6-1-2010

Inference Based on Conditional Moment Inequalities

Donald W.K. Andrews

Xiaoxia Shi

INFERENCE BASED ON CONDITIONAL MOMENT INEQUALITIES

By

Donald W.K. Andrews and Xiaoxia Shi

June 2010

COWLES FOUNDATION DISCUSSION PAPER NO. 1761

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

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

Inference Based on Conditional Moment Inequalities

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

Xiaoxia Shi
Department of Economics
Yale University

November 2007
Revised: June 2010

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-0751517. The authors thank Andrés Aradillas-López, Kees Jan van Garderen, Hidehiko Ichimura, and Konrad Menzel for helpful comments.

Abstract

In this paper, we propose an instrumental variable approach to constructing confidence sets (CS's) for the true parameter in models defined by conditional moment inequalities/equalities. We show that by properly choosing instrument functions, one can transform conditional moment inequalities/equalities into unconditional ones without losing identification power. Based on the unconditional moment inequalities/equalities, we construct CS's by inverting Cramér-von Mises-type or Kolmogorov-Smirnov-type tests. Critical values are obtained using generalized moment selection (GMS) procedures.

We show that the proposed CS's have correct uniform asymptotic coverage probabilities. New methods are required to establish these results because an infinite-dimensional nuisance parameter affects the asymptotic distributions. We show that the tests considered are consistent against all fixed alternatives and have power against some $n^{-1/2}$-local alternatives, though not all such alternatives. Monte Carlo simulations for three different models show that the methods perform well in finite samples.

Keywords: Asymptotic size, asymptotic power, conditional moment inequalities, confidence set, Cramér-von Mises, generalized moment selection, Kolmogorov-Smirnov, moment inequalities.

JEL Classification Numbers: C12, C15.

1 Introduction

This paper considers inference for parameters whose true values are restricted by conditional moment inequalities and/or equalities. The parameters need not be identified. Much of the literature on partially-identified parameters concerns unconditional moment inequalities, see the references given below. However, in many moment inequality models, the inequalities that arise are conditional moments given a vector of covariates X_i. In this case, the construction of a fixed number of unconditional moments requires an arbitrary selection of some functions of X_i. In addition, the selection of such functions leads to information loss that can be substantial. Specifically, the "identified set" based on a chosen set of unconditional moments can be noticeably larger than the identified set based on the conditional moments.[1] With moment inequalities there is a first-order loss in information when moving from conditional to unconditional moments—one loses identifying information. In contrast, with moment equality models, there is only a second-order loss when moving from conditional to unconditional moments—one increases the variance of an estimator and decreases the noncentrality parameter of a test.

This paper provides methods to construct CS's for the true value of the parameter θ by converting conditional moment inequalities into an infinite number of unconditional moment inequalities. This is done using weighting functions $g(X_i)$. We show how to construct a class $\mathcal{G}$ of such functions such that there is no loss in information. We construct Cramér-von Mises-type (CvM) and Kolmogorov-Smirnov-type (KS) test statistics using a function S of the weighted sample moments, which depend on $g \in \mathcal{G}$. For example, the function S can be of the Sum, quasi-likelihood ratio (QLR), or Max form. The KS statistic is given by a supremum over $g \in \mathcal{G}$. The CvM statistic is given by an integral with respect to a probability measure Q on the space $\mathcal{G}$ of g functions. Computation of the CvM test statistics can be carried out by truncation of an infinite sum or simulation of an integral. Asymptotic results are established for both exact and truncated/simulated versions of the test statistic.

For reasons explained below, the choice of critical values is important for all moment inequality tests, and especially so for conditional moment inequalities. Here we consider critical values based on generalized moment selection (GMS), as in Andrews and Soares

[1] By "identified set," we mean the set of parameter values that are consistent with the population moment inequalities/equalities, either unconditional or conditional, given the true distribution of the data.

(2010). For comparative purposes, we also provide results for subsampling critical values and "plug-in asymptotic" (PA) critical values. However, for reasons of accuracy of size and magnitude of power, we recommend GMS critical values over both subsampling and plug-in asymptotic critical values. The GMS critical values can be implemented using an asymptotic Gaussian distribution or the bootstrap.

The main contribution of this paper is to establish the properties of the CS's described above. Our results apply to multiple moment inequalities and/or equalities and vector-valued parameters θ with minimal regularity conditions on the conditional moment functions and the distribution of X_i. For example, no smoothness conditions or even continuity conditions are made on the conditional moment functions as functions of X_i and no conditions are imposed on the distribution of X_i (beyond the boundedness of $2+\delta$ moments of the moment functions). In consequence, the range of moment inequality models for which the methods are applicable is very broad.

The results of the paper are summarized as follows. The paper (i) develops critical values that take account of the issue of moment inequality slackness that arises in finite samples and uniform asymptotics, (ii) proves that the confidence sizes of the CS's are correct asymptotically in a uniform sense, (iii) proves that the proposed CS's yield no information loss (i.e., that the coverage probability for any point outside the identified set converges to zero as $n \to \infty$), (iv) establishes asymptotic local power results for a certain class of $n^{-1/2}$-local alternatives, (v) extends the results to allow for the preliminary estimation of parameters that are identified given knowledge of the parameter of interest θ, as occurs in some game theory examples, and (vi) extends the results to allow for time series observations. A companion paper, Andrews and Shi (2010), generalizes the CS's and extends the asymptotic results to allow for an infinite number of conditional or unconditional moment inequalities, which makes the results applicable to tests of stochastic dominance and conditional stochastic dominance (see Lee and Whang (2009)).

The paper provides Monte Carlo simulation results for three models that exhibit different features. The first model is a quantile selection model. The parameter of interest is a nonparametric quantity, a conditional quantile. Selection effects yield the parameter to be unidentified. We introduce a quantile variant of Manski and Pepper's (2000) monotone instrumental variables condition that provides conditional moment inequalities that bound the conditional quantile.[2] The second model is an interval

[2]Papers in the literature that bound quantiles include Manski (1994), Lee and Melenberg (1998), and Blundell, Gosling, Ichimura, and Meghir (2007), among others. The condition employed here differs

outcome linear regression model with a bivariate parameter of interest. This model is considered in Manski and Tamer (2002).

The third model is a binary entry game model with multiple equilibria. It has ten parameters and a bivariate parameter of interest, which is the competitive effects vector. This model is related to models considered in Andrews, Berry, and Jia (2004), Beresteanu, Molchanov, and Molinari (2009), Galichon and Henry (2009b), and Ciliberto and Tamer (2009). In this model, the eight non-competitive effects parameters are estimated via a preliminary maximum likelihood estimator based on the number of entrants, similar to Berry (1992). These estimators are plugged into a set of moment conditions that includes two moment inequalities and two moment equalities. In this model the competitive effects are point identified and the moment inequalities are used to bring more information to bear. As far as we are aware, no other methods in the literature can handle a model of this sort.

The simulation results compare different forms of the test statistic: CvM versus KS, Sum versus QLR versus Max S function; and different forms of the critical value: GMS based on the asymptotic distribution (GMS/Asy), GMS based on the bootstrap (GMS/Bt), PA/Asy, PA/Bt, and subsampling (Sub). Coverage probabilities (CP's) for points in the identified set are computed and false coverage probabilities (FCP's) for points outside the identified set are computed.[3] In each model, we consider a basecase and variations on the basecase with different sample sizes, true data generating processes, and different GMS tuning parameters.

The CP's of all of the CS's in the first two models are quite good in the sense of being greater than or equal to .944 when the nominal level is .95 in all scenarios considered (except two with a Sub critical value). The CP's of CvM-based CS's with GMS critical values are quite close to .95 in one of two scenarios in the first model and in the second model and are around .98 in the other scenario in the first model. The latter over-coverage is not too surprising because non-similarity on the boundary in finite samples and asymptotically (in a uniform sense) is an inherent feature of good CS's in these contexts, as will be shown in future work. In the third model, CP's of the CS's vary across different true DGP's with CP's being greater than or equal to .95 in most cases except when the competitive effects parameters are large, in which case under-coverage of some CS's is as large as .037.

from the conditions in these papers, although it is closely related to them, see Section 9.

[3] The FCP's are "coverage probability corrected," see Section 9 for details.

Across all three models, the simulation results show that the CvM-based CS's outperform the KS-based CS's in terms of FCP's. The Sum, QLR, and Max versions of the test statistics perform equally well in the first two models, while the Max version performs best in the third model, in terms of FCP's. In all three models, the GMS critical values outperform the PA and Sub critical values in terms of FCP's. The Asy and Bt versions of the GMS critical values perform similarly in the first two models. (The Bt critical values are not computed in the third model because they are expensive computationally.)

Variations on the basecase show a relatively low degree of sensitivity of the CP's and FCP's in most cases.

In sum, in the three models considered, the CvM/Max statistic coupled with the GMS/Asy critical value performed quite well in an absolute sense and best among the CS's considered. In the first two models, use of the Sum or QLR S function or the GMS/Bt critical value produced equally good results.

The literature related to this paper includes numerous papers dealing with unconditional moment inequality models, such as Andrews, Berry, and Jia (2004), Imbens and Manski (2004), Moon and Schorfheide (2006, 2009), Otsu (2006), Pakes, Porter, Ho, and Ishii (2006), Woutersen (2006), Bontemps, Magnac, and Maurin (2007), Canay (2010), Chernozhukov, Hong, and Tamer (2007), Andrews and Jia (2008), Beresteanu, Molchanov, and Molinari (2008), Beresteanu and Molinari (2008), Chiburis (2008), Guggenberger, Hahn, and Kim (2008), Romano and Shaikh (2008, 2010), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Han (2009), Stoye (2009), Andrews and Soares (2010), Bugni (2010), and Canay (2010).

The literature on conditional moment inequalities is smaller and more recent. The present paper and the following papers have been written over more or less the same time period: Chernozhukov, Lee, and Rosen (2008), Fan (2008), Kim (2008), and Menzel (2008). An earlier paper than these by Galichon and Henry (2009a) considers a related testing problem with an infinite number unconditional moment inequalities of a particular type. The test statistic considered by Kim (2008) is the closest to that considered here. He considers subsampling critical values. The approach of Chernozhukov, Lee, and Rosen (2008) is different from that considered here. They consider tests based on nonparametric estimators such as kernels and sieves. Their results apply to scalar conditional lower and upper bounds on a parameter. Menzel's (2008) approach is different again. He investigates tests based on a finite number of moment inequalities in which the number of inequalities increases with the sample size. None of the other papers in the literature that treat conditional moment inequalities provide contributions (ii)-(vi) listed above.

Papers that convert conditional moments into an infinite number of unconditional moments in point identified models include Bierens (1982), Bierens and Ploberger (1997), Chen and Fan (1999), Dominguez and Lobato (2004), and Khan and Tamer (2009), among others.

In addition to reporting a CS, it often is useful to report an estimated set. A CS accompanied by an estimated set reveals how much of the volume of the CS is due to randomness and how much is due to a large identified set. It is well-known that typical set estimators suffer from an inward-bias problem, e.g., see Haile and Tamer (2003) and Chernozhukov, Lee, and Rosen (2008). The reason is that an estimated boundary often behaves like the minimum or maximum of multiple random variables.

A simple solution to the inward-bias problem is to exploit the method of constructing median-unbiased estimators from confidence bounds with confidence level 1/2, e.g., see Lehmann (1959, Sec. 3.5). The CS's in this paper applied with confidence level 1/2 are half-median-unbiased estimated sets. That is, the probability of including a point or any sequence of points in the identified set is greater than or equal to 1/2 with probability that converges to one. This property follows immediately from the uniform coverage probability results for the CS's. The level 1/2 CS, however, is not necessarily median-unbiased in two directions.[4] Nevertheless, this set is guaranteed not to be inward-median biased. Chernozhukov, Lee, and Rosen (2008) also provide bias reduction methods for set estimators.

The results of the paper are stated for the case where the parameter of interest, θ, is finite-dimensional. However, all of the results except the local power results also hold for infinite-dimensional parameters θ. Computation of a CS is noticeably more difficult in the infinite-dimensional case.

The CS's constructed in the paper provide model specification tests of the conditional moment inequality model. One rejects the model if a nominal $1-\alpha$ CS is empty. The results of the paper for CS's imply that this test has asymptotic size less than or equal to α (with the inequality possibly being strict), e.g., see Andrews and Guggenberger

[4] That is, the probability of including points outside the identified set is not necessarily less than or equal to 1/2 with probability that goes to one. This is because lower and upper confidence bounds on the boundary of an identified set do not necessarily coincide.

(2009) for details of the argument.

As noted above, the determination of appropriate critical values plays a major role in all moment inequality tests. This is because the null distribution of a test statistic depends greatly on the slackness, or lack thereof, of the different moment inequalities. The slackness represents a nuisance parameter that appears under the null hypothesis. With conditional moment inequalities, slackness comes in the form of a function, which is an infinite-dimensional parameter. In consequence, the issues that arise due to slackness are exacerbated in conditional moment inequality models compared to unconditional moment inequality models.

The effect of slackness in the moment inequalities causes a discontinuity in the pointwise asymptotic distribution of typical test statistics. This occurs because if a moment inequality is binding, then it affects the pointwise asymptotic distribution of the test statistic, but if the moment inequality is not binding, then the asymptotic distribution of the test statistic is the same as if this moment inequality did not enter the test statistic at all. However, in finite samples there is no discontinuity in the distribution of the test statistic. If a moment inequality is slack by a small amount, the finite sample distribution of the test statistic differs little from when it is binding. In finite samples what matters is how close or distant moment inequalities are to binding, not whether they are binding or not binding. The latter is a potentially misleading distinction obtained by focusing on pointwise asymptotics and is divorced from the finite-sample properties of the test statistic.

In the case of conditional moment inequalities, for some value(s) of x, an inequality that is binding for $X_i = x$ is not binding for some value of x' that is arbitrarily close to x, provided the inequality is not binding for all x and is a smooth function of x. In consequence, one obtains an extreme form of discontinuity of the pointwise asymptotic distribution in which two moment inequalities are arbitrarily close to one another but pointwise asymptotics say that one inequality is irrelevant but the other is not.

The upshot of the discussion above is that pointwise asymptotics do not provide good approximations to the finite-sample properties of test statistics in moment inequality models, especially conditional models. The problem is that pointwise asymptotics do not provide uniform approximations. Depending on the sample size, different values of the "slackness function" cause problems—no matter how large is the sample size. In consequence, pointwise asymptotics fail to detect the potential problems. For issues concerning uniformity of asymptotics in other econometric models, see, e.g., Kabaila

(1995), Leeb and Pötscher (2005), Mikusheva (2007), and Andrews and Guggenberger (2010) (AG).

To ensure that a test (or CS) has good finite-sample size properties one needs to establish asymptotic results that hold uniformly over potential true distributions. AG and Andrews, Cheng, and Guggenberger (2009) (ACG) show that in certain problems one can establish uniform asymptotic results by determining the asymptotic behavior of a statistic and its critical value under particular drifting sequences of true distributions. These results apply to unconditional moment inequality models, see Andrews and Guggenberger (2009) and Andrews and Soares (2010). However, they do not apply to conditional moment inequality models. The reason is that the AG and ACG results require that the asymptotic distribution of the statistic only depends on a finite-dimensional parameter. In the unconditional moment inequality case, this is the vector of moment inequality slackness values.[5] However, with conditional moment inequalities, the nuisance parameter is a vector of slackness functions, which is infinite-dimensional.

The main technical contribution of this paper is to introduce a new method of proving uniformity results that applies to cases in which an infinite-dimensional nuisance parameter appears in the problem. The method is to establish an approximation to the sample size n distribution of the test statistic by a function of a Gaussian distribution where the function depends on the true slackness functions for the given sample size n and the approximation is uniform over all possible true slackness functions.[6] Then, one shows that the data-dependent critical value (the GMS critical value in the present case) is less than or equal to the $1-\alpha$ quantile of the given function of the Gaussian process with probability that goes to one uniformly over all potential true distributions. This establishes that the CS has correct size, greater than or equal to $1-\alpha$, asymptotically and provides the justification for its use. Under a mild distributional continuity condition, one obtains that the asymptotic size equals $1-\alpha$.

The remainder of the paper is organized as follows. Section 2 introduces the moment inequality/equality model. Section 3 specifies the class of test statistics that is considered. Section 4 defines GMS CS's. Section 5 establishes the uniform asymptotic

[5] It also depends on the variance matrix of the moment functions, but the latter does not cause uniformity problems and is not an issue because it can be estimated consistently.

[6] Uniformity is obtained without any regularity conditions in terms of smoothness, uniform continuity, or even continuity of the conditional moment functions as functions of X_i. This is important because the slackness functions are normalized by an increasing function of n which typically would cause violation of uniform continuity or uniform bounds on the derivatives of smooth functions even if the underlying conditional moment inequality functions were smooth in X_i.

coverage properties of GMS and PA CS's. Section 6 establishes the consistency of GMS and PA tests against all fixed alternatives. This implies that GMS and PA CS's do not include any point outside the identified set with probability that goes to one. Section 7 shows that GMS and PA tests have power against some $n^{-1/2}$-local alternatives. Section 8 considers models in which preliminary consistent estimators of identified parameters are plugged into the moment inequalities/equalities. It also considers time series observations. Section 9 provides the Monte Carlo simulation results.

Appendix A provides proofs of the uniform asymptotic coverage probability results for GMS and PA CS's. Appendices B-E are given in the Supplement to this paper, Andrews and Shi (2009) (AS). Appendix B provides (i) results for KS tests and CS's, (ii) the extension of the results of the paper to truncated/simulated CvM tests and CS's, (iii) an illustration of the verification of the assumptions used for the local alternative results, (iv) an illustration of (serious) uniformity problems that arise with the Kolmogorov-Smirnov test unless the critical value is chosen carefully, (v) an illustration of problems with pointwise asymptotics, and (vi) asymptotic coverage probability results for subsampling CS's under drifting sequences of distributions. Appendix C gives proofs of the results stated in the paper, but not given in Appendix A. Appendix D provides proofs of the results stated in Appendix B. Appendix E provides a proof of some empirical process results that are used in Appendices A, C, and D. Appendix F provides some additional material concerning the Monte Carlo simulation results of Section 9.

2 Conditional Moment Inequalities/Equalities

2.1 Model

The conditional moment inequality/equality model is defined as follows. We suppose there exists a true parameter $\theta_0 \in \Theta \subset R^{d_\theta}$ that satisfies the moment conditions:

$$\begin{aligned} &E_{F_0}(m_j\left(W_i, \theta_0\right)|X_i) \geq 0 \text{ a.s. } [F_{X,0}] \text{ for } j = 1, ..., p \text{ and} \\ &E_{F_0}(m_j\left(W_i, \theta_0\right)|X_i) = 0 \text{ a.s. } [F_{X,0}] \text{ for } j = p+1, ..., p+v, \end{aligned} \tag{2.1}$$

where $m_j(\cdot, \theta)$, $j = 1, ..., p+v$ are (known) real-valued moment functions, $\{W_i = (Y_i', X_i')' : i \leq n\}$ are observed i.i.d. random vectors with distribution F_0, $F_{X,0}$ is the marginal distribution of X_i, $X_i \in R^{d_x}$, $Y_i \in R^{d_y}$, and $W_i \in R^{d_w}$ $(= R^{d_y + d_x})$.

We are interested in constructing CS's for the true parameter θ_0. However, we do not assume that θ_0 is point identified. Knowledge of $E_{F_0}(m_j(W_i,\theta)\,|X_i)$ for all $\theta \in \Theta$ does not necessarily identify θ_0. Even knowledge of F_0 does not necessarily point identify θ_0.[7] The model, however, restricts the true parameter value to a set called the *identified set* (which could be a singleton). The identified set is

$$\Theta_{F_0} = \{\theta \in \Theta : (2.1) \text{ holds with } \theta \text{ in place of } \theta_0\}. \tag{2.2}$$

Let (θ, F) denote generic values of the parameter and distribution. Let $\mathcal{F}$ denote the parameter space for (θ_0, F_0). By definition, $\mathcal{F}$ is a collection of (θ, F) such that

$$\begin{aligned}
&\text{(i)} \;\; \theta \in \Theta,\\
&\text{(ii)} \;\; \{W_i : i \geq 1\} \text{ are i.i.d. under } F,\\
&\text{(iii)} \;\; E_F(m_j(W_i,\theta)\,|X_i) \geq 0 \text{ a.s. } [F_X] \text{ for } j = 1, ..., p,\\
&\text{(iv)} \;\; E_F(m_j(W_i,\theta)\,|X_i) = 0 \text{ a.s. } [F_X] \text{ for } j = p+1, ..., k,\\
&\text{(v)} \;\; 0 < Var_F(m_j(W_i,\theta)) < \infty \text{ for } j = 1, ..., k, \text{ and}\\
&\text{(vi)} \;\; E_F\left|m_j(W_i,\theta)/\sigma_{F,j}(\theta)\right|^{2+\delta} \leq B \text{ for } j = 1, ..., k,
\end{aligned} \tag{2.3}$$

for some $B < \infty$ and $\delta > 0$, where $k = p + v$, F_X is the marginal distribution of X_i under F, and $\sigma^2_{F,j}(\theta) = Var_F(m_j(W_i,\theta))$.[8] The k-vector of moment functions is denoted

$$m(W_i,\theta) = (m_1(W_i,\theta), ..., m_k(W_i,\theta))'. \tag{2.4}$$

2.2 Confidence Sets

We are interested in CS's that cover the true value θ_0 with probability greater than or equal to $1-\alpha$ for $\alpha \in (0,1)$. As is standard, we construct such CS's by inverting tests of the null hypothesis that θ is the true value for each $\theta \in \Theta$. Let $T_n(\theta)$ be a test statistic and $c_{n,1-\alpha}(\theta)$ be a corresponding critical value for a test with nominal significance level

[7] It makes sense to speak of a "true" parameter θ_0 in the present context because (i) there may exist restrictions not included in the moment inequalities/equalities in (2.1) that point identify θ_0, but for some reason are not available or are not utilized, and/or (ii) there may exist additional variables not included in W_i which, if observed, would lead to point identification of θ_0. Given such restrictions and/or variables, the true parameter θ_0 is uniquely defined even if it is not point identified by (2.1).

[8] Additional restrictions can be placed on $\mathcal{F}$ and the results of the paper still hold. For example, one could specify that the support of X_i is the same for all F for which $(\theta, F) \in \mathcal{F}$.

α. Then, a nominal level $1-\alpha$ CS for the true value θ_0 is

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_{n,1-\alpha}(\theta)\}. \tag{2.5}$$

3 Test Statistics

3.1 General Form of the Test Statistic

Here we define the test statistic $T_n(\theta)$ that is used to construct a CS. We transform the conditional moment inequalities/equalities into equivalent unconditional moment inequalities/equalities by choosing appropriate weighting functions, i.e., instruments. Then, we construct a test statistic based on the unconditional moment conditions.

The unconditional moment conditions are of the form:

$$\begin{aligned} &E_{F_0} m_j\left(W_i, \theta_0\right) g_j\left(X_i\right) \geq 0 \text{ for } j = 1, ..., p \text{ and} \\ &E_{F_0} m_j\left(W_i, \theta_0\right) g_j\left(X_i\right) = 0 \text{ for } j = p+1, ..., k, \text{ for } g = (g_1, ..., g_k)' \in \mathcal{G}, \end{aligned} \tag{3.1}$$

where $g = (g_1, ..., g_k)'$ are instruments that depend on the conditioning variables X_i and $\mathcal{G}$ is a collection of instruments. Typically $\mathcal{G}$ contains an infinite number of elements.

The identified set $\Theta_{F_0}(\mathcal{G})$ of the model defined by (3.1) is

$$\Theta_{F_0}(\mathcal{G}) = \{\theta \in \Theta : (3.1) \text{ holds with } \theta \text{ in place of } \theta_0\}. \tag{3.2}$$

The collection $\mathcal{G}$ is chosen so that $\Theta_{F_0}(\mathcal{G}) = \Theta_{F_0}$, defined in (2.2). Section 3.3 provides conditions for this equality and gives examples of instrument sets $\mathcal{G}$ that satisfy the conditions.

