NONPARAMETRIC INFERENCE BASED ON CONDITIONAL MOMENT INEQUALITIES

By

Donald W. K. Andrews and Xiaoxia Shi

December 2011
Revised February 2013

COWLES FOUNDATION DISCUSSION PAPER NO. 1840R

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

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

Nonparametric Inference Based on Conditional Moment Inequalities

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

Xiaoxia Shi
Department of Economics
University of Wisconsin at Madison

June 2010

Revised: February 2013

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-0751517 and SES-1058376.

Abstract

This paper develops methods of inference for nonparametric and semiparametric parameters defined by conditional moment inequalities and/or equalities. The parameters need not be identified. Confidence sets and tests are introduced. The correct uniform asymptotic size of these procedures is established. The false coverage probabilities and power of the CS's and tests are established for fixed alternatives and some local alternatives. Finite-sample simulation results are given for a nonparametric conditional quantile model with censoring and a nonparametric conditional treatment effect model. The recommended CS/test uses a Cramér-von-Mises-type test statistic and employs a generalized moment selection critical value.

Keywords: Asymptotic size, kernel, local power, moment inequalities, nonparametric inference, partial identification.

JEL Classification Numbers: C12, C15.

1 Introduction

This paper considers inference for nonparametric and semiparametric parameters defined by conditional moment inequalities and/or equalities. The moments are conditional on X_i a.s. and $Z_i = z_0$ for some random vectors X_i and Z_i. The parameters need not be identified. Due to the conditioning on Z_i at a single point z_0, the parameter considered is a nonparametric or semiparametric parameter (which varies with z_0). Due to the conditioning on X_i a.s., the moment conditions are typical conditional moments which involve an infinite number of restrictions.

Examples covered by the results of this paper include: a nonparametric conditional distribution with selection, a nonparametric conditional quantile with selection, an interval-outcome partially-linear regression, an interval-outcome nonparametric regression, a semiparametric discrete-choice model with multiple equilibria, a nonparametric revealed preference model, tests of a variety of functional inequalities, including nonparametric average treatment effects for certain sub-populations, and nonparametric binary Roy models, as in Henry and Mourifié (2012).

As far as we are aware, the only other paper in the literature that covers the examples described above is Chernozhukov, Lee, and Rosen (2013) (CLR). In this paper, we employ statistics that are akin to Bierens (1982)-type model specification test statistics. In contrast, CLR employ statistics that are akin to Härdle and Mammen (1993)-type model specification statistics, which are based on nonparametric regression estimators. These approaches have different strengths and weaknesses. Specifically, the tests proposed in this paper have higher power against moment functions that are flatter (but not necessarily completely flat) as a function of x, whereas the CLR tests have higher power against moment functions that are more curved. This is shown by the finite-sample simulations reported here and the asymptotic local power results reported in the Appendix, see Andrews and Shi (2013a).

We provide confidence sets (CS's) and tests concerning the true parameter. The class of test statistics used in this paper are like those used in Andrews and Guggenberger (2009), which are extended in Andrews and Shi (2013b,c) (AS1, AS2) to handle moment conditions that are conditional on X_i a.s. Here the test statistics are extended further to cover moment conditions that are conditional on $Z_i = z_0$ as well. The latter conditioning is accomplished using kernel smoothing. The critical values considered here are generalized moment selection (GMS) and plug-in asymptotic (PA) critical values, as in Andrews and Soares (2010), which are extended to cover conditional moment inequalities, as in AS1 and AS2.

The results of the paper are analogous to those in AS1 and AS2. In particular, we establish the correct uniform asymptotic size of the CS's and tests. We also determine the asymptotic behavior of the CS's and tests under fixed alternatives and some local alternatives.

We provide finite-sample simulation results for two models: a nonparametric conditional quantile model with selection and a nonparametric conditional treatment effect model. The conclusions from the finite-sample results are similar in many respects to those from Andrews and Soares (2010), Andrews and Barwick (2012), AS1, and AS2. Cramér-von-Mises (CvM) versions of the CS's and tests out-perform Kolmogorov-Smirnov (KS) versions in terms of false-coverage probabilities (FCP's) and power and have similar size properties. Likewise, GMS critical values out-perform PA critical values according to the same criteria. The "Gaussian asymptotic" versions of the critical values perform similarly to the bootstrap versions in terms of size, FCP's, and power. The finite-sample sizes of the CvM/GMS CS's and tests are close to their nominal size. The CS's and tests show some sensitivity to the nonparametric smoothing parameter employed, but not much sensitivity to other tuning parameters.

In the simulation results for these two models, the CI's and tests proposed in this paper are found to have more robust size properties than the series and local linear CLR procedures. The CI's and tests proposed in this paper are found to have higher power (and lower FCP's) for flat bound functions and lower power (and higher FCP's) for peaked bound functions compared to the CLR procedures.

We note that the results given here also apply to nonparametric models based on moments that are unconditional on X_i but conditional on $Z_i = z_0$. The results also cover the case where different moment functions depend on different sub-vectors of X_i, e.g., as occurs in some panel models.[1] In addition, the results can be extended to the case of an infinite number of moment functions along the lines of Andrews and Shi (2010).

The technical results in this paper differ from those in AS1 and AS2 because (i) the conditional moment inequalities (when evaluated at the true parameter) do not necessarily hold for values Z_i that are in a neighborhood of z_0, but do not equal z_0, and (ii) the sample moments do not satisfy a functional CLT with $n^{1/2}$-norming due to local

[1] This holds because the functions $g_1(x), ..., g_k(x)$, which multiply the moment functions indexed by $1, ..., k$, need not be the same, see (6.1) of Andrews and Shi (2012).

smoothing, and, hence, need to be normalized using their standard deviations which are $o(1)$ as $n \to \infty$.

Now, we discuss the related literature. The literature on inference based on *unconditional* moment inequalities for parameters that are partially identified is now quite large. For brevity, we do not give references here. See Andrews and Soares (2010) for references. The literature on inference for partially-identified models based on *conditional* moment inequalities includes AS1, AS2, CLR, Fan and Park (2007), Fan (2008), Kim (2008), Ponomareva (2010), Armstrong (2011a,b), Beresteanu, Molchanov, and Molinari (2011), Chetverikov (2011), Hsu (2011), Lee, Song, and Whang (2011), and Aradillas-López, Gandhi, and Quint (2012). Khan and Tamer (2009) considers conditional moment inequalities in a point-identified model. Galichon and Henry (2009) considers a testing problem with an infinite number of unconditional moment inequalities of a particular type. Menzel (2009) investigates tests based on a finite number of moment inequalities in which the number of inequalities increases with the sample size.

Of these papers, the only one that allows for conditioning on $Z_i = z_0$, which is the key feature of the present paper, is CLR. As noted above, the forms of the tests considered here and in CLR differ. Other differences are as follows. The assumptions given here are primitive, whereas those in CLR are high-level. The present paper provides uniform asymptotic size results, whereas CLR does not.

The remainder of the paper is organized as follows. Section 2 describes the nonparametric model and discusses six examples covered by the model. Section 3 introduces the test statistics and critical values, establishes the correct asymptotic size (in uniform sense) of the CS's, and establishes the power of the tests against fixed alternatives. Section 4 provides Monte Carlo simulation results for two models.

An Appendix provides proofs of all of the results stated in the paper. For brevity, the Appendix is given in Andrews and Shi (2013a). The results in the Appendix allow for a much broader range of test statistics than is considered in the paper. Specifically, the results cover a wide variety of kernel functions K, test statistic functions S, instrumental functions $g \in \mathcal{G}$, and weight measures Q. The Appendix provides two sets of results for local alternatives. The first set considers $(nb^{d_{\mathbf{z}}})^{-1/2}$-local alternatives, for which the bound functions are asymptotically flat near their minimum, where b denotes a bandwidth parameter and d_z denotes the dimension of Z_i. The tests proposed in this paper have non-trivial power against such alternatives, whereas the tests of CLR do not. The second set considers a_n-local alternatives, for which the bound functions are asymptotically non-flat near their minimum. Here, $a_n \to 0$ as $n \to \infty$ at a rate slower than $(nb^{d_{\mathbf{z}}})^{-1/2}$. For such alternatives, if the functions are sufficiently curved then the CLR tests have higher asymptotic local power than the tests considered here. On the other hand, if the functions are less curved, then the tests proposed here have higher asymptotic power than the CLR tests. The Appendix also gives some additional simulation results for the two models considered in the paper.

2 Nonparametric Conditional Moment Inequalities and Equalities

2.1 Model

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

$$\begin{aligned} &E_{F_0}(m_j\left(W_i, \theta_0\right)|X_i, Z_i = z_0) \geq 0 \text{ a.s. } [F_{X,0}] \text{ for } j = 1, ..., p \text{ and} \\ &E_{F_0}(m_j\left(W_i, \theta_0\right)|X_i, Z_i = z_0) = 0 \text{ a.s. } [F_{X,0}] \text{ for } j = p+1, ..., p+v, \end{aligned} \tag{2.1}$$

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

The object of interest is a CS for the true parameter θ_0. We do not assume that θ_0 is point identified. However, the model restricts the true parameter value to the *identified set* (which could be a singleton) that is defined as follows:

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

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

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

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

2.2 Examples

In this section, we provide several examples in which the nonparametric conditional moment inequality/equality model arises. Note that Examples 2 and 6 below, for a conditional quantile bound and a conditional treatment effect, respectively, are used in a simulation study in Section 4.

Example 1 (Conditional Distribution with Censoring). The first example is a missing data example. The observations are i.i.d. Let Y_i^* be a variable that is subject to censoring: it is observed only for observations i with $D_i = 1$ and not for observations with $D_i = 0$. Let Z_i be a vector of covariates and X_i be a vector of excluded instruments that are independent of Y_i^* conditional on Z_i. Then, the conditional distribution of Y_i^* given Z_i, denoted $F_{Y^*|Z}$, satisfies: for fixed $y_0 \in R$ and $z_0 \in Supp(Z_i)$,

$$\begin{aligned} E(1\{Y_i^* \leq y_0, D_i = 1\} + 1\{D_i = 0\} - F_{Y_1^*|Z_1}(y_0|z_0)|X_i, Z_i = z_0) \geq 0 \\ E(F_{Y^*|Z}(y_0|z_0) - 1\{Y_i^* \leq y_0, D_i = 1\}|X_i, Z_i = z_0) \geq 0. \end{aligned} \tag{2.4}$$

This model fits into the general model (2.1) with $\theta_0 = F_{Y^*|Z}(y_0|z_0)$, $m_1(W_i, \theta_0) = 1\{Y_i^* \leq y_0, D_i = 1\} + 1\{D_i = 0\} - \theta_0$ and $m_2(W_i, \theta_0) = \theta_0 - 1\{Y_i^* \leq y_0, D_i = 1\}$.

