INFERENCE FOR PARAMETERS DEFINED
BY MOMENT INEQUALITIES:
A RECOMMENDED MOMENT SELECTION PROCEDURE

By

Donald W. K. Andrews and Panle Jia Barwick

COWLES FOUNDATION PAPER NO. 1372

INFERENCE FOR PARAMETERS DEFINED BY MOMENT INEQUALITIES: A RECOMMENDED MOMENT SELECTION PROCEDURE

By Donald W. K. Andrews and Panle Jia Barwick[1]

This paper is concerned with tests and confidence intervals for parameters that are not necessarily point 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, 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.

1. INTRODUCTION

THIS PAPER CONSIDERS INFERENCE in (unconditional) moment inequality models with parameters that need not be point 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 undersized 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.[2]

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.[3] 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. For more than ten moment inequalities, the required constant, $\eta_2(p)$ (defined below), can be simulated as described in Section S7.5 of the Supplemental Material. However, the accuracy of this method diminishes as p increases.

We show that with independent and identically distributed (i.i.d.) observations, bootstrap critical values outperform 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-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 (unconditional) moment inequality tests, determine a recommended test, and provide data-dependent tuning parameters.

Other related papers in the literature (referenced below) include Hansen (2005), Fan and Park (2007), Rosen (2008), Chiburis (2009), Bugni (2010), Canay (2010), and Bugni, Canay, and Guggenberger (2012).

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.

The Supplemental Material (Andrews and Barwick (2012)) provides material that includes (i) the asymptotic results that are utilized in this paper, (ii) de-

---

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

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

tails concerning the numerical results given here, and (iii) additional numerical results.

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

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

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

where $\{m_j(\cdot, \theta) : j = 1, \dots, p\}$ 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, that is, 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 Section S9.1 of the Supplemental Material. In addition, the asymptotic results in Section S5 of the Supplemental Material 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_0 is

$$(2.2) \qquad \mathrm{CS}_n = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta)\},$$

where $c_n(\theta)$ is a data-dependent critical value.[4]

We now describe the recommended test statistic and critical value. The justifications for these recommendations are described below and are given in detail in Sections S4 and S5 of the Supplemental Material. The recommended test statistic is AQLR statistic, $T_{\mathrm{AQLR},n}(\theta)$, that is a function of the sample moment conditions, $n^{1/2}\overline{m}_n(\theta)$, and an estimator of their asymptotic variance, $\widehat{\Sigma}_n(\theta)$:

$$(2.3) \qquad T_{\mathrm{AQLR},n}(\theta) = S_{2A}\big(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)\big)$$

$$= \inf_{t \in R^p_{+,\infty}} \big(n^{1/2}\overline{m}_n(\theta) - t\big)' \widetilde{\Sigma}_n^{-1}(\theta)\big(n^{1/2}\overline{m}_n(\theta) - t\big), \quad \text{where}$$

$$\overline{m}_n(\theta) = \big(\overline{m}_{n,1}(\theta), \dots, \overline{m}_{n,p}(\theta)\big)',$$

---

[4]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 $<$.

$$\overline{m}_{n,j}(\theta) = n^{-1} \sum_{i=1}^{n} m_j(W_i, \theta) \quad \text{for} \quad j \le p,$$

$$\widetilde{\Sigma}_n(\theta) = \widehat{\Sigma}_n(\theta) + \max\{\varepsilon - \det(\widehat{\Omega}_n(\theta)), 0\}\widehat{D}_n(\theta), \quad \varepsilon = .012,$$

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

and $\text{Diag}(\Sigma)$ denotes the diagonal matrix based on the matrix Σ.[5] Note that the weight matrix $\widetilde{\Sigma}_n(\theta)$ depends only on $\widehat{\Sigma}_n(\theta)$ and hence $T_{\text{AQLR},n}(\theta)$ 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 quasi-likelihood ratio (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 equivariant to scale changes in the moment functions. The matrix $\widehat{\Omega}_n(\theta)$ is the correlation matrix that corresponds to $\widehat{\Sigma}_n(\theta)$.

