Inference based on many conditional moment inequalities

Donald W.K. Andrews [a,*], Xiaoxia Shi [b]

[a] *Cowles Foundation, Yale University, P.O. Box 208281, New Haven, CT 06520-8281, United States*
[b] *Department of Economics, University of Wisconsin-Madison, United States*

ARTICLE INFO

ABSTRACT

We construct confidence sets for models defined by many conditional moment inequalities/equalities. The number of conditional moment restrictions can be up to infinitely many. To deal with the vast number of moment restrictions, we exploit the manageability (Pollard (1990)) of the class of moment functions. We verify this condition in five examples from the recent partial identification literature.

The confidence sets are shown to have correct uniform asymptotic size 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-coefficient binary-outcome example support the theoretical results.

© 2016 Elsevier B.V. All rights reserved.

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 et al. (2010), (4) ordered-choice models with endogeneity and instruments, see Chesher and Smoliński (2012), and (5) discrete games identified by revealed preference, see Pakes et al. (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 in 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 covering numbers of the functional classes that arise. The upper bounds in the first two examples are derived by bounding the pseudo-dimensions 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.

---

* Corresponding author. Fax: +1 203 432 6167.
E-mail address: donald.andrews@yale.edu (D.W.K. Andrews).

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 et al. (Unpublished Manuscript), which studies models defined by many moment inequalities. They construct two types of tests, one based on a fixed critical value derived using a moderate deviation inequality, and the other based on a bootstrap critical value derived using distributional approximation theory for suprema of empirical processes developed in Chernozhukov et al. (Unpublished Manuscript, 2014b). Both are based on a supremum-type test statistic, which is similar to, but different from, the KS statistic considered here. In one Monte Carlo example considered here, the one-step and two-step versions of their tests do not perform as well as the MCMI tests proposed in this paper. In the other Monte Carlo example considered here, their two-step bootstrap-based method performs better than the MCMI methods proposed in this paper at a large sample size, but not as well at smaller sample sizes.

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 Kahn and Tamer (2009), Chetverikov (Unpublished Manuscript), AS1, Armstrong and Chan (2013), Chernozhukov et al. (2013), Gandhi et al. (Unpublished Manuscript), Lee et al. (2013), Andrews and Shi (2014), and Armstrong (2014a,b, 2015). Galichon and Henry (2009) provide 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 et al. (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 et al. (2014), and tests of likelihood ratio (or density) ordering, see Beare and Moon (2015), Beare and Shi (Unpublished Manuscript), 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 MCMI 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. An Appendix available online provides proofs and some additional simulation results.

For notational simplicity, throughout the paper, we let $(a_i)_{i=1}^n$ denote the n-vector $(a_1, \ldots, a_n)'$ 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:

$$E_{F_0}[m_j(W_i, \theta_0, \tau)|X_i] \geq 0 \text{ a.s. for } j = 1, \ldots, p \text{ and}$$
$$E_{F_0}[m_j(W_i, \theta_0, \tau)|X_i] = 0 \text{ a.s. for } j = p+1, \ldots, p+v, \ \forall \tau \in \mathcal{T},$$
(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 sub-vector of W_i, and $m(w, \theta, \tau) := (m_1(w, \theta, \tau), \ldots, m_{p+v}(w, \theta, \tau))'$ is a vector of known moment functions.[1]

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

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

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

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

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, 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:

$$E_{F_0}[G_s(Y_2, \tau) - G_s(Y_1, \tau)|X] \geq 0 \text{ a.s. } \forall \tau \in \mathcal{T}, \text{ where}$$

$$G_s(y, \tau) := (\tau - y)^{s-1} 1\{y \leq \tau\} \quad (2.4)$$

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 et al. (Unpublished Manuscript). Stochastic dominance relationships also have been used in the study of auctions, e.g., in Guerre et al. (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

---

[1] The requirement that X_i is a sub-vector of W_i does not preclude X_i from containing excluded instruments because $m(W_i, \theta, \tau)$ is not required to vary with every element of W_i.

Y_1 sth order stochastically dominates Y_2 and Y_2 dominates Y_3 conditional on X. The corresponding conditional moment inequalities to be tested are as follows,

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

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):

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

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:

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

where

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

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:

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

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 et al. (2010) (BMM) establish a framework to characterize the sharp identified set for a general class of incomplete models with convex moment predictions using 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 sub-vector 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

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

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

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

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

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

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 Smoliński (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 non-separable model

$$Y = h(Z, U), \quad (2.13)$$

where $Y \in \{1, 2, \ldots, M\}$ and $Z \in \{z_1, \ldots, z_K\}$, 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 $m = 1, \ldots, M$ and $z \in \{z_1, \ldots, z_K\}$:

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

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

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

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

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

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 Smoliński (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 et al. (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, 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

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

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 that a functional central limit theorem (CLT) holds, 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. This concept also is used in AS1 to regulate the complexity of the set of instrumental functions. The manageability condition could be replaced by some other condition from the literature that is sufficient for a functional central limit theorem to hold for the appropriate quantities.

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) $|m_j(w, \theta, \tau)/\sigma_{F,j}(\theta)| \leq M(w), \forall w \in R^{d_w}, \forall j = 1, \ldots, k,$
$\forall \tau \in \mathcal{T}$, 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 $j = 1, \ldots, p, \forall \tau \in \mathcal{T}$,
  (b) $E_F[m_j(W_i, \theta, \tau)|X_i] = 0$ a.s. $[F_X]$ for $j = p+1, \ldots, k, \forall \tau \in \mathcal{T}$.

