AQLR/t-Test/κAuto Test with Normal (AQLR/Nm) and Bootstrap-Based (AQLR/Bt) Critical Values and ELR/t-Test/κAuto Test with Bootstrap-Based (ELR/Bt) Critical Values

Test	Dist	H_0/H_1	$p=2$	
			$\Omega_{Sg,Neg}$	$\Omega_{Sg,Pos}$
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

Table S-VII. κ Values That Maximize (Size-Corrected) Asymptotic Average Power: MMM, Max, SumMax, & AQLR Statistics; t-Test, $\varphi^{(2)}$, $\varphi^{(3)}$, $\varphi^{(4)}$, & MMSC Critical Values[1]

	Crit.	$p = 10$			$p = 4$			$p = 2$		
Stat.	Val.	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†	.6†	.0†	2.4$^{\diamond}$	1.0*	.2*	2.0*	1.2*	.2*
AQLR	$\varphi^{(3)}$	12.5†	2.3†	1.1†	9.0$^{\diamond}$	2.8*	1.4*	10.0*	1.4*	1.2*
AQLR	$\varphi^{(4)}$	2.7†	1.4†	.2†	2.5$^{\diamond}$	1.4*	.4*	2.2*	1.9*	.2*
AQLR	MMSC	5.3†	1.1†	.2†	5.7	1.4	.8	2.8	1.7	.6

[1] All cases not marked with a *, $^{\diamond}$, or † are based on (40000, 40000, 40000) critical-value, size-correction, and rejection-probability repetitions.

*Results are based on (5000, 5000, 5000) repetitions.

$^{\diamond}$Results are based on (2000, 2000, 2000) repetitions.

†Results are based on (1000, 1000, 1000) repetitions.

Table S-VIII. Comparisons of Asymptotic Maximum Null Rejection Probabilities: Max, SumMax, & AQLR Statistics; t-Test, $\varphi^{(2)}$, $\varphi^{(3)}$, & $\varphi^{(4)}$ Critical Values with κ=Best[1] & $\eta = 0$

	Crit.	Tuning	$p = 10$			$p = 4$			$p = 2$		
Stat.	Val.	Par. κ	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†	.102†	.066†	.064$^\diamondsuit$	.057*	.052*	.062*	.059*	.054*
AQLR	$\varphi^{(3)}$	Best	.113†	.111†	.066†	.098$^\diamondsuit$	.063*	.052*	.072*	.068*	.055*
AQLR	$\varphi^{(4)}$	Best	.088†	.089†	.066†	.066$^\diamondsuit$	.057*	.052*	.062*	.058*	.056*
AQLR	t-Test	Best	.058	.061	.051	.053	.058	.051	.053	.053	.051
AQLR	MMSC	Best	.088†	.097†	.066†	.055	.058	.051	.052	.053	.051

[1] All cases not marked with a *, $^\diamondsuit$, or † are based on (40000, 40000, 40000) critical-value, size-correction, and rejection-probability repetitions.

*Results are based on (5000, 5000, 5000) repetitions.

$^\diamondsuit$Results are based on (2000, 2000, 2000) repetitions.

†Results are based on (1000, 1000, 1000) repetitions.

6.5 Comparison of $(\mathbf{S}, \varphi)$ Functions: 19 Ω Matrices

Here we compare the MMM/t-Test/κBest, AQLR/t-Test/κBest, AQLR/t-Test/κAuto, & AQLR/MMSC/κBest tests. This section is quite similar to Section 4 of AJ1 except that 19 Ω matrices are considered here, rather than 3, and fewer tests are considered.[28] The 19 Ω matrices are the same as those considered in Table S-I in Section 6.1 and are defined in Section 7.2 below.

The qualitative results reported in AJ1 are found in Table S-IX to apply as well to the broader range of Ω matrices that are considered.

TABLE S-IX. Asymptotic Power Comparisons (Size-Corrected) for 19 Ω Matrices: MMM & AQLR Statistics; t-Test & MMSC Critical Values with κ=Best & κAuto[1]

(a) $p = 10$

Stat.	Crit. Val.	κ	$\delta(\Omega)$: -.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
			$\delta(\Omega)$: 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	

[1] κ=Best denotes the κ value that maximizes asymptotic average power. The results are based on (40000, 40000, 40000) critical-value, size-correction, and rejection-probability repetitions for $p = 2$, 4, and 10.

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

TABLE S-IX (Cont.)

(b) $p = 4$

Stat.	Crit. Val.	κ	$\delta(\Omega)$: -.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

			$\delta(\Omega)$: 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

(c) $p = 2$

Stat.	Crit. Val.	κ	$\delta(\Omega)$: -.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

			$\delta(\Omega)$: 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

6.6 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, φ) = (MMM, t-Test), (AQLR, t-Test), and (AQLR, MMSC). The GMS tests depend on a tuning parameter κ $(= \kappa_n)$ that does not depend on Ω. We consider the values κ=2.35 and κ=1.87. The former corresponds to the BIC choice $\kappa_n = (\ln n)^{1/2}$ for $n = 250$ and the latter corresponds to the 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 $\Omega = \Omega_{Neg}, \Omega_{Zero}, \Omega_{Pos}$. The critical values are 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.[29] The power results are size-corrected.

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

The AQLR/MMSC test shows substantial over-rejection whenever $p = 10$ or $\Omega = \Omega_{Neg}$ for both $\kappa = 1.87$ and 2.35. For example, the MNRP for the AQLR/MMSC/κ=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 over-rejection asymptotically under fixed κ asymptotics for some Ω matrices. However, some GMS tests with κ=2.35 perform fairly well and over-reject by a relatively small amount. The recommended RMS test performs well. It shows no sign of

[29] 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-X. Asymptotic MNRP Comparisons for Nominal .05 Tests: MMM & AQLR Statistics; t-Test & MMSC Critical Values with κ=2.35, κ=1.87, & κAuto

Stat.	Crit. Val.	Tuning Par. κ	$p=10$ Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	$p=4$ Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	$p=2$ 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†	.081†	.064†	.111	.052	.051	.057	.051	.050
AQLR	MMSC	1.87	.173†	.082†	.064†	.119	.054	.051	.075	.052	.050
AQLR	t-Test	Auto	.044	.046	.038	.047	.049	.047	.051	.051	.050

†These results are based on (1000, 1000) critical-value and rejection-probability repetitions. All other results are based on (40000, 40000) repetitions.

over rejection.

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 κ=2.35 and κ=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/κ=2.35 and AQLR/t-Test/κ=1.87 have power that is similar to that of the recommended RMS test, but lower on average. The GMS tests AQLR/MMSC/κ=2.35 and AQLR/MMSC/κ=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/κ=2.35, (ii) the recommended RMS test performs similarly to this GMS test, but has slightly higher power on average and does not over-reject under the null hypothesis, and (iii) the recommended RMS test out-performs the other GMS tests considered by a noticeable margin in terms of asymptotic MNRP and/or power.

Table S-XI. Asymptotic Power Comparisons (Size-Corrected) for Nominal .05 Tests: MMM & AQLR Statistics; PA, t-Test, & MMSC Critical Values with κ=2.35, κ=1.87, & κAuto

	Crit.	Tuning	$p = 10$			$p = 4$			$p = 2$		
Stat.	Val.	Par. κ	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	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†	.60†	.74†	.56	.68	.75	.64	.68	.70
AQLR	MMSC	1.87	.44†	.63†	.76†	.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

†These results are based on (1000, 1000, 1000) critical-value, size-correction, and rejection-probability repetitions. All other results are based on (40000, 40000, 40000) repetitions.

6.7 Additional Asymptotic MNRP & Power Results

Table S-XII reports asymptotic MNRP results for some tests that are not considered in AJ1 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.[30] (Each of these PA critical values is computed using all null mean vectors μ which consist of $0's$ and $\infty's$.) The critical values are found to be 5.07, 7.99, and 16.2 for $p = 2$, 4, and 10, respectively.

[30] For any given value of $\delta = \delta(\Omega)$, these 43 matrices are defined just as the 19 Toeplitz matrices are defined in Section 7.2. 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: MMM, Max, SumMax, & AQLR Statistics; PA, t-Test, $\varphi^{(2)}$, $\varphi^{(3)}$, $\varphi^{(4)}$, & MMSC Critical Values with κ=Best, κ=2.35, & κ=1.87; & $\eta = 0$[1]

	Crit.	Tuning	$p = 10$			$p = 4$			$p = 2$		
Stat.	Val.	Par. κ	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	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

Table S-XII. (Cont.)

