INFERENCE FOR PARAMETERS DEFINED BY MOMENT INEQUALITIES USING GENERALIZED MOMENT SELECTION

BY

DONALD W. K. ANDREWS and GUSTAVO SOARES

COWLES FOUNDATION PAPER NO. 1291

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

2010

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

Econometrica, Vol. 78, No. 1 (January, 2010), 119–157

INFERENCE FOR PARAMETERS DEFINED BY MOMENT INEQUALITIES USING GENERALIZED MOMENT SELECTION

BY DONALD W. K. ANDREWS AND GUSTAVO SOARES[1]

The topic of this paper is inference in models in which parameters are defined by moment inequalities and/or equalities. The parameters may or may not be identified. This paper introduces a new class of confidence sets and tests based on generalized moment selection (GMS). GMS procedures are shown to have correct asymptotic size in a uniform sense and are shown not to be asymptotically conservative.

The power of GMS tests is compared to that of subsampling, m out of n bootstrap, and "plug-in asymptotic" (PA) tests. The latter three procedures are the only general procedures in the literature that have been shown to have correct asymptotic size (in a uniform sense) for the moment inequality/equality model. GMS tests are shown to have asymptotic power that dominates that of subsampling, m out of n bootstrap, and PA tests. Subsampling and m out of n bootstrap tests are shown to have asymptotic power that dominates that of PA tests.

KEYWORDS: Asymptotic size, asymptotic power, confidence set, exact size, generalized moment selection, m out of n bootstrap, subsampling, moment inequalities, moment selection, test.

1. INTRODUCTION

THIS PAPER CONSIDERS INFERENCE in models in which parameters are defined by moment inequalities and/or equalities. The parameters need not be identified. Numerous examples of such models are now available in the literature, for example, see Manski and Tamer (2002), Imbens and Manski (2004), Andrews, Berry, and Jia (2004), Pakes, Porter, Ishii, and Ho (2004), Moon and Schorfheide (2006), Chernozhukov, Hong, and Tamer (2007) (CHT), and Ciliberto and Tamer (2009).

The paper introduces confidence sets (CS's) based on a method called *generalized moment selection* (GMS). The CS's considered in the paper are obtained by inverting tests that are of an Anderson–Rubin type. This method was first considered in the moment inequality context by CHT.

In this paper, we analyze GMS critical values. We note that the choice of critical value is much more important in moment inequality/equality models than in most models. In most models, the choice of critical value does not affect the first-order asymptotic properties of a test or CS. In the moment inequality/equality model, however, it does, and the effect can be large.

The results of the paper hold for a broad class of test statistics including modified method of moments (MMM) statistics, Gaussian quasilikelihood ratio (QLR) statistics, generalized empirical likelihood ratio (GEL) statistics,

[1]Andrews gratefully acknowledges the research support of the National Science Foundation via Grants SES-0417911 and SES-0751517. The authors thank three referees, a co-editor, Ivan Canay, Victor Chernozhukov, Jörg Stoye, and especially Patrik Guggenberger for comments.

 DOI: 10.3982/ECTA7502

and a variety of others. The results apply to CS's for the true parameter, as in Imbens and Manski (2004), rather than for the identified set (i.e., the set of points that are consistent with the population moment inequalities/equalities), as in CHT. We focus on CS's for the true parameter because answers to policy questions typically depend on the true parameter rather than on the identified set.

Subsampling CS's for the moment inequality/equality model are considered in CHT, Andrews and Guggenberger (2009b) (hereafter AG4), and Romano and Shaikh (2008, 2010). "Plug-in asymptotic" (PA) CS's are widely used in the literature on multivariate one-sided tests and CS's. They are considered in the moment inequality/equality model in AG4 and a variant of them is considered in Rosen (2008).

Here we introduce GMS critical values. Briefly, the idea behind GMS critical values is as follows. The $1-\alpha$ quantile of the finite-sample null distribution of a typical test statistic depends heavily on the extent to which the moment inequalities are binding (i.e., are close to being equalities). In consequence, the asymptotic null distribution of the test statistic under a suitable drifting sequence of parameters depends heavily on a nuisance parameter $h=(h_1,\ldots,h_p)'$, whose jth element $h_j \in [0,\infty]$ indexes the extent to which the jth moment inequality is binding. For a suitable class of test statistics, the larger is h, the smaller is the asymptotic null distribution in a stochastic sense. This is key for obtaining procedures that are uniformly asymptotically valid.

The parameter h cannot be estimated consistently in a uniform sense, but one can use the sample moment inequalities to estimate or test how close h is to 0_p. A computationally simple procedure is to use inequality-by-inequality t-tests to test whether $h_j=0$ for $j=1,\ldots,p$. If a test rejects $h_j=0$, then that inequality is removed from the asymptotic null distribution that is used to calculate the critical value. The t-tests have to be designed so that the probability of incorrectly omitting a moment inequality from the asymptotic distribution is asymptotically negligible. Continuous/smooth versions of such procedures can be employed in which moment inequalities are not "in or out," but are "more in or more out" depending on the magnitude of the t statistics.

Another type of GMS procedure is based on a modified moment selection criterion (MMSC), which is an information-type criterion analogous to the Akaike information criterion (AIC), Bayesian information criterion (BIC), and Hannan–Quinn information criterion (HQIC) model selection criteria; see Hannan and Quinn (1979) regarding HQIC. Andrews (1999a) used an information-type moment selection criterion to determine which moment equalities are invalid in a standard moment equality model. Here we employ one-sided versions of such procedures to determine which moment inequalities are not binding. In contrast to inequality-by-inequality t-tests, the MMSC jointly determines which moment inequalities to select and takes account of correlations between sample moment inequalities.

The results of the paper cover a broad class of GMS procedures that includes all of those discussed above. Section 4.2 gives a step-by-step description of how to calculate GMS procedures.

In this paper, we show that GMS critical values yield uniformly asymptotically valid CS's and tests. These results hold for both independent and identically distributed (i.i.d.) and dependent observations. We also show that GMS procedures are not asymptotically conservative. They are asymptotically nonsimilar, but are less so than subsampling and PA procedures.

The volume of a CS that is based on inverting a test depends on the power of the test; see Pratt (1961). Thus, power is important for both tests and CS's. We determine and compare the power of GMS, subsampling, and PA tests. CHT and Beresteanu and Molinari (2008) also provided some asymptotic local power results for testing procedures in models with partially identified parameters. Otsu (2006), Bugni (2007a, 2007b), and Canay (2007) considered asymptotic power against fixed alternatives. Tests typically have asymptotic power equal to 1 against such alternatives.

We investigate the asymptotic power of GMS, subsampling, and PA tests for local and nonlocal alternatives. Such alternatives are more complicated in the moment inequality/equality model than in most models. The reason is that some inequalities may be violated while others may be satisfied as equalities, as inequalities that are relatively close to being equalities, and/or as inequalities that are far from being equalities. Furthermore, depending upon the particular alternative hypothesis scenario considered, the data-dependent critical values behave differently asymptotically. We derive the asymptotic power of the tests under the complete range of alternatives from $n^{-1/2}$ local, to more distant local, through to fixed alternatives for each of the different moment inequalities and equalities that appear in the model.

We show that (under reasonable assumptions) GMS tests are as powerful asymptotically as subsampling and PA tests with strictly greater power in certain scenarios. The asymptotic power differences can be substantial. Furthermore, we show that subsampling tests are as powerful asymptotically as PA tests with greater power in certain scenarios. m out of n bootstrap tests have the same asymptotic properties as subsampling tests (at least in i.i.d. scenarios when $m = o(n^{1/2})$; see Politis, Romano, and Wolf (1999, p. 48)).

GMS tests are shown to be strictly more powerful asymptotically than subsampling tests whenever (i) at least one population moment inequality is satisfied under the alternative and differs from an equality by an amount that is $O(b^{-1/2})$ and is larger than $O(\kappa_n n^{-1/2})$, where b is the subsample size and κ_n is a GMS constant such as $\kappa_n = (\ln n)^{1/2}$, and (ii) the GMS and subsampling critical values do not have the degenerate probability limit of 0. Note that good choices of b and κ_n in terms of size and power satisfy $b \approx n^\eta$ for some $\eta \in (0, 1)$ and $\kappa_n = o(n^\varepsilon)$ $\forall \varepsilon > 0$, so that condition (i) holds.

GMS and subsampling tests are shown to be strictly more powerful asymptotically than PA tests whenever at least one population moment inequality is satisfied under the alternative, and differs from an equality by an amount that is larger than $O(\kappa_n n^{-1/2})$ for GMS tests and is larger than $o(b^{-1/2})$ for subsampling tests.

The paper shows that (pure) generalized empirical likelihood (GEL) tests, which are based on fixed critical values, are dominated in terms of asymptotic power by GMS and subsampling tests based on a QLR or GEL test statistic.

The paper reports some finite-sample size and power results obtained via simulation for tests based on GMS, subsampling, and PA critical values. The GMS critical values are found to deliver very good null rejection probabilities and power in the scenarios considered. They are found to outperform the subsampling and PA critical values by a substantial margin, especially for larger values of p, the number of moment inequalities. Additional simulation results are reported in the Supplemental Material (Andrews and Soares (2010)) and in Andrews and Jia (2008).

The determination of a best test statistic/GMS procedure is difficult because uniformly best choices do not exist. Nevertheless, it is possible to make comparisons based on all-around performance. Doing so is beyond the scope of the present paper and is the subject of research reported in Andrews and Jia (2008). In the latter paper, the QLR test statistic combined with the GMS procedure based on t-tests is found to work very well in practice and hence is recommended.

Bootstrap versions of GMS critical values are obtained by replacing the multivariate normal random vector that appears in the asymptotic distribution by a bootstrap distribution based on the recentered sample moments. The block bootstrap can be employed in time series contexts. GMS bootstrap critical values, however, do not yield higher-order improvements, because the asymptotic null distribution is not asymptotically pivotal. Bugni (2007a, 2007b) and Canay (2007) considered particular types of bootstrap GMS critical values. Andrews and Jia (2008) found that the bootstrap version of the GMS critical values outperforms the asymptotic normal version in i.i.d. scenarios and hence is recommended.

The paper introduces GMS model specification tests based on the GMS tests discussed above. These tests are shown to be uniformly asymptotically valid. They can be asymptotically conservative.

We now discuss related literature. Bugni (2007a, 2007b) showed that a particular type of GMS test (based on $\varphi^{(1)}$ defined below) has more accurate pointwise asymptotic size than a (recentered) subsampling test. Such results should extend to all GMS tests and to asymptotic size defined in a uniform sense. Given that they do, GMS tests have both asymptotic power and size advantages over subsampling tests. The relatively low accuracy of the size of subsampling tests and CS's in many models is well known in the literature. We are not aware of any other papers or scenarios where the asymptotic power of subsampling tests has been shown to be dominated by other procedures.

GMS critical values based on $\varphi^{(1)}$, defined below, are a variant of the Wald test procedure in Andrews (1999b, Sec. 6.4; 2000, Sec. 4) for the problem of inference when a parameter is on or near a boundary. The 2003 working paper version of CHT discusses a bootstrap version of the GMS method based on $\varphi^{(1)}$ in the context of the interval outcome model. Soares (2005) analyzed the properties of GMS critical values based on $\varphi^{(1)}$ and introduced GMS critical values based on the function $\varphi^{(5)}$. The present paper supplants Soares (2005). CHT mentioned critical values of GMS type based on $\varphi^{(1)}$; see their Remark 4.5. GMS critical values of types $\varphi^{(2)}$–$\varphi^{(4)}$ were considered by the authors in January 2007. Galichon and Henry (2009) consider a set selection method that is analogous to GMS based on $\varphi^{(1)}$. Bugni (2007a, 2007b) considered GMS critical values based on $\varphi^{(1)}$. His work was done independently of, but subsequently to, Soares (2005). Canay (2007) independently considered GMS critical values based on $\varphi^{(3)}$. Bugni (2007a, 2007b) and Canay (2007) focused on bootstrap versions of the GMS critical values.

Other papers in the literature that consider inference with moment inequalities include Andrews, Berry, and Jia (2004), Pakes et al. (2004), Romano and Shaikh (2008, 2010), Moon and Schorfheide (2006), Otsu (2006), Woutersen (2006), Bontemps, Magnac, and Maurin (2007), Bugni (2007a, 2007b), Canay (2007), CHT, Fan and Park (2007), Beresteanu, Molchanov, and Molinari (2008), Beresteanu and Molinari (2008), Guggenberger, Hahn, and Kim (2008), Rosen (2008), AG4, Andrews and Han (2009), and Stoye (2009).

The remainder of the paper is organized as follows. Section 2 describes the moment inequality/equality model. Section 3 introduces the class of test statistics that is considered and states assumptions. Section 4 introduces the class of GMS CS's. Section 5 introduces GMS model specification tests. Sections 6 and 7 define subsampling CS's and PA CS's, respectively. Section 8 determines and compares the $n^{-1/2}$-local alternative power of GMS, subsampling, and PA tests. Section 9 considers the power of these tests against more distant alternatives. Section 10 discusses extensions to GEL test statistics and preliminary estimation of identified parameters. Section 11 provides the simulation results. The Appendix contains some assumptions concerning the test statistics considered, an alternative parametrization of the moment inequality/equality model, and the treatment of dependent observations. The proofs of all results are given in the Supplemental Material (Andrews and Soares (2010)).

For notational simplicity, throughout the paper we write partitioned column vectors as $h = (h_1, h_2)$, rather than $h = (h_1', h_2')'$. Let $R_+ = \{x \in R : x \geq 0\}$, $R_{+,\infty} = R_+ \cup \{+\infty\}$, $R_{[+\infty]} = R \cup \{+\infty\}$, $R_{[\pm\infty]} = R \cup \{\pm\infty\}$, $K^p = K \times \cdots \times K$ (with p copies) for any set K, and $\infty^p = (+\infty, \ldots, +\infty)'$ (with p copies). All limits are as $n \to \infty$ unless specified otherwise. Let pd abbreviate positive definite. Let $\mathrm{cl}(\Psi)$ denote the closure of a set Ψ. We let AG1 abbreviate Andrews and Guggenberger (2010b).

2. MOMENT INEQUALITY MODEL

We now introduce the moment inequality/equality model. The true value θ_0 ($\in \Theta \subset R^d$) is assumed to satisfy the moment conditions:

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

where $\{m_j(\cdot, \theta) : j = 1, \dots, k\}$ 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. The observed sample is $\{W_i : i \leq n\}$. A key feature of the model is that the true value θ_0 is not necessarily identified. That is, knowledge of $E_{F_0} m_j(W_i, \theta)$ for $j = 1, \dots, k$ for all $\theta \in \Theta$ does not necessarily imply knowledge of θ_0. In fact, even knowledge of F_0 does not necessarily imply knowledge of the true value θ_0. More information than is available in $\{W_i : i \leq n\}$ may be needed to identify the true parameter θ_0.