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

$$(2.4) \qquad \widehat{\Sigma}_n(\theta) = n^{-1} \sum_{i=1}^{n} \big(m(W_i, \theta) - \overline{m}_n(\theta)\big)\big(m(W_i, \theta) - \overline{m}_n(\theta)\big)', \quad \text{where}$$

$$m(W_i, \theta) = \big(m_1(W_i, \theta), \dots, m_p(W_i, \theta)\big)'.$$

With temporally dependent observations or when a preliminary estimator of a parameter τ appears, a different definition of $\widehat{\Sigma}_n(\theta)$ often is required; see Section S9.1 of the Supplemental Material. For example, with dependent observations, a heteroskedasticity and autocorrelation consistent (HAC) estimator may be required.

The test statistic $T_{\text{AQLR},n}(\theta)$ 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.[6]

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

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

---

[5]The constant $\varepsilon = .012$ was determined numerically based on an average asymptotic power criterion. See Section S6.2 of the Supplemental Material for details.

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

where $c_n(\theta, \widehat{\kappa})$ is the $1 - \alpha$ quantile of a bootstrap (or "asymptotic normal") distribution of a moment selection version of $T_{\mathrm{AQLR},n}(\theta)$ and $\widehat{\eta}$ is a data-dependent size-correction factor. For i.i.d. data, we recommend using a non-parametric 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 $r = 1, \ldots, R$ 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.[7] The p-vectors of recentered bootstrap sample moments and $p \times p$ bootstrap weight matrices for $r = 1, \ldots, R$ are defined by

$$(2.6) \quad m^*_{n,r}(\theta) = n^{1/2}\big(\overline{m}^*_{n,r}(\theta) - \overline{m}_n(\theta)\big),$$
$$\widetilde{\Sigma}^*_{n,r}(\theta) = \widehat{\Sigma}^*_{n,r}(\theta) + \max\big\{\varepsilon - \det\big(\widehat{\Omega}^*_{n,r}(\theta)\big), 0\big\}\widehat{D}^*_{n,r}(\theta), \quad \text{where}$$
$$\varepsilon = .012,$$
$$\widehat{D}^*_{n,r}(\theta) = \mathrm{Diag}\big(\widehat{\Sigma}^*_{n,r}(\theta)\big),$$
$$\widehat{\Omega}^*_{n,r}(\theta) = \widehat{D}^*_{n,r}(\theta)^{-1/2}\widehat{\Sigma}^*_{n,r}(\theta)\widehat{D}^*_{n,r}(\theta)^{-1/2}.$$

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 $j = 1, \ldots, p$. The jth moment inequality is selected if

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

where $\widehat{\sigma}^2_{n,j}(\theta)$ is the (j, j) element of $\widehat{\Sigma}_n(\theta)$ for $j = 1, \ldots, p$ 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.

---

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

For $r = 1, \ldots, R$, let $m^*_{n,r}(\theta, \widehat{p})$ denote the $\widehat{p}$-subvector of $m^*_{n,r}(\theta)$ that includes the $\widehat{p}$ selected moment inequalities, where $\widehat{p}$ may depend on θ.[8,9] Analogously, let $\widehat{\Sigma}^*_{n,r}(\theta, \widehat{p})$ denote the $\widehat{p} \times \widehat{p}$ submatrix 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

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

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 $\widehat{p} \times \widehat{p}$ submatrix of $\widehat{\Sigma}_n(\theta)$ that consists of the $\widehat{p}$ selected moment inequalities, $Z^*_r \sim$ i.i.d. $N(0_{\widehat{p}}, I_{\widehat{p}})$ for $r = 1, \ldots, R$, and $\{Z^*_r : r = 1, \ldots, R\}$ are independent of $\{W_i : i \le n\}$ 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

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

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

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

This data-dependent method of selecting κ and η is referred to as the κauto method.

Table I provides values of $\kappa(\delta)$, $\eta_1(\delta)$, and $\eta_2(p)$ for $\delta \in [-1, 1]$ and $p \in \{2, 3, \ldots, 10\}$ for tests with level $\alpha = .05$ and CS's with level $1 - \alpha = .95$. Section S6.9 of the Supplemental Material 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

---

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

[9]By definition, $\widehat{p} \ge 1$, that is, 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).

