Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities

By

Donald W. K. Andrews and Patrik Guggenberger

July 2007

COWLES FOUNDATION DISCUSSION PAPER NO. 1620

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

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

Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

Patrik Guggenberger
Department of Economics
UCLA

November 2005
Revised: July 2007

[1] Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-0417911. Guggenberger gratefully acknowledges research support from a faculty research grant from UCLA in 2005. For helpful comments, we thank Victor Chernozukhov, Azeem Shaikh, and the participants at various seminars and conferences at which the paper was presented. Some of the results in this paper first appeared in Andrews, D. W. K. and P. Guggenberger (2005) The Limit of Finite-Sample Size and a Problem with Subsampling, Cowles Foundation Discussion Paper No. 1606, Yale University. They have been removed from that paper and are not being considered for publication elsewhere.

Abstract

This paper considers inference for parameters defined by moment inequalities and equalities. The parameters need not be identified. For a specified class of test statistics, this paper establishes the uniform asymptotic validity of subsampling, m out of n bootstrap, and "plug-in asymptotic" tests and confidence intervals for such parameters. Establishing uniform asymptotic validity is crucial in moment inequality problems because the test statistics of interest have discontinuities in their pointwise asymptotic distributions.

The size results are quite general because they hold without specifying the particular form of the moment conditions—only $2 + \delta$ moments finite are required. The results allow for i.i.d. and dependent observations and for preliminary consistent estimation of identified parameters.

Keywords: Asymptotic size, confidence set, exact size, m out of n bootstrap, subsampling, moment inequalities.

JEL Classification Numbers: C12, C15.

1 Introduction

In this paper, we consider a confidence set (CS) for a true parameter θ_0 ($\in \Theta \subset R^d$) whose value is bounded by moment inequalities and equalities. The true parameter need not be identified. There are now numerous examples in the literature that fit into this framework. One way that moment inequalities arise in economic models is from the necessary conditions for Nash equilibria, e.g., see Ciliberto and Tamer (2003), Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2004), and Bajari, Benkard, and Levin (2008). Moment inequalities also can arise from sufficient conditions for Nash equilibria, e.g., see Ciliberto and Tamer (2003). Another way they arise is from data censoring, e.g., when a continuous variable is only observed to lie in an interval, see Manski and Tamer (2002).

We consider a CS that is obtained by inverting a test that is based on a generalized method of moments-type (GMM) criterion function. The method is of Anderson-Rubin-type and was first considered in the moment inequality context by Chernozhukov, Hong, and Tamer (2008) (CHT). CHT obtain critical values via subsampling. The present paper shows that for a broad class of test statistics subsampling CS's for the true parameter are *uniformly* asymptotically valid. The results hold for any specification of the moment functions (subject to $2+\delta$ moments being finite). The paper also shows that subsampling CS's are not asymptotically conservative. Conditions on the form of the test statistic are given such that validity holds. For example, the results hold for statistics given by the sum of squared negative parts of the normalized sample moment conditions, Gaussian quasi-likelihood ratio statistics (also referred to as modified minimum distance statistics), generalized empirical likelihood (GEL) statistics, and a number of other statistics considered in the literature.[2] As far as we know, the results of this paper are the first results available in the literature that establish uniform asymptotic validity of a method of inference for a *general* class of partially-identified models. (See below for a discussion of other methods in the literature.)

Romano and Shaikh (2005a,b) also provide results concerning the uniform asymptotic validity of subsampling in the moment inequality context. They provide a high-level condition under which subsampling provides uniformly valid inference asymptotically. This condition needs to be verified separately for each specification of the moment functions considered and each type of test statistic considered. Verification is not trivial. They verify the condition for one-sample and two-sample means problems for the sum of squared negative parts statistic. In contrast, as stated above, the results of this paper hold for a broad class of test statistics and for any specification of the moment functions (subject to $2+\delta$ moments being finite).[3]

[2] The lack of a uniformly most powerful test even in a Gaussian location testing problem with a multivariate one-sided null hypothesis (which is a special case of the nonlinear moment inequality model considered here) indicates that it is not possible to unambiguously rank the different statistics that one might use. However, some choices have better all around properties than others.

[3] The general results on subsampling given in Andrews and Guggenberger (2005a) were done independently of, and at about the same time as, Romano and Shaikh (2005a,b). Romano and Shaikh

The results of this paper 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), as in CHT. The reason for this focus is that policy questions based on a structural model in which parameters are restricted by moment inequalities depend on the true parameter, rather than on the identified set. A CS for the identified set typically leads to conservative inference when interest is in the true parameter.

We stress that asymptotic validity of a CS requires the verification of uniformity in the asymptotic results.[4] That is, it requires verification that the limit as $n \to \infty$ of the exact size of a CS is as large as the nominal level. This involves taking the supremum over the distributions that may generate the data before taking the limit as $n \to \infty$. In regular models uniformity is often ignored (rightfully) because it holds under reasonable conditions and hence verification is just a technical exercise. This is not the case in the moment inequality model. The reason is that test statistics in this model have pointwise asymptotic distributions that are discontinuous in the true distribution that generates the data—a moment inequality enters the pointwise asymptotic distribution if and only if it holds as an equality. But, this sharp discontinuity is not a feature of the finite sample distribution. Discontinuities of this type are responsible for the problems that arise with weak instruments, near integrated processes, post-model selection inference, and parameters that are near a boundary. In such cases, (standard) bootstrap methods typically are not asymptotically valid. Furthermore, Andrews and Guggenberger (2005a,b,c,d) and Mikusheva (2008) show that even subsampling and m out of n bootstrap methods often fail to be asymptotically valid. The results of this paper show that such problems do not arise in the moment inequality example when subsampling is applied to an appropriate test statistic (and suitable moments exist). This is not true for all statistics. For example, subsampling the endpoints of the estimated set in the moment inequality model is not uniformly asymptotically valid.

The standard method in the literature for obtaining critical values for tests for multivariate one-sided null hypotheses is to use the least favorable asymptotic null distribution evaluated at a consistent estimator of the asymptotic variance matrix.[5] We refer to such tests as plug-in asymptotic (PA) tests. We show that a CS based on a PA test is uniformly asymptotically valid under similar conditions to those for subsampling. The PA critical values are at least as large as the subsampling critical values asymptotically, and in some cases strictly larger, which implies that subsampling CS's can be smaller than PA CS's. The PA CS is not asymptotically conservative

(2005a,b) established validity of subsampling based on moment inequalities in one-sample and two-sample means problems before the results of this paper were obtained.

[4]The uniformity issue arises whether one is interested in a confidence set for the true parameter or for the identified set. The issue is uniformity over the true distribution generating the data, not uniformity of coverage of all the points in the identified set.

[5]Such critical values can be calculated by computing the appropriate bound from a weighted chi-square distribution or by simulating from the least-favorable asymptotic distribution given the estimated variance matrix.

provided there are no restrictions on the moment inequalities such that satisfaction of one inequality implies violation of another. But, such restrictions do arise in some examples, e.g., see Rosen (2005).

Model specification tests are easily constructed based on subsampling or PA CS's. One rejects correct model specification if the CS is empty. Uniform asymptotic validity of such a test follows immediately from the properties of the CS. But, these tests may be asymptotically conservative. See Guggenberger, Hahn, and Kim (2007) for a different test of model specification based on moment inequalities.

Under stated high-level conditions, our results also apply to the case where preliminary consistent estimators of identified parameters are plugged-in to the sample moment functions. This can be quite useful to reduce the dimension of the parameter under test. For brevity we do not verify the high-level conditions. See Soares (2005) for more primitive conditions.

The asymptotic results given in this paper for subsampling tests also apply to m out of n bootstrap tests with i.i.d. observations provided $b^2/n \to 0$.[6] This is because subsampling based on subsamples of size b can be viewed as bootstrapping without replacement, which is not too different from bootstrapping with replacement when b^2/n is small.[7] The subsampling results apply to both i.i.d. and time series observations, whereas the m out of n bootstrap results apply only to i.i.d. observations.

We now discuss the related literature. Besides Romano and Shaikh (2005a,b), the only other results in the literature that establish uniform validity of a method for inference with moment inequalities are those of Imbens and Manski (2004), Woutersen (2006), and Stoye (2007), and those of Soares (2005) and Andrews and Soares (2007) using moment selection methods. The results of Imbens and Manski (2004) and Woutersen (2006) are quite restrictive because their Assumption 1 requires (i) super-efficiency of the implicit estimator of the length of the identified interval, which holds only in quite special cases, see Stoye (2007), and (ii) joint asymptotic normality of lower and upper bound estimators $(\widehat{\theta}_\ell, \widehat{\theta}_u)$ of the identified interval. Joint estimation of the identified set typically does not yield estimators $(\widehat{\theta}_\ell, \widehat{\theta}_u)$ that satisfy asymptotic normality (even univariate asymptotic normality), see Andrews (2005, Sections 5.2 and 6.1) for simple examples. In consequence, their results do not apply to parameters defined by moment inequalities in general. The results of Stoye (2007) that circumvent the super-efficiency condition also are quite restrictive because they assume asymptotic normality of $(\widehat{\theta}_\ell, \widehat{\theta}_u)$. Soares' (2005) and Andrews and Soares' (2007) results are obtained using the approach in this paper. Research on the power of tests in the moment inequality model is underway, see Andrews and Soares (2007).

[6]The m out of n bootstrap uses a bootstrap sample of size m when the full sample size is n, where $m \to \infty$ and $m/n \to 0$ as $n \to \infty$.

[7]In an i.i.d. scenario, the distribution of a subsample of size b is the same as the conditional distribution of a nonparametric bootstrap sample of size b conditional on there being no duplicates of observations in the bootstrap sample. If $b^2/n \to 0$, then the probability of no duplicates goes to one as $n \to \infty$, see Politis, Romano, and Wolf (1999, p. 48). In consequence, b out of n bootstrap tests and subsampling tests have the same first-order asymptotic properties.

Other papers in the literature that consider inference with moment inequalities include: CHT, Andrews, Berry, and Jia (2004, 2007), Pakes, Porter, Ho, and Ishii (2004), Moon and Schorfheide (2004), Rosen (2005), Beresteanu and Molinari (2006), Galichon and Henry (2006), Bugni (2007), and Canay (2007). To date, none of these methods has been shown to be uniformly asymptotically valid. Some of these methods have the disadvantage of being asymptotically conservative (which leads to a larger CS than desired) either all of the time or some of the time. This is true of the methods in Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2004), Rosen (2005), and Galichon and Henry (2006). The computational requirements for the different methods vary. For some methods this is a comparative advantage.

The results in this paper use the general results given in Andrews and Guggenberger (2005a) (hereafter AG1) and generalize these results in two directions that are useful in the moment inequality model and in other models. First, we relax the (partial) product space assumption on the parameter space that is employed in AG1 (see Assumption A in AG1). By doing so, the results applied to the moment inequality model allow for cases in which different moment conditions are related, e.g., one moment inequality cannot hold as an equality if some other one does. Restrictions of this type arise frequently in models with data censoring, e.g., see Rosen (2005). Second, the results provide a larger lower bound on the asymptotic size (defined to be the limit of finite-sample size) of a CS than the results in AG1. In many models, both bounds reduce to the same value and equal the upper bound. However, in the moment inequality model, the lower bound given in AG1 is not sharp whereas the lower bound given here is. Finally, the results of AG1 are for tests, whereas the results given here are for CS's. This requires uniformity of the results with respect to the parameter of interest as well as with respect to nuisance parameters. For tests the former is not required because the parameter of interest is fixed by the null hypothesis.

The general approach to uniformity given here and the way of setting up the moment inequality model to establish uniform results should be useful for analyzing the asymptotic size of CS's that employ critical values that are not based on subsampling.

The remainder of the paper is organized as follows. Section 2 discusses the issue of uniformity. Section 3 describes the moment inequality/equality model. Section 4 states the assumptions. Sections 5 and 6 introduce subsampling CS's and PA CS's, respectively, and show that these CS's are uniformly asymptotically valid for a specified class of test statistics. Section 7 introduces model specification tests. Section 8 discusses extensions. Section 9 provides general results for the asymptotic size of subsampling CS's. An Appendix contains proofs of the results.

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 $R_+ = \{x \in R : x \geq 0\}$, $R_{+,\infty} = R_+ \cup \{+\infty\}$, $R_{[+\infty]} = R \cup \{+\infty\}$, $K^p = K \times ... \times K$ (with p copies) for any set K, $\infty^p = (+\infty, ..., +\infty)'$ (with p copies). Let 0_k denote a k-vector of zeros. All limits are as $n \to \infty$. Let "pd" abbreviate "positive definite." We let AG2 abbreviate Andrews and Guggenberger (2005b).

2 Uniformity

We are interested in a CS whose exact (finite-sample) size is close to its nominal level. By definition, the exact size of the CS is the supremum of its coverage probability over distributions that may generate the data. We use asymptotics to provide an approximation to the exact size. Such an approximation is not necessarily accurate for the exact size if the asymptotic results are not uniform over the distributions that may generate the data. Thus, pointwise asymptotic results are insufficient to asymptotically validate a CS unless they hold uniformly.

When a statistic has a discontinuity in its asymptotic distribution, but not in its finite-sample distribution, pointwise asymptotics do not hold uniformly. The manifestation of this is that asymptotic distributions arise under drifting sequences of parameters that do not arise under pointwise asymptotics. Furthermore, data-dependent critical values may have probability limits under drifting sequences that are different from their probability limits under pointwise asymptotics. This is exactly what happens for tests and CS's in the moment inequality model. Given that pointwise asymptotics do not consider the full range of asymptotic behavior of the CS (which reflects the full range of its finite-sample behavior), asymptotic validity of the CS cannot be established by its behavior under pointwise asymptotics. To determine the limit of the exact size and establish uniform validity, one needs to consider drifting sequences of parameters.

How serious are uniformity issues when a test statistic has a pointwise asymptotic distribution that is discontinuous in the distribution that generates the data? The answer is that they can be very serious. For example, in the weak instrument context, Dufour (1997) has shown that the exact size of the usual nominal 5% test based on the two-stage least squares (2SLS) estimator equals 100%. In a first-order autoregressive (AR(1)) model, the nominal 95% two-sided confidence interval for the autoregressive coefficient $\rho \in (-1, 1)$ based on the usual normal critical value has asymptotic size equal to 70% when an intercept is included in the model and 39% if an intercept and time trend are included, see AG2.[8] In post-model selection inference, Kabaila (1995) shows that a standard nominal 95% confidence interval based on a post-model selection estimator utilizing a consistent model selection procedure has asymptotic size 0%, see Leeb and Pötscher (2005) for related results. All of these problems with standard inference are due to a lack of uniformity.

In problems in which a lack of uniformity arises the (standard) bootstrap typically is inconsistent. For example, for the parameter near a boundary case, see Andrews (2000). In the literature on the bootstrap, the usual prescription when the bootstrap is inconsistent is to use the m out of n bootstrap or subsampling, see AG1 for references. Politis and Romano (1994) show that subsampling is consistent under very weak conditions, also see Politis, Romano, and Wolf (1999). Similarly, the m out of n bootstrap is consistent under weak conditions. These results, however, are *pointwise* asymptotic

[8] These results and the ones given below are based on simulation of the formula for asymptotic size and hence are accurate up to simulation error.

results. They are not uniform results.

Andrews and Guggenberger (2005a,b,c,d) show that subsampling and the m out of n bootstrap are not necessarily asymptotically valid in a uniform sense. Also see Mikusheva (2008). Furthermore, the problem can be serious. For example, in the weak IV case, a nominal 5% equal-tailed two-sided subsampling test based on the 2SLS estimator has "adjusted" asymptotic size of 30% and exact size of 29% when $n = 120$, the subsample size b is 12, and 5 IVs are used, see Andrews and Guggenberger (2005c).[9] Furthermore, the exact size gets worse as $n \to \infty$ and the (unadjusted) asymptotic size is 82%. Similarly, in the AR(1) model, a nominal 95% equal-tailed two-sided subsampling confidence interval has adjusted asymptotic size of 86% and exact size of 87% when $n = 130$, the subsample size is $b = 12$, and an intercept is included in the model, see AG2. Again, the exact size gets worse as $n \to \infty$ and the (unadjusted) asymptotic size is 60%. Subsampling in the post-consistent model selection example does not solve the uniformity problem. Andrews and Guggenberger (2005d) shows that the asymptotic size of a nominal 95% confidence interval in a simple location model is actually 0%.

In the moment inequality model, uniformity issues arise for some procedures when the identified set is sufficiently small that there is a non-negligible probability of obtaining an estimated set that consists of a singleton. This scenario is of considerable empirical relevance. For example, this is the situation that arises in Andrews, Berry, and Jia (2004) and in both examples in Pakes, Porter, Ho, and Ishii (2004). Note that the identified set does not have to be a singleton for the problem to arise, it just has to be sufficiently small. Problems of this sort arise with the bootstrap applied to the interval endpoints, see Andrews (2005), with subsampling applied to the interval endpoints, and with the procedure in Pakes, Porter, Ho, and Ishii (2004) based on the pointwise asymptotic distribution of interval endpoints.[10]

3 Confidence Sets Based on Moment Inequalities

The moment inequality/equality model is defined as follows. We suppose there exists a true value θ_0 ($\in \Theta \subset R^d$) that satisfies the moment conditions:

$$\begin{aligned} &E_{F_0} m_j(W_i,\theta_0) \geq 0 \text{ for } j = 1, ..., p \text{ and} \\ &E_{F_0} m_j(W_i,\theta_0) = 0 \text{ for } j = p+1, ..., p+v, \end{aligned} \tag{3.1}$$

where $\{m_j(\cdot,\theta) : j = 1, ..., p+v\}$ are (known) real-valued moment functions and $\{W_i : i \geq 1\}$ are observed i.i.d. or stationary random vectors with joint distribution F_0.