Stat.	Crit. Val.	Tuning Par. κ	$p=10$ Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	$p=4$ Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	$p=2$ Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
AQLR	$\varphi^{(2)}$	Best	$.092^{\dagger}$	$.102^{\dagger}$	$.066^{\dagger}$	$.064^{\diamond}$	$.057^{*}$	$.052^{*}$	$.062^{*}$	$.059^{*}$	$.054^{*}$
AQLR	$\varphi^{(2)}$	2.35	$.090^{\dagger}$	$.081^{\dagger}$	$.065^{\dagger}$	$.058^{\diamond}$	$.057^{*}$	$.052^{*}$	$.062^{*}$	$.056^{*}$	$.053^{*}$
AQLR	$\varphi^{(2)}$	1.87	$.098^{\dagger}$	$.081^{\dagger}$	$.065^{\dagger}$	$.066^{\diamond}$	$.057^{*}$	$.052^{*}$	$.062^{*}$	$.056^{*}$	$.053^{*}$
AQLR	$\varphi^{(3)}$	Best	$.113^{\dagger}$	$.111^{\dagger}$	$.066^{\dagger}$	$.098^{\diamond}$	$.063^{*}$	$.052^{*}$	$.072^{*}$	$.068^{*}$	$.055^{*}$
AQLR	$\varphi^{(3)}$	2.35	$.245^{\dagger}$	$.111^{\dagger}$	$.065^{\dagger}$	$.153^{\diamond}$	$.065^{*}$	$.052^{*}$	$.118^{*}$	$.062^{*}$	$.054^{*}$
AQLR	$\varphi^{(3)}$	1.87	$.262^{\dagger}$	$.114^{\dagger}$	$.065^{\dagger}$	$.162^{\diamond}$	$.068^{*}$	$.052^{*}$	$.127^{*}$	$.065^{*}$	$.054^{*}$
AQLR	$\varphi^{(4)}$	Best	$.088^{\dagger}$	$.089^{\dagger}$	$.066^{\dagger}$	$.066^{\diamond}$	$.057^{*}$	$.052^{*}$	$.062^{*}$	$.058^{*}$	$.056^{*}$
AQLR	$\varphi^{(4)}$	2.35	$.092^{\dagger}$	$.081^{\dagger}$	$.065^{\dagger}$	$.062^{\diamond}$	$.057^{*}$	$.052^{*}$	$.062^{*}$	$.056^{*}$	$.053^{*}$
AQLR	$\varphi^{(4)}$	1.87	$.105^{\dagger}$	$.082^{\dagger}$	$.065^{\dagger}$	$.077^{\diamond}$	$.057^{*}$	$.052^{*}$	$.074^{*}$	$.058^{*}$	$.053^{*}$
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	$.088^{\dagger}$	$.097^{\dagger}$	$.066^{\dagger}$	.055	.058	.051	.052	.053	.051
AQLR	MMSC	2.35	$.148^{\dagger}$	$.081^{\dagger}$	$.064^{\dagger}$	.111	.052	.051	.057	.051	.050
AQLR	MMSC	1.87	$.173^{\dagger}$	$.082^{\dagger}$	$.064^{\dagger}$	.119	.054	.051	.075	.052	.050

[1]κ=Best denotes the κ value that maximizes asymptotic average power. Unless stated otherwise, results are based on (40000, 40000) critical-value and rejection-probability repetitions.

*Results are based on (5000, 5000) repetitions.

$^{\diamond}$Results are based on (2000, 2000) repetitions.

†Results are based on (1000, 1000) repetitions.

Table S-XIII. Asymptotic Power Comparisons (Size-Corrected) of Nominal .05 Tests: MMM, Max, SumMax, & AQLR Statistics; t-Test, $\varphi^{(2)}$, $\varphi^{(3)}$, $\varphi^{(4)}$, & MMSC Critical Values with κ=Best, κ=2.35, κ=1.87, & κAuto[1]

	Crit.	Tuning	$p = 10$			$p = 4$			$p = 2$		
Stat.	Val.	Par. κ	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	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

Table S-XIII. (Cont.)

	Crit.	Tuning	$p = 10$			$p = 4$			$p = 2$		
Stat.	Val.	Par. κ	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
AQLR	$\varphi^{(2)}$	Best	$.51^{\dagger}$	$.65^{\dagger}$	$.81^{\dagger}$	$.60^{\diamond}$	$.69^{*}$	$.78^{*}$	$.66^{*}$	$.69^{*}$	$.72^{*}$
AQLR	$\varphi^{(2)}$	2.35	$.50^{\dagger}$	$.58^{\dagger}$	$.77^{\dagger}$	$.60^{\diamond}$	$.65^{*}$	$.75^{*}$	$.64^{*}$	$.68^{*}$	$.70^{*}$
AQLR	$\varphi^{(2)}$	1.87	$.50^{\dagger}$	$.60^{\dagger}$	$.78^{\dagger}$	$.60^{\diamond}$	$.66^{*}$	$.76^{*}$	$.64^{*}$	$.68^{*}$	$.70^{*}$
AQLR	$\varphi^{(3)}$	Best	$.43^{\dagger}$	$.63^{\dagger}$	$.81^{\dagger}$	$.55^{\diamond}$	$.68^{*}$	$.78^{*}$	$.61^{*}$	$.69^{*}$	$.72^{*}$
AQLR	$\varphi^{(3)}$	2.35	$.36^{\dagger}$	$.63^{\dagger}$	$.80^{\dagger}$	$.52^{\diamond}$	$.68^{*}$	$.77^{*}$	$.59^{*}$	$.68^{*}$	$.72^{*}$
AQLR	$\varphi^{(3)}$	1.87	$.36^{\dagger}$	$.63^{\dagger}$	$.81^{\dagger}$	$.52^{\diamond}$	$.68^{*}$	$.77^{*}$	$.59^{*}$	$.69^{*}$	$.72^{*}$
AQLR	$\varphi^{(4)}$	Best	$.51^{\dagger}$	$.65^{\dagger}$	$.81^{\dagger}$	$.60^{\diamond}$	$.70^{*}$	$.78^{*}$	$.66^{*}$	$.69^{*}$	$.72^{*}$
AQLR	$\varphi^{(4)}$	2.35	$.51^{\dagger}$	$.60^{\dagger}$	$.78^{\dagger}$	$.60^{\diamond}$	$.66^{*}$	$.75^{*}$	$.66^{*}$	$.69^{*}$	$.70^{*}$
AQLR	$\varphi^{(4)}$	1.87	$.51^{\dagger}$	$.63^{\dagger}$	$.79^{\dagger}$	$.58^{\diamond}$	$.68^{*}$	$.76^{*}$	$.61^{*}$	$.69^{*}$	$.71^{*}$
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	$.56^{\dagger}$	$.66^{\dagger}$	$.81^{\dagger}$	.63	.69	.78	.65	.69	.73
AQLR	MMSC	2.35	$.46^{\dagger}$	$.60^{\dagger}$	$.74^{\dagger}$	.56	.68	.75	.64	.68	.70
AQLR	MMSC	1.87	$.44^{\dagger}$	$.63^{\dagger}$	$.76^{\dagger}$	.54	.69	.76	.59	.69	.71
Power	Envelope	-	.85	.85	.85	.80	.80	.80	.75	.75	.75

[1] κ=Best denotes the κ value that is best in terms of asymptotic average power. Unless stated otherwise, results are based on (40000, 40000, 40000) critical-value, size-correction, and rejection-probability repetitions.

*Results are based on (5000, 5000, 5000) repetitions.

$^{\diamond}$Results are based on (2000, 2000, 2000) repetitions.

†Results are based on (1000, 1000, 1000) repetitions.

6.8 Comparative Computation Times

As reported in the paper, to compute the recommended bootstrap RMS test, i.e., AQLR/t-Test/κAuto/Boot, using 10,000 critical-value simulation repetitions takes 1.3, 1.7, 3.2, 8.4, 17.2, and 52.0 seconds when $p = 2$, 4, 10, 20, 30, and 50, respectively, and $n = 250$ using a PC with a 3.4 GHz processor. For the asymptotic normal version of the recommended bootstrap RMS test, i.e., AQLR/t-Test/κAuto/Norm, the times are .25, .31, .71, 2.4, 6.1, and 21.8 seconds, respectively.

In contrast, to compute the bootstrap version of the MMM/t-Test/κ=2.35 test using 10,000 critical-value simulation repetitions takes .86, .98, 2.0, 5.9, 11.6, and 28.4 seconds when $p = 2$, 4, 10, 20, 30, and 50, respectively, and $n = 250$. For the asymptotic normal version of the MMM/t-Test/κ=2.35 test, the times are .008, .010, .029, .060, .090, and .18 seconds, respectively. Note that the computation times are not affected by whether κ is taken to be κAuto or κ=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 20 to 85 times) than asymptotic normal version of the AQLR-based test. (This is because the 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), 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/κ=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.

