INFERENCE BASED ON CONDITIONAL MOMENT INEQUALITIES

By

Donald W.K. Andrews and Xiaoxia Shi

June 2010
Revised May 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1761RR

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
University of Wisconsin, Madison

November 2007
Revised: May 2012

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-0751517 and SES-1058376. The authors thank the co-editor, four referees, 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 typically have power against $n^{-1/2}$-local alternatives to some, but not all, sequences of distributions in the null hypothesis. Monte Carlo simulations for five 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 a finite number 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.[1,2]

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.

The choice of critical values is important for all moment inequality tests. Here we consider critical values based on generalized moment selection (GMS), as in Andrews and Soares (2010).[3] The GMS critical values can be implemented using the asymptotic

[1] The "identified set" is 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.

[2] There is a potential first-order loss in information when moving from conditional to unconditional moments with moment inequalities because of partial identification. That is, the size of the identified set typically increases. In contrast, if point-identification holds, as with most moment equality models, there is only a second-order loss in information when moving from conditional to unconditional moments—one increases the variance of an estimator and decreases the noncentrality parameter of a test.

[3] 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 PA critical values.

Gaussian distribution or the bootstrap.

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 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.[4] Then, one shows that the data-dependent critical value (the GMS critical value in the present case) is greater 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 (with equality for some true distributions). See Section 5.1 for reasons why uniform asymptotic results are crucial for conditional moment inequality models.

Compared to Andrews and Soares (2010), the present paper treats an infinite number of unconditional moments, rather than a finite number. In consequence, the form of the test statistics considered here is somewhat different and the method of establishing uniform asymptotic results is quite different.

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

[4] 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.

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.[5]

The paper and Supplement provide simulation results for a quantile selection model, a binary entry-game model with multiple equilibria, an intersection bound model, a mean selection model, and an interval-outcome linear regression model. In the entry game model, an important feature of our approach is that nuisance parameters that are identified given the null value of the parameter of interest are concentrated out, which reduces the dimensionality of the problem. No other approach in the literature does this.

Across the five models, the simulation results show that the CvM-based CS's outperform the KS-based CS's in terms of false coverage probabilities (FCP's) in almost all cases. The Sum, QLR, and Max versions of the test statistics perform equally well in terms of FCP's in four of the models, while the Max version performs best in the entry game model. The GMS critical values outperform the plug-in asymptotic and subsampling critical values in terms of FCP's in almost all cases considered. The asymptotic and bootstrap versions of the GMS critical values perform similarly in all cases considered.[6] Variations on the base case show a relatively low degree of sensitivity of the coverage probabilities and FCP's in most cases.

In sum, in the five models considered, the CvM/Max statistic coupled with the GMS/Asy critical value perform quite well in an absolute sense and best among the CS's considered. Computation of a test based on this statistic/critical value takes .20 seconds in the base case configuration of the quantile selection model using GAUSS9.0 on a PC with 3.12 Ghz processor. For the entry game model it takes .55 seconds.

In the quantile selection model, we compare the finite-sample performance of the CI based on the CvM/Max statistic and GMS/Asy critical value with the series and local linear-based CI's proposed in Chernozhukov, Lee, and Rosen (2008) (CLR) and

[5] In a model in which it is assumed that the matrix of partial derivatives of the conditional moment functions, viewed as a function of the conditioning vector, say x, is bounded away from zero (at its minimum point given the null value θ_0), the tests may not have power against any $n^{-1/2}$-local alternatives. However, we are not aware of any models in the literature where one could justify such an assumption. For example, in a parametric model for a game with multiple equilibria, it amounts to requiring that all conditioning variables enter the equilibrium conditions or the equilibrium selection rule with coefficients that are bounded away from zero. This does not seem to be a reasonable assumption to impose a priori.

[6] The bootstrap critical values are not computed in the entry game model because they are computationally expensive in this model.

the integrated nonparametric kernel-based CI proposed in Lee, Song, and Whang (2011) (LSW). We consider three different parameter bound functions: flat, kinked, and peaked and three sample sizes $n = 100$, 250, and 500. For the quantile selection model, the CI proposed in this paper (denoted AS) and the LSW CI have good CP performances in all cases (i.e., $\geq .95$ for a nominal 95% CI). The CLR-series CI under-covers for $n = 100$ (i.e., its minimal CP over the three bound cases considered is .889), but has good CP's for $n = 250$ and 500. The CLR-local linear CI under-covers somewhat for all sample sizes. Its minimal CP's over the three bound cases are .855, .916, and .927 for $n = 100$, 250, and 500, respectively. The AS CI has the best FCP performance in the flat bound case and the kinked bound case for $n = 250$ and 500. The CLR CI's have the best FCP performance in the peaked bound case and the kinked bound case with $n = 100$. The LSW CI has worse (higher) FCP's than those of the AS CI in all nine cases considered. Analogous comparisons are made for the mean selection model and the results are roughly similar, see Supplemental Appendix F for details.

In the intersection bound model, the CP's of the AS CI's and the LSW CI are found to be robust to bound functions that have very steep slopes. In contrast, the CLR-series CI exhibits severe under-coverage for all sample sizes considered (viz., $n = 100$, 250, 500, and 1000) and the CLR-local linear CI exhibits substantial under-coverage for sample sizes $n = 100$ and $n = 250$ but reasonable coverage for larger sample sizes.

We expect the tests introduced in this paper to exhibit a curse of dimensionality (with respect to the dimension, d_X, of the conditioning variable X_i) in terms of their power for local alternatives for which the test does not have $n^{-1/2}$-local power. In addition, computation becomes more burdensome when the number of functions g considered increases. In such cases, one needs to be less ambitious when specifying the functions g. We provide some practical recommendations for doing so in Section 9.

In addition to reporting a CS or test, 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 CLR. 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 asymptotically half-median-unbiased estimated sets. That is, the limit infimum of the probability of including a point or any sequence of points in the identified set is greater than or equal to 1/2. This property follows immediately from the uniform asymptotic coverage probability results for the CS's. The level 1/2 CS, however, is not necessarily asymptotically median-unbiased in two directions.[7] Nevertheless, this set is guaranteed not to be asymptotically inward-median biased. CLR also provide bias-reduction methods for set estimators.

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, 2012), Otsu (2006), Pakes, Porter, Ho, and Ishii (2006), Woutersen (2006), Canay (2010), Chernozhukov, Hong, and Tamer (2007), 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), Canay (2010), Andrews and Barwick (2012), and Bontemps, Magnac, and Maurin (2012).

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: CLR, Fan (2008), Kim (2008), and Menzel (2008). An earlier paper by Khan and Tamer (2009) considers moment inequalities in a point-identified model. An earlier paper by Galichon and Henry (2009) 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 test statistics considered by CLR are akin to Härdle and Mammen (1993)-type model specification statistics, which are based on nonparametric regression estimators. In contrast, the test statistics considered here are akin to Bierens (1982)-type statistics used for consistent model specification tests. These approaches have different strengths and weaknesses. Menzel (2008) investigates tests based on a finite number of moment inequalities in which the number of inequalities increases with the sample size. None of the papers above that treat conditional moment inequalities provide contributions (ii) and (iv)-(vi) listed above. Pakes (2010) discusses models that generate conditional moment inequalities.

More recent contributions to the literature on conditional moment inequalities in-

[7] That is, the limit supremum of the probability of including a point or a sequence of points on the boundary of the identified set is not necessarily less than or equal to 1/2.

clude Beresteanu, Molchanov, and Molinari (2010), who provide sharp identification regions for a class of game theory models and corresponding CS's using their support function approach combined with the methods introduced in this paper; Aradillas-López, Gandhi, and Quint (2010), who provide CI's for parameters in an auction model; LSW, who construct CS's based on L^p integrated nonparametric kernel estimators; Ponomareva (2010), who uses nonparametric kernel estimators; Armstrong (2011a,b), who provides rate of convergence results for estimators based on weighted and non-weighted KS-based tests and associated inference methods; Chetverikov (2011), who considers statistics based on kernel estimators with many values of the bandwidth parameter, and Hsu (2011), who provides tests for conditional treatment effects using the methods introduced in this paper. For point-identified models, papers that convert conditional moments into an infinite number of unconditional moments include Bierens (1982), Bierens and Ploberger (1997), Chen and Fan (1999), Dominguez and Lobato (2004), and Khan and Tamer (2009), among others.

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 (2009) for details of the argument.

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, conditional stochastic dominance, and conditional treatment effects, see Lee and Whang (2009). Andrews and Shi (2011) extends the results to allow for nonparametric parameters of interest, such as the value of a function at a point.

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 coverage properties of GMS and PA CS's. Section 6 establishes the consistency of GMS and PA tests against all fixed alternatives. 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 gives a step-by-step description of how to calculate the tests. Section 10 provides the Monte Carlo simulation results.

Supplemental Appendix A provides proofs of the uniform asymptotic coverage probability results for GMS and PA CS's. Supplemental 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 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. Supplemental Appendix C gives proofs of the results stated in the paper, but not given in Supplemental Appendix A. Supplemental Appendix D provides proofs of the results stated in Supplemental Appendix B. Supplemental Appendix E provides a proof of some empirical process results that are used in Supplemental Appendices A, C, and D. Supplemental Appendix F provides the simulation results for the mean selection and interval-outcome regression models and some additional material concerning the Monte Carlo simulation results of Section 10.