3. Tests and confidence sets

In this section, we describe the MCMI test statistics. 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:

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

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

$$\overline{m}_n(\theta, \tau, g) := n^{-1} \sum_{i=1}^n m(W_i, \theta, \tau, g) \text{ for } g \in \mathcal{G}_{\text{c-cube}} \text{ and}$$

$$m(W_i, \theta, \tau, g) := (m_1(W_i, \theta, \tau)g_1(X_i), \ldots, m_k(W_i, \theta, \tau)g_k(X_i))'. \tag{3.2}$$

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 $\widehat{\Sigma}_n(\theta, \tau, g)$:

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

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 $\widehat{\sigma}_{n,j}^{-1}(\theta)m_j(W_i, \theta, \tau)$ and $\sigma_{F,j}^{-1}(\theta)m_j(W_i, \theta, \tau)$ 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 $\zeta > 0$, $\sup_{(\theta, F) \in \mathcal{F}} Pr(\max_{j \leq k} |\widehat{\sigma}_{n,j}^2(\theta)/\sigma_{F,j}^2(\theta) - 1| > \zeta) \to 0$.

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 non-transformed 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

$X_i := \Phi(\widehat{\Sigma}_{X,n}^{-1/2}(X_i^{\dagger} - \overline{X}_n^{\dagger}))$, where $\Phi(x) := (\Phi(x_1), \ldots, \Phi(x_{d_X}))'$

for $x = (x_1, \ldots, x_{d_X})'$,

$\widehat{\Sigma}_{X,n} := n^{-1}\Sigma_{i=1}^n(X_i^{\dagger} - \overline{X}_n^{\dagger})(X_i^{\dagger} - \overline{X}_n^{\dagger})'$, and $\overline{X}_n^{\dagger} := n^{-1}\Sigma_{i=1}^n X_i^{\dagger}$. (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:

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

where $\mathcal{C}_{\text{c-cube}} := \left\{ C_{ar} := \prod_{u=1}^{d_x}((a_u - 1)/(2r), a_u/(2r)] \right.$

$\in [0, 1]^{d_x} : a = (a_1, \ldots, a_{d_x})', a_u \in \{1, 2, \ldots, 2r\}$

for $u = 1, \ldots, d_x$ and $r = r_0, r_0 + 1, \ldots \left. \right\}$ (3.6)

for some positive integer r_0 and $1_k := (1, \ldots, 1)' \in R^k$.[2] 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 MCMI 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

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

where $S = S_1, S_2, S_3$, or S_4 as defined in (3.9), $(r^2 + 100)^{-1}$ 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

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

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 focuses on CvM statistics and all results stated concern CvM statistics. Similar results hold for KS statistics.[3]

The functions $S_1–S_4$ are defined by

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

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

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

$$S_4(m, \Sigma) := \inf_{t=(t_1', 0_v')': t_1 \in R_{+,\infty}^p} (m - t)'(m - t)$$

$$= \sum_{j=1}^p [m_j]_-^2 + \sum_{j=p+1}^{p+v} m_j^2, \quad (3.9)$$

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 $R_{+,\infty}^p := R_{+,\infty} \times \cdots \times R_{+,\infty}$ 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 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

In this section we define critical values based on bootstrap simulations for the MCMI test statistics. The critical values are of the generalized moment selection (GMS) type, and are obtained via the following steps.[4]

Step 1. Compute the GMS function $\overline{\varphi}_n(\theta, \tau, g_{a,r})$ 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

$\xi_n(\theta, \tau, g) := \kappa_n^{-1} n^{1/2} \overline{D}_n^{-1/2}(\theta, \tau, g)\overline{m}_n(\theta, \tau, g)$, where

$\overline{D}_n(\theta, \tau, g) := Diag(\overline{\Sigma}_n(\theta, \tau, g))$, $\kappa_n := (0.3 \ln(n))^{1/2}$, (4.1)

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 $j = 1, \ldots, p$. It is shrunk towards zero via κ_n^{-1} to ensure that one does not over-estimate the slackness.

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

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

$B_n := (0.4 \ln(n)/\ln \ln(n))^{1/2}$, (4.2)

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

---

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

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

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.[5]

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 $\overline{\Sigma}_{n,s}^*(\theta, \tau, g_{a,r})$.[6] When standardizing the instrumental variables for the bootstrap sample, the original sample mean and sample covariance matrix are used for re-centering and rescaling. Using the bootstrap sample mean and covariance matrix for re-centering and rescaling should yield similar results.

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 $\{\overline{T}_{n,r_{1,n},s}^*(\theta) : s = 1, \ldots, B\}$ plus η, where η is an infinitesimal positive number that facilitate the proofs but is inconsequential and can be set to zero in practice.

The MCMI CvM (or KS) 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 MCMI CvM 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 MCMI KS 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, and B_n above are the same as those used in AS1, AS2, and Andrews and Shi (2014). These choices are based on some experimentation (in the simulation results reported in AS1 and AS2). They work well in all seven of the simulation examples in those papers as well as in the two simulation examples in this paper. The asymptotic results reported in the Appendix allow for other choices. The robustness of the finite-sample properties of the tests to the choice of these tuning parameters is documented in the Appendix for the two simulation examples considered in this paper.

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 $N_g := \sum_{r=1}^{r_{1,n}}(2r)^{d_X}$.

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).

5. Correct asymptotic size

In this section, we show that the CS defined above has correct asymptotic size.

5.1. Main result

First, we introduce some additional notation. Define the asymptotic variance–covariance kernel, $\{h_{2,F}(\theta, \tau, g, \tau^\dagger, g^\dagger) : (\tau, g), (\tau^\dagger, g^\dagger) \in \mathcal{T} \times \mathcal{G}_{c\text{-cube}}\}$, 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

$$h_{2,F}(\theta, \tau, g, \tau^\dagger, g^\dagger) := D_F^{-1/2}(\theta)\Sigma_F(\theta, \tau, g, \tau^\dagger, g^\dagger)D_F^{-1/2}(\theta),$$
where
$$\Sigma_F(\theta, \tau, g, \tau^\dagger, g^\dagger) := Cov_F(m(W_i, \theta, \tau, g), m(W_i, \theta, \tau^\dagger, g^\dagger)), \quad (5.1)$$
$$D_F(\theta) := Diag(\sigma_{F,1}^2(\theta), \ldots, \sigma_{F,k}^2(\theta)),$$

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

Define the set of variance–covariance kernels

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

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 $(\mathcal{T} \times \mathcal{G}_{c\text{-cube}})^2$, which is a superset of $\mathcal{H}_2$, we use the uniform metric d defined by

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

Correct asymptotic size is established in the following theorem. The theorem is implied by Lemmas D.1 and D.2 in online Appendix D, where the lemmas are also proved. We provide a brief sketch of the proof in the next subsection, highlighting the difference with the analogous result in AS1. The role of η is also explained in the next subsection.