6.9 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, the Table 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 AJ1) 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 AJ1 and above, see AJ1 for their definitions. Table S-XIV considers h_1 vectors that consist of $0's$ and $\infty's$. (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 and the moment selection procedure detects this with probability one 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 one, 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 "# of Zeros in h_1" and through the sub-matrix 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 sub-matrix of Ω of dimension equal to the "# of Zeros in h_1" is the same (provided $p \geq$"# of Zeros in h_1"). In consequence, the results of Table S-XIV hold for any value of p. For example, if $p = 20$, $\Omega = \Omega_{Neg}$, and the "# 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 $p = 15$, $\Omega = \Omega_{Neg}$, and the "# 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 not negligible. For example, for $p = 10$, $\Omega = \Omega_{Zero} = I_{10}$, 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.

Table S-XIV. Mean and Standard Deviation of the Asymptotic Distribution of the Data-Dependent RMS Critical Values Excluding the Size-Correction Factor $\widehat{\eta}$[1]

	Ω_{Neg}		Ω_{Zero}		Ω_{Pos}	
# of	Mean	SD	Mean	SD	Mean.	SD
Zero's in h_1	$c_n(\theta, \widehat{\kappa})$	$c_n(\theta, \widehat{\kappa})$	$c_n(\theta, \widehat{\kappa})$	$c_n(\theta, \widehat{\kappa})$	$c_n(\theta, \widehat{\kappa})$	$c_n(\theta, \widehat{\kappa})$
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

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

7 Details Concerning the Numerical Results

This section contains the following: (i) the definition of the μ vectors used in AJ1 (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 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 of AJ1, and (v) a description of the numerical computation of $\eta_2(p)$, which is part of the recommended size-correction function $\eta(\cdot)$.

7.1 μ Vectors

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

$$\begin{aligned}
\mathcal{M}_2(I_2) &= \{(-2.309, 0), (-2.309, 1), (-2.309, 2), (-2.309, 3), \\
&\quad\;(-2.309, 4), (-2.309, 7), (-1.6263, -1.6263)\}, \\
\mathcal{M}_2(\Omega_{Neg}) &= \{(-1.001, 0), (-1.804, 1), (-2.303, 2), (-2.309, 3), \\
&\quad\;(-2.309, 4), (-2.309, 7), (-0.5165, -0.5165)\}, \\
\mathcal{M}_2(\Omega_{Pos}) &= \mathcal{M}_2(I_2) \text{ except the last vector is } (-2.0040, -2.0040).
\end{aligned} \tag{7.1}$$

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

$$\begin{aligned}
&\mathcal{M}_4(\Omega) \\
&= \{(-\mu_1, -\mu_1, 1, 1), (-\mu_2, -\mu_2, 2, 2), (-\mu_3, -\mu_3, 3, 3), (-\mu_4, -\mu_4, 4, 4), (-\mu_5, -\mu_5, 7, 7), \\
&\quad (-\mu_6, -\mu_6, 1, 7), (-\mu_7, -\mu_7, 2, 7), (-\mu_8, -\mu_8, 3, 7), (-\mu_9, -\mu_9, 4, 7), \\
&\quad (-\mu_{10}, 1, 1, 1), (-\mu_{11}, 2, 2, 2), (-\mu_{12}, 3, 3, 3), (-\mu_{13}, 4, 4, 4), (-\mu_{14}, 7, 7, 7), \\
&\quad (-\mu_{15}, 1, 1, 7), (-\mu_{16}, 2, 2, 7), (-\mu_{17}, 3, 3, 7), (-\mu_{18}, 4, 4, 7), (-\mu_{19}, -\mu_{19}, 0, 0), \\
&\quad (-\mu_{20}, 0, 0, 0), (-\mu_{21}, 25, 25, 25), (-\mu_{22}, -\mu_{22}, 25, 25), (-\mu_{23}, -\mu_{23}, -\mu_{23}, 25), \\
&\quad (-\mu_{24}, -\mu_{24}, -\mu_{24}, -\mu_{24})\},
\end{aligned} \tag{7.2}$$

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

For $p = 4$, the μ vectors in $\mathcal{M}_4(\Omega_{Neg})$ are defined by (7.2) and the following: $\mu_1 = 0.5505$, $\mu_j = 0.5526$ for $j = 2, ..., 5$, $\mu_6 = 0.5505$, $\mu_j = 0.5526$ for $j = 7, 8, 9$, $\mu_{10} = 1.8814$, $\mu_{11} = 2.4283$, $\mu_j = 2.4705$ for $j = 12, 13, 14, 17, 18, 21$, $\mu_{15} = 1.8814$, $\mu_{16} = 2.4283$, $\mu_{19} = 0.3176$, $\mu_{20} = 0.8624$, $\mu_{22} = 0.5526$, $\mu_{23} = 0.2607$, $\mu_{24} = 0.1756$.

For $p = 4$, the μ vectors in $\mathcal{M}_4(\Omega_{Pos})$ are defined by (7.2) and the following: $\mu_j = 2.4047$ for $j = 1, ..., 9, 19, 22$; $\mu_j = 2.4705$ for $j = 10, ..., 18, 20, 21$; $\mu_{23} = 2.2628$; 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:

$$
\begin{aligned}
&\mathcal{M}_{10}(\Omega) \\
&= \{(-\mu_1, -\mu_1, 1, ..., 1), (-\mu_2, -\mu_2, 2, ..., 2), (-\mu_3, -\mu_3, 3, ..., 3), (-\mu_4, -\mu_4, 4, ..., 4), \\
&\quad (-\mu_5, -\mu_5, 7, ..., 7), (-\mu_6, -\mu_6, 1, 1, 1, 7, ..., 7), (-\mu_7, -\mu_7, 2, 2, 2, 7, ..., 7), \\
&\quad (-\mu_8, -\mu_8, 3, 3, 3, 7, ..., 7), (-\mu_9, -\mu_9, 4, 4, 4, 7, ..., 7), (-\mu_{10}, -\mu_{10}, -\mu_{10}, -\mu_{10}, 1, ..., 1), \\
&\quad (-\mu_{11}, -\mu_{11}, -\mu_{11}, -\mu_{11}, 2, ..., 2), (-\mu_{12}, -\mu_{12}, -\mu_{12}, -\mu_{12}, 3, ..., 3), \\
&\quad (-\mu_{13}, -\mu_{13}, -\mu_{13}, -\mu_{13}, 4, ..., 4), (-\mu_{14}, -\mu_{14}, -\mu_{14}, -\mu_{14}, 7, ..., 7), \\
&\quad (-\mu_{15}, -\mu_{15}, -\mu_{15}, -\mu_{15}, 1, 1, 1, 7, 7, 7), (-\mu_{16}, -\mu_{16}, -\mu_{16}, -\mu_{16}, 2, 2, 2, 7, 7, 7), \\
&\quad (-\mu_{17}, -\mu_{17}, -\mu_{17}, -\mu_{17}, 3, 3, 3, 7, 7, 7), (-\mu_{18}, -\mu_{18}, -\mu_{18}, -\mu_{18}, 4, 4, 4, 7, 7, 7), \\
&\quad (-\mu_{19}, 1, ..., 1), (-\mu_{20}, 2, ..., 2), (-\mu_{21}, 3, ..., 3), (-\mu_{22}, 4, ..., 4), (-\mu_{23}, 7, ..., 7), \\
&\quad (-\mu_{24}, 1, 1, 1, 7, ..., 7), (-\mu_{25}, 2, 2, 2, 7, ..., 7), (-\mu_{26}, 3, 3, 3, 7, ..., 7), (-\mu_{27}, 4, 4, 4, 7, ..., 7), \\
&\quad (-\mu_{28}, -\mu_{28}, 0, ..., 0), (-\mu_{29}, -\mu_{29}, -\mu_{29}, -\mu_{29}, 0, ..., 0), (-\mu_{30}, 0, ..., 0), \\
&\quad (-\mu_{31}, 25, ..., 25), (-\mu_{32}, -\mu_{32}, 25, ..., 25), (-\mu_{33}, -\mu_{33}, -\mu_{33}, 25, ..., 25), \\
&\quad (-\mu_{34}, -\mu_{34}, -\mu_{34}, -\mu_{34}, 25, ..., 25), (-\mu_{35}, -\mu_{35}, -\mu_{35}, -\mu_{35}, -\mu_{35}, 25, ..., 25), \\
&\quad (-\mu_{36}, ..., -\mu_{36}, 25, 25, 25, 25), (-\mu_{37}, ..., -\mu_{37}, 25, 25, 25), (-\mu_{38}, ..., -\mu_{38}, 25, 25), \\
&\quad (-\mu_{39}, ..., -\mu_{39}, 25), (-\mu_{40}, ..., -\mu_{40})\}. \qquad (7.3)
\end{aligned}
$$

For $p = 10$, the μ vectors in $\mathcal{M}_{10}(I_{10})$ are defined by (7.3) and the following: $\mu_j = 1.8927$ for $j = 1, ..., 9, 28, 32$ $\mu_j = 1.3360$ for $j = 10, ..., 18, 29, 34$, $\mu_j = 2.6817$ for $j = 19, ..., 27, 30, 31$, $\mu_{33} = 1.5463$, $\mu_{35} = 1.1963$, $\mu_{36} = 1.0893$, $\mu_{37} = 1.0099$, $\mu_{38} = 0.9465$, $\mu_{39} = 0.8882$, and $\mu_{40} = 0.8440$.

