# INFERENCE FOR PARAMETERS DEFINED BY MOMENT INEQUALITIES USING GENERALIZED MOMENT SELECTION

**By**

**Donald W.K. Andrews and Gustavo Soares**

**October 2007**

**COWLES FOUNDATION DISCUSSION PAPER NO. 1631**

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

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

# Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection


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

Gustavo Soares
*Department of Economics*
*Yale University*


May 2007
Revised: October 2007


*Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-0417911. The authors thank Victor Chernozhukov for comments.

## Abstract

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.

*JEL Classification Numbers*: C12, C15.

# 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, e.g., see Manski and Tamer (2002), Imbens and Manski (2004), Ciliberto and Tamer (2003), Andrews, Berry, and Jia (2004), Pakes, Porter, Ishii, and Ho (2004), Moon and Schorfheide (2006), and Chernozhukov, Hong, and Tamer (2007) (CHT).

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. CHT focuses on subsampling and "plug-in asymptotic" (PA) critical values. Here we introduce and 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 quasi-likelihood ratio (QLR) statistics, generalized empirical likelihood ratio (GEL) statistics, 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 (2005d) (hereafter AG4), and Romano and Shaikh (2005a,b). "Plug-in asymptotic" 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 CHT and AG4 and a variant of them is considered in Rosen (2005).

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 $h = (h_1, ..., h_p)'$, whose $j$th element $h_j \in [0, \infty]$ indexes the extent to which the $j$th 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 $j = 1, ..., p$. 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 AIC, BIC, and HQIC model selection criteria, see Hannan and Quinn (1979) regarding HQIC. Andrews (1999) uses 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. In this paper, we show that GMS critical values yield uniformly asymptotically valid CS's and tests. These results hold for both i.i.d. and dependent observations. We also show that GMS procedures are not asymptotically conservative. They are asymptotically non-similar, 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. Thus, power is important for both tests and CS's. We determine and compare the power of GMS, subsampling, and PA tests. To date, no general asymptotic power analysis is available for any tests in the moment inequality literature. Otsu (2006), Bugni (2007a,b), and Canay (2007) consider asymptotic power against fixed alternatives, but tests typically have asymptotic power equal to one against such alternatives.

We investigate the asymptotic power of GMS, subsampling, and PA tests for local and non-local 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 $1/n^{1/2}$-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 $\kappa_n$ is a GMS constant such as $\kappa_n = (2 \ln \ln n)^{1/2}$, and (ii) the GMS and subsampling critical values do not have the degenerate probability limit of 0. Typically $b \approx n^{\zeta}$ for some $\zeta \in (0,1)$ such as $\zeta = 1/2$ and $\kappa_n = o(n^{\varepsilon})\ \forall\ \varepsilon > 0$.

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 reports finite-sample size and power results obtained via simulation for a missing-data model considered in Imbens and Manski (2004) and an interval-outcome regression model considered in Manski and Tamer (2002). The results show good size properties of the GMS procedures considered. The results also indicate that the asymptotic power comparisons described above are reflected in the finite-sample power performance of GMS and subsampling tests in the cases considered.

The paper shows that 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 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 ongoing research to be reported in Andrews, Berry, and Jia (2007). To date, we find that the MMM and QLR test statistics combined with the GMS procedures based on $t$ tests or the MMSC work well in practice. The QLR/MMSC combination has some advantageous performance properties, but is computationally relatively demanding. The MMM/$t$ tests combination has computational advantages.

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,b) and Canay (2007) consider particular types of bootstrap GMS critical values.

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,b) shows that a particular type of a GMS test (based on $\varphi^{(1)}$ defined below) has more accurate pointwise asymptotic size than a 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.

Other papers in the literature that consider inference with moment inequalities include: CHT, Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2004), AG4, Romano and Shaikh (2005a,b), Rosen (2005), Beresteanu and Molinari (2006), Galichon and Henry (2006), Moon and Schorfheide (2006), Otsu (2006), Woutersen (2006), Bugni (2007a,b), Canay (2007), Guggenberger, Hahn, and Kim (2007), and Stoye (2007).[1]

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

[1]GMS critical values based on $\varphi^{(1)}$ and $\varphi^{(5)}$, defined below, were introduced in Soares (2005). The present paper supplants Soares (2005). GMS critical values of types $\varphi^{(2)} - \varphi^{(4)}$ were considered by the authors in January 2007. CHT mentions critical values of GMS type based on $\varphi^{(1)}$, see their Remark 4.5. The 2003 working paper version of CHT also discusses a bootstrap version of the GMS method based on $\varphi^{(1)}$ in the context of the interval outcome model. Galichon and Henry (2006), independently of Soares (2005), consider a set selection method that is analogous to GMS based on $\varphi^{(1)}$. Bugni (2007a,b) considers GMS critical values based on $\varphi^{(1)}$. His work was done independently of, but subsequently to, Soares (2005). Canay (2007) independently considers GMS critical values based on $\varphi^{(3)}$. Bugni (2007a,b) and Canay (2007) focus on bootstrap versions of the GMS critical values.

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 $1/n^{1/2}$-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. Appendix A contains proofs of all of the results except for Theorem 4, which is proved in Appendix B.

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 $R_+ = \{x \in R : x \geq 0\}$, $R_{+,\infty} = R_+ \cup \{+\infty\}$, $R_{[+\infty]} = R \cup \{+\infty\}$, $R_{[\pm\infty]} = R \cup \{\pm\infty\}$, $K^p = K \times ... \times K$ (with $p$ copies) for any set $K$, $\infty^p = (+\infty, ..., +\infty)'$ (with $p$ copies). All limits are as $n \to \infty$ unless specified otherwise. Let "pd" abbreviate "positive definite." Let $cl(\Psi)$ denote the closure of a set $\Psi$. We let AG1 abbreviate Andrews and Guggenberger (2005a).

# 2 Moment Inequality Model

We now introduce the moment inequality/equality model. The true value $\theta_0$ ($\in \Theta \subset R^d$) is assumed to satisfy the moment conditions:

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

where $\{m_j(\cdot, \theta) : j = 1, ..., k\}$ are known real-valued moment functions, $k = p+v$, and $\{W_i : i \geq 1\}$ are i.i.d. or stationary random vectors with joint distribution $F_0$. The observed sample is $\{W_i : i \leq n\}$. A key feature of the model is that the true value $\theta_0$ is not necessarily identified. That is, knowledge of $E_{F_0} m_j(W_i, \theta)$ for $j = 1, ..., k$ for all $\theta \in \Theta$ does not necessarily imply knowledge of $\theta_0$. In fact, even knowledge of $F_0$ does not necessarily imply knowledge of the true value $\theta_0$. More information than is available in $\{W_i : i \leq n\}$ may be needed to identify the true parameter $\theta_0$.

Note that both moment inequalities and moment equalities arise in the entry game models considered in Ciliberto and Tamer (2003) 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, e.g., 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 $\theta_0$.

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

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

where $\Psi$ 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 A, see Section 12.2.

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 $\theta$ is

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_{1-\alpha}(\theta)\}, \tag{2.3}$$

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

$$ExCS_n = \inf_{(\theta,F)\in\mathcal{F}} P_F(T_n(\theta) \leq c_{1-\alpha}(\theta)) \text{ and } AsyCS = \liminf_{n\to\infty} ExCS_n, \tag{2.4}$$

respectively. The definition of $AsyCS$ takes the "sup" before the "lim." This builds uniformity over $(\theta, F)$ 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 (2005a,b,c) and Mikusheva (2007) show 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 "sup") can be very misleading in some models. See AG4 for further discussion.

The exact and asymptotic maximum coverage probabilities are

$$
\begin{aligned}
ExMaxCP_n &= \sup_{(\theta,F)\in\mathcal{F}} P_F(T_n(\theta) \leq c_{1-\alpha}(\theta)) \text{ and} \\
AsyMaxCP &= \limsup_{n\to\infty} ExMaxCP_n,
\end{aligned} \tag{2.5}
$$

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

respectively. The magnitude of asymptotic non-similarity of the CS is measured by the difference $AsyMaxCP - AsyCS$.

# 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.

## 3.1 Form of the Test Statistics

The sample moment functions are

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

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

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

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

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

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

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

and $\sigma_j^2$ is the $j$th diagonal element of $\Sigma$. With the function $S_1$, the parameter space $\Psi$ for the correlation matrices in condition (vi) of (2.2) is $\Psi = \Psi_1$, where $\Psi_1$ contain 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. This type of statistic has been considered in CHT, AG4, Romano and Shaikh (2005a,b), and Soares (2005). Note that $S_1$ normalizes the moment functions by dividing by $\sigma_j$ in each summand. One could consider a function without this normalization, as in Pakes, Porter, Ho, and Ishii (2004), 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

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

With this function, the parameter space $\Psi$ in (2.2) is $\Psi = \Psi_2$, where $\Psi_2$ contains all $k \times k$ correlation matrices whose determinant is greater than or equal to $\varepsilon$ for some $\varepsilon > 0$.[4] 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 (2005). 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.

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

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

where $[m_{(j)}/\sigma_{(j)}]_-^2$ denotes the $j$th largest value among $\{[m_\ell/\sigma_\ell]_-^2 : \ell = 1, ..., p\}$ 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, Berry, and Jia (2007).

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

[4] 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 the absolute value to a different positive power (such as one), 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, Berry, and Jia (2007).

## 3.2 Test Statistic Assumptions

Next, we state the most important assumptions concerning the function $S$, viz. Assumptions 1, 3, and 6. For ease of reading, technical assumptions (mostly continuity and strictly-increasing assumptions on asymptotic distribution functions (df's)), viz., Assumptions 2, 4, 5, and 7, are stated in Appendix A. We show below that the functions $S_1$–$S_3$ automatically satisfy Assumption 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 $\Psi$ is as in condition (vi) of (2.2).[5] 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 non-increasing 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) $S(m, \Sigma) = S(\Delta m, \Delta \Sigma \Delta)$ for all $m \in R^k$, $\Sigma \in R^{k \times k}$, and pd diagonal $\Delta \in R^{k \times k}$.

(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 $\Omega \in \Psi$.[6]

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

[5] For dependent observations, $\Psi$ is as in condition (v) of (12.2) in Appendix A.

[6] In Assumption 1(d) (and in Assumption 4(b) in Appendix A), $S(m, \Omega)$ and $c(\Omega, 1-\alpha)$ are viewed as functions defined on the space of all correlation matrices. By definition, $c(\Omega, 1-\alpha)$ is continuous in $\Omega$ 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 $|c_{\Omega^*}(1-\alpha) - c_{\Omega}(1-\alpha)| < \eta$.

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

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 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. Assumptions 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 non-negative.

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 $\tau > 0$ in $m$. This is used to show that the test based on $S$ has asymptotic power equal to one 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

## 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 $j = 1, ..., p$. 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., $E_F m_j(W_i, \theta_0) >> 0$ for

$j = 2, ..., p$. 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

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

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

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

$h_1 \in R^p_{+,\infty}$, and $\Omega_0$ is a $k \times k$ correlation matrix. This result holds by (4.2), the central limit theorem, and convergence in probability of the sample correlation matrix, see the proof of Thm. 1 of AG4. The $p$-vector $h_1$ is the limit of $(n^{1/2}E_{F_n}m_1(W_i,\theta_0)/\sigma_{F_n,1}(\theta_0), ..., n^{1/2}E_{F_n}m_p(W_i,\theta_0)/\sigma_{F_n,p}(\theta_0))'$ 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 $\Omega_0$ can be consistently estimated, but the "$1/n^{1/2}$-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

$\Omega_0$ 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

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

evaluated at $\theta = \theta_0$, where $\{\kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \to \infty$. As discussed below, by the law of the iterated logarithm, a suitable choice of $\kappa_n$ often is

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

Let $\xi_{n,j}(\theta)$, $h_{1,j}$, and $[\Omega_0^{1/2} Z^*]_j$ denote the $j$th elements of $\xi_n(\theta)$, $h_1$, and $\Omega_0^{1/2} Z^*$, respectively, for $j = 1, ..., p$. 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 $\kappa_n$ factor) and the desired replacement of $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is $\infty$.

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 $j = 1, ..., p$, 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 $\varphi_j$ satisfies (i) $\varphi_j(\xi, \Omega) = 0$ for all $\xi = (\xi_1, ..., \xi_k)' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$ 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 $\xi_* = (\xi_{*,1}, ..., \xi_{*,k})' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$ 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 $\xi_{n,j}(\theta_0)$ if 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 $j = 1, ..., p$. For $j = p+1, ..., k$, 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 $\varphi_j$ include

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

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

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 $\infty$ 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 $\widehat{\sigma}_{n,j}(\theta_0)$ is the $(j,j)$ element of $\widehat{\Sigma}_n(\theta_0)$ for $j = 1, ..., p$. If one takes the critical value to satisfy $\kappa_n = (2 \ln \ln n)^{1/2}$, then by the law of the iterated logarithm the probability is zero that (4.6) occurs for infinitely many $n$ when the true mean $E_F m_j(W_i, \theta_0)$ equals 0. That is, the probability of falsely failing to select the $j$th moment condition for infinitely many $n$ is zero. The GMS procedure based on $\varphi^{(1)}$ is the same as the Wald test procedure in Andrews (1999, 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 non-decreasing 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 $\psi$ 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 $\varphi^{(1)}$ and $\varphi^{(2)}$ as functions of $\xi_j$ for $j = 1, ..., p$.

The functions $\varphi^{(r)}$ for $r = 1, ..., 4$ all exhibit "element by element" determination of $\varphi_j^{(r)}(\xi, \Omega)$ because the latter depends only on $\xi_j$. 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 non-diagonal, 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. It is related to the information criterion-based moment selection criteria (MSC) considered in Andrews (1999) for a different moment selection problem. We refer to $\varphi^{(5)}$ as the modified MSC (MMSC) $\varphi$ function. It is computationally more expensive than the $\varphi^{(r)}$ functions considered above. We use the function $\varphi^{(5)}$ in the simulations reported in Section 11 below.

Define $c = (c_1, ..., c_k)'$ to be a selection $k$-vector of $0's$ and $1's$. If $c_j = 1$, the $j$th moment condition is selected; if $c_j = 0$, it is not selected. The moment equality functions are always selected, so $c_j = 1$ for $j = p+1, ..., k$. Let $|c| = \sum_{j=1}^{k} c_j$. For $\xi \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, define $c \cdot \xi = (c_1 \xi_1, ..., c_k \xi_k)' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, where $c_j \xi_j = 0$ if $c_j = 0$ and $\xi_j = \infty$. Let $\mathcal{C}$ denote the parameter space for the selection vectors. 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, see Rosen (2005) for a discussion of examples of this sort, then this can be built into the definition of $\mathcal{C}$. Let $\eta(\cdot)$ 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

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

Note the minus sign that appears in the first argument of the $S$ function. This ensures that a large *positive* value of $\xi_j$ yields a large value of $S(-c \cdot \xi, \Omega)$ when $c_j = 1$, as desired. Since $\eta(\cdot)$ is increasing, $-\eta(|c|)$ 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 increasing in the number of selected moments. For $j = 1, ..., p$, define

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

Using Assumptions 1(b) and 6,

$$\begin{aligned} &\kappa_n^\tau \left( S(-c \cdot \xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)) - \eta(|c|) \right) \\ &= S(-c \cdot n^{1/2}\overline{m}_n(\theta_0), \widehat{\Sigma}_n(\theta_0)) - \eta(|c|)\kappa_n^\tau, \end{aligned} \tag{4.9}$$

where $\tau$ 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 (rhs) of (4.9) over $\mathcal{C}$. The rhs of (4.9) is analogous to the BIC and HQIC criteria considered in the model selection literature in which case $\eta(x) = x$, $\kappa_n = (\log n)^{1/2}$ for BIC, $\kappa_n = (Q \ln \ln n)^{1/2}$ for some $Q \geq 2$ for HQIC, and $\tau = 2$ (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 $\eta(x) = x$, the function $\varphi^{(5)}$ reduces to $\varphi^{(1)}$.

Returning now to the general case, given a choice of function $\varphi(\xi, \Omega) = (\varphi_1(\xi, \Omega), ..., \varphi_p(\xi, \Omega), 0, ..., 0)' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, 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

$$S_n(\theta_0, Z^*) = S\left(\widehat{\Omega}_n^{1/2}(\theta_0) Z^* + \varphi\Big(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)\Big), \widehat{\Omega}_n(\theta_0)\right), \tag{4.10}$$

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

$$\widehat{c}_n(\theta_0, 1-\alpha) = \inf\left\{x \in R : P\left(S_n(\theta_0, Z^*) \leq x\right) \geq 1-\alpha\right\}, \tag{4.11}$$

where $P(S_n(\theta_0, Z^*) \leq x)$ denotes the conditional df at $x$ of $S_n(\theta_0, Z^*)$ 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 variables $\{Z_r^* : r = 1, ..., R\}$ 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 $\{S_n(\theta_0, Z_r^*) : r = 1, ..., R\}$, where $R$ is large.