We construct test statistics based on (3.1). The sample moment functions are

$$\begin{aligned} \overline{m}_n(\theta, g) &= n^{-1} \sum_{i=1}^{n} m(W_i, \theta, g) \text{ for } g \in \mathcal{G}, \text{ where} \\ m(W_i, \theta, g) &= \begin{pmatrix} m_1(W_i, \theta) g_1(X_i) \\ m_2(W_i, \theta) g_2(X_i) \\ \vdots \\ m_k(W_i, \theta) g_k(X_i) \end{pmatrix} \text{ for } g \in \mathcal{G}. \end{aligned} \tag{3.3}$$

The sample variance-covariance matrix of $n^{1/2}\overline{m}_n(\theta, g)$ is

$$\widehat{\Sigma}_n(\theta, g) = n^{-1} \sum_{i=1}^{n} \left(m(W_i, \theta, g) - \overline{m}_n(\theta, g)\right) \left(m(W_i, \theta, g) - \overline{m}_n(\theta, g)\right)'. \qquad (3.4)$$

The matrix $\widehat{\Sigma}_n(\theta, g)$ may be singular or near singular with non-negligible probability for some $g \in \mathcal{G}$. This is undesirable because the inverse of $\widehat{\Sigma}_n(\theta, g)$ needs to be consistent for its population counterpart uniformly over $g \in \mathcal{G}$ for the test statistics considered below. In consequence, we employ a modification of $\widehat{\Sigma}_n(\theta, g)$, denoted $\overline{\Sigma}_n(\theta, g)$, such that $\det(\overline{\Sigma}_n(\theta, g))$ is bounded away from zero. Different choices of $\overline{\Sigma}_n(\theta, g)$ are possible. Here we use

$$\overline{\Sigma}_n(\theta, g) = \widehat{\Sigma}_n(\theta, g) + \varepsilon \cdot Diag(\widehat{\Sigma}_n(\theta, 1_k)) \text{ for } g \in \mathcal{G} \qquad (3.5)$$

for some fixed $\varepsilon > 0$. Specifically, in the simulations in Section 9, we use $\varepsilon = 5/100$. By design, $\overline{\Sigma}_n(\theta, g)$ is a linear combination of two scale equivariant functions and thus is scale equivariant. (That is, multiplying the moment functions $m(W_i, \theta)$ by a diagonal matrix, D, changes $\overline{\Sigma}_n(\theta, g)$ into $D\overline{\Sigma}_n(\theta, g)D$.) This yields a test statistic that is invariant to rescaling of the moment functions $m(W_i, \theta)$, which is an important property.

The test statistic $T_n(\theta)$ is either a Cramér-von Mises-type (CvM) or Kolmogorov-Smirnov-type (KS) statistic. The CvM statistic is

$$T_n(\theta) = \int S(n^{1/2}\overline{m}_n(\theta, g), \overline{\Sigma}_n(\theta, g))dQ(g), \qquad (3.6)$$

where S is a non-negative function, Q is a weight function (i.e., probability measure) on $\mathcal{G}$, and the integral is over $\mathcal{G}$. The functions S and Q are discussed in Sections 3.2 and 3.4 below, respectively.

The Kolmogorov-Smirnov-type (KS) statistic is

$$T_n(\theta) = \sup_{g \in \mathcal{G}} S(n^{1/2}\overline{m}_n(\theta, g), \overline{\Sigma}_n(\theta, g)). \qquad (3.7)$$

For brevity, in the text of the paper, the discussion focusses on CvM statistics and all results stated concern CvM statistics. Appendix B of AS gives detailed results for KS statistics.

3.2 Function S

To permit comparisons, we establish results in this paper for a broad family of functions S that satisfy certain conditions stated below. We now introduce three functions that satisfy these conditions. The first is the modified method of moments (MMM) or Sum function:

$$S_1(m, \Sigma) = \sum_{j=1}^{p} [m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v} [m_j/\sigma_j]^2, \tag{3.8}$$

where m_j is the jth element of the vector m, σ_j^2 is the jth diagonal element of the matrix Σ, and $[x]_- = -x$ if $x < 0$ and $[x]_- = 0$ if $x \geq 0$.

The second function S is the quasi-likelihood ratio (QLR) function:

$$S_2(m, \Sigma) = \inf_{t=(t_1', 0_v')': t_1 \in R_{+,\infty}^p} (m - t)' \Sigma^{-1} (m - t). \tag{3.9}$$

The third function S is a "maximum" (Max) function. Used in conjunction with the KS form of the test statistic, this S function yields a pure KS-type test statistic:

$$S_3(m, \Sigma) = \max\{[m_1/\sigma_1]_-^2, ..., [m_p/\sigma_p]_-^2, (m_{p+1}/\sigma_{p+1})^2, ..., (m_{p+v}/\sigma_{p+v})^2\}. \tag{3.10}$$

The function S_2 is more costly to compute than S_1 and S_3.

Let $m_I = (m_1, ..., m_p)'$ and $m_{II} = (m_{p+1}, ..., m_k)'$. Let Δ be the set of $k \times k$ positive-definite diagonal matrices. Let $\mathcal{W}$ be the set of $k \times k$ positive-definite matrices. Let (i) $R_{[+\infty]}$, (ii) R_+, and (iii) $R_{+,\infty}$ denote the sets of scalars that are (i) real or $+\infty$, (ii) non-negative (and finite), and (iii) non-negative or $+\infty$, respectively. Let $\mathcal{S} = \{(m, \Sigma) : m \in R_{[+\infty]}^p \times R^v, \Sigma \in \mathcal{W}\}$, where $R_{[+\infty]}^p = R_{[+\infty]} \times ... \times R_{[+\infty]}$ with p copies.

We consider functions S that satisfy the following conditions.

Assumption S1. $\forall (m, \Sigma) \in \mathcal{S}$,

(a) $S(Dm, D\Sigma D) = S(m, \Sigma)$ $\forall D \in \Delta$,

(b) $S(m_I, m_{II}, \Sigma)$ is non-increasing in each element of m_I,

(c) $S(m, \Sigma) \geq 0$,

(d) S is continuous, and

(e) $S(m, \Sigma + \Sigma_1) \leq S(m, \Sigma)$ for all $k \times k$ positive semi-definite matrices Σ_1.

It is worth pointing out that Assumption S1(d) requires S to be continuous in m at all points m in the extended vector space $R_{[+\infty]}^p \times R^v$, not only for points in R^{p+v}.

Assumption S2. $S(m,\Sigma)$ is uniformly continuous in the sense that, for all $m_0 \in R^k$ and all pd Σ_0, $\sup_{\mu \in R_+^p \times \{0\}^v} |S(m+\mu,\Sigma) - S(m_0+\mu,\Sigma_0)| \to 0$ as $(m,\Sigma) \to (m_0,\Sigma_0)$.[9]

The following two assumptions are used only to establish the power properties of tests.

Assumption S3. $S(m,\Sigma) > 0$ if and only if $m_j < 0$ for some $j = 1,...,p$ or $m_j \neq 0$ for some $j = p+1,...,k$, where $m = (m_1,...,m_k)'$ and $\Sigma \in \mathcal{W}$.

Assumption S4. For some $\chi > 0$, $S(am,\Sigma) = a^\chi S(m,\Sigma)$ for all scalars $a > 0$, $m \in R^k$, and $\Sigma \in \mathcal{W}$.

Assumptions S1-S4 are not restrictive as is shown by the following result.

Lemma 1. *The functions S_1, S_2, and S_3 satisfy Assumptions* S1-S4.

3.3 Instruments

When considering consistent specification tests based on conditional moment *equalities*, see Bierens (1982) and Bierens and Ploberger (1997), a wide variety of different types of functions g can be employed without loss of information, see Stinchcombe and White (1998). With conditional moment *inequalities*, however, it is much more difficult to distill the information in the moments because of the one-sided feature of the inequalities.

The collection of instruments $\mathcal{G}$ needs to satisfy the following condition in order for the unconditional moments $\{E_F m(W_i,\theta,g) : g \in \mathcal{G}\}$ to incorporate the same information as the conditional moments $\{E_F(m(W_i,\theta)|X_i = x) : x \in R^{d_x}\}$.

For any $\theta \in \Theta$ and any distribution F with $E_F||m(W_i,\theta)|| < \infty$, let

$$\begin{aligned}\mathcal{X}_F(\theta) = \{x \in R^{d_x} : \ & E_F(m_j(W_i,\theta)|X_i = x) < 0 \text{ for some } j \leq p \text{ or} \\ & E_F(m_j(W_i,\theta)|X_i = x) \neq 0 \text{ for some } j = p+1,...,k\}. \qquad (3.11)\end{aligned}$$

Assumption CI. For any $\theta \in \Theta$ and distribution F for which $E_F||m(W_i,\theta)|| < \infty$ and

[9] It is important that the supremum is only over μ vectors with non-negative elements μ_j for $j \leq p$. Without this restriction on the μ vectors, Assumption S2 would not hold for typical S functions of interest.

$P_F(X_i \in \mathcal{X}_F(\theta)) > 0$, there exists some $g \in \mathcal{G}$ such that

$$E_F m_j(W_i, \theta) g_j(X_i) < 0 \text{ for some } j \le p \text{ or}$$
$$E_F m_j(W_i, \theta) g_j(X_i) \ne 0 \text{ for some } j = p+1, ..., k.$$

Note that CI abbreviates "conditionally identified." The following simple Lemma indicates the importance of Assumption CI.

Lemma 2. *Assumption* CI *implies that* $\Theta_F(\mathcal{G}) = \Theta_F$ *for all* F *with* $\sup_{\theta \in \Theta} E_F ||m(W_i, \theta)|| < \infty$.

Collections $\mathcal{G}$ that satisfy Assumption CI contain non-negative functions whose supports are cubes, boxes, or bounded sets with other shapes whose supports are arbitrarily small, see below.

Next, we state a "manageability" condition that regulates the complexity of $\mathcal{G}$. It ensures that $\{n^{1/2}(\overline{m}_n(\theta, g) - E_{F_n}\overline{m}_n(\theta, g)) : g \in \mathcal{G}\}$ satisfies a functional central limit theorem (FCLT) under drifting sequences of distributions $\{F_n : n \ge 1\}$. The latter is utilized in the proof of the uniform coverage probability results for the CS's. The manageability condition is from Pollard (1990) and is defined and explained in Appendix E of AS.

Assumption M. (a) $0 \le g_j(x) \le G(x)\ \forall x \in R^{d_x}, \forall j \le k, \forall g \in \mathcal{G}$, for some envelope function $G(x)$,

(b) $E_F G^{\delta_1}(X_i) \le C$ for all F such that $(\theta, F) \in \mathcal{F}$ for some $\theta \in \Theta$, for some $C < \infty$, and for some $\delta_1 > 4/\delta + 2$, where $W_i = (Y_i', X_i')' \sim F$ and δ is as in the definition of $\mathcal{F}$ in (2.3), and

(c) the processes $\{g_j(X_{n,i}) : g \in \mathcal{G},\ i \le n, n \ge 1\}$ are manageable with respect to the envelope function $G(X_{n,i})$ for $j = 1, ..., k$, where $\{X_{n,i} : i \le n, n \ge 1\}$ is a row-wise i.i.d. triangular array with $X_{n,i} \sim F_{X,n}$ and $F_{X,n}$ is the distribution of $X_{n,i}$ under F_n for some $(\theta_n, F_n) \in \mathcal{F}$ for $n \ge 1$.[10]

Now we give two examples of collections of functions $\mathcal{G}$ that satisfy Assumptions CI and M. Appendix B of AS gives three additional examples, one of which is based on

[10] The asymptotic results given below hold with Assumption M replaced by any alternative assumption that is sufficient to obtain the requisite empirical process results, see Assumption EP in Section 8.

B-splines.

Example 1. (Countable Hypercubes). Suppose X_i is transformed via a one-to-one mapping so that each of its elements lies in $[0,1]$. There is no loss in information in doing so. Section 9 and Appendix B of AS provide examples of how this can be done.

Consider the class of indicator functions of cubes with side lengths $(2r)^{-1}$ for all large positive integers r that partition $[0,1]^{d_x}$ for each r. This class is countable:

$$\begin{aligned}
\mathcal{G}_{c\text{-}cube} &= \{g(x) : g(x) = 1(x \in C) \cdot 1_k \text{ for } C \in \mathcal{C}_{c\text{-}cube}\}, \text{ where} \\
\mathcal{C}_{c\text{-}cube} &= \Bigg\{ C_{a,r} = \prod_{u=1}^{d_x} ((a_u - 1)/(2r), a_u/(2r)] \in [0,1]^{d_x} : a = (a_1, ..., a_{d_x})' \\
&\qquad a_u \in \{1, 2, ..., 2r\} \text{ for } u = 1, ..., d_x \text{ and } r = r_0, r_0 + 1, ... \Bigg\} \qquad (3.12)
\end{aligned}$$

for some positive integer r_0.[11] The terminology "*c-cube*" abbreviates countable cubes. Note that $C_{a,r}$ is a hypercube in $[0,1]^{d_x}$ with smallest vertex indexed by a and side lengths equal to $(2r)^{-1}$.

The class of countable cubes $\mathcal{G}_{c\text{-}cube}$ leads to a test statistic $T_n(\theta)$ for which the integral over $\mathcal{G}$ reduces to a sum.

Example 2 (Boxes). Let

$$\begin{aligned}
\mathcal{G}_{box} &= \{g : g(x) = 1(x \in C) \cdot 1_k \text{ for } C \in \mathcal{C}_{box}\}, \text{ where} \qquad (3.13) \\
\mathcal{C}_{box} &= \left\{ C_{x,r} = \prod_{u=1}^{d_x} (x_u - r_u, x_u + r_u] \in R^{d_x} : x_u \in R,\ r_u \in (0, \bar{r})\ \forall u \leq d_x \right\},
\end{aligned}$$

$x = (x_1, ..., x_{d_x})'$, $r = (r_1, ..., r_{d_x})'$, $\bar{r} \in (0, \infty]$, and 1_k is a k-vector of ones. The set $\mathcal{C}_{box}$ contains boxes (i.e., hyper-rectangles or orthotopes) in R^{d_x} with centers at $x \in R^{d_x}$ and side lengths less than $2\bar{r}$.

When the support of X_i, denoted $Supp(X_i)$, is a known subset of R^{d_x}, one can replace $x_u \in R\ \forall u \leq d_x$ in (3.13) by $x \in CH(Supp(X_i))$, where $CH(A)$ denotes the convex hull of A. Sometimes, it is convenient to transform the elements of X_i into $[0,1]$ via strictly increasing transformations as in Example 1 above. If the X_i's are transformed in this way, then R in (3.13) is replaced by $[0,1]$.

Both of the sets $\mathcal{G}$ discussed above can be used with continuous and/or discrete

[11] When $a_u = 1$, the left endpoint of the interval $(0, 1/(2r)]$ is included in the interval.

regressors.

The following result establishes Assumptions CI and M for $\mathcal{G}_{c\text{-}cube}$ and $\mathcal{G}_{box}$.

Lemma 3. *For any moment function $m(W_i, \theta)$, Assumptions* CI *and* M *hold with $\mathcal{G} = \mathcal{G}_{c\text{-}cube}$ and $\mathcal{G}_{box}$.*

The proof of Lemma 3 is given in Appendix C of AS.

Moment Equalities. The sets $\mathcal{G}$ introduced above use the same functions for the moment inequalities and equalities, i.e., g is of the form $g^* \cdot 1_k$, where g^* is a real-valued function. It is possible to use different functions for the moment equalities than for the inequalities. One can take $g = (g^{(1)\prime}, g^{(2)\prime})' \in \mathcal{G}^{(1)} \times \mathcal{G}^{(2)}$, where $g^{(1)}$ is an R^p-valued function in some set $\mathcal{G}^{(1)}$ and $g^{(2)}$ is an R^v-valued function in some set $\mathcal{G}^{(2)}$. Any "generically comprehensively revealing" class of functions $\mathcal{G}^{(2)}$, see Stinchcombe and White (1998), leads to a set $\mathcal{G}$ that satisfies Assumption CI provided one uses a suitable class of functions $\mathcal{G}^{(1)}$ (such as any of those defined above with 1_k replaced by 1_p). For brevity, we do not provide further details.

3.4 Weight Function Q

The weight function Q can be any probability measure on $\mathcal{G}$ whose support is $\mathcal{G}$. This support condition is needed to ensure that no functions $g \in \mathcal{G}$, which might have set-identifying power, are "ignored" by the test statistic $T_n(\theta)$. Without such a condition, a CS based on $T_n(\theta)$ would not necessarily shrink to the identified set as $n \to \infty$. Section 6 below introduces the support condition formally and shows that the probability measures Q considered here satisfy it.

We now specify two examples of weight functions Q. Three others are specified in Appendix B of AS.

Weight Function Q for $\mathcal{G}_{c\text{-}cube}$. There is a one-to-one mapping $\Pi_{c\text{-}cube} : \mathcal{G}_{c\text{-}cube} \to AR = \{(a, r) : a \in \{1, ..., 2r\}^{d_x} \text{ and } r = r_0, r_0 + 1, ...\}$. Let Q_{AR} be a probability measure on AR. One can take $Q = \Pi^{-1}_{c\text{-}cube} Q_{AR}$. A natural choice of measure Q_{AR} is uniform on $a \in \{1, ..., 2r\}^{d_x}$ conditional on r combined with a distribution for r that has some probability mass function $\{w(r) : r = r_0, r_0 + 1, ...\}$. This yields the test statistic to be

$$T_n(\theta) = \sum_{r=r_0}^{\infty} w(r) \sum_{a \in \{1,...,2r\}^{d_x}} (2r)^{-d_x} S(n^{1/2}\overline{m}_n(\theta, g_{a,r}), \overline{\Sigma}_n(\theta, g_{a,r})), \tag{3.14}$$

where $g_{a,r}(x) = 1(x \in C_{a,r}) \cdot 1_k$ for $C_{a,r} \in \mathcal{C}_{c\text{-}cube}$.

Weight Function Q for $\mathcal{G}_{box}$. There is a one-to-one mapping $\Pi_{box} : \mathcal{G}_{box} \to \{(x,r) \in R^{d_x} \times (0,\bar{r})^{d_x}\}$. Let Q^* be a probability measure on $\{(x,r) \in R^{d_x} \times (0,\bar{r})^{d_x}\}$. Then, $\Pi_{box}^{-1}Q^*$ is a probability measure on $\mathcal{G}_{box}$. One can take $Q = \Pi_{box}^{-1}Q^*$. Any probability measure on $R^{d_x} \times (0,\bar{r})^{d_x}$ whose support contains $\mathcal{G}_{box}$ is a valid candidate for Q^*. If $Supp(X_i)$ is known, R^{d_x} can be replaced by the convex hull of $Supp(X_i)$. One choice is to transform each regressor to lie in $[0,1]$ and to take Q^* to be the uniform distribution on $[0,1]^{d_x} \times (0,\bar{r})^{d_x}$, i.e., $Unif([0,1]^{d_x} \times (0,\bar{r})^{d_x})$. In this case, the test statistic becomes

$$T_n(\theta) = \int_{[0,1]^{d_x}} \int_{(0,\bar{r})^{d_x}} S(n^{1/2}\overline{m}_n(\theta, g_{x,r}), \overline{\Sigma}_n(\theta, g_{x,r}))\bar{r}^{-d_x} dr dx, \tag{3.15}$$

where $g_{x,r}(y) = 1(y \in C_{x,r}) \cdot 1_k$ and $C_{x,r}$ denotes the box centered at $x \in [0,1]^{d_x}$ with side lengths $2r \in (0, 2\bar{r})^{d_x}$.

3.5 Computation of Sums, Integrals, and Suprema

The test statistics $T_n(\theta)$ given in (3.14) and (3.15) involve an infinite sum and an integral with respect to Q. Analogous infinite sums and integrals appear in the definitions of the critical values given below. These infinite sums and integrals can be approximated by truncation, simulation, or quasi-Monte Carlo (QMC) methods. If $\mathcal{G}$ is countable, let $\{g_1, ..., g_{s_n}\}$ denote the first s_n functions g that appear in the infinite sum that defines $T_n(\theta)$. Alternatively, let $\{g_1, ..., g_{s_n}\}$ be s_n i.i.d. functions drawn from $\mathcal{G}$ according to the distribution Q. Or, let $\{g_1, ..., g_{s_n}\}$ be the first s_n terms in a QMC approximation of the integral wrt Q. Then, an approximate test statistic obtained by truncation, simulation, or QMC methods is

$$\overline{T}_{n,s_n}(\theta) = \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S(n^{1/2}\overline{m}_n(\theta, g_\ell), \overline{\Sigma}_n(\theta, g_\ell)), \tag{3.16}$$

where $w_{Q,n}(\ell) = Q(\{g_\ell\})$ when an infinite sum is truncated, $w_{Q,n}(\ell) = s_n^{-1}$ when $\{g_1, ..., g_{s_n}\}$ are i.i.d. draws from $\mathcal{G}$ according to Q, and $w_{Q,n}(\ell)$ is a suitable weight when a QMC method is used. For example, in (3.14), the outer sum can be truncated at $r_{1,n}$, in which case, $s_n = \sum_{r=r_0}^{r_{1,n}} (2r)^{d_X}$ and $w_{Q,n}(\ell) = w(r)(2r)^{-d_x}$ for ℓ such that g_ℓ corresponds to $g_{a,r}$ for some a. In (3.15), the integral over (x,r) can be replaced by an average over $\ell = 1, ..., s_n$, the uniform density $\bar{r}^{-d_x}$ deleted, and $g_{x,r}$ replaced by g_{x_ℓ, r_ℓ}, where $\{(x_\ell, r_\ell) : \ell = 1, ..., s_n\}$ are i.i.d. with a $Unif([0,1]^{d_x} \times (0, \bar{r})^{d_x})$ distribution.

In Appendix B of AS, we show that truncation at s_n, simulation based on s_n simulation repetitions, or QMC approximation based on s_n terms, where $s_n \to \infty$ as $n \to \infty$, is sufficient to maintain the asymptotic validity of the tests and CS's as well as the asymptotic power results under fixed alternatives and most of the results under $n^{-1/2}$-local alternatives.

The KS form of the test statistic requires the computation of a supremum over $g \in \mathcal{G}$. For computational ease, this can be replaced by a supremum over $g \in \mathcal{G}_n$, where $\mathcal{G}_n \uparrow \mathcal{G}$ as $n \to \infty$, in the test statistic and in the definition of the critical value (defined below). The asymptotic results for KS tests given in Appendix B of AS show that the use of $\mathcal{G}_n$ in place of $\mathcal{G}$ does not affect the asymptotic properties of the test.

4 GMS Confidence Sets

4.1 GMS Critical Values

In this section, we define GMS critical values and CS's.

It is shown in Section 5 below that when θ is in the identified set the "uniform asymptotic distribution" of $T_n(\theta)$ is the distribution of $T(h_n)$, where $h_n = (h_{1,n}, h_2)$, $h_{1,n}(\cdot)$ is a function from $\mathcal{G}$ to $R^p_{+,\infty} \times \{0\}^v$ that depends on the slackness of the moment inequalities and on n, and $h_2(\cdot,\cdot)$ is a $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$. For $h = (h_1, h_2)$, define

$$T(h) = \int S(\nu_{h_2}(g) + h_1(g), h_2(g,g) + \varepsilon I_k) dQ(g), \tag{4.1}$$

where

$$\{\nu_{h_2}(g) : g \in \mathcal{G}\} \tag{4.2}$$

is a mean zero R^k-valued Gaussian process with covariance kernel $h_2(\cdot,\cdot)$ on $\mathcal{G} \times \mathcal{G}$, $h_1(\cdot)$ is a function from $\mathcal{G}$ to $R^p_{+,\infty} \times \{0\}^v$, and ε is as in the definition of $\overline{\Sigma}_n(\theta, g)$ in (3.5).[12] The definition of $T(h)$ in (4.1) applies to CvM test statistics. For the KS test statistic, one replaces $\int ... \, dQ(g)$ by $\sup_{g \in \mathcal{G}} ...$.