For $p = 10$, the μ vectors in $\mathcal{M}_{10}(\Omega_{Neg})$ are defined by (7.3) and the following: $\mu_j = 0.6016$ for $j = 1, ..., 9$, $\mu_j = 0.3475$ for $j = 10, ..., 18$, $\mu_{19} = 1.9847$, $\mu_{20} = 2.5835$,

$\mu_j = 2.6817$ 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 $\mathcal{M}_{10}(\Omega_{Pos})$ are defined by (7.3) and the following: $\mu_j = 2.6227$ for $j = 1, ..., 9$, $\mu_j = 2.4676$ for $j = 10, ..., 18$, $\mu_j = 2.6817$ for $j = 19, ..., 27$, $\mu_{28} = 2.6227$, $\mu_{29} = 2.4676$, $\mu_{30} = 2.6817$, $\mu_{31} = 2.6817$, $\mu_{32} = 2.6227$, $\mu_{33} = 2.5401$, $\mu_{34} = 2.4676$, $\mu_{35} = 2.4005$, $\mu_{36} = 2.3140$, $\mu_{37} = 2.2846$, $\mu_{38} = 2.2565$, $\mu_{39} = 2.2343$, and $\mu_{40} = 2.2066$.

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

7.2 Automatic κ Power Assessment Details

The 19 matrices Ω that are considered in Table S-I in Section 6.1.2 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 $\delta \geq 0$, $\rho = (\delta, ..., \delta)$. 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 AJ1 (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 $\Omega = Diag^{-1/2}(BB')BB' \times Diag^{-1/2}(BB')$, where B is a p by 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 $\mathcal{M}_p(\Omega_{Neg})$, $\mathcal{M}_p(\Omega_{Zero})$, and $\mathcal{M}_p(\Omega_{Pos})$. Specifically, for a given matrix Ω, $\mathcal{M}_p(\Omega)$ is defined by: (i) $\mathcal{M}_p(\Omega) = \mathcal{M}_p(\Omega_{Neg})$ if $\delta(\Omega) \in [-1.0, -.90]$, (ii) if $\delta(\Omega) \in [-.9, 0]$, $\mathcal{M}_p(\Omega) = \{\mu : \mu = (1 + \delta/.9)\mu_{Zero,j} - (\delta/.9)\mu_{Neg,j}$ for $j = 1, ..., J_p\}$, where $\mu_{Zero,j}$ denotes the jth element of $\mathcal{M}_p(\Omega_{Zero})$ and analogously for $\mathcal{M}_p(\Omega_{Neg})$ and $\mathcal{M}_p(\Omega_{Pos})$ and J_p denotes the numbers of elements in $\mathcal{M}_p(\Omega_{Zero})$, (iii) if $\delta(\Omega) \in [0, .5]$, $\mathcal{M}_p(\Omega) = \{\mu : \mu = (1 - \delta/.5)\mu_{Zero,j} + (\delta/.5)\mu_{Pos,j}$ for $j = 1, ..., J_p\}$, and (iv) if $\delta(\Omega) \in [0.5, 1.0]$,

$\mathcal{M}_p(\Omega) = \mathcal{M}_p(\Omega_{Pos})$.

7.3 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 is H_0 is as in (5.3). 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 $N(\mu_{Alt}, \Sigma)$. We take the specified asymptotic null distribution to be $N(\mu_{Null}, \Sigma)$, where μ_{Null} is defined to minimize $(\mu - \mu_{Alt})'\Sigma^{-1}(\mu - \mu_{Alt})$ 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.

7.4 Computation of 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 of AJ1. For all of the RMS tests in Table II of AJ1, 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 $\{.0, .2, ..., 3.6, 3.8, 4.2\}$ 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}: $\{4.9, 5.1, ..., 6.5\}$, and for $p = 10$ and Ω_{Neg}: $\{4.1, 4.4, ..., 6.5\}$. For the AQLR/$\varphi^{(3)}$ test, the following alternative grids are used in special cases: for $p = 2$ and Ω_{Neg}: $\{5.0, 5.5, ..., 10.5\}$, for $p = 4$ and Ω_{Neg}: $\{3.5, 4.0, ..., 10.5\}$, and for $p = 10$ and Ω_{Neg}: $\{11.5, 12.0, ..., 14.0\}$.

7.5 Numerical Computation of $\eta_2(\mathbf{p})$

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

$$\begin{aligned}&\left\{S_{2A}\left(\Omega^{1/2}Z_r + (h_1, 0_v), \Omega\right) - q_{S_{2A}}\left(\varphi^{(1)}\left(\kappa^{-1}(\Omega)[\Omega^{1/2}Z_r + (h_1, 0_v)], \Omega\right), \Omega\right)\right.\\ &\left.+\eta_1(\delta(\Omega)) : r = 1, ..., R\right\},\end{aligned} \tag{7.4}$$

where $Z_r \sim$ i.i.d. $N(0_k, I_k)$ for $r = 1, ..., R$, 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

$$CP(h_1, \Omega, \eta_1(\delta(\Omega)) + \eta_{h_1,\Omega}) = 1 - \alpha. \tag{7.5}$$

The same simulated random variables $\{Z_r : r = 1, ..., R\}$ are used for all (h_1, Ω) considered. The critical value $q_{S_{2A}}\left(\varphi^{(1)}\left(\kappa^{-1}(\Omega)[\Omega^{1/2}Z_r + (h_1, 0_v)], \Omega\right), \Omega\right)$ in (7.4) 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 $\infty's$. By considering a variety of subcases, we find that size is (essentially) attained for $\mu \in \mathcal{E}_1$, see Section 7.6 below.[31] 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

[31] 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.

set $\Psi^* \subset \Psi$. Given this, $\eta_2(p) \in R$ is defined to be

$$\sup_{h_1 \in \mathcal{E}_1, \Omega \in \Psi^*} \eta_{h_1,\Omega}. \tag{7.6}$$

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 $.99$,[32] and (ii) 500 randomly generated matrices Ω that are generated by $\Omega = Corr(V)$, 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 over $\eta_{h_1,\Omega}$ over Ψ. Since the same underlying random variables $\{Z_r : r = 1, ..., R\}$ are used for each (h_1, Ω) considered, an empirical process 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

$$\inf_{h_1 \in \mathcal{E}_1, \Omega \in \Psi} CP(h_1, \Omega, \eta_1(\delta(\Omega)) + \eta_2(p)) = 1 - \alpha. \tag{7.7}$$

7.6 Maximization Over $\boldsymbol{\mu}$ Vectors in the Null Hypothesis

7.6.1 Computation of $\boldsymbol{\eta}_2(\mathbf{p})$

Next, 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 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 $\mathcal{E}_1$.[33] These results are for the 43 fixed Toeplitz variance matrices that are described in Section 7.5. The results are given in Table S-XVII.

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

[32] For any given value of $\delta = \delta(\Omega)$, these 43 matrices are defined just as the 19 Toeplitz matrices are defined in Section 7.2. 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.

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

randomly generated μ vectors (with the same distribution as in the previous paragraph) plus $\mathcal{E}_1$.[34] These results are given in Table S-XVIII.

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 computation of $\eta_2(p)$ for $p = 3, ..., 10$.[35] 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 $p = 3, ..., 10$.[36] For these two variance matrices and $p = 3, ..., 10$, 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-XIX.

Fourth, in Table S-XX, 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 $\boldsymbol{\mu}$ 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.

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

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

[35] 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 $\infty's$ and $\eta_2(p)$ is set equal to 0. This matrix is found to be I_p for 7 of the 8 values of p and within .0001 of being LF for the other case. So, for simplicity, we take $\Omega_{LF_1} = I_p$ for $p = 3, ..., 10$.

[36] 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 $\infty's$ and $\eta_2(p)$ is set equal to 0.

$= \{0, 1, 2.5, 4\}$; and for $p = 8$, $\#grid = 3$, and GridVec $= \{0, 2.5, 3.5\}$.

(2) 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 non-decreasing, and (iii) contain at least one zero. For example, if $p = 4$ and GridVec $= \{0, 1, 2, 3, 4\}$, then $\#grid = 5$ and the μ vectors are of the form $(0, 0, 0, 0), ..., (0, 0, 2, 3)$, $(0, 0, 2, 4)$, $(0, 0, 3, 4), ..., (0, 4, 4, 4)$. 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 $= \{0, 2, 4, 6\}$.