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:

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

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.[8]

[8]It makes sense to speak of a "true" parameter θ_0 in the present context because (i) there may

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

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

for some $B < \infty$ and $\delta > 0$, where F_X is the marginal distribution of X_i under F and $\sigma^2_{F,j}(\theta) = Var_F(m_j(W_i,\theta))$.[9] Let $k = p+v$. 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 α. Then, a nominal level $1-\alpha$ CS for the true value θ_0 is

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

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).

[9] 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 $(\theta, F) \in \mathcal{F}$.

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:

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

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 set $\mathcal{G}$ 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

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

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

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

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

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

for some fixed $\varepsilon > 0$. See Section 9, for suitable choices of ε and other tuning parameters given below. 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 $m(W_i, \theta)$.

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

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

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

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

For brevity, in the text of the paper, the discussion and results focus on CvM statistics. Supplemental Appendix B 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:

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

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 $\mathcal{S} = \{(m,\Sigma) : m \in (-\infty,\infty]^p \times R^v,\ \Sigma\in\mathcal{W}\}$.

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 $R^p_{[+\infty]}\times R^v$, 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 $\Sigma_0\in\mathcal{W}$, $\sup_{\mu\in[0,\infty)^p\times\{0\}^v}|S(m+\mu,\Sigma)-S(m_0+\mu,\Sigma_0)|\to 0$ as $(m,\Sigma)\to(m_0,\Sigma_0)$.[10]

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 $\Sigma\in\mathcal{W}$.

[10]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.

Assumptions S1-S4 allow for natural choices like S_1, S_2, and S_3.

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. Here we show how this can be done and provide proofs that it can be without loss of information.

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

$$\begin{aligned}& E_F m_j(W_i, \theta) g_j(X_i) < 0 \text{ for some } j \leq p \text{ or} \\ & E_F m_j(W_i, \theta) g_j(X_i) \neq 0 \text{ for some } j = p+1, ..., k.\end{aligned}$$

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

Lemma 2. *Assumption* CI *implies* $\Theta_F(\mathcal{G}) = \Theta_F$ $\forall F$ *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. Below we construct tests that use the unconditional moments based on $\mathcal{G}$ and that incorporate all of the information in the conditional moments. To do so, we need to make sure that the tests do not ignore some of the functions in $\mathcal{G}$. Assumption Q, introduced below, plays this role. Assumption Q ensures that for every $\theta \notin \Theta_F$ there is a positive measure set of functions $g \in \mathcal{G}$ for which $E_F m(W_i, \theta) g(W_i) < 0$, so that the tests incorporate all of the information based on the conditional moments.

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 under drifting sequences of distributions $\{F_n : n \geq 1\}$. The latter is used in the proof of the uniform coverage probability results for the CS's. The manageability condition is from Pollard (1990). It is defined in Supplemental Appendix E.

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

(b) $E_F G^{\delta_1}(X_i) \leq C$ 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 $\{g_j(X_{n,i}) : g \in \mathcal{G},\ i \leq n, n \geq 1\}$ are manageable with respect to the envelope function $G(X_{n,i})$ for $j = 1, ..., k$, where $\{X_{n,i} : i \leq n, n \geq 1\}$ 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 $n \geq 1$.[11]

Now we give two examples of collections of functions $\mathcal{G}$ that satisfy Assumptions CI and M. Supplemental Appendix B gives three additional examples.

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 Supplemental Appendix B 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:

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

for some positive integer r_0.[12] The terminology "*c-cube*" abbreviates countable cubes.

[11] 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.

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

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

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

$x = (x_1, ..., x_{d_x})'$, $r = (r_1, ..., r_{d_x})'$, $\bar{r} \in (0, \infty]$, 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 $x \in conv(Supp(X_i))$, where $conv(A)$ 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 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 with $\mathcal{G} = \mathcal{G}_{box}$.*

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 Supplemental Appendix B.

Weight Function Q for $\mathcal{G}_{c\text{-}cube}$. There is a one-to-one mapping $\Pi_{c\text{-}cube} : \mathcal{G}_{c\text{-}cube} \to AR = \{(a, r) : a \in \{1, ..., 2r\}^{d_x}$ and $r = r_0, r_0 + 1, ...\}$. Let Q_{AR} be a probability measure on AR. One can take $Q = \Pi_{c\text{-}cube}^{-1} Q_{AR}$. 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 $\Pi_{box} : \mathcal{G}_{box} \to XR = \{(x, r) \in R^{d_x} \times (0, \bar{r})^{d_x}\}$. Let Q_{XR} be a probability measure on XR. Then, $\Pi_{box}^{-1} Q_{XR}$ is a probability measure on $\mathcal{G}_{box}$. One can take $Q = \Pi_{box}^{-1} Q_{XR}$. 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_{XR}. 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_{XR} 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 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 quasi-Monte Carlo approximation of the integral wrt Q. Then, an approximate test statistic obtained by truncation, simulation, or quasi-Monte Carlo 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 quasi-Monte Carlo 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 Supplemental Appendix B, we show that truncation at s_n, simulation based on s_n simulation repetitions, or quasi-Monte Carlo approximation based on s_n terms, where $s_n \to \infty$ as $n \to \infty$, is sufficient to maintain the asymptotic validity of the CvM 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. Truncation may affect the local power of CvM tests against non-$n^{-1/2}$-local alternatives. (Because we do not consider such alternatives in this paper, we do not give a definiteness statement regarding this.)

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 Supplemental Appendix B show that the use of $\mathcal{G}_n$ in place of $\mathcal{G}$ does not affect the asymptotic properties of the test reported there.

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 $[0,\infty]^p \times \{0\}^v$ 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 $[0,\infty]^p \times \{0\}^v$, and ε is as in the definition of $\overline{\Sigma}_n(\theta, g)$ in (3.5).[13] The definition of $T(h)$ in (4.1) applies to CvM test statistics. For the KS test statistic, one replaces $\int ... \, dQ(g)$ by $\sup_{g \in \mathcal{G}} ...$.

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

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

denote the $1-\alpha$ quantile of $T(h)$. For notational simplicity, we often write $c_0(h, 1-\alpha)$ as $c_0(h_1, h_2, 1-\alpha)$ when $h = (h_1, h_2)$. 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 unknown 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.

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

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* 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 constant η 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:

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

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 quasi-Monte Carlo quantity, as discussed in Section 3.5, 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 h_2 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, (i) $\nu^*_n(g) = \widehat{D}_n(\theta)^{-1/2} n^{-1/2} \sum_{i=1}^n (m(W^*_i, \theta, g) - \overline{m}_n(\theta, g))$, where $\{W^*_i : i \leq n\}$ is an i.i.d. bootstrap sample drawn from the empirical distribution of $\{W_i : i \leq n\}$, (ii) $\widehat{\Sigma}^*_n(\theta, g, g^*)$ are defined as in (4.5) with W^*_i in place of W_i, and (iii) $\widehat{h}^*_{2,n}(\theta, g, g^*) = \widehat{D}_n(\theta)^{-1/2} \widehat{\Sigma}^*_n(\theta, g, g^*) \widehat{D}_n(\theta)^{-1/2}$. 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 quasi-Monte Carlo 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 quasi-Monte Carlo 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 $\boldsymbol{\varphi}_{\mathbf{n}}(\boldsymbol{\theta})$

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 $(\theta, F) \in \mathcal{F}$. 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 $\varphi_n(\theta, g) = (\varphi_{n,1}(\theta, g), ..., \varphi_{n,p}(\theta, g), 0, ..., 0)' \in R^k$ by

$$\varphi_{n,j}(\theta, g) = B_n 1(\xi_{n,j}(\theta, g) > 1) \text{ for } j \leq p. \tag{4.10}$$

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

(b) $\kappa_n \to \infty$ and $B_n/\kappa_n \to 0$ 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. See Section 9, for specific choices of κ_n and B_n that satisfy Assumption GMS1.

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 Supplemental Appendix B. 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 10 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, $0_{\mathcal{G}}$ 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 approximate CvM statistics and critical values, defined in Section 3.5, are provided in Supplemental Appendix B. Supplemental Appendix B also gives results for KS tests.

5.1 Motivation for Uniform Asymptotics

The choice of critical values is important for moment inequality tests 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, e.g., see Andrews and Soares (2010, Sections 1 and 4.1). With conditional moment inequalities, slackness comes in the form of a function, which is an infinite-dimensional parameter, whereas with unconditional moment inequalities it is a finite-dimensional parameter.

Potential slackness in the moment inequalities causes a discontinuity in the pointwise asymptotic distribution of typical test statistics. With conditional moment inequalities, one obtains an extreme form of discontinuity of the pointwise asymptotic distribution because two moment inequalities can be arbitrarily close to one another but pointwise asymptotics say that one inequality is irrelevant—because it is infinitessimally slack, but the other is not—because it is binding. In finite samples there is no discontinuity in the distribution of the test statistic. Hence, pointwise asymptotics do not provide good approximations to the finite-sample properties of test statistics in moment inequality models, especially conditional models. Uniform asymptotics are required.