[12] The sample paths of $\nu_{h_2}(\cdot)$ are concentrated on the set $U^k_\rho(\mathcal{G})$ of bounded uniformly ρ-continuous R^k-valued functions on $\mathcal{G}$, where ρ is defined in Appendix A.

We are interested in tests of nominal level α and CS's of nominal level $1-\alpha$. Let

$$c_0(h, 1-\alpha) \ (= c_0(h_1, h_2, 1-\alpha)) \tag{4.3}$$

denote the $1-\alpha$ quantile of $T(h)$. If $h_n = (h_{1,n}, h_2)$ was known, we would use $c_0(h_n, 1-\alpha)$ as the critical value for the test statistic $T_n(\theta)$. However, h_n is not known and $h_{1,n}$ cannot be consistently estimated. In consequence, we replace h_2 in $c_0(h_{1,n}, h_2, 1-\alpha)$ by a uniformly consistent estimator $\widehat{h}_{2,n}(\theta)$ $(= \widehat{h}_{2,n}(\theta, \cdot, \cdot))$ of the covariance kernel h_2 and we replace $h_{1,n}$ by a data-dependent GMS function $\varphi_n(\theta)$ $(= \varphi_n(\theta, \cdot))$ on $\mathcal{G}$ that is constructed to be less than or equal to $h_{1,n}(g)$ for all $g \in \mathcal{G}$ with probability that goes to one as $n \to \infty$. Because $S(m, \Sigma)$ is non-increasing in m_I by Assumption S1(b), where $m = (m_I', m_{II}')'$, the latter property yields a test whose asymptotic level is less than or equal to the nominal level α. (It is arbitrarily close to α for certain $(\theta, F) \in \mathcal{F}$.) The quantities $\widehat{h}_{2,n}(\theta)$ and $\varphi_n(\theta)$ are defined below.

The nominal $1-\alpha$ GMS critical value is defined to be

$$c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha) = c_0(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) + \eta, \tag{4.4}$$

where $\eta > 0$ is an arbitrarily small positive constant, e.g., .001. A nominal $1-\alpha$ GMS CS is given by (2.5) with the critical value $c_{n,1-\alpha}(\theta)$ equal to $c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$.

The constant η is an *infinitesimal uniformity factor* (IUF) that is employed to circumvent problems that arise due to the presence of the infinite-dimensional nuisance parameter $h_{1,n}$ that affects the distribution of the test statistic in both small and large samples. The IUF obviates the need for complicated and difficult-to-verify uniform continuity and strictly-increasing conditions on the large sample distribution functions of the test statistic.

The sample covariance kernel $\widehat{h}_{2,n}(\theta)$ $(= \widehat{h}_{2,n}(\theta, \cdot, \cdot)$ is defined by:

$$\begin{aligned}
\widehat{h}_{2,n}(\theta, g, g^*) &= \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta, g, g^*)\widehat{D}_n^{-1/2}(\theta), \text{ where} \\
\widehat{\Sigma}_n(\theta, g, g^*) &= n^{-1}\sum_{i=1}^n (m(W_i, \theta, g) - \overline{m}_n(\theta, g))\,(m(W_i, \theta, g^*) - \overline{m}_n(\theta, g^*))' \text{ and} \\
\widehat{D}_n(\theta) &= Diag(\widehat{\Sigma}_n(\theta, 1_k, 1_k)).
\end{aligned} \tag{4.5}$$

Note that $\widehat{\Sigma}_n(\theta, g)$, defined in (3.4), equals $\widehat{\Sigma}_n(\theta, g, g)$ and $\widehat{D}_n(\theta)$ is the sample variance-covariance matrix of $n^{-1/2}\sum_{i=1}^n m(W_i, \theta)$.

The quantity $\varphi_n(\theta)$ is defined in Section 4.4 below.

4.2 GMS Critical Values for Approximate Test Statistics

When the test statistic is approximated via a truncated sum, simulated integral, or QMC quantity, as discussed in Section 3.4, the statistic $T(h)$ in Section 4.1 is replaced by

$$\overline{T}_{s_n}(h) = \sum_{\ell=1}^{s_n} w_{Q,n}(\ell) S(\nu_{h_2}(g_\ell) + h_1(g_\ell), h_2(g_\ell, g_\ell) + \varepsilon I_k), \tag{4.6}$$

where $\{g_\ell : \ell = 1, ..., s_n\}$ are the same functions $\{g_1, ..., g_{s_n}\}$ that appear in the approximate statistic $\overline{T}_{n,s_n}(\theta)$. We call the critical value obtained using $\overline{T}_{s_n}(h)$ an approximate GMS (A-GMS) critical value.

Let $c_{0,s_n}(h, 1-\alpha)$ denote the $1-\alpha$ quantile of $\overline{T}_{s_n}(h)$ for fixed $\{g_1, ..., g_{s_n}\}$. The A-GMS critical value is defined to be

$$c_{s_n}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha) = c_{0,s_n}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) + \eta. \tag{4.7}$$

This critical value is a quantile that can be computed by simulation as follows. Let $\{\overline{T}_{s_n,\tau}(h) : \tau = 1, ..., \tau_{reps}\}$ be τ_{reps} i.i.d. random variables each with the same distribution as $\overline{T}_{s_n}(h)$ and each with the same functions $\{g_1, ..., g_{s_n}\}$, where $h = (h_1, h_2)$ is evaluated at $(\varphi_n(\theta), \widehat{h}_{2,n}(\theta))$. Then, the A-GMS critical value, $c_{s_n}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$, is the $1-\alpha+\eta$ sample quantile of $\{\overline{T}_{s_n,\tau}(\varphi_n(\theta), \widehat{h}_{2,n}(\theta)) : \tau = 1, ..., \tau_{reps}\}$ plus η for very small $\eta > 0$ and large τ_{reps}.

When constructing a CS, one carries out multiple tests with a different θ value specified in the null hypothesis for each test. When doing so, we recommend using the same $\{g_1, ..., g_{s_n}\}$ functions for each θ value considered (although this is not necessary for the asymptotic results to hold).

4.3 Bootstrap GMS Critical Values

Bootstrap versions of the GMS critical value in (4.4) and the A-GMS critical value in (4.7) can be employed. The bootstrap GMS critical value is

$$c^*(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha) = c^*_0(\varphi_n(\theta), \widehat{h}^*_{2,n}(\theta), 1-\alpha+\eta) + \eta, \tag{4.8}$$

where $c_0^*(h, 1-\alpha)$ is the $1-\alpha$ quantile of $T^*(h)$ and $T^*(h)$ is defined as in (4.1) but with $\{\nu_{h_2}(g) : g \in \mathcal{G}\}$ and $\widehat{h}_{2,n}(\theta)$ replaced by the bootstrap empirical process $\{\nu_n^*(g) : g \in \mathcal{G}\}$ and the bootstrap covariance kernel $\widehat{h}_{2,n}^*(\theta)$, respectively. By definition, $\nu_n^*(g) = n^{-1/2}\sum_{i=1}^n (m(W_i^*, \theta, g) - \overline{m}_n(\theta, g))$, where $\{W_i^* : i \leq n\}$ is an i.i.d. bootstrap sample drawn from the empirical distribution of $\{W_i : i \leq n\}$. Also, $\widehat{h}_{2,n}^*(\theta, g, g^*)$, $\widehat{\Sigma}_n^*(\theta, g, g^*)$, and $\widehat{D}_n^*(\theta)$ are defined as in (4.5) with W_i^* in place of W_i. Note that $\widehat{h}_{2,n}^*(\theta, g, g^*)$ only enters $c(\varphi_n(\theta), \widehat{h}_{2,n}^*(\theta), 1-\alpha)$ via functions (g, g^*) such that $g = g^*$.

When the test statistic, $\overline{T}_{n,s_n}(\theta)$, is a truncated sum, simulated integral, or QMC quantity, a bootstrap A-GMS critical value can be employed. It is defined analogously to the bootstrap GMS critical value but with $T^*(h)$ replaced by $T_{s_n}^*(h)$, where $T_{s_n}^*(h)$ has the same definition as $T^*(h)$ except that a truncated sum, simulated integral, or QMC quantity, appears in place of the integral with respect to Q, as in Section 4.2. The same functions $\{g_1, ..., g_{s_n}\}$ are used in all bootstrap critical value calculations as in the test statistic $\overline{T}_{n,s_n}(\theta)$.

4.4 Definition of $\boldsymbol{\varphi_n(\theta)}$

Next, we define $\varphi_n(\theta)$. As discussed above, $\varphi_n(\theta)$ is constructed such that $\varphi_n(\theta, g) \leq h_{1,n}(g)$ $\forall g \in \mathcal{G}$ with probability that goes to one as $n \to \infty$ uniformly over (θ, F). Let

$$\xi_n(\theta, g) = \kappa_n^{-1} n^{1/2} \overline{D}_n^{-1/2}(\theta, g)\overline{m}_n(\theta, g), \text{ where } \overline{D}_n(\theta, g) = Diag(\overline{\Sigma}_n(\theta, g)), \tag{4.9}$$

$\overline{\Sigma}_n(\theta, g)$ is defined in (3.5), and $\{\kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \to \infty$. The jth element of $\xi_n(\theta, g)$, denoted $\xi_{n,j}(\theta, g)$, measures the slackness of the moment inequality $E_F m_j(W_i, \theta, g) \geq 0$ for $j = 1, ..., p$.

Define $\varphi_n(\theta, g) = (\varphi_{n,1}(\theta, g), ..., \varphi_{n,p}(\theta, g), 0, ..., 0)' \in R^k$ via, for $j \leq p$,

$$\begin{aligned}
\varphi_{n,j}(\theta, g) &= 0 && \text{if } \xi_{n,j}(\theta, g) \leq 1\\
\varphi_{n,j}(\theta, g) &= \overline{h}_{2,n,j}(\theta, g)^{1/2} B_n && \text{if } \xi_{n,j}(\theta, g) > 1, \text{ where}\\
\overline{h}_{2,n}(\theta, g) &= \widehat{D}_n^{-1/2}(\theta)\overline{\Sigma}_n(\theta, g)\widehat{D}_n^{-1/2}(\theta) && \text{and } \overline{h}_{2,n,j}(\theta, g) = [\overline{h}_{2,n}(\theta, g)]_{jj}.
\end{aligned} \tag{4.10}$$

Assumption GMS1. (a) $\varphi_n(\theta, g)$ satisfies (4.10), where $\{B_n : n \geq 1\}$ is a non-decreasing sequence of positive constants, and

(b) for some $\zeta > 1$, $\kappa_n - \zeta B_n \to \infty$ as $n \to \infty$.

The constants $\{B_n : n \geq 1\}$ in Assumption GMS1 need not diverge to infinity for the GMS CS to have asymptotic size greater than or equal to $1-\alpha$. However, for the GMS CS not to be asymptotically conservative, B_n must diverge to ∞, see Assumption GMS2(b) below. In the simulations in Section 9, we use $\kappa_n = (0.3\ln(n))^{1/2}$ and $B_n = (0.4\ln(n)/\ln\ln(n))^{1/2}$, which satisfy Assumption GMS1.

The multiplicand $\overline{h}_{2,n,j}(\theta, g)^{1/2}$ in (4.10) is an "ε-adjusted" standard deviation estimator for the jth normalized sample moment based on g. It provides a suitable scaling for $\varphi_n(\theta, g)$.

4.5 "Plug-in Asymptotic" Confidence Sets

Next, for comparative purposes, we define plug-in asymptotic (PA) critical values. Subsampling critical values are defined and analyzed in Appendix B of AS. We strongly recommend GMS critical values over PA and subsampling critical values because (i) GMS tests are shown to be more powerful than PA tests asymptotically, see Comment 2 to Theorem 4 below, (ii) it should be possible to show that GMS tests have higher power than subsampling tests asymptotically and smaller errors in null rejection probabilities asymptotically by using arguments similar to those in Andrews and Soares (2010) and Bugni (2010), respectively, and (iii) the finite-sample simulations in Section 9 show better performance by GMS critical values than PA and subsampling critical values.

PA critical values are obtained from the asymptotic null distribution that arises when all conditional moment inequalities hold as equalities a.s. The PA critical value is

$$c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha) = c_0(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha+\eta) + \eta, \tag{4.11}$$

where η is an arbitrarily small positive constant (i.e., an IUF), $0_{\mathcal{G}}$ denotes the R^k-valued function on $\mathcal{G}$ that is identically $(0, ..., 0)' \in R^k$, and $\widehat{h}_{2,n}(\theta)$ is defined in (4.5). The nominal $1-\alpha$ PA CS is given by (2.5) with the critical value $c_{n,1-\alpha}(\theta)$ equal to $c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta), 1-\alpha)$.

Bootstrap PA, A-PA, and bootstrap A-PA critical values are defined analogously to their GMS counterparts in Sections 4.2 and 4.3.

5 Uniform Asymptotic Coverage Probabilities

In this section, we show that GMS and PA CS's have correct uniform asymptotic coverage probabilities. The results of this section and those in Sections 6-8 below are for CvM statistics based on integrals with respect to Q. Extensions of these results to A-CvM statistics and critical values are provided in Appendix B of AS. Appendix B also gives results for KS tests.

5.1 Notation

In order to establish uniform asymptotic coverage probability results, we now introduce notation for the population analogues of the sample quantities that appear in (4.5). Define

$$\begin{aligned} h_{2,F}(\theta,g,g^*) &= D_F^{-1/2}(\theta)\Sigma_F(\theta,g,g^*)D_F^{-1/2}(\theta) \\ &= Cov_F\left(D_F^{-1/2}(\theta)m(W_i,\theta,g), D_F^{-1/2}(\theta)m(W_i,\theta,g^*)\right), \\ \Sigma_F(\theta,g,g^*) &= Cov_F(m(W_i,\theta,g), m(W_i,\theta,g^*)), \text{ and} \\ D_F(\theta) &= Diag(\Sigma_F(\theta,1_k,1_k)) \ (= Diag(Var_F(m(W_i,\theta)))). \end{aligned} \tag{5.1}$$

To determine the asymptotic distribution of $T_n(\theta)$, we write $T_n(\theta)$ as a function of the following quantities:

$$\begin{aligned} h_{1,n,F}(\theta,g) &= n^{1/2}D_F^{-1/2}(\theta)E_F m(W_i,\theta,g), \\ h_{n,F}(\theta,g,g^*) &= (h_{1,n,F}(\theta,g), h_{2,F}(\theta,g,g^*)), \\ \widehat{h}_{2,n,F}(\theta,g,g^*) &= D_F^{-1/2}(\theta)\widehat{\Sigma}_n(\theta,g,g^*)D_F^{-1/2}(\theta), \\ \overline{h}_{2,n,F}(\theta,g) &= \widehat{h}_{2,n,F}(\theta,g,g) + \varepsilon\widehat{h}_{2,n,F}(\theta,1_k,1_k) \ (= D_F^{-1/2}(\theta)\overline{\Sigma}_n(\theta,g)D_F^{-1/2}(\theta)), \text{ and} \\ \nu_{n,F}(\theta,g) &= n^{-1/2}\sum_{i=1}^n D_F^{-1/2}(\theta)[m(W_i,\theta,g) - E_F m(W_i,\theta,g)]. \end{aligned} \tag{5.2}$$

As defined, (i) $h_{1,n,F}(\theta,g)$ is a k-vector of normalized means of the moment functions $m(W_i,\theta,g)$ for $g \in \mathcal{G}$, which measure the slackness of the population moment conditions under (θ,F), (ii) $h_{n,F}(\theta,g,g^*)$ contains the normalized means of $D_F^{-1/2}(\theta)m(W_i,\theta,g)$ and the covariances of $D_F^{-1/2}(\theta)m(W_i,\theta,g)$ and $D_F^{-1/2}(\theta)m(W_i,\theta,g^*)$, (iii) $\widehat{h}_{2,n,F}(\theta,g,g^*)$ and $\overline{h}_{2,n,F}(\theta,g)$ are hybrid quantities—part population, part sample—based on $\widehat{\Sigma}_n(\theta,g,g^*)$ and $\overline{\Sigma}_n(\theta,g)$, respectively, and (iv) $\nu_{n,F}(\theta,g)$ is the sample average of $D_F^{-1/2}(\theta)m(W_i,\theta,g)$ normalized to have mean zero and variance that does not depend on n.

Note that $\nu_{n,F}(\theta,\cdot)$ is an empirical process indexed by $g \in \mathcal{G}$ with covariance kernel given by $h_{2,F}(\theta,g,g^*)$.

The normalized sample moments $n^{1/2}\overline{m}_n(\theta, g)$ can be written as

$$n^{1/2}\overline{m}_n(\theta, g) = D_F^{1/2}(\theta)(\nu_{n,F}(\theta, g) + h_{1,n,F}(\theta, g)). \tag{5.3}$$

The test statistic $T_n(\theta)$, defined in (3.6), can be written as

$$T_n(\theta) = \int S(\nu_{n,F}(\theta, g) + h_{1,n,F}(\theta, g), \overline{h}_{2,n,F}(\theta, g))dQ(g). \tag{5.4}$$

Note the close resemblance between $T_n(\theta)$ and $T(h)$ (defined in (4.1)).

Let $\mathcal{H}_1$ denote the set of all functions from $\mathcal{G}$ to $R^p_{+,\infty} \times \{0\}^v$. Let

$$\begin{aligned} \mathcal{H}_2 &= \{h_{2,F}(\theta, \cdot, \cdot) : (\theta, F) \in \mathcal{F}\} \text{ and} \\ \mathcal{H} &= \mathcal{H}_1 \times \mathcal{H}_2. \end{aligned} \tag{5.5}$$

On the space of $k \times k$-matrix-valued covariance kernels on $\mathcal{G} \times \mathcal{G}$, which is a superset of $\mathcal{H}_2$, we use the metric d defined by

$$d(h_2^{(1)}, h_2^{(2)}) = \sup_{g, g^* \in \mathcal{G}} ||h_2^{(1)}(g, g^*) - h_2^{(2)}(g, g^*)||. \tag{5.6}$$

For notational simplicity, for any function of the form $b_F(\theta, g)$ for $g \in \mathcal{G}$, let $b_F(\theta)$ denote the function $b_F(\theta, \cdot)$ on $\mathcal{G}$. Correspondingly, for any function of the form $b_F(\theta, g, g^*)$ for $g, g^* \in \mathcal{G}$, let $b_F(\theta)$ denote the function $b_F(\theta, \cdot, \cdot)$ on $\mathcal{G}^2$.

5.2 Uniform Asymptotic Distribution of the Test Statistic

The following Theorem provides a uniform asymptotic distributional result for the test statistic $T_n(\theta)$. It is used to establish uniform asymptotic coverage probability results for GMS and PA CS's.

Theorem 1. *Suppose Assumptions* M, S1, *and* S2 *hold. Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$, *all constants* $x_{h_{n,F}(\theta)} \in R$ *that may depend on* (θ, F) *and* n *through* $h_{n,F}(\theta)$, *and all* $\delta > 0$, *we have*

(a) $\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} \left[P_F(T_n(\theta) > x_{h_{n,F}(\theta)}) - P(T(h_{n,F}(\theta)) + \delta > x_{h_{n,F}(\theta)})\right] \leq 0,$

(b) $\liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} \left[P_F(T_n(\theta) > x_{h_{n,F}(\theta)}) - P(T(h_{n,F}(\theta)) - \delta > x_{h_{n,F}(\theta)})\right] \geq 0,$

where $T(h) = \int S(\nu_{h_2}(g) + h_1(g), h_2(g) + \varepsilon I_k)dQ(g)$ *and* $\nu_{h_2}(\cdot)$ *is the Gaussian process defined in* (4.2).

Comments. 1. Theorem 1 gives a uniform asymptotic approximation to the distribution function of $T_n(\theta)$. Uniformity holds without *any* restrictions on the normalized mean (i.e., moment inequality slackness) functions $\{h_{1,n,F_n}(\theta_n) : n \geq 1\}$. In particular, Theorem 1 does not require $\{h_{1,n,F_n}(\theta_n) : n \geq 1\}$ to converge as $n \to \infty$ or to belong to a compact set. The Theorem does not require that $T_n(\theta)$ has a unique asymptotic distribution under any sequence $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$. These are novel features of Theorem 1.

2. The supremum and infimum in Theorem 1 are over compact sets of covariance kernels $\mathcal{H}_{2,cpt}$, rather than the parameter space $\mathcal{H}_2$. This is not particularly problematic because the potential asymptotic size problems that arise in moment inequality models are due to the pointwise discontinuity of the asymptotic distribution of the test statistic as a function of the means of the moment inequality functions, not as a function of the covariances between different moment inequalities.

3. Theorem 1 is proved using an almost sure representation argument and the bounded convergence theorem (BCT). The continuous mapping theorem does not apply because (i) $T_n(\theta)$ does not converge in distribution uniformly over $(\theta, F) \in \mathcal{F}$ and (ii) $h_{1,n,F}(\theta, g)$ typically does not converge uniformly over $g \in \mathcal{G}$ even in cases where it has a pointwise limit for all $g \in \mathcal{G}$.

5.3 An Additional GMS Assumption

The following assumption is not needed for GMS CS's to have uniform asymptotic coverage probability greater than or equal to $1 - \alpha$. It is used, however, to show that GMS CS's are not asymptotically conservative. (Note that typically GMS and PA CS's are asymptotically non-similar.) For $(\theta, F) \in \mathcal{F}$ and $j = 1, ..., k$, define $h_{1,\infty,F}(\theta)$ to have jth element equal to ∞ if $E_F m_j(W_i, \theta, g) > 0$ and $j \leq p$ and 0 otherwise. Let $h_{\infty,F}(\theta) = (h_{1,\infty,F}(\theta), h_{2,F}(\theta))$.

Assumption GMS2. (a) For some $(\theta_c, F_c) \in \mathcal{F}$, the distribution function of $T(h_{\infty,F_c}(\theta_c))$ is continuous and strictly increasing at its $1-\alpha$ quantile plus δ, viz., $c_0(h_{\infty,F_c}(\theta_c), 1-\alpha) + \delta$, for all $\delta > 0$ sufficiently small and $\delta = 0$,

(b) $B_n \to \infty$ as $n \to \infty$, and

(c) $n^{1/2}/\kappa_n \to \infty$ as $n \to \infty$.

Assumption GMS2(a) is not restrictive. For example, it holds for typical choices of S and Q for any (θ_c, F_c) for which $Q(\{g \in \mathcal{G} : h_{1,\infty,F_c}(\theta_c, g) = 0\}) > 0$. Assumption GMS2(c) is satisfied by typical choices of κ_n, such as $\kappa_n = (0.3 \ln n)^{1/2}$.

5.4 Uniform Asymptotic Coverage Probability Results

The following Theorem gives uniform asymptotic coverage probability results for GMS and PA CS's.

Theorem 2. *Suppose Assumptions* M, S1, *and* S2 *hold and Assumption* GMS1 *also holds when considering GMS CS's. Then, for every compact subset $\mathcal{H}_{2,cpt}$ of $\mathcal{H}_2$, GMS and PA confidence sets CS_n satisfy*

(a) $\liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F(\theta \in CS_n) \geq 1-\alpha$ *and*

(b) *if Assumption* GMS2 *also holds and $h_{2,F_c}(\theta_c) \in \mathcal{H}_{2,cpt}$ (for $(\theta_c, F_c) \in \mathcal{F}$ as in Assumption* GMS2), *then the GMS confidence set satisfies*

$$\lim_{\eta\to 0} \liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F(\theta \in CS_n) = 1-\alpha,$$

where η is as in the definition of $c(h, 1-\alpha)$.