(3) 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 non-decreasing, 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 $(0, 0, 0, 0), (0, 0, 0, 1), ..., (0, 0, 3, 3)$, $(0, 0, 4, 4), (0, 1, 1, 1)...$, $(0, 4, 4, 4)$. The number of such vectors is $(p - 1) * (\#grid - 1) + 1$.

The GridVec vector used with the small partial grid is: $\forall p = 2, ..., 10$, $\#grid =$ 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-XVII, S-XVIII, and S-XIX provide the results. Table S-XVII 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. Table S-XVIII 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-XIX 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.[37]

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 $R^p_{+,\infty}$.

Potential Effects of Inaccuracy in $\boldsymbol{\eta_2}(\mathbf{p})$ Next, we report the potential effects of inaccuracy in the calculation of $\eta_2(p)$. Table S-XX 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 (40000, 40000) critical value and null rejection probability repetitions.

Table S-XX shows that even relatively large percentage changes in $\eta_2(p)$ have fairly small effects on the MNRP's.

7.6.2 Computation of MNRP's for Tests Based on Best Kappa Values

Table II of AJ1 reports asymptotic power comparisons for tests using (infeasible) critical values that employ the asymptotically best κ values (κ=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-XVII and S-XVIII, but for tests other than the AQLR/t-Test/κAuto test. We compute results for a subset of the cases considered in Tables S-XVII and S-XVIII.[38] (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

[37]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.

[38]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.

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-XVII and S-XVIII 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 (40000, 40000, 40000) for the critical values, size-correction, and rejection probabilities, respectively, in all cases considered. For the slow tests, (10000, 10000, 10000) 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 the AQLR/MMSC/κBest test, (40000, 40000, 40000) repetitions are used for $p = 2$ and 4 and (10000, 10000, 10000) 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/$\varphi^{(j)}/\kappa$Best 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 κ=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.

Table S-XVII. Differences in Nominal .05 Asymptotic MNRP's Due to Different Sets of Mean Vectors μ Used in the Computations with 43 Toeplitz Variance Matrices: $\mathcal{E}_1$ Versus a Full Grid, a Large Partial Grid, a Small Partial Grid, and 1000 Random μ Vectors Plus $\mathcal{E}_1$

p	(a) $\mathcal{E}_1$ Versus Full Grid & $\mathcal{E}_1$ Max Diff Over 43 Var Matrices	(b) $\mathcal{E}_1$ Versus Large Partial Grid & $\mathcal{E}_1$ Max Diff Over 43 Var Matrices	(c) $\mathcal{E}_1$ Versus Small Partial Grid & $\mathcal{E}_1$ Max Diff Over 43 Var matrices	(d) $\mathcal{E}_1$ Versus 1000 Random μ & $\mathcal{E}_1$ Max Diff Over 43 Var matrices
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

Table S-XVIII. Differences in Nominal .05 Asymptotic MNRP's Due to Different Sets of Mean Vectors μ Used in the Computations: $\mathcal{E}_1$ Versus 1000 Random μ Vectors Plus $\mathcal{E}_1$ with 260 Random Variance Matrices

p	$\mathcal{E}_1$ Versus 1000 Random μ Vectors Plus $\mathcal{E}_1$ — Max Diff Over 260 Random Variance Matrices
3	.0000
4	.0000
5	.0000
6	.0000
7	.0000
8	.0025
9	.0026
10	.0024

Table S-XIX. 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 $\mathcal{E}_1$ with 2 Variance Matrices

	$\mathcal{E}_1$ Versus 100,000 Random μ Vectors & $\mathcal{E}_1$	
p	Difference for $\Omega = \Omega_{LF_1}$	Difference for $\Omega = \Omega_{LF_2}$
3	.0000	.0000
4	.0000	.0000
5	.0000	.0000
6	.0000	.0000
7	.0000	.0000
8	.0000	.0000
9	.0000	.0000
10	.0000	.0000

Table S-XX. 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

8 Computer Programs

This section lists the GAUSS computer programs that were used to carry out the numerical results reported in AJ1 and above.

- 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. They include: (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 of AJ1, (iv) asymptotic 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-XX.

- 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 Section 6.1.2.

- rmsprg_fs_short_final: This program was not used to compute any of the results reported in AJ1 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 $\infty's$ and (ii) these μ vectors plus randomly generated μ vectors, as reported in Table S-XVIII and Section 7.6.2.

- sizediffprg22_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 $\infty's$ 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-XVII, the first column of results in Table S-XIX, and Section 7.6.2.

- 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 $p = 3, ..., 10$. These results are reported in the last column of Table S-XIX.

9 Alternative Parametrization and Proofs

This section provides proofs of the results given in Section 5. In addition, the first subsection gives an alternative parametrization of the moment inequality/equality model to that given in (2.1). 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.

9.1 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 above.

As stated above, the true value θ_0 $(\in \Theta \subset R^d)$ is assumed to satisfy the moment conditions in (2.1). 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 $m_j(W_i, \theta) = m_j(W_i, \theta, \tau_0)$, $m(W_i, \theta) = (m_1(W_i, \theta, \tau_0), ..., m_k(W_i, \theta, \tau_0))'$, $\overline{m}_{n,j}(\theta) = n^{-1} \sum_{i=1}^n m_j(W_i, \theta, \widehat{\tau}_n(\theta))$, and $\overline{m}_n(\theta) = (\overline{m}_{n,1}(\theta), ..., \overline{m}_{n,k}(\theta))'$. (Hence, in this case, $\overline{m}_n(\theta) \neq n^{-1} \sum_{i=1}^n m(W_i, \theta)$.)

We define $\gamma_1 = (\gamma_{1,1}, ..., \gamma_{1,p})' \in R_+^p$ by writing the moment inequalities in (2.1) as moment equalities:

$$\sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0 \text{ for } j = 1, ..., p, \tag{9.1}$$

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 $\Omega = \Omega(\theta, F) = AsyCorr_F(n^{1/2}\overline{m}_n(\theta))$ 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, $\sigma_{F,j}^2(\theta) = \lim_{n\to\infty} Var_F(n^{1/2}\overline{m}_{n,j}(\theta))$ and $\Omega(\theta, F) = \lim_{n\to\infty} Corr_F$ $(n^{1/2}\overline{m}_n(\theta))$, where $Var_F(n^{1/2}\overline{m}_{n,j}(\theta))$ and $Corr_F(n^{1/2}\overline{m}_n(\theta))$ 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 $\gamma_2 = (\gamma_{2,1}, \gamma_{2,2}) = (\theta, vech_*(\Omega(\theta, F))) \in R^q$, where $vech_*(\Omega)$ 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 (2.2), γ_1 satisfies (9.1), $\gamma_2 = (\theta, vech_*(\Omega(\theta, F)))$, 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:

$$
\begin{aligned}
&\text{(i) } \theta \in \Theta, \\
&\text{(ii) } \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0 \text{ for } j = 1, ..., p, \\
&\text{(iii) } E_F m_j(W_i, \theta) = 0 \text{ for } j = p+1, ..., k, \\
&\text{(iv) } \sigma_{F,j}^2(\theta) = AsyVar_F\left(n^{1/2}\overline{m}_{n,j}(\theta)\right) \text{ exists and lies in } (0, \infty) \text{ for } j = 1, ..., k, \\
&\text{(v) } AsyCorr_F\left(n^{1/2}\overline{m}_n(\theta)\right) \text{ exists and equals } \Omega_{\gamma_{2,2}} \in \Psi, \text{ and} \\
&\text{(vi) } \{W_i : i \geq 1\} \text{ are stationary under } F,
\end{aligned} \tag{9.2}
$$

where $\gamma_1 = (\gamma_{1,1}, ..., \gamma_{1,p})'$ and $\Omega_{\gamma_{2,2}}$ is the $k \times k$ correlation matrix determined by $\gamma_{2,2}$.[39] Furthermore, Γ must be restricted by enough additional conditions such that under any sequence $\{\gamma_{n,h} = (\gamma_{n,h,1}, (\theta_{n,h}, vech_*\ (\Omega_{n,h})),\ F_{n,h}) : n \geq 1\}$ of parameters in Γ that satisfies $n^{1/2}\gamma_{n,h,1} \to h_1$ and $(\theta_{n,h}, vech_*(\Omega_{n,h})) \to h_2 = (h_{2,1}, h_{2,2})$ for some $h = (h_1, h_2) \in R^p_{+,\infty} \times R^q_{[\pm\infty]}$, we have