A model similar to this one is used in Blundell, Gosling, Ichimura, and Meghir (2007) to study the distribution of female wages. In their study, Y_i^* is the potential wage of woman i, D_i is the dummy for employment status, Z_i are demographic variables, and X_i is non-wage income. Parametric and nonparametric versions of this example are discussed in CLR. Notice that the parametric version can be estimated using AS1. □

Example 2 (Conditional Quantile with Censoring). In some cases, it is more useful to bound the conditional quantiles of Y_i^*, rather than its conditional distribution. Again, suppose the observations are i.i.d. Let $q_{Y^*|Z}(\tau|z_0)$ denote the τ quantile of Y_i^* given $Z_i = z_0$. Then under the conditional quantile independence assumption: $q_{Y^*|Z,X}(\tau|z_0, x) = q_{Y^*|Z}(\tau|z_0)$ for all $x \in Supp(X)$. The quantile satisfies: for fixed

$\tau \in (0,1)$ and $z_0 \in Supp(Z)$,

$$E(1\{Y_i^* \leq q_{Y^*|Z}(\tau|z_0), D_i = 1\} + 1\{D_i = 0\} - \tau | X_i, Z_i = z_0) \geq 0$$
$$E(\tau - 1\{Y_i^* \leq q_{Y^*|Z}(\tau|z_0), D_i = 1\} | X_i, Z_i = z_0) \geq 0. \quad (2.5)$$

This model fits into the general model (2.1) with $\theta_0 = q_{Y^*|Z}(\tau|z_0)$, $m_1(W_i, \theta_0) = 1\{Y_i^* \leq \theta_0, D_i = 1\} + 1\{D_i = 0\} - \tau$ and $m_2(W_i, \theta_0) = \tau - 1\{Y_i^* \leq \theta_0, D_i = 1\}$.

If the conditional quantile independence assumption is replaced with the quantile monotone instrumental variable (QMIV) assumption in AS1, then Example 2 becomes a nonparametric version of the quantile selection example considered in AS1. □

Example 3 (Interval-Outcome Partially-Linear Regression). This example is a partially-linear interval-outcome regression model. Let Y_i^* be a latent dependent variable and $Y_i^* = X_i'\beta_0 + \psi_0(Z_i) + \varepsilon$, $E(\varepsilon|X_i, Z_i) = 0$ a.s., where (X_i, Z_i) are exogenous regressors some of which may be excluded from the regression. The latent variable Y_i^* is known to lie in the observed interval $[Y_i^l, Y_i^u]$. Then, the following moment inequalities hold for fixed $z_0 \in Supp(Z_1)$:

$$E(Y_i^u - X_i'\beta_0 - \psi_0(z_0) | X_i, Z_i = z_0) \geq 0 \text{ and}$$
$$E(X_i'\beta_0 + \psi_0(z_0) - Y_i^l | X_i, Z_i = z_0) \geq 0 \quad (2.6)$$

This model fits into the general model (2.1) with $\theta_0 = (\beta_0, \psi_0(z_0))$, $W_i = (Y_i^u, Y_i^l, X_i, Z_i)$, $m_1(W_i, \theta_0) = Y_i^u - X_i'\beta_0 - \psi_0(z_0)$, and $m_2(W_i, \theta_0) = X_i'\beta_0 + \psi_0(z_0) - Y_i^l$.

Example 3 is a partially-linear version of the interval-outcome regression model considered in Manski and Tamer (2002) and widely discussed in the moment inequality literature (e.g., see Chernozhukov, Hong and Tamer (2007), Beresteanu and Molinari (2008), Ponomareva and Tamer (2012), and AS2). Allowing some of the regressors to enter the regression function nonparametrically makes the model less prone to misspecification.

If the linear term $X_i'\beta_0$ does not appear in the model, then the model is an interval-outcome nonparametric regression model. The results of this paper apply to this model as well. However, a linear term $X_i'\beta_0$ often is used in practice to reduce the curse of dimensionality (e.g., see Tamer (2008)). □

Example 4 (Semiparametric Discrete Choice Model with Multiple Equilibria). Consider an entry game with two potential entrants, $j = 1, 2$, and possible multiple equilibria. For notational simplicity, we suppress the observation index i for $i = 1, ..., n$. The payoff from not entering the market is normalized to zero for both players. The payoff from entering is assumed to be $\pi_j = \beta_{j0} X + \psi_{j0}(Z) - \delta_{j0} D_{-j} - \varepsilon_j$, where D_{-j} is a dummy that equals one if the other player enters the market, $\delta_{j0} > 0$ is the competition effect, ε_j is the part of the payoff that is observable to both players but unobservable to the econometrician, and (X, Z) is a vector of firm or market characteristics. Let $F(\varepsilon_1, \varepsilon_2; \alpha_0)$ be the joint distribution function of $(\varepsilon_1, \varepsilon_2)$, which is known up to the finite-dimensional parameter α_0. Let F_1 and F_2 denote the marginal distributions of ε_1 and ε_2 respectively. Let D_j be the dummy that equals one if player j enters the market. Suppose that it is a simultaneous-move static game. Then, following Andrews, Berry and Jia (2004) and Ciliberto and Tamer (2009), we can summarize the game by moment inequalities/equalities:

$$\begin{aligned} E(P_{00}(X, \theta_0) - (1 - D_1)(1 - D_2)|X, Z = z_0) &= 0, \\ E(P_{11}(X, \theta_0) - D_1 D_2 |X, Z = z_0) &= 0, \\ E(P_{10}(X, \theta_0) - D_1(1 - D_2)|X, Z = z_0) &\geq 0, \text{ and} \\ E(P_{01}(X, \theta_0) - D_2(1 - D_1)|X, Z = z_0) &\geq 0, \end{aligned} \tag{2.7}$$

where $\theta_0 = (\psi_{10}(z_0), \psi_{20}(z_0), \beta_{10}, \beta_{20}, \alpha_0, \delta_{10}, \delta_{20})$ and

$$\begin{aligned} & P_{00}(X, \theta) = \\ & 1 - F_1(\beta_1 X + \psi_1(z)) - F_2(\beta_2 X + \psi_2(z)) + F(\beta_1 X + \psi_1(z_0), \beta_2 X + \psi_2(z_0)), \\ & P_{11}(X, \theta) = F(\beta_1 X + \psi_1(z_0) - \delta_1, \beta_2 X + \psi_2(z_0) - \delta_2), \\ & P_{10}(X, \theta) = F_1(\beta_1 X + \psi_1(z_0)) - F(\beta_1 X + \psi_1(z_0), \beta_2 X + \psi_2(z_0) - \delta_2), \text{ and} \end{aligned}$$

$$P_{01}(X, \theta) = F_2(\beta_2 X + \psi_2(z_0)) - F(\beta_1 X + \psi_1(z_0) - \delta_1, \beta_2 X + \psi_2(z_0)). \tag{2.8}$$

In Andrews, Berry and Jia (2004) and Ciliberto and Tamer (2009), ψ_{j0} for $j = 1, 2$ are assumed to be linear functions of z_0. The linear functional form may be restrictive in many applications. It can be shown that the linear form is not essential for the identification of the model (e.g., see Bajari, Hong, and Ryan (2010)). Our method enables one to carry out inference about the parameters while allowing for nonparametric ψ_{j0} for $j = 1, 2$. $\square$

Example 5 (Revealed Preference Model). Consider a multiple-agent discrete choice model with J players, where each player j has a choice set A_j. Again, for notational simplicity, we suppress the i subscript. Let $\pi(a_j, a_{-j}, W)$ be the payoff of agent j that depends on his own action a_j, his opponents action a_{-j}, and his own and opponents' characteristics W. Let I_j be the information set of player j at the time of his decision. Rationality of the agents implies the following basic rule of action:

$$\sup_{a_j \in A_j} E(\pi(a_j, a_{-j}, W)|I_j) \leq E(\pi(a_j^*, a_{-j}, W)|I_j) \tag{2.9}$$

for $j = 1, ..., J$, where a_j^* is the observed action taken by j. For simplicity assume that the players move simultaneously so that the players do not respond to changes in other players' actions. Suppose that the econometrician models the payoff by $r(a_j, a_{-j}, W)$ and

$$r(a_j, a_{-j}, W) = E(\pi(a_j, a_{-j}, W)|I_j) + v_1(a_j) + v_2(a_j), \tag{2.10}$$

where the error $v_1(a_j)$ is unobservable to both the agents and the econometrician, while $v_2(a_j)$ is observable to the agents but not to the econometrician. Pakes (2010) proposes several assumptions on v_1 and v_2 that guarantee that (2.9) implies a moment inequality model of the following form:

$$E(r(a_j^*, a_{-j}, W) - r(a_j, a_{-j}, W)|W) \geq 0 \ \forall a_j \in A_j. \tag{2.11}$$

The model falls into our framework if we parametrize r as follows:

$$r(a_j^*, a_{-j}, W) - r(a_j, a_{-j}, W) = G(a_j^*, a_j, a_{-j}, \beta_0, X, \psi_0(Z)), \tag{2.12}$$

where X and Z are subvectors of W and G is a known function. □

In this paper, we construct confidence sets by inverting tests of the null hypothesis that θ is the true value for different $\theta \in \Theta$. The basis of the method is the test for the null hypothesis that the conditional moment inequalities/equalities (evaluated at θ) are valid. Clearly, such a test can be used directly to evaluate the validity of certain conditional moment inequalities/equalities as described in Example 6, which follows.

Example 6 (Functional Inequalities). Tests constructed in this paper are suitable for testing functional inequalities of the form:

$$
\begin{aligned}
&H_0 : u_j(x, z_0) \geq 0 \text{ for } z_0 \in \mathcal{Z} \text{ and all } (x, j) \in \mathcal{X} \times \{1, ..., p\}, \text{ where} \\
&u_j(x, z) = E(m_j(W_i)|X_i = x, Z_i = z)
\end{aligned}
\tag{2.13}
$$

and the observations $\{W_i = (Y_i, X_i, Z_i) : i \leq n\}$ are from a stationary process. When the Z_i variable is not present, the model reduces to that considered in Lee, Song and Whang (2011).[2] The current model allows one to specify the inequality hypotheses for a subpopulation with characteristic $Z_i = z_0$. Each of Lee, Song, and Whang's (2011) examples extend straightforwardly to our framework. An illustration of the extension is now given for the conditional treatment effect example.

Consider a controlled experiment, where treatment is randomly assigned to a group of subjects. Each subject is assigned the treatment with known probability $\pi(X_i, Z_i)$, where (X_i, Z_i) are the observed characteristics of the subject.[3] The researcher observes the treatment status $D_i \in \{1, 0\}$ and the outcomes $y_i(1)$ if treated and $y_i(0)$ if not treated. That is, the researcher observes D_i and $Y_i = D_i y_i(1) + (1 - D_i)y_i(0)$. The treatment effect for the ith individual is the difference between $y_i(1)$ and $y_i(0)$. The researcher is interested in testing if the average treatment effect given $X_i = x$ is positive for all $x \in \mathcal{X}$ for the subpopulation with characteristic $Z_i = z_0$. Then, our test for the hypotheses in (2.13) can be applied with $p = 1$ and

$$
m(W_i) = \frac{D_i Y_i}{\pi(X_i, Z_i)} - \frac{(1 - D_i)Y_i}{1 - \pi(X_i, Z_i)}, \tag{2.14}
$$

where $W_i = (Y_i, D_i, X_i, Z_i)$ and no parameter θ appears in the problem. □

2.3 Parameter Space

Let (θ, F) denote generic values of the parameter and distribution. Let $\mathcal{F}$ denote the parameter space for (θ_0, F_0). To specify $\mathcal{F}$ we need to introduce some notation.

Let $F_{Y|x,z}$ denote the conditional distribution of Y_i given $X_i = x$ and $Z_i = z$ under (θ, F). Let $F_{X|z}$ denote the conditional distribution of X_i given $Z_i = z$ under (θ, F). Let F_Z and F_X denote the marginal distributions of Z_i and X_i, respectively, under (θ, F).

