INFERENCE BASED ON MANY CONDITIONAL MOMENT INEQUALITIES

By

Donald W. K. Andrews and Xiaoxia Shi

July 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 2010

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

http://cowles. yale.edu/

Inference Based on Many Conditional Moment Inequalities

Donald W. K. Andrews
Cowles Foundation
Yale University

Xiaoxia Shi
Department of Economics
University of Wisconsin-Madison

June 2010; revised: June 27, 2015*

Abstract

In this paper, we construct confidence sets for models defined by many conditional moment inequalities/equalities. The conditional moment restrictions in the models can be finite, countably infinite, or uncountably infinite. To deal with the complication brought about by the vast number of moment restrictions, we exploit the manageability (Pollard (1990)) of the class of moment functions. We verify the manageability condition in five examples from the recent partial identification literature.

The proposed confidence sets are shown to have correct asymptotic size in a uniform sense and to exclude parameter values outside the identified set with probability approaching one. Monte Carlo experiments for a conditional stochastic dominance example and a random-coefficients binary-outcome example support the theoretical results.

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-1058376 and SES-1355504. Shi gratefully acknowledges the research support of the University of Wisconsin-Madison Graduate School with funding from the Wisconsin Alumni Research Foundation. The first version of the results of this paper were given in Section 9 of Andrews and Shi (2009).

1 Introduction

In this paper, we extend the results in Andrews and Shi (2013a, b) (AS1, AS2) to cover models defined by many conditional moment inequalities and/or equalities ("MCMI" in short). The number of conditional moment inequalities/equalities can be countable or uncountable. Examples of models covered by the results include (1) conditional stochastic dominance, (2) random-coefficients binary-outcome models with instrumental variables, see Chesher and Rosen (2014), (3) convex moment prediction models, see Beresteanu, Molchanov, and Molinari (2010), (4) ordered-choice models with endogeneity and instruments, see Chesher and Smolinski (2012), and (5) discrete games identified by revealed preference, see Pakes, Porter, Ho, and Ishii (2015).

The main feature of an MCMI model is that the number of moment restrictions implied by the model is doubly "many." First, there are many (countable or uncountable) conditional moment restrictions, and second each conditional moment restriction implies infinitely many moment conditions. As in AS1 and AS2, we transform each conditional moment restriction into infinitely many unconditional ones using instrumental functions. After the transformation, the unconditional moment functions of the model form a class that is indexed by both the instrumental functions and the indices of the conditional moment restrictions. We exploit a manageability assumption on the class of conditional moment functions. With this assumption, we show that the class of transformed unconditional moment inequalities/equalities is also manageable and, in consequence, can be treated similarly to those to AS1 and AS2.

Thus, the manageability assumption on the class of conditional moment functions is crucial for our theoretical framework. This assumption is verified in the examples by deriving upper bounds on the cover numbers of the functional classes that arise. The upper bounds in the first two examples are derived by bounding the pseudodimensions of the functional classes. In the third example, they are derived using the Lipschitz continuity of the moment functions with respect to the index. These three examples are representative of cases where there are a continuum of conditional moment inequality/equalities. In the fourth and the fifth examples, the numbers of conditional moment inequalities/equalities are countable. For countable functional classes, we treat their elements as sequences and impose decreasing weights on them. The weights guarantee an appropriate bound for the covering numbers.

We note that the approach in this paper also is applicable to models defined by many unconditional moment inequalities/equalities. For such models, one simply omits the step that transforms the conditional moments restrictions into unconditional ones using instrumental functions.

This paper belongs to the moment inequality literature, which is now quite large. The most closely related paper is Chernozhukov, Chetverikov, and Kato (2014), which studies models defined by many moment inequalities. Their framework is different from ours in that (1) their number of moment restrictions increases with the sample size and is finite for each fixed sample size, and (2) they do not assume or exploit the correlation structure between the many moments. Regarding the mechanics of the approaches, their MB (multiplier bootstrap) test is similar to our KS (Kolmogorov-Smirnov) test, while their SN (self-normalizing) test and our CvM (Cramer-von Mises) test are unique to each paper. Like this paper, Delgado and Escanciano (2013) consider tests for conditional stochastic dominance. They take a different approach from the approach in this paper.

Papers in the literature that consider conditional moment inequalities, but not MCMI, include Khan and Tamer (2009), Chetverikov (2012), AS1, Armstrong and Chan (2013), Chernozhukov, Lee, and Rosen (2013), Gandhi, Lu, and Shi (2013), Lee, Song, and Whang (2013), Andrews and Shi (2014), and Armstrong (2014a,b, 2015). Galichon and Henry (2009) provides related results. Papers in the literature that test a continuum of unconditional moment inequalities include papers on testing stochastic dominance and stochastic monotonicity, see Linton, Song, and Whang (2010) and references therein. Papers in the literature that test a continuum of inequalities that are not moment inequalities and, hence, to which the tests in this paper do not apply, include tests of Lorenz dominance, see Dardanoni and Forcina (1999) and Barrett, Donald, and Bhattacharya (2014), and tests of likelihood ratio (or density) ordering, see Beare and Moon (2015), Beare and Shi (2015), and references therein.

The rest of the paper is organized as follows. Section 2 specifies the model and describes the examples. Section 3 introduces the test statistics and confidence sets. Section 4 defines the critical values and gives a step-by-step guide for implementation. Section 5 shows the uniform asymptotic size of the proposed tests and confidence sets in the general setup. Section 6 gives the power results. Sections 7-9 verify the conditions imposed in Sections 5 and 6 for each of the examples. Sections 7 and 8 also provide finite-sample Monte Carlo results for the problem of testing conditional stochastic dominance and for the random-coefficients binary-outcome model with instruments. Section 10 concludes.

For notational simplicity, throughout the paper, we let $(a_i)_{i=1}^n$ denote the n-vector $(a_1, ..., a_n)'$ for $a_i \in R$. We let $A := B$ denote that A equals B by definition or assumption.

2 Many Conditional Moment Inequalities/Equalities

2.1 Models

The models considered in this paper are of the following general form:

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

where $\mathcal{T}$ is a set of indices that may contain an infinite number of elements, θ_0 is the unknown true parameter value that belongs to a parameter space $\Theta \subset R^{d_\theta}$, the observations $\{W_i : i \leq n\}$ are i.i.d., F_0 is the unknown true distribution of W_i, X_i is a subvector of W_i, and $m(w, \theta, \tau) := (m_1(w, \theta, \tau), ..., m_{p+v}(w, \theta, \tau))'$ is a vector of known moment functions.

In contrast, the parameter $\tau \in \mathcal{T}$ does not appear in the moment inequality/equality models considered in AS1 and AS2.

The object of interest is θ_0, which is not assumed to be point identified. The model restricts θ_0 to the *identified set* (which could be a singleton), which is defined by

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

We are interested in confidence sets (CS's) that cover the true value θ_0 with probability greater than or equal to $1 - \alpha$ for $\alpha \in (0, 1)$. 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

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

At each $\theta \in \Theta$, we test the validity of the moment conditions in (2.1) with θ_0 replaced with

θ. The tests are of interest in their own right when (i) there is no parameter to estimate in the moment conditions, as in Example 1 below, or (ii) the validity of the moment conditions at a given θ has policy implications.

2.2 Examples

Models of the form described in (2.1) arise in many empirically relevant situations. Below are some examples.

Example 1 (Conditional Stochastic Dominance). Let $W := (Y_1, Y_2, X)$. Some economic theories imply that the distribution of Y_1 stochastically dominates that of Y_2 conditional on X. For an integer $s \geq 1$, the sth-order conditional stochastic dominance of Y_1 over Y_2 can be written as conditional moment inequalities:

$$\begin{aligned} &E_{F_0}\left[G_s(Y_2,\tau) - G_s(Y_1,\tau)|X\right] \geq 0 \text{ a.s. } \forall \tau \in \mathcal{T}, \text{ where} \\ &G_s(y,\tau) := (\tau - y)^{s-1}1\{y \leq \tau\} \end{aligned} \tag{2.4}$$

and $\mathcal{T}$ contains the supports of Y_1 and Y_2. The tests developed below are directly applicable in this example without being inverted into a confidence set.

Stochastic dominance relationships have been used in income and welfare analysis, for example, in Anderson (1996, 2004), Davidson and Duclos (2000), and Bishop, Zeager, and Zheng (2011). Stochastic dominance relationships also have been used in the study of auctions, e.g., in Guerre, Perrigne, and Vuong (2009). Conditional stochastic dominance implies that the relationship holds for every subgroup of the population defined by X and is useful in all of these applications. See Delgado and Escanciano (2013) for a different approach to testing conditional stochastic dominance from the one considered here.

Sometimes, one may be interested in the conditional stochastic dominance relationship among multiple distributions. For example, for $W = (Y_1, Y_2, Y_3, X)$, one would like to know whether Y_1 s-th order stochastically dominates Y_2 and Y_2 over Y_3 conditional on X. The corresponding conditional moment inequalities to be tested are as follows,

$$\begin{aligned} &E_{F_0}[G_s(Y_2,\tau) - G_s(Y_1,\tau)|X] \geq 0 \text{ and} \\ &E_{F_0}[G_s(Y_3,\tau) - G_s(Y_2,\tau)|X] \geq 0 \text{ a.s. } \forall \tau \in \mathcal{T}, \end{aligned} \tag{2.5}$$

where $\mathcal{T}$ contains the supports of Y_1, Y_2, and Y_3. For example, the comparison of multiple distributions has been considered in Dardanoni and Forcina (1999) for Lorenz dominance.

Example 2 (Random-Coefficients Binary-Outcome Models with Instrumental Variables). Consider the random-coefficients binary-outcome model with instrumental variables (IV's) studied in Chesher and Rosen (2014) (CR):

$$Y_1 = 1\{\beta_0 + X_1'\beta_1 + Y_2'\beta_2 \geq 0\}, \tag{2.6}$$

where $\beta := (\beta_0, \beta_1', \beta_2')'$ are random coefficients that belong to the space R^{d_β}. The covariate vector X_1 is assumed to be exogenous (i.e., independent of β), while the covariate vector Y_2 may be endogenous. Let X_2 be a vector of instrumental variables that is independent of β. Suppose the parameter of interest is the marginal distribution of β, denoted by F_β. Theorem 1 of CR implies that under their Assumptions A1-A3, the sharp identified set for F_β is defined by the following moment inequalities:

$$E_{F_0}[F_\beta(\mathcal{S}) - 1\{S(Y_1, Y_2, X_1) \subset \mathcal{S}\}|X_1, X_2] \geq 0 \text{ a.s. } \forall \mathcal{S} \in \mathbf{S}, \tag{2.7}$$

where

$$\begin{aligned} S(y_1, y_2, x_1) &:= cl\{b = (b_0, b_1', b_2')' \in R^{d_\beta} : y_1 = 1\{b_0 + x_1'b_1 + y_2'b_2 \geq 0\}\}, \\ \mathbf{S} &:= \{cl(\cup_{c \in \mathcal{C}} H(c)) : \mathcal{C} \subset R^{d_\beta}\}, \\ H(c) &:= \{b \in R^{d_\beta} : b'c \geq 0\} \text{ for } c \in R^{d_\beta}, \end{aligned} \tag{2.8}$$

cl denotes "closure," and $H(c)$ is the half-space orthogonal to $c \in R^{d_\beta}$.