Note that both moment inequalities and moment equalities arise in the entry game models considered in Ciliberto and Tamer (2009) and Andrews, Berry, and Jia (2004), and in the macroeconomic model in Moon and Schorfheide (2006). There are numerous models where only moment inequalities arise; for example, see Manski and Tamer (2002) and Imbens and Manski (2004). There are also unidentified models in which only moment equalities arise; see CHT for references.

We are interested in 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

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

where Ψ is a set of $k \times k$ correlation matrices specified below, and $M < \infty$ and $\delta > 0$ are constants. For expositional convenience, we specify $\mathcal{F}$ for dependent observations in the Appendix, Section A.2.

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

$$\text{(2.3)} \qquad \text{CS}_n = \{\theta \in \Theta : T_n(\theta) \le c_{1-\alpha}(\theta)\},$$

where $c_{1-\alpha}(\theta)$ is a critical value.[2] We consider GMS, subsampling, and plug-in asymptotic critical values. These are data-dependent critical values and their probability limits, when they exist, typically depend on the true distribution generating the data.

The exact and asymptotic confidence sizes of CS_n are

$$\text{(2.4)} \qquad \text{ExCS}_n = \inf_{(\theta,F)\in\mathcal{F}} P_F(T_n(\theta) \le c_{1-\alpha}(\theta)) \quad \text{and} \quad \text{AsyCS} = \liminf_{n\to\infty} \text{ExCS}_n,$$

respectively. The definition of AsyCS takes the $\inf_{(\theta,F)\in\mathcal{F}}$ before the $\lim_{n\to\infty}$. 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 CS's. Andrews and Guggenberger (2009a, 2010a, 2010b) and Mikusheva (2007) showed that when a test statistic has a discontinuity in its limit distribution, as occurs in the moment inequality/equality model, pointwise asymptotics (in which one takes the lim before the inf) can be very misleading in some models. See AG4 for further discussion.

The exact and asymptotic maximum coverage probabilities are

$$\text{(2.5)} \qquad \text{ExMaxCP}_n = \sup_{(\theta,F)\in\mathcal{F}} P_F(T_n(\theta) \le c_{1-\alpha}(\theta)),$$

$$\text{AsyMaxCP} = \limsup_{n\to\infty} \text{ExMaxCP}_n,$$

respectively. The magnitude of asymptotic nonsimilarity of the CS is measured by the difference AsyMaxCP − AsyCS.

If interest is in a subvector, say β, of θ, then confidence sets for β can be constructed via projection. That is, one takes the CS to be $\{\beta \in R^{d_\beta}:$ for some $\lambda \in R^{d-d_\beta}$, $(\beta', \lambda')' \in \text{CS}_n\}$. By a standard argument, if CS_n is a CS for θ with asymptotic size greater than or equal to $1-\alpha$, then this CS for β has the same property. Typically, however, a CS for β constructed in this way has an asymptotic size that is strictly greater than $1-\alpha$, which implies that it is asymptotically conservative.

3. TEST STATISTICS

In this section, we define the main class of test statistics $T_n(\theta)$ that we consider. GEL statistics are discussed in Section 10 below.

[2]It is important that the inequality in the definition of CS_n is less than or equal to, not less than. When θ is in the interior of the identified set, it is often the case that $T_n(\theta) = 0$ and $c_{1-\alpha}(\theta) = 0$.

3.1. *Form of the Test Statistics*

The sample moment functions are

$$\overline{m}_n(\theta) = (\overline{m}_{n,1}(\theta), \ldots, \overline{m}_{n,k}(\theta))', \quad \text{where} \tag{3.1}$$
$$\overline{m}_{n,j}(\theta) = n^{-1} \sum_{i=1}^{n} m_j(W_i, \theta) \quad \text{for} \quad j = 1, \ldots, k.$$

Let $\widehat{\Sigma}_n(\theta)$ be an estimator of the asymptotic variance, $\Sigma(\theta)$, of $n^{1/2}\overline{m}_n(\theta)$. When the observations are i.i.d., we take

$$\widehat{\Sigma}_n(\theta) = n^{-1} \sum_{i=1}^{n} (m(W_i, \theta) - \overline{m}_n(\theta))(m(W_i, \theta) - \overline{m}_n(\theta))', \quad \text{where} \tag{3.2}$$
$$m(W_i, \theta) = (m_1(W_i, \theta), \ldots, m_k(W_i, \theta))'.$$

With temporally dependent observations, a different definition of $\widehat{\Sigma}_n(\theta)$ often is required. For example, a heteroskedasticity and autocorrelation consistent (HAC) estimator may be required.

The statistic $T_n(\theta)$ is defined to be of the form

$$T_n(\theta) = S\big(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)\big), \tag{3.3}$$

where S is a real function on $R^p_{[+\infty]} \times R^v \times \mathcal{V}_{k\times k}$, where $\mathcal{V}_{k\times k}$ 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 function S is required to satisfy Assumptions 1–6 stated below. We now give several examples of functions that do so.

First, consider the MMM test function $S = S_1$ defined by

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

and σ_j^2 is the jth diagonal element of Σ. With the function S_1, the parameter space Ψ for the correlation matrices in condition (vi) of (2.2) is $\Psi = \Psi_1$, where Ψ_1 contains all $k \times k$ correlation matrices.[3] The function S_1 yields a test statistic that gives positive weight to moment inequalities only when they are violated.

[3]Note that with temporally dependent observations, Ψ is the parameter space for the limiting correlation matrix, $\lim_{n\to\infty} \mathrm{Corr}_F(n^{1/2}\overline{m}_n(\theta))$.

This type of statistic has been considered in Romano and Shaikh (2008, 2010), Soares (2005), CHT, and AG4. Note that S_1 normalizes the moment functions by dividing by σ_j in each summand. One could consider a function without this normalization, as in Pakes et al. (2004) and Romano and Shaikh (2008, 2010), but the resulting statistic is not invariant to rescaling of the moment conditions and, hence, is not likely to have good properties in terms of the volume of its CS. We use the function S_1 in the simulations reported in Section 11 below.

Second, we consider a QLR test function defined by

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

With this function, the parameter space Ψ in (2.2) is $\Psi=\Psi_2$, where Ψ_2 contains all $k\times k$ correlation matrices whose determinant is greater than or equal to ε for some $\varepsilon>0$.[4,5] This type of 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)), as well as papers in the moment inequality literature (see Rosen (2008)). We note that GEL test statistics behave asymptotically (to the first order) under the null and alternative hypotheses like the statistic $T_n(\theta)$ based on S_2; see Section 10 below and AG4.

The requirement that $\Psi=\Psi_2$ for S_2 is restrictive in some cases, such as when two moment inequalities have correlation equal to 1 in absolute value. In such cases, one can alter the definition of S_2 in (3.5) by replacing Σ by $\Sigma+\varepsilon\,\mathrm{Diag}(\Sigma)$ for some $\varepsilon>0$, where $\mathrm{Diag}(\Sigma)$ denotes the $k\times k$ diagonal matrix whose diagonal elements equal those of Σ. With this alteration, one can take $\Psi=\Psi_1$.

For a test with power directed against alternatives with p_1 ($<p$) moment inequalities violated, the following function is suitable:

$$(3.6)\qquad S_3(m,\Sigma)=\sum_{j=1}^{p_1}\big[m_{(j)}/\sigma_{(j)}\big]_-^2+\sum_{j=p+1}^{p+v}(m_j/\sigma_j)^2,$$

where $[m_{(j)}/\sigma_{(j)}]_-^2$ denotes the jth largest value among $\{[m_\ell/\sigma_\ell]_-^2:\ell=1,\dots,p\}$ and $p_1<p$ is some specified integer. The function S_3 satisfies (2.2) with $\Psi=\Psi_1$. The function S_3 is considered in Andrews and Jia (2008).

[4]The condition that $\Psi=\Psi_2$ for the function S_2 is used in the proofs of various asymptotic results. This condition may be just a consequence of the method of proof. It may not actually be needed.

[5]The definition of $S_2(m,\Sigma)$ takes the infimum over $t_1\in R^p_{+,\infty}$, rather than over $t_1\in R^p_+$. For calculation of the test statistic based on S_2, using the latter gives an equivalent value. To obtain the correct asymptotic distribution, however, the former definition is required because it leads to continuity at infinity of S_2 when some elements of m may equal infinity. For example, suppose $k=p=1$. In this case, when $m\in R_+$, $\inf_{t_1\in R_{+,\infty}}(m-t_1)^2=\inf_{t_1\in R_+}(m-t_1)^2=0$. However, when $m=\infty$, $\inf_{t_1\in R_{+,\infty}}(m-t_1)^2=0$, but $\inf_{t_1\in R_+}(m-t_1)^2=\infty$.

Other examples of test functions S that satisfy Assumptions 1–6 are variations of S_1 and S_3 with the step function $[x]_-$ replaced by a smooth function, with the square replaced by the absolute value to a different positive power (such as 1), or with weights added.

It is difficult to compare the performance of one test function S with another function without specifying the critical values to be used. Most critical values, such as the GMS, subsampling, and PA critical values considered here, are data dependent and have limits as $n \to \infty$ that depend on the distribution of the observations. For a given test function S, a different test is obtained for each type of critical value employed and the differences do not vanish asymptotically. The relative performances of different functions S are considered elsewhere; see Andrews and Jia (2008).

3.2. *Test Statistic Assumptions*

Next, we state the most important assumptions concerning the function S, namely, Assumptions 1, 3, and 6. For ease of reading, technical assumptions (mostly continuity and strictly-increasing assumptions on asymptotic distribution functions (df's)), namely, Assumptions 2, 4, 5, and 7, are stated in the Appendix. We show below that the functions S_1–S_3 automatically satisfy Assumptions 1–6. Assumption 7 is not restrictive.

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 the assumptions below, the set Ψ is as in condition (vi) of (2.2).[6] For p-vectors m_1 and m_1^*, $m_1 < m_1^*$ means that $m_1 \leq m_1^*$ and at least one inequality in the p-vector of inequalities holds strictly.

ASSUMPTION 1: (a) *$S((m_1, m_2), \Sigma)$ is nonincreasing in m_1 for all $m_1 \in R^p$, $m_2 \in R^v$, and variance matrices $\Sigma \in R^{k\times k}$.*

(b) *$S(m, \Sigma) = S(Dm, D\Sigma D)$ for all $m \in R^k$, $\Sigma \in R^{k\times k}$, and pd diagonal $D \in R^{k\times k}$.*

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

(d) *$S(m, \Omega)$ is continuous at all $m \in R^p_{[+\infty]} \times R^v$ and $\Omega \in \Psi$.*[7]

ASSUMPTION 3: *$S(m, \Omega) > 0$ if and only if $m_j < 0$ for some $j = 1, \dots, p$ or $m_j \neq 0$ for some $j = p+1, \dots, k$, where $m = (m_1, \dots, m_k)'$ and $\Omega \in \Psi$.*

ASSUMPTION 6: *For some $\chi > 0$, $S(am, \Omega) = a^\chi S(m, \Omega)$ for all scalars $a > 0$, $m \in R^k$, and $\Omega \in \Psi$.*

[6]For dependent observations, Ψ is as in condition (v) of (A.2) in the Appendix.

[7]In Assumption 1(d) (and in Assumption 4(b) in the Appendix), $S(m, \Omega)$ and $c(\Omega, 1-\alpha)$ are viewed as functions defined on the space of all correlation matrices Ψ_1. By definition, $c(\Omega, 1-\alpha)$ is continuous in Ω uniformly for $\Omega \in \Psi$ if for all $\eta > 0$, there exists $\delta > 0$ such that whenever $\|\Omega^* - \Omega\| < \delta$ for $\Omega^* \in \Psi_1$ and $\Omega \in \Psi$, we have $|c_{\Omega^*}(1-\alpha) - c_\Omega(1-\alpha)| < \eta$.

Assumptions 1–6 are shown in Lemma 1 below not to be restrictive. Assumption 1(a) is the key assumption that is needed to ensure that GMS and subsampling CS's have correct asymptotic size. Assumption 1(b) is a natural assumption that specifies that the test statistic is invariant to the scale of each sample moment. Assumption 1(b) and 1(d) are conditions that enable one to determine the asymptotic properties of $T_n(\theta)$. Assumption 1(c) normalizes the test statistic to be nonnegative.

Assumption 3 implies that a positive value of $S(m, \Omega)$ only occurs if some inequality or equality is violated. Assumption 3 implies that $S(\infty^p, \Sigma) = 0$ when $v = 0$. Assumption 6 requires S to be homogeneous of degree $\chi > 0$ in m. This is used to show that the test based on S has asymptotic power equal to 1 against fixed alternatives.

LEMMA 1: *The functions* $S_1(m, \Sigma)$–$S_3(m, \Sigma)$ *satisfy Assumptions* 1–6 *with* $\Psi = \Psi_1$ *for* $S_1(m, \Sigma)$ *and* $S_3(m, \Sigma)$, *and with* $\Psi = \Psi_2$ *for* $S_2(m, \Sigma)$.

4. GENERALIZED MOMENT SELECTION

This section is concerned with critical values for use with the test statistics introduced in Section 3.

4.1. *Description of the GMS Method*

We start by motivating the GMS method. Consider the null hypothesis $H_0 : \theta = \theta_0$. The finite-sample null distribution of $T_n(\theta_0)$ depends continuously on the degree of *slackness* of the moment inequalities. That is, it depends on how much greater than zero is $E_F m_j(W_i, \theta_0)$ for $j = 1, \dots, p$. Under Assumption 1(a), the least favorable case (at least asymptotically) can be shown to be the case where there is no slackness—each of the moments is zero. That is, the distribution of $T_n(\theta_0)$ is stochastically largest over distributions in the null hypothesis when the inequality moments equal zero. One way to construct a critical value for $T_n(\theta_0)$, then, is to take the $1 - \alpha$ quantile of the distribution (or asymptotic distribution) of $T_n(\theta_0)$ when the inequality moments all equal zero. This yields a test with correct (asymptotic) size, but its power properties are poor against many alternatives of interest.

The reason for its poor power is that the least favorable critical value is relatively large. This is especially true if the number of moment inequalities, p, is large. For example, consider power against an alternative for which only the first moment inequality is violated (i.e., $E_F m_1(W_i, \theta_0) < 0$) and the last $p - 1$ moment inequalities are satisfied by a wide margin (i.e., $E_F m_j(W_i, \theta_0) \gg 0$ for $j = 2, \dots, p$). Then the last $p - 1$ moment inequalities have little or no effect on the value of the test statistic $T_n(\theta_0)$. (This holds for typical test statistics and is implied by Assumption 3.) Yet, the critical value *does* depend on the existence of the last $p - 1$ moment inequalities and is much larger than it would be if these moment inequalities were absent. In consequence, the test has significantly lower power than if the last $p-1$ moment inequalities were absent.

The idea behind *generalized moment selection* is to use the data to determine whether a given moment inequality is satisfied and is far from being an equality, and if so to take the critical value to be smaller than otherwise—both under the null and under the alternative. Of course, in doing so, one has to make sure that the (asymptotic) size of the resulting test is correct. We use the sample moment functions to estimate or test whether the population moment inequalities are close to or far from being equalities.