Methods for establishing uniform asymptotics given in Andrews and Guggenberger (2009, 2010) only apply to finite-dimensional nuisance parameters, and hence, are not applicable to conditional moment inequality models. The same is true of the method in Mikusheva (2007). Linton, Song, and Whang (2010) establish uniform asymptotic results in a model where the nuisance parameter is infinite dimensional. However, their results rely on a complicated condition that is hard to verify. For issues concerning uniformity of asymptotics in other econometric models, see Kabaila (1995), Leeb and Pötscher (2005), Mikusheva (2007), and Andrews and Guggenberger (2010).

5.2 Notation

In order to establish uniform asymptotic coverage probability results, we now introduce notation for the population analogues of the sample quantities 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 $[0, \infty]^p \times \{0\}^v$. Let

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

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 $\mathcal{G}^2$.

5.3 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*

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

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

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 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. 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 $g \in \mathcal{G}$.

5.4 Uniform Asymptotic Coverage Probability Results

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

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.[14] 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, we verify that it holds when S is the Sum or Max function, $Q(\{g \in \mathcal{G} : h_{1,\infty,F_c}(\theta_c, g) = 0\}) > 0$, and $\alpha < 1/2$, see Section 13.3 in Supplemental Appendix B. (We conjecture that it also holds when S is the QLR function under these conditions, but we do not have a proof.) Assumption GMS2(c) is satisfied by typical choices of κ_n, such as $\kappa_n = (0.3 \ln n)^{1/2}$.

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

[14] Andrews (2012) shows that even in the simple case of a finite number of unconditional moment conditions tests that are asymptotically similar (in a uniform sense) exist but have very poor power. Hence, asymptotic similarity of tests and CS's in moment inequality models is not a desirable property. See Hirano and Porter (2012) for related results.

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 $(\theta_c, F_c) \in \mathcal{F}$.[15] 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 approximate CvM and KS tests are given in Supplemental Appendix B.

4. Theorem 2 is stated for the case where the parameter of interest, θ, is finite-dimensional. However, Theorem 2 and all of the results below except the local power results also hold for infinite-dimensional parameters θ. However, computation of a CS is noticeably more difficult in the infinite-dimensional case.

5. 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) $P_{F_0}(X_i \in$

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

$\mathcal{X}_{F_0}(\theta_*)) > 0$, 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. Assumption 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. Supplemental Appendix B 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[16]

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

7 Power Against Some $n^{-1/2}$-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.

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.

[16]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.

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

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

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, 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 $h_{1,j}(g) \in [0, \infty]$ for $j = 1, ..., p$ and all $g \in \mathcal{G}$,

(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:

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

If $h_{1,j}(g) = \infty$, then by definition $h_{1,j}(g) + y = \infty$ for any $y \in R$. We have $h_1(g) + \Pi_0(g)\lambda \in (-\infty,\infty]^p \times R^v$. 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 Supplemental Appendix B 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 Supplemental Appendix B 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 13.6 in Supplemental Appendix B when $P_{F_0}(X_i \in \mathcal{X}_{zero}) = 0$.

As discussed in Section 13.6, 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

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

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

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

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

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

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 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.[17] Assumption LA5(b) typically holds because of the absolute continuity of the Gaussian random variables $\nu_{h_2}(g)$ that enter $T(\varphi(\pi_1), h_2)$.[18]

[17] 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 $x = \pi_1(g)$.

[18] 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 $J_{h,\lambda}(c(\varphi(\pi_1),$

The following assumption is used only for PA tests.

Assumption LA6. The distribution function 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,*

(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

---

$h_2, 1-\alpha)-)$, respectively, in Theorem 4(a), where the latter are the limits from the left and right, respectively, of $J_{h,\lambda}(x)$ at $x = c(\varphi(\pi_1), h_2, 1-\alpha)$.

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 10 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 expectation 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 $\widehat{\tau}_n(W_i, \theta) : R^{d_w} \times \Theta \to R^{d_\tau}$ 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.[19]

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

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

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. 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 $n \rightarrow \infty$ 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 $\lim_{n\rightarrow\infty} \sup_{g,g^* \in \mathcal{G}} ||h_{2,F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*)|| = 0$ 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) $\sup_{g,g^* \in \mathcal{G}} ||\widehat{h}_{2,a_n,F_{a_n}}(\theta_{a_n}, g, g^*) - h_2(g, g^*)|| \rightarrow_p 0$ as $n \rightarrow \infty$.

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 $\widehat{\tau}_n(\theta)$.[20] (The proof is the same as that given for Theorem 2 in Supplemental 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, $\widehat{\Sigma}_n(\theta, g)$

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

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.[21]

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.

9 Computation

In this section, we describe how the tests introduced in this paper are computed. For specificity, we focus on tests based on countable cubes and approximate GMS critical values in an i.i.d. context. We describe both the asymptotic distribution and bootstrap implementations of the critical values.

Step 1. Compute the test statistic:

(a) Transform each regressor to lie in $[0, 1]$. Let $X_i^\dagger \in R^{d_X}$ denote the untransformed regressor vector. In the simulations reported below, we transform $X_i^\dagger$ via a shift and rotation and then an application of the standard normal distribution function. Specifically, first compute $\widehat{\Sigma}_{X,n} = n^{-1}\Sigma_{i=1}^n (X_i^\dagger - \overline{X}_n^\dagger)(X_i^\dagger - \overline{X}_n^\dagger)'$, where $\overline{X}_n^\dagger = n^{-1}\Sigma_{i=1}^n X_i^\dagger$. Then, let $X_i = \Phi(\widehat{\Sigma}_{X,n}^{-1/2}(X_i^\dagger - \overline{X}_n^\dagger))$, where $\Phi(x) = (\Phi(x_1), ..., \Phi(x_{d_X}))'$ for $x = (x_1, ..., x_{d_X})' \in R^{d_X}$ and $\Phi(x_j)$ is the standard normal distribution function at x_j for $x_j \in R$.

(b) Specify the functions g. For countable cubes, the functions are $g_{a,r}(x) = 1(x \in C_{a,r})1_k$ for $C_{a,r} \in \mathcal{C}_{c\text{-}cube}$, where $C_{a,r}$ and $\mathcal{C}_{c\text{-}cube}$ are defined in (3.12).

(c) Specify the weight function Q_{AR}. In the simulations, we take it to be uniform on $a \in \{1, ..., 2r\}^{d_x}$ given r, combined with $w(r) = (r^2 + 100)^{-1}$ for $r = 1, ..., r_{1,n}$. (See below regarding the choice of $r_{1,n}$.)

(d) Compute the CvM test statistic, which is defined by

$$\overline{T}_{n,r_{1,n}}(\theta) = \sum_{r=1}^{r_{1,n}} (r^2 + 100)^{-1} \sum_{a \in \{1,...,2r\}^{d_X}} (2r)^{-d_x} S(n^{1/2}\overline{m}_n(\theta, g_{a,r}), \overline{\Sigma}_n(\theta, g_{a,r})), \qquad (9.4)$$

where $S = S_1, S_2$, or S_3, as defined in (3.8)-(3.10), and $\overline{m}_n(\theta, g_{a,r})$ and $\overline{\Sigma}_n(\theta, g_{a,r})$ are

[21] 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.

defined in (3.3)-(3.5) with $\varepsilon = .05$. Alternatively, compute the KS statistic, which is $\sup_{g_{a,r} \in \mathcal{G}_{c\text{-}cube}} S(n^{1/2}\overline{m}_n(\theta, g_{a,r}), \overline{\Sigma}_n(\theta, g_{a,r}))$.

Step 2. Compute the GMS critical value based on the asymptotic distribution:

(a) Compute $\varphi_n(\theta, g_{a,r})$, as defined in (4.10), for $(a, r) \in AR$. We recommend taking $\kappa_n = (0.3 \ln(n))^{1/2}$ and $B_n = (0.4 \ln(n)/ \ln\ln(n))^{1/2}$.

(b) Simulate a $(kN_g) \times \tau_{reps}$ matrix Z of standard normal random variables, where k is the dimension of $m(W_i, \theta)$, $N_g = \sum_{r=1}^{r_{1,n}} (2r)^{d_X}$ is the number of g functions employed in Step 1(d), and τ_{reps} is the number of simulation repetitions used to simulate the asymptotic Gaussian process.

(c) Compute the $(kN_g) \times (kN_g)$ covariance matrix $\widehat{h}_{2,n,mat}(\theta)$ whose elements are the covariances $\widehat{h}_{2,n}(\theta, g_{a,r}, g^*_{a,r})$ defined in (4.5) for functions $g_{a,r}, g^*_{a,r}$ as in Step 1(b), where $a \in \{1, ..., 2r\}^{d_X}$ and $r = 1, ..., r_{1,n}$.

(d) Compute the $(kN_g) \times \tau_{reps}$ matrix $\widehat{\nu}_n(\theta) = \widehat{h}^{1/2}_{2,n,mat}(\theta)Z$. Let $\widehat{\nu}_{n,j}(\theta, g_{a,r})$ denote the element of $\widehat{\nu}_n$ that corresponds to the row indexed by $g_{a,r}$ and column j for $j = 1, ..., \tau_{reps}$.

(e) For $j = 1, ..., \tau_{reps}$, compute the test statistic $\overline{T}_{n,r_{1,n},j}(\theta)$ just as $\overline{T}_{n,r_{1,n}}(\theta)$ is computed in Step 1(d) but with $n^{1/2}\overline{m}_n(\theta, g_{a,r})$ replaced by $\widehat{\nu}_{n,j}(\theta, g_{a,r}) + \varphi_n(\theta, g_{a,r})$.