Theorem 5.1. *Suppose Assumption* SIG1 *holds. For any compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$, *the MCMI confidence set* CS_n *satisfies*

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

Comments. 1. Theorem 5.1 shows that the MCMI CS has 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.

5.2. Sketch of the proof of Theorem 5.1 and the role of η

A sketch of the proof of Theorem 5.1. The theorem is proved using several steps. While the steps are the same as those used to prove the analogous result (Theorem 2(a)) in AS1, notational modifications, and occasionally more substantial modifications to the arguments that complete each step are needed.

First, we use the compactness of $\mathcal{H}_{2,cpt}$ and the definitions of infimum and lim inf to write $\liminf_{n \to \infty} \inf_{\substack{(\theta,F) \in \mathcal{F}: \\ h_{2,F}(\theta) \in \mathcal{H}_{2,cpt}}} P_F(\theta \in CS_n)$

as $\lim_{n \to \infty} P_{F_{a_n}}(\overline{T}_{a_n,r_{1,a_n}}(\theta_{a_n}) \leq c_{a_n,1-\alpha}^{GMS,*}(\theta_{a_n}))$, where $\{a_n\}_{n \geq 1}$ is a

---

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

[6] 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.

subsequence of $\{n\}$, and $\{(\theta_{a_n}, F_{a_n})\}$ is a sequence in $\mathcal{F}$ such that $h_{2,F_{a_n}}(\theta_{a_n}) \to h_2$ for some $h_2 \in \mathcal{H}_{2,cpt}$. This step is the same as the analogous step in the proof of Theorem 2(a) of AS1.

Next, we show an asymptotic distributional approximation for $\overline{T}_{a_n,r_{1,a_n}}(\theta_{a_n})$:

$$\liminf_{n \to \infty} \left[Pr_{F_{a_n}}(\overline{T}_{a_n,r_{1,a_n}}(\theta_{a_n}) \leq x + c) - Pr(\overline{T}_{a_n,F_{a_n}}(\theta_{a_n}) \leq x) \right]$$
$$\geq 0, \tag{5.4}$$

for any $x \in R$ and $c > 0$, where

$$\overline{T}_{n,F_n}(\theta_n)$$
$$= \sup_{\tau \in \mathcal{T}} \sum_{r=1}^{r_{1,n}} (r^2 + 100)^{-1} \sum_{a \in \{1,\dots,2r\}^{d_X}} (2r)^{-d_X} \times S(\nu_{h_{2,F_n}(\theta_n)}(\tau, g_{a,r})$$
$$+ h_{1,n,F_n}(\theta_n, \tau, g_{a,r}), h_{2,F_n}^{\varepsilon}(\theta_n, \tau, g_{a,r})), \tag{5.5}$$

$\nu_{h_{2,F_n}(\theta_n)}$ is a Gaussian process indexed by $(\tau, g_{a,r})$ with variance–covariance kernel $h_{2,F_n}(\theta_n, \tau_1, g_1, \tau_2, g_2)$, $h_{1,n,F_n}(\theta_n, \tau, g) = n^{1/2} E_{F_n} m(W_i, \theta_n, \tau, g)$, and $h_{2,F_n}^{\varepsilon}(\theta_n, \tau, g) = h_{2,F_n}(\theta_n, \tau, g, \tau, g) + \varepsilon I_k$. The approximation (5.4) is proved using the weak convergence of the empirical process $\{a_n^{1/2}[\overline{m}_{a_n}(\theta_{a_n}, \tau, g) - E_{F_{a_n}} m(W_i, \theta_{a_n}, \tau, g)] : (\tau, g) \in \mathcal{T} \times \mathcal{G}\}$ to the Gaussian process, the in-probability convergence of $\{\overline{\Sigma}_{a_n}(\theta_{a_n}, \tau, g) : (\tau, g) \in \mathcal{T} \times \mathcal{G}\}$, and the continuity of the S function. This step is similar to the analogous step in the proof of Theorem 2(a) in AS1 (which is composed of the proofs of Theorem 1 and Lemma A1 in AS2), but substantively differs from the latter in two places:

- Lemma A1 of AS1 establishes the weak convergence for the empirical process involved in AS1, which is indexed by g only. In the present paper, the empirical process is indexed by both τ and g. To account for the double index, we present and prove a lemma (Lemma D.5) that takes advantage of a stability formula for covering numbers.

- The proof of Theorem 2(a) of AS1 employs a dominated convergence argument that is suitable for the pure CvM-type test statistic considered in AS1. On the other hand, we consider a KS-CvM hybrid statistic that takes a supremum over τ and integrates over g, for which the dominated convergence argument does not apply. Instead, we rewrite Assumption S2 of AS1, which is the continuity assumption on S, in an equivalent but more convenient form, and use that to establish sup-norm convergence. Detailed arguments are given at the end of the proof of Theorem D.3 in the Appendix.