Using Assumption 1(b), we can write

$$\begin{aligned}(4.1)\qquad T_n(\theta) &= S\big(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)\big)\\ &= S\big(\widehat{D}_n^{-1/2}(\theta)n^{1/2}\overline{m}_n(\theta), \widehat{\Omega}_n(\theta)\big), \quad \text{where}\\ \widehat{D}_n(\theta) &= \mathrm{Diag}(\widehat{\Sigma}_n(\theta)) \quad \text{and} \quad \widehat{\Omega}_n(\theta) = \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta)\widehat{D}_n^{-1/2}(\theta).\end{aligned}$$

Thus, the test statistic $T_n(\theta)$ depends only on the normalized sample moments and the sample correlation matrix. Under an appropriate sequence of null distributions $\{F_n : n \geq 1\}$, the asymptotic null distribution of $T_n(\theta_0)$ is that of

$$(4.2)\qquad S\big(\Omega_0^{1/2}Z^* + (h_1, 0_v), \Omega_0\big), \quad \text{where} \quad Z^* \sim N(0_k, I_k),$$

$h_1 \in R^p_{+,\infty}$, and Ω_0 is a $k \times k$ correlation matrix. This result holds by (4.1), the central limit theorem, and the convergence in probability of the sample correlation matrix; see the proof of Theorem 1 of AG4. The p-vector h_1 is the limit of $(n^{1/2}E_{F_n}m_1(W_i, \theta_0)/\sigma_{F_n,1}(\theta_0), \ldots, n^{1/2}E_{F_n}m_p(W_i, \theta_0)/\sigma_{F_n,p}(\theta_0))'$ under the null distributions $\{F_n : n \geq 1\}$. By considering suitable sequences of distributions F_n that depend on n, rather than a fixed distribution F, we obtain an asymptotic distribution that depends continuously on the degree of slackness of the population moment inequalities via the parameter h_1 $(\geq 0_p)$. This reflects the finite-sample situation.

Note that the correlation matrix Ω_0 can be consistently estimated, but the $n^{-1/2}$-local asymptotic mean parameter h_1 cannot be (uniformly) consistently estimated. It is the latter property that makes it challenging to determine a critical value that yields a test with correct asymptotic size and good power properties.

The GMS critical value is defined to be 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. We measure the degree of slackness of the moment inequalities via

$$(4.3)\qquad \xi_n(\theta) = \kappa_n^{-1}n^{1/2}\widehat{D}_n^{-1/2}(\theta)\overline{m}_n(\theta) \in R^k$$

evaluated at $\theta = \theta_0$, where $\{\kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \to \infty$. A suitable choice of κ_n is the BIC choice

$$\kappa_n = (\ln n)^{1/2}. \tag{4.4}$$

The law of the iterated logarithm choice, $\kappa_n = (2 \ln \ln n)^{1/2}$, also is possible, but the simulations reported in Section 11 below indicate that the BIC choice is preferable. CHT also suggested using the BIC value.

Let $\xi_{n,j}(\theta)$, $h_{1,j}$, and $[\Omega_0^{1/2} Z^*]_j$ denote the jth elements of $\xi_n(\theta)$, h_1, and $\Omega_0^{1/2} Z^*$, respectively, for $j = 1, \ldots, p$. When $\xi_{n,j}(\theta_0)$ 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 $\xi_{n,j}(\theta_0)$ is large, this indicates $h_{1,j}$ is quite large (where the adjective "quite" is due to the κ_n factor) and the desired replacement of $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is ∞.

We replace $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ by $\varphi_j(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$ for $j = 1, \ldots, p$, where $\varphi_j : (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi \to R_{[\pm\infty]}$ is a function that is chosen to deliver the properties described above. Suppose φ_j satisfies (i) $\varphi_j(\xi, \Omega) = 0$ for all $\xi = (\xi_1, \ldots, \xi_k)' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$ with $\xi_j = 0$ and all $\Omega \in \Psi$, and (ii) $\varphi_j(\xi, \Omega) \to \infty$ as $(\xi, \Omega) \to (\xi_*, \Omega_*)$ for all $\xi_* = (\xi_{*,1}, \ldots, \xi_{*,k})' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$ with $\xi_{*,j} = \infty$ and all $\Omega_* \in \Psi$, where $\xi \in R^k$ and $\Omega \in \Psi$. In this case, if $\xi_{n,j}(\theta_0) = 0$, then $\varphi_j(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)) = 0$ and $h_{1,j}$ is replaced by 0, as desired. On the other hand, if $\xi_{n,j}(\theta_0)$ is large, condition (ii) implies that $\varphi_j(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$ is large and $h_{1,j}$ is replaced by a large value, as desired, for $j = 1, \ldots, p$. For $j = p+1, \ldots, k$, we define $\varphi_j(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)) = 0$ because no $h_{1,j}$ term appears in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$.

Examples of functions φ_j include

$$\varphi_j^{(1)}(\xi, \Omega) = \begin{cases} 0, & \text{if } \xi_j \leq 1 \\ \infty, & \text{if } \xi_j > 1, \end{cases} \qquad \varphi_j^{(2)}(\xi, \Omega) = \psi(\xi_j), \tag{4.5}$$
$$\varphi_j^{(3)}(\xi, \Omega) = [\xi_j]_+, \quad \text{and} \quad \varphi_j^{(4)}(\xi, \Omega) = \xi_j$$

for $j = 1, \ldots, p$, where ψ is defined below. Let $\varphi^{(r)}(\xi, \Omega) = (\varphi_1^{(r)}(\xi, \Omega), \ldots, \varphi_p^{(r)}(\xi, \Omega), 0, \ldots, 0)' \in R^p_{[\pm\infty]} \times \{0\}^v$ for $r = 1, \ldots, 4$.

The function $\varphi^{(1)}$ generates a "moment selection t-test" procedure. Using $\varphi^{(1)}$, $h_{1,j}$ is replaced in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ by ∞ if $\xi_{n,j}(\theta_0) > 1$ and by 0 otherwise. Note that $\xi_{n,j}(\theta_0) > 1$ is equivalent to

$$\frac{n^{1/2} \overline{m}_{n,j}(\theta_0)}{\widehat{\sigma}_{n,j}(\theta_0)} > \kappa_n, \tag{4.6}$$

where $\widehat{\sigma}^2_{n,j}(\theta_0)$ is the (j, j) element of $\widehat{\Sigma}_n(\theta_0)$ for $j = 1, \ldots, p$. The GMS procedure based on $\varphi^{(1)}$ is the same as the Wald test procedure in Andrews (1999b, Sec. 6.4; 2000, Sec. 4) for the related problem of inference when a parameter is on or near a boundary.

The function $\varphi^{(2)}$ in (4.5) depends on a nondecreasing function $\psi(x)$ that satisfies $\psi(x) = 0$ if $x \leq a_L$, $\psi(x) \in [0, \infty]$ if $a_L < x < a_U$, and $\psi(x) = \infty$ if $x > a_U$ for some $0 < a_L \leq a_U \leq \infty$. A key condition is that $a_L > 0$; see Assumption GMS1(a) below. The function $\varphi^{(2)}$ is a continuous version of $\varphi^{(1)}$ when ψ is taken to be continuous on R (where continuity at a_U means that $\lim_{x \to a_U} \psi(x) = \infty$).

The functions $\varphi^{(3)}$ and $\varphi^{(4)}$ exhibit a less steep rate of increase than $\varphi^{(1)}$ as a function of ξ_j for $j = 1, \ldots, p$.

The functions $\varphi^{(r)}$ for $r = 1, \ldots, 4$ all exhibit "element-by-element" determination of $\varphi_j^{(r)}(\xi, \Omega)$ because the latter depends only on ξ_j. This has significant computational advantages because $\varphi_j^{(r)}(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$ is very easy to compute. On the other hand, when $\widehat{\Omega}_n(\theta_0)$ is nondiagonal, the whole vector $\xi_n(\theta_0)$ contains information about the magnitude of $h_{1,j}$. We now introduce a function $\varphi^{(5)}$ that exploits this information (at least for certain choices of function S such as S_2). It is related to the information criterion-based moment selection criteria (MSC) considered in Andrews (1999a) 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 $\varphi^{(r)}$ functions considered above.

Define $c = (c_1, \ldots, c_k)'$ 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, that is, $c_j = 1$ for $j = p+1, \ldots, k$. Let $|c| = \sum_{j=1}^{k} c_j$. For $\xi \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, define $c \cdot \xi = (c_1 \xi_1, \ldots, c_k \xi_k)' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, 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. In many cases, $\mathcal{C} = \{0, 1\}^p \times \{1\}^v$. However, if there is a priori information that one moment inequality cannot hold as an equality if some other does and the sum of the degrees of slackness of the two moment inequalities is bounded away from zero over all admissible distributions, then this can be built into the definition of $\mathcal{C}$; see Rosen (2008) for a discussion of examples of this sort. Let $\zeta(\cdot)$ be a strictly increasing real function on R_+. Given $(\xi, \Omega) \in (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi$, the selected moment vector $c(\xi, \Omega) \in \mathcal{C}$ is the vector in $\mathcal{C}$ that minimizes the MMSC defined by

$$S(-c \cdot \xi, \Omega) - \zeta(|c|). \tag{4.7}$$

Note the minus sign that appears in the first argument of the S function. This 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. Hence, the minimizing selection vector $c(\xi, \Omega)$ trades off the minimization of $S(-c \cdot \xi, \Omega)$, which is achieved by selecting few moment functions, with the maximization of the bonus term, which is decreasing in the number of selected moments. For $j = 1, \ldots, p$, define

$$(4.8)\qquad \varphi_j^{(5)}(\xi, \Omega) = \begin{cases} 0 & \text{if } c_j(\xi, \Omega) = 1, \\ \infty & \text{if } c_j(\xi, \Omega) = 0. \end{cases}$$

Using Assumptions 1(b) and 6,

$$(4.9)\qquad \begin{aligned} &\kappa_n^{\chi}\big(S(-c \cdot \xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)) - \zeta(|c|)\big) \\ &\quad = S\big(-c \cdot n^{1/2}\overline{m}_n(\theta_0), \widehat{\Sigma}_n(\theta_0)\big) - \zeta(|c|)\kappa_n^{\chi}, \end{aligned}$$

where χ is as in Assumption 6. In consequence, the MMSC selection vector $c(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$ minimizes both the left-hand and right-hand sides (r.h.s.) of (4.9) over $\mathcal{C}$. The r.h.s. of (4.9) is analogous to the BIC and HQIC criteria considered in the model selection literature in which case $\zeta(x) = x$, $\kappa_n = (\ln n)^{1/2}$ for BIC, $\kappa_n = (Q \ln\ln n)^{1/2}$ for some $Q \geq 2$ for HQIC, and $\chi = 2$ (which holds for the functions S_1–S_3). Note that some calculations show that when $\widehat{\Omega}_n(\theta_0)$ is diagonal, $S = S_1$ or S_2, and $\zeta(x) = x$, the function $\varphi^{(5)}$ reduces to $\varphi^{(1)}$.

Returning now to the general case, given a choice of function $\varphi(\xi, \Omega) = (\varphi_1(\xi, \Omega), \ldots, \varphi_p(\xi, \Omega), 0, \ldots, 0)' \in R^p_{[+\infty]} \times \{0\}^v$, the replacement for the k-vector $(h_1, 0_v)$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is $\varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$. Thus, the GMS critical value, $\widehat{c}_n(\theta_0, 1-\alpha)$, is the $1-\alpha$ quantile of

$$(4.10)\qquad L_n(\theta_0, Z^*) = S\big(\widehat{\Omega}_n^{1/2}(\theta_0) Z^* + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_n(\theta_0)\big),$$

where $Z^* \sim N(0_k, I_k)$ and Z^* is independent of $\{W_i : i \geq 1\}$. That is,

$$(4.11)\qquad \widehat{c}_n(\theta_0, 1-\alpha) = \inf\big\{x \in R : P(L_n(\theta_0, Z^*) \leq x) \geq 1-\alpha\big\},$$

where $P(L_n(\theta_0, Z^*) \leq x)$ denotes the conditional df at x of $L_n(\theta_0, Z^*)$ given $(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$. One can compute $\widehat{c}_n(\theta_0, 1-\alpha)$ by simulating R i.i.d. random vectors $\{Z_r^* : r = 1, \ldots, R\}$ with $Z_r^* \sim N(0_k, I_k)$ and taking $\widehat{c}_n(\theta_0, 1-\alpha)$ to be the $1-\alpha$ sample quantile of $\{L_n(\theta_0, Z_r^*) : r = 1, \ldots, R\}$, where R is large.

A bootstrap version of the GMS critical value is obtained by replacing $L_n(\theta_0, Z^*)$ in (4.11) by

$$(4.12)\qquad S\big(M_n^*(\theta_0) + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_n^*(\theta_0)\big),$$

where $M_n^*(\theta)$ is a recentered bootstrapped version of $n^{1/2}\widehat{D}_n^{-1/2}(\theta)\overline{m}_n(\theta)$ and $\widehat{\Omega}_n^*(\theta)$ is a bootstrapped version of $\widehat{\Omega}_n(\theta)$ (defined as follows). Let $\{W_i^* : i \leq n\}$ be a bootstrap sample, such as a nonparametric i.i.d. bootstrap sample in an i.i.d. scenario or a block bootstrap sample in a time series scenario. By definition,

$$(4.13)\qquad M_n^*(\theta) = n^{1/2}(\widehat{D}_n^*(\theta))^{-1/2}(\overline{m}_n^*(\theta) - \overline{m}_n(\theta)),$$

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

$$\overline{m}_n^*(\theta) = n^{-1}\sum_{i=1}^{n} m(W_i^*, \theta), \quad \widehat{D}_n^*(\theta) = \mathrm{Diag}(\widehat{\Sigma}_n^*(\theta)),$$

and $\widehat{\Sigma}_n^*(\theta)$ is defined in the same manner as $\widehat{\Sigma}_n(\theta)$ is defined (e.g., as in (3.2) in the i.i.d. case) with W_i^* in place of W_i. One can compute the bootstrap critical value by simulating R bootstrap samples $\{\{W_{i,r}^* : i \le n\} : r = 1, \ldots, R\}$ (i.i.d. across samples), computing $\{(M_{n,r}^*(\theta_0), \widehat{\Omega}_{n,r}^*(\theta_0)) : r = 1, \ldots, R\}$ (defined as in (4.13)), and taking the bootstrap critical value to be the $1-\alpha$ sample quantile of $\{S(M_{n,r}^*(\theta_0) + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_{n,r}^*(\theta_0)) : r = 1, \ldots, R\}$, where R is large.

For the asymptotic results given below to hold with a bootstrap GMS critical value, one needs that $M_n^*(\theta_{n,h}) \to_d \Omega_0^{1/2} Z^*$ under certain triangular arrays of true distributions and true parameters $\theta_{n,h}$, where Ω_0 is a $k \times k$ correlation matrix and Z^* is as in (4.10).[8] This can be established for the nonparametric i.i.d. and block bootstraps using fairly standard arguments. For brevity, we do not do so here.