Often one may wish to parameterize F_β by assuming $F_\beta(\cdot) = F_\beta(\cdot; \theta)$ for a known distribution function $F_\beta(\cdot; \cdot)$ and an unknown finite-dimensional parameter $\theta \in \Theta$. Then, the sharp identified set for θ is defined by the moment inequalities:

$$E_{F_0}[F_\beta(\mathcal{S}, \theta) - 1\{S(Y_1, Y_2, X_1) \subset \mathcal{S}\}|X_1, X_2] \geq 0 \text{ a.s. } \forall \mathcal{S} \in \mathbf{S}. \tag{2.9}$$

This fits into the framework of (2.1) with $W = (Y_1, Y_2', X_1', X_2')'$, $X = (X_1', X_2')'$, $\tau = \mathcal{S}$, $\mathcal{T} = \mathbf{S}$, $p = 1$, $v = 0$, and $m(w, \theta, \tau) = F_\beta(\mathcal{S}, \theta) - 1\{S(y_1, y_2, x_1) \subset \mathcal{S}\}$.

Example 3 (Convex Moment Prediction Models–Support Function Approach).

Beresteanu, Molchanov, and Molinari (2010) (BMM) establish a framework to characterize the sharp identified set for a general class of incomplete models with convex moment predictions using the random set theory. Examples of such models include static, simultaneous move, finite games with complete or incomplete information in the presence of multiple equilibria, best linear prediction models with interval outcome and/or regressor data, and random utility models of multinomial choice with interval regressor data. BMM show that the sharp identified set for these models can be characterized by a continuum of conditional moment inequalities using the support function of the set. For parameter inference, BMM suggest applying the procedure in this paper and they verify the high-level assumptions in an earlier version of this paper in two examples. Here, we describe their identification framework briefly.

Consider a model based on an observed random vector W and an unobserved random vector V. The model maps each value of (W, V) to a closed set $Q_\theta(W, V) \subseteq R^d$, where θ is the model parameter that belongs to a parameter space Θ, and d is a positive integer. Let X be a subvector of W with support contained in $\mathcal{X}$ and let $q(x) : \mathcal{X} \to R^d$ be a known function. Suppose (W, V) and W take values in some sets $\mathcal{WV}$ and $\mathcal{W}$, respectively. BMM assume that the sharp identified set of θ implied by the model is

$$\Theta_I = \{\theta \in \Theta : q(X) \in \mathbb{E}_{F_0}[Q_\theta(W, V)|X] \; a.s. \; [X]\}, \tag{2.10}$$

where $\mathbb{E}_{F_0}[\cdot]$ stands for the Aumann expectation of the random set inside the square brackets under the true distribution F_0 of (W, V). BMM show that the event $q(X) \in \mathbb{E}_{F_0}[Q_\theta(W, V)|X]$ can be written equivalently as the following set of moment inequalities

$$E_{F_0}[h(Q_\theta(W, V), u) - u'q(X)|X] \geq 0 \; a.s.[X], \; \forall u \in R^d \; s.t. \; \|u\| \leq 1, \tag{2.11}$$

where $h(Q, u)$ is the support function of Q in the direction given by u, that is, $h(Q, u) = \sup_{q \in Q} q'u$.

The inequalities (2.11) do not fall immediately into our general framework because of the unobservable V. However, in applications, one typically has that either $Q_\theta(W, V) = Q_\theta(W)$ (so that V does not appear in (2.11)) or the distribution of V given X (denoted $F_{V|X}(v|x; \theta)$) is known to the researcher up to an unknown parameter θ. In the former case, (2.11) fits the form of (2.1). In the latter case, we write (2.11) as

$$E_{F_0}\left[\int h(Q_\theta(W,v),u)dF_{V|X}(v|X;\theta)-u'q(X)|X\right]\geq 0\ a.s.\ [X],\ \forall u\in R^d\ s.t.\ \|u\|\leq 1, \tag{2.12}$$

which fits the form of (2.1). The former case includes the best linear predictor example in BMM, and the latter case includes the entry game example in BMM.

Example 4 (IV Ordered-Choice Models). Chesher and Smolinski (2012) show that the sharp identified set for a nonparametric single equation instrumental variable (SEIV) model with ordered outcome and discrete endogenous regressors can be characterized by a finite, but potentially very large, number of moment inequalities. Consider the nonseparable model

$$Y=h\left(Z,U\right), \tag{2.13}$$

where $Y\in\{1,2,...,M\}$ and $Z\in\{z_1,...,z_K\}$, the error term U is normalized to be uniformly distributed in $[0,1]$. Assume that there is a vector of instrumental variables X that is independent of U. Then, one has a SEIV model. Further, assume that h is weakly increasing in U. Then, h has a threshold crossing representation: for $m=1,...,M$ and $z\in\{z_1,...,z_K\}$:

$$h\left(z,u\right)=m \text{ if } u\in\left(h_{m-1}(z),h_m(z)\right] \tag{2.14}$$

for some constants $0=h_0(z)<\cdots<h_M(z)=1$. Thus, estimating $h\left(z,u\right)$ amounts to estimating the $J=(M-1)K$ threshold parameters $\gamma=(\gamma_{11},...,\gamma_{(M-1)1},...,\gamma_{1K},...,\gamma_{(M-1)K})'$, where

$$\gamma_{mk}=h_m\left(z_k\right)\ \forall m=1,...,M-1,\ \forall k=1,...,K. \tag{2.15}$$

Chesher and Smolinski (2012) show that the sharp identified set for γ can be characterized by the following moment inequalities

$$
\begin{aligned}
E_{F_0}\left[\gamma_{\ell s}-\sum_{k=1}^{K}\sum_{m=1}^{M-1}1\{Y=m,Z=z_k,\gamma_{mk}\leq\gamma_{\ell s}\}\middle|X\right]\geq 0 \text{ a.s. } [X] \text{ and}\\
E_{F_0}\left[\sum_{k=1}^{K}\sum_{m=1}^{M-1}1\{Y=m,Z=z_k,\gamma_{(m-1)k}<\gamma_{\ell s}\}-\gamma_{\ell s}\middle|X\right]\geq 0 \text{ a.s. } [X]\ \forall\ell\leq M-1,\ \forall s\leq K,\\
E_{F_0}\left[\gamma_{\ell s}-\gamma_{ms}-1\{m<Y\leq\ell,Z=z_s\}\middle|X\right]\geq 0 \text{ a.s. } [X]\\
\forall\ell>m,\ \forall\ell,m\leq M-1,\ \forall s\leq K.
\end{aligned}
\tag{2.16}
$$

We arrange the above $N := 2(M-1)K+(M-2)(M-1)K/2$ inequalities into a column, and index them by τ for $\tau = 1,\ldots,N$. Let $W = (Y,X,Z')'$ and let $m(W,\gamma,\tau)$ be the expression inside the conditional expectation in the τth inequality. Then, this example falls into the framework of (2.1) with $\theta = \gamma$.

One may wish to parameterize the threshold functions γ via $\gamma = \Gamma(\theta)$. In that case, the same set of moment inequalities as above defines the sharp identified set for θ. For example, Chesher and Smolinski (2012) show that, for the linear ordered-probit model,

$$\gamma_{mk} := h_m(z_k) = \Phi(c_m - a_1 z_k)\ \forall m = 1,\ldots,M-1,\ \forall k = 1,\ldots,K, \tag{2.17}$$

where $c_1,\ldots,c_{M-1}$ are the threshold values, a_1 is the slope parameter, and $\Phi(\cdot)$ is the standard normal distribution function.

Example 5 (Revealed Preference Approach in Discrete Games). Pakes, Porter, Ho, and Ishii (2015) formalize the idea of using the revealed preference principle to estimate games in which a finite number of players have a discrete set of actions to choose from. Observing the players' equilibrium play, the econometrician can write down moment inequalities that are implied by the revealed preference principle. These moment inequalities allow one to estimate the structural parameters without solving for the equilibrium. Here we describe a simplified version of their framework.

Suppose that all players make decisions based on the same information set and the econometrician observes the information set. Players make decisions based on expected utility maximization. Suppose there are J players and each player has a feasible action set A_j that is discrete (i.e., finite or countably infinite). Let $\pi_j(a_j, a_{-j}, Z; \theta)$ be the utility of player j given her own action a_j and her opponents' actions a_{-j} and the covariates Z. Let X be a subvector of Z that generates the information set of the players. Let the boldfaced $\mathbf{a}_j$ and $\mathbf{a}_{-j}$ be the observed actions of player j and her opponents. The function π_j is known up to the finite dimensional parameter θ. Then, the moment inequalities are

$$E_{F_0}\left(\pi_j\left(\mathbf{a}_j, \mathbf{a}_{-j}, Z; \theta\right) - \pi_j\left(a_j', \mathbf{a}_{-j}, Z; \theta\right)\middle| X\right) \geq 0 \ \forall a_j' \in A_j, \ \forall j = 1, ..., J. \tag{2.18}$$

When J is large or the number of elements in A_j is large, there are many (possibly countably infinitely many) conditional moment inequalities.

2.3 Parameter Space

Let (θ, F) denote a generic value of the parameter and the distribution of W_i. Let $\mathcal{F}$ denote the parameter space for the true values (θ_0, F_0), which satisfy the conditional moment inequalities and equalities. To specify $\mathcal{F}$, we first introduce some additional notation. For each distribution F, let F_X denote the marginal distribution of X_i under F. Let $k := p + v$.

Below, we employ a "manageability" condition that regulates the complexity of $\mathcal{T}$. It ensures a functional central limit theorem (CLT) result, which is used in the proof of the uniform coverage probability results for the CS's. The concept of manageability is from Pollard (1990) and is defined in Section B.3 of the Appendix.

The test consistency results given below apply to (θ, F) pairs that do not satisfy the conditional moment inequalities and equalities. For this reason, we introduce a set $\mathcal{F}_+$ that is a superset of $\mathcal{F}$ and does not impose the inequalities and equalities. Let $\mathcal{F}_+$ be some collection of (θ, F) that satisfy the following parameter space (PS) Assumptions PS1 and PS2 for given constants $\delta > 0$ and $C_1 < \infty$ and given deterministic function of (θ, F): $\sigma_F(\theta) := (\sigma_{F,1}(\theta), \ldots, \sigma_{F,k}(\theta))'$. The function $\sigma_F(\theta)$ is useful for the standardization of certain forms of the test statistic, and is specified in greater detail in sections below.

Assumption PS1. For any $(\theta, F) \in \mathcal{F}_+$,

(a) $\theta \in \Theta$,

(b) $\{W_i : i \geq 1\}$ are i.i.d. under F,

(c) $\sigma_{F,j}(\theta) > 0$, $\forall j = 1, \ldots, k$,

(d) $|m_j(w, \theta, \tau)/\sigma_{F,j}(\theta)| \leq M(w)$, $\forall w \in R^{d_w}$, $\forall j = 1, ..., k$, $\forall \tau \in \mathcal{T}$, for some function $M : R^{d_w} \to [0, \infty)$, and

(e) $E_F M^{2+\delta}(W_i) \leq C_1$.

Assumption PS2. For all sequences $\{(\theta_n, F_n) \in \mathcal{F}_+ : n \geq 1\}$, the triangular array of processes $\{(m_j(W_{n,i}, \theta_n, \tau)/\sigma_{F_n,j}(\theta_n))_{j=1}^k : \tau \in \mathcal{T}, i \leq n, n \geq 1\}$ is manageable with respect to the envelopes $\{M(W_{n,i}) : i \leq n, n \geq 1\}$, where $\{W_{n,i} : i \leq n, n \geq 1\}$ is a row-wise i.i.d. triangular array with $W_{n,i} \sim F_n$ $\forall i \leq n, n \geq 1$.