The result (5.4) implies immediately that

$$\liminf_{n \to \infty} Pr_{F_{a_n}}(\overline{T}_{a_n,r_{1,a_n}}(\theta_{a_n}) \leq \overline{c}_{a_n,1-\alpha}(\theta_{a_n}) + c) \geq 1 - \alpha, \tag{5.6}$$

for any $c > 0$, where $\overline{c}_{n,1-\alpha}(\theta_n)$ is the $1 - \alpha$ quantile of $\overline{T}_{n,F_n}(\theta_n)$. The rest of the proof shows that the bootstrap critical value satisfies

$$\limsup_{n \to \infty} Pr_{F_{a_n}}(c_{a_n,1-\alpha+c}^{GMS,0}(\theta_{a_n}) \leq \overline{c}_{a_n,1-\alpha}(\theta_{a_n}) - c_1) = 0, \tag{5.7}$$

for any positive constants c, c_1, where $c_{a_n,1-\alpha+c}^{GMS,0}(\theta_{a_n})$ is defined as $c_{a_n,1-\alpha+c}^{GMS,*}(\theta_{a_n})$ is defined except with $\eta = 0$. This step is similar to the analogous step in the proof of Theorem 2(a) of AS1. However, AS1 gives explicit arguments only for the asymptotic approximation critical value and not for the bootstrap critical value. In the present paper, we prove bootstrap validity explicitly. The arguments for the bootstrap are given in Lemma D.4 in online Appendix D. ∎

Next we explain the role of η in the above proof. First, note that the infinitesimal number η is added to two places in the critical value: to the conditional quantile and to the confidence level. The η added to the conditional quantile is needed due to the c in (5.4) and (5.6). To make these two equations hold with $c = 0$, one would need to establish a uniform (over n) anti-concentration bound for the distribution of $\overline{T}_{a_n,F_{a_n}}(\theta_{a_n})$. While such a bound has been derived in Chernozhukov et al. (Unpublished Manuscript,U, 2014b) for the supremum of a Gaussian process, it is to our knowledge not available for $\overline{T}_{a_n,F_{a_n}}(\theta_{a_n})$, which is not a supremum of a Gaussian process even for the KS test statistic.

The η added to the confidence level is due to the c in (5.7). There are two ways to eliminate this η. One is by imposing a uniform (over n) lower bound on the slope of the distribution function of $\overline{T}_{a_n,F_{a_n}}(\theta_{a_n})$ around its $1 - \alpha$ quantile. This would make (5.7) hold with $c = 0$. However, such a bound is difficult to verify. Another way is to strengthen (5.7) so that c is replaced by $c_n \to 0$, as done in Chernozhukov et al. (Unpublished Manuscript). This would require either a Berry–Esseen type distributional convergence rate result for the empirical process $\{a_n^{1/2}[\overline{m}_{a_n}(\theta_{a_n}, \tau, g) - E_{F_{a_n}} m(W_i, \theta_{a_n}, \tau, g)] : (\tau, g) \in \mathcal{T} \times \mathcal{G}\}$ or such a result for the KS or CvM test statistic. Neither is available to our knowledge.

6. Power against fixed alternatives

We now show that the power of the MCMI test converges to one as $n \to \infty$ for all fixed alternatives (for which Assumptions PS1 and PS2 hold). This implies that for any fixed distribution F_0 and any parameter value θ_* *not* in the identified set Θ_{F_0}, the MCMI CS excludes θ_* with probability approaching one. In this sense, the MCMI CS based on $T_n(\theta)$ fully exploits 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.[7]

The null hypothesis is

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

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

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

The next assumption, Assumption MFA, states that violations of the conditional moment inequalities or equalities occur for the null

---

[7] This holds because the identified set based on a finite number of moment inequalities typically is larger than the identified 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.

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

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 the MCMI test is consistent against all fixed alternatives that satisfy Assumption MFA.

Theorem 6.1. *Suppose Assumptions* MFA *and* SIG2 *hold. Then, the* MCMI *test satisfies*

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

Theorem 6.1 is implied by Theorem E.1 in online Appendix E, where the latter is proved. The proof is composed of two parts. First, we show that $n^{-1} T_n(\theta_*)$ converges in probability to a positive quantity, and second, we show that the critical value is $O_p(1)$. The first part combines the proofs of the fixed alternative results for the KS and CvM cases in AS1 and AS2. The second part is the same as the analogous part in AS2 up to notational changes.

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

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

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 $\{(Y_{1,i}, Y_{2,i}, X_i')' : i \geq 1\}$ are i.i.d. under F. Then, $\mathcal{F}_+$ satisfies Assumptions* PS1 *and* PS2 *with $M(w) = 1$, $\delta > 0$, and $C_1 = 1$.*

The proof of the lemma is given in online Appendix F. The core part of the proof is the verification of Assumption PS2, which is done via the pseudo-dimension bound on the covering numbers of the set $\{1\{y_2 \leq \tau\} - 1\{y_1 \leq \tau\} : \tau \in \mathcal{T}\}$ and the fact that the pseudo-dimension of the set is at most one (by Lemma 4.4 of Pollard (1990)).

7.1.2. Higher-order stochastic dominance

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

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

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

For this example, we use $\sigma^2_{F,1}(0) = E_F[(Y_1 - E(Y_1))^2] + E_F[(Y_2 - E(Y_2))^2]$ and $\widehat{\sigma}^2_{n,1}(0) = n^{-1} \sum_{i=1}^{n}[(Y_{1,i} - \overline{Y}_{1,n})^2 + (Y_{2,i} - \overline{Y}_{2,n})^2]$, 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) *$\{(Y_{1,i}, Y_{2,i}, X_i')' : i \geq 1\}$ are i.i.d. under F,* (iii) *$\sigma^2_{F,1}(0) \geq \underline{\sigma}^2$, 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.*

The verification of Assumption PS2 in this case also uses the pseudo-dimension bound on the covering numbers. Unlike in Lemma 7.1, the pseudo-dimension of the set of standardized moment functions is not obvious. We prove that the pseudo-dimension is at most one.