TABLE I

MOMENT SELECTION TUNING PARAMETERS $\kappa(\delta)$ AND SIZE-CORRECTION FACTORS $\eta_1(\delta)$ AND $\eta_2(p)$ FOR $\alpha = .05$[a]

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

[a]The values in this table are obtained by simulating asymptotic formulae using 40,000 critical-value and 40,000 rejection-probability simulation repetitions; see Section S7.5 of the Supplemental Material for details.

size-correction factor $\widehat{\eta}$ typically is small compared to $c_n(\theta, \widehat{\kappa})$, but not negligible.[10]

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 provide $\eta_2(p)$ values only for $p \leq 10$. For the correlation matrices, we consider both a fixed grid and randomly generated matrices; see Sections S7.5 and S7.6 of the Supplemental Material for details. For the null mean vectors $\mu \in R_{+,\infty}^p$, computation of the $\eta_2(p)$ values is carried out initially for mean vectors that consist only of 0's and ∞'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) 1000 or 100,000 randomly generated μ vectors (depending on the variance matrix). The differences are found to be .0000 in most cases and small ($\leq$.0026) in all cases; see Section

---

[10]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.

S7.6 of the Supplemental Material for details. These results indicate, although do not establish unequivocally, that the maxima over $\mu \in R_{+,\infty}^p$ are obtained at μ vectors that consist only of 0's and ∞'s.

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 (2.3) and (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 recentered bootstrap sample moments, covariance matrices, and weight matrices $\{(m_{n,r}^*(\theta, \widehat{p}), \widehat{\Sigma}_{n,r}^*(\theta, \widehat{p}), \widetilde{\Sigma}_{n,r}^*(\theta, \widehat{p})) : r = 1, \ldots, R\}$, defined in (2.6) with the nonselected moment inequalities omitted, (vii) $c_n(\theta, \widehat{\kappa})$, which is the .95 sample quantile of $\{S_{2A}(m_{n,r}^*(\theta, \widehat{p}), \widehat{\Sigma}_{n,r}^*(\theta, \widehat{p})) : r = 1, \ldots, R\}$ (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 subvector and submatrix of $\widehat{\Sigma}_n^{1/2}(\theta, \widehat{p}) Z_r^*$ 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 10,000 bootstrap repetitions takes .34, .39, and .86 seconds when $p = 2$, 4, and 10, respectively, and $n = 250$ using GAUSS on a PC with a 3.2-GHz processor. For the asymptotic normal version, the times are .08, .09, and .16 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/$\kappa = 2.35$ 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.[11]

3. TEST STATISTICS AND CRITICAL VALUES

We now describe the justification for the recommended RMS test. Details are given in Sections S4 and S5 of the Supplemental Material. The test statistics

---

[11]The asymptotic normal version of the MMM/t-test/$\kappa = 2.35$ 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 be used for this purpose. The computation times for the asymptotic normal version of the MMM/t-test/$\kappa = 2.35$ test using 10,000 critical-value simulations are .003, .004, and .009 seconds when $p = 2$, 4, and 10, respectively.

$T_n(\theta)$ that we consider are of the form

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

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:

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

where $[x]_- = \min\{x, 0\}$, $m = (m_1, \ldots, m_p)'$, σ_j^2 is the jth diagonal element of Σ, $[m_{(j)}/\sigma_{(j)}]_-^2$ denotes the jth largest value among $\{[m_\ell/\sigma_\ell]_-^2 : \ell = 1, \ldots, p\}$, and $p_1 < p$ is some specified integer.[12–14] The MMM statistic S_1 has been used by Pakes, Porter, Ho, and Ishii (2004), Hansen (2005), 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.[15]