[2] Note that the model is also covered by AS1 when Z_i is not present.

[3] The function $p(x, z)$ can be a constant. In this case, the assignment does not depend on observed or unobserved characteristics.

Let μ_X and μ_Y denote some measures on R^{d_x} and R^{d_y} (that do not depend on (θ, F)), with supports $\mathcal{Y}$ and $\mathcal{X}$, respectively. Let $\mathcal{Z}_0$ denote some neighborhood of z_0. Let μ_{Leb} denote Lebesgue measure on $\mathcal{Z}_0 \subset R^{d_z}$.

Define

$$\begin{aligned} m_F(\theta, x, z) &= E_F(m(W_i, \theta)|X_i = x, Z_i = z)f(z|x), \\ \Sigma_F(\theta, x, z) &= E_F(m(W_i, \theta)m(W_i, \theta)'|X_i = x, Z_i = z)f(z|x), \text{ and} \\ \sigma^2_{F,j}(\theta, z) &= E_F(m_j^2(W_i, \theta)|Z_i = z)f(z) \text{ for } j \leq k, \end{aligned} \tag{2.15}$$

where $k = p + v$, $f(z|x)$ is the conditional density with respect to Lebesgue measure of Z_i given $X_i = x$ and $f(z)$ is the density of Z_i wrt Lebesgue measure μ_{Leb} on $\mathcal{Z}_0$, defined in Assumption PS2 below.

The parameter space $\mathcal{F}$ is defined to be the collection of (θ, F) that satisfy the following parameter space (PS) assumptions, which define the model precisely.

Assumption PS1. (a) $\theta \in \Theta$,

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

(c) $E_F(m_j(W_i, \theta)|X_i, Z_i = z_0) \geq 0$ a.s. $[F_X]$ for $j = 1, ..., p$, and

(d) $E_F(m_j(W_i, \theta)|X_i, Z_i = z_0) = 0$ a.s. $[F_X]$ for $j = p + 1, ..., k$, where $k = p + v$.

Assumption PS2. (a) F_Z restricted to $z \in \mathcal{Z}_0$ is absolutely continuous wrt μ_{Leb} with density $f(z)$ $\forall z \in \mathcal{Z}_0$,

(b) F_X is absolutely continuous wrt μ_X with density $f(x)$ $\forall x \in \mathcal{X}$,

(c) $F_{Y|x,z}$ is absolutely continuous wrt μ_Y with density $f(y|x, z)$ $\forall (y, x, z) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{Z}_0$,

(d) $F_{Z|x}$ is absolutely continuous wrt μ_{Leb} on $\mathcal{Z}_0$ with density $f(z|x)$ $\forall (z, x) \in \mathcal{Z}_0 \times \mathcal{X}$, and

(e) $F_{X|z}$ is absolutely continuous wrt μ_X on R^{d_x} with density $f(x|z)$ $\forall (x, z) \in \mathcal{X} \times \mathcal{Z}_0$.

Let $\{C_\ell : \ell \leq 4\}$ be some finite constants and $\{\delta_j : j \leq k\}$ be some positive constants that do not depend on (θ, F).

Assumption PS3. (a) $\sigma^2_{F,j}(\theta, z_0) \geq \delta_j$,

(b) $m_F(\theta, x, z)$ is twice continuously differentiable in z on $\mathcal{Z}_0$ $\forall x \in \mathcal{X}$ with $\int L_m(x) \times f(x) d\mu_X(x) \leq C_1$, where $L_m(x) = \sup_{z \in \mathcal{Z}_0} \|(\partial^2/\partial z \partial z')m_F(\theta, x, z)\|$,

(c) $\sup_{z \in \mathcal{Z}_0} \int ||m_F(\theta, x, z)|| f(x, z) d\mu_X(x) \leq C_2$,

(d) $\Sigma_F(\theta, x, z)$ is Lipschitz continuous in z at z_0 on $\mathcal{Z}_0$ $\forall x \in \mathcal{X}$, i.e., $||\Sigma_F(\theta, x, z) - \Sigma_F(\theta, x, z_0)|| \leq L_\Sigma(x)||z||$, and $\int L_\Sigma(x) f(x) d\mu_X(x) \leq C_3$, and

(e) $E_F\left(|m_j(W_i, \theta)|^4 \middle| Z_i = z\right) f(z) \leq C_4$ $\forall z \in \mathcal{Z}_0$ $\forall j \leq k$.

Assumptions PS1(c) and (d) are the key partial-identification conditions of the model. Assumption PS2 specifies some absolute continuity conditions. Assumptions PS2(a) and (d) impose absolute continuity wrt Lebesgue measure of F_Z and $F_{Z|x}$ in a neighborhood of z_0. This is not restrictive because if F_Z and $F_{Z|x}$ have point mass at z_0, then the results of AS1 cover the model. Assumptions PS2(b), (c), and (e) are not very restrictive because the absolute continuity is wrt arbitrary measures μ_X and μ_Y, so the conditions allow for continuous, discrete, and mixed random variables. Assumption PS3 bounds some variances away from zero and imposes some smoothness and moment conditions. The smoothness conditions are on expectations, not on the underlying functions themselves, which makes them relatively weak.

Let $f(y, x, z) = f(y|x, z) f(x|z) f(z)$ and $f(x, z) = f(x|z) f(z)$.

The k-vector of moment functions is denoted

$$m(W_i, \theta) = (m_1(W_i, \theta), ..., m_k(W_i, \theta))'. \tag{2.16}$$

3 Tests and Confidence Sets

3.1 Test Statistics

Here we define the test statistic $T_n(\theta)$ that is used to construct a CS. We transform the conditional moment inequalities/equalities given X_i and $Z_i = z_0$ into equivalent conditional moment inequalities/equalities given only $Z_i = z_0$ by choosing appropriate weighting functions of X_i, i.e., X_i-instruments. Then, we construct a test statistic based on kernel averages of the instrumented moment conditions over Z_i values that lie in a neighborhood of z_0.

The instrumented conditional moment conditions given $Z_i = z_0$ are of the form:

$$\begin{aligned} &E_{F_0}(m_j(W_i, \theta_0) g_j(X_i) | Z_i = z_0) \geq 0 \text{ for } j = 1, ..., p \text{ and} \\ &E_{F_0}(m_j(W_i, \theta_0) g_j(X_i) | Z_i = z_0) = 0 \text{ for } j = p+1, ..., k, \text{ for } g = (g_1, ..., g_k)' \in \mathcal{G}_{c\text{-}cube}, \end{aligned} \tag{3.1}$$

where $g = (g_1, ..., g_k)'$ are instruments that depend on the conditioning variables X_i

and $\mathcal{G}_{c\text{-}cube}$ is a collection of instruments defined in (3.6) below. The collection $\mathcal{G}_{c\text{-}cube}$ is chosen so that there is no loss in information.

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

$$\begin{aligned} \overline{m}_n(\theta, g) &= n^{-1} \sum_{i=1}^{n} m(W_i, \theta, g, b) \text{ for } g \in \mathcal{G}_{c\text{-}cube}, \text{ where} \\ m(W_i, \theta, g, b) &= b^{-d_z/2} K_b(Z_i) m(W_i, \theta, g), \\ K_b(Z_i) &= 0.75 \max\{1 - ((Z_i - z_0)/b)^2, 0\}, \\ m(W_i, \theta, g) &= \begin{pmatrix} m_1(W_i, \theta) g_1(X_i) \\ m_2(W_i, \theta) g_2(X_i) \\ \vdots \\ m_k(W_i, \theta) g_k(X_i) \end{pmatrix} \text{ for } g \in \mathcal{G}_{c\text{-}cube}, \end{aligned} \tag{3.2}$$

and $b > 0$ is a scalar bandwidth parameter for which $b = b_n = o(n^{-1/(4+d_z)})$ and $nb^{d_z} \to \infty$ as $n \to \infty$.[4] In the scalar Z_i case, we take $b = b^0 n^{-2/7}$, where $b^0 = 4.68\hat{\sigma}_z$ and $\hat{\sigma}_z$ is the estimated standard deviation of Z_i.[5,6] The kernel employed in (3.2) is the Epanechnikov kernel. For notational simplicity, we omit the dependence of $\overline{m}_n(\theta, g)$ (and various other quantities below) on b.

Note that the normalization $b^{-d_z/2}$ that appears in $m(W_i, \theta, g, b)$ yields $m(W_i, \theta, g, b)$ to have a variance matrix that is $O(1)$, but not $o(1)$. In fact, under the conditions given below, $Var_F(m(W_i, \theta, g, b)) \to Var_F(m(W_i, \theta, g)|Z_i = z_0) f(z_0)$ as $n \to \infty$ under $(\theta, F) \in \mathcal{F}$.

If the sample average $\overline{m}_n(\theta, g)$ is divided by the scalar $n^{-1} \sum_{i=1}^{n} b^{-d_z/2} K_b(Z_i)$ it becomes the Nadaraya-Watson nonparametric kernel estimator of $E(m(W_i, \theta, g)|Z_i = z_0)$. We omit this divisor because doing so simplifies the statistic and has no effect on the test defined below.[7]

[4] The conditions on b are standard assumptions in the nonparametric density and regression literature. When these conditions are applied to a nonparametric regression or density estimator, the first condition implies that the bias of the estimator goes to zero faster than the variance (and is the weakest condition for which this holds) and the second condition implies that the estimator is asymptotically normal (because it implies that b goes to zero sufficiently slowly that a Lindeberg condition holds).

[5] The bandwidth b is under-smoothed due to the factor $n^{-2/7}$, which is the same as in Chernozhukov, Lee, and Rosen (2008), rather than $n^{-1/5}$. It is somewhat arbitrary, but seems to work well in practice.

[6] The definition of $\overline{m}_n(\theta, g)$ in (3.2) is the same as the definition of $\overline{m}_n(\theta, g)$ in AS1 except for the multiplicand $b^{-d_z/2} K_b(Z_i)$ in $m(W_i, \theta, g, b)$.

[7] This holds because division by $n^{-1} \sum_{i=1}^{n} b^{-d_z/2} K_b(Z_i)$ rescales the test statistic and critical value identically and in consequence the rescaling cancels out.

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

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

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

$$\overline{\Sigma}_n(\theta, g) = \widehat{\Sigma}_n(\theta, g) + \varepsilon \cdot Diag(\widehat{\Sigma}_n(\theta, 1_k)) \text{ for } g \in \mathcal{G}_{c\text{-}cube} \text{ for } \varepsilon = 5/100. \qquad (3.4)$$

By design, $\overline{\Sigma}_n(\theta, g)$ is a linear combination of two scale equivariant functions and hence is scale equivariant.[8] This yields a test statistic that is invariant to rescaling of the moment functions $m(W_i, \theta)$, which is an important property.