(f) Take the critical value to be the $1 - \alpha + \eta$ sample quantile of the simulated test statistics $\{\overline{T}_{n,r_{1,n},j}(\theta) : j = 1, ..., \tau_{reps}\}$ plus η, where η is a very small positive constant, such as 10^{-6}. In the simulations, we obtain the same results with $\eta = 0$ as with 10^{-6}.

For the bootstrap version of the critical value, Steps 2(b)-2(e) are replaced by the following steps:

Step 2$_{boot}$. **(b)** Generate B bootstrap samples $\{W^*_{i,b} : i = 1, ..., n\}$ for $b = 1, ..., B$ using the standard nonparametric i.i.d. bootstrap. That is, draw $W^*_{i,b}$ from the empirical distribution of $\{W_\ell : \ell = 1, ..., n\}$ independently across i and b.

(c) For each bootstrap sample, transform the regressors as in Step 1(a) and compute $\overline{m}^*_{n,b}(\theta, g_{a,r})$ and $\overline{\Sigma}^*_{n,b}(\theta, g_{a,r})$ just as $\overline{m}_n(\theta, g_{a,r})$ and $\overline{\Sigma}_n(\theta, g_{a,r})$ are computed, but with the bootstrap sample in place of the original sample.

(d) For each bootstrap sample, compute the bootstrap test statistic $\overline{T}^*_{n,r_{1,n},b}(\theta)$ as $\overline{T}_{n,r_{1,n}}(\theta)$ is computed in Step 1(d) but with $n^{1/2}\overline{m}_n(\theta, g_{a,r})$ replaced by $\widehat{D}_n(\theta)^{-1/2}$ $n^{1/2}(\overline{m}^*_{n,b}(\theta, g_{a,r}) - \overline{m}_n(\theta, g_{a,r})) + \varphi_n(\theta, g_{a,r})$ and with $\overline{\Sigma}_n(\theta, g_{a,r})$ replaced by $\widehat{D}_n(\theta)^{-1/2}$ $\overline{\Sigma}^*_{n,b}(\theta, g_{a,r})\widehat{D}_n(\theta)^{-1/2}$, where $\widehat{D}_n(\theta) = Diag(\widehat{\Sigma}_n(\theta, 1_k, 1_k))$.

(e) Take the critical value to be the $1 - \alpha + \eta$ sample quantile of the bootstrap test statistics $\{\overline{T}^*_{n,r_{1,n},b}(\theta) : b = 1, ..., B\}$ plus η, where η is a very small positive constant, such as 10^{-6}. In the simulations, we obtain the same results with $\eta = 0$ as with 10^{-6}.

The choices of ε, κ_n, and B_n above are based on some experimentation.[22] Smaller values of ε, such as $\varepsilon = .01$, do not perform as well if the expected number of observations per cube (for some cubes) is small, say 15 or less.

For the quantile selection and interval-outcome models, in which X_i is a scalar, we take $r_{1,n} = 7$ when $n = 250$ and obtain quite similar results for $r_{1,n} = 5$, 9, and 11. For the entry game model, in which bivariate regressor indices appear, we take $r_{1,n} = 3$ when $n = 500$ and obtain similar results for $r_{1,n} = 2$ and 4. Based on the simulation results, we recommend taking $r_{1,n}$ so that the expected number of observations in the smallest cubes is between 10 and 20 (when $\varepsilon = .05$). For example, with $(n, d_X, r_{1,n}) = (250, 1, 7)$, $(500, 2, 3)$, and $(1000, 3, 2)$, the expected number of observations in the smallest cells are 17.9, 13.9, and 15.6, respectively.

Note that the number of cubes with side-edge length indexed by r is $(2r)^{d_X}$, where d_X denotes the dimension of the covariate X_i. The computation time is approximately linear in the number of cubes. Hence, it is linear in $\sum_{r=1}^{r_{1,n}} (2r)^{d_X}$.

In Step 1(a), when there are discrete variables in X_i, the sets $C_{a,r}$ can be formed by taking interactions of each value of the discrete variable(s) with cubes based on the other variable(s).[23]

When the dimension, d_X, of X_i is greater than three (or equal to three with n small, say less than 750), the number of cubes is too large to be practical and the expected number of observations per cube is too small, even if $r_{1,n}$ is small. In such cases, we suggest replacing the sets $C_{a,r}$ above with sets that are rectangles with sub-intervals of $[0, 1]$ in 2 dimensions (equal to the two-dimensional cubes in $\mathcal{C}_{c\text{-}cube}$ when $d_X = 2$) and $[0, 1]$ in the other dimensions, and constructing such sets using all possible combinations of 2 dimensions out of d_X dimensions. For example, if $d_X = 6$, then there are $6!/(4!2!) = 15$ combinations of 2 dimensions out of 6. For each choice of 2 dimensions there are 20 cubes if $(r_0, r_{1,n}) = (1, 2)$ and 56 cubes if $(r_0, r_{1,n}) = (1, 3)$, which yields totals of 300 and 840 cubes, respectively, when $d_X = 6$.[24] If the dimension 2 above is increased to 3, 4, ... as $n \to \infty$, then there is no loss in information asymptotically.

[22] These values are the base case values used in the simulations reported below.

[23] See Example 5 in the second subsection of Supplemental Appendix B for details.

[24] For example, with $n = 500$ and $r_{1,n} = 2$, the expected number of observations per cube is 125 or 31.3 depending on the cube. With $n = 1000$ and $r_{1,n} = 3$, the expected number of observations per cube is 250, 62.5, or 15.6. These expected numbers hold for any value of d_X. Computation time is proportional to $(d_X!/(d_X!2!)) \cdot \sum_{r=1}^{r_{1,n}} (2r)^{d_X}$.

10 Monte Carlo Simulations

This section provides simulation evidence concerning the finite-sample properties of the CI's introduced in the paper. We consider five models: a quantile selection model, an intersection bound model, an entry game model with multiple equilibria, a mean selection model, and an interval-outcome linear regression model. For brevity, the results for the fourth and fifth models are reported in Supplemental Appendix F. The results for the fifth model are remarkably similar to those for the "flat bound" version of the quantile selection model, in spite of the substantial differences between the models. The results for the fourth model are similar to those for the quantile selection model.

In all models, we compare different versions of the CI's introduced in the paper. In the quantile selection, intersection bound, and mean selection models, we compare one of the CI's introduced in the paper with CI's introduced in CLR and LSW.

10.1 Tests Considered in the Simulations

In the simulation results reported below, 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.

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, as defined in Section 9. Both asymptotic normal and bootstrap versions of these tests are computed.

In all models we consider the PA/Asy and GMS/Asy critical values. We also consider the PA/Bt, GMS/Bt, and Sub critical values in the quantile selection model and interval-outcome regression model. The critical values are simulated using 5001 repetitions (for each original sample repetition).[25] The "base case" values of κ_n, B_n, and ε for the GMS critical values are specified in Section 9 and are used in all four 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

[25] 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 quantity 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. It is the .95 sample quantile of the subsample statistics, each of which is defined exactly as the full sample statistic is defined but using the subsample in place of the full sample. The number of subsamples considered is 5001. They are drawn randomly without replacement.

repetitions used to compute CP's and FCP's is 5000 for all cases. This yields a simulation standard error of .0031.

We also report results for the CLR-series, CLR-local linear, and LSW CI's. These CI's are computed as described in CLR and LSW. Section 17.1.3 of Supplemental Appendix F provides details.[26] The L^1 version of the LSW CI is employed. The critical values and CP/FCP's are simulated using 5001 and 5000 repetitions, respectively, except when stated otherwise.[27]

The reported FCP's are "CP-corrected" by employing a critical value that yields a CP equal to .95 at the closest point of the identified set (for the same data generating process and the same sample size as used when computing the FCP) if the CP at the closest point is less than .95.[28] 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.

10.2 Quantile Selection Model

10.2.1 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. Selection yields the conditional quantile to be unidentified. We use a *quantile* monotone instrumental variable (QMIV) condition that is a variant of Manski and Pepper's (2000) Monotone Instrumental Variable (MIV) condition to obtain bounds on the conditional quantile. (The MIV condition 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 distribution functions that define the quantiles are naturally bounded between 0 and 1.

[26] For the CLR and LSW CI's, we use the code graciously provided by CLR and LSW. In the quantile selection model, the two-sided CLR CI's are constructed following the method in Example C of the 2011 version of CLR. The CLR CI's use estimated contact sets.

[27] The LSW critical value is not simulated. It uses a standard normal critical value.

[28] Note that FCP's are determined using the same data generating process as CP's. The only difference is the null value θ being considered is not in the identified set with FCP's, whereas it is with CP's.

Other papers that bound quantiles using the natural bounds of distribution functions include Manski (1994), Lee and Melenberg (1998), Blundell, Gosling, Ichimura, and Meghir (2007), and Giustinelli (2010). The QMIV condition differs from the conditions in these papers, except Giustinelli (2010), although it is closely related to them.[29] Giustinelli (2010) derives bounds on unconditional quantiles with a finite-support IV, whereas we consider bounds on conditional quantiles with a continuous (or discrete) IV.