The 2003 working paper version of CHT discusses a bootstrap version of the GMS critical value based on $\varphi^{(1)}$ in the context of the interval outcome regression model. CHT mentions critical values of GMS type based on $\varphi^{(1)}$; see their Remark 4.5. Bugni (2007a, 2007b) and Canay (2007) provided results regarding the pointwise asymptotic null properties of nonparametric i.i.d. bootstrap procedures applied with $\varphi^{(1)}$ and $\varphi^{(3)}$, respectively. Note that GMS bootstrap critical values do not generate higher-order improvements in the present context because the asymptotic null distribution of the test statistic $T_n(\theta)$ is not asymptotically pivotal. Fan and Park (2007) considered a critical value based on a function that is analogous to $\varphi^{(1)}$ except with ∞ replaced by $\kappa_n \xi_j$ and with ξ_j replaced by $\xi_j \widehat{\sigma}_{n,j}(\theta)$. The latter makes their procedure lack invariance to the scaling of the moment functions, which is not desirable. If Fan and Park's (2007) φ function is altered to be scale invariant, then the test based on it has the same asymptotic properties under the null and local alternatives as the test based on $\varphi^{(1)}$ because $\kappa_n \to \infty$.

4.2. *Step-by-Step Calculation of GMS Tests and CIs*

Here we describe the steps in the calculation of a nominal level α GMS test of $H_0 : \theta = \theta_0$. First, we describe the bootstrap version of the GMS procedure based on $(S_2, \varphi^{(1)})$, which is the recommended procedure for i.i.d. observa-

[8]More specifically, this convergence must hold under any sequence of distributions $\{\gamma_n : n \ge 1\}$ defined just above (A.3) in the Appendix (in which case $\Omega_0 = \Omega_{h_{2,2}}$), the convergence needs to be joint with that in (A.3) of the Appendix, and the convergence must hold with $\{n\}$ replaced by any subsequence $\{w_n\}$ of sample sizes.

tions. (For non-i.i.d. observations, we recommend using $(S_2, \varphi^{(1)})$, but the asymptotic version may perform as well as a bootstrap version.)

Compute (i) the sample moments $\overline{m}_n(\theta_0)$, (ii) the sample variance estimator $\widehat{\Sigma}_n(\theta_0)$, and (iii) the test statistic $T_n(\theta_0) = S_2(n^{1/2}\overline{m}_n(\theta_0), \widehat{\Sigma}_n(\theta_0))$. Next, to determine the GMS critical value $\widehat{c}_n(\theta_0, 1-\alpha)$, use the following steps: (iv) simulate R bootstrap samples each of size n, i.i.d. across bootstrap samples, denoted $\{\{W_{i,r}^* : i \leq n\} : r = 1, \ldots, R\}$ (according to a bootstrap procedure that is suitable for the observations under consideration, such as a nonparametric i.i.d. bootstrap for i.i.d. observations and a block bootstrap for time series), where the number of bootstrap simulations R is large, say 1000 or more; (v) compute $\{(M_{n,r}^*(\theta_0), \widehat{\Omega}_{n,r}^*(\theta_0)) : r = 1, \ldots, R\}$ (defined in (4.13) with $W_{i,r}^*$ in place of W_i^*); (vi) determine whether $n^{1/2}\overline{m}_{n,j}(\theta_0)/\widehat{\sigma}_{n,j}(\theta_0) > \kappa_n = (\ln n)^{1/2}$ for $j = 1, \ldots, p$, where $\widehat{\sigma}_{n,j}^2(\theta_0)$ is the (j, j) element of $\widehat{\Sigma}_n(\theta_0)$; (vii) eliminate the elements in $(M_{n,r}^*(\theta_0), \widehat{\Omega}_{n,r}^*(\theta_0))$ for all $r = 1, \ldots, R$ that correspond to the moment conditions that satisfy the condition in (vi), with the resulting quantities denoted by $(M_{n,r}^{**}(\theta_0), \widehat{\Omega}_{n,r}^{**}(\theta_0))$ for $r = 1, \ldots, R$; (viii) take the critical value $\widehat{c}_n(\theta_0, 1-\alpha)$ to be the $1-\alpha$ sample quantile of $\{S_2(M_{n,r}^{**}(\theta_0), \widehat{\Omega}_{n,r}^{**}(\theta_0)) : r = 1, \ldots, R\}$. The GMS test rejects $H_0 : \theta = \theta_0$ if $T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)$.

A GMS CS is obtained by inverting tests of $H_0 : \theta = \theta_0$ for $\theta_0 \in \Theta$. A GMS CS can be calculated by employing a grid search (or some more sophisticated) algorithm using the method described above to calculate whether $T_n(\theta_0) \leq \widehat{c}_n(\theta_0, 1-\alpha)$, which implies that θ_0 should be included in the CS.

For a general choice of (S, φ), the asymptotic version of the GMS test is computed as follows. Compute (i) the sample moments $\overline{m}_n(\theta_0)$, (ii) the sample variance estimator $\widehat{\Sigma}_n(\theta_0)$, and (iii) the test statistic $T_n(\theta_0) = S(n^{1/2} \times \overline{m}_n(\theta_0), \widehat{\Sigma}_n(\theta_0))$. Next, to determine the critical value $\widehat{c}_n(\theta_0, 1-\alpha)$, compute (iv) $\widehat{\Omega}_n(\theta_0) = \mathrm{Diag}^{-1/2}(\widehat{\Sigma}_n(\theta_0))\widehat{\Sigma}_n(\theta_0)\,\mathrm{Diag}^{-1/2}(\widehat{\Sigma}_n(\theta_0))$, (v) $\xi_n(\theta_0) = \kappa_n^{-1} n^{1/2}\,\mathrm{Diag}^{-1/2}(\widehat{\Sigma}_n(\theta_0))\overline{m}_n(\theta_0)$, where $\kappa_n = (\ln n)^{1/2}$, and (vi) $\varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$; (vii) simulate R i.i.d. random vectors $\{Z_r^* : r = 1, \ldots, R\}$ with $Z_r^* \sim N(0_k, I_k)$, where R is large; and (viii) take $\widehat{c}_n(\theta_0, 1-\alpha)$ to be the $1-\alpha$ sample quantile of $\{S(\widehat{\Omega}_n^{1/2}(\theta_0)Z_r^* + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_n(\theta_0)) : r = 1, \ldots, R\}$. The GMS test rejects $H_0 : \theta = \theta_0$ if $T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)$.

The bootstrap version of the GMS test with general choice of (S, φ) replaces steps (vii) and (viii) with the following: (vii*) simulate R bootstrap samples each of size n, i.i.d. across bootstrap samples, denoted $\{\{W_{i,r}^* : i \leq n\} : r = 1, \ldots, R\}$, (viii*) compute $\{(M_{n,r}^*(\theta_0), \widehat{\Omega}_{n,r}^*(\theta_0)) : r = 1, \ldots, R\}$ (defined in (4.13) with $W_{i,r}^*$ in place of W_i^*), and (ix*) take the critical value $\widehat{c}_n(\theta_0, 1-\alpha)$ to be the $1-\alpha$ sample quantile of $\{S(M_{n,r}^*(\theta_0) + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_{n,r}^*(\theta_0)) : r = 1, \ldots, R\}$. When $S = S_2$ and $\varphi = \varphi^{(1)}$, this procedure is equivalent to that for the GMS test described three paragraphs above.

4.3. *Assumptions*

Next we state assumptions on the function φ and the constants $\{\kappa_n : n \geq 1\}$ that define a GMS procedure. The first two assumptions are used to show that GMS CS's and tests have correct asymptotic size.

ASSUMPTION GMS1: (a) *$\varphi_j(\xi, \Omega)$ is continuous at all $(\xi, \Omega) \in (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi$ with $\xi_j = 0$, where $\xi = (\xi_1, \ldots, \xi_k)'$, for $j = 1, \ldots, p$.*

(b) *$\varphi_j(\xi, \Omega) = 0$ for all $(\xi, \Omega) \in (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi$ with $\xi_j = 0$, where $\xi = (\xi_1, \ldots, \xi_k)'$, for $j = 1, \ldots, p$.*

(c) *$\varphi_j(\xi, \Omega) = 0$ for all $j = p+1, \ldots, k$ for all $(\xi, \Omega) \in (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi$.*

ASSUMPTION GMS2: $\kappa_n \to \infty$.

Assumptions GMS1 and GMS2 are not restrictive. For example, the functions $\varphi^{(1)}$–$\varphi^{(4)}$ satisfy Assumption GMS1 and $\kappa_n = (\ln n)^{1/2}$ satisfies Assumption GMS2. Assumption GMS1 also holds for $\varphi^{(5)}$ for all functions S that satisfy Assumption 1(d), which includes S_1–S_3; see the Supplemental Material (Andrews and Soares (2010)) for a proof.

The next two assumptions are used in conjunction with Assumptions GMS1 and GMS2 to show that GMS CS's and tests are not asymptotically conservative. They also are used to determine the formula for the asymptotic power of GMS tests against $n^{-1/2}$-local alternatives.

ASSUMPTION GMS3: *$\varphi_j(\xi, \Omega) \to \infty$ as $(\xi, \Omega) \to (\xi_*, \Omega_*)$ for all $(\xi_*, \Omega_*) \in R^p_{[+\infty]} \times R^v_{[\pm\infty]} \times \mathrm{cl}(\Psi)$ with $\xi_{*,j} = \infty$, where $\xi_* = (\xi_{*,1}, \ldots, \xi_{*,k})'$, for $j = 1, \ldots, p$.*

ASSUMPTION GMS4: $\kappa_n^{-1} n^{1/2} \to \infty$.

Assumptions GMS3 and GMS4 are not restrictive and are satisfied by $\varphi^{(1)}$–$\varphi^{(4)}$ and $\kappa_n = (\ln n)^{1/2}$. Assumption GMS3 also holds for $\varphi^{(5)}$ for all functions S that satisfy Assumption 1(d) and for which $S(-c \cdot \xi, \Omega) \to \infty$ as $(\xi, \Omega) \to (\xi_*, \Omega_*)$ whenever $c_j = 1$, see the Supplemental Material for a proof. The latter holds for the test functions S_1–S_3.

The next two assumptions are used in conjunction with Assumptions GMS2 and GMS3 to show that GMS tests dominate subsampling tests (based on a subsample size b) in terms of $n^{-1/2}$-local asymptotic power.

ASSUMPTION GMS5: *$\kappa_n^{-1}(n/b)^{1/2} \to \infty$, where $b = b_n$ is the subsample size.*

ASSUMPTION GMS6: *$\varphi_j(\xi, \Omega) \geq 0$ for all $(\xi, \Omega) \in (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi$ for $j = 1, \ldots, p$.*

Assumption GMS5 holds for all reasonable choices of κ_n and b. For example, for $\kappa_n = (\ln n)^{1/2}$, Assumption GMS5 holds for $b = n^\eta$ for any $\eta \in (0, 1)$. Any reasonable choice of b satisfies the latter condition. Note that for recentered subsampling tests, the optimal value of η in terms of size is $2/3$; see Bugni (2007a, 2007b). When $\eta = 2/3$, Assumption GMS5 holds easily for κ_n as above. Assumption GMS5 fails when $\kappa_n = (\ln n)^{1/2}$ (or some other logarithmic function) only if b is larger than $O(n^\eta)$ for all $\eta \in (0, 1)$ and the latter yields a recentered subsampling test whose error in the null rejection probability is very large—of order larger than $O(n^{-\varepsilon})$ for all $\varepsilon > 0$. For a non-recentered subsampling test, it yields a test whose power against $n^{1/2}$-local alternatives converges from below and very slowly to its asymptotic local power. The reason is that if b is larger than $O(n^\eta)$ for all $\eta \in (0, 1)$, then under the alternative, the subsampling critical value mimics the $1 - \alpha$ quantile of the alternative distribution of the test statistic unless n is very, very large, because the subsample size is almost equal to that of the full-sample statistic unless n is very, very large. Hence, the finite-sample power of the subsampling test is poor in sample sizes that are of interest in practice.

Assumption GMS6 is satisfied by the functions $\varphi^{(1)}$–$\varphi^{(5)}$ except for $\varphi^{(4)}$. Hence, it is slightly restrictive.

The last assumption is used to show that GMS tests are consistent against alternatives that are more distant from the null than $n^{-1/2}$-local alternatives.

ASSUMPTION GMS7: *$\varphi_j(\xi, \Omega) \geq \min\{\xi_j, 0\}$ for all $(\xi, \Omega) \in (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi$ for $j = 1, \dots, p$.*

Assumption GMS7 is not restrictive. For example, it is satisfied by $\varphi^{(1)}$–$\varphi^{(5)}$.

Next we introduce a condition that depends on the model, not on the GMS method, and is only used when showing that GMS CS's have AsyMaxCP $= 1$ when $v = 0$.

ASSUMPTION M: *For some $(\theta, F) \in \mathcal{F}$, $E_F m_j(W_i, \theta) > 0$ for all $j = 1, \dots, p$.*

Assumption M typically holds if the identified set (i.e., the set of parameter values θ that satisfy the population moment inequalities and equalities under F) has a nonempty interior for some data-generating process included in the model.

4.4. *Asymptotic Size Results*

The following theorem applies to i.i.d. observations, in which case $\mathcal{F}$ is as defined in (2.2), and to dependent observations, in which case for brevity $\mathcal{F}$ is as defined in (A.2) and (A.3) in the Appendix.

THEOREM 1: *Suppose Assumptions 1–3, GMS1, and GMS2 hold and $0 < \alpha < 1/2$. Then the nominal level $1 - \alpha$ GMS CS based on $T_n(\theta)$ satisfies the following statements*:

(a) AsyCS $\geq 1-\alpha$.

(b) AsyCS $= 1-\alpha$ *if Assumptions* GMS3, GMS4, *and* 7 *also hold*.

(c) AsyMaxCP $= 1$ *if* $v = 0$ (*i.e.*, *no moment equalities appear*) *and Assumption* M *also holds*.

COMMENTS: (i) Theorem 1(a) shows that a GMS CS is asymptotically valid in a uniform sense. Theorem 1(b) shows it is not asymptotically conservative. Theorem 1(c) shows it is not asymptotically similar.

(ii) Theorem 1 places no assumptions on the moment functions $m(W_i, \theta)$ beyond the existence of mild moment conditions that appear in the definition of $\mathcal{F}$. Thus, the results apply to moment conditions based on instruments that are weak. (The reason is that the test statistics considered are of the Anderson–Rubin type.)

(iii) Theorem 1 holds even when there are restrictions on the moment inequalities such that when one moment inequality holds as an equality, then another moment inequality cannot. Restrictions of this sort arise in some models, such as models with interval outcomes (e.g., see Rosen (2008)).

(iv) The proof of Theorem 1 and all other results below are given in the Supplemental Material (Andrews and Soares (2010)).