Comments. 1. Theorem 2(a) shows that GMS and PA CS's have correct uniform asymptotic size over compact sets of covariance kernels. Theorem 2(b) shows that GMS CS's are at most infinitesimally conservative asymptotically. The uniformity results hold whether the moment conditions involve "weak" or "strong" instrumental variables.

2. An analogue of Theorem 2(b) holds for PA CS's if Assumption GMS2(a) holds and $E_{F_c}(m_j(W_i, \theta_c)|X_i) = 0$ a.s. for $j \leq p$ (i.e., if the conditional moment inequalities hold as equalities a.s.) under some $(\theta_c, F_c) \in \mathcal{F}$.[13] However, the latter condition is restrictive—it fails in many applications.

3. Theorem 2 applies to CvM tests based on integrals with respect to a probability measure Q. Extensions to A-CvM and KS tests are given in Appendix B of AS.

4. Comments 1 and 2 to Theorem 1 also apply to Theorem 2.

6 Power Against Fixed Alternatives

We now show that the power of GMS and PA tests converges to one as $n \to \infty$ for all fixed alternatives (for which the moment functions have $2+\delta$ moments finite). Thus, both tests are consistent tests. This implies that for any fixed distribution F_0 and any parameter value θ_* *not* in the identified set Θ_{F_0}, the GMS and PA CS's do not include θ_* with probability approaching one. In this sense, GMS and PA CS's based on $T_n(\theta)$ fully exploit the conditional moment inequalities and equalities. CS's based on a finite number of unconditional moment inequalities and equalities do not have this property.

The null hypothesis is

$$\begin{aligned} H_0 : E_{F_0}(m_j(W_i, \theta_*)|X_i) &\geq 0 \text{ a.s. } [F_{X,0}] \text{ for } j = 1, ..., p \text{ and} \\ E_{F_0}(m_j(W_i, \theta_*)|X_i) &= 0 \text{ a.s. } [F_{X,0}] \text{ for } j = p+1, ..., k, \end{aligned} \tag{6.1}$$

where θ_* denotes the null parameter value and F_0 denotes the fixed true distribution of the data. The alternative is $H_1 : H_0$ does not hold. The following assumption specifies the properties of fixed alternatives (FA).

Assumption FA. The value $\theta_* \in \Theta$ and the true distribution F_0 satisfy: (a) $P_{F_0}(X_i \in \mathcal{X}_{F_0}(\theta_*)) > 0$, where $\mathcal{X}_{F_0}(\theta_*)$ is defined in (3.11), (b) $\{W_i : i \geq 1\}$ are i.i.d. under F_0, (c) $Var_{F_0}(m_j(W_i, \theta_*)) > 0$ for $j = 1, ..., k$, (d) $E_{F_0}||m(W_i, \theta_*)||^{2+\delta} < \infty$ for some $\delta > 0$, and (e) Assumption M holds with F_0 in place of F and F_n in Assumptions M(b) and M(c), respectively.

Assumption FA(a) states that violations of the conditional moment inequalities or equalities occur for the null parameter θ_* for X_i values in some set with positive probability under F_0. Thus, under Assumption FA(a), the moment conditions specified in (6.1) do not hold. Assumptions FA(b)-(d) are standard i.i.d. and moment assumptions. As-

[13] The proof follows easily from results given in the proof of Theorem 2(b).

sumption FA(e) holds for $\mathcal{G}_{c\text{-}cube}$ and $\mathcal{G}_{box}$ because $\mathcal{C}_{c\text{-}cube}$ and $\mathcal{C}_{box}$ are Vapnik-Cervonenkis classes of sets.

For $g \in \mathcal{G}$, define

$$\begin{aligned} m_j^*(g) &= E_{F_0} m_j(W_i, \theta_*) g_j(X_i) / \sigma_{F_0,j}(\theta_*) \text{ and} \\ \beta(g) &= \max\{-m_1^*(g), ..., -m_p^*(g), |m_{p+1}^*(g)|, ..., |m_k^*(g)|\}. \end{aligned} \tag{6.2}$$

Under Assumptions FA(a) and CI, $\beta(g_0) > 0$ for some $g_0 \in \mathcal{G}$.

For a test based on $T_n(\theta)$ to have power against all fixed alternatives, the weighting function Q cannot "ignore" any elements $g \in \mathcal{G}$, because such elements may have identifying power for the identified set. This requirement is captured in the following assumption, which is shown in Lemma 4 to hold for the two probability measures Q considered in Section 3.4.

Let $F_{X,0}$ denote the distribution of X_i under F_0. Define the pseudo-metric ρ_X on $\mathcal{G}$ by

$$\rho_X(g, g^*) = (E_{F_{X,0}} ||g(X_i) - g^*(X_i)||^2)^{1/2} \text{ for } g, g^* \in \mathcal{G}. \tag{6.3}$$

Let $\mathcal{B}_{\rho_X}(g, \delta)$ denote an open ρ_X-ball in $\mathcal{G}$ centered at g with radius δ.

Assumption Q. The support of Q under the pseudo-metric ρ_X is $\mathcal{G}$. That is, for all $\delta > 0$, $Q(\mathcal{B}_{\rho_X}(g, \delta)) > 0$ for all $g \in \mathcal{G}$.

The next result establishes Assumption Q for the probability measures Q on $\mathcal{G}_{c\text{-}cube}$ and $\mathcal{G}_{box}$ discussed in Section 3.4 above. Appendix B of AS provides analogous results for three other choices of Q and $\mathcal{G}$.

Lemma 4. *Assumption* Q *holds for the weight functions*:

(a) $Q_a = \Pi_{c\text{-}cube}^{-1} Q_{AR}$ *on* $\mathcal{G}_{c\text{-}cube}$, *where* Q_{AR} *is uniform on* $a \in \{1, ..., 2r\}^{d_x}$ *conditional on* r *and* r *has some probability mass function* $\{w(r) : r = r_0, r_0 + 1, ...\}$ *with* $w(r) > 0$ *for all* r *and*

(b) $Q_b = \Pi_{box}^{-1} Unif([0,1]^{d_x} \times (0, \bar{r})^{d_x})$ *on* $\mathcal{G}_{box}$ *with the centers of the boxes in* $[0,1]^{d_x}$.

Comment. The uniform distribution that appears in both specifications of Q in the Lemma could be replaced by another distribution and the results of the Lemma still hold provided the other distribution has the same support.

The following Theorem shows that GMS and PA tests are consistent against all fixed alternatives.

Theorem 3. *Under Assumptions* FA, CI, Q, S1, S3, *and* S4,

(a) $\lim_{n\to\infty} P_{F_0}(T_n(\theta_*) > c(\varphi_n(\theta_*), \widehat{h}_{2,n}(\theta_*), 1-\alpha)) = 1$ *and*

(b) $\lim_{n\to\infty} P_{F_0}(T_n(\theta_*) > c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_*), 1-\alpha)) = 1.$

Comment. Theorem 3 implies the following for GMS and PA CS's: Suppose $(\theta_0, F_0) \in \mathcal{F}$ for some $\theta_0 \in \Theta$, θ_* ($\in \Theta$) is not in the identified set Θ_{F_0} (defined in (2.2)), and Assumptions FA(c), FA(d), CI, M, S1, S3, and S4 hold, then for GMS and PA CS's we have

$$\lim_{n\to\infty} P_{F_0}(\theta_* \in CS_n) = 0.^{14} \tag{6.4}$$

7 Power Against $\mathbf{n^{-1/2}}$-Local Alternatives

In this section, we show that GMS and PA tests have power against certain, but not all, $n^{-1/2}$-local alternatives. This holds even though these tests fully exploit the information in the conditional moment restrictions, which is of an infinite-dimensional nature. These testing results have immediate implications for the volume of CS's, see Pratt (1961).

We show that a GMS test has asymptotic power that is greater than or equal to that of a PA test (based on the same test statistic) under all alternatives with strict inequality in certain scenarios. Although we do not do so here, arguments analogous to those in Andrews and Soares (2010) could be used to show that a GMS test's power is greater than or equal to that of a subsampling test with strictly greater power in certain scenarios.

For given $\theta_{n,*} \in \Theta$ for $n \geq 1$, we consider tests of

$$\begin{aligned} H_0 : E_{F_n} m_j(W_i, \theta_{n,*}) &\geq 0 \text{ for } j = 1, ..., p, \\ E_{F_n} m_j(W_i, \theta_{n,*}) &= 0 \text{ for } j = p+1, ..., k, \end{aligned} \tag{7.1}$$

and $(\theta_{n,*}, F_n) \in \mathcal{F}$, where F_n denotes the true distribution of the data. The null values $\theta_{n,*}$ are allowed to drift with n or be fixed for all n. Drifting $\theta_{n,*}$ values are of interest because they allow one to consider the case of a fixed identified set, say Θ_0, and to derive the asymptotic probability that parameter values $\theta_{n,*}$ that are not in the identified set, but drift toward it at rate $n^{-1/2}$, are excluded from a GMS or PA CS. In this scenario,

[14] This holds because $\theta_* \notin \Theta_{F_0}$ implies Assumption FA(a) holds, $(\theta_0, F_0) \in \mathcal{F}$ implies Assumption FA(b) holds, and Assumption M with $F = F_n = F_0$ implies Assumption FA(e) holds.

the sequence of true distributions are ones that yield Θ_0 to be the identified set, i.e., $F_n \in \mathcal{F}_0 = \{F : \Theta_F = \Theta_0\}$.

The true parameters and distributions are denoted (θ_n, F_n). We consider the Kolmogorov-Smirnov metric on the space of distributions F.

The $n^{-1/2}$-local alternatives are defined as follows.

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

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

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

(c) $d(h_{2,F_n}(\theta_n), h_{2,F_0}(\theta_0)) \to 0$ and $d(h_{2,F_n}(\theta_{n,*}), h_{2,F_0}(\theta_0)) \to 0$ as $n \to \infty$ (where d is defined in (5.6)),

(d) $Var_{F_n}(m_j(W_i, \theta_{n,*})) > 0$ for $j = 1, ..., k$, for $n \geq 1$, and

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

Assumption LA2. The $k \times d$ matrix $\Pi_F(\theta, g) = (\partial/\partial\theta')[D_F^{-1/2}(\theta) E_F m(W_i, \theta, g)]$ exists and is continuous in (θ, F) for all (θ, F) in a neighborhood of (θ_0, F_0) for all $g \in \mathcal{G}$.

For notational simplicity, we let h_2 abbreviate $h_{2,F_0}(\theta_0)$ throughout this section. Assumption LA1(a) states that the true values $\{\theta_n : n \geq 1\}$ are $n^{-1/2}$-local to the null values $\{\theta_{n,*} : n \geq 1\}$. Assumption LA1(b) specifies the asymptotic behavior of the (normalized) moment inequality functions when evaluated at the true values $\{\theta_n : n \geq 1\}$. Under the true values, these (normalized) moment inequality functions are non-negative. Assumption LA1(c) specifies the asymptotic behavior of the covariance kernels $\{h_{2,F_n}(\theta_n, \cdot, \cdot) : n \geq 1\}$ and $\{h_{2,F_n}(\theta_{n,*}, \cdot, \cdot) : n \geq 1\}$. Assumptions LA1(d) and LA1(e) are standard. Assumption LA2 is a smoothness condition on the normalized expected moment functions. Given the smoothing properties of the expectation operator, this condition is not restrictive.

Under Assumptions LA1 and LA2, we show that the moment inequality functions evaluated at the null values $\{\theta_{n,*} : n \geq 1\}$ satisfy:

$$\begin{aligned} &\lim_{n \to \infty} n^{1/2} D_{F_n}^{-1/2}(\theta_{n,*}) E_{F_n} m(W_i, \theta_{n,*}, g) = h_1(g) + \Pi_0(g)\lambda \in R^k, \text{ where} \\ &h_1(g) = (h_{1,1}(g), ..., h_{1,p}(g), 0, ..., 0)' \in R^k \text{ and } \Pi_0(g) = \Pi_{F_0}(\theta_0, g). \end{aligned} \tag{7.2}$$

If $h_{1,j}(g) = \infty$, then by definition $h_{1,j}(g) + y = \infty$ for any $y \in R$. We have $h_1(g) + \Pi_0(g)\lambda \in R^p_{[+\infty]} \times R^v$. Let $\Pi_{0,j}(g)$ denote the jth row of $\Pi_0(g)$ written as a column d_θ-vector for $j = 1, ..., k$.

The null hypothesis, defined in (7.1), does not hold (at least for n large) when the following assumption holds.

Assumption LA3. For some $g \in \mathcal{G}$, $h_{1,j}(g) + \Pi_{0,j}(g)'\lambda < 0$ for some $j = 1, ..., p$ or $\Pi_{0,j}(g)'\lambda \neq 0$ for some $j = p+1, ..., k$.

Under the following assumption, if $\lambda = \beta\lambda_0$ for some $\beta > 0$ and some $\lambda_0 \in R^{d_\theta}$, then the power of GMS and PA tests against the perturbation λ is arbitrarily close to one for β arbitrarily large:

Assumption LA3$'$. $Q(\{g \in \mathcal{G} : h_{1,j}(g) < \infty$ and $\Pi_{0,j}(g)'\lambda_0 < 0$ for some $j = 1, ..., p$ or $\Pi_{0,j}(g)'\lambda_0 \neq 0$ for some $j = p+1, ..., k\}) > 0$.

Assumption LA3$'$ requires that either (i) the moment equalities detect violations of the null hypothesis for g functions in a set with positive Q measure or (ii) the moment inequalities are not too far from being binding, i.e., $h_{1,j}(g) < \infty$, and the perturbation λ_0 occurs in a direction that yields moment inequality violations for g functions in a set with positive Q measure.

Assumption LA3 is employed with the KS test. It is weaker than Assumption LA3$'$. It is shown in Appendix B of AS that if Assumption LA3 holds with $\lambda = \beta\lambda_0$ (and some other assumptions), then the power of KS-GMS and KS-PA tests against the perturbation λ is arbitrarily close to one for β arbitrarily large.

In Appendix B of AS we illustrate the verification of Assumptions LA1-LA3 and LA3$'$ in a simple example. In this example, $v = 0$, $h_{1,j}(g) < \infty\ \forall g \in \mathcal{G}$, and $\Pi_{0,j}(g) = -Eg(X_i)$ $\forall g \in \mathcal{G}$, so $\Pi_{0,j}(g)'\lambda_0 < 0$ in Assumption LA3$'$ $\forall g \in \mathcal{G}$ with $Eg(X_i) > 0$ for all $\lambda_0 > 0$.

Assumptions LA3 and LA3$'$ can fail to hold even when the null hypothesis is violated. This typically happens if the true parameter/true distribution is fixed, i.e., $(\theta_n, F_n) = (\theta_0, F_0) \in \mathcal{F}$ for all n in Assumption LA1(a), the null hypothesis parameter $\theta_{n,*}$ drifts with n as in Assumption LA1(a), and $P_{F_0}(X_i \in \mathcal{X}_{zero}) = 0$, where $\mathcal{X}_{zero} = \{x \in R^{d_x} : E_{F_0}(m(W_i, \theta_0)|X_i = x) = 0\}$. In such cases, typically $h_{1,j}(g) = \infty\ \forall g \in \mathcal{G}$ (because the conditional moment inequalities are non-binding with probability one), Assumptions LA3 and LA3$'$ fail, and Theorem 4 below shows that GMS and PA tests have trivial asymptotic power against such $n^{-1/2}$-local alternatives. For example, this occurs in the example of Section 12.5 in Appendix B of AS when $P_{F_0}(X_i \in \mathcal{X}_{zero}) = 0$.

As discussed in Section 12.5, asymptotic results based on a fixed true distribution provide poor approximations when $P_{F_0}(X_i \in \mathcal{X}_{zero}) = 0$. Hence, one can argue that it makes sense to consider local alternatives for sequences of true distributions $\{F_n : n \geq 1\}$ for which $h_{1,j}(g) < \infty$ for a non-negligible set of $g \in \mathcal{G}$, as in Assumption LA3$'$, because such sequences are the ones for which the asymptotics provide good finite-sample approximations. For such sequences, GMS and PA tests have non-trivial power against $n^{-1/2}$-local alternatives, as shown in Theorem 4 below.

Nevertheless, local-alternative power results can be used for multiple purposes and for some purposes, one may want to consider local-alternatives other than those that satisfy Assumptions LA3 and LA3$'$.

The asymptotic distribution of $T_n(\theta_{n,*})$ under $n^{-1/2}$-local alternatives is shown to be $J_{h,\lambda}$. By definition, $J_{h,\lambda}$ is the distribution of

$$T(h_1 + \Pi_0\lambda, h_2) = \int S(\nu_{h_2}(g) + h_1(g) + \Pi_0(g)\lambda, h_2(g) + \varepsilon I_k) dQ(g), \qquad (7.3)$$

where $h = (h_1, h_2)$, Π_0 denotes $\Pi_0(\cdot)$, and $\nu_{h_2}(\cdot)$ is a mean zero Gaussian process with covariance kernel $h_2 = h_{2,F_0}(\theta_0)$. For notational simplicity, the dependence of $J_{h,\lambda}$ on Π_0 is suppressed.

Next, we introduce two assumptions, viz., Assumptions LA4 and LA5, that are used only for GMS tests in the context of local alternatives. The population analogues of $\overline{\Sigma}_n(\theta, g)$ and its diagonal matrix are

$$\overline{\Sigma}_F(\theta, g) = \Sigma_F(\theta, g, g) + \varepsilon\Sigma_F(\theta, 1_k, 1_k) \text{ and } \overline{D}_F(\theta, g) = Diag(\overline{\Sigma}_F(\theta, g)), \qquad (7.4)$$

where $\Sigma_F(\theta, g, g)$ is defined in (5.1). Let $\overline{\sigma}_{F,j}(\theta, g)$ denote the square-root of the (j, j) element of $\overline{\Sigma}_F(\theta, g)$.

Assumption LA4. $\kappa_n^{-1} n^{1/2} E_{F_n} m_j(W_i, \theta_n, g)/\overline{\sigma}_{F_n,j}(\theta_n, g) \to \pi_{1,j}(g)$ for some $\pi_{1,j}(g) \in R_{+,\infty}$ for $j = 1, ..., p$ and $g \in \mathcal{G}$.

In Assumption LA4 the functions are evaluated at the true value θ_n, not at the null value $\theta_{n,*}$, and $(\theta_n, F_n) \in \mathcal{F}$. In consequence, the moment functions in Assumption LA4 satisfy the moment inequalities and $\pi_{1,j}(g) \geq 0$ for all $j = 1, ..., p$ and $g \in \mathcal{G}$. Note that $0 \leq \pi_{1,j}(g) \leq h_{1,j}(g)$ for all $j = 1, ..., p$ and all $g \in \mathcal{G}$ (by Assumption LA1(b) and $\kappa_n \to \infty$.)

Let $\pi_1(g) = (\pi_{1,1}(g), ..., \pi_{1,p}(g), 0, ..., 0)' \in R^p_{+,\infty} \times \{0\}^v$. Let $c_0(\varphi(\pi_1), h_2, 1-\alpha)$ denote the $1-\alpha$ quantile of

$$T(\varphi(\pi_1), h_2) = \int S(\nu_{h_2}(g) + \varphi(\pi_1(g)), h_2(g) + \varepsilon I_k) dQ(g), \text{ where}$$
$$\varphi(\pi_1(g)) = (\varphi(\pi_{1,1}(g)), ..., \varphi(\pi_{1,p}(g)), 0, ..., 0)' \in R^k \text{ and}$$
$$\varphi(x) = 0 \text{ if } x \leq 1 \text{ and } \varphi(x) = \infty \text{ if } x > 1. \tag{7.5}$$

Let $\varphi(\pi_1)$ denote $\varphi(\pi_1(\cdot))$. The probability limit of the GMS critical value $c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)$ is shown below to be $c(\varphi(\pi_1), h_2, 1-\alpha) = c_0(\varphi(\pi_1), h_2, 1-\alpha+\eta) + \eta$.

Assumption LA5. (a) $Q(\mathcal{G}_\varphi) = 1$, where $\mathcal{G}_\varphi = \{g \in \mathcal{G} : \pi_{1,j}(g) \neq 1 \text{ for } j = 1, ..., p\}$, and

(b) the distribution function (df) of $T(\varphi(\pi_1), h_2)$ is continuous and strictly increasing at $x = c(\varphi(\pi_1),\, h_2, 1-\alpha)$, where $h_2 = h_{2,F_0}(\theta_0)$.

The value 1 that appears in $\mathcal{G}_\varphi$ in Assumption LA5(a) is the discontinuity point of φ. Assumption LA5(a) implies that the $n^{-1/2}$-local power formulae given below do not apply to certain "discontinuity vectors" $\pi_1(\cdot)$, but this is not particularly restrictive.[15] Assumption LA5(b) typically holds because of the absolute continuity of the Gaussian random variables $\nu_{h_2}(g)$ that enter $T(\varphi(\pi_1), h_2)$.[16]

The following assumption is used only for PA tests.

Assumption LA6. The df of $T(0_\mathcal{G}, h_2)$ is continuous and strictly increasing at $x = c(0_\mathcal{G}, h_2, 1-\alpha)$, where $h_2 = h_{2,F_0}(\theta_0)$.

The probability limit of the PA critical value is shown to be $c(0_\mathcal{G}, h_2, 1-\alpha) = c_0(0_\mathcal{G}, h_2, 1-\alpha+\eta)+\eta$, where $c_0(0_\mathcal{G}, h_2, 1-\alpha)$ denotes the $1-\alpha$ quantile of $J_{(0_\mathcal{G},h_2),0_{d_\theta}}$.

Theorem 4. *Under Assumptions* M, S1, S2, *and* LA1-LA2,

(a) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(\varphi_n(\theta_{n,*}), \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)) = 1 - J_{h,\lambda}(c(\varphi(\pi_1),\, h_2, 1-\alpha))$ *provided Assumptions* GMS1, LA4, *and* LA5 *also hold,*

[15] Assumption LA5(a) is not particularly restrictive because in cases where it fails, one can obtain lower and upper bounds on the local asymptotic power of GMS tests by replacing $c(\varphi(\pi_1), h_2, 1-\alpha)$ by $c(\varphi(\pi_1-), h_2, 1-\alpha)$ and $c(\varphi(\pi_1+), h_2, 1-\alpha)$, respectively, in Theorem 4(a). By definition, $\varphi(\pi_1-) = \varphi(\pi_1(\cdot)-)$ and $\varphi(\pi_1(g)-)$ is the limit from the left of $\varphi(x)$ at $x = \pi_1(g)$. Likewise $\varphi(\pi_1+) = \varphi(\pi_1(\cdot)+)$ and $\varphi(\pi_1(g)+)$ is the limit from the right of $\varphi(x)$ at $x = \pi_1(g)$.

[16] If Assumption LA5(b) fails, one can obtain lower and upper bounds on the local asymptotic power of GMS tests by replacing $J_{h,\lambda}(c(\varphi(\pi_1),\, h_2, 1-\alpha))$ by $J_{h,\lambda}(c(\varphi(\pi_1),\, h_2, 1-\alpha)+)$ and $J_{h,\lambda}(c(\varphi(\pi_1),\, h_2, 1-\alpha)-)$, respectively, in Theorem 4(a), where the latter are the limits from the left and right, respectively, of $J_{h,\lambda}(x)$ at $x = c(\varphi(\pi_1), h_2, 1-\alpha)$.

(b) $\lim_{n\to\infty} P_{F_n}(T_n(\theta_{n,*}) > c(0_{\mathcal{G}}, \widehat{h}_{2,n}(\theta_{n,*}), 1-\alpha)) = 1 - J_{h,\lambda}(c(0_{\mathcal{G}}, h_2, 1-\alpha))$ *provided Assumption* LA6 *also holds, and*

(c) $\lim_{\beta\to\infty}[1 - J_{h,\beta\lambda_0}(c(\varphi(\pi_1), h_2, 1-\alpha))] = \lim_{\beta\to\infty}[1 - J_{h,\beta\lambda_0}(c(0_{\mathcal{G}}, h_2, 1-\alpha))] = 1$ *provided Assumptions* LA3$'$*,* S3*, and* S4 *hold.*

Comments. 1. Theorem 4(a) and 4(b) provide the $n^{-1/2}$-local alternative power function of the GMS and PA tests, respectively. Theorem 4(c) shows that the asymptotic power of GMS and PA tests is arbitrarily close to one if the $n^{-1/2}$-local alternative parameter $\lambda = \beta\lambda_0$ is sufficiently large in the sense that its scale β is large.