A bootstrap version of the GMS critical value is obtained by replacing $\widehat{\Omega}_n^{1/2}(\theta) Z^*$ in (4.10) by a re-centered bootstrapped version of $n^{1/2} \widehat{D}_n^{-1/2}(\theta) \overline{m}_n(\theta)$, denoted by $BT_n^*(\theta)$. 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,

$$BT_n^*(\theta) = n^{1/2}\left(\widehat{D}_n^*(\theta)\right)^{-1/2}\left(\overline{m}_n^*(\theta) - \overline{m}_n(\theta)\right), \text{ where}$$
$$\overline{m}_n^*(\theta) = n^{-1}\sum_{i=1}^n m(W_i^*, \theta),\ \widehat{D}_n^*(\theta) = Diag(\widehat{\Sigma}_n^*(\theta)), \tag{4.12}$$

and $\widehat{\Sigma}_n^*(\theta)$ is defined 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 $W_i$. For the asymptotic results given below to hold with a bootstrap GMS critical value, one needs that $BT_n^*(\theta_{n,h}) \rightarrow_d \Omega_0^{1/2} Z^*$ under certain triangular arrays of true distributions and true parameters $\theta_{n,h}$, where $\Omega_0$ is a $k \times k$ correlation matrix and $Z^*$ is as in (4.10).[7] 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. Bugni (2007a,b) and Canay (2007) provide 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.

## 4.2 Assumptions

Next we state assumptions on the function $\varphi$ 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 $\xi = (\xi_1, ..., \xi_k)'$, for $j = 1, ..., p$.

(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 $\xi = (\xi_1, ..., \xi_k)'$, for $j = 1, ..., p$.

**Assumption GMS2.** $\kappa_n \rightarrow \infty$.

[7] More specifically, this convergence must hold under any sequence of distributions $\{\gamma_n : n \geq 1\}$ defined just above (12.3) in the Appendix (in which case $\Omega_0 = \Omega_{h_{2,2}}$), the convergence needs to be joint with that in (12.3) of the Appendix, and the convergence must hold with $\{n\}$ replaced by any subsequence $\{w_n\}$ of sample sizes.

Assumptions GMS1 and GMS2 are not restrictive. For example, the functions $\varphi^{(1)} - \varphi^{(4)}$ satisfy Assumption GMS1 and $\kappa_n = (2 \ln \ln n)^{1/2}$ 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 Appendix B 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 $1/n^{1/2}$-local alternatives.

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

**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 $\kappa_n = (2 \ln \ln n)^{1/2}$. 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 Appendix B 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 $1/n^{1/2}$-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 $j = 1, ..., p$.

Assumption GMS5 holds for all reasonable choices of $\kappa_n$ and $b$. For example, for $\kappa_n = (2 \ln \ln n)^{1/2}$, Assumption GMS5 holds for $b = n^{\zeta}$ for any $\zeta \in (0, 1)$. Any reasonable choice of $b$ satisfies the latter condition. 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 $1/n^{1/2}$-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 $j = 1, ..., p$.

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

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

### 4.3 Asymptotic Size Results

The following Theorem applies to i.i.d. observations, in which case $\mathcal{F}$ is defined in (2.2), and to dependent observations, in which case for brevity $\mathcal{F}$ is defined in (12.2)-(12.3) in Appendix A.

**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*

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

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

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

**Comments. 1.** 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.

**2.** Theorem 1 places no assumptions on the moment functions $m(W_i, \theta)$ beyond the existence of mild moments conditions that appear in the definition of $\mathcal{F}$.

**3.** 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 (2005).

## 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 $(\theta, F)$). 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 over-identifying 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 of correct model specification, the limit as $n \to \infty$ of the finite-sample size of the GMS model specification test is $\leq \alpha$ 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 $\alpha$ under the assumptions of Theorem 1(b).[8] That is, the GMS model specification test may be asymptotically conservative.

# 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 $j$th subsample of size $b$ rather than the full sample for $j = 1, ..., q_n$. The empirical df and $1 - \alpha$ sample quantile of $\{T_{n,b,j}(\theta) : j = 1, ..., q_n\}$ are

$$U_{n,b}(\theta, x) = q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta) \leq x) \text{ for } x \in R \text{ and}$$
$$c_{n,b}(\theta, 1-\alpha) = \inf\{x \in R : U_{n,b}(\theta, x) \geq 1 - \alpha\}. \quad (6.1)$$

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 "re-centered" subsample statistics and subsample statistics based on full-sample variance matrices, see AG4. The resulting subsampling CS's have the same asymptotic size and power properties (to first order) as those defined above.

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.

[8] The reason is that when the null of correct model specification holds and $(\theta_0, F_0)$ 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 $\theta \neq \theta_0$.

# 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 $\Omega$ of the moment functions. We analyze plug-in asymptotic (PA) critical values that are determined by the least-favorable asymptotic null distribution for given $\Omega$ evaluated at a consistent estimator of $\Omega$. Such critical values have been considered for many years in the literature on multivariate one-sided tests, see Silvapulle and Sen (2005) for references. CHT and AG4 consider them in the context of the moment inequality literature. Rosen (2005) considers variations of PA critical values that make adjustments in the case where it is known that if one moment inequality holds as an equality then another cannot.

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

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

AG4 shows 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 $1/n^{1/2}$-local alternatives. In the next section we consider "distant alternatives," which differ from the null by more than $O(1/n^{1/2})$ and may 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 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 $\theta_0$, we consider tests of

$$H_0 : E_F m_j(W_i, \theta_0) \geq 0 \text{ for } j = 1, ..., p \text{ and}$$
$$E_F m_j(W_i, \theta_0) = 0 \text{ for } j = p + 1, ..., k, \quad (8.1)$$

where $F$ 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 $(\theta, F)$ 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

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

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

**Assumption LA1.** The true parameters $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ satisfy:

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

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

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

**Assumption LA2.** The $k \times d$ matrix $\Pi(\theta, F) = (\partial/\partial\theta')[D^{-1/2}(\theta, F) E_F m(W_i, \theta)]$ exists and is continuous in $(\theta, F)$ for all $(\theta, F)$ in a neighborhood of $(\theta_0, F_0)$.

Assumption LA1(a) specifies that the true values $\{\theta_n : n \geq 1\}$ are local to the null value $\theta_0$. 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 non-negative. Assumptions LA1(a) and (c) imply that $\Omega(\theta_0, F_n)$ exists and $\Omega(\theta_0, F_n) \to \Omega_0 = \Omega(\theta_0, F_0)$.

The asymptotic distribution of the test statistic $T_n(\theta_0)$ under $1/n^{1/2}$-local alternatives depends on the limit of the (normalized) moment inequality functions when evaluated at the null value $\theta_0$ because $T_n(\theta_0)$ is evaluated at $\theta_0$. Under Assumptions LA1 and LA2, we show that

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

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

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 $j = 1, ..., p$ or $\Pi'_{0,j}\lambda \neq 0$ for some $j = p + 1, ..., k$.

The asymptotic distribution of $T_n(\theta_0)$ under $1/n^{1/2}$-local alternatives is shown to be $J_{h_1,\lambda}$, where $J_{h_1,\lambda}$ is defined by

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

for $Z^* \sim N(0_k, I_k)$. 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 $j = 1, ..., p$.

Note that in Assumption LA4 the functions are evaluated at the true value $\theta_n$, not at the null value $\theta_0$, 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 $j = 1, ..., p$.

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

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

Below the probability limit of the GMS critical value $\widehat{c}_n(\theta_0, 1-\alpha)$ is shown to be $c_{\pi_1}(\varphi, 1-\alpha)$.

The following assumption is used to obtain the $1/n^{1/2}$-local alternative power function of the GMS test. Let $C(\varphi) = \{\widetilde{\pi}_1 = (\widetilde{\pi}_1, ..., \widetilde{\pi}_p)' \in R^p_{[+\infty]}$ : for $j = 1, ..., p$, either $\widetilde{\pi}_{1,j} = \infty$ or $\varphi_j(\xi, \Omega) \to \varphi_j((\widetilde{\pi}_1, 0_v), \Omega_0)$ as $(\xi, \Omega_0) \to ((\widetilde{\pi}_1, 0_v), \Omega_0)\}$. Roughly speaking, $C(\varphi)$ is the set of $\widetilde{\pi}_1$ vectors for which $\varphi$ is continuous at $((\widetilde{\pi}_1, 0_v), \Omega_0)$. For example, $C(\varphi^{(1)}) = \{\widetilde{\pi}_1 \in R^p_{[+\infty]} : \widetilde{\pi}_{1,j} \neq 1$ for $j = 1, ..., p\}$, $C(\varphi^{(2)}) = R^p_{[+\infty]}$ provided $\psi$ 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 $C(\varphi^{(5)}) = \{\pi_1 \in R^p_{[+\infty]} : S(-c \cdot (\widetilde{\pi}_1, 0_v), \Omega_0) - \eta(|c|)$ 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 $1/n^{1/2}$-local power formulae given below do not apply to certain "discontinuity vectors" $\pi_1 = (\pi_{1,1}, ..., \pi_{1,p})'$. However, this does not affect the power comparisons between GMS, subsampling, and PA tests, because Assumption LA5 is not needed for those results. They hold for all $\pi_1$ 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 $j = 1, ..., p$.

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 $g_1 = (g_{1,1}, ..., g_{1,p})'$. Note that $0_p \leq g_1 \leq \pi_1 \leq h_1$.[9] The probability limit of the subsampling critical value is shown to depend on

$$\lim_{n\to\infty} b^{1/2}D^{-1/2}(\theta_0, F_n)E_{F_n}m(W_i,\theta_0) = (g_1, 0_v) \in R^k. \quad (8.6)$$

Note that $(g_1, 0_v) \in R^p_{+,\infty} \times \{0_v\}$. Thus, elements of $(g_1, 0_v)$ are necessarily non-negative. 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 $J_{g_1,0_d}$. 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 $J_{0_p,0_d}$.

**Theorem 2** *Under Assumptions* 1-5 *and* LA1-LA2,

(a) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)) = J_{h_1,\lambda}(c_{\pi_1}(\varphi, 1-\alpha))$ *provided Assumptions* GMS2, GMS3, LA4, *and* LA5 *hold,*

(b) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha)) = J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha))$ *provided Assumption* LA6 *holds, and*

(c) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c(\widehat{\Omega}_n(\theta_0), 1-\alpha)) = J_{h_1,\lambda}(c_{0_p,0_d}(1-\alpha))$.

**Comments. 1.** Theorem 2(a) provides the $1/n^{1/2}$-local alternative power function of the GMS test. The probability limit of the GMS critical value $\widehat{c}_n(\theta_0, 1-\alpha)$ under $1/n^{1/2}$-local alternatives is $c_{\pi_1}(\varphi, 1-\alpha)$, see Lemma 4(b) in Appendix A. Theorem 2(b) and (c) provide the $1/n^{1/2}$-local alternative power function of the subsampling and PA tests.

**2.** 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 non-similarity of the GMS, subsampling, and PA tests. See Section 11 for numerical results concerning finite-sample non-similarity.

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

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

**Theorem 3** *Under Assumptions* 1-5, LA1-LA4, LA6, GMS2-GMS3, *and* GMS5-GMS6,

(a) $\liminf_{n\to\infty} P_{F_n}(T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)) \geq \lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha))$ *with strict inequality whenever* $g_{1,j} < \infty$ *and* $\pi_{1,j} = \infty$ *for some* $j = 1, ..., p$ *and* $c_{g_1, 0_d}(1-\alpha) > 0$,

(b) $\liminf_{n\to\infty} P_{F_n}(T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)) \geq \lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c(\widehat{\Omega}_n(\theta_0), 1-\alpha))$ *with strict inequality whenever* $\pi_{1,j} = \infty$ *for some* $j = 1, ..., p$, *and*

(c) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha)) \geq \lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c(\widehat{\Omega}_n(\theta_0), 1-\alpha))$ *with strict inequality whenever* $g_1 > 0_p$, *where Assumptions* GMS2-GMS3, GMS5-GMS6, *and* LA4 *are not needed for this result.*

**Comments. 1.** 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 $c_{g_1, 0_d}(1-\alpha) > 0$.[10] 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 $\Omega_0$) which occurs when all moment inequalities hold as equalities.

**2.** 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.

**3.** 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 $1/n^{1/2}$-local alternatives.

[10] 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.

**4.** 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 alternative considered is such that the first inequality constraint is violated and $h_{1,1} + \Pi_{0,1}'\lambda \in (-\infty, 0)$, but the other $j = 2, ..., p$ inequality constraints are not violated and differ from being equalities by magnitudes 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 $j = 2, ..., p$. Let $\mu_1 = h_{1,1} + \Pi_{0,1}'\lambda$. 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 $Z^* = (Z_1^*, ..., Z_p^*)' \sim N(0_p, I_p)$, because $Z_j^* + \infty = \infty$ for $j = 2, ..., p$.

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 $j = 2, ..., p$, $\pi_{1,j} = \infty$ 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 non-violated moment inequalities, $p-1$, increases. Just one non-violated moment inequality, $p = 2$, increases the critical value from 2.71 to 4.25.

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

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

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 $\mu_1$. The first value of $\mu_1$ is zero, so the null hypothesis holds. 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 non-similarity of these tests. The last three values of $\mu_1$ 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 $p$.

TABLE I. Asymptotic Critical Values and Power of the Nominal .05 GMS Test Compared to Subsampling and PA Tests for Several Values of $\mu_1$ for Certain Alternatives

| | | Critical Values | Asy. Null Rej. Prob. | Asy. Power | | |
|---|---|---|---|---|---|---|
| | | | $\mu_1$ | | | |
| | p | $z_\alpha(p)$ | 0.00 | -1.645 | -2.170 | -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 |

**5.** 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 3 except that the $j = 2, ..., p$ 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 $j = 2, ..., p$ and $J_{g_1,0_d}$ is the distribution of $[Z_1^*]_-^2$ because $g_1 = (0, \infty, ..., \infty)'$. Hence, the probability limit of the subsampling critical value, $c_{g_1,0_d}(1 - \alpha)$, equals that of the GMS critical value and $AsyPow_{Sub}(\mu_1) = AsyPow_{GMS}(\mu_1)$. Everything else is the same as in Comment 4. Hence, in the present scenario, Table I applies but with the results for the subsampling test given by those of the GMS test.

**6.** 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, viz., $J_{h_1,\lambda}(c_{\pi_1}(\varphi 1-\alpha))$, $J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha))$, and $J_{h_1,\lambda}(c_{0_p,0_d}(1-\alpha))$, respectively, are less than $\alpha$. This occurs because these tests are not asymptotically similar on the boundary of the null hypothesis.[11] Lack of asymptotic

[11] For example, with a subsampling test, when a moment inequality is satisfied under the

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.

**7.** Rosen (2005) introduces 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.

# 9 Power Against Distant Alternatives

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

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

$$\begin{aligned} m^*_{n,j} &= E_{F_n} m_j(W_i,\theta_0)/\sigma_{F_n,j}(\theta_0) \text{ and} \\ \beta_n &= \max\{-m^*_{n,1},...,-m^*_{n,p},|m^*_{n,p+1}|,...,|m^*_{n,k}|\}. \end{aligned} \tag{9.1}$$

**Assumption DA.** (a) $n^{1/2}\beta_n \to \infty$.

(b) $\Omega(\theta_0,F_n) \to \Omega_1$ 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 non-negativity condition and is not $o(n^{-1/2})$ for $j=1,...,p$ or some moment equality term $m^*_{n,j}$ violates the zero condition and is not $o(n^{-1/2})$ for $j=p+1,...,k$. In contrast to Assumption DA, under Assumptions LA1-LA3 above, $n^{1/2}\beta_n \to \max\{-h_{1,1}-\Pi'_{0,1}\lambda,...,-h_{1,p}-\Pi'_{0,p}\lambda,|\Pi'_{0,p+1}\lambda|,...,|\Pi'_{0,k}\lambda|\} < \infty$.

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

---

alternative but is not sufficiently far from being an equality (i.e., is $O(b^{-1/2})$), then the subsampling critical value at best only *partially* takes this into account and, in consequence, does not deliver an asymptotically similar test.

**Theorem 4** *Under Assumptions* 1, 3, 6, *and* DA,

(a) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)) = 1$ *provided Assumption* GMS7 *holds,*

(b) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha)) = 1,$ *and*

(c) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c(\widehat{\Omega}_n(\theta_0), 1-\alpha)) = 1.$

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

**2.** The proof of Theorem 4 is in Appendix B.

# 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. (iii) The "pure" GEL test that uses a constant critical value (equal to $c_{GEL}(1-\alpha) = \sup_{\Omega\in\Psi_2} c(\Omega, 1-\alpha)$, 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 $\tau$, 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 $\tau$ exists when $\theta_0$ is the true value of $\theta$. This requires that $\tau$ 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 $\tau$ is replaced by the estimator $\widehat{\tau}_n(\theta)$, 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 (12.3) of Appendix A. These are high-level conditions that essentially just require that $\overline{m}_{n,j}(\theta, \widehat{\tau}_n(\theta))$ 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 Experiments

## 11.1 Introduction

In this section, we use simulation to investigate the finite-sample properties of GMS CS's and to compare them to some other methods in the literature. We consider the coverage probabilities (CP's) of the CS's for points in and not in the identified set. For points on the boundary of the identified set and for which all inequalities are binding (i.e., hold as equalities), the CP's should be close to the nominal level $1-\alpha$. For points on the boundary of the identified set and for which some inequality is not binding, the CP's should be greater than or equal to $1-\alpha$. Probabilities for these points indicate the non-similarity on the boundary of the CS's. For points in the interior of the identified set, the CP's should be greater than $1-\alpha$. For points that are not in the identified set, the CP's should be less than $1-\alpha$—the smaller, the better.