The functions g that we consider are hypercubes on $[0, 1]^{d_X}$. 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 untransformed IV vector and we let X_i denote the transformed IV vector. We transform $X_i^\dagger$ via a shift and rotation and then an application of the standard normal distribution function $\Phi(x)$. Specifically, let

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

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

[8] That is, multiplying the moment functions $m(W_i, \theta)$ by a diagonal matrix, D, changes $\overline{\Sigma}_n(\theta, g)$ into $D\overline{\Sigma}_n(\theta, g)D$.

is countable:

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

for some positive integer r_0.[9] 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

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

where $S = S_1, S_2$, or S_3, as defined in (3.9) below. The asymptotic size and consistency results for the CS's and tests based on $\overline{T}_{n,r_{1,n}}(\theta)$ 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

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

where $\mathcal{G}_{c\text{-}cube,r_{1,n}} = \{g_{a,r} \in \mathcal{G}_{c\text{-}cube} : r \le r_{1,n}\}$. For brevity, the discussion in this paper focusses on CvM statistics and all results stated concern CvM statistics. Similar results hold for KS statistics.[10]

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

[10] 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_2, and S_3 are defined by

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

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

3.2 Critical Values

3.2.1 GMS Critical Values

In this section we define two GMS critical values. The first is based on the asymptotic distribution. The second is a bootstrap version of the first. Both require simulation.

We first describe how to compute the GMS critical value that is based on the asymptotic null distribution of the test statistic.

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

$$\begin{aligned} \xi_n(\theta,g) &= \kappa_n^{-1} n^{1/2} \overline{D}_n^{-1/2}(\theta,g)\overline{m}_n(\theta,g), \text{ where} \\ \overline{D}_n(\theta,g) &= Diag(\overline{\Sigma}_n(\theta,g)),\ \kappa_n = (0.3\ln(n))^{1/2}, \end{aligned} \tag{3.10}$$

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

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

$$\begin{aligned} &\overline{\varphi}_{n,j}(\theta,g) = \overline{\Sigma}_{n,j}^{1/2}(\theta,g) B_n 1(\xi_{n,j}(\theta,g) > 1) \text{ and} \\ &B_n = (0.4\ln(n)/\ln\ln(n))^{1/2}, \end{aligned} \tag{3.11}$$

where $\widehat{\Sigma}_{n,j}(\theta, g)$ and $\overline{\Sigma}_{n,j}(\theta, g)$ denote the (j,j) elements of $\widehat{\Sigma}_n(\theta, g)$ and $\overline{\Sigma}_n(\theta, g)$, respectively.

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

Step 3. Compute the $(kN_g) \times (kN_g)$ covariance matrix $\widehat{\Sigma}_{n,mat}(\theta)$. Its elements are the covariances $\widehat{\Sigma}_n(\theta, g_{a,r}, g_{a,r}^*)$ for $a \in \{1, ..., 2r\}^{d_X}$ and $r = 1, ..., r_{1,n}$, which are defined as follows. For $g = g_{a,r}$ and $g^* = g_{a,r}^*$, let

$$\widehat{\Sigma}_n(\theta, g, g^*) = n^{-1} \sum_{i=1}^{n} \left(m(W_i, \theta, g, b) - \overline{m}_n(\theta, g)\right) \left(m(W_i, \theta, g^*, b) - \overline{m}_n(\theta, g^*)\right)'. \quad (3.12)$$

Note that $\widehat{\Sigma}_n(\theta, g)$, defined in (3.3), equals $\widehat{\Sigma}_n(\theta, g, g)$.

Step 4. Compute the $(kN_g) \times \tau_{reps}$ matrix $\overline{\nu}_n(\theta) = \widehat{\Sigma}_{n,mat}^{1/2}(\theta) Z$. Let $\overline{\nu}_{n,\tau}(\theta, g_{a,r})$ denote the k dimensional sub-vector of $\overline{\nu}_n$ that corresponds to the k rows indexed by $g_{a,r}$ and column τ for $\tau = 1, ..., \tau_{reps}$.

Step 5. For $\tau = 1, ..., \tau_{reps}$, compute the simulated test statistic $\overline{T}_{n,r_{1,n},\tau}(\theta)$ just 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 $n^{1/2}\overline{m}_n(\theta, g_{a,r})$ replaced by $\overline{\nu}_{n,j}(\theta, g_{a,r}) + \overline{\varphi}_n(\theta, g_{a,r})$.

Step 6. Take the critical value $c_{n,1-\alpha}^{GMS,Asy}(\theta)$ to be the $1 - \alpha + \eta$ sample quantile of the simulated test statistics $\{\overline{T}_{n,r_{1,n},\tau}(\theta) : \tau = 1, ..., \tau_{reps}\}$ plus η, where $\eta = 10^{-6}$.[11] v

For the bootstrap version of the GMS critical value, Steps 2 and 4-6 are replaced by the following steps:

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

Step 4$_{boot}$. For each bootstrap sample, transform the regressors as in (3.5) (using

[11] The description of the GMS critical values given here is a little different (and simpler) than in AS1 and in the asymptotic results given in the Appendix. However, their properties are the same. In AS1, $\overline{\varphi}_{n,j}(\theta, g)$ is multiplied by $\widehat{\Sigma}_{n,j}^{-1/2}(\theta, 1_k)$ for $j \le p$ and $\widehat{\Sigma}_n(\theta, g, g^*)$ is replaced by $\widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta, g, g^*)\widehat{D}_n^{-1/2}(\theta)$, where $\widehat{D}_n(\theta) = Diag(\widehat{\Sigma}_n(\theta, 1_k))$. This has no effect on the distribution of $\overline{T}_{n,r_{1,n},\tau}(\theta)$ (conditionally on the sample or unconditionally) because (i) $S_j(m, \Sigma) = S_j(Dm, D\Sigma D)$ for any pd diagonal $k \times k$ matrix D for $j = 1, 2, 3$ and (ii) $Var_{|\{W_i\}}(((1_{N_g \times N_g} \otimes \widehat{D}_n^{-1/2}(\theta))\widehat{\Sigma}_{n,mat}(\theta) \times (1_{N_g \times N_g} \otimes \widehat{D}_n^{-1/2}(\theta)))^{1/2} Z_\tau) = Var_{|\{W_i\}}((1_{N_g \times N_g} \otimes \widehat{D}_n^{-1/2}(\theta))\widehat{\Sigma}_{n,mat}^{1/2}(\theta) Z_\tau)$, where $Var_{|\{W_i\}}(\cdot)$ denotes the conditional variance given the sample $\{W_i : i \le n\}$ and Z_τ denotes the τth column of Z.

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

Step 5$_{boot}$. For each bootstrap sample, compute the bootstrap test statistic $\overline{T}^*_{n,r_{1,n},\tau}(\theta)$ as $\overline{T}^{CvM}_{n,r_{1,n}}(\theta)$ (or $\overline{T}^{KS}_{n,r_{1,n}}(\theta)$) is computed in (3.7) (or (3.8)) but with $n^{1/2}\overline{m}_n(\theta, g_{a,r})$ replaced by $n^{1/2}(\overline{m}^*_{n,\tau}(\theta, g_{a,r}) - \overline{m}_n(\theta, g_{a,r})) + \overline{\varphi}_n(\theta, g_{a,r})$ and with $\overline{\Sigma}_n(\theta, g_{a,r})$ replaced by $\overline{\Sigma}^*_{n,\tau}(\theta, g_{a,r})$.

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

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

3.2.2 Plug-in Asymptotic Critical Values

Next, for comparative purposes, we define 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 Section 4 show better performance by GMS critical values than PA and subsampling critical values.

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

The nominal $1-\alpha$ PA CS is given by (2.3) with $T_n(\theta) = \overline{T}^{CvM}_{n,r_{1,n}}(\theta)$ (or $\overline{T}^{KS}_{n,r_{1,n}}(\theta)$) 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 $\overline{T}^{CvM}_{n,r_{1,n}}(\theta_*)$ (or $\overline{T}^{KS}_{n,r_{1,n}}(\theta_*)$) 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 $\overline{\varphi}_{n,j}(\theta, g) \geq 0$ for all $g \in \mathcal{G}$ 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 $\ell = 1, 2, 3$). 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 $E_{F_n}(m_j(W_i, \theta_{n,*})|X_i, Z_i = z_0)$ for some $j = 1, ..., p$ is bounded away from zero as $n \to \infty$ with positive X_i probability.

3.3 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 notation. We define the asymptotic covariance kernel, $\{h_{2,F}(\theta, g, g^*) : g, g^* \in \mathcal{G}\}$, of $n^{1/2}\overline{m}_n(\theta, g)$ after normalization via a diagonal matrix $D_F^{-1/2}(\theta, z_0)$. Define[12]

$$\begin{aligned} h_{2,F}(\theta, g, g^*) &= D_F^{-1/2}(\theta, z_0)\Sigma_F(\theta, g, g^*, z_0)D_F^{-1/2}(\theta, z_0), \text{ where} \\ \Sigma_F(\theta, g, g^*, z) &= E_F(m(W_i, \theta, g)m(W_i, \theta, g^*)'|Z_i = z)f(z) \text{ and} \\ D_F(\theta, z) &= Diag(\Sigma_F(\theta, 1_k, 1_k, z)) \; (= Diag(E_F(m(W_i, \theta)m(W_i, \theta)'|Z_i = z)f(z))). \end{aligned} \tag{3.13}$$

For simplicity, let $h_{2,F}(\theta)$ abbreviate $\{h_{2,F}(\theta, g, g^*) : g, g^* \in \mathcal{G}_{c\text{-}cube}\}$.

Define

$$\mathcal{H}_2 = \{h_{2,F}(\theta) : (\theta, F) \in \mathcal{F}\}. \tag{3.14}$$

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

$$d(h_2^{(1)}, h_2^{(2)}) = \sup_{g, g^* \in \mathcal{G}_{c\text{-}cube}} ||h_2^{(1)}(g, g^*) - h_2^{(2)}(g, g^*)||. \tag{3.15}$$

[12]Note that $D_F(\theta, z) = Diag(\sigma^2_{F,1}(\theta, z), ..., \sigma^2_{F,k}(\theta, z))$, where $\sigma^2_{F,j}(\theta, z)$ is defined in (2.15). Also note that the means, $E_F m(W_i, \theta, g)$, $E_F m(W_i, \theta, g^*)$, and $E_F m(W_i, \theta)$, are not subtracted off in the definitions of $\Sigma_F(\theta, g, g^*, z)$ and $D_F(\theta, z)$. The reason is that the population means of the sample-size n quantities based on $m(W_i, \theta, g, b)$ are smaller than the second moments by an order of magnitude and, hence, are asymptotically negligible. See Lemmas AN6 and AN7 in the Appendix.

Correct asymptotic size is established in the following theorem.

Theorem N1. *For every compact subset* $\mathcal{H}_{2,cpt}$ *of* $\mathcal{H}_2$, GMS *and* PA *confidence sets* CS_n *satisfy*

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

(b) GMS *confidence sets based on the MMM and Max functions,* S_1 *and* S_3, *satisfy*

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

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

Comments. 1. Theorem N1(a) shows that GMS and PA CS's have correct uniform asymptotic size over compact sets of covariance kernels. Theorem N1(b) shows that GMS CS's based on S_1 and S_3 are at most infinitesimally conservative asymptotically (i.e., their asymptotic size is infinitessimally close to their nominal size). The uniformity results hold whether the moment conditions involve "weak" or "strong" instrumental variables X_i.

2. Theorem N1(b) also holds for GMS CS's based on the QLR function S_2 provided the asymptotic distribution function of the test statistic under some fixed $(\theta_c,F_c)\in\mathcal{F}$ with $h_{2,F_c}(\theta_c)\in\mathcal{H}_{2,cpt}$ is continuous and strictly increasing at its $1-\alpha$ quantile plus δ for all $\delta>0$ sufficiently small and $\delta=0$.[13] This condition likely holds in most models, but it is hard to give primitive conditions under which it holds.