5. GMS MODEL SPECIFICATION TESTS

Tests of model specification can be constructed using the GMS CS introduced above. The null hypothesis of interest is that (2.1) holds for some parameter $\theta_0 \in \Theta$ (with additional conditions imposed by the parameter space for (θ, F)). By definition, the GMS test rejects the model specification if $T_n(\theta)$ exceeds the GMS critical value $\widehat{c}_n(\theta, 1-\alpha)$ for all $\theta \in \Theta$. Equivalently, it rejects if the GMS CS is empty. The idea behind such a test is the same as for the J test of overidentifying restrictions in GMM; see Hansen (1982).

When the model of (2.1) is correctly specified, the GMS CS includes the true value with asymptotic probability $1-\alpha$ (or greater) uniformly over the parameter space. Thus, under the null hypothesis of correct model specification, the limit as $n \to \infty$ of the finite-sample size of the GMS model specification test is less than or equal to α under the assumptions of Theorem 1(a). In other words, the asymptotic size of this specification test is valid uniformly over the parameter space.

Note that the asymptotic size of the GMS model specification test is not necessarily equal to α under the assumptions of Theorem 1(b).[9] That is, the GMS model specification test may be asymptotically conservative.

[9]The reason is that when the null of correct model specification holds and (θ_0, F_0) is the truth, the GMS test may fail to reject the null even when $T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)$ because $T_n(\theta) \leq \widehat{c}_n(\theta, 1-\alpha)$ for some $\theta \neq \theta_0$.

6. SUBSAMPLING CONFIDENCE SETS

The volume of a CS is directly related to the power of the tests used in its construction. Below we compare the power of GMS tests to that of subsampling and PA tests. In this section and the following one we define subsampling and PA CS's.

We now define subsampling critical values and CS's. Let $b = b_n$ denote the subsample size when the full-sample size is n. We assume $b \to \infty$ and $b/n \to 0$ as $n \to \infty$ (here and below). The number of subsamples of size b considered is q_n. With i.i.d. observations, there are $q_n = n!/((n-b)!b!)$ subsamples of size b. With time series observations, there are $q_n = n - b + 1$ subsamples, each based on b consecutive observations.

Let $T_{n,b,j}(\theta)$ be a subsample statistic defined exactly as $T_n(\theta)$ is defined but based on the jth subsample of size b rather than the full sample for $j = 1, \ldots, q_n$. The empirical df and the $1-\alpha$ sample quantile of $\{T_{n,b,j}(\theta) : j = 1, \ldots, q_n\}$ are

$$U_{n,b}(\theta, x) = q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta) \leq x) \quad \text{for} \quad x \in R, \tag{6.1}$$

$$c_{n,b}(\theta, 1-\alpha) = \inf\{x \in R : U_{n,b}(\theta, x) \geq 1 - \alpha\}.$$

The subsampling test rejects $H_0 : \theta = \theta_0$ if $T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha)$. The nominal level $1-\alpha$ subsampling CS is given by (2.3) with $c_{1-\alpha}(\theta) = c_{n,b}(\theta, 1-\alpha)$.

One also can define "recentered" subsample statistics by defining $T_{n,b,j}(\theta)$ using $b^{1/2}(\overline{m}_{n,b,j}(\theta) - \overline{m}_n(\theta))$, rather than $b^{1/2}\overline{m}_{n,b,j}(\theta)$, in place of $n^{1/2}\overline{m}_n(\theta)$ in (3.3), where $\overline{m}_{n,b,j}(\theta)$ is the average of the moment conditions over the observations in the jth subsample; see AG4.

It is shown in AG4 that under Assumptions 1–3 and $0 < \alpha < 1/2$, the nominal level $1-\alpha$ subsampling CS based on $T_n(\theta)$ satisfies (a) AsyCS $\geq 1-\alpha$, (b) AsyCS $= 1-\alpha$ if Assumption 7 also holds, and (c) AsyMaxCP $= 1$ if $v = 0$ (i.e., no moment equalities appear) and Assumption M also holds.

7. PLUG-IN ASYMPTOTIC CONFIDENCE SETS

Now we discuss CS's based on a PA critical value. The least favorable asymptotic null distributions of the statistic $T_n(\theta)$ are shown in AG4 to be those for which the moment inequalities hold as equalities. These distributions depend on the correlation matrix Ω of the moment functions. We analyze plug-in asymptotic (PA) critical values that are determined by the least favorable asymptotic null distribution for given Ω evaluated at a consistent estimator of Ω. Such critical values have been considered for many years in the literature on multivariate one-sided tests; see Silvapulle and Sen (2005) for references. AG4 considered them in the context of the moment inequality literature. Rosen (2008) considered 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.

Let $c(\Omega, 1-\alpha)$ denote the $1-\alpha$ quantile of $S(Z, \Omega)$, where $Z \sim N(0_k, \Omega)$. This is the $1-\alpha$ quantile of the asymptotic null distribution of $T_n(\theta)$ when the moment inequalities hold as equalities.

The nominal $1-\alpha$ PA CS is given by (2.3) with critical value $c_{1-\alpha}(\theta)$ equal to

$$(7.1) \qquad c(\widehat{\Omega}_n(\theta), 1-\alpha).$$

AG4 showed that if Assumptions 1 and 4 hold and $0 < \alpha < 1/2$, then the nominal level $1-\alpha$ PA CS based on $T_n(\theta)$ satisfies AsyCS $\geq 1-\alpha$.

8. LOCAL ALTERNATIVE POWER COMPARISONS

In this section and the next, we compare the power of GMS, subsampling, and PA tests. These results have immediate implications for the volume of CS's 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. Here we analyze the power of tests against $n^{-1/2}$-local alternatives. In the next section we consider "distant alternatives," which differ from the null by more than $O(n^{-1/2})$ and may be fixed or local.

We show that a GMS test has asymptotic power that is greater than or equal to that of a subsampling or PA test (based on the same test statistic) under all alternatives. We show that a GMS test's power is *strictly greater* than that of a subsampling test in the scenario stated in the Introduction. In addition, we show that GMS and subsampling tests have asymptotic power that is greater than or equal to that of a PA test with strictly greater power in the scenarios stated in the Introduction.

For given $\theta_{n,*}$, we consider tests of

$$(8.1) \qquad H_0 : E_{F_n} m_j(W_i, \theta_{n,*}) \begin{cases} \geq 0 & \text{for} \quad j = 1, \dots, p, \\ = 0 & \text{for} \quad j = p+1, \dots, k, \end{cases}$$

where F_n denotes the true distribution of the data, versus $H_1 : H_0$ does not hold. For brevity, we only give results for the case of i.i.d. observations. (The results can be extended to dependent observations, and the advantage of GMS tests over subsampling and PA tests also holds with dependent observations.) The parameter space $\mathcal{F}$ for (θ, F) is assumed to satisfy (2.2).

With i.i.d. observations, F denotes the distribution of W_i. We consider the Kolmogorov–Smirnov metric on the space of distributions F. Let

$$(8.2) \qquad D(\theta, F) = \text{Diag}\{\sigma^2_{F,1}(\theta), \dots, \sigma^2_{F,k}(\theta)\}, \quad \Omega(\theta, F) = \text{Corr}_F(m(W_i, \theta)).$$

We now introduce the $n^{-1/2}$-local alternatives that are considered.

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

(a) $\theta_n = \theta_{n,*} - \lambda n^{-1/2}(1 + o(1))$ *for some* $\lambda \in R^d$, $\theta_{n,*} \to \theta_0$, *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* $j = 1, \ldots, p$.

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

Assumption LA1(a) specifies that the true values $\{\theta_n : n \geq 1\}$ are local to the null values $\{\theta_{n,*} : n \geq 1\}$. Assumption LA1(b) specifies the asymptotic behavior of the (normalized) moment inequality functions when evaluated at the true parameter values $\{\theta_n : n \geq 1\}$. Under the true values, these (normalized) moment inequalities are nonnegative. Assumption LA1(a) and (c) imply that $\Omega(\theta_{n,*}, F_n)$ exists and $\Omega(\theta_{n,*}, F_n) \to \Omega_0 = \Omega(\theta_0, F_0)$.

The asymptotic distribution of the test statistic $T_n(\theta_{n,*})$ under $n^{-1/2}$-local alternatives depends on the limit of the (normalized) moment inequality functions when evaluated at the null value $\theta_{n,*}$ because $T_n(\theta_{n,*})$ is evaluated at $\theta_{n,*}$. Under Assumptions LA1 and LA2, we show that

$$(8.3) \qquad \lim_{n\to\infty} n^{1/2} D^{-1/2}(\theta_{n,*}, F_n) E_{F_n} m(W_i, \theta_{n,*}) = (h_1, 0_v) + \Pi_0 \lambda \in R^k, \quad \text{where}$$
$$h_1 = (h_{1,1}, \ldots, h_{1,p})' \quad \text{and} \quad \Pi_0 = \Pi(\theta_0, F_0).$$

By definition, if $h_{1,j} = \infty$, then $h_{1,j} + y = \infty$ for any $y \in R$. Let $\Pi_{0,j}$ denote the jth row of Π_0 written as a column d-vector for $j = 1, \ldots, k$. Note that $(h_1, 0_v) + \Pi_0 \lambda \in R^p_{[+\infty]} \times R^v$.

The following assumption states that the true distribution of the data F_n is in the alternative, not the null (for n large).

ASSUMPTION LA3: $h_{1,j} + \Pi'_{0,j}\lambda < 0$ *for some* $j = 1, \ldots, p$ *or* $\Pi'_{0,j}\lambda \neq 0$ *for some* $j = p+1, \ldots, k$.

The following is a simple example to illustrate Assumptions LA1–LA3. Suppose $m(W_i, \theta) = W_i - \theta$, $E_F m(W_i, \theta) \geq 0$, and $\text{Var}_F(m(W_i, \theta)) = 1$ for all $(\theta, F) \in \mathcal{F}$. Then $p = 1$, $v = 0$, and $D(\theta, F) = 1$. Consider a sequence of true parameters/distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ that satisfy $\theta_n = \theta_{n,*} - \lambda n^{-1/2}$, $E_{F_n} W_i = \theta_n + h_1 n^{-1/2}$ for some $\theta_{n,*}, \lambda \in R$, and $h_1 \geq 0$, and $\theta_{n,*} \to \theta_0$. Then Assumption LA1(a) holds and in Assumption LA1(b), we have $n^{1/2} E_{F_n} m_j(W_i, \theta_n)/\sigma_{F_n,j}(\theta_n) = n^{1/2}(E_{F_n} W_i - \theta_n) = h_1 \geq 0$ for all n (using $\sigma_{F_n,j}(\theta_n) = 1$). So, Assumption LA1(b) also holds. We have $\Pi(\theta, F) = -1$ for all (θ, F). Hence, in Assumption LA3, $h_1 + \Pi_0 \lambda = h_1 - \lambda$, which is negative whenever $\lambda > h_1$. Hence, if the null value $\theta_{n,*}$ deviates from the true value θ_n

by enough (i.e., if $\theta_{n,*} - \theta_n = \lambda n^{-1/2}$ is large enough) relative to the magnitude of the slackness of the moment condition (i.e., $E_{F_n} W_i - \theta_n = h_1 n^{-1/2}$), then the null hypothesis is violated for all n and Assumption LA3 holds.

The asymptotic distribution of $T_n(\theta_{n,*})$ under $n^{-1/2}$-local alternatives is shown to be $J_{h_1,\lambda}$, where $J_{h_1,\lambda}$ is defined by

$$(8.4) \qquad S\big(\Omega_0^{1/2} Z^* + (h_1, 0_v) + \Pi_0 \lambda, \Omega_0\big) \sim J_{h_1,\lambda}$$

for $Z^* \sim N(0_k, I_k)$. For notational simplicity, the dependence of $J_{h_1,\lambda}$ on Ω_0 and Π_0 is suppressed. Let $c_{h_1,\lambda}(1-\alpha)$ denote the $1-\alpha$ quantile of $J_{h_1,\lambda}$.

We now introduce two assumptions that are used for GMS tests only.

ASSUMPTION LA4: *$\kappa_n^{-1} n^{1/2} E_{F_n} m_j(W_i, \theta_n)/\sigma_{F_n,j}(\theta_n) \to \pi_{1,j}$ for some $\pi_{1,j} \in R_{+,\infty}$ for $j = 1, \ldots, p$.*

Note that in Assumption LA4, the functions are evaluated at the true value θ_n, not at the null value $\theta_{n,*}$, and $(\theta_n, F_n) \in \mathcal{F}$. In consequence, the moment functions in Assumption LA4 satisfy the inequalities and $\pi_{1,j} \geq 0$ for all $j = 1, \ldots, p$.

Let $\pi_1 = (\pi_{1,1}, \ldots, \pi_{1,p})'$. Let $c_{\pi_1}(\varphi, 1-\alpha)$ denote the $1-\alpha$ quantile of

$$(8.5) \qquad S\big(\Omega_0^{1/2} Z^* + \varphi((\pi_1, 0_v), \Omega_0), \Omega_0\big), \quad \text{where} \quad Z^* \sim N(0_k, I_k).$$

Below the probability limit of the GMS critical value, $\widehat{c}_n(\theta_{n,*}, 1-\alpha)$ is shown to be $c_{\pi_1}(\varphi, 1-\alpha)$.

The following assumption is used to obtain the $n^{-1/2}$-local alternative power function of the GMS test. Let $C(\varphi) = \{\widetilde{\pi}_1 = (\widetilde{\pi}_{1,1}, \ldots, \widetilde{\pi}_{1,p})' \in R^p_{[+\infty]} : \text{for } j = 1, \ldots, p, \widetilde{\pi}_{1,j} = \infty \text{ or } \varphi_j(\xi, \Omega) \to \varphi_j((\widetilde{\pi}_1, 0_v), \Omega_0) \text{ as } (\xi, \Omega) \to ((\widetilde{\pi}_1, 0_v), \Omega_0)\}$. Roughly speaking, $C(\varphi)$ is the set of $\widetilde{\pi}_1$ vectors for which φ is continuous at $((\widetilde{\pi}_1, 0_v), \Omega_0)$. For example, $C(\varphi^{(1)}) = \{\widetilde{\pi}_1 \in R^p_{[+\infty]} : \widetilde{\pi}_{1,j} \neq 1 \text{ for } j = 1, \ldots, p\}$, $C(\varphi^{(2)}) = R^p_{[+\infty]}$ provided ψ is continuous on $[a_L, a_U]$ (where continuity at a_U means that $\lim_{x \to a_U} \psi(x) = \infty$), $C(\varphi^{(3)}) = R^p_{[+\infty]}$, $C(\varphi^{(4)}) = R^p_{[+\infty]}$, and $C(\varphi^{(5)}) = \{\pi_1 \in R^p_{[+\infty]} : S(-c \cdot (\widetilde{\pi}_1, 0_v), \Omega_0) - \zeta(|c|) \text{ has a unique minimum over } c \in \mathcal{C}\}$.

ASSUMPTION LA5: (a) $\pi_1 \in C(\varphi)$.