The parameter space $\mathcal{F}$ for the conditional moment inequality model is the subset of $\mathcal{F}_+$ that satisfies:

Assumption PS3. (a) $E_F[m_j(W_i, \theta, \tau)|X_i] \geq 0$ a.s. $[F_X]$ for $j = 1, ..., p$, $\forall \tau \in \mathcal{T}$,

(b) $E_F[m_j(W_i, \theta, \tau)|X_i] = 0$ a.s. $[F_X]$ for $j = p+1, ..., k$, $\forall \tau \in \mathcal{T}$.

3 Tests and Confidence Sets

In this section, we describe the test statistic. To do so, we first transform the conditional moment inequalities/equalities into equivalent unconditional ones using instrumental functions. The unconditional moment conditions are as follows:

$$\begin{aligned} &E_{F_0}[m_j(W_i, \theta_0, \tau) g_j(X_i)] \geq 0 \text{ for } j = 1, ..., p \text{ and} \\ &E_{F_0}[m_j(W_i, \theta_0, \tau) g_j(X_i)] = 0 \text{ for } j = p+1, ..., k, \\ &\qquad\qquad \forall \tau \in \mathcal{T} \text{ and } \forall g = (g_1, ..., g_k)' \in \mathcal{G}_{\text{c-cube}}, \end{aligned} \tag{3.1}$$

where g is a vector of instruments that depends on X_i and $\mathcal{G}_{\text{c-cube}}$ is a collection of instrumental functions g defined below.

We construct test statistics based on (3.1). Let the sample moment functions be

$$\begin{aligned} \overline{m}_n(\theta, \tau, g) &:= n^{-1} \sum_{i=1}^n m(W_i, \theta, \tau, g) \text{ for } g \in \mathcal{G}_{\text{c-cube}} \text{ and} \\ m(W_i, \theta, \tau, g) &:= (m_1(W_i, \theta, \tau) g_1(X_i), ..., m_k(W_i, \theta, \tau) g_k(X_i))'. \end{aligned} \tag{3.2}$$

The sample variance matrix of $n^{1/2}\overline{m}_n(\theta, g, \tau)$ is useful for most versions of the test statistic and for the critical values. It is defined as

$$\widehat{\Sigma}_n(\theta, \tau, g) := n^{-1} \sum_{i=1}^n (m(W_i, \theta, \tau, g) - \overline{m}_n(\theta, \tau, g))(m(W_i, \theta, \tau, g) - \overline{m}_n(\theta, \tau, g))'. \tag{3.3}$$

When the sample variance is used, we would like it to be nonsingular because it is used to studentize the sample moment functions. However, the matrix $\widehat{\Sigma}_n(\theta,\tau,g)$ may be singular or nearly singular with non-negligible probability for some (τ,g). Thus, we add a small positive definite matrix to $\widehat{\Sigma}_n(\theta,\tau,g)$:

$$\overline{\Sigma}_n(\theta,\tau,g) := \widehat{\Sigma}_n(\theta,\tau,g) + \varepsilon \cdot \mathrm{Diag}(\widehat{\sigma}^2_{n,1}(\theta),\ldots,\widehat{\sigma}^2_{n,k}(\theta)) \text{ for } (\tau,g) \in \mathcal{T} \times \mathcal{G}_{\text{c-cube}} \text{ and } \varepsilon = 1/20, \tag{3.4}$$

where $\widehat{\sigma}_{n,j}(\theta)$ is a consistent estimator of the $\sigma_{F,j}(\theta)$ introduced just above Assumption PS1.

In practice, if the moment functions have a natural scale (say, being a probability or the difference of two probabilities), one can take $\widehat{\sigma}_{n,j}(\theta) = \sigma_{F,j}(\theta) = 1$ for all j, (θ,F), and n. Otherwise, we recommend taking $\widehat{\sigma}_{n,j}(\theta)$ and $\sigma_{F,j}(\theta)$ such that $\widehat{\sigma}^{-1}_{n,j}(\theta)m_j(W_i,\theta,\tau)$ and $\sigma^{-1}_{F,j}(\theta)m_j(W_i,\theta,\tau)$ are invariant to the rescaling of the moment functions, because this yields a test with the same property. We discuss specific choices for the examples in later sections.

We assume that the estimators $\{\widehat{\sigma}_{n,j}(\theta) : j \leq k\}$ satisfy the following uniform consistency condition.

Assumption SIG1. For all $\zeta > 0$, $\sup_{(\theta,F)\in\mathcal{F}} \Pr\left(\max_{j\leq k} |\widehat{\sigma}^2_{n,j}(\theta)/\sigma^2_{F,j}(\theta) - 1| > \zeta\right) \to 0$.

The functions g that we consider are hypercubes in $[0,1]^{d_X}$. Hence, we transform each element of X_i to lie in $[0,1]$. (There is no loss in information in doing so.) For notational convenience, we suppose $X_i^\dagger \in R^{d_X}$ denotes the nontransformed IV vector and we let X_i denote the transformed IV vector. We transform $X_i^\dagger$ via a shift and rotation and then apply the standard normal distribution function $\Phi(x)$. Specifically, let

$$\begin{aligned} &X_i := \Phi(\widehat{\Sigma}^{-1/2}_{X,n}(X_i^\dagger - \overline{X}_n^\dagger)), \text{ where } \Phi(x) := (\Phi(x_1),...,\Phi(x_{d_X}))' \text{ for } x = (x_1,...,x_{d_X})', \\ &\widehat{\Sigma}_{X,n} := n^{-1}\Sigma_{i=1}^n (X_i^\dagger - \overline{X}_n^\dagger)(X_i^\dagger - \overline{X}_n^\dagger)', \text{ and } \overline{X}_n^\dagger := n^{-1}\Sigma_{i=1}^n X_i^\dagger. \end{aligned} \tag{3.5}$$

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

$$\mathcal{G}_{\text{c-cube}} := \{g_{a,r} : g_{a,r}(x) := 1\{x \in C_{a,r}\} \cdot 1_k \text{ for } C_{a,r} \in \mathcal{C}_{\text{c-cube}}\}, \text{ where}$$
$$\mathcal{C}_{\text{c-cube}} := \Big\{ C_{ar} := \prod_{u=1}^{d_x} ((a_u - 1)/(2r), a_u/(2r)] \in [0,1]^{d_x} : a = (a_1, ..., a_{d_x})'$$
$$a_u \in \{1, 2, ..., 2r\} \text{ for } u = 1, ..., d_x \text{ and } r = r_0, r_0 + 1, ... \Big\} \tag{3.6}$$

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

The test statistic $\overline{T}_{n,r_{1,n}}(\theta)$ is either a Cramér-von-Mises-type (CvM) or Kolmogorov-Smirnov-type (KS) statistic. The CvM statistic is

$$\overline{T}_{n,r_{1,n}}(\theta) := \sup_{\tau \in \mathcal{T}} \sum_{r=1}^{r_{1,n}} (r^2 + 100)^{-1} \sum_{a \in \{1,...,2r\}^{d_X}} (2r)^{-d_x} S(n^{1/2} \overline{m}_n(\theta, \tau, g_{a,r}), \overline{\Sigma}_n(\theta, \tau, g_{a,r})), \tag{3.7}$$

where $S = S_1, S_2,\ S_3$, or S_4 as defined in (3.9) below, $(r^2 + 100)^{-1}$ is a weight function, and $r_{1,n}$ is a truncation parameter. The asymptotic size and consistency results for the CS's and tests based on $\overline{T}_{n,r_{1,n}}(\theta)$ allow for more general forms of the weight function and hold whether $r_{1,n} = \infty$ or $r_{1,n} < \infty$ and $r_{1,n} \to \infty$ as $n \to \infty$. (No rate at which $r_{1,n} \to \infty$ is needed for these results.) For computational tractability, we typically take $r_{1,n} < \infty$.

The Kolmogorov-Smirnov-type (KS) statistic is

$$\overline{T}_{n,r_{1,n}}(\theta) := \sup_{\tau \in \mathcal{T}} \sup_{g_{a,r} \in \mathcal{G}_{\text{c-cube},r_{1,n}}} S(n^{1/2} \overline{m}_n(\theta, \tau, g_{a,r}), \overline{\Sigma}_n(\theta, \tau, g_{a,r})), \tag{3.8}$$

where $\mathcal{G}_{\text{c-cube},r_{1,n}} = \{g_{a,r} \in \mathcal{G}_{\text{c-cube}} : r \leq r_{1,n}\}$. For brevity, the discussion in this paper focusses on CvM statistics and all results stated concern CvM statistics. Similar results hold for KS statistics.[2]

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

[2] Such results can be established by extending the results given in Section 13.1 of Appendix B of AS2 and proved in Section 15.1 of Appendix D of AS2.

The functions S_1-S_4 are defined by

$$
\begin{aligned}
S_1(m,\Sigma) &:= \sum_{j=1}^{p} [m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v} [m_j/\sigma_j]^2, \\
S_2(m,\Sigma) &:= \inf_{t=(t_1',0_v')':t_1\in R_{+,\infty}^p} (m-t)'\Sigma^{-1}(m-t), \\
S_3(m,\Sigma) &:= \max\{[m_1/\sigma_1]_-^2, ..., [m_p/\sigma_p]_-^2, (m_{p+1}/\sigma_{p+1})^2, ..., (m_{p+v}/\sigma_{p+v})^2\}, \text{ and} \\
S_4(m,\Sigma) &:= \inf_{t=(t_1',0_v')':t_1\in R_{+,\infty}^p} (m-t)'(m-t) = \sum_{j=1}^{p}[m_j]_-^2 + \sum_{j=p+1}^{p+v} m_j^2, \qquad (3.9)
\end{aligned}
$$

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 $x \geq 0$, $R_{+,\infty} := \{x \in R : x \geq 0\} \cup \{+\infty\}$, and $R_{+,\infty}^p := R_{+,\infty} \times ... \times R_{+,\infty}$ with p copies. The functions S_1, S_2, and S_3 are referred to as the modified method of moments (MMM) or Sum function, the quasi-likelihood ratio (QLR) function, and the Max function, respectively. The function S_4 is referred to as the identity-weighted MMM function. The test statistic based on S_4 is not invariant to scale changes of the moment functions, which may be a disadvantage in some examples. But, in other examples (e.g., Examples 2 and 4 above and the $s = 1$ case of Example 1), the moment functions are naturally on a probability scale (i.e., they take values in $[-1, 1]$) and scale invariance is not an issue. In such cases, S_4 is a desirable choice for its simplicity.

4 Critical Values

4.1 GMS Critical Values

In this section we define GMS critical values based on bootstrap simulations. The critical value is obtained via the following steps.