3. As in AS1, an analogue of Theorem N1(b) holds for PA CS's if $E_{F_c}(m_j(W_i,\theta_c)|X_i, Z_i=z_0)=0$ a.s. for $j\leq p$ (i.e., if the conditional moment inequalities hold as equalities a.s.) under some $(\theta_c,F_c)\in\mathcal{F}$. However, the latter condition is restrictive—it fails in many applications.

4. The proofs in the Appendix cover asymptotic critical values, but not bootstrap critical values. Extending the results to cover bootstrap critical values just requires a suitable bootstrap empirical process result. For brevity, we do not give such a result. 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

[13] This condition is Assumption GMS2(a) in Section 7.4 of the Appendix.

converge in probability to nonrandom hypercubes (in an L^2 sense). Again, for brevity, we do not do so. Finally, the asymptotic results cover non-data dependent bandwidths, as is typical in the nonparametric and semiparametric literature.

3.4 Power Against Fixed Alternatives

We now show that the power of GMS and PA tests converges to one as $n \to \infty$ for all fixed alternatives (for which the moment functions have $4+\delta$ moments finite). Thus, both tests are consistent tests. This implies that for any fixed distribution F_0 and any parameter value θ_* *not* in the identified set Θ_{F_0}, the GMS and PA CS's do not include θ_* with probability approaching one. In this sense, GMS and PA CS's based on $T_n(\theta)$ fully exploit the conditional moment inequalities and equalities. CS's based on a finite number of unconditional moment inequalities and equalities do not have this property.

The null hypothesis is

$$\begin{aligned} H_0 : E_{F_0}(m_j(W_i, \theta_*)|X_i, Z_i = z_0) &\geq 0 \text{ a.s. } [F_{X,0}] \text{ for } j = 1, ..., p \text{ and} \\ E_{F_0}(m_j(W_i, \theta_*)|X_i, Z_i = z_0) &= 0 \text{ a.s. } [F_{X,0}] \text{ for } j = p+1, ..., k, \end{aligned} \tag{3.16}$$

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 $\mathcal{F}_+$ denote all (θ, F) that satisfy Assumptions PS1-PS3 that define $\mathcal{F}$ except Assumptions PS1(c) and (d) (which impose the conditional moment inequalities and equalities). As defined, $\mathcal{F} \subset \mathcal{F}_+$. 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, $\mathcal{X}_F(\theta)$, 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 $E_F(||m(W_i, \theta)||\ |Z_i = z_0) < \infty$, let

$$\begin{aligned} \mathcal{X}_F(\theta) = \{x \in R^{d_x} : &E_F(m_j(W_i, \theta)\,|X_i = x, Z_i = z_0) < 0 \text{ for some } j \leq p \text{ or} \\ &E_F(m_j(W_i, \theta)\,|X_i = x, Z_i = z_0) \neq 0 \text{ for some } j = p+1, ..., k\}. \end{aligned} \tag{3.17}$$

The next assumption, Assumption NFA, 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 conditional probability given $Z_i = z_0$ under F_0. Thus, under Assumption

NFA, the moment conditions specified in (3.16) do not hold.

Assumption NFA. The null value $\theta_* \in \Theta$ and the true distribution F_0 satisfy: (a) $P_{F_0}(X_i \in \mathcal{X}_{F_0}(\theta_*)|Z_i = z_0) > 0$, where $\mathcal{X}_{F_0}(\theta_*)$ is defined in (3.17), and (b) $(\theta_*, F_0) \in \mathcal{F}_+$.

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

Theorem AN2. *Suppose Assumption* NFA *holds. Then,*

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

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

Comment 4 to Theorem AN1 applies also to Theorem AN2.

4 Monte Carlo Simulations

This section provides simulation evidence concerning the finite-sample properties of the confidence intervals (CI's) and tests introduced in the paper. We consider two models: a quantile selection model and a conditional treatment effect model. In the quantile selection model, we compare different versions of the CI's introduced in the paper. In the conditional treatment effect model, the tests are used directly (rather than to construct CI's), and we compare different versions of the tests. In both models, we provide comparisons of the proposed procedures with the series and local linear procedures in CLR.

4.1 Confidence Intervals and Tests Considered

To be specific, we compare different test statistics and critical values in terms of their coverage probabilities (CP's) for points in the identified set and their false coverage probabilities (FCP's) for points outside the identified set in the quantile selection model. We compare different test statistics and critical values in terms of their rejection probabilities under the null (NRP's) and under alternatives (ARP's) in the conditional treatment effect model. Obviously, one wants FCP's (ARP's) to be as small (large) as possible. FCP's are directly related to the power of the tests used to constructed the CI and are related to the length of the CI, see Pratt (1961).

The following test statistics are considered: (i) CvM/Sum, (ii) CvM/QLR, (iii) CvM/Max, (iv) KS/Sum, (v) KS/QLR, and (vi) KS/Max, as defined in Section 3. In the conditional treatment effect model, different choices of the S function (Sum, QLR and Max) coincide because there is only one conditional moment inequality. We thus do not distinguish them in the results. Asymptotic normal, bootstrap, and subsampling critical values are computed. In particular, we consider PA/Asy, PA/Bt, GMS/Asy, GMS/Bt, and Sub critical values.[14] The critical values are simulated using 5001 repetitions (for each original sample repetition). The base case values of κ_n, B_n, and ε for the GMS critical values are specified as follows and are used in both models: $\kappa_n = \sqrt{0.3\log(n)}$, $B_n = \sqrt{0.4\log(n)/\log(\log(n))}$, and $\varepsilon = 5/100$. Additional results are reported for variations of these values. The base case sample size is 250. Some additional results are reported for $n = 100$ and 500. The subsample size is 20 when the sample size is 250. Results are reported for nominal 0.95 CI's and 0.05 tests. The number of simulation repetitions used to compute CP's and FCP's is 5000 for all cases. This yields a simulation standard error of 0.0031.

In the first model, the reported FCP's are "CP-corrected" by employing a critical value that yields a CP equal to 0.95 at the closest point of the identified set if the CP at the closest point is less than 0.95. If the CP at the closest point is greater than 0.95, then no CP correction is carried out. The reason for this "asymmetric" CP correction is that CS's may have CP's greater than 0.95 for points in the identified set, even asymptotically, in the present context and one does not want to reward over-coverage of points in the identified set by CP correcting the critical values when making comparisons of FCP's. In the second model, the ARP's are "NRP-corrected" analogously.

We use the Epanechnikov kernel and the bandwidth $b = b^0 n^{-2/7}$ described in the paragraph containing (3.2) for both simulation examples. For comparative purposes, some results are also reported for $b = 0.5b^0 n^{-2/7}$ and $b = 2b^0 n^{-2/7}$.

We provide simulation comparisons of our CS's and tests with those of CLR. To implement the CLR tests, we follow Example C of CLR. For the quantile selection model, for each θ, we use $\beta(x, z, \theta)$ defined in (4.4) of CLR as the auxiliary bound function, use $\beta_l(z, \theta) = \min_{x \in \mathcal{X}} \beta(x, z, \theta)$ as the auxiliary parameter, and test $H_0 : \beta_l(z, \theta) \geq 0$ against $\beta_l(z, \theta) < 0$. (The CLR CI's are obtained by inverting the CLR tests.) For the treatment

[14]The Sum, QLR, and Max statistics use the functions S_1, S_2, and S_3, respectively. The PA/Asy and PA/Bt critical values are based on the asymptotic distribution and bootstrap, respectively, and likewise for the GMS/Asy and GMS/Bt critical values. The quantity η is set to 0 because its value, provided it is sufficiently small, has no effect in these models. Sub denotes a (non-recentered) subsampling critical value. It is the 0.95 sample quantile of the subsample statistics, each of which is defined exactly as the full sample statistic is defined but using the subsample in place of the full sample. The number of subsamples considered is 5001. They are drawn randomly without replacement.

effect model, described below, we use $\beta(x,z) = E[Y_iD_i/p - Y_i(1-D_i)/(1-p)|(X_i,Z_i) = (x,z)]$ as the auxiliary bound function, use $\beta_l(z) = \min_{x\in\mathcal{X}}\beta(x,z)$ as the auxiliary parameter, and test $H_0 : \beta_l(z) \geq 0$ against $\beta_l(z) < 0$.

We implement both the series and local linear versions of CLR's test. We use their GAUSS code and follow the implementation instructions in CLR whenever possible. The models considered here, however, are more complicated than in CLR's examples because the nonparametric estimation of $\beta(x,z,\theta)$ involves two regressors X_i and Z_i. The latter poses new questions about the choices of knots in the series approximation and the choices of bandwidths in the local linear approximation. For the series version, we use tensor product B-splines and allow different numbers of knots for X_i and Z_i. The number of knots is the (integer part of the) number chosen by cross-validation multiplied by $\sqrt{n^{-1/5}n^{2/7}}$. The multiplicative factor is used to obtain undersmoothing. For the local linear version, we use the optimal bandwidth formula given for multivariate local linear regression by Yang and Tschernig (1999) (Equation A.1), and use the same plug-in rule as CLR's rule-of-thumb bandwidth to plug in the estimated quantities. The resulting plug-in bandwidth is then multiplied by $\sqrt{n^{-1/5}n^{2/7}}$ to obtain undersmoothing. The CLR CS's and tests employ an estimated contact set.

4.2 Nonparametric Quantile Selection

This model extends the quantile selection model in AS1. We are interested in the conditional τ-quantile of a treatment response given the value of covariates X_i and Z_i. The results also apply to other types of response variables with selection. As in AS1, X_i is assumed to satisfy the quantile monotone instrumental variable (QMIV) assumption. In this paper, we add an additional covariate Z_i that does not necessarily satisfy the QMIV assumption. The results of AS1 do not cover such a model.

The model setup is as follows. The observations are i.i.d. Let $y_i(t) \in \mathcal{Y}$ be individual i's "conjectured" response variable given treatment $t \in \mathcal{T}$. Let T_i be the realization of the treatment for individual i. The observed outcome variable is $Y_i = y_i(T_i)$. Let X_i be a covariate whose support contains an ordered set $\mathcal{X}$. Let Z_i be another covariate. We observe $W_i = (Y_i, X_i, Z_i, T_i)$. The parameter of interest, θ, is the conditional τ-quantile of $y_i(t)$ given $(X_i, Z_i) = (x_0, z_0)$ for some $t \in \mathcal{T}$, some $x_0 \in \mathcal{X}$, and some $z_0 \in \mathcal{Z}$, which is denoted $Q_{y_i(t)|X_i,Z_i}(\tau|x_0, z_0)$. We assume the conditional distribution of $y_i(t)$ given $(X_i, Z_i) = (x, z_0)$ is absolutely continuous at its τ-quantile for all $x \in \mathcal{X}$. We assume that X_i satisfies the QMIV assumption given $Z_i = z_0$, i.e., $Q_{y_i(t)|X_i,Z_i}(\tau|x_1, z_0) \leq Q_{y_i(t)|X_i,Z_i}(\tau|x_2, z_0)$ for all $x_1 \leq x_2$.

AS1 describes four empirical problems that fit in their quantile selection model. All of those problems fit in the nonparametric quantile selection model considered here if one or more of the covariates is not a QMIV.