$$
\begin{aligned}
&\text{(vii) } A_n = (A_{n,1}, ..., A_{n,k})' \to_d Z_{h_{2,2}} \sim N(0_k, \Omega_{h_{2,2}}) \text{ as } n \to \infty, \text{ where} \\
&A_{n,j} = n^{1/2}\left(\overline{m}_{n,j}(\theta_{n,h}) - E_{F_{n,h}} m_j(W_i, \theta_{n,h})\right)/\sigma_{F_{n,h},j}(\theta_{n,h}), \\
&\text{(viii) } \widehat{\sigma}_{n,j}(\theta_{n,h})/\sigma_{F_{n,h},j}(\theta_{n,h}) \to_p 1 \text{ as } n \to \infty \text{ for } j = 1, ..., k, \\
&\text{(ix) } \widehat{D}_n^{-1/2}(\theta_{n,h})\widehat{\Sigma}_n(\theta_{n,h})\widehat{D}_n^{-1/2}(\theta_{n,h}) \to_p \Omega_{h_{2,2}} \text{ as } n \to \infty, \text{ and} \\
&\text{(x) conditions (vii)-(ix) hold for all subsequences } \{w_n\} \text{ in place of } \{n\},
\end{aligned} \tag{9.3}
$$

where $\Omega_{h_{2,2}}$ is the $k \times k$ correlation matrix for which $vech_*(\Omega_{h_{2,2}}) = h_{2,2}$, $\widehat{\sigma}^2_{n,j}(\theta) = [\widehat{\Sigma}_n(\theta)]_{jj}$ for $1 \leq j \leq k$ and $\widehat{D}_n(\theta) = Diag\{\widehat{\sigma}^2_{n,1}(\theta), ..., \widehat{\sigma}^2_{n,k}(\theta)\}$ $(= Diag(\widehat{\Sigma}_n(\theta)))$.[40,41]

For example, for i.i.d. observations, conditions (i)-(vi) in (2.2) imply conditions (i)-(vi) in (9.2). Furthermore, conditions (i)-(vi) in (2.2) plus the definition of $\widehat{\Sigma}_n(\theta)$ in

[39] In Andrews and Guggenberger (2009), a strong mixing condition is imposed in condition (vi) of (9.2). This condition is used to verify Assumption E0 in that paper and is not needed with RMS critical values.

[40] When a preliminary estimator $\widehat{\tau}_n(\theta)$ appears, $A_{n,j}$ can be written equivalently as $n^{1/2}\left(n^{-1}\sum_{i=1}^n m_j(W_i, \theta_{n,h}, \widehat{\tau}_n(\theta_{n,h})) - E_{F_{n,h}} m_j(W_i, \theta_{n,h}, \tau_0)\right)/\sigma_{F_{n,h},j}(\theta_{n,h})$, 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 $n^{1/2}(\overline{m}_{n,j}(\theta_{n,h}) - E_{F_{n,h}}\overline{m}_{n,j}(\theta_{n,h}))/\sigma_{F_{n,h},j}(\theta_{n,h})$.

[41] Condition (x) of (9.3) requires that conditions (vii)-(ix) must hold under any sequence of parameters $\{\gamma_{w_n,h} : n \geq 1\}$ that satisfies the conditions preceding (9.3) with n replaced by w_n.

(3.2) and the additional condition (vii) in (2.2) imply conditions (vii)-(x) in (9.3). 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 (9.2) that ensure that Γ is such that any sequences $\{\gamma_{w_n,h} : n \geq 1\}$ in Γ satisfy (9.3). For brevity, we do not do so here.

We now specify the set Δ, defined in (4.13), in the parametrization introduced above. Define

$$\begin{aligned} H = \{ & h \in R^p_{[\pm\infty]} \times R^q_{[\pm\infty]} : \exists \text{ a subsequence } \{w_n\} \text{ of } \{n\} \text{ and a sequence} \\ & \{\gamma_{w_n,h} \in \Gamma : n \geq 1\} \text{ for which } w_n^{1/2}\gamma_{w_n,h,1} \to h_1 \text{ and } \gamma_{w_n,h,2} \to h_2\}. \end{aligned} \tag{9.4}$$

Then, Δ can be written equivalently as

$$\begin{aligned} \Delta = \{ & (h_1, \Omega_{h_{2,2}}) \in R^p_{+,\infty} \times cl(\Psi) : h = (h_1, h_{2,1}, h_{2,2}) \in H \\ & \text{for some } h_{2,1} \in cl(\Theta), \text{ where } h_{2,2} = vech_*(\Omega_{h_{2,2}})\}. \end{aligned} \tag{9.5}$$

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

9.2 Proofs

The proof of Theorem 1 above uses the following Lemmas. Let

$$CP_n(\gamma) = P_\gamma(T_n(\theta) \leq c_n(\theta)). \tag{9.6}$$

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

Lemma 4 *Suppose Assumptions* S, φ, κ, *and* $\eta1$ *hold. Let* $\{\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3}) : n \geq 1\}$ *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* $h_{2,2}$. *Then,*

(a) $CP_n(\gamma_{n,h}) \to [CP(h_1, \Omega_{h_{2,2}}, \eta(\Omega_{h_{2,2}})-), CP(h_1, \Omega_{h_{2,2}}, \eta(\Omega_{h_{2,2}}))]$ *and*

(b) *for any subsequence* $\{w_n : n \geq 1\}$ *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 $\{\gamma_n^* = (\gamma_{n,1}^*, \gamma_{n,2}^*, \gamma_{n,3}^*) \in \Gamma : n \geq 1\}$ be a sequence such that $\liminf_{n\to\infty} CP_n(\gamma_n^*) = \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} CP_n(\gamma)$ $(= AsyCS)$. Such a sequence always exists. Let $\{u_n : n \geq 1\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} CP_{u_n}(\gamma_{u_n}^*)$ exists and equals $\liminf_{n\to\infty} CP_n(\gamma_n^*) = AsyCS$. Such a subsequence always exists.

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

$$w_n^{1/2}\gamma_{w_n,1,j}^* \to h_{1,j}^* \text{ for some } h_{1,j}^* \in R_+. \tag{9.7}$$

If (2) holds, then for some subsequence $\{w_n\}$ of $\{u_n\}$,

$$w_n^{1/2}\gamma_{w_n,1,j}^* \to h_{1,j}^*, \text{ where } h_{1,j}^* = \infty. \tag{9.8}$$

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

$$\gamma_{w_n,2}^* \to h_2^* \text{ for some } h_2^* \in \mathrm{cl}(\Gamma_2). \tag{9.9}$$

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 $j = 1, ..., p$ either (9.7) or (9.8) applies and (9.9) 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 (9.4)), and (iv) $\lim_{n\to\infty} CP_{w_n}(\gamma_{w_n}^*) = AsyCS$. Hence, by Lemma 4(b),

$$\begin{aligned} AsyCS = \lim_{n\to\infty} CP_{w_n}(\gamma_{w_n}^*) &\geq CP(h_1^*, \Omega_{h_{2,2}^*}, \eta(\Omega_{h_{2,2}^*})-) \\ &\geq \inf_{(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta(\Omega)-), \end{aligned} \tag{9.10}$$

where the second inequality holds because $(h_1^*, \Omega_{h_{2,2}^*}) \in \Delta$ by the definition of Δ in (9.5).

Next, by the definition of Δ in (9.5), 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,

$$\begin{aligned} AsyCS =& \liminf_{n\to\infty} \inf_{(\theta,F)\in\mathcal{F}} P_F(T_n(\theta) \leq c_n(\theta)) \\ &\leq \liminf_{n\to\infty} CP_{t_n}(\gamma_{t_n,h}) \\ &\leq CP(h_1, \Omega_{h_{2,2}}, \eta(\Omega_{h_{2,2}})), \end{aligned} \tag{9.11}$$

where the second inequality holds by Lemma 4(b). Since (9.11) holds for all $(h_1, \Omega_{h_{2,2}}) \in \Delta$, we have

$$AsyCS \leq \inf_{(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta(\Omega)). \tag{9.12}$$

Combining (9.10) and (9.12) 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 $\eta 3$. $\square$

Proof of Lemma 4. For notational simplicity, let Ω_0 denote $\Omega_{h_{2,2}}$. To establish part (a), we show below that

$$\begin{pmatrix} T_n(\theta_{n,h}) \\ c_n(\theta_{n,h}) \end{pmatrix} \to_d \begin{pmatrix} S\left(Z + (h_1, 0_v), \Omega_0\right) \\ q_S\left(\varphi\left(\kappa^{-1}(\Omega_0)[Z + (h_1, 0_v)], \Omega_0\right), \Omega_0\right) + \eta(\Omega_0) \end{pmatrix} \text{ as } n \to \infty \tag{9.13}$$

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

$$\begin{aligned} & P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_n(\theta_{n,h}) + x) \\ \to\ & P\left(S\left(Z + (h_1, 0_v), \Omega_0\right) \leq q_S\left(\varphi\left(\kappa^{-1}(\Omega_0)[Z + (h_1, 0_v)], \Omega_0\right), \Omega_0\right) + \eta(\Omega_0) + x\right) \\ =\ & CP(h_1, \Omega_0, \eta(\Omega_0) + x). \end{aligned} \tag{9.14}$$

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

$$\begin{aligned}
&\limsup_{n\to\infty} P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_n(\theta_{n,h})) \\
&\leq \lim_{x\downarrow 0} \limsup_{n\to\infty} P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_n(\theta_{n,h}) + x) \\
&= \lim_{x\downarrow 0} CP(h_1, \Omega_0, \eta(\Omega_0) + x) \\
&= CP(h_1, \Omega_0, \eta(\Omega_0)), \qquad (9.15)
\end{aligned}$$

where the first equality holds by (9.14) and the second equality holds because $CP(h_1, \Omega_0, \eta(\Omega_0) + x)$ is a df and hence is right-continuous. Analogously,

$$\begin{aligned}
\liminf_{n\to\infty} P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_n(\theta_{n,h})) &\geq \lim_{x\downarrow 0} CP(h_1, \Omega_0, \eta(\Omega_0) - x) \\
&= CP(h_1, \Omega_0, \eta(\Omega_0)-), \qquad (9.16)
\end{aligned}$$

where the equality holds by definition. Equations (9.15) and (9.16) combine to establish part (a).