Step 1. Compute $\overline{\varphi}_n(\theta,\tau,g_{a,r})$ for $(\tau,g_{a,r}) \in \mathcal{T} \times \mathcal{G}_{\text{c-cube},r_{1,n}}$, where $\overline{\varphi}_n(\theta,g_{a,r})$ is defined as follows. For $g = g_{a,r}$, let

$$
\begin{aligned}
\xi_n(\theta,\tau,g) &:= \kappa_n^{-1} n^{1/2} \overline{D}_n^{-1/2}(\theta,\tau,g)\overline{m}_n(\theta,\tau,g), \text{ where} \\
\overline{D}_n(\theta,\tau,g) &:= Diag(\overline{\Sigma}_n(\theta,\tau,g)),\ \kappa_n := (0.3\ln(n))^{1/2}, \qquad (4.1)
\end{aligned}
$$

and $\overline{\Sigma}_n(\theta,\tau,g)$ is defined in (3.4). The jth element of $\xi_n(\theta,\tau,g)$, denoted $\xi_{n,j}(\theta,\tau,g)$, measures the slackness of the moment inequality $E_F m_j(W_i, \theta, \tau, g) \geq 0$ for $j = 1, ..., p$. It is shrunk towards zero via κ_n^{-1} to ensure that one does not over-estimate the slackness.

Define $\overline{\varphi}_n(\theta, \tau, g) := (\overline{\varphi}_{n,1}(\theta, \tau, g), ..., \overline{\varphi}_{n,p}(\theta, \tau, g), 0, ..., 0)' \in R^k$ by

$$\overline{\varphi}_{n,j}(\theta, \tau, g) := \overline{\Sigma}_{n,j}^{1/2}(\theta, \tau, g) B_n 1\{\xi_{n,j}(\theta, \tau, g) > 1\} \text{ for } j \leq p \text{ and}$$
$$B_n := (0.4 \ln(n) / \ln\ln(n))^{1/2}, \tag{4.2}$$

where $\overline{\Sigma}_{n,j}(\theta, \tau, g)$ denotes the (j, j) element of $\overline{\Sigma}_n(\theta, \tau, g)$.

Step 2. Generate B bootstrap samples $\{W_{i,s}^* : i = 1, ..., n\}$ for $s = 1, ..., B$ using the standard nonparametric i.i.d. bootstrap. That is, draw $W_{i,s}^*$ randomly with replacement from $\{W_\ell : \ell = 1, ..., n\}$ for $i = 1, ..., n$ and $s = 1, ..., B$.

Step 3. For each bootstrap sample, transform the regressors as in (3.5) (using the bootstrap sample in place of the original sample) and compute $\overline{m}_{n,s}^*(\theta, \tau, g_{a,r})$ and $\overline{\Sigma}_{n,s}^*(\theta, \tau, g_{a,r})$ just as $\overline{m}_n(\theta, \tau, g_{a,r})$ and $\overline{\Sigma}_n(\theta, \tau, g_{a,r})$ are computed, but with the bootstrap sample in place of the original sample.[3]

Step 4. For each bootstrap sample, compute the bootstrap test statistic $\overline{T}_{n,r_{1,n},s}^*(\theta)$ as $\overline{T}_{n,r_{1,n}}^{CvM}(\theta)$ (or $\overline{T}_{n,r_{1,n}}^{KS}(\theta)$) is computed in (3.7) (or (3.8)) but with $n^{1/2}\overline{m}_n(\theta, \tau, g_{a,r})$ replaced by $n^{1/2}(\overline{m}_{n,s}^*(\theta, \tau, g_{a,r}) - \overline{m}_n(\theta, \tau, g_{a,r})) + \overline{\varphi}_n(\theta, \tau, g_{a,r})$ and with $\overline{\Sigma}_n(\theta, \tau, g_{a,r})$ replaced by $\overline{\Sigma}_{n,s}^*(\theta, \tau, g_{a,r})$.[4]

Step 5. Take the bootstrap GMS critical value $c_{n,1-\alpha}^{GMS,*}(\theta)$ to be the $1-\alpha+\eta$ sample quantile of the bootstrap test statistics $\{\overline{T}_{n,r_{1,n},s}^*(\theta) : s = 1, ..., B\}$ plus η, where $\eta = 10^{-6}$.

The CvM (or KS) GMS CS is defined in (2.3) with $T_n(\theta) = \overline{T}_{n,r_{1,n}}^{CvM}(\theta)$ (or $\overline{T}_{n,r_{1,n}}^{KS}(\theta)$) and $c_{n,1-\alpha}(\theta) = c_{n,1-\alpha}^{GMS,*}(\theta)$. The CvM GMS test of $H_0 : \theta = \theta_*$ rejects H_0 if $\overline{T}_{n,r_{1,n}}^{CvM}(\theta_*) > c_{n,1-\alpha}^{GMS,*}(\theta_*)$. The KS GMS test is defined likewise using $\overline{T}_{n,r_{1,n}}^{KS}(\theta_*)$ and the KS GMS critical value.

The choices of ε, κ_n, B_n, and η above are based on some experimentation (in the simulation results reported in AS1 and AS2). The asymptotic results reported in the Appendix allow for other choices.

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

[3]If the test statistic uses function S_4 defined above, $\overline{\Sigma}_n^*(\theta, \tau, g_{a,r})$ does need to be computed.

[4]If the function S_4 is used, $\overline{\Sigma}_n(\theta, \tau, g_{a,r})$ does not appear in the test statistic, and thus $\overline{\Sigma}_n^*(\theta, \tau, g_{a,r})$ does not enter the calculation of the bootstrap statistic.

of cubes. Hence, it is linear in $N_g := \sum_{r=1}^{r_{1,n}} (2r)^{d_X}$.

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).

4.2 Plug-in Asymptotic Critical Values

Next, for comparative purposes, we define bootstrap plug-in asymptotic (PA) critical values. Subsampling critical values also can be considered, see Appendix B of AS2 for details. We strongly recommend GMS critical values over PA and subsampling critical values for the same reasons as given in AS1 plus the fact that the finite-sample simulations in Sections 7.2 and 8.2 show better performance by GMS critical values than PA critical values, and better or equal performance by GMS critical values comparing to subsampling ones.

Bootstrap PA critical values (denoted by $c_{n,1-\alpha}^{PA,*}(\theta)$) are based on the least-favorable asymptotic null distribution with an estimator of the unknown covariance kernel plugged-in. They are computed just as the (bootstrap) GMS critical values are, but with $\overline{\varphi}_n(\theta, \tau, g_{a,r}) = 0^k$ $(\in R^k)$.

The nominal $1-\alpha$ PA CS is given by (2.3) with $T_n(\theta) = \overline{T}_{n,r_{1,n}}^{CvM}(\theta)$ (or $\overline{T}_{n,r_{1,n}}^{KS}(\theta)$) and the critical value $c_{n,1-\alpha}(\theta)$ equal to the PA critical value. The CvM (or KS) PA test of $H_0 : \theta = \theta_*$ rejects H_0 if $\overline{T}_{n,r_{1,n}}^{CvM}(\theta_*)$ (or $\overline{T}_{n,r_{1,n}}^{KS}(\theta_*)$) exceeds the CvM (or KS) PA critical value evaluated at $\theta = \theta_*$.

PA critical values are greater than or equal to GMS critical values for all n (because $\overline{\varphi}_{n,j}(\theta, \tau, g) \geq 0$ for all $(\tau, g) \in \mathcal{T} \times \mathcal{G}$ for $j \leq p$ and $S_\ell(m, \Sigma)$ is non-increasing in $m_I \in R^p$, where $m := (m_I', m_{II}')'$), for $\ell = 1, 2, 3, 4$). Hence, the asymptotic local power of a GMS test is greater than or equal to that of a PA test for all local alternatives. Strict inequality typically occurs whenever the conditional moment inequality $E_{F_n}(m_j(W_i, \theta_*, \tau)|X_i)$ is bounded away from zero as $n \to \infty$ with positive X_i probability for some $j = 1, ..., p$ and some $\tau \in \mathcal{T}$.

5 Correct Asymptotic Size

In this section, we show that GMS and PA CS's have correct asymptotic size (in a uniform sense).

First, we introduce some additional notation. We define the asymptotic covariance kernel, $\{h_{2,F}(\theta, \tau, g, \tau^\dagger, g^\dagger) : (\tau, g), (\tau^\dagger, g^\dagger) \in \mathcal{T} \times \mathcal{G}_{\text{c-cube}}\}$, of $n^{1/2}\overline{m}_n(\theta, \tau, g)$ after normalization via a diagonal matrix $D_F^{-1/2}(\theta)$. That is, we define

$$\begin{aligned} h_{2,F}(\theta,\tau,g,\tau^\dagger,g^\dagger) &:= D_F^{-1/2}(\theta)\Sigma_F(\theta,\tau,g,\tau^\dagger,g^\dagger)D_F^{-1/2}(\theta), \text{ where} \\ \Sigma_F(\theta,\tau,g,\tau^\dagger,g^\dagger) &:= Cov_F((m(W_i,\theta,\tau,g),m(W_i,\theta,\tau^\dagger,g^\dagger)), \\ D_F(\theta) &:= Diag(\sigma^2_{F,1}(\theta),\ldots,\sigma^2_{F,k}(\theta)), \end{aligned} \tag{5.1}$$

and $\{\sigma_{F,j}(\theta) : j = 1,\ldots,k\}$ are specified just before Assumption PS1. For simplicity, let $h_{2,F}(\theta)$ abbreviate $\{h_{2,F}(\theta,\tau,g,\tau^\dagger,g^\dagger) : (\tau,g),(\tau^\dagger,g^\dagger) \in \mathcal{T}\times\mathcal{G}_{\text{c-cube}}\}$.

Define

$$\mathcal{H}_2 := \{h_{2,F}(\theta) : (\theta,F) \in \mathcal{F}\}, \tag{5.2}$$

where, as defined at the end of Section 2, $\mathcal{F}$ is the subset of $\mathcal{F}_+$ that satisfies Assumption PS3. On the space of $k\times k$ matrix-valued covariance kernels on $(\mathcal{T}\times\mathcal{G}_{\text{c-cube}})^2$, which is a superset of $\mathcal{H}_2$, we use the uniform metric d defined by

$$d(h_2^{(1)},h_2^{(2)}) := \sup_{(\tau,g),(\tau^\dagger,g^\dagger)\in\mathcal{T}\times\mathcal{G}_{\text{c-cube}}} \|h_2^{(1)}(\tau,g,\tau^\dagger,g^\dagger) - h_2^{(2)}(\tau,g,\tau^\dagger,g^\dagger)\|. \tag{5.3}$$

Correct asymptotic size is established in the following theorem.

Theorem 5.1 *Suppose Assumption* SIG1 *holds. For any compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$, *the* GMS *and* PA *confidence sets* CS_n *satisfy*

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

Comments. 1. Theorem 5.1 shows that GMS and PA CS's have correct uniform asymptotic size over compact sets of covariance kernels. The uniformity results hold whether the moment conditions involve "weak" or "strong" IV's X_i. That is, weak identification of the parameter θ due to a low correlation between X_i and the functions $m_j(W_i,\theta,\tau)$ does not affect the uniformity results.

2. The proofs in the Appendix take the transformation of the IV's to be non-data dependent. One could extend the results to allow for data-dependence by considering random hypercubes as in Pollard (1979) and Andrews (1988). These results show that one obtains the same asymptotic results with random hypercubes as with nonrandom hypercubes that converge in probability to nonrandom hypercubes (in an L^2 sense). For brevity, we do not do so.