[9] The "adjusted" asymptotic size is defined in Andrews and Guggenberger (2005b). It is based on a formula for the asymptotic size that is adjusted to take into account the ratio of the subsample size, b, to the full-sample size, n, that is actually used in a given problem. In many cases, the adjusted asymptotic size is found to be more accurate than the usual "unadjusted" asymptotic size.

[10] Also note that the probability of obtaining a singleton set does not have to be large to have adverse effects on some procedures because errors in tests or confidence intervals with probability .05 are what is typically relevant.

The true value θ_0 is not necessarily identified. Thus, knowledge of $E_{F_0} m_j(W_i,\theta)$ for all $\theta \in \Theta$ does not necessarily imply knowledge of θ_0. Furthermore, even knowledge of F_0 itself does not necessarily imply knowledge of the true value θ_0. It may require more information than is available in the observed sample $\{W_i : i \leq n\}$ to identify the true parameter θ_0. We are interested in CS's for the true value θ_0.

Let

$$m(W_i, \theta) = (m_1(W_i, \theta), ..., m_k(W_i, \theta))', \tag{3.2}$$

where $k = p + v$. Let (θ, F) denote generic values of the parameters. For i.i.d. observations, the parameter space $\mathcal{F}$ for (θ, F) is the set of all (θ, F) that satisfy:

$$\begin{aligned}
&\text{(i) } E_F m_j(W_i, \theta) \geq 0 \text{ for } j = 1, ..., p, \\
&\text{(ii) } E_F m_j(W_i, \theta) = 0 \text{ for } j = p+1, ..., k, \\
&\text{(iii) } \{W_i : i \geq 1\} \text{ are i.i.d. under } F, \\
&\text{(iv) } \sigma^2_{F,j}(\theta) = Var_F(m_j(W_i, \theta)) \in (0, \infty) \text{ for } j = 1, ..., k, \\
&\text{(v) } Corr_F(m(W_i, \theta)) \in \Psi, \text{ and} \\
&\text{(vi) } E_F |m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq M \text{ for } j = 1, ..., k,
\end{aligned} \tag{3.3}$$

where Ψ is a specified set of $k \times k$ correlation matrices, see below, and $M < \infty$ and $\delta > 0$ are fixed constants.[11] For expositional convenience, we specify $\mathcal{F}$ for dependent observations in the Appendix, see Section 10.1.

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

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_{1-\alpha}(\theta)\}, \tag{3.4}$$

where $c_{1-\alpha}(\theta)$ is a critical value. We consider subsampling and "plug-in asymptotic" critical values below.

The exact and asymptotic confidence sizes of CS_n are

$$ExCS_n = \inf_{(\theta,F) \in \mathcal{F}} P_F(T_n(\theta) \leq c_{1-\alpha}(\theta)) \text{ and } AsyCS = \liminf_{n \to \infty} ExCS_n, \tag{3.5}$$

respectively. The exact maximum coverage probability and the asymptotic maximum coverage probability are

$$ExMaxCP_n = \sup_{(\theta,F) \in \mathcal{F}} P_F(T_n(\theta) \leq c_{1-\alpha}(\theta)) \text{ and } AsyMaxCP = \limsup_{n \to \infty} ExMaxCP_n. \tag{3.6}$$

The difference $AsyMaxCP - AsyCS$ measures the magnitude of asymptotic non-similarity of the CS.

The definition of $AsyCS$ in (3.5) takes the " sup " before the " lim ." In consequence, uniformity over (θ, F) is built into the definition of $AsyCS$. Uniformity is necessary

[11] The moment condition (vi) could be relaxed slightly to uniform integrability of second moments.

for the asymptotic size to provide a good approximation to the finite-sample size of CS's. Andrews and Guggenberger (2005a,b,c,d) show that when a test statistic has a discontinuity in its limit distribution, as occurs in the moment inequality model, pointwise asymptotics (in which one takes the "lim" before the "sup") can be very misleading in some models.

We consider a general class of test statistics $T_n(\theta)$ that are defined as follows. The sample moment functions are

$$\begin{aligned} \overline{m}_n(\theta) &= (\overline{m}_{n,1}(\theta), ..., \overline{m}_{n,k}(\theta))', \text{ where} \\ \overline{m}_{n,j}(\theta) &= n^{-1} \sum_{i=1}^{n} m_j(W_i, \theta) \text{ for } j = 1, ..., k. \end{aligned} \quad (3.7)$$

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

$$\widehat{\Sigma}_n(\theta) = n^{-1} \sum_{i=1}^{n} (m(W_i, \theta) - \overline{m}_n(\theta))(m(W_i, \theta) - \overline{m}_n(\theta))'. \quad (3.8)$$

When the observations are dependent, $\widehat{\Sigma}_n(\theta)$ must take this into account. A heteroskedasticity and autocorrelation consistent (HAC) estimator may be required.

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

$$T_n(\theta) = S(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)), \quad (3.9)$$

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 1-4 stated in Section 4 below. Examples of functions that do so are now defined.

The first test function S that we consider is

$$\begin{aligned} S_1(m, \Sigma) &= \sum_{j=1}^{p} [m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v} (m_j/\sigma_j)^2, \text{ where} \\ [x]_- &= \begin{cases} x & \text{if } x < 0 \\ 0 & \text{if } x \geq 0, \end{cases} \quad m = (m_1, ..., m_k)', \end{aligned} \quad (3.10)$$

and σ_j^2 is the jth diagonal element of Σ. With this function, the parameter space Ψ for the correlation matrices in condition (v) of (3.3) is not restricted. That is, (3.3) holds with $\Psi = \Psi_1$, where Ψ_1 contains all $k \times k$ correlation matrices.[12] The function S_1 leads to the test statistic

$$T_n(\theta) = n \sum_{j=1}^{p} [\overline{m}_{n,j}(\theta)/\widehat{\sigma}_{n,j}(\theta)]_-^2 + n \sum_{j=p+1}^{p+v} (\overline{m}_{n,j}(\theta)/\widehat{\sigma}_{n,j}(\theta))^2, \quad (3.11)$$

[12] With dependent observations, Ψ is the parameter space for the limiting correlation matrix, $\lim_{n\to\infty} Corr_F(n^{1/2}\overline{m}_n(\theta))$.

where $\widehat{\sigma}_{n,j}^2(\theta) = [\widehat{\Sigma}_n(\theta)]_{jj}$. This is an Anderson–Rubin-type GMM statistic that gives positive weight to moment inequalities only when they are violated. This type of statistic has been considered in CHT and Romano and Shaikh (2005a,b).

The second test function is a Gaussian quasi-likelihood ratio (or minimum distance) function defined by

$$S_2(m, \Sigma) = \inf_{t=(t_1, 0_v): t_1 \in R^p_{+,\infty}} (m-t)'\Sigma^{-1}(m-t). \tag{3.12}$$

With this function, we restrict the parameter space Ψ in (3.3). In particular, we take $\Psi = \Psi_2$, where Ψ_2 contains all $k \times k$ correlation matrices whose determinant is greater than or equal to ε for some $\varepsilon > 0$.[13,14] This type of statistic has been considered in numerous 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 Manski and Tamer (2002) and Rosen (2005).

The following function yields a test with particularly good power against alternatives with p_1 ($\le p$) moment inequalities violated, the following function is suitable:

$$S_3(m, \Sigma) = \sum_{j=1}^{p_1} [m_{(j)}/\sigma_{(j)}]_-^2 + \sum_{j=p+1}^{p+v} (m_j/\sigma_j)^2, \tag{3.13}$$

where $[m_{(j)}/\sigma_{(j)}]_-^2$ denotes the jth largest value among $\{[m_\ell/\sigma_\ell]_-^2 : \ell = 1, ..., p\}$ and p_1 is some specified integer. The function S_3 satisfies (3.3) with $\Psi = \Psi_1$. The function S_3 is considered in Andrews, Berry, and Jia (2007). Note that the function S_1 is a special case of S_3.

Other test functions S can be considered that satisfy Assumptions 1-4. For example, one could alter S_1 or S_3 by replacing the step function $[x]_-$ by a smooth function, by replacing the square by the absolute value to a different positive power (such as one), or by adding weights.

Generally it is not possible to compare the performance of one test function/statistic with that of another without specifying the critical values to be used. The reason is that most critical values, such as the 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. Hence, a given test statistic generates different tests depending on the critical values employed and the differences do not vanish asymptotically.

[13] If Σ is singular, we define S_2 using the Moore-Penrose inverse Σ^+ in place of Σ^{-1} in (3.12). With some work, it may be possible to extend the results given below for the function $S = S_2$ to the case where $\Psi = \Psi_1$.

[14] 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 be 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 $\inf_{t_1 \in R_+} (m - t_1)^2 = \infty$.

The test statistics based on the functions S_1 and S_3 are easier to compute than those based on S_2 because the former are simple functions of the data, whereas the latter involve minimization over $t_1 \in R^p_{+,\infty}$. Computation of S_2 requires solving quadratic programming problems. This can be done quickly. But, many computations of the test statistic are required to construct a CS, especially if one is using resampling methods, because (i) one needs to compute tests for an arbitrarily large number of null parameter values θ_0 in order to construct a CS, (ii) in most cases a different critical value is needed for each null value, and (iii) each critical value requires numerous computations of the test statistic if resampling methods are employed. On the other hand, the function S_2 employs information about the correlation matrix $\Omega = D^{-1/2}\Sigma D^{-1/2}$, which has power advantages in some cases, whereas S_1 and S_3 do not.

One also could consider a test statistic that is the same as S_1 but without the division by σ_j in each summand. Pakes, Porter, Ho, and Ishii (2004) consider a test statistic of this form. In this case, the asymptotic validity results given below for subsampling and for "plug-in asymptotic" methods can be shown to hold provided $\sigma^2_{F,j}(\theta)$ is bounded away from zero in condition (iv) of (3.3). This test statistic is not recommended, however, because it is not invariant to rescaling of the moment conditions and, hence, is not likely to have good properties in terms of the volume of confidence sets. (In fact, in their empirical applications, Pakes, Porter, Ho, and Ishii (2004) find that it is desirable to consider an alternative test statistic to the one they first propose that roughly standardizes the variances of the moment conditions.)

4 Assumptions

In this section we state Assumptions 1-4 concerning the function S and show that the functions S_1–S_3 satisfy them. We also state some assumptions that are not needed for the main results given below, but are used for some peripheral results.

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 $y \in R^p \times B$ implies that $G(y) \to G(x)$ as $\to \infty$. In the assumptions below, the set Ψ is as in condition (v) of (3.3).[15] For p-vectors m_1, 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 $\Omega \in \Psi$.

Assumption 2. For all $h_1 \in R^p_{+,\infty}$ all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the distribution function (df) of $S(Z + (h_1, 0_v), \Omega)$ at $x \in R$ is

[15] For temporally dependent observations, Ψ is as in condition (iv) of (10.2) in the Appendix.

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

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

In Assumption 2, if an element of h_1 equals $+\infty$, then by definition the corresponding element of $Z + (h_1, 0_v)$ equals $+\infty$.[16] Assumptions 1-3 are used for subsampling CS's. Assumptions 1 and 4 are used for PA CS's.

Assumptions 1-4 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. Assumptions 2 and 3 are used to show that certain asymptotic df's satisfy suitable continuity/strictly-increasing properties. These properties ensure that the subsampling 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 3 implies that $S(\infty^p, \Sigma) = 0$ when $v = 0$. Assumption 4 is a mild continuity assumption.

Lemma 1 *The functions $S_1(m, \Sigma)$–$S_3(m, \Sigma)$ satisfy Assumptions* 1-4 *with $\Psi = \Psi_1$ for $S_1(m, \Sigma)$ and $S_3(m, \Sigma)$ and with $\Psi = \Psi_2$ for $S_2(m, \Sigma)$.*

Comment. In Lemma 1, the function S_2 requires the correlation matrices to be bounded away from singularity, whereas none of the other functions require this.

Next we introduce three conditions that are not needed to show that subsampling and PA CS's are asymptotically valid (i.e., $AsyCS \geq 1-\alpha$). Rather, the first and third conditions are used to show that subsampling and PA CS's, respectively, are not asymptotically conservative (i.e., $AsyCS \not> 1-\alpha$). The second condition is used when showing that subsampling CS's have $AsyMaxCP = 1$ when $v = 0$.

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 C1. 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 $Z \sim N(0_k, \Omega(\theta, F))$.

[16] In Assumptions 1(d) and 4(b), $S(m, \Omega)$ 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 Ω uniformly for $\Omega \in \Psi$ if for all $\eta > 0$ there exists $\delta > 0$ such that whenever $||\Omega^* - \Omega|| < \delta$ for $\Omega^* \in \Psi_1$ and $\Omega \in \Psi$ we have $|c_{\Omega^*}(1-\alpha) - c_\Omega(1-\alpha)| < \eta$.

Assumption C2. For some $(\theta, F) \in \mathcal{F}$, $E_F m_j(W_i, \theta) > 0$ for $j = 1, ..., p$.

Assumption C3. For some $(\theta, F) \in \mathcal{F}$ with $h_1(\theta, F) = 0_p$, the df of $S(Z, \Omega(\theta, F))$ is continuous at its $1 - \alpha$ quantile, where $Z \sim N(0_k, \Omega(\theta, F))$.

Assumption C1 is a very weak continuity condition. (Hence, subsampling CS's typically are not asymptotically conservative.) Assumption C2 typically holds if the identified set is not a singleton. Assumption C3 holds quite generally if there are no restrictions relating the expectation of one moment function to that of another. But, if such restrictions exist, then Assumption C3 fails and the PA CS is asymptotically conservative. (Assumption C3 fails when there are restrictions because there is no $(\theta, F) \in \mathcal{F}$ with $h_1(\theta, F) = 0_p$.) For example, Assumption C3 fails in a regression model in which one only observes the integer part of a latent dependent variable.

5 Subsampling Confidence Sets

We now define subsampling critical values and CS's. Let b denote the subsample size when the full-sample size is n. We assume $b \to \infty$ and $b/n \to 0$ as $n \to \infty$ (throughout the paper). The choice of b is discussed in the subsampling literature, e.g., see Politis, Romano, and Wolf (1999). We do not discuss it further here. (It is beyond the scope of this paper.) The number of different subsamples of size b is q_n. With i.i.d. observations, there are $q_n = n!/((n-b)!b!)$ different subsamples of size b. With time series observations, there are $q_n = n - b + 1$ subsamples each consisting of b consecutive observations.

The subsample statistics used to construct the subsampling critical value are $\{T_{n,b,j}(\theta) : j = 1, ..., q_n\}$, where $T_{n,b,j}(\theta)$ is a subsample statistic defined exactly as $T_n(\theta)$ is defined but based on the jth subsample of size b rather than the full sample. The empirical df and $1 - \alpha$ sample quantile of $\{T_{n,b,j}(\theta) : j = 1, ..., q_n\}$ are

$$\begin{aligned} U_{n,b}(\theta, x) &= q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta) \leq x) \text{ for } x \in R \text{ and} \\ c_{n,b}(\theta, 1-\alpha) &= \inf\{x \in R : U_{n,b}(\theta, x) \geq 1 - \alpha\}. \end{aligned} \tag{5.1}$$

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 (3.4) with $c_{1-\alpha}(\theta) = c_{n,b}(\theta, 1-\alpha)$.

The following Theorem applies to i.i.d. observations, in which case $\mathcal{F}$ is defined in (3.3), and to dependent observations, in which case for brevity $\mathcal{F}$ is defined in (10.2)-(10.3) in the Appendix.

Theorem 1 *Suppose Assumptions* 1-3 *hold and* $0 < \alpha < 1/2$. *Then, the nominal level* $1 - \alpha$ *subsampling CS based on* $T_n(\theta)$ *satisfies*

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

(b) $AsyCS = 1 - \alpha$ *if Assumption* C1 *also holds, and*

(c) $AsyMaxCP = 1$ *if* $v = 0$ (*i.e., no moment equalities appear*) *and Assumption* C2 *also holds.*

Comments. 1. An important feature of Theorem 1 is that no assumptions are placed on the moment functions $m(W_i, \theta)$ beyond the existence of mild moment conditions (e.g., $2 + \delta$ moments finite in the i.i.d. case) that appear in the definition of $\mathcal{F}$ and Assumption C2 that is used in Theorem 1(c).

2. The asymptotic distribution of $T_n(\theta)$ differs depending on the sequence of true values $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ considered. The Appendix provides explicit expressions for $AsyCS$ and $AsyMaxCP$ in terms of these asymptotic distributions, see (10.13) and the definitions in (10.10) and (9.3). Hence, $AsyMaxCP$ can be evaluated in cases in which $v \geq 1$.

3. The results of Theorem 1 hold 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 a variety of models. For example, they arise in a location model with interval outcomes. In this model, y_i is observed, y_i^* and u_i are not observed, $y_i^* = \theta_0 + u_i$ for $i = 1, ..., n$, $y_i = [y_i^*]$ (i.e., y_i equals the integer part of y_i^*), and u_i has mean zero. The interval outcome $[y_i, y_i + 1]$ necessarily includes the unobserved outcome variable y_i^*. Two moment inequalities that place bounds on θ_0 are (i) $-E_{\theta_0} y_i + \theta_0 \geq 0$ and (ii) $E_{\theta_0} y_i + 1 - \theta_0 \geq 0$. Obviously, both inequalities cannot simultaneously hold as equalities. Subsampling automatically takes this into account and generates a (data-dependent) critical value that is smaller than what one would obtain if no functional relationship existed between the two moment functions. This yields a CS that is smaller than otherwise, as is desirable.