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

$$(3.3) \qquad \varphi_j^{(1)}(\xi, \Omega) = \begin{cases} 0, & \text{if } \xi_j \leq 1, \\ \infty, & \text{if } \xi_j > 1, \end{cases}$$

$$\varphi_j^{(2)}(\xi, \Omega) = \big[\kappa(\xi_j - 1)\big]_+, \quad \varphi_j^{(3)}(\xi, \Omega) = [\xi_j]_+,$$

$$\varphi_j^{(4)}(\xi, \Omega) = \kappa \xi_j 1(\xi_j > 1), \quad \text{and} \quad \varphi_j^{(0)}(\xi, \Omega) = 0$$

for $j \leq p$, where $[x]_+ = \max\{x, 0\}$, $\xi = (\xi_1, \ldots, \xi_p)'$, Ω is a $p \times p$ correlation matrix, and κ in $\varphi_j^{(2)}$ and $\varphi_j^{(4)}$ is the tuning parameter κ. Let $\varphi(\xi, \Omega) = (\varphi_1(\xi, \Omega), \ldots, \varphi_p(\xi, \Omega))'$ (for any $\varphi_j(\xi, \Omega)$ as in (3.3)). CHT, AS, and Bugni (2010) consider the function $\varphi^{(1)}$; Hansen (2005) and Canay (2010) consider

---

[12]When constructing a CS, a natural choice for p_1 is the dimension d of θ; see the discussion following (3.5) below.

[13]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.

[14]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.

[15]Personal communication.

$\varphi^{(2)}$; AS consider $\varphi^{(3)}$; and Fan and Park (2007) use a non-scale-invariant version of $\varphi^{(4)}$.[16]

The function $\varphi^{(1)}$ generates the recommended "moment selection t-test" procedure of (2.7); see Section S4.2 of the Supplemental Material for details. The function $\varphi^{(0)}$ generates a critical value based on the least favorable distribution (for which all moment inequalities are binding) evaluated at an estimator of the true variance matrix Σ. It depends only 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 $\kappa(\Omega)$ 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. Under asymptotics for which $\kappa \to \infty$ as $n \to \infty$, RMS critical values require size correction via a (data-dependent) quantity $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$, where $\eta(\Omega)$ is an R_+-valued function.

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 is smaller.[17] Third, for the recommended (S, φ) functions,

[16]Hansen's (2005) test is developed in the context of testing for a model's superior predictive ability. The structure of the latter testing problem is analogous to the testing problem considered here. In consequence, the recommended test in this paper also can be used to test for a model's superior predictive ability and it provides a data-dependent tuning parameter for that testing problem (that is chosen to maximize an average power criterion).

[17]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.

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 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 Sections S5 and S9 of the Supplemental Material. 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

$$(3.4) \qquad \text{AsyPow}(\mu, \Omega, S, \varphi, \kappa, \eta)$$

$$= P\Big[S(\Omega^{1/2} Z^* + \mu, \Omega) > q_S\big(\varphi\big(\kappa^{-1}[\Omega^{1/2} Z^* + \mu], \Omega\big), \Omega\big) + \eta \Big],$$

where $\mu \in (R \cup \{+\infty\})^p$ 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 Section S5.2 of the Supplemental Material to be $\text{AsyPow}(\mu, \Omega(\theta), S, \varphi, \kappa(\Omega(\theta)), \eta(\Omega(\theta)))$, 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

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

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 nonbinding 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 Section S7.4 of the Supplemental Material 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 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 Section S5 of the Supplemental Material, 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 descriptions: 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 $\rho = (.9, .8, .7, .6, .5, \ldots, .5)$ 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 Section S7.1 of the Supplemental Material. 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 Section S6.3 of the Supplemental Material, 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 functions $(S, \varphi) = (\text{MMM}, \text{PA}), (\text{MMM}, t\text{-test}), (\text{Max}, \text{PA}), (\text{Max}, t\text{-test}), (\text{SumMax}, \text{PA}), (\text{SumMax}, t\text{-test}), (\text{AQLR}, \text{PA}), (\text{AQLR}, t\text{-test}), (\text{AQLR}, \varphi^{(3)}), (\text{AQLR}, \varphi^{(4)}),$ and $(\text{AQLR}, \text{MMSC}).$[18] We also consider the "pure ELR" test, for which Canay (2010) established 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).[19] 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 $\kappa = \text{best}$. The best κ values are determined numerically using grid search; see Section S7.4 of the Supplemental Material 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, for the AQLR/t-test test, the best κ values for $(\Omega_{\text{Neg}}, \Omega_{\text{Zero}}, \Omega_{\text{Pos}})$ 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.[20,21] The critical values, size-correction factors, and power results are each calculated using 40,000 simula-

---

[18]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.