The model setup above implies the following conditional moment inequalities:

$$\begin{aligned} E\left(1(X_i \leq x_0)[1(Y_i \leq \theta, T_i = t) + 1(T_i \neq t) - \tau]|X_i, Z_i = z_0\right) \geq 0 \text{ a.s. and} \\ E\left(1(X_i \geq x_0)[\tau - 1(Y_i \leq \theta, T_i = t)]|X_i, Z_i = z_0\right) \geq 0 \text{ a.s.} \end{aligned} \tag{4.1}$$

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

$$\begin{aligned} y_i(1) &= \mu(X_i, Z_i) + \sigma\left(X_i, Z_i\right) u_i, \text{ where } \partial\mu\left(x, z\right)/\partial x \geq 0 \text{ and } \sigma\left(x, z\right) \geq 0, \\ T_i &= 1\{L\left(X_i, Z_i\right) + \varepsilon_i \geq 0\}, \text{ where } \partial L\left(x, z\right)/\partial x \geq 0, \\ X_i, Z_i &\sim Unif[0,2], \ (\varepsilon_i, u_i) \sim N(0, I_2), \ (X_i, Z_i) \perp (\varepsilon_i, u_i), \ X_i \perp Z_i, \\ Y_i &= y_i(T_i), \text{ and } t = 1. \end{aligned} \tag{4.2}$$

The variable $y_i(0)$ is irrelevant (because Y_i enters the moment inequalities in (4.1) only through $1(Y_i \leq \theta, T_i = t)$) and, hence, is left undefined. With this DGP, X_i satisfies the QMIV assumption for any $\tau \in (0, 1)$ and Z_i might not. We consider the median: $\tau = 0.5$. We focus on the conditional median of $y_i(1)$ given $(X_i, Z_i) = (1.5, 1.0)$, i.e., $\theta = Q_{y_i(1)|X_i,Z_i}(0.5|x_0, z_0)$ with $(x_0, z_0) = (1.5, 1.0)$.

Some algebra shows that the conditional moment inequalities in (4.1) imply:

$$\begin{aligned} \theta \geq \underline{\theta}(x, z_0) := \mu(x, z_0) + \sigma\left(x, z_0\right) \Phi^{-1}\left(1 - \left[2\Phi\left(L\left(x, z_0\right)\right)\right]^{-1}\right) \text{ for } x \leq 1.5 \text{ and} \\ \theta \leq \bar{\theta}\left(x, z_0\right) := \mu(x, z_0) + \sigma\left(x, z_0\right) \Phi^{-1}\left(\left[2\Phi\left(L\left(x, z_0\right)\right)\right]^{-1}\right) \text{ for } x \geq 1.5. \end{aligned} \tag{4.3}$$

We call $\underline{\theta}(x, z_0)$ and $\bar{\theta}\left(x, z_0\right)$ the lower and upper bound functions on θ, respectively. The identified set for the quantile selection model is $\left[\sup_{x \leq x_0} \underline{\theta}(x, z_0), \ \inf_{x \geq x_0} \bar{\theta}\left(x, z_0\right)\right]$. The shape of the lower and upper bound functions depends on the μ, σ, and L functions. We consider three specifications, one that yields flat bound functions, another that yields kinked bound functions, and a third that yields peaked bound functions.[15]

[15] For the flat bound DGP, $\mu(x, z) = 2$, $\sigma\left(x, z\right) = 1$, and $L\left(x, z\right) = 1$ for $x, z \in [0, 2]$. In this case, $\underline{\theta}(x, z) = 2 + \Phi^{-1}\left(1 - \left[2\Phi\left(1\right)\right]^{-1}\right)$ for $x \leq 1.5$ and $\bar{\theta}\left(x, z\right) = 2 + \Phi^{-1}\left(\left[2\Phi\left(1\right)\right]^{-1}\right)$ for $x > 1.5$. For

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

$$\beta(x,z,\theta)=\begin{cases} E\left(1(Y_i\leq\theta,T_i=1)+1(T_i\neq 1)-0.5|\,(X_i,Z_i)=(x,z)\right) & \text{if } x<1.5 \\ E\left(0.5-1(Y_i\leq\theta,T_i=1)|\,(X_i,Z_i)=(x,z)\right) & \text{if } x\geq 1.5. \end{cases} \tag{4.4}$$

The conditional moment functions as functions of x at $z=z_0$ are flat, kinked and peaked under the three specifications of μ, σ, and L functions, respectively. The functions as a function of z at each x also possess those three shapes at the point $z=z_0$ depending on the specification.

4.2.1 g Functions

The g functions employed by the test statistics are indicator functions of hypercubes in $[0,1]$, i.e., intervals, as in AS1. The regressor X_i is transformed via the method described in Section 9 in AS1 to lie in $(0,1)$.[16] The hypercubes have side-edge lengths $(2r)^{-1}$ for $r=r_0,...,r_1$, where $r_0=1$ and the base case value of r_1 is 3.[17] The base case number of hypercubes is 12. We also report results for $r_1=2$, 4, which yield 6, and 20 hypercubes, respectively.

Note that we use a smaller value of r_1 as the base-case value in this paper than in AS1. This is because the test statistic for a nonparametric parameter of interest depends only on observations local to $Z_i=z_0$, which is a fraction of the full sample. For example, the Epanechnikov kernel gives positive weight only to observations within distance b to z_0. When $n=250$ and $Z\sim Unif[0,2]$, observations that receive positive weight lie in an interval centered at z_0 of length about $2b=9.36\sigma_Z n^{-2/7}\approx 0.64$, which is 32% of the support of Z_i. This interval on average contains 80 effective observations when $n=250$. Thus, the finest cube when $r_1=3$ contains $80/6\approx 13$ effective observations. On the

the kinked bound DGP, $\mu(x,z)=(x\wedge 1)+(z\wedge 1)$, $\sigma(x,z)=(x+z)/2$, $L(x,z)=x\wedge 1$, $\underline{\theta}(x,z)=(x\wedge 1)+(z\wedge 1)+(x+z)\cdot\Phi^{-1}\left(1-[2\Phi(x\wedge 1)]^{-1}\right)/2$ for $x\leq 1.5$, and $\bar{\theta}(x,z)=(x\wedge 1)+(z\wedge 1)+(x+z)\cdot\Phi^{-1}\left([2\Phi(x\wedge 1)]^{-1}\right)/2$ for $x>1.5$. For the peaked bound function, $\mu(x,z)=(x\wedge 1)+(z\wedge 1)$, $\sigma(x,z)=\left(x^5+z^5\right)/2$, $L(x,z)=x\wedge 1$, $\underline{\theta}(x,z)=(x\wedge 1)+(z\wedge 1)+\left(x^5+z^5\right)\Phi^{-1}\left(1-[2\Phi(x\wedge 1)]^{-1}\right)/2$ for $x\leq 1.5$, and $\bar{\theta}(x,z)=(x\wedge 1)+(z\wedge 1)+\left(x^5+z^5\right)\Phi^{-1}\left([2\Phi(x\wedge 1)]^{-1}\right)/2$ for $x>1.5$.

[16] This method takes the transformed regressor to be $\Phi((X_i-\overline{X}_n)/\sigma_{X,n})$, where $\overline{X}_n$ and $\sigma_{X,n}$ are the sample mean and standard deviation of X_i and $\Phi(\cdot)$ is the standard normal distribution function.

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

other hand, the finest cube when $r_1 = 7$ contains only $80/14 \approx 5.7$ effective observations. For this reason, a value of r_1 that is smaller than that used in AS1 leads to better CP and FCP performance of the CS's in the nonparametric model.

4.2.2 Simulation Results: Confidence Intervals Proposed in This Paper

Tables I-III report CP's and CP-corrected FCP's for a variety of test statistics and critical values proposed in this paper for a range of cases. The CP's are for the lower endpoint of the identified interval in Tables I-III and for the flat, kinked, and peaked bound functions. FCP's are for points below the lower endpoint.[18]

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

Table I shows that all of the CI's have coverage probabilities greater than or equal to 0.95 for all three specifications of the bound functions. The PA/Asy CI's have noticeably larger over-coverage than the GMS/Asy CI's in all cases. The GMS/Asy CI's have CP's close to 0.95 with the flat bound DGP and larger than 0.95 with the other two DGP's. The CP's are not sensitive to the choice of the test statistics.

The FCP results in Table 1 show (i) a clear advantage of the GMS-based CI's over the PA-based ones, (ii) a clear advantage of the CvM-based CI's over the KS-based ones, and (iii) little difference between the test statistic functions: Sum, QLR or Max. The comparison holds for all three types of DGP's.

Table II compares the critical values PA/Asy, PA/Bt, GMS/Asy, GMS/Asy, and Sub. The results show little difference in CP's and FCP's between the Asy and Bt

[18] Note that the DGP is the same for FCP's as for CP's, just the value θ that is to be covered is different. For the lower endpoint of the identified set, FCP's are computed for θ equal to $\sup_{x \leq 1.5} \underline{\theta}(x, 1) - c \times (250/n)^{5/14}$, where $c = .34$, $.78$, and 1.1 in the flat, kinked, and peaked bound cases, respectively. These points are chosen to yield similar values for the FCP's across the different cases considered.

Table I. Nonparametric Quantile Selection Model: Base-Case Test Statistic Comparisons

(a) Coverage Probabilities (nominal 95%)

DGP	Statistic: / Crit Val	CvM/ Sum	CvM/ QLR	CvM/ Max	KS/ Sum	KS/ QLR	KS/ Max
Flat Bound	PA/Asy	.974	.974	.971	.968	.968	.963
	GMS/Asy	.953	.953	.951	.955	.955	.953
Kinked Bound	PA/Asy	.998	.998	.997	.995	.995	.995
	GMS/Asy	.990	.990	.989	.989	.989	.987
Peaked Bound	PA/Asy	.998	.998	.997	.995	.995	.996
	GMS/Asy	.992	.992	.991	.991	.991	.991

(b) False Coverage Probabilities (Coverage Probability Corrected)

DGP	Crit Val	CvM/ Sum	CvM/ QLR	CvM/ Max	KS/ Sum	KS/ QLR	KS/ Max
Flat Bound	PA/Asy	.57	.57	.54	.67	.67	.64
	GMS/Asy	.45	.45	.45	.61	.61	.60
Kinked Bound	PA/Asy	.67	.67	.65	.67	.67	.64
	GMS/Asy	.49	.49	.49	.57	.57	.57
Peaked Bound	PA/Asy	.57	.57	.55	.60	.60	.56
	GMS/Asy	.50	.50	.49	.55	.55	.53

versions of the CI's regardless of the DGP specification or the test statistic choice (CvM or KS).[19]

[19] Hall (1993) shows that undersmoothing or bias correction is necessary for consistency of the bootstrap. Undersmoothing is employed in this paper. Hall (1993) also shows that in the context of nonparametric curve estimation, the bootstrap has advantages over the Gaussian approximation in providing a uniform confidence band for the curve. This result does not shed light on the relative performance of Asy and Bt-based tests in this paper because (i) the test statistics are not asymptotically pivotal in the present context, whereas they are in the situation consider in Hall (1993), and (ii) we consider inference at just one point ($Z = z_0$) of the curve.