2. We have $c(\varphi(\pi_1), h_2, 1-\alpha) \le c(0_{\mathcal{G}}, h_2, 1-\alpha)$ (because $\varphi(\pi_1(g)) \ge 0$ for all $g \in \mathcal{G}$ and $S(m, \Sigma)$ is non-increasing in m_I by Assumption S1(b), where $m = (m_I', m_{II}')'$). Hence, the asymptotic local power of a GMS test is greater than or equal to that of a PA test. Strict inequality holds whenever $\pi_1(\cdot)$ is such that $Q(\{g \in \mathcal{G} : \varphi(\pi_1(g)) > 0\}) > 0$. The latter typically occurs whenever the conditional moment inequality $E_{F_n}(m_j(W_i, \theta_{n,*}) |X_i)$ for some $j = 1, ..., p$ is bounded away from zero as $n \to \infty$ with positive X_i probability.

3. The results of Theorem 4 hold under the null hypothesis as well as under the alternative. The results under the null quantify the degree of asymptotic non-similarity of the GMS and PA tests.

4. Suppose the assumptions of Theorem 4 hold and each distribution F_n generates the same identified set, call it $\Theta_0 = \Theta_{F_n}$ $\forall n \ge 1$. Then, Theorem 4(a) implies that the asymptotic probability that a GMS CS includes, $\theta_{n,*}$, which lies within $O(n^{-1/2})$ of the identified set, is $J_{h,\lambda}(c(\varphi(\pi_1),\ h_2, 1-\alpha))$. If $\lambda = \beta\lambda_0$ and Assumptions LA3$'$, S3, and S4 also hold, then $\theta_{n,*}$ is not in Θ_0 (at least for β large) and the asymptotic probability that a GMS or PA CS includes $\theta_{n,*}$ is arbitrarily close to zero for β arbitrarily large by Theorem 4(c). Analogous results hold for PA CS's.

8 Preliminary Consistent Estimation of Identified Parameters and Time Series

In this section, we consider the case in which the moment functions in (2.4) depend on a parameter τ as well as θ and a preliminary consistent estimator, $\widehat{\tau}_n(\theta)$, of τ is available when θ is the true value. (This requires that τ is identified given the true value θ.) For example, this situation often arises with game theory models, as in the third model considered in Section 9 below. The parameter τ may be finite dimensional or infinite dimensional. As pointed out to us by A. Aradillas-López, infinite-dimensional parameters arise as expectations functions in some game theory models. Later in the section, we also consider the case where $\{W_i : i \leq n\}$ are time series observations.

Suppose the moment functions are of the form $m_j(W_i, \theta, \tau)$ and the model specifies that (2.1) holds with $m_j(W_i, \theta, \tau_F(\theta))$ in place of $m_j(W_i, \theta)$ for $j \leq k$ for some $\tau_F(\theta)$ that may depend on θ and F.

The normalized sample moment functions are of the form

$$n^{1/2}\overline{m}_n(\theta, g) = n^{-1/2} \sum_{i=1}^{n} m(W_i, \theta, \widehat{\tau}_n(\theta), g). \tag{8.1}$$

In the infinite-dimensional case, $m(W_i, \theta, \widehat{\tau}_n(\theta), g)$ can be of the form $m^*(W_i, \theta, \widehat{\tau}_n(W_i, \theta)$ $, g)$, where $\widehat{\tau}_n(W_i, \theta) : R^{d_w} \times \Theta \rightarrow R^{d_\tau}$ for some $d_\tau < \infty$.

Given (8.1), the quantity $\Sigma_F(\theta, g, g^*)$ in (5.1) denotes the asymptotic covariance of $n^{1/2}\overline{m}_n(\theta, \widehat{\tau}_n(\theta), g)$ and $n^{1/2}\overline{m}_n(\theta, \widehat{\tau}_n(\theta), g^*)$ under (θ, F), rather than $Cov_F(m(W_i, \theta, g),$ $m(W_i, \theta, g^*))$. Correspondingly, $\widehat{\Sigma}_n(\theta, g, g^*)$ is not defined by (4.5) but is taken to be an estimator of $\Sigma_F(\theta, g, g^*)$ that is consistent under (θ, F). With these adjusted definitions of $\overline{m}_n(\theta, g)$ and $\widehat{\Sigma}_n(\theta, g, g^*)$, the test statistic $T_n(\theta)$ and GMS or PA critical value $c_{n,1-\alpha}(\theta)$ are defined in the same way as above.[17]

For example, when τ is finite dimensional, the preliminary estimator $\widehat{\tau}_n(\theta)$ is chosen to satisfy:

$$n^{1/2}(\widehat{\tau}_n(\theta) - \tau_F(\theta)) \rightarrow_d Z_F \text{ as } n \rightarrow \infty \text{ under } (\theta, F) \in \mathcal{F}, \tag{8.2}$$

for some normally distributed random vector Z_F with mean zero.

The normalized sample moments can be written as

$$\begin{aligned} n^{1/2}\overline{m}_n(\theta, g) &= D_F^{1/2}(\theta)(\nu_{n,F}(\theta, g) + h_{1,n,F}(\theta, g)), \text{ where} \\ \nu_{n,F}(\theta, g) &= n^{-1/2} \sum_{i=1}^{n} D_F^{-1/2}(\theta)[m(W_i, \theta, \widehat{\tau}_n(\theta), g) - E_F m(W_i, \theta, \tau_F(\theta), g)], \\ h_{1,n,F}(\theta, g) &= n^{1/2} D_F^{-1/2}(\theta) E_F m(W_i, \theta, \tau_F(\theta), g). \end{aligned} \tag{8.3}$$

In place of Assumption M, we use the following empirical process (EP) assumption.

[17] When computing bootstrap critical values, one needs to bootstrap the estimator $\widehat{\tau}_n(\theta)$ as well as the observations $\{W_i : i \leq n\}$.

Let $\Rightarrow$ denote weak convergence. Let $\{a_n : n \geq 1\}$ denote a subsequence of $\{n\}$.

Assumption EP. (a) For some specification of the parameter space $\mathcal{F}$ that imposes the conditional moment inequalities and equalities and all $(\theta, F) \in \mathcal{F}$, $\nu_{n,F}(\theta, \cdot) \Rightarrow \nu_{h_{2,F}(\theta)}(\cdot)$ as $n \to \infty$ under (θ, F), for some mean zero Gaussian process $\nu_{h_{2,F}(\theta)}(\cdot)$ on $\mathcal{G}$ with covariance kernel $h_{2,F}(\theta)$ on $\mathcal{G} \times \mathcal{G}$ and bounded uniformly ρ-continuous sample paths a.s. for some pseudo-metric ρ on $\mathcal{G}$.

(b) For any subsequence $\{(\theta_{a_n}, F_{a_n}) \in \mathcal{F} : n \geq 1\}$ for which $\lim_{n\to\infty} \sup_{g,g^* \in \mathcal{G}} ||h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*)|| = 0$ for some $k \times k$ matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$, we have (i) $\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot) \Rightarrow \nu_{h_2}(\cdot)$ and (ii) $\sup_{g,g^* \in \mathcal{G}} ||\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*)|| \to_p 0$ as $n \to \infty$.

The quantity $\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g, g^*)$ is defined as in previous sections but with $\widehat{\Sigma}_n(\theta, g, g^*)$ and $\Sigma_F(\theta, g, g^*)$ defined as in this section.

With Assumption EP in place of Assumption M, the results of Theorem 2 hold when the GMS or PA CS depends on a preliminary estimator $\widehat{\tau}_n(\theta)$.[18] (The proof is the same as that given for Theorem 2 in Appendices A and C with Assumption EP replacing the results of Lemma A1.)

Next, we consider time series observations $\{W_i : i \leq n\}$. Let the moment conditions and sample moments be defined as in (2.3) and (3.3), but do not impose the definitions of $\mathcal{F}$ and $\widehat{\Sigma}_n(\theta, g)$ in (2.3) and (3.4). Instead, define $\widehat{\Sigma}_n(\theta, g)$ in a way that is suitable for the temporal dependence properties of $\{m(W_i, \theta, g) : i \leq n\}$. For example, $\widehat{\Sigma}_n(\theta, g)$ might need to be defined to be a heteroskedasticity and autocorrelation consistent (HAC) variance estimator. Or, if $\{m(W_i, \theta) : i \leq n\}$ have zero autocorrelations under (θ, F), define $\widehat{\Sigma}_n(\theta, g)$ as in (3.4). Given these definitions of $\overline{m}_n(\theta, g)$ and $\widehat{\Sigma}_n(\theta, g)$, define the test statistic $T_n(\theta)$ and GMS or PA critical value $c_{n,1-\alpha}(\theta)$ as in previous sections.[19]

Define $\nu_{n,F}(\theta, g)$ as in (5.2). Now, with Assumption EP in place of Assumption M, the results of Theorem 2 hold with time series observations.

Note that Assumption EP also can be used when the observations are independent but not identically distributed.

[18] Equation (8.2) is only needed for this result in order to verify Assumption EP(a) in the finite-dimensional τ case.

[19] With bootstrap critical values, the bootstrap employed needs to take account of the time series structure of the observations. For example, a block bootstrap does so.

9 Monte Carlo Simulations

9.1 Description of the Tests

In this section, we provide simulation evidence concerning the finite sample properties of the tests introduced in this paper. We consider three models: a quantile selection model, an interval outcome regression model, and an entry game model with multiple equilibria.

We compare different test statistics and critical values in terms of their coverage probabilities (CP's) for points in the identified set and their false coverage probabilities (FCP's) for points outside the identified set. Obviously, one wants FCP's to be as small as possible. FCP's are directly related to the power of the tests used to constructed the CS and are related to the volume of the CS, see Pratt (1961).

The following test statistics are considered: (i) CvM/Sum, (ii) CvM/QLR, (iii) CvM/Max, (iv) KS/Sum, (v) KS/QLR, and (vi) KS/Max. In all three models countable hypercubes and truncated versions of the test statistics are employed. (More details are given below.) The weights $\{w(r) : r = r_0, ...\}$ employed by the CvM statistics, see (3.14), are proportional to $(r^2 + 100)^{-1}$ for a cube with side-edge length indexed by r, for $r = r_0, \ldots$. The number of boxes with side-edge length indexed by r is $(2r)^{d_X}$, where d_X denotes the dimension of the covariate X_i. The weights are normalized to sum to one, but this does not affect the results.

In all three models we consider the PA/Asy and GMS/Asy critical values. In the first two models we also consider the PA/Bt, GMS/Bt, and Sub critical values. The critical values are simulated using 5001 repetitions (for each original sample repetition).[20] The GMS critical value is based on $\kappa_{n,bc} = (0.3\ln(n))^{1/2}$, $B_{n,bc} = (0.4\ln(n)/\ln\ln(n))^{1/2}$, and $\varepsilon = 5/100$, where bc abbreviates "basecase." The same basecase values $\kappa_{n,bc}$, $B_{n,bc}$, and ε are used in all three models. Additional results are reported for variations of these values. The subsample size is 20 when the sample size is 250. Results are reported for nominal 95% CS's. The number of simulation repetitions used to compute CP's and FCP's is 5000 for all cases. This yields a simulation standard error of .0031.

The reported FCP's are "CP corrected" by employing a critical value that yields a

[20] The Sum, QLR, and Max statistics use the functions S_1, S_2, and S_3, respectively. The PA/Asy and PA/Bt critical values are based on the asymptotic distribution and bootstrap, respectively, and likewise for the GMS/Asy and GMS/Bt critical values. The IUF η is set to 0 because its value, provided it is sufficiently small, has no effect in these models. Sub denotes a (non-recentered) subsampling critical value. The number of subsamples considered is 5001. They are drawn randomly without replacement.

CP equal to .95 at the closest point of the identified set if the CP at the closest point is less than .95. If the CP at the closest point is greater than .95, then no CP correction is carried out. The reason for this "asymmetric" CP correction is that CS's may have CP's greater than .95 for points in the identified set, even asymptotically, in the present context and one does not want to reward over-coverage of points in the identified set by CP correcting the critical values when making comparisons of FCP's.

9.2 Quantile Selection Model

9.2.1 Description of the Model

In this model we are interested in the conditional τ-quantile of a treatment response given the value of a covariate X_i. The results also apply to conditional quantiles of arbitrary responses that are subject to selection. We introduce a *quantile* monotone instrumental variable (QMIV) condition that is a variant of Manski and Pepper's (2000) Monotone Instrumental Variable (MIV) condition. (The latter applies when the parameter of interest is a conditional *mean* of a treatment response.) A nice feature of the QMIV condition is that non-trivial bounds are obtained without assuming that the support of the response variable is bounded, which is restrictive in some applications. The nontrivial bounds result from the fact that the df's that define the quantiles are naturally bounded between 0 and 1.

Other papers that bound quantiles using the natural bounds of df's include Manski (1994), Lee and Melenberg (1998), and Blundell, Gosling, Ichimura, and Meghir (2007). The QMIV condition differs from the conditions in these papers, although it is closely related to them.[21]

The model set-up is quite similar to that in Manski and Pepper (2000). The observations are i.i.d. for $i = 1, ..., n$. Let $y_i(t) \in \mathcal{Y}$ be individual i's "conjectured" response variable given treatment $t \in \mathcal{T}$. Let T_i be the realization of the treatment for individual i. The observed outcome variable is $Y_i = y_i(T_i)$. Let X_i be a covariate whose support

[21] Manski (1994, pp. 149-153) establishes the worst case quantile bounds, which do not impose any restrictions. Lee and Melenberg (1998, p. 30) and Blundell, Gosling, Ichimura, and Meghir (2007, pp. 330-331) provide quantile bounds based on the assumption of monotonicity in the selection variable T_i (which is binary in their contexts), which is a quantile analogue of Manski and Pepper's (2000) monotone treatment selection condition, as well as bounds based on exclusion restrictions. In addition, Blundell, Gosling, Ichimura, and Meghir (2007, pp. 332-333) employ a monotonicity assumption that is close to the QMIV assumption, but their assumption is imposed on the whole conditional distribution of $y_i(t)$ given X_i, rather than on a single conditional quantile, and they do not explicitly bound quantiles.

contains an ordered set $\mathcal{X}$. We observe $W_i = (Y_i, X_i)$. The parameter of interest, θ, is the conditional τ-quantile of $y_i(t)$ given $X_i = x_0$ for some $t \in \mathcal{T}$ and some $x_0 \in \mathcal{X}$, which is denoted $Q_{y_i(t)|X_i}(\tau|x_0)$. We assume the conditional distribution of $y_i(t)$ given $X_i = x$ is absolutely continuous at its τ-quantile for all $x \in \mathcal{X}$.

For examples, one could have: (i) $y_i(t)$ is conjectured wages of individual i for t years of schooling, T_i is realized years of schooling, and X_i is measured ability or wealth, (ii) $y_i(t)$ is conjectured wages when individual i is employed, say $t = 1$, X_i is measured ability or wealth, and selection occurs due to elastic labor supply, (iii) $y_i(t)$ is consumer durable expenditures when a durable purchase is conjectured, X_i is income or non-durable expenditures, and selection occurs because an individual may decide not to purchase a durable, and (iv) $y_i(t)$ is some health response of individual i given treatment t, T_i is the realized treatment, which may be non-randomized or randomized but subject to imperfect compliance, and X_i is some characteristic of individual i, such as weight, blood pressure, etc.

The quantile monotone IV assumption is as follows:

Assumption QMIV. The covariate X_i satisfies: for some $t \in T$ and all $(x_1, x_2) \in \mathcal{X}^2$ such that $x_1 \leq x_2$,

$$Q_{y_i(t)|X_i}(\tau|x_1) \leq Q_{y_i(t)|X_i}(\tau|x_2),$$

where $\tau \in (0,1)$, $\mathcal{X}$ is some ordered subset of the support of X_i, and $Q_{y_i(t)|X_i}(\tau|x)$ is the quantile function of $y_i(t)$ conditional on $X_i = x$.[22]

This assumption may be suitable in the applications mentioned above.

Given Assumption QMIV, we have: for $(x, x_0) \in \mathcal{X}^2$ with $x \leq x_0$,

$$\begin{aligned}
\tau &= P\left(y_i(t) \leq Q_{y_i(t)|X_i}(\tau|x)|X_i = x\right) \\
&\leq P\left(y_i(t) \leq \theta|X_i = x\right) \\
&= P\left(y_i(t) \leq \theta \;\&\; T_i = t|X_i = x\right) + P\left(y_i(t) \leq \theta \;\&\; T_i \neq t|X_i = x\right) \\
&\leq P\left(Y_i \leq \theta \;\&\; T_i = t|X_i = x\right) + P\left(T_i \neq t|X_i = x\right), \qquad (9.1)
\end{aligned}$$

[22] The "τ-quantile monotone IV" terminology follows that of Manski and Pepper (2000). Alternatively, it could be called a "τ-quantile monotone covariate."

Assumption QMIV can be extended to the case where additional (non-monotone) covariates arise, say Z_i. In this case, the QMIV condition becomes $Q_{y_i(t)|Z_i,X_i}(\tau|z, x_1) \leq Q_{y_i(t)|Z_i,X_i}(\tau|z, x_2)$ when $x_1 \leq x_2$ for all z in some subset $\mathcal{Z}$ of the support of Z_i. Also, as in Manski and Pepper (2000), the assumption QMIV is applicable if $\mathcal{X}$ is only a partially-ordered set.

where first equality holds by the definition of the τ-quantile $Q_{y_i(t)|X_i}(\tau|x)$, the first inequality holds by Assumption QMIV, and the second inequality holds because $Y_i = y_i(T_i)$ and $P(A \cap B) \leq P(B)$.

Analogously, for $(x, x_0) \in \mathcal{X}^2$ with $x \geq x_0$,

$$\begin{aligned}
\tau &= P\left(y_i(t) \leq Q_{y_i(t)|X_i}(\tau|x)|X_i = x\right) \\
&\geq P\left(y_i(t) \leq \theta|X_i = x\right) \\
&= P\left(y_i(t) \leq \theta \text{ \& } T_i = t|X_i = x\right) + P\left(y_i(t) \leq \theta \text{ \& } T_i \neq t|X_i = x\right) \\
&\geq P\left(Y_i \leq \theta \text{ \& } T_i = t|X_i = x\right), \qquad (9.2)
\end{aligned}$$

where the first inequality holds by Assumption QMIV and the second inequality holds because $P(A) \geq 0$.

The inequalities in (9.1) and (9.2) impose sharp bounds on θ. They can be rewritten as conditional moment inequalities:

$$\begin{aligned}
E\left(1(X_i \leq x_0)[1(Y_i \leq \theta, T_i = t) + 1(T_i \neq t) - \tau]|X_i\right) \geq 0 \text{ a.s. and} \\
E\left(1(X_i \geq x_0)[\tau - 1(Y_i \leq \theta, T_i = t)]|X_i\right) \geq 0 \text{ a.s.} \qquad (9.3)
\end{aligned}$$

For the purposes of the Monte Carlo simulations, we consider the following data generating process (DGP):

$$\begin{aligned}
y_i(1) &= \mu(X_i) + \sigma\left(X_i\right) u_i, \text{ where } \partial\mu\left(x\right)/\partial x \geq 0 \text{ and } \sigma\left(x\right) \geq 0, \\
T_i &= 1\{\varphi\left(X_i\right) + \varepsilon_i \geq 0\}, \text{ where } \partial\varphi\left(x\right)/\partial x \geq 0, \\
X_i &\sim Unif[0, 2], \ (\varepsilon_i, u_i) \sim N(0, I_2), \ X_i \perp (\varepsilon_i, u_i), \\
Y_i &= y_i(T_i), \text{ and } t = 1. \qquad (9.4)
\end{aligned}$$

The variable $y_i(0)$ is irrelevant (because Y_i enters the moment inequalities in (9.3) only through $1(Y_i \leq \theta, T_i = t)$) and, hence, is left undefined.

Under the DGP above, X_i satisfies the QMIV assumption for any $\tau \in (0, 1)$. We consider the median: $\tau = 0.5$. We focus on the conditional median of $y_i(1)$ given $X_i = 1.5$, i.e., $\theta = Q_{y_i(1)|X_i}(0.5|1.5)$ and $x_0 = 1.5$.

Some algebra shows that the conditional moment inequalities in (9.3) imply:

$$\begin{aligned}\theta &\geq \underline{\theta}(x) := \mu(x) + \sigma(x)\,\Phi^{-1}\left(1 - [2\Phi(\varphi(x))]^{-1}\right) \text{ for } x \leq 1.5 \text{ and} \\ \theta &\leq \bar{\theta}(x) := \mu(x) + \sigma(x)\,\Phi^{-1}\left([2\Phi(\varphi(x))]^{-1}\right) \text{ for } x \geq 1.5.\end{aligned} \tag{9.5}$$

We call $\underline{\theta}(x)$ and $\bar{\theta}(x)$ the lower and upper bound functions on θ, respectively. The identified set for the quantile selection model is

$$\left[\sup_{x \leq x_0} \underline{\theta}(x),\ \ \inf_{x \geq x_0} \bar{\theta}(x)\right]. \tag{9.6}$$

The shape of the lower and upper bound functions depends on the shape of the φ, μ, and σ functions. We consider two specifications, one that yields flat bound functions and the other that yields kinky bound functions.

Under the flat bound DGP, $\mu(x) = 2$, $\sigma(x) = 1$, and $\varphi(x) = 1\ \forall x \in [0, 2]$. In this case,

$$\begin{aligned}\underline{\theta}(x) &= 2 + \Phi^{-1}\left(1 - [2\Phi(1)]^{-1}\right) \text{ for } x \leq 1.5 \text{ and} \\ \bar{\theta}(x) &= 2 + \Phi^{-1}\left([2\Phi(1)]^{-1}\right) \text{ for } x \geq 1.5.\end{aligned} \tag{9.7}$$

Figure 1 shows the flat bound functions. The solid line is the lower bound function $\underline{\theta}(x)$, and the dashed line is the upper bound function $\bar{\theta}(x)$. Note that $\underline{\theta}(x)$ is defined only for $x \in [0, 1.5]$ and $\bar{\theta}(x)$ only for $x \in [1.5, 1]$.

Under the kinky bound DGP, $\mu(x) = 2(x \wedge 1)$, $\sigma(x) = x$, $\varphi(x) = x \wedge 1$.[23] In this case,

$$\begin{aligned}\underline{\theta}(x) &= 2(x \wedge 1) + x \cdot \Phi^{-1}\left(1 - [2\Phi(x \wedge 1)]^{-1}\right) \text{ for } x \leq 1.5 \text{ and} \\ \bar{\theta}(x) &= 2(x \wedge 1) + x \cdot \Phi^{-1}\left([2\Phi(x \wedge 1)]^{-1}\right) \text{ for } x \geq 1.5.\end{aligned} \tag{9.8}$$

[23] The kinky shaped μ and φ functions are the same as in the simulation example in Chernozhukov, Lee, and Rosen (2008).

Figure 2 shows the kinky bound functions.