[19]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 Section S6.3.3 of the Supplemental Material 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.

[20]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.

[21]The maximum null rejection probability calculations used in the size correction were calculated using μ vectors that consist of 0's and ∞'s. Then additional calculations were carried out to determine whether the maximum over $\mu \in R_{+,\infty}^p$ is attained at such a μ vector in each case. No evidence was found to suggest otherwise. See Sections S7.5 and S7.6 of the Supplemental Material for details.

TABLE II

ASYMPTOTIC AVERAGE POWER COMPARISONS (SIZE-CORRECTED)[a]

Statistic	Critical Value	Tuning Param. κ	$p = 10$			$p = 4$			$p = 2$		
			Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
MMM	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[b]	.65[b]	.81[b]	.60[c]	.69[d]	.78[d]	.66[d]	.69[d]	.72[d]
AQLR	$\varphi^{(3)}$	Best	.43[b]	.63[b]	.81[b]	.55[c]	.68[d]	.78[d]	.61[d]	.69[d]	.72[d]
AQLR	$\varphi^{(4)}$	Best	.51[b]	.65[b]	.81[b]	.60[c]	.70[d]	.78[d]	.66[d]	.69[d]	.72[d]
AQLR	MMSC	Best	.56[b]	.66[b]	.81[b]	.63	.69	.78	.65	.69	.73
Power	Envelope	—	.85	.85	.85	.80	.80	.80	.75	.75	.75

[a] κ = Best denotes the κ value that maximizes asymptotic average power. κ = Auto denotes the data-dependent method of choosing κ described in (2.9)–(2.10). Unless otherwise noted, all cases are based on (40,000, 40,000, 40,000) critical-value, size-correction, and power repetitions, respectively.

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

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

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

tion 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 almost 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} when $p = 10$. This 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 on 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 $\kappa = $ best) are AQLR/t-test and AQLR/MMSC, followed closely by AQLR/$\varphi^{(2)}$ and AQLR/$\varphi^{(4)}$. Each of these tests outperforms 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 data-dependent κ values. See Sections S6.4 and S6.7 of the Supplemental Material for additional numerical results for these tests.

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 "unidirectional" envelope; see Section S7.3 of the Supplemental Material for details. One does not expect a test that is designed to perform well for multidirectional alternatives to be on, or close to, the unidirectional envelope. In fact, it is surprising how close the AQLR/t-test/κbest test is to the power envelope when $\Omega = \Omega_{\text{Pos}}$. As expected, the larger is p, the greater is the difference between the power of a test designed for p-directional alternatives and the unidirectional power envelope.

When the sample correlation matrix is singular, the quasi-likelihood ratio (QLR) test statistic can be defined using the Moore–Penrose (MP) generalized inverse in the definition of the weighting matrix. Let MP-QLR denote this

statistic. For the case of singular correlation matrices, Section S6.3 of the Supplemental Material 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.[22] For example, for $p = 10$, its power ranges from 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-deviation asymptotic optimality criterion is not a suitable criterion in this context.[23]

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-deviation 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 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 are size-corrected, a by-product of determining the

---

[22]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 the Supplemental Material.

[23]In our view, the large-deviation asymptotic optimality criterion is not appropriate when comparing tests with substantially different asymptotic properties under nonlarge deviations. In particular, this criterion is questionable when the alternative hypothesis is multidimensional 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.

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.[24] If true, then the size of the AQLR/t-test test depends on the two-dimensional submatrices 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 submatrix of Ω is the one with the smallest correlation. Hence, the value of κ that makes the size of the test equal to α for a small value of η is (approximately) a function of Ω through $\delta(\Omega)$ 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, that is, $\eta_1(\delta)$ 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). See Sections S6.1 and S7.4 of the Supplemental Material 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

---

[24]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 with dimensions 3×3 or greater has restrictions on its correlations that are imposed by the positive semidefiniteness property. If all of the correlations are equal, they cannot be arbitrarily close to -1. In contrast, with a two-dimensional variance matrix, the correlation can be arbitrarily close to -1.

method does in approximating the best κ (viz., $\kappa^*(\Omega)$).[25] 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 ($= 3 \times 19$) 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 three (p, Ω) combinations, power that is lower by .01 for two combinations, and power that is lower by .02 for one combination.