The subsample statistic can be defined using a re-centering and the uniform asymptotic validity results go through with some additional effort. The re-centered subsample statistic $\widehat{T}_{n,b,j}(\theta)$ is defined to be

$$\widehat{T}_{n,b,j}(\theta) = S(b^{1/2}(\overline{m}_{n,b,j}(\theta) - [\overline{m}_n(\theta)]_-^*), \widehat{\Sigma}_{n,b,j}(\theta)), \tag{5.2}$$

where $\overline{m}_{n,b,j}(\theta)$ is the sample average based on the observations in the jth subsample, $[x]_-^* = ([x_1]_-, ..., [x_p]_-, x_{p+1}, ..., x_k)'$ for $x \in R^k$, and $\widehat{\Sigma}_{n,b,j}(\theta)$ is the variance matrix estimator based on the observations in the jth subsample. Chernozhukov and Fernandez-Val (2005) considers a re-centered subsampling method in the context of inference for quantile processes. The argument behind re-centering is that the re-centered subsample statistic does not diverge in probability to infinity under fixed alternatives. Hence, it may have better power properties than the subsampling test that does not use re-centering. On the other hand, the simulation results in Linton, Maasoumi, and Whang (2005) for a different testing problem (viz., tests of stochastic dominance) show that the re-centered subsampling test performs substantially worse in terms of both size and power than the subsampling test without re-centering.

The subsample statistic also can be defined using a full sample variance estimator and the uniform asymptotic validity results go through with some additional effort.

The subsample statistic $\widehat{T}_{n,b,j}(\theta)$ is defined to be of the form

$$\widehat{T}_{n,b,j}(\theta) = S(b^{1/2}\overline{m}_{n,b,j}(\theta), \widehat{\Sigma}_n(\theta)), \tag{5.3}$$

The argument in favor of using a full sample estimator of Σ in the subsample statistics is that a subsample estimator is not needed for asymptotic validity and the full-sample estimator is more accurate, which may lead to better size and power properties in finite samples.

One can use a full-sample variance estimator in the re-centered subsample statistic. In this case, the subsample statistic is defined to be

$$\widehat{T}_{n,b,j}(\theta) = S(b^{1/2}(\overline{m}_{n,b,j}(\theta) - [\overline{m}_n(\theta)]_-^*), \widehat{\Sigma}_n(\theta)). \tag{5.4}$$

6 Plug-in Asymptotic Confidence Sets

Next, we discuss CS's based on an asymptotic critical value. The least-favorable asymptotic null distributions of the statistic $T_n(\theta)$ are shown to be those for which the moment inequalities hold as equalities. These distributions depend on the (asymptotic) correlation matrix, Ω, of the moment functions. We consider plug-in asymptotic (PA) critical values that are obtained from the least-favorable asymptotic null distribution evaluated at a consistent estimator of Ω. Critical values of this type have long been considered in the literature on multivariate one-sided tests, see Silvapulle and Sen (2005) for references. They have been considered in the moment inequality literature by Rosen (2005). We exploit results in AG2 for "plug-in size-corrected fixed critical values" to obtain the asymptotic results given in this section.

As above, 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. Define

$$\widehat{\Omega}_n(\theta) = \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta)\widehat{D}_n^{-1/2}(\theta), \tag{6.1}$$

where $\widehat{D}_n(\theta) = Diag(\widehat{\Sigma}_n(\theta))$ and $\widehat{\Sigma}_n(\theta)$ is defined in (3.8) for i.i.d. observations and is a consistent estimator of $\lim_{n\to\infty} Var(n^{1/2}\overline{m}_n(\theta))$ for dependent observations.

The nominal $1-\alpha$ PA CS is given by (3.4) with critical value $c_{1-\alpha}(\theta)$ equal to

$$c(\widehat{\Omega}_n(\theta), 1-\alpha). \tag{6.2}$$

Theorem 2 *Suppose Assumptions* 1 *and* 4 *hold and* $0 < \alpha < 1/2$. *Then, the nominal level* $1-\alpha$ PA CS *based on* $T_n(\theta)$ *satisfies*

(a) $AsyCS \geq 1-\alpha$ *and*

(b) $AsyCS = 1-\alpha$ *provided Assumption* C3 *also holds.*

Comment. Theorem 2(a) 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. However, Theorem 2(b) does not hold in this case because Assumption C3 fails. The PA critical value does not automatically take functional relationships between the moment functions into account as the subsampling critical value does. The PA critical value is larger than necessary and the PA CS is asymptotically conservative in this scenario. Thus, subsampling CS's have advantages over PA CS's in this scenario.

7 Model Specification Tests

Tests of model specification can be constructed using the subsampling and PA CS's introduced in Sections 5 and 6. The null hypothesis of interest is that there exists a parameter $\theta_0 \in \Theta$ such that (3.1) holds (with additional conditions specified by the parameter space for (θ, F), such as those in (3.3) or those given in the Appendix for temporally dependent observations). The idea of such specification tests is the same as for the J test of over-identifying restrictions in GMM, see Hansen (1982). With the J test, one rejects the null hypothesis of correct model specification if the GMM criterion function evaluated at the GMM estimator is sufficiently large. In the moment inequality/equality model, the analogue of the GMM criterion function is the test statistic $T_n(\theta)$. By definition, the subsampling test rejects the model specification if $T_n(\theta)$ exceeds the subsampling critical value $c_{n,b}(\theta, 1-\alpha)$ for all $\theta \in \Theta$. Equivalently, it rejects if the subsampling CS is empty. The PA model specification test is analogous with the PA critical value in place of the subsampling critical value.

If the model specified in (3.1) is correctly specified, then the subsampling CS and the PA CS contain the true value with asymptotic probability $1-\alpha$ (or greater) uniformly over the parameter space. Hence, under the null hypothesis of correct model specification, the limit as $n \to \infty$ of the finite-sample size of the subsampling and PA model specification tests are $\leq \alpha$ under the assumptions of Theorems 1(a) and 2(a), respectively. Note that the model specification tests may be asymptotically conservative (i.e., have asymptotic size $< \alpha$) even when the assumptions of part (b) of those Theorems hold. As discussed above, it is crucial that the asymptotic sizes of these tests are shown to be valid uniformly over the parameter space because the present testing scenario is one in which the test statistic $T_n(\theta)$ has a limit distribution that is discontinuous in the parameters.

8 Extensions

8.1 Generalized Empirical Likelihood Statistics

Here we consider CS's for parameters in the moment inequality/equality model based on a generalized empirical likelihood (GEL) test statistic, $T_n^{GEL}(\theta)$, rather than a test statistic of the form $T_n(\theta)$ in (3.9). In the context of moment equalities, Smith (1997) considers GEL statistics. In the context of moment inequalities, Soares (2006) considers GEL statistics and Moon and Schorfheide (2004) and Canay (2007) consider empirical likelihood statistics. In the Appendix, we show that the asymptotic distribution of $T_n^{GEL}(\theta)$ is the same as that of the statistic $T_n(\theta)$ in (3.9) based on $S_2(m, \Sigma)$ in (3.12) when the observations come from a row-wise i.i.d. triangular array.[17] In consequence, by the same argument as for the latter statistic, GEL-based subsampling and PA CS's based on $T_n^{GEL}(\theta)$ have correct asymptotic size.

For $t \in R_+^p$, define

$$m_i(t, \theta) = m(W_i, \theta) - (t, 0_v). \tag{8.1}$$

The vector t can be viewed as an additional nuisance parameter that captures the slackness in the first p moment inequalities. The minimum distance (MD) formulation of the empirical likelihood (EL) statistic for inference under moment equalities and inequalities is given by

$$L_{EL}(\theta, t) = \sup_{\pi} \{ \textstyle\prod_{i=1}^n \pi_i : \pi_i \geq 0, \ \sum_{i=1}^n \pi_i = 1, \ \sum_{i=1}^n \pi_i m_i(t, \theta) = 0 \}, \tag{8.2}$$

where $\pi = (\pi_1, ..., \pi_n)'$. Under weak additional assumptions, the MD formulation of the EL estimator $\widehat{\theta}_{EL} = \arg\max_{\theta \in \Theta} \sup_{t \in R_+^p} L_{EL}(\theta, t)$ can be re-expressed (equivalently) as the solution to a saddlepoint problem $\widehat{\theta}_{EL} = \arg\min_{\theta \in \Theta} \inf_{t \in R_+^p} \sup_{\lambda \in \widehat{\Lambda}_n(t,\theta)} 2 \sum_{i=1}^n \ln(1 - \lambda' m_i(t, \theta))$, where $\widehat{\Lambda}_n(t, \theta)$ is defined in (8.3) below. We consider the GEL generalization of this saddlepoint problem and work with the statistic

$$\begin{aligned} T_n^{GEL}(\theta) &= \inf_{t \in R_+^p} \sup_{\lambda \in \widehat{\Lambda}_n(t,\theta)} n \widehat{P}_\rho(t, \theta, \lambda), \text{ where} \\ \widehat{P}_\rho(t, \theta, \lambda) &= 2n^{-1} \textstyle\sum_{i=1}^n (\rho(\lambda' m_i(t, \theta)) - \rho(0)), \\ \widehat{\Lambda}_n(t, \theta) &= \{\lambda \in R^k : \lambda' m_i(t, \theta) \in Q \text{ for } i = 1, ..., n\}, \end{aligned} \tag{8.3}$$

Q is an open interval of the real line that contains 0, and $\rho : Q \to R$ is a concave function that is twice continuously differentiable on a neighborhood of 0 with first and second derivatives at 0 normalized to equal -1. For $\rho(x) = \ln(1 - x)$ we obtain the EL estimator; for $\rho(x) = -(1 + x)^2/2$, we obtain the continuous updating estimator; and for $\rho(x) = -\exp x$, we obtain the exponential tilting estimator. In the Appendix we show that under the i.i.d. setup of (3.3) and Assumption GEL the equivalents of Theorems 1 and 2 hold for CS's based on $T_n^{GEL}(\theta)$ rather than $T_n(\theta)$ in (3.9).

8.2 Preliminary Estimation of Identified Parameters

Suppose the population moment functions are of the form $E_F m_j(W_i, \theta_0, \tau_0) \geq 0$ for $j = 1, ..., p$ and $E_F m_j(W_i, \theta_0, \tau_0) = 0$ for $j = p + 1, ..., k$, where τ_0 is a parameter for which a preliminary asymptotically normal estimator $\widehat{\tau}_n(\theta_0)$ exists. Of course, this typically requires that τ_0 is identified. Soares (2005) considers this scenario in some detail.

[17] This result holds under Assumption GEL (stated in the Appendix) under sequences of parameters $\{\gamma_{w_n,h} : n \geq 1\}$ as in Assumption B0 below.

The sample moment functions in this case are of the form $\overline{m}_{n,j}(\theta) = \overline{m}_{n,j}(\theta, \widehat{\tau}_n(\theta))$. The asymptotic variance of $n^{1/2}\overline{m}_{n,j}(\theta)$ is different when τ_0 is replaced by the $\widehat{\tau}_n(\theta)$, but otherwise the theoretical treatment of this model is the same. In fact, Theorems 1 and 2 hold in this case using the conditions given in (10.3) of the Appendix. 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).

9 General Results for Subsampling Confidence Sets

This section provides general results for CS's. These results are used in the Appendix to prove Theorem 1 for subsample CS's in the moment inequality model.

Let $R_\infty = R \cup \{\pm\infty\}$.

9.1 Definition of Confidence Sets

We consider CS's for a parameter $\theta \in R^d$ when nuisance parameters $\eta \in R^s$ and $\gamma_3 \in \mathcal{T}_3$ may appear, where $\mathcal{T}_3$ is an arbitrary, possibly infinite-dimensional, space. We obtain CS's for θ by inverting tests based on a test statistic $T_n(\theta_0)$ for testing the null hypothesis $H_0 : \theta = \theta_0$. Fixed and subsampling critical values are considered. Let Θ $(\subset R^d)$ denote the parameter space for θ. The CS for θ is defined as in (3.4). The focus of this section is on the behavior of CS's when the asymptotic distribution of $T_n(\theta)$ depends on the parameters (θ, η) and is discontinuous at some value(s) of (θ, η).

We partition θ and η into (θ_1, θ_2) and (η_1, η_2), where $\theta_j \in R^{d_j}$ and $\eta_j \in R^{s_j}$ for $j = 1, 2$. By definition (made precise below), $\gamma_1 = (\theta_1, \eta_1)$ are parameters that determine how close the asymptotic distribution of $T_n(\theta)$ is to a point of discontinuity and $\gamma_2 = (\theta_2, \eta_2)$ are parameters that do not do so, but still may affect the asymptotic distribution of $T_n(\theta)$. The parameter γ_3 does not affect the asymptotic distribution of $T_n(\theta)$. Define $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, where $\gamma_1 \in R^p$, $\gamma_2 \in R^q$, $p = d_1 + s_1$, and $q = p_2 + s_2$. Let Γ denote the parameter space for γ. In most models, either no parameter θ_1 or θ_2 appears (i.e., $d_1 = 0$ or $d_2 = 0$). For example, in the moment inequality model, $d_1 = 0$.

$ExCS_n$, $AsyCS$, $ExMaxCP_n$, and $AsyMaxCP$ are defined as in (3.5) and (3.6) with $\gamma \in \Gamma$ in place of $(\theta, F) \in \mathcal{F}$.

9.2 Critical Values

A test rejects the null hypothesis when $T_n(\theta_0)$ exceeds some critical value. We consider two types of critical values for use with the test statistic $T_n(\theta_0)$. The first is a *fixed critical value* (FCV) and is denoted $c_{Fix}(\theta_0, 1-\alpha)$, where $\alpha \in (0, 1)$ is the nominal size of the FCV test. The FCV test rejects H_0 when $T_n(\theta_0) > c_{Fix}(\theta_0, 1-\alpha)$. A common choice is $c_{Fix}(\theta_0, 1-\alpha) = c_\infty(1-\alpha)$, where $c_\infty(1-\alpha)$ denotes the $1-\alpha$ quantile of J_∞ and J_∞ is the asymptotic null distribution of $T_n(\theta_0)$ when γ is fixed and is not a point of discontinuity.

The second type of critical value that we consider is a subsampling critical value. Let b and q_n be as in Section 5. The subsample statistics that are used to construct the subsampling critical value are denoted by $\{\widehat{T}_{n,b,j}(\theta_0) : j = 1, ..., q_n\}$ when testing $H_0 : \theta = \theta_0$.

Let $\{T_{n,b,j}(\theta) : j = 1, ..., q_n\}$ be subsample statistics that are defined exactly as $T_n(\theta)$ is defined, but are based on subsamples of size b rather than the full sample.

In most cases, the subsample statistics $\{\widehat{T}_{n,b,j}(\theta_0) : j = 1, ..., q_n\}$ are defined to satisfy one or the other of the following assumptions.

Assumption Sub1. $\widehat{T}_{n,b,j}(\theta_0) = T_{n,b,j}(\widehat{\theta}_n)$ for all $j \leq q_n$, where $\widehat{\theta}_n$ is an estimator of θ.

Assumption Sub2. $\widehat{T}_{n,b,j}(\theta_0) = T_{n,b,j}(\theta_0)$ for all $j \leq q_n$.

The estimator $\widehat{\theta}_n$ in Assumption Sub1 usually is chosen to be an estimator that is consistent under both the null and alternative hypotheses. In the moment inequality example, the subsample statistics are defined such that Assumption Sub2 holds—because we do not assume that θ is identified and hence no consistent estimator $\widehat{\theta}_n$ is available.

Let $L_{n,b}(\theta, x)$ and $c_{n,b}(\theta, 1-\alpha)$ denote the empirical df and $1-\alpha$ sample quantile, respectively, of the subsample statistics $\{\widehat{T}_{n,b,j}(\theta) : j = 1, ..., q_n\}$. By definition,

$$\begin{aligned} L_{n,b}(\theta, x) &= q_n^{-1} \sum_{j=1}^{q_n} 1(\widehat{T}_{n,b,j}(\theta) \leq x) \text{ for } x \in R \text{ and} \\ c_{n,b}(\theta, 1-\alpha) &= \inf\{x \in R : L_{n,b}(\theta, x) \geq 1-\alpha\}. \end{aligned} \tag{9.1}$$

The subsampling test rejects $H_0 : \theta = \theta_0$ if $T_n(\theta_0) > c_{n,b}(\theta_0, 1-\alpha)$.

9.3 Parameter Space

The parameter space for γ is Γ.

Assumption A0. $\Gamma \subset \{(\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in R^p, \gamma_2 \in R^q, \gamma_3 \in \mathcal{T}_3\}$.

In contrast to Assumption A of AG1, Assumption A0 does not require γ_1 to lie in a product space in R^p and does not require (γ_1, γ_2) to lie in a product space of the form $\Gamma_1 \times \Gamma_2$ for some $\Gamma_1 \subset R^p$ and $\Gamma_2 \subset R^q$. The latter product space condition is typically violated in the moment inequality model. The former product space condition is sometimes violated in the moment inequality model. The relaxation of Assumption A of AG1 to Assumption A0 is a substantial contribution of this paper. It is useful in a variety of models beyond the moment inequality model.

9.4 Convergence Assumption

For an arbitrary distribution G, let $G(\cdot)$ denote the df of G, let $G(x-)$ denote the limit from the left of $G(\cdot)$ at x, and let $C(G)$ denote the set of continuity points of $G(\cdot)$. Define the $1-\alpha$ quantile, $q(1-\alpha)$, of a distribution G by $q(1-\alpha) = \inf\{x \in R :$

$G(x) \geq 1-\alpha\}$. The distribution J_h considered below is the distribution of a proper random variable that is finite with probability one.

Let $r > 0$ denote a *rate of convergence index* such that when the true parameter γ_1 satisfies $n^r\gamma_1 \to h_1$, then the test statistic $T_n(\theta_0)$ has an asymptotic distribution that depends on the localization parameter h_1 (see Assumption B0 below). In most examples, including the moment inequality example, $r = 1/2$. For a given model, we assume there is a single fixed $r > 0$.