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 contains an ordered set $\mathcal{X}$. We observe $W_i = (Y_i, X_i, T_i)$. 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

[29] 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.

on $X_i = x$.[30]

This assumption may be suitable in the applications mentioned above.

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

$$\begin{aligned}
\tau &= P\left(y_i(t) \leq Q_{y_i(t)|X_i}(\tau|x)|X_i = x\right) \leq P\left(y_i(t) \leq \theta|X_i = x\right) \\
&= P\left(y_i(t) \leq \theta \;\&\; T_i = t|X_i = x\right) + P\left(y_i(t) \leq \theta \;\&\; T_i \neq t|X_i = x\right) \\
&\leq P\left(Y_i \leq \theta \;\&\; T_i = t|X_i = x\right) + P\left(T_i \neq t|X_i = x\right), \qquad (10.1)
\end{aligned}$$

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 $x \geq x_0$,

$$\begin{aligned}
\tau &= P\left(y_i(t) \leq Q_{y_i(t)|X_i}(\tau|x)|X_i = x\right) \geq P\left(y_i(t) \leq \theta|X_i = x\right) \\
&= P\left(y_i(t) \leq \theta \;\&\; T_i = t|X_i = x\right) + P\left(y_i(t) \leq \theta \;\&\; T_i \neq t|X_i = x\right) \\
&\geq P\left(Y_i \leq \theta \;\&\; T_i = t|X_i = x\right), \qquad (10.2)
\end{aligned}$$

where the first and second inequalities hold by Assumption QMIV and $P(A) \geq 0$.

The inequalities in (10.1) and (10.2) impose sharp bounds on θ. They can be rewritten as conditional moment inequalities:

$$\begin{aligned}
E\left(1(X_i \leq x_0)[1(Y_i \leq \theta, T_i = t) + 1(T_i \neq t) - \tau]|X_i\right) &\geq 0 \text{ a.s. and} \\
E\left(1(X_i \geq x_0)[\tau - 1(Y_i \leq \theta, T_i = t)]|X_i\right) &\geq 0 \text{ a.s.} \qquad (10.3)
\end{aligned}$$

For the simulations, we consider the following data generating process (DGP):

$$\begin{aligned}
y_i(1) &= \mu(X_i) + \sigma\left(X_i\right) u_i, \text{ where } \partial\mu\left(x\right)/\partial x \geq 0 \text{ and } \sigma\left(x\right) \geq 0, \\
T_i &= 1\{L\left(X_i\right) + \varepsilon_i \geq 0\}, \text{ where } \partial L\left(x\right)/\partial x \geq 0, \; X_i \sim Unif[0, 2], \\
(\varepsilon_i, u_i) &\sim N(0, I_2), \; X_i \perp (\varepsilon_i, u_i), \; Y_i = y_i(T_i), \text{ and } t = 1. \qquad (10.4)
\end{aligned}$$

[30] 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.

The variable $y_i(0)$ is irrelevant (because Y_i enters the moment inequalities in (10.3) only through $1(Y_i \leq \theta, T_i = t)$) and, hence, is left undefined. With this DGP, 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 (10.3) imply:

$$\begin{aligned} \theta &\geq \underline{\theta}(x) := \mu(x) + \sigma(x)\,\Phi^{-1}\left(1 - [2\Phi(L(x))]^{-1}\right) \text{ for } x \leq 1.5 \text{ and} \\ \theta &\leq \bar{\theta}(x) := \mu(x) + \sigma(x)\,\Phi^{-1}\left([2\Phi(L(x))]^{-1}\right) \text{ for } x \geq 1.5. \end{aligned} \tag{10.5}$$

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 $\left[\sup_{x \leq x_0} \underline{\theta}(x),\ \inf_{x \geq x_0} \bar{\theta}(x)\right]$. The shape of the lower and upper bound functions depends on the μ, σ, and L functions. We consider three specifications, one that yields flat bound functions, another that yields kinked bound functions, and a third that yields peaked bound functions.[31]

The CP or FCP performance of a CI at a particular value θ depends on the shape of the conditional moment functions, as functions of x, evaluated at θ. In the present model, the conditional moment functions are

$$\beta(x,\theta) = \begin{cases} E\left(1(Y_i \leq \theta, T_i = 1) + 1(T_i \neq 1) - 0.5 | X_i = x\right) & \text{if } x < 1.5 \\ E\left(\tau - 1(Y_i \leq \theta, T_i = 1) | X_i = x\right) & \text{if } x \geq 1.5. \end{cases} \tag{10.6}$$

Figure 1 shows the bound functions and conditional moment functions for the flat, kinked, and peaked cases. The bound functions are given in the upper row. Note that $\underline{\theta}(x)$ is defined only for $x \in [0, 1.5]$ and $\bar{\theta}(x)$ only for $x \in [1.5, 1]$. The conditional moment functions are given in the lower row. The latter are evaluated at the value of θ that yields the lower endpoint of the identified interval.[32]

We consider a base case sample size of $n = 250$. We also report a few results for

[31] For the flat bound DGP, $\mu(x) = 2$, $\sigma(x) = 1$, and $L(x) = 1$ for $x \in [0,2]$. In this case, $\underline{\theta}(x) = 2 + \Phi^{-1}\left(1 - [2\Phi(1)]^{-1}\right)$ for $x \leq 1.5$ and $\bar{\theta}(x) = 2 + \Phi^{-1}\left([2\Phi(1)]^{-1}\right)$ for $x > 1.5$. For the kinked bound DGP, $\mu(x) = 2(x \wedge 1)$, $\sigma(x) = x$, $L(x) = x \wedge 1$, $\underline{\theta}(x) = 2(x \wedge 1) + x \cdot \Phi^{-1}\left(1 - [2\Phi(x \wedge 1)]^{-1}\right)$ for $x \leq 1.5$, and $\bar{\theta}(x) = 2(x \wedge 1) + x \cdot \Phi^{-1}\left([2\Phi(x \wedge 1)]^{-1}\right)$ for $x > 1.5$. The kinked μ and L functions are the same as in the simulation example in Chernozhukov, Lee, and Rosen (2008). For the peaked bound function, $\mu(x) = 2(x \wedge 1)$, $\sigma(x) = x^5$, $L(x) = x \wedge 1$, $\underline{\theta}(x) = 2(x \wedge 1) + x^5\Phi^{-1}\left(1 - [2\Phi(x \wedge 1)]^{-1}\right)$ for $x \leq 1.5$, and $\bar{\theta}(x) = 2(x \wedge 1) + x^5\Phi^{-1}\left([2\Phi(x \wedge 1)]^{-1}\right)$ for $x > 1.5$.

[32] See Supplemental Appendix F for conditional-moment-function figures with θ evaluated at the point at which the FCP's are computed.

$n = 100$, 500, and 1000.

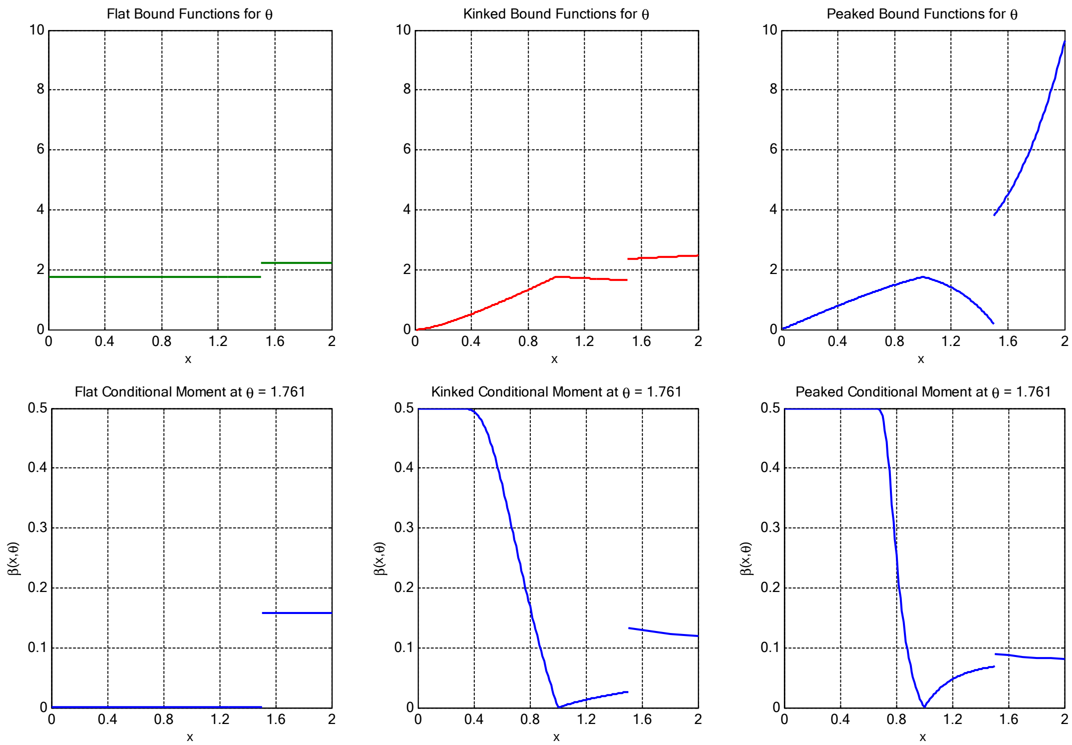

Figure 1. Three Bound Functions on θ and Three Corresponding Conditional Moment Functions for the Quantile Selection Model

10.2.2 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 $X_i \sim U[0,2]$. The regressor X_i is transformed via the method described in Section 9 to lie in $(0,1)$.[33] The hypercubes have side-edge lengths $(2r)^{-1}$ for $r = r_0, ..., r_1$, where $r_0 = 1$ and the base case value of r_1 is 7.[34] The base case number of hypercubes is 56. We also report results for $r_1 = 5$, 9, and 11, which yield 30, 90, and 132 hypercubes, respectively. With $n = 250$ and $r_1 = 7$, the expected number of observations per cube is 125, 62.5, ..., 20.8, or 17.9 depending on the cube. With $n = 250$ and $r_1 = 11$, the expected number also

[33]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 distribution function.