We consider two very simple models. The first is a particular case of the missing-data model considered in Imbens and Manski (2004) (IM). In this model, there is one parameter, two moment inequalities, and no moment equalities. We consider the GMS CS based on the MMM test statistic (i.e., the test function $S_1$) with the MMSC function $\varphi^{(5)}$ with $\eta(x) = x$ and $\kappa_n = (2.01 \ln \ln n)^{1/2}$ (i.e., the HQIC MMSC procedure).[12] We compare the GMS CS to the CS introduced by IM for this model, see IM for its definition, and to the subsampling CS based on the MMM test statistic and with subsample size $b = floor(n^{1/2})$. (We mention, but do not report, results for other values of $b$.) Rosen (2005) provides a comparison of the finite-sample properties of his proposed CS with that of IM.

The second model considered is the interval-outcome regression model of Manski and Tamer (2002). In this model, there are two parameters, two moment

[12] We take $\widehat{\sigma}^2_{n,j}(\theta) = n^{-1}\sum_{i=1}^n m_j^2(W_i,\theta)$ for $j = 1,...,k$, rather than $\widehat{\sigma}^2_{n,j}(\theta) = n^{-1}\sum_{i=1}^n (m_j(W_i,\theta) - \overline{m}_{n,j}(\theta))^2$. Results for the latter are similar.

inequalities, and no moment equalities. We compare the same GMS and subsampling procedures as defined above. (The IM CS does not apply to this model.)

All results reported are for CS's with nominal level .95. For both models, we report results for $n = 100$, $500$, and $1,000$. We take the number of simulation repetitions, $R$, to be $20,000$ for the GMS and IM CS's and $5,000$ for the subsampling CS's.[13] The reported CP's are the relative frequencies of coverage over the $R$ repetitions.

## 11.2 Missing-Data Model

In this model, $W_i = (Y_i, D_i)$ are i.i.d. for $i = 1, ..., n$ with $Y_i \sim U[0,1]$, $D_i \sim Bern[.85]$, and $Y_i$ and $D_i$ independent. The observations are $\{(Y_i D_i, D_i) : i \leq n\}$. Thus, $Y_i$ is not observed when $D_i = 0$. The parameter of interest is $\theta = EY_i$. Given that data are missing, the parameter $\theta$ is not identified. Two moment inequality functions used to bound $\theta$ are

$$\begin{pmatrix} m_1(W_i, \theta) \\ m_2(W_i, \theta) \end{pmatrix} = \begin{pmatrix} \theta - Y_i D_i \\ (1-\theta) - (1-Y_i) D_i \end{pmatrix}. \qquad (11.1)$$

When $\theta$ is the true parameter, we have $Em_1(W_i, \theta) = \theta - EY_i D_i \geq \theta - EY_i = 0$, where the inequality holds because $Y_i \geq 0$ and $D_i \leq 1$ and $Em_2(W_i, \theta) = (1-\theta) - E(1-Y_i)D_i \geq (1-\theta) - E(1-Y_i) = 0$, where the inequality holds because $1 - Y_i \geq 0$ and $D_i \leq 1$. Hence, the functions in (11.1) satisfy two moment inequalities.

For the data-generating process above, the identified set $[\theta_L, \theta_U]$ is $[.425, .575]$.[14] We consider the CP's of the CS's for the values $\theta_L = .425$, $\theta = .5$, and $\theta_H = .575$, which are in the identified set, and for the values $.9 \times \theta_L$ and $1.1 \times \theta_U$, which are not in the identified set.

Table II reports the CP's of the GMS, IM, and subsampling CS's with nominal level 95%. The Table shows that for $\theta$ values in the identified set the performance of the GMS and IM CS's is excellent for all sample sizes. Probabilities for the GMS CS for boundary $\theta$ points range from .948 to .951. In contrast, the subsampling CS over-covers by a noticeable amount for all sample sizes. Probabilities for the subsampling CS for boundary $\theta$ points range from .971 to .990. (This over-coverage is a finite-sample phenomenon because the subsampling asymptotic CP is .95 at both boundaries.) All three CS's cover $\theta = .5$, which lies in the interior

[13] The subsampling CS's are more computationally intensive than the moment selection CS's. Only $5,000$ repetitions are used for the moment selection results of Table III for $\theta$ not in the identified set.

[14] The identified set is determined by $\theta_L - EY_i D_i = 0$, i.e., $\theta_L = .5 \times .85 = .425$, and $(1-\theta_U) - E(1-Y_i)D_i = 0$, i.e., $\theta_U = 1 - E(1-Y_i)ED_i = 1 - .5 \times .85 = .575$.

of the identified set and is far from either boundary, with probability one. This is in accord with the asymptotic results.

TABLE II. Missing-Data Model: Finite-Sample Coverage Probabilities of Nominal 95% Confidence Intervals

| | | Coverage Probabilities for $\theta$ Values: | | | | |
|---|---|---|---|---|---|---|
| | | $\theta$ Values in Identified Set | | | $\theta$ Values Not in Identified Set | |
| $n$ | Type of Confidence Int. | $\theta_L = .425$ | $\theta = .5$ | $\theta_H = .575$ | $0.9 \times \theta_L$ | $1.1 \times \theta_H$ |
| | GMS | .951 | 1.0 | .948 | .619 | .445 |
| 100 | Imbens/Manski | .946 | 1.0 | .951 | .637 | .439 |
| | Subsampling | .981 | 1.0 | .990 | .791 | .667 |
| | GMS | .951 | 1.0 | .951 | .095 | .010 |
| 500 | Imbens/Manski | .949 | 1.0 | .950 | .094 | .007 |
| | Subsampling | .975 | 1.0 | .971 | .145 | .032 |
| | GMS | .951 | 1.0 | .949 | .006 | .000 |
| 1,000 | Imbens/Manski | .953 | 1.0 | .950 | .005 | .000 |
| | Subsampling | .972 | 1.0 | .971 | .008 | .000 |

For $\theta$ points not in the identified set, we want the CP of a CS to be as close to zero as possible. (A lower CP for such points translates into a shorter and more informative CS.) Table II shows that the GMS CS covers points not in the identified set with substantially lower probability than the subsampling CS when $n = 100$ (viz., .619 versus .791 and .445 versus .667) and with slightly lower probability for $n = 500$ (viz., .095 versus .145 and .010 versus .030). This is consistent with the asymptotic power comparisons given in Section 8. For points not in the identified set, the GMS and IM CS's have comparable CP's. For $n = 500$ and $1,000$, the CP's of all three CS's is sufficiently low that the differences between them are small.

As has been reported in other scenarios, subsampling CP's are sensitive to the choice of the subsample size $b$. Additional simulation results not reported here show that for smaller $b$ the subsampling CP's for $\theta$ in the identified set become slightly closer to the nominal level, while for larger subsample sizes they become closer to one. For $\theta$ not in the identified set, smaller $b$ reduces the subsampling CP's slightly and larger $b$ increases them slightly.

### 11.3 Interval-Outcome Regression Model

This model is a regression model with unobserved dependent variable $Y_i$:

$$Y_i = \theta_1 + Z_i\theta_2 + U_i, \tag{11.2}$$

where $(Z_i, U_i)$ are i.i.d. for $i = 1, ..., n$, $Z_i \sim N(1,1)$, $U_i \sim N(0,1)$, and $\theta = (\theta_1, \theta_2)$. The observations are $\{(Y_i^L, Y_i^H, Z_i) : i \leq n\}$, where $Y_i^L = floor(Y_i)$, $Y_i^H = ceil(Y_i)$, and so, $Y_i^L \leq Y_i \leq Y_i^H$ a.s. The parameter $\theta$ is not identified because $Y_i$ is not observed. The two moment inequality functions are

$$\begin{pmatrix} m_1(W_i, \theta) \\ m_2(W_i, \theta) \end{pmatrix} = \begin{pmatrix} \theta_1 + Z_i\theta_2 - Y_i^L \\ (Y_i^H - \theta_1 - Z_i\theta_2)Z_i^2 \end{pmatrix}. \tag{11.3}$$

When $\theta$ is the true parameter value, we have $Em_1(W_i, \theta) = \theta_1 + Z_i\theta_2 - EY_i^L \geq \theta_1 + Z_i\theta_2 - EY_i = 0$ and $Em_2(W_i, \theta) = E(Y_i^H - \theta_1 - Z_i\theta_2)Z_i^2 \geq E(Y_i - \theta_1 - Z_i\theta_2)Z_i^2 = 0.$[15] Thus, the functions in (11.3) satisfy two moment inequalities.

We consider the case where the true parameter is $\theta = (1, 1)$. In this case, the identified set consists of the $(\theta_1, \theta_2)$ values that satisfy:[16]

$$\theta_1 + \theta_2 \geq 1.5 \text{ and } 2\theta_1 + 4\theta_2 \leq 7. \tag{11.4}$$

We consider the CP's of the CS's for the $\theta$ values $(-.5, 2)$, $(1.5, 0)$, $(1, 1.25)$, and $(1, 1)$, which are all in the identified set. The point $(-.5, 2)$ is on the boundary of the identified set with both moment inequalities binding; $(1.5, 0)$ and $(1, 1.25)$ are on the boundary of the identified set with only one inequality binding in each case; and $(1, 1)$ is in the interior of the identified set. We also consider the CP's of the CS's for the $\theta$ values $(1.35, 0)$ and $(1, 1.375)$, which are not in the identified set. The point $(1.35, 0)$ violates the first inequality in (11.4) and satisfies the second. The reverse is true for the point $(1, 1.375)$.

Table III reports CP's for the interval-outcome regression model. Table III shows that the GMS CS performs very well at $\theta = (-.5, 2)$ (at which both inequalities are binding) and at $\theta = (1.5, 0)$ (at which only the first inequality is binding) with CP's ranging between .948 and .953. Its CP's at $\theta = (1, 1.25)$ (at which only the second inequality is binding) are somewhat higher with CP's

[15] In the second moment function, $Y_i^H - \theta_1 - Z_i\theta_2$ is multiplied by $Z_i^2$ to avoid perfect colinearity with $\theta_1 + Z_i\theta_2 - Y_i^L$ since $Y_i^H = Y_i^L + 1$ by definition. We do not consider optimal choices of moment functions for this model because such choices are not known and the results are only illustrative anyway.

[16] The identified set is determined by $\theta_1 + EZ_i\theta_2 - EY_i^L \geq 0$, i.e., $\theta_1 + \theta_2 \geq 1.5$, where $EY_i^L \approx 1.5$ (by numerical calculation), and $E(Y_i^H - \theta_1 - Z_i\theta_2)Z_i^2 \geq 0$, i.e., $EY_i^H Z_i^2 - EZ_i^2\theta_1 - EZ_i^3\theta_2 \geq 0$, where $EY_i^H Z_i^2 \approx 7$ (by numerical calculation), $EZ_i^2 = 2$, and $EZ_i^3 = 4$.

ranging between .957 and .963. Over-coverage in this case is not necessarily a finite-sample phenomenon because the CS is not asymptotically similar on the boundary of the identified set. For points on the boundary of the identified set, the CP's of the subsampling CS are not quite as good as for the GMS CS. At $\theta = (-.5, 2)$ they vary between .940 and .963; at $\theta = (1.5, 0)$, they vary between .972 and .986; at $\theta = (1, 1.25)$ they are comparable to those of the GMS CS. Both CS's cover the point $\theta = (1, 1)$ (which is in the interior of the identified set and not close to a boundary) with probability one. This is expected given that the asymptotic CP is one.

TABLE III. Interval-Outcome Regression Model: Finite-Sample Coverage Probabilities of Nominal 95% Confidence Sets for $(\theta_1, \theta_2)$

| | | Coverage Probabilities for $(\theta_1, \theta_2)$ Values: | | | | | |
|---|---|---|---|---|---|---|---|
| | | $(\theta_1, \theta_2)$ in Identified Set | | | | $(\theta_1, \theta_2)$ Not in Identified Set | |
| $n$ | Type of Confidence Int. | $(-.5, 2)$ | $(1.5, 0)$ | $(1, 1.25)$ | $(1, 1)$ | $(1.35, 0)$ | $(1, 1.375)$ |
| 100 | GMS | .953 | .948 | .963 | 1.0 | .719 | .626 |
| | Subsampling | .963 | .986 | .962 | 1.0 | .851 | .638 |
| 500 | GMS | .953 | .950 | .959 | 1.0 | .236 | .061 |
| | Subsampling | .944 | .980 | .962 | 1.0 | .375 | .067 |
| 1,000 | GMS | .951 | .951 | .957 | 1.0 | .056 | .002 |
| | Subsampling | .940 | .972 | .957 | 1.0 | .106 | .002 |

Next, we consider $\theta$ points not in the identified set. Table III shows that the GMS CS has noticeably lower CP at $\theta = (1.35, 0)$ than the subsampling CS (viz., .719 versus .851 and .236 versus .375). For $\theta = (1, 1.375)$, the two CS's have comparable CP's. These results are consistent with the power results of Section 8, which show that the GMS test has higher power at some points and equal power at other points compared to the subsampling test.

Similar comments regarding the sensitivity of the subsampling results to $b$ apply in this model as in the missing-data model.

In sum, the simulation results of this section are in accord with the asymptotic results. They show that the GMS CS has advantages relative to the subsampling CS. The GMS CS has CP's that are (i) closer to the nominal level and less non-similar on the boundary of the identified set and (ii) lower for points outside the identified set.

# 12 APPENDIX A

In Appendix A, 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. Lastly, we prove the results stated in the paper except for that of Theorem 4, which is given in Appendix B.

## 12.1 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 $\Omega$ 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) $P\left(S(Z + (m_1, 0_v), \Omega) \leq x\right) < P\left(S(Z + (m_1^*, 0_v), \Omega) \leq x\right)$ 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 $j = 1, ..., p$. Let $h_1(\theta, F) = (h_{1,1}(\theta, F), ..., h_{1,p}(\theta, F))'$ and $\Omega(\theta, F) = \lim_{n \to \infty} Corr_F(n^{1/2} \overline{m}_n(\theta))$.

**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 effected by a jump in a df. Assumption 4 is a mild continuity assumption. Assumption 5 is used for the $1/n^{1/2}$-local power results. Assumption 5(a) is a continuity/strictly increasing df condition that is the same as Assumption 2(a) except that $\ell$ can take negative values. Assumption 5(b) is a stochastically strictly-increasing condition. With a non-strict 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).

## 12.2 Alternative Parametrization and Dependent Observations

In this section we specify a one-to-one mapping between the parameters $(\theta,F)$ with parameter space $\mathcal{F}$ and a new parameter $\gamma=(\gamma_1,\gamma_2,\gamma_3)$ with corresponding parameter space $\Gamma$. We define $\gamma_1=(\gamma_{1,1},...,\gamma_{1,p})'\in R_+^p$ by writing the moment inequalities in (2.1) as moment equalities:

$$\sigma_{F,j}^{-1}(\theta)E_Fm_j(W_i,\theta)-\gamma_{1,j}=0 \text{ for } j=1,...,p, \tag{12.1}$$

where $\sigma_{F,j}(\theta)=\lim_{n\to\infty}Var_F^{1/2}(n^{1/2}\overline{m}_{n,j}(\theta))$ and $F$ is the distribution of the data. Let $\Omega=\Omega(\theta,F)=\lim_{n\to\infty}Corr_F(n^{1/2}\overline{m}_n(\theta))$, where $Corr_F(n^{1/2}\overline{m}_n(\theta))$ denotes the $k\times k$ correlation matrix of $n^{1/2}\overline{m}_n(\theta)$. (We only consider distributions $F$ for which the previous limits exist, see conditions (iv) and (v) of (12.2) below.) Let $\gamma_2=(\gamma_{2,1},\gamma_{2,2})=(\theta,vech_*(\Omega(\theta,F)))\in R^q$, where $vech_*(\Omega)$ denotes the vector of elements of $\Omega$ 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 $\tau_0$), we define $m_j(W_i,\theta)=m_j(W_i,\theta,\tau_0)$ and $\overline{m}_n(\theta)=\overline{m}_n(\theta,\widehat{\tau}_n(\theta))$.