Let $\{w_n : n \geq 1\}$ denote some subsequence of $\{n\}$. Given $\{w_n\}$, we consider sequences of parameters with the following properties.

Definition of $\{\gamma_{w_n,h} : n \geq 1\}$**:** Given $r > 0$ and $h = (h_1, h_2) \in R^p_\infty \times R^q_\infty$, let $\{\gamma_{w_n,h} = (\gamma_{w_n,h,1}, \gamma_{w_n,h,2}, \gamma_{w_n,h,3}) : n \geq 1\}$ denote a sequence of parameters in Γ for which $w_n^r\gamma_{w_n,h,1} \to h_1$, $\gamma_{w_n,h,2} \to h_2$, $\gamma_{w_n,h} = ((\theta_{w_n,h,1}, \eta_{w_n,h,1}), (\theta_{w_n,h,2}, \eta_{w_n,h,2}), \gamma_{w_n,h,3})$ and $\theta_{w_n,h} = (\theta_{w_n,h,1}, \theta_{w_n,h,2})$ if such a sequence exists.

Define

$$H = \{h \in R^p_\infty \times R^q_\infty : \exists \text{ a subsequence } \{w_n\} \text{ and a sequence } \{\gamma_{w_n,h} : n \geq 1\}\}. \quad (9.2)$$

The sequence $\{\gamma_{w_n,h} : n \geq 1\}$ is defined such that under $\{\gamma_{w_n,h} : n \geq 1\}$, the asymptotic distribution of $T_{w_n}(\theta_{w_n,h})$ depends on h and only h.

Assumption B0. For some $r > 0$, all $h \in H$, all subsequences $\{w_n\}$ of $\{n\}$, all sequences $\{\gamma_{w_n,h} : n \geq 1\}$, and some distributions J_h, $T_{w_n}(\theta_{w_n,h}) \to_d J_h$ under $\{\gamma_{w_n,h} : n \geq 1\}$, where $\gamma_{w_n,h} = ((\theta_{w_nh,1}, \eta_{w_n,h,1}), (\theta_{w_n,h,2}, \eta_{w_n,h,2}), \gamma_{w_n,h,3})$ and $\theta_{w_n,h} = (\theta_{w_n,h,1}, \theta_{w_n,h,2})$.

Assumption B0 is a strengthening of Assumption B of AG1 to cover subsequences $\{w_n\}$ rather than just sequences $\{n\}$. Also it differs slightly from Assumption B of AG1 because it applies to CS's rather than tests. Although more complicated, Assumption B0 is usually not more difficult to verify than Assumption B. When Assumption A of AG1 holds, Assumptions B and B0 are equivalent, see the proof of (30) of AG1. Assumption B0 holds in a wide variety of examples of interest, see below for the moment inequality model and Andrews and Guggenberger (2005a,b,c,d) for other models.

9.5 Subsampling Assumptions

Theorem 3 below shows that the asymptotic size of a subsampling CS is determined by the asymptotic distributions of the full-sample statistic $T_{w_n}(\theta_{w_n,h})$ and the subsample statistic $T_{w_n,b_{w_n},j}(\theta_{w_n,h})$ under certain parameter sequences $\{\gamma_{w_n,g,h} : n \geq 1\}$. By Assumption B0, the asymptotic distribution of $T_{w_n}(\theta_{w_n,h})$ is J_h. The asymptotic distribution of $T_{w_n,b_{w_n},j}(\theta_{w_n,h})$ under $\{\gamma_{w_n,g,h} : n \geq 1\}$ is shown to be J_g for $g \in H$.

Definition of $\{\gamma_{w_n,g,h} : n \geq 1\}$**:** Given $r > 0$, $g = (g_1, g_2) \in R^p_\infty \times R^q_\infty$, and $h = (h_1, h_2) \in R^p_\infty \times R^q_\infty$ with $g_2 = h_2$, let $\{\gamma_{w_n,g,h} = (\gamma_{w_n,g,h,1}, \gamma_{w_n,g,h,2}, \gamma_{w_n,g,h,3}) : n \geq 1\}$ denote a sequence of parameters in Γ for which $w_n^r\gamma_{w_n,g,h,1} \to h_1$, $b_{w_n}^r\gamma_{w_n,g,h,1} \to$

g_1, $\gamma_{w_n,g,h,2} \to h_2$, $\gamma_{w_n,g,h} = ((\theta_{w_n,g,h,1}, \eta_{w_n,g,h,1}), (\theta_{w_n,g,h,2}, \eta_{w_n,g,h,2}), \gamma_{w_n,g,h,3})$ and $\theta_{w_n,g,h} = (\theta_{w_n,g,h,1}, \theta_{w_n,g,h,2})$ if such a sequence exists.

By definition, a sequence $\{\gamma_{w_n,g,h} : n \geq 1\}$ also is of the form $\{\gamma_{w_n,h} : n \geq 1\}$.

The index set of the asymptotic distributions of $T_{w_n}(\theta_{w_n,h})$ and $T_{w_n,b_{w_n},j}(\theta_{w_n,h})$ under sequences $\{\gamma_{w_n,g,h} : n \geq 1\}$ is denoted by GH. By definition,

$$\begin{aligned} GH = \{&(g,h) \in (R_\infty^p \times R_\infty^q)^2 : \exists \text{ a subsequence } \{w_n\} \text{ and} \\ &\text{a sequence } \{\gamma_{w_n,g,h} : n \geq 1\}\}. \end{aligned} \tag{9.3}$$

By definition of $\{\gamma_{w_n,g,h} : n \geq 1\}$ and Assumption C below (i.e., $b/n \to 0$), for all $(g,h) = ((g_1,g_2),(h_1,h_2)) \in GH$, we have $g_2 = h_2$ and $|g_{1,j}| \leq |h_{1,j}|$ for $j = 1, ..., p$, where $g_1 = (g_{1,1}, ..., g_{1,p})'$ and $h_1 = (h_{1,1}, ..., h_{1,p})'$.

For subsampling CS's, we require the following additional assumptions:

Assumption C. (i) $b \to \infty$ and (ii) $b/n \to 0$.

Assumption D. (i) $\{T_{n,b,j}(\theta) : j = 1, ..., q_n\}$ are identically distributed under any $\gamma \in \Gamma$ for all $n \geq 1$ and (ii) $T_{n,b,j}(\theta)$ and $T_b(\theta)$ have the same distribution under any $\gamma \in \Gamma$ for all $n \geq 1$, where $\theta = (\theta_1, \theta_2)$ and $\gamma = ((\theta_1, \eta_1), (\theta_2, \eta_2), \gamma_3)$.

Assumption E0. For all subsequences $\{w_n\}$ of $\{n\}$ and all sequences $\{\gamma_{w_n,g,h} \in \Gamma : n \geq 1\}$, $U_{w_n,b_{w_n}}(\theta_{w_n,g,h}, x) - E_{\gamma_{w_n,g,h}} U_{w_n,b_{w_n}}(\theta_{w_n,g,h}, x) \to_p 0$ under $\{\gamma_{w_n,g,h} : n \geq 1\}$ for all $x \in R$, where $\theta_{w_n,g,h} = ((\theta_{w_n,g,h,1}, \theta_{w_n,g,h,2})$ and $\gamma_{w_n,g,h} = ((\theta_{w_n,g,h,1}, \eta_{w_n,g,h,1}), (\theta_{w_n,g,h,2}, \eta_{w_n,g,h,2}), \gamma_{w_n,g,h,3})$.

Assumption F. For all $\varepsilon > 0$ and $h \in H$, $J_h(c_h(1-\alpha) + \varepsilon) > 1 - \alpha$, where $c_h(1-\alpha)$ is the $1-\alpha$ quantile of J_h.

Assumption G0. For all $h = (h_1, h_2) \in H$, all subsequences $\{w_n\}$ of $\{n\}$, and all sequences $\{\gamma_{w_n,g,h} : n \geq 1\}$, if $U_{w_n,b_{w_n}}(\theta_{w_n,g,h}, x) \to_p J_g(x)$ under $\{\gamma_{w_n,g,h} : n \geq 1\}$ for all $x \in C(J_g)$, then $L_{w_n,b_{w_n}}(\theta_{w_n,g,h}, x) - U_{w_n,b_{w_n}}(\theta_{w_n,g,h}, x) \to_p 0$ under $\{\gamma_{w_n,g,h} : n \geq 1\}$ for all $x \in C(J_g)$, where $\theta_{w_n,g,h} = ((\theta_{w_n,g,h,1}, \theta_{w_n,g,h,2})$ and $\gamma_{w_n,g,h} = ((\theta_{w_n,g,h,1}, \eta_{w_n,g,h,1}), (\theta_{w_n,g,h,2}, \eta_{w_n,g,h,2}), \gamma_{w_n,g,h,3})$.

Assumptions C, D, and F are the same as in AG1. Assumptions E0 and G0 are extensions of Assumptions E and G of AG1 to cover subsequences $\{w_n\}$ rather than just full sequences $\{n\}$. Assumptions C and D are standard assumptions in the subsampling literature and are not restrictive. Assumption D necessarily holds when the observations are i.i.d. or stationary and the subsamples are constructed in the usual way. Assumption E0 holds automatically when the observations are i.i.d. for each fixed $\gamma \in \Gamma$ or are stationary strong mixing for each fixed $\gamma \in \Gamma$ and $\sup_{\gamma \in \Gamma} \alpha_\gamma(j) \to 0$ as $j \to \infty$, where $\{\alpha_\gamma(j) : j \geq 1\}$ are the strong mixing numbers of the observations when the true parameter is γ, see AG1. Assumption F is not restrictive. It holds in all of the examples considered in Andrews and Guggenberger (2005a,b,c,d). Assumption G0 holds automatically when Assumption Sub2 holds, as occurs with the moment inequality subsample statistics considered in this paper. In AG1, sufficient conditions for Assumption G are given when Assumption Sub1 holds. These can be extended to provide sufficient conditions for Assumption G0.

9.6 Asymptotic Results

Theorem 3 below is a CS analogue of the testing results of Theorem 1 of AG1 but with two improvements that are needed in the moment inequality example. The first improvement is that the product space form of Γ_1 and $\Gamma_1 \times \Gamma_2$ is eliminated (Assumption A of AG1 is replaced by Assumption A0.) This extension is useful in many models. The price to pay for this extension is the more complicated form of GH here than in AG1 and the more complicated forms of Assumptions B0, E0, and G0, which involve subsequences $\{w_n\}$, than Assumptions B, E, and G of AG1.

The second improvement is that Theorem 3 provides a larger lower bound on $AsyCS$ than does the straight analogue of Theorem 1 of AG1. In most examples, continuity of $J_h(x)$ at suitable values of (h, x) yields the lower and upper bounds given in Theorem 1 of AG1 to be equal and, hence, the latter delivers the precise value of asymptotic size. This continuity does not hold in the moment inequality example when $v = 0$. We introduce an improvement that is applicable in models in which $J_h(x)$ has a discontinuity at $x = c_g(1-\alpha)$ for some $(g, h) \in GH$ and the test statistic and the subsample statistics have a common lower bound on their support for all $n \geq 1$. The improvement is possible because the test statistic and the subsampling critical values cannot be smaller than the lower bound.

Let GH^* be the set of points $(g, h) \in GH$ such that for all sequences $\{\gamma_{w_n,g,h} : n \geq 1\}$, we have

$$\liminf_{n\to\infty} P_{\gamma_{w_n,g,h}}(T_{w_n}(\theta_{w_n,g,h}) \leq c_{w_n,b_{w_n}}(\theta_{w_n,g,h}, 1-\alpha)) \geq J_h(c_g(1-\alpha)). \tag{9.4}$$

The improved lower bound on $AsyCS$ for subsampling CS's is

$$Min^{-}_{CS,Sub}(\alpha) = \min\left\{\inf_{(g,h)\in GH\backslash GH^*} J_h(c_g(1-\alpha)-), \inf_{(g,h)\in GH^*} J_h(c_g(1-\alpha))\right\} \tag{9.5}$$

(where infimum over a null set is defined to be ∞). Clearly, $Min^{-}_{CS,Sub}(\alpha) \geq \inf_{(g,h)\in GH}$ $J_h(c_g(1-\alpha)-)$, where the latter is the lower bound on $AsyCS$ without the improvement. Define $Max^{-}_{CS,Sub}(\alpha)$ analogously to $Min^{-}_{CS,Sub}(\alpha)$ with min and inf replaced by max and sup.

Sufficient conditions for (g, h) to be in GH^* are that for all sequences $\{\gamma_{w_n,g,h} : n \geq 1\}$, (a) there exists a finite non-stochastic lower bound LB_h such that the subsample statistics are $\geq LB_h$ a.s. under $\{\gamma_{w_n,g,h} : n \geq 1\}$, (b) $J_h(LB_h) \geq J_h(c_g(1-\alpha))$, and (c) $\liminf_{n\to\infty} P_{\gamma_{w_n,g,h}}(T_{w_n}(\theta_{w_n,g,h}) \leq LB_h) \geq J_h(LB_h)$. (Conditions (a)-(c) imply (9.4) because $\liminf_{n\to\infty} P_{\gamma_{w_n,g,h}}(T_{w_n}(\theta_{w_n,g,h}) \leq c_{w_n,b_{w_n}}(\theta_{w_n,g,h}, 1-\alpha)) \geq \liminf_{n\to\infty} P_{\gamma_{w_n,g,h}}(T_{w_n}(\theta_{w_n,g,h}) \leq LB_h) \geq J_h(LB_h) \geq J_h(c_g(1-\alpha))$.)

The main results of this section are the following.

Theorem 3 (a) *Suppose Assumptions* A0 *and* B0 *hold. Then, an* FCV CS *satisfies*

$$AsyCS \in [\inf_{h \in H} J_h(c_{Fix}(1-\alpha)-), \inf_{h \in H} J_h(c_{Fix}(1-\alpha))] \textit{ and}$$
$$AsyMaxCP \in [\sup_{h \in H} J_h(c_{Fix}(1-\alpha)-), \sup_{h \in H} J_h(c_{Fix}(1-\alpha))].$$

(b) *Suppose Assumptions* A0, B0, C, D, E0, F, *and* G0 *hold. Then, a subsampling* CS *satisfies*

$$AsyCS \in [Min^{-}_{CS,Sub}(\alpha), \inf_{(g,h) \in GH} J_h(c_g(1-\alpha))] \textit{ and}$$
$$AsyMaxCP \in [Max^{-}_{CS,Sub}(\alpha), \sup_{(g,h) \in GH} J_h(c_g(1-\alpha))].$$

Comments. 1. Theorem 3(b) is used below to prove Theorem 1.

2. When the parameter space Γ takes on a partial product-space form as in Assumption A of AG1, then the forms of the localization parameter spaces H and GH can be made more explicit and the results of Theorem 3 hold under a somewhat simpler assumption than Assumption B0. See the Appendix for details.

APPENDIX

In the Appendix, we first show how the moment inequality model fits into the general framework for CS's introduced in Section 9. Next, we prove the main results stated in the paper for the moment inequality model and, in particular, we use the general result Theorem 3 to prove Theorem 1. Then, we provide results for GEL test statistics. Finally, we prove Theorem 3.

10 Moment Inequality Model

10.1 Specification of Parameters

In this section we specify a one-to-one mapping between the parameters (θ, F) in the moment inequality model and the parameter $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ that appears in the general results of Section 9. We define $\gamma_1 = (\gamma_{1,1}, ..., \gamma_{1,p})' \in R_+^p$ by writing the moment inequalities in (3.1) as moment equalities:

$$\sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0 \text{ for } j = 1, ..., p, \tag{10.1}$$

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 (iii) and (iv) of (10.2) 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 Ω that lie below the main diagonal, $q = d + k(k-1)/2$, and $\gamma_3 = F$. For the case described in Section 8.2 (where the sample moment functions depend on a preliminary estimator $\widehat{\tau}_n(\theta)$ of an identified parameter vector τ_0), we have $m_j(W_i, \theta) = m_j(W_i, \theta, \tau_0)$ 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 Γ for γ in the moment inequality example is defined by $\Gamma = \{\gamma = (\gamma_1, \gamma_2, \gamma_3) :$ for some $(\theta, F) \in \mathcal{F}$, where $\mathcal{F}$ is defined in (3.3), γ_1 satisfies (10.1), $\gamma_2 = (\theta, vech_*(\Omega(\theta, F)))$, 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 Γ for the moment inequality model using a set of high-level conditions. To verify the high-level conditions using primitive conditions one has to specify an estimator $\widehat{\Sigma}_n(\theta)$ of the asymptotic variance matrix $\Sigma(\theta)$ of $n^{1/2}\overline{m}_n(\theta)$. For brevity, we do not do so here. Since there is a one-to-one mapping from γ to (θ, F), Γ also defines the parameter space $\mathcal{F}$ of (θ, F). Let Ψ be a specified set of $k \times k$ correlation matrices. Let $\{\overline{\alpha}(j) : j \geq 1\}$ be a sequence of non-negative numbers that satisfies $\overline{\alpha}(j) \to 0$ as $j \to \infty$. The parameter space Γ is defined to include parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3) = (\gamma_1, (\theta, \gamma_{2,2}), F)$ that satisfy:

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

(ii) $E_F m_j(W_i,\theta) = 0$ for $j = p+1,...,k$,

(iii) $\sigma^2_{F,j}(\theta) = \lim_{n\to\infty} Var_F(n^{1/2}\overline{m}_{n,j}(\theta))$ exists and lies in $(0,\infty)$ for $j = 1,...,k$,

(iv) $\lim_{n\to\infty} Corr_F(n^{1/2}\overline{m}_n(\theta))$ exists and equals $\Omega_{\gamma_{2,2}} \in \Psi$,

(v) $\{W_i : i \geq 1\}$ are stationary & strong mixing under F with strong mixing numbers $\alpha_F(j) \leq \overline{\alpha}(j)$ for all $j \geq 1$, (10.2)

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

(vi) $A_n = (A_{n,1},...,A_{n,k})' \to_d Z_{h_{2,2}} \sim N(0_k, \Omega_{h_{2,2}})$ as $n \to \infty$, where $A_{n,j} = n^{1/2}(\overline{m}_{n,j}(\theta_{n,h}) - E_{F_{n,h}}\overline{m}_{n,j}(\theta_{n,h}))/\sigma_{F_{n,h},j}(\theta_{n,h})$,

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

(viii) $\widehat{D}_n^{-1/2}(\theta_{n,h})\widehat{\Sigma}_n(\theta_{n,h})\widehat{D}_n^{-1/2}(\theta_{n,h}) \to_p \Omega_{h_{2,2}}$ as $n \to \infty$, and (10.3)

(ix) conditions (vi)-(viii) hold for all subsequences $\{w_n\}$ in place of $\{n\}$,

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

For example, for i.i.d. observations, conditions (i)-(v) of (3.3) imply conditions (i)-(v) of (10.2). Furthermore, conditions (i)-(v) of (3.3) plus the definition of $\widehat{\Sigma}_n(\theta)$ in (3.8) and the additional condition (vi) of (3.3) imply conditions (vi)-(ix) of (10.3).