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

(a) Coverage Probabilities (nominal 95%)

DGP	Critical Value: / Statistic	PA/Asy	PA/Bt	GMS/Asy	GMS/Bt	Sub
Flat Bound	CvM/Max	.971	.971	.951	.948	.963
	KS/Max	.963	.963	.953	.948	.909
Kinked Bound	CvM/Max	.997	.998	.989	.988	.990
	KS/Max	.995	.996	.987	.986	.959
Peaked Bound	CvM/Max	.997	.997	.991	.990	.991
	KS/Max	.996	.996	.991	.990	.968

(b) False Coverage Probabilities (Coverage Probability Corrected)

DGP	Statistic	PA/Asy	PA/Bt	GMS/Asy	GMS/Bt	Sub
Flat Bound	CvM/Max	.54	.55	.45	.44	.53
	KS/Max	.64	.66	.60	.57	.66
Kinked Bound	CvM/Max	.65	.66	.49	.47	.51
	KS/Max	.64	.67	.57	.53	.40
Peaked Bound	CvM/Max	.55	.54	.49	.47	.51
	KS/Max	.56	.55	.53	.49	.39

The GMS critical values noticeably outperform the PA counterparts in terms of FCP's. The CvM/Max test statistic coupled with the GMS/Asy or GMS/Bt critical values outperforms all other combinations in terms of FCP's in all cases.

Table III provides results for the CvM/Max and KS/Max statistics coupled with the GMS/Asy critical values for several variations of the base case. The table shows that the CI's perform similarly at different sample sizes, with different choices of cells and

Table III. Nonparametric Quantile Selection Model with Flat-Bound: Variations on the Base Case

		(a) Coverage Probabilities		(b) False Cov Probs (CPcor)	
	Statistic:	CvM/Max	KS/Max	CvM/Max	KS/Max
Case	Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Base Case: $(n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7})$		.951	.953	.45	.60
$n = 100$		.950	.956	.46	.61
$n = 500$		.950	.953	.44	.59
$r_1 = 2$		.951	.950	.44	.56
$r_1 = 4$		.952	.961	.45	.63
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.948	.947	.46	.61
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.967	.961	.48	.62
$\varepsilon = 1/100$		.949	.953	.45	.63
$b = 0.5b^0 n^{-2/7}$		.960	.963	.68	.77
$b = 2b^0 n^{-2/7}$		.950	.948	.19	.34
$\alpha = .5$		.525	.516	.045	.072
$\alpha = .5$ & $n = 500$		.517	.519	.042	.070

with a smaller ε.[20] There is some sensitivity to the magnitude of the GMS tuning parameters (κ_n, B_n)—doubling their values increases both the CP's and the FCP's, but halving their values does not decrease the CP's much below 0.95. There is more sensitivity to the kernel bandwidth—a larger bandwidth reduces the FCP drastically while keeping the CP at around 0.95 and a smaller bandwidth does the opposite. This result is closely related to the flatness of the bound. The bound is completely flat on the support of Z_i. It is more efficient to use more of the data information by using a larger bandwidth. This phenomenon does not occur with the kinked bound and the peaked

[20] The θ values at which the FCP's are computed differs from the lower endpoint of the identified set by a distance that depends on $(nb)^{-1/2}$. Table III suggests that the "local alternatives" that give equal FCP's converge to the null hypothesis at a rate that is slightly faster than $(nb)^{-1/2}$ for sample sizes n in the range 100 to 500.

bound as shown in Tables A1 and A2 in the Appendix, see Andrews and Shi (2013a).

The last two rows of Table III show that a CI based on $\alpha = 0.5$ provides a good choice for an estimator of the identified set. For example, the lower endpoint estimator based on the CvM/Max-GMS/Asy CS with $\alpha = 0.5$ is close to being median-unbiased. It is less than the lower bound with probability 0.525 and exceeds it with probability 0.475 when $n = 250$.

The FCP's reported in Tables I-III are computed at different θ values (outside the identified set) with the three different bound functions. This is done to ensure that the FCP's lie in a meaningful range. However, it is also of interest to consider the same θ value for all three bounds and, hence, to see how the shape of the bound function affects FCP's. For the CvM/Max/GMS/Asy CI, the FCP's computed for $\theta = 0.78$ are .02, .49, and .81 for the flat, kinked, and peaked bound functions, respectively. Thus, the FCP's are best (lowest) for the flat bound and highest (worst) for the peaked bound function.

In summary, we find that the CI's based on the CvM/Max statistic with the GMS/Asy critical value perform the best of those proposed in this paper in the quantile selection example considered. Equally good are the CI's that use the Sum or QLR statistic in place of the Max statistic and the GMS/Bt critical value in place of the GMS/Asy critical value. The CP's and FCP's of the CvM/Max-GMS/Asy CI's are quite good over a range of sample sizes. The findings echo those in AS1 in the parametric quantile selection example.

4.2.3 Simulation Results: Comparisons with CLR Confidence Intervals

Table IV reports comparisons of CP's and FCP's of the CI's proposed in this paper, denoted by AS, with the series and local linear versions of the CI's proposed in CLR. The AS CI's use the Max S function and GMS/Asy critical values. The CLR CI's are described in Section 4.1. The data generating processes considered are the same as in Table I.

Table IV shows that the nominal 95% AS CI's have good finite sample CP's, being .951 or greater in all cases. In contrast, the series and local linear CLR CI's under cover in the flat bound case with CP's being .895 and .860, respectively. The FCP's of the AS CI's are noticeably less than those of the CLR CI's in the flat bound case. The opposite is true in the peaked bound case. In the kinked bound case, the AS and CLR CI's have similar FCP's. This is consistent with the theoretical asymptotic power comparisons in Section 11 of the Appendix, see Andrews and Shi (2013a).

Table IV. Nonparametric Quantile Selection Model: CP and FCP Comparisons of AS and CLR Confidence Intervals

	CP (nominal 95%)				FCP (CP-Corrected)			
	AS		CLR		AS		CLR	
DGP:	CvM	KS	Series	Loc Lin	CvM	KS	Series	Loc Lin
Flat	.951	.953	.895	.860	.45	.60	.78	.75
Kinked	.989	.987	.967	.964	.49	.57	.56	.51
Peaked	.991	.991	.963	.956	.55	.53	.44	.30

In sum, the CvM/Max-GMS/Asy CI has more robust null rejection probabilities than the CLR CI's. Its FCP's are better (i.e., lower) for the flat bound function and worse (i.e., higher) for the peaked bound function.

4.3 Conditional Treatment Effects

In this example, we illustrate how the proposed method can be used to test functional inequality hypotheses.

We are interested in the effect of a randomly assigned binary treatment (D_i) conditional on covariates X_i and Z_i. The outcome variable of interest, Y_i is a mixture of two potential outcomes $y_i(1)$ and $y_i(0)$: $Y_i = D_i y_i(1) + (1 - D_i) y_i(0)$. The difference $y_i(1) - y(0)$ is the effect of treatment on individual i. The treatment effect for every individual cannot be identified (even partially) because $y_i(1)$ and $y_i(0)$ are never observed simultaneously. Thus, one often focuses on the average treatment effect of a chosen group of individuals with certain observed characteristics. The chosen group of individuals that we consider here is individuals with $Z_i = z_0 \in \mathcal{Z}$ and $X_i \in \mathcal{X}$, where $\mathcal{Z}$ and $\mathcal{X}$ are the supports of Z_i and X_i, respectively. We test the hypothesis:

$$E[y_i(1) - y_i(0)|(X_i, Z_i) = (x, z_0)] \geq 0 \text{ for all } x \in \mathcal{X}. \tag{4.5}$$

The framework can be extended to treatments with any finite number of treatment values. If the X_i variable is not present, the problem is a trivial case of (2.1) where $\mathcal{X}$ is a singleton. If the Z_i variable is not present, the problem fits in the framework of AS1 and Lee, Song, and Whang (2009). The nonparametric method proposed in this paper allows us to focus on a particular value of Z_i.

Examples of the above hypothesis include: (i) whether a certain drug reduces blood pressure for people of all ages and genders ($X_i = $ (age, gender)) whose body mass index (Z_i) is at certain level (z_0); (ii) whether students of a certain IQ score ($Z_i = z_0$) do better in smaller classes than in bigger classes regardless of their parents' income (X_i); and (iii) whether group liability discourages default better than individual liability in a micro-loan program for villages of all sizes (X_i) and certain average income level ($Z_i = z_0$).

The model setup is as follows. We assume that D_i is randomly assigned and $\Pr(D_i = 1) = \pi \in (0, 1)$.[21] Then,

$$E[y_i(1) - y_i(0)|(X_i, Z_i) = (x, z_0)] = E\left[\frac{Y_i D_i}{\pi} - \frac{Y_i(1 - D_i)}{1 - \pi}|(X_i, Z_i) = (x, z_0)\right]. \quad (4.6)$$

Then, the hypothesis (4.5) is equivalent to testing if $\theta = 0$ is in the identified set of the following moment inequality model:

$$E\left[\frac{Y_i D_i}{\pi} - \frac{Y_i(1 - D_i)}{1 - \pi} - \theta|(X_i, Z_i) = (x, z_0)\right] \geq 0 \text{ for all } x \in \mathcal{X}. \quad (4.7)$$

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

$$\begin{aligned} y_i(0) &= 0,\ y_i(1) = \mu(X_i, Z_i) + u_i,\ D_i = 1\{\varepsilon_i \geq 0\}, \\ &X_i \sim Unif[0, 2],\ Z_i \sim Unif[-1, 1],\ (\varepsilon_i, u_i) \sim N(0, I_2), \\ (X_i, Z_i) &\perp (\varepsilon_i, u_i), \text{ and } X_i \perp Z_i. \end{aligned} \quad (4.8)$$

The function $\mu(x, z)$ is the conditional treatment effect function at $(X_i, Z_i) = (x, z)$. We focus on $z_0 = 0$.

Three different $\mu(x, z)$ functions are considered, which are flat, kinked, and tilted as a function of z, respectively. They are: $\mu_1(x, z) = -a$, $\mu_2(x, z) = |x| + |z| - a$, and $\mu_3(x, z) = \log(z+1) - a$, where a is a constant. The hypothesis (4.5) holds if $a = 0$ and is

[21] It is easy to allow for "selection on observables," i.e., to allow D_i to depend on X_i and Z_i. E.g., see Imbens (2004).

violated if $a > 0$. The functions μ_1 and μ_2 do not change sign in a neighborhood around z_0, whereas the tilted function μ_3 changes sign in any neighborhood of z_0 if $a = 0$.

Notice that there is only one conditional moment inequality in this model (i.e., $p = 1$ and $v = 0$). In consequence, the different S-functions, i.e. Sum, Max and QLR, are identical to each other and we do not distinguish them in the results reported below.

4.3.1 g Functions

The g functions employed by the test statistics are indicator functions of hypercubes in $[0, 1]$, i.e., intervals, as in the example above. The regressor X_i is transformed to lie in $(0, 1)$ by the same method as in the example above. The hypercubes have side-edge lengths $(2r)^{-1}$ for $r = r_0, ..., r_1$, where $r_0 = 1$ and the base case value of r_1 is 3. The base case number of hypercubes is 12. We also report results for $r_1 = 2$ and 4, which yield 6 and 20 hypercubes, respectively.

4.3.2 Simulation Results: Tests Proposed in This Paper

Tables V and VI report NRP's and ARP's, respectively, for a variety of test statistics and critical values proposed in this paper for a range of cases. The NRP's are for $a = 0$ and the ARP's are for $a > 0$.[22]

Table V provides comparisons of the PA/Asy, PA/Bt, GMS/Asy, GMS/Bt, and Sub critical values for the CvM and KS test statistics. Table VI provides robustness results for the CvM and KS test statistics in the flat bound case. Table VI shows the degree of sensitivity of the results to (i) the sample size, n, (ii) the number of cubes employed, as indexed by r_1, (iii) the choice of (κ_n, B_n) for the GMS/Asy critical values, (iv) the value of ε, upon which the variance estimator $\overline{\Sigma}_n(\theta, g)$ depends, and (v) the bandwidth b.