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. That is, we test the null hypothesis:

$$E_{F_0}[1\{Y_2 \leq \tau\} - 1\{Y_1 \leq \tau\}|X] \geq 0 \text{ a.s. } \forall \tau \in \mathcal{T} \equiv R, \tag{7.3}$$

where Y_1, Y_2, X are scalar random variables. We consider the MCMI tests proposed above based on the CvM and KS test statistics combined with the GMS critical value. For comparative purposes, we also consider the CvM and KS test statistics combined with sub-sampling critical values, as well as the two-step multiplier bootstrap method (CCK-MB) and the two-step empirical bootstrap method (CCK-EB) proposed in Chernozhukov et al. (Unpublished Manuscript), CCK hereafter).[8]

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:

$$Y_1 = \exp(\sigma_1(X)Z_1 + \mu_1(X)) \text{ and}$$
$$Y_2 = \exp(\sigma_2(X)Z_2 + \mu_2(X)), \tag{7.4}$$

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 $\mu_1(X) = c_1 X + c_3$, $\sigma_1(X) = c_2 X + c_4$, $\mu_2(X) = 0.85$, 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 dominates Y_2 conditional on X holds, while with $\mathbf{c}_C$ and $\mathbf{c}_D$, the null hypothesis is

---

[8] These methods have the best power among the six one-step and two-step methods proposed in CCK. The three-step methods proposed in CCK are not applicable in this model.
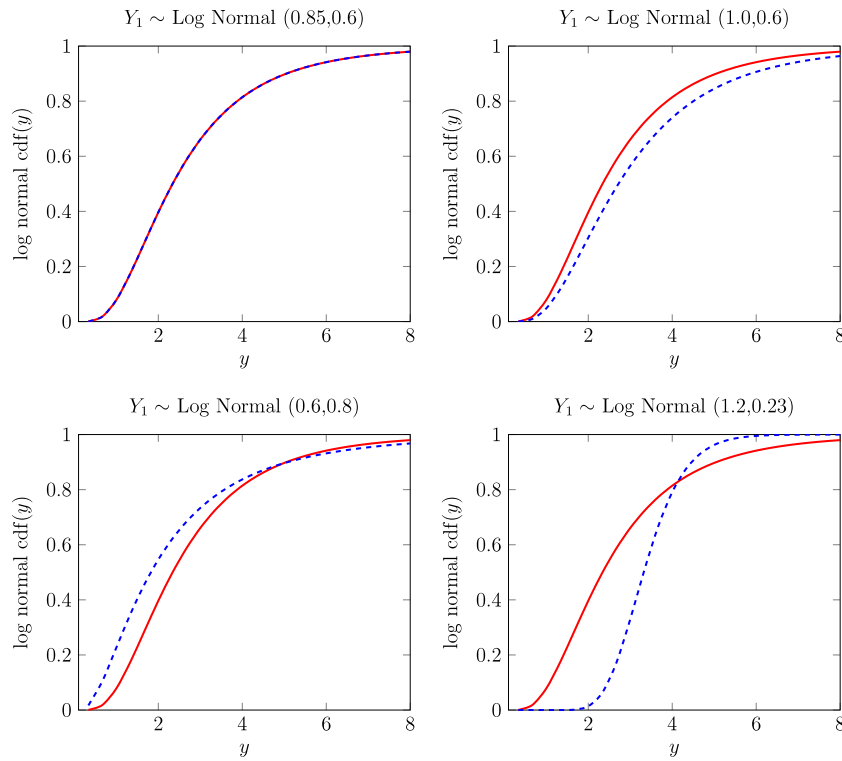
Fig. 1. Conditional CDF's of Y_1 (dashed blue) and Y_2 (solid red) given $X = 1$. In all graphs, $Y_2 \sim$ Log Normal (0.85, 0.6).

violated. To visualize the nature of the DGPs, we draw in Fig. 1 the conditional cdf's of Y_1 and Y_2 given $X = 1$ at these four **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 all tests considered, we consider a base case with the sample size $n = 250$, the hypercube parameter $r_{1,n} = 3$, and $N_\tau = 25$. Then, for comparison, we also consider three variations of the base case where each differs from the base case in only one dimension.[9] We set η to zero in all cases for our methods. For the sub-sampling critical values, we use a subsample size of 20. For the CCK methods, we take the tuning parameters from CCK's Monte Carlo simulations.

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 the highest power. The KS/GMS test has somewhat worse power perhaps due to the DGP designs. The CvM/Sub-sampling test has greater over-rejections than, and comparable power to, the CvM/GMS test,

while the KS/Sub-sampling test exhibit severe over-rejections. The CCK tests have good size control, but somewhat lower power than the KS/GMS test and significantly lower (size-corrected) power than the CvM/GMS test.

8. Example 2: Random-coefficients binary-outcome models with instrumental variables

We focus on the model given in (2.9), and restated here for the reader's convenience:

$$E_{F_0}[F_\beta(\mathcal{S}, \theta) - 1\{S(Y_1, Y_2, X_1) \subset \mathcal{S}\}|X_1, X_2] \geq 0 \text{ a.s. } \forall \mathcal{S} \in \mathbf{S}, \quad (8.1)$$

where Y_1 is the binary dependent variable, Y_2 is a d_2-dimensional endogenous covariate, X_1 is a d_1-dimensional exogenous covariate, and X_2 is a vector of instruments.

8.1. Verification of assumptions

We first note that, when $d_1 + d_2 > 1$, the manageability assumption, 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 $n \to \infty$. 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: $\mathbf{S}_m := \{\cup_{j=1}^m H(c_j) : c_j \in R^{d_\beta} \setminus \{0^{d_\beta}\}\}$. 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 $m = 2$.

---

[9] More variations are considered in the additional Monte Carlo exercise in online Appendix G.

Table 1
Null rejection probabilities for nominal 0.05 first-order stochastic dominance tests.

	CvM/GMS	KS/GMS	CvM/Sub	KS/Sub	CCK-MB	CCK-EB
Null 1: $(c_1, c_2, c_3, c_4) = (0, 0, 0.85, 0.6)$						
Base case:						
$(n = 250, r_{1,n} = 3, N_\tau = 25)$	0.057	0.064	0.071	0.213	0.035	0.018
$n = 500$	0.049	0.052	0.079	0.212	0.032	0.029
$r_{1,n} = 4$	0.059	0.055	0.098	0.282	0.027	0.010
$N_\tau = 30$	0.062	0.068	0.085	0.239	0.034	0.014
Null 2: $(c_1, c_2, c_3, c_4) = (0.15, 0, 0.85, 0.6)$						
Base case:						
$(n = 250, r_{1,n} = 3, N_\tau = 25)$	0.014	0.019	0.029	0.131	0.011	0.006
$n = 500$	0.009	0.014	0.017	0.089	0.013	0.010
$r_{1,n} = 4$	0.014	0.019	0.039	0.192	0.007	0.006
$N_\tau = 30$	0.018	0.019	0.037	0.137	0.011	0.007

Note: For computation reasons, not all subsamples are used. The bootstrap and sub-sampling critical values both use 1000 repetitions to simulate the critical values. The two-step version of CCK's MB and EB methods are used.