For i.i.d. observations (and no preliminary estimator $\widehat{\tau}_n(\theta)$), the parameter space for $\gamma$ 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), $\gamma_1$ satisfies (12.1), $\gamma_2=(\theta,vech_*(\Omega(\theta,F)))$, 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 $\Gamma$ 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 $\gamma$ to $(\theta,F)$, $\Gamma$ also defines the parameter space $\mathcal{F}$ of $(\theta,F)$. Let $\Psi$ be a specified set of $k\times k$

correlation matrices. The parameter space $\Gamma$ is defined to include parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3) = (\gamma_1, (\theta, \gamma_{2,2}), F)$ that satisfy:

$$
\begin{aligned}
&\text{(i) } \theta \in \Theta, \\
&\text{(ii) } \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0 \text{ for } j = 1, ..., p, \\
&\text{(iii) } E_F m_j(W_i, \theta) = 0 \text{ for } j = p+1, ..., k, \\
&\text{(iv) } \sigma_{F,j}^2(\theta) = \lim_{n\to\infty} Var_F(n^{1/2}\overline{m}_{n,j}(\theta)) \text{ exists and lies in } (0,\infty) \text{ for } j = 1, ..., k, \\
&\text{(v) } \lim_{n\to\infty} Corr_F(n^{1/2}\overline{m}_n(\theta)) \text{ exists and equals } \Omega_{\gamma_{2,2}} \in \Psi, \\
&\text{(vi) } \{W_i : i \geq 1\} \text{ are stationary under } F,
\end{aligned} \tag{12.2}
$$

where $\gamma_1 = (\gamma_{1,1}, ..., \gamma_{1,p})'$ and $\Omega_{\gamma_{2,2}}$ is the $k \times k$ correlation matrix determined by $\gamma_{2,2}$.[17] Furthermore, $\Gamma$ must be restricted by enough additional conditions such that under any sequence $\{\gamma_{n,h} = (\gamma_{n,h,1}, (\theta_{n,h}, vech_*\ (\Omega_{n,h})),\ F_{n,h}) : n \geq 1\}$ of parameters in $\Gamma$ that satisfies $n^{1/2}\gamma_{n,h,1} \to h_1$ and $(\theta_{n,h}, vech_*(\Omega_{n,h})) \to h_2 = (h_{2,1}, h_{2,2})$ for some $h = (h_1, h_2) \in R_{+,\infty}^p \times R_{[\pm\infty]}^q$, we have

$$
\begin{aligned}
&\text{(vii) } A_n = (A_{n,1}, ..., A_{n,k})' \to_d Z_{h_{2,2}} \sim N(0_k, \Omega_{h_{2,2}}) \text{ as } n \to \infty, \text{ where} \\
&A_{n,j} = n^{1/2}(\overline{m}_{n,j}(\theta_{n,h}) - E_{F_{n,h}}\overline{m}_{n,j}(\theta_{n,h}))/\sigma_{F_{n,h},j}(\theta_{n,h}), \\
&\text{(viii) } \widehat{\sigma}_{n,j}(\theta_{n,h})/\sigma_{F_{n,h},j}(\theta_{n,h}) \to_p 1 \text{ as } n \to \infty \text{ for } j = 1, ..., k, \\
&\text{(ix) } \widehat{D}_n^{-1/2}(\theta_{n,h})\widehat{\Sigma}_n(\theta_{n,h})\widehat{D}_n^{-1/2}(\theta_{n,h}) \to_p \Omega_{h_{2,2}} \text{ as } n \to \infty, \text{ and} \\
&\text{(x) conditions (vii)-(ix) hold for all subsequences } \{w_n\} \text{ in place of } \{n\},
\end{aligned} \tag{12.3}
$$

where $\Omega_{h_{2,2}}$ is the $k \times k$ correlation matrix for which $vech_*(\Omega_{h_{2,2}}) = h_{2,2}$, $\widehat{\sigma}_{n,j}^2(\theta) = [\widehat{\Sigma}_n(\theta)]_{jj}$ for $1 \leq j \leq k$ and $\widehat{D}_n(\theta) = Diag\{\widehat{\sigma}_{n,1}^2(\theta), ..., \widehat{\sigma}_{n,k}^2(\theta)\}$ $(= Diag(\widehat{\Sigma}_n(\theta)))$.[18]

For example, for i.i.d. observations, conditions (i)-(vi) of (2.2) imply conditions (i)-(vi) of (12.2). 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 (12.3). For details of the proof, see 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 (12.2) that ensure that $\Gamma$ is such that any sequences $\{\gamma_{n,h} : n \geq 1\}$ in $\Gamma$ satisfy (12.3). For brevity, we do not do so here.

[17] In AG4, a strong mixing condition is imposed in condition (v) of (12.2). 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.

[18] Condition (x) of (12.3) requires that conditions (vii)-(ix) must hold under any sequence of parameters $\{\gamma_{w_n,h} : n \geq 1\}$ that satisfies the conditions preceding (12.3) with $n$ replaced by $w_n$.

## 12.3 Proof of the Asymptotic Size Result for GMS

The proof of Lemma 1 is given at the end of this subsection.

The following Lemmas are used in the proof of Theorem 1. The first Lemma uses the following notation. Suppose $\pi = (\pi_1, \pi_2) \in R^p_{+,\infty} \times R^q_{[\pm\infty]}$, where $\pi_2 = (\pi_{2,1}, \pi_{2,2})$ and $\pi_{2,2} = vech_*(\Omega_{\pi_{2,2}})$ for some $k \times k$ correlation matrix $\Omega_{\pi_{2,2}}$. Given $\pi$, define $\pi^*_{1,j} = \infty$ if $\pi_{1,j} > 0$ and $\pi^*_{1,j} = 0$ if $\pi_{1,j} = 0$ for $j = 1, ..., p$ and let $\pi^*_1 = (\pi^*_{1,1}, ..., \pi^*_{1,p})'$. Define $\pi^* = (\pi^*_1, \pi_2)$ and let $c_{\pi^*}(1-\alpha)$ denote the $1-\alpha$ quantile of $S(\Omega^{1/2}_{\pi_{2,2}} Z^* + (\pi^*_1, 0_v), \Omega_{\pi_{2,2}})$, where $Z^* \sim N(0_k, I_k)$ and by definition if $\pi^*_{1,j} = \infty$ then the $j$th element of $\Omega^{1/2}_{\pi_{2,2}} Z^* + (\pi^*_1, 0_v)$ equals $\infty$ for $j = 1, ..., p$.

**Lemma 2** *Suppose Assumptions* 1-3, GMS1, *and* GMS2 *hold and* $0 < \alpha < 1/2$. *Let* $\{\gamma_{n,h} = (\gamma_{n,h,1}, \gamma_{n,h,2}, \gamma_{n,h,3}) : n \geq 1\}$ *be a sequence of points in* $\Gamma$ *that satisfies* (i) $n^{1/2}\gamma_{n,h,1} \to h_1$ *for some* $h_1 \in R^p_{+,\infty}$, (ii) $\kappa_n^{-1} n^{1/2} \gamma_{n,h,1} \to \pi_1$ *for some* $\pi_1 \in R^p_{+,\infty}$, *and* (iii) $\gamma_{n,h,2} \to h_2$ *for some* $h_2 \in R^q_{[\pm\infty]}$. *Let* $h = (h_1, h_2)$, $\pi = (\pi_1, \pi_2)$, *and* $\pi_2 = h_2$. *Then,*

(a) $\widehat{c}_n(\theta_{n,h}, 1-\alpha) \geq c^*_n$ *a.s. for all* $n$ *for a sequence of random variables* $\{c^*_n : n \geq 1\}$ *that satisfies* $c^*_n \to_p c_{\pi^*}(1-\alpha)$ *under* $\{\gamma_{n,h} : n \geq 1\}$, *where* $\gamma_{n,h,2} = (\theta_{n,h}, \gamma_{n,h,2,2})$,

(b) $\liminf_{n\to\infty} P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq \widehat{c}_n(\theta_{n,h}, 1-\alpha)) \geq 1-\alpha$, *and*

(c) *for any subsequence* $\{w_n : n \geq 1\}$ *of* $\{n\}$, *the results of parts* (a) *and* (b) *hold with* $w_n$ *in place of* $n$ *provided conditions* (i)-(iii) *above hold with* $w_n$ *in place of* $n$.

**Lemma 3** *Suppose Assumptions* 1-3, 7, *and* GMS1-GMS4 *hold and* $0 < \alpha < 1/2$. *Let* $(\theta^*, F^*)$ *be an element of* $\mathcal{F}$ *for which Assumption* 7 *applies, let* $\gamma^*$ *be the value in* $\Gamma$ *that corresponds to* $(\theta^*, F^*) \in \mathcal{F}$, *and let* $h^* = (h^*_1, h^*_2)$ *be defined by* $h^*_1 = (h^*_{1,1}, ..., h^*_{1,p})'$, *where* $h^*_{1,j} = \infty$ *if* $\gamma^*_{1,j} > 0$ *and* $h^*_{1,j} = 0$ *if* $\gamma^*_{1,j} = 0$ *for* $j = 1, ..., p$, *and* $h^*_2 = \gamma^*_2$. *Let* $c_{h^*}(1-\alpha)$ *denote the* $1-\alpha$ *quantile of the distribution of* $S(\Omega^{1/2}_{h^*_{2,2}} Z^* + (h^*_1, 0_v), \Omega_{h^*_{2,2}})$. *When the true distribution is determined by* $\gamma^*$ *for all* $n$, *we have*

(a) $\widehat{c}_n(\theta^*, 1-\alpha) \to_p c_{h^*}(1-\alpha)$ *and*

(b) $\lim_{n\to\infty} P_{\gamma^*}(T_n(\theta^*) \leq \widehat{c}_n(\theta^*, 1-\alpha)) = 1-\alpha$.

**Proof of Theorem 1.** First, we prove part (a). Let $CP_n(\gamma) = P_\gamma(T_n(\theta) \leq \widehat{c}_n(\theta, 1-\alpha))$, where $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, $\gamma_2 = (\gamma_{2,1}, \gamma_{2,2})$, and $\gamma_{2,1} = \theta$. Let $\{\gamma^*_n = (\gamma^*_{n,1}, \gamma^*_{n,2}, \gamma^*_{n,3}) \in \Gamma : n \geq 1\}$ be a sequence such that $\liminf_{n\to\infty} CP_n(\gamma^*_n) = \liminf_{n\to\infty} \inf_{\gamma\in\Gamma} CP_n(\gamma)$ $(= AsyCS)$. Such a sequence always exists. Let $\{u_n : n \geq 1\}$ be a subsequence of $\{n\}$ such that $\lim_{n\to\infty} CP_{u_n}(\gamma^*_{u_n})$ exists and equals $\liminf_{n\to\infty} CP_n(\gamma^*_n) = AsyCS$. Such a subsequence always exists.

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

$$\begin{aligned} &\kappa_{w_n}^{-1} w_n^{1/2}\gamma^*_{w_n,1,j} \to 0 \text{ and} \\ &\quad w_n^{1/2}\gamma^*_{w_n,1,j} \to h_{1,j} \text{ for some } h_{1,j} \in R_+. \end{aligned} \quad (12.4)$$

If (2) holds, then either 2(a) $\limsup_{n\to\infty} \kappa_{u_n}^{-1} u_n^{1/2}\gamma^*_{u_n,1,j} < \infty$ or 2(b) $\limsup_{n\to\infty} \kappa_{u_n}^{-1} u_n^{1/2}\gamma^*_{u_n,1,j} = \infty$. If 2(a) holds, then for some subsequence $\{w_n\}$ of $\{u_n\}$,

$$\begin{aligned} &\kappa_{w_n}^{-1} w_n^{1/2}\gamma^*_{w_n,1,j} \to \pi_{1,j} \text{ for some } \pi_{1,j} \in R_+ \text{ and} \\ &\quad w_n^{1/2}\gamma^*_{w_n,1,j} \to h_{1,j}, \text{ where } h_{1,j} = \infty. \end{aligned} \quad (12.5)$$

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

$$\begin{aligned} &\kappa_{w_n}^{-1} w_n^{1/2}\gamma^*_{w_n,1,j} \to \pi_{1,j}, \text{ where } \pi_{1,j} = \infty, \text{ and} \\ &\quad w_n^{1/2}\gamma^*_{w_n,1,j} \to h_{1,j}, \text{ where } h_{1,j} = \infty. \end{aligned} \quad (12.6)$$

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

$$\gamma^*_{w_n,2} \to h_2 \text{ for some } h_2 \in \mathrm{cl}(\Gamma_2). \quad (12.7)$$

By taking successive subsequences over the $p$ components of $\gamma^*_{u_n,1}$ and $\gamma^*_{u_n,2}$, we find that there exists a subsequence $\{w_n\}$ of $\{u_n\}$ such that for each $j = 1, ..., p$ exactly one of the cases (12.4)-(12.6) applies and (12.7) holds. In consequence, conditions (i)-(iii) of Lemma 2 hold. Hence,

$$\liminf_{n\to\infty} CP_{w_n}(\gamma^*_{w_n}) \geq 1 - \alpha \quad (12.8)$$

by Lemma 2. Since $\lim_{n\to\infty} CP_{u_n}(\gamma^*_{u_n}) = AsyCS$ and $\{w_n\}$ is a subsequence of $\{u_n\}$, we have $\lim_{n\to\infty} CP_{w_n}(\gamma^*_{w_n}) = AsyCS$. This and (12.8) yield the result of part (a).

Part (b) follows from part (a) and Lemma 3(b) because $AsyCS \leq \lim_{n\to\infty} P_{\gamma^*}(T_n(\theta^*) \leq \widehat{c}_n(\theta^*, 1-\alpha))$.

Now, we prove part (c) of the Theorem. By assumption, $v = 0$. Under Assumption M, the sequence of constant true values $\{(\theta^*, F^*) \in \mathcal{F} : n \geq 1\}$ satisfies $n^{1/2}E_{F^*}m_j(W_i, \theta^*)/\sigma_{F^*,j}(\theta^*) \to \infty$ for $j = 1, ..., p$. Let $\gamma^* = (\gamma_1^*, \gamma_2^*, F^*) \in \Gamma$ correspond to $(\theta^*, F^*) \in \mathcal{F}$, where $\gamma_2^* = (\theta^*, \gamma^*_{2,2})$. Define $h^* = (\infty^p, \gamma_2^*)$. As in the proof of part (b) of Lemma 2 below, we have $T_n(\theta^*) \to_d J_{h^*}$ under $\{\gamma^*\}$, where

$J_{h^*}$ is the distribution of $S(Z_{h^*_{2,2}} + (h^*_1, 0_v), \Omega_{h^*_{2,2}})$, where $Z_{h^*_{2,2}} \sim N(0_k, \Omega_{h^*_{2,2}})$. Furthermore, $J_{h^*}(x) = 1$ for $x \geq 0$ because $S(Z_{h^*_{2,2}} + \infty^p, \Omega_{h^*_{2,2}}) = S(\infty^p, \Omega_{h^*_{2,2}}) = 0$ by Assumption 3. Using these results, we obtain

$$\begin{aligned} AsyMaxCP &\geq \limsup_{n\to\infty} P_{\gamma^*}(T_n(\theta^*) \leq \widehat{c}_n(\theta^*, 1-\alpha)) \\ &\geq \limsup_{n\to\infty} P_{\gamma^*}(T_n(\theta^*) \leq 0) = J_{h^*}(0) = 1, \end{aligned} \tag{12.9}$$

where the first inequality follows from the definition of $AsyMaxCP$ and the second inequality holds by Assumption 1(c). □

**Proof of Lemma 2.** First, suppose $c_{\pi^*}(1-\alpha) = 0$. In this case, define $c^*_n = 0$ and we have

$$c_n(\widehat{\theta}_{n,h}, 1-\alpha) \geq c^*_n \to_p c_{\pi^*}(1-\alpha), \tag{12.10}$$

where the inequality holds by Assumption 1(c), which establishes part (a) for this case.

Next, suppose $c_{\pi^*}(1-\alpha) > 0$. For $(\xi, \Omega) \in R^k \times \Psi$, let $\varphi^*(\xi, \Omega)$ denote the $k$-vector whose $j$th element is

$$\varphi^*_j(\xi, \Omega) = \begin{cases} \varphi_j(\xi, \Omega) & \text{if } \pi_{1,j} = 0 \text{ and } j = 1, ..., p \\ \infty & \text{if } \pi_{1,j} > 0 \text{ and } j = 1, ..., p \\ 0 & \text{if } j = p+1, ..., k. \end{cases} \tag{12.11}$$

By construction,

$$\varphi^*(\xi_n(\theta_{n,h}), \widehat{\Omega}_n(\theta_{n,h})) \geq \varphi(\xi_n(\theta_{n,h}), \widehat{\Omega}_n(\theta_{n,h})) \text{ a.s.}[Z^*] \text{ for all } n. \tag{12.12}$$

Let $c^*_n$ denote the $1-\alpha$ quantile of the df of $S(\widehat{\Omega}^{1/2}_n(\theta_{n,h})Z^* + \varphi^*(\xi_n(\theta_{n,h}), \widehat{\Omega}_n(\theta_{n,h})), \widehat{\Omega}_n(\theta_{n,h}))$, where $Z^*$ is random and $(\xi_n(\theta_{n,h}), \widehat{\Omega}_n(\theta_{n,h}))$ is fixed. Then, $\widehat{c}_n(\theta_{n,h}, 1-\alpha) \geq c^*_n$ a.s. for all $n$ by (12.12) and Assumption 1(a).