Table V shows that tests with the Asy versions of both the PA and GMS critical values have NRP's less than or equal to the nominal level 0.05 with the flat bound and kinked bound DGP's. The tilted bound DGP is a difficult case for NRP control because the conditional mean function changes sign at $z = z_0$ and the integral of the mean function over any symmetric neighborhood around z_0 is negative under the DGP with $a = 0$. With this difficult DGP, tests with Asy critical values using the KS statistic have

[22] Note that, contrary to the previous simulation example, the DGP is different for the NRP's and for the ARP's. The null hypothesis stays the same. ARP's are computed for a equal to $c \times (250/n)^{5/14}$, where $c = 0.25, 1.05$, and 0.25 in the flat, kinked, and tilted bound cases, respectively. These points are chosen to yield similar values for the ARP's across the different cases considered.

Table V. Nonparametric Conditional Treatment Effect Model: Base-Case Comparisons

	(a) Null Rejection Probabilities (nominal 5%)					
	Critical Value:	PA/Asy	PA/Bt	GMS/Asy	GMS/Bt	Sub
DGP	Statistic					
Flat Bound	CvM	.040	.054	.044	.063	.106
	KS	.028	.039	.031	.046	.231
Kinked Bound	CvM	.000	.000	.000	.000	.000
	KS	.000	.000	.000	.000	.002
Tilted Bound	CvM	.066	.085	.072	.094	.148
	KS	.044	.057	.047	.064	.280
	(b) Rejection Probabilities under H_1 (Null Rejection Probability Corrected)					
Flat Bound	CvM	.50	.57	.51	.54	.52
	KS	.30	.42	.30	.42	.35
Kinked Bound	CvM	.32	.24	.52	.59	.63
	KS	.37	.19	.49	.53	.79
Tilted Bound	CvM	.53	.54	.53	.53	.52
	KS	.36	.46	.36	.44	.35

NRP's less than or equal to 0.05 and tests using the CvM statistic have NRP's slightly above 0.05. The tests using Bt critical values have noticeably greater over-rejection compared to their counterparts using Asy critical values. The tests using subsampling critical values with either the CvM or KS test statistic appear unreliable: their NRP's exceed 0.05 by a substantial amount with not only the tilted bound DGP but also the flat bound DGP.

The ARP comparison in Table V shows (i) a clear advantage of CvM-based tests

Table VI. Nonparametric Conditional Treatment Effect Model with Flat Bound: Variations on the Base Case

		(a) Null Rejection Probabilities (nominal 5%)		(b) Rej. Probs. under H_1 (NRP-corrected)	
	Statistic:	CvM	KS	CvM	KS
Case	Crit Val:	GMS/Asy	GMS/Asy	GMS/Asy	GMS/Asy
Base Case: $(n = 250, r_1 = 3,$ $\varepsilon = 0.05, b = b^0 n^{-2/7})$		.044	.031	.51	.30
$n = 100$		.047	.026	.50	.26
$n = 500$		.048	.037	.53	.34
$r_1 = 2$		.047	.040	.51	.36
$r_1 = 4$		.044	.024	.50	.26
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.052	.037	.51	.31
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.040	.028	.50	.30
$\varepsilon = 1/100$		.046	.027	.51	.25
$b = 0.5b^0 n^{-2/7}$		.041	.020	.28	.14
$b = 2b^0 n^{-2/7}$		.049	.043	.78	.57

over KS-based tests, and (ii) clearly better performance of GMS-based tests compared to PA-based ones with the kinked bound DGP and similar performance of GMS and PA critical values with the flat and the tilted bound DGP's.

Table VI provides results for the CvM and KS statistics coupled with the GMS/Asy critical values for several variations of the base case with the flat bound function. Analogous results for the kinked and tilted bound functions are given in Tables A3 and A4 in the Appendix, see Andrews and Shi (2013a). The results in Table VI show little sensitivity to the sample size and a smaller ε for the CvM-based test. The ARP performance of the KS-based test improves noticeably with the sample size, but stays much worse than that of the CvM-based test at all three sample sizes considered. There is some sensitivity to the number of cubes and the magnitude of the GMS tuning parameters (κ_n, B_n). Increasing the number of cubes or increasing (κ_n, B_n) reduces both the NRP's and the ARP's. As in the quantile selection example, there is some sensitivity to the bandwidth. A larger bandwidth leads to higher ARP's but still keeps the NRP's below 0.05. As discussed in the quantile selection example, this is closely related to the flatness of the bound and the same phenomenon does not occur with the other types of bounds, see Tables A3 and A4 in the Appendix, see Andrews and Shi (2013a).

The ARP results reported in Tables V and VI are computed under DGP's with different a values ($a > 0$) with the three different bound functions. For the CvM/Max/GMS /Asy test, the ARP's computed for the same value $a = 0.25$ for all three bound functions are .51, .00, and .53 for the flat, kinked, and peaked bound functions, respectively. Thus, the power is highest for the flat and tilted bound functions and worst for the kinked bound function.

In conclusion, the comparison between test statistics and critical values is largely consistent with the quantile selection example, with the CvM-GMS/Asy couple performing the best both in terms of NRP's and ARP's. The CvM-GMS/Bt couple has somewhat worse NRP than CvM-GMS/Asy. The performance of CvM-GMS/Asy is quite good over a range of sample sizes.

4.3.3 Simulation Results: Comparisons with CLR Tests

Next, we compare NRP's and ARP's of the tests proposed in this paper with those of the series and local linear tests in CLR. The sample size is $n = 250$. The parameter values at which the NRP's and ARP's are calculated are the same as in Table V. The tests proposed in this paper, denoted AS, use the GMS/Asy critical values.

The results are reported in Table VII. The nominal 5% CvM AS test over-rejects somewhat in the tilted bound case with a NRP of .072. Its NRP in the flat and kinked bounded cases is less than 0.05. Both CLR tests over reject the null considerably in the flat and tilted bound cases. Specifically, the NRP's of the series CLR test are .103 and .104, respectively, while those of the local linear CLR test are .177 and .185, respectively. The power of the CvM AS test is substantially higher than that of the two CLR tests in the flat and tilted bound cases (being .51 versus .17 and .18 in the flat bound case and .53 versus .16 and .16 in the tilted bound case). In the kinked bound case, the power of the CvM AS test exceeds that of the series CLR test, but is lower than that of the local linear CLR test.

Table VII. Nonparametric Conditional Treatment Effect Model: NRP and ARP Comparisons of AS and CLR Tests

	NRP (nominal 5%)				ARP (NRP-Corrected)			
	AS		CLR		AS		CLR	
DGP:	CvM	KS	Series	Loc Lin	CvM	KS	Series	Loc Lin
Flat	.044	.031	.103	.177	.51	.30	.17	.18
Kinked	.000	.000	.011	.025	.52	.49	.37	.61
Tilted	.072	.047	.104	.185	.53	.36	.16	.16

References

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

Andrews, D. W. K., Barwick, P., J., 2012. Inference for parameters defined by moment inequalities: a recommended moment selection procedure. Econometrica 80, 2805–2826.

Andrews, D. W. K., Guggenberger, P., 2009. Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. Econometric Theory 25, 669–709.

Andrews, D. W. K., Shi, X., 2010. Inference based on (possibly infinitely) many conditional moment inequalities. Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K., Shi, X., 2013a. Appendix to "Nonparametric inference based on conditional moment inequalities." Unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K., Shi, X., 2013b. Inference based on conditional moment inequalities. Econometrica 81, forthcoming.

Andrews, D. W. K., Shi, X., 2013c. Supplement to "inference based on conditional moment inequalities." Econometrica 81, forthcoming.

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

Aradillas-López, A., Gandhi, A. Quint, D., 2012. Identification and testing in ascending auctions with unobserved heterogeneity. Econometrica 80, forthcoming.

Armstrong, T. B., 2011a. Asymptotically exact inference in conditional moment inequality models. Unpublished manuscript, Department of Economics, Stanford University.

Armstrong, T. B., 2011b. Weighted KS statistics for inference on conditional moment inequalities. Unpublished manuscript, Department of Economics, Stanford University.

Beresteanu, A., Molchanov, I., Molinari, F., 2011. Sharp identification regions in models with convex moment predictions. Econometrica 79, 1785–1821.

Bierens, H., 1982. Consistent model specification tests. Journal of Econometrics 20, 105–134.

Blundell, R. , Gosling, A., Ichimura, H., Meghir, C., 2007. Changes in the distribution of male and female wages accounting for employment composition using bounds. Econometrica 75, 323–363.

Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75, 1243–1284.

Chernozhukov, V., Lee, S., Rosen, A. M., 2013. Intersection bounds: estimation and inference. Econometrica 81, forthcoming.

Chetverikov, V., 2011. Adaptive test of conditional moment inequalities. Unpublished manuscript, Department of Economics, M.I.T.

Ciliberto, F., Tamer, E., 2009. Market structure and multiple equilibria in the airline industry. Econometrica 77, 1791–1828.

Fan, Y., 2008. Confidence sets for parameters defined by conditional moment inequalities/equalities. Unpublished manuscript, Department of Economics, Vanderbilt University.

Fan, Y., Park, S. S., 2011. Bias-correction and confidence sets under IV and MIV assumptions. Unpublished manuscript, Department of Economics, Vanderbilt University.

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

Hall, P., 1993. On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. Journal of the Royal Statistical Society Series B (Methodological) 55, 291–304.

Hall, P., Schucany, W. R., 1989. A local cross-validation algorithm. Statistics and Probability Letters 8, 109–117.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.

Henry, M., Mourifié, I. 2012. Sharp bounds in the binary Roy model. Unpublished manuscript, Department of Economics, University of Montreal.

Hsu, Y.-C., 2011. Consistent tests for conditional treatment effects. Unpublished manuscript, Department of Economics, University of Missouri at Columbia.

Imbens, G., 2004. Nonparametric estimation of average treatment effects under exogeneity: a review. Review of Economics and Statistics 86, 4–29.

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

Kim, K., 2008. Set estimation and inference with models characterized by conditional moment inequalities. Unpublished manuscript, Department of Economics, University of Minnesota.

Lee, S., Song, K., Whang, Y.-J., 2011. Testing functional inequalities. CEMMAP Working Paper CWP12/11, Institute for Fiscal Studies, University College London.

Menzel, K., 2009. Consistent estimation with many moment inequalities. Unpublished manuscript, Department of Economics, NYU.

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

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

Ponomareva, M., 2010. Inference in models defined by conditional moment inequalities with continuous covariates. Unpublished manuscript, Department of Economics, University of Western Ontario.

Ponomareva, M., Tamer, E. , 2012. Misspecification in moment inequality models: back to moment equalities? Econometrics Journal 15, forthcoming.

Pratt, J. W., 1961. Length of confidence intervals. Journal of the American Statistical Association 56, 547–567.

Tamer, E., 2008. Partial Linear Model. In: Durlauf, S. N., Blume, L. E. (Eds.), The New Palgrave Dictionary of Economics. Palgrave Macmillan, New York.