[34]For simplicity, we let r_1 denote $r_{1,n}$ here and below.

can equal 12.5 or 11.4. With $n = 100$ and $r_1 = 7$, the expected number is $50, 25, ..., 8.3$, or 7.3.

10.2.3 Simulation Results: Confidence Intervals Proposed in This Paper

Tables I-III report CP's and CP-corrected FCP's for a variety of test statistics and critical values proposed in this paper for a range of cases. All CI's considered are two-sided CI's for the true value. The CP's are for the lower endpoint of the identified interval in Tables I-III for the flat, kinked, and peaked bound functions.[35] FCP's are for points below the lower endpoint.[36]

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 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 results in Table III 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 CI's have CP's greater than or equal to .95 with flat, kinked, and peaked bound DGP's. The PA/Asy critical values lead to noticeably larger over-coverage than the GMS/Asy critical values with flat and kinked bound DGP's. 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 kinked and peaked 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 GMS/Asy critical values over PA/Asy critical values, (ii) a clear advantage of CvM-based CI's over KS-based CI's in an overall sense when the GMS/Asy critical values are employed, and (iii) little

[35] Supplemental Appendix F provides additional results for the upper endpoints. The results are similar in many respects.

[36] 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, FCP's are computed for θ equal to $\underline{\theta}(1) - c \times sqrt(250/n)$, where $c = .25$, .58, and .61 in the flat, kinked, and peaked bound cases, respectively. These points are chosen to yield similar values for the FCP's across the different cases considered.

Table I. Quantile Selection Model: Base Case 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
Kinked Bound	PA/Asy	.999	.999	.999	.994	.994	.994
	GMS/Asy	.983	.983	.983	.985	.985	.984
Peaked Bound	PA/Asy	1.000	1.000	1.000	.997	.997	.997
	GMS/Asy	.997	.997	.997	.991	.991	.990

(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
Kinked Bound	PA/Asy	.65	.65	.62	.68	.68	.67
	GMS/Asy	.35	.35	.34	.53	.53	.52
Peaked Bound	PA/Asy	.70	.71	.68	.48	.48	.47
	GMS/Asy	.43	.43	.41	.39	.39	.38

difference between the test statistic functions: Sum, QLR, and Max.

Table II compares the critical values PA/Asy, PA/Bt, GMS/Asy, GMS/Asy, and Sub. The results show little differences 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 outperform the PA critical values in terms of FCP's. When using the GMS/Asy or GMS/Bt critical values, the CvM/Max statistic yields lower FCP's than the KS/Max statistic except in the peaked bound case, in which case the difference is relatively small. For the CvM/Max statistic, the GMS critical values also noticeably outperform the Sub critical values in terms of FCP's. However, in the peaked design case, the lowest FCP's are obtained by the KS/Max statistic with the Sub critical value.

Table III provides results for the CvM/Max and KS/Max statistics coupled with the

Table II. Quantile Selection Model: Base Case Critical Value Comparisons

(a) Coverage Probabilities

DGP	Statistic \ Critical Value:	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
Kinked Bound	CvM/Max	.999	.999	.983	.982	.993
	KS/Max	.994	1.00	.984	.982	.950
Peaked Bound	CvM/Max	1.00	1.00	.997	.997	.999
	KS/Max	.997	.998	.990	.990	.965

(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
Kinked Bound	CvM/Max	.62	.64	.34	.33	.47
	KS/Max	.67	.72	.52	.50	.47
Peaked Bound	CvM/Max	.68	.69	.41	.40	.48
	KS/Max	.47	.51	.38	.36	.28

GMS/Asy critical values for several variations of the base case. Table III shows that these CS's perform quite similarly for different sample sizes, different numbers of cubes, and a smaller constant ε.[37] There is some sensitivity to the magnitude of the GMS 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

[37] 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 Base Case

	(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
Base Case ($n = 250, r_1 = 7, \varepsilon = 5/100$)	.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

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 .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 overall 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.

10.2.4 Simulation Results: Comparisons with CLR and LSW Confidence Intervals

Table IV provides comparisons of the CvM/Max/GMS/Asy CI (denoted in this section by AS) with the CLR-series, CLR-local linear, and LSW CI's. Results are reported for the flat, kinked, and peaked bound functions, for the base case sample size

Table IV. Quantile Selection Model: Comparisons of Confidence Intervals Proposed in This Paper with Those Proposed in CLR and LSW

CS	CP (95%)			FCP (corrected)			CP (50%)		
	flat	kink	peak	flat	kink	peak	flat	kink	peak
$n = 100$									
CvM/Max/GMS/Asy	.957	.981	.989	.40	.34	.47	.52	.69	.73
CLR-series	.889	.954	.945	.69	.35	.19	.54	.73	.71
CLR-local linear	.855	.949	.961	.66	.31	.16	.43	.73	.77
LSW	.976	1.000	1.000	.53	.68	.47	.70	.96	.97
$n = 250$									
CvM/Max/GMS/Asy	.951	.983	.997	.37	.34	.41	.52	.72	.82
CLR-series	.939	.972	.979	.65	.39	.18	.57	.80	.79
CLR-local linear	.916	.973	.987	.58	.41	.21	.47	.79	.85
LSW	.979	1.000	1.000	.53	.84	.72	.73	.99	.99
$n = 500$									
CvM/Max/GMS/Asy	.954	.984	.998	.36	.39	.72	.51	.74	.88
CLR-series	.950	.987	.989	.65	.44	.33	.59	.73	.84
CLR-local linear	.927	.985	.996	.62	.49	.47	.50	.80	.91
LSW	.986	1.000	1.000	.54	.92	.95	.75	.99	1.000

250, and for sample sizes 100 and 500.

Table IV shows that the CP performances of the nominal 95% AS and LSW CI's are good (i.e., greater than or equal to .95) for all bound functions and all sample sizes.[38] The CLR-series CI has good CP performance for $n = 250$ and 500, but not for $n = 100$ (in which case its CP is .889 in the flat bound design, which implies that its finite-sample size is less than or equal to .889). The CLR-local linear CI has poor finite-sample size for $n = 100$ (since its CP equals .855 in the flat bound case). For $n = 250$ and 500, its CP's are still low in the flat bound case, being .916 and .927, respectively.

The AS CI has the best (lowest) FCP performance in the flat bound cases for all three sample sizes. The CLR-local linear and CLR-series CI's have the best FCP's in the peaked bound case for all three sample sizes. The AS FCP's are slightly better (lower) than those of the CLR CI's overall in the kinked bound case, with AS performing best with $n = 250$ and $n = 500$ and CLR-local linear performing best with $n = 100$. The

[38] Note that a CP that exceeds .95 is, in and of itself, good. It is only bad if it causes higher FCP's. The latter shows up in the discussion of FCP's, not CP's.

LSW FCP's are noticeably worse (higher) than the AS and CLR FCP's in the kinked bound case.[39,40] The LSW CI has worse FCP's than those of the AS CI in all nine cases considered. This is probably due to its choice of critical value, which is essentially a least favorable critical value.

The 50% AS, CLR-series, and LSW CI's are half-median-unbiased in all of the scenarios considered. The 50% CLR-local linear CI's are "inward biased" in the flat bound case for sample sizes $n = 100$ and 250 with CP's of .43 and .47, respectively (rather than CP's that are greater than or equal to .50). In the flat bound case, the AS CI's are fairly close to being median-unbiased with coverage probabilities of .52, .52, and .51 for the three sample sizes. For the kinked and peaked bound cases, all of the CI's have CP's that exceed .50 by a substantial margin. For all bound functions, the LSW CI's are the farthest from being median unbiased.[41]

In the quantile selection model, the LSW CI's are the quickest CI's to compute, followed by the CLR-series and AS CI's, which are followed by the CLR-local linear CI when $n = 250$ and $n = 500$ and are equalled by it for $n = 100$. Specifically, to compute 5000 tests using 5001 critical value repetitions, the times in minutes (using a 3.33 Ghz processor running GAUSS 6.0) for $n = 100$, 250, and 500 are: .1, .3, .5 for LSW; 10, 11, 12 for CLR-series; 13, 13, 13 for AS; and 12, 28, 62 for CLR-local linear.

[39] The CP correction used in the FCP results in Table IV and elsewhere does not provide complete size correction because it corrects the CP only based on the data generating process (DGP) considered for the particular FCP calculation. More complete finite-sample size correction can be obtained by applying the size correction constants computed for the least favorable DGP considered (which is the flat bound case) when computing the FCP's for the flat, kinked, and peaked bound cases.