Lemma 2 *The parameter space* Γ *for i.i.d. observations* (*defined in* (3.3)) *is such that conditions* (i)-(ix) *of* (10.2)-(10.3) *hold when* $\widehat{\Sigma}_n(\theta)$ *is defined by* (3.8).

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)-(v) in (10.2) that ensure that Γ is such that any sequences $\{\gamma_{n,h} : n \geq 1\}$ in Γ satisfy (10.3). For brevity, we do not do so here. Note that the strong mixing assumption in condition (v) of (10.2) is used to verify Assumption E0.

10.2 Proofs for the Moment Inequality Model

Proof of Theorem 1. We prove Theorem 1 for the moment inequality/equality model by showing (i) Assumptions A0, B0, C, D, E0, F, and G0 hold and hence Theorem 3 applies, (ii) $Min^-_{CS,Sub}(\alpha) = \inf_{(g,h)\in GH}\ J_h(c_g(1-\alpha)) \geq 1-\alpha$, (iii) $\inf_{(g,h)\in GH}$ $J_h(c_g(1-\alpha)) \leq 1-\alpha$, and (iv) $\sup_{(g,h)\in GH} J_h(c_g(1-\alpha)) = 1$ when $v = 0$.

[18] Condition (ix) of (10.3) requires that conditions (vi)-(viii) must hold under any sequence of parameters $\{\gamma_{w_n,h} : n \geq 1\}$ that satisfies the conditions preceding (10.3) with n replaced by w_n.

Assumption A0 holds with Γ defined as in Section 10.1. Assumption B0 is verified with $r = 1/2$ as follows. Using Assumption 1(b), we have

$$T_n(\theta) = S\Big(\widehat{D}_n^{-1/2}(\theta)n^{1/2}\overline{m}_n(\theta),\ \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta)\widehat{D}_n^{-1/2}(\theta)\Big)\,. \tag{10.4}$$

For i.i.d. or dependent observations, (10.3) holds (using Lemma 2 for i.i.d. observations). By (10.3), the jth element of $\widehat{D}_n^{-1/2}(\theta_{n,h})n^{1/2}\overline{m}_n(\theta_{n,h})$ equals $(1+o_p(1))(A_{n,j} + n^{1/2}\gamma_{n,h,1,j})$, where $\gamma_{n,h,1} = (\gamma_{n,h,1,1},\ ...,\gamma_{n,h,1,p})'$ and by definition $\gamma_{n,h,1,j} = 0$ for $j = p+1, ..., k$. Condition (vi) of (10.3) and the definition of $\{\gamma_{n,h} : n \geq 1\}$ imply that if $h_{1,j} = \infty$ and $j \leq p$, where $h_1 = (h_{1,1}, ..., h_{1,p})'$, then $A_{n,j} + n^{1/2}\gamma_{n,h,1,j} \rightarrow_p \infty$ under $\{\gamma_{n,h} : n \geq 1\}$. In consequence, if any element of h_1 equals ∞, $\widehat{D}_n^{-1/2}(\theta_{n,h})n^{1/2}\overline{m}_n(\theta_{n,h})$ does not converge in distribution (to a proper finite random vector) and the continuous mapping theorem cannot be applied to obtain the asymptotic distribution of the right-hand side of (10.4).

To circumvent these problems, we consider a k-vector-valued function of $\widehat{D}_n^{-1/2}(\theta_{n,h})n^{1/2}\overline{m}_n(\theta_{n,h})$ that converges in distribution whether or not some elements of h_1 equal ∞. Then, we write the right-hand side of (10.4) as a continuous function of this k-vector and apply the continuous mapping theorem. Let $G(\cdot)$ be a strictly increasing continuous df on R, such as the standard normal df. For $j \leq k$, we have

$$G_{n,j} = G\left(\widehat{\sigma}_{n,j}^{-1}(\theta_{n,h})n^{1/2}\overline{m}_{n,j}(\theta_{n,h})\right) = G\left(\widehat{\sigma}_{n,j}^{-1}(\theta_{n,h})\sigma_{F_{n,h},j}(\theta_{n,h})\left[A_{n,j} + n^{1/2}\gamma_{n,h,1,j}\right]\right). \tag{10.5}$$

Let $Z_{h_{2,2}} = (Z_{h_{2,2},1}, ..., Z_{h_{2,2},k})' \sim N(0_k, \Omega_{h_{2,2}})$. Define $h_{1,j} = 0$ for $j = p+1, ..., k$. If $j \leq p$ and $h_{1,j} < \infty$ or if $j = p+1, ..., k$, then

$$G_{n,j} \rightarrow_d G\left(Z_{h_{2,2},j} + h_{1,j}\right) \tag{10.6}$$

by (10.5), conditions (vi) and (vii) of (10.3), and the continuous mapping theorem. If $j \leq p$ and $h_{1,j} = \infty$, then

$$G_{n,j} = G\left(\widehat{\sigma}_{n,j}^{-1}(\theta_{n,h})n^{1/2}\overline{m}_{n,j}(\theta_{n,h})\right) \rightarrow_p 1 \tag{10.7}$$

by (10.5), $A_{n,j} = O_p(1)$, and $G(x) \rightarrow 1$ as $x \rightarrow \infty$. The results in (10.6)-(10.7) hold jointly and combine to give

$$G_n = (G_{n,1}, ..., G_{n,k})' \rightarrow_d (G(Z_{h_{2,2},1} + h_{1,1}), ..., G(Z_{h_{2,2},k} + h_{1,k}))' = G_\infty, \tag{10.8}$$

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

Let G^{-1} denote the inverse of G. For $x = (x_1, ..., x_k)' \in R^p_{[+\infty]} \times R^v$, let $G_{(k)}(x) = (G(x_1), ..., G(x_k))' \in (0, 1]^p \times (0, 1)^v$. For $y = (y_1, ..., y_k)' \in (0, 1]^p \times (0, 1)^v$, let $G^{-1}_{(k)}(y) = (G^{-1}(y_1), ..., G^{-1}(y_k))' \in R^p_{[+\infty]} \times R^v$. Define S^* as

$$S^*(y, \Omega) = S(G^{-1}_{(k)}(y), \Omega) \tag{10.9}$$

for $y \in (0,1]^p \times (0,1)^v$ and $\Omega \in \Psi$. Assumption 1(d) implies that $S^*(y,\Omega)$ is continuous at all (y,Ω) for $y \in (0,1]^p \times (0,1)^v$ and $\Omega \in \Psi$. We now have

$$\begin{aligned} T_n(\theta_{n,h}) &= S\Big(G_{(k)}^{-1}(G_n),\ \widehat{D}_n^{-1/2}(\theta_{n,h})\widehat{\Sigma}_n(\theta_{n,h})\widehat{D}_n^{-1/2}(\theta_{n,h})\Big) \\ &= S^*\Big(G_n,\ \widehat{D}_n^{-1/2}(\theta_{n,h})\widehat{\Sigma}_n(\theta_{n,h})\widehat{D}_n^{-1/2}(\theta_{n,h})\Big) \\ &\to_d S^*(G_\infty, \Omega_{h_{2,2}}) \\ &= S(G_{(k)}^{-1}(G_\infty), \Omega_{h_{2,2}}) \\ &= S(Z_{h_{2,2}} + (h_1, 0_v), \Omega_{h_{2,2}}), \\ &\sim J_h, \end{aligned} \tag{10.10}$$

where the first equality holds by (10.4) and the definition of $G_{(k)}^{-1}(G_n)$, the second and third equalities hold by the definition of S^*, the convergence holds by (10.8), condition (viii) of (10.3), and the continuous mapping theorem, the last equality holds by the definitions of $G_{(k)}^{-1}$ and G_∞ and the definition that if $h_{1,j} = \infty$, then the corresponding element of $Z_{h_{2,2}} + (h_1, 0_v)$ equals ∞, and the last line gives the definition of J_h (where $h = (h_1, h_2)$, $h_2 = (h_{2,1}, h_{2,2})$, $h_{2,1} \in R^d$ is arbitrary because it does not appear in $S(Z_{h_{2,2}} + (h_1, 0_v), \Omega_{h_{2,2}})$, and $h_{2,2} = vech_*(\Omega_{h_{2,2}})$). By the same argument but using condition (ix) of (10.3) in place of conditions (vi)-(viii), the result of (10.10) holds with $\{w_n\}$ in place of $\{n\}$ for any subsequence $\{w_n\}$. Hence, Assumption B0 holds with J_h defined as in (10.10).

Assumption C is assumed in Section 5. Assumption D holds by stationarity and the standard definition of subsample statistics in the i.i.d. and dependent cases. Assumption E0 holds for i.i.d. and stationary strong mixing observations by the remarks at the end of Section 9.5 using condition (v) of (10.2).

Next, we verify Assumption F. When $v = 0$ and $h_1 = \infty^p$, the limit random variable in (10.10) is $S(Z_{h_{2,2}} + \infty^p, \Omega_{h_{2,2}}) = S(\infty^p, \Omega_{h_{2,2}}) = 0$ using Assumption 3. In consequence, $J_h(x) = 1$ for all $x \geq 0$, $c_h(1-\alpha) = 0$, and Assumption F holds for $\alpha > 0$. Now, suppose $v \geq 1$ or $h_1 \neq \infty^p$. Then, by Assumption 2(b), $J_h(x)$ is strictly increasing for $x > 0$. Using this, we have (i) if $c_h(1-\alpha) > 0$, then $J_h(x)$ is strictly increasing at $x = c_h(1-\alpha)$ and Assumption F holds, (ii) if $c_h(1-\alpha) = 0$, then $J_h(0) \geq 1-\alpha$ (by the definition of $c_h(1-\alpha)$), (iii) if $c_h(1-\alpha) = 0$ and $J_h(0) \geq 1-\alpha$, then $J_h(x) > 1-\alpha$ for all $x > 0$ and Assumption F holds (otherwise, $J_h(x) = 1-\alpha$ for some $x > 0$ and $J_h(x/2) = 1-\alpha$ since J_h is non-decreasing, which contradicts the fact that $J_h(x)$ is strictly increasing for $x > 0$). Hence, Assumption F holds. Assumption G0 holds automatically because the subsampling procedure satisfies Assumption Sub2.

Given that Assumptions A0, B0, C, D, E0, F, and G0 hold, the result of Theorem 3(b) holds, i.e., $AsyCS \in [Min_{CS,Sub}^-(\alpha), \inf_{(g,h)\in GH} J_h(c_g(1-\alpha))]$.

We now prove Theorem 1(a) by showing that $Min_{CS,Sub}^-(\alpha) \geq 1-\alpha$. First, by Assumption 1(a), for $0 \leq g_1 \leq h_1 \in R_{+,\infty}^p$ and all $(g,h) \in GH$, we have

$$S(Z_{h_{2,2}} + (g_1, 0_v), \Omega_{h_{2,2}}) \geq S(Z_{h_{2,2}} + (h_1, 0_v), \Omega_{h_{2,2}}),$$

$$
\begin{aligned}
c_g(1-\alpha) &\geq c_h(1-\alpha), \text{ and} \\
J_h(c_g(1-\alpha)) &\geq J_h(c_h(1-\alpha)). \qquad (10.11)
\end{aligned}
$$

Next, we show that $GH^* = GH$. Given $(g,h) \in GH$, suppose $c_g(1-\alpha) > 0$. Then, $J_h(c_g(1-\alpha)-) = J_h(c_g(1-\alpha))$ because $J_h(x)$ is continuous for all $x > 0$ by Assumption 2(a). This and Lemma 6(f) of AG1 (which holds under Assumptions A0, B0, C, D, E0, F, and G0 by the proof of Theorem 3 in Section 11) establish (9.4). Hence, $(g,h) \in GH^*$.

Now, suppose $c_g(1-\alpha) = 0$. (Assumption 1(c) rules out $c_g(1-\alpha) < 0$.) This implies that $c_h(1-\alpha) = 0$ by (10.11). 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 show $(g,h) \in GH^*$ by verifying conditions (a)-(c) in the paragraph preceding Theorem 3. Condition (a) holds with $LB_h = 0$ by Assumption 1(c). Condition (b) holds because $LB_h = c_g(1-\alpha) = 0$. Next we show condition (c). Under $\{\gamma_{n,h} : n \geq 1\}$ and with $v = 0$, we have

$$
\begin{aligned}
& P_{\gamma_{n,h}}(T_n(\theta_{n,h}) \leq 0) \\
= \ & P_{\gamma_{n,h}}(n^{1/2}\overline{m}_{n,j}(\theta_{n,h})/\sigma_{F_{n,h},j}(\theta_{n,h}) \geq \zeta \text{ for all } j = 1, ..., p) \\
= \ & P_{\gamma_{n,h}}(A_{n,j} + n^{1/2}\gamma_{n,h,1,j}) \geq \zeta \text{ for all } j = 1, ..., p) \\
\to \ & P(Z_{h_{2,2},j} + h_{1,j} \geq \zeta \text{ for all } j = 1, ..., p) \\
= \ & P(S(Z_{h_{2,2}} + h_1, \Omega_{h_{2,2}}) \leq 0) \\
= \ & J_h(0), \qquad (10.12)
\end{aligned}
$$

where the first and third equalities hold by (10.4) and Assumption 3, the second equality and the convergence hold by (10.3), and the last equality holds by the definition of J_h given in (10.10) and $v = 0$. The same argument holds with $\{\gamma_{w_n,g,h} : n \geq 1\}$ in place of $\{\gamma_{n,h} : n \geq 1\}$. Hence, (10.12) completes the verification of condition (c) and concludes the proof that $GH^* = GH$.

For subsampling CS's, we now have

$$
AsyCS = \inf_{(g,h) \in GH} J_h(c_g(1-\alpha)) \geq \inf_{h \in H} J_h(c_h(1-\alpha)) \geq 1-\alpha, \qquad (10.13)
$$

where the first equality holds by Theorem 3 and $GH^* = GH$, the first inequality holds by (10.11), and the second inequality holds by the definition of $c_h(1-\alpha)$. This establishes Theorem 1(a). (Note that $AsyMaxCP$ is given by the second expression in (10.13) with "sup" in place of "inf."

Next, let (θ^*, F^*) be an element of $\mathcal{F}$ for which Assumption C1 applies and let γ^* be the value in Γ that corresponds to $(\theta^*, F^*) \in \mathcal{F}$. Define $h^* = (h_1^*, h_2^*)$, where $h_1^* = h_1(\theta^*, F^*)$, $h_{2,1}^* = \theta^* \in R^d$, and $h_{2,2}^* = vech_*(\Omega(\theta^*, F^*))$. We have $(h^*, h^*) \in GH$ because the sequence $\{\gamma_{w_n,g,h} : n \geq 1\}$ defined by $\gamma_{w_n,g,h} = \gamma^*$ for all $n \geq 1$ leads to the point $(h^*, h^*) \in GH$ by the definition of GH given in (9.3). By Assumption C1, $J_{h^*}(c_{h^*}(1-\alpha)) = 1-\alpha$. In consequence, we have

$$
AsyCS = \inf_{(g,h) \in GH} J_h(c_g(1-\alpha)) \leq J_{h^*}(c_{h^*}(1-\alpha)) = 1-\alpha. \qquad (10.14)
$$

Combining (10.13) and (10.14) completes the proof of Theorem 1(b).

Now, we prove Theorem 1(c). By assumption, $v = 0$. Assumption C2 guarantees the existence of $(\theta^*, F^*) \in \mathcal{F}$ for which $E_{F^*} m_j(W_i, \theta^*)/\sigma_{F^*,j}(\theta^*) > 0$ for $j = 1, ..., p$. The sequence of constant true values $\{(\theta^*, F^*) \in \mathcal{F} : n \geq 1\}$ satisfies $n^{1/2} E_{F^*} m_j(W_i, \theta^*)/\sigma_{F^*,j}(\theta^*) \to \infty$ and $b^{1/2} E_{F^*} m_j(W_i, \theta^*)/\sigma_{F^*,j}(\theta_n) \to \infty$ for all $j = 1, ..., p$. Let $\gamma^* = (\gamma_1^*, (\theta^*, \gamma_{2,2}^*), F^*) \in \Gamma$ correspond to $(\theta^*, F^*) \in \mathcal{F}$. Define $g^* = h^* = (\infty^p, (\theta^*, \gamma_{2,2}^*))$. Then, $(g^*, h^*) \in GH$ and $h_1^* = \infty^p$. We have $J_{h^*}(x) = 1$ for $x \geq 0$ because $S(Z_{h_{2,2}^*} + \infty^p, \Omega_{h_{2,2}^*}) = S(\infty^p, \Omega_{h_{2,2}^*}) = 0$ using Assumption 3 and $c_{g^*}(1-\alpha) \geq 0$ by Assumption 1(c). Hence, $J_{h^*}(c_{g^*}(1-\alpha)) = 1$. Given the result above that $GH^* = GH$, we have $Max^-_{CS,Sub}(\alpha) = \sup_{(g,h) \in GH} J_h(c_g(1-\alpha)) \geq J_{h^*}(c_{g^*}(1-\alpha)) = 1$. $\square$

Proof of Theorem 2. We prove Theorem 2 using the confidence set analogue of Theorem 2 of AG2 discussed in Section 8 of AG2. (Note that the PA CS considered in this paper is an example of the PSC-FCV CS considered in AG2.) Theorem 2 of AG2 can be extended to hold with Assumptions A0, B0, E0, and G0 in place of Assumptions A, B, E, and G using the same arguments as in the proof of Theorem 3 below. The definitions of H and GH are then given by (9.2) and (9.3) of this paper. For the case of PSC-FCV tests, which are relevant here, only Assumptions A, B, L(i), N, and OF are needed for Theorem 2 of AG2. Hence, we only need to verify Assumptions A0, B0, L(i), N, and OF here and the set GH is not relevant.