Next, we prove (9.13). Using Assumption S(a), we have

$$T_n(\theta) = S\left(\widehat{D}_n^{-1/2}(\theta) n^{1/2} \overline{m}_n(\theta),\ \widehat{D}_n^{-1/2}(\theta) \widehat{\Sigma}_n(\theta) \widehat{D}_n^{-1/2}(\theta)\right). \qquad (9.17)$$

For i.i.d. or dependent observations with or without preliminary estimators of identified parameters, (9.3) 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 (9.3) holds for i.i.d. observations). By (9.3), 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 $\gamma_{n,h,1} = (\gamma_{n,h,1,1}, ..., \gamma_{n,h,1,p})'$ and by definition $\gamma_{n,h,1,j} = 0$ for $j = p+1, ..., k$. If $h_{1,j} = \infty$ and $j \leq p$, where $h_1 = (h_{1,1}, ..., h_{1,p})'$, then $A_{n,j} + n^{1/2}\gamma_{n,h,1,j} \to_p \infty$ under $\{\gamma_{n,h} : n \geq 1\}$ by condition (vii) of (9.3) and the definition of $\{\gamma_{n,h} : n \geq 1\}$. 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 (9.17) or of the RMS critical value, which is defined by

$$c_n(\theta) = q_S\left(\varphi\left(\xi_n(\theta), \widehat{\Omega}_n(\theta)\right), \widehat{\Omega}_n(\theta)\right) + \eta(\widehat{\Omega}_n(\theta)). \qquad (9.18)$$

To circumvent these problems, we consider k-vector-valued functions of $\widehat{D}_n^{-1/2}(\theta_{n,h}) \times n^{1/2}\overline{m}_n(\theta_{n,h})$ 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 (9.17) and (9.18) as continuous functions of these k-vectors and apply the continuous mapping theorem. Let $G(\cdot)$ be a strictly increasing continuous df on R, such as the standard normal df.

For $j \leq k$, we have

$$\begin{aligned} G_{\kappa,n,j} &= G(\xi_{n,j}(\theta_{n,h})) = G\left(\kappa^{-1}(\widehat{\Omega}_n(\theta_{n,h}))\widehat{\sigma}_{n,j}^{-1}(\theta_{n,h})n^{1/2}\overline{m}_{n,j}(\theta_{n,h})\right) \\ &= G\left(\kappa^{-1}(\widehat{\Omega}_n(\theta_{n,h}))\widehat{\sigma}_{n,j}^{-1}(\theta_{n,h})\sigma_{F_{n,h},j}(\theta_{n,h})\left[A_{n,j} + n^{1/2}\gamma_{n,h,1,j}\right]\right), \end{aligned} \tag{9.19}$$

where $A_{n,j}$ is defined in (9.3) and by definition $\gamma_{n,h,1,j} = 0$ for $j = p+1, ..., k$.

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

$$G_{\kappa,n,j} \to_d G\left(\kappa^{-1}(\Omega_0)[Z_j + h_{1,j}]\right) \tag{9.20}$$

using (9.19), conditions (vii) and (viii) of (9.3) (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 (9.3) (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

$$G_{\kappa,n,j} \to_p 1 \tag{9.21}$$

using (9.19), $A_{n,j} = O_p(1)$, $\kappa^{-1}(\widehat{\Omega}_n(\theta_{n,h})) \to_p \kappa^{-1}(\Omega_0) > 0$, and $G(x) \to 1$ as $x \to \infty$. The results in (9.20)-(9.21) hold jointly and combine to give

$$\begin{aligned} G_{\kappa,n} &= (G_{\kappa,n,1}, ..., G_{\kappa,n,k})' \to_d G_{\kappa,\infty}, \text{ where} \\ G_{\kappa,\infty} &= (G(\kappa^{-1}(\Omega_0)[Z_1 + h_{1,1}]), ..., G(\kappa^{-1}(\Omega_0)[Z_k + h_{1,k}]))' \end{aligned} \tag{9.22}$$

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

$$\widetilde{q}_{S,\varphi}(z, \Omega) = q_S\left(\varphi(G_{(k)}^{-1}(z), \Omega), \Omega\right) \tag{9.23}$$

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

$$\begin{aligned} \mathcal{Z}((h_1,0_v),\Omega_0) &= \left\{ z \in (0,1]^p \times (0,1)^v : G_{(k)}^{-1}(z) \in \Xi((h_1,0_v),\Omega) \right\} \text{ and} \\ P\left(G_{\kappa,\infty} \in \mathcal{Z}((h_1,0_v),\Omega_0)\right) &= P\left(\kappa^{-1}(\Omega_0)[Z+(h_1,0_v)] \in \Xi((h_1,0_v),\Omega_0)\right) \\ &= 1, \end{aligned} \tag{9.24}$$

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

We now have

$$\begin{aligned} c_n(\theta_{n,h}) &= q_S\left(\varphi(\xi_n(\theta_{n,h}), \widehat{\Omega}_n(\theta_{n,h})), \widehat{\Omega}_n(\theta_{n,h})\right) + \eta(\widehat{\Omega}_n(\theta_{n,h})) \\ &= q_S\left(\varphi(G_{(k)}^{-1}(G_{\kappa,n}), \widehat{\Omega}_n(\theta_{n,h})), \widehat{\Omega}_n(\theta_{n,h})\right) + \eta(\widehat{\Omega}_n(\theta_{n,h})) \\ &= \widetilde{q}_{S,\varphi}\left(G_{\kappa,n}, \widehat{\Omega}_n(\theta_{n,h})\right) + \eta(\widehat{\Omega}_n(\theta_{n,h})) \\ &\to_d \widetilde{q}_{S,\varphi}\left(G_{\kappa,\infty}, \Omega_0\right) + \eta(\Omega_0) \\ &= q_S\left(\varphi(G_{(k)}^{-1}(G_{\kappa,\infty}), \Omega_0), \Omega_0\right) + \eta(\Omega_0) \\ &= q_S\left(\varphi(\kappa^{-1}(\Omega_0)[Z+(h_1,0_v)], \Omega_0), \Omega_0\right) + \eta(\Omega_0), \end{aligned} \tag{9.25}$$

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 $G_{(k)}^{-1}(\cdot)$, the third and fourth equalities hold by the definition of $\widetilde{q}_{S,\varphi}(\cdot,\cdot)$, the convergence holds by (9.22), condition (ix) of (9.3), Assumption $\eta 1$, and the continuous mapping theorem using (9.24), the last equality holds by the definitions of $G_{\kappa,\infty}$ and $G_{(k)}^{-1}(\cdot)$ 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 (9.19)-(9.25) to show that

$$T_n(\theta_{n,h}) \to_d S(Z + (h_1, 0_v), \Omega_0). \tag{9.26}$$

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 $\widetilde{q}_{S,\varphi}(z,\Omega) = q_S(\varphi(G_{(k)}^{-1}(z),\Omega),\Omega)$ is replaced by $\widetilde{S}(z,\Omega) = S(G_{(k)}^{-1}(z),\Omega)$, and (iv) the continuity argument in the paragraph containing (9.24) 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 (9.25) and (9.26) is joint because the two results can be obtained by a single application of the continuous mapping theorem. Hence, the verification of (9.13) is complete and part (a) is proved.