For the CLR-series CI with $n = 100$, more complete finite-sample size correction for the three DGP's considered (flat, kinked, peaked) yields size-corrected FCP's for the kinked and peaked cases of .57 and .35, respectively (and no change from the Table IV value of .69 for the flat case). For $n = 250$, the corresponding values are .44 and .21 for the kinked and peaked cases. For the CLR-local linear CI, the more completely corrected FCP's are .57 and .38 for the kinked and peaked cases for $n = 100$, .51 and .29 for $n = 250$, and .57 and .54 for $n = 500$. With more complete size-correction, the AS CI out-performs the CLR-series and CLR-local linear CI's in terms of FCP's in the kinked case for all three sample sizes.

[40] A referee suggested using a hybrid version of the CI method proposed here and that proposed in CLR. Such an approach is possible, but it is beyond the scope of this paper.

[41] The FCP performances of one-sided AS, CLR, and LSW CI's in a mean selection model are roughly similar to those of the two-sided CI's in the quantile selection model (with $n = 250$), see Supplemental Appendix F. In the mean selection model, the minimal CP over the three bound functions is .947 for AS, .918 for CLR-series, .930 for CLR-local linear, and .940 for LSW. The FCP's of the AS and LSW CI's are best in the flat bound case by a large amount over the CLR CI's (.37 for AS versus .68 for CLR-linear). The FCP's of the CLR CI's are better than those of the AS CI by a smaller amount in the kinked and peaked cases (.35 and .38 for AS, .31 and .34 for CLR-series, .30 and .30 for CLR-local linear, and 1.00 and .87 for LSW in the kinked and peaked bound cases, respectively).

10.3 Intersection Bound Example

Next, we carry out some simulations to assess the CP robustness of the AS, CLR, and LSW CI's to bound functions with steep slopes. We consider the same intersection bound model as in the 2011 version of CLR but with two different bound functions. We consider the single moment inequality $E(\theta - Y|X) \geq 0$ a.s. The DGP with the first bound function is

$$Y = L\phi(X^{10}) + u, \tag{10.7}$$

where $X \sim Unif[-2, 2]$, $u = \min\{3, \max\{-3, \sigma^2 v\}\}$, and $v \sim N(0, 1)$. The function $\phi(X^{10})$ yields a near plateau-shaped bound function similar to a smoothed version of $\phi(0)1(X \in [-1, 1])$. The second DGP replaces $\phi(X^{10})$ by $\max\{\phi((X - 1.5)^{10}), \phi((X + 1.5)^{10})\}$, which results in a near double-plateau-shaped bound function similar for $X \in [-2, 2]$ to a smoothed version of $\phi(0)1(X \in [-2, -0.5] \cup [0.5, 2])$. For both DGP's, the identified set is $[L\phi(0), \infty)$. We consider one-sided CI's for θ of the form $[\widehat{lb}_n, \infty)$. We compute CP's at $\theta = L\phi(0)$ and FCP's at $\theta = L\phi(0) - 0.02$.[42] We consider $(L, \sigma) = (1, 0.1)$ and $(5, 0.1)$.[43] We report results for the CvM/Max/GMS/Asy, KS/Max/GMS/Asy CI's, CLR-series, CLR-local linear, and LSW CI's.[44]

The results use 5000 CP/FCP simulation repetitions and 5001 critical value repetitions for each CP/FCP simulation repetition. The results are reported in Table V. In Table V, DGP1 and DGP2 denote the single-plateau DGP with $(L, \sigma) = (1, 0.1)$ and $(5, 0.1)$, respectively. DGP 3 and DGP4 denote the double-plateau DGP with $(L, \sigma) = (1, 0.1)$ and $(5, 0.1)$, respectively.

Table V shows that the CLR-series CI has severe under-coverage for all sample sizes under both the single- and double-plateau DGP's, which suggests that the CLR-series CI may be unreliable when the bound function has very steep slopes (even if it is perfectly smooth). The CLR-local linear CI has severe under-coverage at the smallest sample size, substantial under-coverage for $n = 250$, but the CP's improve as sample size grows large. On the other hand, both versions of the AS CI's and the LSW CI never under-cover.

In terms of (CP-corrected) FCP's, the CvM AS CI is best in DGP3 for all sample sizes and is best in DGP1 and DGP4 for all sample sizes except the smallest, while the

[42] The FCP's are CP-corrected if there is under-coverage, as in the quantile selection model.

[43] These are the same values as in CLR, but for brevity we do not report results for $(L, \sigma) = (0, 0.1)$, which yields the same DGP as in CLR, and $(L, \sigma) = (5, 0.01)$, which is a rather extreme case.

[44] For the AS CI's, the Sum, QLR, and Max test statistics coincide in this example because $k = 1$. The CLR CI's use estimated contact sets.

Table V. Comparison of Nominal 95% AS, CLR, and LSW CI's with Plateau-Bound Functions

		CP (95%)					FCP (CP-corrected)				
		AS		CLR		LSW	AS		CLR		LSW
	n=	CvM	KS	series	loc.lin		CvM	KS	series	loc.lin	
DGP1	100	.986	.986	.734	.804	1.00	.84	.89	.83	.83	1.0
	250	.975	.973	.734	.893	1.00	.57	.67	.75	.69	1.0
	500	.975	.970	.525	.925	1.00	.25	.37	.66	.50	1.0
	1000	.971	.966	.090	.935	1.00	.03	.07	.38	.26	1.0
DGP2	100	1.00	1.00	.207	.713	1.00	1.0	1.0	.90	.89	1.0
	250	1.00	1.00	.057	.856	1.00	1.0	1.0	.87	.73	1.0
	500	1.00	1.00	.004	.908	1.00	.97	.99	.84	.56	1.0
	1000	1.00	1.00	.000	.927	1.00	.70	.89	.72	.33	1.0
DGP3	100	.970	.969	.736	.721	1.00	.70	.79	.83	.84	1.0
	250	.969	.964	.665	.854	1.00	.30	.46	.75	.66	1.0
	500	.963	.957	.436	.900	1.00	.06	.15	.66	.47	1.0
	1000	.969	.963	.089	.927	1.00	.00	.01	.43	.23	1.0
DGP4	100	.998	.999	.241	.655	1.00	.95	.99	.90	.88	1.0
	250	.997	.998	.021	.826	1.00	.66	.83	.89	.70	1.0
	500	.994	.994	.000	.890	1.00	.23	.42	.86	.51	1.0
	1000	.994	.991	.000	.918	1.00	.01	.04	.79	.29	1.0

CLR-local linear CI is best in DGP2 for all sample sizes. The CvM AS CI dominates the KS AS CI in terms of FCP's, and the CLR-local linear CI dominates the CLR-series CI. The LSW CI is dominated by each of the four other CI's in terms of FCP's.

10.4 Entry Game Model

10.4.1 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, see Tamer (2003). In consequence, two conditional moment inequalities and two conditional moment equalities arise. Andrews, Berry, and Jia (2004), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2010), and Galichon and Henry (2011) also consider moment inequalities and equalities in models of this sort.

Following the approach in Section 8, eight non-competitive effects parameters are estimated via a preliminary maximum likelihood estimator based on the number of entrants, similar to Bresnahan and Reiss (1991) and Berry (1992). These estimators are plugged into a set of moment conditions that includes two moment inequalities and two moment equalities.

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

$$\begin{aligned} &X_{i,b}'\tau_b + U_{i,b} \text{ if the other player does not enter and} \\ &X_{i,b}'\tau_b - \theta_b + U_{i,b} \text{ if the other player enters,} \end{aligned} \tag{10.8}$$

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) used to generate the data is a maximum profit ESR (which is unknown to the econometrician and not used by the CS's). 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 when the data is generated by 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

$$
\begin{aligned}
m_1(W_i,\theta,\tau) &= P(X_{i,1}'\tau_1 + U_{i,1} \geq 0,\ X_{i,2}'\tau_2 - \theta_2 + U_{i,2} \leq 0|X_i) - 1(Y_i = (1,0)) \\
&= \Phi(X_{i,1}'\tau_1)\Phi(-X_{i,2}'\tau_2 + \theta_2) - 1(Y_i = (1,0)) \text{ and} \\
m_2(W_i,\theta,\tau) &= P(X_{i,1}'\tau_1 - \theta_1 + U_{i,1} \leq 0,\ X_{i,2}'\tau_2 + U_{i,2} \geq 0|X_i) - 1(Y_i = (0,1)), \\
&= \Phi(-X_{i,1}'\tau_1 + \theta_1)\Phi(X_{i,2}'\tau_2) - 1(Y_i = (0,1)). \qquad (10.9)
\end{aligned}
$$

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

$$
\begin{aligned}
m_3(W_i,\theta,\tau) &= 1(Y_i = (1,1)) - P(X_{i,1}'\tau_1 - \theta_1 + U_{i,1} \geq 0,\ X_{i,2}'\tau_2 - \theta_2 + U_{i,2} \geq 0|X_i), \\
&= 1(Y_i = (1,1)) - \Phi(X_{i,1}'\tau_1 - \theta_1)\Phi(X_{i,2}'\tau_2 - \theta_2), \text{ and} \\
m_4(W_i,\theta,\tau) &= 1(Y_i = (0,0)) - P(X_{i,1}'\tau_1 + U_{i,1} \leq 0,\ X_{i,2}'\tau_2 + U_{i,2} \leq 0|X_i) \\
&= 1(Y_i = (0,0)) - \Phi(-X_{i,1}'\tau_1)\Phi(-X_{i,2}'\tau_2). \qquad (10.10)
\end{aligned}
$$

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 $\{(1(Y_i = (0,0)), 1(Y_i = (1,1)), X_{i,1}, X_{i,2}) : i \leq n\}$.[45]

The model described above is point identified under suitable conditions because τ is identified by the second conditional 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. See Tamer (2003) for some sufficient conditions for point identification.[46] Although the model is point identified, considerable additional information about θ and τ is provided by the moment inequalities in (10.9), as pointed out by Tamer (2003). We exploit this information using the methods employed here.