Table 2
Size-corrected power for nominal 0.05 first-order stochastic dominance tests.

	CvM/GMS	KS/GMS	CvM/Sub	KS/Sub	CCK-MB	CCK-EB
Alternative 1: $(c_1, c_2, c_3, c_4) = (-0.25, 0.2, 0.85, 0.6)$						
Base case:						
$(n = 250, r_{1,n} = 3, N_\tau = 25)$	0.505	0.379	0.463	0.281	0.301	0.210
$n = 500$	0.809	0.689	0.806	0.603	0.596	0.525
$r_{1,n} = 4$	0.509	0.367	0.475	0.272	0.254	0.148
$N_\tau = 30$	0.470	0.405	0.443	0.297	0.309	0.202
Alternative 2: $(c_1, c_2, c_3, c_4) = (0.35, 0, 0.85, 0.23)$						
Base case:						
$(n = 250, r_{1,n} = 3, N_\tau = 25)$	0.581	0.295	0.622	0.346	0.204	0.178
$n = 500$	0.942	0.768	0.946	0.767	0.670	0.665
$r_{1,n} = 4$	0.609	0.246	0.643	0.335	0.168	0.131
$N_\tau = 30$	0.539	0.309	0.598	0.350	0.208	0.172

Note: The bootstrap and sub-sampling critical values use 1000 repetitions to simulate the critical values. Size correction is carried out using the null DGP with $(c_1, c_2, c_3, c_4) = (0, 0, 0.85, 0.6)$.

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

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

$$E_F[F_\beta(\delta, \theta) - 1\{S(Y_1, Y_2, X_1) \subset \delta\}|X_1, X_2] \geq 0 \text{ a.s. } \forall \delta \in S_m. \quad (8.2)$$

We first write $S(y_1, y_2, x_1)$ in the canonical form of a half-space:

$$S(y_1, y_2, x_1)$$
$$= cl\{b = (b_0, b_1', b_2')' \in R^{d_\beta} : y_1 = 1\{b_0 + b_1'x_1 + b_2'y_2 \geq 0\}\}$$
$$= H((y_1 - 1/2)(1, x_1', y_2')'). \quad (8.3)$$

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

Lemma 8.1. For any $c_1, \ldots, c_m \in R^{d_\beta} \setminus \{0^{d_\beta}\}$ (not necessarily distinct from each other), there exists a $d_\beta \times M$ real matrix $B(c_1, \ldots, c_m)$ with $M = \max_{j \in \{1, \ldots, d_\beta\}} \left[\binom{m}{\min\{j,m\}-1} + 2(d_\beta - j) \right]$ such that, for any $\bar{c} \in R^{d_\beta} \setminus \{0^{d_\beta}\}$, 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 $\lambda_1, \ldots, \lambda_m \geq 0$, and
  (c) $B(c_1, \ldots, c_m)'\bar{c} \geq 0^M$.

The lemma implies that the conditional moment inequality model (8.2) has the following equivalent representation:

$$E_F[F_\beta(\delta(\tau), \theta) - 1\{(Y_1 - 1/2)B(\tau)'(1, X_1', Y_2')' \geq 0\}|X_1, X_2]$$
$$\geq 0 \text{ a.s. } \forall \tau \in \mathcal{T}, \quad (8.4)$$

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

The equivalent representation just given is instrumental in proving the lemma below, which verifies the high-level conditions for this example. Note that in this example, $p = k = 1$. We use $\sigma_{F,1}(\theta) = \hat{\sigma}_{n,1}(\theta) = 1$ for all (θ, F) because the moment function has a natural scale. Hence, Assumptions SIG1 and SIG2 hold.

Lemma 8.2. For the model in (8.2), 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 $C_1 = 1$.

The main part of the proof of Lemma 8.2 is the verification of Assumption PS2. To verify this assumption, we use a pseudo-dimension bound for covering numbers (specifically, Lemma 4.1 of Pollard (1990)). We show that the pseudo-dimension is finite by applying Lemma 4.4 of Pollard (1990) to the equivalent representation in (8.4).[10]

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),

---

[10] Note that the representation (8.4) 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 $c_1, \ldots, c_m$. Algebraic representations of facets of convex polyhedral cones are complicated.

and no exogenous regressors. That is,

$$Y_1 = 1\{\beta_0 + \beta_1 Y_2 < 0\} \text{ with } (\beta_0, \beta_1) \perp X. \tag{8.5}$$

Further, we take β_0 and β_1 to be jointly normally distributed: $\beta_0 = \alpha_0 + U_0$ and $\beta_1 = \alpha_1 + U_1$, where

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \gamma_0 \\ \gamma_0 & \gamma_1 + \gamma_0^2 \end{pmatrix} \right).$$

Thus, the model contains the unknown parameter $\theta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)'$. CR show that the sharp identified set for θ is characterized by the following conditional moment inequalities:

$$E_{F_0}[F_\beta(\mathscr{s}, \theta) - 1\{S(Y_1, Y_2) \subset \mathscr{s}\}|X] \geq 0 \text{ a.s. } \forall \mathscr{s} \in \mathbf{S}, \tag{8.6}$$

where, using the half-space notation $H(\cdot)$ defined in (2.8),

$$S(y_1, y_2) = H((y_1 - 1/2)(1, y_2)'),$$
$$\mathbf{S} = \{\mathscr{s}_{\tau_1, \tau_2} = H(\cos \tau_1, \sin \tau_1) \cup H(\cos \tau_2, \sin \tau_2) :$$
$$0 \leq \tau_1 \leq \tau_2 \leq 2\pi\},$$
$$F_\beta(\mathscr{s}_{\tau_1, \tau_2}, \theta) = 1 - \Phi(-m_1, -m_2, \rho), \tag{8.7}$$

$\Phi(x_1, x_2, \rho)$ is cdf of the bivariate normal $N(0, [1, \rho; \rho, 1])$, m_j is the mean divided by the standard deviation of $\beta_0 \cos \tau_j + \beta_1 \sin \tau_j$, for $j = 1, 2$, and ρ is the correlation coefficient between $\beta_0 \cos \tau_1 + \beta_1 \sin \tau_1$ and $\beta_0 \cos \tau_2 + \beta_1 \sin \tau_2$.