We now show that $c^*_n \to_p c_{\pi^*}(1-\alpha) > 0$. Under $\{\gamma_{n,h} : n \geq 1\}$, we have

$$\begin{aligned} &\kappa_n^{-1} n^{1/2} \widehat{D}_n^{-1/2}(\theta)\overline{m}_n(\theta_{n,h}) \\ &= \kappa_n^{-1} \widehat{D}_n^{-1/2}(\theta_{n,h}) D^{1/2}(\theta_{n,h}, F_{n,h}) \left(A_n + (n^{1/2}\gamma_{n,h,1}, 0_v)\right) \\ &= o_p(1) + (I_k + o_p(1))(\kappa_n^{-1} n^{1/2}\gamma_{n,h,1}, 0_v) \to_p (\pi_1, 0_v), \end{aligned} \tag{12.13}$$

where the first equality holds by the definitions of $\gamma_{n,h,1}$ and $A_n$ in (12.1) and (12.3), the second equality holds using $\kappa_n \to \infty$ and conditions (vii) and (viii) of (12.3), which apply by conditions (i) and (iii) of the Lemma, and the convergence holds using condition (ii) of the Lemma. This and condition (ix) of (12.3) yield: under $\{\gamma_{n,h} : n \geq 1\}$,

$$(\xi_n(\theta_{n,h}), \widehat{\Omega}_n(\theta_{n,h})) \rightarrow_p ((\pi_1, 0_v), \Omega_{\pi_{2,2}}). \tag{12.14}$$

Consider $j$ for which $\pi_{1,j} = 0$ and $j = 1, ..., p$. For notational simplicity, let $\Omega_0$ denote $\Omega_{\pi_{2,2}}$. Then, as $(\xi, \Omega) \rightarrow ((\pi_1, 0_v), \Omega_0)$,

$$\varphi_j^*(\xi, \Omega) = \varphi_j(\xi, \Omega) \rightarrow \varphi_j((\pi_1, 0_v), \Omega_0) = 0 \text{ a.s.}[Z^*], \tag{12.15}$$

where the first equality holds by (12.11), the convergence holds by Assumption GMS1(a), and the last equality holds by Assumption GMS1(b).

Next, consider $j$ for which $\pi_{1,j} > 0$ for some $j = 1, ..., p$. Then, $\varphi_j^*(\xi, \Omega) = \infty$ by the definition in (12.11). For $j = p+1, ..., k$, $\varphi_j^*(\xi, \Omega) = 0$ by definition. These results, (12.15), and Assumption 1(d) give: for $x > 0$, as $(\xi, \Omega) \rightarrow ((\pi_1, 0_v), \Omega_0)$,

$$\begin{aligned} S(\Omega^{1/2}Z^* + \varphi^*(\xi, \Omega), \Omega) &\rightarrow S(\Omega_0^{1/2}Z^* + \varphi^*((\pi_1, 0_v), \Omega_0), \Omega_0) \text{ a.s.}[Z^*], \\ 1\left(S(\Omega^{1/2}Z^* + \varphi^*(\xi, \Omega), \Omega) \leq x\right) &\rightarrow 1\left(S(\Omega_0^{1/2}Z^* + \varphi^*((\pi_1, 0_v), \Omega_0), \Omega_0) \leq x\right) \\ &\qquad \text{a.s.}[Z^*], \text{ and} \qquad (12.16) \\ P\left(S(\Omega^{1/2}Z^* + \varphi^*(\xi, \Omega), \Omega) \leq x\right) &\rightarrow P\left(S(\Omega_0^{1/2}Z^* + \varphi^*((\pi_1, 0_v), \Omega_0), \Omega_0) \leq x\right). \end{aligned}$$

The third convergence result of (12.16) holds by the second result and the bounded convergence theorem. The second convergence result of (12.16) follows from the first result provided

$$P\left(S(Z + \varphi^*((\pi_1, 0_v), \Omega_0), \Omega_0) = x\right) = P\left(S(Z + (\pi_1^*, 0_v), \Omega_0) = x\right) = 0, \tag{12.17}$$

where $Z = \Omega_0^{1/2}Z^* \sim N(0_k, \Omega_0)$. The first equality in (12.17) holds because $[(\pi_1^*, 0_v)]_j = \infty$ if $\pi_{1,j}^* = \infty$ by definition and $[(\pi_1^*, 0_v)]_j = 0 = \varphi_j((\pi_1, 0_v), \Omega_0) = \varphi_j^*((\pi_1, 0_v), \Omega_0)$ if $\pi_{1,j}^* = \pi_{1,j} = 0$ using Assumption GMS1(b) and (12.11). The second equality in (12.17) holds because the df of $S(Z+(\pi_1^*, 0_v), \Omega_0)$ is continuous and strictly increasing for $x > 0$ by Assumptions 2(a) and 2(b) unless $v = 0$ and $\pi_1^* = \infty^p$. The latter does not hold because, if $v = 0$ and $\pi_1^* = \infty^p$, then $S(Z + (\pi_1^*, 0_v), \Omega_0) = 0$ and $c_{\pi^*}(1 - \alpha) = 0$), which contradicts the assumption that $c_{\pi^*}(1 - \alpha) > 0$.

In sum, (12.16) shows that $P(S(\Omega^{1/2}Z^* + \varphi^*(\xi, \Omega), \Omega) \leq x)$ is a continuous function of $(\xi, \Omega)$ at $((\pi_1, 0_v), \Omega_0)$. This, (12.14), and Slutsky's Theorem combine to give: under $\{\gamma_{n,h} : n \geq 1\}$,

$$\begin{aligned} L_n(x) &= P\left(S(\widehat{\Omega}_n^{1/2}(\theta_{n,h})Z^* + \varphi^*(\xi_n(\theta_{n,h}), \widehat{\Omega}_n(\theta_{n,h})), \widehat{\Omega}_n(\theta_{n,h})) \leq x\right) \\ &\rightarrow_p P\left(S(\Omega_0^{1/2}Z^* + \varphi^*((\pi_1, 0_v), \Omega_0), \Omega_0) \leq x\right) = L(x) \end{aligned} \tag{12.18}$$

for all $x > 0$, where $P(\cdot)$ denotes conditional probability given $(\xi_n(\theta_{n,h}), \widehat{\Omega}_n(\theta_{n,h}))$ in (12.18) and hence is a random probability. By definition, $c_n^*$ is the $1-\alpha$ quantile of $L_n(x)$ and $c_{\pi^*}(1-\alpha)$ is the $1-\alpha$ quantile of $L(x)$. By Lemma 5 of AG1, given that (12.18) holds for all $x$ in a neighborhood of $c_{\pi^*}(1-\alpha) > 0$ and $L(x)$ is continuous and strictly increasing at $x = c_{\pi^*}(1-\alpha)$ (see the previous paragraph), we have $c_n^* \to_p c_{\pi^*}(1-\alpha)$. This completes the proof of part (a).

Part (b) is proved as follows. First, conditions (i) and (ii) of the Lemma imply that if $\pi_{1,j} > 0$ then $h_{1,j} = \infty$ and $\pi_{1,j}^* = \infty$, and if $\pi_{1,j} = 0$ then $h_{1,j} \geq 0$ and $\pi_{1,j}^* = 0$. Thus, we have

$$\begin{aligned} \pi_1^* &\leq h_1, \\ S(\Omega_{h_{2,2}}^{1/2} Z^* + (\pi_1^*, 0_v), \Omega_{h_{2,2}}) &\geq S(\Omega_{h_{2,2}}^{1/2} Z^* + (h_1, 0_v), \Omega_{h_{2,2}}), \text{ and} \\ c_{\pi^*}(1-\alpha) &\geq c_h(1-\alpha), \end{aligned} \tag{12.19}$$

where $c_h(1-\alpha)$ denotes the $1-\alpha$ quantile of $S(\Omega_{h_{2,2}}^{1/2} Z^* + (h_1, 0_v), \Omega_{h_{2,2}})$, the second inequality holds by the first inequality and Assumption 1(a), and the third inequality holds by the second.

Second, by the verification of Assumption B0 in AG4, we have

$$T_n(\theta_{n,h}) \to_d J_h \text{ under } \{\gamma_{n,h}\}, \tag{12.20}$$

where $J_h$ is the distribution of $S(\Omega_{h_{2,2}}^{1/2} Z^* + (h_1, 0_v), \Omega_{h_{2,2}})$. This result is obtained by using Assumption 1(b) to write

$$T_n(\theta_{n,h}) = S\Big(\widehat{D}_n^{-1/2}(\theta_{n,h}) n^{1/2} \overline{m}_n(\theta_{n,h}),\ \widehat{D}_n^{-1/2}(\theta_{n,h}) \widehat{\Sigma}_n(\theta_{n,h}) \widehat{D}_n^{-1/2}(\theta_{n,h})\Big). \tag{12.21}$$

If any element of $h_1$ equals $\infty$, then it can be shown using (12.3) that the corresponding element of $\widehat{D}_n^{-1/2}(\theta_{n,h}) n^{1/2} \overline{m}_n(\theta_{n,h})$ diverges in probability to $\infty$. Hence, $\widehat{D}_n^{-1/2}(\theta_{n,h}) n^{1/2} \overline{m}_n(\theta_{n,h})$ does not converge in distribution to a proper finite random vector and the continuous mapping theorem cannot be applied to obtain the asymptotic distribution of the right-hand side of (12.21). The verification of Assumption B0 in AG4 avoids this problem by (i) considering a transformation of $\widehat{D}_n^{-1/2}(\theta_{n,h}) n^{1/2} \overline{m}_n(\theta_{n,h})$ that converges in distribution even if some elements of $h_1$ equal $\infty$, (ii) writing the right-hand side of (12.21) as a continuous function of this transformation, and (iii) applying the continuous mapping theorem to the transformation.

We now have

$$\liminf_{n\to\infty} P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq \widehat{c}_n(\theta_{n,h}, 1-\alpha))$$

$$\geq \liminf_{n\to\infty} P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq c_n^*)$$
$$\geq J_h(c_{\pi^*}(1-\alpha)-), \tag{12.22}$$

where $J_h(x-)$ denotes the limit from the left of $J_h(\cdot)$ at $x$, the first inequality holds because $\widehat{c}_n(\theta_{n,h}, 1-\alpha) \geq c_n^*$ a.s. and the second inequality holds by part (a) of the Lemma and (12.20).

Suppose $c_{\pi^*}(1-\alpha) > 0$, then

$$J_h(c_{\pi^*}(1-\alpha)-) = J_h(c_{\pi^*}(1-\alpha)) \geq 1-\alpha \tag{12.23}$$

and part (b) of the Lemma holds, where the equality holds because $J_h(x)$ is continuous for all $x > 0$ by Assumption 2(a) and the inequality holds by (12.19).

Next, suppose $c_{\pi^*}(1-\alpha) = 0$. This implies that $c_h(1-\alpha) = 0$ by (12.19) and Assumption 1(c). The conditions $c_h(1-\alpha) = 0$ and $0 < \alpha < 1/2$ are consistent with Assumption 2(c) only if $v = 0$. Given $v = 0$, under $\{\gamma_{n,h} : n \geq 1\}$, we have

$$\begin{aligned}
& P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq 0) \\
= \; & P_{\gamma_{n,h}}(n^{1/2}\overline{m}_{n,j}(\theta_{n,h})/\sigma_{F_{n,h},j}(\theta_{n,h}) \geq 0 \text{ for all } j = 1, ..., p) \\
= \; & P_{\gamma_{n,h}}(A_{n,j} + n^{1/2}\gamma_{n,h,1,j}) \geq 0 \text{ for all } j = 1, ..., p) \\
\to \; & P([\Omega_{h_{2,2}}^{1/2} Z^*]_j + h_{1,j} \geq 0 \text{ for all } j = 1, ..., p) \\
= \; & P(S(\Omega_{h_{2,2}}^{1/2} Z^* + h_1, \Omega_{h_{2,2}}) \leq 0) \\
= \; & J_h(0) \geq J_h(c_h(1-\alpha)) \geq 1-\alpha,
\end{aligned} \tag{12.24}$$

where the first equality holds by Assumptions 1(b) and 3, the second equality and the convergence hold by (12.3), the third equality holds by Assumption 3, the fourth equality holds by the definition of $J_h$, the first inequality holds because $c_h(1-\alpha) \geq 0$ (note that $c_h(1-\alpha) = 0$ here, but the argument in (12.24) is applied below to a case in which one only knows that $c_h(1-\alpha) \geq 0$), and the second inequality holds by the definition of $c_h(1-\alpha)$. This completes the proof of part (b).

The proof of part (c) is the same as that for parts (a) and (b) with $w_n$ in place of $n$. □

**Proof of Lemma 3.** Conditions (i)-(iii) of Lemma 2 hold with $\gamma_{n,h} = \gamma^*$ for all $n$, $h = h^*$, and $\pi = h^*$ because $\kappa_n^{-1} n^{1/2} \to \infty$ by Assumption GMS4. Each element of $\pi_1$ is either zero or infinity. Thus, the vector $\pi^*$ that depends on $\pi$ and is defined preceding Lemma 2 equals $\pi$. Now, (12.14) in the proof of Lemma 2 applies with $\theta_{n,h} = \theta^*$, $\pi_1 = h_1^*$, and $\Omega_{\pi_{2,2}} = \Omega_{h_{2,2}^*}$.

Equation (12.15) applies (with the first quantity on the left-hand side deleted) for all $j = 1, ..., p$ for which $\pi_{1,j} = 0$. In addition, we have: as $(\xi, \Omega) \rightarrow ((\pi_1, 0_v), \Omega_{h_{2,2}^*})$, $\varphi_j(\xi, \Omega) \rightarrow \infty$ a.s.$[Z^*]$ for all $j = 1, ..., p$ for which $\pi_{1,j} = \infty$ by Assumption GMS3. Given these results, (12.16) and (12.17) hold with $\varphi^*(\xi, \Omega)$ and $S(\Omega_0^{1/2} Z^* + \varphi^*((\pi_1, 0_v), \Omega_0), \Omega_0)$ replaced by $\varphi(\xi, \Omega)$ and $S(\Omega_0^{1/2} Z^* + \varphi((h_1^*, 0_v), \Omega_{h_{2,2}^*}), \Omega_{h_{2,2}^*})$, respectively, and the second equality in (12.17) holds because $c_{h^*}(1-\alpha) = c_{\pi^*}(1-\alpha) > 0$. (The case $c_{h^*}(1-\alpha) = 0$ does not occur because the df of $S(\Omega_{h_{2,2}^*}^{1/2} Z^* + (h_1^*, 0_v), \Omega_{h_{2,2}^*})$ at $x < 0$ is zero by Assumption 1(c), the df at $x = c_{h^*}(1-\alpha) = 0$ is zero by continuity (Assumption 7), the latter implies that the df is less than $1-\alpha$ for $x > 0$, and the latter implies that $c_{h^*}(1-\alpha) > 0$.) The remainder of the proof of part (a) is the same as that given in (12.18) but with $\widehat{c}_n(\theta^*, 1-\alpha))$ in place of $c_n^*$.

To prove part (b), we note that the asymptotic distribution of $T_n(\theta^*)$ is $S(\Omega_{h_{2,2}^*}^{1/2} Z^* + (h_1^*, 0_v), \Omega_{h_{2,2}^*})$ under $\{\gamma^* : n \geq 1\}$ by the verification of Assumption B0 in AG1, see (12.20) and the discussion following it. The df of $S(\Omega_{h_{2,2}^*}^{1/2} Z^* + (h_1^*, 0_v), \Omega_{h_{2,2}^*})$ is continuous and strictly increasing at $c_{h^*}(1-\alpha) > 0$ by Assumptions 2(a) and 2(b) unless $v = 0$ and $h_1^* = \infty^p$. The latter does not hold by the argument given in the proof of Lemma 2 just below (12.17) because $c_{h^*}(1-\alpha) > 0$. These results and $\widehat{c}_n(\theta^*, 1-\alpha) \rightarrow_p c_{h^*}(1-\alpha)$ establish part (b). $\square$

**Proof of Lemma 1.** Assumptions 1-4 hold by Lemma 1 of AG4. For $S_1$, Assumption 5(a) holds by the same arguments as for Assumption 2 given in the proof of Lemma 1 of AG4. Assumption 5(b) holds with a non-strict inequality by Assumption 1(a) and the fact that $Z + (m_1, 0_v)$ is stochastically strictly increasing in $m_1 \in R_{+,\infty}^p$. Assumption 5(b) holds for $S_1$ with a strict inequality because $S_1(Z + (m_1^*, 0_v), \Omega)$ is strictly stochastically less than $S_1(Z + (m_1, 0_v), \Omega)$ on $R_+$ for $m_1 < m_1^*$. Assumption 6 holds immediately for $S_1$ with $\tau = 2$.

For $S_2$, Assumptions 5(a)(i) and 5(a)(ii) hold by the same arguments as for Assumptions 2(a) and 2(b) given in the proof of Lemma 1 of AG4. Assumption 5(b) holds for $S_2$ by the same argument as for $S_1$. Assumption 6 holds immediately for $S_2$ with $\tau = 2$. The verification of Assumptions 1-6 for $S_3$ is essentially the same as that for $S_1$. $\square$

## 12.4 Proofs for Local Alternatives

Theorem 2 follows immediately from Lemmas 4 and 5 below. Theorem 3(a) and (c) do likewise from Lemmas 5-8. Theorem 3(b) follows from Lemmas 6 and 7, where one takes $g_1 = 0_p$ in Lemma 7 and one notes that $c_{0_p,0_d}(1-\alpha) > 0$ by Assumption 2(c) and $\alpha \in (0, 1/2)$.