We show that the gains from exploiting the moment inequalities are substantial by

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

[46] Tamer (2003) uses a large support condition on one regressor in each index $X_{i,1}'\tau_1$ and $X_{i,2}'\tau_2$ to obtain point identification. However, this is just a sufficient condition. It seems that identification is likely to hold in many cases under much less stringent conditions on the distribution of the regressors. See Supplemental Appendix F for further discussion.

comparing the finite-sample FCP's of the tests introduced in this paper with those of Wald, Lagrange multiplier, and likelihood ratio CS's based on the ML estimator which groups the outcomes $(0,1)$ and $(1,0)$, as in Bresnahan and Reiss (1991) and Berry (1992).

We consider a base case sample size of $n = 500$, as well as $n = 250$ and 1000.

10.4.2 g Functions

We take the functions g to be hypercubes in R^2. They are functions of the 2-vector $X_i^\dagger = (X_{i,1}^{\dagger\prime}, X_{i,2}^{\dagger\prime})' = (X_{i,1}'\widehat{\tau}_{n,1}(\theta), X_{i,2}'\widehat{\tau}_{n,2}(\theta))'$. The vector $X_i^\dagger$ 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 $X_i^\dagger$). Then, it is transformed to lie in $[0,1]^2$ by applying the standard normal distribution function $\Phi(\cdot)$ element by element.

The hypercubes have side-edge lengths $(2r)^{-1}$ for $r = r_0, ..., r_1$, where $r_0 = 1$ and the base case value of r_1 is 3. The base case number of hypercubes is 56. We also report results for $r_1 = 2$ and 4, which yield 20 and 120 hypercubes, respectively. With $n = 500$ and $r_1 = 3$, the expected number of observations per cube is 125, 31.3, or 13.9 depending on the cube. With $n = 500$ and $r_1 = 4$, the expected number also can equal 7.8. With $n = 250$ and $r_1 = 3$, the expected number is 25, 15.6, or 6.9.

10.4.3 Entry Game Simulation Results I

Tables VI and VII 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 similar 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 VI provides CP's and FCP's for competitive effect θ values ranging from $(0,0)$ to $(3,1)$.[47] Table VI 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 is clearly the best statistic in terms of FCP's. The CP's of the CvM/Max CS are good for many θ

[47] 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 VI. 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)							
$(\theta_1, \theta_2) = (0, 0)$		.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

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 VII provides results for variations on the base case 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 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.

Table VII. Entry Game Model: Variations on the Base Case $(\theta_1, \theta_2) = (1, 1)$

		(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
Base Case ($n = 500, r_1 = 3, \varepsilon = 5/100$)		**.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 last two rows of Table VII 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.

10.4.4 Entry Game Simulation Results II

Next, we compare the finite-sample (CP-corrected) FCP's of two CS's introduced in this paper with the FCP's of three CS's that do not exploit the moment inequalities. Figure 2 graphs the FCP's of the CvM/Max and KS/Max CS's using the GMS/Asy critical values (with the base case values of the tuning parameters). It also graphs the FCP's of the Wald, Lagrange multiplier, and likelihood ratio CS's based on the ML estimator that groups the outcomes $(1,0)$ and $(0,1)$ (which ignore the moment inequalities). The sample size is $n = 500$ and the true values of (θ_1, θ_2) are $(1,1)$. The horizontal axis in Figure 2 gives the distance between the true value of θ_1, i.e., $\theta_{1,0} = 1$, and the null value of θ_1, i.e., $\theta_{1,null}$. The distance for the corresponding values of θ_2 is taken to be the same.[48]

As $\theta_{1,0} - \theta_{1,null}$ increases, the FCP's decrease for all CS's, as expected. Figure 2 shows that the CS's that exploit the moment inequalities have far better (lower) FCP's. Specifically, to obtain a FCP equal to p for any p in $[0.75, 0.0]$, the distance of a parameter from the identified set needs to be three times as far or farther when using the Wald, LM, or LR CS as compared to the CvM/Max or KS/Max CS. Thus, we conclude that the CS's introduced here, which exploit the moment inequalities and equalities, are noticeably superior to those that just employ the moment equalities.

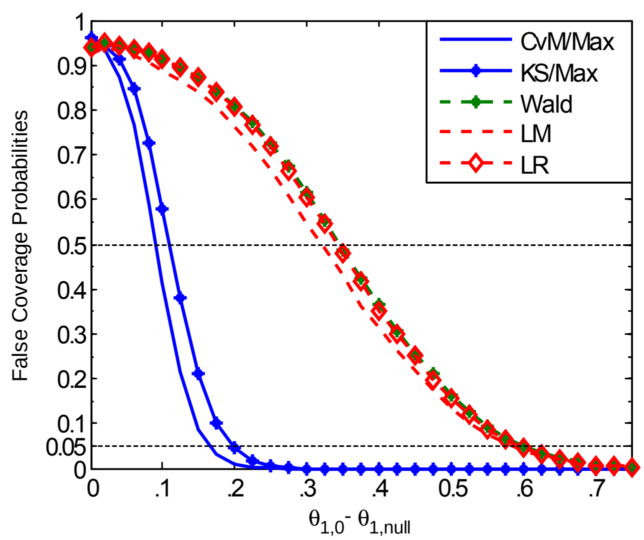

Figure 2. False Coverage Probabilities of Several Nominal 95% Confidence Sets in the Entry Game Model.

[48]Hence, the Euclidean distance between points outside the identified set and points on the boundary of the identified set are proportional to the distances on the horizontal axis in Figure 2.

References

ANDREWS, D. W. K. (2012): "Similar-on-the-Boundary Tests for Moment Inequalities Exist, But Have Very Poor Power," *Econometric Theory,* forthcoming. Also available as Cowles Foundation Discussion Paper No. 1815R, Yale University.

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. 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. J. BARWICK (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica,* 2012, 80, forthcoming. Also available as Andrews and Jia (2008), Cowles Foundation Discussion Paper No. 1676R, 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.

———(2011): "Nonparametric Inference Based on Conditional Moment Inequalities," Cowles Foundation Discussion Paper No. 1840, 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.

Aradillas-Lopez, A., A. Ghandi, and D. Quint (2010): "Identification and Testing in Ascending Auctions with Unobserved Heterogeneity," unpublished manuscript, Department of Economics, University of Wisconsin, Madison.

Armstrong, T. B. (2011a): "Asymptotically Exact Inference in Conditional Moment Inequality Models," unpublished manuscript, Department of Economics, Stanford University.

———(2011b): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," unpublished manuscript, Department of Economics, Stanford University.

Beresteanu, A., I. Molchanov, and F. Molinari (2010): "Sharp Identification Regions in Models with Convex Moment Predictions," CEMMAP Working Paper CWP25/10, 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 (2012): "Set Identified Linear Models," *Econometrica,* 80, forthcoming.

Bresnahan, T. F. and P. C. Reiss (1991): "Empirical Models of Discrete Games," *Journal of Econometrics,* 48, 57-81.

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,* 156, 408-425.

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.

Chetverikov, V. (2011): "Adaptive Test of Conditional Moment Inequalities," unpublished manuscript, Department of Economics, M.I.T.

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

———(2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies,* 78, 1264-1298.

Giustinelli, P. (2010): "Nonparametric Bounds on Quantiles under Monotonicity Assumptions: With an Application to the Italian Education Returns," *Journal of Applied Econometrics,* Wiley Online Library.

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.

Hardle, W. and E. Mammen (1993): "Comparing Nonparametric Versus Parametric Regression Fits," *Annals of Statistics,* 21, 1926-1947.

Hirano, K. and J. Porter (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica,* 80, forthcoming.

Hsu, Y.-C. (2011): "Consistent Tests for Conditional Treatment Effects," Department of Economics, University of Missouri at Columbia.

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

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

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., K. Song, and Y.-J. Whang (2011): "Testing Functional Inequalities," CEMMAP Working Paper CWP12/11, Institute for Fiscal Studies, University College London.

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.

Linton, O., K. Song, and Y.-J. Whang (2010): "An Improved Bootstrap Test of Stochastic Dominance," *Journal of Econometrics,* 154, 186-202.

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.

———(2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica,* 80, forthcoming.

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

Pakes, A. (2010): "Alternative Models for Moment Inequalities," *Econometrica,* 78, 1783-1822.

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.

Ponomareva, M. (2010): "Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates," unpublished manuscript, Department of Economics, University of Western Ontario.

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.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies,* 70, 147-165.

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