(b) *The df of $S(\Omega_0^{1/2} Z^* + \varphi((\pi_1, 0_v), \Omega_0), \Omega_0)$ is continuous and strictly increasing at $x = c_{\pi_1}(\varphi, 1-\alpha)$.*

Assumption LA5(a) implies that the $n^{-1/2}$-local power formulae given below do not apply to certain "discontinuity vectors" $\pi_1 = (\pi_{1,1}, \ldots, \pi_{1,p})'$. However, this does not affect the power comparisons between GMS, subsampling, and PA tests, because Assumption LA5 is not needed for those results. The power comparisons hold for all π_1 vectors.

We now introduce an assumption that is used for subsampling tests only.

ASSUMPTION LA6: *$b^{1/2}E_{F_n}m_j(W_i, \theta_n)/\sigma_{F_n,j}(\theta_n) \to g_{1,j}$ for some $g_{1,j} \in R_{+,\infty}$ for $j = 1, \ldots, p$.*

Assumption LA6 is not restrictive. It specifies the limit of the (normalized) moment inequality functions when evaluated at the true parameter values $\{\theta_n : n \geq 1\}$ and when scaled by the square root of the subsample size $b^{1/2}$.

Define $g_1 = (g_{1,1}, \ldots, g_{1,p})'$. Note that $0_p \leq g_1 \leq \pi_1 \leq h_1$.[10] The probability limit of the subsampling critical value is shown to depend on

$$(8.6) \qquad \lim_{n\to\infty} b^{1/2}D^{-1/2}(\theta_{n,*}, F_n)E_{F_n}m(W_i, \theta_{n,*}) = (g_1, 0_v) \in R^k_{+,\infty}.$$

Note that $(g_1, 0_v) \in R^p_{+,\infty} \times \{0_v\}$. Thus, elements of $(g_1, 0_v)$ are necessarily nonnegative. The probability limit of the subsampling critical value is shown to be $c_{g_1,0_d}(1-\alpha)$, which denotes the $1-\alpha$ quantile of $J_{g_1,0_d}$ (where $J_{g_1,0_d}$ equals $J_{h_1,\lambda}$ with $h_1 = g_1$ and $\lambda = 0$). The probability limit of the PA critical value is shown to be $c_{0_p,0_d}(1-\alpha)$, which is the $1-\alpha$ quantile of $J_{0_p,0_d}$ (and also can be written as $c(\Omega_0, 1-\alpha)$ using the notation introduced just above (7.1)).

THEOREM 2: *Under Assumptions 1–5, LA1, and LA2, the following statements hold*:

(a) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > \widehat{c}_n(\theta_{n,*}, 1-\alpha)) = 1 - J_{h_1,\lambda}(c_{\pi_1}(\varphi, 1-\alpha))$ *provided Assumptions GMS2, GMS3, LA4, and LA5 hold*.

(b) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c_{n,b}(\theta_{n,*}, 1-\alpha)) = 1 - J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha))$ *provided Assumption LA6 holds*.

(c) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(\widehat{\Omega}_n(\theta_{n,*}), 1-\alpha)) = 1 - J_{h_1,\lambda}(c_{0_p,0_d}(1-\alpha))$.

COMMENTS: (i) Theorem 2(a) provides the $n^{-1/2}$-local alternative power function of the GMS test. The probability limit of the GMS critical value $\widehat{c}_n(\theta_{n,*}, 1-\alpha)$ under $n^{-1/2}$-local alternatives is $c_{\pi_1}(\varphi, 1-\alpha)$. Theorem 2(b) and (c) provide the $n^{-1/2}$-local alternative power function of the subsampling and PA tests.

(ii) The results of Theorem 2 hold under the null hypothesis as well as under the alternative. The results under the null quantify the degree of asymptotic nonsimilarity of the GMS, subsampling, and PA tests.

The next result provides power comparisons of GMS, subsampling, and PA tests.

THEOREM 3: *Under Assumptions 1–5, LA1–LA4, LA6, GMS2, GMS3, GMS5, and GMS6, the following statements hold*:

(a) $\liminf_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > \widehat{c}_n(\theta_{n,*}, 1-\alpha)) \geq \lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c_{n,b}(\theta_{n,*}, 1-\alpha))$ *with strict inequality whenever $g_{1,j} < \infty$ and $\pi_{1,j} = \infty$ for some $j = 1, \ldots, p$ and $c_{g_1,0_d}(1-\alpha) > 0$.*

[10]This holds by condition (ii) of (2.2) (since $(\theta_n, F_n) \in \mathcal{F}$), Assumptions LA1(b), LA6, and GMS5, and $b/n \to 0$.

(b) $\liminf_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > \widehat{c}_n(\theta_{n,*}, 1-\alpha)) \geq \lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(\widehat{\Omega}_n(\theta_{n,*}), 1-\alpha))$ *with strict inequality whenever* $\pi_{1,j} = \infty$ *for some* $j = 1, \ldots, p$.

(c) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c_{n,b}(\theta_{n,*}, 1-\alpha)) \geq \lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(\widehat{\Omega}_n(\theta_{n,*}), 1-\alpha))$ *with strict inequality whenever* $g_1 > 0_p$, *where Assumptions* GMS2, GMS3, GMS5, GMS6, *and* LA4 *are not needed for this result*.

COMMENTS: (i) Theorem 3(a) and (b) show that a GMS test based on a given test statistic has asymptotic power greater than or equal to that of subsampling and PA tests based on the same test statistic. For GMS versus subsampling tests, the inequality is strict whenever one or more moment inequality is satisfied and has a magnitude that is $o(b^{-1/2})$, and is larger than $O(\kappa_n n^{-1/2})$ and $c_{g_1,0_d}(1-\alpha) > 0$.[11] For GMS versus PA tests, the inequality is strict whenever one or more moment inequality is satisfied and has a magnitude that is larger than $O(\kappa_n n^{-1/2})$.

The reason the GMS test has higher power in these cases is that its (data-dependent) critical value is smaller asymptotically than the subsampling and PA critical values. It is smaller because when some moment inequality is satisfied under the alternative and is sufficiently far from being an equality (specifically, is larger than $O(\kappa_n n^{-1/2})$), then the GMS critical value takes this into account and delivers a critical value that is suitable for the case where this moment inequality is omitted. On the other hand, in the scenarios specified, the subsampling critical value does not take this into account, and in all scenarios the PA critical value is based on the least favorable distribution (for given Ω_0) which occurs when all moment inequalities hold as equalities.

(ii) Theorem 3(c) shows that the subsampling test has asymptotic power greater than or equal to that of the PA test for all local alternatives and is more powerful asymptotically than the PA test for many local alternatives. The reason is that when some moment inequality is satisfied under the alternative and is sufficiently far from being an equality (specifically, is larger than $o(b^{-1/2})$), then the subsampling critical value automatically takes this (at least partially) into account and delivers a smaller critical value than the PA critical value.

(iii) The comparison of the power of GMS tests and subsampling tests given in Theorem 3(a) does not impose Assumption LA5. Hence, the comparison holds for all $n^{-1/2}$-local alternatives.

(iv) We now show that the difference in power between the GMS test and the subsampling and PA tests can be quite large. Suppose there are no equality constraints (i.e., $v = 0$) and the distribution considered is such that the first inequality constraint may or may not be violated, but the other $j = 2, \ldots, p$ inequality constraints are not violated and differ from being equalities by mag-

[11] For most test functions S, $c_{g_1,0_d}(1-\alpha) > 0$ whenever one or more of the moment inequalities is violated asymptotically, so the latter condition holds under local alternatives.

nitudes that are $o(b^{-1/2})$ and are larger than $O(\kappa_n n^{-1/2})$. In this case, $g_{1,j}=0$, $h_{1,j}=\pi_{1,j}=\infty$, and $h_{1,j}+\Pi'_{0,j}\lambda=\infty$ for $j=2,\ldots,p$. Let $\mu_1=h_{1,1}+\Pi'_{0,1}\lambda$. If $\mu_1\in(-\infty,0)$, the first inequality constraint is violated asymptotically. If the null hypothesis is true (for all n large), then $\theta_n=\theta_{n,*}$, $\lambda=0$, and $\mu_1=h_{1,1}\geq 0$. Since $|\mu_1|<\infty$, we have $|h_{1,1}|<\infty$ and $g_{1,1}=0$. Thus, $g_1=0_p$. For simplicity, suppose $\Omega_0=I_p$. In this case, the asymptotic powers of the tests based on the functions S_1 and S_2 are the same, so we consider the S_1 test statistic. The asymptotic distribution $J_{h_1,\lambda}$ in this case is the distribution of

$$\sum_{j=1}^{p}[Z_j^*+h_{1,j}+\Pi'_{0,j}\lambda]_-^2=[Z_1^*+\mu_1]_-^2, \tag{8.7}$$

where $Z^*=(Z_1^*,\ldots,Z_p^*)'\sim N(0_p,I_p)$, because $Z_j^*+\infty=\infty$ for $j=2,\ldots,p$.

The probability limit of the GMS critical value, $c_{\pi_1}(\varphi,1-\alpha)$, is the $1-\alpha$ quantile of $[Z_1^*]_-^2$ which equals $z_{1-\alpha}^2$, where $z_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal distribution. This holds using (8.5) because $\pi_{1,1}=0$ and Assumption GMS1(b) imply that $\varphi_1((\pi_1,0_v),\Omega_0)=0$, and for $j=2,\ldots,p$, $\pi_{1,j}=\infty$ and Assumption GMS3 imply that $\varphi_j((\pi_1,0_v),\Omega_0)=\infty$. On the other hand, $J_{g_1,0_d}=J_{0_p,0_d}$ is the distribution of $\sum_{j=1}^{p}[Z_j^*]_-^2$. Hence, the probability limit of the subsampling and PA critical values, $c_{0_p,0_d}(1-\alpha)$, is the $1-\alpha$ quantile of $\sum_{j=1}^{p}[Z_j^*]_-^2$, call it $z_\alpha(p)$. Clearly, $z_\alpha(1)=z_{1-\alpha}^2$, $z_\alpha(p)>z_{1-\alpha}^2$ for $p\geq 2$, and the difference is strictly increasing in p.

Table I provides the value of $z_\alpha(p)$ for $\alpha=.05$ and several values of p. One sees that the critical value of the subsampling and PA tests increases substantially as the number of nonviolated moment inequalities, $p-1$, increases. Just one nonviolated moment inequality (i.e., $p=2$) increases the critical value from 2.71 to 4.25.

TABLE I

ASYMPTOTIC CRITICAL VALUES AND POWER OF THE NOMINAL .05 GMS TEST COMPARED TO SUBSAMPLING AND PA TESTS

			μ_1			
	p	Critical Values $z_\alpha(p)$	Asy. Null Rej. Prob. .00	Asy. Power −1.645	Asy. Power −2.170	Asy. Power −2.930
GMS test	All p	2.71	.050	.50	.70	.90
Sub & PA tests	2	4.25	.020	.34	.54	.81
	3	5.43	.010	.25	.44	.73
	4	6.34	.005	.18	.35	.65
	5	7.49	.003	.14	.29	.58
	10	11.83	.000	.04	.10	.31
	20	19.28	.000	.00	.01	.07

By Theorem 2, the asymptotic powers of the GMS, subsampling, and PA tests in the present scenario are

$$\begin{aligned}(8.8)\qquad &\mathrm{AsyPow}_{\mathrm{GMS}}(\mu_1) = P([Z_1^* + \mu_1]_-^2 > z_{1-\alpha}^2) = \Phi(-\mu_1 - z_{1-\alpha}),\\ &\mathrm{AsyPow}_{\mathrm{Sub}}(\mu_1) = \mathrm{AsyPow}_{\mathrm{PA}}(\mu_1)\\ &\qquad = P([Z_1^* + \mu_1]_-^2 > z_\alpha(p)) = \Phi\big(-\mu_1 - z_\alpha^{1/2}(p)\big),\end{aligned}$$

respectively. Table I reports the asymptotic power of the GMS test and the subsampling and PA tests, where the power of the latter depends on p, for four values of μ_1. (Note that the asymptotic size of each test is α, so that no asymptotic size correction is needed before comparing the asymptotic powers of the tests.) The first value of μ_1 is zero, which corresponds to a distribution in the null hypothesis. In this case, the asymptotic rejection rate of the GMS test is precisely .05, while that of the subsampling and PA tests is much less than .05 due to the asymptotic nonsimilarity of these tests. The last three values of μ_1 are negative, which correspond to distributions in the alternative. Table I shows that the power of the GMS test is substantially higher than that of the subsampling and PA tests even when $p = 2$ and the difference increases with p.

(v) The difference in powers of the subsampling and PA tests can be as large as the differences illustrated in Table I between GMS and PA tests. Consider the same scenario as in comment (iv) except that the $j = 2, \ldots, p$ inequality constraints differ from being equalities by a magnitude that is greater than $O(b^{-1/2})$. In this case, $g_{1,j} = \infty$ for $j = 2, \ldots, p$ and $J_{g_1,0_d}$ is the distribution of $[Z_1^*]_-^2$ because $g_1 = (0, \infty, \ldots, \infty)'$. Hence, the probability limit of the subsampling critical value, $c_{g_1,0_d}(1-\alpha)$, equals that of the GMS critical value and $\mathrm{AsyPow}_{\mathrm{Sub}}(\mu_1) = \mathrm{AsyPow}_{\mathrm{GMS}}(\mu_1)$. Everything else is the same as in comment (iv). Hence, in the present scenario, Table I applies but with the results for the subsampling test given by those of the GMS test.

(vi) The GMS, subsampling, and PA tests are not asymptotically unbiased. That is, there exist local alternatives for which the asymptotic rejection probabilities of the tests, namely $1 - J_{h_1,\lambda}(c_{\pi_1}(\varphi, 1-\alpha))$, $1 - J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha))$, and $1 - J_{h_1,\lambda}(c_{0_p,0_d}(1-\alpha))$, respectively, are less than α (e.g., see Table I with $p = 10$ or 20). This occurs because these tests are not asymptotically similar on the boundary of the null hypothesis. Lack of asymptotic unbiasedness is a common feature of tests of multivariate one-sided hypotheses, so this property of GMS, subsampling, and PA tests in the moment inequality example is not surprising.

(vii) Rosen (2008) introduced a critical value method that is a variant of the PA critical value. His method has the advantage of being simple computationally. However, it sacrifices power relative to GMS critical values in two respects. First, an upper bound on the $1-\alpha$ quantile of the asymptotic null distribution is employed. Second, in models in which some moment inequality can be slack without another being binding, his procedure yields larger critical values than GMS critical values because it does not use the data to detect slack inequalities. His procedure only adjusts for slack moment inequalities when it is known that if some inequality is binding, then some other necessarily cannot be.

(viii) For moment conditions based on weak instruments, the results of Theorem 2 still hold. But, the power comparisons of Theorem 3 do not because $\Pi_{0,j}'\lambda = 0$ for all $j = 1, \ldots, k$ in this case and so Assumption LA3 does not hold. With weak instruments, all of the tests have power less than or equal to α against $n^{-1/2}$-local alternatives, as is expected.