6 Power Against Fixed Alternatives

We now show that the powers of GMS and PA tests converge to one as $n \to \infty$ for all fixed alternatives (for which Assumptions PS1 and PS2 hold). 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 exclude θ_* with probability approaching one. In this sense, GMS and PA CS's based on $T_n(\theta)$ fully exploit the infinite number of conditional moment inequalities/equalities. CS's based on a finite number of unconditional moment inequalities/equalities do not have this property.[5]

The null hypothesis is

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

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

Let $\mathcal{F}_+$ be as defined in Section 2.3. Note that $\mathcal{F}_+$ includes (θ, F) pairs for which θ lies outside of the identified set Θ_F as well as all values in the identified set.

The set $\mathcal{X}_F(\theta, \tau)$ of values x for which the moment inequalities or equalities evaluated at θ are violated under F is defined as follows. For any $\theta \in \Theta$ and any distribution F with $E_F[\|m(W_i, \theta, \tau)\|] < \infty$, let

$$\begin{aligned} \mathcal{X}_F(\theta, \tau) := \{ & x \in R^{d_x} : E_F[m_j(W_i, \theta, \tau) | X_i = x] < 0 \text{ for some } j \leq p \text{ or} \\ & E_F[m_j(W_i, \theta, \tau) | X_i = x] \neq 0 \text{ for some } j = p+1, ..., k \}. \end{aligned} \quad (6.2)$$

[5] This holds because the identified set based on a finite number of moment inequalities typically is larger than the identifed set based on all the conditional moment inequalities. In consequence, CI's based on a finite number of inequalities include points in the difference between these two identified sets with probability whose limit infimum as $n \to \infty$ is $1 - \alpha$ or larger even though these points are not in the identified set based on the conditional moment inequalities.

The next assumption, Assumption MFA, states that violations of the conditional moment inequalities or equalities occur for the null parameter θ_* for X_i values in a set with positive probability under F_0 for some $\tau \in \mathcal{T}$. Thus, under Assumption MFA, the moment conditions specified in (6.1) do not hold.

Assumption MFA. The null value $\theta_* \in \Theta$ and the true distribution F_0 satisfy: (a) for some $\tau_* \in \mathcal{T}$, $P_{F_0}(X_i \in \mathcal{X}_{F_0}(\theta_*, \tau_*)) > 0$ and (b) $(\theta_*, F_0) \in \mathcal{F}_+$.

We employ the following assumption on the weights $\{\widehat{\sigma}^2_{n,j}(\theta) : j \leq k, n \geq 1\}$.

Assumption SIG2. For all $\zeta > 0$, $\Pr_{F_0}\left(\max_{j \leq k} |\widehat{\sigma}^2_{n,j}(\theta_*)/\sigma^2_{F_0,j}(\theta_*) - 1| > \zeta\right) \to 0$.

Note that Assumption SIG2 is not implied by Assumption SIG1 because (θ_*, F_0) does not belong to $\mathcal{F}$.

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

Theorem 6.1 *Suppose Assumptions* MFA *and* SIG2 *hold. Then,*

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

(b) $\lim_{n\to\infty} P_{F_0}(T_n(\theta_*) > c^{PA,*}_{n,1-\alpha}(\theta_*)) = 1$.

7 Example 1: Conditional Stochastic Dominance

In this section, we apply the general theory developed above to Example 1. We first establish primitive sufficient conditions for Assumptions PS1 and PS2 for this example, and then carry out a simple Monte Carlo experiment for testing first-order stochastic dominance.

7.1 Verification of Assumptions

We treat the first-order stochastic dominance case separately in our discussion from the higher-order stochastic dominance case because it allows for weaker assumptions on the distributions of Y_1 and Y_2.

7.1.1 First-Order Stochastic Dominance

Recall that the conditional moment inequalities implied by first-order conditional stochastic dominance are

$$E_{F_0}[1\{Y_2 \le \tau\} - 1\{Y_1 \le \tau\}|X] \ge 0 \text{ a.s. } \forall \tau \in \mathcal{T}. \tag{7.1}$$

The moment conditions for this model do not depend on a parameter θ. Hence, to fit the notation with that of the general theory, we set $\Theta = \{0\}$ (without loss of generality). Also observe that $p = k = 1$ in this example.

For this example, we use $\sigma_{F,1}(0) = \widehat{\sigma}_{n,1}(0) = 1$ for all F because the moment function has a natural scale. Hence, Assumptions SIG1 and SIG2 hold.

Lemma 7.1 *Let $\mathcal{F}_+$ be the set of $(0, F)$ such that $\{(Y_{1,i}, Y_{2,i}, X_i')' : i \ge 1\}$ are i.i.d. under F. Then, $\mathcal{F}_+$ satisfies Assumptions* PS1 and PS2 *with $M(w) = 1$, $\delta > 0$, and $C_1 = 1$.*

7.1.2 Higher-Order Stochastic Dominance

The conditional moment inequalities implied by sth-order conditional stochastic dominance for $s > 1$ are

$$E_{F_0}[(\tau - Y_2)^{s-1} 1\{Y_2 \le \tau\} - (\tau - Y_1)^{s-1} 1\{Y_1 \le \tau\}|X] \ge 0 \text{ a.s. } \forall \tau \in \mathcal{T}. \tag{7.2}$$

As above, we set $\Theta = \{0\}$. In this example, $p = k = 1$.

For this example, we use $\sigma_{F,1}^2(0) = E_F[(Y_1 - E(Y_1))^2] + E_F[(Y_2 - E(Y_2))^2]$ and $\widehat{\sigma}_{n,1}^2(0) = n^{-1}\sum_{i=1}^n [(Y_{1,i} - \overline{Y}_{1,n})^2 + (Y_{2,i} - \overline{Y}_{2,n})^2]$, where $\overline{Y}_{j,n} = n^{-1}\sum_{i=1}^n Y_{j,n}$ for $j = 1, 2$.

Lemma 7.2 *Suppose $s > 1$. Let $\underline{\sigma} > 0$ and $B \in (0, \infty)$ be constants. Let $\mathcal{F}_+$ be the set of (θ, F) for which* (i) *$\theta \in \Theta$,* (ii) *$\{(Y_{1,i}, Y_{2,i}, X_i')' : i \ge 1\}$ are i.i.d. under F,* (iii) *$\sigma_{F,1}^2(0) \ge \underline{\sigma}^2$, and* (iv) *$\mathcal{T} \subset [-B, B]$. Then,*

(a) *$\mathcal{F}_+$ satisfies Assumptions* PS1 and PS2 *with $M(w) = [(B - y_2)^{s-1} + (B - y_1)^{s-1}]/\sigma_{F,1}(0)$, $\delta > 0$, and $C_1 = 2^{s(2+\delta)} B^{(s-1)(2+\delta)} \underline{\sigma}^{-(2+\delta)}$, and*

(b) *Assumptions* SIG1 *and* SIG2 *hold.*

7.2 Monte Carlo Results

In this subsection, we report Monte Carlo results for testing the first-order conditional stochastic dominance between the conditional distributions of Y_1 and Y_2 given X. We consider the tests proposed above based on the CvM and KS test statistics combined with the GMS and PA critical values. For comparative purposes, we also consider the CvM and KS test statistics combined with subsampling critical values.

In this example, we take the instrument X to have the uniform $[0, 1]$ distribution and take Y_1 and Y_2 to have log-normal distributions given X:

$$\begin{aligned} Y_1 &= \exp(\sigma_1(X)Z_1 + \mu_1(X)) \text{ and} \\ Y_2 &= \exp(\sigma_2(X)Z_2 + \mu_2(X)), \end{aligned} \tag{7.3}$$

where $\sigma_1(X)$, $\mu_1(X)$, $\sigma_2(X)$, and $\mu_2(X)$ determine whether and how the null hypothesis that Y_1 first-order stochastically dominates Y_2 given X is violated.

To generate the simulated data, we let $\mu_1(X) = c_1X+c_3$, $\sigma_1(X) = c_2X+c_4$, $\mu_2(X) = 0.85$, and $\sigma_2(X) = 0.6$. These data-generating processes (DGPs) are adapted from Barrett and Donald (2003). Four values of $\mathbf{c} := (c_1, c_2, c_3, c_4)$ are considered: $\mathbf{c}_A = (0, 0, 0.85, 0.6)$, $\mathbf{c}_B = (0.15, 0, 0.85, 0.6)$, $\mathbf{c}_C = (-0.25, 0.2, 0.85, 0.6)$, and $\mathbf{c}_D = (0.35, 0, 0.85, 0.23)$. With $\mathbf{c}_A$ and $\mathbf{c}_B$, the null that Y_1 first-order stochastically dominate Y_2 conditional on X holds, while with $\mathbf{c}_C$ and $\mathbf{c}_D$, the null hypothesis is violated. To visualize the nature of the DGPs, we draw in Figure 1 the conditional cdf's of Y_1 and Y_2 given $X = 1$ at these four $\mathbf{c}$ values.

Note that with $\mathbf{c}_A$, Y_1 and Y_2 have identical distributions conditional on X. In this case, all of the moment inequalities are binding. The test should ideally have rejection probability equal to its nominal level in this boundary case. For this reason, we use this DGP to size-correct the rejection probabilities under the two alternative DGP's $\mathbf{c}_C$ and $\mathbf{c}_D$.

In the implementation of the tests, we compute the supremum over $\mathcal{T}$ by discretization. Specifically, we approximate $\mathcal{T}$ by N_τ points in $\mathcal{T}$ for a positive integer N_τ. The N_τ points on $\mathcal{T}$ are chosen as follows: first pool the n observations of Y_1 and those of Y_2 to get a sample of size $2n$. Then use as grid points the $1/(N_\tau + 1), 2/(N_\tau + 1), \ldots, N_\tau/(N_\tau + 1)$ percentiles of this $2n$ sample.

For the sample size and the tuning parameters of the test, we consider a base case with the sample size $n = 250$, the hypercube parameter $r_{1,n} = 3$, and $N_\tau = 10$. Then, for

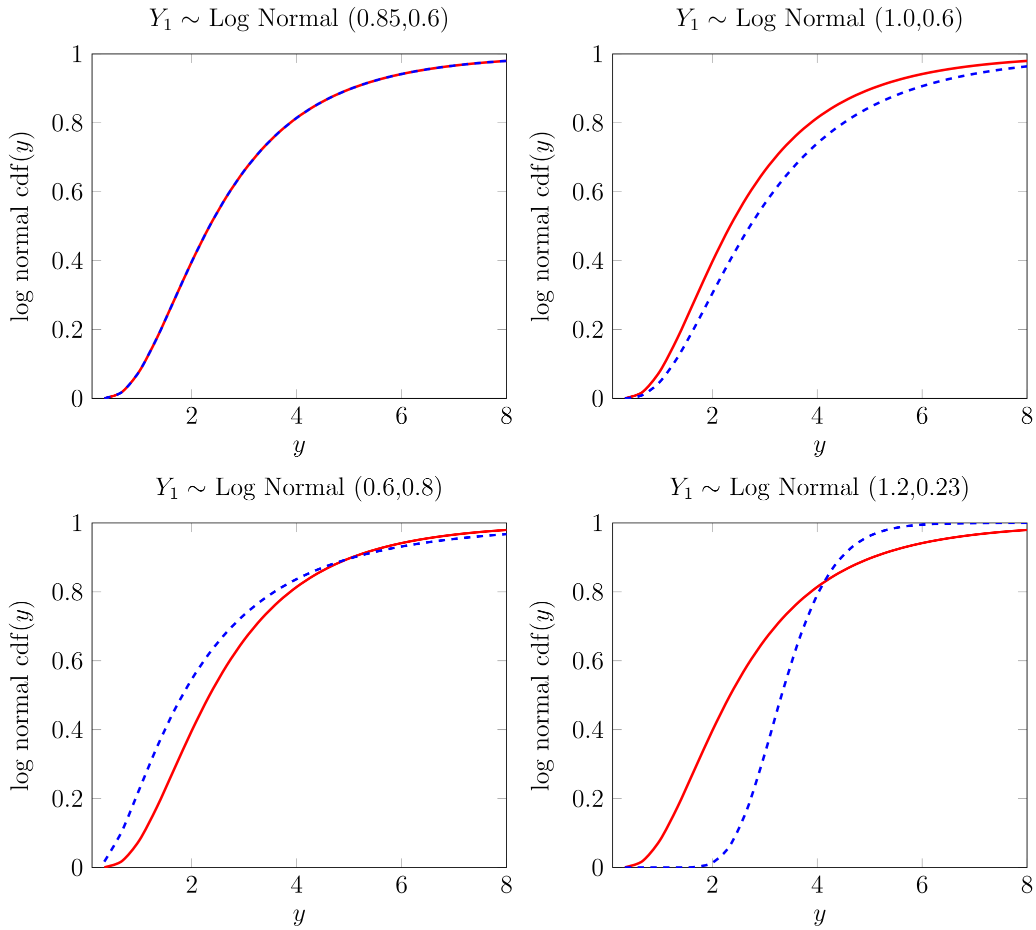

Figure 1: Conditional CDF's of Y_1 (dashed blue) and Y_2 (solid red) given $X = 1$. In all graphs, $Y_2 \sim$ Log Normal $(0.85, 0.6)$.

comparison, we also consider three variations of the base case where each differs from the base case in only one dimension.