In the present case, the quantity $cv_{h_2}(1-\alpha)$ in (5.1) of AG2 satisfies

$$cv_{h_2}(1-\alpha) = \sup_{h_1 \in H_1} c_{(h_1,h_2)}(1-\alpha) \leq c_{(0_p,h_2)}(1-\alpha), \tag{10.15}$$

where $H_1 = \{h_1 \in R_\infty^p : h = (h_1, h_2) \in H \text{ for some } h_2 \in R_\infty^q\}$, the equality is by definition, and the inequality holds by (10.11) because (10.11) holds for all $g = (0_p, h_2)$, $h = (h_1, h_2)$, and $h_1 \in R_{+,\infty}^p$ whether or not $(g, h) \in GH$ (which is not necessarily the case here). (Note that the inequality in (10.15) is not necessarily an equality because 0_p is not necessarily in H_1.) Hence, the critical value $c(\widehat{\Omega}_n(\theta), 1-\alpha)$ in (6.2) is greater than or equal to the critical value $cv_{\widehat{\gamma}_{n,2}}(1-\alpha)$ of AG2 and Theorem 2 of AG2 yields $AsyCS \geq 1-\alpha$. Under Assumption C3, the inequality in (10.15) holds as an equality because $(0_p, h_2) \in H$ for some h_2 (by a similar argument to that preceding (10.14)) and so $0_p \in H_1$. In this case, $c(\widehat{\Omega}_n(\theta), 1-\alpha)$ equals the critical value $cv_{\widehat{\gamma}_{n,2}}(1-\alpha)$ of AG2 and Theorem 2 of AG2 yields $AsyCS = 1-\alpha$.

Now, it suffices to show that Assumptions L(i), N, and OF of AG2 hold because Assumptions A0 and B0 hold by the proof of Theorem 1 above. Assumption L(i) holds, i.e., $\sup_{\Omega \in \Psi} c(\Omega, 1-\alpha) < \infty$ because $c(\Omega, 1-\alpha)$ is a uniformly continuous function (by Assumption 4(b)) on the subset Ψ of the compact set Ψ_1 of all $k \times k$ correlation matrices and hence is bounded on Ψ. For dependent observations, Assumption N holds by condition (viii) of (10.3) (which holds for i.i.d. observations by Lemma 2) and the definition of $\{\gamma_{n,h} : n \geq 1\}$—which implies that $\gamma_{n,h,2} \to h_2$. Assumptions OF(i) and

OF(iii) hold by Assumptions 4(b) and 4(a), respectively. Assumption OF(ii) holds with $h_1^* = 0_p$. $\square$

Proof of Lemma 1. For S_1, Assumptions 1 and 3 hold immediately with $\zeta = 0$ in Assumption 3. Assumption 2(a) holds because (i) if $v \geq 1$, the summand $\sum_{j=p+1}^{p+v} Z_j^2$ is absolutely continuous, where $Z = (Z_1, ..., Z_k)'$, (ii) if $v = 0$ and $h_1 \neq \infty^p$, the summands $[Z_j + h_{1,j}]_-^2$ are absolutely continuous for $x > 0$ for all $j = 1, ..., p$ such that $h_{1,j} < \infty$, and (iii) if $v = 0$ and $h_1 = \infty^p$, $S_1(Z + h_1, \Omega) = 0$ and its df equals 1 for all $x > 0$. Assumption 2(b) holds because (i) if $v \geq 1$, the summand $\sum_{j=p+1}^{p+v} Z_j^2$ has positive density on R_+, each summand $[Z_j + h_{1,j}]_-^2$ for which $h_{1,j} < \infty$ (of which there may be none) has positive density on R_+, and so does the sum and (ii) if $v = 0$ and $h_1 \neq \infty^p$, each summand $[Z_j + h_{1,j}]_-^2$ for which $h_{1,j} < \infty$ (of which there is at least one) has positive density on R_+ and the sum does as well. Assumption 2(c) holds because if $v \geq 1$, $P(S_1(Z + (h_1, 0_v), \Omega) \leq 0) \leq P(\sum_{j=p+1}^{p+v} Z_j^2 \leq 0) = 0$, and if $h_1 = 0$ and $v = 0$, $P(S_1(Z, \Omega) \leq 0) \leq P([Z_j]_-^2 \leq 0) = P(Z_j \geq 0) = 1/2$ where the inequality holds for any $j \leq p$. Assumption 4(a) holds by the same argument as for Assumption 2(a). Assumption 4(b) holds because $c(\Omega, 1 - \alpha)$ is continuous at each $\Omega \in \Psi$ and $\Psi = \Psi_1$ is compact. To see the former, let $\{\Omega_N : N \geq 1\}$ be a sequence of correlation matrices such that $\Omega_N \to \Omega$ as $N \to \infty$. We need to show that $c(\Omega_N, 1-\alpha) \to c(\Omega, 1-\alpha)$. Denote by f_N and f the df's of $S_1(Z_N, \Omega_N)$ and $S_1(Z, \Omega)$, respectively, where $Z_N \sim N(0_k, \Omega_N)$ and $Z \sim N(0_k, \Omega)$. By Assumption 2(b), f is increasing for $x > 0$ (because $h_1 = 0_p$, not ∞^p, in this case). By Assumption 2(c) we have $c(\Omega, 1 - \alpha) > 0$ and it follows that f is increasing at $c(\Omega, 1 - \alpha)$. This implies that $c(\Omega_N, 1 - \alpha) \to c(\Omega, 1 - \alpha)$ because by $S_1(Z, \Omega) = \sum_{j=1}^{p} [Z_j]_-^2 + \sum_{j=p+1}^{p+v} Z_j^2$ we have $\sup_{x \in R} |f_N(x) - f(x)| \to 0$.

For S_2, Assumptions 1(b)-(c) and 3 hold immediately with $\zeta = 0$ in Assumption 3. Assumption 1(d) holds straightforwardly using the specification of Ψ_2, which bounds the determinant of the correlation matrix Ω away from zero. Assumption 1(a) holds because for $x \in R^p$ with $x \geq 0$, we have

$$\begin{aligned} S_2((m_1 + x, m_2), \Sigma) &= \inf_{t_1 \in R_{+,\infty}^p} \begin{pmatrix} m_1 + x - t_1 \\ m_2 \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} m_1 + x - t_1 \\ m_2 \end{pmatrix} \\ &= \inf_{t_1 \in R_{+,\infty}^p - x} \begin{pmatrix} m_1 - t_1 \\ m_2 \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} m_1 - t_1 \\ m_2 \end{pmatrix} \\ &\leq \inf_{t_1 \in R_{+,\infty}^p} \begin{pmatrix} m_1 - t_1 \\ m_2 \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} m_1 - t_1 \\ m_2 \end{pmatrix} \\ &= S_2((m_1, m_2, \Sigma). \end{aligned} \quad (10.16)$$

To show Assumption 2(c), first suppose $h_1 = 0_p$, then

$$S_2(Z, \Omega) = \inf_{t=(t_1, 0_v): t_1 \in R_{+,\infty}^p} (Z - t)' \Omega^{-1} (Z - t) \quad (10.17)$$

$$= \inf_{t=(t_1, 0_v): t_1 \in R_{+,\infty}^p} (Z^* - Bt)'(Z^* - Bt) = \inf_{t_1 \in R_{+,\infty}^p} (Z^* - B_1 t_1)'(Z^* - B_1 t_1),$$

where $Z \sim N(0_k, \Omega)$, $B'B = \Omega^{-1}$, $Z^* = BZ \sim N(0_k, I_k)$, $B = [B_1 : B_2]$, and B_1 is $k \times p$ and full rank $p \leq k$. The right-hand side of (10.17) is zero only if $Z^* = B_1 t_1$ for some $t_1 \in R^p_{+,\infty}$. The latter holds with probability zero if $k > p$ and with probability $\leq 1/2$ if $k = p$, which verifies Assumption 2(c) for $h_1 = 0_p$. Next, suppose $v \geq 1$, without loss of generality (wlog) we can assume $||h_1|| < \infty$ (because if some element of h_1 equals infinity then the infimum in $S_2(Z, \Omega)$ is obtained by taking the corresponding element of t_1 equal to infinity). Then, using (10.17), we have

$$S_2(Z + (h_1, 0_v), \Omega) = \inf_{t_1 \in R^p_{+,\infty}} (Z^* - B_1 t_1)'(Z^* - B_1 t_1), \tag{10.18}$$

where $Z^* = B(Z + (h_1, 0_v)) \sim N(B_1 h_1, I_k)$ and B and B_1 are as above. As above, $S_2(Z, \Omega) = 0$ is zero only if $Z^* = B_1 t_1$ for some $t_1 \in R^p_{+,\infty}$. The support of Z^* is R^k, whereas $\{B_1 t_1 : t_1 \geq 0_p\}$ is a subset of a p-dimensional linear subspace of R^k. Since $v = k - p > 0$, $S_2(Z, \Omega) = 0$ with probability zero.

Next, we show that Assumptions 2(a) and 2(b) hold for S_2. If $v = 0$ and $h_1 = \infty^p$, then $S_2(Z + h_1, \Omega) = 0$, $J_h(x) = 1$ for all $x > 0$, Assumption 2(a) holds, and Assumption 2(b) does not impose any restriction. Otherwise, $v \geq 1$ or $h_1 \neq \infty^p$. As above, wlog we can assume $||h_1|| < \infty$ (because "$v \geq 1$ or $h_1 \neq \infty^p$" imply that at least one element of Z remains after setting to zero all those elements $Z_j + h_{1,j} - t_{1,j}$ for which $h_{1,j} = \infty$). Equation (10.18) holds in the present case (whether or not $v \geq 1$). The random variable $S_2(Z + (h_1, 0_v), \Omega)$ in (10.18) has support R_+ and is absolutely continuous. Hence, Assumptions 2(a) and 2(b) hold. Assumption 4(a) holds by the same argument as for Assumption 2(a).

To show Assumption 4(b) for S_2, first we show continuity of $c(\Omega, 1-\alpha)$ at a fixed $\Omega \in \Psi_2$. Let $\{\Omega_N : N \geq 1\}$ be a sequence of correlation matrices not necessarily in Ψ_2 such that $\Omega_N \to \Omega$ as $N \to \infty$. We need to show that $c(\Omega_N, 1-\alpha) \to c(\Omega, 1-\alpha)$. Denote by f_N and f the df's of $S_2(Z_N, \Omega_N)$ and $S_2(Z, \Omega)$, respectively, where $Z_N \sim N(0_k, \Omega_N)$ and $Z \sim N(0_k, \Omega)$. By Assumption 2(b), f is increasing for $x > 0$ (because $h_1 = 0_p$, not ∞^p, in this case). By Assumption 2(c) we have $c(\Omega, 1-\alpha) > 0$ and it follows that f is increasing at $c(\Omega, 1-\alpha)$. This implies that $c(\Omega_N, 1-\alpha) \to c(\Omega, 1-\alpha)$ because by (10.18) with $h_1 = 0_p$ we have $\sup_{x \in R} |f_N(x) - f(x)| \to 0$.

Next, choose $\delta > 0$ small enough that for the compact set $\Psi_2^* = \{\Omega \in \Psi : \det(\Omega) \geq \varepsilon/2\}$ (where $\varepsilon > 0$ is as in the definition of Ψ_2) it holds that for every $\Omega_2 \in \Psi_2$ we have $\{\Omega \in \Psi : ||\Omega_2 - \Omega|| < \delta\} \subset \Psi_2^*$. By the argument in the preceding paragraph, $c(\Omega, 1-\alpha)$ is continuous on Ψ_2^* as a function on Ψ_2^* and thus is uniformly continuous on Ψ_2^*. This implies that $c(\Omega, 1-\alpha)$ is uniformly continuous on Ψ_2 as a function on $c(\Omega, 1-\alpha)$.

The proof for S_3 is essentially the same as that for S_1. $\square$

Proof of Lemma 2. Condition (i) of (10.2) holds by the definition of $\gamma_{1,j}$ in (10.1) for $j = 1, ..., p$ using condition (i) of (3.3). Conditions (ii), (iii), (iv), and (v) of (10.2) hold by conditions (ii), (iii) & (iv), (iii) & (v), and (iii) of (3.3), respectively.

Condition (vi) of (10.3) holds by the combination of the Cramér-Wold device and the Liapounov triangular array CLT for row-wise i.i.d. random variables with mean zero and variance one by conditions (iii) and (vi) of (3.3). Conditions (vii) and (viii) of (10.3) hold by standard arguments using a weak law of large numbers for row-wise i.i.d. random variables with variance one by conditions (iii) and (vi) of (3.3). Condition (ix) of (10.3) holds by the same argument as for conditions (vi)-(viii) of (10.3). □

10.3 Results for GEL Statistics

Here we prove that Theorems 1 and 2 hold for CS's based on the GEL statistic $T_n^{GEL}(\theta)$ rather than $T_n(\theta)$, provided Assumption GEL below holds, which requires that the observations are i.i.d. for each fixed $(\theta, F) \in \mathcal{F}$. It suffices to show that for any sequence $\{\gamma_{w_n,h} : n \geq 1\}$ for $h = (h_1, h_2) \in R^p_{+,\infty} \times R^q_\infty$ and corresponding $\{(\theta_{w_n,h}, F_{w_n,h}) \in \mathcal{F} : n \geq 1\}$, we have

$$T_{w_n}^{GEL}(\theta_{w_n,h}) - T_{w_n}(\theta_{w_n,h}) = o_p(1). \tag{10.19}$$

The result in (10.19) implies that $T_{w_n}^{GEL}(\theta_{w_n,h})$ satisfies Assumption B0. The remainder of the proofs are the same as the proofs of Theorems 1 and 2.

We use the following notation. Let

$$\begin{aligned} t_{n,h} &= E_{F_{n,h}} m(W_i, \theta_{n,h}), \\ \widehat{t}_n &= \arg\min_{t \in R^p_+} \sup_{\lambda \in \widehat{\Lambda}_n(t,\theta_{n,h})} n\widehat{P}_\rho(t, \theta_{n,h}, \lambda), \text{ if it exists,} \\ \widehat{m}_n(t) &= n^{-1} \textstyle\sum_{i=1}^n m_i(t, \theta_{n,h}), \\ \widehat{\Delta}(t) &= n^{-1} \textstyle\sum_{i=1}^n m_i(t, \theta_{n,h}) m_i(t, \theta_{n,h})', \text{ and} \\ \Lambda_n &= \{\lambda \in R^k : ||\lambda|| \leq n^{-1/(2+\delta/2)}\} \end{aligned} \tag{10.20}$$

for $\delta > 0$ as in condition (vi) of (3.3). Let $\rho_j(x) = (\partial^j \rho / \partial x^j)(x)$ for $j = 1, 2$. Let "w.a.p.1" abbreviate "with probability that approaches one as $n \to \infty$."

We make the following assumption.

Assumption GEL. (a) The observations are i.i.d. for each fixed $(\theta, F) \in \mathcal{F}$. (b) Part (iv) of (3.3) is strengthened to $Var_F(m_j(W_i, \theta)) \in [\varepsilon_*, M_*]$ for some $\varepsilon_* > 0$ and $M_* < \infty$ for $j = 1, ..., k$. (c) The parameter space Ψ in (3.3) equals Ψ_2. (d) For any sequence $\{\gamma_{w_n,h} : n \geq 1\}$ and corresponding $\{(\theta_{w_n,h}, F_{w_n,h}) \in \mathcal{F} : n \geq 1\}$, $\widehat{t}_{w_n} = \arg\min_{t \in R^p_+} \sup_{\lambda \in \widehat{\Lambda}_{w_n}(t,\theta_{w_n,h})} w_n \widehat{P}_\rho(t, \theta_{w_n,h}, \lambda)$ and $t^*_{w_n} = \arg\min_{t \in R^p_+} w_n(\overline{m}_{w_n}(\theta_{w_n,h}) - (t, 0_v))' \widehat{\Sigma}^{-1}_{w_n}(\theta_{w_n,h})(\overline{m}_{w_n}(\theta_{w_n,h}) - (t, 0_v))$ exist and satisfy $\sup_{n \geq 1} ||\widehat{t}_{w_n}|| \leq K$ and $\sup_{n \geq 1}$ $||t^*_{w_n}|| \leq K$ w.p.a.1. for some constant $K < \infty$, where $\widehat{\Sigma}_{w_n}(\theta)$ is defined in (3.8).

The proof of (10.19) uses a similar approach to that in Newey and Smith (2004). The proof uses the following four Lemmas.