In each of Lemmas 4-8, the parameter space $\mathcal{F}$ for $(\theta, F)$ is assumed to satisfy (2.2) (which implies that the observations are i.i.d.). In the Lemmas that involve subsampling, it is assumed that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. Let $Z \sim N(0_k, \Omega_0)$.

**Lemma 4** *Under Assumptions* 1-3, 5(a), GMS2, GMS3, LA1–LA2, LA4, *and* LA5,
(a) $T_n(\theta_0) \to_d S(Z + (h_1, 0_v) + \Pi_0 \lambda, \Omega_0) \sim J_{h_1,\lambda}$,
(b) $\widehat{c}_n(\theta_0, 1-\alpha) \to_p c_{\pi_1}(\varphi, 1-\alpha)$ *and*
(c) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)) = J_{h_1,\lambda}(c_{\pi_1}(\varphi, 1-\alpha))$.

**Lemma 5** *Under Assumptions* 1-3, 5(a), LA1–LA2, *and* LA6,
(a) $T_n(\theta_0) \to_d S(Z + (h_1, 0_v) + \Pi_0 \lambda, \Omega_0) \sim J_{h_1,\lambda}$,
(b) $T_b(\theta_0) \to_d S(Z + (g_1, 0_v), \Omega_0) \sim J_{g_1,0_d}$,
(c) $c_{n,b}(\theta_0, 1-\alpha) \to c_{g_1,0_d}(1-\alpha)$, *and*
(d) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha)) = J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha))$.

**Lemma 6** *Under Assumptions* 1, 4, 5(a), *and* LA1–LA2,
(a) $T_n(\theta_0) \to_d S(Z + (h_1, 0_v) + \Pi_0 \lambda, \Omega_0) \sim J_{h_1,\lambda}$,
(b) $c(\widehat{\Omega}_n(\theta_0), 1-\alpha) \to c_{0_p,0_d}(1-\alpha)$, *and*
(c) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c(\widehat{\Omega}_n(\theta_0), 1-\alpha)) = J_{h_1,\lambda}(c_{0_p,0_d}(1-\alpha))$.

The next Lemma uses the following notation. Let $g_1 = (g_{1,1}, ..., g_{1,p})'$ be as in Assumption LA6. Let $\pi_{1,j}^{**} = \infty$ if $\pi_{1,j} = \infty$ and let $\pi_{1,j}^{**} = 0$ if $\pi_{1,j} < \infty$ for $j = 1, ..., p$. As defined, $\pi_1^{**} = (\pi_{1,1}^{**}, ..., \pi_{1,p}^{**})' \leq h_1$. Let $\pi^{**} = (\pi_1^{**}, \pi_2^{**}) \in R_{+,\infty}^p \times R_{[\pm\infty]}^q$, where $\pi_2^{**} = (\pi_{2,1}^{**}, \pi_{2,2}^{**})$, $\pi_{2,1}^{**} = \theta_0$, where $\theta_0$ is as in Assumption LA1(a), and $\pi_{2,2}^{**} = vech_*(\Omega_0)$ for the $k \times k$ correlation matrix $\Omega_0 = \Omega(\theta_0, F_0)$ determined by Assumption LA1(a). Let $c_{\pi_1^{**}}(1-\alpha)$ denote the $1-\alpha$ quantile of $S(\Omega_0^{1/2} Z^* + (\pi_1^{**}, 0_v), \Omega_0)$, where $Z^* \sim N(0_k, I_k)$ and by definition if $\pi_{1,j}^{**} = \infty$ then the $j$th element of $\Omega_0^{1/2} Z^* + (\pi_1^{**}, 0_v)$ equals $\infty$ for $j = 1, ..., p$.

**Lemma 7** *Under Assumptions* 1-3, 5(a), LA1–LA4, LA6, GMS2, GMS3, GMS5, *and* GMS6,
(a) *if* $c_{\pi_1^{**}}(1-\alpha) > 0$, $\widehat{c}_n(\theta_0, 1-\alpha) \leq c_n^{**}$ *a.s. for all* $n$ *for a sequence of random variables* $\{c_n^{**} : n \geq 1\}$ *that satisfies* $c_n^{**} \to_p c_{\pi_1^{**}}(1-\alpha)$,
(b) $\liminf_{n\to\infty} P_{F_n}(T_n(\theta_0) > \widehat{c}_n(\theta_0, 1-\alpha)) \geq J_{h_1,\lambda}(c_{\pi_1^{**}}(1-\alpha))$,
(c) $c_{g_1,0_d}(1-\alpha) \geq c_{\pi_1^{**}}(1-\alpha)$ *with strict inequality whenever* $g_{1,j} < \infty$ *and* $\pi_{1,j} = \infty$ *for some* $j = 1, ..., p$ *and* $c_{g_1,0_d}(1-\alpha) > 0$, *and*
(d) $J_{h_1,\lambda}(c_{\pi_1^{**}}(1-\alpha)) \geq J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha))$ *with strict inequality whenever* $g_{1,j} < \infty$ *and* $\pi_{1,j} = \infty$ *for some* $j = 1, ..., p$ *and* $c_{g_1,0_d}(1-\alpha) > 0$.

**Lemma 8** *Under Assumptions* 1-5, LA1–LA3, *and* LA6,

(a) $c_{0_p,0_d}(1-\alpha) \geq c_{g_1,0_d}(1-\alpha)$,

(b) $J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha)) \geq J_{h_1,\lambda}(c_{0_p,0_d}(1-\alpha))$,

(c) $c_{0_p,0_d}(1-\alpha) > c_{g_1,0_d}(1-\alpha)$ *unless* $g_1 = 0_p$, *and*

(d) $J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha)) > J_{h_1,\lambda}(c_{0_p,0_d}(1-\alpha))$ *unless* $g_1 = 0_p$.

**Proof of Lemma 4.** To prove part (a), by element-by-element mean-value expansions about $\theta = \theta_n$ and Assumptions LA1, LA2, and LA4, we obtain

$$\begin{aligned} D^{-1/2}(\theta_0, F_n)E_{F_n}m(W_i, \theta_0) &= D^{-1/2}(\theta_n, F_n)E_{F_n}m(W_i, \theta_n) \\ &\quad + \Pi(\theta_n^*, F_n)(\theta_0 - \theta_n), \\ n^{1/2}D^{-1/2}(\theta_0, F_n)E_{F_n}m(W_i, \theta_0) &\rightarrow (h_1, 0_v) + \Pi_0\lambda, \qquad (12.25) \end{aligned}$$

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

Next, under $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ as in Assumption LA1, we have

$$\begin{aligned} &\text{(i) } A_n^0 = (A_{n,1}^0, ..., A_{n,k}^0)' \rightarrow_d Z \sim N(0_k, \Omega_0) \text{ as } n \rightarrow \infty, \text{ where} \\ &A_{n,j}^0 = n^{1/2}(\overline{m}_{n,j}(\theta_0) - E_{F_n}\overline{m}_{n,j}(\theta_0))/\sigma_{F_n,j}(\theta_0), \\ &\text{(ii) } \widehat{\sigma}_{n,j}(\theta_0)/\sigma_{F_n,j}(\theta_0) \rightarrow_p 1 \text{ as } n \rightarrow \infty \text{ for } j = 1, ..., k, \text{ and} \\ &\text{(iii) } \widehat{D}_n^{-1/2}(\theta_0)\widehat{\Sigma}_n(\theta_0)\widehat{D}_n^{-1/2}(\theta_0) \rightarrow_p \Omega_0 \text{ as } n \rightarrow \infty, \qquad (12.26) \end{aligned}$$

where result (i) holds by the Cramér-Wold device and the Liapounov triangular array CLT for row-wise i.i.d. random variables with mean zero and variance one using condition (iv) of (2.2) and Assumptions LA1(a) and LA1(c), and results (ii) and (iii) hold by standard arguments using a weak law of large numbers for row-wise i.i.d. random variables with variance one by condition (iv) of (2.2) and Assumptions LA1(a) and LA1(c). Note that results (i)-(iii) of (12.26) do not hold by (12.3) because the functions are evaluated at $\theta_0$ but the true value is $\theta_n$.

For the same reason as described above following (12.21), to obtain the asymptotic distribution of $T_n(\theta_0)$ we use the same type of argument as in the verification of Assumption B0 in AG4. Let $G(\cdot)$ be a strictly increasing continuous df on $R$, such as the standard normal df. Using (12.25) and (12.26), for $j = 1, ..., k$, we have

$$\begin{aligned} G_{n,j}^0 &= G\left(\widehat{\sigma}_{n,j}^{-1}(\theta_0)n^{1/2}\overline{m}_{n,j}(\theta_0)\right) \\ &= G\left(\widehat{\sigma}_{n,j}^{-1}(\theta_0)\sigma_{F_n,j}(\theta_0)\left[A_{n,j}^0 + n^{1/2}\sigma_{F_n,j}^{-1}(\theta_0)E_{F_n}m_j(W_i, \theta_0)\right]\right), \\ G_{n,j}^0 &\rightarrow_p 1 \text{ if } j \leq p \text{ and } h_{1,j} = \infty, \qquad (12.27) \\ G_{n,j}^0 &\rightarrow_d G\left(Z_j + h_{1,j} + \Pi_{0,j}'\lambda\right) \text{ if } j \leq p \text{ and } h_{1,j} < \infty, \\ G_{n,j}^0 &\rightarrow_d G\left(Z_j + \Pi_{0,j}'\lambda\right) \text{ if } j = p+1, ..., k, \\ G_n^0 &= (G_{n,1}^0, ..., G_{n,k}^0) \rightarrow_d G_\infty^0 = (G(Z_1 + h_{1,1} + \Pi_{0,1}'\lambda), ..., G(Z_k + \Pi_{0,k}'))', \end{aligned}$$

where $Z = (Z_1, ..., Z_k)'$ and $Z_j + h_{1,j} + \Pi'_{0,j}\lambda = \infty$ by definition if $h_{1,j} = \infty$. Now, the same argument as in the verification of Assumption B0 in AG4 gives

$$T_n(\theta_0) \to_d S(Z + (h_1, 0_v) + \Pi_0\lambda, \Omega_0) \sim J_{h_1,\lambda}. \tag{12.28}$$

In short, the idea behind the argument is to write the right-hand side of (12.21) as a continuous function of $G_n^0$ and $\widehat{D}_n^{-1/2}(\theta_{n,h})\widehat{\Sigma}_n(\theta_{n,h})\widehat{D}_n^{-1/2}(\theta_{n,h})$ and apply the continuous mapping theorem. This completes the proof of part (a).

To prove part (b), by the mean-value expansions in (12.25), Assumptions LA1(a), LA2, and LA4, and $\kappa_n \to \infty$, we obtain

$$\kappa_n^{-1} n^{1/2} D^{-1/2}(\theta_0, F_n) E_{F_n} m(W_i, \theta_0) \to (\pi_1, 0_v). \tag{12.29}$$

This leads to

$$\begin{aligned} & \kappa_n^{-1} n^{1/2} \widehat{D}_n^{-1/2}(\theta_0)\overline{m}_n(\theta_0) \\ = \; & \kappa_n^{-1} \widehat{D}_n^{-1/2}(\theta_0) D^{1/2}(\theta_0, F_n) \left(A_n^0 + n^{1/2} D^{-1/2}(\theta_0, F_n) E_{F_n} m(W_i, \theta_0)\right) \\ \to_p \; & (\pi_1, 0_v), \end{aligned} \tag{12.30}$$

where the equality holds by the definition of $A_n^0$ in (12.26) and the convergence holds by (12.29), conditions (i) and (ii) of (12.26), and $\kappa_n \to \infty$. Equation (12.30) and condition (iii) of (12.26) yield

$$(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)) \to_p ((\pi_1, 0_v), \Omega_0). \tag{12.31}$$

For $j = 1, ..., k$, as $(\xi, \Omega) \to ((\pi_1, 0_v), \Omega_0)$,

$$\varphi_j(\xi, \Omega) \to \varphi_j((\pi_1, 0_v), \Omega_0) \text{ a.s.}[Z^*], \tag{12.32}$$

because by Assumption LA5(a) $\pi_1 \in C(\varphi)$, which yields (12.32) by Assumption GMS3 if $\pi_{1,j} = \infty$ and yields (12.32) by the definition of $C(\varphi)$ otherwise.

Assumption 1(d) and (12.32) give: for $x$ in a neighborhood of $c_{\pi_1}(\varphi, 1-\alpha)$, as $(\xi, \Omega) \to ((\pi_1, 0_v), \Omega_0)$,

$$\begin{aligned} S(\Omega^{1/2}Z^* + \varphi(\xi, \Omega), \Omega) &\to S(\Omega_0^{1/2}Z^* + \varphi((\pi_1, 0_v), \Omega_0), \Omega_0) \text{ a.s.}[Z^*], \\ 1\left(S(\Omega^{1/2}Z^* + \varphi(\xi, \Omega), \Omega) \le x\right) &\to 1\left(S(\Omega_0^{1/2}Z^* + \varphi((\pi_1, 0_v), \Omega_0), \Omega_0) \le x\right) \\ & \quad \text{a.s.}[Z^*], \text{ and} \\ P\left(S(\Omega^{1/2}Z^* + \varphi(\xi, \Omega), \Omega) \le x\right) &\to P\left(S(\Omega_0^{1/2}Z^* + \varphi((\pi_1, 0_v), \Omega_0), \Omega_0) \le x\right). \end{aligned} \tag{12.33}$$

The third convergence result of (12.33) holds by the second result and the bounded convergence theorem, and the second convergence result of (12.33) follows from the first result provided $P(S(\Omega_0^{1/2}Z^* + \varphi((\pi_1, 0_v), \Omega_0), \Omega_0) = x) = 0$, which holds by Assumption LA5(b).

Given (12.31) and (12.33), the remainder of the proof of part (b) is the same as that given in the paragraph containing (12.18) using Lemma 5 of AG1.

Part (c) of the Lemma holds by parts (a) and (b) and Assumption LA5(b). $\square$

**Proof of Lemma 5.** Part (a) holds by Lemma 4(a).

To prove part (b), by the mean-value expansions in (12.25), Assumptions LA1(a), LA2, and LA6, and $b/n \to 0$, we obtain

$$b^{1/2} D^{-1/2}(\theta_0, F_n) E_{F_n} m(W_i, \theta_0) \to (g_1, 0_v). \tag{12.34}$$

Using (12.34) and an analogous argument to that given in the proof of Lemma 4(a) with $n^{1/2}$ replaced by $b^{1/2}$ in (12.27), we have

$$T_b(\theta_0) \to_d S(Z + (g_1, 0_v), \Omega_0) \sim J_{g_1, 0_d}, \tag{12.35}$$

which proves part (b).

To establish Lemma 5(c) and (d), we apply Lemma 5 of AG1. We verify conditions (i)-(iii) of Lemma 5 of AG1 as follows. Lemma 5(a) of the present paper implies condition (ii). To verify condition (i), Lemma 5(b) of the present paper and identical distributions for $\{W_i : i \leq n\}$ imply that

$$E_{F_n} U_{n,b}(\theta_0, x) = P_{F_n}(T_b(\theta_0) \leq x) \to J_{g_1, 0_d}(x) \tag{12.36}$$

for all continuity points $x$ of $J_{g_1, 0_d}$. In addition, $Var_{F_n}(U_{n,b}(\theta_0, x)) \to 0$ by a U-statistic inequality of Hoeffding, as in Politis, Romano, and Wolf (1999, p. 44), using the i.i.d. property of $\{W_i : i \leq n\}$ and the boundedness of $U_{n,b}(\theta_0, x)$. This and (12.36) give

$$U_{n,b}(\theta_0, x) \to_p J_{g_1, 0_d}(x) \tag{12.37}$$

for all continuity points $x$ of $J_{g_1, 0_d}$, which verifies condition (i) of Lemma 5 of AG1.

To verify condition (iii) of Lemma 5 of AG1, we need to show

$$J_{g_1, 0_d}(c_{g_1, 0_d}(1 - \alpha) + \varepsilon) > 1 - \alpha \text{ for all } \varepsilon > 0. \tag{12.38}$$

When $v = 0$ and $g_1 = \infty^p$, $S(Z + (g_1, 0_v), \Omega_0) = S(\infty^p, \Omega_0) = 0$ using Assumption 3. In consequence, $J_{g_1, 0_d}(x) = 1$ for all $x \geq 0$, $c_{g_1, 0_d}(1 - \alpha) = 0$, and (12.38) holds for $\alpha > 0$. Now, suppose $v \geq 1$ or $g_1 \neq \infty^p$. Then, by Assumption 2(b), $J_{g_1, 0_d}(x)$ is strictly increasing for $x > 0$. Using this, we have (i) if $c_{g_1, 0_d}(1 - \alpha) > 0$, then $J_{g_1,0_d}(x)$ is strictly increasing at $x = c_{g_1,0_d}(1-\alpha)$ and (12.38) holds, (ii) if $c_{g_1,0_d}(1-\alpha) = 0$, then $J_{g_1,0_d}(0) \geq 1-\alpha$ (by the definition of $c_{g_1,0_d}(1-\alpha)$), (iii) if $c_{g_1,0_d}(1-\alpha) = 0$ and $J_{g_1,0_d}(0) \geq 1-\alpha$, then $J_{g_1,0_d}(x) > 1-\alpha$ for all $x > 0$ and (12.38) holds (otherwise, $J_{g_1,0_d}(x) = 1-\alpha$ for some $x > 0$ and $J_{g_1,0_d}(x/2) = 1-\alpha$ since $J_{g_1,0_d}$ is non-decreasing, which contradicts the fact that $J_{g_1,0_d}(x)$ is strictly increasing for $x > 0$). Hence, (12.38) holds.