Next, we prove part (b). By the same argument as above but using condition (x) of (9.3) in place of conditions (vii)-(ix), the results of (9.25) and 9.26 hold with $\{w_n\}$ in place of $\{n\}$ for any subsequence $\{w_n\}$. Hence, (9.13) and (9.14) hold with the same changes, which implies that part (b) holds. $\square$

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 df 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 $\varepsilon_U > 0$,

$$P\left(S(\Omega_0^{1/2} Z^* + \beta_0, \Omega_0) \leq q_S(\beta_0, \Omega_0) + \varepsilon\right) = 1 - \alpha + \varepsilon_U. \tag{9.27}$$

The df 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

$$\begin{aligned}
&\text{(a) } S(\Omega^{1/2} Z^* + \beta, \Omega) \to S(\Omega_0^{1/2} Z^* + \beta_0, \Omega_0) \text{ a.s.,}\\
&\text{(b) } 1\left(S(\Omega^{1/2} Z^* + \beta, \Omega) \leq x\right) \to 1\left(S(\Omega_0^{1/2} Z^* + \beta_0, \Omega_0) \leq x\right) \text{ a.s.}\\
&\quad \text{except if } S(\Omega_0^{1/2} Z^* + \beta_0, \Omega_0) = x,\\
&\text{(c) } P\left(S(\Omega_0^{1/2} Z^* + \beta_0, \Omega_0) = x\right) = 0, \text{ and}\\
&\text{(d) the indicator function is bounded,}
\end{aligned} \tag{9.28}$$

where (a) holds by Assumption S(c), (b) holds by (a), and (c) holds because the df 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

$$\begin{aligned}
&\big|P\left(S(\Omega^{1/2} Z^* + \beta, \Omega) \leq q_S(\beta_0, \Omega_0) + \varepsilon\right)\\
&- P\left(S(\Omega_0^{1/2} Z^* + \beta_0, \Omega_0) \leq q_S(\beta_0, \Omega_0) + \varepsilon\right)\big| < \varepsilon_U/2.
\end{aligned} \tag{9.29}$$

Equations (9.27) and (9.29) imply that

$$P\left(S(\Omega^{1/2}Z^* + \beta, \Omega) \le q_S(\beta_0, \Omega_0) + \varepsilon\right) \ge 1 - \alpha + \varepsilon_U/2. \tag{9.30}$$

The definition of a quantile and (9.30) imply that

$$q_S(\beta, \Omega) \le q_S(\beta_0, \Omega_0) + \varepsilon. \tag{9.31}$$

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 df 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

$$P(S(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0) \le \varepsilon) = 1 - \alpha + \varepsilon_1. \tag{9.32}$$

Because the df of $S(\Omega^{1/2}Z^* + \beta, \Omega)$ at $\varepsilon > 0$ is continuous in (β, Ω) by (9.28), for all (β, Ω) sufficiently close to (β_0, Ω_0), we have

$$\left| P\left(S(\Omega^{1/2}Z^* + \beta, \Omega) \le \varepsilon\right) - P\left(S(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0) \le \varepsilon\right)\right| < \varepsilon_1/2. \tag{9.33}$$

Equations (9.32) and (9.33) imply

$$P\left(S(\Omega^{1/2}Z^* + \beta, \Omega) \le \varepsilon\right) \ge 1 - \alpha. \tag{9.34}$$

This and the definition of a quantile imply that $q_S(\beta, \Omega) \le \varepsilon$. Since $q_S(\beta, \Omega) \ge 0$ 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 Assumptions S(b) and S(d). This and the continuity in (β, Ω) at (β_0, Ω_0) of the df of $S(\Omega^{1/2}Z^* + \beta, \Omega)$ at $x > 0$, which holds by (9.28), give: for all $x > 0$,

$$\lim_{(\beta,\Omega)\to(\beta_0,\Omega_0)} P\left(S(\Omega^{1/2}Z^* + \beta, \Omega) \le x\right) = P\left(S(\Omega_0^{1/2}Z^* + \beta_0, \Omega_0) \le x\right) = 1. \tag{9.35}$$

Equation (9.35) implies that given any $x > 0$ for all (β, Ω) sufficiently close to (β_0, Ω_0), the df 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. $\square$

Proof of Lemma 2. Assumption LA3(a) holds by the Liapounov triangular array CLT for row-wise i.i.d. random variables with mean zero and variance one using Assumptions LA1(a), LA1(c), and LA3* and the Cramér-Wold device. Assumptions LA3(b) and LA3(c) hold by standard arguments using a weak law of large numbers for row-wise i.i.d. random variables with variance one using Assumptions LA1(a), LA1(c), and LA3*. Note that Assumption LA3 does not follow from (9.3) because in Assumption LA3 the functions are evaluated at θ_0, which is not the true value (unless $\lambda = 0$). $\square$

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 (9.13) 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

$$\begin{aligned} D^{-1/2}(\theta_0, F_n)E_{F_n}m(W_i, \theta_0) &= D^{-1/2}(\theta_n, F_n)E_{F_n}m(W_i, \theta_n) \\ &\quad + \Pi(\theta_n^*, F_n)(\theta_0 - \theta_n), \\ n^{1/2}D^{-1/2}(\theta_0, F_n)E_{F_n}m(W_i, \theta_0) &\to (h_1, 0_v) + \Pi_0\lambda, \end{aligned} \tag{9.36}$$

where $D(\theta, F) = Diag\{\sigma^2_{F,1}(\theta), ..., \sigma^2_{F,k}(\theta)\}$, 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 (9.17), 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 df on R, such as the standard normal df. Using (9.36), 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 $j = 1, ..., k$, we have

$$\begin{aligned}
G^0_{\kappa,n,j} &= G\left(\kappa^{-1}(\widehat{\Omega}_n(\theta_0))\widehat{\sigma}^{-1}_{n,j}(\theta_0)n^{1/2}\overline{m}_{n,j}(\theta_0)\right) \\
&= G\left(\kappa^{-1}(\widehat{\Omega}_n(\theta_0))\widehat{\sigma}^{-1}_{n,j}(\theta_0)\sigma_{F_n,j}(\theta_0)\left[A^0_{n,j} + n^{1/2}\sigma^{-1}_{F_n,j}(\theta_0)E_{F_n}m_j(W_i,\theta_0)\right]\right), \\
G^0_{\kappa,n,j} &\to_p 1 \text{ if } j \leq p \text{ and } h_{1,j} = \infty, \qquad (9.37)\\
G^0_{\kappa,n,j} &\to_d G\left(\kappa^{-1}(\Omega(\theta_0))[Z_j + h_{1,j} + \Pi'_{0,j}\lambda]\right) \text{ if } j \leq p \text{ and } h_{1,j} < \infty, \\
G^0_{\kappa,n,j} &\to_d G\left(\kappa^{-1}(\Omega(\theta_0))[Z_j + \Pi'_{0,j}\lambda]\right) \text{ if } j = p+1, ..., k, \\
G^0_{\kappa,n} &= (G^0_{\kappa,n,1}, ..., G^0_{\kappa,n,k}) \to_d G^0_{\kappa,\infty} = \\
&\quad (G(\kappa^{-1}(\Omega(\theta_0))[Z_1 + h_{1,1} + \Pi'_{0,1}\lambda]), ..., G(\kappa^{-1}(\Omega(\theta_0))[Z_k + \Pi'_{0,k}\lambda]))',
\end{aligned}$$

where $Z = (Z_1, ..., Z_k)'$ and $Z_j + h_{1,j} + \Pi'_{0,j}\lambda = \infty$ by definition if $h_{1,j} = \infty$. Now, the same argument as in (9.23)-(9.25) of the proof of Lemma 4(a) gives

$$c_n(\theta_0) \to_d q_S\left(\varphi(\kappa^{-1}(\Omega_0)[Z + (h_1, 0_v) + \Pi_0\lambda], \Omega_0), \Omega_0\right) + \eta(\Omega_0). \qquad (9.38)$$

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 (9.26) in the proof of Lemma 4(a), we obtain

$$T_n(\theta_0) \to_d S([Z + (h_1, 0_v) + \Pi_0\lambda], \Omega_0). \qquad (9.39)$$

Furthermore, the convergence in (9.38) and (9.39) is joint, which establishes that (9.13) 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 (9.14)-(9.16) in the proof of Lemma 4(a) with $(h_1, 0_v)$ replaced by $(h_1, 0_v) + \Pi_0\lambda$ and $CP(h_1, \Omega_0, \eta(\Omega_0))$ replaced by $AsyPow(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0))$. □

References

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

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.

Andrews, D. W. K. and P. Jia (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," unpublished manuscript, Cowles Foundation, Yale University.

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.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica,* 78, 119-157.

Bugni, F. (2010): Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica,* 78, 735-753.

Canay, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics,* 156, 408-425.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica,* 75, 1243-1284.

Fan, Y. and S. Park (2007): "Confidence Sets for Some Partially Identified Parameters," unpublished manuscript, Department of Economics, Vanderbilt University.

Imbens, G. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica,* 72, 1845-1857.

Müller, U. and M. Watson (2008): "Low-Frequency Robust Cointegration Testing," *Econometrica,* 76, 979-1016.

Pratt, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association,* 56, 541-567.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics,* 146, 107-117.

Silvapulle, M. J. and P. K. Sen (2005): *Constrained Statistical Inference.* New York: Wiley.