Figure 1. Flat bound functions

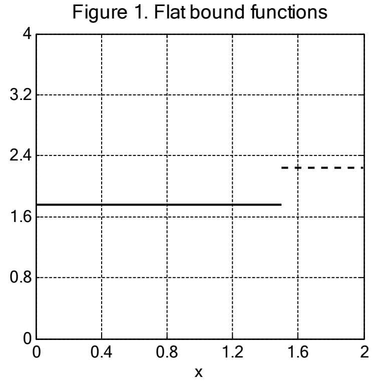
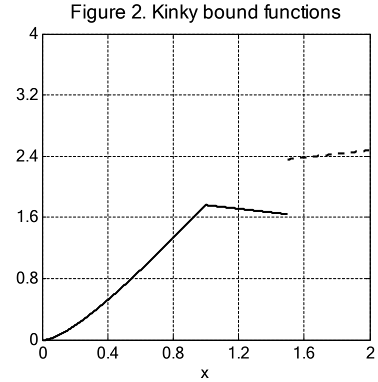


Figure 2. Kinky bound functions



9.2.2 g Functions

The g functions employed by the test statistics are indicator functions of hypercubes in $[0,1]$, i.e., intervals. It is not assumed that the researcher knows that $X_i \sim U[0,2]$. Hence, the regressor X_i is transformed via a general method to lie in $(0,1)$. This method takes the transformed regressor to be $\Phi((X_i - \overline{X}_n)/\sigma_{X,n})$, where $\overline{X}_n$ and $\sigma_{X,n}$ are the sample mean and standard deviations of X_i and $\Phi(\cdot)$ is the standard normal df. The hypercubes have side-edge lengths $(2r)^{-1}$ for $r = r_0, ..., r_1$, where $r_0 = 1$ and the basecase value of r_1 is 7. The basecase number of hypercubes is 56. We also report results for $r_1 = 5$, 9, and 11, which yield 30, 90, and 132 hypercubes, respectively.

9.2.3 Simulation Results

Tables I-III report CP's and CP-corrected FCP's for a variety of test statistics and critical values for a range of cases. The CP's are for the lower endpoint of the identified interval in Tables I-III. (Appendix F of AS provides additional results for the upper endpoints.) FCP's are for points below the lower endpoint.[24]

Table I provides comparisons of different test statistics when each statistic is coupled with PA/Asy and GMS/Asy critical values. Table II provides comparisons of the

[24]Note that the DGP is the same for FCP's as for CP's, just the value θ that is to be covered is different. For the lower endpoint of the identified set and the flat bound, FCP's are computed for θ equal to $\underline{\theta}(1) - 0.25 \times sqrt(250/n)$. For the lower endpoint with the kinky bound, FCP's are computed for θ equal to $\underline{\theta}(1) - 0.58 \times sqrt(250/n)$. These points are chosen to yield similar values for the FCP's across the different cases considered.

Table I. Quantile Selection Model: Basecase Test Statistic Comparisons

(a) Coverage Probabilities

DGP	Statistic: / Crit Val	CvM/Sum	CvM/QLR	CvM/Max	KS/Sum	KS/QLR	KS/Max
Flat Bound	PA/Asy	.979	.979	.976	.972	.972	.970
	GMS/Asy	.953	.953	.951	.963	.963	.960
Kinky Bound	PA/Asy	.999	.999	.999	.994	.994	.994
	GMS/Asy	.983	.983	.983	.985	.985	.984

(b) False Coverage Probabilities (coverage probability corrected)

DGP	Crit Val	CvM/Sum	CvM/QLR	CvM/Max	KS/Sum	KS/QLR	KS/Max
Flat Bound	PA/Asy	.51	.50	.48	.68	.67	.66
	GMS/Asy	.37	.37	.37	.60	.60	.59
Kinky Bound	PA/Asy	.65	.65	.62	.68	.68	.67
	GMS/Asy	.35	.35	.34	.53	.53	.52

* These results are for the lower endpoint of the identified interval.

PA/Asy, PA/Bt, GMS/Asy, GMS/Bt, and Sub critical values for the CvM/Max and KS/Max test statistics. Table III provides robustness results for the CvM/Max and KS/Max statistics coupled with GMS/Asy critical values. The Table III results show the degree of sensitivity of the results to (i) the sample size, n, (ii) the number of cubes employed, as indexed by r_1, (iii) the choice of (κ_n, B_n) for the GMS/Asy critical values, and (iv) the value of ε, upon which the variance estimator $\overline{\Sigma}_n(\theta, g)$ depends. Table III also reports results for confidence intervals with nominal level .5, which yield asymptotically half-median unbiased estimates of the lower endpoint.

Table I shows that all tests have CP's greater than or equal to .95 with flat and kinky bound DGP's. The PA/Asy critical values lead to noticeably larger over-coverage than the GMS/Asy critical values. The GMS/Asy critical values lead to CP's that are close to .95 with the flat bound DGP and larger than .95 with the kinky bound DGP. The CP results are not sensitive to the choice of test statistic function: Sum, QLR, or Max. They are only marginally sensitive to the choice of test statistic form: CvM or KS.

The FCP results of Table I show (i) a clear advantage of CvM-based CI's over

Table II. Quantile Selection Model: Basecase Critical Value Comparisons*

(a) Coverage Probabilities

DGP	Critical Value: / Statistic	PA/Asy	PA/Bt	GMS/Asy	GMS/Bt	Sub
Flat Bound	CvM/Max	.976	.977	.951	.950	.983
	KS/Max	.970	.973	.960	.959	.942
Kinky Bound	CvM/Max	.999	.999	.983	.982	.993
	KS/Max	.994	1.00	.984	.982	.950

(b) False Coverage Probabilities (coverage probability corrected)

DGP	Statistic	PA/Asy	PA/Bt	GMS/Asy	GMS/Bt	Sub
Flat Bound	CvM/Max	.48	.49	.37	.36	.57
	KS/Max	.66	.69	.59	.57	.69
Kinky Bound	CvM/Max	.62	.64	.34	.33	.47
	KS/Max	.67	.72	.52	.50	.47

* These results are for the lower endpoint of the identified interval.

KS-based CI's, (ii) a clear advantage of GMS/Asy critical values over PA/Asy critical values, and (iii) little difference between the test statistic functions: Sum, QLR, and Max. These results hold for both the flat and kinky bound DGP's.

Table II compares the critical values PA/Asy, PA/Bt, GMS/Asy, GMS/Asy, and Sub. The results show little difference in terms of CP's and FCP's between the Asy and Bt versions of the PA and GMS critical values in most cases. The GMS critical values noticeably out-perform the PA critical values in terms of FCP's. For the CvM/Max statistic, which is the better statistic of the two considered, the GMS critical values also noticeably out-perform the Sub critical values in terms of FCP's.

Table III provides results for the CvM/Max and KS/Max statistics coupled with the GMS/Asy critical values for several variations of the basecase. The table shows that these CS's perform quite similarly for different sample sizes, different numbers of cubes, and a smaller constant ε.[25] There is some sensitivity to the magnitude of the GMS

[25] The θ value at which the FCP's are computed differs from the lower endpoint of the identified set by a distance that depends on $n^{-1/2}$. Hence, Table III suggests that the "local alternatives" that give equal FCP's decline with n at a rate that is slightly faster than $n^{-1/2}$ over the range $n = 100$ to 1000.

Table III. Quantile Selection Model with Flat Bound: Variations on the Basecase*

		(a) Coverage Probabilities		(b) False Cov Probs (CPcor)	
	Statistic:	CvM/Max	KS/Max	CvM/Max	KS/Max
Case	Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Basecase ($n = 250, r_1 = 7, \varepsilon = 5/100$)		.951	.960	.37	.59
$n = 100$		.957	.968	.40	.64
$n = 500$		.954	.955	.36	.58
$n = 1000$		.948	.948	.34	.57
$r_1 = 5$		.949	.954	.36	.56
$r_1 = 9$		.951	.963	.37	.61
$r_1 = 11$		.951	.966	.37	.63
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.948	.954	.38	.58
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.967	.968	.38	.63
$\varepsilon = 1/100$		.949	.957	.37	.64
$\alpha = .5$		.518	.539	.03	.08
$\alpha = .5$ & $n = 500$		.513	.531	.03	.07

* These results are for the lower endpoint of the identified interval.

tuning parameters, (κ_n, B_n)—doubling their values increases CP's, but halving their values does not decrease their CP's below .95. Across the range of cases considered the CvM-based CS's out perform the KS-based CS's in terms of FCP's and are comparable in terms of CP's.

The last two rows of Table III show that a CS based on $\alpha = .5$ provides a good choice for an estimator of the identified set. For example, the lower endpoint estimator based on the CvM/Max-GMS/Asy CS with $\alpha = .5$ is close to being median-unbiased. It is less than the lower bound with probability is .518 and exceeds it with probability .482 when $n = 250$.

In conclusion, we find that the CS based on the CvM/Max statistic with the GMS/Asy critical value performs best in the quantile selection models considered. Equally good are the CS's that use the Sum or QLR statistic in place of the Max statistic and the GMS/Bt critical value in place of the GMS/Asy critical value. The CP's and FCP's of the CvM/Max-GMS/Asy CS are quite good over a range of sample sizes.

9.3 Interval Outcome Regression Model

9.3.1 Description of Model

This model has been considered by Manski and Tamer (2002, Sec. 4.5). It is a regression model where the outcome variable Y_i^* is partially observed:

$$Y_i^* = \theta_1 + X_i\theta_2 + U_i, \text{ where } E(U_i|X_i) = 0 \text{ a.s., for } i = 1, ..., n. \tag{9.9}$$

One observes X_i and an interval $[Y_{L,i}, Y_{U,i}]$ that contains Y_i^*: $Y_{L,i} = [Y_i]$ and $Y_{U,i} = [Y_i] + 1$, where $[x]$ denotes the integer part of x. Thus, $Y_i^* \in [Y_{L,i}, Y_{U,i}]$.

It is straightforward to see that the following conditional moment inequalities hold in this model:

$$\begin{aligned} &E(\theta_1 + X_i\theta_2 - Y_{L,i}|X_i) \geq 0 \text{ a.s. and} \\ &E(Y_{U,i} - \theta_1 - X_i\theta_2|X_i) \geq 0 \text{ a.s.} \end{aligned} \tag{9.10}$$

In the simulation experiment, we take the true parameters to be $(\theta_1, \theta_2) = (1, 1)$ (without loss of generality), $X_i \sim U[0, 1]$, and $U_i \sim N(0, 1)$. We consider a basecase sample size of $n = 250$, as well as $n = 100$, 500, and 1000.

The parameter $\theta = (\theta_1, \theta_2)$ is not identified. The identified set is a parallelogram in (θ_1, θ_2) space with vertices at $(.5, 1), (.5, 2), (1.5, 0)$, and $(1.5, 1)$. (Appendix F of AS provides a figure that illustrates the identified set.) By symmetry, CP's of CS's are the same for the points $(.5, 1)$ and $(1.5, 1)$. Also, they are the same for $(.5, 2)$ and $(1.5, 0)$. We focus on CP's at the corner point $(.5, 1)$, which is in the identified set, and at points close to $(.5, 1)$ but outside the identified set.[26] The corner point $(.5, 1)$ is of interest because it is a point in the identified set where CP's of CS's typically are strictly less than one. Due to the features of the model, the CP's of CS's typically equal one (or essentially equal one) at interior points, non-corner boundary points, and the corner points $(.5, 2)$ and $(1.5, 0)$.

[26] Specifically, the θ values outside the identified set are given by $\theta_1 = 0.5 - 0.075 \times (500/n)^{1/2}$ and $\theta_2 = 1.0 - 0.050 \times (500/n)^{1/2}$. These θ values are selected so that the FCP's of the CS's take values in an interesting range for all values of n considered.

9.3.2 g Functions

The g functions employed by the test statistics are indicator functions of hypercubes in $[0,1]$. It is not assumed that the researcher knows that $X_i \sim U[0,1]$ and so the regressor X_i is transformed via the same method as in the quantile selection model to lie in $(0,1)$.

9.3.3 Simulation Results

Tables IV-VI provide results for the interval outcome regression model that are analogous to the results in Tables I-III for the quantile selection model. In spite of the differences in the models—the former is linear and parametric with a bivariate parameter, while the latter is nonparametric with a scalar parameter—the results are similar.

Table IV shows that the CvM/Max statistic combined with the GMS/Asy critical value has CP's that are very close to the nominal level .95. Its FCP's are noticeably lower than those for CS's that use the KS form or PA-based critical values. The CvM/Sum-GMS/Asy and CvM/QLR-GMS/Asy CS's perform equally well as the Max version. Table V shows that the results for the Asy and Bt versions of the critical values are quite similar for the CvM/Max-GMS CS, which is the best CS. The Sub critical value yields substantial under-coverage for the KS/Max statistic. The Sub critical values are dominated by the GMS critical values in terms of FCP's.

Table VI shows that the CS's do not exhibit much sensitivity to the sample size or the number of cubes employed. It also shows that at the non-corner boundary point $\theta = (1.0, 0.5)$ and the corner point $\theta = (1.5, 0)$, all CP's are (essentially) equal to one.[27] Lastly, Table VI shows that the lower endpoint estimator based on the CvM/Max-GMS/Asy CS with $\alpha = .5$ is close to being median-unbiased, as in the quantile selection model. It is less than the lower bound with probability is .472 and exceeds it with probability .528 when $n = 250$.

We conclude that the preferred CS for this model is of the CvM form, combined with the Max, Sum, or QLR function, and uses a GMS critical value, either Asy or Bt.

[27] This is due to the fact that the CP's at these points are linked to their CP's at the corner point $\theta = (0.5, 1.0)$ given the linear structure of the model. If the CP is reduced at the two former points (by reducing the critical value), the CP at the latter point is very much reduced and the CS does not have the desired size.

Table IV. Interval Outcome Regression Model: Basecase Test Statistic Comparisons

(a) Coverage Probabilities

Critical Value	Statistic:	CvM/Sum	CvM/QLR	CvM/Max	KS/Sum	KS/QLR	KS/Max
PA/Asy		.990	.993	.990	.989	.990	.989
GMS/Asy		.950	.950	.950	.963	.963	.963

(b) False Coverage Probabilities (coverage probability corrected)

PA/Asy		.62	.66	.61	.78	.80	.78
GMS/Asy		.37	.37	.37	.61	.61	.61

Table V. Interval Outcome Regression Model: Basecase Critical Value Comparisons

(a) Coverage Probabilities

Statistic	Critical Value:	PA/Asy	PA/Bt	GMS/Asy	GMS/Bt	Sub
CvM/Max		.990	.995	.950	.941	.963
KS/Max		.989	.999	.963	.953	.890

(b) False Coverage Probabilities (coverage probability corrected)

CvM/Max		.61	.69	.37	.38	.45
KS/Max		.78	.96	.61	.54	.66

9.4 Entry Game Model

9.4.1 Description of the Model

This model is a complete information simultaneous game (entry model) with two players and n i.i.d. plays of the game. We consider Nash equilibria in pure strategies. Due to the possibility of multiple equilibria, the model is incomplete. In consequence,

Table VI. Interval Outcome Regression Model: Variations on the Basecase

	(a) Coverage Probabilities		(b) False Cov Probs (CPcor)	
Statistic:	CvM/Max	KS/Max	CvM/Max	KS/Max
Case — Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Basecase ($n = 250, r_1 = 7, \varepsilon = 5/100$)	.950	.963	.37	.61
$n = 100$	.949	.970	.39	.66
$n = 500$	.950	.956	.37	.60
$n = 1000$	.954	.955	.37	.60
$r_1 = 5$ (30 cubes)	.949	.961	.37	.59
$r_1 = 9$ (90 cubes)	.951	.965	.37	.63
$r_1 = 11$ (132 cubes)	.950	.968	.38	.64
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$	.944	.961	.40	.62
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$	.958	.973	.39	.65
$\varepsilon = 1/100$	.946	.966	.39	.69
$(\theta_1, \theta_2) = (1.0, 0.5)$	.999	.996	.91	.92
$(\theta_1, \theta_2) = (1.5, 0.0)$	1.000	.996	.99	.97
$\alpha = .5$	.472	.481	.03	.08
$\alpha = .5$ & $n = 500$	.478	.500	.03	.07

two conditional moment inequalities and two conditional moment equalities arise. Andrews, Berry, and Jia (2004), Beresteanu, Molchanov, and Molinari (2009), Galichon and Henry (2009b), and Ciliberto and Tamer (2009) also consider moment inequalities and equalities in models of this sort.

We consider the case where the two players' utility/profits depend linearly on vectors of covariates, $X_{i.1}$ and $X_{i,2}$, with corresponding parameters τ_1 and τ_2. A scalar parameter θ_1 indexes the competitive effect on player 1 of entry by player 2. Correspondingly, θ_2 indexes the competitive effect on player 2 of entry by player 1.

Specifically, for player $b = 1, 2$, player b's utility/profits are given by

$$\begin{aligned} &X_{i,b}'\tau_b + U_{i,b} \text{ if the other player does not enter and} \\ &X_{i,b}'\tau_b - \theta_b + U_{i,b} \text{ if the other player enters,} \end{aligned} \tag{9.11}$$

where $U_{i,b}$ is an idiosyncratic error known to both players, but unobserved by the econometrician. The random variables observed by the econometrician are the covariates $X_{i,1} \in R^4$ and $X_{i,2} \in R^4$ and the outcome variables $Y_{i,1}$ and $Y_{i,2}$, where $Y_{i,b}$ equals 1 if player b enters and 0 otherwise for $b = 1, 2$. The unknown parameters are $\theta = (\theta_1, \theta_2)' \in [0, \infty)^2$, and $\tau = (\tau_1', \tau_2')' \in R^8$. Let $Y_i = (Y_{i,1}, Y_{i,2})$ and $X_i = (X_{i,1}', X_{i,2}')'$.

The covariate vector $X_{i,b}$ equals $(1, X_{i,b,2}, X_{i,b,3}, X_i^*)' \in R^4$, where $X_{i,b,2}$ has a Bern(p) distribution with $p = 1/2$, $X_{i,b,3}$ has a $N(0, 1)$ distribution, X_i^* has a $N(0, 1)$ distribution and is the same for $b = 1, 2$. The idiosyncratic error $U_{i,b}$ has a $N(0, 1)$ distribution. All random variables are independent of each other. Except when specified otherwise, the equilibrium selection rule (ESR) employed is a maximum profit ESR (which is unknown to the econometrician). That is, if Y_i could be either $(1, 0)$ or $(0, 1)$ in equilibrium, then it is $(1, 0)$ if player 1's monopoly profit exceeds that of player 2 and is $(0, 1)$ otherwise. We also provide some results for a "player 1 first" ESR in which $Y_i = (1, 0)$ whenever Y_i could be either $(1, 0)$ or $(0, 1)$ in equilibrium.

The moment inequality functions are

$$\begin{aligned} m_1(W_i, \theta, \tau) &= P(X_{i,1}'\tau_1 + U_{i,1} \geq 0,\ X_{i,2}'\tau_2 - \theta_2 + U_{i,2} \leq 0 | X_i) - 1(Y_i = (1, 0)) \\ &= \Phi(X_{i,1}'\tau_1)\Phi(-X_{i,2}'\tau_2 + \theta_2) - 1(Y_i = (1, 0)) \text{ and} \\ m_2(W_i, \theta, \tau) &= P(X_{i,1}'\tau_1 - \theta_1 + U_{i,1} \leq 0,\ X_{i,2}'\tau_2 + U_{i,2} \geq 0 | X_i) - 1(Y_i = (0, 1)), \\ &= \Phi(-X_{i,1}'\tau_1 + \theta_1)\Phi(X_{i,2}'\tau_2) - 1(Y_i = (0, 1)). \end{aligned} \tag{9.12}$$

We have $E(m_1(W_i, \theta_0, \tau_0)|X_i) \geq 0$ a.s., where θ_0 and τ_0 denote the true values, because given X_i a necessary condition for $Y_i = (1, 0)$ is $X_{i,1}'\tau_1 + U_{i,1} \geq 0$ and $X_{i,2}'\tau_2 - \theta_2 + U_{i,2} \leq 0$. However, this condition is not sufficient for $Y_i = (1, 0)$ because some sample realizations with $Y_i = (0, 1)$ also may satisfy this condition. An analogous argument leads to $E(m_2(W_i, \theta_0, \tau_0)|X_i) \geq 0$ a.s.

The two moment equality functions are

$$\begin{aligned} m_3(W_i, \theta, \tau) &= 1(Y_i = (1, 1)) - P(X_{i,1}'\tau_1 - \theta_1 + U_{i,1} \geq 0,\ X_{i,2}'\tau_2 - \theta_2 + U_{i,2} \geq 0 | X_i), \\ &= 1(Y_i = (1, 1)) - \Phi(X_{i,1}'\tau_1 - \theta_1)\Phi(X_{i,2}'\tau_2 - \theta_2), \text{ and} \\ m_4(W_i, \theta, \tau) &= 1(Y_i = (0, 0)) - P(X_{i,1}'\tau_1 + U_{i,1} \leq 0,\ X_{i,2}'\tau_2 + U_{i,2} \leq 0 | X_i) \\ &= 1(Y_i = (0, 0)) - \Phi(-X_{i,1}'\tau_1)\Phi(-X_{i,2}'\tau_2). \end{aligned} \tag{9.13}$$

We employ a preliminary estimator of τ given θ, as in Section 8. In particular, we use the probit ML estimator $\widehat{\tau}_n(\theta) = (\widehat{\tau}_{n,1}(\theta)', \widehat{\tau}_{n,2}(\theta)')'$ of $\tau = (\tau_1', \tau_2')'$ given θ based on the observations $\{(1(Y_i = (0,0)), 1(Y_i = (1,1)), X_{i,1}, X_{i,2}) : i \leq n\}$.[28]

The model described above is point identified because τ is identified by the second moment equality $m_4(W_i, \theta, \tau)$ and θ is identified by the first moment equality $m_3(W_i, \theta, \tau)$ given that τ is identified. However, additional information about θ and τ is provided by the moment inequalities in (9.12), which we exploit by the methods employed here.

9.4.2 g Functions

We take the functions g to be hypercubes in R^2. They are functions of the 2-vector $X_i^* = (X_{i,1}^{*\prime}, X_{i,2}^{*\prime})' = (X_{i,1}'\widehat{\tau}_{n,1}(\theta), X_{i,2}'\widehat{\tau}_{n,2}(\theta)'$. The vector X_i^* is transformed first to have sample mean equal to zero and sample variance matrix equal to I_2 (by multiplication by the inverse of the upper-triangular Cholesky decomposition of the sample covariance matrix of X_i^*). Then, it is transformed to lie in $[0,1]^2$ by applying the standard normal df $\Phi(\cdot)$ element by element.

The hypercubes have side-edge lengths $(2r)^{-1}$ for $r = r_0, ..., r_1$, where $r_0 = 1$ and the basecase value of r_1 is 3. The basecase number of hypercubes is 56. We also report results for $r_1 = 2$ and 4, which yield 20 and 120 hypercubes, respectively.

9.4.3 Simulation Results

Tables VII and VIII provide results for the entry game model. Results are provided for GMS/Asy critical values only because (i) PA/Asy critical values are found to provide identical results and (ii) bootstrap and subsampling critical values are computationally quite costly because they require computation of the bootstrap or subsample ML estimator for each repetition of the critical value calculations.

Table VII provides CP's and FCP's for competitive effect θ values ranging from $(0,0)$ to $(3,1)$.[29] Table VII shows that the CP's for all CS's vary as θ varies with values ranging from .913 to .987. The QLR-based CS's tend to have higher CP's than the Sum- and Max-based CS's. The CvM/Max statistic dominates all other statistics except the CvM/QLR statistic in terms of FCP's. In addition, CvM/Max dominates CvM/QLR—in most cases by a substantial margin—except for $\theta = (2,2)$ or $(3,1)$. Hence, CvM/Max

[28] See Appendix F of AS for the specification of the log likelihood function and its gradient.

[29] The θ values for which FCP's are computed are given by $\theta_1 - .1 \times sqrt(500/n)$ and $\theta_2 - .1 \times sqrt(500/n)$, where (θ_1, θ_2) is the true parameter vector.