To generate the data, we let

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

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). Let $\theta = (0, -1, -1, 1)'$, and $\delta := (\delta_1, \delta_2, \delta_3, \delta_4)' = (1, 0.577, -0.577, 0.577)'.$[11]

We compute the probabilities that the CS's for θ cover given values of θ. For the given values of θ, we consider $\theta = (0, \alpha_1, -1, 1)'$, where α_1 runs from -1 to 1.4. Note that $(0, -1, 1)$ is the true value of $(\alpha_0, \gamma_0, \gamma_1)$, and -1 is the true value of α_1. Thus, $(0, -1, -1, 1)$ is in the identified set, and should ideally be covered by the CS's with at least the nominal coverage probability. Numerical calculation of the boundary of the identified set shows that $(0, \alpha_1, -1, 1)$ is outside the identified set for any $\alpha_1 > -0.8274$ and, hence, it is desirable that CS's cover such α_1 values with low probabilities.[12]

We consider CS's based on the CvM and KS statistics and the GMS and sub-sampling critical values. For comparative purposes, we also consider the two-step CCK-EB and two-step CCK-SN (self-normalizing) based CS's.[13] For all CS's, we choose $r_{1,n} = 3$ and

approximate $\mathcal{T}$ by grid points.[14] For the GMS CS's, we set η to zero. For the sub-sampling CS's, we set the subsample size to 20. For the CCK-EB CS, we take the tuning parameter values from CCK's Monte Carlo simulations. We use 1001 repetitions to simulate the bootstrap critical values and we use 1001 subsamples to construct the sub-sampling critical values. We employ 1000 Monte Carlo repetitions to obtain the simulated coverage probabilities of given points of θ.

Fig. 2 provides coverage probability graphs for sample sizes $n = 250, 500, 1000$, and 2000. As the figure shows, the coverage probabilities of the CS's equal one at the boundary ($\alpha_1 = -0.8274$) of the identified set for all of the CS's except the CCK-EB CS for the case of $n = 250$. This is probably due to the fact that the boundary of the identified set is determined by X values in a set with Lebesgue measure zero. For $n = 250$, the coverage probability of the CCK-EB CS is closer to the nominal size 0.95 at the boundary of the identified set than the other CS's, but its coverage probabilities decrease more slowly than those of the other CS's as α_1 deviates from the identified set (i.e., as α_1 increases beyond -0.8274).

The coverage probabilities of all of the CS's for points outside the identified set decrease with the sample size (with the exception of the CCK CS's for points close to the identified set) and with the magnitude of the deviation from the identified set, as expected. The best performing CS's (lowest curves) are the KS/GMS and KS/sub-sampling CS's at $n = 500$, where the coverage probability curves of the KS/GMS and KS/Sub-sampling CS's overlap completely and form the lowest curve in the graph. At $n = 1000$, the KS/GMS and the CCK-EB curves overlap and form the lowest curve in the graph. At $n = 2000$, the CCK-EB performs better than the other CS's. For each of the four sample sizes considered, the CvM-based CS's do not perform as well as the other CS's.

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 et al. (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* $M(w)$ $\forall(w, v) \in \mathcal{WV}$, (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.*

---

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

[12] Specifically, the way we compute the boundary is as follows. First we construct the criterion function $Q(\theta) = \min_{x \in \mathcal{X}_{N_x}} \min_{\tau \in \mathcal{T}_{N_\tau}} F_\beta(\mathscr{s}(\tau), \theta) - E[1\{S(Y_1, Y_2) \subseteq \mathscr{s}_{\tau_1, \tau_2}\}|X = x]$, 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 footnote 14, $F_\beta(\mathscr{s}_\tau, \theta)$ is computed using the bivariate-normal cdf function in Aptech Gauss, and $E[1\{S(Y_1, Y_2) \subseteq \mathscr{s}_{\tau_1, \tau_2}\}|X = x]$ 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 $\alpha_1 > -1$ that makes $Q(\alpha_0, \alpha_1, \gamma_0, \gamma_1)$ 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.

[13] The two-step CCK methods perform better than the one-step CCK methods in this example. The performance of the CCK-MB method lies in between that of the CCK-EB and the CCK-SN. We do not consider the three-step CCK methods for two reasons. First, those methods require the derivative of $F_\beta(\mathscr{s}, \theta)$ with respect to θ, the analytical form of which is complicated because $F_\beta(\mathscr{s}, \theta)$ is a quadrant probability of a bivariate normal with both the mean and the variance–covariance matrix dependent on θ. Second, the potential gain of using the three-step CCK methods is likely small because, for every $\mathscr{s}$, we expect $F_\beta(\mathscr{s}, \theta)$ to depend strongly on the mean and the variance of the bivariate normal, and hence on θ. CCK also do not provide any simulation results for their three-step methods.