9. POWER AGAINST DISTANT ALTERNATIVES

Next we consider power against alternatives that are more distant from the null than $n^{-1/2}$-local alternatives. For all such alternatives, the powers of GMS, subsampling, and PA tests are shown to converge to 1 as $n \to \infty$. Thus, all three tests are consistent tests.

The following assumption specifies the properties of "distant alternatives" (DA), which include fixed alternatives and local alternatives that deviate from the null hypothesis by more than $O(n^{-1/2})$. Define

$$\begin{aligned}(9.1)\qquad & m_{n,j}^* = E_{F_n} m_j(W_i, \theta_{n,*})/\sigma_{F_n,j}(\theta_{n,*}),\\ & \beta_n = \max\{-m_{n,1}^*, \ldots, -m_{n,p}^*, |m_{n,p+1}^*|, \ldots, |m_{n,k}^*|\}.\end{aligned}$$

ASSUMPTION DA: (a) $n^{1/2}\beta_n \to \infty$.
(b) $\Omega(\theta_{n,*}, F_n) \to \Omega_1$ *for some* $k \times k$ *correlation matrix* $\Omega_1 \in \Psi$.

Assumption DA(a) requires that some moment inequality term $m_{n,j}^*$ violates the nonnegativity condition and is not $o(n^{-1/2})$ for $j = 1, \ldots, p$ or some moment equality term $m_{n,j}^*$ violates the zero condition and is not $o(n^{-1/2})$ for $j = p+1, \ldots, k$. In contrast to Assumption DA, under Assumptions LA1–LA3 above, $n^{1/2}\beta_n \to \max\{-h_{1,1} - \Pi_{0,1}'\lambda, \ldots, -h_{1,p} - \Pi_{0,p}'\lambda, |\Pi_{0,p+1}'\lambda|, \ldots, |\Pi_{0,k}'\lambda|\} < \infty$.

As in Section 8, we consider i.i.d. observations and $\mathcal{F}$ satisfies (2.2).

THEOREM 4: *Under Assumptions* 1, 3, 6, *and* DA, *we can make the following assertions*:

(a) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > \widehat{c}_n(\theta_{n,*}, 1-\alpha)) = 1$ *provided Assumption* GMS7 *holds*.

(b) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c_{n,b}(\theta_{n,*}, 1-\alpha)) = 1$.

(c) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(\widehat{\Omega}_n(\theta_{n,*}), 1-\alpha)) = 1$.

COMMENT: Theorem 4 shows that GMS, subsampling, and PA tests are consistent against all fixed alternatives and all non-$n^{-1/2}$-local alternatives.

10. EXTENSIONS

10.1. *Generalized Empirical Likelihood Statistics*

We now discuss CS's based on generalized empirical likelihood (GEL) test statistics. For definitions and regularity conditions concerning GEL test statistics, see AG4. The asymptotic distribution of a GEL test statistic (under any drifting sequence of parameters) is the same as that of the QLR test statistic; see AG4 for a proof. Given the structure of the proofs below, this implies that all of the asymptotic results stated above for QLR tests also hold for GEL tests.

Specifically, under the assumptions of Theorems 1–4, we have (i) GEL CS's based on GMS critical values have correct size asymptotically, (ii) GEL tests based on GMS critical values have asymptotic power greater than or equal to that of GEL tests based on subsampling or PA critical values with strictly greater power in certain scenarios, and (iii) the "pure" GEL test that uses a constant critical value (equal to $c_{\text{GEL}}(1-\alpha) = \sup_{\Omega \in \Psi_2} c(\Omega, 1-\alpha)$, where $c(\Omega, 1-\alpha)$ is as defined above using the function S_2) is dominated asymptotically by various alternative tests. Such tests include tests constructed from a GEL or QLR test statistic combined with GMS, subsampling, or PA critical values. The results of (iii) indicate that there are notable drawbacks to the asymptotic optimality criteria based on large deviation probabilities considered by Otsu (2006) and Canay (2007).

10.2. *Preliminary Estimation of Identified Parameters*

Here we consider the case where the moment functions in (2.2) depend on a parameter τ (i.e., 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 when θ_0 is the true value of θ. This requires that τ is identified. The sample moment functions in this case are of the form $\overline{m}_{n,j}(\theta) = \overline{m}_{n,j}(\theta, \widehat{\tau}_n(\theta))$. The asymptotic variance of $n^{1/2}\overline{m}_{n,j}(\theta)$ is different when τ is replaced by the estimator $\widehat{\tau}_n(\theta)$ and so $\widehat{\Sigma}_n(\theta)$ needs to be defined accordingly, but otherwise the theoretical treatment of this model is the same as that given above. In fact, Theorem 1 holds in this case using the conditions given in (A.3) of the Appendix. These are high-level conditions that essentially just require that $n^{-1}\sum_{i=1}^{n} m_{n,j}(W_i, \theta, \widehat{\tau}_n(\theta))$ is asymptotically normal (after suitable normalization).

Furthermore, the power comparisons in Section 8, which are stated for i.i.d. observations and no preliminary estimated parameters, can be extended to the case of preliminary estimated parameters. Thus, in this case too, GMS tests have power advantages over subsampling, and PA tests and subsampling tests have power advantages over PA tests.

11. MONTE CARLO EXPERIMENT

11.1. *Experimental Design*

In this section, we provide finite-sample comparisons of the maximum null rejection probability (MNRP) over different null mean vectors and MNRP-corrected power of GMS, recentered subsampling, and PA tests.[12] (MNRP is defined precisely below.) For each test, we consider the QLR test statistic. For the GMS tests, we use the $\varphi^{(1)}$ critical value function, which yields the t-test selection method. We use two values of κ_n, the BIC value $\kappa_n = (\ln n)^{1/2}$ and the law of the iterated logarithm (LIL) value $\kappa_n = (2\ln\ln n)^{1/2}$, which yield $\kappa_n = 2.35$ and $\kappa_n = 1.85$, respectively, when $n = 250$, which is the sample size considered here. We provide results for the bootstrap and asymptotic versions of the GMS test. The GMS tests are denoted by GMS/Boot1 and GMS/Asy1, which use the BIC value of κ_n, and GMS/Boot2 and GMS/Asy2, which use the LIL value of κ_n. (The focus on $(S_2, \varphi^{(1)})$ is based on results in Andrews and Jia (2008) that compare different choices of (S, φ) in terms of asymptotic size and power.)

The subsampling and PA tests considered here also employ the QLR test statistic. Hence, the tests differ only in the way in which the critical value is calculated. Bugni (2007a, 2007b) showed that taking b of order $n^{2/3}$ minimizes the error in the null rejection probability for the recentered subsampling test. In consequence, we use subsample sizes $b = .75n^{2/3}$, $n^{2/3}$, and $1.25n^{2/3}$, which for $n = 250$ yields $b = 30$, 40, and 50, respectively. These subsampling tests are denoted Sub1, Sub2, and Sub3, respectively.

We consider the case in which no equalities arise (i.e., $v = 0$) and the number of inequalities, p, is 2, 4, or 10. For given θ, the null hypothesis is $H_0 : Em(W_i, \theta) \geq 0_p$ for some given moment functions $m(W_i, \theta)$ and the alternative hypothesis is that H_0 does not hold. We consider a general formulation of the testing problem of interest which does not require the specification of a particular form for $m(W_i, \theta)$, as in Andrews and Jia (2008). The finite-sample properties of tests of H_0 depend on $m(W_i, \theta)$ only through (i) $\mu = Em(W_i, \theta)$, (ii) $\Omega = \mathrm{Corr}(m(W_i, \theta))$, and (iii) the distribution of the mean zero, variance I_p random vector $Z^\dagger = \mathrm{Var}^{-1/2}(m(W_i, \theta))(m(W_i, \theta) - Em(W_i, \theta))$. We consider the case in which $Z^\dagger \sim N(0_p, I_p)$. We consider three representative correlation matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}, which exhibit negative, zero, and positive correlations, respectively. By definition, MNRP denotes the maximum null rejection probability over mean vectors in H_0 given the correlation matrix Ω_{Neg}, Ω_{Zero}, or Ω_{Pos} and under the assumption of normally-distributed moment inequalities.[13]

[12]We consider recentered subsampling tests because non-recentered subsampling tests are found to underreject the null hypothesis substantially for sample sizes of 250 and 1000. Even for a sample size of 5000, there is some underrejection. For details, see the Supplemental Material.

[13]The MNRP of a test is the same as the size of the test except that the MNRP is for a fixed correlation matrix and distribution—in the present case a normal distribution—whereas the size is given by the maximum over all allowable correlation matrices and distributions.

Specifically, 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 given by a $p - 1$ vector ρ. 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}.

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 $Z^\dagger$. In large samples, the impact of the distribution of $Z^\dagger$ vanishes because of the central limit theorem (CLT).

For all of the tests considered below, calculations for a subset of the cases considered show without exception that the maximum null rejection probabilities occur for mean vectors μ whose elements are 0's and ∞'s. In consequence, the MNRP results are obtained by computing the maximum null rejection probabilities over μ vectors with this form.

For the power comparisons given here, we compare MNRP-corrected power. By this we mean that the critical values are adjusted by a constant so that the MNRP equals the nominal level .05.

The power comparisons are made based on average power over certain sets, $\mathcal{M}_p(\Omega)$, of vectors μ in the alternative (i.e., $\mu \ngeq 0_p$). For $p = 2$, the set of μ vectors $\mathcal{M}_2(\Omega)$ includes seven elements and is of the form $\mathcal{M}_2(\Omega) = \{(-\mu_1, 0), (-\mu_2, 1), (-\mu_3, 2), (-\mu_4, 3), (-\mu_5, 4), (-\mu_6, 7), (-\mu_7, -\mu_7)\}$, where $\mu_j > 0$ depends on Ω for $j = 1, \ldots, 7$ and is such that the finite-sample power envelope (for known Ω) is .73 at each $\mu \in \mathcal{M}_2(\Omega)$ (see Andrews and Jia (2008) for more details). For brevity, the values of μ_j are given in the Supplemental Material. For $p = 4$ and 10, $\mathcal{M}_4(\Omega)$ includes 24 and 40 elements, respectively. For brevity, the complete specifications of $\mathcal{M}_4(\Omega)$ and $\mathcal{M}_{10}(\Omega)$ are given in the Supplemental Material. The sets $\mathcal{M}_4(\Omega)$ and $\mathcal{M}_{10}(\Omega)$ are defined such that the finite-sample power envelope (for known Ω) is (approximately) .79 and .84, respectively, at each $\mu \in \mathcal{M}_p(\Omega)$ for $p = 4$ and 10.

The simulation results are based on 2500 repetitions for the calculation of the GMS, subsampling, and PA critical values, 2500 simulation repetitions for the finite-sample MNRP results, and 1000 simulation repetitions for the finite-sample MNRP-corrected power results.

11.2. *Simulation Results*

Table II provides the MNRP and power results. The table shows that the GMS/Boot1 and GMS/Asy1 tests have better MNRP and power properties than the GMS/Boot2 and GMS/Asy2 tests. The GMS1 tests have good MNRPs in all cases. For example, the GMS/Boot1 test has MNRP in the interval [.048, .065] for all cases considered. The GMS2 tests tend to overreject the null somewhat with Ω_{Neg}. The GMS/Boot1 test has slightly higher power than the GMS/Asy1 test. Hence, the GMS/Boot1 test performs the best of the GMS tests in terms of MNRP and power by a slight margin over the GMS/Asy1 test.

TABLE II

FINITE-SAMPLE MNRP'S AND MNRP-CORRECTED POWER OF NOMINAL .05 TESTS BASED ON THE QLR TEST STATISTIC COMBINED WITH GMS, RECENTERED SUBSAMPLING (SUB), AND PA CRITICAL VALUES FOR SAMPLE SIZE $n = 250$

		Ω_{Neg}		Ω_{Zero}		Ω_{Pos}	
Number of Moment Inequalities	Critical Value	MNRP	Avg. Power	MNRP	Avg. Power	MNRP	Avg. Power
2	GMS/Boot1	.054	.66	.048	.71	.048	.73
	GMS/Asy1	.047	.66	.041	.71	.046	.72
	GMS/Boot2	.073	.61	.052	.71	.048	.73
	GMS/Asy2	.066	.61	.043	.70	.046	.73
	PA	.040	.56	.039	.61	.046	.66
	Sub1	.050	.60	.050	.65	.051	.68
	Sub2	.061	.60	.061	.65	.060	.67
	Sub3	.071	.60	.068	.65	.068	.68
	P. Envelope	—	.73	—	.73	—-	.73
4	GMS/Boot1	.065	.58	.051	.68	.051	.76
	GMS/Asy1	.064	.57	.052	.66	.052	.76
	GMS/Boot2	.080	.53	.053	.69	.051	.76
	GMS/Asy2	.079	.53	.055	.67	.044	.76
	PA	.050	.43	.045	.52	.045	.69
	Sub1	.046	.43	.047	.53	.051	.71
	Sub2	.060	.43	.062	.53	.062	.71
	Sub3	.077	.43	.075	.53	.068	.71
	P. Envelope	—	.79	—	.79	—	.77
10	GMS/Boot1	.059	.58	.050	.65	.051	.78
	GMS/Asy1	.064	.56	.054	.63	.048	.78
	GMS/Boot2	.075	.54	.051	.67	.051	.79
	GMS/Asy2	.083	.52	.058	.64	.048	.79
	PA	.055	.29	.055	.37	.043	.66
	Sub1	.010	.27	.017	.36	.049	.68
	Sub2	.030	.27	.040	.35	.058	.69
	Sub3	.052	.28	.060	.35	.067	.69
	P. Envelope	—	.85	—	.84	—	.83

The power of the GMS/Boot1 test is substantially greater than that of the PA test. The relative advantage is increasing in p. The average power differences for $p = 2$, 4, and 10 are .09, .13, and .23, respectively, where the average is over Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}.

The best subsampling test in terms of MNRP is Sub1. The MNRP's of Sub1 are quite good except for $p = 10$ with Ω_{Neg} and Ω_{Zero}, in which case Sub1 dramatically underrejects with MNRP's of .010 and .017. The MNRP-corrected power of Sub1, Sub2, and Sub3 is the same and hence does not depend on b.

The power gains of the GMS/Boot1 test over the subsampling tests are quite similar to those of the GMS/Boot1 test over the PA test, although they are a bit smaller for $p = 2$. The power gains are substantial, especially for $p = 4, 10$. For example, in the most extreme case, where $p = 10$ and $\Omega = \Omega_{\text{Neg}}$, the GMS/Boot1 and Sub1 tests have power .58 and .27, respectively.

The differences between the power of the GMS/Boot1 test and the power envelope increase quickly in p. This is because the GMS/Boot1 test is a p-directional test, whereas the power envelope is attained by a unidirectional test. The differences are small for $p = 2$, but quite large for $p = 10$ when $\Omega = \Omega_{\text{Neg}}$ and Ω_{Pos}. The differences in power are decreasing as one moves from Ω_{Neg} to Ω_{Zero} to Ω_{Pos}. Even for $p = 10$, the difference for Ω_{Pos} is only .05, which is remarkably small.