Table VII. Entry Game Model: Test Statistic Comparisons for Different Competitive Effects Parameters (θ_1, θ_2)

(a) Coverage Probabilities

Case	Statistic:	CvM/Sum	CvM/QLR	CvM/Max	KS/Sum	KS/QLR	KS/Max
$(\theta_1, \theta_2) = (0, 0)$		.979	.972	.980	.977	.975	.985
$(\theta_1, \theta_2) = (1, 0)$		.961	.980	.965	.959	.983	.972
$(\theta_1, \theta_2) = (1, 1)$		.961	.985	.961	.955	.985	.962
$(\theta_1, \theta_2) = (2, 0)$		.935	.982	.935	.944	.984	.952
$(\theta_1, \theta_2) = (2, 1)$		.943	.974	.940	.953	.987	.947
$(\theta_1, \theta_2) = (3, 0)$		.921	.975	.915	.938	.935	.984
$(\theta_1, \theta_2) = (2, 2)$		.928	.942	.913	.943	.972	.922
$(\theta_1, \theta_2) = (3, 1)$		.928	.950	.918	.949	.973	.932

(b) False Coverage Probabilities (coverage probability corrected)

Case	Statistic:	CvM/Sum	CvM/QLR	CvM/Max	KS/Sum	KS/QLR	KS/Max
$(\theta_1, \theta_2) = (0, 0)$		.76	.99	.59	.91	.99	.83
$(\theta_1, \theta_2) = (1, 0)$		.60	.99	.42	.83	.66	.99
$(\theta_1, \theta_2) = (1, 1)$		.62	.96	.41	.82	.99	.58
$(\theta_1, \theta_2) = (2, 0)$		.51	.83	.35	.66	.96	.47
$(\theta_1, \theta_2) = (2, 1)$		.57	.57	.38	.69	.82	.44
$(\theta_1, \theta_2) = (3, 0)$		.49	.41	.36	.61	.43	.64
$(\theta_1, \theta_2) = (2, 2)$		.59	.34	.39	.65	.42	.49
$(\theta_1, \theta_2) = (3, 1)$		.57	.27	.39	.65	.47	.44

is clearly the best statistic in terms of FCP's. The CP's of the CvM/Max CS are good for many θ values, but they are low for relatively large θ values. For $\theta = (3, 0)$, $(2, 2)$, and $(3, 1)$, its CP's are .915, .913, and .918, respectively. This is a "small" sample effect—for $n = 1000$, this CS has CP's for these three cases equal to .934, .951, and .952, respectively.

Table VIII provides results for variations on the basecase θ value of $(1, 1)$ for the CvM/Max and KS/Max statistics combined with GMS/Asy critical values. The CP's and FCP's of the CvM/Max CS increase with n. They are not sensitive to the number of hypercubes. There is some sensitivity to the magnitude of (κ_n, B_n), but it is relatively small. There is noticeable sensitivity of the CP of the KS/Max CS to ε, but less so for

Table VIII. Entry Game Model: Variations on the Basecase $(\theta_1, \theta_2) = (1, 1)$				
	(a) Coverage Probabilities		(b) False Cov Probs (CPcor)	
Statistic:	CvM/Max	KS/Max	CvM/Max	KS/Max
Case / Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Basecase $(n = 500, r_1 = 3, \varepsilon = 5/100)$	**.961**	**.962**	**.41**	**.58**
$n = 250$	.948	.963	.39	.56
$n = 1000$	.979	.968	.52	.65
$r_1 = 2$ (20 cubes)	.962	.956	.41	.55
$r_1 = 4$ (120 cubes)	.962	.964	.42	.59
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$	.954	.959	.39	.57
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$	.967	.962	.42	.58
$\varepsilon = 1/100$	.926	.873	.32	.66
Reg'r Variances = 2	.964	.968	.54	.71
Reg'r Variances = 1/2	.963	.966	.29	.43
Player 1 First Eq Sel Rule	.955	.957	.39	.57
$\alpha = .5$	.610	.620	.05	.13
$\alpha = .5$ & $n = 1000$	.695	.650	.06	.16

the CvM/Max CS. There is relatively little sensitivity of CP's to changes in the DGP via changes in the regressor variances (of $X_{i,b,2}$ and $X_{i,b,3}$ for $b = 1, 2$) or a change in the equilibrium selection rule to player 1 first.

The last two rows of Table VIII provide results for estimators of the identified set based on CS's with $\alpha = .5$. The two CS's considered are half-median unbiased. For example, the CvM/Max-GMS/Asy CS with $\alpha = .5$ covers the true value with probability .610, which exceeds .5, when $n = 500$.

In conclusion, in the entry game model we prefer the CvM/Max-GMS/Asy CS over other CS's considered because of its the clear superiority in terms of FCP's even though it under-covers somewhat for large values of the competitive effects vector θ.

10 Appendix A

In this Appendix, we prove Theorems 1 and 2(a). Proofs of the other results stated in the paper are given in Appendix C in AS.

The following Lemma is used in the proofs of Theorems 1, 2, 3, and 4. It establishes a functional CLT and uniform LLN for certain independent non-identically distributed empirical processes.

Let h_2 denote a $k \times k$-matrix-valued covariance kernel on $\mathcal{G} \times \mathcal{G}$ (such as an element of $\mathcal{H}_2$).

Definition SubSeq($\mathbf{h_2}$). $SubSeq(h_2)$ is the set of subsequences $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\}$, where $\{a_n : n \geq 1\}$ is some subsequence of $\{n\}$, for which

$$\text{(i)} \lim_{n \to \infty} \sup_{g, g^* \in \mathcal{G}} ||h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*)|| = 0,$$

(ii) $\theta_{a_n} \in \Theta$, (iii) $\{W_i : i \geq 1\}$ are i.i.d. under F_{a_n}, (iv) $Var_{F_{a_n}}(m_j(W_i, \theta_{a_n})) > 0$ for $j = 1, ..., k$, for $n \geq 1$, (v) $\sup_{n \geq 1} E_{F_{a_n}} |m_j(W_i, \theta_{a_n}) / \sigma_{F_{a_n}, j}(\theta_{a_n})|^{2+\delta} < \infty$ for $j = 1, ..., k$, for some $\delta > 0$, and (vi) Assumption M holds with F_{a_n} in place of F and F_n in Assumptions M(b) and M(c), respectively.

The sample paths of the Gaussian process $\nu_{h_2}(\cdot)$, which is defined in (4.2) and appears in the following Lemma, are bounded and uniformly ρ-continuous a.s. The pseudo-metric ρ on $\mathcal{G}$ is a pseudo-metric commonly used in the empirical process literature:

$$\rho^2(g, g^*) = tr\left(h_2(g, g) - h_2(g, g^*) - h_2(g^*, g) + h_2(g^*, g^*)\right). \tag{10.1}$$

For $h_2(\cdot, \cdot) = h_{2,F}(\theta, \cdot, \cdot)$, where $(\theta, F) \in \mathcal{F}$, this metric can be written equivalently as

$$\begin{aligned} \rho^2(g, g^*) &= E_F ||D_F^{-1/2}(\theta)[\widetilde{m}(W_i, \theta, g) - \widetilde{m}(W_i, \theta, g^*)]||^2, \text{ where} \\ \widetilde{m}(W_i, \theta, g) &= m(W_i, \theta, g) - E_F m(W_i, \theta, g). \end{aligned} \tag{10.2}$$

Lemma A1. *For any subsequence* $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$,
(a) $\nu_{a_n, F_{a_n}}(\theta_{a_n}, \cdot) \Rightarrow \nu_{h_2}(\cdot)$ *as* $n \to \infty$ (*as processes indexed by* $g \in \mathcal{G}$), *and*
(b) $\sup_{g, g^* \in \mathcal{G}} ||\widehat{h}_{2, a_n, F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*)|| \to_p 0$ *as* $n \to \infty$.

Comments. 1. The proof of Lemma A1 is given in Appendix E of AS. Part (a) is proved by establishing the manageability of $\{m(W_i, \theta_{a_n}, g) - E_{F_{a_n}} m(W_i, \theta_{a_n}, g) : g \in \mathcal{G}\}$

and by establishing a functional central limit theorem (FCLT) for R^k-valued i.n.i.d. empirical processes with the pseudo-metric ρ by using the FCLT in Pollard (1990, Thm. 10.2) for real-valued empirical processes. Part (b) is proved using a maximal inequality given in Pollard (1990, (7.10)).

2. To obtain uniform asymptotic coverage probability results for CS's, Lemma A1 is applied with $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ and $h_2 \in \mathcal{H}_2$. In this case, conditions (ii)-(vi) in the definition of $SubSeq(h_2)$ hold automatically by the definition of $\mathcal{F}$. To obtain power results under fixed and local alternatives, Lemma A1 is applied with $(\theta_{a_n}, F_{a_n}) \notin \mathcal{F}$ for all $n \geq 1$ and h_2 may or may not be in $\mathcal{H}_2$.

Proof of Theorem 1. First, we prove part (a). Let $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ be a sequence for which $h_{2,F_n}(\theta_n) \in \mathcal{H}_{2,cpt}$ for all $n \geq 1$ and the term in square brackets in Theorem 1(a) evaluated at (θ_n, F_n) differs from its supremum over $(\theta, F) \in \mathcal{F}$ with $h_{2,F}(\theta) \in \mathcal{H}_{2,cpt}$ by δ_n or less, where $0 < \delta_n \to 0$ as $n \to \infty$. Such a sequence always exists. To prove part (a), it suffices to show that part (a) holds with the supremum deleted and with (θ, F) replaced by (θ_n, F_n).

By the compactness of $\mathcal{H}_{2,cpt}$, given any subsequence $\{u_n : n \geq 1\}$ of $\{n\}$, there exists a subsubsequence $\{a_n : n \geq 1\}$ for which $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ as $n \to \infty$ for some $\theta_0 \in \Theta$, where d is defined in (5.6), and some $h_{2,0} \in \mathcal{H}_{2,cpt}$. This and $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ implies that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$.

Now, by Lemma A1, we have

$$\begin{pmatrix} \nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot) \\ \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, \cdot) \end{pmatrix} \Rightarrow \begin{pmatrix} \nu_{h_{2,0}}(\cdot) \\ h_{2,0}(\cdot) \end{pmatrix} \text{ as } n \to \infty \tag{10.3}$$

as stochastic processes on $\mathcal{G}$, where $\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g) = \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g, g)$ and $h_{2,0}(g) = h_{2,0}(g, g)$.

Given this, by the almost sure representation theorem, e.g., see Pollard (1990, Thm. 9.4), there exists a probability space and random quantities $\tilde{\nu}_{a_n}(\cdot)$, $\tilde{h}_{2,a_n}(\cdot)$, $\tilde{\nu}_0(\cdot)$, and $\tilde{h}_2(\cdot)$ defined on it such that (i) $(\tilde{\nu}_{a_n}(\cdot), \tilde{h}_{2,a_n}(\cdot))$ has the same distribution as $(\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot), \widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, \cdot))$, (ii) $(\tilde{\nu}_0(\cdot), \tilde{h}_2(\cdot))$ has the same distribution as $(\nu_{h_{2,0}}(\cdot), h_{2,0}(\cdot))$, and

$$\text{(iii) } \sup_{g \in \mathcal{G}} \left\| \begin{pmatrix} \tilde{\nu}_{a_n}(g) \\ \tilde{h}_{2,a_n}(g) \end{pmatrix} - \begin{pmatrix} \tilde{\nu}_0(g) \\ \tilde{h}_2(g) \end{pmatrix} \right\| \to 0 \text{ as } n \to \infty \text{ a.s.} \tag{10.4}$$

Because $h_{2,0}(\cdot)$ is deterministic, condition (ii) implies that $\tilde{h}_2(\cdot) = h_{2,0}(\cdot)$ a.s.

Define

$$\begin{aligned}
\tilde{h}^{\varepsilon}_{2,a_n}(\cdot) &= \tilde{h}_{2,a_n}(\cdot) + \varepsilon \cdot Diag(\tilde{h}_{2,a_n}(1_k)), \\
\tilde{T}_{a_n} &= \int S(\tilde{\nu}_{a_n}(g) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), \tilde{h}^{\varepsilon}_{2,a_n}(g))dQ(g), \\
h^{\varepsilon}_{2,0}(\cdot) &= h_{2,0}(\cdot) + \varepsilon I_k, \text{ and} \\
\tilde{T}_{a_n,0} &= \int S\left(\tilde{\nu}_0(g) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), h^{\varepsilon}_{2,0}(g)\right) dQ(g).
\end{aligned} \tag{10.5}$$

By construction, $\tilde{T}_{a_n}$ and $T_{a_n}(\theta_{a_n})$ have the same distribution, and $\tilde{T}_{a_n,0}$ and $T(h_{a_n,F_{a_n}}(\theta_{a_n}))$ have the same distribution for all $n \geq 1$.

Hence, to prove part (a), it suffices to show that

$$A = \limsup_{n\to\infty} \left[P_{F_{a_n}}(\tilde{T}_{a_n} > x_{h_{a_n,F_{a_n}}(\theta_{a_n})}) - P(\tilde{T}_{a_n,0} + \delta > x_{h_{a_n,F_{a_n}}(\theta_{a_n})})\right] \leq 0. \tag{10.6}$$

Below we show that

$$\tilde{T}_{a_n} - \tilde{T}_{a_n,0} \to 0 \text{ as } n \to \infty \text{ a.s.} \tag{10.7}$$

Let

$$\begin{aligned}
\widetilde{\Delta}_n &= 1(\tilde{T}_{a_n,0} + (\tilde{T}_{a_n} - \tilde{T}_{a_n,0}) > x_{h_{a_n,F_{a_n}}(\theta_{a_n})}) - 1(\tilde{T}_{a_n,0} + \delta > x_{h_{a_n,F_{a_n}}(\theta_{a_n})}) \\
&= \widetilde{\Delta}^+_n - \widetilde{\Delta}^-_n, \text{ where} \\
\widetilde{\Delta}^+_n &= \max\{\widetilde{\Delta}_n, 0\} \in [0,1] \text{ and } \widetilde{\Delta}^-_n = \max\{-\widetilde{\Delta}_n, 0\} \in [0,1].
\end{aligned} \tag{10.8}$$

By (10.7) and $\delta > 0$, $\lim_{n\to\infty} \widetilde{\Delta}^+_n = 0$ a.s. Hence, by the BCT,

$$\begin{aligned}
\lim_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}^+_n &= 0 \text{ and} \\
A = \limsup_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}_n &= \limsup_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}^+_n - \liminf_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}^-_n \\
&= -\liminf_{n\to\infty} E_{F_{a_n}} \widetilde{\Delta}^-_n \leq 0.
\end{aligned} \tag{10.9}$$

Hence, (10.6) holds and the proof of part (a) is complete, except for (10.7).

To prove part (b), analogous results to (10.6), (10.8), and (10.9) hold by analogous arguments.

It remains to show (10.7). We do so by fixing a sample path ω and using the bounded convergence theorem (because $\tilde{T}_{a_n}$ and $\tilde{T}_{a_n,0}$ are both integrals over $g \in \mathcal{G}$ with respect to the measure Q). Let $\tilde{\Omega}$ be the collection of all $\omega \in \Omega$ such that $(\tilde{\nu}_{a_n}(g), \tilde{h}_{2,a_n}(g))(\omega)$ converges to $(\tilde{\nu}_0(g), h_{2,0}(g))(\omega)$ uniformly over $g \in \mathcal{G}$ as $n \to \infty$ and $\sup_{g\in\mathcal{G}} \|\tilde{\nu}_0(g)(\omega)\| < \infty$. By (10.4) and $\tilde{h}_2(\cdot) = h_{2,0}(\cdot)$ a.s., $P(\tilde{\Omega}) = 1$. Consider a fixed $\omega \in \tilde{\Omega}$. By Assumption S2 and (10.4), for all $g \in \mathcal{G}$,

$$\sup_{\mu \in R_+^p \times \{0\}^v} \left| S\left(\tilde{\nu}_{a_n}(g)(\omega) + \mu, \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)\right) - S\left(\tilde{\nu}_0(g)(\omega) + \mu, h^{\varepsilon}_{2,0}(g)\right)\right| \to 0 \tag{10.10}$$

as $n \to \infty$ a.s. Thus, for all $g \in \mathcal{G}$ and all $\omega \in \widetilde{\Omega}$,

$$\begin{aligned} &S\left(\tilde{\nu}_{a_n}(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)\right) \\ &\quad - S\left(\tilde{\nu}_0(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), h^{\varepsilon}_{2,0}(g)\right) \\ &\to 0 \text{ as } n \to \infty. \end{aligned} \tag{10.11}$$

Next, we show that for fixed $\omega \in \tilde{\Omega}$ the first summand on the left-hand side of (10.11) is bounded by a constant. Let $0 < \chi < 1$. By (10.4), there exists $N < \infty$ such that for all $n \geq N$,

$$\sup_{g\in\mathcal{G}} \|\tilde{\nu}_{a_n}(g)(\omega) - \tilde{\nu}_0(g)(\omega)\| < \chi \text{ and } \left\|Diag(\tilde{h}_{2,a_n}(1_k))(\omega) - I_k\right\| < \chi \tag{10.12}$$

using the fact that $Diag(h_{2,0}(1_k)) = I_k$ by construction. Let $B_\chi(\omega) = \sup_{g\in\mathcal{G}} ||\tilde{\nu}_0(g)(\omega)|| + \chi$. Then, for all $n \geq N$,

$$\sup_{g\in\mathcal{G}} \|\tilde{\nu}_{a_n}(g)(\omega)\| \leq B_\chi(\omega) < \infty. \tag{10.13}$$

First, consider the case where no moment equalities are present, i.e., $v = 0$ and $k = p$. In this case, for $n \geq N$, we have: for all $g \in \mathcal{G}$,

$$\begin{aligned} 0 &\leq S(\tilde{\nu}_{a_n}(g)(\omega) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)) \\ &\leq S(\tilde{\nu}_{a_n}(g)(\omega), \tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)) \\ &\leq S(-B_\chi(\omega)1_p, \varepsilon \cdot Diag(\tilde{h}_{2,a_n}(1_p))) \\ &\leq S(-B_\chi(\omega)1_p, \varepsilon(1-\chi)I_p), \end{aligned} \tag{10.14}$$

where the first inequality holds by Assumption S1(c), the second inequality holds by

Assumption S1(b) and $h_{1,a_n,F_{a_n}}(\theta_{a_n}, g) \geq 0_p$ (which holds because $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$), the third inequality holds by Assumption S1(b) and (10.13) as well as by Assumption S1(e) and the definition of $\tilde{h}^{\varepsilon}_{2,a_n}(g)(\omega)$ in (10.5), and the last inequality holds by Assumption S1(e) and (10.12). For fixed $\omega \in \tilde{\Omega}$, the constant $S(-B_\chi(\omega)1_p, \varepsilon(1-\chi)I_p)$ bounds the first summand on the left-hand side of (10.11) for all $n \geq N$.

For the case where $v > 0$, the third inequality in (10.14) needs to be altered because $S(m, \Sigma)$ is not assumed to be non-increasing in m_{II}, where $m = (m_I', m_{II}')'$. In this case, for the bound with respect to the last v elements of $\tilde{\nu}_{a_n}(g)(\omega)$, denoted by $\tilde{\nu}_{a_n,II}(g)(\omega)$, we use the continuity condition on $S(m, \Sigma)$, i.e., Assumption S1(d), which yields uniform continuity of $S(-B_\chi(\omega)1_p, m_{II}, \varepsilon(1-\chi)I_k)$ over the compact set $\{m_{II} : ||m_{II}|| \leq B_\chi(\omega) < \infty\}$ and delivers a finite bound because $\sup_{g \in \mathcal{G}, n \geq 1} ||\tilde{\nu}_{a_n,II}(g)(\omega)|| \leq B_\chi(\omega)$.

By an analogous but simpler argument, for fixed $\omega \in \tilde{\Omega}$, the second summand on the left-hand side of (10.11) is bounded by a constant.

Hence, the conditions of the BCT hold and for fixed $\omega \in \tilde{\Omega}$, $\tilde{T}_{a_n}(\omega) - \tilde{T}_{a_n,0}(\omega) \to 0$ as $n \to \infty$. Thus, (10.7) holds and the proof is complete. $\square$

For GMS CS's, Theorem 2(a) follows immediately from the following three Lemmas. The PA critical value is a GMS critical value with $\varphi_n(x) = 0$ for all $x \in R$ and this function $\varphi_n(x)$ satisfies Assumption GMS1 (though not Assumption GMS2(b)). Hence, Theorem 2(a) for GMS CS's covers PA CS's.

Lemma A2. *Suppose Assumptions* M, S1, *and* S2 *hold. Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$ *and all* $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\left(T_n(\theta) > c_0(h_{n,F}(\theta), 1-\alpha) + \delta\right) \leq \alpha.$$

Lemma A3. *Suppose Assumptions* M, S1, *and* GMS1 *hold. Then, for every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$,

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\left(c(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha) < c(h_{1,n,F}(\theta), \widehat{h}_{2,n}(\theta), 1-\alpha)\right) = 0.$$

Lemma A4. *Suppose Assumptions* M, S1, *and* S2 *hold. Then, for every compact* *subset $\mathcal{H}_{2,cpt}$ of $\mathcal{H}_2$ and for all $0<\delta<\eta$ (where η is as in the definition of $c(h,1-\alpha)$),*

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\Big(c(h_{1,n,F}(\theta),\widehat{h}_{2,n}(\theta),1-\alpha)<c_0(h_{1,n,F}(\theta),h_{2,F}(\theta),1-\alpha)+\delta\Big)=0.$$

The following Lemma is used in the proof of Lemma A4.

Lemma A5. *Suppose Assumptions* M, S1, *and* S2 *hold. Let $\{h_{2,n}:n\geq 1\}$ and $\{h_{2,n}^*:n\geq 1\}$ be any two sequences of $k\times k$-valued covariance kernels on $\mathcal{G}\times\mathcal{G}$ such that $d(h_{2,n},h_{2,n}^*)\to 0$ and $d(h_{2,n},h_{2,0})\to 0$ for some $k\times k$-valued covariance kernel $h_{2,0}$ on $\mathcal{G}\times\mathcal{G}$. Then, for all $\eta_1>0$ and all $\delta>0$,*

$$\liminf_{n\to\infty}\inf_{h_1\in\mathcal{H}_1}\big[c_0(h_1,h_{2,n},1-\alpha+\eta_1)+\delta-c_0(h_1,h_{2,n}^*,1-\alpha)\big]\geq 0.$$

Proof of Lemma A2. For all $\delta>0$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\left(T_n(\theta)>c_0(h_{n,F}(\theta),1-\alpha)+\delta\right)\\
&\leq \limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} \left[P_F\left(T_n(\theta)>c_0(h_{n,F}(\theta),1-\alpha)+\delta\right)\right.\\
&\qquad\qquad\left. -P\left(T(h_{n,F}(\theta))>c_0(h_{n,F}(\theta),1-\alpha)\right)\right]\\
&\quad+\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P\left(T(h_{n,F}(\theta))>c_0(h_{n,F}(\theta),1-\alpha)\right)\\
&\leq 0+\alpha,
\end{aligned}\tag{10.15}$$

where the second inequality holds by Theorem 1(a) with $x_{h_{n,F}(\theta)}=c_0(h_{n,F}(\theta),1-\alpha)+\delta$ and by the definition of the quantile $c_0(h_{n,F}(\theta),1-\alpha)$ of $T(h_{n,F}(\theta))$. $\square$

Proof of Lemma A3. Let $\{(\theta_n,F_n)\in\mathcal{F}:n\geq 1\}$ be a sequence for which $h_{2,F_n}(\theta_n)\in\mathcal{H}_{2,cpt}$ and the probability in the statement of the Lemma evaluated at (θ_n,F_n) differs from its supremum over $(\theta,F)\in\mathcal{F}$ (with $h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}$) by δ_n or less, where $0<\delta_n\to 0$ as $n\to\infty$. Such a sequence always exists. It suffices to show

$$\lim_{n\to\infty} P_{F_n}\left(c(\varphi_n(\theta_n),\widehat{h}_{2,n}(\theta_n),1-\alpha)<c(h_{1,n,F_n}(\theta_n),\widehat{h}_{2,n}(\theta_n),1-\alpha)\right)=0. \tag{10.16}$$

By the compactness of $\mathcal{H}_{2,cpt}$, given any subsequence $\{u_n:n\geq 1\}$ of $\{n\}$, there exists a subsubsequence $\{a_n : n \geq 1\}$ for which $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ as $n \to \infty$ for some $h_{2,0} \in \mathcal{H}_{2,cpt}$. This and $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ for all $n \geq 1$ implies that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$. Hence, it suffices to show

$$\lim_{n\to\infty} P_{F_{a_n}} \left(c(\varphi_{a_n}(\theta_{a_n}), \widehat{h}_{2,a_n}(\theta_{a_n}), 1-\alpha) < c(h_{1,a_n,F_{a_n}}(\theta_{a_n}), \widehat{h}_{2,a_n}(\theta_{a_n}), 1-\alpha) \right) = 0 \tag{10.17}$$

for $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$.