The fourth group consists of 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 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 Section S6.3.3 of the Supplemental Material 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

---

[25]For brevity, details of the numerical results are given in Section S6.1.2 of the Supplemental Material.

TABLE III
FINITE-SAMPLE MAXIMUM NULL REJECTION PROBABILITIES AND (SIZE-CORRECTED)
AVERAGE POWER OF NOMINAL .05 TESTS[a]

Test	Distribution	H_0/H_1	$p = 10$			$p = 4$			$p = 2$		
			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_3^2	H_0	.136	.153	.068	.093	.101	.062	.085	.087	.080
AQLR/Bt	χ_3^2	H_0	.062	.066	.057	.050	.055	.050	.054	.053	.056
ELR/Bt	χ_3^2	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_3^2	H_1	.37	.42	.72	.48	.53	.71	.56	.57	.61
AQLR/Bt	χ_3^2	H_1	.43	.51	.72	.53	.57	.70	.57	.59	.62
ELR/Bt	χ_3^2	H_1	.41	.47	.70	.53	.56	.70	.56	.59	.62

[a]Asymptotic normal critical values are denoted Nm; bootstrap-based critical values are denoted Bt. All of the tests use t-test/κauto RMS critical values.

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 0 variance I_p random vector $Z^{\dagger} = \text{Var}^{-1/2}(m(W_i, \theta))(m(W_i, \theta) - Em(W_i, \theta))$ to be i.i.d. across elements and consider three distributions for the elements: standard normal (i.e., $N(0, 1)$), t_3, and chi-squared with 3 degrees of freedom χ_3^2. All of these distributions are centered and scaled to have mean 0 and variance 1. 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.[26]

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 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 overrejects 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_3^2, the AQLR normal test overrejects the null hypothesis substantially with its MNRP's being in the range [.068, .153]. The fact that overrejection 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 overrejection over the 27 cases is .012, whereas it is .005 for the AQLR bootstrap test.

For the $N(0, 1)$, t, and χ_3^2 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 10,000 bootstrap repetitions takes 3.1, 3.8, and 5.6 seconds when $p = 2$, 4, and 10, respectively, and $n = 250$ using GAUSS on a PC with a 3.2-GHz processor. This is slower than the AQLR/t-test/κauto bootstrap test (see the end of Section 2) by factors of 9.1, 9.7, and 6.5.

Section S6.3.2 of the Supplemental Material 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

---

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

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 Section S6.3.3 of the Supplemental Material 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 almost surely, 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 outperforms 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. [2814]

ANDREWS, D. W. K., AND P. J. BARWICK (2012): "Supplement to 'Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure'," *Econometrica Supplemental Material*, 80, http://www.econometricsociety.org/ecta/Supmat/8166_miscellaneous.pdf; http://www.econometricsociety.org/ecta/Supmat/8166_data_and_programs-1.zip; http://www.econometricsociety.org/ecta/Supmat/8166_data_and_programs-2.zip. [2806]

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. [2805]

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157. [2805]

BUGNI, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753. [2805,2806,2813,2815]

BUGNI, F. A., I. A. CANAY, AND P. GUGGENBERGER (2012): "Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models," *Econometrica*, 80, 1741–1768. [2806]

CANAY, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408–425. [2806,2813,2816,2817]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [2805]

CHIBURIS, R. (2009): "Approximately Most Powerful Tests for Moment Inequalities," Unpublished Manuscript, Department of Economics, University of Texas, Austin. [2806]

FAN, Y., AND S. PARK (2007): "Confidence Sets for Some Partially Identified Parameters," Unpublished Manuscript, Department of Economics, Vanderbilt University. [2806,2813,2814]

HANSEN, P. (2005): "A Test for Superior Predictive Ability," *Journal of Business & Economic Statistics*, 23, 365–380. [2806,2813,2814]

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [2805]

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2004): "Applications of Moment Inequalities," Unpublished Working Paper, Department of Economics, Harvard University. [2813]

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)*, 138, 2786–2807. [2805,2813]

ROSEN, A. M. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117. [2806,2813]

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

and

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