In conclusion, the finite-sample simulations reported here indicate that the GMS/Boot1 test has good MNRP for the cases considered and good power relative to the PA and subsampling tests that are considered. In consequence, the GMS/Boot1 test is the recommended test. It is the bootstrap version of the GMS test based on the QLR test statistic, the t-test moment selection critical value, and the tuning parameter $\kappa_n = (\ln n)^{1/2}$.

APPENDIX

In this Appendix, we start by stating some assumptions on the test statistic function S. Next, we give an alternative parametrization of the moment inequality/equality model to that of Section 2. The new parametrization is conducive to the calculation of the asymptotic properties of CS's and tests. It was first used in AG4. We also specify the parameter space for the case of dependent observations. Proofs of the results of the paper are given in the Supplemental Material (Andrews and Soares (2010)).

A.1. *Test Statistic Assumptions*

The following assumptions concern the test statistic function S.

ASSUMPTION 2: *For all $h_1 \in R^p_{+,\infty}$, all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the df of $S(Z + (h_1, 0_v), \Omega)$ at $x \in R$ is* (a) *continuous for $x > 0$,* (b) *strictly increasing for $x > 0$ unless $v = 0$ and $h_1 = \infty^p$, and* (c) *less than or equal to $1/2$ at $x = 0$ whenever $v \geq 1$ or $h_1 = 0_p$.*

ASSUMPTION 4: (a) *The df of $S(Z, \Omega)$ is continuous at its $1 - \alpha$ quantile, $c(\Omega, 1 - \alpha)$, for all $\Omega \in \Psi$, where $Z \sim N(0_k, \Omega)$ and $\alpha \in (0, 1/2)$.*

(b) *$c(\Omega, 1 - \alpha)$ is continuous in Ω uniformly for $\Omega \in \Psi$.*

ASSUMPTION 5: (a) *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 $x > 0$.*

(b) *$P(S(Z + (m_1, 0_v), \Omega) \leq x) < P(S(Z + (m_1^*, 0_v), \Omega) \leq x)$ for all $x > 0$ for all $m_1, m_1^* \in R^p_{+,\infty}$ with $m_1 < m_1^*$.*

For $(\theta, F) \in \mathcal{F}$, *define* $h_{1,j}(\theta, F) = \infty$ *if* $E_F m_j(W_i, \theta) > 0$ *and* $h_{1,j}(\theta, F) = 0$ *if* $E_F m_j(W_i, \theta) = 0$ *for* $j = 1, \ldots, p$. *Let* $h_1(\theta, F) = (h_{1,1}(\theta, F), \ldots, h_{1,p}(\theta, F))'$ *and* $\Omega(\theta, F) = \lim_{n\to\infty} \mathrm{Corr}_F(n^{1/2}\overline{m}_n(\theta))$.

ASSUMPTION 7: *For some* $(\theta, F) \in \mathcal{F}$, *the df of* $S(Z + (h_1(\theta, F), 0_v), \Omega(\theta, F))$ *is continuous at its* $1 - \alpha$ *quantile, where* $Z \sim N(0_k, \Omega(\theta, F))$.

In Assumption 2, if an element of h_1 equals $+\infty$, then by definition the corresponding element of $Z + (h_1, 0_v)$ equals $+\infty$.

Assumption 2 is used to show that certain asymptotic df's satisfy suitable continuity/strictly increasing properties. These properties ensure that the GMS critical value converges in probability to a constant and the CS has asymptotic size that is not affected by a jump in a df. Assumption 4 is a mild continuity assumption. Assumption 5 is used for the $n^{-1/2}$-local power results. Assumption 5(a) is a continuity/strictly increasing df condition that is the same as Assumption 2(a) except that ℓ can take negative values. Assumption 5(b) is a stochastically strictly increasing condition. With a nonstrict inequality, it is implied by Assumption 1(a). Assumption 7 is used to show that GMS CS's are not asymptotically conservative (i.e., AsyCS $\not> 1 - \alpha$). It is a very weak continuity condition. If the $1 - \alpha$ quantile of $S(Z + (h_1(\theta, F), 0_v), \Omega(\theta, F))$ is positive for some $(\theta, F) \in \mathcal{F}$, which holds quite generally, Assumption 7 is implied by Assumption 2(a). For example, Assumption 7 holds for $S = S_1$ or S_2 whenever (i) $E_F m_j(W_i, \theta) = 0$ for some $j \leq p$ for some $(\theta, F) \in \mathcal{F}$ or (ii) $v \geq 1$ (which holds if an equality is present). It is hard to envision cases of interest where condition (i) fails.

A.2. *Alternative Parametrization and Dependent Observations*

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

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

where $\sigma_{F,j}^2(\theta) = \mathrm{AsyVar}_F(n^{1/2}\overline{m}_{n,j}(\theta))$ denotes the variance of the asymptotic distribution of $n^{1/2}\overline{m}_{n,j}(\theta)$ when the true parameter is θ and the true distribution of the data is F. Let $\Omega = \Omega(\theta, F) = \mathrm{AsyCorr}_F(n^{1/2}\overline{m}_n(\theta))$, where $\mathrm{AsyCorr}_F(n^{1/2}\overline{m}_n(\theta))$ denotes the correlation matrix of the asymptotic distribution of $n^{1/2}\overline{m}_n(\theta)$ when the true parameter is θ and the true distribution of the data is F. (We only consider (θ, F) for which these asymptotic variances and correlation matrices exist, see conditions (iv) and (v) of (A.2) below.) When no preliminary parameter τ is estimated $\sigma_{F,j}^2(\theta) = \lim_{n\to\infty} \mathrm{Var}_F(n^{1/2}\overline{m}_{n,j}(\theta))$

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

For the case described in Section 10.2 (where the sample moment functions depend on a preliminary estimator $\widehat{\tau}_n(\theta)$ of an identified parameter vector τ_0), we define $m_j(W_i, \theta) = m_j(W_i, \theta, \tau_0), m(W_i, \theta) = (m_1(W_i, \theta, \tau_0), \ldots, m_k(W_i, \theta, \tau_0))', \overline{m}_{n,j}(\theta) = n^{-1}\sum_{i=1}^n m_j(W_i, \theta, \widehat{\tau}_n(\theta))$, and $\overline{m}_n(\theta) = (\overline{m}_{n,1}(\theta), \ldots, \overline{m}_{n,k}(\theta))'$. (Hence, in this case, $\overline{m}_n(\theta) \neq n^{-1}\sum_{i=1}^n m(W_i, \theta)$.)

For i.i.d. observations (and no preliminary estimator $\widehat{\tau}_n(\theta)$), 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 (A.1), $\gamma_2 = (\theta, \mathrm{vech}_*(\Omega(\theta, F)))$, and $\gamma_3 = F\}$.

For dependent observations and for sample moment functions that depend on preliminary estimators of identified parameters, 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

(A.2)
(i) $\theta \in \Theta$,
(ii) $\sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0$ for $j = 1, \ldots, p$,
(iii) $E_F m_j(W_i, \theta) = 0$ for $j = p+1, \ldots, k$,
(iv) $\sigma_{F,j}^2(\theta) = \mathrm{AsyVar}_F\big(n^{1/2}\overline{m}_{n,j}(\theta)\big)$ exists and lies in $(0, \infty)$ for $j = 1, \ldots, k$,
(v) $\mathrm{AsyCorr}_F\big(n^{1/2}\overline{m}_n(\theta)\big)$ exists and equals $\Omega_{\gamma_{2,2}} \in \Psi$,
(vi) $\{W_i : i \geq 1\}$ are stationary under F,

where $\gamma_1 = (\gamma_{1,1}, \ldots, \gamma_{1,p})'$ and $\Omega_{\gamma_{2,2}}$ is the $k \times k$ correlation matrix determined by $\gamma_{2,2}$.[14] Furthermore, Γ must be restricted by enough additional conditions such that, under any sequence $\{\gamma_{n,h} = (\gamma_{n,h,1}, (\theta_{n,h}, \mathrm{vech}_*(\Omega_{n,h})), F_{n,h}) : n \geq 1\}$ of parameters in Γ that satisfies $n^{1/2}\gamma_{n,h,1} \to h_1$ and $(\theta_{n,h}, \mathrm{vech}_*(\Omega_{n,h})) \to h_2 =$

[14] In AG4, a strong mixing condition is imposed in condition (vi) of (A.2). This condition is used to verify Assumption E0 in that paper and is not needed with GMS critical values. To extend the subsampling power results of the paper to dependent observations, this assumption needs to be imposed.

$(h_{2,1}, h_{2,2})$ for some $h = (h_1, h_2) \in R^p_{+,\infty} \times R^q_{[\pm\infty]}$, we have

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

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

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

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

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

where $\Omega_{h_{2,2}}$ is the $k \times k$ correlation matrix for which $\text{vech}_*(\Omega_{h_{2,2}}) = h_{2,2}$, $\widehat{\sigma}^2_{n,j}(\theta) = [\widehat{\Sigma}_n(\theta)]_{jj}$ for $1 \leq j \leq k$, and $\widehat{D}_n(\theta) = \text{Diag}\{\widehat{\sigma}^2_{n,1}(\theta), \ldots, \widehat{\sigma}^2_{n,k}(\theta)\}$ $(= \text{Diag}(\widehat{\Sigma}_n(\theta)))$.[15,16]

For example, for i.i.d. observations, conditions (i)–(vi) of (2.2) imply conditions (i)–(vi) of (A.2). Furthermore, conditions (i)–(vi) of (2.2) plus the definition of $\widehat{\Sigma}_n(\theta)$ in (3.2) and the additional condition (vii) of (2.2) imply conditions (vii)–(x) of (A.3). For a proof, see Lemma 2 of AG4.

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

REFERENCES

ANDREWS, D. W. K. (1999a): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–564. [120,132]

——— (1999b): "Estimation When a Parameter Is on a Boundary," *Econometrica*, 67, 1341–1383. [123,131]

——— (2000): "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space," *Econometrica*, 68, 399–405. [123,132]

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

[16]Condition (x) of (A.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 (A.3) with n replaced by w_n.

ANDREWS, D. W. K., AND P. GUGGENBERGER (2009a): "Hybrid and Size-Corrected Subsampling Methods," *Econometrica*, 77, 721–762. [125]

——— (2009b): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709. [120]

——— (2010a): "Applications of Subsampling, Hybrid, and Size-Correction Methods," *Journal of Econometrics* (forthcoming). [125]

——— (2010b): "Asymptotic Size and a Problem With Subsampling and With the *m* Out of *n* Bootstrap," *Econometric Theory*, 26 (forthcoming). [123,125]

ANDREWS, D. W. K., AND S. HAN (2009): "Invalidity of the Bootstrap and *m* Out of *n* Bootstrap for Interval Endpoints," *Econometrics Journal*, 12, S172–S199. [123]

ANDREWS, D. W. K., AND P. JIA (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," Discussion Paper 1676, Cowles Foundation, Yale University. [122,127,128,149,150]

ANDREWS, D. W. K., AND G. SOARES (2010): "Supplement to 'Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection'," *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/7502_Proofs.pdf; http://www.econometricsociety.org/ecta/Supmat/7502_data and programs.zip. [122,123,136, 138,152]

ANDREWS, D. W. K., S. BERRY, AND P. JIA (2004): "Confidence Regions for Parameters in Discrete Games With Multiple Equilibria, With an Application to Discount Chain Store Location," Unpublished Manuscript, Cowles Foundation, Yale University. [119,123,124]

BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814. [121]

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2008): "Sharp Identification Regions in Games," CEMMAP Working Paper CWP15/08, Institute for Fiscal Studies, UCL. [123]

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2007): "Set Identified Linear Models," Unpublished Manuscript, Toulouse School of Economics. [123]

BUGNI, F. (2007a): "Bootstrap Inference in Partially Identified Models," Unpublished Manuscript, Department of Economics, Northwestern University. [121-123,134,137,149]

——— (2007b): "Bootstrap Inference in Partially Identified Models: Pointwise Construction," Unpublished Manuscript, Department of Economics, Northwestern University. [121-123,134, 137,149]

CANAY, I. A. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Unpublished Manuscript, Department of Economics, University of Wisconsin. [121-123,134,148]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [119]

CILIBERTO, F., AND E. TAMER (2009): "Market Structure and Multiple Equilibrium in Airline Markets," *Econometrica*, 77, 1791–1828. [119,124]

FAN, Y., AND S. PARK (2007): "Confidence Sets for Some Partially Identified Parameters," Unpublished Manuscript, Department of Economics, Vanderbilt University. [123,134]

GALICHON, A., AND M. HENRY (2009): "A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics* (forthcoming). [123]

GUGGENBERGER, P., J. HAHN, AND K. KIM (2008): "Specification Testing Under Moment Inequalities," *Economics Letters*, 99, 375–378. [123]

HANNAN, E. J., AND B. G. QUINN (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, Ser. B*, 41, 190–195. [120]

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [138]

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [119,120,124]

KUDO, A. (1963): "A Multivariate Analog of a One-Sided Test," *Biometrika*, 50, 403–418. [127]

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions With Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546. [119,124]
MIKUSHEVA, A. (2007): "Uniform Inference in Autoregressive Models," *Econometrica*, 75, 1411–1452. [125]
MOON, H. R., AND F. SCHORFHEIDE (2006): "Boosting Your Instruments: Estimation With Overidentifying Inequality Moment Conditions," Unpublished Working Paper, Department of Economics, University of Southern California. [119,123,124]
OTSU, T. (2006): "Large Deviation Optimal Inference for Set Identified Moment Inequality Models," Unpublished Manuscript, Cowles Foundation, Yale University. [121,123,148]
PAKES, A., J. PORTER, K. HO, AND J. ISHII (2004): "Applications of Moment Inequalities," Unpublished Working Paper, Department of Economics, Harvard University. [119,123,127]
POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer. [121]
PRATT, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56, 541–567. [121]
ROMANO, J. P., AND A. M. SHAIKH (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Inference and Planning*, 138, 2786–2807. [120,123,127]
——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211. [120,123,127]
ROSEN, A. M. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117. [120,123,127,132,138, 139,146]
SILVAPULLE, M. J., AND P. K. SEN (2005): *Constrained Statistical Inference*. New York: Wiley. [127,139]
SOARES, G. (2005): "Inference With Inequality Moment Constraints," Unpublished Working Paper, Department of Economics, Yale University. [123,127]
STOYE, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315. [123]
WOUTERSEN, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters," Unpublished Manuscript, Department of Economics, Johns Hopkins University. [123]

Cowles Foundation for Research in Economics, Dept. of Economics, Yale University, P.O. Box 208281, Yale Station, New Haven, CT 06520-8281, U.S.A.; Donald.andrews@yale.edu

and

Dept. of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268, U.S.A.; gsoares@aya.yale.edu.

Manuscript received October, 2007; final revision received February, 2009.