[14] We consider N_{τ_2} equally-spaced grid points for τ_2 in $[0, 2\pi]$, and grid points for τ_1 in $[0, \tau_2]$ with the same spacing. We set $N_{\tau_2} = 15$ for our CS's, which results in 120 points in $\{(\tau_1, \tau_2) \in [0, 2\pi] : \tau_1 \leq \tau_2\}$. We set $N_{\tau_2} = 14$ for the CCK-EB CS, because when $N_{\tau_2} = 15$, some of the moments have very small variance, which causes the CCK-EB CS to have a zero coverage probability for the true value.
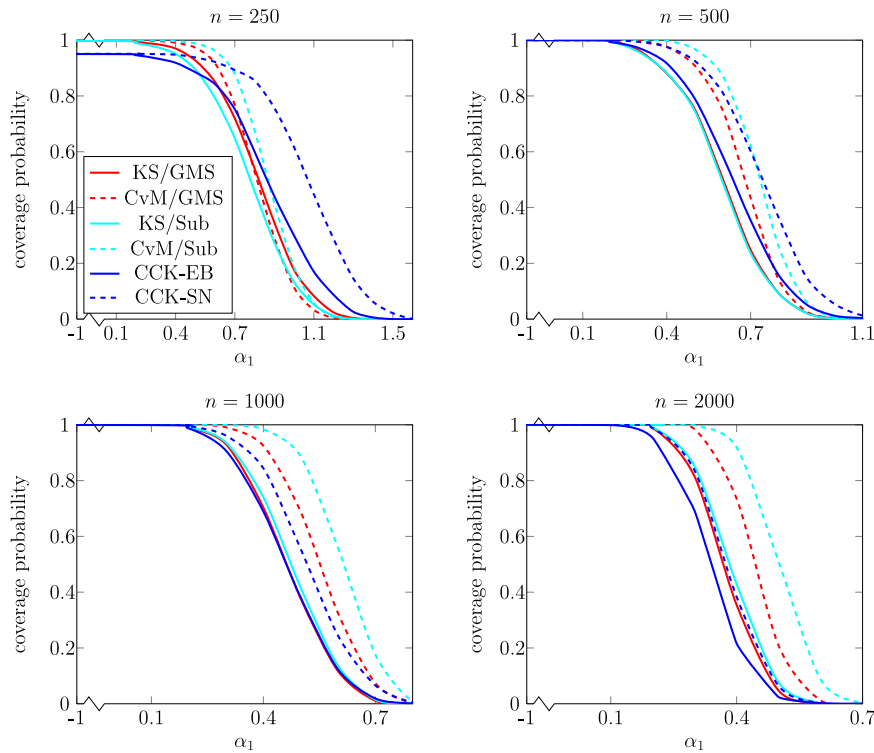
Fig. 2. Coverage probabilities in the IV random-coefficients binary-outcome model. (Nominal size =0.95, $(\alpha_0, \gamma_0, \gamma_1) = (0, -1, 1)$, and $(\alpha_0, \alpha_1, \gamma_0, \gamma_1)$ is in the identified set if and only if $\alpha_1 \leq -0.8274$.)

The verification of Assumption PS2 in this case relies on a direct calculation of the covering numbers using the Lipschitz continuity of the moment function with respect to the index u.

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:

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

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

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

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

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

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

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

$\underline{\sigma}^2$ *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.*

The verification of Assumption PS2 in this case relies on a direct calculation of the covering numbers. The covering numbers are properly bounded due to the decreasing weight $w_{\mathcal{T}}(\tau)$.

We note that the weighting scheme requires an ordering of the moment conditions. A natural ordering of the moment conditions is often available. For example, in Example 4, suppose that M (the number of values that the dependent variable Y can take) is small but the number of values that Z can take is large, one natural order of the moment conditions is according to the empirical probability $Z = z_s$, while treating moment conditions with the same s but different ℓ, m (indices for the value of Y) as ties in the ordering. In Example 5, one can order the actions according to how close they are to the optimal (observed) action. A similar ordering may be used for the dynamic model of imperfect competition in Example 3 of CCK.

When there are no ties in the ordering, an example of the weight is $w_{\mathcal{T}}(\tau) = \tau^{-b}$ for some $b > 0$. Then $\lambda_{\mathcal{T}}(\xi) = \xi^{-1/b}$ and

$$\int_0^1 \sqrt{\log(\xi^{-1/b})}d\xi = \sqrt{1/b} \int_0^1 \sqrt{\log(\xi^{-1})}d\xi$$

$$= b^{-1/2} \int_0^\infty 2x^2 e^{-x^2} dx < \infty, \quad (9.4)$$

where the last equality holds by change of variables with $x = \sqrt{\log(\xi^{-1})}$ (or, equivalently, $\xi = e^{-x^2}$). When there are ties, one can consider the tied moment conditions as one, assign decreasing weights as just described, and give equal weights to the tied observations.

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.

Acknowledgments

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2016.09.010.

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. J. Econometrics 122, 1–26.

Andrews, D.W.K., 1988. Chi-square diagnostic tests for econometric models: theory. Econometrica 56, 1419–1453.

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. J. Econometrics 179, 31–45. Appendix of proofs available as Cowles Foundation Discussion Paper No. 1840RR, Yale University, 2011.

Andrews, D.W.K., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78, 119–157.

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. J. Econometrics 181, 92–116.

Armstrong, T.B., 2015. Asymptotically exact inference in conditional moment inequality models. J. 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. J. Bus. Econom. Statist. 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., 2013. Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors. Department of Economics, MIT. Unpublished Manuscript.

Chernozhukov, V., Chetverikov, D., Kato, K., 2014a. Testing Many Moment Inequalities. Department of Economics, MIT. Unpublished Manuscript.

Chernozhukov, V., Chetverikov, D., Kato, K., 2014b. Anti-Concentration and Honest Adaptive Confidence Bands. Ann. Statist. 42, 1787–1818.

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. Econom. J. 17, S1–S19.

Chesher, A., Smoliński, K., 2012. IV models of ordered choice. J. 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. Econom. J. 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. J. Bus. Econom. Statist. 31, 16–28.

Galichon, A., Henry, M., 2009. A test of non-identifying restrictions and confidence regions for partially identified parameters. J. 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. J. Econometrics 152, 104–119.

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

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

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

Pollard, D., 1979. General chi-square goodness-of-fit tests with data-dependent cells. Z. Wahrscheinlichkeitstheor. 50, 317–331.

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