Simulated rejection probabilities based on 1000 simulation repetitions are reported in Tables 1 and 2. Table 1 reports the rejection probabilities under the two null DGP's and Table 2 reports the size-corrected rejection probabilities under the two alternative DGP's.

As the tables show, the CvM/GMS test performs the best overall in that it has the most accurate size and relatively high power. The KS/GMS test has somewhat worse power perhaps due to the DGP designs. The tests based on PA critical values do not perform as well in terms of power in the second alternative. The CvM/Subsampling test has slightly greater over-rejection than, and comparable power to, the CvM/GMS test. On the other hand, the KS/Subsampling test exhibits substantial over-rejection in an absolute sense and noticeably lower power than the KS/GMS test.

Table 1: Null Rejection Probabilities for Nominal .05 First-Order Stochastic Dominance Tests

	CvM/GMS	KS/GMS	CvM/PA	KS/PA	CvM/Sub	KS/Sub
	Null 1: $(c_1, c_2, c_3, c_4) = (0, 0, 0.85, 0.6)$					
Basecase: $(n = 250,\ r_{1,n} = 3,\ N_\tau = 10)$	.059	.070	.052	.059	.072	.198
$n = 500$	.040	.047	.037	.043	.066	.176
$r_{1,n} = 4$	.062	.068	.054	.057	.080	.243
$N_\tau = 15$	.065	.066	.058	.055	.069	.205
	Null 2: $(c_1, c_2, c_3, c_4) = (0.15, 0, 0.85, 0.6)$					
Base case: $(n = 250,\ r_{1,n} = 3,\ N_\tau = 10)$	.018	.020	.011	.014	.027	.107
$n = 500$	.005	.010	.002	.006	.014	.079
$r_{1,n} = 4$	.019	.022	.012	.014	.030	.141
$N_\tau = 15$	.016	.020	.012	.014	.023	.120

Note: The subsampling critical values use a subsample size of 20 and, for computation reasons, not all subsamples are used. The bootstrap and subsampling critical values both use 1000 repetitions to simulate the critical values.

Table 2: Size-corrected Alternative Rejection Probabilities for Nominal .05 First-Order Stochastic Dominance Tests

	CvM/GMS	KS/GMS	CvM/PA	KS/PA	CvM/Sub	KS/Sub
	Alternative 1: $(c_1, c_2, c_3, c_4) = (-0.25, 0.2, 0.85, 0.6)$					
Basecase: $(n = 250,\ r_{1,n} = 3,\ N_\tau = 10)$	.482	.349	.497	.363	.450	.280
$n = 500$	.803	.673	.798	.671	.799	.581
$r_{1,n} = 4$	.490	.354	.507	.367	.450	.267
$N_\tau = 15$	.487	.384	.502	.395	.456	.316
	Alternative 2: $(c_1, c_2, c_3, c_4) = (0.35, 0, 0.85, 0.23)$					
Base case: $(n = 250,\ r_{1,n} = 3,\ N_\tau = 10)$	.482	.244	.280	.128	.525	.311
$n = 500$	.866	.707	.717	.539	.879	.718
$r_{1,n} = 4$	.478	.232	.269	.109	.516	.300
$N_\tau = 15$	.512	.300	.299	.141	.599	.354

Note: The subsampling critical values use a subsample size of 20. The bootstrap and subsampling critical values use 1000 repetitions to simulate the critical values. Size correction is carried out using the null DGP with $(c_1, c_2, c_3, c_4) = (0, 0, 0.85, 0.6)$.

8 Example 2: Random-Coefficients Binary-Outcome Models with Instrumental Variables

We focus on the parametrized version of the model given in (2.9).

8.1 Verification of Assumptions

Let d_1 and d_2 denote the dimensions of the exogenous covariate X_1 and the endogenous covariate Y_2, respectively. We treat the case where $d_1 + d_2 = 1$ separately because it allows for simpler arguments than the general case.

8.1.1 Single Covariate Model

Suppose $d_1 + d_2 = 1$. That is, there is only one covariate in the binary choice model. This covariate could be either exogenous or endogenous. There is no restriction on the instrument vector X. This case includes Example 1 in CR.

For notational simplicity, let the single covariate be denoted by Z (with a generic realization denoted by z) and its coefficient be denoted by β_z (with a generic realization denoted by b_z). Then, the set $S(y_1, y_2, x_1)$ simplifies to

$$S(y_1, z) = cl\{b = (b_0, b_z)' \in R^2 : y_1 = 1\{b_0 + b_z z \geq 0\}\}. \tag{8.1}$$

In fact, $S(y_1, z)$ can be rewritten as a half-space:

$$S(y_1, z) = \{b \in R^2 : (y_1 - 1/2)b_0 + (y_1 - 1/2)zb_z \geq 0\} = H(y_1 - 1/2, (y_1 - 1/2)z). \tag{8.2}$$

The following lemma gives a convenient representation for $H(c)$.

Lemma 8.1 *For any* $c \in R^2 \backslash \{0^2\}$,

$$H(c) = \{(\rho \cos(a), \rho \sin(a)) : \rho \geq 0, a \in [\phi(c), \phi(c) + \pi]\},$$

where $\phi(c) : R^2 \backslash \{0^2\} \to [0, 2\pi)$ *is defined by* $\phi(c) := \arctan(-c_1/c_2)$ *for* $c_1 \leq 0, c_2 > 0$; $\phi(c) := \pi/2$ *for* $c_1 < 0, c_2 = 0$; $\phi(c) := \arctan(-c_1/c_2) + \pi$ *for* $c_2 < 0$; $\phi(c) := 3\pi/2$ *for* $c_1 > 0, c_2 = 0$; *and* $\phi(c) := \arctan(-c_1/c_2) + 2\pi$ *for* $c_1 > 0, c_2 > 0$.

The following is an immediate corollary of (8.2) and Lemma 8.1.

Corollary 8.2 *For any subset $\mathcal{C}$ of $R^2 \backslash \{0^2\}$, we have*

(a) $cl(\cup_{c \in \mathcal{C}} H(c)) = \{(\rho\cos(a), \rho\sin(a)) : \rho \geq 0, a \in [\inf_{c \in \mathcal{C}} \phi(c), \sup_{c \in \mathcal{C}} \phi(c) + \pi]\}$, *and*

(b) $1\{S(Y_1, Z) \subset cl(\cup_{c \in \mathcal{C}} H(c))\} = 1\{A \in [\inf_{c \in \mathcal{C}} \phi(c), \sup_{c \in \mathcal{C}} \phi(c)]\}$, *where* $A := \phi((Y_1 - 1/2)(1, Z)')$.

As a result, any $\mathcal{S} \in \mathbf{S}$ can be written as $\mathcal{S}(\tau)$ for some $\tau \in \mathcal{T}$, where $\mathcal{T} = \{(\tau_1, \tau_2)' \in [0, 2\pi)^2 : \tau_1 \leq \tau_2\}$ and

$$\mathcal{S}(\tau) = \{(\rho\cos(a), \rho\sin(a)) : \rho \geq 0, a \in [\tau_1, \tau_2 + \pi]\}. \tag{8.3}$$

Thus, the moment inequality model (2.9) can be written as

$$E_{F_0}[F_\beta(\mathcal{S}(\tau), \theta) - 1\{A \in [\tau_1, \tau_2]\}|X] \geq 0 \text{ a.s. } \forall \tau = (\tau_1, \tau_2)' \in \mathcal{T}. \tag{8.4}$$

Now we are ready to verify Assumptions PS1 and PS2 for this model.

Observe that $p = k = 1$ in this example. For this example, we use $\sigma_{F,1}(\theta) = \widehat{\sigma}_{n,1}(\theta) = 1$ for all (θ, F) because the moment function has a natural scale. Hence, Assumptions SIG1 and SIG2 hold.

Lemma 8.3 *Let $\mathcal{F}_+$ be the set of (θ, F) such that $\theta \in \Theta$ and $\{(Y_{1,i}, Y_{2,i}, X_i')' : i \geq 1\}$ are i.i.d. under F. Then $\mathcal{F}_+$ satisfies Assumptions* PS1 and PS2 *with $M(w) = 1$, $\delta > 0$, and $C_1 = 1$.*

8.1.2 Multi-Covariate Model

Now we consider the case without the restriction that $d_1 + d_2 = 1$. When $d_1 + d_2 > 1$, Assumption PS2 does not hold in general because the Vapnik-Chervonenkis (VC) dimension of the set $\{(1\{S(Y_{1,i}, Y_{2,i}, X_{1,i}) \subset \mathcal{S}\})_{i=1}^n : \mathcal{S} \in \mathbf{S}\}$ typically diverges to infinity as $n \to \infty$, which violates the manageability condition. Thus, we need to restrict attention to a subset of $\mathbf{S}$. Fortunately, in many applications, restriction to an appropriate subset of $\mathbf{S}$ (specified below) does not affect the set identification power of the model. We apply our general theory to such applications.

For a positive integer m, we consider subsets of $\mathbf{S}$ of the form: $\mathbf{S}_m := \{\cup_{j=1}^m H(c_j) : c_j \in R^{d_\beta} \backslash \{0^{d_\beta}\}\}$. That is, $\mathbf{S}_m$ is the collection of at most m unions of half-spaces in R^{d_β} through the origin. Let $\Theta_F(\mathbf{S}_m) := \{\theta \in \Theta : E_F[F_\beta(\mathcal{S}, \theta) - 1\{S(Y_1, Y_2, X_1) \subset \mathcal{S}\}|X_1, X_2] \geq 0 \text{ a.s. } \forall \mathcal{S} \in \mathbf{S}_m\}$. Define $\Theta_F(\mathbf{S})$ analogously. The applications we consider are required to satisfy the following assumption. This assumption is satisfied in Example 2 of CR with $m = 2$ and Example 3 of CR with $m = 4$. This assumption is always satisfied when $d_1 + d_2 = 1$ because in that case $\mathbf{S}_m = \mathbf{S}$ for $m = 2$.