Lemma 3 *Suppose Assumption GEL holds. For any sequence* $\{\gamma_{w_n,h} : n \geq 1\}$ *and corresponding* $\{(\theta_{w_n,h}, F_{w_n,h}) \in \mathcal{F} : n \geq 1\}$, *there exist constants* $K < \infty$ *and* $\varepsilon > 0$ *such that* (a) $\lambda_{\min}(\widehat{\Delta}(t_{w_n,h})) \geq \varepsilon$ *w.p.a.1 and* $w_n^{-1} \sum_{i=1}^{w_n} ||m_i(t_{w_n,h}, \theta_{w_n,h}) m_i(t_{w_n,h}, \theta_{w_n,h})'|| = O_p(1)$, (b) *for every random sequence* $\{\widetilde{t}_{w_n} \in R_+^p : n \geq 1\}$ *with* $\sup_{n \geq 1} ||\widetilde{t}_{w_n}|| \leq K$ *w.p.a.1,* $\max_{1 \leq i \leq w_n} ||m_i(\widetilde{t}_{w_n}, \theta_{w_n,h})|| = O_p(w_n^{1/(2+\delta)})$ *and* $\lambda_{\max}(\widehat{\Delta}(\widetilde{t}_{w_n})) \leq K$ *w.p.a.1,* (c) $\widehat{m}_{w_n}(t_{w_n,h}) = O_p(w_n^{-1/2})$, *and* (d) $w_n^{-1} \sum_{i=1}^{w_n} ||m(W_i, \theta_{w_n,h})|| = O_p(1)$ *and* $||t_{w_n,h}|| \leq K$.

The next three Lemmas are analogues of Lemmas 7-9 of Guggenberger and Smith (2005). They all hold under a given sequence $\{\gamma_{n,h} : n \geq 1\}$ and corresponding $\{(\theta_{n,h}, F_{n,h}) \in \mathcal{F} : n \geq 1\}$ and Assumption GEL.

Lemma 4 *Assume that for a (possibly random) sequence* $\{t_n \in R_+^p : n \geq 1\}$ *we have* $\max_{1 \leq i \leq n} ||m_i(t_n, \theta_{n,h})|| = O_p(n^{1/(2+\delta)})$. *Then,* $\sup_{\lambda \in \Lambda_n, 1 \leq i \leq n} |\lambda' m_i(t_n, \theta_{n,h})| \to_p 0$ *and* $\Lambda_n \subset \widehat{\Lambda}_n(t_n, \theta_{n,h})$ *w.p.a.1.*

Lemma 5 *Assume that for a (possibly random) sequence* $\{t_n \in R_+^p : n \geq 1\}$ *we have* $\max_{1 \leq i \leq n} ||m_i(t_n, \theta_{n,h})|| = O_p(n^{1/(2+\delta)})$, $\lambda_{\min}(\widehat{\Delta}(t_n)) \geq \varepsilon$ *w.p.a.1 for an* $\varepsilon > 0$, *and* $\widehat{m}_n(t_n) = O_p(n^{-1/2})$. *Then,* $\lambda(t_n, \theta_{n,h}) \in \widehat{\Lambda}_n(t_n, \theta_{n,h})$ *satisfying* $\widehat{P}_\rho(t_n, \theta_{n,h}, \lambda(t_n, \theta_{n,h})) = \sup_{\lambda \in \widehat{\Lambda}_n(t_n, \theta_{n,h})} \widehat{P}_\rho(t_n, \theta_{n,h}, \lambda)$ *exists w.p.a.1,* $\lambda(t_n, \theta_{n,h}) = O_p(n^{-1/2})$, *and* $\sup_{\lambda \in \widehat{\Lambda}_n(t_n, \theta_{n,h})} \widehat{P}_\rho(t_n, \theta_{n,h}, \lambda) = O_p(n^{-1})$.

Lemma 6 *Suppose* $\widehat{t}_n$ *(defined in* (10.20)*) exists w.p.a.1,* $\max_{1 \leq i \leq n} ||m_i(\widehat{t}_n, \theta_{n,h})|| = O_p(n^{1/(2+\delta)})$, $\lambda_{\max}(\widehat{\Delta}(\widehat{t}_n)) \leq K$ *w.p.a.1 for some* $K < \infty$, *and* $\sup_{\lambda \in \widehat{\Lambda}_n(t_{n,h}, \theta_{n,h})} \widehat{P}(t_{n,h}, \theta_{h,h}, \lambda) = O_p(n^{-1})$. *Then,* $\widehat{m}_n(\widehat{t}_n) = O_p(n^{-1/2})$.

Comment. Lemmas 4-6 hold for any subsequence $\{w_n : n \geq 1\}$ in place of $\{n\}$.

Proof of (10.19). For notational simplicity we use n instead of w_n in the proof. We use the abbreviations $\widehat{P}_\rho(t, \lambda) = \widehat{P}_\rho(t, \theta_{n,h}, \lambda)$ and $m_i(t) = m_i(t, \theta_{n,h})$. Using Lemma 3, Lemma 5 with $t_n = t_{n,h}$ yields $\sup_{\lambda \in \widehat{\Lambda}_n(t_{n,h}, \theta_{n,h})} \widehat{P}_\rho(t_{n,h}, \lambda) = O_p(n^{-1})$. Using this result, Assumption GEL(d), and Lemma 3(b), Lemma 6 gives $\widehat{m}_n(\widehat{t}_n) = O_p(n^{-1/2})$. Therefore,

$$\begin{aligned} O_p(n^{-1/2}) = \widehat{m}_n(\widehat{t}_n) &= n^{-1} \textstyle\sum_{i=1}^n m(W_i, \theta_{n,h}) - (t_{n,h}, 0_v) + (t_{n,h}, 0_v) - (\widehat{t}_n, 0_v) \\ &= o_p(1) + (t_{n,h}, 0_v) - (\widehat{t}_n, 0_v) \end{aligned} \tag{10.21}$$

and hence $\widehat{t}_n - t_{n,h} \to_p 0_p$. By Lemma 3(a) we have $\lambda_{\min}(\widehat{\Delta}(t_{w_n,h})) \geq \varepsilon$ w.p.a.1 and it was just shown that $\widehat{t}_n - t_{n,h} \to_p 0_p$. These two statements imply $\lambda_{\min}(\widehat{\Delta}(\widehat{t}_n)) \geq \varepsilon$ w.p.a.1 using the technicalities in Lemma 3(d) and simply multiplying out. By Lemma 3(b), the second part of Lemma 3(d), and $\widehat{t}_n - t_{n,h} \to_p 0_p$, we have $\max_{1 \leq i \leq n} ||m_i(\widehat{t}_n, \theta_{n,h})||$

$= O_p(n^{1/(2+\delta)})$. Using this, we apply Lemma 5 with $t_n = \widehat{t}_n$ and conclude that $\lambda_n \equiv \lambda(\widehat{t}_n, \theta_{n,h}) \in \widehat{\Lambda}_n(\widehat{t}_n, \theta_{n,h})$ satisfying $\widehat{P}_\rho(\widehat{t}_n, \lambda_n) = \sup_{\lambda \in \widehat{\Lambda}_n(\widehat{t}_n, \theta_{n,h})} \widehat{P}_\rho(\widehat{t}_n, \lambda)$ exists w.p.a.1 and $\lambda_n = O_p(n^{-1/2})$. Therefore, the first-order conditions

$$n^{-1} \textstyle\sum_{i=1}^n \rho_1(\lambda_n' m_i(\widehat{t}_n)) m_i(\widehat{t}_n) = 0_k \tag{10.22}$$

hold w.p.a.1. Expanding the first-order conditions in λ around 0_k, there exists a mean value $\widetilde{\lambda}_n$ between 0_k and λ_n (that may be different for each row) such that w.p.a.1

$$\begin{aligned} 0_k &= -\widehat{m}_n(\widehat{t}_n) + [n^{-1} \textstyle\sum_{i=1}^n \rho_2(\widetilde{\lambda}_n' m_i(\widehat{t}_n)) m_i(\widehat{t}_n) m_i(\widehat{t}_n)'] \lambda_n \\ &= -\widehat{m}(\widehat{t}_n) - \widehat{\Delta}_n \lambda_n, \end{aligned} \tag{10.23}$$

where the matrix $\widehat{\Delta}_n$ has been defined implicitly. Because $\lambda_n = O_p(n^{-1/2})$, $\max_{1 \le i \le n} ||m_i(\widehat{t}_n, \theta_{n,h})|| = O_p(n^{1/(2+\delta)})$, and $\rho_2(0) = -1$, we have $\max_{1 \le i \le n} |\rho_2(\widetilde{\lambda}_n' m_i(\widehat{t}_n)) + 1| \to_p 0$. Thus, by Lemma 3(a), $\widehat{\Delta}_n - \widehat{\Delta}(\widehat{t}_n) \to_p 0_{k \times k}$. In addition, by the argument above, $\lambda_{\min}(\widehat{\Delta}(\widehat{t}_n)) \ge \varepsilon$ w.p.a.1. In consequence, $\widehat{\Delta}_n$ is invertible w.p.a.1 and

$$\lambda_n = -\widehat{\Delta}_n^{-1} \widehat{m}_n(\widehat{t}_n) \tag{10.24}$$

w.p.a.1. Inserting this into a second-order Taylor expansion for $\widehat{P}_\rho(\widehat{t}_n, \lambda)$ with mean value λ_n^*, it follows that w.p.a.1

$$\begin{aligned} \widehat{P}_\rho(\widehat{t}_n, \lambda_n) &= -2\lambda_n' \widehat{m}_n(\widehat{t}_n) + \lambda_n' [n^{-1} \textstyle\sum_{i=1}^n \rho_2(\lambda_n^{*\prime} m_i(\widehat{t}_n)) m_i(\widehat{t}_n) m_i(\widehat{t}_n)'] \lambda_n \\ &= 2\widehat{m}_n(\widehat{t}_n)' \widehat{\Delta}_n^{-1} \widehat{m}_n(\widehat{t}_n) - \widehat{m}_n(\widehat{t}_n)' \widehat{\Delta}_n^{-1} \widetilde{\Delta}_n \widehat{\Delta}_n^{-1} \widehat{m}_n(\widehat{t}_n), \end{aligned} \tag{10.25}$$

where $\widetilde{\Delta}_n \equiv n^{-1} \sum_{i=1}^n \rho_2(\lambda_n^{*\prime} m_i(\widehat{t}_n)) m_i(\widehat{t}_n) m_i(\widehat{t}_n)'$ satisfies $\widetilde{\Delta}_n - \widehat{\Delta}_n \to_p 0$ by the same argument as used above to show $\widehat{\Delta}_n - \widehat{\Delta}(\widehat{t}_n) \to_p 0_{k \times k}$. Therefore, up to $o_p(1)$ terms, we have

$$\begin{aligned} T_n^{GEL}(\theta_{n,h}) &= n\widehat{m}_n(\widehat{t}_n)' \widehat{\Delta}(\widehat{t}_n)^{-1} \widehat{m}_n(\widehat{t}_n) \\ &= n(\overline{m}_n(\theta_{n,h}) - (\widehat{t}_n, 0_v))' \widehat{\Delta}(\widehat{t}_n)^{-1} (\overline{m}_n(\theta_{n,h}) - (\widehat{t}_n, 0_v)) \\ &= n(\overline{m}_n(\theta_{n,h}) - (\widehat{t}_n, 0_v))' \widehat{\Sigma}_n(\theta_{n,h})^{-1} (\overline{m}_n(\theta_{n,h}) - (\widehat{t}_n, 0_v)) \\ &= \min_{t \in R_{+,\infty}^p} n(\overline{m}_n(\theta_{n,h}) - (t, 0_v)) \widehat{\Sigma}_n(\theta_{n,h})^{-1} (\overline{m}_n(\theta_{n,h}) - (t, 0_v)) \\ &= S_2(n^{1/2} \overline{m}_n(\theta_{n,h}), \widehat{\Sigma}_n(\theta_{n,h})), \end{aligned} \tag{10.26}$$

where the third equality holds because $\widehat{\Sigma}_n(\theta_{n,h}) = n^{-1} \sum_{i=1}^n (m(W_i, \theta_{n,h}) - E\overline{m}_n(\theta_{n,h}))(m(W_i, \theta_{n,h}) - E\overline{m}_n(\theta_{n,h}))' + o_p(1)$ and $t_{n,h} - \widehat{t}_n \to_p 0_p$. The second to last equality holds by the following argument. Denote by t_n^* the minimizing $t \in R_{+,\infty}^p$ in the second to last line of (10.26). We have to show that

$$\begin{aligned} &n(\overline{m}_n(\theta_{n,h}) - (\widehat{t}_n, 0_v))' \widehat{\Sigma}_n(\theta_{n,h})^{-1} (\overline{m}_n(\theta_{n,h}) - (\widehat{t}_n, 0_v)) \\ &-n(\overline{m}_n(\theta_{n,h}) - (t_n^*, 0_v)) \widehat{\Sigma}_n(\theta_{n,h})^{-1} (\overline{m}_n(\theta_{n,h}) - (t_n^*, 0_v)) \end{aligned} \tag{10.27}$$

is $o_p(1)$. If this does not hold, then it could not be the case that $\widehat{m}_n(t_n^*) = O_p(n^{-1/2})$ because if the latter were true, then the argument in (10.21)-(10.26) could be applied to t_n^* instead of $\widehat{t}_n$ yielding $T_n^{GEL}(\theta_{n,h}) = n(\overline{m}_n(\theta_{n,h}) - (t_n^*, 0_v))'\widehat{\Sigma}_n(\theta_{n,h})^{-1}(\overline{m}_n(\theta_{n,h}) - (t_n^*, 0_v))$ w.p.a.1, which is a contradiction. Therefore, $\widehat{m}_n(t_n^*)$ is not $O_p(n^{-1/2})$. But then $T_n(\theta_{n,h}) = \min_{t \in R^p_{+,\infty}} n(\overline{m}_n(\theta_{n,h}) - (t, 0_v))\ \widehat{\Sigma}_n(\theta_{n,h})^{-1}(\overline{m}_n(\theta_{n,h}) - (t, 0_v))$ cannot be $O_p(1)$ because $\lambda_{\min}(\widehat{\Sigma}_n(\theta_{n,h})^{-1}) \geq \varepsilon$ w.p.a.1 by the second part of Lemma 3(b). Therefore, we get a contradiction to (10.10) (where the latter shows that $T_n(\theta_{n,h}) \to_d J_h$ and thus $T_n(\theta_{n,h}) = O_p(1)$). Therefore, the expression in (10.27) is indeed $o_p(1)$. $\square$

Proof of Lemma 3. The first part of Lemma 3(a) holds because (i) $\widehat{\Delta}(t_{w_n,h}) - E_{F_{w_n}}\widehat{\Delta}(t_{w_n,h}) = o_p(1)$ by a weak LLN for row-wise i.i.d. random variables given Assumptions GEL(a) and (b) and condition (vi) of (3.3) and (ii) $\lambda_{\min}(E_{F_{w_n}}\widehat{\Delta}(t_{w_n,h})) \geq \varepsilon$ for some $\varepsilon > 0$ for all n by Assumptions GEL(b) and (c). The second part of Lemma 3(a) and the first part of Lemma 3(d) hold by a weak LLNs, Assumptions GEL(a) and (b), and condition (vi) of (3.3). The first part of Lemma 3(b) holds using Assumptions GEL(a) and (b) and condition (vi) of (3.3), e.g., see Guggenberger and Smith (2005, eqn. (2.4)). The second part of Lemma 3(b) holds by a weak LLN, Assumptions GEL(a) and (b), and condition (vi) of (3.3). Lemma 3(c) holds by a Liapounov CLT for a row-wise i.i.d. triangular array of random variables applied to $\{m_i(t_{w_n,h}, \theta_{w_n,h}) : i = 1, ..., n; n \geq 1\}$ using Assumptions GEL(a) and (b), condition (vi) of (3.3), and the fact that $m_i(t_{w_n,h}, \theta_{w_n,h})$ has mean zero given the definition of $t_{n,h}$. The second part of Lemma 3(d) holds by Assumption GEL(b) and condition (vi) of (3.3). $\square$

Proof of Lemma 4. The result of the Lemma follows from $\sup_{\lambda \in \Lambda_n, 1 \leq i \leq n} |\lambda' m_i(t_n, \theta_{n,h})| \leq O_p(n^{-1/(2+\delta/2)} n^{1/(2+\delta)}) = o_p(1)$. $\square$