Lemma 5 of AG1 establishes Lemma 5(c) of the present paper and shows that $\lim_{n\to\infty} P_{F_n}(T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha)) \in [J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha)-), J_{h_1,\lambda}(c_{g_1,0_d}(1-\alpha))]$. If $c_{g_1,0_d}(1-\alpha) > 0$, then by Assumption 5(a)(i) $J_{h_1,\lambda}$ is continuous at $c_{g_1,0_d}(1-\alpha)$ and the result of Lemma 5(d) holds. Assumption 1(c) implies that $c_{0_p,0_d}(1-\alpha) \geq 0$. The conditions $c_{g_1,0_d}(1-\alpha) = 0$ and $0 < \alpha < 1/2$ are consistent with Assumption 2(c) only if $v = 0$. Given $v = 0$ and $c_{g_1,0_d}(1-\alpha) = 0$, we use the argument given in (12.24) to establish Lemma 5(d) with $\theta_{n,h}$, $P_{\gamma_{n,h}}$, $S(\Omega_{h_{2,2}}^{1/2} Z^* + h_1, \Omega_{h_{2,2}})$, $J_h$, and $c_h(1-\alpha)$ replaced by $\theta_0$, $P_{F_n}$, $S(Z + h_1 + \Pi_0\lambda, \Omega_0)$, $J_{h_1,\lambda}$, and $c_{h_1,\lambda}(1-\alpha)$, respectively. $\square$

**Proof of Lemma 6.** Part (a) holds by Lemma 4(a) because Assumptions 2 and 3 are not used in the proof of Lemma 4(a).

By standard arguments using a weak law of large numbers for row-wise i.i.d. triangular arrays and Assumption LA1(c), we have

$$\begin{aligned}
&D^{-1/2}(\theta_0, F_n)\widehat{\Sigma}_n(\theta_0)D^{-1/2}(\theta_0, F_n) \\
&-D^{-1/2}(\theta_0, F_n)Var_{F_n}(m(W_i, \theta_0))D^{-1/2}(\theta_0, F_n) \to_p 0_{k\times k} \text{ and} \\
&\widehat{D}_n^{-1/2}(\theta_0)D^{1/2}(\theta_0, F_n) - I_k \to_p 0_{k\times k}. \qquad (12.39)
\end{aligned}$$

In consequence, $\widehat{\Omega}_n(\theta_0) - \Omega(\theta_0, F_n) \to_p 0_{k\times k}$. This and Assumptions LA1(a) and LA1(c) give $\widehat{\Omega}_n(\theta_0) \to_p \Omega_0$. The latter and Assumption 4(b) yield $c(\widehat{\Omega}_n(\theta_0), 1-\alpha) \to_p c(\Omega_0, 1-\alpha)$. This establishes Lemma 6(b) because $c(\Omega_0, 1-\alpha) = c_{0_p,0_d}(1-\alpha)$ by definition.

If $c_{0_p,0_d}(1-\alpha) > 0$, Lemma 6(c) holds by parts (a) and (b) and Assumption 5(a)(i). Assumption 1(c) implies that $c_{0_p,0_d}(1-\alpha) \geq 0$. The conditions $c_{0_p,0_d}(1-\alpha) = 0$ and $0 < \alpha < 1/2$ are consistent with Assumption 2(c) only if $v = 0$ (because $c_{0_p,0_d}(1-\alpha)$ is the $1-\alpha$ quantile of $S(Z, \Omega_0)$). Given $v = 0$ and $c_{0_p,0_d}(1-\alpha) = 0$, Lemma 6(c) holds by the same argument as used to prove Lemma 5(d) when $c_{g_1,0_d}(1-\alpha) = 0$. $\square$

**Proof of Lemma 7.** First we prove part (a). By assumption, $c_{\pi_1^{**}}(1-\alpha) > 0$. For $(\xi, \Omega) \in R^k \times \Psi$, let $\varphi^{**}(\xi, \Omega)$ denote the $k$-vector whose $j$th element is

$$\varphi_j^{**}(\xi, \Omega) = \begin{cases} 0 & \text{if } \pi_{1,j} < \infty \text{ and } j = 1, ..., p \\ \varphi_j(\xi, \Omega) & \text{if } \pi_{1,j} = \infty \text{ and } j = 1, ..., p \\ 0 & \text{if } j = p+1, ..., k. \end{cases} \qquad (12.40)$$

By Assumption GMS6, $\varphi_j(\xi,\Omega) \geq 0$ for $j \leq p$. Hence, $\varphi^{**}(\xi,\Omega) \leq \varphi(\xi,\Omega)$ and

$$\varphi^{**}(\xi_n(\theta_n), \widehat{\Omega}_n(\theta_n)) \leq \varphi(\xi_n(\theta_n), \widehat{\Omega}_n(\theta_n)) \text{ a.s.}[Z^*] \text{ for all } n. \quad (12.41)$$

Let $c_n^{**}$ denote the $1-\alpha$ quantile of the conditional df of $S(\widehat{\Omega}_n^{1/2}(\theta_n)Z^* + \varphi^{**}(\xi_n(\theta_n), \widehat{\Omega}_n(\theta_n)), \widehat{\Omega}_n(\theta_n))$ given $(\xi_n(\theta_n), \widehat{\Omega}_n(\theta_n))$. Then, $\widehat{c}_n(\theta_n, 1-\alpha) \leq c_n^{**}$ a.s. for all $n$ by (12.41) and Assumption 1(a). By the same argument as in the proof of Lemma 4, (12.31) holds.

For $j = 1, ..., p$ with $\pi_{1,j} < \infty$, $\varphi_j^{**}(\xi,\Omega) = 0$. For $j = 1, ..., p$ with $\pi_{1,j} = \infty$, as $(\xi,\Omega) \to ((\pi_1, 0_v), \Omega_0)$, we have $\varphi_j(\xi,\Omega) \to \infty$ a.s.$[Z^*]$ by Assumption GMS3. These results can be written as

$$\varphi_j^{**}(\xi,\Omega) \to \pi_{1,j}^{**} \text{ a.s.}[Z^*] \quad (12.42)$$

for $j = 1, ..., p$ by the definition of $\pi_{1,j}^{**}$.

Assumption 1(d) and (12.42) give: for $x$ in a neighborhood of $c_{\pi_1^{**}}(1-\alpha)$, as $(\xi,\Omega) \to ((\pi_1, 0_v), \Omega_0)$,

$$\begin{aligned} S(\Omega^{1/2}Z^* + \varphi^{**}(\xi,\Omega),\Omega) &\to S(\Omega^{1/2}Z^* + (\pi_1^{**}, 0_v), \Omega_0) \text{ a.s.}[Z^*], \\ 1\left(S(\Omega^{1/2}Z^* + \varphi^{**}(\xi,\Omega),\Omega) \leq x\right) &\to 1\left(S(\Omega^{1/2}Z^* + (\pi_1^{**}, 0_v), \Omega_0) \leq x\right) \text{ a.s.}[Z^*], \\ P\left(S(\Omega^{1/2}Z^* + \varphi^{**}(\xi,\Omega),\Omega) \leq x\right) &\to P\left(S(\Omega^{1/2}Z^* + (\pi_1^{**}, 0_v), \Omega_0) \leq x\right). \quad (12.43) \end{aligned}$$

The third convergence result of (12.43) holds by the second result and the bounded convergence theorem. The second convergence result of (12.43) follows from the first result provided $P(S(\Omega^{1/2}Z^* + (\pi_1^{**}, 0_v), \Omega_0) = x) = 0$, which holds because $c_{\pi_1^{**}}(1-\alpha) > 0$ for the same reason as the second equality in (12.17) holds.

Given (12.31) and (12.43), the remainder of the proof of part (a) is the same as that given in the paragraph containing (12.18) using Lemma 5 of AG1.

Now we prove part (b). If $c_{\pi_1^{**}}(1-\alpha) > 0$, part (b) of the Lemma holds by part (a), Lemma 4(a) (i.e., $T_n(\theta_0) \to_d J_{h_1,\lambda}$), and Assumption 5(a)(i).

Next, we prove part (b) for the case where $c_{\pi_1^{**}}(1-\alpha) = 0$. We have

$$\liminf_{n\to\infty} P_{\gamma_n}(T_n(\theta_n) \leq c_n^{**}) \geq \liminf_{n\to\infty} P_{\gamma_n}(T_n(\theta_n) \leq 0) \quad (12.44)$$

because $c_n^{**} \geq 0$ by Assumption 1(c). By the definition of $\pi_1^{**}$, we have $\pi_1^{**} \leq h_1$. As in (12.19), this implies that $c_h(1-\alpha) \leq c_{\pi^{**}}(1-\alpha)$ and hence $c_h(1-\alpha) = 0$ using Assumption 1(c). The conditions $c_h(1-\alpha) = 0$ and $0 < \alpha < 1/2$ are consistent with Assumption 2(c) only if $v = 0$. Given $v = 0$, we use the same argument as given in (12.24) with $\gamma_{n,h}$, $A_{n,j}$, $h_{1,j}$, $J_h$, and $c_h(1-\alpha)$ replaced by $\gamma_n$, $A_{n,j}^0$, $h_{1,j} + \Pi_{0,j}'\lambda$, $J_{h_1,\lambda}$, and $c_{h_1,\lambda}(1-\alpha)$, respectively, where $A_n^0$ is defined in

(12.26), to show that the right-hand side in (12.44) is greater than or equal to $1-\alpha$. This completes the proof of part (b).

When the inequality is not strict, part (c) holds because (i) $\pi_1^{**} \geq g_1$, which holds because if $\pi_{1,j} = \infty$ then $\pi_{1,j}^{**} = \infty$ and if $\pi_{1,j} < \infty$ then $g_{1,j} = 0$ by Assumption LA4 and GMS5 and $\pi_{1,j}^{**} = 0$ by definition, (ii) $S(\Omega_0^{1/2} Z^* + (\pi_1^{**}, 0_v), \Omega_0) \leq S(Z + (g_1, 0_v), \Omega_0)$ a.s. by (i) and Assumption 1(a), and (iii) the corresponding quantiles satisfy $c_{\pi_1^{**}}(1-\alpha) \leq c_{g_1,0_d}(1-\alpha)$ by (ii).

Next, we show part (c) holds with a strict inequality when $c_{g_1,0_d}(1-\alpha) > 0$ and $g_{1,j} < \infty$ and $\pi_{1,j} = \infty$ for some $j = 1, ..., p$. The latter implies that $\pi_1^{**} > g_1$. Given $\pi_1^{**} > g_1$ and $c_{g_1,0_d}(1-\alpha) > 0$, Assumption 5(b) implies that

$$\begin{aligned} &P(S(Z + (\pi_1^{**}, 0_v), \Omega) \leq c_{g_1,0_d}(1-\alpha)) \\ &> P(S(Z + (g_1, 0_v), \Omega) \leq c_{g_1,0_d}(1-\alpha)) \geq 1-\alpha, \end{aligned} \tag{12.45}$$

where $Z \sim N(0_k, \Omega)$. If "$v = 0$ and $\pi_1^{**} = \infty^p$" does not hold, then the df of $S(Z + (\pi_1^{**}, 0_v), \Omega)$ is strictly increasing for $x > 0$ by Assumption 2(b). This and (12.45) imply that $c_{\pi_1^{**}}(1-\alpha) < c_{g_1,0_d}(1-\alpha)$. If $v = 0$ and $\pi_1^{**} = \infty^p$, then $S(Z + (\pi_1^{**}, 0_v), \Omega)) = S(Z + \infty^p, \Omega)) = 0$ by Assumption 1(c) and $c_{\pi_1^{**}}(1-\alpha) = 0 < c_{g_1,0_d}(1-\alpha)$ and the proof of part (c) is complete.

Part (d) follows immediately from part (c) when the inequality is not strict. When $c_{g_1,0_d}(1-\alpha) > 0$ and $g_{1,j} < \infty$ and $\pi_{1,j} = \infty$ for some $j = 1, ..., p$, part (c) holds with a strict inequality. The latter, $c_{\pi_1^{**}}(1-\alpha) \geq 0$ (which holds by Assumption 1(c)), and $J_{h_1,\lambda}(x)$ is strictly increasing for $x > 0$ (which holds by Assumption 5(a)(ii) because the caveat in Assumption 5(a)(ii) that "$v = 0$ and $\ell = \infty^p$ does not occur" holds by Assumption LA3) imply that part (d) holds with a strict inequality. □

**Proof of Lemma 8.** Part (a) holds because for $0_p \leq g_1 \in R_{+,\infty}^p$, we have

$$S(Z + (0_p, 0_v), \Omega_0) \geq S(Z + (g_1, 0_v), \Omega_0) \tag{12.46}$$

by Assumption 1(a). Part (b) follows from part (a). To prove Lemma 8(c), note that $c_{0_p,0_d}(1-\alpha) > 0$ by Assumption 2(c) and $\alpha \in (0, 1/2)$. This, Assumptions 2(a) and 5(b), and $g_1 > 0_p$ imply that

$$\begin{aligned} 1-\alpha &= P(S(Z + (0_p, 0_v), \Omega_0) \leq c_{0_p,0_d}(1-\alpha)) \\ &< P(S(Z + (g_1, 0_v), \Omega_0) \leq c_{0_p,0_d}(1-\alpha)), \end{aligned} \tag{12.47}$$

where $Z \sim N(0_k, \Omega_0)$. The latter and Assumption 2(a) prove part (c).

Lemma 8(d) holds by part (c), $c_{g_1,0_d}(1-\alpha) \geq 0$ (which holds by Assumption 1(c)), and Assumption 5(a)(ii) (because the caveat in Assumption 5(a)(ii) that "$v = 0$ and $\ell = \infty^p$ does not occur" holds by Assumption LA3). □

# 13 APPENDIX B

Appendix B contains the proof of Theorem 4 and the verification of Assumptions GMS1, GMS3, GMS6, and GMS7 for $\varphi^{(5)}$.

**Proof of Theorem 4.** It suffices to show that for any subsequence $\{t_n\}$ of $\{n\}$ there exists a sub-subsequence $\{s_n\}$ such that $\lim_{n\to\infty} P_{F_{s_n}}(T_{s_n}(\theta_0) > c_{1-\alpha}) = 1$, where $c_{1-\alpha} = \widehat{c}_n(\theta_0, 1-\alpha)$, $c_{n,b}(\theta_0, 1-\alpha)$, or $c(\widehat{\Omega}_n(\theta_0), 1-\alpha)$. We can take the subsequence $\{s_n\}$ to be such that $m^*_{s_n,j}/\beta_{s_n} \to e_j$ for some $e_j \in [-1,\infty]$ for $j = 1, ..., k$ because $\{m^*_{n,j}/\beta_n : n \geq 1\}$ is a sequence of points in the set $[-1,\infty]$ by the definition of $\beta_n$. For notational simplicity, we establish the former result with $s_n$ replaced by $n$ and by a subsequence argument assume without loss of generality (wlog) that

$$m^*_{n,j}/\beta_n \to e_j \text{ for some } e_j \in [-1,\infty] \text{ for } j = 1, ..., k. \tag{13.1}$$

The following is used in the proofs of parts (a)-(c). We have

$$\begin{aligned}
& (n^{1/2}\beta_n)^{-\tau} T_n(\theta_0) \\
= \; & (n^{1/2}\beta_n)^{-\tau} S\Big(\widehat{D}_n^{-1/2}(\theta_0) n^{1/2}\overline{m}_n(\theta_0),\ \widehat{D}_n^{-1/2}(\theta_0)\widehat{\Sigma}_n(\theta_0)\widehat{D}_n^{-1/2}(\theta_0)\Big) \\
= \; & (n^{1/2}\beta_n)^{-\tau} S\Big(\widehat{D}_n^{-1/2}(\theta_0) D^{1/2}(\theta_0, F_n)(A_n^0 + n^{1/2} m_n^*),\ \Omega_1 + o_p(1)\Big) \\
= \; & S(o_p(1) + m_n^*/\beta_n,\ \Omega_1 + o_p(1)) \\
\to_p \; & S\,(e,\ \Omega_1) > 0,
\end{aligned} \tag{13.2}$$

where $A_n^0$ is defined in (12.26), $m_n^* = (m^*_{n,1}, ..., m^*_{n,k})'$, $e = (e_1, ..., e_k)'$, the first equality uses Assumption 1(b), the second equality holds by the definitions of $A_n^0$, $m_n^*$, and $D(\theta_0, F_n)$ and by (12.26) (with $\Omega_1$ in place of $\Omega_0$ and with Assumption DA(b) used in place of Assumption LA1(a) in the proof of (12.26)), the third equality holds by Assumptions 6 and DA(a) and (12.26) (with the same adjustments as above), the convergence holds by Assumption 1(d) and (13.1), and the inequality holds by Assumption 3 because for some $j^* \leq k$ the $j^*$th element of $e$, $e_{j^*}$, has absolute value equal to one and is negative if $j^* \leq p$, which implies that $e_{j^*} < 0$ if $j^* \leq p$ and $e_{j^*} \neq 0$ if $j^* \geq p+1$.