Assumption V1. $\Theta_{F_0}(\mathbf{S}_m) = \Theta_{F_0}(\mathbf{S})$.

Under this assumption, we can base inference on the conditional moment inequality model:

$$E_F[F_\beta(\mathcal{S}, \theta) - 1\{S(Y_1, Y_2, X_1) \subset \mathcal{S}\}|X_1, X_2] \geq 0 \text{ a.s. } \forall \mathcal{S} \in \mathbf{S}_m. \tag{8.5}$$

As in the single covariate case, we first write $S(y_1, y_2, x_1)$ in the canonical form of a half-space:

$$\begin{aligned} S(y_1, y_2, x_1) &= cl\{b = (b_0, b_1', b_2')' \in R^{d_\beta} : y_1 = 1\{b_0 + b_1' x_1 + b_2' y_2 \geq 0\}\} \\ &= H((y_1 - 1/2)(1, x_1', y_2')'). \end{aligned} \tag{8.6}$$

The following lemma yields a convenient representation of the event $\{S(Y_1, Y_2, X_1) \subset \mathcal{S}\}$ for $\mathcal{S} \in \mathbf{S}_m$.

Lemma 8.4 *For any $c_1, ..., c_m \in R^{d_\beta} \backslash \{0^{d_\beta}\}$ (not necessarily distinct from each other), there exists a $d_\beta \times M$ real matrix $B(c_1, ..., c_m)$ with $M = \max_{j \in \{1,...,d_\beta\}} \left[\binom{m}{\min\{j,m\}-1} + 2(d_\beta - j) \right]$ such that, for any $\bar{c} \in R^{d_\beta} \backslash \{0^{d_\beta}\}$, the following statements are equivalent*:

(a) $H(\bar{c}) \subset \cup_{j=1}^m H(c_j)$,

(b) $\bar{c} = \sum_{j=1}^m \lambda_j c_j$ *for some* $\lambda_1, ..., \lambda_m \geq 0,$ *and*

(c) $B(c_1, ..., c_m)' \bar{c} \geq 0^M$.

The lemma implies that the conditional moment inequality model (8.5) has the following equivalent representation:

$$E_F[F_\beta(\mathcal{S}(\tau), \theta) - 1\{(Y_1 - 1/2) B(\tau)'(1, X_1', Y_2')' \geq 0\}|X_1, X_2] \geq 0 \text{ a.s. } \forall \tau \in \mathcal{T}, \tag{8.7}$$

where $\mathcal{T} = \{\tau = (c_1, ..., c_m) : c_1, ..., c_m \in R^{d_\beta} \backslash \{0^{d_\beta}\}\}$, $B(\tau) := B(c_1, ..., c_m)$, and $\mathcal{S}(\tau) = \cup_{j=1}^m H(c_j)$.

Because the VC dimension of the set $\{(1\{(Y_{1,i} - 1/2)(1, X'_{1,i}, Y'_{2,i})B \geq 0\})_{i=1}^n : B = [b^1, \cdots, b^M]$ for $b^j \in R^{d_\beta}\}$ is bounded ($\leq M$), we can verify Assumption PS2 for the model (8.5) through the equivalent representation in (8.7).[6] As in the single covariate case, $p = k = 1$ and we use $\sigma_{F,1}(\theta) = \widehat{\sigma}_{n,1}(\theta) = 1$ for all (θ, F) because the moment function has a natural scale. Hence, Assumptions SIG1 and SIG2 hold.

Lemma 8.5 *For the model in* (8.5)*, let* $\mathcal{F}_+$ *be the set of* (θ, F) *such that* $\theta \in \Theta$ *and* $\{(Y_{1,i}, Y'_{2,i}, X'_i)' : i \geq 1\}$ *are i.i.d. under* F*. Then* $\mathcal{F}_+$ *satisfies Assumptions* PS1 and PS2 *with* $M(w) = 1$*,* $\delta > 0$*, and* $C_1 = 1$*.*

8.2 Monte Carlo Results

In this subsection, we report Monte Carlo results for a binary choice model similar to the numerical example in CR. The model has one endogenous regressor (Y_2), one instrument variable (X), and no exogenous regressors. That is,

$$Y_1 = 1\{\beta_0 + \beta_1 Y_2 < 0\} \text{ with } (\beta_0, \beta_1) \perp X.$$

Further assume that β_0 and β_1 are jointly normal: $\beta_0 = \alpha_0 + U_0$ and $\beta_1 = \alpha_1 + U_1$, where

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \gamma_0 \\ \gamma_0 & \gamma_1 + \gamma_0 \end{pmatrix} \right).$$

Thus, the model contains the unknown parameter $\theta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)'$.

For the data generating processes, we consider

$$Y_2 = \delta_1 X + \delta_2 U_0 + \delta_3 U_1 + \delta_4 V, \tag{8.8}$$

where $X \sim N(0, 1)$ is independent of (U_0, U_1) and $V \sim N(0, 1)$ is independent of (X, U_0, U_1).

[6] Note that the representation (8.7) is simply a technical device useful for the theory and for intuitive understanding, and is *not* needed in practice. Thus, we do not need to know the form of the mapping $B(\cdot)$. This is important because its form is typically complicated. Mathematically, each column of B is the polar of a facet of the convex (pointed) polyhedral cone spanned by $c_1, ..., c_m$. Algebraic representations of facets of convex polyhedral cones are complicated.

Let $\theta = (0, -1, -1, 1)'$, and $\delta := (\delta_1, \delta_2, \delta_3, \delta_4)' = (1, 0.577, -0.577, 0.577)'$.[7]

We consider the test proposed above based on the CvM and KS test statistics and the GMS, PA, and subsampling critical values. In the implementation of the tests, we choose $r_{1,n} = 3$ and approximate $\mathcal{T}$ by grid points.[8] In the implementation of the subsampling test, we set the subsample size to 20. We use 1001 repetitions to simulate the bootstrap and the subsampling critical values, and carry out 1000 Monte Carlo repetitions to obtain the simulated coverage probabilities of given points of θ.

To choose the points to cover, we fix $\alpha_0, \gamma_0, \gamma_1$ at their true values, and shift α_1 up to obtain a boundary point of the identified set. There is no analytical solution to the boundary for this example, and thus we computed the boundary point numerically. Computation shows that the upper boundary of the identified set for α_1 (when the rest of the parameters are fixed at the true value) is approximately -0.8274.[9]

Figure 2 provides coverage probability graphs for sample sizes $n = 250$ and $n = 500$. As the figure shows, the coverage probabilities of the CS's equal one not only at the boundary of the identified set, but also for points to the right of the boundary and outside the identified set. This is probably because the boundary of the identified set is determined by X values in a set with Lebesgue measure zero. The coverage probabilities of the CS's for points outside the identified set decrease with the sample size, as expected. The best performing CS's are the KS/GMS and the KS/subsampling CS's. (In the figure for $n = 500$, the graphs for these two CS's are on top of each other and form the left-most curve in the figure.) The CvM/GMS CS performs better than the subsampling CvM CS. The PA CS's perform uniformly worse than the GMS and the subsampling CS's in terms of coverage probabilities for points outside the identified set.

[7] This δ value is the weak-identification specification in CR. Since identification strength is irrelevant for evaluating the property of our test, we focus on this weak-identification specification and do not consider the other specification.

[8] Because $\mathcal{T}$ is the collection of unions of two half-spaces through the origin, the elements in $\mathcal{T}$ correspond one-to-one to $\{(t_1, t_2) \in [0, 2\pi] : t_1 \leq t_2]\}$, where t_1 and t_2 are the polar angles of the coefficient vector of the half-spaces forming the union. We consider N_{t_2} equally-spaced grid points for t_2 in $[0, 2\pi]$, and grid points for t_1 in $[0, t_2]$ with the same spacing. We set $N_{t_2} = 10$, which results in 55 points on $\{(t_1, t_2) \in [0, 2\pi] : t_1 \leq t_2]\}$.

[9] Specifically, the way we compute the boundary is as follows. First we construct the criterion function $Q(\theta) = \min_{x \in \mathcal{X}_{n_x}} \min_{\tau \in \mathcal{T}_{n_\tau}} F_\beta(\mathcal{S}(\tau), \theta) - \widehat{E}[1\{A \in [\tau_1, \tau_2]\}|X = x]$, where $\mathcal{X}_{N_x}$ is the set of $N_x = 20$ equally-spaced grid points in the interval $[-4, 4]$, $\mathcal{T}_{N_\tau}$ is the approximation of $\mathcal{T}$ described in the previous footnote, $F_\beta(\mathcal{S}_\tau, \theta)$ is computed using the bivariate-normal cdf function in Aptech Gauss, and $\widehat{E}[1\{A \in [\tau_1, \tau_2]\}|X = x]$ is computed using i.i.d. Monte Carlo simulations with 10^7 simulation repetitions. Then we fix $\alpha_0, \gamma_0, \gamma_1$ at their true values, and search for $a_1 > -1$ that makes $Q(\alpha_0, a_1, \gamma_0, \gamma_1)$ zero. The function $Q(\alpha_0, \cdot, \gamma_0, \gamma_1)$ appears to be monotonically decreasing in the range $[-1, 2]$ and changes signs from one end point to the other.

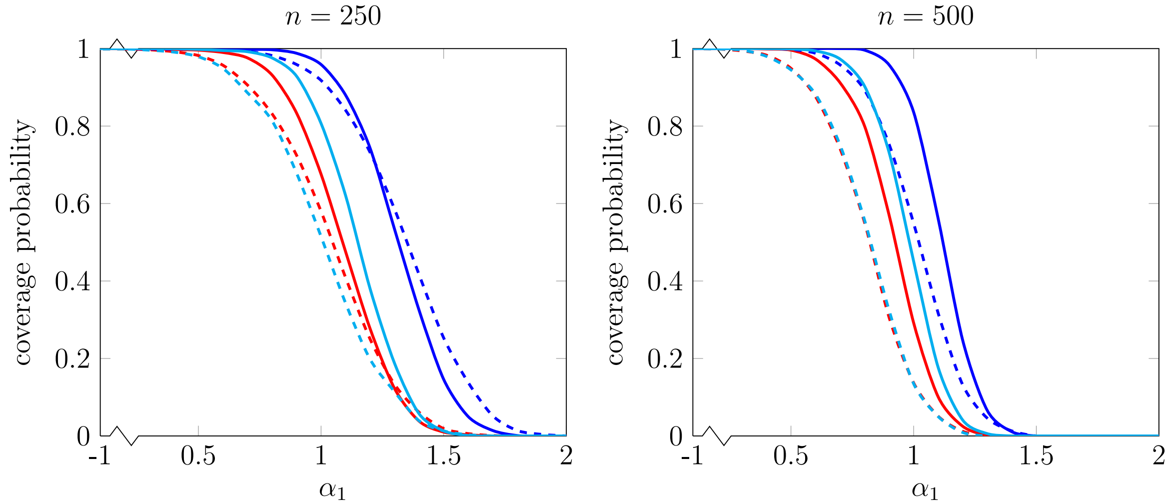

Figure 2: Coverage Probabilities in the IV Random-Coefficients Binary-Outcome Model. (Solid lines: CvM statistic; dashed lines: KS statistic; red: GMS critical value; blue: PA critical value; cyan: subsampling critical value. Nominal size = .95. $(\alpha_0, \gamma_0, \gamma_1) = (0, -1, 1)$.)