By Lemma A1(a), for $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_2)$, we have

$$\nu_{a_n,F_{a_n}}(\theta_{a_n}, \cdot) \Rightarrow \nu_{h_{2,0}}(\cdot) \text{ as } n \to \infty. \tag{10.18}$$

We now show that for all sequences $\tau_n \to \infty$ as $n \to \infty$, we have

$$\lim_{n\to\infty} P_{F_{a_n}} \left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)| > \tau_{a_n} \right) = 0, \tag{10.19}$$

where $\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)$ denotes the jth element of $\nu_{a_n,F_{a_n}}(\theta_{a_n}, g)$. We show this by noting that (10.18) and the continuous mapping theorem give: $\forall \tau > 0$,

$$\lim_{n\to\infty} P_{F_{a_n}} \left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{a_n,F_{a_n},j}(\theta_{a_n}, g)| > \tau \right) = P \left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{h_{2,0},j}(g)| > \tau \right), \tag{10.20}$$

where $\nu_{h_{2,0},j}(g)$ denotes the jth element of $\nu_{h_{2,0}}(g)$. In addition, the sample paths of $\nu_{h_{2,0},j}(\cdot)$ are bounded a.s., which yields $1\left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{h_{2,0},j}(g)| > \tau\right) \to 0$ as $\tau \to \infty$ a.s. Hence, by the bounded convergence theorem,

$$\lim_{\tau\to\infty} P \left(\sup_{g\in\mathcal{G}, j\leq p} |\nu_{h_{2,0},j}(g)| > \tau \right) = 0. \tag{10.21}$$

Equations (10.20) and (10.21) imply (10.19).

Next, we have

$$\begin{aligned}
\xi_{a_n}(\theta_{a_n}, g) &= \kappa_{a_n}^{-1} \left(\overline{D}_{a_n}^{-1/2}(\theta_{a_n}, g) D_{F_{a_n}}^{1/2}(\theta_{a_n}) \right) a_n^{1/2} D_{F_{a_n}}^{-1/2}(\theta_{a_n}) \overline{m}_{a_n}(\theta_{a_n}, g) \\
&= \kappa_{a_n}^{-1} Diag^{-1/2}(\overline{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g)) Diag^{1/2}(\overline{h}_{2,a_n}(\theta_{a_n}, g)) \\
&\quad \times Diag^{-1/2}(\overline{h}_{2,a_n}(\theta_{a_n}, g)) \left(\nu_{a_n,F_{a_n}}(\theta_{a_n}, g) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g) \right) \\
&= \kappa_{a_n}^{-1} \left((I_k + o_p(1))\, Diag^{-1/2}(\overline{h}_{2,a_n}(\theta_{a_n}, g)) \left(\nu_{a_n,F_{a_n}}(\theta_{a_n}, g) + h_{1,a_n,F_{a_n}}(\theta_{a_n}, g) \right) \right),
\end{aligned} \tag{10.22}$$

where the second equality holds by the definitions of $\overline{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g)$, $\nu_{a_n,F_{a_n}}(\theta_{a_n},g)$, and $h_{1,a_n,F_{a_n}}(\theta_{a_n},g)$ in (5.2) and $\overline{D}_n(\theta,g) = Diag(\overline{\Sigma}_n(\theta,g))$ and the third equality holds by Lemma A1(b) using the fact that $\overline{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g)$ is a function of $\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g)$, see (5.2), and Definition SubSeq(h_2).

Let $\tau_n = (\kappa_n/\zeta - B_n)/2$. By Assumption GMS1(b), $\tau_n = (\kappa_n - \zeta B_n)/2\zeta \to \infty$ as $n \to \infty$. Also,

$$(\kappa_n/\zeta - \tau_n) - B_n = (\kappa_n/\zeta + B_n)/2 - B_n = \tau_n \to \infty \text{ as } n \to \infty. \tag{10.23}$$

For τ_n defined in this way, we have

$$\begin{aligned}
& P_{F_{a_n}}\left(c(\varphi_{a_n}(\theta_{a_n}),\widehat{h}_{2,a_n}(\theta_{a_n}),1-\alpha) < c(h_{1,a_n,F_{a_n}}(\theta_{a_n}),\widehat{h}_{2,a_n}(\theta_{a_n}),1-\alpha)\right) \\
&\le P_{F_{a_n}}\left(\varphi_{a_n,j}(\theta_{a_n},g) > h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) \text{ for some } j \le p, \text{ some } g \in \mathcal{G}\right) \\
&\le P_{F_{a_n}}\left(\begin{array}{l} \xi_{a_n,j}(\theta_{a_n},g) > 1 \ \&\ h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) < \overline{h}_{2,a_n,F_{a_n},j}(\theta_{a_n},g)^{1/2}B_{a_n} \\ \text{for some } j \le p, \text{ some } g \in \mathcal{G} \end{array}\right) \\
&\le P_{F_{a_n}}\left(\begin{array}{l} (1+o_p(1))[\overline{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n},g)\nu_{a_n,F_{a_n},j}(\theta_{a_n},g) + \overline{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g)] \\ > \kappa_{a_n} \ \&\ h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) < \overline{h}_{2,a_n,F_{a_n},j}(\theta_{a_n},g)^{1/2}B_{a_n} \text{ for some } j \le p, \text{ some } g \in \mathcal{G} \end{array}\right) \\
&\le P_{F_{a_n}}\left(\begin{array}{l} (1+o_p(1))[\tau_{a_n} + \overline{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g)] > \kappa_{a_n} \ \& \\ \overline{h}^{-1/2}_{2,a_n,F_{a_n}}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) < B_{a_n} \text{ for some } j \le p, \text{ some } g \in \mathcal{G} \end{array}\right) \\
&\quad + P_{F_{a_n}}\left(\sup_{g \in \mathcal{G}, j \le p} |\overline{h}^{-1/2}_{2,a_n,F_{a_n},j}(\theta_{a_n},g)\nu_{a_n,F_{a_n},j}(\theta_{a_n},g)| > \tau_{a_n}\right) \\
&\le P_{F_{a_n}}\left(\begin{array}{l} \overline{h}^{-1/2}_{2,a_n,F_{a_n}}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) > \kappa_{a_n}/\zeta - \tau_{a_n} \ \& \\ \overline{h}^{-1/2}_{2,a_n,F_{a_n}}(\theta_{a_n},g)h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) < B_{a_n} \text{for some } j \le p, \text{ some } g \in \mathcal{G} \end{array}\right) + o(1) \\
&= o(1),
\end{aligned} \tag{10.24}$$

where the first inequality holds because $c_0(h, 1-\alpha+\eta)$ and $c(h, 1-\alpha)$ are non-increasing in the first p elements of h_1 by Assumption S1(b), the second inequality holds because $(\theta_{a_n}, F_{a_n}) \in \mathcal{F}$ implies that $h_{1,a_n,F_{a_n},j}(\theta_{a_n},g) \ge 0\ \forall j \le p, \forall g \in \mathcal{G}$ and Assumption GMS1(a) implies that (i) $\varphi_{a_n,j}(\theta_{a_n},g) = 0 \le h_{1,a_n,F_{a_n}}(\theta_{a_n},g)$ whenever $\xi_{a_n,j}(\theta_{a_n},g) \le 1$ and (ii) $\varphi_{a_n,j}(\theta_{a_n},g) \le \overline{h}_{2,a_n,F_{a_n},j}(\theta_{a_n},g)^{1/2}B_{a_n}$ a.s. $\forall j \le p, \forall g \in \mathcal{G}$, the third inequality holds by (10.22), the fourth inequality holds because $P(A) \le P(A \cap B) + P(B^c)$, the last inequality holds with probability that goes to one as $n \to \infty$ (wp$\to$ 1) because $\kappa_{a_n}/(1+o_p(1)) > \kappa_{a_n}/\zeta$ wp$\to$ 1 for $\zeta > 1$ and using (10.19) with τ_{a_n} replaced by

$\varepsilon^{1/2}\tau_{a_n}/2$ because $\overline{h}_{2,a_n,F_{a_n},j}^{-1/2}(\theta_{a_n},g) \leq \varepsilon^{-1/2}h_{2,j}^{-1/2}(1_k,1_k)(1+o_p(1)) = \varepsilon^{-1/2}(1+o_p(1))$ by Lemma A1(b) and (5.2), and the equality holds using (10.23).

Hence, (10.17) holds and the Lemma is proved. $\square$

Proof of Lemma A4. The result of the Lemma is equivalent to

$$\lim_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}:\\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} P_F\Big(c_0(h_{1,n,F}(\theta),\widehat{h}_{2,n}(\theta),1-\alpha+\eta) < c_0(h_{1,n,F}(\theta),h_{2,F}(\theta),1-\alpha)-\varepsilon^*\Big) = 0, \tag{10.25}$$

where $\varepsilon^* = \eta - \delta > 0$. By considering a sequence $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ that is within $\delta_n \to 0$ of the supremum in (10.25) for all $n \geq 1$, it suffices to show that

$$\lim_{n\to\infty} P_{F_n}\Big(c_0(h_{1,n,F_n}(\theta_n),\widehat{h}_{2,n}(\theta_n),1-\alpha+\eta) < c_0(h_{1,n,F_n}(\theta_n),h_{2,F_n}(\theta_n),1-\alpha)-\varepsilon^*\Big) = 0. \tag{10.26}$$

Given any subsequence $\{u_n\}$ of $\{n\}$, there exists a subsubsequence $\{a_n\}$ such that $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ as $n \to \infty$ for some $h_{2,0} \in \mathcal{H}_{2,cpt}$ because $h_{2,F_n}(\theta_n) \in \mathcal{H}_{2,cpt}$. Hence, it suffices to show that (10.26) holds with a_n in place of n.

The condition $d(h_{2,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to 0$ and $(\theta_n, F_n) \in \mathcal{F}$ for all $n \geq 1$ imply that $\{(\theta_{a_n}, F_{a_n}) : n \geq 1\} \in SubSeq(h_{2,0})$. Hence, by Lemma A1(b), $d(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}), h_{2,0}) \to_p 0$ as $n \to \infty$. Furthermore,

$$\begin{aligned}
&\widehat{h}_{2,a_n}(\theta_{a_n},g,g^*)\\
&= \widehat{D}_{a_n}^{-1/2}(\theta_{a_n})\widehat{\Sigma}_{a_n}(\theta_{a_n},g,g^*)\widehat{D}_{a_n}^{-1/2}(\theta_{a_n})\\
&= Diag(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},1_k))^{-1/2}\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},g,g^*)Diag(\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n},1_k))^{-1/2}.
\end{aligned} \tag{10.27}$$

Hence, $d(\widehat{h}_{2,a_n}(\theta_{a_n}), h_{2,0}) \to_p 0$ as $n \to \infty$. Given this, using the almost sure representation theorem as above, we can construct $\{\tilde{h}_{2,a_n}(g,g^*) : g, g^* \in \mathcal{G}\}$ such that $d(\tilde{h}_{2,a_n}, h_{2,0}) \to 0$ as $n \to \infty$ a.s. and $\tilde{h}_{2,a_n}$ and $\widehat{h}_{2,a_n}(\theta_{a_n})$ have the same distribution under (θ_{a_n}, F_{a_n}) for all $n \geq 1$.

For fixed ω in the underlying probability space such that $d(\tilde{h}_{2,a_n}(\cdot,\cdot)(\omega), h_{2,0}) \to 0$ as $n \to \infty$, Lemma A5 with $h_{2,n} = \tilde{h}_{2,a_n}(\omega)$ $(= \tilde{h}_{2,a_n}(\cdot,\cdot)(\omega))$, $h_{2,n}^* = h_{2,F_{a_n}}(\theta_{a_n})$, $h_{2,0} = h_{2,0}$, and $\eta_1 = \eta$ gives: for all $\delta > 0$,

$$\liminf_{n\to\infty} \Big[c_0(h_{1,a_n,F_{a_n}}(\theta_{a_n}), \tilde{h}_{2,a_n}(\omega), 1-\alpha+\eta) + \delta \\ -c_0(h_{1,a_n,F_{a_n}}(\theta_{a_n}), h_{2,F_{a_n}}(\theta_{a_n}), 1-\alpha)\Big] \geq 0. \tag{10.28}$$

Equation (10.28) holds a.s. This implies that (10.26) holds with a_n in place of n because (i) $\tilde{h}_{2,a_n}$ and $\widehat{h}_{2,a_n}(\theta_{a_n})$ have the same distribution for all $n \geq 1$ and (ii) for any sequence of sets $\{A_n : n \geq 1\}$, $P(A_n\ ev.)$ $(= P(\cup_{m=1}^{\infty} \cap_{k=m}^{\infty} A_k)) = 1$ (where ev. abbreviates eventually) implies that $P(A_n) \to 1$ as $n \to \infty$. $\square$

Proof of Lemma A5. Below we show that for $\{h_{2,n}\}$ and $\{h_{2,n}^*\}$ as in the statement of the Lemma, for all constants $x_{h_1,h_{2,n}^*} \in R$ that may depend on $h_1 \in \mathcal{H}_1$ and $h_{2,n}^*$, and all $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \Big[P(T(h_1,h_{2,n}) \leq x_{h_1,h_{2,n}^*}) - P(T(h_1,h_{2,n}^*) \leq x_{h_1,h_{2,n}^*} + \delta)\Big] \leq 0. \tag{10.29}$$

Note that this result is similar to those of Theorem 1.

We use (10.29) to obtain: for all $\delta > 0$ and $\eta_1 > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} P(T(h_1,h_{2,n}) \leq c_0(h_1,h_{2,n}^*,1-\alpha) - \delta) \\
&\leq \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \big[P(T(h_1,h_{2,n}) \leq c_0(h_1,h_{2,n}^*,1-\alpha) - \delta) \\
&\qquad\qquad -P(T(h_1,h_{2,n}^*) \leq c_0(h_1,h_{2,n}^*,1-\alpha) - \delta/2)\big] \\
&\quad +\limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} P(T(h_1,h_{2,n}^*) \leq c_0(h_1,h_{2,n}^*,1-\alpha) - \delta/2) \\
&\leq 0 + 1 - \alpha \\
&< 1 - \alpha + \eta_1,
\end{aligned} \tag{10.30}$$

where the second inequality holds by (10.29) with $\delta/2$ in place of δ and $x_{h_1,h_{2,n}^*} = c_0(h_1,h_{2,n}^*,1-\alpha) - \delta$ and by the definition of the $1-\alpha$ quantile of $T(h_1,h_{2,n}^*)$.

We now use (10.30) to show by contradiction that the result of the Lemma holds. Suppose the result of the Lemma does not hold. Then, there exist constants $\delta > 0$ and $\varepsilon^* > 0$, a subsequence $\{a_n : n \geq 1\}$, and a sequence $\{h_{1,a_n} \in \mathcal{H}_1 : n \geq 1\}$ such that

$$\lim_{n\to\infty} \left[c_0(h_{1,a_n},h_{2,a_n},1-\alpha+\eta_1) + \delta - c_0(h_{1,a_n},h_{2,a_n}^*,1-\alpha)\right] \leq -\varepsilon^* < 0. \tag{10.31}$$

Using this and (10.30), we have

$$
\begin{aligned}
&\limsup_{n\to\infty} P(T(h_{1,a_n}, h_{2,a_n}) \le c_0(h_{1,a_n}, h_{2,a_n}, 1-\alpha+\eta_1)+\delta) \\
&\le \limsup_{n\to\infty} P(T(h_{1,a_n}, h_{2,a_n}) \le c_0(h_{1,a_n}, h^*_{2,a_n}, 1-\alpha) - \varepsilon^*/2) \\
&\le \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} P(T(h_1, h_{2,a_n}) \le c_0(h_1, h^*_{2,a_n}, 1-\alpha) - \varepsilon^*/2) \\
&< 1-\alpha+\eta_1, \qquad (10.32)
\end{aligned}
$$

where the first inequality holds by (10.31) and the last inequality holds by (10.30) with $\varepsilon^*/2$ in place of δ.

Equation (10.32) is a contradiction to (10.31) because the left-hand side quantity in (10.32) (without the $\limsup_{n\to\infty}$) is greater than or equal to $1-\alpha+\eta_1$ for all $n \ge 1$ by the definition of the $1-\alpha+\eta_1$ quantile $c_0(h_{1,a_n}, h_{2,a_n}, 1-\alpha+\eta_1)$ of $T(h_{1,a_n}, h_{2,a_n})$. This completes the proof of the Lemma except for establishing (10.29).

To establish (10.29), we write

$$
\begin{aligned}
&\limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \Big[P(T(h_1, h_{2,n}) \le x_{h_1,h^*_{2,n}}) - P(T(h_1, h^*_{2,n}) \le x_{h_1,h^*_{2,n}} + \delta)\Big] \qquad (10.33) \\
&\le \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \Big[P(T(h_1, h_{2,n}) \le x_{h_1,h^*_{2,n}}) - P(T(h_1, h_{2,0}) \le x_{h_1,h^*_{2,n}} + \delta/2)\Big] \\
&+ \limsup_{n\to\infty} \sup_{h_1\in\mathcal{H}_1} \Big[P(T(h_1, h_{2,0}) \le x_{h_1,h^*_{2,n}} + \delta/2) - P(T(h_1, h^*_{2,n}) \le x_{h_1,h^*_{2,n}} + \delta)\Big].
\end{aligned}
$$

The first summand on the right-hand side of (10.33) is less than or equal to 0 by the same argument as used to prove Theorem 1(a) with $\nu_{a_n,F_{a_n}}(\theta_{a_n},\cdot)$ replaced by $\nu_{h_{2,a_n}}(\cdot)$ in (10.3), where $\nu_{h_{2,a_n}}(\cdot)$ is defined in (4.2), because $d(h_{2,a_n}, h_{2,0}) \to 0$ as $n \to \infty$ implies that the Gaussian processes $\nu_{h_{2,a_n}}(\cdot) \Rightarrow \nu_{h_{2,0}}(\cdot)$ as $n \to \infty$. This argument uses Assumption S2.

Similarly, the second summand on the right-hand side of (10.33) is less than or equal to 0 by an argument analogous to that for Theorem 1(b). Hence, (10.29) is established, which completes the proof. $\square$

References

ANDREWS, D. W. K., S. T. BERRY, AND P. JIA (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," unpublished manuscript, Cowles Foundation, Yale University.

ANDREWS, D. W. K., X. CHENG, AND P. GUGGENBERGER (2009): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," unpublished manuscript, Cowles Foundation, Yale University.

ANDREWS, D. W. K. AND P. GUGGENBERGER (2009): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory,* 25, 669-709.

———(2010): "Asymptotic Size and a Problem with Subsampling and with the m Out of n Bootstrap," *Econometric Theory,* 26, 426-468.

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

ANDREWS, D. W. K. AND P. JIA (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," Cowles Foundation Discussion Paper No. 1676, Yale University.

ANDREWS, D. W. K. AND X. SHI (2009): "Supplement to 'Inference Based on Conditional Moment Inequalities,'" unpublished manuscript, Cowles Foundation, Yale University.

———(2010): "Inference Based on (Possibly Infinitely) Many Conditional Moment Inequalities,'" unpublished manuscript, Cowles Foundation, Yale University.

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

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

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica,* 76, 763-814.

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica,* 60, 889-917.

Bierens, H. (1982): "Consistent Model Specification Tests," *Journal of Econometrics,* 20, 105-134.

Bierens, H. and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica,* 65, 1129-1152.

Blundell, R. , A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica,* 75, 323-363.

Bontemps, C., T. Magnac, and E. Maurin (2007): "Set Identified Linear Models," unpublished manuscript, Toulouse School of Economics.

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

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

Chen, X. and Y. Fan (1999): "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics,* 91, 373-401.

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

Chernozhukov, V., S. Lee, and A. Rozen (2008): "Inference with Intersection Bounds," unpublished manuscript, Department of Economics, University College London.

Chiburis, R. C. (2008): "Approximately Most Powerful Tests for Moment Inequalities," unpublished manuscript, Department of Economics, Princeton University.

Ciliberto, F. and E. T. Tamer (2009): "Market Structure and Multiple Equilibria in the Airline Industry," *Econometrica,* 77, 1791-1828.

Dominguez, M. and I. Lobato (2004): "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica,* 72, 1601-1615.

Fan, Y. (2008): "Confidence Sets for Parameters Defined by Conditional Moment Inequalities/Equalities," unpublished manuscript, Vanderbilt University.

Galichon, A. and M. Henry (2009a): "A Test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics,* 152, 186-196.

———(2009b): "Set Identification in Models with Multiple Equilibria," unpublished working paper, available at SSRN: http:\\ssrn.com\abstract=1134762.

Guggenberger, P., J. Hahn, and K. Kim (2008): "Specification Testing Under Moment Inequalities," *Economics Letters,* 99, 375-378.

Haile, P. A. and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy,* 111, 1-51.

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

Kahn, S. and E. Tamer (2009): "Inference on Randomly Censored Regression Models Using Conditional Moment Inequalities," *Journal of Econometrics,* 152, 104-119.

Kim, K. (2008): "Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities," unpublished manuscript, University of Minnesota.

Lee, M.-J. and B. Melenberg (1998): "Bounding Quantiles in Sample Selection Models," *Economics Letters,* 61, 29-35.

Lee, S. and Y.-J. Whang (2009): "Nonparametric Tests of Conditional Treatment Effects," unpublished manuscript, Department of Economics, Seoul National University.

Leeb, H. and B. M. Potscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory,* 21, 21–59.

LEHMANN, E. L. (1959): *Testing Statistical Hypotheses.* New York: Wiley.

KABAILA, P. (1995): "The Effect of Model Selection on Confidence Regions and Prediction Regions," *Econometric Theory,* 11, 537–549.

MANSKI, C. F. (1994): "The Selection Problem," Chap. 4 in *Advances in Econometrics*: Sixth World Congress, ed. by C. A. Sims. Cambridge: Cambridge University Press.

MANSKI, C. F. AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica,* 68, 997-1010.

MANSKI, C. F. AND E. TAMER (2002): "Inference on Regression with Interval Data on a Regressor or Outcome," *Econometrica,* 70, 519-546.

MENZEL, K. (2008): "Estimation and Inference with Many Moment Inequalities," unpublished manuscript, MIT.

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

MOON, H. R. AND F. SCHORFHEIDE (2006): "Boosting Your Instruments: Estimation with Overidentifying Inequality Moment Conditions," unpublished working paper, Department of Economics, University of Southern California.

———(2009): "Bayesian and Frequentist Inference in Partially Identified Models," unpublished manuscript, Department of Economics, University of Pennsylvania.

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

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2006): "Moment Inequalities and Their Application," unpublished manuscript, Department of Economics, Harvard University.

POLLARD, D. (1990): *Empirical Process Theory and Application, NSF-CBMS Regional Conference Series in Probability and Statistics,* Vol. II. Institute of Mathematical Statistics.

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

Romano, J. P. and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Inference and Planning (Special Issue in Honor of T. W. Anderson),* 138, 2786-2807.

———(2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica,* 78, 169-211.

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

Stinchcombe, M. and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory,* 14, 295-325.

Stoye, J. (2010): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica,* 77, 1299-1315.

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