We prove part (b) first. By another subsequence argument, we can assume wlog that $\lim_{n\to\infty} b^{1/2}\beta_n$ exists and (13.1) holds. We consider two cases: (i) $\lim_{n\to\infty} b^{1/2}\beta_n = \infty$ and (ii) $\lim_{n\to\infty} b^{1/2}\beta_n \in [0,\infty)$. When case (i) holds, the same argument as used to show (13.2) gives

$$(b^{1/2}\beta_n)^{-\tau} T_b(\theta_0) \to_p S\,(e,\ \Omega_1)\,, \tag{13.3}$$

where $\beta_n$ appears, not $\beta_b$, because $m_n^*/\beta_n \to e$ under $\{F_n : n \geq 1\}$. Equation (13.3) and $b/n \to 0$ imply that $T_b^\dagger(\theta_0) = (n^{1/2}\beta_n)^{-\tau} T_b(\theta_0) \to_p 0$.

Define $U_{n,b}^\dagger(\theta_0, x)$ as $U_{n,b}(\theta_0, x)$ is defined but with $T_{n,b,j}^\dagger(\theta_0) = (n^{1/2}\beta_n)^{-\tau} \times T_{n,b,j}(\theta_0)$ in place of $T_{n,b,j}(\theta_0)$. Using the result of the previous paragraph, we have $E_{F_n} U_{n,b}^\dagger(\theta_0, x) = P_{F_n}(T_b^\dagger(\theta_0) \leq x) \to 0$ for $x < 0$ and $\to 1$ for $x > 0$. In addition, $Var_{F_n}(U_{n,b}^\dagger(\theta_0, x)) \to 0$ by Hoeffding's U-statistic inequality for bounded i.i.d. random variables, see Politis, Romano, and Wolf (1999, p. 44). Hence, $U_{n,b}^\dagger(\theta_0, x) \to_p 0$ for $x < 0$ and $\to_p 1$ for $x > 0$. This and Lemma 5(a) of AG1 imply that $c_{n,b}^\dagger(\theta_0, 1-\alpha) \to_p 0$, where $c_{n,b}^\dagger(\theta_0, 1-\alpha)$ is the $1-\alpha$ quantile of the rescaled subsample statistics $\{T_{n,b,j}^\dagger(\theta_0) : j = 1, ..., q_n\}$. The latter result and (13.2) give

$$
\begin{aligned}
& P_{F_n}(T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha)) \\
= \; & P_{F_n}((n^{1/2}\beta_n)^{-\tau} T_n(\theta_0) > (n^{1/2}\beta_n)^{-\tau} c_{n,b}(\theta_0, 1-\alpha)) \\
= \; & P_{F_n}((n^{1/2}\beta_n)^{-\tau} T_n(\theta_0) > c_{n,b}^\dagger(\theta_0, 1-\alpha)), \\
\to \; & P(S(e, \Omega_1) > 0) = 1, \qquad (13.4)
\end{aligned}
$$

where the second equality holds because $c_{n,b}^\dagger(\theta_0, 1-\alpha) = (n^{1/2}\beta_n)^{-\tau} c_{n,b}(\theta_0, 1-\alpha)$ by the scale equivariance of quantiles and the last equality holds because $S(e, \Omega_1) > 0$ by (13.2).

Next, suppose case (ii) holds. Then, the same argument as used to show (13.2) but with $(n^{1/2}\beta_n)^{-\tau}$ deleted gives

$$
T_b(\theta_0) = S\big(O_p(1) + (b^{1/2}\beta_n)\beta_n^{-1} m_n^*,\ \Omega_1 + o_p(1)\big) = O_p(1), \qquad (13.5)
$$

where the second equality uses Assumption 1(a). Hence, $T_b^\dagger(\theta_0) = (n^{1/2}\beta_n)^{-\tau} \times T_b(\theta_0) \to_p 0$. Given this, the remainder of the proof is the same as in case (i).

Next, we prove part (c). We have $\widehat{\Omega}_n(\theta_0) \to_p \Omega_1$ because (12.39) holds by the argument given for (12.39) but using condition (vii) of (2.2) and Assumption DA(b). This and Assumption 4(b) imply that $c(\widehat{\Omega}_n(\theta_0), 1-\alpha) \to_p c(\Omega_1, 1-\alpha)$. Combining the latter with (13.2) and (13.4) with $c_{n,b}(\theta_0, 1-\alpha))$ replaced by $c(\widehat{\Omega}_n(\theta_0), 1-\alpha)$ in (13.4) gives the desired result.

Finally, we prove part (a). By (13.2) and the first equality of (13.4) with $c_{n,b}(\theta_0, 1-\alpha))$ replaced by $\widehat{c}_n(\theta_0, 1-\alpha)$, it suffices to show that $(n^{1/2}\beta_n)^{-\tau}\widehat{c}_n(\theta_0, 1-\alpha) = o_p(1)$.

Let $A_n^0$ be defined as in (12.26). We have

$$
\begin{aligned}
& (n^{1/2}\beta_n)^{-1}\kappa_n^{-1} n^{1/2} \widehat{D}_n^{-1/2}(\theta_0)\overline{m}_n(\theta_0) \\
= \; & (n^{1/2}\beta_n)^{-1}\kappa_n^{-1} \left(\widehat{D}_n^{-1/2}(\theta_0) D^{1/2}(\theta_0, F_n)\right) \left(A_n^0 + n^{1/2} m_n^*\right) \\
= \; & o_p(1) + \kappa_n^{-1}(m_n^*/\beta_n)(1 + o_p(1)), \qquad (13.6)
\end{aligned}
$$

where the second equality uses (12.26) with $\Omega_0$ replaced by $\Omega_1$ (using Assumption DA(b)), $\kappa_n \to \infty$, and $n^{1/2}\beta_n \to \infty$. By the definition of $\beta_n$, $m^*_{n,j}/\beta_n \in [-1,\infty)$ for $j = 1, ..., k$ for all $n$. By a subsequence argument, wlog we assume $\kappa_n^{-1} m^*_{n,j}/\beta_n \to \eta_j \in [0,\infty]$ for $j = 1, ..., k$. This and (13.6) give

$$\begin{aligned}
&(n^{1/2}\beta_n)^{-1}\xi_n(\theta_0) \to_p \eta = (\eta_1, ..., \eta_k)' \in R^p_{+,\infty} \times R^v \text{ and} \\
&\Phi_{n,1} \equiv (n^{1/2}\beta_n)^{-1} \left(\min\{\xi_{n,1}(\theta_0), 0\}, ..., \min\{\xi_{n,p}(\theta_0), 0\}, 0, ..., 0\right)' \to_p 0_k,
\end{aligned} \tag{13.7}$$

where $\xi_n(\theta_0) = (\xi_{n,1}(\theta_0), , ..., \xi_{n,k}(\theta_0))'$.

Using Assumption 6, we have

$$\begin{aligned}
&(n^{1/2}\beta_n)^{-\tau} S\left(\widehat{\Omega}_n^{1/2}(\theta_0)Z^* + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_n(\theta_0)\right) \\
&= S\left((n^{1/2}\beta_n)^{-1}\left[\widehat{\Omega}_n^{1/2}(\theta_0)Z^* + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))\right], \widehat{\Omega}_n(\theta_0)\right) \\
&\leq S\left(\Phi_{n,2}Z^* + \Phi_{n,1}, \widehat{\Omega}_n(\theta_0)\right),
\end{aligned} \tag{13.8}$$

where $\Phi_{n,2} \equiv (n^{1/2}\beta_n)^{-1}\widehat{\Omega}_n^{1/2}(\theta_0)$ ($\in R^{k\times k}$) and the inequality holds by Assumptions 1(a) and GMS7. We have $\Phi_{n,2} = o_p(1)$ by (12.26) and Assumption DA(a). Let $\widetilde{c}_n$ denote the $1-\alpha$ quantile of $S(\Phi_{n,2}Z^* + \Phi_{n,1}, \widehat{\Omega}_n(\theta_0))$ in (13.8).

By (13.8), $(n^{1/2}\beta_n)^{-\tau}\widehat{c}_n(1-\alpha) \leq \widetilde{c}_n$. Hence, it suffices to show that $\widetilde{c}_n = o_p(1)$. To do so, we use a similar argument to that in (12.16). For $x > 0$, as $(\xi, \Omega_a, \Omega_b) \to (0_k, 0_{k\times k}, \Omega_1)$, we have

$$\begin{aligned}
&S(\Omega_a^{1/2}Z^* + \xi, \Omega_b) \to S(0_k, \Omega_1) = 0 \text{ a.s.}[Z^*], \\
&1(S(\Omega_a^{1/2}Z^* + \xi, \Omega_b) \leq x) \to 1(0 \leq x) \text{ a.s.}[Z^*], \text{ and} \\
&P(S(\Omega_a^{1/2}Z^* + \xi, \Omega_b) \leq x) \to 1,
\end{aligned} \tag{13.9}$$

where the equality in the first line uses Assumption 3, the second convergence result follows from the first result for $x > 0$, and the third convergence result holds by the second result and the bounded convergence theorem. The third result of (13.9), $(\Phi_{n,1}, \Phi_{n,2}, \widehat{\Omega}_n(\theta_0)) \to_p (0_k, 0_{k\times k}, \Omega_1)$ (which uses (13.7)), and Slutsky's Theorem give

$$P(S(\Phi_{n,2}Z^* + \Phi_{n,1}, \widehat{\Omega}_n(\theta_0)) \leq x) \to_p 1 \text{ for all } x > 0, \tag{13.10}$$

where $P(\cdot)$ denotes the distribution of $Z^*$ conditional on $(\Phi_{n,1}, \Phi_{n,2}, \widehat{\Omega}_n(\theta_0))$. By Assumption 1(c), the probability limit in (13.10) is zero for all $x < 0$. These results and Lemma 5(a) of AG1 imply that $\widetilde{c}_n \to_p 0$, where $\widetilde{c}_n$ is the $1-\alpha$ quantile of the (random) df in (13.10). This completes the proof of part (a). □

We now verify Assumptions GMS1, GMS3, GMS6, and GMS7 for $\varphi^{(5)}$. Assumption GMS1(b) holds for $\varphi^{(5)}$ if $c_j(\xi,\Omega) = 1$ whenever the $j$th element of $\xi$ equals 0 by the definition of $\varphi^{(5)}$. If the $j$th element of $\xi$ equals zero, $c_j(\xi,\Omega)$ does not enter the criterion function $S(-c\cdot\xi,\Omega) - \eta(|c|)$. In consequence, the criterion function is minimized by taking $c_j(\xi,\Omega) = 1$ because $\eta(\cdot)$ is strictly increasing. Hence, Assumption GMS1(b) holds for $\varphi^{(5)}$.

We show Assumption GMS1(a) holds for $\varphi^{(5)}$ (provided $S$ satisfies Assumption 1(d)) by showing that if $(\xi_{[r]},\Omega_{[r]}) \to (\xi,\Omega)$ as $r\to\infty$ and $\xi_j = 0$, then $c_j(\xi_{[r]},\Omega_{[r]}) = 1$ for $r$ sufficiently large. By Assumption 1(d), $S$ is continuous at $(\xi,\Omega)$. Hence, $\lim_{r\to\infty} S(-c\cdot\xi_{[r]},\Omega_{[r]}) \to S(-c\cdot\xi,\Omega)$ as $r\to\infty$. The limit $S(-c\cdot\xi,\Omega)$ does not depend on $c_j$ because $\xi_j = 0$. Given $\varepsilon > 0$, there exists an $r^*$ sufficiently large that $|S(-c\cdot\xi_{[r]},\Omega_{[r]}) - S(-c\cdot\xi,\Omega)| \leq \varepsilon$ for all $c\in\mathcal{C}$ and all $r \geq r^*$. Hence, the first term of the selection criterion, $S(-c\cdot\xi,\Omega)$, is reduced by at most $\varepsilon$ if $c_j$ is changed from 1 to 0, where $c = (c_1,...,c_k)'$. On the other hand, the second term of the selection criterion, $-\eta(|c|)$, is increased by $\eta(|c|+1) - \eta(|c|) > 0$. Taking $\varepsilon < \inf_{c\in\mathcal{C}}(\eta(|c|+1) - \eta(|c|))$ implies that the selection criterion is minimized by $c_j(\xi_{[r]},\Omega_{[r]}) = 1$ for all $r \geq r^*$. Hence, Assumption GMS1(a) holds for $\varphi^{(5)}$.

Next we verify Assumption GMS3 for $\varphi^{(5)}$ for all functions $S$ for which $S(-c\cdot\xi,\Omega) \to \infty$ as $(\xi,\Omega) \to (\xi_*,\Omega_*)$ whenever $c_j = 1$. For any $c_* \in \mathcal{C}$ with $c_{*\ell} = 0$ for all $\ell$ such that $\xi_{*\ell} = \infty$ we have $S(-c_*\cdot\xi,\Omega) \leq K$ as $(\xi,\Omega) \to (\xi_*,\Omega_*)$ for some $K < \infty$ by Assumption 1(d). Hence, some $c_* = (c_{*1},...,c_{*k})' \in \mathcal{C}$ with $c_{*j} = 0$ is selected over any $c = (c_1,...,c_k)' \in \mathcal{C}$ with $c_j = 1$ as $(\xi,\Omega) \to (\xi_*,\Omega_*)$. This gives $c_j(\xi,\Omega) = 0$ and $\varphi_j(\xi,\Omega) = \infty$ (using the definition of $\varphi^{(5)}$) as $(\xi,\Omega) \to (\xi_*,\Omega_*)$.

Assumptions GMS6 and GMS7 hold immediately for $\varphi^{(5)}$ by its definition.

# References


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

——— (1999): "Estimation When a Parameter Is on a Boundary," *Econometrica,* 67, 1341-1383.

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

Andrews, D. W. K., Berry, S., and Jia, P. (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.

——— (2007): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria," manuscript in preparation, Cowles Foundation, Yale University.

Andrews, D. W. K. and P. Guggenberger (2005a): "The Limit of Exact Size and a Problem with Subsampling and with the $m$ Out of $n$ Bootstrap," unpublished manuscript, Cowles Foundation, Yale University.

——— (2005b): "Hybrid and Size-corrected Subsampling Methods," unpublished manuscript, Cowles Foundation, Yale University.

——— (2005c): "Applications of Subsampling, Hybrid, and Size-Correction Methods," Cowles Foundation Discussion Paper No. 1608, Yale University.

——— (2005d): "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," Cowles Foundation Discussion Paper No. 1620, Yale University.

Beresteanu, A. and F. Molinari (2006): "Asymptotic Properties for a Class of Partially Identified Models," Working Paper No. 2006-04, Department of Economics, Duke University.

Bugni, F. (2007a): "Bootstrap Inference in Partially Identified Models," unpublished manuscript, Department of Economics, Northwestern University.

——— (2007b): "Bootstrap Inference in Partially Identified Models: Pointwise Construction," unpublished manuscript, Department of Economics, Northwestern University.

Canay, I. A. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," unpublished manuscript, Department of Economics, University of Wisconsin.

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

Ciliberto, F. and E. Tamer (2003): "Market Structure and Market Equilibrium in Airline Markets," unpublished manuscript, Department of Economics, Princeton University.

Galichon, A. and M. Henry (2006): "Inference in Incomplete Models," unpublished manuscript, Department of Economics, Columbia University.

Guggenberger, P., J. Hahn, and K. Kim (2007): "Specification Testing Under Moment Inequalities," unpublished manuscript, Department of Economics, UCLA.

Hannan, E. J. and B. G. Quinn (1979): "The Determination of the Order of an Autoregression," *Journal of Royal Statistical Society,* Series B, 41, 190-195.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica,* 50, 1029-1054.

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

Kudo, A. (1963): "A Multivariate Analog of a One-Sided Test," *Biometrika,* 59, 403-418.

Manski, C. F. and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica,* 70, 519-546.

Mikusheva, A. (2007): "Uniform Inference in Autoregressive Models," *Econometrica,* 75, 1411-1452.

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.

Otsu, T. (2006): "Large Deviation Optimal Inference for Set Identified Moment Inequality Models, unpublished manuscript, Cowles Foundation, Yale University.

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

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling.* New York: Springer.

Romano, J. P. and A. M. Shaikh (2005a): "Inference for Identifiable Parameters in Partially Identified Econometric Models," unpublished working paper, Department of Economics, University of Chicago.

——— (2005b): "Inference for the Identified Set in Partially Identified Econometric Models," unpublished working paper, Department of Economics, University of Chicago.

Rosen, A. M. (2005): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," unpublished working paper, Department of Economics, Northwestern University.

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

Soares, G. (2005): "Inference with Inequality Moment Constraints," unpublished working paper, Department of Economics, Yale University.

Stoye, J. (2007): "More on Confidence Intervals for Partially Identified Parameters," unpublished manuscript, Department of Economics, New York University.

Woutersen, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters," unpublished manuscript, Department of Economics, Johns Hopkins University.