Proof of Lemma 5. We use the abbreviations $\widehat{P}_\rho(\lambda) = \widehat{P}_\rho(t_n, \theta_{n,h}, \lambda)$ and $m_i = m_i(t_n, \theta_{n,h})$ in this proof. Let $\lambda_n \in \Lambda_n$ be such that $\widehat{P}_\rho(\lambda_n) = \max_{\lambda \in \Lambda_n} \widehat{P}_\rho(\lambda)$. Such a $\lambda_n \in \Lambda_n$ exists w.p.a.1 because a continuous function takes on its maximum on a compact set and by Lemma 4, $\widehat{P}_\rho(\cdot)$ is twice continuously differentiable, i.e., C^2, on some open neighborhood of Λ_n w.p.a.1. We now show that actually $\widehat{P}_\rho(\lambda_n) = \sup_{\lambda \in \widehat{\Lambda}_n(t_n, \theta_{n,h})} \widehat{P}_\rho(\lambda)$ w.p.a.1, which then proves the first part of the Lemma. By a second-order Taylor expansion around $\lambda = 0$, there is a λ_n^* on the line segment joining 0_k and λ_n such that for some positive constants C_1 and C_2, we have

$$\begin{aligned} 0 &= \widehat{P}_\rho(0) \leq \widehat{P}_\rho(\lambda_n) \\ &= -2\lambda_n'\widehat{m}_n(t_n) + \lambda_n'[n^{-1}\textstyle\sum_{i=1}^n \rho_2(\lambda_n^{*\prime} m_i) m_i m_i{}'] \lambda_n \\ &\leq -2\lambda_n'\widehat{m}_n(t_n) - C_1 \lambda_n' \widehat{\Delta}(t_n)\lambda_n \leq 2||\lambda_n||\ ||\widehat{m}_n(t_n)|| - C_2||\lambda_n||^2 \end{aligned} \tag{10.28}$$

w.p.a.1, where the first inequality follows because $\max_{1 \leq i \leq n} \rho_2(\lambda_n^{*\prime} m_i) < -1/2$ w.p.a.1 by Lemma 4, continuity of $\rho_2(\cdot)$ at zero, and $\rho_2(0) = -1$. The last inequality follows from $\lambda_{\min}(\widehat{\Delta}(t_n)) \geq \varepsilon > 0$ w.p.a.1. Now, (10.28) implies that $(C_2/2)||\lambda_n|| \leq ||\widehat{m}_n(t_n)||$ w.p.a.1, the latter being $O_p(n^{-1/2})$. Therefore, $\lambda_n \in int(\Lambda_n)$ w.p.a.1. Hence, the first-order conditions for an interior maximum $\partial \widehat{P}_\rho(\lambda)/\partial\lambda = 0$ hold at $\lambda = \lambda_n$ w.p.a.1. By Lemma 4, $\lambda_n \in \widehat{\Lambda}_n(t_n, \theta_{n,h})$ w.p.a.1 and thus by concavity of $\widehat{P}_\rho(\lambda)$ and convexity of $\widehat{\Lambda}_n(t_n, \theta_{n,h})$ it follows that $\widehat{P}_\rho(\lambda_n) = \sup_{\lambda \in \widehat{\Lambda}_n(t_n,\theta_{n,h})} \widehat{P}_\rho(\lambda)$ w.p.a.1, which implies the first part of the lemma. From above, $\lambda(t_n, \theta_{n,h}) = \lambda_n = O_p(n^{-1/2})$. Thus, the second part of the Lemma holds. This, $||\widehat{m}_n(t_n)|| = O_p(n^{-1/2})$, and (10.28) give the third part of the Lemma. □

Proof of Lemma 6. We use the abbreviations $\widehat{P}_\rho(t, \lambda) = \widehat{P}_\rho(t, \theta_{n,h}, \lambda)$ and $m_i(t) = m_i(t, \theta_{n,h})$ in this proof. Without loss of generality, $\widehat{m}_n(\widehat{t}_n) \neq 0$ can be assumed. Define $\underline{\lambda}_n = -n^{-1/2}\widehat{m}_n(\widehat{t}_n)/||\widehat{m}_n(\widehat{t}_n)||$. Note that $\underline{\lambda}_n \in \Lambda_n$ and thus $\underline{\lambda}_n \in \widehat{\Lambda}_n(\widehat{t}_n, \theta_{n,h})$ w.p.a.1 by Lemma 4 (applied with $t_n = \widehat{t}_n$). By a second-order Taylor expansion around $\lambda = 0$, there is a $\widetilde{\lambda}_n$ on the line segment joining 0_k and $\underline{\lambda}_n$, such that for some positive constants C_1 and C_2, we have

$$\begin{aligned}
\widehat{P}_\rho(\widehat{t}_n, \underline{\lambda}_n) &= -2\underline{\lambda}_n'\widehat{m}_n(\widehat{t}_n) + \underline{\lambda}_n'[n^{-1}\textstyle\sum_{i=1}^n \rho_2(\widetilde{\lambda}_n' m_i(\widehat{t}_n))m_i(\widehat{t}_n)m_i(\widehat{t}_n)']\underline{\lambda}_n \\
&\geq 2n^{-1/2}||\widehat{m}_n(\widehat{t}_n)|| - C_1\underline{\lambda}_n'\widehat{\Delta}(\widehat{t}_n)\underline{\lambda}_n \\
&\geq 2n^{-1/2}||\widehat{m}_n(\widehat{t}_n)|| - C_2 n^{-1} \qquad (10.29)
\end{aligned}$$

w.p.a.1, where the first inequality follows from $\min_{1\leq i\leq n} \rho_2(\widetilde{\lambda}_n' m_i(\widehat{t}_n)) \geq -1.5$ w.p.a.1, which is implied by Lemma 4. The second inequality follows by $\lambda_{\max}(\widehat{\Delta}(\widehat{t}_n)) \leq K < \infty$ w.p.a.1. The definition of $\widehat{t}_n$ implies

$$\widehat{P}_\rho(\widehat{t}_n, \underline{\lambda}_n) \leq \sup_{\lambda \in \widehat{\Lambda}_n(\widehat{t}_n,\theta_{n,h})} \widehat{P}(\widehat{t}_n, \lambda) \leq \sup_{\lambda \in \widehat{\Lambda}_n(t_{n,h},\theta_{n,h})} \widehat{P}(t_{n,h}, \lambda) = O_p(n^{-1}). \qquad (10.30)$$

Combining equations (10.29) and (10.30) implies $n^{-1/2}||\widehat{m}_n(\widehat{t}_n)|| = O_p(n^{-1})$. □

11 General Results

This section is concerned with the general results of Section 9. First, we state a Corollary to Theorem 3 that applies when the parameter space Γ takes on a partial product-space form, as in Assumption A of AG1. In this case, the form of H and GH can be made more explicit and the results of Theorem 3 hold under an assumption that eliminates the subsequences that appear in Assumption B0. Second, we prove Theorem 3.

Let $\lfloor$ denote the left endpoint of an interval that may be open or closed at the left end. Define $\rfloor$ analogously for the right endpoint. The following assumption implies Assumption A0. Let $R_- = \{x \in R : x \leq 0\}$ and $R_{-,\infty} = R_- \cup \{-\infty\}$.

Assumption A. (i) For some $\Gamma_1 \subset R^p$, $\Gamma_2 \subset R^q$, and $\Gamma_3(\gamma_1, \gamma_2) \subset \mathcal{T}_3$, which may depend on γ_1 and γ_2, Γ satisfies

$$\Gamma = \{(\gamma_1, \gamma_2, \gamma_3) : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3(\gamma_1, \gamma_2)\}. \qquad (11.31)$$

and (ii) $\Gamma_1 = \prod_{m=1}^{p} \Gamma_{1,m}$, where $\Gamma_{1,m} = \lfloor \gamma_{1,m}^{\ell}, \gamma_{1,m}^{u} \rfloor$ for some $-\infty \leq \gamma_{1,m}^{\ell} < \gamma_{1,m}^{u} \leq \infty$ that satisfy $\gamma_{1,m}^{\ell} \leq 0 \leq \gamma_{1,m}^{u}$ for $m = 1, ..., p$.

Under Assumption A, it follows that

$$H = H_1 \times H_2,\ H_1 = \prod_{m=1}^{p} \begin{cases} R_{+,\infty} & \text{if } \gamma_{1,m}^{\ell} = 0 \\ R_{-,\infty} & \text{if } \gamma_{1,m}^{u} = 0 \\ R_{\infty} & \text{if } \gamma_{1,m}^{\ell} < 0 \text{ and } \gamma_{1,m}^{u} > 0, \end{cases} \qquad H_2 = \mathrm{cl}(\Gamma_2), \tag{11.32}$$

where $\mathrm{cl}(\Gamma_2)$ is the closure of Γ_2 with respect to R_{∞}^{q}. For example, if $p = 1$, $\gamma_{1,1}^{\ell} = 0$, and $\Gamma_2 = R^q$, then $H_1 = R_{+,\infty}$, $H_2 = R_{\infty}^{q}$, and $H = R_{+,\infty} \times R_{\infty}^{q}$.

Under Assumption A, the set GH reduces to

$$\begin{aligned} GH = \{&(g,h) \in H \times H : g = (g_1, g_2),\ h = (h_1, h_2),\ g_2 = h_2,\ \text{and for} \\ &m = 1, ..., p,\ \text{(i) } g_{1,m} = 0 \text{ if } |h_{1,m}| < \infty,\ \text{(ii) } g_{1,m} \in R_{+,\infty} \text{ if } h_{1,m} \\ &= +\infty,\ \text{and (iii) } g_{1,m} \in R_{-,\infty} \text{ if } h_{1,m} = -\infty\}, \end{aligned} \tag{11.33}$$

where $g_1 = (g_{1,1}, ..., g_{1,p})' \in H_1$ and $h_1 = (h_{1,1}, ..., h_{1,p})' \in H_1$. Note that for $(g,h) \in GH$, we have $|g_{1,m}| \leq |h_{1,m}|$ for all $m = 1, ..., p$.

Given Assumption A, the following weakened version of Assumption B0 is sufficient.

Assumption B′. For some $r > 0$, all $h \in H$, all sequences $\{\gamma_{n,h} : n \geq 1\}$, and some distributions J_h, $T_n(\theta_{n,h}) \to_d J_h$ under $\{\gamma_{n,h} : n \geq 1\}$, where $\gamma_{n,h} = ((\theta_{n,h,1}, \eta_{n,h,1}), (\theta_{n,h,2}, \eta_{n,h,2}), \gamma_{n,h,3})$ and $\theta_{n,h} = (\theta_{n,h,1}, \theta_{n,h,2})$.

Assumption B′ is the same as Assumption B of AG1 except that it applies to CS's rather than tests, so that $T_n(\theta)$ is evaluated at $\theta_{n,h}$ rather than at the null value θ_0 and $\gamma_{n,h}$ is defined differently.

We have the following Corollary to Theorem 3.

Corollary 1 *Theorem* 3 *holds with H and GH defined in* (11.32) *and* (11.33), *respectively, with Assumptions* A0 *and* B0 *replaced by Assumptions* A *and* B′.

Comments. Assumption B′ is simpler and weaker than Assumption B0. But typically the work needed to verify these assumptions and the strength of the assumptions is almost the same. Hence, the main advantage of Corollary 1 is that when Assumption A holds one has the explicit forms for H and GH given in (11.32) and (11.33).

2. Corollary 1 is proved using the proof of Theorem 3 coupled with the argument given in (30)-(31) of the proof of Lemma 6 of AG1.

Proof of Theorem 3. The proof of the results of Theorem 3 for $AsyCS$ is analogous to that of Theorem 1 of AG1 with the following changes: $AsySz(\theta_0)$ is replaced by $1 - AsyCS$, probabilities $P_{\theta,\gamma}(\cdot)$ and expectations $E_{\theta,\gamma}(\cdot)$ are replaced by $P_{\gamma}(\cdot)$ and

$E_\gamma(\cdot)$, respectively, because θ is a subvector of γ, the test statistic $T_n(\theta_0)$ is replaced by $T_n(\theta_n)$ throughout, where θ_n denotes the true value of θ which may depend on n, and one makes use of the fact that $\inf_{h\in H} J_h(c_{Fix}(1-\alpha)-) = 1 - Max^-_{Fix}(\alpha)$, $\inf_{(g,h)\in GH} J_h(c_g(1-\alpha)) = 1 - Max_{Sub}(\alpha)$, etc., where $Max^-_{Fix}(\alpha)$ and $Max_{Sub}(\alpha)$ are defined in AG1. The proof of the results for $AsyMaxCP$ is quite similar to those for $AsyCS$ and hence is not discussed.

The replacement of Assumptions A, B, E, and G of AG1 by Assumptions A0, B0, E0, and G0 requires the following changes in the proof of Theorem 1 of AG1. First, we show that the results of Lemma 6(a)-(f) of AG1 hold with $\{\gamma_{w_n} : n \geq 1\}$ equal to $\{\gamma_{w_n,g,h} : n \geq 1\}$ (defined in this paper) under the assumptions of Theorem 3 of this paper. Lemma 6(a) of AG1 (i.e., $(g,h) \in GH$) holds by the definition of GH in this paper. The proof of Lemma 6(b) is the same as in AG1 (noting that $\{\gamma_{w_n} : n \geq 1\}$ in Lemma 6 is of the form $\{\gamma_{w_n,g,h} : n \geq 1\}$ considered in this paper) but with (30) holding by Assumption B0 rather than by the proof given in AG1. The proof of Lemma 6(c) is much simpler than that in AG1. By Assumption E0, $U_{w_n,b_{w_n}}(x) - E_{\gamma_{w_n,g,h}} U_{w_n,b_{w_n}}(x) \to_p 0$ under $\{\gamma_{w_n,g,h} : n \geq 1\}$, so (31) of AG1 is not needed. This result and the result of Lemma 6(b) yield Lemma 6(c). Similarly, the proof of Lemma 6(d) is much simpler than that in AG1. The result of Lemma 6(d) holds immediately by Assumption G0 and the result of Lemma 6(c). The proof of Lemma 6(e)-(f) of AG1 is the same as in AG1 but with the result of (II) stated in the proof holding by Assumption B0 rather than by the proof given in AG1.

Second, in the proof of Theorem 1 of AG1, (34) holds by the definition of GH (which guarantees that for each $(g,h) \in GH$ there is a sequence $\{\gamma_{w_n,g,h} : n \geq 1\}$) and the result of Lemma 6(f) of AG1. The remainder of the proof holds using the modified version of Lemma 6 (which holds under Assumptions A0, B0, C, D, E0, F, and G0) with the only change being that $h_2 \in cl(\Gamma_2)$ in (38) is replaced by $h_2 \in R^q_\infty$. This concludes the adjustment of the proof of Theorem 1 of AG1 to take account of the change in assumptions.

The improvement to the lower bound on $AsyCS$ for subsampling CS's is obtained as follows. If the assumption is added to Lemma 5 of AG1 that $\liminf_{n\to\infty} P(T_n \leq c_n) \geq G_T(c_\infty)$, then the Lemma yields the stronger conclusion that $P(T_n \leq c_n) \to G_T(c_\infty)$. This follows directly from the proof of Lemma 5(b) of AG1. Therefore, for any $(g,h) \in GH^*$ and any sequence $\{\gamma_{w_n,g,h} : n \geq 1\}$, the proof of Lemma 6(f) of AG1 yields the stronger conclusion that $P_{\gamma_{w_n,g,h}}(T_{w_n}(\theta_{w_n,g,h}) \leq c_{w_n,b_{w_n}}(\theta_{w_n,g,h}, 1-\alpha)) \to J_h(c_g(1-\alpha))$. Combining this with the proof of Theorem 1(b) of AG1 establishes the lower bound $Min^-_{CS,Sub}(\alpha)$ to $AsyCS$ given in the Theorem. $\square$

References

Andrews, D. W. K. (2000) Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space. *Econometrica* 68, 399–405.

Andrews, D. W. K. (2005) Inconsistency of Bootstrapping Interval Endpoints Defined by Moment Inequalities. Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K., S. Berry, & P. Jia (2004) Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location. Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K., S. Berry, & P. Jia (2007) Confidence Regions for Parameters in Discrete Games with Multiple Equilibria. Manuscript in preparation, Cowles Foundation, Yale University.

Andrews, D. W. K. & 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.

Andrews, D. W. K. & P. Guggenberger (2005b) Hybrid and Size-corrected Subsampling Methods. Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K. & P. Guggenberger (2005c) Applications of Subsampling, Hybrid, and Size-Correction Methods. Cowles Foundation Discussion Paper No. 1608, Yale University.

Andrews, D. W. K. & P. Guggenberger (2005d) Invalidity of Subsampling Inference Based on Post-Consistent Model Selection Estimators. Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K. & G. Soares (2007) Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection. Manuscript in preparation, Cowles Foundation, Yale University.

Bajari, P., L. Benkard, & J. Levin (2008) Estimating Dynamic Models of Imperfect Competition. *Econometrica* 76, forthcoming.

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

Bugni, F. (2007) Bootstrap Inference in Partially Identified Models. 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. & I. Fernandez-Val (2005): Inference on Quantile Regression Process, An Alternative. *Sankhya.*

Chernozhukov, V., H. Hong, & E. Tamer (2008) Parameter Set Inference in a Class of Econometric Models. *Econometrica* 76, forthcoming.

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

Dufour, J.-M. (1997) Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica* 65, 1365-1387.

Galichon, A. & M. Henry (2006) Inference in Incomplete Models. Unpublished manuscript, Department of Economics, Columbia University.

Guggenberger, P., J. Hahn, & K. Kim (2007) Specification Testing Under Moment Inequalities. Unpublished manuscript, Department of Economics, UCLA.

Guggenberger, P. & R.J. Smith (2005) Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification. *Econometric Theory* 21, 667–709.

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

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

Kabaila, P. (1995) The Effect of Model Selection on Confidence Regions and Prediction Regions. *Econometric Theory* 11, 537-549.

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

Leeb, H. & B. M. Pötscher (2005) Model Selection and Inference: Facts and Fiction. *Econometric Theory* 21, 21-59.

Linton, O. E. Maasoumi, & Y.-J. Whang (2005). Consistent Testing for Stochastic Dominance under General Sampling Schemes. *Review of Economic Studies* 72, 735-765.

Mikusheva, A. (2008) Uniform Inferences in Autoregressive Models. *Econometrica* 76, forthcoming.

Moon, H. R. & F. Schorfheide (2004) Empirical Likelihood Estimation with Inequality Moment Conditions. Unpublished working paper, Department of Economics, University of Southern California.

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

Newey, W. K. & R. J. Smith (2004) Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* 72, 219-255.

Otsu, T. (2006) Large Deviation Optimal Inference for Set Identified Moment Inequality Models. Unpublished working paper, Cowles Foundation, Yale University.

Pakes, A., J. Porter, K. Ho, & J. Ishii (2004) Applications of Moment Inequalities. Unpublished working paper, Department of Economics, Harvard University.

Politis, D. N. & J. P. Romano (1994) Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions. *Annals of Statistics* 22, 2031-2050.

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

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

Romano, J. P. & A. M. Shaikh (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. & 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.

Smith, R. J. (1997) Alternative Semi–Parametric Likelihood Approaches to Generalized Method of Moments Estimation. *Economic Journal* 107, 503–519.

Soares, G. (2006) Inference for Partially Identified Models 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.