9 Examples 3-5

In this section, we verify the high-level assumptions for Examples 3, 4, and 5.

9.1 Example 3: Convex Moment Prediction Models–Support Function Approach

As mentioned above, Beresteanu, Molchanov, and Molinari (2010) verify a version of the high-level conditions given in an earlier version of our paper for the best linear predictor and entry-game applications of this example. In this subsection, we verify our current high-level conditions for the general BMM framework in (2.12).

We focus on the moment inequality model in (2.12) because it includes the case where $Q_\theta(W, V) = Q_\theta(W)$ as a special case. For this model, $p = k = 1$. For simplicity, we take $\widehat{\sigma}_{n,1}(\theta) = \sigma_{F,1}(\theta) = 1$ for all (θ, F) and all n, and hence Assumptions SIG1 and SIG2 hold. Alternatively, one could choose $\sigma_{F,1}(\theta)$ and $\widehat{\sigma}_{n,1}(\theta)$ that are scale equivariant in the spirit of those in Section 7.1.2.

Lemma 9.1 *For the model in* (2.12)*, let* $\mathcal{F}_+$ *be the set of* (θ, F) *such that* (i) $\theta \in \Theta$*,* (ii) $\{W_i : i \geq 1\}$ *are i.i.d. under* F*,* (iii) $Q_\theta(w, v) \subseteq \{q \in R^d : \|q\| \leq M(w)/2\}$ *for some measurable function* $M(w)$ $\forall (w, v) \in \mathcal{W}\mathcal{V}$*,* (iv) $\|q(x)\| \leq M(w)/2$ $\forall x \in \mathcal{X}$*,* $\forall w \in \mathcal{W}$*, and*

(v) $E_F[M(W)^{2+\delta}] \leq C_1$ *for some* $\delta > 0$ *and* $C_1 < \infty$. *Then,* $\mathcal{F}_+$ *satisfies Assumptions* PS1 *and* PS2 *with* $M(w)$, δ, *and* C_1 *as defined immediately above.*

9.2 Examples 4 and 5: Countable Conditional Moment Inequalities

In this subsection, we verify the high-level assumptions for models with countably many conditional moment inequalities. Examples 4 and 5 are of this type.

Suppose that the identification theory implies the following moment inequality model:

$$E_{F_0}[\widetilde{m}(W,\theta,\tau)|X] \geq 0, \text{ for } \tau = 1,2,3,\ldots, \tag{9.1}$$

where $\widetilde{m}(W,\theta,\tau)$ is a real-valued moment function. For example, these moment conditions could be the ones in (2.16) or (2.18).

In general, the raw moment functions $\widetilde{m}(W,\theta,\tau)$ may not satisfy Assumption PS2. Thus, we rescale them with weights that decrease with τ. Let $w_{\mathcal{T}}(\tau) : [1,\infty) \to (0,1]$ be a strictly decreasing, positive, weight function with inverse function $\lambda_{\mathcal{T}}(\xi) : (0,1] \to [1,\infty)$ that satisfies $\int_0^1 \sqrt{\log(\lambda_{\mathcal{T}}(\xi))}d\xi < \infty$. Then, we let

$$m(W,\theta,\tau) = w_{\mathcal{T}}(\tau)\widetilde{m}(W,\theta,\tau) \ \forall \tau = 1,2,\ldots\ . \tag{9.2}$$

In consequence, the moment inequality model (9.1) is equivalent to

$$E_{F_0}[m(W,\theta,\tau)|X] \geq 0 \ \forall \tau = 1,2,\ldots\ . \tag{9.3}$$

We verify the high-level assumptions given above for this rescaled form of the moment inequalities.

For example, if $w_{\mathcal{T}}(\tau) = \tau^{-b}$ for some $b > 0$, then $\lambda_{\mathcal{T}}(\xi) = \xi^{-1/b}$ and

$$\int_0^1 \sqrt{\log(\xi^{-1/b})}d\xi = \sqrt{1/b}\int_0^1 \sqrt{\log(\xi^{-1})}d\xi = b^{-1/2}\int_0^\infty 2x^2 e^{-x^2}dx < \infty, \tag{9.4}$$

where the last equality holds by change of variables with $x = \sqrt{\log(\xi^{-1})}$ (or, equivalently, $\xi = e^{-x^2}$). For general purposes, we recommend $b = 1/2$.

For this model, $p = k = 1$, and we use $\sigma^2_{F,1}(\theta) = Var_F(m(W, \theta, 1))$ and $\widehat{\sigma}^2_{n,1}(\theta) = n^{-1}\sum_{i=1}^n [m(W_i, \theta, 1) - \overline{m}_n(\theta, 1)]^2$, where $\overline{m}_n(\theta, 1) = n^{-1}\sum_{i=1}^n m(W_i, \theta, 1)$.

Lemma 9.2 *For the model in* (9.3)*, let* $\mathcal{F}_+$ *be the set of* (θ, F) *such that* (i) $\theta \in \Theta$, (ii) $\{W_i : i \geq 1\}$ *are i.i.d. under* F*,* (iii) $\sigma^2_{F,1}(\theta) \geq \underline{\sigma}^2$ *for some constant* $\underline{\sigma}^2 > 0$, (iv) $|\widetilde{m}(w, \theta, \tau)| \leq B(w)$ $\forall w \in \mathcal{W}$*,* $\forall \tau \in \mathcal{T}$*,* $\forall \tau \in \Theta$*, for some measurable function* $B(w)$*, and* (v) $E[(B(W)/\underline{\sigma})^{2+\delta}] \leq C_1$ *for some* $\delta > 0$ *and* $C_1 < \infty$*. Let* $w_{\mathcal{T}}(\tau)$ *be a weight function that satisfies the definition above. Then,*

(a) $\mathcal{F}_+$ *satisfies Assumptions* PS1 *and* PS2 *with* $M(w) = B(w)/\underline{\sigma}$ *and with* C_1 *and* δ *defined immediately above, and*

(b) *Assumptions* SIG1 *and* SIG2 *hold.*

10 Conclusion

In this paper, we construct confidence sets for models defined by many conditional moment inequalities/equalities. The conditional moment restrictions in the models can be finite, countably infinite, or uncountably infinite. To deal with the complication brought about by the vast number of moment restrictions, we exploit the manageability (Pollard (1990)) of the class of moment functions. We verify the manageability condition in five examples from the recent partial-identification literature.

The proposed confidence sets are constructed by inverting joint tests that employ all of the moment restrictions. The confidence sets are shown to have correct asymptotic size in a uniform sense and to exclude parameter values outside the identified set with probability approaching one. Monte Carlo experiments for a conditional stochastic dominance example and a random-coefficients binary-outcome example support the theoretical results.

References

Anderson, G., 1996. Nonparametric tests of stochastic dominance in income distributions. Econometrica 64, 1183–1193.

Anderson, G., 2004. Toward an empirical analysis of polarization. Journal of Econometrics 122, 1–26.

Andrews, D. W. K., Shi, X., 2009. Inference based on conditional moment inequalities. Cowles Foundation, Yale University. Unpublished manuscript.

Andrews, D. W. K., Shi, X., 2013a. Inference based on conditional moment inequalities. Econometrica 81, 609–666.

Andrews, D. W. K., Shi, X., 2013b. Supplemental material to 'Inference based on conditional moment inequalities.' Available at Econometrica Supplemental Material 81, http://www.econometricsociety.org/ecta/supmat/9370_Proofs.pdf.

Andrews, D. W. K., Shi, X., 2014. Nonparametric inference based on conditional moment inequalities. Journal of Econometrics 179, 31–45. Appendix of proofs available as Cowles Foundation Discussion Paper No. 1840RR, Yale University, 2011.

Armstrong, T. B., 2014a. On the choice of test statistic for conditional moment inequalities. Cowles Foundation Discussion Paper No. 1960, Yale University.

Armstrong, T. B., 2014b. Weighted KS statistics for inference on conditional moment inequalities. Journal of Econometrics 181, 92–116.

Armstrong, T. B., 2015. Asymptotically exact inference in conditional moment inequality models. Journal of Econometrics 186, 51–85.

Armstrong, T. B., Chan, H. P., 2013. Multiscale adaptive inference on conditional moment inequalities. Cowles Foundation Discussion Paper No. 1885R, Yale University.

Barrett, G. F., Donald, S. G., 2003. Consistent tests for stochastic dominance. Econometrica 71, 71–104.

Barrett, G. F., Donald, S. G., Bhattacharya, D., 2014. Consistent nonparametric tests for Lorenz dominance. Journal of Business and Economic Statistics 32, 1–13.

Beare, B. K., Moon, J.-M., 2015. Nonparametric tests of density ratio ordering. Econometric Theory 31, 471–492.

Beare, B. K., Shi, X., 2015. An improved bootstrap test of density ratio ordering. Department of Economics, University of Wisconsin. Unpublished manuscript.

Beresteanu, A., Molchanov, I., Molinari, F., 2010. Sharp identification regions in models with convex predictions. Econometrica 79, 1785–1821.

Bishop, J. A., Zeager, L. A., Zheng, B., 2011. Interdistributional inequality, stochastic dominance, and poverty. Department of Economics, East Carolina University. Unpublished manuscript.

Chernozhukov, V., Chetverikov, D., Kato, K., 2014. Testing many moment inequalities. Department of Economics, MIT. Unpublished manuscript.

Chernozhukov, V., Lee, S., Rosen, A., 2013. Intersection bounds: estimation and inference. Econometrica 81, 667–737.

Chesher, A., Rosen, A. M., 2014. An instrumental variable random-coefficients model for binary outcomes. Econometrics Journal 17, S1–S19.

Chesher, A., Smoliński, K., 2012. IV models of ordered choice. Journal of Econometrics 166, 33–48.

Chetverikov, D., 2012. Adaptive test of conditional moment inequalities. Department of Economics, UCLA. Unpublished manuscript.

Dardanoni, V., Forcina, A., 1999. Inference for Lorenz curve orderings. Econometrics Journal 2, 49–75.

Davidson, R., Duclos, J.-Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica 68, 1435–1464.

Delgado, M. A., Escanciano, J. C., 2013. Conditional stochastic dominance testing. Journal of Business and Economic Statistics 31, 16–28.

Galichon, A., Henry, M., 2009. A test of non-identifying restrictions and confidence regions for partially identified parameters. Journal of Econometrics 152, 186–196.

Gandhi, A., Lu, Z., Shi, X., 2013. Estimating demand for differentiated products with error in market shares. Department of Economics, University of Wisconsin. Unpublished manuscript.

Guerre, E., Perrigne, I., Vuong, Q., 2009. Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. Econometrica 77, 1193–1227.

Kahn, S., Tamer, E., 2009. Inference on endogenously censored regression models using conditional moment inequalities. Journal of Econometrics 152, 104–119.

Lee, S., Song, K., Whang, Y.-J., 2013. Testing functional inequalities. Journal of Econometrics 172, 14–32.

Linton, O., Song, K., Whang, Y.-J., 2010. An improved bootstrap test of stochastic dominance. Journal of Econometrics 154, 186–202.

Pakes, A., Porter, J., Ho, K., Ishii, J., 2015. Moment inequalities and their application. Econometrica 83, 315–334.

Pollard, D., 1990. Empirical Process Theory and Application, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. II. Institute of Mathematical Statistics.